EVOLVING TO A NEW GENOMICS SEGMENTATION BASE

by

Camelia R. Taheri

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor in
Business Administration

Charlotte

2024

Approved by:

_____
Dr. Laura J. Stanley

_____
Dr. Shaoyin Du

_____
Dr. Reginald Silver

_____
Dr. Chandra Subramaniam

ABSTRACT

CAMELIA R. TAHERI. Evolving to a New Genomics Segmentation Base.
(Under the direction of DR. LAURA J. STANLEY)

Advances in behavioral genetics provide a game-changing paradigm shift in the development of an accurate framework for a more precise marketing segmentation strategy. Genetics can explain most of the systematic variation between individuals, continuity of behavioral and personality traits, as well as 50% of the variance in human traits. Leveraging that all human behaviors are influenced in some way by the individual's genetic constitution, a theoretical framework is presented for the definition of a new segmentation base called "Genomics Segmentation". Moreover, we empirically showed the applicability of the new Genomics Segmentation through a K-mean clustering analysis of the alcohol consumption market using 7 different polygenic scores related to personality and cognitive traits. This study increases the predictive power of consumer behavior and marketing segmentation leveraging molecular genetics and 150 years of behavioral genetics replicable findings. It presents for the first-time fundamental principles from behavioral genetics to lay the ground for genomics marketing and the transformation of segmentation strategies. It proposes the segmentation of markets through the genetic propensity of consumers. It not only highlights embryonic research in genomics marketing but also shows the practical application of genomics segmentation through the usage of molecular genetics to create clusters and understand consumption patterns of each subset.

ACKNOWLEDGEMENTS

This dissertation is dedicated to my family. Thank you for your unconditional love and support. To my dad, Prof. Mojtaba Taheri (phD), mi inspiración, who this year is celebrating from heaven. To my husband, Andres Ayala, mi chiquito, who kept supporting me very patiently reminding me there is light at the end of the tunnel. To my mom, mi apoyo incondicional, who has always believed in me and supported me on every project no matter the craziness. To my brother, Cyrus Taheri, mi hermano, who was always a phone call away whenever I needed help or just needed to take a break. And of course, to my two beautiful daughters, Isabel and Camila Ayala-Taheri, mis princesas, who have cheered for me in every step and have inspired me to keep going until the end.

To my amazing Committee Chair, Dr. Laura J. Stanley. Thank you for sharing your expertise, listening to me during the hard times, and guiding me to be able to continue and finish in time. Also, I am highly grateful for my committee members, Dr. Shaoyin Du, Dr. Reginald Silver, and Dr. Chandra Subramaniam. Thank you for your time, support and wisdom inside and outside of the classroom.

A special thanks to Dr. Reginald Silver and Dr. Franz Kellermanns for recruiting me into this program. Thank you for always being one email or call away. It made a big difference having your guidance, your professionalism and support during my time in the DBA program.

I am grateful for Cohort 5. Sixteen incredible individuals with whom I enjoyed and learned during our weekend classes, therapeutic Zoom meetings, and joyful dinners. I am especially grateful for Anupam Raina and Kerrie A. Williams with whom I shared great laughter, inspiring wisdom, and months of drama. And of course, thank you to all my family and friends for your support, help, knowledge, and warmth friendship and love. ¡Muchas Gracias!.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIO - Activity interests and opinion

DNA - Deoxyribonucleic acid

DTC - Direct-to-consumer genetic testing

EA - Educational Attainment

FFT – Five Factor Theory

FPs - Foundational Premises

GINA - Genetic Information Nondiscrimination Act

GWAS - Genome-wide association studies

HRS - Health and Retirement Studies

IQ - Intelligence quotient

IR Model - Inductive Realist Model

LOVS - List of values

LPA - Latent Profile Analysis

MANOVA - Multivariate analysis of variance

PGS - Polygenic Scores

SNPs - Single nucleotide polymorphism

VALS - Value and lifestyle

WCSS – Within Clusters Sum of Squares

CHAPTER I. INTRODUCTION

Since Smith's (1956) pioneering article, segmentation has become a central concept in marketing strategy and firm competitive advantage. As markets grow, adapt, and change, consumer patterns are more heterogeneous and widely scattered in their buying requirements. Any attempt to satisfy everyone - the mass market - will inevitably fail. In today's imperfect competitive marketplace, goods and services would "find their markets of maximum potential as a result of recognition of differences in the requirements of market segments" (Smith, 1956 p.7). Consequently, the need to meet "human wants more accurately than the competition" (Dickson and Ginter 1987, p.2) has become a central need for building competitive advantage and surviving the marketplace. Over the years, different forms of segmentation bases have emerged. The most common segmentation studies involve geographic segmentation (1910s), demographic segmentation (1920s), psychographic segmentation (1930s), and behavioral segmentation (1960s). Depending on factors related to product type, consumer demand, distribution method, media availability, and buyer motivation, segmentation bases are selected and often combined (Goyat 2011).

Since the 1990s, behavioral genetics have found that genes are "the essence of individuality" (Plomin 2019, p.12). In the past 100 years (see Table 1. Genomics and Consumer Behavior), advances in genetics including the first sequencing of the human genome (2001), accessibility to next-generation sequencing technologies (e.g., whole genome sequencing WGS, 2005), the creation of polygenic risk scores (PRS) (2009) and the development of a gene-editing method based on CRISPR-Cas9 (2012) have expanded the study of the inheritance tendency of human behaviors and traits to understand population stratification (ancestral and regional differences within the population) (Bryc et al. 2015), assortative mating (mating patterns)

(Robinson et al. 2017; Vandenberg 1972), consumer decisions (Cesarin et al. 2012; Loewen and Dawes 2012; Miller et al. 2012; Simonson and Sela 2011), consumer negative behaviors (Deak, Miller and Gizer 2019; Li et al. 2017; Saunders et al. 2022; Verhulst, Neale and Kendler 2014), human personality (Sanchez-Roige et al. 2017; Turkheimer, Pettersson and Horn 2014), entrepreneurial tendencies (Rietveld, Slob and Thurik 2021), intelligence (Bouchard and McGue 1981; Plomin and Deary 2015; Plomin and von Stumm 2018), financial risk taking (Cesarin et al 2010; Cronqvist and Siegel 2014; Dreber et al 2008) and educational attainment (Lee et al. 2018) (see Table 1. Genomics and Consumer Behavior). Furthermore, for nearly 20 years, the exponential growth in genome sequencing technology has exceeded Moore's law computer revolution impacting industries at scales never seen since the industrial revolution of the 19th century (Boyle 2020; Burke 2012). Today, there is no doubt that all human behavior is influenced in some way by the individual's genetic constitution (Johnston and Edwards 2002; Plomin 2019; Turkheimer 2000). Quantitative genetics, through advances in twin and adoption studies, and molecular genetics, through the whole sequencing of DNA, have transformed the power and precision of research in the field of genetics. This transformation has expanded the understanding of individuals' heterogeneity, continuity and change in human development, and the interaction and correlation between genes and environment (Plomin and von Stumm 2022). Genomics is a game changer in today's marketplace as "all human behavioral traits are heritable" (first law of behavior genetics, Turkheimer 2000, p.160) and "the effect of being raised in the same family (the environment) is smaller than the effect of genes" (second law of behavior genetics, Turkheimer 2000, p.160).

1.1 Motivation and Research Gap

Firms have entered new dynamic markets highlighted by technology breakthroughs, the flux of global markets, and aggressive innovative competitors (d'Aveni 1995). The marketplace has seen an increase in disposable income and a widening of tastes and needs. Therefore, firms have had to improve their strategies to go above and beyond mass production and mass distribution (Gunter and Furnham 1992). The need for market segmentation and positioning have become two of the most central concepts in marketing management (Kotler 1991; Wyner 2016). Firms expanded their understanding of consumer needs by learning and developing strategies and tactics for product differentiation and market segmentation. Consequently, with major advances using economic and behavioral theories combined with sophisticated analytical techniques, firms have improved their understanding of the marketplace and the creation of market segments that better respond to consumer demand functions (Dickson and Ginter 1987).

The identification of numerous variables including demographic characteristics and past behaviors have been used to create subsets that may or may not translate into an accurate representation of the market segment (Dickson and Ginter 1987). Those market segmentation strategies are founded on traditional bases of segmentation such as demographic, geographic, and psychographic visible and identifiable factors (Gunter 2016) that used individually have become less and less useful (Firat and Shultz 1997). Altogether, most of these techniques rely only on descriptive factors of past purchases (Haley 1968); and past data is not an efficient predictor of possible future trends in buying behaviors (Goyat 2011). Consequently, as most segmentation methods rely on the understanding of "demand (which) results from the interaction of a person with his or her environment" (Dickson 1982, p.56), they provide only a partial picture of the

consumer and they do not predict tendencies for future buying behavior which are central to marketers (Haley 1968).

The majority of the time individuals are unaware of their desires until they are presented with them (Businessweek 2011). Consumers preconceived notions and established patterns distort their understanding and evaluation of what they need, want, or believe (Riquelme 2001). It is helpful to conduct market research - focus groups, interviews, and surveys - however, merely asking customers what they want doesn't always provide unique and inspiring customer insights (Gibson 2015). For experts to spot innovation, they need to be proactive and find a balance between customer-centric marketing and changing thinking patterns. Innovative thinking involves anticipating driving forces with the use of one's intellectual and creative capacities instead of reactive solutions to consumers' demands (Gibson 2015). With the expansion of marketing intelligence due to the "unprecedented volume, velocity and variety of primary data available" (Erevelles, Fukawa, and Swayne 2016, p.897) from individual consumers, better known as Big Data, firms, and researchers have the potential to generate strong business and market impact (Fan, Lau and Zhao 2015). Combining the growth and applicability of Big Data with 50 years of behavioral genetics findings, researchers and practitioners can capture and predict tendencies in consumer behavior from birth to the individual's lifespan. Today, it is known that 50% of variance in psychological traits and personality are explained by inherited DNA (Plomin 2019). Genetics typically explains half of the correlation between traits, behaviors, and environment (see Table 1. Genomics and Consumer Behavior) which is crucial data for the identification of segments in the marketplace and the development of targeted marketing strategies (Daviet, Nave, and Wind 2022). However, the usage and understanding of genomics data for marketing research, segmentation, and positions has been singled out from the field.

Studies in the marketing field have just been looking at one-half of the tendencies related to environmental factors and interaction. There are only two studies showing the potential of genomics data in marketing and behavioral attitudes. Daviet, Nave, and Wind (2022) and Simonson and Sela (2011) have presented the important role of genetics and heritability in understanding consumer patterns with the potential to identify more promising target consumer segments.

"Genetics provides most of the systematic variation between us, environmental effects are random, and our chosen environments show genetic influence" (Plomin 2019, p.94). Furthermore, literature shows how adoptive children are similar to their biological parents, to whom they are genetically related, than to their adoptive parents, to whom they are environmentally related. As longitudinal genetic studies have shown, the same genes affect the same traits from age to age contributing to continuity. For example, 80% of the phenotypic stability of personality is due to genetic factors and changes in personality are largely due to environmental factors (McGue, Bacon, and Lykken 1993). Those environmental factors are the central aspects of the current strategic marketing segmentation. It is what researchers have been focusing on but the main issue is its randomized nature. On the other hand, nature (our genetic makeup) provides a systematic variation that can be understood from the first years of an individual. By using advances in behavioral genetics, marketers have the opportunity to understand both sides of the picture - the systematic variation from the genetic makeup of each individual and the variation depending on the environmental aspects that surround that individual.

In summary, although there are many studies examining marketing segmentation, we can see how most of the studies look at past data painting a partial picture that does not support

accurate prediction of future buying behaviors. Moreover, over 120 years of segmentation research have looked at factors related to the 50% variance coming from the environment without considering the other 50% variance coming from the genetic propensities of individuals. Finally, understanding that genetic data can provide systematic information on the personality and psychological traits of consumers, it is central for marketers to look at the advances in behavioral genetics to improve accuracy in the understanding of consumers and the segmentation of markets.

1.2 Research Goals

Understanding the heritability influence of genetic variation is crucial to developing a more accurate framework for the interpretation of consumer behavioral patterns and the development of subsets for market segmentation. Generating market intelligence with the usage of genetic Big Data could produce the most fine-grained segmentation to 'profile' each individual for one-to-one marketing. The "application of science to marketing problems" (Smith 1956, p.8) is needed for the advancement of the field. As the contemporary consumer is looking for changing patterns of value, images, meaning, and experiences in life and roles in society (Firat and Shultz 1997), the planning and selection of proper marketing strategies require an integrated approach within market segmentation (Smith 1956). Recognizing the advancements in direct-to-consumer genetic testing (DTC- GT) and genome-wide association studies (GWAS), marketers have the opportunity to combine over 50 years of behavioral research, advances in technology including blockchain and Big Data with traditional market segmentation strategies to accurately create precise and recognizable subsets of the marketplace and have a more rational adjustment of product offerings and marketing mixes.

Strategic market segmentation involves the understanding of "what, when, where, how and why of demand" (Dickson 1982, p. 56) to effectively break down the market into actionable customer segments (Goyat 2011). However, marketers have failed to benefit from years of research in the fields of genomics medicine, genetic psychology, nutrigenetics, behavior genetics, and genetic technology. With the idea to increase discussion and relevance in the field of marketing, a theoretical framework is proposed with the introduction of a new marketing segmentation base called Genomics Segmentation. This paper presents a preliminary theoretical model using advances in molecular genetics and polygenic scores (PGS), identified to drive behavior and lifestyle tendencies, for the development of a marketing genomics segmentation.

"Heterogeneity of demand is natural" (Hunt and Arnett 2004 p.10). Researchers and firms know that understanding the heterogeneity of consumers is critical for strategic marketing and a "determinant of competitive advantage" (Dickson and Ginter 1987 p.5). Genetic factors have the potential to reveal biological mechanisms that could regulate consumer's emotional, physiological, and cognitive processes (Daviet, Nave, and Wind 2022), so marketers can build a more accurate understanding of virtually any information about an individual that could not be explored in the past. Overall, incorporating genomics consumer information for the segmentation of markets will enhance the precision of marketing strategies and will heighten the personalization of services to achieve the power of one.

This article aims to enhance theory development in the field of marketing by introducing a new model for strategic segmentation. Considering this is the first time behavioral genetics might be incorporated into marketing segmentation, we begin the conversation with a literature review of both concepts combined in this study: marketing segmentation and behavioral genetics. We follow the literature review with the hypothesis development: the creation of a new

segmentation process called Genomics Segmentation. After reviewing the methodological model, we present in detail the empirical process using polygenic scores to create clusters of the marketplace. It is followed by a comparison of subsets on the basis of their consumption differences and commonalities with the aim of identifying relevant niche markets. In the concluding discussion, we reiterate the importance of incorporating the advances of behavioral genetics into strategic marketing; we also show the possible applicability of genomics segmentation as another segmentation base that can be combined with traditional segmentation strategies; and finally, we highlight ethical implications of genetic data management for the field of marketing.

1.3 Potential Seminal Nature of the Research

The study introduces an approach that marries two strategic concepts into a single system. One concept is behavioral genetic data for the understanding of psychological and behavioral traits, and the other is market segmentation for the development of marketing strategies. Genomics segmentation is based on the premise that "all psychological traits show significance and substantial genetic influence" (Plomin et al. 2016 p.4) which Turkheimer (2000) describes as the first law of behavioral genetics. To be more specific, we can create marketing intelligence from quantitative and molecular genetics which has the power of explaining 50% of the variance in behavioral traits and personality of consumers from birth through their lifespan. Also, it is important to understand that "no traits are 100% heritable" (Plomin et al. 2016 p.5) and a high level of heritability does not imply that the environment has no influence on the development of the trait (Neisser et al. 1996). Following the premise that individuals can select or create their own environment, genomics segmentation will take into account the genetic predisposition that influences learning, personality traits, cognitive abilities, the development of

attitudes, and the decision-making process while considering the active and reactive genotype-environment correlation (Plomin, DeFries, and Loehlin 1977).

It is important to highlight that genomics segmentation should not be considered (1) predictive in nature; (2) the only determinant of consumer behavior; and (3) that genes and environment operate separately. On the opposite end, knowing that "a typical human behavior trait is associated with very many genetic variants" (Chabris et al. 2015 p.304), genomics segmentation applies PGS, a combination of thousands of single-nucleotide polymorphisms (SNPs) identified from genome-wide association studies (Plomin and von Stumm 2022), to probabilistically predict tendencies and create subsets of the marketplace that are highly homogeneous within groups and heterogeneous between groups. Those clusters are created through the combination of PGS looking to conform groups of individuals with similar genetic tendencies. PGS predict through correlations, so we must remember that we are not implying causality. However, PGS have high predictive power, and, as they are inherited at the moment of conception, their predictive power moves from the moment we are born to adulthood. The study proposes that by combining the predictive power of genomics segmentation with other traditional segmentation bases, academic researchers and practitioners can better capture the homogeneity and heterogeneity of markets for a precise personalization of services.

Genomics segmentation serves as a foundation to extend research on genetic propensities and their influence on the transformation of marketing strategies. This new segmentation base not only impacts academic research in marketing but is also relevant for the application of a more precise segmentation in the practice of marketing. Furthermore, this article shows the lack of research in the current literature and potentially shows fundamental principles in behavioral genetics applicable to consumer marketing and strategic segmentation.

1.4 Limitations

Genomics segmentation has considerable limitations. First, there is a lack of systematic access to genomics data specifically to understand individuals' psychological traits and behavior. The most known databases such as Health and Retirement Study (HRS), UK BioBank, and All of Us Research Workbench have been mainly created to understand individual genetic propensities and environmental exposure to the development of diseases and health knowledge generation. Second, there is very little research about the role of genomics on consumer behavior and marketing overall. As reviewed in this study, there have been only a few studies in the marketing field addressing the advances of genomics and consumer patterns. Third, genomics segmentation proposes the utilization of PGS as the main mechanism for segmentation. PGS are mechanisms in constant update as new findings and knowledge are created in the field of genomics. Consequently, ongoing research findings could demonstrate considerable differences in the SNPs and PGS created and used today. Moreover, there is an overreliance on samples of European ancestry for the creation of most PGS which are used in this study. Finally, genomics segmentation could imply an expensive method for segmentation strategy as access to data and technical expertise could involve the need for high-cost investment for marketing teams.

1.5 Ethical Considerations

Genomics data in marketing as with any other rich data initiative will benefit individuals and societies. A precise understanding of genetic propensities could provide early intervention on physical, psychological, and behavioral conditions in the medical, nutrition, and well-being areas. However, we know genomics data is too personal. It is the most personal information one can share. Consequently, concerns related to privacy, security, trust, and ethical manipulation of

data are brought to the table as they need to be properly addressed by all stakeholders - academics, government agencies, practitioners, and policymakers.

Today, consumers have the disposition to share their most personal data in exchange for better and more precise services that match their deepest wants and needs. Nevertheless, "genetic data is identifiable, predictive of virtually every aspect of one's life" (Daviet, Nave, and Wind 2022, p.17), as well as the life of family members. For example, individuals can track family members through the use of forensic DNA and public genealogy databases. In this same way, marketers can have access to publicly available genetic data for the understanding of individuals and their relatives. These opportunities raise great concerns related to breach of trust and privacy as individuals who have not opted-in to share information could be profiled and targeted. Currently, high volumes of genetic data are already available for open access and others can be purchased. The main concern is not if firms are using genetic data to understand and profile consumers (because they are), the concern is now that genomics private entities are already exploiting these tactics, we need stakeholders to work together for the identification of best practices and protection of the consumers' information and best interests. As genomics' advances and usage are becoming a common practice in fields of healthcare, technology, psychology, nutrition, and sociology, together with the undeniable interest of consumers for direct-to-consumer (DTC) genetic testing and GWAS, it is inevitable that firms and researchers have the opportunity to understand with higher accuracy demand functions in the marketplace.

"DNA profiles are unique to each individual on Earth" (Zaaijer et al. 2017, p. 1). With access to 60 to 300 informative SNPs, the identity of an anonymous sample can be inferred. Despite the relevance of genetic data's rich content and easily identifiable information of individuals, most privacy regulations in the field involve information consent mechanisms

created and managed by each of the private entities. To avoid stigmatization and discrimination based on genetic information, governance mechanisms, and technical global safeguards are crucial to increase trust and protect research participants and patients (Borry et al 2018).

As of 2021, it is estimated that 100 million individuals have completed a direct-to-consumer (DTC) genetic test according to the American Medical Association (Henry 2021). Individuals are curious about their ancestry, genealogy, and possible risk for disease. However, they tend to overestimate results and recommendations due to the belief that every genetic-based information has some type of scientific evidence (Daviet, Nave, and Wind 2022). When genetic findings and processes are used for non-medical issues and they do not have present direct health risk for individuals, the Food and Drug Administration is not the regulatory body; instead, the Federal Trade Commission is the responsible institution for regulating possible deception of consumers due to marketing and product offerings. One of the main conversations concerns the possible genetic misinformation and discrimination of individuals by insurance companies and employers. While each country has its own management practices for genetic discrimination and misinformation, in the United States the 2008 Genetic Information Nondiscrimination Act (GINA) regulates health insurance coverage and premiums to avoid any possible initiative of overcharging or coverage denial due solely to genetic results (Wilson and Nicholls 2015). However, this leaves consumers without protection and guidance in everyday life interactions, long-term care, and possible disabilities (Henry 2021).

CHAPTER II. LITERATURE REVIEW

2.1 Marketing Segmentation

"Heterogeneity of demand is natural" (Hunt and Arnett 2004, p.10). Today, the marketplace is characterized by imperfect competition and has "become increasingly crowded with large numbers of brands fighting for market ascendancy" (Gunter 2016, p.2). As marketing experts are responsible for defining the overall marketing strategies that converge individual market demands with the variety of products and services of the firm (Smith 1956), the understanding of the heterogeneity of consumers is critical for strategic marketing and a "determinant of competitive advantage" (Dickson and Ginter 1987, p.5). The heterogeneity in demand comes from the different desires, tastes, preferences, and customs of consumers' ways of satisfying their wants and needs (Smith 1956). Marketing "segmentation helps to homogenize market heterogeneity" (Dibb, Stern & Wensley 2002, p. 113). Through the process of market segmentation, the overall market demand is disaggregated into subsets with different demand functions. Furthermore, these subsets are integrated by individuals (within-group) with similar characteristics and functions, and between-groups, there are more accentuated differences (Dickson and Ginter, 1987).

Since the introduction of segmentation by Smith (1956), rich literature has developed and presented different techniques to effectively segment the marketplace (Goyat 2011). The process of segmenting the marketplace is framed based on observable characteristics and non-observable data from demand functions. Understanding the "what, when, where, how and why of demand" (Dickson 1982, p. 56) is used to effectively break down the market into actionable customer segments (Goyat 2011). The observable and non-observable characteristics are analyzed to determine similarities within geographic, demographic, psychographic and behavioral

characteristics of the consumers. These traditional bases for market segmentation (geographic, demographic, psychographic, and behavioral) follow 2 general categories: physical attributes of consumers, and psychological attributes of consumers (Gunter and Furnham 1992; Gunter 2016) and we are addressing them below.

2.1.1 Segmentation based on Physical Attributes of Consumers

Demographic and geographic segmentation were the most common basis for segmentation in and before the 20th century focusing on mass marketing strategies as media advertising platforms were limited (Gunter 2016). Geographic segmentation was a natural strategy as the boundaries of geographical distance were common and easy to identify (Snellman 2000). The approach divides the market into units based on cities, counties, regions, states, urban, suburban, and rural settings, neighborhood classifications, and TV regions with the notion that where people live, work and play could differ in climate, time zones and culture which have an impact on their consumer patterns (Gunter and Furnham 1992). Most of the subdivide is determined by political, economic, or physical boundaries (Gunter 2016). However, with the increase in mobility, immigration, and universal communication, boundaries for geographic market segments have blurred. This has also been heightened by technology's increase in the globalization of markets and the micro-targeting of customers (Erevelles et al. 1996). On the other hand, demographic segmentation has been the most common method for segmenting markets based on socio-economic factors that are central characteristics of any human population. Using a combination of physical attributes such as age, gender, family size, income level, education level, and ethnic origin, people are categorized (Gunter 2016). For example, the understanding of individuals' differences based on family background, education level, job type, neighborhood, and type of dwelling and how they are linked to economic circumstances and

lifestyle are traditionally used for market segmentation. Even though these classifiers involve social, cultural, and psychological characteristics, they are generally mainly identified by physical features (Gunter 2016). Demographic segmentation does not explain to brands and manufacturers the specificities of what products to make or the type of products to sell in each market segment. Other modes of segmentation were needed for similar brands and products that differ so little could find their own niches in the marketplace (Yankelovich 1964).

These two bases, geographic and demographic segmentation, are the most adopted form of market segmentation due to their defined scale of measurements, as well as the easy process to identify, understand, and apply to classify consumers (Beane and Ennis 1987, Gunter and Furnham 1992; Plummer 1974). However, their biggest criticism involves the lack of richness and how most of the time they have the need to be complemented with other types of data as many "studies have shown demographic variables to be poor predictors of behavior" (Haley 1968, p. 30). Traditional demographic traits no longer present enough data from consumers to serve as the basis of marketing strategies (Yankelovich and Meer 2006). Furthermore, with the increase of competition in the marketplace, marketers and brands have the need to understand subjective differences in consumers' attitudes, needs, wants and values to identify specific niches in the market (Yankelovich 1964).

Today, there are many geo-demographic frameworks for the segmentation of markets. The ACORN (A Classification of Residential Neighborhoods) system is the most common and accounts for the socioeconomic characteristics and family structure of neighborhoods to describe possible lifestyle patterns (Gunter and Furnham 1992). Another traditional method is called MOSAIC which produces a similar residence-type classification of markets combining financial information of families and customers (Gunter 2016). Continuing the need to improve the

predictability of demographic segmentation, Research Services Limited combined socio-economic factors as a system for consumer grouping called Sagacity (Tynan and Drayton 1987). Sagacity is a classification system that combines descriptive features of consumers to predict behavioral patterns. Sagacity proposes four main stages of life cycle for a consumer which apply to the female and male population separately and are based on income and occupation: (1) The dependent, (2) The pre-family, (3) The family, and (4) The later (Gunter and Furnham 1992).

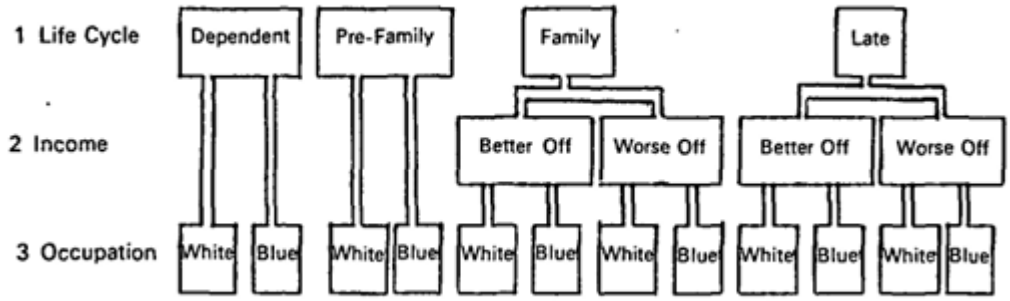


Figure 1. The Sagacity model and twelve Sagacity segments (Tynan and Drayton 1987)

Studies have demonstrated that demographic factors yield a richer insight about consumer patterns if they are combined (Tynan and Drayton 1987). Some have even demonstrated weak explanatory power (Diamantopoulos et al. 2003). For example, age or disposable income are important classifiers of population, however, people of the same age or with the same annual income have different interests as most “have reached different life stages” (p.11) where priorities are defined based on getting married, having children or finding a partner (Gunter 2016). Lifestyle classifiers have had to be updated as society evolves, and individuals' life patterns change. In the 20th century, and specifically in the last few decades, new family structures have emerged improving the performance of consumer classifiers for the prediction of consumer expenditures (Gunter 2016). Thus, the combination of characteristics of consumers

through cluster analysis has been a common practice in the field of marketing especially for market segmentation and the identification of consumers responses to marketing stimuli (Beane and Ennis 1987).

2.1.2 Segmentation based on Psychological Attributes of Individuals

Psychological segmentation bases "focus on the psychological and product orientations of consumers" (Gunter and Furnham 1992, p. 20). These types of segmentation bases look for the understanding of the main motives for consumers to engage with activities, products, and services. They provide a different perspective to marketers from the demographic segmentation studying the reasons behind and how consumers use, act, and feel about a specific product or service and a specific brand. They validate the predictive potential of physical attribute classifiers, as they go above and beyond the visible behaviors and try to understand the way of thinking and the decision-making process of consumers (Gunter 2016). The most common bases involve psychographic, benefit, and behavioral segmentation. Psychographic segmentation involves the strategic study of lifestyle patterns developed from everyday dynamics and interactions of living in a society (Plummer 1974; Vyncke 2002) including social, cultural, and economic forces that influence the individuals' qualities (Gunter and Furnham 1992). Benefit segmentation is based on understanding the benefits and value people are seeking from purchasing and consuming a specific product or service (Haley 1968). And, behavioral segmentation examines consumers' behaviors related to buying patterns and usage of products and services (Gunter 2016).

2.1.2.1 Psychographic Segmentation

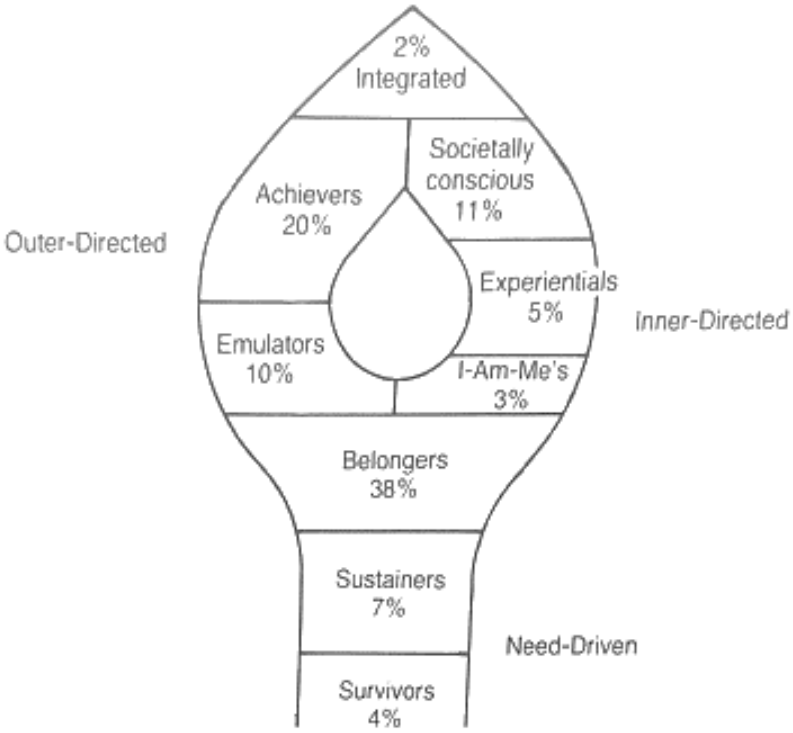
As a response to the need for better insights into the consumer markets, in the 1960s, marketers began to expand the demographic information of consumers to the lifestyle of

individuals. The first concept of psychographics is credited to William Lazer (1963) who defined lifestyle as the different dynamics and patterns coming from the interaction of a society. The psychographic classifiers "look at the inner person rather than the outward expression of the person" (Beane and Ennis 1987, p.22). It has contributed to the understanding of consumer markets based on specific product categories (Ziff 1971) and shopping orientation (Darden and Reynolds 1971). Aspects of motivation research and personality factors were applied and showed a lack of validity due to small samples and consistently low correlations with consumer behavior. Consequently, experts combined the strengths of these psychological approaches to classify consumers based on lifestyle, values, and personality profiles and created the psychographic market segmentation base (Gunter and Furnham 1992).

Lifestyle patterns start with the analysis of people (not products) combining demographic characteristics with the richness of psychological characteristics related to activities, interests, and opinions (AIO) (Plummer 1974). AIO involves the measurement of what people do and the anticipation of possible future events through the analysis of people's actions (activity), their excitement and attention (interest), as well as their written and verbal responses (opinion) (Darden and Reynolds 1974). Each AIO category is measured with the presentation of general and specific statements with 5-point Likert scales for individuals to select whether they agree, or disagree with various factors and situations. Several studies during the 1970s and 1980s showed how using spacing AIO statements can help profile consumers and how their lifestyle relates to specific behaviors (Cosmas 1982; Wells and Tigert 1971).

Another traditional approach of lifestyle research for market segmentation is based on values as they are "guiding principles in people's lives" (Vyncke 2002, p. 271). As defined by Schwartz (1994), values are a hierarchical set of beliefs that guide the evaluation and selection of

individuals, events, and actions based on a desirable outcome or mode of conduct. The most common segmentation models combine Maslow's (1954) hierarchy of needs with social theories. The Values and Lifestyles (VALS) approach developed by Mitchell (1983) classifies consumers into nine specific VALS segments grouped into four basic categories of values and lifestyle: Need-Driven, Outer-Directed, Inner-Directed, and Integrated. And, the List of Values (LOV) model developed by the University of Michigan Survey Research Center (1976) (Gunter and Furnham 1992), classifies people based on values of life's major roles. LOV includes a list of nine values: (1) security, (2) sense of belongings, (3) being well respected, (4) fun and enjoyment of life, (5) warm relationship with others, (6) self-respect, (7) sense of accomplishment, (8) self-fulfillment, and (9) excitement from which individuals need to select the two most important values (Kahle and Kennedy 1988). The "impact of VALS has been widespread and dramatic" as many companies such as AT&T, New York Times, Penthouse, Boeing, and American Motors and popular press have praised and used this approach (Kahle, Beatty, and Homer 1986, p. 405). These value segmentation approaches presented a broader scope than the attitudes-based approach in AIO measures, as well as they use a smaller group of values than the hundred statements itemized in the AIO constructs (Vynke 2002). Consequently, it is obvious that VALS and LOV have similarities in their classification of consumers, but at the same time, several studies have demonstrated that LOV method has some advantages over VALS in the prediction of consumer behavior (Kahle, Beatty, and Homer 1986; Kahle and Kennedy, 1988).



Adapted from: Arnold Mitchell. *The Nine American Lifestyles: Who We Are and Where We Are Going.* New York: Macmillan, 1983 and the Values and Lifestyles (VALS) Program, SRI International, Menlo Park, California.

Figure 2 - The original VALS typology (Gunter and Furnham 1992)

Personality research is another approach of psychographic segmentation that has a long history going back to Sigmund Freud and motivational researchers after the 1945s (Gunter and Furnham 1992). Personality theory examines a set of explanations of cognitive, affective, and behavioral dispositions of individuals. These personality traits are descriptors of people based on stable patterns of thoughts, behaviors, and emotions (Gunter 2016; McCrae and Costa 2003). The most common taxonomy has been the Five-Factor Model of Personality as it brought order and consistency to an endless list of specific traits (McCrae and Costa 2003). The Five Factor Model presents a quantitative assessment of phenotypic traits based on how they look like and how they can be recognized (Costa and McCrae 1999). This model presents five robust factors labeled Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C)

(Costa and McCrae 1999). However, the approach uses pre-set tests and measurements that have revealed weak and inconsistent relationships between consumer behavior and personality traits. The most widely used test is the 'Big Five' personality model which has been applied to research in social media (Azucar, Marengo, and Settanni 2018), parenting (Prinzie et al. 2009), entrepreneurial status (Zhao and Seibert 2006), resilience (Oshio et al. 2018); academic performance (O'Connor and Paunonen 2007); prejudice (Sibley and Duckitt 2008) and many other areas.

