

COMBINING USER PROFILING AND SEGMENTATION, AND TEXT ANALYSIS FOR TARGETED AD CAMPAIGNS

Introduction

This project delves into the power of combining user profiling, segmentation, and text analysis to extract valuable customer insights from user data. The project utilizes a sample dataset containing user profiles and their online behavior, including potential text data sources like:

- **Top Interests:** text associated with primary interests of the user.

Goals

- **Develop user profiles** that capture key demographic and online behavior characteristics.
- **Segment users** into distinct groups based on shared characteristics using clustering techniques.
- **Analyze user-generated text data** associated with different user segments to understand their preferences, opinions, and pain points.
- **Identify patterns and trends** within user segments using text analysis techniques.
- **Leverage user segmentation and text analysis insights** to develop targeted marketing campaigns that achieve:
 - **Increased Relevancy:** Content and promotions become more relevant to users' specific needs and interests, leading to higher engagement.
 - **Improved Conversion Rates:** Tailored messaging resonates better with users, potentially increasing conversion rates for desired actions.
 - **Enhanced User Experience:** By addressing the preferences of various user segments, the overall user experience improves.

Data Description

The project utilizes a sample dataset containing user profiles with the features:

- User ID (unique identifier)
- Demographics (Age, Gender, Location, Language, Education Level)
- Online Behavior (Likes and Reactions, Followed Accounts, Device Usage, Time Spent Online, Click-Through Rates, Conversion Rates, Ad Interaction Time)
- Income Level
- Top Interests (text data)

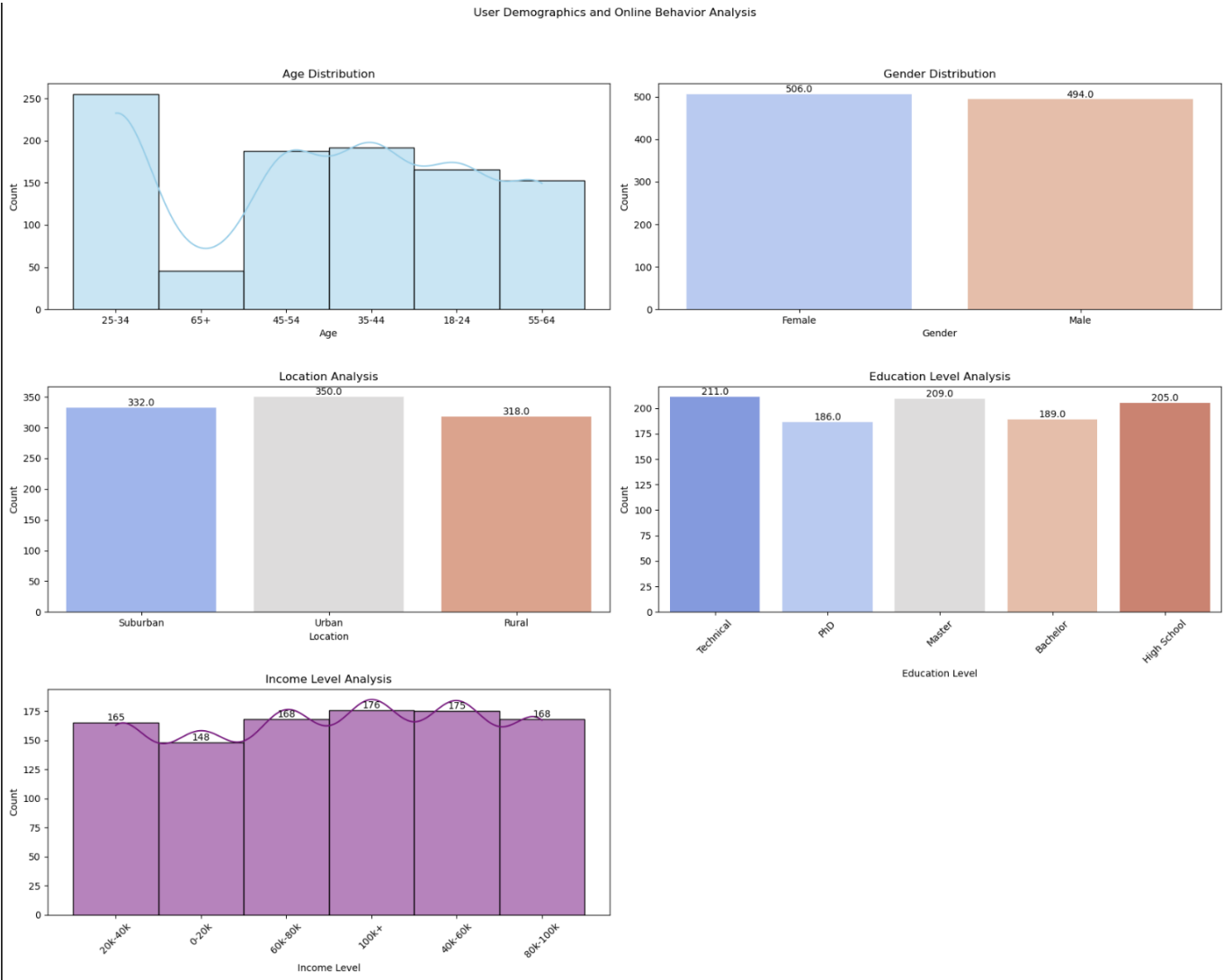
Below is a comprehensive overview of each column present in the dataset.

