Motivation: Why are you working on this project?

Firstly, as an avid music enthusiast, I am genuinely interested in gaining a deeper understanding of my own listening habits, favorite genres, and preferred artists. Exploring my Spotify data allows me to unravel patterns and trends that reflect my evolving musical tastes over the course of the year.

Beyond my personal interest in music, I am also eager to further develop my data exploration and analysis skills. This project provides a practical application of various data science techniques, allowing me to enhance my proficiency in data visualization, statistical analysis, and potentially machine learning applications.

Moreover, I see this endeavor as an opportunity to customize my Spotify recommendations. By delving into the nuances of my music preferences, I aim to optimize my listening experience and discover new artists and genres that align with my evolving tastes.

Data source: Where did you get this data? How did you collect it?

I got it by requesting for my own data (2014-2023) from Spotify itself and also, using the Spotify Web API (Spotify for developers) which I generated with my personal account. With the two datasets, I interchangeably utilise them by getting the data from my own account then extracting additional information (eg. audio features and genres) from the API.

Data analysis: Techniques used, different stages of the analysis

Introduction:

(1) Data Cleaning and Preprocessing: Handling missing values, Data type conversions, Converting date strings to datetime objects, Filtering and selecting relevant columns or rows.

Content:

(1) Data Visualization: Utilizing matplotlib and seaborn for creating various types of plots: bar charts, scatter plots, line plots, & employing WordCloud for visualizing word frequencies.

(2) API Interaction: Interacting with the Spotify API using Spotipy library by retrieving information about artists, tracks, and audio features.

(3) Feature Engineering: Creating new features, such as 'genres' and 'main_genre'. Binning and classifying data into groups based on specific criteria.

(4) Error Handling: Handling exceptions and errors during API calls and emoving or dropping rows with missing values.

(5) Text Processing: Extracting languages from track names and utilizing natural language processing libraries (langdetect) for language detection.

(6) Classification and Categorization: Classifying artists into genres (K-pop, Pop, Other).

(7) Statistical Analysis/ Descriptive statistics: Calculating sum, mean, and sorting and ranking data based on specific criteria.

(8) Time Series Analysis: Analyzing trends over time, e.g., monthly playback hours and grouping data by month and visualizing time series data.

These techniques collectively help in exploring, understanding, and gaining insights from the provided data, demonstrating a comprehensive approach to data analysis.

Part 1: Listening Habits

(1) Minimum Listening Hours:

During weeks where I recorded a minimum of 2 hours of Spotify usage, a consistent pattern emerges. Notably, these weeks coincide with periods of holiday, providing a clear correlation between leisure time and music consumption.

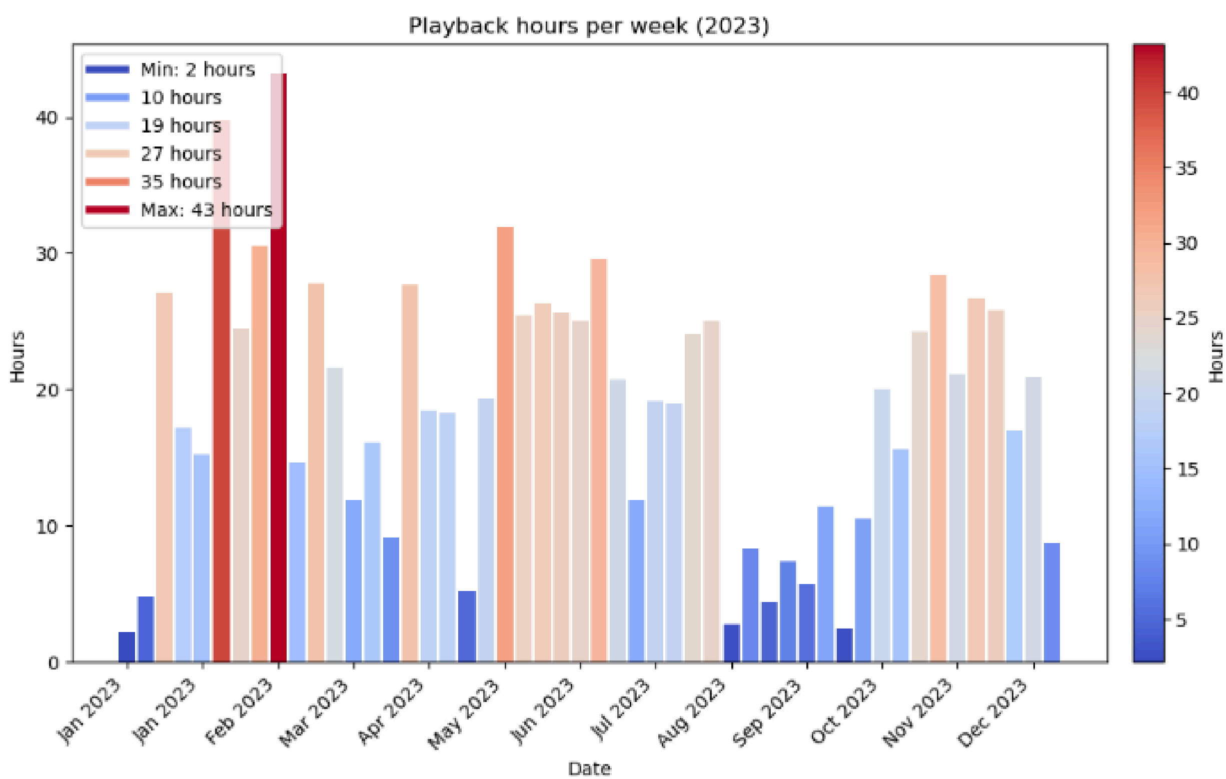
An exception occurs in September, where a dip in listening hours is observed. This anomaly is attributed to the transition period associated with my exchange to Türkiye, suggesting that the relocation process likely took precedence over leisure activities such as music listening.