2.1.2.2 Behavioral Segmentation

Behavioral segmentation identifies purchasing behavioral features such as product orientation to segment consumers and expands the predictive power of physical attributes. The main strategy focuses on examining past data on purchasing consumption to identify consumers' product usage, places of purchase, and complementary purchases (Gunter 2016). According to Kotler and Armstrong (1994), many marketers prefer behavioral segmentation variables as the starting point to build a complete segmentation strategy. The most common practice has been the segmentation based on product usage popularized by Twedt (1964). Also known as the volume of product usage segmentation was based on finding purchase concentrations. One of the most important findings has been that 50% of the customers generally account for more than 80% of the consumption (Twedt 1962). Many times, segmentation based on the usage of products has been expanded with variables on product benefit (benefit segmentation), as well as analysis of consumer's attitudes and beliefs (psychographic segmentation) (Gunter 2016).

2.1.2.3 Benefits Segmentation

Another approach to market segmentation was introduced by Haley (1968) trying to identify causal factors (rather than descriptives) related to the benefits individuals seek by

consuming a specific product. The main premise of benefit segmentation lies in the idea that benefits sought by consumers determine their behavior (Gunter and Furnham 1992). It is the "internal psychological component to consumers' involvement with product or services" that goes beyond the specific usage (Gunter 2016, p.15). Operationally, this is a complex process as it is based on the specific definition of what consumers think and value (Haley 1968). The technique has been described as pragmatic and managerial and has been highly used for market and advertising planning of existing brands (Haley 1984; Tynan and Drayton 2010).

As "no brand can expect to appeal to all consumers" (Haley 1968, p.34), segmentation strategies are useful because they move our decisions beyond overall scores and averages and reveal specific differences and tendencies so that brands and services can create actionable plans (Plummer 1974). Most of the above segmentation bases should be combined to really provide a more extensive and meaningful market intelligence (Gunter and Furnham 1992). Overall, psychological segmentation bases have increased the predictive power of physical attributes segmentations, however, in general, their tendency to anticipate consumer behavior has "been far from impressive" (Lesser and Hughes 1986).

2.1.3 A Deeper Dive on Segmentation based on Consumer Traits

Segmentation initiatives to understand and distinguish consumers are commonly based on people's characteristics and personality traits (Myers, Sen and Alexandrov 2010). Researchers have been trying to understand and define human traits for years to learn more about consumer behavior (Baumgartner 2002). Specifically, the understanding of the relationship between personality variables and consumer behavior "has existed since the importance of marketing was first recognized" (Haugtvedt, Petty and Cacioppo 1982, p.231). *Trait theory* state that individuals can be "characterized in terms of relatively enduring patterns of thoughts, feelings, and actions"

(p. 140); those are traits consistent in some degree across situations and can be quantitatively assessed (Costa and McCrae 1999). Consequently, human traits have been studied by behavioral scientists and marketers for many years starting with Francis Galton (1875) with the understanding of human intelligence; Gordon Allport and Floyd Allport (1921) with the measurement of personality traits; and Hans Eysenck (1976) with the creation of the Eysenck personality inventory.

Consumer behaviors are not random, they involve a combination of organized and meaningful responses to the world (Hansen 1969). Theoretical and empirical studies have shown that personality and cognitive traits are linked to consumer choices (Foxall 1995) and can affect people's shopping preferences, self-control, decision making, emotions and interaction with others (Leung and Law 2010). From one side, personality traits refer to "consistent patterns in the way individuals behave, feel and think" (Cervone and Pervin, 2022 p. 238), and on the other side cognitive abilities are related to intellectual, broadminded, and logical qualities of individuals (Hoyers, MacInnis and Pieters 2012).

As personality traits and cognitive abilities are crucial for marketing strategy, marketers and psychologists have been trying to create quantitative measurements to assess those traits and understand consumer behavior. McCrae and Costa (1999; 2003) have been pioneers in the field of personality and traits with the development of the NEO personality Inventory and the Five Factor Model (FFM). The FFM uses the Big Five taxonomy to capture at a larger scale the similarities among different personality systems (Widiger 2016). The first two factors: Neuroticism and Extraversion from personality theory were identified by Hans Eysenck (1916-1997). After several years, Costa and McCrae (1999) presented the Five-Factor Theory of Personality and since then it has been adopted as a universal system. The five factors are defined

as: *Neuroticism* (N), the tendency to experiencing emotional instability, negative emotions, pessimism, and helplessness (Cervone and Pervin 2022); *Extraversion* (E), the tendency to experiencing positive emotions, talkativeness, sociability and assertiveness (McCrae and Costa 2003); *Openness* (O) to experience involves curiosity, creativity and independent thinking (Cervone and Pervin 2022); *Agreeableness* (A), tendency to be helpful, cooperative, trusting and forgiving (Parks-Leduc, Fieldman and Bardi 2015); and *Conscientiousness* (C), tendency to be responsible, organized, dependable, efficient and achievement-oriented (Parks-Leduc, Fieldman and Bardi 2015). Figure 3 presents examples more in detail for the characteristics and definition of each personality trait from the Five Factor Model.

Figure 3. Some Examples of FFT Personality System Components (Costa and McCrae 1999)

Numerous researchers from different fields have been able to replicate findings sustaining the theory of the five basic dimensions of personality (Matzler et al. 2005). The Five

TABLE 5.1. Some Examples of FFT Personality System Components

Basic tendencies	Characteristic adaptations	Objective biography
Neuroticism		
N3: Depression (a tendency to experience dysphoric effect—sadness, hopelessness, guilt)	Low self-esteem, irrational perfectionistic beliefs, pessimistic attitudes	"Betty" (very high N3) feels guilty about her low-prestige job (Bruehl, 1994).
Extraversion		
E2: Gregariousness (a preference for companionship and social stimulation)	Social skills, numerous friendships, enterprising vocational interests, participation in sports, club memberships	J.-J. Rousseau (very low E2) leaves Paris for the countryside (McCrae, 1996).
Openness to Experience		
O4: Actions (a need for variety, novelty, and change)	Interest in travel, many different hobbies, knowledge of foreign cuisine, diverse vocational interests, friends who share tastes	Diane Ackerman (high O4) cruises the Antarctic (McCrae, 1993–1994).
Agreeableness		
A4: Compliance (a willingness to defer to others during interpersonal conflict)	Forgiving attitudes, belief in cooperation, inoffensive language, reputation as a pushover.	Case 3 (very low A4) throws things at her husband during a fight (Costa & McCrae, 1992b).
Conscientiousness		
C4: Achievement Striving (strong sense of purpose and high aspiration levels)	Leadership skills, long-term plans, organized support network, technical expertise	Richard Nixon (very high C4) runs for President (Costa & McCrae, in press).

Factors of personality has been linked to many human behaviors such as political attitudes (Gerber et al. 2011; Gerber et al. 2012); positive and negative consumption and advertising emotions (Larsen and Katelaar 1991; Matzler et al. 2005; Mooradian 1996); active use of social

media (Goldsmith and Hofacker 2013); impulse buying (Wang et al. 2020); green and sustainable consumption (Duong 2022; Ribeiro et al. 2016); and fashion consumption (Sarah, Roy, Sethuraman 2016). For example, following the finding that neuroticism has been associated with negative affect (Costa and McCrae 1980), Larsen and Katelaar (1991) and Matzler et al. (2005) shows how neurotic consumers experience distress and negative self-satisfaction emotions. On the other side, extraversion has been associated with positive affect and positive self-satisfaction emotions (Matzler et al. 2005) including enjoying social interactions of all kinds (Lyons et al. 2016). Consequently, extraverted individuals are considered assertive people with natural leadership skills that have no problem expressing their feelings and desires (McCrae and Costa 2003). Overall, it has been found better life outcomes for individuals with high levels of extraversion, agreeableness, conscientiousness, and emotional stability than individuals with opposite traits related to neuroticism, introversion, closedness and anxiety (Widiger 2016).

Negative behaviors such as alcohol consumption and smoking have been associated with personality traits (Hakulinen and Jokela 2018). Alcohol consumption has been found to be positively correlated with sociability and extraversion and negatively correlated with conscientiousness and willingness to confirm (Cook et al. 1998). Supporting these findings, different meta-analyses have found that individuals higher in neuroticism, higher in extraversion and lower in agreeableness were associated with alcohol problems (Hakulinen et al. 2015; Turiano et al. 2012), smoking and physical inactivity (Hakulinen and Jokela 2018).

Further research in the neuroscience field has shown the biological basis of the Big Five personality traits. For example, extraversion uses a region of the brain involve in processing reward information; in the case of neuroticism, it is associated with brain regions related to threat, punishment, and negative affect (DeYoung et al. 2010). In reference to agreeableness, it

showed how it is covaried with volume in regions that process information related to intentions and mental states; and for conscientiousness is associated to regions involved in planning and voluntary control of behavior (DeYoung et al. 2010). These personality traits have been shown to be the result of a combination of nature and nurture (Bouchard 1994; Chipuer et al. 1993; Loehlin 1992). As we will see detailed in the behavioral genetic section, research found that 50% of the variance in personality traits is explained by a person's genes (Briley and Tucker-Drob 2014) demonstrating the nature aspect of personality.

Cognition is a stable trait that influences how people engage and enjoy activities (Cacioppo and Petty 1982). Literature shows that cognitive abilities and traits are generally related to intelligence, abstract abilities, and problem-solving skills (Brunello and Schlotter 2011). In more detail, cognitive abilities refer to the process to integrate complex information to decide or make a choice (Hoyer, MacInnis and Pieters 2012). The study of the individuals' disposition to engage and enjoy thinking emerged early in history together with personality traits in the social psychology field. Pioneer studies include Maslow's (1943) work on motivation, Asch's (1957) social psychology on intuition and peer pressure, and Sarnoff and Katz's (1954) study of motivational bases of attitude change (Cacioppo and Petty 1982). These cognitive traits are mainly measured through standardized "national and international tests taken by students or adults" (Brunello and Schlotter 2011, p. 4). For example, academic achievement and analytical skills are some of the most measured cognitive traits.

Cognition has been strongly correlated with other important marketing variables such as personality, attitudes, behavior, and choice. For example, a study based on attitude theory presents that cognition precedes attitudes which precedes behavior in the decision-making process (Ray 1973). On the other hand, the cognitive decision-making model presents that

consumers make decisions based on attributes with compensatory or non-compensatory information to accept or eliminate a brand (Hoyer, MacInnis and Pieters 2012). In addition, research has recognized that cognition significantly predicts satisfaction judgements as a function of evaluating expectations and performance (Oliver 1980; Homburg et al. 2006).

Additional predictions and influence in consumer behavior and choice have been found, especially related to how cognitive skills influence schooling decisions, employment, work experience and choice of occupation (Heckman, Stixrud and Urzua 2006). Also, literature related to alcohol consumption shows how cognitive-processing capacity is affected resulting in a reduction of emotional and behavioral responses (Curtin et al. 2001). Extending on this argument, Kim et al. (2012) shows how despite the known negative effects of large consumption of alcohol, low to moderate alcohol intake protects against cognitive decline and dementia specifically in the elderly population.

Expanding on cognition and personality traits, literature shows how cognitive functions are inversely related to anxiety trait (Jaiswal et al. 2018) and depressive symptoms, and how personality traits could influence the course and response of depression (Klein, Kotov, and Buffer 2011) and anxiety (Kotov et al. 2010). Anxiety is defined as a negative emotion related to a goal-incongruent future outcome (Lin, MacInnis and Eisingerich 2020); and depressive symptoms are emotions that negatively affect how people think, feel, and interact (Alcoforado, Melo and Alcoforado 2022). Both traits have been associated with low extraversion and high levels of neuroticism (Kotov et al. 2010). Furthermore, these “emotional dispositions” (p.26) have been found important in the decision process to buy or adopt a new product as they push decision makers toward low-risk and low-reward choices (Cohen, Pham and Andrade 2018). Also, these traits influence consumer’s price perception (Suri and Monroe 2001), and as they

interfere with individual's reasoning and cognition, they also influence the judgement process of consumers (Cohen, Pham and Andrade 2018).

As we have seen, personality and cognitive abilities are linked, and they influence each other (Jaiswal et al. 2018). Some even considered cognition as a personality variable that could moderate the resistance of newly formed attitudes towards an advertised message (Haugtvedt and Petty 1992). Consequently, this dissertation concentrates on the most known personality traits: neuroticism and extraversion; as well as important cognitive traits and emotions such as general cognition, educational attainment, subjective well-being, anxiety, and depression that together influence consumer choice and behavior.

2.2 Relevance of Alcohol Consumption Market in Marketing

The market for alcohol consumption is commonplace in many areas of the world. It is projected to generate a total revenue of more than $1.17 billion in 2024 based on The Market Insight Report from Statista (2023). Based on the World Health Organization (2019) more than half of those older than 15 years old in Europe, Americas and Western Pacific Regions are drinkers. Moreover, advertising expenditures and corporate social responsibility campaigns are high and widespread in the industry (Jernigan and Ross 2020) and they are known to shape and reinforce alcohol use and attitudes in consumers (World Health Organization 2019). Consequently, the alcohol consumption market represents an extensive and diverse customer base with great concerns for alcohol-related harm worldwide. Research has found how the promotion of alcohol is associated with health consequences (Babor 2010); but on the other hand, the benefits of light and moderated alcohol consumption have been documented for providing stress and depression symptoms reduction, mood enhancement and improvement of cognitive functions in elderly (Kim et al. 2012).

Negative behaviors such as alcohol consumption and smoking are strongly correlated with personality and cognitive traits (Hakulinen and Jokela 2018). As literature shows, alcohol consumption has a negative correlation with introversion, self-control, and positive correlation with sociability and extraversion (Cook et al. 1998). Expanding on these findings, a meta-analysis of over 72,949 individuals showed that extraverted individuals have an increased probability of risky alcohol consumption (Hakulinen et al. 2015). Moreover, neuroticism is the strongest predictor within the personality traits associated with many forms of negative traits and disorders such as substance abuse (including binge drinking), depression, and anxiety disorder (Adan, Forero and Navarro 2017: Kushner, Sher, and Beitman 1990).

In addition, alcohol consumption has been in the center of the nature-nurture discussion as abundant evidence has demonstrated how the risk of alcoholism has a complex genetic background (Edenberg and Foroud 2013). Alcohol consumption is highly polygenic and two genes, specifically, are related to the metabolism of alcohol (ADH1B and ALDH2) which have been identified as the strongest functional variants to protect against alcoholism (Blum et al. 1990; Edenberg and Foroud 2013; Tawa, Hall and Lohoff 2016).

As we have seen the importance of traits and alcohol consumption, in this dissertation we present an empirical segmentation through genetic tendencies of these traits and the understanding of their relationship with alcohol consumption. The study adds empirical findings to better understand negative behaviors that have strong genetic influence and could be identified and even prevented early in life.

2.3 Segmentation Approach: *a priori and post hoc*

As detailed previously, an abundance of segmentation bases has presented the complexity of attributes and characteristics considered in segmenting a consumer market. However, two

traditional approaches to segmentation have been used to group all these alternative bases for segmentation – *a priori* (prescriptive) *and post hoc* (exploratory) – (Green 1977). A *priori,* or "commonsense" segmentation (Dolnicar and Grun 2008), involves the usage of commonly observed variables such as demographic, psychographic, and geographic characteristics. *A priori* method consists of using the person's background variables to understand the different consumption and interest functions (Moore 1980) such as their favorite soda brand. Statistical models are used to determine if groups with different interests in specific soda brands differ in terms of demographics or lifestyle profiles. For example, Wittink and Montgomery (1980) found that marital status influences the preference for travel. The study specifically used 'single or married' status as a segmentation variable to understand the attributes of business travel (Wittink and Montgomery 1980). The most important aspect of *a priori* segmentation is that the researcher selects the descriptors that will be used to cluster the market in advance or according to prior knowledge (Green 1977). Consequently, the biggest problem is that *a priori* segmentation does not address unobserved heterogeneity (Bart et al. 2005).

In the case of *post hoc* segmentation, known as data-driven segmentation, provides segments based on multivariate relationships such as purchasing behaviors or attitudes (Green 1977). In this type of segmentation, the researcher does not know the number of clusters or the specific size of each group until the cluster analysis is completed; the researcher selects the variables that will be used to cluster consumers such as product benefits, behaviors, or motivations. Post hoc segmentation "relies only on empirically delineated segments" (p.17), producing more comprehensive results (Formica and Uysal 1998). For example, Lopes, Boubeta, and Mallou (2009) uses post hoc segmentation to understand the complexity of tourists' preferences identifying five different segments more precisely. In this dissertation, we will be

concentrating on a post hoc segmentation analysis combining PGS to create clusters that will help us understand homogeneity in genetic tendencies of consumers and compare those subgroups' preferences and consumption tendencies for more precise marketing strategies.

2.4 Behavioral Genetics

"There is no escape from the conclusion that nature prevails enormously over nurture" (Galton 1891, p. 576)

The interaction and relationship between nature and nurture is the central premise for behavioral genetics. For over 100 years, behavioral genetics has been studying the individual's genetic composition and its relationship with environmental factors for the development and influence of complex human behavioral traits (Plomin, DeFries, and McClearn 2008) (See Table 1. Genomics and Consumer Behavior). Starting in 1891, Galton's book *Hereditary Genius* introduces the possibility that "human's natural abilities are derived by inheritance" (p.1) presenting a highly controversial argument that human intelligence is inherited from parents. Further research demonstrated the bottom line that "all human behavioral traits are heritable" (Turkheimer 2000, p.160) an argument that became the first law of behavioral genetics. Many studies have supported this first law (see Table 1. Genomics and Consumer Behavior) including the heritability of the 'Big Five" personality factors (Bouchard 1994; Eaves and Eysenk 1975; Eaves et al. 1989; Kato and Pedersen 2005; Li et al. 2017; Loehlin and Nichols 1976; Sanchez-Roige et al. 2018), political orientations and voting (Hatemi et al. 2009; Loewen and Dawes 2012), financial risk (Dreber et al. 2008; Cesarin et al. 2009; Cesarin et al. 2010; Cesarin et al. 2012; Cronqvist and Siegel 2014), entrepreneurship (Rietveld, Slob, and Thurik 2021; Zao and Seibert 2006), drug addiction, alcohol abuse, smoking tendencies, promiscuity, and divorce likelihood (Deak, Miller, and Gizer 2019; Garcia et al. 2010; Verhulst, Neale, and Kendler 2014;

Li et al. 2017; Saunders et al. 2022), adoption of mobile technology (Miller et al. 2012) and procrastination and impulsivity (Gustavson et al. 2014).

Table 1. Genomics and Consumer Behavior

Article	Year	Method	Data	Dependent Variable	Finding
Mendel 1866	1866	Molecular Genetics	Experiments - study of pea plants	Experiment in plants	Demonstrated actions of invisible factors (now called genes) in predict the traits of an organism. Heredity is the process when a parent passes down one gene for each trait. There is a gene expression when the factor is a dominant gene.
F. Galton 1875	1875	Twin Studies	Twin pairs	Likeness of twins	History of Twins. "There is no escape from the conclusion that nature prevails enormously over nurture" (p. 576)
Jones 1928	1928	Family Study	Parent-child study	Intelligence	Correlation between parent and children resemblance in mental test traits.
Merriman 1924	1924	Family study	Siblings and Twins	Intelligence	Resemblance of twins. (1) There is no significant difference when comparing twin resemblance based on environmental aspects. (2) Higher correlation, intellectual and physical similarity for like-sex pairs compared to unlike-sex twins and siblings overall.
Folling 1934	1934	Molecular Genetics		Phenylketonuria	Single gene of phenylketonuria
Newman, Freeman and Holzinger 1937	1937	Twin Studies	Identical and fraternal twins reared apart	Intelligence	On identical twins reared apart, environment has low effect on physical traits, more effects on ability scores, and the most on personality and temperament.
Skodak and Skeels 1949	1949	Twin and Adoption	One hundred adopted children	Intelligence	Early or late agency placement of adopted children in stable environments have an impact on IQ performance of children
Watson and Crick 1953	1953	Molecular Genetics		Double Helix	Discovery of the molecular structure (the double helix) of the DNA
Jensen 1967	1967	Family study	Between and within families	Intelligence	Most postnatal environmental factors, except extreme isolation, do not remarkably affect IQ

Table 2. Genomics and Consumer Behavior (continued)

Bouchard and McGue 1981	1981	Twin and Adoption	Review of 111 studies	Intelligence	Determination of IQ is partially die to genetics. However, not enough information for precise strength of the effect.
Bouchard, Lykken, McGue and Tellegen 1990	1990	Twin and Adoption	100 sets of reared apart twins	Cognitive abilities	Monozygotic twins reared apart or together are similar personality, temperament, leisure interests, occupation, and social attitudes.
Loehlin 1992	1992	Twin and Adoption		Personality	Genes and environment contribute to the development of personality.
Pedersen, Plomin, Nesselroade and McClearn 1992	1992	Twin and Adoption	302 pair of twins	Cognitive abilities	Heritability of cognitive abilities in twins is much higher than heritability found earlier in life
MacDonald, Ambrose, Duyao, Myers, Lin, Srinidhi, BArnes, Taylor, James, and Groot 1993	1993	Molecular Genetics		Huntington's disease	Single-gene Huntington's disease
Chipuer, Plomin, Pedersen, McClearn and Nesselroade 1993	1993	Twin and Adoption	400 pairs at average age of 59 years old	Personality	Personality characteristics mediate genetic influence on environmental measures. Neuroticism and extraversion have substantial genetic influence on personal growth and family environment. Genetics influence personality and affects perception of family environment.
Bouchard 1994	1994	Twin and Adoption	Minnesota study of twins reared apart (MISTRA) and Minnesota Twin Registry	Personality	Study the link between genes and our evolutionary process. Each individual picks and chooses from a range of stimuli based on genotype and creates a unique set of experiences. Study compares Loehlin (1992) study with their own study reaching the same estimate of genetic influence of over 40% on personality.
Plomin, Fulker, Corley and DeFries 1997	1997	Twin and Adoption	245 adopted children, their biological parents and adopted parents	Cognitive abilities	In early childhood, adopted children/adoptive parents showed resemblance. Children reach adolescence, resemblance increases between adopted children/biological parents.

Table 1. Genomics and Consumer Behavior (continued)

Turkheimer 2000	2000	Theory Development		First laws of behavioral genetics	Proposes the three laws of behavior genetics: "(1) All human behavioral traits are heritable; (2) The effect of being raised in the same family is smaller than the effect of genes; (3) A substantial portion of the variation in complex human behavioral traits is not accounted for by the effects of genes or families." (p.160)
Kato and Pedersen 2005	2005	Twin and Adoption	1339 individual twins	Personality	Genetic influence was found for all three coping strategies except avoidance in men. The shared environment in twins had limited influence.
Dreber et al 2008	2008	Twin and Adoption	Twin design	Financial risk taking	Presence of 7-repeat allele (7R+) in the dopamine receptor D4 in young men accounts for 20% of heritable variance in financial risk taking.
Purcell, Wray, Stone, Visscher, O'Donovan, Sullivan et al. 2009	2009	Molecular Genetics		Schizophrenia	First PGS study on schizophrenia
Cesarin et al 2009	2009	Twin and Adoption	Twin design comparing MZ and DZ twins (Swedish Twin Registry)	Financial risk taking	Genes explaining 20% of phenotypic variation for risk taking and giving.
Hatemi et al. 2009	2009	Twin and Adoption	Gene-environment transmission of political attitudes	Political tendencies	Childhood and adolescence, major effect of environmental influences for political attitudes. In early adulthood (20s), political attitude has high genetic influence and remains stable all the way through adulthood.
Cesarin et al 2010	2010	Twin and Adoption	Twin design comparing MZ and DZ twins (Swedish Twin Registry)	Financial risk taking	Genes account for almost 25% of the variation in an individual's financial risk portfolio.
Vernulst, Hatemi and Martin 2010	2010	Twin and Adoption	Causal relationship between personality and left-right ideology using attitude dimensions	Personality	Covariance between social attitudes and psychoticism comes from genetics. Females' covariance on social desirability is largely due to environmental factors. Psychoticism is highly related to conservative and social desirability is related to liberal ideologies.

Table 1. Genomics and Consumer Behavior (continued)

Garcia et al., 2010	2010	Molecular Genetics	DRD4 VNTR (cheating gene)	Sexual promiscuity and infidelity	Higher promiscuous sexual behavior ('one-night stand') from those individuals with at least one 7-repeat allele (7R+). Reported 50% increase in sexual infidelity cases.
Simonson and Sela 2011	2011	Twin and Adoption	Twin study design (SRI International Northern California Twin Registry)	Consumer judgment and decision making (JDM)	Large heritable component for selection of risk and loss. Choosing a compromise option and preference for utilitarian goods is largely genetically driven. Discounting, variety seeking do not show much heritable component.
Miller et al. 2012	2012	Twin and Adoption	Twin study design (Brisbane, Australia) (ACE model)	Consumer adoption and usage of mobile technology	Mobile phone usage has moderate heritability (34% to 60%) for talk and text frequency. Negative relationship between increased intelligence and decreased mobile phone usage.
Cesarin et al. 2012	2012	Twin and Adoption	Twin design comparing MZ and DZ twins (Swedish Twin Registry) (Survey SALTY)	Risk taking	Genes have a moderate effect over conjunction fallacy, loss aversion, default bias. 20-30% of individual variations comes from genetic differences.
Loewen and Dawes, 2012	2012	Twin and Adoption	Twin study design (MacArthur Foundation's Survey of Midlife Development in the United States (MIDUS) (ACE model)	Voting	Large genetic heritability for the degree of sense of voting is a duty.
Verhulst, Neale and Kendler, 2014	2014	Twin and Adoption	Meta-analysis of twin and adoption studies	Alcohol use disorder	AUD is 50% heritable approximately. Familial aggregation of AUDs is also contributed by environmental factors.
Cronqvist and Siegel 2014	2014	Twin and Adoption	Twin design comparing MZ and DZ twins (Swedish Twin Registry)	Financial risk taking	Genes explain up to 45% of differences between investors. Genes determine to an extent behaviors of persistent investment.
Gustavson et al., 2014	2014	Twin and Adoption	Twin study design (Colorado Longitudinal Twin Study)	Procrastination and impulsivity	Genes explain 46% of procrastination and 49% of impulsivity.

Table 1. Genomics and Consumer Behavior (continued)

Pedersen et al 2015	2015	Twin and Adoption	Twin study design (Danish Twin Register connected to Danish – Scandinavian Donation and Transfusion Database)	Blood donation	Altruism is determined by genetic factors. 53% of propensity to become blood donors comes from genetic factors (28% for environmental and nonshared environment 18%).
Chabris, Lee, Cesarini, Benjamin and Laibson 2015	2015	Theory Development		Fourth law of behavioral genetics	Propose the Fourth Law of Behavioral Genetics: "A typical human behavioral traits is associated with very many genetic variants, each of which accounts for a very small percentage of the behavioral variability" (p.305)
Plomin, DeFries, Knopik & Neiderhiser 2016	2016	Theory Development		10 replicated findings from behavioral genetic	Describes 10 replicated findings from behavioral genetic research. The most important aspect is that both genetics and environment have strong effects on psychological differences. Also, not all traits are 100% heritable, so environment represents an important piece in the genetic puzzle. Continuity in personality and traits comes from genes.
Giri, Zhang and Lu, 2016	2016	Molecular Genetics	GWAS and NGS Technologies. SNP specific for early-onset (EOAD) and late-onset (LOAD) Alzheimer's	Alzheimer's disease	60%-80% of late-onset (LOAD) Alzheimer's disease is genetically heritable.
Okbay, Baselmans, De Neve, Turley et al. 2016	2016	Molecular Genetics	GWAS of three phenotypes	Subjective well-being, depressive symptom, and neuroticism	Three phenotypes are highly heritable. Identified 3 genetic variants associated with subjective well-being, 2 genetic variants associated with depressive symptoms, and 11 genetic variants associated with neuroticism.
Van den Berg, de Moor, Verweij, Krueger et al. 2016	2016	Molecular Genetics	Meta-analysis of GWAS	Extraversion	Found the polygenic nature of extraversion.

Table 1. Genomics and Consumer Behavior (continued)

Li, Savage, Kendler, Hickman et al. 2017	2017	Molecular Genetics	Two population samples: Twin Cohort (FinnTwin12) and Longitudinal Study of Parents and Children (ALSPAC) with PRS and alcohol use problems assessed at age 16 and personality dimensions at age 13	Alcohol use disorder and personality	PGS have low predictability of early-age alcohol use problems, and it is mediated by sensation seeking. PGS may directly influence sensation seeking, affecting the development of alcohol problems in adolescence.
Pirastu, Joshi, de Vries, Cornelis et al. 2017	2017	Twin and Adoption		Hair loss	39% of variance in hair loss (alopecia) in men is explained by SNPs
Sanchez-Roige, Gray, MacKillop, Chen and Palmer 2018	2018	Molecular Genetics	Analysis of meta-analysis	Genetics of human personality	Meta-analysis of study designs for genetics of human personality
Lee, Wedow, Okbay, Kong et al. 2018	2018	Molecular Genetics	Large-scale genetic association analysis (PRS)	Educational attainment	Sample of 1.1 million individuals. Multi-phenotype analysis explain 11-13% of variance in education attainment. Three related cognitive phenotypes explain 7-10% of variance in cognitive performance
Davies, Lam, Harris, Trampush et al. 2018	2018	Molecular Genetics	Cognitive and genetic data from UK Biobank and CHARGE and COGENT consortia	General cognition	Found 148 genome-wide significant associated with general cognition, and 709 genes. 4.3% variance in general cognition function is explained by polygenic scores.
Duncan et al. 2019	2019	Molecular Genetics	Review of all polygenic score studies - mainly use European ancestry data.	Over representation of European ancestry samples in human genetics research	Predictive performance of European ancestry-derived polygenic scores is lower when applying it over non-European ancestry samples. Large-scale GWAS are needed with a diverse population.
Deak, Miller and Gizer, 2019	2019	Molecular Genetics	Literature review - see figure AUD in tab	Alcohol use disorder AUD	50% of AUD can be explained by genetic heritability. GWAS highlighted the need of molecular genetic research on AUD to increase identification of loci related to AUD risk.
Chaudhury et al 2019	2019	Molecular Genetics	Polygenic risk scores (PRS) using PRSice-2	Alzheimer's disease	PRS model could predict 61% of conversion for people with MCI to LOAD.

Table 1. Genomics and Consumer Behavior (continued)

Schmitz, Abbondanza and Paracchini, 2020	2020	Molecular Genetics	ALSPAC data base (longitudinal cohort of general population living in Bristol, UK)	Hearing measures and cognitive abilities at behavioral level	Better hearing associated with higher cognitive performance and socio-economic status. Negative genetic correlation between low hearing and polygenic score for reading ability. Higher educational attainment associated with better hearing.
Plomin & von Stumm 2022	2022	Theory Development		Polygenic scores	There should be a high priority to increase the predictive power of polygenic scores to prevent problems and identify early warnings. Their power to predict are groundbreaking.
Tanigawa et al, 2022	2022	Molecular Genetics	PRS prediction from UK Biobank	Assessment of polygenic risk score (PRS) prediction across +1500 phenotypes/traits	SNPs heritability estimate and the predictive power of polygenic scores are highly correlated to explain certain phenotypes/traits.
Saunders et al., 2022	2022	Molecular Genetics	PRS (homogeneous effect in genes ADH1B and CACNA1B)	Alcohol use and tobacco	Global genetic diversity in tobacco and alcohol use (global ancestry to find genomic loci associated with tobacco and alcohol use). Generic variants of tobacco and alcohol use have a effect size similar across ancestries. Also, environmental factors - such as culture and public health policies affect tobacco and alcohol usage
Daviet, Nave and Wind 2022	2022	Theory Development		Genetic data overall	Potential impact of genetics in the field of marketing
Vukasovic and Bratko 2015	2015	Twin and Adoption	Meta-analysis of Twin and adoption studies	Heritability of personality	Genetics explains 40% of individual variances in personality. 60% of the variance is due to environmental factors. Moderator effect of study design: Twin studies showing higher estimates (.47) compared to family and adoption studies (.22)
Plomin 2023	2023	Theory Development			Highlights milestones in the field of behavioral genetics passing through quantitative and molecular genetics, GWAS and polygenic scores.

The term heritability in behavioral genetics has a very specific and narrow definition. It refers to “the proportion of phenotypic variance that can be accounted for by genetic differences among individuals” (Knopik et al. 2013, p.87). It is generally estimated from the correlation between relatives because parents and offsprings are 50% genetically similar. As a result, it is assumed that similarities and differences in psychological traits between family members could be related to nature (genetics) or nurture (environment). Knowing that nature is inevitable, it has been challenging to identify which percentage comes from nurture, from the aspects you can hear and feel as part of family life (Plomin 2019). Heritability is not a constant number; it is a statistic that describes a trait or tendency of a particular population at a specific point in time considering the combination of genetic and environmental influences (Plomin 2019). Behavioral genetics uses twin and family adoption studies (quantitative genetics), and specific gene identification (molecular genetics) as strategies to understand the genetic and environmental influence over behavioral and psychological traits.

Quantitative genetics has been the most used method involving adoption and twin studies. Family and adoption studies provide a “direct test of nurture” (Plomin 2019, p.13). It is a social experiment where relatives who share nature (genetics) live in different nurture environments with non-genetic related parents. The twin method is a biological experiment using identical and fraternal twins. Identical or monozygotic (MZ) twins inherit the same DNA, and it is assumed the MZ twins reared apart or reared together are almost similar demonstrating that similarity comes from nature (genetics) and not nurture (environment). Molecular genetics, on the other hand, studies the structure, function, and interaction of genes at the molecular level. It started with the hunt for single genes responsible for specific disorders and complex psychological traits (Plomin, Owen and McGuffin 1994). A few single-gene disorders were

identified such as Huntington's disease until linkage analysis, the proposed double-helix structure of DNA (Watson and Crick 1953), and the mapping of the human genome provided access to multiple DNA variants and molecular genetic techniques. However, the process of genotyping was slow and expensive until genome-wide association studies, in the twenty-first century, created the opportunity to genotype hundreds of DNA microarrays (better called SNPs) of an individual in a quick, inexpensive, and accurate way (Plomin 2023). GWA studies opened the door to the study of highly polygenic traits.

2.4.1 Quantitative Genetic: Family Adoption and Twins Studies

Quantitative genetic studies 'genetic' siblings and 'environmental' siblings, as well as parent-child resemblance through the family and adoption method and twins method. It started in 1875 with Galton's publication "The History of Twins, as a Criterion of the Relative Power of Nature and Nurture" ((Loehlin and Nichols 1976) where he introduced the study of twins to learn the influence of heritability and environment on the development of individuals. Moreover, Galton's work on IQ inspired studies of heritability using family and adoption studies since the 1920s. For example, Jones (1928) and Merriman (1924) studied the resemblance of parent-child and siblings respectively in mental test traits. Substantial genetic influence on IQ was found as the research continued in the 1940s with the first study of identical and fraternal twins reared apart (Newman, Freeman & Holzinger 1937) and the adoption study of 100 children incorporating birth parents in the sample (Skodak and Skeels 1949). In 1967, another controversial study looked at the possibility of boosting IQ by looking at genetic and environmental influences and was able to find that most postnatal environmental factors, except extreme isolation, do not remarkably affect IQ (Jensen 1967). In this study, Jensen (1967) calls

out for a diversity in the educational system (instead of uniformity) to improve education for the disadvantaged (individuals).