- **User ID:** A unique identifier assigned to each user.
- **Age:** The age group of the user, categorized into ranges.
- **Gender:** The gender identity of the user.
- **Location:** Describes the type of location where the user resides, categorized as Urban, Suburban, or Rural.

- **Language:** Indicates the primary language spoken by the user.
- **Education Level:** Reflects the highest level of education attained by the user.
- **Likes and Reactions:** Represents the total count of likes and reactions expressed by the user.
- **Followed Accounts:** Indicates the number of accounts that the user follows.
- **Device Usage:** Specifies the primary device utilized by the user to access the platform, including options such as Mobile, Desktop, or Tablet.
- **Time Spent Online (hrs/weekday):** The average duration spent by the user online during weekdays, measured in hours.
- **Time Spent Online (hrs/weekend):** The average duration spent by the user online during weekends, measured in hours.
- **Click-Through Rates (CTR):** Represents the percentage of ad impressions that result in user clicks.
- **Conversion Rates:** Indicates the percentage of clicks on ads that lead to desired conversions or actions.
- **Ad Interaction Time (sec):** The average duration of user interaction with ads, measured in seconds.
- **Income Level:** Reflects the income bracket or level of the user.
- **Top Interests:** Highlights the primary areas of interest or hobbies of the user.

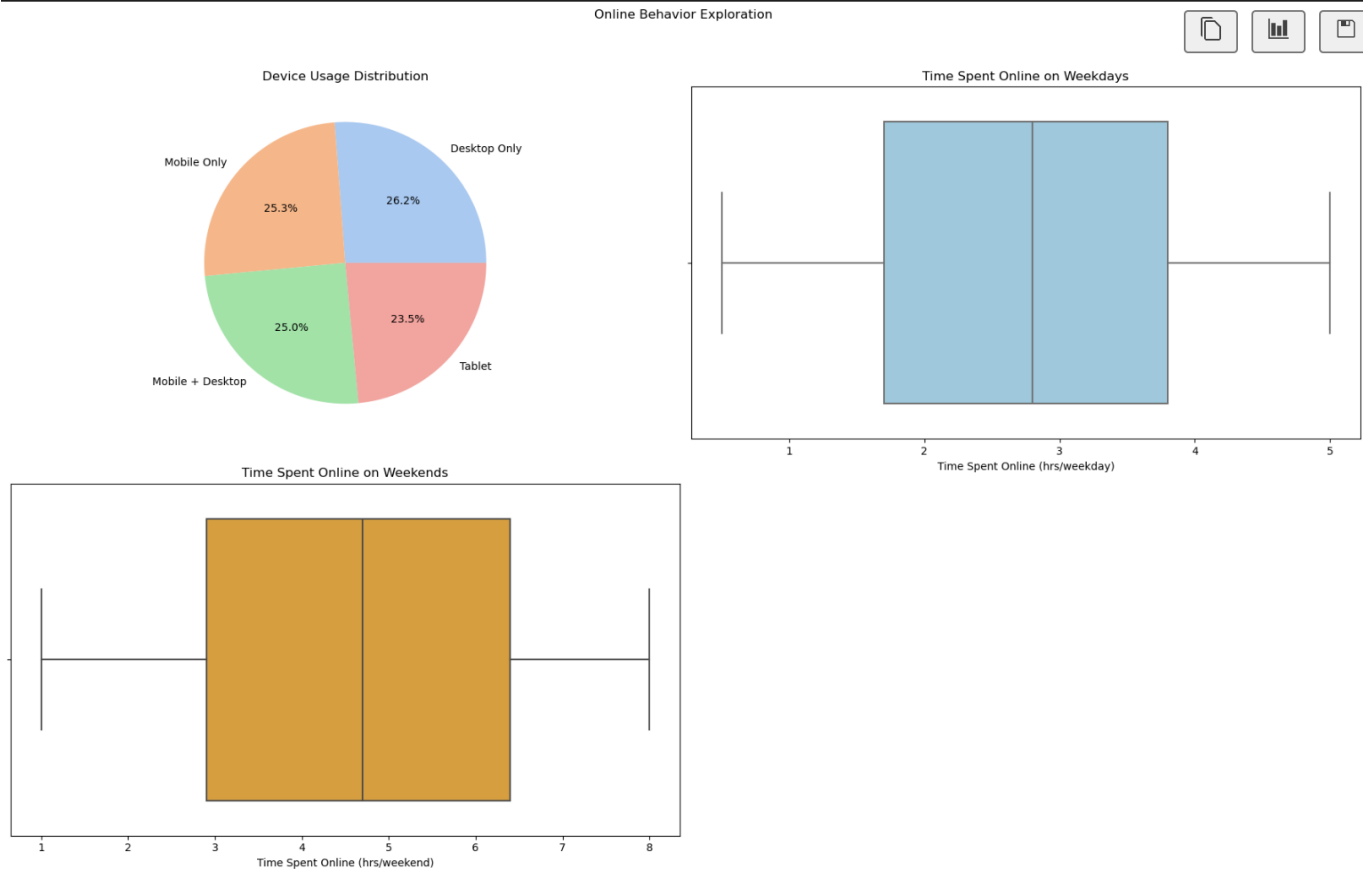
EDA

User Demographics



- **Age:** The user base skews slightly younger, with the largest group falling within the 25-34 age range (255 users). This is followed by the 35-44 (192 users) and 45-54 (188 users) age groups. Teenagers (18-24) and users above 55 are less represented (166 and 153 users respectively).
- **Gender:** The gender distribution is relatively balanced, with a slight majority of females (506) compared to males (494).