(2) Maximum Listening Hours:

The week with the highest Spotify usage, amounting to 43 hours, falls in the first week of February. This spike can be attributed to the intense preparations for a performance at SMU (Singapore Management University, my home university) during this period. The demand for focus and concentration and the need to prepare for the performance probably raised my music listening, especially more of the song I performed.

(3) Consistent Peaks in June:

Noteworthy is the consistent peak in listening hours throughout June. This pattern is attributed to the preparation for another dance performance scheduled for mid-July 2023. The heightened activity during this period suggests that music played a crucial role in my preparation and rehearsal routines.



(1) Night Owl Listening Habits:

The consistent trend of music listening until the early morning hours aligns with my self-identification as a night owl. It suggests that late-night hours serve as a prime time for me to engage with music, indicating a preference for nocturnal leisure and relaxation.

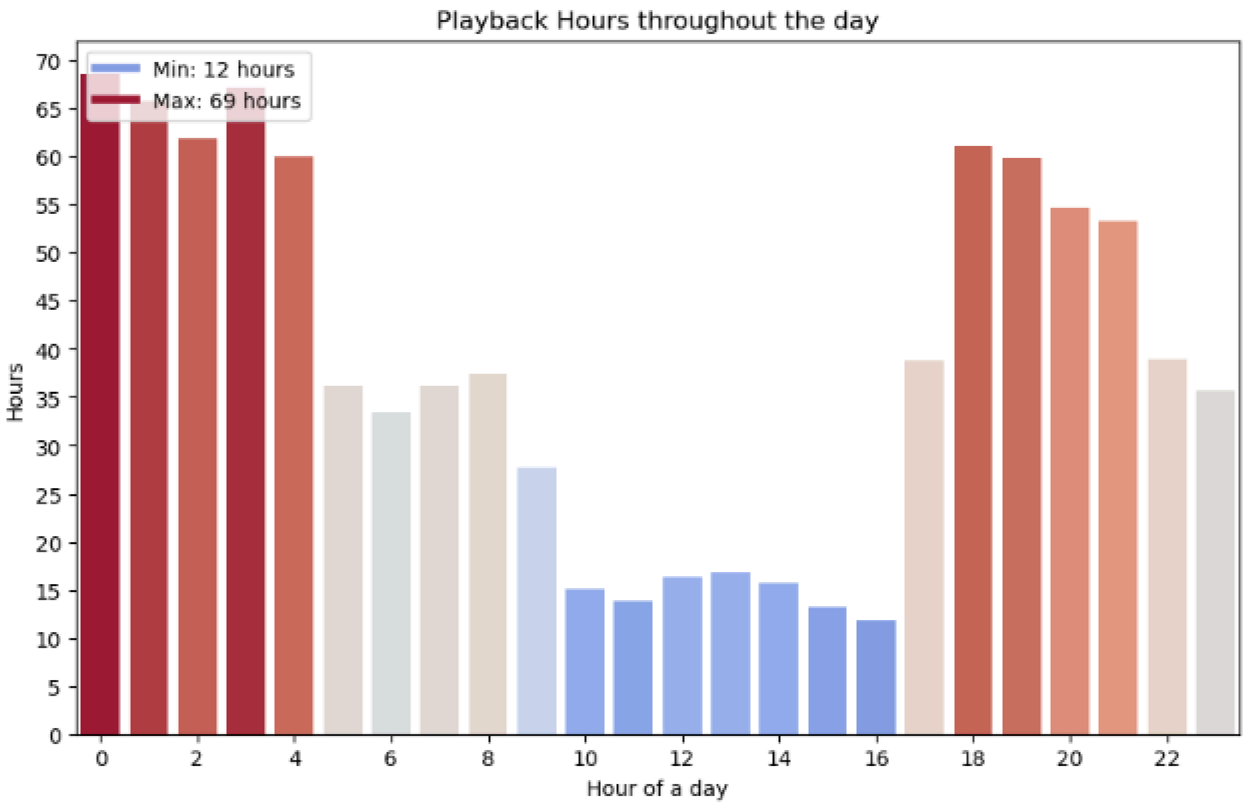
(2) Work and Music Integration:

The period from 5 pm to 4 am emerges as a significant window for music consumption. This timeframe corresponds with my engagement in school assignments, implying that music serves as a companion during work hours. This finding supports the notion that music plays a role in enhancing focus and productivity during academic pursuits.

(3) Sleep Schedule and Music Usage:

Notably, the data indicates that my minimum listening hours occur around 4 pm. This aligns with the time when I typically make my way back home. It is plausible that the decline in music usage during this period is influenced by the transition from public transportation to a more relaxed home environment, where other activities may take precedence over listening to music.

The data also suggests that I tend to wind down and eventually sleep from 5 am onwards. This aligns with a decline in Spotify usage during these hours, indicating a transition from active engagement with music to a more restful state.