Starting in the 1970s, research in quantitative genetics grew exponentially with the creation of the Behavior Genetics Association which launched the first journal *Behavior Genetics,* and the expansion of studies in different realms within behavioral traits (Plomin 2023). Within the 1970s and 1980s studies about the major dimensions of personality and behavior in the development of individuals were strongly expanded. Four large bodies of personality data studying adult fraternal and identical twins are found (Eaves et al. 1989). Studies in the London region concluded that both genetic and environmental factors contribute to personality differences (Eaves and Eysenck 1975; Eaves and Eysenck 1977) and that personality traits could become more pronounced with advancing age (Eaves and Eysenck 1976). Similar results were found in the US region and Australian studies. They found that personality test scores between relatives are similar due to genetic factors and not cultural/environmental transmission (Loehlin and Nichols 1976; Martin and Jardine 1986). Also, they expanded the findings by identifying that the proportion of genetic variance in personality traits is different based on sex (Martine and Jardine 1986). Using a very large sample, a Swedish study expanded the same findings stating that “half of the phenotypic variation may be attributed to genetic factors” (Floderus-Myrhed, Pedersen, and Rasmuson 1980, p.153).

Quantitative genetic methods drove decades of research through semi-experimental situations showing that resemblance between relatives increases with genetic relatedness, as well as confirming family resemblance between adopted individuals reared apart (Knopik et al. 2013). The greatest example is one of the longest and largest ongoing genetic studies called the Colorado Adoption Project (Knopik et al. 2013). The project included a 20-year longitudinal

study about the resemblances of cognitive and verbal abilities between adopted children, their biological parents, and adoptive parents. Findings showed a slight resemblance between adopted children and adoptive parents in early childhood only and an increased resemblance between adopted children and their biological parents as children reach adolescence and adulthood (Plomin et al. 1997). Other studies such as the Minnesota Study of Twins Reared Apart (Bouchard et al. 1990) and the ongoing Swedish Adoption/Twin Study of Aging (SATSA) (Kato and Pedersen 2005; Pedersen et al. 1992) had similar findings comparing monozygotic twins reared apart. These twins have similar cognitive and psychological abilities as monozygotic twins growing together confirmed that a shared family environment has little influence. As a result of many of these studies, Turkheimer (2000) proposed the Second Law of Behavioral Genetics: "The effect of being raised in the same family is smaller than the effect of genes" (p. 160). Another important finding from the twin and family-adoption studies highlighted that environmental influence works differently than what was known before (Plomin 2023). The study of Loehlin and Nichols (1976) of 850 sets of twins growing up in the same family covering aspects of personality, abilities, and interest supported the finding that environmental effects operate randomly, and it is highly situational.

2.4.2 Molecular Genetics

"The instructions for assembling every organism on the planet... are all specified in DNA sequences that can be translated into digital information" (Lander and Weinberg 2000, p. 1777)

Molecular genetics study the underlying biological mechanism of heredity through the linkage of genes as a result of the mapping of the human genome (Knopik et al. 2023). The initial studies were focused on mapping genes responsible for single-gene mutations (Plomin 2023) related to monogenic disorders (Gray, Campbell, and Spurr 2000). The most common

single-gene effects were found for phenylketonuria, a common metabolic disorder (Fölling 1934), and Huntington's disease (Huntington 1872; MacDonald et al. 1993). By the 1940s, the power of DNA on heritability was clear. However, it wasn't until the revolutionary discovery of the molecular structure of the DNA (the double helix) by Watson and Crick (1953) that a door was opened for the study of genetics at the molecular level (Bansal 2003). Most of the studies at the end of the twentieth century focused on the development of methods to sequence DNA's nucleotide bases. Attempts to understand complex traits and diseases through linkage mapping, allelic association, and the Human Genome Project (that started in 1990) provided fundamental advances in the understanding of the human blueprint. The biggest challenge in this era focused on finding how to use molecular genetics to identify genes that through the interaction with multiple other genes and non-genetic factors (environment) influence complex traits. Linkage mapping was used to "determine the relative position of genes along a chromosome" (Donis-Keller et al. 1987, p. 319). The complete genetic linkage maps were essential for the study of mutations and the determination of locus order for the initial effort to map diseases (Dausset et al. 1990; Donis-Keller et al. 1987). Moreover, experiments pivot to allelic association studies to understand the statistical relationship between alleles and observed phenotypes. Several studies reported allelic association of specific gene receptors in alcoholism (Blum et al. 1990) and late-onset Alzheimer's disease (Corder et al. 1994). They tried to explain the effect size of functional genes on those specific traits or diseases. However, most of these studies based on candidate-gene approaches failed to replicate and identify effect sizes.

The Human Genome Project was critical for the advancement of global science when in 2003 reached 90% sequencing of the human genome (NIH 2022). After over a decade, researchers were able to determine the exact order of the DNA bases (the As, Cs, Gs, and Ts

segments of the DNA). The complete sequencing of the human genome gave the tools for the creation of the genome-wide association study (GWAS). GWAS identifies genotyping of specific pre-selected variants for the understanding of traits. However, access to data with sufficient size to run a full-powered GWAS on specific complex traits was still expensive and time-consuming beyond the capacity of most researchers (Uffelmann et al. 2021). The biggest advancements were made possible by the assessment of associations between thousands of DNA variants through the creation of SNPs. Those SNP markers (single nucleotide polymorphism) are an "individual nucleotide base difference between two DNA sequences", also, they are generally stable and do not change between generations (Edwards et al. 2007). SNPs are an important source of molecular information for the understanding of marker-trait gene associations and the assessment of individuals' genetic relationships. Due to their low mutation rate, SNPs are an excellent tool for information related to complex genetic traits (Edwards et al. 2007).

"SNP chips paved the way for GWA analyses" (Plomin 2023, p.79). With the expansion of genome-wide association studies (GWAS), there was a shift away from the study of monogenic disorders due to the clear understanding that single genes do not have major effects (Gray, Campbell, and Spurr 2000). Instead, the study of complex polygenic traits and diseases that represented a bigger social burden was the priority. The combination of hundreds of SNPs in polygenic score analysis started to gain interest in assessing explanatory power. Polygenic scores resulted from the weighted sum of thousands of SNPs based on their effect size on the association with a specific trait (Dudbridge 2013; Plomin and von Stumm 2022). Since 2009, with the first study using polygenic scores to understand the possible genetic risk of schizophrenia, hundreds and thousands of polygenic scores have been created. That initial study (Purcell et al. 2009) found molecular genetic evidence for a polygenic component to the risk of

schizophrenia and an association with the risk of bipolar disorder too “establishing a common polygenic basis to those conditions" (Dudbridge 2013, p.1). Larger GWAS together with next-generation sequencing technologies providing fast and cost-effective sequencing strategies (Giri, Zhang, and Lu 2016) marked the milestone for the transformation of behavioral science as polygenic scores are introduced and publicly available through the PGS Catalog (Plomin 2023). This PGS Catalog is an open resource database of PGS created to promote PGS reproducibility, as well as the increased access, usage, and reusability of PGS facilitating further research in predictability abilities and validity of findings (Lambert et al. 2021).

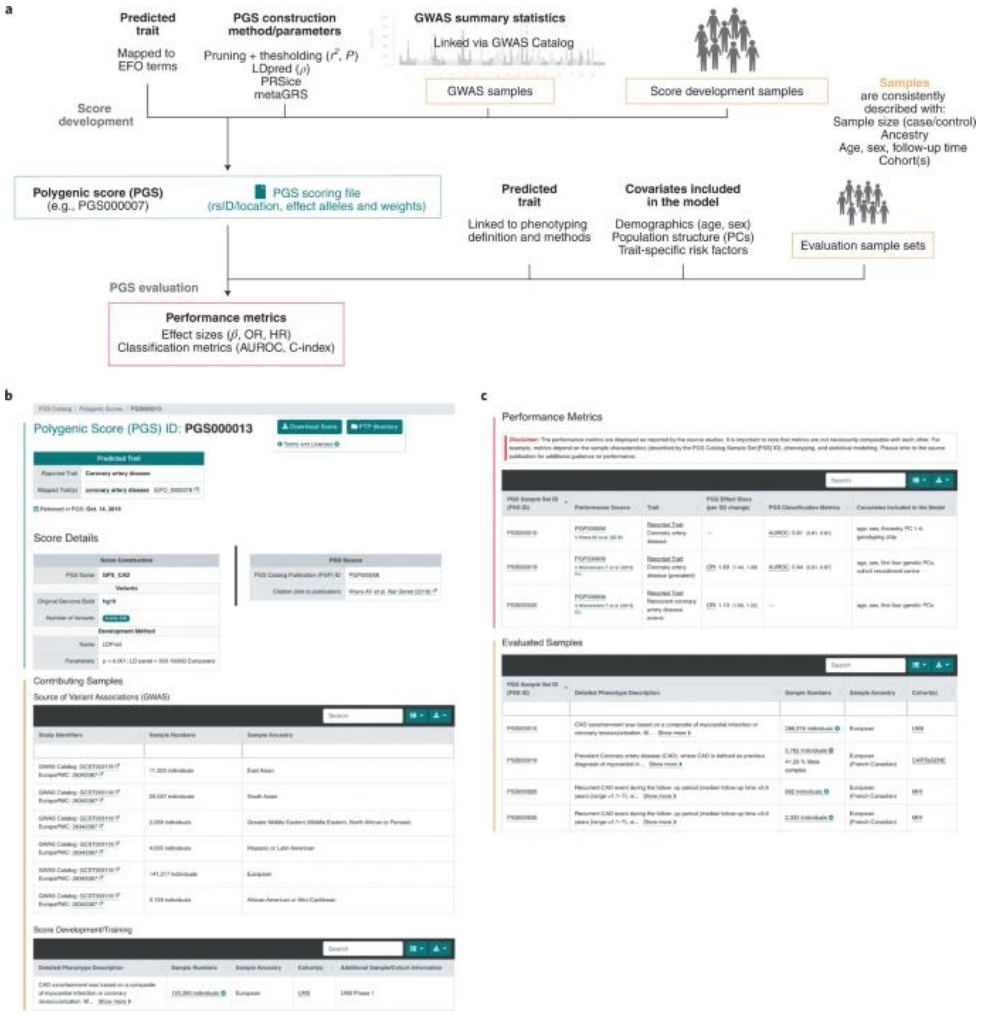


Figure 4. Common aspects of PGS analyses that are captured and displayed in the PGS Catalog.

GWAS demonstrated that the biggest effect for any behavioral and biological traits was smaller than anyone considered (Plomin 2023; Visscher et al. 2017). The highly polygenic nature of heritability was a shock of discovery as many studies of diseases and cognitive traits using a combination of SNPs found small effects. Some examples studied Schizophrenia (Purcell et al. 2009; Visscher et al 2017), diabetes and auto-immune diseases (Visscher et al. 2017), educational attainment (Rietveld et al. 2013), sexual promiscuity and infidelity (Garcia et al. 2010) and childhood intelligence (Benyamin et al. 2014). However, with the expansion of sample sizes and the identification of hundreds and thousands of SNPs related to behavioral

traits, many GWA meta-analyses showed up with increasing explanatory power. For example, a study of 1.1 million individuals identified a combination of SNPs that can explain between 11-16% of the variance in educational attainment and 7-10% variance in cognitive abilities (Lee et al. 2018). A following study with 3 million individuals had very similar results identifying a PGS index that can explain 12-16% of education attainment variance (Okbay et al. 2022). The same for GWA meta-analyses of personality and cognitive traits such as the explanation of 10% variance in neuroticism traits (Okbay et al. 2016a), 5% variance for extraversion (van den Berg et al. 2016), 28-29% variance for general cognition abilities (Davies et al. 2018) and 5% variance for subjective well-being (Okbay et al. 2016a).

2.5 Genomics in Strategic Marketing

The utilization of genetic data in the marketing field has been highly embryonic. It started with the accessibility of consumer genetic data through the accelerated growth of direct-to-consumer (DTC) genetic testing. The demand for DTC genetic testing has grown exponentially as consumers want to go outside their regular healthcare provider (Majumder, Guerrini, and McGuire 2021) and personally uncover forgotten family histories (e.g., ancestry) and assess predisposition for diseases (Daviet, Nave, and Wind 2022). However, geneticists, public health advocates, and governmental institutions have been raising concerns related to the quality and accuracy of the tests, the adequacy of the information provided, and the possible misleading claims that can drive individuals to harmful healthcare decisions (Hogarth, Javitt, and Melzer 2008). Despite these concerns, according to the American Medicine Association (AMA), it has been estimated that 100 million individuals have undergone genetic testing by 2021 (Henry 2021).

Most initiatives looking to leverage genetic data for marketing purposes have come from private entities. The majority have been trying to offer innovative, but not scientifically reliable, services and marketing promotions related to family ancestry. Companies such as Spotify for example announced in 2018 that it would genetically curate playlists for users that upload their genetic data into the system (Herrman 2018); Airbnb connects individuals' genetic heritage with cultural trips and experiences (Knowledge at Wharton 2020); and Vinome promises a personalized wine experience customized to your specific DNA (Robbins 2016). Further offerings keep appearing offering personalized services such as cosmetics and wellness packages based on preferences identified in the client's genome, or discounts in airline tickets equivalent to the percentage of the user's heritage (Kadiri 2022). Moreover, other companies are leveraging genetics to service and recruit young generations as they offer love matching and finding the perfect partner through genetic pairing (e.g., DNA Romance) (Mansky 2018).

In the field of marketing research, just a few studies have brought to the forefront the importance of genetic data and the possible applicability of behavioral genetics in the transformation of marketing. As presented by Daviet, Nave, and Wind (2022) discoveries in behavioral genetics and advances in molecular genetics have undoubtedly opened the door for the leverage of genetic influence in the refinement of consumer behavior theory, marketing research, marketing strategy, and segmentation. Starting with the understanding that individual differences have a heritable component, Simonson and Sela (2011) identified a pattern of genetic influence for certain consumer choices including asymmetric dominance effect and common consumption preferences related to cars, chocolate, music, and movies. It defined inherited tendencies as "constructive predispositions" (p. 962) that help explain some consumer behavior irrationalities that contradict economic theory and value maximization (Simonson and Sela

2011). It has also been defined as the "next significant trend in targeting and personalization" (p324) of marketing strategies and services (Kadiri 2022).

New technologies have increased access to genetic data and new analytical methods that have fundamentally altered strategic marketing and firms' decision-making (Hoffman et al. 2021). The understanding of genetic data through quantitative and molecular genetics can refine marketing theories revealing the nature of relationships and biological mechanisms for a better understanding of behaviors and traits (Daviet, Nave, and Wind 2022). Examples of applications of genetic data in advanced marketing practices such as gene-based segmentation to divide the market into stable, distinct, and identifiable subsets; advanced targeting using PGS for accurate predictions; or new product development tailoring services and products to individuals' DNA propensities are a few.

2.6 Potential Genetic Data for Marketing Segmentation

There are different types of genetic data that have been used in the study of traits in the field of behavioral genetics. As described before, the study of genetic heritability is based on quantitative genetics and molecular genetics. Below, we will be describing the different types of genetic data, their functionalities, and limitations. Finally, we will examine how polygenic scores have opened the door to millions of possibilities for understanding genetic influences at the individual level.

Table 2. Types of Genetic Data

Type of Data	Definition	Literature
Twin and Adoption Family Studies	Comparison between monozygotic and dizygotic twins, as well as family members on the phenotypic and environmental variance to identify heritable effects. (Assumption: Individuals closely genetically related tend to be more similar on a measured trait due to genetic heritability.)	Simonson and Sela (2011) Knopik et al. (2013)
SNPs	Single-nucleotide polymorphism is a genetic variant in a single base position in the DNA. A mutation in a single nucleotide. (e.g. APOE located in chromosome 11 commonly related to longevity trait)	Gunter (2023) Knopik et al. (2013)
Polygenic Scores (PGS)	Sums of thousands of DNA differences (SNPs) identified and weighted by their association with a specific target trait.	Plomin (2019) Plomin and von Stumm (2022)

2.6.1 Twin and Family Adoption Studies

The most common method to estimate heritability is through the use of family data. Using basic laws of heredity, family studies compare phenotypic similarity (measured) with genetic tendencies (gene-specific) of related individuals (Knopik et al 2013). The comparison of monozygotic and dizygotic twins is one of the most common approaches to estimating heritability; it provides an estimate of variance in observed traits accounted by genetics, “shared” (growing in the same family), and “unshared” environment (life experiences) (Simonson and Sela 2011). The adoption study is another approach for the study of relatives who are genetically similar but do not share environmental influences. One of the simplest versions of adoption studies is the one using monozygotic twins reared apart where “the correlation directly estimates heritability” (Knopik et al 2013, p. 383). However, one of the biggest mysteries in family studies is related to how children growing up in the same family are so different (Plomin 2011). Consequently, there are concerns related to the assumption of equal environments as parents and siblings may treat individuals differently (Simonson and Sela 2011). Family studies use a

variance-covariance matrix to estimate relationships between genetically related individuals sharing or not sharing environmental factors. The estimation represents a generalizability of the percentage of heritability of a possible trait overall in a sample population. It does not differentiate between individuals which is needed when estimating subsets of consumers for segmentation.

Table 3. Coefficients of Genetic Relatedness

Related Pair	Proportion of Additive Genetic Variation Shared	Proportion of Dominant Genetic Variation Shared
Parent and offspring (PO)	1/2	0
Half siblings (HS)	1/4	0
Full siblings (FS)	1/2	¼
Nonidentical twins (DZ)	1/2	¼
Identical twins (MZ)	1	1

Source: Knopik et al. (2013, p.380)

2.6.2 Single- Nucleotide Polymorphism (SNPs)

SNPs are the most common genetic variation among individuals. They are a genomics variant at a single base position in the DNA and they might influence health, disease, drug response, and other traits (Gunter 2023). It is a variation of a single letter of one of the nucleotide bases (A, C, G, T) in the DNA. Candidate gene approach consists of testing these genetic variants that have been identified to be associated with a specific outcome (Rietveld, Slob, and Thurik 2021). The genetic variants are selected from GWAS, a hypothesis-free analysis, which has identified the predictive power of over 1 million individual SNPs spread across the genome. Those SNPs have an association with a specific trait. For example, Nicolaou et al. (2011) was the first study to report an association of the SNP DRD3 gene with entrepreneurial behaviors. The

study selected this SNP because in studies of qualities in entrepreneurs, aspects of novelty seeking, sensation seeking, and ADHD were identified and the SNP DRD3 is related to those traits. The candidate gene approach follows a classic process of justifying a relationship and then testing the hypothesis. However, most studies using the candidate gene approach have failed to replicate (Daviet, Nave, and Wind 2022). Furthermore, it has been demonstrated that traits are influenced by many SNPs (polygenicity) (Plomin 2019). Consequently, studies that have used SNPs have had very small effect sizes, not giving us the possibility to truly understand the effect of genetics on each consumer market.

Gene	Gene Name	Chr	Start	End	Trait Studied for Association with the Gene	References	Number of SNPs Imputed from 1000 Genomes	Number SNPs Imputed from 1000 Genomes with INFO>0.8
ADAPT78 (DSCR1; RCAN1)	regulator of calcineurin 1	21	35880440	35992441	longevity		1099	969
ADIPOQ	adiponectin, C1Q and collagen domain containing	3	186555463	186581252	longevity	18	191	140
AKT1	v-**akt** murine thymoma viral oncogene homolog 1	14	105230686	105267088	longevity	6	306	226
APOC1	apolipoprotein C-I	19	45412504	45427606	longevity	1,7	104	61
APOC3	apolipoprotein C-III	11	116695422	116708788	longevity		138	114
APOE	apolipoprotein E	19	45404011	45417650	longevity	1,7, 14	85	62
CAT	catalase	11	34455472	34498609	longevity		253	216
CAMKIV (CAMK4)	calcium/calmodulin-dependent protein kinase IV	5	110554351	110835584	longevity	3,7	2170	1716

Figure 5. Example of SNPs for longevity. HRS “Candidate Gene for Longevity”. November 2014.

2.6.3 Polygenic Scores (PGS)

Polygenic scores have been created following the same process as polygenic risk scores (PRS). As a result of the enormous amount of genomics data, polygenic risk scores combine

genomics variants associated with complex diseases and statistically estimate how the collection of a person's variants could affect their risk or tendency to a certain disease (National Human Genome Research Institute 2020). Polygenic scores follow the same process of combining thousands of single-nucleotide polymorphisms (SNPs) associated with a specific psychological or behavioral trait and have become one of the fast-growing areas of research in behavioral science (Plomin and von Stumm 2022). As GWA studies double their samples, the predictive power of polygenic scores has substantially increased (Plomin 2019). These polygenic scores "estimate heritability, infer genetic overlap between traits, and predict phenotypes" (Choi, Mak, and O'Reilly 2020, p.1). They present the correlation of a combination of SNPs with a specific trait. They present a propensity; they do not imply causation.

Currently, polygenic scores are one of the main indicators of genetic trait tendencies and risks. It is recommended that polygenic score data is standardized as it shows gradations of the genetic propensity for an individual about a specific trait (Davidson et al. 2021). The weighted sum represents low tendency (PGS less than or equal to -1), average (PGS between -1 and +1), and high propensity (PGS greater or equal to +1) (Davidson et al. 2021). Consequently, the predictive power of polygenic scores is groundbreaking as it has enhanced the precision of genetics, as well as provided the tools to directly assess each person's genetic propensities (Plomin and von Stumm 2022).

2.7 Genomics Segmentation: The Research Question

After a century of research in behavioral genetics, we can see that the DNA revolution has made it possible to "predict individual differences in behavior from early in life" (Plomin 2023, p. 75). The argument that complex behaviors and traits have genetic influence is not new (Tesser 1993). The same occurs for the need of firms to segment their audience to be able to

identify and satisfy customer needs and achieve competitive advantage (Judd, Owens and Self 1989).

As literature supports the first law of behavioral genetics: “All human behavioral traits are heritable” (Turkheimer 2000, p.160), and the consistent finding that heritability accounts “for half of the variance of psychological traits” (Plomin et al. 2016, p.5), there is no doubt of the power of genes on individuals’ behaviors. Thus, in this study, we want to leverage genetic data to understand the systematic variation and continuity of traits and personality (McGue, Bacon, and Lykken 1993). Those human genetic characteristics have been continuously studied outside the field of marketing and could be leverage in strategic marketing for the creation of clusters from those customers’ genetic propensities.

2.7.1 PGS as main attributes for Genomics Segmentation

Scientists have been able to estimate genetic correlations between large samples of SNPs from unrelated individuals to understand phenotypic similarities (Uffelmann et al. 2021). With the aggregation of those SNPs into PGS important milestones have been reached in the DNA revolution (Plomin 2019). PGS have been used to identify individuals’ predictive tendencies and early propensities for behaviors and traits such as educational attainment and intelligence (Plomin 2023). For example, in a study of 1.1 million individuals, it was identified that 1,271 SNPs taken together in a PGS could explain 11-13% and 7-10% of the variance in educational attainment and cognitive abilities respectively (Lee et al. 2018). In the case of personality traits, GWA studies have found that personality traits are highly polygenic (Sanchez-Roige et al. 2017) and two-thirds of the variance in measured personality traits come from genetic influence (Bouchard 1994).

"The amount of genomics data available to researchers is increasing by a factor of 10 every year" (Berger, Daniels, and Yu 2016, p.1) which demonstrates how advances in DNA sequencing are outpacing advances in computer power (Moore's law) which doubles every 2 years (November 2018). This claim presented by many, including the National Human Genome Research Institute (NHGRI), demonstrates the incredible relevance and expectations in the understanding and applicability of the field of genomics. Moreover, the growth of genetic data involves the expansion of meta-analyses and molecular genetic studies which continuously identify new SNPs and aggregate them into improved PGS resulting in exponentially higher predictive power (Lee et al. 2018). With the rise of new technology, the accessibility of genetic data, and the expansion of analytical methods, marketing experts need to start thinking about the endless possibilities of consumers consented genetic data for the improvement of strategic segmentation, and product development (Hoffman et al. 2022), as well as more accurate predictions of early purchasing patterns (Daviet, Nave, and Wind 2022).

We understand PGS will "never predict complex traits with perfect precision" (Plomin and von Stumm 2022, p.50). However, PGS provide a quantitative measure of genetic predispositions (Davidson et al. 2021). The perfect examples are the multiple GWA meta-analyses using sample sizes of over thousands and million cases that have shown a higher percentage of genetic predictability of variations in cognitive and personality traits such as educational attainment (Lee et al. 2018; Selzam et al. 2019), cognitive abilities (Davies et al. 2018), extraversion (van den Berg et al. 2016), neuroticism, subjective well-being and depressive symptoms (Okbay et al. 2016a).

Hence, we know that genomics data is a "door-opener for nascent customer hyper-segmentation metrics based on hyper-personalization" (Ivanova-Kadiri 2023, p.623). Following

the premise that the deeper our knowledge and understanding of our customers, the better we, marketers, can convey and promote brands and products (Plummer 1974), the study proposes a *post hoc* segmentation approach leveraging this new genetic constructs, PGS, to create clusters with different levels of genetic propensities and better understand their consumption patterns. Genomics segmentation will be a data-driven strategy that combines PGS as main attributes to create homogeneous clusters of the marketplace. As descriptive factors of past tendencies, commonly used in traditional segmentation bases, only provide a partial view of consumer tendencies without the possibility of predicting future trends (Firat and Shultz 1997; Haley 1968), the study builds on over 100 years of behavioral genetics and strategic segmentation theory to understand the unaccounted 30-50% cognitive and personality traits explained by genetic makeup.

2.7.2 Genomics Segmentation: Creation of Clusters Combining PGS for Personality and Cognitive Traits

For firms to touch the heart of consumers and build stronger relationships with customers, "they need to consider the customer's human nature" (Myers, Sen and Alexandrov 2010, p. 4). Longitudinal genetic studies of 'nature' have shown that age-to-age phenotypic stability comes from genetics, and changes over time are primarily a result of the environment (Plomin et al. 2016). As the literature review showed above, personality and cognitive traits are consistent and stable throughout a period of over 40 years (Soldz and Vaillant 1999). Thus, genomics segmentation provides a redefinition of the key variables used to define segments of the marketplace, by using PGS for personality traits such as extraversion and neuroticism, and cognitive traits and emotions such as general cognition, educational attainment, subjective well-being, anxiety, and depression symptoms. Using a combination of PGS for the formation of

clusters will help us better understand the marketplace early in the customer journey and early in the life of consumers.

These PGS present high, medium, and low genetic tendencies for each individual. Segments are created through clustering analysis integrating the combination of different levels of genetic propensities per specific trait. For example, if we combine PGS for the Big Five personality traits, clusters should follow the findings of longitudinal traditional personality research where individuals high in neuroticism tend to be unhappy, are susceptible to depressive symptoms, high levels of anxiety and have highly impulsive behaviors (Wallace, Newman and Bachorowski 1991). On the other side, extraversion is primarily characterized by high levels of sociability and impulsivity (Cetola and Prinkey 1986). Considering negative behavior tendencies, research shows that individuals high in extraversion, high in neuroticism and low in conscientiousness have a higher risk of heavier alcohol consumption over time, while high levels of agreeableness and low openness to experience are related to abstinence of alcohol consumption (Hakulinen et al 2015; Li et al. 2017). In an effort to apply all these findings to the marketing field, we study the usage of PGS for cognitive and personality traits in segmentation strategies specifically PGS for educational attainment, general cognition, extraversion, neuroticism, depressive symptoms, anxiety, and subjective well-being.

Moreover, effectiveness of traditional segmentation strategies based on personality traits has been considered limited as it is very difficult for marketers to measure consumers' personality types, reaching individuals with specific traits and recognizing the immediate effects on consumer behavior (Judd, Owens and Self 1989). Consequently, the use of genetic data will expand the effectiveness of the study of personality and cognitive traits to segment a market, as these biological data are consistent and stable over individuals' lifespan.

Based on the above details, the first hypothesis proposes:

H1: "With the combination of polygenic scores, clusters will emerge characterized by different levels of extraversion, neuroticism, educational attainment, general cognition, subjective well-being, depressive symptoms, and anxiety tendencies.

2.7.3 Genomics Segmentation of the Alcohol Market

As we want to demonstrate the applicability of PGS for the explanation and segmentation of consumer markets, literature shows that the alcohol consumption market is heavily influenced by personality and cognitive traits especially health-related behaviors such as smoking, drinking alcohol, or exercising (Hakulinen and Jokela 2018). Alcohol consumption and alcohol usage disorder have been identified as highly polygenic especially because of the many genes associated with alcohol metabolism (Edenberg and Foroud 2013). In addition, alcohol consumption is not only highly correlated to different diseases but is also associated with different traits (Edenberg and Foroud 2013) such as sociability and extraversion (Cook et al. 1998), anxiety and phobias (Kushner, Sher, and Beitman 1990) and level of education (Tomkins et al. 2007).

Consequently, we are using PGS variables to possibly capture the influence of genetic data over the patterns of consumption of alcohol. Specifically, we selected PGS for *neuroticism* as literature shows that there is a high correlation between neurotic personality and highly impulsive, hostile, and pessimistic behaviors (Costa and McCrae 1999) such as smoking, alcohol consumption, and physical inactivity (Hakulinen and Jokela 2018). In the case of the PGS of *extraversion,* studies show that extraverted individuals experience positive emotions with highly sociable and energetic behaviors (Costa and McCrae 1999; DeNeve and Cooper 1998) including high and increasing alcohol consumption (Hakulinen et al. 2015). The same goes for PGS for

general cognition which is expected to follow results that recognize its association with memory, reasoning, processing speed, spatial ability, and organized thinking. It has been found that high levels of general cognition tend to increase the activity and engagement of individuals (Davies et al. 2018), as well as moderate the risk of alcoholism (Finn and Hall 2004). Regarding PGS for *educational attainment*, research shows that the number of years of schooling completed provides more cognitive resources to process complex information and make decisions. Moreover, education is a key determinant of occupation and income level, as well as it is one of the main factors for consumer's income potential and consumption patterns (Hoyer, MacInnis, and Pieters (2012). Thus, education level has been found to be strongly associated with hazardous levels of drinking when the number of years of education is low. *Subjective well-being* is also a human trait that has a high correlation with alcohol consumption. It has been found that people are happier at the moment of drinking, showing a short-term reward (Geiger and MacKerron (2016); in addition, better cognition and well-being have been associated with moderate levels of alcohol consumption (Lang et al. 2007).

Hence, based on the above literature, we are expecting our study to follow very similar research findings related to alcohol consumption. Overall, we are trying to show the applicability of PGS through the combination of several ones, to provide a deeper understanding of each consumer segment in the alcohol consumption market and strategically develop better-personalized marketing offerings. Specifically, our second hypothesis is:

H2: Clusters created will be differentially related to alcohol consumption.

CHAPTER III. METHODOLOGY

The dissertation proposed for the first time the applicability of genetic data for the creation of clusters and the understanding of consumer genetic profiles for marketing purposes. Polygenic scores (PGS) were independent variables in this study which include PGS for general cognition, PGS for extraversion, PGS for neuroticism, PGS for educational attainment, PGS for subjective well-being, PGS for depressive symptoms, PGS for anxiety. PGS are the weighted sum of multiple genetic variants (SNPs) to provide a quantitative measure of genetic predispositions (Davidson et al. 2021). The participants' information related to age in years, education level, income level, and ethnicity served as control variables for all hypotheses. Moreover, the dependent variable, level of alcohol consumption, was measured by the number of days in a week the participants consume alcohol.

Despite the lack of research in the area of genomics marketing, in our previous chapters, we demonstrated substantial literature from other disciplines on the advancements of behavioral genetics. The literature presented in Chapter 2 supports the theoretical framework for the development of a genomics segmentation base. The purpose of Chapter 3 is to explore the empirical applicability of the genomics segmentation base through a cluster analysis model. A detailed explanation of the study's design, sample size, specific variables and statistical model are presented below.

3.1 Research Design

The study proposed a quantitative research methodology for analyzing the genetic data and consumption patterns obtained from the Health and Retirement Study (HRS), sponsored by the National Institute of Aging (grant number NIA U01AG009740) and conducted by the University of Michigan. The unit of analysis for the model to be tested is group-level. The study

presented the combination of PGS for cognitive and personality traits as a basis or indicators to create distinct clusters of homogeneous individuals, each uniquely predictive of consumption patterns. Furthermore, an analysis of covariance is used to understand the similarities and differences between and within clusters related to alcohol consumption patterns. The study was designed to extend the theoretical framework and presented the empirical testing of the assumption that PGS could produce a more systematic market segmentation approach as it is known that genetic data is stable since the moment of inception.

Cluster analysis is a multivariate technique widely used by academics and practitioners in the marketing field (Rao and Wang 1995). In marketing, its primary use has been for market segmentation, basically to empirically identify homogeneous groups of products, consumers, or occasions for further analysis and to understand buyer behaviors (Punj and Stewart 1983). This analysis used a statistical method for classification with no prior assumptions (Punj and Stewart 1983). With its simple structure, cluster analysis shows relationships that might not be revealed otherwise (Hair et al. 2019).

K-means clustering is used to define groupings of observations with the minimal possible total intra-cluster (within-cluster) variation. The within-cluster variation is the sum of squared distances Euclidean distances between items and its corresponding centroid:

$$W(C_k) = \sum_{x_i \epsilon C_k} (x_i - \mu_k)^2$$

Where:

- x_i is an observation point belonging to cluster C_k

- μ_k correspond to the mean value of the points assigned to cluster C_k

The equation above shows how each observation (x_i) is assigned to a specific cluster minimizing the sum of squares (SS) distance of the observation to the assigned cluster centers (μ_k). The total within-cluster sum of square is the measurement that shows the level of compactness (i.e. goodness) of the clustering. In the study, we looked into the smallest possible total within-cluster sum of square. Some of the limitations of k-means cluster include (1) favoring clusters of relatively equal size (Leiter and Maslach 2016); (2) no agreed rules to determine the number of cluster (Punj and Stewart 1983); (3) the number of clusters are decided from a data-based criteria to subjectively identify the distance between variables and the cluster mean; and (4) subgroups created could be skewed if variables have large ranges (Stanley, Kellermanns and Zellweger 2017).

The research design identified the significant difference between each cluster using an analysis of variance (ANOVA) and an analysis of covariance (ANCOVA) to evaluate the validity of cluster solutions (Tuma, Decker, and Scholz 2011). An ANOVA is a statistical technique that examines if two or more groups come from populations with equal means (Hair et al. 2019). ANCOVA is a similar technique utilized to compare the means of two or more groups of samples but in this case, it takes into consideration covariates (Hair et al. 2019). Covariates or control variables bring the advantage of extracting "extraneous influences from the dependent variable" (p. 394) and increasing variance within clusters (Hair et al. 2019).

Due to the nature of the hypotheses and based on the literature, cluster analysis and ANCOVA had been the most appropriate methods for analyzing the data in this study (Punj and Stewart 1983; Reutterer and Dan 2021; Tuma, Decker, and Scholz 2011). In summary, as the

first study of its kind in genomics marketing, the research aimed to identify the empirical applicability of polygenic scores to create clusters integrated by a combination of consumers with similar genetic propensities and consumption patterns.

3.1.1 Data Source and Sample

The study used secondary data from the Health and Retirement Study (HRS) sponsored by the National Institute of Aging (grant number NIA U01AG009740) and conducted by the University of Michigan. The University of Michigan Health and Retirement Study (HRS) is a longitudinal panel study that surveys a sample of approximately 20,000 people in the United States addressing important questions about the challenges and opportunities of health, retirement, and aging. The data was extracted from 2 different databases of the HRS data products – Polygenic Score Data, and 2020 HRS Core. The Polygenic Score Data has two samples: the European ancestry database with 12,090 individuals' genetic data and the African ancestry database with 3,100 individuals' genetic data. The Polygenic Score Database was constructed for a variety of phenotypes from HRS respondents who provided salivary DNA between 2006 and 2012. Those PGS are based on single, replicated GWAS and have been updated as new large GWAS are published for new or existing phenotypes. The 2020 HRS Core a national longitudinal study of the health, economic, marital, family status and support systems of older Americans. The data was collected between March 2020 and May 2021.

Databases are combined by unique identifiers. Observations in the sample have a unique household identifier number (HHID) and a unique respondent person identifier number (RPN). These identifiers do not change across waves of surveys and interviews. Consequently, each observation has a unique overall identifier by combining both the household identifier and the respondent person identifier number.

