Survey Analysis

95-851 Making Products Count:

Data Science for Product Managers – Final Project

Spring 2024

Terry Wei, Hsin-Li (Cindy) Kan,

Yvonne Hsu, Gautam Devadiga

Instructed by Professor David Steier

Table of contents

1. Executive Summary

Problem Statement: As YouTube TV ventures into the streaming market, identifying the target audience and developing tailored engagement strategies are crucial. The project leverages Deloitte's 2014 and 2016 Digital Democracy Survey data to address these challenges, focusing on digital device preferences, streaming habits, and demographic insights (Deloitte, n.d.).

Solution and Objectives: Through comprehensive data cleaning and exploratory analysis, we've discovered key demographics (peak age at 65, income range of $50,000 to $100,000, and high employment rate) and preferences (laptops, smartphones, and TVs as top entertainment devices). Principal Component Analysis (PCA) and clustering techniques helped us identify two distinct viewer clusters, enabling targeted strategies for engaging younger viewers primarily using computers and mature viewers favoring TV.

Value Proposition: Our analysis culminates in a dual-strategy approach:

- Younger Segment: Focus on interactive content and second-screen experiences to captivate computer users, alongside value-oriented packages, flexible payment options, and loyalty programs.
- Mature Segment: Emphasize high-quality broadcasts and a simplified user interface for TV watchers, coupled with exclusive events and priority customer service for the affluent.

Conclusion: This project underscores the importance of a data-driven approach to market entry for YouTube TV, ensuring product offerings and marketing strategies resonate with distinct viewer segments. By addressing the unique needs and preferences of each cluster, YouTube TV can enhance user engagement, satisfaction, and loyalty, securing a competitive edge in the streaming service industry. The next steps involve implementing these strategies and monitoring their impact on market performance.

2. Introduction

Our study as YouTube TV product managers is motivated by the need to strategically negotiate the competitive streaming video services market. With the support of data from the Deloitte Digital Democracy Surveys, which covered the years 2014 and 2016 (Deloitte, n.d.), we aim to develop a strong go-to-market plan for YouTube TV, which debuted in the market in 2017 (Boorstin, 2017). The surveys provide a plethora of information on the usage of streaming video services, preferences for digital devices, and other topics, which lays a strong basis for answering our main questions about target audience identification and engagement tactics. This analysis is essential to establishing YouTube TV as a leader in the changing media consumption trends of the digital age.

3. Problem statement

As we stand on the brink of launching YouTube TV into the competitive streaming market, two critical questions underpin our strategy: Who should we target? and How should we treat them? The Deloitte Digital Democracy Surveys offer a rich dataset to dissect viewer habits, preferences, and technology usage trends across various demographics. Our analysis aims to:

1. Identify the ideal target audience for YouTube TV, drawing on survey responses related to digital device preferences, streaming video service utilization, and content consumption behaviors.
2. Develop tailored engagement strategies for the identified target segments, focusing on content offerings, platform features, and marketing messages that resonate with their unique preferences and habits.

These questions are central to crafting a go-to-market strategy that ensures YouTube TV not only enters the market but does so with a significant impact, appealing directly to the needs and wants of its potential users.

4. Benefits

The strategic analysis and resultant go-to-market strategy for YouTube TV, informed by the 2014 and 2016 Deloitte Digital Democracy Surveys, are poised to deliver multifaceted benefits:

1. For YouTube TV: Enhanced market positioning and differentiation by targeting the most receptive audience segments with tailored content and features. This approach will accelerate user adoption and loyalty, ensuring a strong foothold in the competitive streaming service market.
2. For Users: A highly personalized viewing experience that aligns with individual preferences, device usage, and content consumption habits, enhancing overall satisfaction and engagement with the platform.

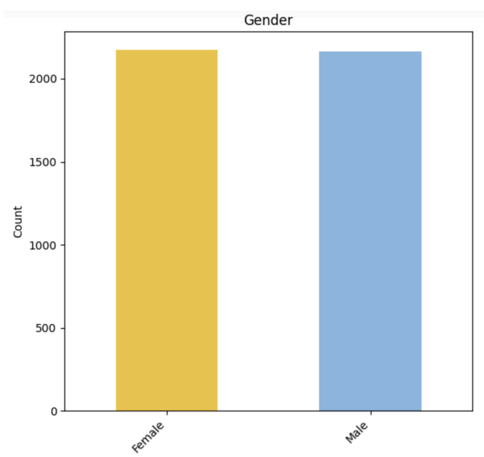
3. For Advertisers and Content Creators: Improved targeting and engagement metrics, as YouTube TV's data-driven strategy enables more effective matching of content with viewer interests, potentially leading to higher conversion rates and audience retention.

With a win-win outcome for YouTube TV, its audience, and its partners, this comprehensive benefit structure emphasizes the value of using in-depth market insights to guide strategic decisions.

5. Exploratory data analysis

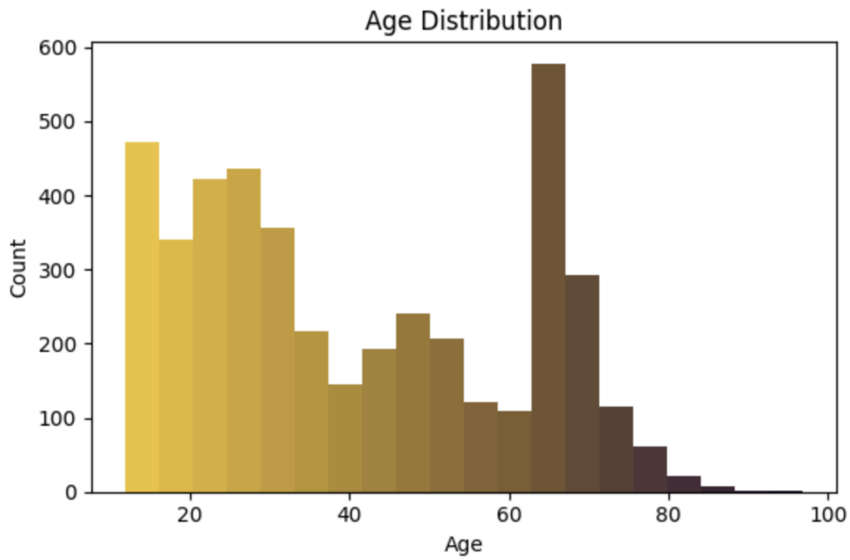
First stage, we analyze the demographic features of survey takers.