- **Location:** Geographically, the user base seems evenly distributed across urban (350 users), suburban (332 users), and rural (318 users) areas.
- **Education Level:** A significant portion of users hold advanced degrees (Master's - 209 users, PhD - 186 users). Technical degrees (211 users) and Bachelor's degrees (189 users) are also well-represented. High school graduates (205 users) form the remaining segment.
- **Income Level:** The income distribution suggests a user base with moderate to high purchasing power. The largest group falls within the 40k-60k and 60k-80k income brackets (175 and 168 users respectively). A sizeable portion also earns above 100k annually (176 users). While income levels decrease at the lower and higher ends of the spectrum (20k-40k: 165 users, 0-20k: 148 users), they still hold a presence.

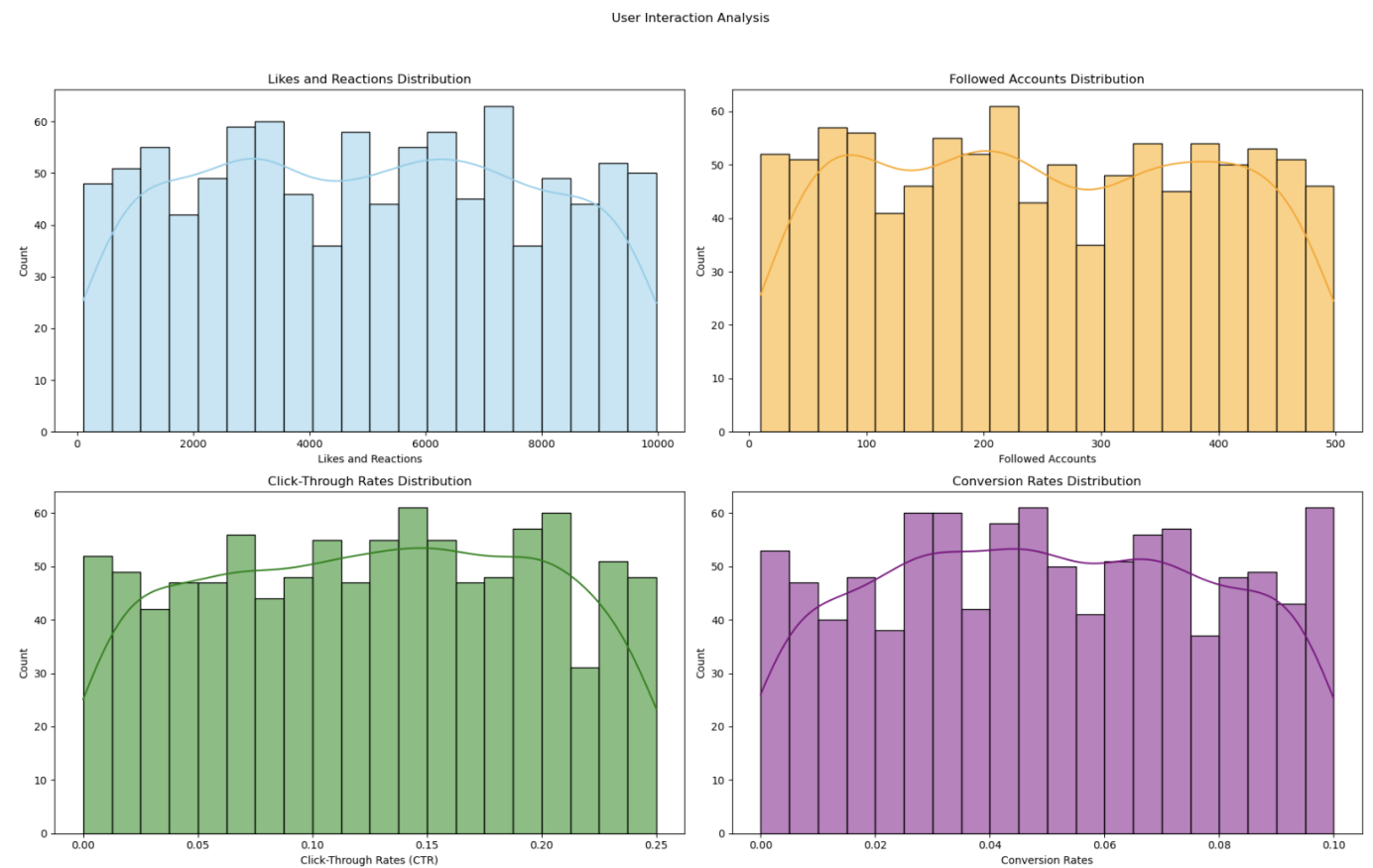
Online Behavior Exploration



- **Device Usage Distribution:** The pie chart illustrates the distribution of users across different device categories. It reveals that the user base is fairly evenly split among desktop-only, mobile-only, and a combination of mobile and desktop devices, each representing approximately a quarter of the total users. Tablet usage is slightly less common, with around 23.5% of users accessing the platform through this device.
- **Time Spent Online Analysis:** The box plots display the distribution of time spent online by users on both weekdays and weekends. For weekdays, the median time spent online is approximately 2.8 hours, with a relatively narrow interquartile range (IQR) from 1.7 to 3.8 hours. On weekends, users

tend to spend more time online, with a median of 4.7 hours and a wider IQR from 2.9 to 6.4 hours. The plots highlight the variability in online behavior between weekdays and weekends, with users generally spending more time online during leisure periods.