3.1.2 Definition of Key Variables

3.1.2.1 Independent Variables: Polygenic Scores

The study used 7 different PGS within cognitive and personality traits for the definition of clusters.

- *PGS for Extraversion*: Extraversion is a personality trait relatively stable and associated with psychosocial, lifestyle, and positive health outcomes (van den Berg et al. 2016). It has been associated with traits of sociability and impulsiveness (Eaves and Eysenck 1975). This PGS was created by the HRS using the GWA meta-analysis from van den Berg et al. (2016) which identified 74 SNPs that combined into a PGS could estimate around 5% of the variance in the extraversion trait. Overall, studies found that extraversion is a highly polygenic trait (van den Berg et al. 2016).
- *PGS for Neuroticism*: Neuroticism is a stable personality trait that arises early in life (Barnhofer, Duggan, and Griffith 2011). It involves experiencing relatively intense negative emotions and it is highly associated with psychiatric disorders related to emotional dysregulation, substance abuse, depression, and anxiety (De Moor et al. 2015). This PGS was created by the HRS using the GWA study conducted by the Social Science Genetic Association Consortium (SSGAC) as part of the subjective well-being study (Okbay et al. 2016a). This study yielded 11 lead SNPs for the explanation of 10% variance for neuroticism personality traits. Other studies such as De Moor et al. (2015) reported a proportion of 15% of variance explained in neuroticism.
- *PGS for General Cognition*: General cognitive ability includes verbal, spatial, memory and processing speed (Plomin and Spinath 2002) and it is associated with IQ (Pedersen, Plomin, and McClearn 1994). This PGS was created by the HRS following the GWAS

from Davies et al. (2018). This study identified 13 significant SNPs that can explain between 28-29% of the variance in general cognitive functions. Another study from Davies et al. (2015) found that there is a genetic overlap between general cognitive function, reaction time, brain structure, eyesight, and longevity.

- *PGS for Subjective Well-being*: Subjective well-being is a trait related to life satisfaction; it involves factors that lead people to "subjectively experience their life as worthwhile and rewarding" (Diener, Oishi, and Tay 2018, p.253). This PGS was created by the HRS following the GWAS from Okbay et al. (2016a). The study identified 3 lead SNPs that can explain almost 5% of the variance in subjective well-being.
- *PGS for Depressive Symptoms*: Depression is a cognitive trait characterized by introversion, passivity, cheerlessness, pessimism, and self-criticism (Klein, Kotov, and Bufferd 2011). This PGS was created by the HRS following the 2016 auxiliary GWAS conducted by the Social Science Genetic Association Consortium (SSGAC) as part of the subjective well-being GWAS from Okbay et al. (2016a). The study identified 2 loci associated with depressive symptoms and could explain almost 5% of the variance. Moreover, Okbay et al. (2016a) identified a highly genetic correlation between depressive symptoms and neuroticism.
- *PGS for Educational Attainment*: Educational attainment is cognitive function highly heritable and calculated by number of years of schooling completed (Okbay et al. 2016b) This PGS was created by the HRS following the large-scale GWAS which identified 1,271 SNPs that together can explained between 11-13% of the variance in educational attainment (Lee et al. 2018). Moreover, this study identified that there are phenotypes

highly genetically correlated with educational attainment: those are cognitive (test) performance and math abilities (Lee et al. 2018).

- *PGS for Anxiety*: Anxiety is an emotional trait and a core construct in all personality theories that involves the subjective experience of fear, negative and unpleasant emotions (Endler and Kocovski 2001). It is a multidimensional trait highly related to depressive symptoms and neuroticism (Barnhofer, Duggan, and Griffith 2011; Kotov et al. 2010) This PGS was created by the HRS following the GWA meta-analysis from Otowa et al. (2016) that identified 6.5 million SNPs and found a variance explained of 7-9% for anxiety.

  In Table 5 below, we present a brief of each independent variable.

Table 4. Independent Variables - Polygenic Scores (PGS)

PGS	Definition	Meta-analysis results	Explanation of variance	GWA study
PGS for Educational Attainment (3) (EA)	EA is measured as number of years of schooling completed	Over 1.1 million individuals identified significant SNPs in 1,271 loci.	Explain 11-13% of the variance in educational attainment	Created from the 2018 study conducted by the Social Science Genetic Association Consortium (SSGAC). Lee et al. (2018)
PGS for Neuroticism	Highly impulsive, hostile, and pessimistic personality trait associated with negative behaviors, anxiety and fear.	170,911 observations which yielded 11 lead SNPs.	Estimated at almost 10% of variance	Created from a 2016 auxiliary GWAS conducted by the Social Science Genetic Association Consortium (SSGAC) as part of the subjective well-being GWAS. Okbay et al. (2016a)
PGS for Extraversion	Stable and heritable personality trait associated with psychosocial, lifestyle, and positive health outcomes	63,030 individuals and 74 SNPs were identified.	Estimated 5% of variance in extraversion is explained.	2016 study conducted by the Genetics of Personality Consortium (GPC). van den Berg et al. (2016)
PGS for General Cognition	General cognitive ability includes reasoning, memory, processing speed and spatial ability	53,949 individuals and 13 SNPs were significant.	PGS explained 28% to 29% of variance in general cognitive function.	Davies et al. (2018)
PGS for Subjective Wellbeing	Trait related to life satisfaction, positive affect, and happiness.	298,420 observations combining phenotype of life satisfaction and positive affect. Identified 3 lead SNPs.	Estimated at almost 5% of variance	Okbay et al. (2016a)
PGS for Depressive Symptoms	Symptoms characterized by negative affect, anxiety, low energy, pessimism, bodily aches and pains.	180,866 observations and identified 2 loci associated with depressive symptoms	Estimated at almost 5% of variance	Created from a 2016 auxiliary GWAS conducted by the Social Science Genetic Association Consortium (SSGAC) as part of the subjective well-being GWAS. Okbay et al. (2016a)
PGS for Anxiety	Anxiety is an emotional trait related to fear, phobias, and negative emotions.	18,000 individuals using 6.5 million imputed SNPs	Explain between 7-9% of variance	Otowa et al. (2016)

3.2.2.2 Dependent Variables: Levels of alcohol consumption

The dependent variable is related to the alcohol consumption market. Data for alcohol consumption level comes from the 2020 HRS Core questionnaire. The 2020 Core HRS is part of the national longitudinal study of the economic, health, marital, and family status of older

Americans from the Health and Retirement Study (HRS). The data was collected from March 2020 through May 2021.

The questions are pulled from the "Physical Health" section of the survey. Below are the details of the question:

- *DV* - Question: In the last three months, on average, how many days per week have you had any alcohol to drink? (For example, beer, wine or any drink containing liquor). Answers include: 0 to 7 days a week.

In Table 6 below, we present a full description of the variable used from the 2020 HRS Core database.

Table 5. Dependent Variables - Alcohol Consumption

Variable Name	Variable Code	Definition	Survey
DV - # of days per week drinking alcohol	R13DRINKD	How many days per week you normally drink	2020 HRS Core

3.2.2.3 Control Variables

We have identified specific variables that could affect the results of this study, but they are not part of the scope, so they warrant some special attention as control variables. Control variables or covariates should be correlated to the dependent variable (Hair et al. 2019) Taking into consideration that our data source, the Health and Retirement Study (HRS), concentrates on a population of retired and seniors, it is highly important to take into account *age* as a control variable. Moreover, we have added specific variables that have some kind of correlation with the dependent variable – level of alcohol consumption. Those variables are level of education, income level, and ethnicity.

- *C1 - Age in Years*: Age is a continuous variable as count age in years. Our research sample has a mean of 70 years old, so results are going to be limited to a certain population only and not be able to be generalizable to all population.
- *C2 - Years of Education:* Years of education is also a continuous variable that goes from 0 years of education to 17 years of education. Answers are continuous and it can be classified as: 0-11 years: no degree; 12-13: HS/GED; 14-15: AA; 16-17: BA. It is a highly important variable as literature shows its implications on consumer expending, consumption patterns, employment level, and disposable income.
- *C3 – Income level*: Amount from wages and salary income received over the last calendar year before taxes and other deductions. Variable is measured in nominal dollars. Answers include a positive number from 0 to millions; 998 represents "don't know", and 999 represents "refused to answer".
- *C4 – Ethnicity*: We have databases representing two different ethnicities – European and African. The HRS recommends keeping them separate and running the studies separately because PGS have been created specifically from European samples. Consequently, PGS might have a different explanation power for the African ancestry sample. The study compared genetic tendencies between each ancestry for each PGS, and the statistical model is run with both datasets together and separate.

3.2.3 Summary

This dissertation investigated the applicability of a new genomics segmentation base using PGS to create clusters and better understand consumer genetic profiles. Those clusters are integrated by a combination of similar genetic propensities and an analysis of covariance presented a comparison of alcohol consumption levels.

One of the most difficult aspects in a segmentation study is the translation of the results into marketing strategies (Wind 1978). Many times, segmentation strategy is seen as an analytic tool only to look for homogeneous responses to marketing stimuli. However, strategic segments should be sufficiently distinctive on customer criteria and behavior within a period of time to justify the additional costs and efforts of a personalized strategy (Rao and Wang 1995). Consequently, this chapter offered a detailed discussion of the research design and empirical applicability utilized in this dissertation. Chapter 4 presents the study findings, descriptive results, and conclusions of the data analysis, accompanied by the corresponding tables and figures to explain the relationship between each of the variables. Furthermore, findings are interpreted in Chapter 5 expanding on the limitations and suggestions for further research for this study, as well as highlighting ethical implications for the new genomics segmentation base. Overall, this cross-sectional study showed how polygenic scores provide distinctiveness in the profile of consumers into clusters to create more targeted and personalized initiatives.

CHAPTER IV. RESULTS

The dissertation presented a new genomic segmentation base through the empirical applicability of polygenic scores for the creation of clusters and the definition of profiles of consumers based on genetic tendencies and level of alcohol consumption. This chapter laid out the results from empirically testing each of the hypotheses in the research model. The concept model used in this research is depicted below:

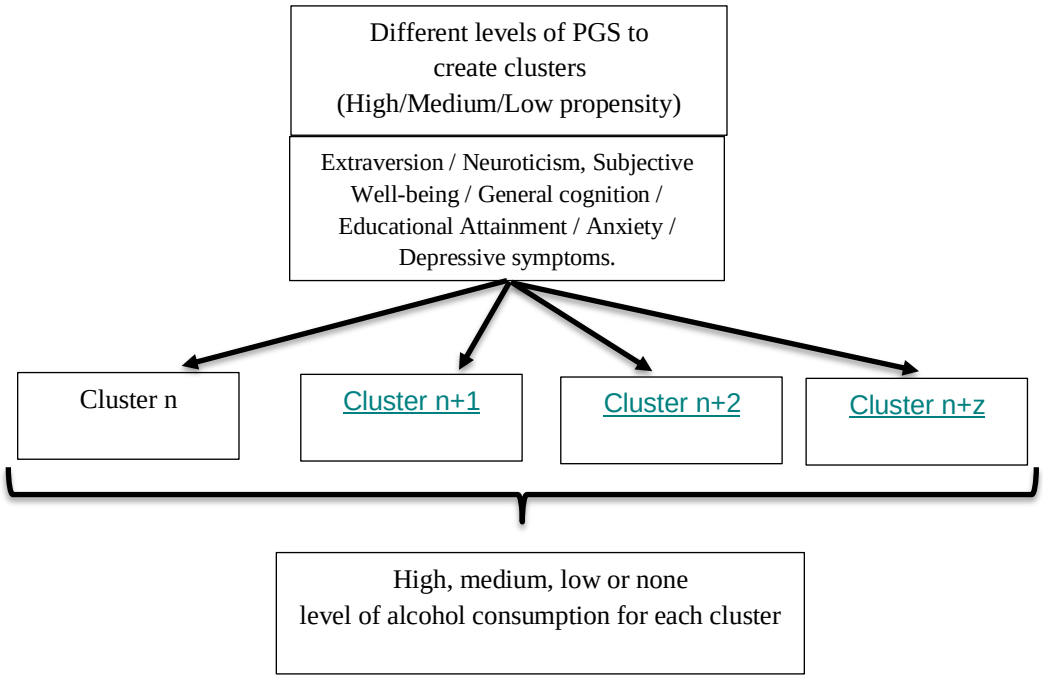


Figure 6. Conceptual Model

This chapter provided the four stages of the empirical analysis. First, a specific description of the sample, including details on missing data is presented. Second, a descriptive analysis of the sample profile and bivariate correlation analysis are performed. Third, K-means cluster analysis was used to group the observations into homogenous clusters in terms of PGS.

Thereafter, ANOVA was used to identify statistical differences between clusters in terms of PGS for personality and cognitive traits. Fourth, ANOVA and ANCOVAs are presented to identify statistical differences between clusters in terms of the dependent variable, level of alcohol consumption testing the main effect (ANOVA) and incorporating covariates (ANCOVA).

4.1 Preliminary Analysis

The two databases, Polygenic Score Data and the 2020 HRS Core, had been imported from the Health and Retirement Study HRS conducted by the University of Michigan. For the merging of the databases, a unique number identifier is used. This unique number identifier combined the household identifier and the respondent person identifier that did not change across waves of surveys and interviews. Data for the independent variables, PGS, was obtained from the Polygenic Score Data from individuals who provided salivary DNA between 2006 and 2012. Data for dependent and control variables was obtained from the 2020 HRS Core from the 2020 Core questionnaire. It is important to highlight that the Polygenic Score Data was provided in two separate data bases grouped by ancestry – European American and African American ancestry.

Data sets were examined for missing values. If missing values were identified, the complete row was removed. There are several techniques to manage missing values including removing the complete row, or imputing the missing values based on mean, media, or mode imputations. The technique is determined based on the impact of missing data. Missing data is important to considered as it may impact the sample size and lead to erroneous or biased results (Hair et al. 2019). In this case, we applied list-wise deletion as there were excessive levels of missing values. Observations with missing values for level of education and number of days drinking per week were removed. Also, as we can see on Table 6 and Table 7, the dependent

variable – number of days drinking per week should be a 1 to 7 days a week; there is a number 8 that represents those individuals that answered, 'I don't know'. Consequently, those observations were removed as the variable is a continuous variable measuring number of days drinking per week and the answers are between 0 days and 7 days.

The European ancestry database includes a total of 3,533 observations when we combined the PGS database with the 2020 HRS Core database and cleaned the data. The African ancestry database includes 810 observations when we combined the PGS database with the 2020 HRS Core database and cleaned the data.

Table 6. Preliminary Descriptive Statistic for European Ancestry Data

Descriptive Statistics for European Ancestry Data

						Skewness		Kurtosis	
Variables	**N**	**Min**	**Max**	**Mean**	**Std. Deviation**	**Statistic**	**Std. Error**	**Statistic**	**Std. Error**
PGS for Neuroticism	3568	-3.71	3.26	-0.0373	0.99999	-0.026	0.041	-0.065	0.082
PGS for Subj. Wellbeing	3568	-3.85	3.30	0.0007	1.00486	-0.048	0.041	0.179	0.082
PGS for Depressive Symptoms	3568	-3.34	3.43	-0.0500	1.00522	0.045	0.041	-0.093	0.082
PGS for Extraversion	3568	-3.58	3.51	0.0166	1.00918	-0.114	0.041	-0.072	0.082
PGS for Educational Attainment	3568	-3.14	3.37	0.0871	1.00032	0.052	0.041	-0.135	0.082
PGS for Anxiety	3568	-3.49	3.12	0.0021	0.98017	-0.101	0.041	-0.061	0.082
PGS for General Cognition	3568	-3.41	4.20	0.0805	1.00164	0.084	0.041	0.059	0.082
Age in years	3568	41	102	72.47	9.368	0.206	0.041	-0.581	0.082
# of Days Drinking per Week	3568	0	**8**	2.53	2.543	0.720	0.041	-0.919	0.082
Income Level	992	0	9999999	644361.14	2334600.579	3.759	0.078	12.181	0.155
Level of Education	3568	0	17	14.15	2.239	-0.444	0.041	0.347	0.082
Valid N (listwise)	992								

Table 7. Preliminary Descriptive Statistic for African Ancestry Data

Descriptive Statistics for African Ancestry Data

Variable	N	Min	Max	Mean	Std. Deviation	Skewness Statistic	Skewness Std. Error	Kurtosis Statistic	Kurtosis Std. Error
PGS for Neuroticism	816	-3.15	3.09	-0.0238	0.99583	-0.122	0.086	0.032	0.171
PGS for Subj. Wellbeing	816	-3.17	3.17	-0.0185	1.00853	0.042	0.086	0.017	0.171
PGS for Depressive Symptoms	816	-3.01	3.36	-0.0535	1.00192	0.066	0.086	-0.235	0.171
PGS for Extraversion	816	-3.98	2.73	-0.0496	1.00313	0.055	0.086	-0.166	0.171
PGS for Educational Attainment	816	-2.86	3.37	0.0200	0.97197	0.094	0.086	0.193	0.171
PGS for Anxiety	816	-4.05	3.08	0.0251	0.99242	-0.044	0.086	0.526	0.171
PGS for General Cognition	816	-3.40	4.08	0.0732	1.00671	0.123	0.086	0.358	0.171
Age in years	816	32	98	67.67	8.133	0.249	0.086	1.069	0.171
# of Days Drinking per Week	816	0	**8**	1.66	1.875	1.430	0.086	1.660	0.171
Income Level	274	0	9999999	1858827.01	3853515.818	1.653	0.147	0.738	0.293
Level of Education	816	1	17	13.14	2.436	-0.590	0.086	1.409	0.171
Valid N (listwise)	274								

4.2 Descriptive Statistics and Correlation Analysis

After reviewing the missing data, a complete analysis of descriptive statistics was generated for all independent, dependent and control variables for European and African ancestry databases as shown in Table 8, 9, and 10. Most of the PGS for personality and cognitive traits move between -4 to +4. These indicate that the closer to +4, there is a higher level of propensity of that specific trait; the closer to 0 represents a medium tendency of the trait; and the closer to -4, represents very low to no propensity of that specific trait. Moreover, the average age of the sample is 71.5 years old combining the European and African ancestry. The age mean reflected how our sample is mainly integrated by seniors. We must remember the HRS is an entity that focuses on providing data for researchers to understand challenges and opportunities of aging in the senior population. The mean education level is 13.9 which reflects some college

level for the sample. Descriptives also showed how the average income for European ancestry is higher ($64K) than the African ancestry ($39K). Moreover, there are outliers and missing data for the control variable, income level, which will be addressed later in the study.

Table 8. Descriptive Statistics for European Ancestry Final Data

Descriptive Statistics For European Ancestry Final Data

						Skewness		Kurtosis	
Variables	**N**	**Min**	**Max**	**Mean**	**Std. Deviation**	**Statistic**	**Std. Error**	**Statistic**	**Std. Error**
PGS for Neuroticism	3533	-3.71	3.26	-0.0359	0.99857	-0.031	0.041	-0.052	0.082
PGS for Subj. Wellbeing	3533	-3.85	3.30	-0.0008	1.00543	-0.053	0.041	0.164	0.082
PGS for Depressive Symptoms	3533	-3.34	3.43	-0.0488	1.00641	0.043	0.041	-0.092	0.082
PGS for Extraversion	3533	-3.58	3.51	0.0176	1.01009	-0.116	0.041	-0.070	0.082
PGS for Educational Attainment	3533	-3.14	3.37	0.0888	1.00048	0.054	0.041	-0.136	0.082
PGS for Anxiety	3533	-3.49	3.12	0.0033	0.98147	-0.098	0.041	-0.067	0.082
PGS for General Cognition	3533	-3.41	4.20	0.0819	1.00118	0.088	0.041	0.068	0.082
Age in years	3533	41	102	72.41	9.375	0.215	0.041	-0.576	0.082
# of Days Drinking per Week	3533	0	7	2.47	2.496	0.730	0.041	-0.905	0.082
Income Level	989	0	9999999	636155.96	2318937.687	3.792	0.078	12.427	0.155
Level of Education	3533	0	17	14.15	2.235	-0.437	0.041	0.315	0.082
Valid N (listwise)	989								

Table 9. Descriptive Statistics for African Ancestry Final Data

Descriptive Statistics for African Ancestry Final Data									
						Skewness		Kurtosis	
Variables	N	Min	Max	Mean	Std. Deviation	Statistic	Std. Error	Statistic	Std. Error
PGS for Neuroticism	810	-3.15	3.09	-0.0266	0.99453	-0.122	0.086	0.044	0.172
PGS for Subj. Wellbeing	810	-3.17	3.17	-0.0187	1.00959	0.042	0.086	0.021	0.172
PGS for Depressive Symptoms	810	-3.01	3.36	-0.0540	1.00297	0.068	0.086	-0.233	0.172
PGS for Extraversion	810	-3.98	2.73	-0.0467	1.00285	0.050	0.086	-0.156	0.172
PGS for Educational Attainment	810	-2.86	3.37	0.0211	0.96972	0.110	0.086	0.183	0.172
PGS for Anxiety	810	-4.05	3.08	0.0269	0.99379	-0.046	0.086	0.527	0.172
PGS for General Cognition	810	-3.40	4.08	0.0733	1.00798	0.120	0.086	0.358	0.172
Age in years	810	32	98	67.63	8.123	0.253	0.086	1.094	0.172
# of Days Drinking per Week	810	0	7	1.61	1.800	1.381	0.086	1.516	0.172
Income Level	273	0	9999999	1865526.01	3858994.276	1.647	0.147	0.719	0.294
Level of Education	810	1	17	13.16	2.423	-0.587	0.086	1.439	0.172
Valid N (listwise)	273								

Table 10. Descriptive Statistics for European + African Ancestry Data Combined

Descriptive Statistics for European and African Ancestry Combined									
						Skewness		Kurtosis	
Variables	N	Min	Max	Mean	Std. Deviation	Statistic	Std. Error	Statistic	Std. Error
PGS for Neuroticism	4343	-3.710	3.26	-0.034	0.998	-0.048	0.037	-0.036	0.074
PGS for Subj. Wellbeing	4343	-3.850	3.30	-0.004	1.006	-0.035	0.037	0.135	0.074
PGS for Depressive Symptoms	4343	-3.340	3.43	-0.050	1.006	0.048	0.037	-0.119	0.074
PGS for Extraversion	4343	-3.980	3.51	0.006	1.009	-0.085	0.037	-0.093	0.074
PGS for Educational Attainment	4343	-3.140	3.37	0.076	0.995	0.066	0.037	-0.085	0.074
PGS for Anxiety	4343	-4.050	3.12	0.008	0.984	-0.088	0.037	0.047	0.074
PGS for General Cognition	4343	-3.410	4.2	0.081	1.003	0.094	0.037	0.119	0.074
Age in years	4343	32.000	102	71.520	9.341	0.254	0.037	-0.352	0.074
# of Days Drinking per Week	4343	**0.0**	**7.**	2.313	2.405	0.851	0.037	-0.630	0.074
Income Level	1262	0.0	9999999	902097.34	2771622.65	2.980	0.069	6.899	0.138
Level of Education	4343	0.0	17	13.969	2.304	-0.491	0.037	0.656	0.074
Valid N (listwise)	1262								

Table 11. Frequencies statistics for Ancestry variable (A=African, E=European)

Frequency by Ancestry (A= African, E=European)				
Variables	Frequency	Percent	Valid Percent	Cumulative Percent
African Ancestry	810	18.7	18.7	18.7
European Ancestry	3533	81.3	81.3	100.0
Total	4343	100.0	100.0	

Skewness and Kurtosis are within parameters for most variables. Income level is skewed (2.980) as distribution is unbalanced to the left. In the case of kurtosis (6.899), income level showed a leptokurtic distribution. Income level not only needs to be normalized due to outliers, but also includes a great number of missing values (70%) for the overall dataset. Furthermore, removing those missing values and outliers for income level reduces the dataset to less than half.

As the variable is too problematic and will compromise our results due to such a small data set, we have decided to remove this control variable from the study.

The HRS recommends completing the analyses separately by ancestral group because the majority of GWAS used to create the SNP weights come from GWAS with European ancestry groups. Consequently, PGS for other ancestry groups, such as African Americans, may not have the same predictive capacity (Ware et al. 2021). However, we performed an analysis of variance between ancestry data sets to identify if the pattern of behaviors of PGS are different. An ANOVA was conducted to determine similarity of the European and African ancestry databases. Table 12 below shows how differences between the European and African ancestry groups are not statistically significant. Consequently, it is understood that the groups of PGS behave in very similar patterns (there is no difference), thus data sets can be combined for the following empirical models. Note: Separate study of each ancestry data set can be found in Appendix A: Empirical Study per Ancestry.

Table 12. ANOVA Comparison of Ancestry Data

ANOVA Comparison of Ancestry Data						
Variables		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
PGS for Neuroticism	Between Groups	0.057	1	0.057	.057	**0.811**
	Within Groups	4,334.818	4,341	0.999		
	Total	4,334.875	4,342			
PGS for Subj. Wellbeing	Between Groups	0.211	1	0.211	.210	**0.647**
	Within Groups	4,353.069	4,341	1.003		
	Total	4,353.280	4,342			
PGS for Depressive Symptoms	Between Groups	0.018	1	0.018	.018	**0.894**
	Within Groups	4,359.715	4,341	1.004		
	Total	4,359.732	4,342			
PGS for Extraversion	Between Groups	2.678	1	2.678	.675	**0.102**
	Within Groups	4,345.008	4,341	1.001		
	Total	4,347.686	4,342			
PGS for Educational Attainment	Between Groups	3.051	1	3.051	.050	**0.081**
	Within Groups	4,342.760	4,341	1.000		
	Total	4,345.811	4,342			
PGS for Anxiety	Between Groups	0.381	1	0.381	.380	**0.538**
	Within Groups	4,355.422	4,341	1.003		
	Total	4,355.803	4,342			
PGS for General Cognition	Between Groups	0.028	1	0.028	.028	**0.866**
	Within Groups	4,347.756	4,341	1.002		
	Total	4,347.784	4,342			

Bivariate correlations between the variables in the study are provided in Table 13. Variables have been standardized (Z-score) as variables have different scales. Some of the variables are significantly correlated. When analyzing the independent variables, there are many

variables that follow the literature about the correlation between personality and cognitive traits. PGS for neuroticism, PGS for depressive symptoms and PGs for anxiety are positively correlated, and they are negatively correlated to PGS for subjective well-being, PGS for extraversion, PGS for educational attainment and PGs for general cognition. On the other side, PGS for extraversion and PGS for subjective well-being are positively correlated following traditional personality research; as well as PGS for educational attainment and PGS for general cognition positively correlate. Also, PGs for subjective well-being is positive correlated to PGS for general cognition. It is important to highlight that the magnitude of the correlations are low (below 0.5). When looking at the dependent variable, number of days drinking per week has a low positive correlation with PGS for educational attainment, PGS for general cognition and level of education. In the case of control variables, age in years has a low positive correlation with PGS for educational attainment and PGS for general cognition; and a negative low correlation with PGS for anxiety. Level of education has a positive low correlation with PGS for general cognition, PGS for educational attainment, and the DV number of days drinking per week; and a negative low correlation with PGS for depressive symptoms and income level.

Table 13. Pearson Correlation

Pearson Correlation											
Variables	1	2	3	4	5	6	7	8	9	10	11
(1) PGS for Neuroticism	1.00										
(2) PGS for Subj. Wellbeing	**-0.380****	1.00									
(3) PGS for Depressive Symptoms	**0.552****	**-0.305****	1.00								
(4) PGS for Extraversion	**-0.291****	**0.207****	**-0.093****	1.00							
(5) PGS for Educational Attainment	**-0.076****	0.015	**-0.128****	0.029	1.00						
(6) PGS for Anxiety	**0.105****	**-0.099****	**0.094****	**-0.035***	**-0.058****	1.00					
(7) PGS for General Cognition	**-0.053****	**0.047****	**-0.121****	0.001	**0.490****	**-0.159****	1.00				
(8) Age in years	-0.019	0.027	-0.005	0.008	**0.069****	**-0.037***	**0.043****	1.00			
(9) # of Days Drinking per week	-0.024	0.029	-0.016	0.012	**0.061****	-0.001	**0.044****	0.025	1.00		
(10) Income Level	-0.013	0.001	0.007	0.023	0.006	-0.015	0.025	0.042	-0.032	1.00	
(11) Level of Education	-0.030	-0.004	**-0.041****	0.025	**0.237****	-0.027	**0.180****	-0.018	**0.121****	**-0.091****	1.00

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

4.3 Cluster Analysis

A cluster analysis was used to address the two hypotheses presented earlier. A K-means clustering of the PGS is used to create clusters representing homogeneous consumers in terms of their genetic propensity. Since the number of clusters is not known beforehand, the ‘elbow criterion’ is used to identify the best number to apply in the K-means algorithm (Bholowalia and Kumar 2014). The elbow method was created in RStudio to review the percentage of variance explained as a function of the number of clusters. Based on Figure 7, the elbow method, it looks like a three-cluster is the best solution. After three-clusters, we can see the line flatten.

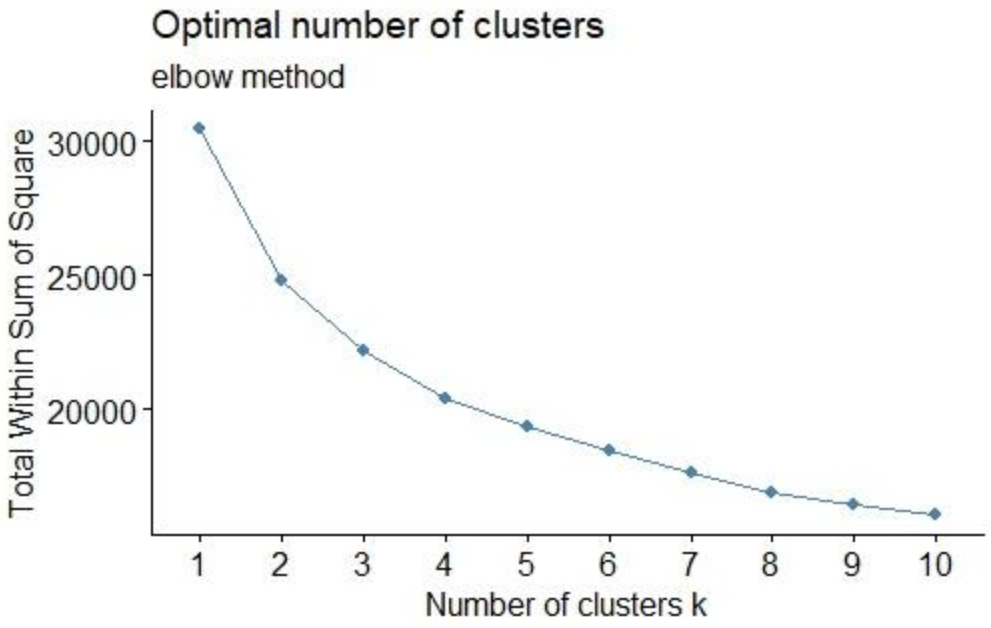


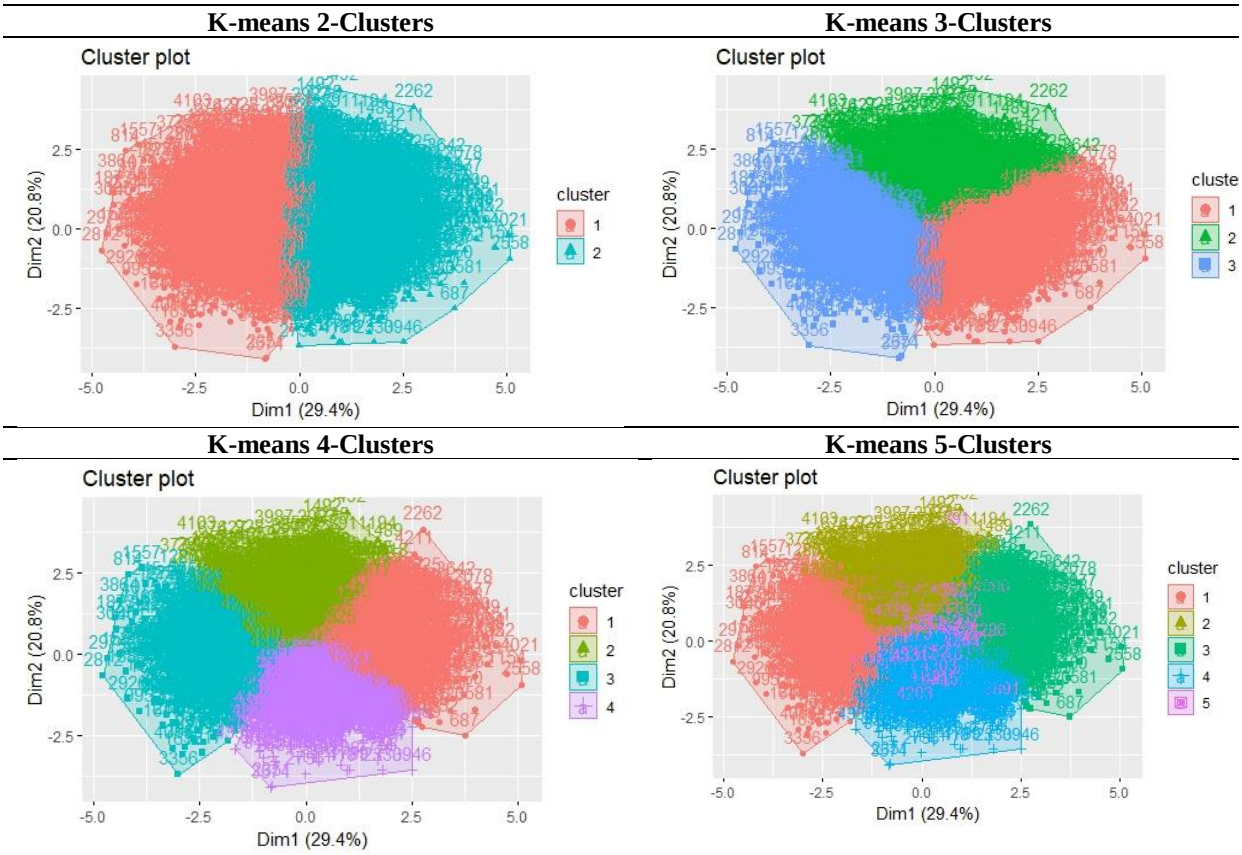
Figure 7. Elbow Method

A robustness check of the three-cluster solution is performed comparing the percentage of change between "within cluster sum of squares" (WCSS) of each possible cluster. The total WCSS measures the compactness (i.e. goodness) of the clustering. It essentially measures the variability of observations within each cluster. The higher the value the greater variability of the observations within cluster. The three-cluster solution has a WCSS of 27.3% which is lower than four and five clusters. Table 14 compares the WCSS, and Figure 8 shows the cluster visual comparison for two-clusters, three-clusters, four-clusters, and five-clusters based on PGS. There is a difference between WCSS of two-clusters and WCSS of three-clusters of 8.8%; a difference between WCSS of three-clusters and WCSS of four-clusters of 5.6%; and a smaller change starting with WCSS of four-clusters and WCSS of five-clusters of below 4%. These calculations confirmed the elbow starts in the three-clusters solution. Looking at Figure 8, the dispersion of observations bleeding over the other clusters can be seen in the four-clusters and five-clusters,

and there is better compactness in the three-clusters solution. Consequently, a three-clusters solution is used for the K-mean cluster analysis.

