THE UNIVERSITY OF CHICAGO

From Free to Premium: A Persona-Based Approach to Understanding Spotify Subscription Upgrades

By

Zihan Zhao

April 2024

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Preceptor: Sabrina Nardin

Sponsor: Ashley Henry Cauly, Devin Singh

Abstract

In the era of digital music streaming, understanding user behavior and preferences is crucial for platforms like Spotify to optimize their services and drive growth. This study aims to investigate the factors influencing premium membership conversion and develop user personas to inform targeted strategies for enhancing user satisfaction and engagement on Spotify. Through a survey of 600 Spotify users, three distinct user segments are identified: free members (59%), free members willing to upgrade to premium (24%), and premium members (17%). Logistic regression analysis reveals key factors influencing premium conversion, including current Spotify plan, premium plan price, and age. To gain a deeper understanding of each user segment, three personas are developed: Nia (free member), Aria (free member with willingness to upgrade), and Sienna (premium member). These personas encapsulate the characteristics, preferences, and pain points of each user segment, providing insights into their needs and expectations from the platform. By leveraging these insights, Spotify can improve user satisfaction, increase premium conversion rates, and maintain a competitive edge in the music streaming industry. This research contributes to the growing body of knowledge on user experience and premium conversion in music streaming services by synthesizing insights from various data sources and methodologies. It bridges the gap between academic research and industry application, providing actionable recommendations for Spotify and other streaming platforms to drive growth and user engagement.

Keywords: music streaming, Spotify, user personas, user behavior, user segmentation

Main Deliverable

1 Summary of Key Findings

1.1 User Personas

Through the analysis of user data and characteristics, three distinct user personas were developed to represent the different segments of Spotify's user base:

1. Nia (Free Member): Nia represents the majority of Spotify users who enjoy the free, ad-supported version of the platform. While satisfied with the basic features, Nia occasionally experiences frustration with ad interruptions and limited skips. Nia's listening habits revolve around discovering new music and creating playlists for various moods and activities.

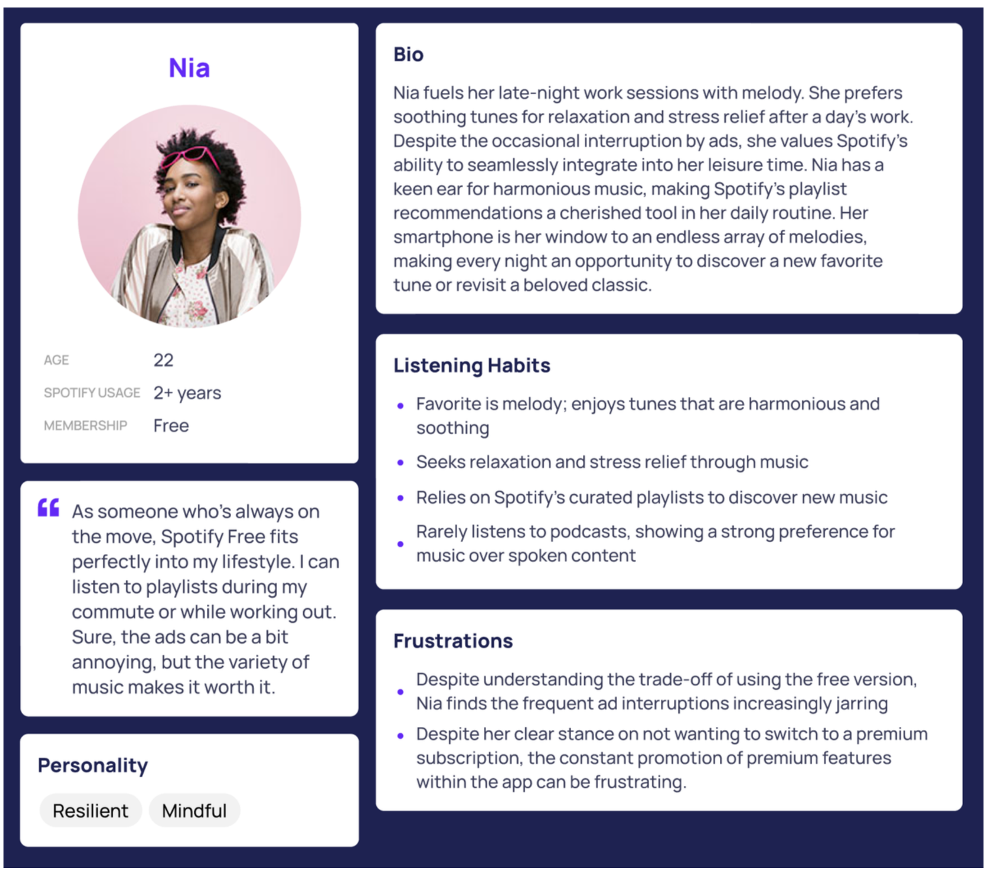

Figure 1: Spotify Free Member Persona - Nia

2. Aria (Free Member with Willingness to Upgrade): Aria embodies the segment of free users who are open to upgrading to a premium subscription. Aria highly values personalized recommendations and curated playlists, but feels that the current free version lacks the full potential of Spotify's features. Aria's willingness to upgrade is driven by a desire for an ad-free experience, offline listening, and enhanced audio quality.