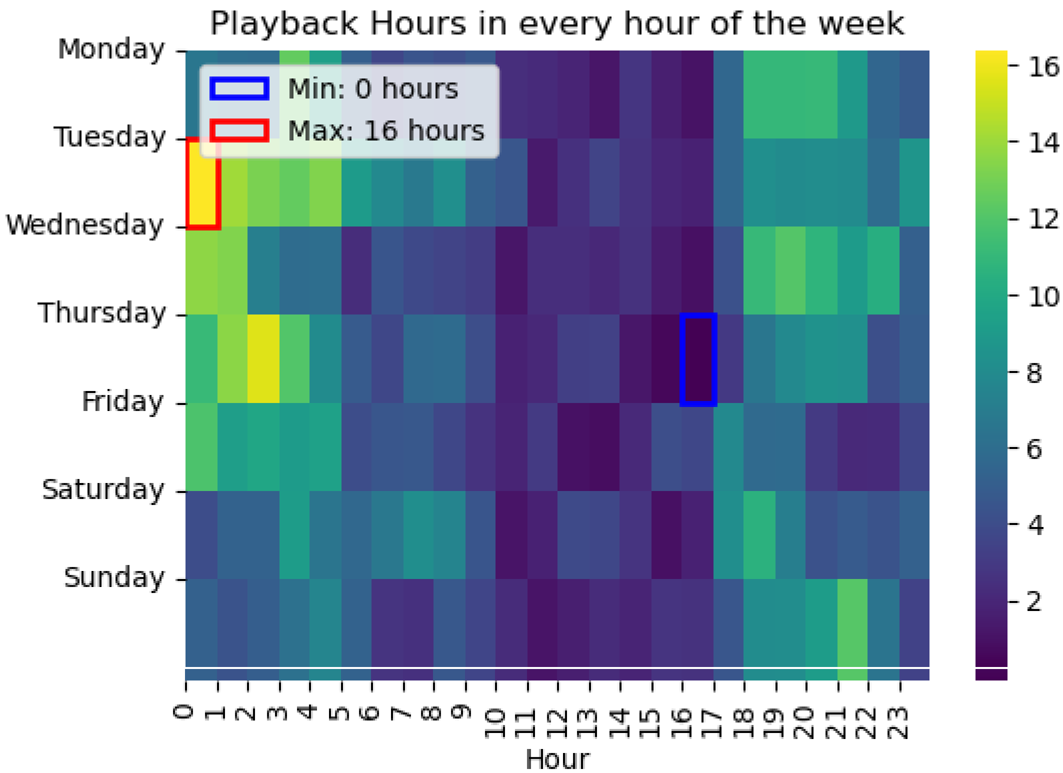
(1) Thursday 5 pm – Class Schedule Impact:

A noticeable dip in music listening on Thursdays at 5 pm correlates with the timing of my classes during the first semester. This decline in Spotify usage during this specific time slot aligns with the reasonable assumption that academic commitments, such as classes, take precedence over leisure activities like listening to music.

(2) Tuesday Midnight – Optimal Listening Window:

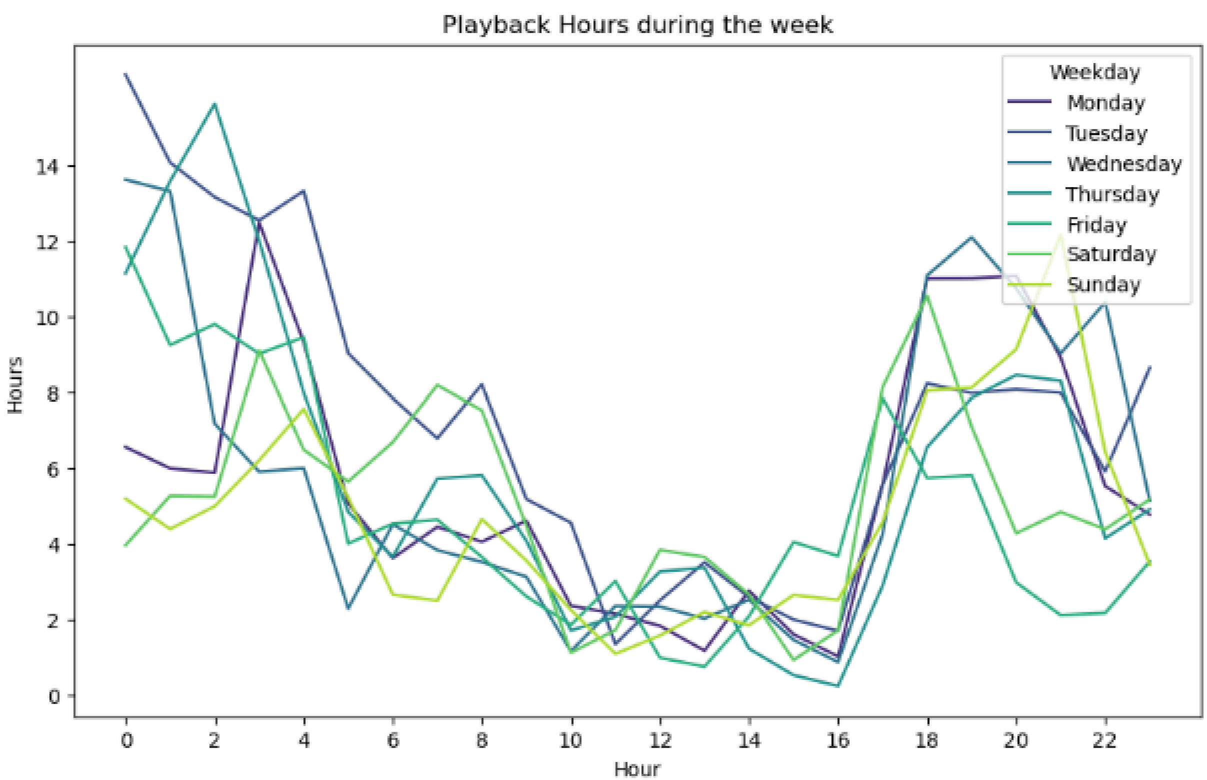
The data highlights Tuesday midnight as the time when I engage most with Spotify. This observation corresponds with the absence of classes on Wednesdays, creating an opportunity for extended activities on Tuesday nights. The lack of early morning commitments on Wednesdays likely allows me to stay up late, contributing to increased music consumption during this time.