User Interaction Analysis



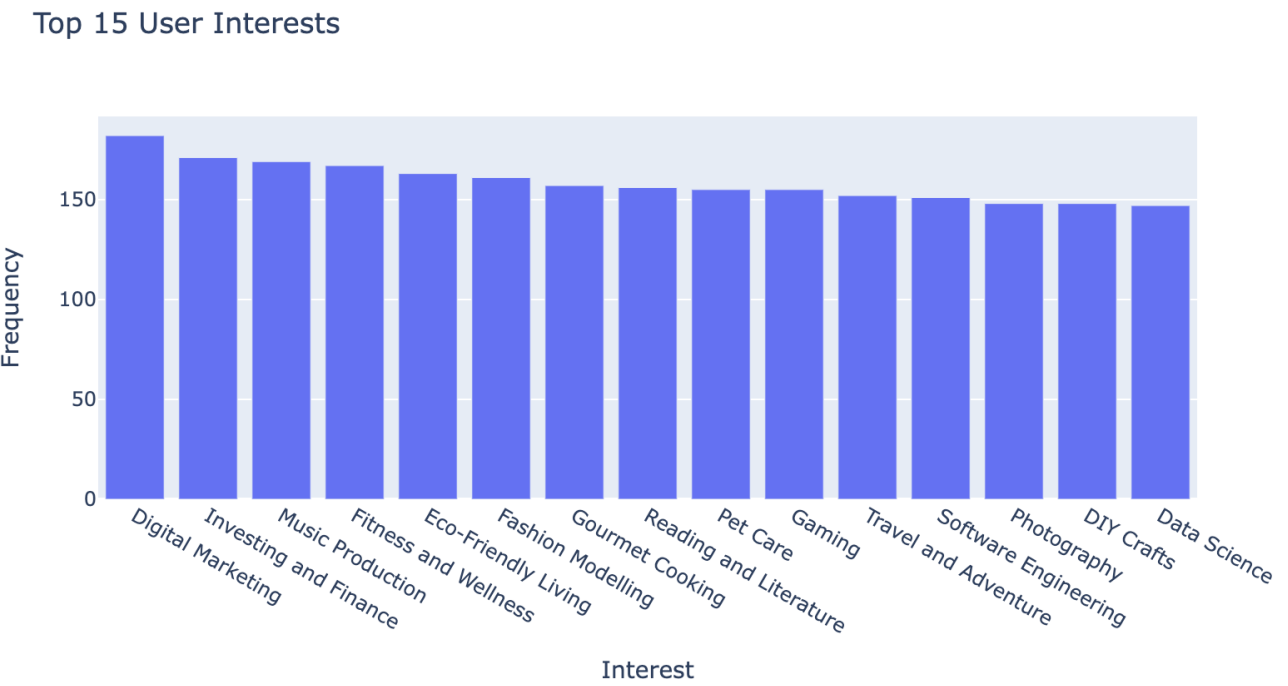
Metric	Mean	Median	Standard Deviation
Likes and Reactions	4997.084	5002.5	2838.494
Followed Accounts	251.438	245.5	141.942
Click-Through Rates (CTR)	0.125	0.128	0.071
Conversion Rates	0.050	0.049	0.029

- **Likes and Reactions:** The distribution of Likes and Reactions appears to be somewhat right-skewed, with a median of 5002.5 and a mean slightly higher (4997.08). This suggests that a larger portion of users generate a moderate to high number of Likes and Reactions, while there might be a smaller group with exceptionally high engagement. The standard deviation (2838.49) indicates a fair amount of variability in user behavior.
- **Followed Accounts:** The distribution of Followed Accounts also seems right-skewed, with a median of 245.5 and a slightly higher mean (251.44). This implies that most users follow a moderate number of accounts, while some follow a significantly higher number. The standard deviation (141.94) reflects some variation in user behavior regarding following accounts.
- **Click-Through Rates (CTR):** The Click-Through Rate (CTR) distribution is likely centered around the median (0.128) with a standard deviation (0.0712). This suggests that a majority of users exhibit click-through behavior within a specific range. However, the presence of a tail on either side of the distribution indicates that some users might have very low or very high CTRs.

- **Conversion Rates:** Similar to CTR, the Conversion Rate distribution is likely centered around the median (0.049) with a standard deviation (0.0287). This implies that most users fall within a specific conversion rate range. However, as with CTR, there might be users with exceptionally low or high conversion rates.

Text Analysis Exploration

Top 15 User Interests



Top 50 User Interests (Word Cloud)



User Profiling and Segmentation for targeted ad campaigns

By leveraging the insights gained from user demographics, online behavior, interaction patterns, and potential interests, we can create user profiles and segment our user base into distinct groups. This approach allows for more targeted ad campaigns, ultimately leading to increased user engagement and conversion rates.

Segmentation Criteria

User segmentation can be based on various criteria, including:

- **Demographics:** Age, gender, income level, education level.
- **Behavioral:** Time spent online, likes and reactions, click-through rates (CTR), conversion rates.
- **Interests:** Analyzing user behavior to identify potential interests and preferences.

Implementing Segmentation

- **Clustering Algorithms:** Techniques like K-Means clustering can be used to group users based on similarities in their attributes.