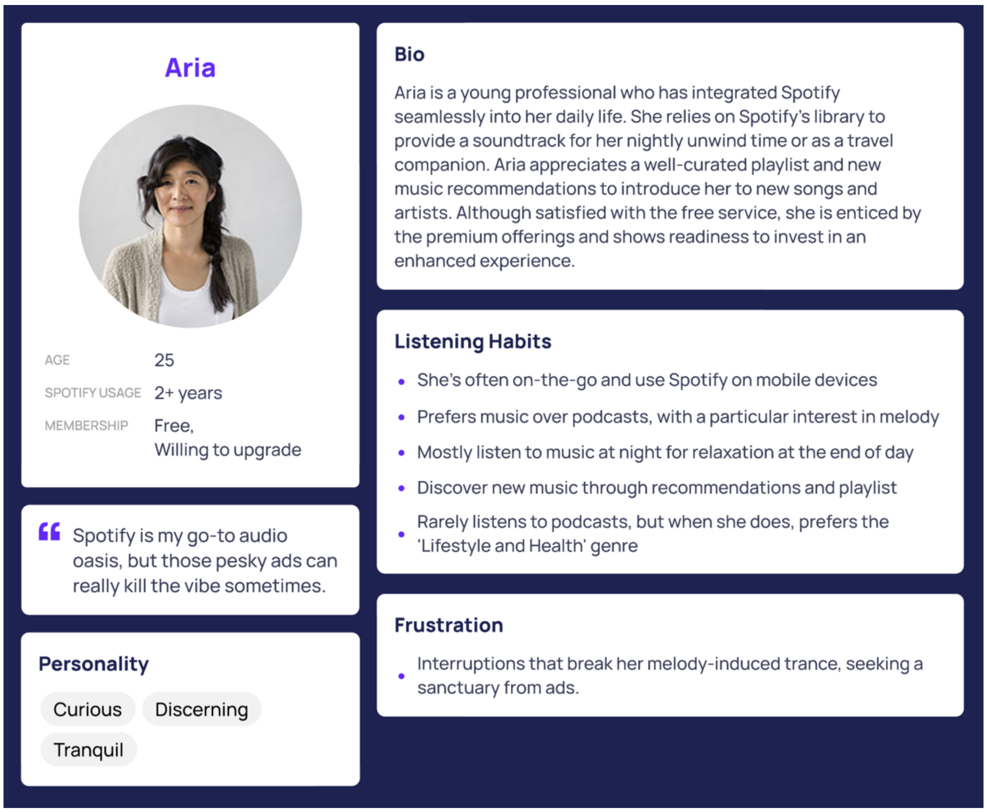

Figure 2: Spotify Free Member with Willingness to Upgrade Persona - Aria

3. Sienna (Premium Member): Sienna represents the premium subscribers who fully embrace Spotify's advanced features and benefits. Sienna highly engages with personalized playlists, enjoys ad-free listening, and frequently uses offline mode for uninterrupted music access. However, Sienna occasionally experiences dissatisfaction with playlist repetitiveness and seeks even more diverse recommendations.

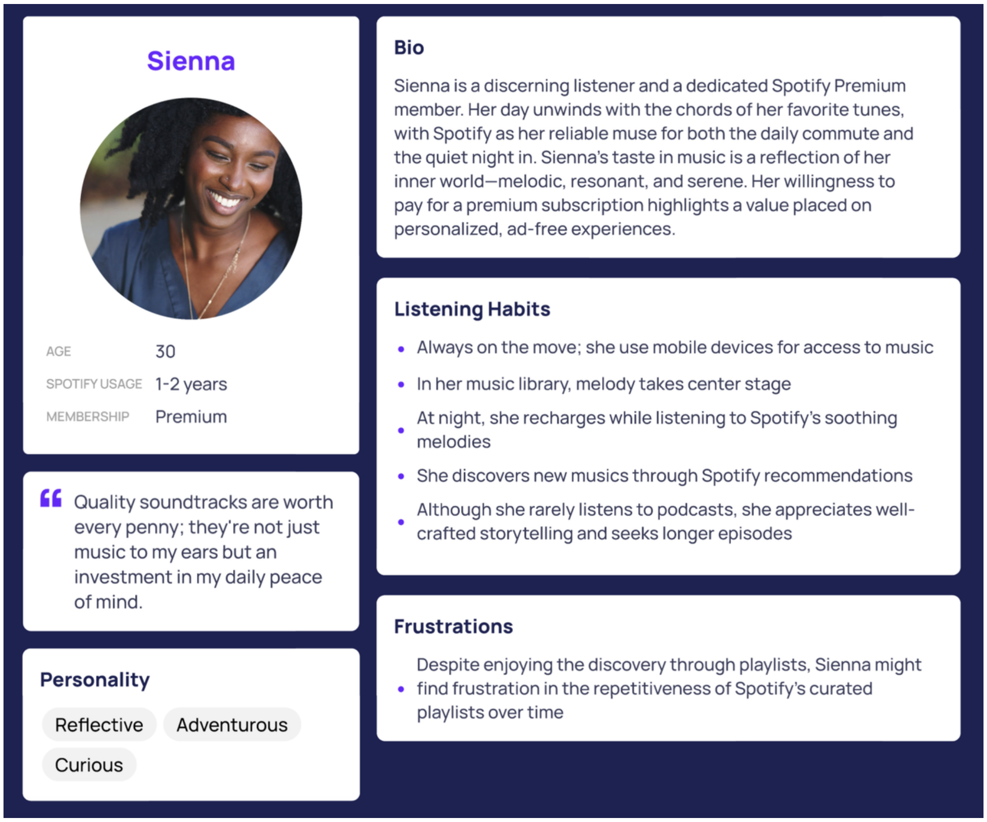

Figure 3: Spotify Premium Member Persona - Sienna

1.2 Key Factors Influencing Premium Conversion

$$
\begin{aligned}
y = & -1.48 \times \text{Age} && (1)\\
& + 1.57 \times \text{Current Spotify plan} && (2)\\
& - 3.51 \times \text{Premium Plan Price} && (3)\\
& + 0.36 \times \text{Favorite Music Content} && (4)\\
& + 0.10 \times \text{Music Listen Frequency} && (5)\\
& + 1.15 \times \text{Music Discovery Method} && (6)\\
& - 0.66 \times \text{Podcast Listen Frequency} && (7)\\
& + 1.60 \times \text{Podcast Satisfaction} && (8)\\
& + 0.79 && (9)
\end{aligned}
$$

The equation predicts the willingness to upgrade to a premium membership (y) based on various factors. A positive coefficient indicates that an increase in the corresponding variable leads to a higher likelihood of upgrading to premium, while a negative coefficient suggests the opposite effect.

According to the equation, the most significant factor influencing premium membership conversion is the current Spotify plan (coefficient: 1.57). Users who are already on a paid plan are more likely to continue their premium subscription in the future. The music discovery method (coefficient: 1.15) also plays a crucial role, suggesting that users who rely on Spotify's features to discover new music are more inclined to upgrade.

Interestingly, the premium plan price (coefficient: -3.51) has a strong negative impact on the willingness to upgrade. This indicates that users are sensitive to the cost of the premium subscription and may be deterred by higher prices. Age (coefficient: -1.48) also has a negative effect, implying that older users are less likely to convert to premium compared to younger users.

The equation also reveals that podcast satisfaction (coefficient: 1.60) positively influences premium conversion. Users who are satisfied with the variety and quality of podcasts available on Spotify are more likely to upgrade. However, podcast listening frequency (coefficient: -0.66) has a negative impact, suggesting that frequent podcast listeners may be less inclined to pay for a premium subscription.