The ability to stay up late on Tuesday nights provides an optimal window for leisure activities, including music listening, as it aligns with the anticipation of a more flexible schedule the following day.



Consistent Tuesday Peaks:

Across all hours of the day, Tuesdays consistently emerge as a day with elevated Spotify usage. This trend is notable as it remains one of the highest among all days of the week at every point in the hourly distribution.



Part 2: Content Analysis

Extended Listening Sessions:

The analysis reveals a consistent pattern of binge listening, where I tend to engage in longer and more extended music sessions rather than shorter intervals.

```
Percentage of time spent binge listening: 83.91%
```

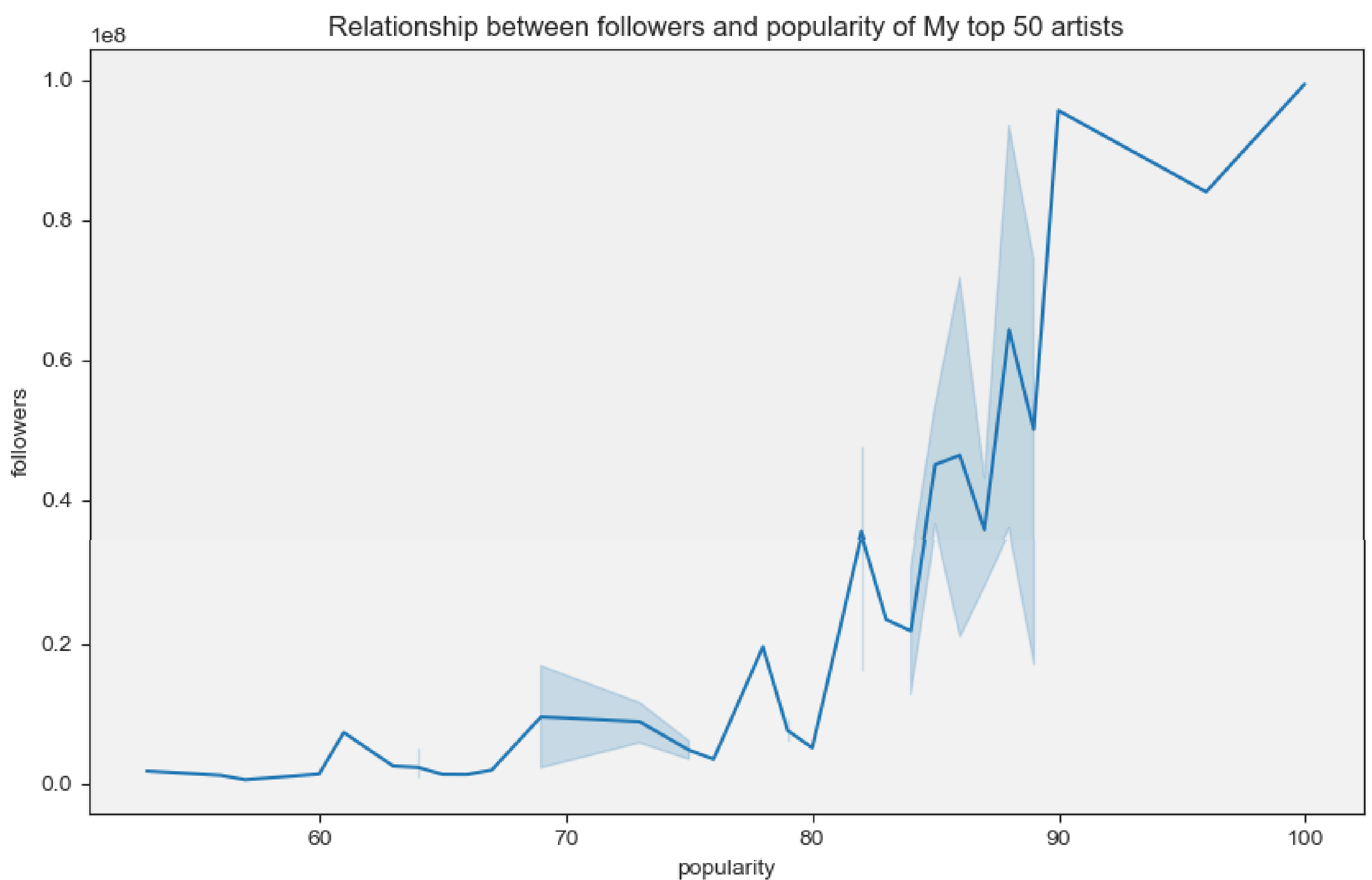
(1) Artist Popularity and Listening Preferences:

The observation that the popularity scores of the artists in my Spotify listening history fall within the range of 50 to 100 suggests a preference for well-known and popular artists. This aligns with the idea that listeners often gravitate towards familiar and mainstream music.