1. Gender



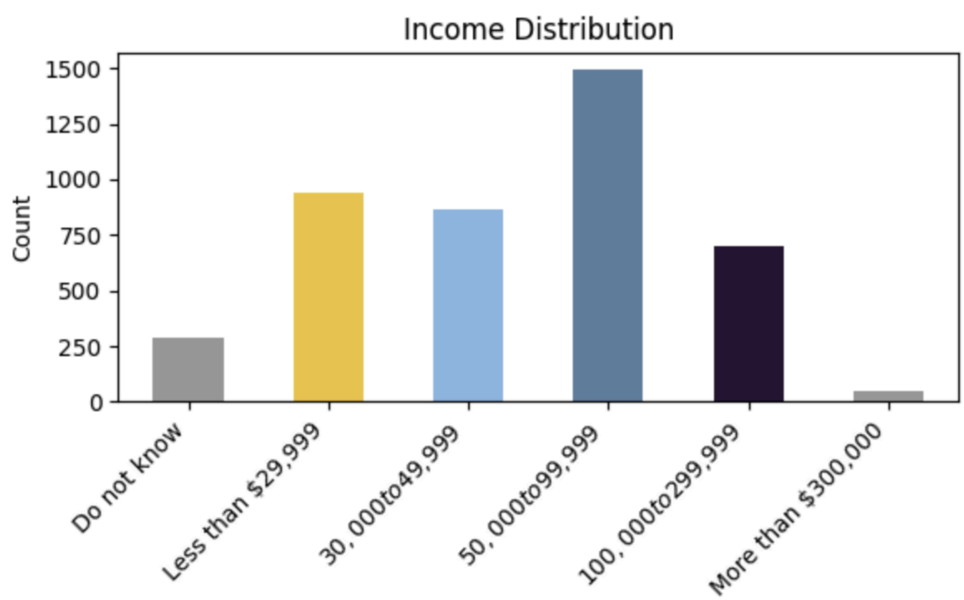
Our dataset presents a balanced distribution across genders, with roughly equal counts for both female and male participants. This parity allows for a more representative analysis of consumer preferences across different genders.

2. Age

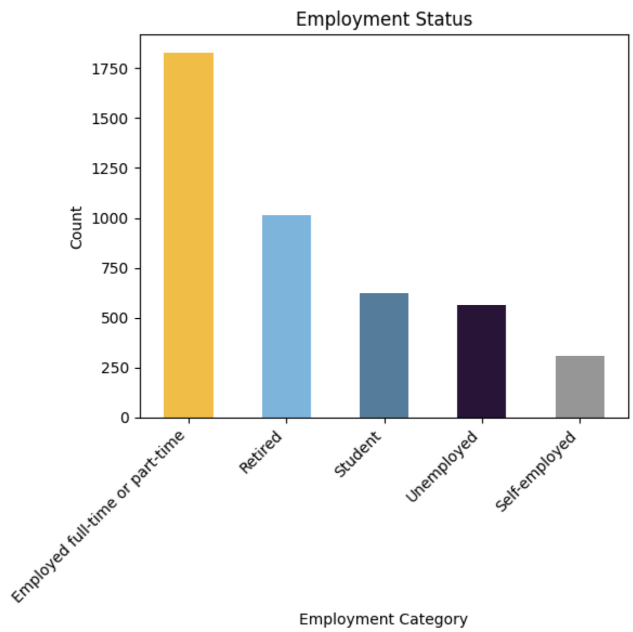


A notable observation within the age distribution is the significant peak in the 65-year-old group. This suggests a substantial representation of individuals who are likely to be recently retired, possibly influencing their consumption patterns and product preferences.

3. Financial Status



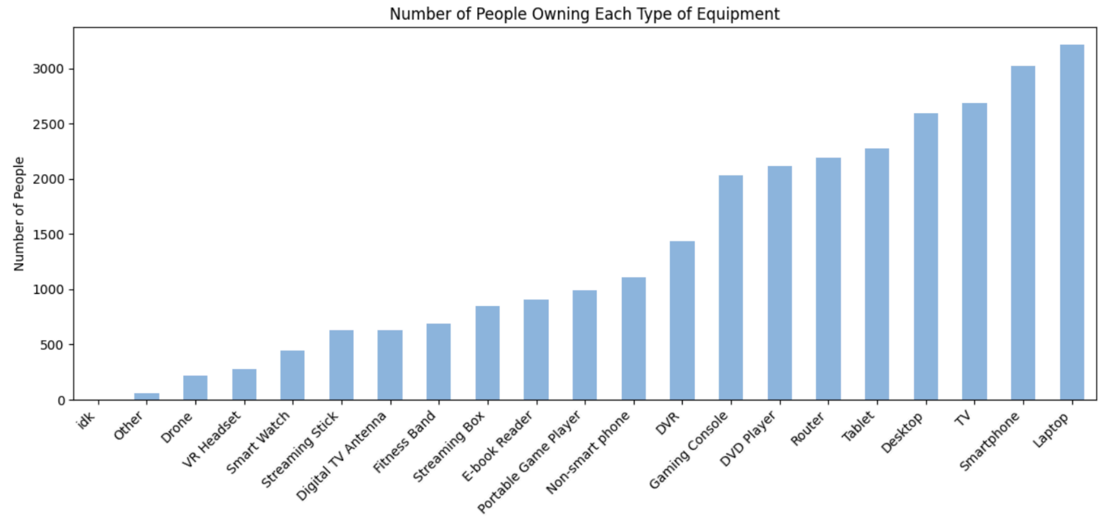
Income levels among our participants predominantly range from $50,000 to $100,000 annually. This income bracket represents the majority, indicating a middle-class demographic with potential spending power for consumer electronics.

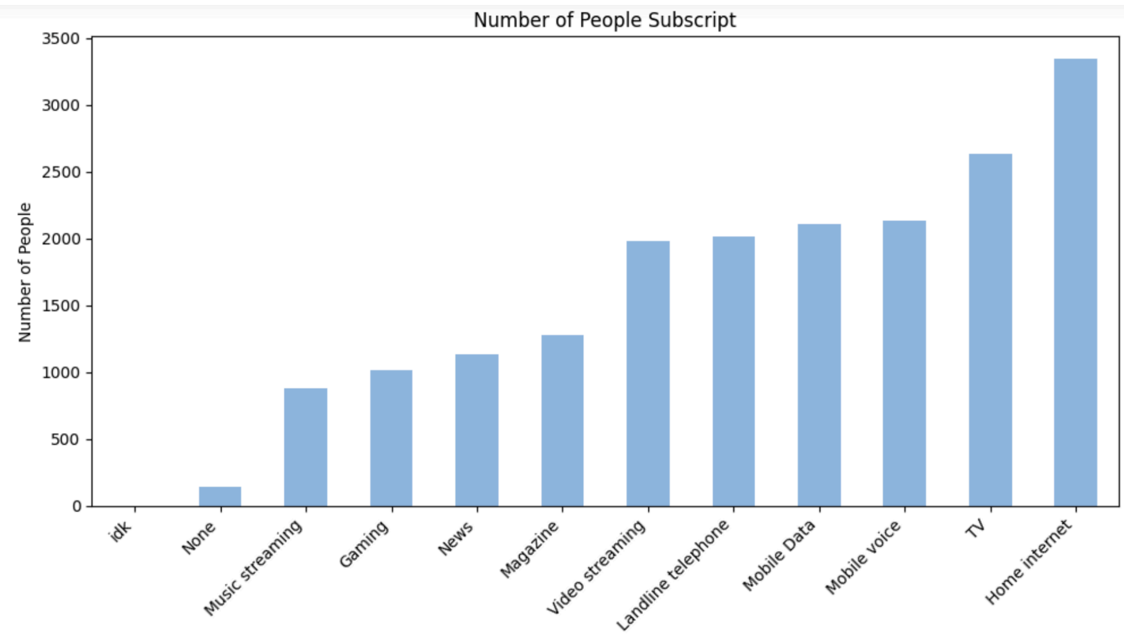


When considering employment status, the data shows that the majority of participants are employed. This status is crucial as it typically correlates with disposable income, which in turn may affect purchasing decisions for entertainment equipment.