KMeans Clustering

The K-Means clustering algorithm successfully segmented the user base into five distinct clusters (0-4). Each cluster exhibits unique characteristics based on the chosen features, including age, gender, income level, online behavior (time spent online, likes, reactions), and engagement metrics. The analysis of these characteristics will help gain valuable insights into the preferences and behaviors of each user segment. These insights will serve as the foundation for developing targeted advertising campaigns tailored to resonate effectively with each segment.

Identifying the commonalities and differences within each cluster

To identify the commonalities and differences within each cluster, the mean and mode values of features for each cluster is computed.

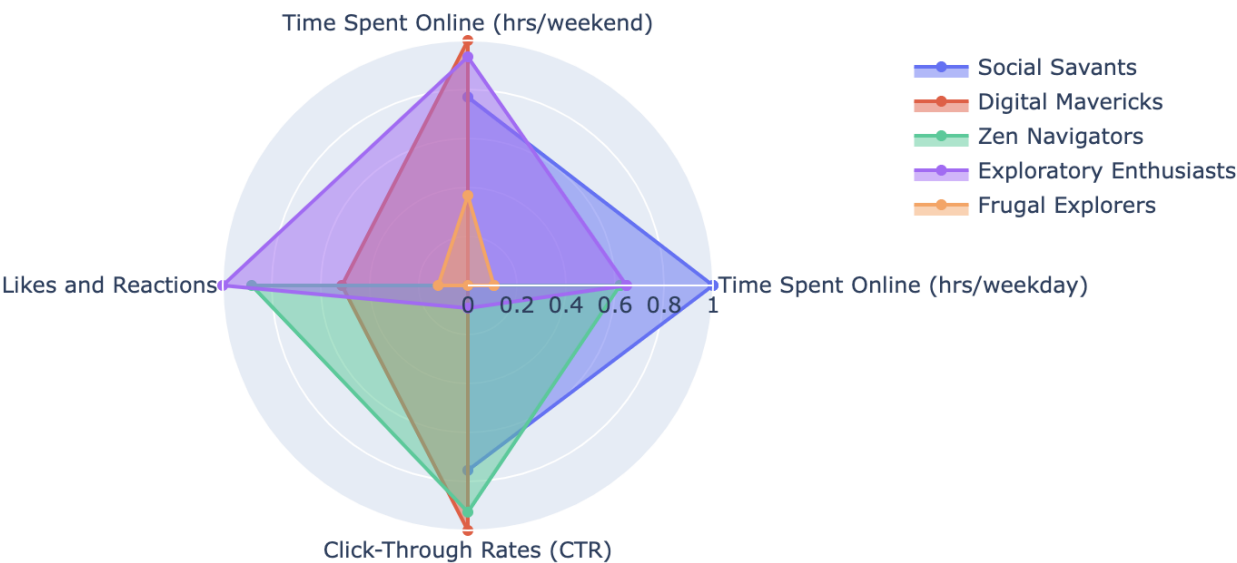
Cluster	Time Spent Online (hrs/weekday)	Time Spent Online (hrs/weekend)	Likes and Reactions	Click-Through Rates (CTR)	Age	Gender	Income Level
0	3.911111	5.212963	2409.620370	0.149588	25-34	Female	80k-100k
1	1.559394	6.002424	5005.121212	0.179836	35-44	Male	80k-100k
2	3.019737	2.584211	6861.587719	0.170614	25-34	Male	20k-40k
3	3.080882	5.774510	7457.602941	0.067971	25-34	Female	100k+
4	1.809626	3.839572	3021.219251	0.056594	45-54	Female	0-20k

Average behavior and characteristics of users within each cluster

- **Cluster 0 - Social Savants:** Users in this cluster, predominantly females aged 25-34 with an income of 80k-100k, spend approximately 3.91 hours on weekdays and 5.21 hours on weekends online. They have moderate levels of likes and reactions and a relatively higher click-through rate (CTR).
- **Cluster 1 - Digital Mavericks:** Mostly males aged 35-44 with an income of 80k-100k, users here spend less time on weekdays online but have longer weekend activity. They show higher levels of likes and reactions and a slightly higher CTR.
- **Cluster 2 - Zen Navigators:** With a majority of males aged 25-34 and an income of 20k-40k, users in this cluster spend a moderate amount of time online on both weekdays and weekends. They have the highest levels of likes and reactions but a relatively lower CTR.
- **Cluster 3 - Exploratory Enthusiasts:** Predominantly females aged 25-34 with a higher income of 100k+, users in this cluster spend a similar amount of time online as Cluster 2 but have higher levels of likes and reactions. However, they have a lower CTR.
- **Cluster 4 - Frugal Explorers:** Users here, mostly females aged 45-54 with an income of 0-20k, spend the least amount of time online on both weekdays and weekends. They have lower levels of likes and reactions and the lowest CTR among all clusters.