Other factors, such as favorite music content (coefficient: 0.36) and music listening frequency (coefficient: 0.10), have relatively smaller positive effects on premium conversion. This indicates that users who prefer specific types of music content and listen to music more frequently are slightly more likely to upgrade.

1.3 Business Recommendations

Based on the insights gathered from the logistic regression analysis, the following recommendations are proposed to enhance user satisfaction and increase the conversion rate from free to premium membership:

1. Highlight Premium Plan Benefits: Emphasize the ad-free listening experience, offline playback, high-quality audio, exclusive content, and features offered by premium subscriptions. Use targeted marketing campaigns to communicate the value proposition of premium plans to different user segments.

2. Improve Music Discovery Features: Invest in advanced algorithms and machine learning techniques to provide highly personalized music recommendations. Curate playlists and stations based on users' listening histories, favorite genres, and moods. Collaborate with artists and influencers to create exclusive content and playlists that encourage users to explore new music.

3. Optimize Pricing and Offer Flexible Plans: Conduct market research to determine the optimal pricing for premium subscriptions based on user preferences and willingness to pay. Consider introducing more flexible pricing options, such as family plans, student discounts, or bundle deals with other services. Offer limited-time promotions or free trials to incentivize users to experience the benefits of premium membership.

4. Enhance Podcast Experience: Invest in exclusive podcast content and partnerships with popular creators to attract and retain podcast listeners. Improve podcast discoverability by implementing advanced search and recommendation algorithms. Provide personalized podcast recommendations based on users' listening behaviors and preferences.

5. Continuously Gather User Feedback and Iterate: Regularly survey users to understand their evolving needs, preferences, and pain points. Analyze user behavior data to identify patterns and trends that can inform product development and marketing strategies. Implement a data-driven approach to continuously refine and optimize the user experience and premium conversion strategies.

1.4 For Detailed Research Report

The research presented in this paper provides an overview of the study's key findings and insights. A more comprehensive and detailed analysis can be found in the full research report: click here.

Literature review

1 Introduction

The advent of music streaming services has reshaped the way we consume and engage with music, ushering in a new era of personalized listening experiences. This disruptive technology has democratized access to vast libraries of music, allowing users unprecedented choice and control over their audio content. Among the pioneers in the music streaming industry, Spotify stands out as a platform that has captivated millions of users worldwide with its extensive catalog and innovative features.

Spotify's success can be attributed to its ability to cater to the diverse preferences and behaviors of its user base, which includes both free, ad-supported members and premium subscribers. By leveraging data-driven insights from its vast user base and deploying advanced algorithms, Spotify has crafted a user experience that seamlessly blends personalization, content discovery, and social connectivity, fostering a deep emotional connection between users and their music. As of 2022, Spotify boasted an impressive 456 million monthly active users, with 195 million of them being premium subscribers (GilPress, 2024), a testament to the platform's ability to attract and retain users.

As a leader in the music streaming industry, Spotify's revenue heavily relies on its premium membership model. While the platform offers a free, ad-supported tier, premium subscriptions represent a significant and sustainable revenue stream that fuels the company's growth and continued innovation. In 2022, Spotify reported premium subscription revenue of €9.6 billion, accounting for 87% of the company's total revenue (GilPress, 2024). The predictability and continuity of premium subscription revenue are vital for Spotify's long-term sustainability and competitiveness within the rapidly evolving music streaming landscape. By fostering a loyal and engaged premium subscriber base, Spotify can mitigate the risks associated with fluctuations in advertising revenue and maintain a steady stream of income to invest in platform enhancements, content acquisition, and user experience improvements. Therefore, understanding its user base, the factors that drive users to adopt premium subscriptions, and developing strategies to encourage conversion from free to paid memberships are crucial for Spotify's continued success and growth in the highly competitive music streaming market.

2 The Freemium Model in Music Streaming

The freemium business model has become a dominant strategy among music streaming platforms, including Spotify. This model offers users access to a basic, ad-supported version of the service for free, while providing the option to upgrade to a premium, paid subscription with additional features and benefits. The freemium approach has been widely adopted in the digital services industry, as it allows companies to attract a large user base and generate revenue through a combination of advertising and subscription fees (Kumar, 2014).

Under the freemium model, music streaming platforms like Spotify provide users with a limited set of features and functionalities in the free version, such as ad-supported listening, limited skips, and basic personalization. The goal is to entice users to try the service and experience its value proposition without any upfront financial commitment. By offering a free tier, platforms can lower the barriers to entry and rapidly expand their user base (Thomes, 2013).

However, the freemium model also presents challenges in terms of converting free users to paid subscribers. While the free version attracts a large number of users, only a small percentage typically upgrades to the premium tier. This conversion rate is crucial for the long-term sustainability and profitability of music streaming platforms (Oestreicher-Singer & Zalmanson, 2013). To overcome this challenge, platforms must focus on strategies that demonstrate the value and benefits of the premium version, such as ad-free listening, offline access, and enhanced personalization.

One key aspect of optimizing the freemium model is understanding user personas. By segmenting users based on their behaviors, preferences, and willingness to pay, music streaming platforms can tailor their marketing messages and product features to specific user groups (Wedel & Kannan, 2016). Moreover, user personas can inform the design of the freemium user experience to encourage premium conversion. Persona-driven insights can guide the development of targeted onboarding processes, in-app messaging, and promotional campaigns that resonate with specific user segments and motivate them to upgrade (Becker & Jaakkola, 2020).

3 User Personas

User personas are research-based representations of different user types, capturing their unique characteristics, behaviors, motivations, and pain points (J. S. Pruitt & Adlin, 2006). In the music streaming industry, user personas have proven to be particularly valuable for understanding and segmenting diverse user groups, enabling more effective product design, marketing strategies, and user experience optimization.

Numerous studies have highlighted the effectiveness of personas in driving user-centric design and decision-making in the technology industry. J. Pruitt and Grudin (2003) demonstrated how personas can facilitate a shared understanding of users within cross-functional teams, leading to more cohesive and user-focused product development efforts.

Within the music streaming industry, user personas have proven to be particularly valuable. Lee and Waterman (2012) explored the use of personas in the design of music discovery features for a streaming service. By developing personas based on user research, they identified distinct user groups with varying music discovery preferences and behaviors, informing the design of personalized music discovery features that enhanced user satisfaction and engagement.