Next stage, we explore product popularity and satisfaction.

1. Product Popularity

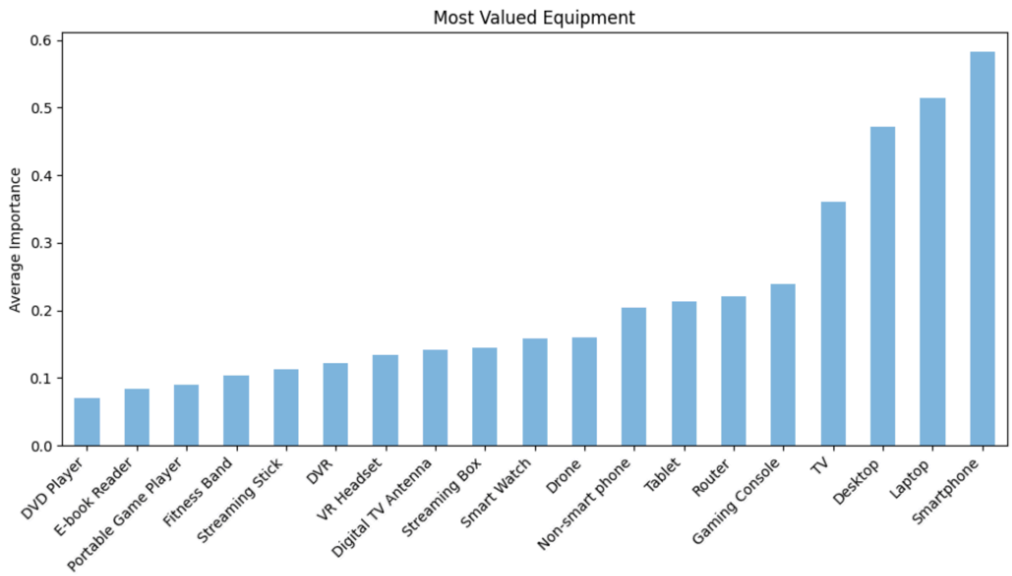


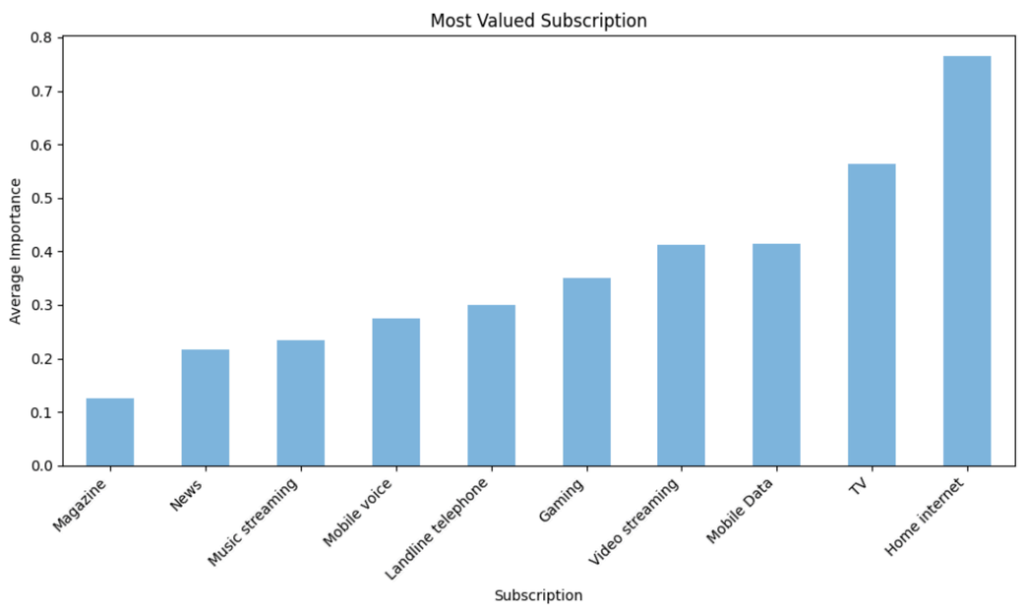


Our survey focus is on entertainment equipment, with laptops, smartphones, and TVs being the top three products owned by the participants. These devices are integral to modern entertainment consumption and reflect the importance of technology in daily life.

Subscription services are also a significant aspect of entertainment, with home internet, TV, and video streaming subscriptions being the most common. The prevalence of these services underscores the shift towards digital and on-demand content consumption. From both graphs, we can see that TV plays an important role in people’s lives.

1. Importance of Owned Products





To gauge the value participants place on their products, we analyzed the importance ratings assigned to each item. By adjusting these ratings for the quantity of owned items, we derive an average importance score for each product.

Interestingly, the importance ranking of TVs decreased when considering this average importance score, falling below that of desktop computers. This suggests that while TVs are widely owned, desktops are regarded as more essential by those who own them.

Conclusion

In summary, our exploratory data analysis reveals a consumer base with balanced gender representation, a significant portion in the retirement age bracket, and a middle-class income profile. The majority are employed, indicating potential purchasing power.

Laptops, smartphones, and TVs are the most commonly owned entertainment devices, while home internet and streaming services are the top subscriptions. The analysis of product importance indicates a higher value placed on desktops over TVs, suggesting a shift in perceived importance amongst these consumers. Although our analysis indicates a higher perceived importance for desktops compared to TVs, it is important to note that televisions still represent a substantial portion of the entertainment market with opportunities for further development.

6. Data Cleaning

The data-cleaning process was critical in preparing the dataset for sophisticated analyses. Our approach was thorough and included several key steps:

Initial Dataset Examination: We started by scrutinizing the dataset to grasp its structure, identify the types of variables present, and detect any clear issues such as missing values or inconsistent entries.

Renaming Columns for Clarity: Many columns were initially named with direct survey questions or cryptic codes, which were not straightforward. We systematically renamed these columns to more accurately reflect the data they contained, facilitating easier navigation and comprehension of the dataset.

Addressing Missing Data: Our dataset had a significant amount of missing data across various variables. This is caused by highly related columns of questions, where participants were only asked about equipment, subscription, and corresponding rankings of what they already obtained, leaving the rest to be Null. We employed a strategic approach to address this, mainly representing all Null to be 0 in binary columns and categorical columns that are to be converted to numerics as well, ensuring the integrity of our dataset was maintained.

Correcting Data Types: Several variables were identified as being in incorrect formats, such as numerical responses stored as strings. We converted these to the correct data types, enabling proper numerical analysis and ensuring categorical variables were appropriately encoded for analytical models.