(2) Positive Relationship Between Popularity and Followers:

A notable correlation is also observed between the popularity of artists and the number of followers they have. In general, artists with higher popularity scores tend to have a larger following. This positive relationship aligns with the conventional understanding that more popular artists attract a larger fan base.

The positive correlation between artist popularity and followers may be indicative of a general trend in the music industry, where widespread recognition translates into increased support and following. Thus, I have no shame but say I do listen to music from popular artists with high followers count in order to not miss out on the up and coming music trends.



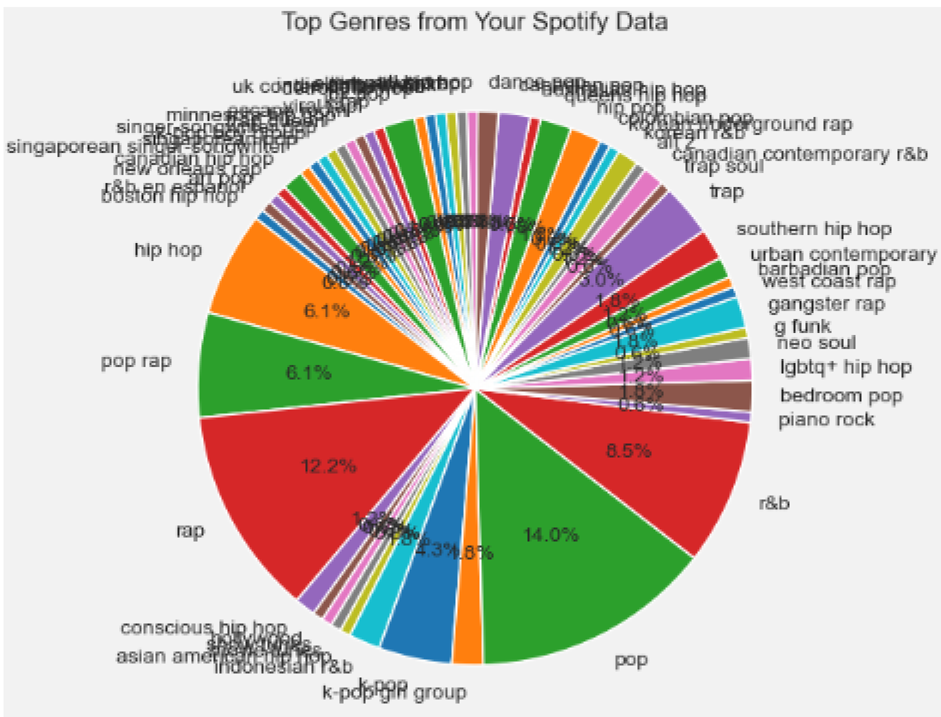
(1) Dominant Genres:

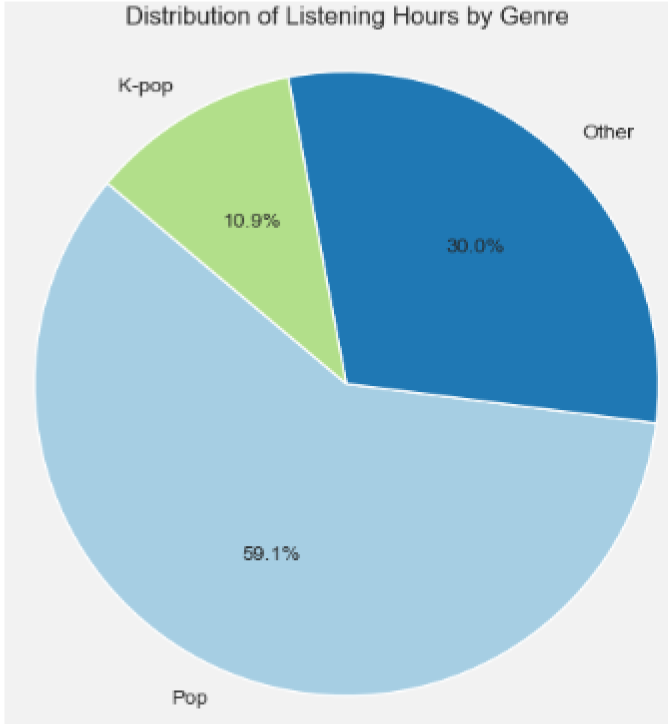
The pie chart indicates that the majority of my listening preferences are attributed to Pop, constituting 60% of my overall music consumption. This suggests a strong affinity for mainstream and popular music within the Pop genre.

K-Pop emerges as another notable genre, accounting for 10.9% of my listening habits. The inclusion of K-Pop underscores a diverse range of musical interests, incorporating global and contemporary genres into my playlist.

(2) Other Genres:

The remaining 29.1% is distributed among various genres, including but not limited to Hip Hop, Trap, Soul and R&B. This diversity in the "Others" category signifies a well-rounded taste, encompassing a mix of urban, contemporary, and international musical styles.





