Spotify User Analysis: ***Faraz Samie | MGT100 Spring 2025***

Data Overview:
The Spotify User Data consisted of 520 customers across 20 behavioral and demographic variables. This dataset presented information about listening habits, subscription preferences, and content consumption patterns, and includes information such as age, gender, device usage, music and podcast preferences which helped create a comprehensive user profiling system. The dataset is available to be downloaded at: ***https://www.kaggle.com/datasets/meeraajayakumar/spotify-user-behavior-dataset***

Methodology and R Script

For the analysis of Spotify users, I used customer segmentation techniques which were highlighted through MGT 100 homework assignments. The R script consisted of popular libraries like tidyverse for data manipulation, cluster analysis packages to perform the segmentation, and visualization tools to help explore the data. The R code uses behavioral scoring like engagement metrics or content preferences from variables within the Spotify dataset. These scores incorporate measures such as willingness scores or music and podcast engagement levels. Before performing the clustering, I created behavioral metrics to enhance accuracy and ensured the data was reliable. This preparation allowed me to apply K-means clustering with optimized parameters. I also evaluated the resulting segments using statistical tests to confirm they were both meaningful and relevant for business purposes.

Clustering Analysis and Segment Discovery
Using K-means clustering, the best option is through repeated assignment of customers to their closest centroids being k=4 clusters as the preferred segmentation method. This process involved randomly selecting centroids, assigning each customer to the nearest centroid, recalculating centroids based on customer groupings, and repeating these steps until results had stabilized. The K-means algorithm identified natural groupings within the user base by measuring similarity across behavioral attributes. When examining the UMAP visualization below, we observe distinct groupings within the four clusters, highlighting both clear separations and some overlap across user segments. This indicates that while certain user groups show consistent behavior, others share overlapping characteristics that reflect similar usage patterns.

Cluster One: 17% of Users | Highly Engaged Power Users:
This cluster stands out with the highest overall engagement across different features. The users within this cluster demonstrate strong willingness to get premium and represent the most active segment of the platform. Ultimately, I found that cluster one should be highly targeted, specifically for music engagement given its high overall activity levels.

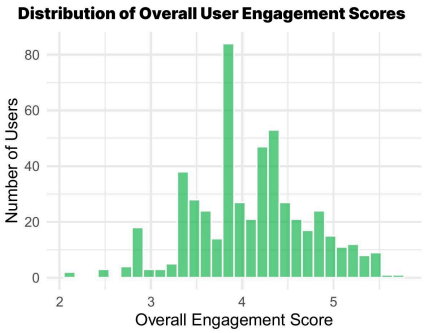
Cluster Two: 31% of Users | Popcast Focused with Strong Engagement:
Cluster two distinguishes itself by having strong podcast engagement. These users show balanced content consumption patterns and demonstrate sophisticated platform usage which suggests that they should receive high targeting specifically for podcasts.

Cluster 3: 22% of Users | Moderate Engagement Users:
Cluster three has moderate overall engagement with different content preferences and a larger share of Casual Listeners.
We can see consistent behavioral traits within this group as shown in the UMAP analysis. Ultimately, this cluster represents a middle ground between high and low engagement patterns.

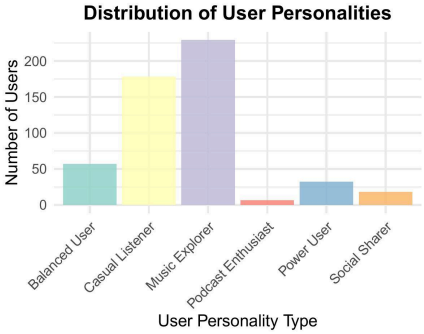
Cluster 4: 30% of User | Low Engagement Casual Users
Cluster four shows the lowest levels across metrics and represents the largest segment, suggesting that it shouldn't be prioritized. However users in the cluster could benefit from initiatives to increase user activity rather than a focus on converting to premium.