Table 14. K-Means Cluster: Within Cluster Sum of Squares (WCSS)

K-means 2-Clusters	K-means 3-Clusters
Within cluster sum of squares by cluster [1] 11648.30 [2] 13150.81 **(total_SS = 18.5%)**	Within cluster sum of squares by cluster [1] 7899.560 [2] 6775.981 [3] 7458.849 **(total_SS = 27.3%)**
K-means 4-Clusters	**K-means 5-Clusters**
Within cluster sum of squares by cluster [1] 4695.117 [2] 5822.353 [3] 4366.773 [4] 5547.869 **(total_SS = 32.9%)**	Within cluster sum of squares by cluster [1] 4033.651 [2] 4530.352 [3] 3590.616 [4] 4003.713 [5] 3118.604 **(total_SS = 36.7%)**



Dim1: Dimension 1; Dim2: Dimension 2; K-means Cluster Analysis

Figure 8. Cluster Visualization Comparison

K-means clustering of the data was performed. Table 15 shows the final centroid means for the three-clusters, and Figure 9 the bar graph of the final centers. Looking at the final cluster means, it is inferred that cluster 1 shows high levels of genetic propensities for wellbeing and extraversion traits; cluster 2 has high levels of genetic propensities for educational attainment and general cognition; and cluster 3 includes high levels of genetic propensities for neuroticism and depressive symptoms.

Table 15. K-means Cluster Centers

K-Means Final Cluster Centers			
Variables	**Cluster**		
	1	**2**	**3**
PGS for Neuroticism	-0.7764	-0.0706	**0.7690**
PGS for Subj. Wellbeing	**0.6274**	-0.0025	-0.5751
PGS for Depressive Symptoms	-0.5949	-0.2083	**0.7197**
PGS for Extraversion	**0.4833**	-0.0141	-0.4287
PGS for Educational Attainment	-0.3437	**0.9002**	-0.4318
PGS for Anxiety	0.0114	-0.3247	**0.2634**
PGS for General Cognition	-0.3973	**0.9162**	-0.3983

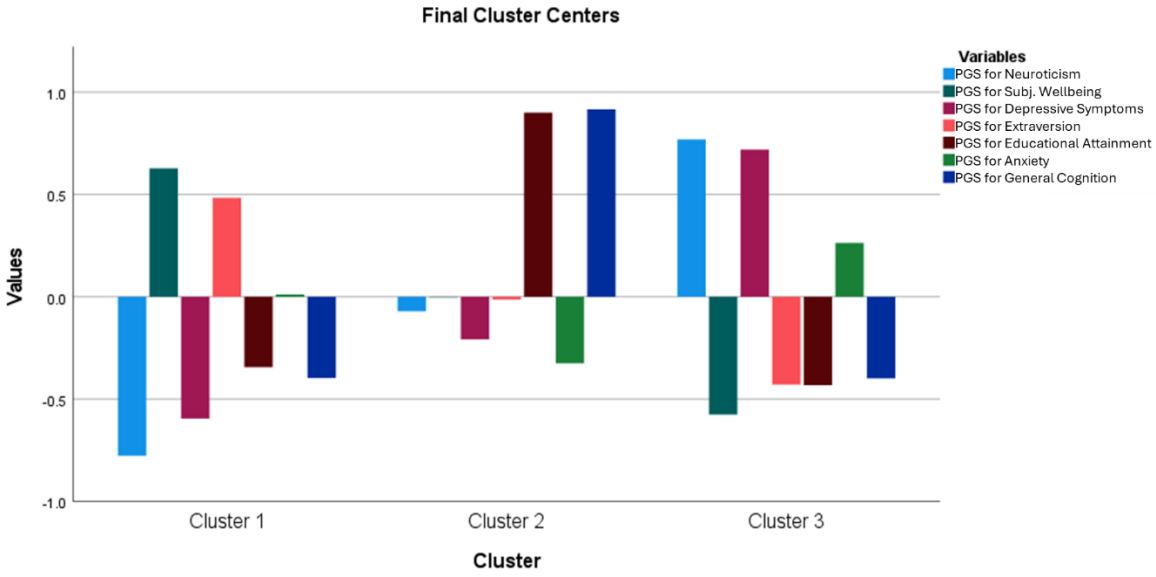


Figure 9. Final Cluster Centers Bar Graph

Table 16. K-means ANOVA of Three-Clusters Created

K-Means ANOVA of Three-Clusters Created						
Variables	**Cluster**		**Error**		**F**	**Sig.**
	Mean Square	**Df**	**Mean Square**	**df**		
PGS for Neuroticism	905.377	2	0.582	4340	1556.715	**0.000**
PGS for Subj. Wellbeing	545.280	2	0.752	4340	725.319	**0.000**
PGS for Depressive Symptoms	693.127	2	0.685	4340	1011.667	**0.000**
PGS for Extraversion	313.834	2	0.857	4340	366.138	**0.000**
PGS for Educational Attainment	767.399	2	0.648	4340	1184.808	**0.000**
PGS for Anxiety	124.494	2	0.946	4340	131.563	**0.000**
PGS for General Cognition	793.247	2	0.636	4340	1246.769	**0.000**

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

A post-hoc study was conducted to confirm significant difference between each cluster means based on PGS. Table 17 shows ANOVA and post-hoc test for the PGS. ANOVA confirmed significant difference between the three clusters. However, we must remember that an ANOVA does not specify which groups are statistically significant different from each other. Thus, a post-hoc test (Bonferroni) showed significant differences between the three-clusters for most of the PGS except PGS for general cognition. Cluster means of PGS for neuroticism are statistically significant different between clusters with high PGS for neuroticism in cluster 3, average PGS in cluster 2, and low PGS in cluster 1. Cluster means of PGS for subjective well-being are statistically significant different between clusters with high PGS for subjective wellbeing in cluster 1, average PGS in cluster 2, and low to none PGS in cluster 3. Cluster means of PGS for depressive symptoms are statistically significant different between clusters with high PGS for depressive symptoms in cluster 3, and low PGS for clusters 1 and 2. Cluster means of PGS for extraversion are statistically significant different between clusters with high PGS foe extraversion for cluster 1, average PGS for cluster 2, and low to none PGS for cluster 3. Cluster

means of PGS for educational attainment are statistically significant different between clusters with high PGS for educational attainment for cluster 2, and low to none PGS for clusters 1 and 3. Cluster means of PGS for anxiety are statistically significant different between clusters with high PGS for anxiety for cluster 3, average PGS for cluster 1, and low PGS for cluster 2. Finally, cluster means of PGS for general cognition are not statistically significant different between all clusters. Cluster 2 for PGS for general cognition is statistically significant different from cluster 1 and 3. However, PGS for general cognition in cluster 3 and 1 are not significantly different (as we can visually confirmed in bar graph from Figure 9).

Table 17. ANOVA and Post-hoc analysis for Three-Clusters Comparison for PGS

ANOVA for PGS						
Variables		**Sum of Squares**	**Df**	**Mean Square**	**F**	**Sig.**
PGS for Neuroticism	Between Groups	1810.754	2	905.377	1556.715	**0.00**
	Within Groups	2524.121	4340	0.582		
	Total	4334.875	4342			
PGS for Subj. Wellbeing	Between Groups	1090.560	2	545.280	725.319	**0.00**
	Within Groups	3262.721	4340	0.752		
	Total	4353.280	4342			
PGS for Depressive Symptoms	Between Groups	1386.254	2	693.127	1011.667	**0.00**
	Within Groups	2973.478	4340	0.685		
	Total	4359.732	4342			
PGS for Extraversion	Between Groups	627.669	2	313.834	366.138	**0.00**
	Within Groups	3720.017	4340	0.857		
	Total	4347.686	4342			
PGS for Educational Attainment	Between Groups	1534.798	2	767.399	1184.808	**0.00**
	Within Groups	2811.013	4340	0.648		
	Total	4345.811	4342			
PGS for Anxiety	Between Groups	248.989	2	124.494	131.563	**0.00**
	Within Groups	4106.814	4340	0.946		
	Total	4355.803	4342			
PGs for General Cognition	Between Groups	1586.494	2	793.247	1246.769	**0.00**
	Within Groups	2761.290	4340	0.636		
	Total	4347.784	4342			

Table 17 (Continued). ANOVA and Post-hoc analysis for Three-Clusters Comparison for PGS

Post-hoc analysis for Three-Clusters Comparison for PGS							
Bonferroni							
Independent Variable	**Cluster number (I)**	**Cluster number (J)**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
						Lower Bound	**Upper Bound**
PGS for Neuroticism	1	2	-.705817*	0.02905	**0.00000**	-0.77538	-0.63626
		3	-1.54539*	0.02777	**0.00000**	-1.61190	-1.47889
	2	1	.705817*	0.02905	**0.00000**	0.63626	0.77538
		3	-.83957*	0.02844	**0.00000**	-0.90769	-0.77147
	3	1	1.54539*	0.02777	**0.00000**	1.47889	1.61190
		2	.83957*	0.02844	**0.00000**	0.77147	0.90769
PGS for Subj. Wellbeing	1	2	.62991*	0.03302	**0.00000**	0.55083	0.70900
		3	1.20249*	0.03157	**0.00000**	1.12688	1.27810
	2	1	-.62991*	0.03302	**0.00000**	-0.70900	-0.55083
		3	.57257*	0.03233	**0.00000**	0.49515	0.65001
	3	1	-1.20249*	0.03157	**0.00000**	-1.27810	-1.12688
		2	-.57257*	0.03233	**0.00000**	-0.65001	-0.49515
PGS for Depressive Symptoms	1	2	-.38664*	0.03153	**0.00000**	-0.46215	-0.31115
		3	-1.31457*	0.03014	**0.00000**	-1.38675	-1.24239
	2	1	.38664*	0.03153	**0.00000**	0.31115	0.46215
		3	-.92792*	0.03087	**0.00000**	-1.00184	-0.85400
	3	1	1.31457*	0.03014	**0.00000**	1.24239	1.38675
		2	.92792*	0.03087	**0.00000**	0.85400	1.00184
PGS for Extraversion	1	2	.49741*	0.03526	**0.00000**	0.41297	0.58187
		3	.91198*	0.03371	**0.00000**	0.83124	0.99272
	2	1	-.49741*	0.03526	**0.00000**	-0.58187	-0.41297
		3	.41456*	0.03452	**0.00000**	0.33188	0.49724
	3	1	-.91198*	0.03371	**0.00000**	-0.99272	-0.83124
		2	-.41456*	0.03452	**0.00000**	-0.49724	-0.33188
PGS for Educational Attainment	1	2	-1.24391*	0.03065	**0.00000**	-1.31733	-1.17051
		3	.08813*	0.02931	**0.00795**	0.01795	0.15832
	2	1	1.24391*	0.03065	**0.00000**	1.17051	1.31733
		3	1.33205*	0.03001	**0.00000**	1.26018	1.40392
	3	1	-.08813*	0.02931	**0.00795**	-0.15832	-0.01795
		2	-1.33205*	0.03001	**0.00000**	-1.40392	-1.26018
PGS for Anxiety	1	2	.33610*	0.03705	**0.00000**	0.24737	0.42483
		3	-.25201*	0.03542	**0.00000**	-0.33685	-0.16719
	2	1	-.33610*	0.03705	**0.00000**	-0.42483	-0.24737
		3	-.58812*	0.03627	**0.00000**	-0.67499	-0.50125
	3	1	.25201*	0.03542	**0.00000**	0.16719	0.33685
		2	.58812*	0.03627	**0.00000**	0.50125	0.67499
PGS for General Cognition	1	2	-1.31350*	0.03038	**0.00000**	-1.38626	-1.24075
		3	0.00102	0.02904	1.00000	-0.06854	0.07058
	2	1	1.31350*	0.03038	**0.00000**	1.24075	1.38626
		3	1.31452*	0.02974	**0.00000**	1.24329	1.38576
	3	1	-0.00102	0.02904	1.00000	-0.07058	0.06854
		2	-1.31452*	0.02974	**0.00000**	-1.38576	-1.24329

*. The mean difference is significant at the 0.05 level.

Hypothesis 1 is supported. Significantly different clusters were created based on PGS. The study looked at clusters with specific high genetic propensities shown by the PGS. Remember PGS show a high tendency when the numbers are closer to +1 (standardized numbers), average tendency close to 0 and low to no propensity close to -1 (standardized

numbers). Consequently, the resulting clusters were named based on the personality and cognitive trait with the highest genetic propensity:

- **Cluster 1: Cheerful traits**. Cluster was statistically significant different from the rest of the clusters. Cheerful cluster showed high genetic propensity for PGS for well-being and PGS for extraversion. It also displayed low to no genetic propensity for PGS for neuroticism, PGS for depressive symptoms, and PGS for educational attainment.
- **Cluster 2: Cognitive traits**. Cluster was significantly different from the rest of the clusters. Cognitive cluster showed high genetic propensity for PGS for educational attainment and PGS for general cognition. It also displayed low to no genetic propensity for PGS for depressive symptoms and anxiety.
- **Cluster 3: Neurotic traits**. Cluster was significantly different to the other clusters. Neurotic cluster showed high genetic propensity for PGS for neuroticism, PGS for depressive symptoms and PGS for anxiety. On the other hand, it also displayed low to no genetic propensity for PGS for subjective well-being, PGS for extraversion, and PGS for educational attainment.

The above results supported hypothesis 1 as significant different clusters emerged from the combination of PGS for the different personality and cognitive traits showing differences of genetic propensities between clusters and similar propensities within clusters.

4.4 Analysis of Variance for Alcohol Consumption (DV)

ANOVA was conducted to identify statistical differences between clusters in terms of the dependent variable: level of alcohol consumption per week. The ANOVA in Table 18 and Figure 10 showed significant difference of means between clusters for level of drinking. In the descriptive table, the mean per cluster is shown. Cluster 2 presented higher levels of alcohol

consumption per week than the other two clusters; cluster 1 had an average consumption of alcohol; and cluster 3 showed low alcohol consumption. The challenge with an ANOVA is that it doesn't show which specific groups are significantly different. Consequently, a post-hoc study presented in Table 19 demonstrated how cluster 2 and cluster 3 were significantly different but cluster 1 was not significantly different to the other clusters (2 and 3).

Table 18. ANOVA for Dependent Variable in Three-Cluster Solution

Descriptives of Dependent Variable within Three-Cluster Solution

DV: # of Days Drinking per Week

Cluster Number	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
1	1443	**-0.00905**	0.98504	0.02593	-0.05991	0.04182	-0.96459	1.88680
2	1320	**0.02630**	1.00569	0.02768	-0.02800	0.08060	-0.96459	1.88680
3	1580	**-0.07540**	0.95076	0.02392	-0.12232	-0.02848	-0.96459	1.88680
Total	4343	-0.02244	0.97980	0.01487	-0.05159	0.00670	-0.96459	1.88680

ANOVA for Dependent Variable Three-Cluster Comparison

DV: # of Days Drinking per Week

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7.826	2	3.913	4.082	**0.017**
Within Groups	4160.533	4340	0.959		
Total	4168.359	4342			

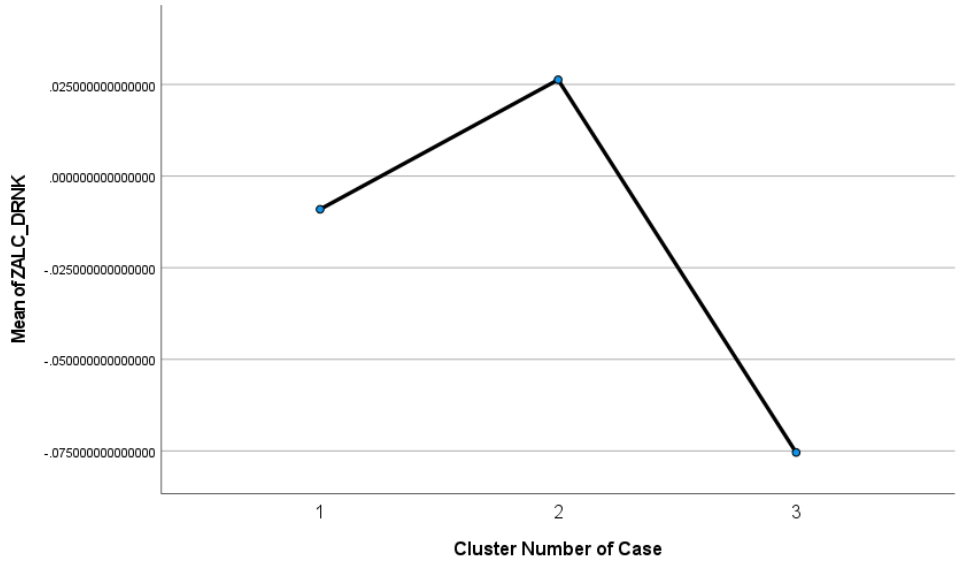


Figure 10. Mean Plot of Alcohol Consumption by Cluster

Table 19. Post-hoc Study for Dependent Variable Three-Cluster Comparison

Post-hoc Test for Dependent Variable in the Three-Cluster Solution						
Dependent Variable: # of Days Drinking per Week						
Bonferroni						
					95% Confidence Interval	
(I) Cluster Number	(J) Cluster Number	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	2	-0.03534	0.03729	1.00000	-0.12465	0.05396
	3	0.06635	0.03565	0.18837	-0.01903	0.15174
2	1	0.03534	0.03729	1.00000	-0.05396	0.12465
	3	.10169*	0.03651	**0.01610**	0.01426	0.18914
3	1	-0.06635	0.03565	0.18837	-0.15174	0.01903
	2	-.10169*	0.03651	**0.01610**	-0.18914	-0.01426
*. The mean difference is significant at the 0.05 level.						

Hypothesis 2 is supported. Based on the results above, three different clusters created are significantly different for levels of alcohol consumption. Cluster 2, the *Cognitive Group*, showed higher levels of alcohol consumption per week meaning that individuals with higher genetic propensities for educational attainment and general cognition consume more alcohol per week than the other groups. On the other side, cluster 3, the *Neurotic Group*, showed low consumption of alcohol per week denoting that individuals with higher genetic propensity for

neuroticism and depressive symptoms drink less alcohol than the other groups. As Table 20 shows, the *Cognitive Group* have almost 15 years of education on average and they are estimated to consume almost 2 and a half drinks per week. In the case of the *Neurotic Group*, individuals have 13.6 years of education on average and are estimated to drink around 2 drinks per week.

Table 20. Final Clusters Characteristics

Variables	**Cluster 1 Cheerful Group**		**Cluster 2 Cognitive Group**		**Cluster 3 Neurotic Group**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
# of Days Drinking per Week	**2.34**	2.41	**2.43**	2.46	**2.18**	2.33
Age in Years	**71.58**	9.51	**71.96**	9.30	**71.08**	9.19
Years of Education	**13.74**	2.27	**14.59**	2.18	**13.64**	2.32
PGS for Neuroticism	-0.81	0.76	-0.10	0.79	0.73	0.73
PGS for Subj. Wellbeing	0.62	0.84	-0.00	0.89	-0.58	0.87
PGS for Depressive Symptoms	-0.64	0.82	-0.26	0.85	0.67	0.81
PGS for Extraversion	0.49	0.91	-0.00	0.92	-0.427	0.95
PGS for Educational Attainment	-0.26	0.80	0.96	0.76	-0.35	0.82
PGS for Anxiety	0.01	0.95	-0.31	0.96	0.26	0.95
PGS for General Cognition	-0.31	0.80	0.99	0.75	-0.32	0.82

The bivariate correlation showed how the control variable education level is correlated to the dependent variable: level of alcohol consumption per week; and how age is not correlated to the dependent variable or education level. Consequently, we conducted an ANCOVA to analyze the level of variance between clusters taking into account these control variables. Table 20 showed how the model and the covariate level of education were statistically significant different between clusters. However, the covariate age in years is not statistically significant different in

the clusters. Shown in Figure 11, as we added age and education level to the mix, cluster 1 increased in alcohol consumption per week above cluster 2. Cluster 2 decreased alcohol consumption and cluster 3 stayed at the same level without covariates.

Table 21. Analysis of Variance with Covariates (ANCOVA)

ANCOVA - Tests of Between-Subjects Effects					
Dependent Variable: # of Days Drinking per Week					
Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	66.866[a]	4	16.716	17.680	**0.000**
Intercept	1.954	1	1.954	2.067	0.151
ZAGE	2.796	1	2.796	2.957	0.086
ZEDUC	56.823	1	56.823	60.099	**0.000**
QCL_1	3.173	2	1.587	1.678	0.187
Error	4101.494	4338	0.945		
Total	4170.547	4343			
Corrected Total	4168.359	4342			
a. R Squared = .016 (Adjusted R Squared = .015)					

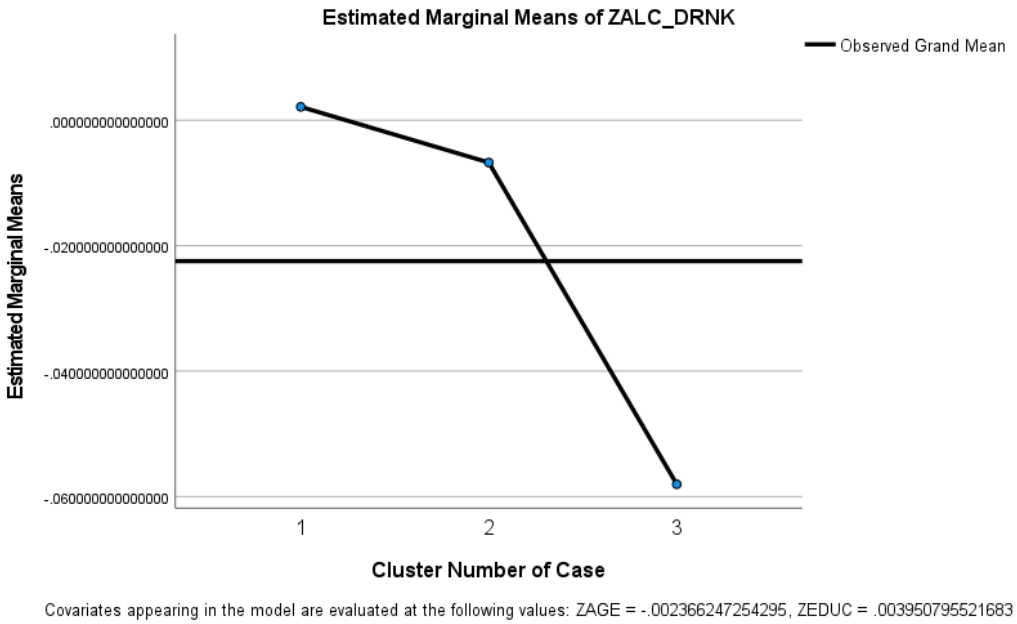


Figure 11. Mean Plot for Alcohol Consumption by Cluster considering Covariates

Overall, the empirical study supports both hypotheses per Table 21.

Table 22. Summary of Hypotheses

Hypotheses	Result
H1: With the combination of polygenic scores, clusters will emerge characterized by different levels of extraversion, neuroticism, educational attainment, general cognition, subjective well-being, depressive symptoms, and anxiety tendencies.	**Supported**
H2: Clusters created will be differentially related to alcohol consumption.	**Supported**

CHAPTER V. DISCUSSION AND CONCLUSION

The study proposes a new segmentation base called Genomics Segmentation using genetic data to create clusters and understand consumer profiles. Chapter 5 provides a discussion of the research findings. First, it presents an overview of the study. Second, it offers a discussion of the hypotheses tested in the research model. Third, it describes the contributions to literature, theory, and marketing practice. Finally, it addresses the limitations, future research suggestions, and final conclusions.

5.1 Overview of the Study

The need to cover consumers' needs and wants more accurately than the competitors has been central to marketing strategies (Dickson and Ginter 1987), thus the overflow of leading researchers presenting the underlying logic and value of segmentation (Dibb 1998; Snellman 2005; Yankelovich and Meer 2006). It is well-known that the first step of a successful customer-driven marketing strategy involves the selection of customers to serve and the identification of the value proposition that best fits those customers (Kotler et al., 2015). However, firms have been facing great challenges as segmentation strategies are founded on traditional bases of segmentation that paint a partial picture and do not have the power to predict future buying behaviors (Haley 1968). Therefore, this study proposed a new segmentation base called Genomics Segmentation, a strategy that relies on genetic data to understanding consumer tendencies and divide the market into homogeneous and stable clusters.

Our genetic makeup explains most of the systematic variation between us and the stability of phenotypes from age to age (Plomin 2019). Literature has clearly shown how genes provide our individuality and can explain half of the variance in our psychological and personality traits (Briley and Tucker-Drob 2014; Plomin 2019; Plomin and von Stumm 2018).

Despite all these findings, literature shows embryonic initiatives within the marketing discipline on the usage of genetic data for the understanding of consumers. Consequently, the purpose of this study was to identify the best way to leverage genetic data, through PGS, for the creation of clusters of the marketplace. The first objective was to present a review of the literature within behavioral genetics and strategic segmentation and identify the gaps. The second objective was to empirically test the utilization of PGS to create significant different clusters of the market and understand consumer tendencies of those subgroups. The chapter continues with a discussion of the findings.

5.2 Research Findings

The current study presented a theoretical framework and empirical applicability of the new genomics segmentation base. The dissertation evaluated two hypotheses related to the creation of significantly different clusters using PGS for personality and cognitive traits, and the understanding of alcohol consumption patterns in those clusters. The literature supported the concept of using PGS to identify individuals' early tendencies for behavior and traits; and the empirical study showed how significant clusters can be created from those PGS understanding systematic variance early in life and early tendencies in the consumer journey.

Hypothesis 1 proposed that the combination of PGS create clusters with significantly different levels of extraversion, neuroticism, educational attainment, general cognition, subjective well-being, depressive symptoms, and anxiety. The study supported this hypothesis. With the great progress on molecular genetics, the explanatory power of PGS has and will continue to increase (Plomin and von Stumm 2022). PGS represent a door-opener for the expansion of marketing personalization and hyper-segmentation (Ivanova-Kadiri 2023). Findings suggested that PGS for personality and cognitive traits behave in line with the traditional

marketing and psychology literature about personality and cognitive traits. From a trait perspective, literature shows that personality and cognitive traits bring a framework to understand the development and operation of psychological mechanism, behaviors, and experiences (Costa and McCrae 1999; Hecjman, Stixrud, and Urzua 2006). For example, it is known that extroverted individuals have positive emotions including high level of subjective well-being (DeNeve and Cooper 1998) which it is shown in the *Cheerful Group 1* where high PGS for extraversion and subjective well-being have been grouped together to create this cluster. Consequently, we assume that the *Cheerful Group 1* follows findings related to the tendency of extraverted and high in subjective well-being individuals that commonly experience positive reactivity and consumption emotions to positive-mood messages and stimulus (Larsen and Ketelaar 1991; Matzler et al. 2005). Also, there is considerable insight indicating that these kinds of individuals have better life outcomes with a more emotional stability and positive outlook of life (Widiger 2016).

In the same line, literature shows how neurotic traits relate to elevated stress, anxiety, and depression tendencies (Barlow et al. 2014; Kotov et al. 2010). The study resulted in a *Neurotic Group 3* with high PGS for neuroticism clustered together with high PGS for depression and PGS for anxiety. Thus, it is assumed that *Neurotic Group 3* follows the characteristics of a neurotic personality structure with high reactivity to negative mood stimulus, and lack of emotional stability (Matzler et al. 2005; Widiger 2016).

A very similar pattern is shown in the *Cognitive Group 2* where high PGS for education attainment and PGS for general cognition were clustered together following the literature that presents a high correlation between years of schooling and cognitive abilities and performance (Davies et a. 2016). Also, results showed how the *Cognitive Group 2* with high cognitive

functions has low levels of anxiety and depressive symptoms following the literature about cognition inversely related to anxiety traits (Jaiswal et al. 2018).

Furthermore, hypothesis 2 proposed that those clusters created from PGS show significant differences on the level of alcohol consumption per week. The study supported the hypothesis. Literature showed a high correlation between education level, sociability, and level of alcohol consumption. Thus, the *Cognitive Group 2* showed the highest level of alcohol consumption per week, in line with literature findings related to education and moderate consumption of alcohol (Makela 1999). Study found that the *Cognitive Group* 2 has an average of almost 15 years of education and consume an average of 2.43 drinks of alcohol per week. In the case of the *Cheerful Group 1*, it also followed the literature as extroverted individuals have an increased probability of risky alcohol consumption (Hakulinen et al. 215). On the other side, results from the *Neurotic Group 3* are not consistent with the literature. It is understood that neurotic and depressive individuals are highly associated and most vulnerable to high levels of alcohol consumption and alcoholism overall (Hakulinen and Jokela 2018; Mulder 2002). However, there is one literature that highlights how high levels of neuroticism can also be beneficial for anticipating negative outcomes and possible risk (Widiger 2016). Consequently, following those findings the *Neurotic Group 3* showed the lowest tendency for alcohol consumption. The study found that the *Neurotic Group 3* has an average of alcohol consumption of 2 drinks per week. The overall findings demonstrated that clusters created from PGS can provide marketers useful and impactful information on consumption tendencies and behaviors. The study used alcohol consumption levels as an example and initial application of the proposed genomics segment base. Central to the study's finding is the possibility of segmenting consumers early in life using genetic data.

Consequently, the study demonstrated the theoretical and empirical applicability of genomics segmentation base using genetic data for the segmentation of markets. On one hand, the study asserted that PGS for personality and cognitive traits follow the consumer psychology literature related to the Big Five personality model and cognitive factors for consumption patterns and behaviors. Clusters created showed similar patterns as personality and cognition literature (Barlow et al. 2014; Costa and McCrae 1999; Jaiswal et al. 2018; Kotov et al. 2010) that highlight how these positive and negative personality traits and cognitive factors affect consumption, impulse buying and advertising emotions (Larsen and Katelaar 1991; Matzler et al. 2005; Wang et al. 2010). On the other hand, possibly due to our sampling being mainly seniors, the alcohol consumption levels differed from the literature specifically on levels of consumption for neurotic and depressive traits. It is widely known that alcohol affects cognitive processing through response inhibition (Curtin et al. 2001); as well as it positively correlates with sociability, extraversion, and subjective well-being (Cook et al. 1998; Geiger and MacKerron 2016). Moreover, literature also showed that people who do not drink are little more withdrawn and less ambitious compared to those who drink in moderation (Cook et al. 1998). Thus, results can be interpreted following the literature that individuals high in extraversion and subjective well-being have higher tendencies of alcohol consumption. On the opposite end, not following the main literature related to neuroticism and negative traits such as anxiety and depression, results showed that these individuals have no tendencies for alcohol consumption maybe following the one literature found that revealed how neurotic tendencies can help individuals identified risky situations (Widiger 2016).

5.3 Contributions

This dissertation introduced a marketing theory by proposing a new genomics segmentation base. It addressed the call for the development of indigenous marketing theory, as well as the use of that theory for empirical research (Rust 2006; Hunt 2020). Marketers have relied on theories from other disciplines, instead "we must develop ideas, concepts and theories whose central focus is the study of marketing in its natural environment" (Zeithaml et al. 2020, p. 49). "Ideally, theories go beyond post hoc interpretation" (p.169) of what experts and academics observe, theories should drive research to present testable hypotheses (Costa and McCrae 1999). Consequently, with the proposal of a new genomics segmentation base, the study has leveraged innovation and presented a new concept within marketing segmentation. Moreover, it has proposed a construct and demonstrated its empirical applicability expanding existing literature and over time and with additional research could become a generalizable strategy for genomics segmentation.

"Publications must seek to ultimately impact both the external audience and also advance the published academic literature" (Shugan 2003, p.3). From a theoretical perspective, this study expanded on the embryonic marketing literature about genomics marketing. The findings added to the research by Simonson and Sela (2011) embracing the genetic dimension in the decision-making process of consumers. Findings showed the importance of molecular genetics in the understanding of consumer tendencies and behavior. Furthermore, it also expanded the research by Daviet, Nave, and Wind (2022) incorporating genetic variables into strategic marketing and their potential applications. The study built on the usage of genomics to infer other segmentation bases and added genetics to predictive empirical models.

The study empirically developed and tested the genomics segmentation base. For practitioners, this research is a door-opener for the applicability of genetic data in marketing segmentation. It provides a roadmap specifically leveraging the power of explained heritability of PGS to enhance the understanding of consumer behavioral tendencies, as well as the segmentation of markets. Firms can potentially use PGS to identify needs and wants early in life and improve brand connection and satisfaction with potential consumers. On the other hand, organizations and government entities can better understand negative behaviors early in the consumer journey with the potential of diminishing those behaviors that harm health and social aspects of communities. For example, non-profit organization and socially conscious brands can look to invest in more value-added content to impact extroverted and high in subjective well-being individuals showing the benefits of drinking non-alcoholic beverages and the threats of drinking alcoholic beverages.

Overall, leveraging genetic data in marketing is going to be revolutionary. It will influence the marketing discipline in numerous ways from academic researchers, educators, government entities, regulators to marketing practitioners and firms. Marketers are able to identify future consumers early enough and before they show the actual behavior or need (Daviet, Nave and Wind 2022). Therefore, managers and researchers are able to estimate individual tendencies without needing to rely on consumer surveys and past behaviors.

5.3.1 Ethical implications

Research in marketing should strive to provide insights to better understand the potential of marketing actions and improve marketing practices (Lehmann, McAlister, and Staelin 2011). Technology and socioeconomic trends have completely and will continue to transform to a truly customer-centered marketing (Rust 2020). However, this transformation presents great

opportunities and existential threats that need to be addressed by marketing academics, regulators, and practitioners. As we have seen, genetic data can provide insight for early intervention when there are high tendencies of negative behaviors. However, it might also facilitate discrimination and the exploitation of vulnerabilities of consumers (Daviet, Nave, and Wind 2022). Thus, marketers need to support the anonymity and preservation of the information complying with all the latest ethical guidelines of data identification and manipulation. Moreover, there is a need for industry standards to guarantee privacy and security of consumer data, and transparency of the companies' processes and underlying scientific claims.

5.4 Limitations and Future Research

Despite the advantages of genomics segmentation in the evolution of marketing strategy, this new segmentation base is prone to several limitations. First, genomics segmentation uses secondary data that can introduce potential limitations due to missing data, unmeasured variables, and data quality because data has not been collected particularly for the research question (Cheng and Phillips 2014). Most of the genetic data available has been gathered with the purpose of expanding on complex human health and precision medicine research. Thus, there are many missing data and variables related to individuals' consumption, traits, and demographic factors. Some of the most common databases for the research community are the Health and Retirement Study (HRS) from the University of Michigan, the All of Us Research Hub and the UK BioBank. As the study uses two datasets from HRS, the sample is concentrated on the senior American population with a median of 71 years old. Moreover, the income level data had a high number of missing values, as well as presented high levels of skewness and kurtosis. If removing missing values and outliers, we would have ended up with less than 30% of the data. The above

details showed examples of the limitations to access genetic and demographic data for a more representative sample.

Second, research used cross-sectional data which only shows a one-time picture of the studied relationship. Cross-sectional studies fail to address the variation of relationships over time and have high potential of biased inferences (Bowen and Wiersema 1999). Despite knowing that genetic data is stable (Edwards et al. 2007; Plomin et al 2016), it is important to expand and confirm findings through longitudinal data. Third, genomics segmentation uses PGS as the main attributes for segmentation, and those PGS are statistical calculations in constant update. With the publication of bigger and better GWAS, more powerful PGS are created (Plomin and von Stumm 2021). Consequently, genomics segmentation is an initial proposal to incorporate genetics in marketing segmentation and will continue to evolve as behavioral scientists increase the predictive power of PGS. Fourth, genomics segmentation implies an expensive method for marketing practitioners. Access to genetic data "is central to the revolution taking place in molecular genetics" (Chee et al. 1996). Despite efforts that have decreased cost and time for accessing genetic data and PGS specifically, genomics segmentation involves a high-cost investment in data and technical expertise. Fifth, genomics segmentation proposed the use of cluster analysis: a technique widely applied for segmentation in marketing (Dibbs 1998) but hard to determine its statistical reliability (Beane and Ennis 1987). There are no properly established ways to compute statistical power for cluster analysis (Dalmaijer, Nord and Astle 2022). However, the study used the elbow criterion to determine cluster numbers for the k-means, as well as performed a robustness check using the total sum of squares by cluster to confirm the cluster numbers. Moreover, it has been found that cluster analysis is superior to factor analysis for marketing segmentation (Beane and Ennis 1987).