One approach to persona development is the Goal-Directed Design method, popularized by Cooper (1999). This method involves conducting in-depth interviews with users to understand their goals, motivations, and behaviors. By synthesizing this qualitative data, researchers can create detailed persona profiles that include a name, photo, background, goals, and a narrative describing their typical interactions with the product.

Another approach is the Jobs-to-Be-Done method, which focuses on understanding the specific jobs or tasks that users want to accomplish using a product (Christensen et al., 2016). By identifying these jobs and the context in which they occur, researchers can create personas that reflect the different ways users engage with a product to achieve their goals.

In addition to these methods, data analytics can provide valuable insights into user behavior and preferences. By analyzing user data such as listening history, playlists, and social interactions, music streaming platforms can identify patterns and segments that inform persona development. For example, Spotify's "Taste Profiles" feature uses machine learning algorithms to analyze users' listening data and create personalized taste profiles that capture their unique music preferences (Ciocca, 2020).

Case studies from other industries illustrate the power of nuanced persona development. In the gaming industry, Electronic Arts (EA) used persona research to redesign its FIFA soccer game franchise. By conducting extensive user research and creating detailed player personas, EA was able to identify key motivations and behaviors that informed the game's design, leading to increased player engagement and sales (King et al., 2019).

Similarly, in the e-commerce industry, Amazon uses persona development to guide its product recommendations and marketing strategies. By analyzing user data and creating personas based on purchasing habits, browsing history, and demographic information, Amazon can tailor its product recommendations and marketing messages to specific user segments, increasing customer satisfaction and loyalty (Xu & Lee, 2020).

In conclusion, user persona development is a crucial aspect of understanding and segmenting users in the music streaming industry. By employing a range of methodologies, including Goal-Directed Design and Jobs-to-Be-Done, music streaming platforms can create personas that capture the psychological, demographic, and behavioral characteristics of their users. These personas can then inform product design, marketing strategies, and user experience optimization, ultimately leading to increased user satisfaction, engagement, and retention.

4 Focus on Premium Conversion

As the sustainability of music streaming platforms relies heavily on premium subscriptions, understanding the factors that influence users' decisions to upgrade is crucial. Several studies have delved into the psychology behind premium service subscriptions, addressing decision-making processes, spending psychology, and perceived value.

Hamari and Keronen (2017) investigated the motivations and attitudes influencing users' willingness to pay for freemium services, including music streaming platforms. Their findings revealed that perceived benefits, such as increased enjoyment and social value, positively influenced users' attitudes towards premium subscriptions, while perceived sacrifices, such as monetary costs and privacy concerns, had a negative impact.

In a study focused on the decision-making processes behind premium subscriptions, Wagner et al. (2014) explored the mental models and cognitive biases that shape users' perceptions of value and their willingness to pay for digital services. The researchers found that users often relied on heuristics and mental shortcuts when evaluating the benefits and costs of premium offerings, leading to biased decisions based on factors such as loss aversion, anchoring effects, and the sunk cost fallacy.

Pricing psychology also plays a crucial role in users' willingness to pay for premium services. Wlömert and Papies (2016) investigated how users' price perceptions, reference points, and sensitivity influence their attitudes towards premium subscriptions. The study found that users' willingness to pay was influenced by factors such as the perceived fairness of the price, the value of the premium features, and the comparison to other streaming services. This highlights the importance of strategic pricing and positioning for music streaming platforms, taking into account users' psychological responses to different price points and framing.

In addition to these factors, the social and emotional aspects of music consumption can also influence users' decisions to upgrade to premium services. Oestreicher-Singer and Zalmanson (2013) explored how social engagement and community participation on music streaming platforms can drive premium conversion. They found that users who actively contribute to the platform, such as by creating playlists, sharing music, and engaging with other users, are more likely to perceive value in premium features and upgrade their subscriptions.

Furthermore, the timing and context of premium offers can impact users' likelihood of upgrading. Kim et al. (2018) investigated how contextual factors, such as users' current mood, listening behavior, and situational context, can influence their receptiveness to premium offers. The study found that users were more likely to consider upgrading when they were in a positive mood, actively engaged with the platform, and in situations where the benefits of premium features were salient (e.g., when listening to music offline or in a noisy environment).

In conclusion, the conversion of free users to premium subscribers is a complex process influenced by a range of psychological, social, and contextual factors. Music streaming platforms should focus on communicating the unique value proposition of their premium offerings, addressing cognitive biases in users' decision-making processes, and employing strategic pricing and positioning. Furthermore, fostering social engagement, encouraging user participation, and delivering personalized and contextually relevant premium offers can help drive premium conversion and support the long-term sustainability of music streaming services.

5 Theoretical Frameworks

The Technology Acceptance Model (TAM) has been widely used to understand user behavior in various digital platforms, including music streaming services. Developed by Davis et al. (1989), TAM posits that perceived usefulness and perceived ease of use are the primary factors influencing users' attitudes towards technology and their intention to use it. In the context of music streaming, perceived usefulness refers to the extent to which users believe that the service will enhance their music listening experience, while perceived ease of use relates to the user-friendliness and accessibility of the platform.

Researchers have applied TAM to investigate the factors that influence users' adoption of premium subscriptions in music streaming services. For example, Sinclair and Tinson (2017) examined how perceived usefulness, such as ad-free listening and offline access, and perceived ease of use, such as intuitive interface and seamless integration, shape users' attitudes towards premium offerings. Their study found that these factors significantly influence users' willingness to pay for premium services, highlighting the importance of designing user-centric features and experiences that align with users' needs and expectations.

Moreover, TAM has been extended to incorporate additional factors that may influence user behavior in music streaming services. For instance, some researchers have included social influence and perceived value as additional predictors of users' attitudes and intentions (Chen & Hu, 2019). By considering these factors, researchers can gain a more comprehensive understanding of the complex dynamics that shape user behavior in music streaming platforms.

Another relevant theoretical framework is the Uses and Gratifications Theory (U&G), which focuses on understanding the reasons why individuals actively seek out and consume specific media or technologies (Katz et al., 1973). U&G posits that users are goal-oriented and actively choose media that fulfill their specific needs and gratifications, such as entertainment, information seeking, social interaction, and personal identity formation.