Outlier Identification and Treatment: Outliers can significantly affect the results of data analysis. We used the Interquartile Range (IQR) method to take a look at the distribution of data in raw numeric columns(referred to EDA part, though not moved any rows), thus reducing the possibility of their influence on our analysis without losing valuable data.

Category Consolidation: For categorical variables with numerous but sparsely populated categories, we consolidated similar categories to simplify our analysis and enhance the statistical power of our models. For data scale harmonization, we made sure that converted numeric values still fall within the interval of 0 to 1 based on their mathematical meanings.

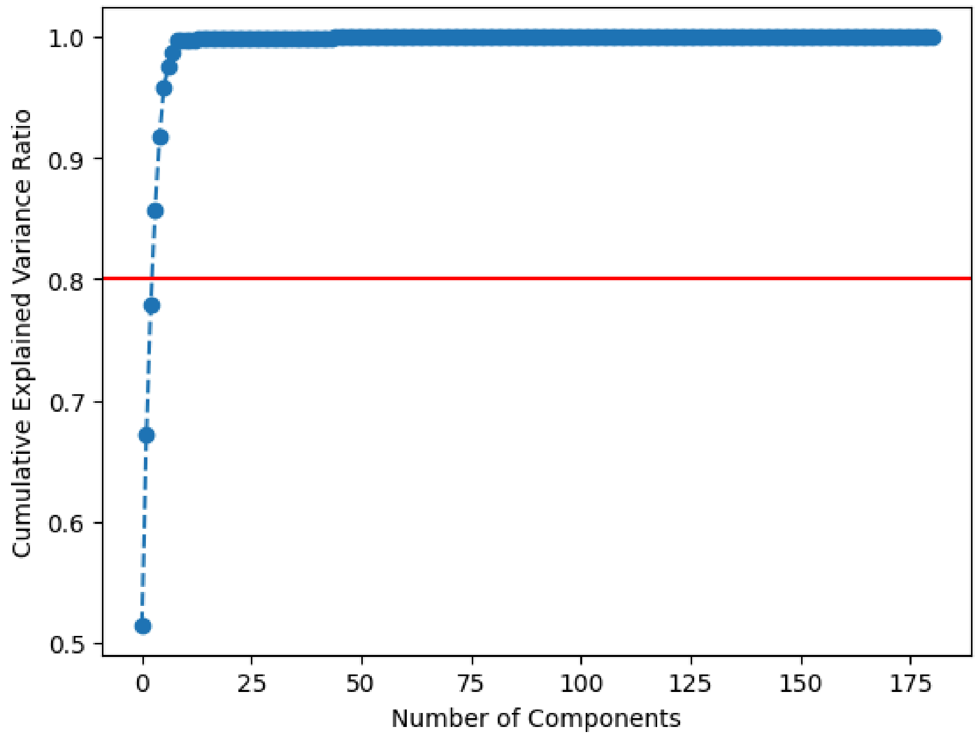
Variable Transformation: Certain variables underwent transformations to normalize their distributions or to better suit the analytical methods applied later. For example, log transformations were applied to variables with highly skewed distributions to approximate normality.

7. Modeling approach

Our modeling approach was designed to uncover latent structures and insights within the data. The steps included:

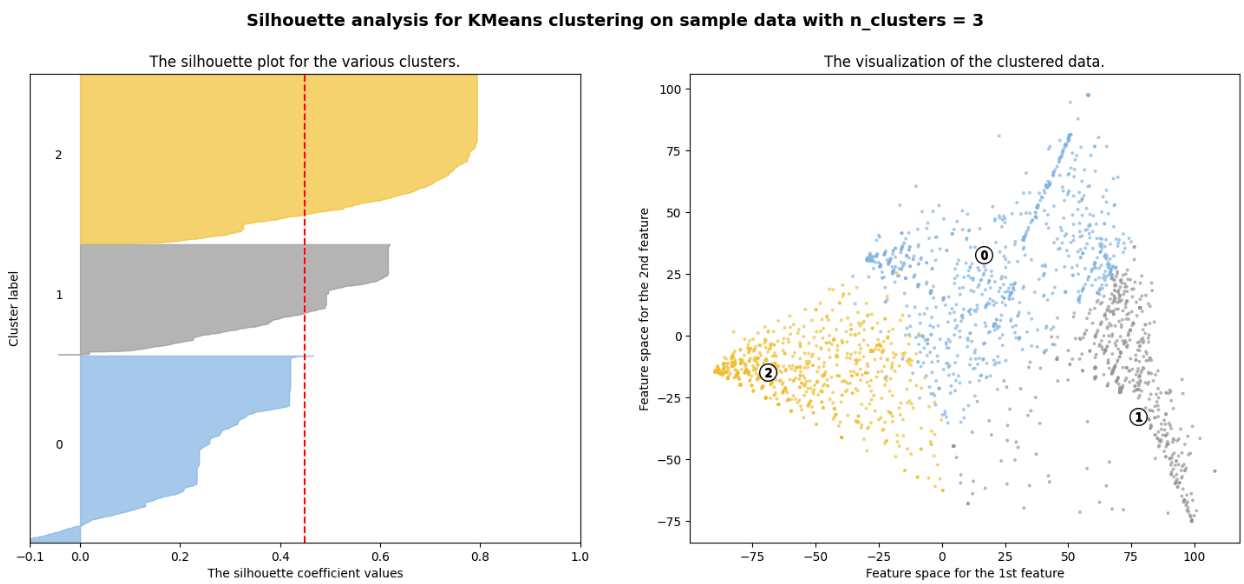
Feature Selection: With a diverse range of variables available, we applied both domain expertise and statistical techniques (like correlation analysis) to identify the most impactful variables for our models, ensuring a focus on the most relevant information.