Reflecting User Segments and Their Distinctive Characteristics via Radar Charts

User Segments Profile



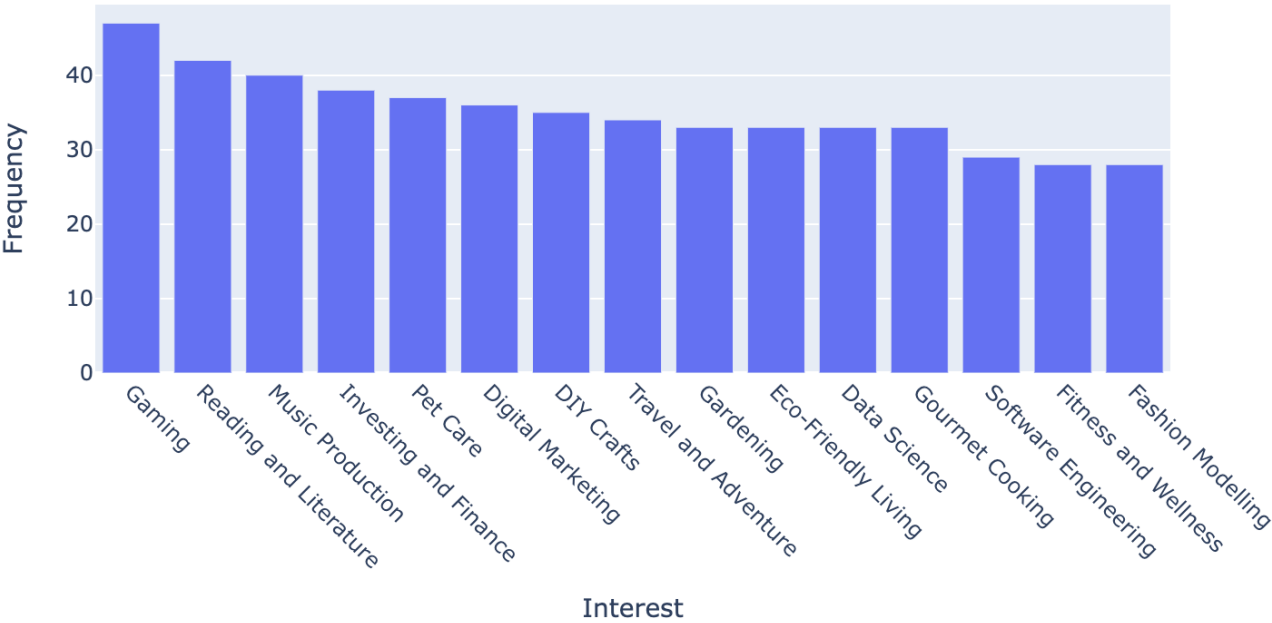
The radar chart provides a comprehensive view of the distinct behavioral patterns exhibited by different user segments. By analyzing key metrics such as time spent online, likes and reactions, and click-through rates across various clusters, marketers gain valuable insights into the preferences and tendencies of each segment. This visualization enables marketers to tailor their advertising campaigns more effectively, aligning content and timing with the unique preferences of each user segment. For instance, campaigns targeting 'Digital Mavericks' may benefit from a data-driven approach, leveraging insights into their high engagement levels and specific interests to optimize ad placement and content. Similarly, understanding the preferences of 'Zen Navigators' can guide marketers in crafting messages that resonate with their

preference for quality interactions and meaningful content. By leveraging these insights, marketers can create more targeted and impactful ad campaigns that drive engagement and conversion rates.

Top Interest Within Clusters

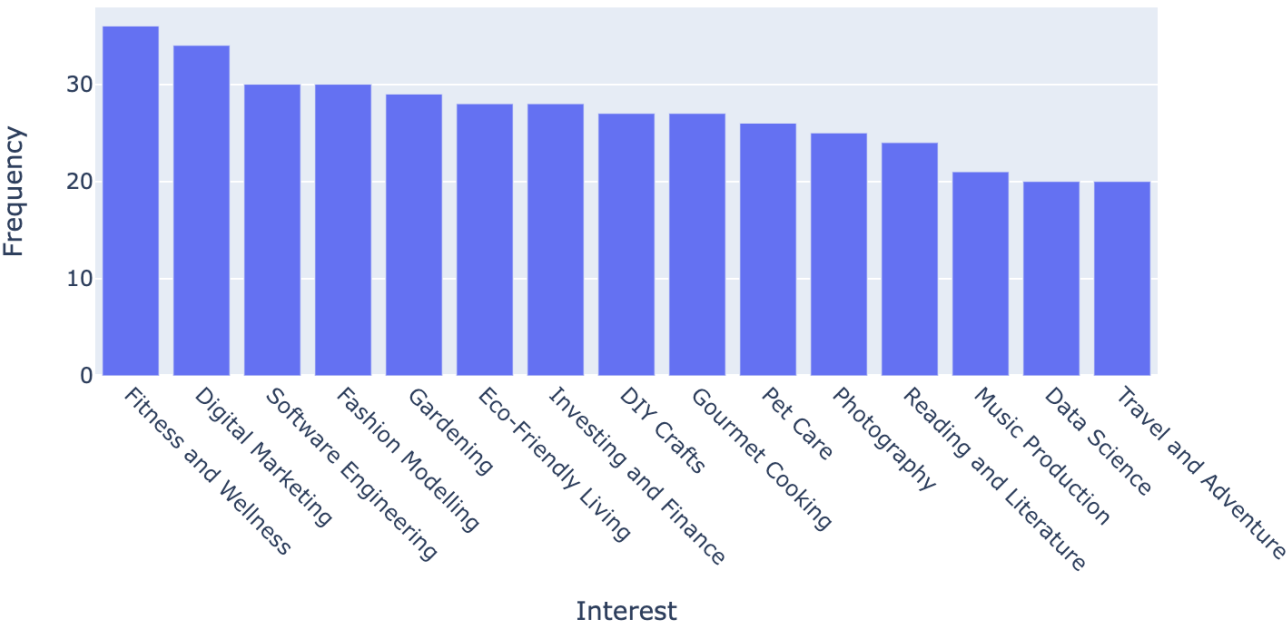
Based on the frequency counts of interests within each cluster, we can observe distinct patterns that provide valuable insights into the preferences and behaviors of different user segments:

Top 15 Interests for Cluster 0



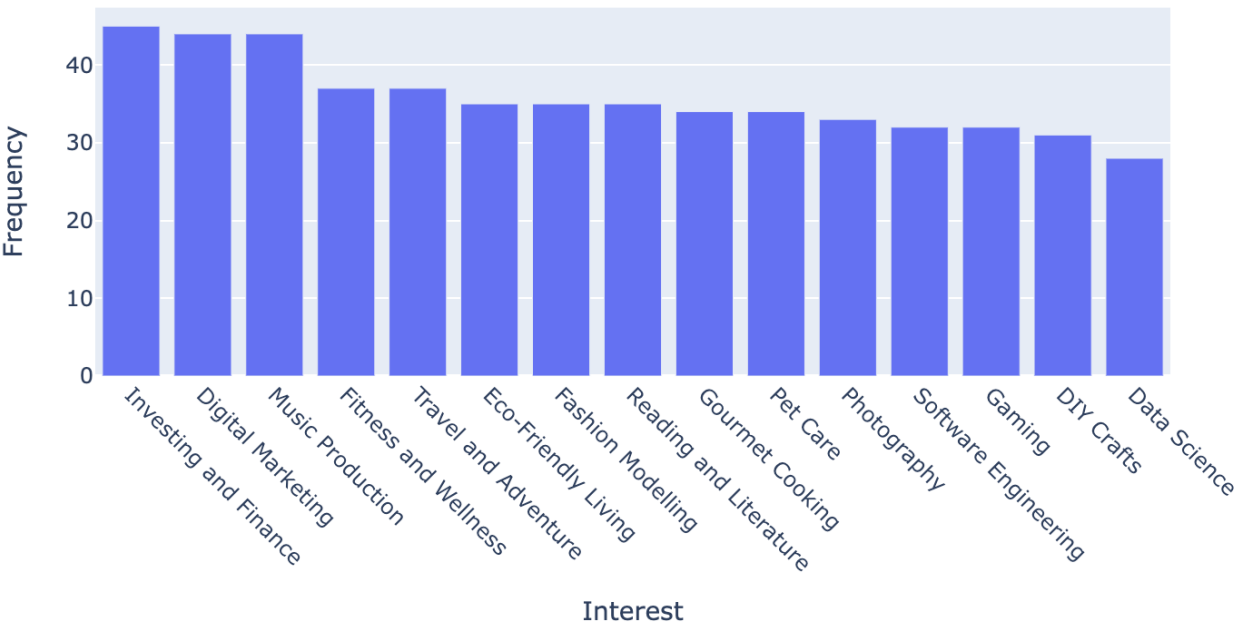
Cluster 0 (Social Savants): This cluster shows a strong affinity towards interests such as gaming, reading and literature, music production, and investing and finance. These users are likely to be engaged in a diverse range of activities, including digital marketing, DIY crafts, and pet care, reflecting their varied interests and hobbies.

Top 15 Interests for Cluster 1



Cluster 1 (Digital Mavericks): Users in this cluster exhibit a keen interest in fitness and wellness, digital marketing, software engineering, and fashion modeling. They are likely to be tech-savvy individuals with a focus on personal development and professional growth, as indicated by their interests in gardening, eco-friendly living, and DIY crafts.