In the context of music streaming, researchers have employed U&G to explore the diverse motivations and gratifications that drive users' engagement with these platforms. For example, Krause et al. (2019) identified four main categories of gratifications sought by music streaming users: entertainment, information seeking, social interaction, and mood management. Their study found that these gratifications were significantly associated with users' satisfaction, loyalty, and willingness to pay for premium features, underscoring the importance of understanding and catering to users' diverse needs and motivations.

By employing TAM and U&G as theoretical frameworks, researchers can gain valuable insights into the psychological, social, and technological factors that shape user behavior in music streaming services. These frameworks provide a foundation for understanding the complex interplay between users' perceptions, motivations, and experiences, and can inform the design, marketing, and management strategies of music streaming platforms.

However, it is important to note that while these frameworks offer valuable insights, they may not fully capture the complexity and specificity of user behavior in music streaming services. As the digital landscape continues to evolve, researchers may need to adapt and extend these frameworks to incorporate new factors and dynamics that emerge in the context of music streaming. Additionally, researchers should consider complementing these theoretical frameworks with empirical studies and data-driven approaches to validate and refine their understanding of user behavior in music streaming platforms.

6 Methodological Diversity

The existing literature on user experience and paid membership conversion in digital platforms has employed a diverse range of research methodologies, reflecting the multifaceted nature of the phenomenon under investigation. Qualitative approaches, such as in-depth interviews and ethnographic studies, have been widely utilized to gain a nuanced understanding of user motivations, behaviors, and experiences (Hagen, 2015). These methods have proven invaluable in capturing rich, contextual data and uncovering underlying patterns and themes that quantitative methods alone may overlook.

On the other hand, quantitative methods, including surveys, usage data analysis, and experimental designs, have provided a complementary perspective by allowing researchers to measure and quantify user preferences, attitudes, and behaviors (Berman, 2018; Sinclair & Tinson, 2017). These approaches have enabled the identification of statistically significant relationships and the prediction of user behavior based on various factors, contributing to a more comprehensive understanding of the phenomenon.

Researchers have increasingly adopted mixed-methods approaches, combining the strengths of both qualitative and quantitative techniques (Wagner et al., 2014). This methodological triangulation has allowed for the corroboration and convergence of findings, enhancing the validity and reliability of the research outcomes.

7 The Present Study's Contribution

Building upon the existing theoretical foundations and methodological approaches, the present research aims to contribute a comprehensive understanding of user personas within Spotify's platform, with a particular focus on membership personas and conversion factors. While theoretical frameworks such as the Technology Acceptance Model (TAM) and Uses and Gratifications Theory (U&G) offer valuable insights into user behavior and technology adoption, this study takes a more data-driven and industry-focused approach to understanding Spotify's user segments and the factors influencing premium conversion.

The TAM, which examines factors like perceived usefulness and perceived ease of use, and the U&G theory, which explores the motivations and gratifications sought by users, provide a conceptual foundation for understanding user attitudes and behaviors in the context of music streaming services. However, given the specific focus of this study on identifying user personas and the factors influencing premium conversion, these frameworks are not directly applied in the research design.

Instead, this study leverages a combination of primary survey data and secondary open-source data to develop a comprehensive understanding of Spotify's user segments and their unique characteristics. By employing descriptive statistics and machine learning techniques, such as logistic regression and persona development, this research seeks to uncover different user personas, their unique traits, and the key factors influencing their willingness to upgrade to a premium subscription.

Through the identification of motivations, preferences, and behaviors of free users, premium users, and those willing to upgrade, this research endeavors to provide actionable insights to Spotify and other music streaming platforms. These insights can inform targeted strategies for enhancing user experiences, optimizing recommendation algorithms, and developing effective marketing campaigns to drive premium membership adoption.

Evaluation of Research Design

1 Data

To gain a comprehensive understanding of Spotify's user behavior and preference, this research used both primary and secondary data sources. The primary data source is a survey conducted using Google Forms, designed to gather information on user behavior, demographics, membership status, and willingness to upgrade to a premium subscription (see appendix). The survey questions were crafted to align with the Spotify User Behavior Dataset, an open-source dataset by Meeraa Jayakumar (2020), which serves as the secondary data source.

The primary survey was conducted online from March 7th to March 20th, 2024, using a convenience sampling method. The survey was distributed among the researcher's personal network, including friends, family, classmates, and acquaintances who were known to be Spotify users. This sampling technique was chosen due to its accessibility and cost-effectiveness, given the limited resources available for the study.

Before sending the survey link, the researcher confirmed through text messages that each potential respondent was either a current Spotify user or had used Spotify in the past. This screening process ensured that the sample consisted entirely of individuals with relevant experience using the platform.

To ensure data completeness, all survey questions were marked as required, preventing respondents from submitting the form with missing data. As a result, no data-cleaning procedures were necessary, and all 80 responses collected were included in the final dataset. The 80 responses collected through the survey were then combined with the 520 response data from the Spotify User Behavior Dataset by Meeraa Jayakumar (2020). By merging the two datasets, a sample of 600 responses was obtained for all later analysis.

Characteristics	Questionnaire Item	Frequency	Percentage (%)
Gender	Female	427	71.17
	Male	143	23.83
	Others	30	5.00
Age	6-12	3	0.50
	12-20	102	17.00
	20-35	449	74.83
	35-60	45	7.50
	60+	1	0.17
Spotify Usage Period	Less than 6 months	112	18.67
	6 months to 1 year	135	22.50
	1 year to 2 years	163	27.17
	More than 2 years	190	31.67
Spotify Plan	Free	470	78.33
	Premium	130	21.67

Figure 4: Descriptive Statistics of Survey Data

Figure 4 presents the demographic characteristics of the 600 respondents. The majority of respondents were female (71.17%), aged between 20 and 35 years old (74.83%), and had been using Spotify for more than 1 year (58.84%). Additionally, 78.33% of respondents were using the free version of Spotify, while 21.67% were premium subscribers.

2 Methods

2.1 User Segmentation and Persona Development

In this study, users were segmented based on their responses to two key survey questions:

1. "Which Spotify subscription plan do you currently have?"
2. "Are you willing to take a premium subscription or willing to continue with premium subscription in future?"

Based on the responses, users were categorized into three distinct segments:

1. **Free Users:** Those who currently use Spotify's free, ad-supported tier.
2. **Premium Users:** Those who are currently subscribed to Spotify's premium plan.
3. **Free Users with Willingness to Upgrade:** Free users who expressed an interest in upgrading to a premium subscription in the future.