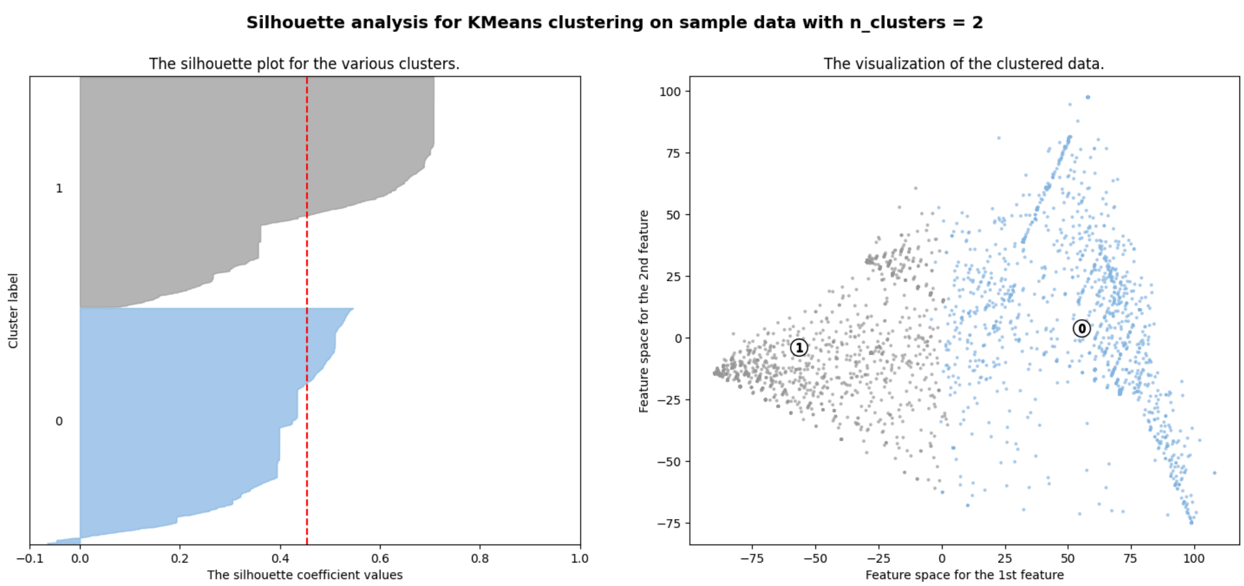
Principal Component Analysis (PCA): PCA was used to reduce the dimensionality of our data as we have nearly 200 columns. After standardizing the variables (mainly age number at this point), we ran PCA and chose 4 components, as they accounted for a significant enough proportion of around 80% of the variance in the dataset. This decision was based on the scree plot analysis, where we observed that the addition of a fifth component provided diminishing returns in explained variance but not necessarily. The chosen components captured essential underlying patterns and allowed for a more manageable, yet comprehensive, analysis.

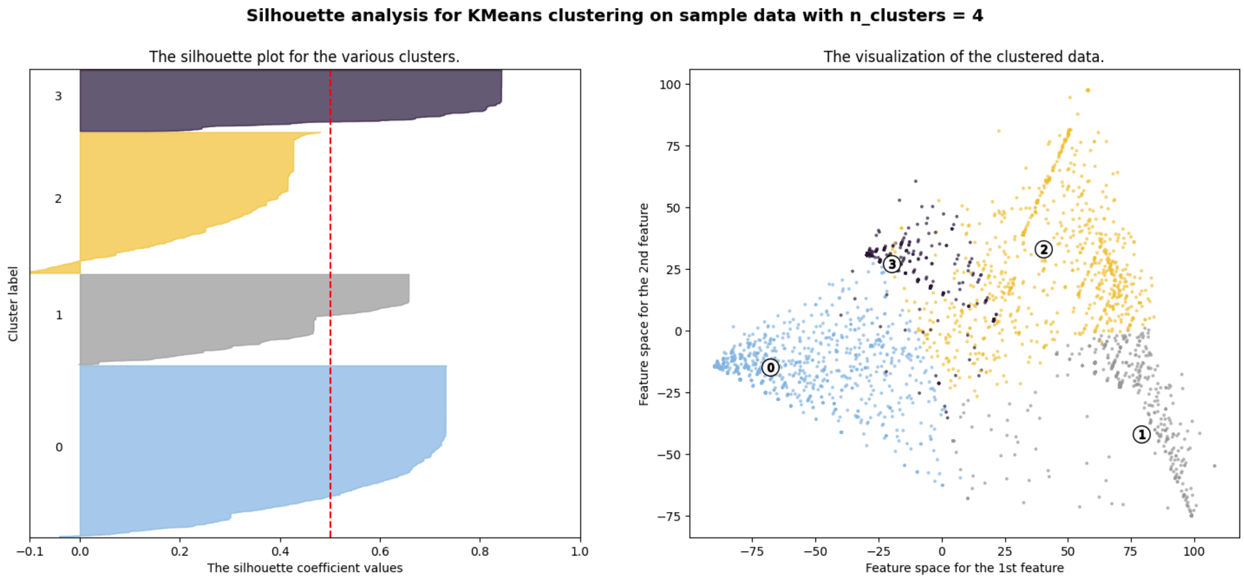


K-Means Clustering with Silhouette Analysis:

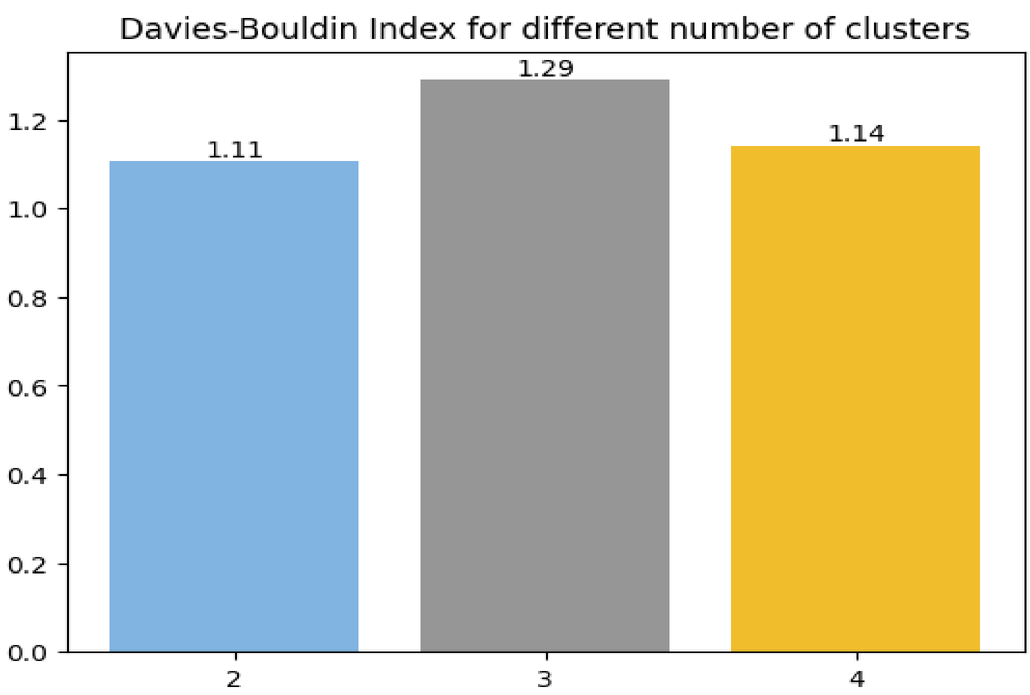
To segment the dataset, we applied K-means clustering. The optimal number of clusters was determined through silhouette analysis, which measures how similar an object is to its cluster compared to other clusters. We used the elbow method, tested a range of cluster numbers, and selected 2 as the optimal number, as it yielded a proper average silhouette score, indicating a clear and meaningful separation between the clusters. When we chose, though 4 clusters seemed to be slightly better in the silhouette score, it yielded a small, potentially unreasonable 4th cluster. This is because this cluster might overlap significantly with others, or the samples within the big 2nd cluster are not closely packed enough.