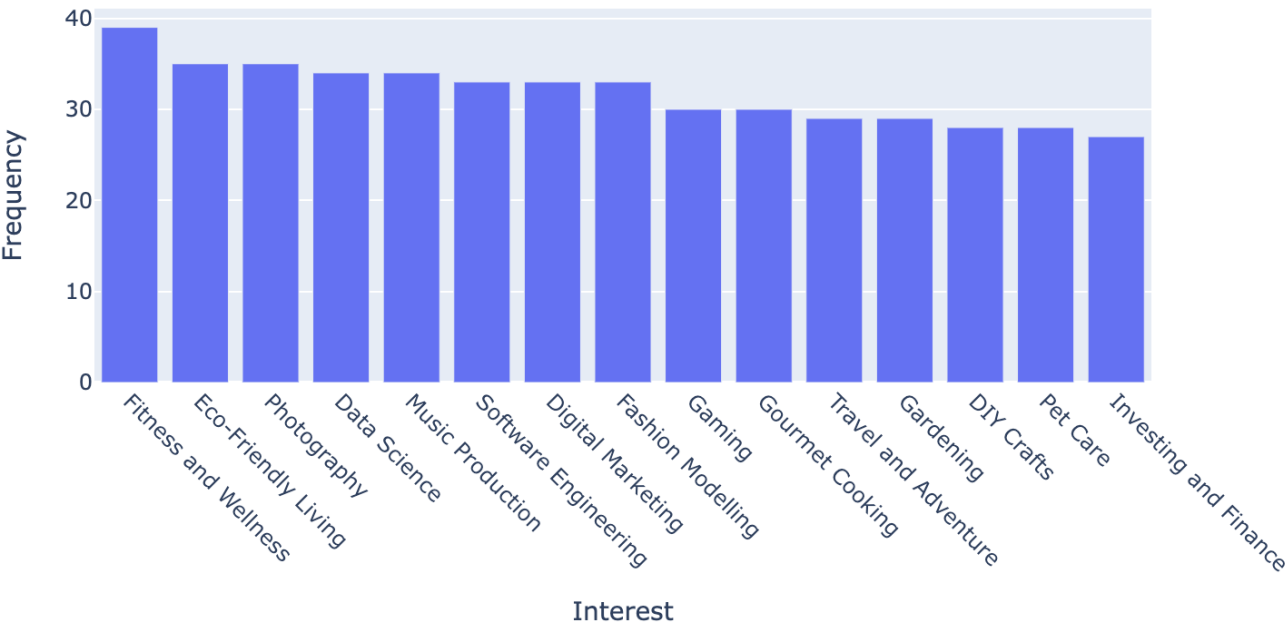
Top 15 Interests for Cluster 2



Cluster 2 (Zen Navigators): This cluster demonstrates a balanced interest in various domains, including investing and finance, digital marketing, music production, and fitness and wellness. These users are likely

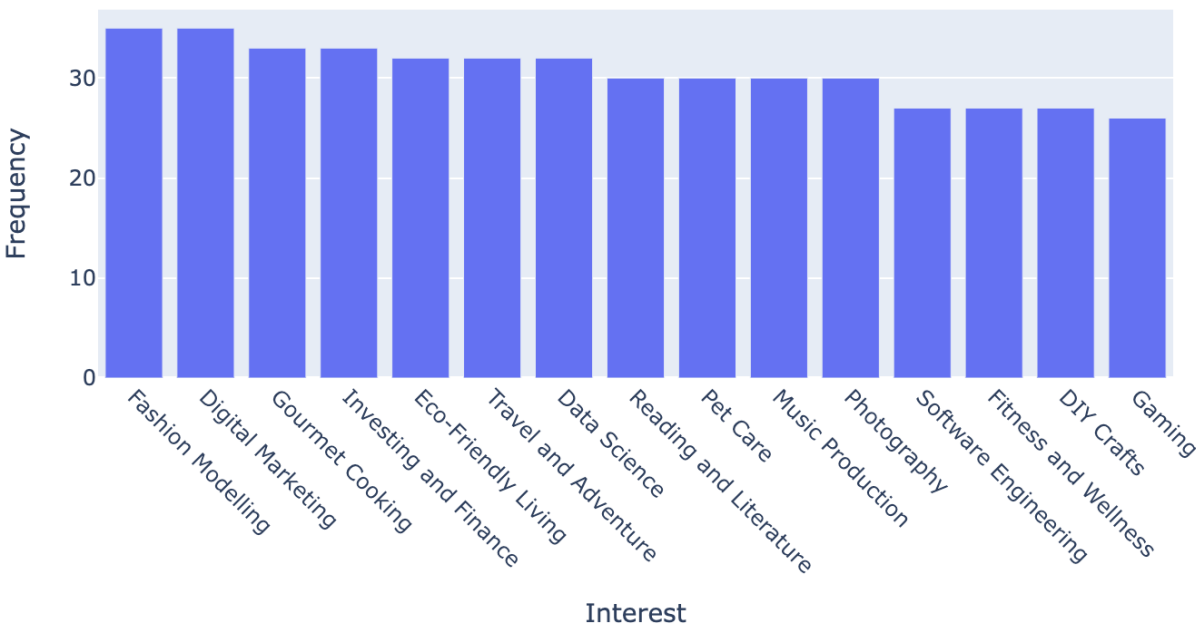
to prioritize personal interests such as gourmet cooking, photography, and reading and literature, reflecting a holistic approach to lifestyle and leisure activities.

Top 15 Interests for Cluster 3



Cluster 3 (Exploratory Enthusiasts): Users in this cluster show a strong interest in fitness and wellness, eco-friendly living, photography, and data science. They are likely to be adventurous individuals who enjoy exploring new topics and experiences, as evidenced by their diverse interests in music production, software engineering, and gaming.

Top 15 Interests for Cluster 4



Cluster 4 (Frugal Explorers): This cluster exhibits interests in fashion modeling, digital marketing, gourmet cooking, and investing and finance. Users in this segment are likely to be budget-conscious individuals who prioritize value-driven activities and hobbies, as reflected in their interests in eco-friendly living, travel and adventure, and DIY crafts.

Summary

This project successfully integrated user profiling, segmentation, and interest analysis to develop targeted advertising campaigns for distinct user segments. Leveraging demographic data, behavioral metrics, and user-generated content, the project gained valuable insights into the preferences and interests of different user groups.

Through user segmentation, five distinct clusters were identified: 'Social Savants', 'Digital Mavericks', 'Zen Navigators', 'Exploratory Enthusiasts', and 'Frugal Explorers', each characterized by unique behavioral patterns and preferences. These insights enabled the creation of tailored advertising strategies that resonate with the diverse needs and preferences of each segment.

The focus on analyzing interests within clusters provided deep insights into user preferences and behaviors. This approach allowed for the development of targeted campaigns that align with the specific interests and behaviors of each user segment.

Conclusion

In conclusion, this project underscores the power of combining advanced analytics techniques to drive targeted advertising campaigns. By understanding the nuances of user behavior and preferences, marketers can optimize their advertising efforts, enhance engagement, and drive better business outcomes. This approach not only improves campaign effectiveness but also strengthens customer relationships, ultimately leading to greater success in the competitive digital landscape.