In addition to considering future research related to the above detailed limitations, future research can examine the applicability of genomics segmentation to other consumption and behavioral patterns such as hedonic or utilitarian preferences, buying a car, vacation interests, social versus antisocial activities, etc. Additional research can compare PGS for the Big Five personalities with the traditional literature of self-reported data for the Big Five. In general, genomics segmentation can be applied to different industries and diverse interests of products and services to expand on the literature and confirm its applicability.

Finally, despite the reduced sample with the income level data and any other demographic data, future research could be considered as a post-hoc study to understand genomics segmentation.

5.5 Conclusions

Genomics segmentation serves as a foundation to extend research on the applicability of molecular genetics in marketing, as well as foster incremental research in the transformation of genomics marketing. Genetic data has been singled out from the marketing literature. Genomics segmentation shows the potential impact on academic research and practitioners by incorporating fundamental principles of behavioral genetics in consumer marketing theory and strategic segmentation. By finding statistically significant clusters created from different PGS for personality and cognitive traits, the study suggests that genomics segmentation is a feasible empirical model for the segmentation of markets. Further research must examine additional examples and factors to apply genetic data in the development of strategic marketing.

REFERENCES

Alcoforado, Daniela Gomes, Francisco Vicente Sales Melo, and Renata Gomes Alcoforado. 2022. 'Depression and consumption habits: a cross-cultural study', *Journal of Consumer Marketing*, 39: 621-31.

Allport, Floyd H, and Gordon W Allport. 1921. 'Personality traits: Their classification and measurement', *The Journal of Abnormal Psychology and Social Psychology*, 16: 6.

Asch, Solomon E. 1957. "An experimental investigation of group influence." In *Symposium on preventive and social psychiatry*, 15-17. Walter Reed Army Institute of Research Washington, DC.

Azucar, Danny, Davide Marengo, and Michele Settanni. 2018. 'Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis', *Personality and individual differences*, 124: 150-59.

Babor, Thomas. 2010. 'Alcohol: no ordinary commodity: research and public policy'.

Bansal, Manju. 2003. 'DNA structure: Revisiting the Watson–Crick double helix', *Current Science*: 1556-63.

Barlow, David H, Kristen K Ellard, Shannon Sauer-Zavala, Jacqueline R Bullis, and Jenna R Carl. 2014. 'The origins of neuroticism', *Perspectives on psychological science*, 9: 481-96.

Barnhofer, Thorsten, Danielle S. Duggan, and James W. Griffith. 2011. 'Dispositional mindfulness moderates the relation between neuroticism and depressive symptoms', *Personality and individual differences*, 51: 958-62.

Bart, Yakov, Venkatesh Shankar, Fareena Sultan, and Glen L Urban. 2005. 'Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study', *Journal of marketing*, 69: 133-52.

Baumgartner, Hans. 2002. 'Toward a personology of the consumer', *Journal of Consumer Research*, 29: 286-92.

Beane, TP, and DM Ennis. 1987. 'Market segmentation: a review', *European journal of marketing*, 21: 20-42.

Benyamin, Beben, BSt Pourcain, Oliver S Davis, Gail Davies, Narelle K Hansell, M-JA Brion, RM Kirkpatrick, Rolieke AM Cents, Sanja Franić, and MB Miller. 2014. 'Childhood intelligence is heritable, highly polygenic and associated with FNBP1L', *Molecular psychiatry*, 19: 253-58.

Berger, Bonnie, Noah M. Daniels, and Y. William Yu. 2016. 'Computational biology in the 21st century: scaling with compressive algorithms', *Commun. ACM*, 59: 72–80.

Bholowalia, Purnima, and Arvind Kumar. 2014. 'EBK-means: A clustering technique based on elbow method and k-means in WSN', *International Journal of Computer Applications*, 105.

Blum, Kenneth, Ernest P Noble, Peter J Sheridan, Anne Montgomery, Terry Ritchie, Pudur Jagadeeswaran, Harou Nogami, Arthur H Briggs, and Jay B Cohn. 1990. 'Allelic association of human dopamine D2 receptor gene in alcoholism', *Jama*, 263: 2055-60.

Borry, Pascal, Heidi Beate Bentzen, Isabelle Budin-Ljøsne, Martina C Cornel, Heidi Carmen Howard, Oliver Feeney, Leigh Jackson, Deborah Mascalzoni, Álvaro Mendes, and Borut Peterlin. 2018. 'The challenges of the expanded availability of genomic information: an agenda-setting paper', *Journal of community genetics*, 9: 103-16.

Bouchard Jr, Thomas J. 1994. 'Genes, environment, and personality', *Science*, 264: 1700-01.

Bouchard Jr, Thomas J, and Matthew McGue. 1981. 'Familial studies of intelligence: A review', *Science*, 212: 1055-59.

Bouchard, Thomas J., David T. Lykken, Matthew McGue, Nancy L. Segal, and Auke Tellegen. 1990. 'Sources of Human Psychological Differences: The Minnesota Study of Twins Reared Apart', *Science*, 250: 223-28.

Boyle, Patrick. 2020. "Microbes and Manufacturing: Moore's Law Meets Biology." In *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2019 Symposium*, 45-54. National Academies Press.

Briley, Daniel A, and Elliot M Tucker-Drob. 2014. 'Genetic and environmental continuity in personality

development: a meta-analysis', *Psychological bulletin*, 140: 1303.
Brunello, Giorgio, and Martin Schlotter. 2011. 'Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems'.
Bryc, Katarzyna, Eric Y Durand, J. Michael Macpherson, David Reich, and Joanna L Mountain. 2015. 'The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States', *The American Journal of Human Genetics*, 96: 37-53.
Burke, A. 2012. "DNA Sequencing Is Now Improving Faster Than Moore's Law! Forbes." In.
Businessweek, Bloomberg. 2011. *Steve Jobs: 1955-2011* (Bloomberg Businessweek).
Cacioppo, John T, and Richard E Petty. 1982. 'The need for cognition', *Journal of personality and social psychology*, 42: 116.
Cervone, Daniel, and Lawrence A Pervin. 2022. *Personality: Theory and research* (John Wiley & Sons).
Cetola, Hank, and Kathleen Prinkey. 1986. 'Introversion-extraversion and loud commercials', *Psychology & Marketing*, 3: 123-32.
Chabris, Christopher F, James J Lee, David Cesarini, Daniel J Benjamin, and David I Laibson. 2015. 'The fourth law of behavior genetics', *Current directions in psychological science*, 24: 304-12.
Chee, Mark, Robert Yang, Earl Hubbell, Anthony Berno, Xiaohua C Huang, David Stern, Jim Winkler, David J Lockhart, Macdonald S Morris, and Stephen PA Fodor. 1996. 'Accessing genetic information with high-density DNA arrays', *Science*, 274: 610-14.
Cheng, Hui G, and Michael R Phillips. 2014. 'Secondary analysis of existing data: opportunities and implementation', *Shanghai archives of psychiatry*, 26: 371.
Chipuer, HM, Robert Plomin, Nancy L Pedersen, Gerald E McClearn, and John R Nesselroade. 1993. 'Genetic influence on family environment: The role of personality', *Developmental psychology*, 29: 110.
Choi, Shing Wan, Timothy Shin-Heng Mak, and Paul F O'Reilly. 2020. 'Tutorial: a guide to performing polygenic risk score analyses', *Nature protocols*, 15: 2759-72.
Cohen, Joel B, Michel Tuan Pham, and Eduardo B Andrade. 2018. 'The nature and role of affect in consumer behavior.' in, *Handbook of consumer psychology* (Routledge).
Cook, Mark, Alison Young, Dean Taylor, and Anthony P. Bedford. 1998. 'Personality correlates of alcohol consumption', *Personality and individual differences*, 24: 641-47.
Corder, EH, Al M Saunders, NJ Risch, WJ Strittmatter, DE Schmechel, PC Gaskell Jr, JB Rimmler, PA Locke, PM Conneally, and KE Schmader. 1994. 'Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease', *Nature genetics*, 7: 180-84.
Cosmas, Stephen C. 1982. 'Life styles and consumption patterns', *Journal of Consumer Research*, 8: 453-55.
Costa, Paul T, and Robert R McCrae. 1999. 'A five-factor theory of personality', *The five-factor model of personality: Theoretical perspectives*, 2: 51-87.
Curtin, John J, Christopher J Patrick, Alan R Lang, John T Cacioppo, and Niels Birbaumer. 2001. 'Alcohol affects emotion through cognition', *Psychological Science*, 12: 527-31.
Dalmaijer, Edwin S, Camilla L Nord, and Duncan E Astle. 2022. 'Statistical power for cluster analysis', *BMC bioinformatics*, 23: 1-28.
Darden, William R, and Fred D Reynolds. 1971. "Shopping orientations and product usage rates." In.: SAGE Publications Sage CA: Los Angeles, CA.
Darden, William R, and Fred D Reynolds. 1974. 'Backward profiling of male innovators', *Journal of marketing research*, 11: 79-85.
Dausset, Jean, Howard Cann, Daniel Cohen, Mark Lathrop, Jean-Marc Lalouel, and Ray White. 1990. 'Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome', *Genomics*, 6: 575-77.
d'Aveni, Richard A. 1995. 'Coping with hypercompetition: Utilizing the new 7S's framework', *Academy of Management Perspectives*, 9: 45-57.
Davidson, Trent, Justin Vinneau-Palarino, Joshua A Goode, and Jason D Boardman. 2021. 'Utilizing genome wide data to highlight the social behavioral pathways to health: the case of obesity and

cardiovascular health among older adults', *Social Science & Medicine*, 273: 113766.

Davies, Gail, Nicola Armstrong, Joshua C Bis, Jan Bressler, Vincent Chouraki, Sudheer Giddaluru, Edith Hofer, Carla A Ibrahim-Verbaas, Mirna Kirin, and J Lahti. 2015. 'Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N= 53 949)', *Molecular psychiatry*, 20: 183-92.

Davies, Gail, Max Lam, Sarah E. Harris, Joey W. Trampush, Michelle Luciano, W. David Hill, Saskia P. Hagenaars, Stuart J. Ritchie, Riccardo E. Marioni, Chloe Fawns-Ritchie, David C. M. Liewald, Judith A. Okely, Ari V. Ahola-Olli, Catriona L. K. Barnes, Lars Bertram, Joshua C. Bis, Katherine E. Burdick, Andrea Christoforou, Pamela DeRosse, Srdjan Djurovic, Thomas Espeseth, Stella Giakoumaki, Sudheer Giddaluru, Daniel E. Gustavson, Caroline Hayward, Edith Hofer, M. Arfan Ikram, Robert Karlsson, Emma Knowles, Jari Lahti, Markus Leber, Shuo Li, Karen A. Mather, Ingrid Melle, Derek Morris, Christopher Oldmeadow, Teemu Palviainen, Antony Payton, Raha Pazoki, Katja Petrovic, Chandra A. Reynolds, Muralidharan Sargurupremraj, Markus Scholz, Jennifer A. Smith, Albert V. Smith, Natalie Terzikhan, Anbupalam Thalamuthu, Stella Trompet, Sven J. van der Lee, Erin B. Ware, B. Gwen Windham, Margaret J. Wright, Jingyun Yang, Jin Yu, David Ames, Najaf Amin, Philippe Amouyel, Ole A. Andreassen, Nicola J. Armstrong, Amelia A. Assareh, John R. Attia, Deborah Attix, Dimitrios Avramopoulos, David A. Bennett, Anne C. Böhmer, Patricia A. Boyle, Henry Brodaty, Harry Campbell, Tyrone D. Cannon, Elizabeth T. Cirulli, Eliza Congdon, Emily Drabant Conley, Janie Corley, Simon R. Cox, Anders M. Dale, Abbas Dehghan, Danielle Dick, Dwight Dickinson, Johan G. Eriksson, Evangelos Evangelou, Jessica D. Faul, Ian Ford, Nelson A. Freimer, He Gao, Ina Giegling, Nathan A. Gillespie, Scott D. Gordon, Rebecca F. Gottesman, Michael E. Griswold, Vilmundur Gudnason, Tamara B. Harris, Annette M. Hartmann, Alex Hatzimanolis, Gerardo Heiss, Elizabeth G. Holliday, Peter K. Joshi, Mika Kähönen, Sharon L. R. Kardia, Ida Karlsson, Luca Kleineidam, David S. Knopman, Nicole A. Kochan, Bettina Konte, John B. Kwok, Stephanie Le Hellard, Teresa Lee, Terho Lehtimäki, Shu-Chen Li, Christina M. Lill, Tian Liu, Marisa Koini, Edythe London, Will T. Longstreth, Oscar L. Lopez, Anu Loukola, Tobias Luck, Astri J. Lundervold, Anders Lundquist, Leo-Pekka Lyytikäinen, Nicholas G. Martin, Grant W. Montgomery, Alison D. Murray, Anna C. Need, Raymond Noordam, Lars Nyberg, William Ollier, Goran Papenberg, Alison Pattie, Ozren Polasek, Russell A. Poldrack, Bruce M. Psaty, Simone Reppermund, Steffi G. Riedel-Heller, Richard J. Rose, Jerome I. Rotter, Panos Roussos, Suvi P. Rovio, Yasaman Saba, Fred W. Sabb, Perminder S. Sachdev, Claudia L. Satizabal, Matthias Schmid, Rodney J. Scott, Matthew A. Scult, Jeannette Simino, P. Eline Slagboom, Nikolaos Smyrnis, Aïcha Soumaré, Nikos C. Stefanis, David J. Stott, Richard E. Straub, Kjetil Sundet, Adele M. Taylor, Kent D. Taylor, Ioanna Tzoulaki, Christophe Tzourio, André Uitterlinden, Veronique Vitart, Aristotle N. Voineskos, Jaakko Kaprio, Michael Wagner, Holger Wagner, Leonie Weinhold, K. Hoyan Wen, Elisabeth Widen, Qiong Yang, Wei Zhao, Hieab H. H. Adams, Dan E. Arking, Robert M. Bilder, Panos Bitsios, Eric Boerwinkle, Ornit Chiba-Falek, Aiden Corvin, Philip L. De Jager, Stéphanie Debette, Gary Donohoe, Paul Elliott, Annette L. Fitzpatrick, Michael Gill, David C. Glahn, Sara Hägg, Narelle K. Hansell, Ahmad R. Hariri, M. Kamran Ikram, J. Wouter Jukema, Eero Vuoksimaa, Matthew C. Keller, William S. Kremen, Lenore Launer, Ulman Lindenberger, Aarno Palotie, Nancy L. Pedersen, Neil Pendleton, David J. Porteous, Katri Räikkönen, Olli T. Raitakari, Alfredo Ramirez, Ivar Reinvang, Igor Rudan, Rujescu Dan, Reinhold Schmidt, Helena Schmidt, Peter W. Schofield, Peter R. Schofield, John M. Starr, Vidar M. Steen, Julian N. Trollor, Steven T. Turner, Cornelia M. Van Duijn, Arno Villringer, Daniel R. Weinberger, David R. Weir, James F. Wilson, Anil Malhotra, Andrew M. McIntosh, Catharine R. Gale, Sudha Seshadri, Thomas H. Mosley, Jan Bressler, Todd Lencz, and Ian J. Deary. 2018. 'Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function', *Nature Communications*, 9: 2098.

Daviet, Remi, Gideon Nave, and Jerry Wind. 2022. 'Genetic data: potential uses and misuses in marketing', *Journal of marketing*, 86: 7-26.

De Moor, Marleen HM, Stéphanie M Van Den Berg, Karin JH Verweij, Robert F Krueger, Michelle Luciano, Alejandro Arias Vasquez, Lindsay K Matteson, Jaime Derringer, Tõnu Esko, and Najaf Amin. 2015. 'Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder', *JAMA Psychiatry*, 72: 642-50.

DeNeve, Kristina M, and Harris Cooper. 1998. 'The happy personality: a meta-analysis of 137 personality traits and subjective well-being', *Psychological bulletin*, 124: 197.

Diamantopoulos, Adamantios, Bodo B Schlegelmilch, Rudolf R Sinkovics, and Greg M Bohlen. 2003. 'Can socio-demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation', *Journal of business research*, 56: 465-80.

Dibb, Sally. 1998. 'Market segmentation: strategies for success', *Marketing Intelligence & Planning*, 16: 394-406.

Dickson, Peter R. 1982. 'Person-situation: Segmentation's missing link', *Journal of marketing*, 46: 56-64.

Dickson, Peter R, and James L Ginter. 1987. 'Market segmentation, product differentiation, and marketing strategy', *Journal of marketing*, 51: 1-10.

Diener, Ed, Shigehiro Oishi, and Louis Tay. 2018. 'Advances in subjective well-being research', *Nature Human Behaviour*, 2: 253-60.

Donis-Keller, Helen, Philip Green, Cynthia Helms, Samuel Cartinhour, Barbara Weiffenbach, Karen Stephens, Tim P Keith, Donald W Bowden, Douglas R Smith, and Eric S Lander. 1987. 'A genetic linkage map of the human genome', *Cell*, 51: 319-37.

Dudbridge, Frank. 2013. 'Power and predictive accuracy of polygenic risk scores', *PLoS genetics*, 9: e1003348.

Duong, Cong Doanh. 2022. 'Big Five personality traits and green consumption: bridging the attitude-intention-behavior gap', *Asia Pacific Journal of Marketing and Logistics*, 34: 1123-44.

Eaves, Lindon, and Hans Eysenck. 1975. 'The nature of extraversion: a genetical analysis', *Journal of personality and social psychology*, 32: 102.

Eaves, Lindon, and Hans Eysenck. 1976. 'Genotype× age interaction for neuroticism', *Behavior genetics*, 6: 359-62.

Eaves, LJ, Hans Jurgen Eysenck, NG Martin, R Jardine, AC Heath, L Feingold, PA Young, and KS Kendler. 1989. *Genes, Culture and Personality: An Empirical Approach* (Academic Press. Harcourt Brace Jovanovich, Publishers).

Eaves, Lindon J, and Hans J Eysenck. 1977. 'A genotype-environmental model for psychoticism', *Advances in Behaviour Research and Therapy*, 1: 5-26.

Edenberg, Howard J., and Tatiana Foroud. 2013. 'Genetics and alcoholism', *Nature Reviews Gastroenterology & Hepatology*, 10: 487-94.

Edwards, David, John W. Forster, David Chagné, and Jacqueline Batley. 2007. 'What Are SNPs?' in Nnadozie C. Oraguzie, Erik H. A. Rikkerink, Susan E. Gardiner and H. Nihal De Silva (eds.), *Association Mapping in Plants* (Springer New York: New York, NY).

Endler, Norman S., and Nancy L. Kocovski. 2001. 'State and trait anxiety revisited', *Journal of Anxiety Disorders*, 15: 231-45.

Erevelles, Sunil, Nobuyuki Fukawa, and Linda Swayne. 2016. 'Big Data consumer analytics and the transformation of marketing', *Journal of business research*, 69: 897-904.

Erevelles, S, E Rolland, C Huntley, C Chih-Fen, and V Horton. 1996. "Geographic Information Systems and Marketing Education: Issues, Challenges & Solutions." In.: The A. Gary Anderson Graduate School of Management. University of California ….

Eysenck, HJ. 1976. 'Manual of the EPQ (Eysenck Personality Inventory)', *San Diego: Educational and industrial testing service*.

Fan, Shaokun, Raymond YK Lau, and J Leon Zhao. 2015. 'Demystifying big data analytics for business intelligence through the lens of marketing mix', *Big Data Research*, 2: 28-32.

Finn, Peter R, and Julie Hall. 2004. 'Cognitive ability and risk for alcoholism: short-term memory capacity and intelligence moderate personality risk for alcohol problems', *Journal of Abnormal Psychology*, 113: 569.

Firat, A Fuat, and Clifford J Shultz. 1997. 'From segmentation to fragmentation: markets and marketing strategy in the postmodern era', *European journal of marketing*, 31: 183-207.

Floderus-Myrhed, Birgitta, Nancy Pedersen, and Ingrid Rasmuson. 1980. 'Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: A study of 12,898 twin pairs', *Behavior genetics*, 10: 153-62.

Fölling, Asbjörn. 1934. 'Über Ausscheidung von Phenylbrenztraubensäure in den Harn als Stoffwechselanomalie in Verbindung mit Imbezillität'.

Foxall, Gordon R. 1995. 'Cognitive styles of consumer initiators', *Technovation*, 15: 269-88.

Galton, Francis. 1875. 'THE HISTORY OF TWINS, AS A CRITERION OF THE RELATIVE POWERS OF NATURE AND NURTURE', *Fraser's Magazine*, 12 566.

Galton, Francis. 1891. *Hereditary Genius. An Inquiry into Its Laws and Consequences* (D. Appleton and Company: New York).

Garcia, Justin R, James MacKillop, Edward L Aller, Ann M Merriwether, David Sloan Wilson, and J Koji Lum. 2010. 'Associations between dopamine D4 receptor gene variation with both infidelity and sexual promiscuity', *PloS one*, 5: e14162.

Geiger, Ben Baumberg, and George MacKerron. 2016. 'Can alcohol make you happy? A subjective wellbeing approach', *Social Science & Medicine*, 156: 184-91.

Gerber, Alan S, Gregory A Huber, David Doherty, and Conor M Dowling. 2012. 'Disagreement and the avoidance of political discussion: Aggregate relationships and differences across personality traits', *American Journal of Political Science*, 56: 849-74.

Gerber, Alan S, Gregory A Huber, David Doherty, Conor M Dowling, Connor Raso, and Shang E Ha. 2011. 'Personality traits and participation in political processes', *The Journal of Politics*, 73: 692-706.

Gibson, Rowan. 2015. *The four lenses of innovation: a power tool for creative thinking* (John Wiley & Sons).

Goyat, Sulekha. 2011. 'The basis of market segmentation: a critical review of literature', *European Journal of Business and Management*, 3: 45-54.

Gray, Ian C, David A Campbell, and Nigel K Spurr. 2000. 'Single nucleotide polymorphisms as tools in human genetics', *Human molecular genetics*, 9: 2403-08.

Green, Paul E. 1977. 'A new approach to market segmentation', *Business Horizons*, 20: 61-73.

Gunter, Barrie. 2016. *The psychology of consumer profiling in a digital age* (Routledge).

Gunter, Barrie, and Adrian Furnham. 1992. *Consumer profiles: An introduction to psychographics* (Routledge: London).

Gunter, Chris. 2023. 'Single Nucleotide Polymorphism (SNPS)', NIH, Accessed 9/27/2023. https://www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms.

Hair Jr, Joseph F; William, Black; Babin, Barry J.; Anderson, Rolph E. 2019. 'Multivariate Data Analysis Joseph F. Hair Jr. William C. Black Barry J. Babin Rolph E. Anderson Eighth Edition'.

Hakulinen, Christian, Marko Elovainio, G David Batty, Marianna Virtanen, Mika Kivimäki, and Markus Jokela. 2015. 'Personality and alcohol consumption: Pooled analysis of 72,949 adults from eight cohort studies', *Drug and alcohol dependence*, 151: 110-14.

Hakulinen, Christian, and Markus Jokela. 2018. 'Chapter 3 - Personality as Determinant of Smoking, Alcohol Consumption, Physical Activity, and Diet Preferences.' in Christoffer Johansen (ed.), *Personality and Disease* (Academic Press: San Diego).

Haley, Russell I. 1968. 'Benefit segmentation: A decision-oriented research tool', *Journal of marketing*, 32: 30-35.

Haley, Russell I. 1984. 'Benefit segmentation–20 years later', *Journal of Consumer Marketing*.

Hansen, Flemming. 1969. 'Consumer choice behavior: An experimental approach', *Journal of marketing research*, 6: 436-43.

Haugtvedt, Curtis P, and Richard E Petty. 1992. 'Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes', *Journal of personality and social psychology*, 63: 308.

Haugtvedt, Curtis P, Richard E Petty, and John T Cacioppo. 1992. 'Need for cognition and advertising: Understanding the role of personality variables in consumer behavior', *Journal of Consumer Psychology*, 1: 239-60.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor economics*, 24: 411-82.

Henry, Tanya Albert. 2021. 'Protect sensitive individual data at risk from DTC genetic tests', *American Medical Association. Retrieved February*, 13: 2022.

Herrman, John. 2018. 'Content for Humans About the Content of Humans', *The New York times*.

Hoffman, Donna L., C. Page Moreau, Stefan Stremersch, and Michel Wedel. 2021. 'The Rise of New Technologies in Marketing: A Framework and Outlook', *Journal of marketing*, 86: 1-6.

Hoffman, Donna L, C Page Moreau, Stefan Stremersch, and Michel Wedel. 2022. "The rise of new technologies in marketing: A framework and outlook." In, 1-6. SAGE Publications Sage CA: Los Angeles, CA.

Hogarth, Stuart, Gail Javitt, and David Melzer. 2008. 'The Current Landscape for Direct-to-Consumer Genetic Testing: Legal, Ethical, and Policy Issues', *Annual review of genomics and human genetics*, 9: 161-82.

Hoyer, Wayne D, Deborah J MacInnis, and Rik Pieters. 2012. *Consumer behavior* (Cengage Learning).

Hunt, Shelby D. 2020. 'Indigenous theory development in marketing: The foundational premises approach', *AMS Review*, 10: 8-17.

Hunt, Shelby D, and Dennis B Arnett. 2004. 'Market segmentation strategy, competitive advantage, and public policy: Grounding segmentation strategy in resource-advantage theory', *Australasian Marketing Journal (AMJ)*, 12: 7-25.

Huntington, George. 1872. 'On chorea'.

IVANOVA-KADIRI, IVELINA. 2023. "CUSTOMER GENETIC DATA FOR BUSINESS: EMPOWERING YOUR GENES FOR SUSTAINABLE PRODUCT DEVELOPMENT." In *7th FEB International Scientific Conference*, 619.

Jaiswal, Satish, Shao-Yang Tsai, Chi-Hung Juan, Wei-Kuang Liang, and Neil G Muggleton. 2018. 'Better cognitive performance is associated with the combination of high trait mindfulness and low trait anxiety', *Frontiers in Psychology*, 9: 627.

Jernigan, D., and C. S. Ross. 2020. 'The Alcohol Marketing Landscape: Alcohol Industry Size, Structure, Strategies, and Public Health Responses', *J Stud Alcohol Drugs Suppl*, Sup 19: 13-25.

Johnston, Timothy D, and Laura Edwards. 2002. 'Genes, interactions, and the development of behavior', *Psychological review*, 109: 26.

Jones, Harold Ellis. 1928. 'A first study of parent-child resemblance in intelligence', *Teachers College Record*, 29: 61-72.

Judd, Vaughan C, Cathleen A Owens, and Donald R Self. 1989. 'Market segmentation of alcoholics: A demographic/psychographic approach', *Health Marketing Quarterly*, 6: 57-92.

Kadiri, Ivelina. 2022. *GENETIC MARKETING: (R)EVOLUTION IN CUSTOMER SEGMENTATION*.

Kahle, Lynn R, Sharon E Beatty, and Pamela Homer. 1986. 'Alternative measurement approaches to consumer values: the list of values (LOV) and values and life style (VALS)', *Journal of Consumer Research*, 13: 405-09.

Kahle, Lynn R, and Patricia Kennedy. 1988. 'Using the list of values (LOV) to understand consumers', *Journal of Services Marketing*, 2: 49-56.

Kato, Kenji, and Nancy L. Pedersen. 2005. 'Personality and Coping: A Study of Twins Reared Apart and Twins Reared Together', *Behavior genetics*, 35: 147-58.

Kim, Jee Wook, Dong Young Lee, Boung Chul Lee, Myung Hun Jung, Hano Kim, Yong Sung Choi, and Ihn-Geun Choi. 2012. 'Alcohol and cognition in the elderly: a review', *Psychiatry Investigation*, 9: 8.

Klein, D. N., R. Kotov, and S. J. Bufferd. 2011. 'Personality and depression: explanatory models and review of the evidence', *Annu Rev Clin Psychol*, 7: 269-95.

Knopik, Valerie S, Jenae M Neiderhiser, John C DeFries, and Robert Plomin. 2013. *Behavioral genetics*

(Worth Publishers, Macmillan Learning New York).
Kotler, P. 1991. "From mass marketing to mass customization, Planning Review, September." In.: October.
Kotler, Philip, and G Armstrong. 1994. *Marketing management, analysis, planning, implementation, and control, Philip Kotler* (London: Prentice-Hall International).
Kotler, Philip, Suzan Burton, Kenneth Deans, Linen Brown, and Gary Armstrong. 2015. *Marketing* (Pearson Higher Education AU).
Kotov, Roman, Wakiza Gamez, Frank Schmidt, and David Watson. 2010. 'Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis', *Psychological bulletin*, 136: 768.
Kushner, Matt G, Kenneth J Sher, and Bernard D Beitman. 1990. 'The relation between alcohol problems and the anxiety disorders', *The American journal of psychiatry*, 147: 685-95.
Lambert, Samuel A, Laurent Gil, Simon Jupp, Scott C Ritchie, Yu Xu, Annalisa Buniello, Aoife McMahon, Gad Abraham, Michael Chapman, and Helen Parkinson. 2021. 'The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation', *Nature genetics*, 53: 420-25.
Lander, Eric S, and Robert A Weinberg. 2000. 'Journey to the center of biology', *Science*, 287: 1777-82.
Lang, Iain, Robert B Wallace, Felicia A Huppert, and David Melzer. 2007. 'Moderate alcohol consumption in older adults is associated with better cognition and well-being than abstinence', *Age and ageing*, 36: 256-61.
Larsen, Randy J, and Timothy Ketelaar. 1991. 'Personality and susceptibility to positive and negative emotional states', *Journal of personality and social psychology*, 61: 132.
Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, and Richard Karlsson Linnér. 2018. 'Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals', *Nature genetics*, 50: 1112-21.
Lehmann, Donald R, Leigh McAlister, and Richard Staelin. 2011. 'Sophistication in research in marketing', *Journal of marketing*, 75: 155-65.
Leiter, Michael P., and Christina Maslach. 2016. 'Latent burnout profiles: A new approach to understanding the burnout experience', *Burnout Research*, 3: 89-100.
Lesser, Jack A, and Marie Adele Hughes. 1986. 'The generalizability of psychographic market segments across geographic locations', *Journal of marketing*, 50: 18-27.
Leung, Rosanna, and Rob Law. 2010. 'A review of personality research in the tourism and hospitality context', *Journal of Travel & Tourism Marketing*, 27: 439-59.
Li, James J, Jeanne E Savage, Kenneth S Kendler, Matthew Hickman, Liam Mahedy, John Macleod, Jaakko Kaprio, Richard J Rose, and Danielle M Dick. 2017. 'Polygenic risk, personality dimensions, and adolescent alcohol use problems: a longitudinal study', *Journal of studies on alcohol and drugs*, 78: 442-51.
Lin, Yu-Ting, Deborah J MacInnis, and Andreas B Eisingerich. 2020. 'Strong anxiety boosts new product adoption when hope is also strong', *Journal of marketing*, 84: 60-78.
Loehlin, John C. 1992. *Genes and environment in personality development* (Sage Publications, Inc).
Loehlin, John C, and Robert C Nichols. 1976. *Heredity, environment, and personality: A study of 850 sets of twins* (University of Texas Press: New York, USA).
Lopes, Sérgio Dominique Ferreira, Antonio Rial Boubeta, and Jesús Varela Mallou. 2009. 'Post hoc tourist segmentation with conjoint and cluster analysis', *PASOS Revista de turismo y patrimonio cultural*, 7: 491-501.
Lyons, Jeffrey, Anand E Sokhey, Scott D McClurg, and Drew Seib. 2016. 'Personality, interpersonal disagreement, and electoral information', *The Journal of Politics*, 78: 806-21.
MacDonald, Marcy E, Christine M Ambrose, Mabel P Duyao, Richard H Myers, Carol Lin, Lakshmi Srinidhi, Glenn Barnes, Sherryl A Taylor, Marianne James, and Nicolet Groot. 1993. 'A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease

chromosomes', *Cell*, 72: 971-83.
Majumder, Mary A, Christi J Guerrini, and Amy L McGuire. 2021. 'Direct-to-consumer genetic testing: value and risk', *Annual Review of Medicine*, 72: 151-66.
Makela, Pia. 1999. 'Alcohol-related mortality as a function of socio-economic status', *Addiction*, 94: 867-86.
Mansky, Jackie. 2018. 'The dubious science of genetics-based dating', *Smithsonian. Com.*
Martin, Nicholas, and Rosemary Jardine. 1986. 'Eysenck's contributions to behaviour genetics.' in, *Hans Eysenck: consensus and controversy* (Routledge).
Matzler, Kurt, Rita Faullant, Birgit Renzl, and Verena Leiter. 2005. 'The relationship between personality traits (extraversion and neuroticism), emotions and customer self-satisfaction', *Innovative Marketing*, 1: 32-39.
McCrae, Robert R, and Paul T Costa. 2003. *Personality in adulthood: A five-factor theory perspective* (Guilford Press).
McGue, Matt, Steven Bacon, and David T Lykken. 1993. 'Personality stability and change in early adulthood: A behavioral genetic analysis', *Developmental psychology*, 29: 96.
Merriman, Curtis. 1924. 'The intellectual resemblance of twins', *Psychological Monographs*, 33: i.
Mooradian, Todd A. 1996. 'Personality and ad-evoked feelings: The case for extraversion and neuroticism', *Journal of the Academy of Marketing Science*, 24: 99-109.
Mulder, Roger T. 2002. 'Alcoholism and personality', *Australian & New Zealand Journal of Psychiatry*, 36: 46-51.
Myers, Susan D, Sandipan Sen, and Aliosha Alexandrov. 2010. 'The moderating effect of personality traits on attitudes toward advertisements: A contingency framework', *Management & Marketing*, 5.
Neisser, Ulric, Gwyneth Boodoo, Thomas J Bouchard Jr, A Wade Boykin, Nathan Brody, Stephen J Ceci, Diane F Halpern, John C Loehlin, Robert Perloff, and Robert J Sternberg. 1996. 'Intelligence: knowns and unknowns', *American psychologist*, 51: 77.
Newman, Horatio Hackett, Frank Nugent Freeman, and Karl John Holzinger. 1937. 'Twins: A study of heredity and environment'.
NIH, National Human Genome Research Institute. 2022. 'Human Genome Project Fact Sheet', NIH. https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project.
November, Joseph. 2018. 'More than Moore's mores: computers, genomics, and the embrace of innovation', *Journal of the History of Biology*, 51: 807-40.
O'Connor, Melissa C, and Sampo V Paunonen. 2007. 'Big Five personality predictors of post-secondary academic performance', *Personality and individual differences*, 43: 971-90.
Okbay, Aysu, Bart M. L. Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G. Nivard, Mark Alan Fontana, S. Fleur W. Meddens, Richard Karlsson Linnér, Cornelius A. Rietveld, Jaime Derringer, Jacob Gratten, James J. Lee, Jimmy Z. Liu, Ronald de Vlaming, Tarunveer S. Ahluwalia, Jadwiga Buchwald, Alana Cavadino, Alexis C. Frazier-Wood, Nicholas A. Furlotte, Victoria Garfield, Marie Henrike Geisel, Juan R. Gonzalez, Saskia Haitjema, Robert Karlsson, Sander W. van der Laan, Karl-Heinz Ladwig, Jari Lahti, Sven J. van der Lee, Penelope A. Lind, Tian Liu, Lindsay Matteson, Evelin Mihailov, Michael B. Miller, Camelia C. Minica, Ilja M. Nolte, Dennis Mook-Kanamori, Peter J. van der Most, Christopher Oldmeadow, Yong Qian, Olli Raitakari, Rajesh Rawal, Anu Realo, Rico Rueedi, Börge Schmidt, Albert V. Smith, Evie Stergiakouli, Toshiko Tanaka, Kent Taylor, Gudmar Thorleifsson, Juho Wedenoja, Juergen Wellmann, Harm-Jan Westra, Sara M. Willems, Wei Zhao, Najaf Amin, Andrew Bakshi, Sven Bergmann, Gyda Bjornsdottir, Patricia A. Boyle, Samantha Cherney, Simon R. Cox, Gail Davies, Oliver S. P. Davis, Jun Ding, Nese Direk, Peter Eibich, Rebecca T. Emeny, Ghazaleh Fatemifar, Jessica D. Faul, Luigi Ferrucci, Andreas J. Forstner, Christian Gieger, Richa Gupta, Tamara B. Harris, Juliette M. Harris, Elizabeth G. Holliday, Jouke-Jan Hottenga, Philip L. De Jager, Marika A. Kaakinen, Eero Kajantie, Ville Karhunen, Ivana Kolcic, Meena Kumari, Lenore J. Launer,

Lude Franke, Ruifang Li-Gao, David C. Liewald, Marisa Koini, Anu Loukola, Pedro Marques-Vidal, Grant W. Montgomery, Miriam A. Mosing, Lavinia Paternoster, Alison Pattie, Katja E. Petrovic, Laura Pulkki-Råback, Lydia Quaye, Katri Räikkönen, Igor Rudan, Rodney J. Scott, Jennifer A. Smith, Angelina R. Sutin, Maciej Trzaskowski, Anna E. Vinkhuyzen, Lei Yu, Delilah Zabaneh, John R. Attia, David A. Bennett, Klaus Berger, Lars Bertram, Dorret I. Boomsma, Harold Snieder, Shun-Chiao Chang, Francesco Cucca, Ian J. Deary, Cornelia M. van Duijn, Johan G. Eriksson, Ute Bültmann, Eco J. C. de Geus, Patrick J. F. Groenen, Vilmundur Gudnason, Torben Hansen, Catharine A. Hartman, Claire M. A. Haworth, Caroline Hayward, Andrew C. Heath, David A. Hinds, Elina Hyppönen, William G. Iacono, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L. R. Kardia, Liisa Keltikangas-Järvinen, Peter Kraft, Laura D. Kubzansky, Terho Lehtimäki, Patrik K. E. Magnusson, Nicholas G. Martin, Matt McGue, Andres Metspalu, Melinda Mills, Renée de Mutsert, Albertine J. Oldehinkel, Gerard Pasterkamp, Nancy L. Pedersen, Robert Plomin, Ozren Polasek, Christine Power, Stephen S. Rich, Frits R. Rosendaal, Hester M. den Ruijter, David Schlessinger, Helena Schmidt, Rauli Svento, Reinhold Schmidt, Behrooz Z. Alizadeh, Thorkild I. A. Sørensen, Tim D. Spector, John M. Starr, Kari Stefansson, Andrew Steptoe, Antonio Terracciano, Unnur Thorsteinsdottir, A. Roy Thurik, Nicholas J. Timpson, Henning Tiemeier, André G. Uitterlinden, Peter Vollenweider, Gert G. Wagner, David R. Weir, Jian Yang, Dalton C. Conley, George Davey Smith, Albert Hofman, Magnus Johannesson, David I. Laibson, Sarah E. Medland, Michelle N. Meyer, Joseph K. Pickrell, Tõnu Esko, Robert F. Krueger, Jonathan P. Beauchamp, Philipp D. Koellinger, Daniel J. Benjamin, Meike Bartels, David Cesarini, and Study LifeLines Cohort. 2016a. 'Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses', *Nature genetics*, 48: 624-33.