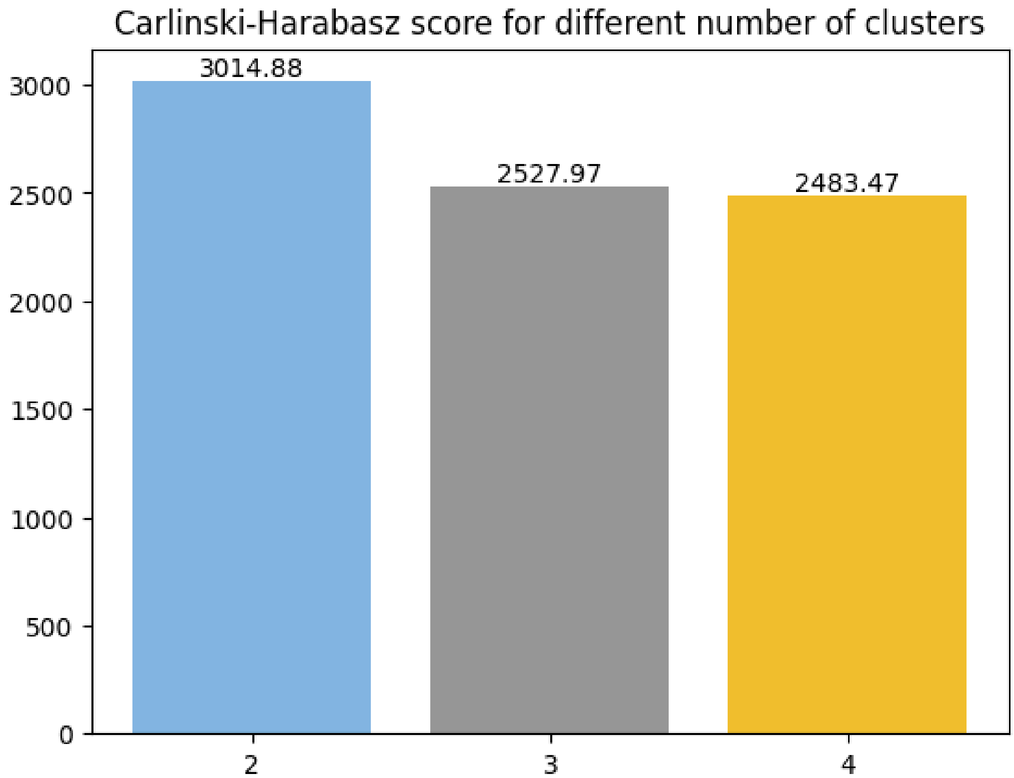






Other metrics: Then we checked the Davies-Bouldin Index and Calinski-Harabasz Score. The DBI is a measure of how well clustering has been done, which is based on a ratio between within-cluster distances and between-cluster distances. A lower DBI value indicates better clustering because it signifies that the clusters are compact (low within-cluster distances) and well-separated (high between-cluster distances). The CH Index is calculated by the ratio of the sum of between-cluster dispersion and within-cluster dispersion for all clusters. A higher CH Index indicates better clustering performance, as it signifies that the clusters are well separated and compact. Among these two results of different numbers of clusters, the number of 2 has the best overall result. Therefore, we decided to go with 2.





Cluster Profiling: We then delved into each cluster to understand its defining characteristics, analyzing central tendencies and distributions of key variables. This profiling was visualized using various plots to illustrate the distinct nature of each cluster.

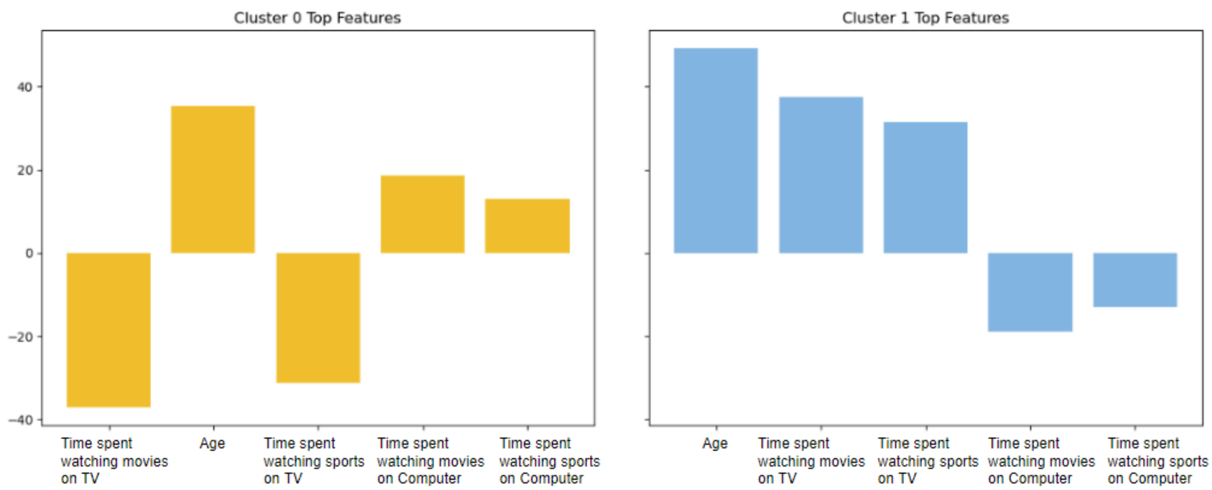
Model Validation: To ensure the robustness of our clustering, we conducted additional validation techniques, including a thorough silhouette analysis for each cluster. This step confirmed that our clusters were well-defined and distinct from each other.

Insight Extraction: The final stage involved interpreting the clusters in the context of our objectives. We explored the implications of the clustering, drawing insights from the unique characteristics of each group and relating them to our research questions.

This detailed approach allowed us to maintain the integrity of our dataset and apply rigorous analytical techniques. As a result, we were able to derive nuanced and actionable insights from our data.