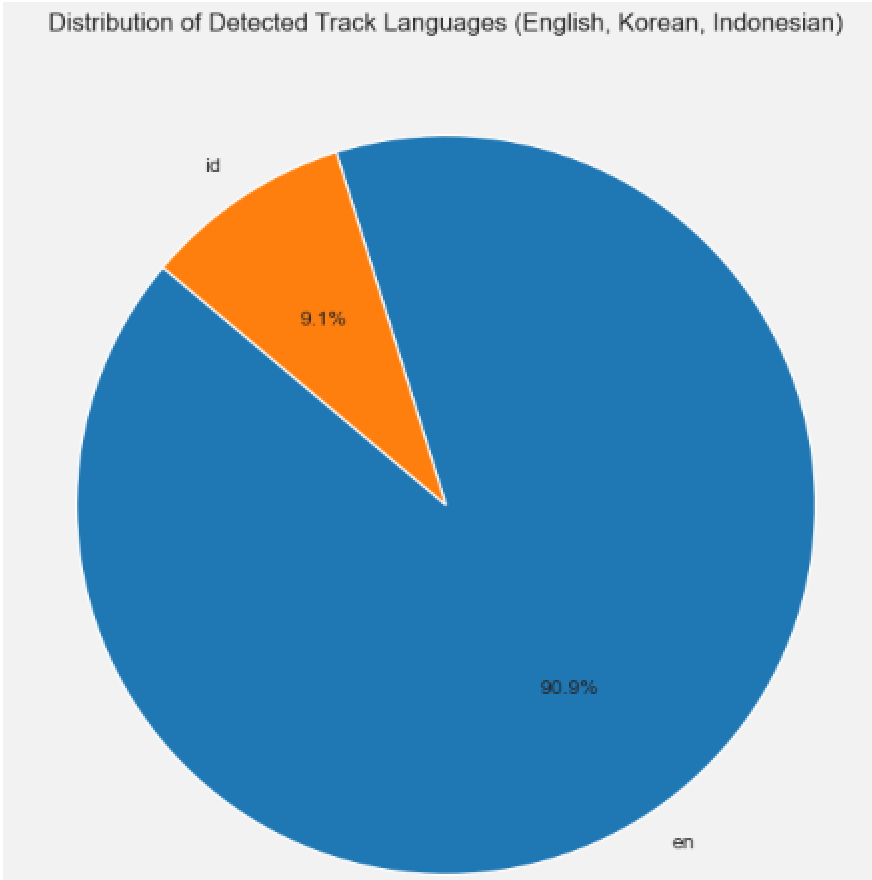
(1) Primary Language Distribution:

English emerges as the dominant language in my Spotify listening habits, constituting a significant 90% of the detected languages. This suggests a strong inclination towards English-language tracks in my music preferences.

The remaining 10% comprises primarily Indonesian, indicating a secondary interest in music from Indonesia.
(P.S. I am a Malay from Singapore. It's essential to note that the detected language may not always perfectly align with the actual language content. While the majority of words in Malay and Indonesian are similar or identical, there are some vocabulary differences due to the influence of other languages. Thus, I will be using Indonesian detection language in my analysis instead of Malay language.)

(2) Influence of K-Pop:

The surprise at the dominance of English over Korean in my listening habits, especially given my avid interest in K-Pop, highlights potential challenges in accurately detecting languages in music. K-Pop often involves a blend of Korean and English, and language detection algorithms may face difficulty discerning the primary language. Thus, there is some limitation in the language detection but nevertheless, this was an interesting finding.



(1) Word Cloud Analysis:

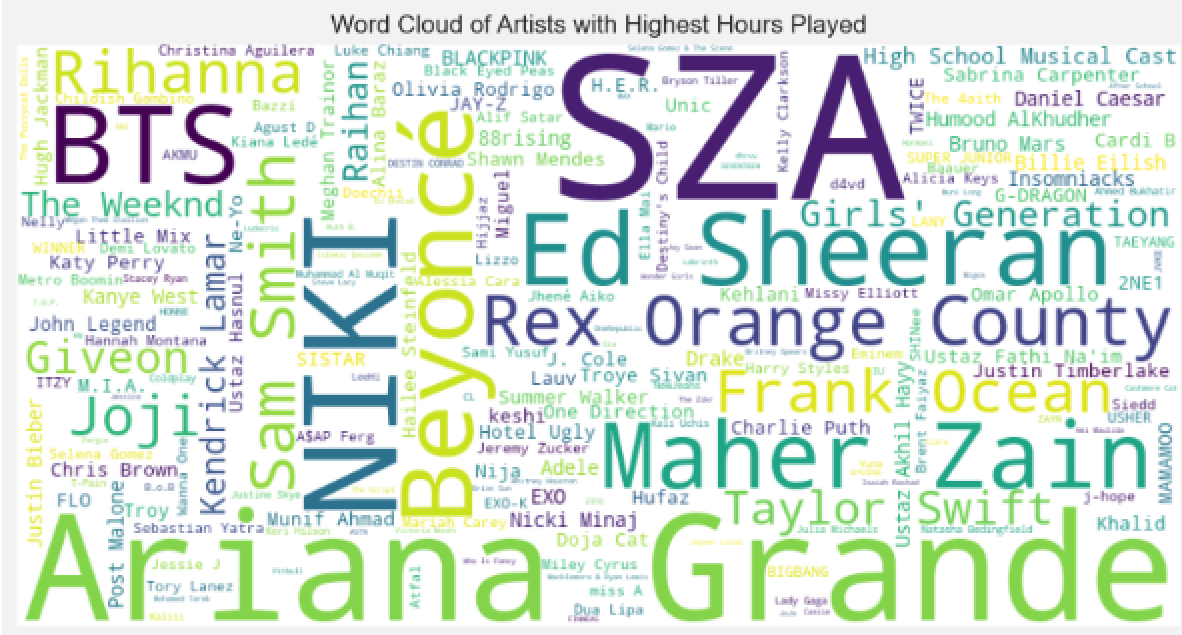
The word cloud visually represents the frequency of artist mentions, highlighting familiar and famous names. This aligns with the subsequent bar chart analysis, emphasizing the prevalence of well-known artists in my Spotify listening history.