After segmenting users, personas were developed for each user category using Figma, a popular design and prototyping tool. This approach aligns with industry standards for user experience research and design (Goltz, 2014). The persona development process involved analyzing the modes of characteristics, preferences, and behaviors of users within each segment. Factors such as age, gender, listening habits, preferred music genres, and engagement with features were examined to identify the most common attributes within each user category.

For instance, if the mode age for free users was found to be 25-30 years old, and the most common preferred music genre was melodic music, these attributes would be incorporated into the persona representing the free user segment. This approach ensures that the personas are grounded in the actual data and reflect the most prevalent characteristics of each user group.

The use of user segmentation and persona development based on survey data has been validated in various studies. For example, Salminen et al. (2018) employed a similar approach to create personas for online video streaming users, using survey data to identify distinct user segments and their characteristics. Similarly, Spiliotopoulos et al. (2020) used survey data to develop personas for Twitter users, demonstrating the effectiveness of this method in capturing user diversity and informing user-centered design.

2.2 Logistic Regression

Logistic regression is a statistical method for analyzing datasets in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (where there are only two possible outcomes). In our study, this method is particularly suitable because the dependent variable, users' willingness to upgrade, is inherently binary.

Unlike linear regression, which predicts a continuous outcome, logistic regression is designed for binary classification tasks. It estimates the likelihood of an event occurring by fitting data to a logistic function, thus providing a probabilistic framework for modeling binary outcome variables. This characteristic of logistic regression makes it more appropriate for our study compared to linear regression, which assumes a linear relationship between independent variables and a continuous dependent variable.

2.2.1 Data Preprocessing

Before building the logistic regression model, the dataset underwent preprocessing to ensure its suitability for analysis. Since 18 out of the 19 predictor variables were categorical, the LabelEncoder() function from the scikit-learn library was used to convert each categorical variable into unique numerical labels. This encoding process allowed the categorical variables to be effectively utilized in the logistic regression model.

2.2.2 Feature Selection

To mitigate potential multicollinearity among the predictor variables, two feature selection techniques were applied. First, a heatmap was plotted to identify highly correlated variables.

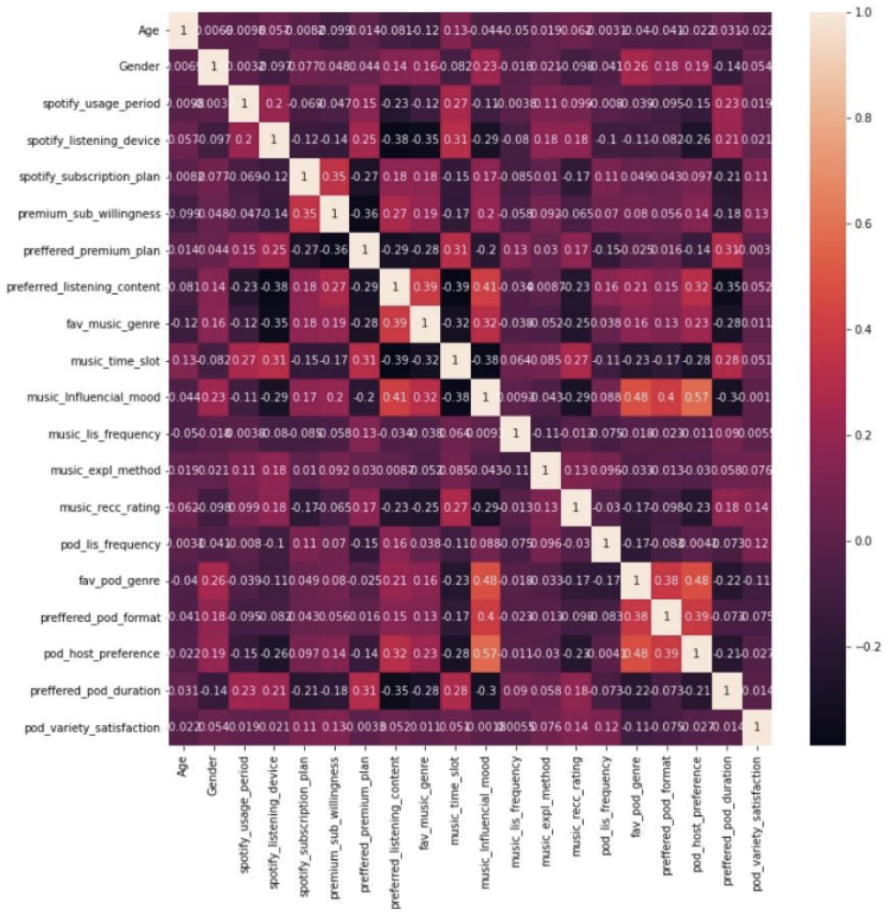

Figure 5: Heatmap of Correlation Coefficients Among Predictor Variables

The heatmap in Figure 5 visualizes the pairwise correlation coefficients among the predictor variables. Each cell in the heatmap represents the correlation between two variables, with darker shades indicating stronger correlations.

Based on the heatmap analysis, the following variables were excluded due to their high correlations with other predictors: 'spotify_listening_device', 'music_time_slot', 'preffered_pod_duration', 'fav_music_genre', 'music_Influencial_mood', 'pod_host_preference', and 'preffered_pod_format'.

Next, the Chi-Square test was employed as an additional feature selection method. This test assesses the independence between each categorical variable and our target variable. Features with p-values exceeding the alpha level of 0.05 were considered to have no significant association with the target and were thus excluded from the model.

By running the Chi-Square test, eight predictor variables were identified as the most informative for the logistic regression model:

- 'Age': Age group of user
- 'spotify_subscription_plan': Current Spotiy plan
- 'preffered_premium_plan': Price willing to pay for subscription
- 'preferred_listening_content': Favorite content
- 'music_lis_frequency': Music listen frequency
- 'music_expl_method': Method of exploring musics
- 'pod_lis_frequency': Podcast listening frequency
- 'pod_variety_satisfaction': Podcast satisfaction score

2.2.3 Model Training and Evaluation

The dataset was split into training and testing sets, with 80% of the data used for training and 20% for testing. The logistic regression model was then instantiated from the scikit-learn library and trained on the training set using the selected predictor variables.