Conclusion & Strategy Implementations
The segmentation analysis reveals that targeted marketing would be most effective with Power Users and Music Explorers, as they show both high activity and higher responsiveness to premium offerings. However, at the same time the dominance of causal and exploratory listening behaviors suggests there is room to convert more passive users into highly engaged participants. This highlights an opportunity to refine subscription offerings across different user groups and to develop targeted content strategies that align with preferences. This highlights an opportunity to refine subscription offerings across different user groups and to develop targeted content strategies that align with preferences. The CDP framework enables data driven targeting and programmatic advertising that can personalize experiences for each identified segment.

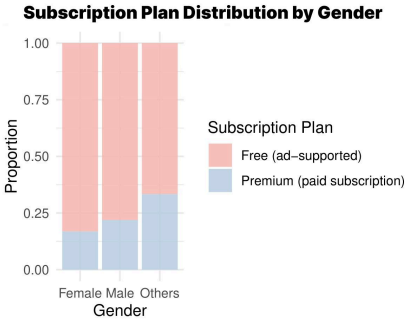
The accompanying charts below provide further insights into the findings of the data. Furthermore, the attached excel sheet further drives into the segmentation results. This analysis conducted in excel further reveals key demographic and behavioral patterns that inform targeted marketing strategies. For instance, device usage data shows that multi device users demonstrate higher premium adoption rates compared to single device users, which helps suggest that cross platform engagement correlates with premium willingness. Ultimately, these findings demonstrate the importance of combining segmentation insights with the CDP framework to improve marketing strategies. By focusing on high potential segments and engaging more casual users, Spotify can optimize both user experience and monetization from its diverse audience.



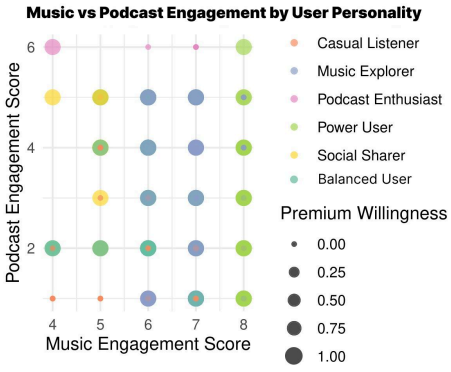
The overall distribution of user engagement scores displays somewhat of a normal distribution pattern. This comes as the majority of users clustering around scores of 4 and 5 with the highest concentration being in the center with a score of 3.8. This suggests that while most users demonstrate moderate engagement with the platform, there is a subset showing significantly higher levels of activity. Overall, the distribution indicates a healthy user base, however there is an opportunity to further enhance engagement across the broader population.



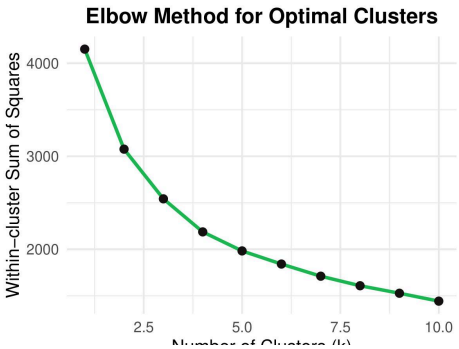
The distribution of user personality helps reveal that Spotify has a diverse ecosystem which comprises six different user types. In particular, Music Explorers make up the largest segment with over 200 users, followed by Casual Listeners with around 175 users and Balanced Users trailing further behind. However other segments such as Podcast Enthusiasts, Power Users, and Social Sharers are smaller segments making up at most 25 users approximately. This distribution highlights Spotify's ability to attract users with diverse content preferences through horizontal differentiation strategies. However, the dominance of causal and exploratory listening behaviors suggests there is room to convert more passive users into highly engaged participants.



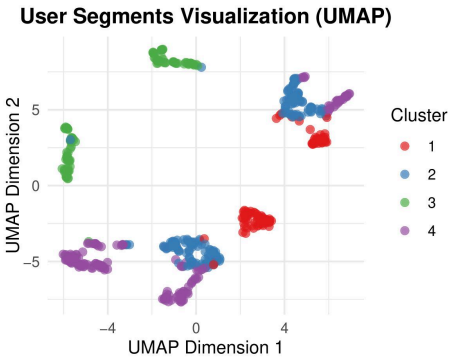
The distribution of gender based subscriptions reveals distinct behavioral tendencies across user segments. In particular, male users account for approximately 22% of the population and show a stronger inclination toward premium subscriptions. Female users represent over 75% of the population, showing a comparatively lower rate of premium adoption. Meanwhile, the Others category shows the most even split between free and premium plans even though it's a small segment with 3%. This distribution shows that current monetization strategies may align more closely with male user preferences.However, the significantly larger proportion of female users suggests that further data harmonization is needed to further understand the barriers to premium adoption. Ultimately, this highlights a strategic opportunity to refine subscription offerings across groups and suggests that premium brand partnerships may resonate particularly well with group segments that are less price sensitive.