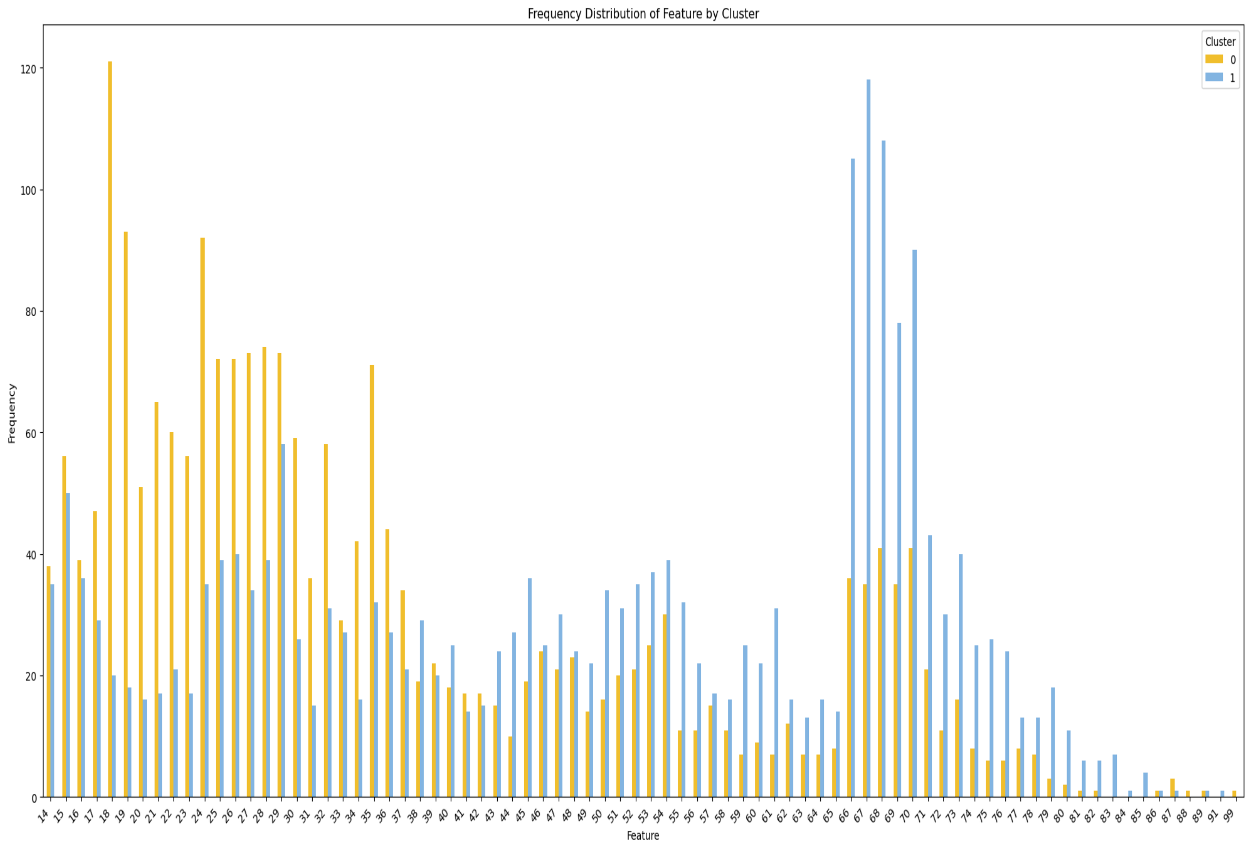
8. Results

Feature Importance: We have looked at the top features for each cluster based on the largest mean differences from the overall mean of the dataset. Larger differences indicate that the feature is more influential in defining the uniqueness of the clusters. Age, % Time spent on watching movies on TV, % Time spent on watching sports on TV, % Time spent on watching movies on Computers, % Time spent on watching movies on computers emerged as important features. While our analysis extended to other viewing platforms such as tablets and phones, these devices showed much smaller differences in usage between the clusters. This indicates that, although tablets and phones are used for media consumption, they do not distinguish between our clusters as significantly as the top five features.



Building on the insights garnered from these distinct features, we delved deeper into our data for a more nuanced understanding. We examined key statistics that were instrumental in shaping our targeted engagement strategy. This analysis was tailored to ensure that our outreach and content are aligned with the specific preferences and behaviors of our audience segments.

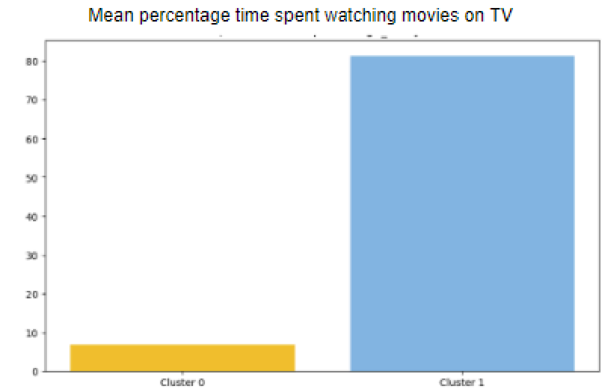
Firstly, in examining the frequency distribution of age within our clusters each cluster has peaks at certain ages where the frequency of viewers is highest. For Cluster 0, the age distribution peaks within the 20-30-year range, indicating a concentration of younger viewers in this segment. Conversely, Cluster 1 displays a predominant age range around 65 years, highlighting a mature audience base.

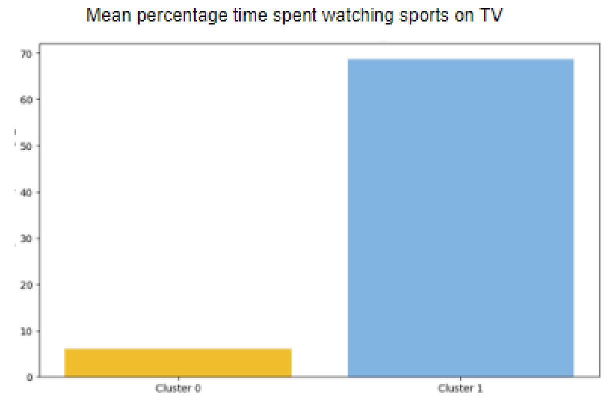


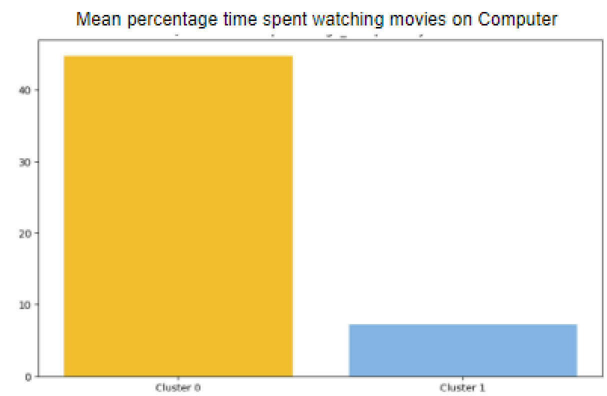
The below charts display the mean percentage of time spent by both clusters on watching movies and sports, both on TV and on a computer.

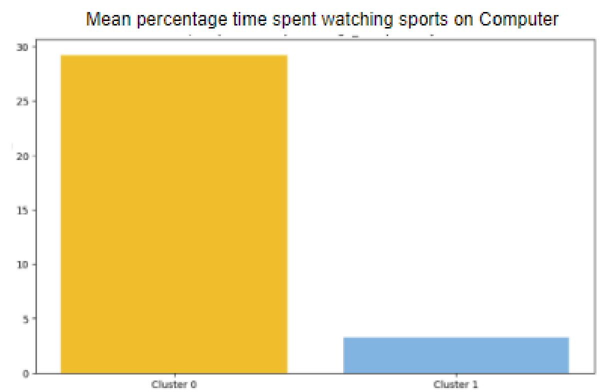
Cluster 0 in yellow seems to be characterized by younger, more tech-savvy individuals who prefer consuming media content on the computer.

Cluster 1, on the other hand, seems to represent an older or more traditional group that prefers watching content on TV. Their engagement with digital platforms for watching movies or sports is considerably lower than Cluster 0.