To evaluate the performance of the trained logistic regression model, several metrics were computed on the test set:

- **Accuracy**: The model achieved an accuracy of 0.7833, indicating that it correctly predicted the upgrade willingness for 78.33% of the users in the test set.

	Precision	Recall	F-1
0	0.79	0.89	0.84
1	0.76	0.59	0.67

Accuracy = 0.783333333333

Figure 6: Model Performance

The precision and F1-score values indicated good performance, while the recall was lower. This suggests that the model may have a tendency to misclassify some users who were actually willing to upgrade as not willing to upgrade (higher false negatives).

- **Confusion Matrix**: A confusion matrix was generated to provide a detailed breakdown of the model's predictions.

	Predicted: No	Predicted: Yes	
Actual: No	68	8	76
Actual: Yes	18	26	44
	86	34	

Figure 7: Confusion Matrix

The matrix revealed that out of the actual negative cases (users not willing to upgrade), the model correctly predicted 68 as negative (true negatives) and incorrectly predicted 8 as positive (false positives). Among the actual positive cases (users willing to upgrade), the model incorrectly predicted 18 as negative (false negatives) and correctly predicted 26 as positive (true positives).

2.2.4 Hyperparameter Tuning

To further improve the model's performance, hyperparameter tuning was conducted using GridSearchCV from the scikit-learn library. GridSearchCV exhaustively searched over a specified parameter grid to find the best combination of hyperparameters that maximized the model's accuracy.

The parameter grid included different values for the regularization strength (C), the penalty type (L1 or L2), and the optimization algorithm (liblinear or saga). The grid search was performed using 5-fold cross-validation to ensure robust hyperparameter selection.

The best hyperparameters identified through the grid search were then used to train the final logistic regression model.

The tuned model demonstrated improved performance, particularly in predicting users who were willing to upgrade (class 1). Both precision and recall for class 1 increased, indicating that the hyperparameter tuning helped the model better capture the characteristics of users likely to upgrade without compromising its performance on the majority class.

	Precision	Recall	F-1
0	0.80	0.89	0.84
1	0.77	0.61	0.68

Accuracy = 0.79166666666

Figure 8: Model Performance after Tuning

2.2.5 Equation & Interpretation

The equation of predicting y (willingness to upgrade) can be represented as:

$$
\begin{aligned}
y = & -1.48 \times \text{Age} && (10)\\
& + 1.57 \times \text{Current Spotify plan} && (11)\\
& - 3.51 \times \text{Premium Plan Price} && (12)\\
& + 0.36 \times \text{Favorite Music Content} && (13)\\
& + 0.10 \times \text{Music Listen Frequency} && (14)\\
& + 1.15 \times \text{Music Discovery Method} && (15)\\
& - 0.66 \times \text{Podcast Listen Frequency} && (16)\\
& + 1.60 \times \text{Podcast Satisfaction} && (17)\\
& + 0.79 && (18)
\end{aligned}
$$

- **Age**: The coefficient of -1.48 suggests that older users are less likely to upgrade to a premium subscription compared to younger users. For each unit increase in age, the log-odds of upgrading decrease by 1.48, keeping other variables constant.

- **Current Spotify Plan**: The coefficient of 1.57 indicates that users who are currently on a paid Spotify plan are more likely to continue their premium subscription in the future. Being on a paid plan increases the log-odds of upgrading by 1.57 compared to being on a free plan, holding other variables constant.

- **Premium Plan Price**: The coefficient of -3.51 suggests that higher premium plan prices have a strong negative impact on the likelihood of upgrading. For each unit increase in the premium plan price, the log-odds of upgrading decrease by 3.51, keeping other variables constant. This highlights the price sensitivity of users when considering a premium subscription.

- **Favorite Music Content**: The coefficient of 0.36 indicates that users who prefer certain types of music content are slightly more likely to upgrade to a premium subscription. A one-unit increase in the preference for specific music content increases the log-odds of upgrading by 0.36, holding other variables constant.

- **Music Listen Frequency**: The coefficient of 0.10 suggests that users who listen to music more frequently on Spotify have a slightly higher likelihood of upgrading. Each unit increase in music listening frequency increases the log-odds of upgrading by 0.10, keeping other variables constant.

- **Music Discovery Method**: The coefficient of 1.15 indicates that the way users discover new music on Spotify has a positive impact on their likelihood of upgrading. Users who rely more on Spotify's music discovery features, such as curated playlists or personalized recommendations, have higher log-odds of upgrading by 1.15 compared to those who use other discovery methods, holding other variables constant.

- **Podcast Listen Frequency**: The coefficient of -0.66 suggests that frequent podcast listeners on Spotify are less likely to upgrade to a premium subscription. Each unit increase in podcast listening frequency decreases the log-odds of upgrading by 0.66, keeping other variables constant.

- **Podcast Satisfaction**: The coefficient of 1.60 indicates that users who are highly satisfied with the podcast content on Spotify are more likely to upgrade to a premium subscription. A one-unit increase in podcast satisfaction increases the log-odds of upgrading by 1.60, holding other variables constant.

The intercept term of 0.79 represents the log-odds of upgrading when all predictor variables are zero. It provides a baseline probability of upgrading in the absence of other factors.

In conclusion, the logistic regression modeling process employed in this study demonstrates a systematic approach to understanding and predicting users' willingness to upgrade to a premium Spotify subscription. By pre-processing the data, selecting informative features through heatmap analysis and the Chi-Square test, training and evaluating the model, and fine-tuning its hyperparameters, we have developed a robust predictive model that achieves an accuracy of 79.17% on the test set.

3 Evaluation

One strength of this study is the combined use of primary survey data and secondary open-source data, providing a more comprehensive understanding of Spotify's diverse user base.

By merging direct user responses with larger-scale dataset insights, the research offers a well-rounded perspective on user preferences and behaviors. Additionally, the application of logistic regression and persona development techniques aligns with industry best practices for user research and experience design, enhancing the practical relevance and applicability of the findings.

However, the study's weaknesses include the lack of more nuanced qualitative insights into users' motivations, pain points, and aspirations. While the survey captured demographic and preference data, incorporating in-depth interviews or focus groups could provide richer context for persona development and user experience optimization.

Moving forward, there are several opportunities for future research in this domain. Conducting longitudinal studies that track user behavior and preferences over time could provide valuable insights into how user needs and motivations evolve, informing the development of more adaptive and personalized user experiences. Exploring the impact of emerging technologies, such as voice assistants and extended reality (XR), on music streaming consumption and user experiences could uncover new opportunities for innovation and differentiation in the industry. Also, investigating the role of social and community features in music streaming platforms could shed light on the potential for fostering deeper user engagement and loyalty through shared experiences and interactions.