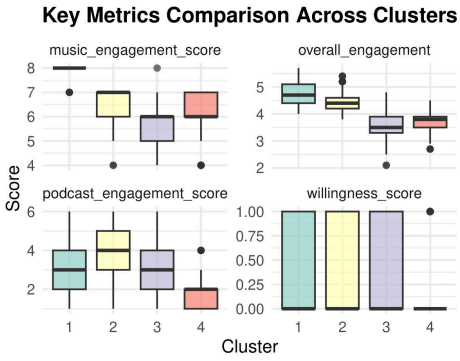
The scatter plot of music vs podcast engagement helps uncover distinct clustering patterns based on user personality types. In particular, Podcast Enthusiasts cluster is in the high podcast engagement region, while Music Explorers tends to be more dispersed with engagement in both music and podcast content. Balanced Users span across both, though they tend to lean slightly higher on music engagement, demonstrating horizontal differentiation based on taste preferences. In particular, Power Users and Music Explorers demonstrate elevated premium willingness scores, which suggest a stronger inclination toward paid features and vertical differentiation opportunities. This pattern indicates that users with high engagement in specific content types are more likely to convert to premium. As a result, targeted marketing through programmatic advertising would be most effective with Power Users and Music Explorers, as they show both high activity and higher receptiveness to premium offerings.



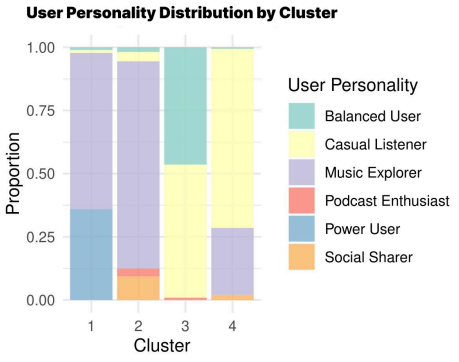
The K-means clustering algorithm can help determine the ideal number of user segments by utilizing Euclidean distance calculations to assign customers to specific centroids. The algorithm grouped each customer with the closest centroid and then recalculated the centroids based on those groupings. From 1 to 4 clusters, the sum of squares drops sharply, indicating significant gains in how well the data is grouped. However, the rate of improvement levels off beyond 4 clusters, which indicates that additional clusters offer little additional value. This suggests that dividing users into four distinct segments is the best course of action as it provides meaningful insights.



The market mapping visualization reveals six different groupings within the four clusters, which highlights how there is both separation and overlap occurring across user segments. Some regions exclusively have Cluster three users, while other zones feature an overlap between Clusters one, two, and four. This distribution suggests that while certain user segments like Cluster three exhibit more consistent behavioral traits, others like clusters two and four share overlapping characteristics. The spread of the clusters show how each group is apart based on their engagement habits and what kind of content they prefer.



A comparison of metrics reveals distinct behavioral patterns across the four user clusters, showing clear differences between each segment. Cluster one stands out with the highest overall engagement across features, while cluster four shows the lowest levels across all metrics. Cluster two is characterized by strong podcast engagement, whereas cluster three displays moderate overall engagement with varied content preferences. Ultimately, this suggests that cluster four shouldn't be prioritized due to low engagement, and that cluster one should be highly targeted particularly for music engagement given its high overall activity. Cluster two should receive high targeting specifically for podcasts.



The distribution of user personalities across clusters reveals how distinct user types align with specific behavioral patterns. Clusters 1 and 2 are mainly made up of Music Explorers, which represents a highly engaged and content focused segment showing strong vertical differentiation potential.. In contrast, Clusters 3 and 4 contain a larger share of Casual Listeners, which indicates lower overall engagement.