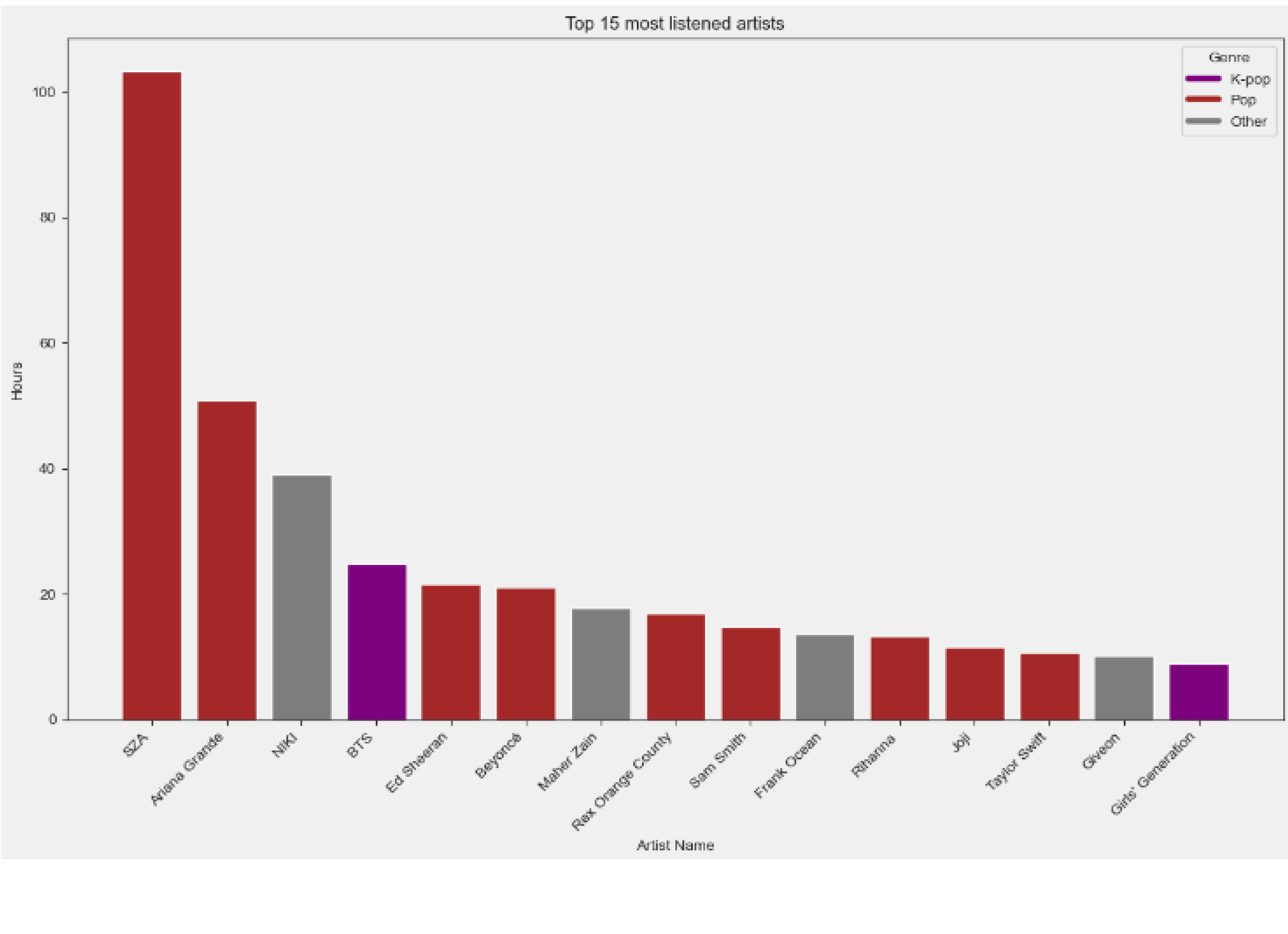
(2) Top 15 Artists Bar Chart:

The bar chart reveals that the majority of my top 15 artists belong to the Pop and K-pop genres, confirming a strong affinity for mainstream and global music trends. This pattern aligns with the recognition of familiar and famous artists in the word cloud.

(3) Discovery of SZA:

The revelation that I have become a SZA fan in the current year, coupled with the realization that I havelistened to her for over 100 hours in 2023, adds an element of surprise. This underscores the dynamic nature of music preferences and the potential for new discoveries, even within genres not previously explored extensively.