4 Data and Code Availability Statement

The datasets, analytical code, and deliverables associated with this academic thesis are publicly available to ensure transparency, peer review, and encourage future research on similar topics. The repository includes all materials necessary to understand, replicate, and extend the findings of this study: Github Repository

References

Becker, L., & Jaakkola, E. (2020). Customer experience: Fundamental premises and implications for research. *Journal of the Academy of Marketing Science*, *48*, 630–648.

Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). Know your customers' jobs to be done. *Harvard business review*, *94*(9), 54–62.

Ciocca, S. (2020, April). How does spotify know you so well? https://medium.com/@sophiaciocca/spotifys-discover-weekly-how-machine-learning-finds-your-new-music-19a41ab76efe

Cooper, A. (1999). *The inmates are running the asylum*. Springer.

Davis, F. D., et al. (1989). Technology acceptance model: Tam. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption*, *205*, 219.

GilPress. (2024, March). Spotify statistics for 2024: Users, demographics. https://whatsthebigdata.com/spotify-statistics/#google_vignette

Goltz, S. (2014). A closer look at personas: What they are and how they work: Part 1. *Smashing magazine*.

Hagen, J. (2015). *Industrial catalysis: A practical approach*. John Wiley & Sons.

Hamari, J., & Keronen, L. (2017). Why do people buy virtual goods: A meta-analysis. *Computers in Human Behavior*, *71*, 59–69.

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *The public opinion quarterly*, *37*(4), 509–523.

Kim, J., Lee, J., & Zo, H. (2018). Toward sustainable freemium software: The roles of user satisfaction and use context. *Journal of Electronic Commerce Research*, *19*(3).

King, D. L., Delfabbro, P. H., Perales, J. C., Deleuze, J., Király, O., Krossbakken, E., & Billieux, J. (2019). Maladaptive player-game relationships in problematic gaming and gaming disorder: A systematic review. *Clinical psychology review*, *73*, 101777.

Kumar, V. (2014). Making" freemium" work. *Harvard business review*, *92*(5), 27–29.

Lee, J. H., & Waterman, N. M. (2012). Understanding user requirements for music information services. *ISMIR*, 253–258.

Oestreicher-Singer, G., & Zalmanson, L. (2013). Content or community? a digital business strategy for content providers in the social age. *MIS quarterly*, 591–616.

Pruitt, J., & Grudin, J. (2003). Personas. *Proceedings of the 2003 conference on Designing for user experiences*. https://doi.org/10.1145/997078.997089

Pruitt, J. S., & Adlin, T. (2006). Overview of the persona lifecycle. *The Persona Lifecycle*, 46–65. https://doi.org/10.1016/b978-012566251-2/50003-4

Salminen, J., Jansen, B. J., An, J., Kwak, H., & Jung, S.-g. (2018). Are personas done? evaluating their usefulness in the age of digital analytics. *Persona Studies*, *4*(2), 47–65.

Sinclair, G., & Tinson, J. (2017). Psychological ownership and music streaming consumption. *Journal of Business Research*, *71*, 1–9.

Spiliotopoulos, D., Margaris, D., & Vassilakis, C. (2020). Data-assisted persona construction using social media data. *Big Data and Cognitive Computing*, *4*(3), 21.

Thomes, T. P. (2013). An economic analysis of online streaming music services. *Information Economics and Policy*, *25*(2), 81–91.

Wagner, T. M., Benlian, A., & Hess, T. (2014). Converting freemium customers from free to premium—the role of the perceived premium fit in the case of music as a service. *Electronic Markets*, *24*, 259–268.

Wedel, M., & Kannan, P. (2016). Marketing analytics for data-rich environments. *Journal of marketing*, *80*(6), 97–121.

Wlömert, N., & Papies, D. (2016). On-demand streaming services and music industry revenues—insights from spotify's market entry. *International Journal of Research in Marketing*, *33*(2), 314–327.

Xu, Y., & Lee, M. J. (2020). Identifying personas in online shopping communities. *Multimodal Technologies and Interaction*, *4*(2), 19.

Appendix

Spotify Usage Survey

Please take a few minutes to complete this survey about your Spotify usage. Your responses will help us understand user preferences and habits.

1. What is your age group?

A. 12-20 B. 20-35 C. 35-60

2. What is your gender?

A. Male B. Female C. Others

3. How long have you been using Spotify?

A. Less than 6 months B. 6 months to 1 year C. 1 year to 2 years D. More than 2 years

4. Which of the following devices do you primarily use to listen to Spotify? (Multiple choices allowed)

A. Smartphone B. Computer or laptop C. Smart speakers or voice assistants

5. Which Spotify subscription plan do you currently have?

A. Free (ad-supported) B. Premium (specify plan if known) C. Individual Plan D. Student Plan E. Family Plan

6. Are you willing to take a premium subscription or continue with a premium subscription in the future?

A. Yes B. No C. Unsure

7. If you are premium or considering premium, what specific plan are you on or considering?

A. Individual Plan - $10.99/month B. Student Plan - $5.99/month C. Family Plan - $16.99/month

8. What do you prefer to listen to more?

A. Music B. Podcasts

9. What genre(s) of music do you enjoy the most? (Multiple choices allowed)

A. Rap B. Melody C. Pop D. Old songs

10. What is your favorite time slot to listen to music?

A. Morning B. Afternoon C. Evening D. Night

11. When do you listen to music? (Multiple choices allowed)

A. Study Hours B. While Traveling C. Office hours D. Workout session E. Leisure time

12. How do you discover new music on Spotify? (Multiple choices allowed)

A. Playlists B. Recommendations

13. How do you rate the Spotify music recommendations?

A. 1 (Poor) B. 2 C. 3 D. 4 E. 5 (Excellent)

14. How often do you listen to podcasts?

A. Daily B. Several times a week C. Once a week D. Rarely E. Never

15. Do you prefer shorter podcast episodes (under 30 minutes) or longer episodes (over 30 minutes)?

A. Shorter B. Longer C. Both

16. Are you satisfied with the variety and availability of podcasts on Spotify?

A. Very satisfied B. Satisfied C. Ok D. Unsatisfied E. Very dissatisfied