Okbay, Aysu, Jonathan P. Beauchamp, Mark Alan Fontana, James J. Lee, Tune H. Pers, Cornelius A. Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S. Fleur W. Meddens, Sven Oskarsson, Joseph K. Pickrell, Kevin Thom, Pascal Timshel, Ronald de Vlaming, Abdel Abdellaoui, Tarunveer S. Ahluwalia, Jonas Bacelis, Clemens Baumbach, Gyda Bjornsdottir, Johannes H. Brandsma, Maria Pina Concas, Jaime Derringer, Nicholas A. Furlotte, Tessel E. Galesloot, Giorgia Girotto, Richa Gupta, Leanne M. Hall, Sarah E. Harris, Edith Hofer, Momoko Horikoshi, Jennifer E. Huffman, Kadri Kaasik, Ioanna P. Kalafati, Robert Karlsson, Augustine Kong, Jari Lahti, Sven J. van der Lee, Christiaan deLeeuw, Penelope A. Lind, Karl-Oskar Lindgren, Tian Liu, Massimo Mangino, Jonathan Marten, Evelin Mihailov, Michael B. Miller, Peter J. van der Most, Christopher Oldmeadow, Antony Payton, Natalia Pervjakova, Wouter J. Peyrot, Yong Qian, Olli Raitakari, Rico Rueedi, Erika Salvi, Börge Schmidt, Katharina E. Schraut, Jianxin Shi, Albert V. Smith, Raymond A. Poot, Beate St Pourcain, Alexander Teumer, Gudmar Thorleifsson, Niek Verweij, Dragana Vuckovic, Juergen Wellmann, Harm-Jan Westra, Jingyun Yang, Wei Zhao, Zhihong Zhu, Behrooz Z. Alizadeh, Najaf Amin, Andrew Bakshi, Sebastian E. Baumeister, Ginevra Biino, Klaus Bønnelykke, Patricia A. Boyle, Harry Campbell, Francesco P. Cappuccio, Gail Davies, Jan-Emmanuel De Neve, Panos Deloukas, Ilja Demuth, Jun Ding, Peter Eibich, Lewin Eisele, Niina Eklund, David M. Evans, Jessica D. Faul, Mary F. Feitosa, Andreas J. Forstner, Ilaria Gandin, Bjarni Gunnarsson, Bjarni V. Halldórsson, Tamara B. Harris, Andrew C. Heath, Lynne J. Hocking, Elizabeth G. Holliday, Georg Homuth, Michael A. Horan, Jouke-Jan Hottenga, Philip L. de Jager, Peter K. Joshi, Astanand Jugessur, Marika A. Kaakinen, Mika Kähönen, Stavroula Kanoni, Liisa Keltigangas-Järvinen, Lambertus A. L. M. Kiemeney, Ivana Kolcic, Seppo Koskinen, Aldi T. Kraja, Martin Kroh, Zoltan Kutalik, Antti Latvala, Lenore J. Launer, Maël P. Lebreton, Douglas F. Levinson, Paul Lichtenstein, Peter Lichtner, David C. M. Liewald, LifeLines Cohort Study, Anu Loukola, Pamela A. Madden, Reedik Mägi, Tomi Mäki-Opas, Riccardo E. Marioni, Pedro Marques-Vidal, Gerardus A. Meddens, George McMahon, Christa Meisinger, Thomas Meitinger, Yusplitri Milaneschi, Lili Milani, Grant W. Montgomery, Ronny Myhre, Christopher P. Nelson, Dale R. Nyholt, William E. R. Ollier, Aarno Palotie, Lavinia Paternoster, Nancy L. Pedersen, Katja E. Petrovic, David J. Porteous, Katri Räikkönen,

Susan M. Ring, Antonietta Robino, Olga Rostapshova, Igor Rudan, Aldo Rustichini, Veikko Salomaa, Alan R. Sanders, Antti-Pekka Sarin, Helena Schmidt, Rodney J. Scott, Blair H. Smith, Jennifer A. Smith, Jan A. Staessen, Elisabeth Steinhagen-Thiessen, Konstantin Strauch, Antonio Terracciano, Martin D. Tobin, Sheila Ulivi, Simona Vaccargiu, Lydia Quaye, Frank J. A. van Rooij, Cristina Venturini, Anna A. E. Vinkhuyzen, Uwe Völker, Henry Völzke, Judith M. Vonk, Diego Vozzi, Johannes Waage, Erin B. Ware, Gonneke Willemsen, John R. Attia, David A. Bennett, Klaus Berger, Lars Bertram, Hans Bisgaard, Dorret I. Boomsma, Ingrid B. Borecki, Ute Bültmann, Christopher F. Chabris, Francesco Cucca, Daniele Cusi, Ian J. Deary, George V. Dedoussis, Cornelia M. van Duijn, Johan G. Eriksson, Barbara Franke, Lude Franke, Paolo Gasparini, Pablo V. Gejman, Christian Gieger, Hans-Jörgen Grabe, Jacob Gratten, Patrick J. F. Groenen, Vilmundur Gudnason, Pim van der Harst, Caroline Hayward, David A. Hinds, Wolfgang Hoffmann, Elina Hyppönen, William G. Iacono, Bo Jacobsson, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L. R. Kardia, Terho Lehtimäki, Steven F. Lehrer, Patrik K. E. Magnusson, Nicholas G. Martin, Matt McGue, Andres Metspalu, Neil Pendleton, Brenda W. J. H. Penninx, Markus Perola, Nicola Pirastu, Mario Pirastu, Ozren Polasek, Danielle Posthuma, Christine Power, Michael A. Province, Nilesh J. Samani, David Schlessinger, Reinhold Schmidt, Thorkild I. A. Sørensen, Tim D. Spector, Kari Stefansson, Unnur Thorsteinsdottir, A. Roy Thurik, Nicholas J. Timpson, Henning Tiemeier, Joyce Y. Tung, André G. Uitterlinden, Veronique Vitart, Peter Vollenweider, David R. Weir, James F. Wilson, Alan F. Wright, Dalton C. Conley, Robert F. Krueger, George Davey Smith, Albert Hofman, David I. Laibson, Sarah E. Medland, Michelle N. Meyer, Jian Yang, Magnus Johannesson, Peter M. Visscher, Tõnu Esko, Philipp D. Koellinger, David Cesarini, and Daniel J. Benjamin. 2016b. 'Genome-wide association study identifies 74 loci associated with educational attainment', *Nature*, 533: 539-42.

Okbay, A., Y. Wu, N. Wang, H. Jayashankar, M. Bennett, S. M. Nehzati, J. Sidorenko, H. Kweon, G. Goldman, T. Gjorgjieva, Y. Jiang, B. Hicks, C. Tian, D. A. Hinds, R. Ahlskog, P. K. E. Magnusson, S. Oskarsson, C. Hayward, A. Campbell, D. J. Porteous, J. Freese, P. Herd, C. Watson, J. Jala, D. Conley, P. D. Koellinger, M. Johannesson, D. Laibson, M. N. Meyer, J. J. Lee, A. Kong, L. Yengo, D. Cesarini, P. Turley, P. M. Visscher, J. P. Beauchamp, D. J. Benjamin, and A. I. Young. 2022. 'Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals', *Nat Genet*, 54: 437-49.

Oliver, Richard L. 1980. 'A cognitive model of the antecedents and consequences of satisfaction decisions', *Journal of marketing research*, 17: 460-69.

Organization, World Health 2019. "Global status report on alcohol and health 2018." In.: World Health Organization.

Oshio, Atsushi, Kanako Taku, Mari Hirano, and Gul Saeed. 2018. 'Resilience and Big Five personality traits: A meta-analysis', *Personality and individual differences*, 127: 54-60.

Otowa, T., K. Hek, M. Lee, E. M. Byrne, S. S. Mirza, M. G. Nivard, T. Bigdeli, S. H. Aggen, D. Adkins, A. Wolen, A. Fanous, M. C. Keller, E. Castelao, Z. Kutalik, S. Van der Auwera, G. Homuth, M. Nauck, A. Teumer, Y. Milaneschi, J. J. Hottenga, N. Direk, A. Hofman, A. Uitterlinden, C. L. Mulder, A. K. Henders, S. E. Medland, S. Gordon, A. C. Heath, P. A. Madden, M. L. Pergadia, P. J. van der Most, I. M. Nolte, F. V. van Oort, C. A. Hartman, A. J. Oldehinkel, M. Preisig, H. J. Grabe, C. M. Middeldorp, B. W. Penninx, D. Boomsma, N. G. Martin, G. Montgomery, B. S. Maher, E. J. van den Oord, N. R. Wray, H. Tiemeier, and J. M. Hettema. 2016. 'Meta-analysis of genome-wide association studies of anxiety disorders', *Mol Psychiatry*, 21: 1391-9.

Parks-Leduc, Laura, Gilad Feldman, and Anat Bardi. 2015. 'Personality traits and personal values: A meta-analysis', *Personality and Social Psychology Review*, 19: 3-29.

Pedersen, Nancy L, Robert Plomin, and Gerald E McClearn. 1994. 'Is there G beyond g?(Is there genetic influence on specific cognitive abilities independent of genetic influence on general cognitive ability?)', *Intelligence*, 18: 133-43.

Pedersen, N.L., R. Plomin, J.R. Nesselroade, and G.E. McClearn. 1992. 'A Quantitative Genetic Analysis of Cognitive Abilities during the Second Half of the Life Span', *Psychological Science*, 3: 346-53.

Pedersen, Ole Birger, Skytthe Axel, Klaus Rostgaard, Christian Erikstrup, Gustaf Edgren, Kaspar René Nielsen, Henrik Ullum, Kirsten Ohm Kyvik, and Henrik Hjalgrim. 2015. 'The heritability of blood donation: a population-based nationwide twin study', *Transfusion*, 55: 2169-74.

Plomin, Robert. 2011. 'Commentary: Why are children in the same family so different? Non-shared environment three decades later', *International journal of epidemiology*, 40: 582-92.

Plomin, Robert. 2019. *Blueprint, with a new afterword: How DNA makes us who we are* (Mit Press).

Plomin, Robert. 2023. 'Celebrating a Century of Research in Behavioral Genetics', *Behavior genetics*, 53: 75-84.

Plomin, Robert, and Ian J Deary. 2015. 'Genetics and intelligence differences: five special findings', *Molecular psychiatry*, 20: 98-108.

Plomin, Robert, John C DeFries, Valerie S Knopik, and Jenae M Neiderhiser. 2016. 'Top 10 replicated findings from behavioral genetics', *Perspectives on psychological science*, 11: 3-23.

Plomin, Robert, John C DeFries, and John C Loehlin. 1977. 'Genotype-environment interaction and correlation in the analysis of human behavior', *Psychological bulletin*, 84: 309.

Plomin, Robert, John C DeFries, and Gerald E McClearn. 2008. *Behavioral genetics* (Macmillan).

Plomin, Robert, David W. Fulker, Robin Corley, and John C DeFries. 1997. 'Nature, Nurture, and Cognitive Development from 1 to 16 Years: A Parent-Offspring Adoption Study', *Psychological Science*, 8: 442-47.

Plomin, Robert, Michael J Owen, and Peter McGuffin. 1994. 'The genetic basis of complex human behaviors', *Science*, 264: 1733-39.

Plomin, Robert, and Frank M. Spinath. 2002. 'Genetics and general cognitive ability (g)', *Trends in Cognitive Sciences*, 6: 169-76.

Plomin, Robert, and Sophie Von Stumm. 2018. 'The new genetics of intelligence', *Nature Reviews Genetics*, 19: 148-59.

Plomin, Robert, and Sophie Von Stumm. 2022. 'Polygenic scores: Prediction versus explanation', *Molecular psychiatry*, 27: 49-52.

Plummer, Joseph T. 1974. 'The concept and application of life style segmentation: The combination of two useful concepts provides a unique and important view of the market', *Journal of marketing*, 38: 33-37.

Prinzie, Peter, Geert Jan JM Stams, Maja Deković, Albert HA Reijntjes, and Jay Belsky. 2009. 'The relations between parents' Big Five personality factors and parenting: A meta-analytic review', *Journal of personality and social psychology*, 97: 351.

Punj, Girish, and David W Stewart. 1983. 'Cluster analysis in marketing research: Review and suggestions for application', *Journal of marketing research*, 20: 134-48.

Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, and P. Sklar. 2009. 'Common polygenic variation contributes to risk of schizophrenia and bipolar disorder', *Nature*, 460: 748-52.

Rao, Chatrathi P, and Zhengyuan Wang. 1995. 'Evaluating alternative segmentation strategies in standardindustrial markets', *European journal of marketing*, 29: 58-75.

Ray, Michael L. 1973. 'Marketing communication and the hierarchy-of-effects', *(No Title)*.

Reutterer, Thomas, and Daniel Dan. 2021. 'Cluster analysis in marketing research.' in, *Handbook of Market Research* (Springer).

Ribeiro, Juliane de Almeida, Ricardo Teixeira Veiga, and Agnaldo Keiti Higuchi. 2016. 'Personality traits and sustainable consumption', *Ribeiro, J., Veiga, R., & Higuchi, A.(2016). Personality Traits and Sustainable Consumption. Revista Brasileira De Marketing*, 15.

Rietveld, Cornelius A, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, and Arpana Agrawal. 2013. 'GWAS of 126,559 individuals identifies genetic variants associated with educational attainment', *Science*, 340: 1467-71.

Rietveld, Cornelius A, Eric AW Slob, and A Roy Thurik. 2021. 'A decade of research on the genetics of entrepreneurship: a review and view ahead', *Small Business Economics*, 57: 1303-17.
Riquelme, Hernan. 2001. 'Do consumers know what they want?', *Journal of Consumer Marketing*, 18: 437-48.
Robbins, Rebecca. 2016. 'Fruity with a hint of double helix: A startup claims to tailor wine to your DNA', Stat News, Accessed 07/30/2023.
Robinson, Matthew R, Aaron Kleinman, Mariaelisa Graff, Anna AE Vinkhuyzen, David Couper, Michael B Miller, Wouter J Peyrot, Abdel Abdellaoui, Brendan P Zietsch, and Ilja M Nolte. 2017. 'Genetic evidence of assortative mating in humans', *Nature Human Behaviour*, 1: 0016.
Rust, Roland T. 2006. 'From the editor: the maturation of marketing as an academic discipline', *Journal of marketing*, 70: 1-2.
Rust, Roland T. 2020. 'The future of marketing', *International Journal of Research in Marketing*, 37: 15-26.
Sanchez-Roige, Sandra, Joshua C Gray, James MacKillop, C-H Chen, and Abraham A Palmer. 2018. 'The genetics of human personality', *Genes, Brain and Behavior*, 17: e12439.
Sarnoff, Irving, and Daniel Katz. 1954. 'The motivational bases of attitude change', *The Journal of Abnormal and Social Psychology*, 49: 115.
Schwartz, Shalom H. 1994. 'Are there universal aspects in the structure and contents of human values?', *Journal of social issues*, 50: 19-45.
Selzam, Saskia, Stuart J Ritchie, Jean-Baptiste Pingault, Chandra A Reynolds, Paul F O'Reilly, and Robert Plomin. 2019. 'Comparing within-and between-family polygenic score prediction', *The American Journal of Human Genetics*, 105: 351-63.
Sibley, Chris G, and John Duckitt. 2008. 'Personality and prejudice: A meta-analysis and theoretical review', *Personality and Social Psychology Review*, 12: 248-79.
Simonson, Itamar, and Aner Sela. 2011. 'On the heritability of consumer decision making: An exploratory approach for studying genetic effects on judgment and choice', *Journal of Consumer Research*, 37: 951-66.
Smith, Wendell R. 1956. 'Product differentiation and market segmentation as alternative marketing strategies', *Journal of marketing*, 21: 3-8.
Snellman, Kaisa. 2000. *From One Segment to a Segment of One-The Evolution of Market Segmentation Theory* (Svenska handelshögskolan).
Soldz, Stephen, and George E Vaillant. 1999. 'The Big Five personality traits and the life course: A 45-year longitudinal study', *Journal of Research in Personality*, 33: 208-32.
Stanley, Laura, Franz W Kellermanns, and Thomas M Zellweger. 2017. 'Latent profile analysis: Understanding family firm profiles', *Family business review*, 30: 84-102.
Statista. 2023. "Alcoholic Drinks: market data and analysis." In, 166. Statista: Statista.
Suri, Rajneesh, and Kent B Monroe. 2001. 'The effects of need for cognition and trait anxiety on price acceptability', *Psychology & Marketing*, 18: 21-42.
Tawa, Elisabeth A., Samuel D. Hall, and Falk W. Lohoff. 2016. 'Overview of the Genetics of Alcohol Use Disorder', *Alcohol and Alcoholism*, 51: 507-14.
Tesser, Abraham. 1993. 'The importance of heritability in psychological research: the case of attitudes', *Psychological review*, 100: 129.
Tomkins, Susannah, Lyudmila Saburova, Nikolay Kiryanov, E Andreev, Michael McKee, Vladimir Shkolnikov, and David A Leon. 2007. 'Prevalence and socio-economic distribution of hazardous patterns of alcohol drinking: study of alcohol consumption in men aged 25–54 years in Izhevsk, Russia', *Addiction*, 102: 544-53.
Tuma, Michael N, Reinhold Decker, and Sören W Scholz. 2011. 'A survey of the challenges and pifalls of cluster analysis application in market segmentation', *International Journal of Market Research*, 53: 391-414.
Turkheimer, Eric. 2000. 'Three Laws of Behavior Genetics and What They Mean', *Current directions in psychological science*, 9: 160-64.

Turkheimer, Eric, Erik Pettersson, and Erin E Horn. 2014. 'A phenotypic null hypothesis for the genetics of personality', *Annual review of psychology*, 65: 515-40.
Twedt, DW. 1964. 'Is the "Heavy User"', *Journal of marketing*, 28: 71-72.
Twedt, Dik Warren. 1962. 'A cash register test of sales effectiveness', *Journal of marketing*, 26: 41-43.
Tynan, A Caroline, and Jennifer Drayton. 1987. 'Market segmentation', *Journal of marketing management*, 2: 301-35.
Uffelmann, Emil, Qin Qin Huang, Nchangwi Syntia Munung, Jantina de Vries, Yukinori Okada, Alicia R. Martin, Hilary C. Martin, Tuuli Lappalainen, and Danielle Posthuma. 2021. 'Genome-wide association studies', *Nature Reviews Methods Primers*, 1: 59.
Van den Berg, Stéphanie M, Marleen HM de Moor, Karin JH Verweij, Robert F Krueger, Michelle Luciano, Alejandro Arias Vasquez, Lindsay K Matteson, Jaime Derringer, Tonu Esko, and Najaf Amin. 2016. 'Meta-analysis of genome-wide association studies for extraversion: findings from the genetics of personality consortium', *Behavior genetics*, 46: 170-82.
Vandenberg, Steven G. 1972. 'Assortative mating, or who marries whom?', *Behavior genetics*, 2: 127-57.
Varadarajan, Rajan. 2010. 'Strategic marketing and marketing strategy: domain, definition, fundamental issues and foundational premises', *Journal of the Academy of Marketing Science*, 38: 119-40.
Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang. 2017. '10 Years of GWAS Discovery: Biology, Function, and Translation', *Am J Hum Genet*, 101: 5-22.
Vyncke, Patrick. 2002. 'Lifestyle segmentation: from attitudes, interests and opinions, to values, aesthetic styles, life visions and media preferences', *European journal of communication*, 17: 445-63.
Wallace, John F, Joseph P Newman, and Jo-Anne Bachorowski. 1991. 'Failures of response modulation: Impulsive behavior in anxious and impulsive individuals', *Journal of Research in Personality*, 25: 23-44.
Ware, Erin; Gard, Arianna; Schmitz, Lauren; Faul, Jessica. 2021. "HRS Polygenic Scores - Release 4.3." In. HRS Document Report: Health and Retirement Study HRS.
Watson, James D, and Francis HC Crick. 1953. "The structure of DNA." In *Cold Spring Harbor symposia on quantitative biology*, 123-31. Cold Spring Harbor Laboratory Press.
Wells, William D, and Douglas J Tigert. 1971. 'ACTIVITIES, INTERESTS AND OPINIONS¹', *Marketing: Critical Perspectives on Business and Management*, 2: 486.
Wharton, Knowledge at. 2020. 'Should your DNA data be used to sell products?', The Wharton School. University of Pensylvania, Accessed 07/30/2023. https://knowledge.wharton.upenn.edu/article/dna-data-used-sell-products/.
Widiger, Thomas A. 2016. 'An integrative model of personality strengths and weaknesses', *The Wiley handbook of positive clinical psychology*: 261-77.
Wilson, Brenda J, and Stuart G Nicholls. 2015. 'The Human Genome Project, and recent advances in personalized genomics', *Risk management and healthcare policy*: 9-20.
Wind, Yoram. 1978. 'Issues and advances in segmentation research', *Journal of marketing research*, 15: 317-37.
Wyner, Gordon. 2016. 'How Segmentation Provides the Roadmap to Success', *Recuperado el*, 30.
Yankelovich, Daniel. 1964. 'New criteria for market segmentation', *Harvard business review*, 42: 83-90.
Yankelovich, Daniel, and David Meer. 2006. 'Rediscovering market segmentation', *Harvard business review*, 84: 122.
Zaaijer, Sophie, Assaf Gordon, Daniel Speyer, Robert Piccone, Simon Cornelis Groen, and Yaniv Erlich. 2017. 'Rapid re-identification of human samples using portable DNA sequencing', *elife*, 6: e27798.
Zeithaml, Valarie A, Bernard J Jaworski, Ajay K Kohli, Kapil R Tuli, Wolfgang Ulaga, and Gerald Zaltman. 2020. 'A theories-in-use approach to building marketing theory', *Journal of marketing*, 84: 32-51.
Zhao, Hao, and Scott E Seibert. 2006. 'The big five personality dimensions and entrepreneurial status: a meta-analytical review', *Journal of applied psychology*, 91: 259.
Ziff, Ruth. 1971. 'Psychographics for market segmentation', *Journal of Advertising Research*, 11: 3-9.

APPENDIX: EMPIRICAL STUDY PER ANCESTRY

In this Appendix, it is provided the empirical model by ancestry. First, the European American ancestry data set and second, the African American ancestry data set.

A1.1 European Ancestry Data Analysis

For the European American ancestry data set, results are presented below:

Table A1. Descriptive Statistics for European ancestry

Descriptive Statistics for European Ancestry

Variables	N	Min	Max	Mean	Std. Deviation	Skewness		Kurtosis	
						Statistic	Std. Error	Statistic	Std. Error
PGS for Neuroticism	3533	-3.71	3.26	-0.0359	0.9985	-0.031	0.041	-0.052	0.082
PGS for Subj. Wellbeing	3533	-3.85	3.30	-0.0007	1.0054	-0.053	0.041	0.164	0.082
PGS for Depressive Symptoms	3533	-3.34	3.43	-0.0488	1.0064	0.043	0.041	-0.092	0.082
PGS for Extraversion	3533	-3.58	3.51	0.0175	1.0100	-0.116	0.041	-0.070	0.082
PGS for Educational Attainment	3533	-3.14	3.37	0.0888	1.0004	0.054	0.041	-0.136	0.082
PGS for Anxiety	3568	-3.49	3.12	0.0032	0.9814	-0.098	0.041	-0.067	0.082
PGS for Gen. Cognition	3533	-3.41	4.20	0.0818	1.0011	0.088	0.041	0.068	0.082
Age in Years	3533	41	102	72.41	9.375	0.215	0.041	-0.576	0.082
# of Days Drinking per Week	3533	0	8	2.47	2.496	0.730	0.041	-0.905	0.082
Income Level	989	0	9999999	636155.96	2318937.68	3.792	0.078	12.427	0.155
Level of Education	3533	0	17	14.15	2.235	-0.437	0.041	0.315	0.082
Valid N (listwise)	989								

Table A2. Pearson Correlation

Pearson Correlation											
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
PGS for Neuroticism	1.00										
PGS for Subj. Wellbeing	**-0.392****	1.00									
PGS for Depressive Symptoms	**0.563****	**-0.320****	1.00								
PGS for Extraversion	**-0.307****	**0.221****	**-0.102****	1.00							
PGS for Educational Attainment	**-0.073****	0.01	**-0.119****	0.03	1.00						
PGS for Anxiety	**0.121****	**-0.112****	**0.109****	**-0.034***	**-0.068****	1.00					
PGS for Gen. Cognition	**-0.039***	**0.045****	**-0.112****	-0.01	**0.511****	**-0.179****	1.00				
Age in Years	-0.02	0.02	-0.02	0.00	**0.068****	**-0.038***	**0.056****	1.00			
# of Days Drinking per Week	-0.02	0.03	-0.02	0.02	**0.070****	0.00	**0.052****	0.01	1.00		
Income Level	-0.01	0.00	0.00	0.01	-0.01	0.01	-0.01	**.078***	-0.01	1.00	
Level of Education	**-0.036***	-0.01	**-0.048****	0.02	**0.273****	-0.03	**0.193****	**-0.053****	**0.126****	-0.02	1.00
**. Correlation is significant at the 0.01 level (2-tailed).											
*. Correlation is significant at the 0.05 level (2-tailed).											

Cluster Analysis for European Ancestry

Based on the elbow method and the total WCSS, a 3-clusters solution was selected.

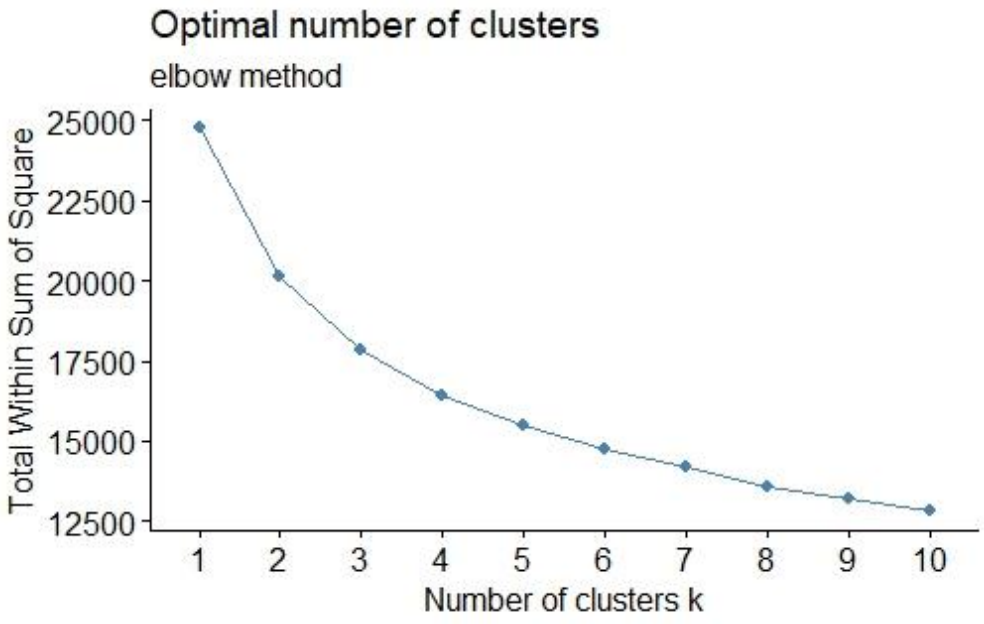
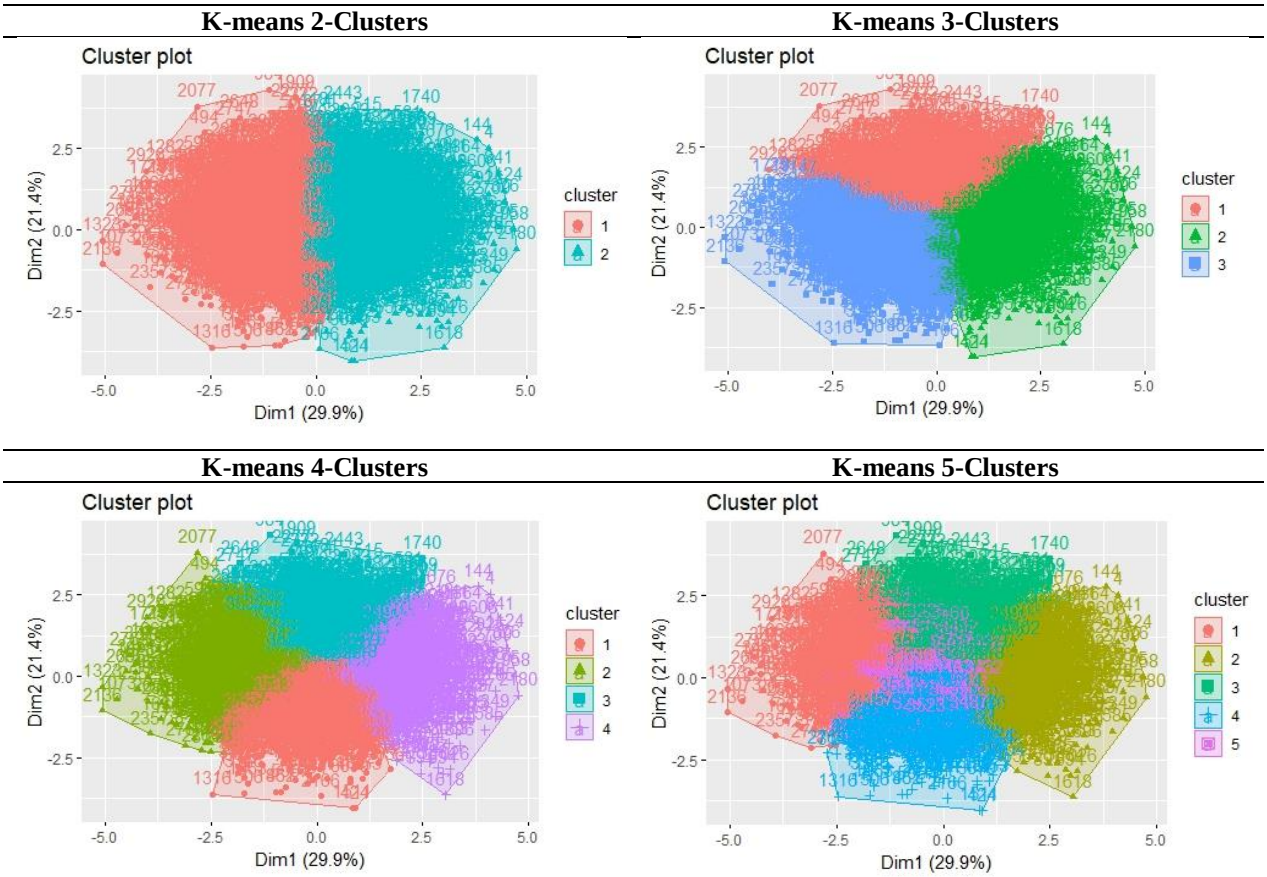


Figure A1. Elbow Method

Table A3. K-Means Cluster: Within Cluster Sum of Squares

K-means 2-Clusters	K-means 3-Clusters
Within cluster sum of squares by cluster: [1] 10577.328 [2]9526.068 **(total_SS = 18.7 %)**	Within cluster sum of squares by cluster: [1] 5731.690 [2] 6039.412 [3] 6084.390 **(total_SS = 27.8 %)**

K-means 4-Clusters	K-means 5-Clusters
Within cluster sum of squares by cluster: [1] 4137.397 [2] 3997.141 [3] 4656.047 [4] 3641.617 **(total_SS = 33.6 %)**	Within cluster sum of squares by cluster: [1] 2793.963 [2] 3456.467 [3] 3673.633 [4] 3115.727 [5] 2453.789 **(total_SS = 37.4 %)**



Dim1: Dimension 1; Dim2: Dimension 2; K-means Cluster Analysis

Figure A2. Cluster Visualization Comparison

Table A3. K-means Final Centers

Final Cluster Centers			
Variables	**Cluster**		
	1	**2**	**3**
PGS for Neuroticism	-0.6866	-0.1552	0.8353
PGS for Subj. Wellbeing	0.5163	0.1908	-0.6936
PGS for Depressive Symptoms	-0.6972	0.0280	0.6865
PGS for Extraversion	0.3263	0.2304	-0.5281
PGS for Educational Attainment	0.5591	-0.8234	0.1517
PGS for Anxiety	-0.3836	0.2833	0.1470
PGS for Gen. Cognition	0.5790	-0.8857	0.1845