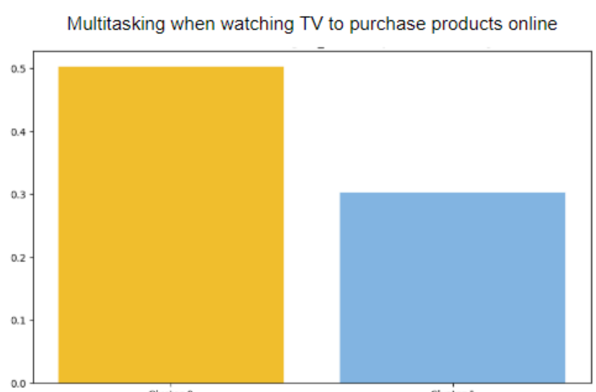


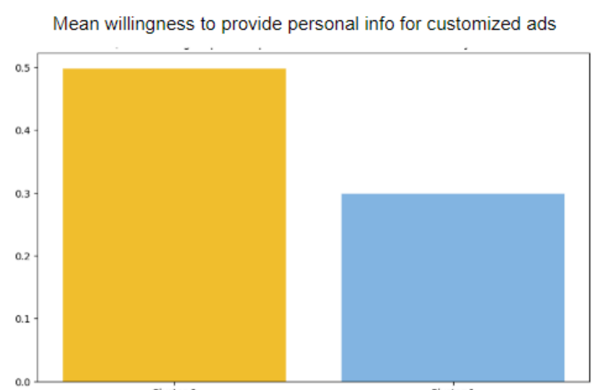






The below chart provides a comparative analysis between the two clusters, specifically describing their behavioral patterns and preferences when it comes to media multitasking and data privacy.





Multitasking when watching TV to purchase products online:

People in Cluster 0 show an inclination towards multitasking, with a notable portion of this group engaging in online shopping while watching TV. This trend suggests a

dynamic viewing habit where television acts not only as a source of entertainment but also as a prompt for online shopping.
Conversely, those in cluster 1 demonstrate a more traditional approach to TV viewing, with a considerably lower likelihood of combining watching TV with online purchasing. This indicates a preference for focused viewing.

Mean willingness to provide personal information for customized ads:

Customers in Cluster 0 show a higher propensity towards sharing personal information to receive customized advertising content. This willingness points towards a group that values personalized experiences and is comfortable with the trade-offs involved in data sharing.
In contrast, those in Cluster 1 exhibit more caution, showing a lower average willingness to share personal information for customized ads. This reflects a more guarded attitude towards privacy and a potential skepticism towards personalized marketing techniques.

Finally, as expected the older group has 150% more retired people and 8% more annual income compared to the younger group.

9. Recommendations

Following our data analysis, we have crafted specialized viewer engagement strategies tailored to align with the distinct preferences of both the younger and more mature audience segments.

Strategies for the Younger Audience Segment:

1. Enhanced Computer Viewing Experience:

   Develop content with interactive elements, such as series that allow viewers to choose their narrative paths, leveraging the interactive capabilities of computers. Introduce second-screen applications or websites to supplement live events, enabling simultaneous interaction and shopping. This also addresses the current trend where viewers primarily engage with movie and sports content, hence we recommend expanding our offerings to include serial content that encourages longer viewing sessions.

2. Financially Incentivized Content Access:

We recommend providing subscription bundles or discounts to present exceptional value. Offering a range of payment options and introducing trial periods to reduce barriers to subscription is another strategy we propose to attract this segment.

3. Personalized Privacy-Considerate Promotions:

   We recommend utilizing targeted online advertising by leveraging user data, while respecting privacy, to serve relevant ads during content streaming. Additionally, leveraging multiple digital platforms for campaigns, such as social media, email, and content marketing on popular websites and streaming services they frequent is proposed. Furthermore, we suggest employing interactive advertising that allows engagement or quick purchases while multitasking, such as shoppable videos or ads with polls and quizzes.

Strategies for the Mature Audience Segment:

1. Optimized Television Viewing Experience:

   We advise concentrating on delivering high-definition or 4K content suitable for larger TV screens to underscore the quality of the viewing experience. Moreover, we propose offering a simplified user interface for smart TVs or set-top boxes, thereby enabling older users to easily navigate and discover content.

2. Engagement for the Affluent Viewer:

   We suggest emphasizing exclusivity, premium quality, and any senior discounts to attract higher-income consumers. Furthermore, hosting special events such as virtual meet-and-greets with actors, and directors, or offering behind-the-scenes glimpses of popular shows could be a great way to engage this group. Also, we propose offering a dedicated customer support line for immediate assistance to address technical issues or content inquiries, which could extend to in-home tech support for products purchased by consumers.

3. Data Privacy and Traditional Advertising:

   We recommend strategizing on Traditional TV Advertising by investing in commercials during programs favored by this demographic. Additionally, advertisement strategies like Direct Response TV (DRTV) enabling viewers to directly purchase products from commercials could be a great way to leverage their higher disposable income. Also, ensuring Privacy Assurance by clearly communicating your company's dedication to privacy and data security in all marketing materials, thereby encouraging them to engage with our content confidently, with the assurance that their privacy is a foremost concern for our company is vital.

10. Conclusion

As the Product Manager of YouTube TV, I am proud to present our strategic approach, shaped by rigorous data analysis, to address the critical questions of targeting and engagement in our market entry strategy. Our journey began with meticulous data cleaning and exploratory data analysis (EDA), followed by Principal Component Analysis (PCA) and clustering to deeply understand our audience and segment them effectively.

Our analysis revealed key demographics: a balanced gender distribution, a notable retirement age group, and a middle-class income bracket.

After PCA and clustering, we separate our survey takers into two clusters. This has guided us to focus on two primary segments: the younger, tech-savvy audience and the more mature viewers.

For the former, we are introducing interactive content and multi-screen experiences, supplemented with value-oriented packages and flexible payment options. For the latter, we emphasize high-quality broadcasts and a simplified user interface, ensuring a premium viewing experience.

Our targeted strategies extend to advertising and customer service, aiming to meet the unique needs and preferences of each segment. For younger viewers, we leverage online and interactive ads, while for mature audiences, we focus on quality content and privacy assurances.

This strategic blueprint is not merely a plan but a commitment to adapt and evolve with our consumers' needs, ensuring YouTube TV offers an unmatched streaming experience. This marks the beginning of a journey where YouTube TV isn't just an option but the preferred choice for diverse viewers seeking exceptional value.

11. References

- Contributions
    - Data cleaning: Terry Wei
    - EDA: Hsin-Li (Cindy) Kan, Yvonne Hsu, Gautam Devadiga
    - Clustering: Terry Wei, Gautam Devadiga
    - Problem brainstorming & Recommendations: All of us
    - Presentation slides development: All of us
    - Report writing: All of us

- Citations

Boorstin, J. (2017, April 5). *YouTube officially launches YouTube TV in select markets*. CNBC. Retrieved February 29, 2024, from https://www.cnbc.com/2017/04/05/youtube-officially-launches-youtube-tv-in-select-markets.html

Deloitte. (n.d.). *Digital Democracy Survey*. Deloitte. Retrieved February 29, 2024, from https://www2.deloitte.com/xe/en/pages/technology-media-and-telecommunications/articles/gx-digital-democracy-survey-generational-media-consumption-trends.html