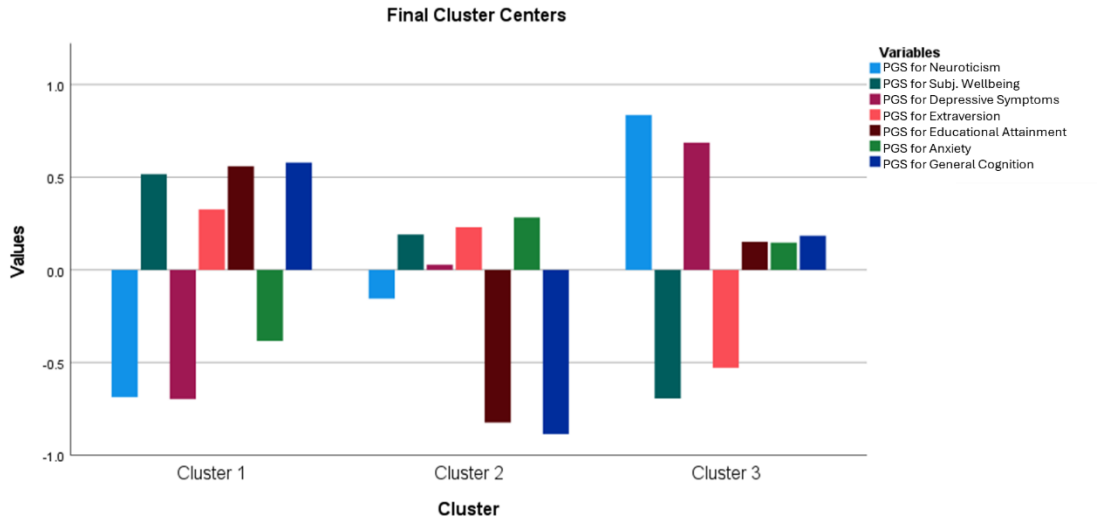


Figure A3. Final Cluster Centers Bar Graph for European Ancestry

Table A4. K-means ANOVA for Three-Clusters Solution

K-Mean ANOVA						
	Cluster		**Error**			
Variables	**Mean Square**	**Df**	**Mean Square**	**df**	**F**	**Sig.**
PGS for Neuroticism	733.727	2	0.582	3530	1260.680	0.000
PGS for Subj. Wellbeing	480.056	2	0.730	3530	657.866	0.000
PGS for Depressive Symptoms	591.652	2	0.668	3530	886.062	0.000
PGS for Extraversion	265.398	2	0.852	3530	311.499	0.000
PGS for Educational Attainment	568.913	2	0.679	3530	838.403	0.000
PGS for Anxiety	147.423	2	0.920	3530	160.297	0.000
PGS for Gen. Cognition	646.316	2	0.633	3530	1020.276	0.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Table A5. ANOVA and Post-hoc Test for PGS

ANOVA for Three-Cluster Solution for PGS						
		Sum of Squares	df	Mean Square	F	Sig.
PGS for Neuroticism	Between Groups	1467.455	2	733.727	1260.680	**0.000**
	Within Groups	2054.492	3530	0.582		
	Total	3521.947	3532			
PGS for Subj. Wellbeing	Between Groups	960.111	2	480.056	657.866	**0.000**
	Within Groups	2575.898	3530	0.730		
	Total	3536.009	3532			
PGS for Depressive Symptoms	Between Groups	1183.304	2	591.652	886.062	**0.000**
	Within Groups	2357.094	3530	0.668		
	Total	3540.398	3532			
PGS for Extraversion	Between Groups	530.795	2	265.398	311.499	**0.000**
	Within Groups	3007.567	3530	0.852		
	Total	3538.363	3532			
PGS for Educational Attainment	Between Groups	1137.826	2	568.913	838.403	**0.000**
	Within Groups	2395.343	3530	0.679		
	Total	3533.170	3532			
PGS for Anxiety	Between Groups	294.846	2	147.423	160.297	**0.000**
	Within Groups	3246.495	3530	0.920		
	Total	3541.341	3532			
PGS for Gen. Cognition	Between Groups	1292.632	2	646.316	1020.276	**0.000**
	Within Groups	2236.156	3530	0.633		
	Total	3528.788	3532			

Post-hoc test for Three-Cluster Solutions for PGS							
Bonferroni							
Independent Variable	Cluster		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
PGS for Neuroticism	1	2	-.531345089539702*	0.0319	**0.0000**	-0.6077	-0.4550
		3	-1.521890687867336*	0.0307	**0.0000**	-1.5954	-1.4484
	2	1	.531345089539702*	0.0319	**0.0000**	0.4550	0.6077
		3	-.990545598327635*	0.0320	**0.0000**	-1.0671	-0.9140
	3	1	1.521890687867337*	0.0307	**0.0000**	1.4484	1.5954
		2	.990545598327635*	0.0320	**0.0000**	0.9140	1.0671
PGS for Subj. Wellbeing	1	2	.325556870544847*	0.0357	**0.0000**	0.2401	0.4110
		3	1.209918303829092*	0.0344	**0.0000**	1.1276	1.2923
	2	1	-.325556870544847*	0.0357	**0.0000**	-0.4110	-0.2401
		3	.884361433284245*	0.0358	**0.0000**	0.7986	0.9701
	3	1	-1.209918303829092*	0.0344	**0.0000**	-1.2923	-1.1276

		2	-.884361433284245*	0.0358	**0.0000**	-0.9701	-0.7986
PGS for Depressive Symptoms	1	2	-.725235276240408*	0.0341	**0.0000**	-0.8070	-0.6435
		3	-1.383692692817927*	0.0329	**0.0000**	-1.4625	-1.3049
	2	1	.725235276240408*	0.0341	**0.0000**	0.6435	0.8070
		3	-.658457416577518*	0.0342	**0.0000**	-0.7405	-0.5764
	3	1	1.383692692817927*	0.0329	**0.0000**	1.3049	1.4625
		2	.658457416577519*	0.0342	**0.0000**	0.5764	0.7405
PGS for Extraversion	1	2	.095905528774832*	0.0386	**0.0387**	0.0036	0.1882
		3	.854402521444882*	0.0371	**0.0000**	0.7654	0.9434
	2	1	-.095905528774832*	0.0386	**0.0387**	-0.1882	-0.0036
		3	.758496992670050*	0.0387	**0.0000**	0.6658	0.8511
	3	1	-.854402521444882*	0.0371	**0.0000**	-0.9434	-0.7654
		2	-.758496992670050*	0.0387	**0.0000**	-0.8511	-0.6658
PGS for Educational Attainment	1	2	1.382518000357015*	0.0344	**0.0000**	1.3001	1.4649
		3	.407330339117949*	0.0332	**0.0000**	0.3279	0.4867
	2	1	-1.382518000357015*	0.0344	**0.0000**	-1.4649	-1.3001
		3	-.975187661239066*	0.0345	**0.0000**	-1.0579	-0.8925
	3	1	-.407330339117949*	0.0332	**0.0000**	-0.4867	-0.3279
		2	.975187661239066*	0.0345	**0.0000**	0.8925	1.0579
PGS for Anxiety	1	2	-.666881245469491*	0.0401	**0.0000**	-0.7628	-0.5709
		3	-.530603631161244*	0.0386	**0.0000**	-0.6230	-0.4382
	2	1	.666881245469491*	0.0401	**0.0000**	0.5709	0.7628
		3	.136277614308248*	0.0402	**0.0021**	0.0400	0.2325
	3	1	.530603631161244*	0.0386	**0.0000**	0.4382	0.6230
		2	-.136277614308247*	0.0402	**0.0021**	-0.2325	-0.0400
PGS for General Cognition	1	2	1.464707639741960*	0.0332	**0.0000**	1.3851	1.5443
		3	.394576238538374*	0.0320	**0.0000**	0.3179	0.4713
	2	1	-1.464707639741960*	0.0332	**0.0000**	-1.5443	-1.3851
		3	-1.070131401203586*	0.0334	**0.0000**	-1.1500	-0.9902
	3	1	-.394576238538374*	0.0320	**0.0000**	-0.4713	-0.3179
		2	1.070131401203586*	0.0334	**0.0000**	0.9902	1.1500

*. The mean difference is significant at the 0.05 level.

Analysis of Variance (ANOVA) for the Dependent Variable

Table A5. ANOVA for Dependent Variable

ANOVA for Dependent Variable					
# of Days Drinking per Week					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	7.120	2	3.560	3.700	**0.025**

Within Groups	3396.219	3530	0.962
Total	3403.339	3532	

Table A6. Post-hoc Study for Dependent Variable

Post-hoc Test for Dependent Variable						
Dependent Variable: # of Days Drinking per Week						
Bonferroni						
(I) Cluster Number	**(J) Cluster Number**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
					Lower Bound	**Upper Bound**
1	2	.10505*	0.0410	**0.0311**	0.0069	0.2032
	3	0.0169	0.0395	1.0000	-0.0776	0.1115
2	1	-0.1051*	0.0410	**0.0311**	-0.2032	-0.0069
	3	-0.0881	0.0411	0.0963	-0.1866	0.0103
3	1	-0.0169	0.0395	1.0000	-0.1115	0.0776
	2	0.0881	0.0411	0.0963	-0.0103	0.1866
*. The mean difference is significant at the 0.05 level.						

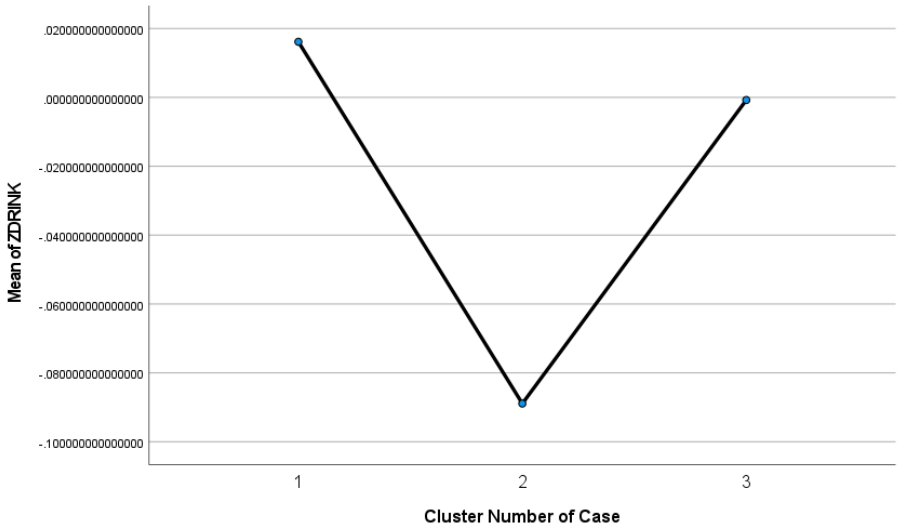


Figure A4. Mean Plot for Dependent Variable in Three-Clusters Solution

Table A7. ANCOVA for Dependent Variable and Covariates

Tests of Between-Subjects Effects
Dependent Variable: # of days drinking per week

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	57.082[a]	4	14.271	15.046	**0.000**
Intercept	1.844	1	1.844	1.944	0.163
ZAGE	1.056	1	1.056	1.113	0.291
ZEDUC	49.634	1	49.634	52.329	**0.000**
QCL_1	1.557	2	0.779	0.821	0.440
Error	3346.257	3528	0.948		
Total	3404.944	3533			
Corrected Total	3403.339	3532			
a. R Squared = .017 (Adjusted R Squared = .016)					

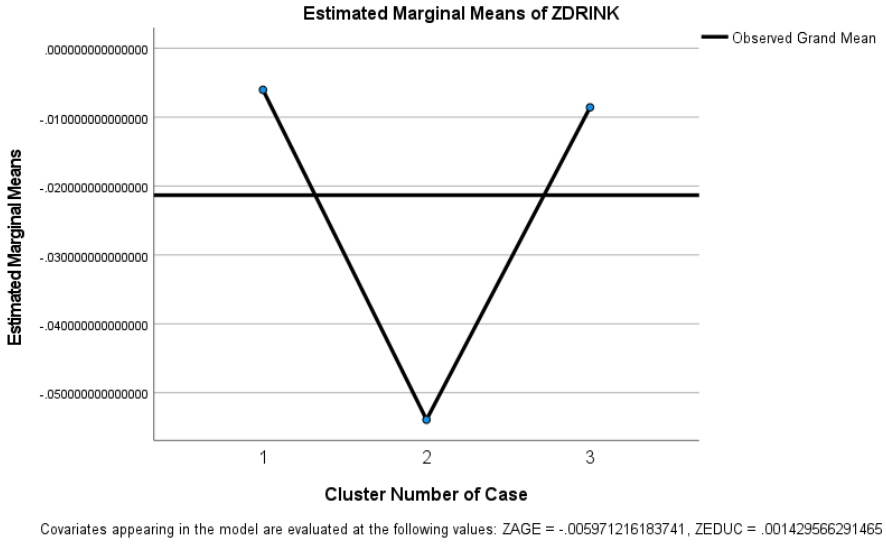


Figure A5. Mean Plot for Dependent Variable in Three-Clusters with Covariates

A1.2 African Ancestry Data Analysis

African ancestry data set analysis is presented below:

Table A7. Descriptive Statistics for African ancestry

Descriptive Statistics for African Ancestry									
						Skewness		Kurtosis	
Variables	**N**	**Min**	**Max**	**Mean**	**Std. Deviation**	**Statistic**	**Std. Error**	**Statistic**	**Std. Error**
PGS for Neuroticism	810	-3.15	3.09	-0.0266	0.99453	-0.122	0.086	0.044	0.172
PGS for Subj. Wellbeing	810	-3.17	3.17	-0.0187	1.00959	0.042	0.086	0.021	0.172
PGS for Depressive Symptoms	810	-3.01	3.36	-0.0540	1.00297	0.068	0.086	-0.233	0.172
PGS for Extraversion	810	-3.98	2.73	-0.0467	1.00285	0.050	0.086	-0.156	0.172
PGS for Educational Attainment	810	-2.86	3.37	0.0211	0.96972	0.110	0.086	0.183	0.172
PGS for Anxiety	810	-4.05	3.08	0.0269	0.99379	-0.046	0.086	0.527	0.172
PGS for Gen. Cognition	810	-3.40	4.08	0.0733	1.00798	0.120	0.086	0.358	0.172
Age in Years	810	32	98	67.63	8.123	0.253	0.086	1.094	0.172
# of Days Drinking per Week	810	0	7	1.61	1.800	1.381	0.086	1.516	0.172
Income Level	273	0	9999999	1865526.01	3858994.276	1.647	0.147	0.719	0.294
Level of Education	810	1	17	13.16	2.423	-0.587	0.086	1.439	0.172
Valid N (listwise)	273								

Table A8. Bivariate Correlation for African ancestry

	Pearson Correlations										
Variables	1	2	3	4	5	6	7	8	9	10	11
PGS for Neuroticism	1.00										
PGS for Subj. Wellbeing	**-0.326****	1.00									
PGS for Depressive Symptoms	**0.503****	**-0.236****	1.00								
PGS for Extraversion	**-0.220****	**0.149****	-0.05	1.00							
PGS for Educational Attainment	**-0.093****	0.04	**-0.173****	0.02	1.00						
PGS for Anxiety	0.03	-0.04	0.03	-0.04	-0.01	1.00					
PGS for Gen. Cognition	**-0.110****	0.06	**-0.157****	0.04	0.397**	**-0.075***	1.00				
Age in Years	-0.01	0.07	0.04	0.04	0.05	-0.02	-0.02	1.00			
# of Days Drinking per Week	-0.03	0.05	-0.02	-0.04	-0.01	0.02	0.00	**-0.113****	1.00		
Income Level	-0.05	0.02	0.02	0.09	0.07	-0.06	0.08	0.05	0.00	1.00	
Level of Education	0.00	0.01	-0.01	0.01	**0.084***	0.00	.137**	-0.06	-0.04	**-.186****	1.00

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Cluster Analysis for African Ancestry

Based on the elbow method and the total within SS, a 3-clusters solution was selected.

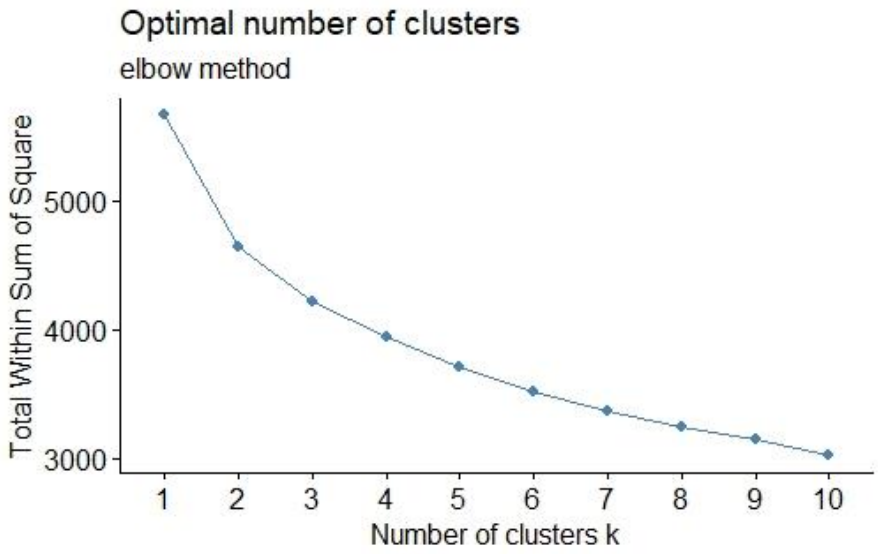
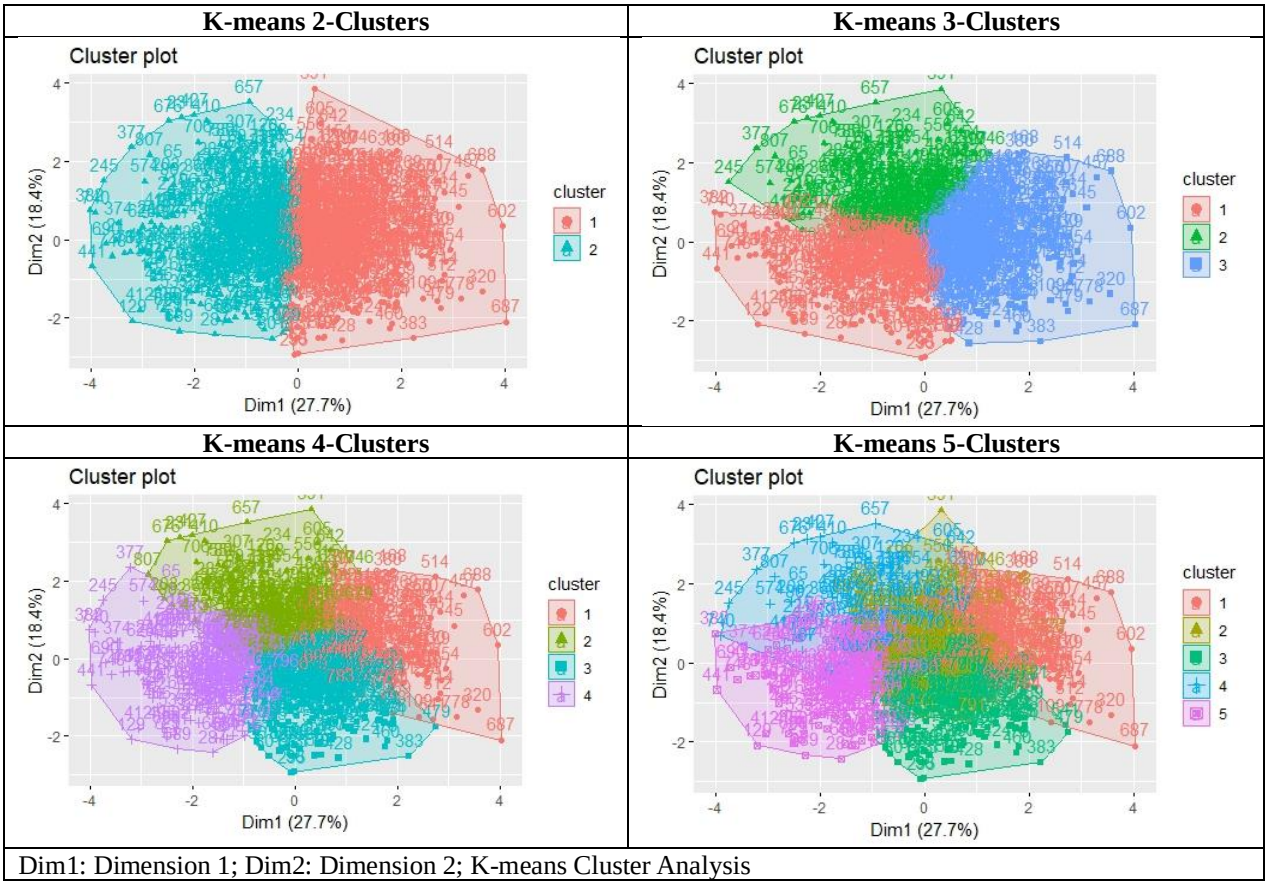


Figure A3. Elbow Method

Table A9. K-Means Cluster: Within Sum of Squares

K-means 2-Clusters	**K-means 3-Clusters**
Within cluster sum of squares by cluster: [1] 2486.431 [2] 2155.320 **(total_SS = 18.1 %)**	Within cluster sum of squares by cluster: [1] 1301.864 [2] 1174.006 [3] 1741.284 **(total_SS = 25.5 %)**
K-means 4-Clusters	**K-means 5-Clusters**
Within cluster sum of squares by cluster: [1] 861.9481 [2] 942.7612 [3] 1084.9757 [4] 1017.9457 **(total_SS = 31.0 %)**	Within cluster sum of squares by cluster: [1] 763.3582 [2] 645.8678 [3] 848.7116 [4] 589.0643 [5] 820.0117 **(total_SS = 35.3 %)**



Dim1: Dimension 1; Dim2: Dimension 2; K-means Cluster Analysis

Figure A4. Cluster Plot Comparison

Table A10. K-means Clustering

Final Cluster Centers			
Variables	**Cluster**		
	1	**2**	**3**
PGS for Neuroticism	**0.6251**	-0.9152	**0.1430**
PGS for Subj. Wellbeing	-0.3191	**0.7168**	-0.3793
PGS for Depressive Symptoms	**0.6766**	-0.6973	-0.1880
PGS for Extraversion	-0.1582	**0.5105**	-0.3634
PGS for Educational Attainment	-0.5686	0.0357	**0.8221**
PGS for Anxiety	0.0303	-0.0667	0.0412
PGS for Gen. Cognition	-0.5518	0.0329	**0.7959**

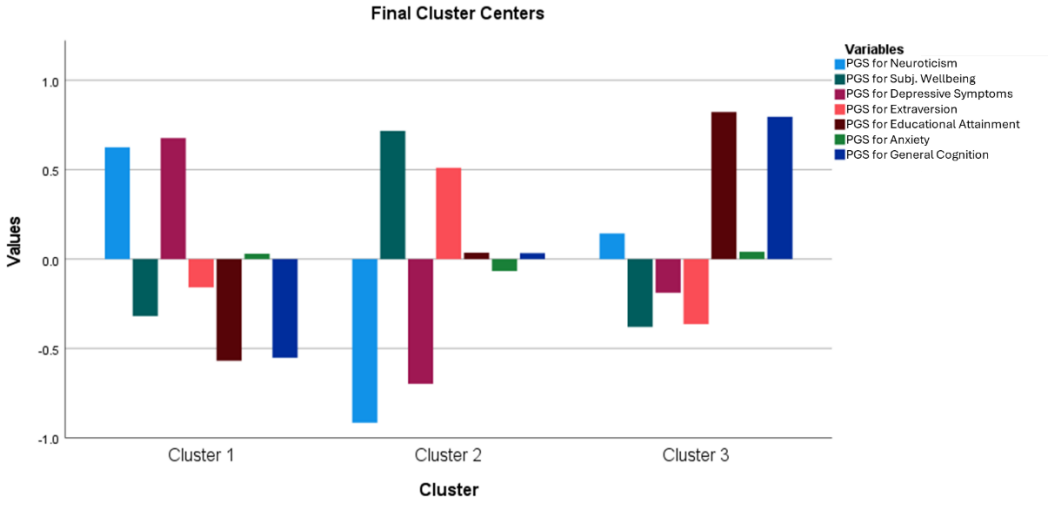


Figure A5. Final Cluster Centers Bar Graph

Table A11. K-mean ANOVA for Three-Clusters Solution

K-mean ANOVA						
Variables	**Cluster**		**Error**		**F**	**Sig.**
	Mean Square	**df**	**Mean Square**	**df**		
PGS for Neuroticism	176.431	2	0.563	807	313.587	**0.000**
PGS for Subj. Wellbeing	99.793	2	0.757	807	131.779	**0.000**
PGS for Depressive Symptoms	143.078	2	0.650	807	220.123	**0.000**
PGS for Extraversion	52.655	2	0.871	807	60.424	**0.000**
PGS for Educational Attainment	127.195	2	0.683	807	186.331	**0.000**
PGS for Anxiety	0.917	2	1.003	807	0.915	0.401
PGS for Gen. Cognition	119.430	2	0.709	807	168.443	**0.000**

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Table A12. ANOVA and Post-hoc Analysis of Three-Cluster Solution for PGS

ANOVA for PGS						
Variables		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
PGS for Neuroticism	Between Groups	352.861	2	176.431	313.587	0.000
	Within Groups	454.036	807	0.563		

	Total	806.897	809			
PGS for Subj. Wellbeing	Between Groups	199.585	2	99.793	131.779	0.000
	Within Groups	611.117	807	0.757		
	Total	810.702	809			
PGS for Depressive Symptoms	Between Groups	286.155	2	143.078	220.123	0.000
	Within Groups	524.542	807	0.650		
	Total	810.697	809			
PGS for Extraversion	Between Groups	105.310	2	52.655	60.424	0.000
	Within Groups	703.244	807	0.871		
	Total	808.554	809			
PGS for Educational Attainment	Between Groups	254.390	2	127.195	186.331	0.000
	Within Groups	550.880	807	0.683		
	Total	805.270	809			
PGS for Anxiety	Between Groups	1.835	2	0.917	0.915	0.401
	Within Groups	809.405	807	1.003		
	Total	811.240	809			
PGS for General Cognition	Between Groups	238.860	2	119.430	168.443	0.000
	Within Groups	572.181	807	0.709		
	Total	811.041	809			

Post-hoc study for PGS

Bonferroni

Independent Variable	**Cluster Numbers (I)**	**Cluster Numbers (J)**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
						Lower Bound	**Upper Bound**
PGS for Neuroticism	1	2	1.540334888369262*	0.0621	**0.0000**	1.3914	1.6892
		3	.482144718128087*	0.0655	**0.0000**	0.3251	0.6392
	2	1	-1.540334888369262*	0.0621	**0.0000**	-1.6892	-1.3914
		3	-1.058190170241175*	0.0688	**0.0000**	-1.2231	-0.8932
	3	1	-.482144718128086*	0.0655	**0.0000**	-0.6392	-0.3251
		2	1.058190170241175*	0.0688	**0.0000**	0.8932	1.2231
PGS for Subj. Wellbeing	1	2	-1.035925931657972*	0.0720	**0.0000**	-1.2087	-0.8632
		3	0.060162743699419	0.0760	1.0000	-0.1220	0.2424
	2	1	1.035925931657972*	0.0720	**0.0000**	0.8632	1.2087
		3	1.096088675357391*	0.0798	**0.0000**	0.9047	1.2875
	3	1	-0.060162743699419	0.0760	1.0000	-0.2424	0.1220
		2	-1.096088675357391*	0.0798	**0.0000**	-1.2875	-0.9047
PGS for Depressive Symptoms	1	2	1.373869560916605*	0.0667	**0.0000**	1.2138	1.5339
		3	.864598955190849*	0.0704	**0.0000**	0.6958	1.0334
	2	1	-1.373869560916605*	0.0667	**0.0000**	-1.5339	-1.2138
		3	-.509270605725756*	0.0739	**0.0000**	-0.6866	-0.3320
	3	1	-.864598955190849*	0.0704	**0.0000**	-1.0334	-0.6958
		2	.509270605725756*	0.0739	**0.0000**	0.3320	0.6866
PGS for Extraversion	1	2	-.668693889679226*	0.0772	**0.0000**	-0.8540	-0.4834
		3	.205138362018008*	0.0815	**0.0360**	0.0097	0.4006
	2	1	.668693889679226*	0.0772	**0.0000**	0.4834	0.8540
		3	.873832251697233*	0.0856	**0.0000**	0.6685	1.0791

	3	1	-.205138362018007*	0.0815	**0.0360**	-0.4006	-0.0097
		2	-.873832251697233*	0.0856	**0.0000**	-1.0791	-0.6685
PGS for Educational Attainment	1	2	-.604341522686996*	0.0684	**0.0000**	-0.7684	-0.4403
		3	-1.390784169430957*	0.0721	**0.0000**	-1.5638	-1.2178
	2	1	.604341522686997*	0.0684	**0.0000**	0.4403	0.7684
		3	-.786442646743961*	0.0757	**0.0000**	-0.9681	-0.6047
	3	1	1.390784169430957*	0.0721	**0.0000**	1.2178	1.5638
		2	.786442646743961*	0.0757	**0.0000**	0.6047	0.9681
PGS for Anxiety	1	2	0.096970952914639	0.0829	0.7269	-0.1018	0.2958
		3	-0.010909210070270	0.0874	1.0000	-0.2206	0.1988
	2	1	-0.096970952914639	0.0829	0.7269	-0.2958	0.1018
		3	-0.107880162984910	0.0918	0.7210	-0.3281	0.1124
	3	1	0.010909210070270	0.0874	1.0000	-0.1988	0.2206
		2	0.107880162984910	0.0918	0.7210	-0.1124	0.3281
PGS for General Cognition	1	2	-.584752367063569*	0.0697	**0.0000**	-0.7519	-0.4176
		3	-1.347707104555951*	0.0735	**0.0000**	-1.5240	-1.1714
	2	1	.584752367063569*	0.0697	**0.0000**	0.4176	0.7519
		3	-.762954737492382*	0.0772	**0.0000**	-0.9481	-0.5778
	3	1	1.347707104555951*	0.0735	**0.0000**	1.1714	1.5240
		2	.762954737492382*	0.0772	**0.0000**	0.5778	0.9481

*. The mean difference is significant at the 0.05 level.

Table A13. ANOVA for Dependent Variable

Descriptives of Dependent Variable within Three-Cluster Solution

DV: # of Days Drinking per Week

Cluster Number	**N**	**Mean**	**Std. Deviation**	**Std. Error**	**95% Confidence Interval for Mean**		**Min**	**Max**
					Lower Bound	**Upper Bound**		
1	330	**-0.0762**	0.9225	0.0508	-0.1761	0.0237	-0.8843	2.8488
2	262	**0.0480**	0.9744	0.0602	-0.0705	0.1665	-0.8843	2.8488
3	218	**-0.0354**	0.9972	0.0675	-0.1685	0.0977	-0.8843	2.8488
Total	810	-0.0251	0.9602	0.0337	-0.0913	0.0412	-0.8843	2.8488

ANOVA for Dependent Variable Three-Cluster Comparison

DV: # of Days Drinking per Week

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.286	2	1.143	1.240	**0.290**
Within Groups	743.573	807	0.921		
Total	745.859	809			

Table A14. Post-hoc test for Dependent Variable

Post-hoc Test for Dependent Variable in the Three-Cluster Solution

Dependent Variable: # of Days Drinking per Week

Bonferroni

(I) Cluster Number	(J) Cluster Number	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-0.12423	0.07943	0.35463	-0.31478	0.06632

	3	-0.04085	0.08378	1.00000	-0.24183	0.16014
2	1	0.12423	0.07943	0.35463	-0.06632	0.31478
	3	0.08338	0.08800	1.00000	-0.12772	0.29448
3	1	0.04085	0.08378	1.00000	-0.16014	0.24183
	2	-0.08338	0.08800	1.00000	-0.29448	0.12772

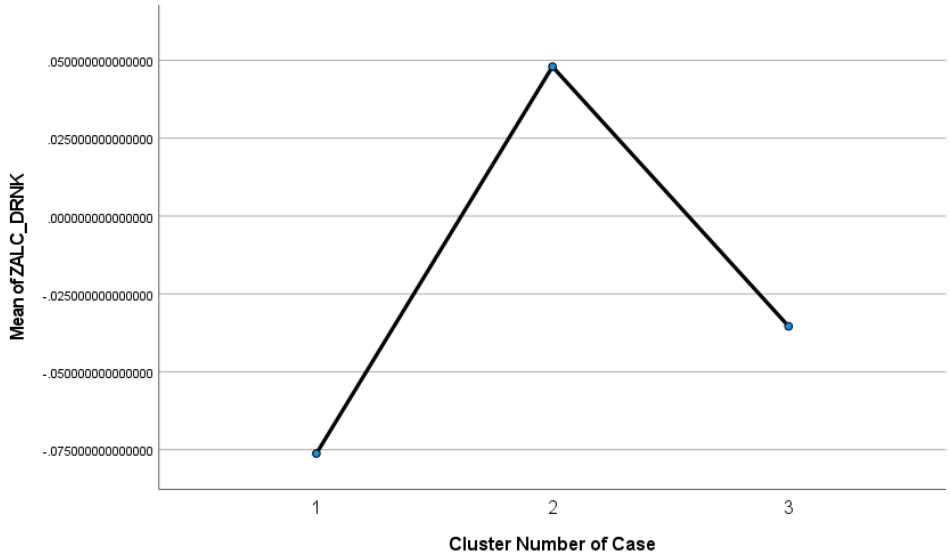


Figure A6. Mean Plot of Dependent Variable in the Three-Clusters

Table A15. ANCOVA for Dependent Variable and Covariates

ANCOVA Tests of Between-Subjects Effects					
Dependent Variable: # of days drinking per week					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	13.810[a]	4	3.453	3.797	**0.005**
Intercept	0.358	1	0.358	0.393	0.531
ZAGE	10.040	1	10.040	11.041	**0.001**
ZEDUC	1.947	1	1.947	2.141	0.144
QCL_1	2.454	2	1.227	1.349	0.260
Error	732.048	805	0.909		
Total	746.367	810			
Corrected Total	745.859	809			
a. R Squared = .019 (Adjusted R Squared = .014)					

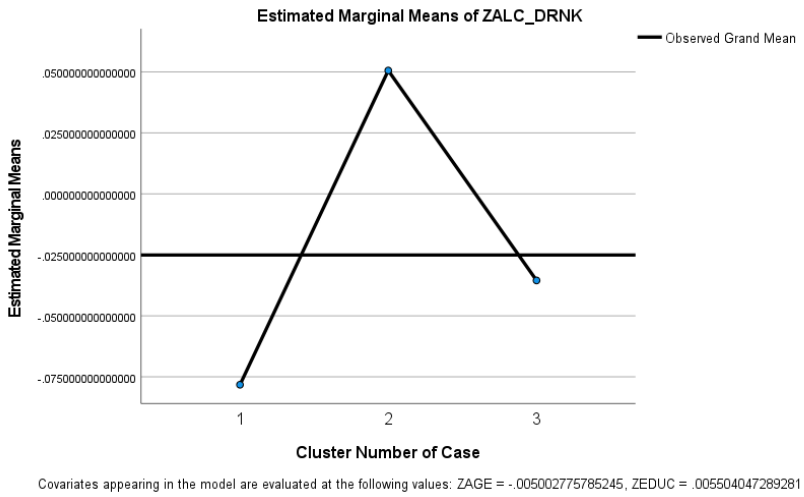


Figure A7. Mean Plot of Dependent Variable in the Three-Clusters with Covariates