Word Cloud of Artists with Highest Hours Played
SZA
Ariana Grande
NIKI
Beyoncé
Ed Sheeran
Rex Orange County
Maher Zain
BTS
Rihanna
Sam Smith
Frank Ocean
Taylor Swift
Joji
Giveon
Girls' Generation
The Weeknd
Kendrick Lamar
High School Musical Cast
Sabrina Carpenter
Daniel Caesar
Humood AlKhudher
Bruno Mars
Billie Eilish
Cardi B
Insomniacks
G-DRAGON
Olivia Rodrigo
BLACKPINK
Christina Aguilera
Luke Chiang
H.E.R.
Unic
88rising
Shawn Mendes
Raihan
Little Mix
Katy Perry
Kanye West
John Legend
Chris Brown
Justin Bieber
Post Malone
Kehlani
Omar Apollo
Charlie Puth
Justin Timberlake
Drake
Troye Sivan
Summer Walker
One Direction
Hotel Ugly
Adele
Nicki Minaj
Doja Cat
Khalid
Miley Cyrus
Dua Lipa

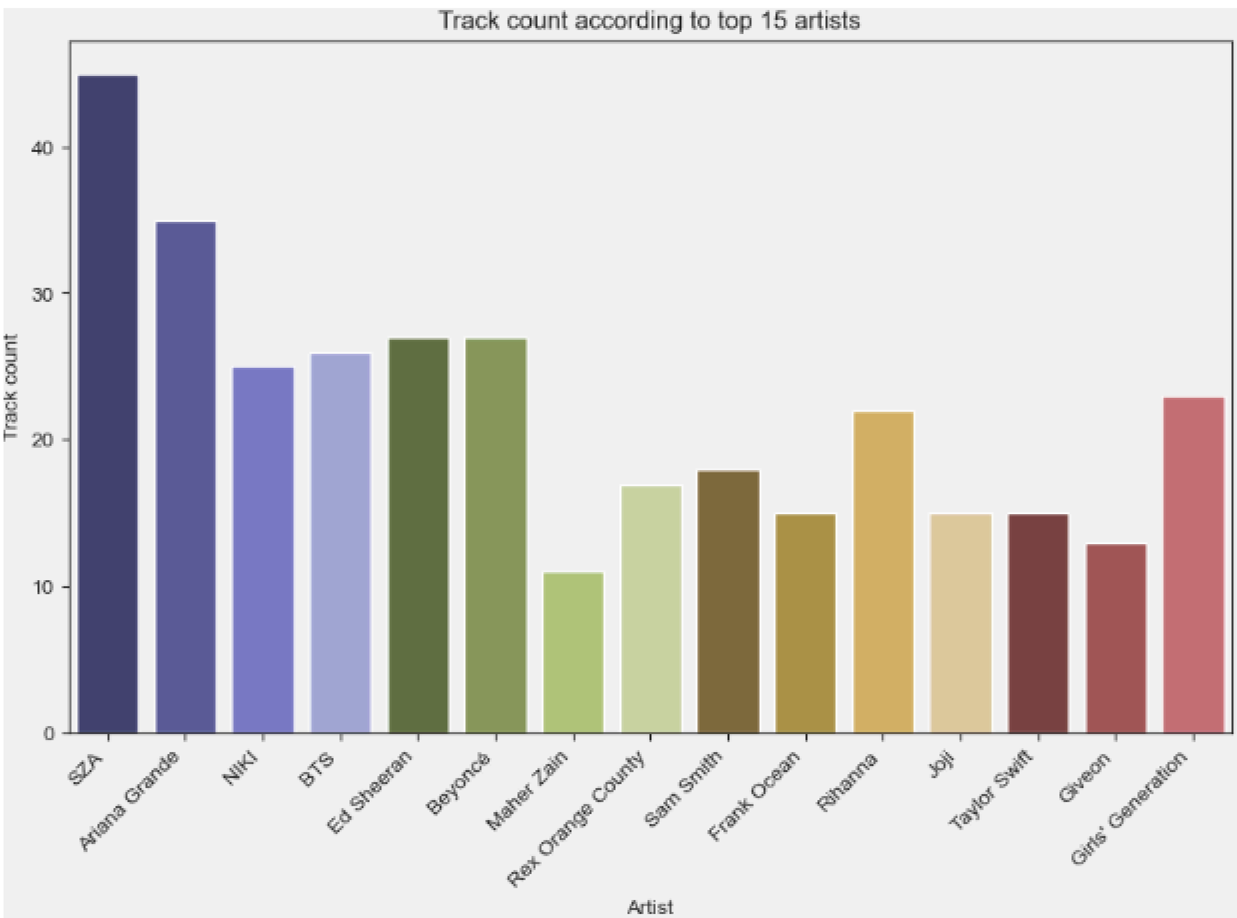

Top 15 most listened artists
Genre
K-pop
Pop
Other
Hours
0
20
40
60
80
100
SZA
Ariana Grande
NIKI
BTS
Ed Sheeran
Beyoncé
Maher Zain
Rex Orange County
Sam Smith
Frank Ocean
Rihanna
Joji
Taylor Swift
Giveon
Girls' Generation
Artist Name

(1) Surprising Maher Zain Ranking:

Maher Zain's notable position as my 7th favorite artist, despite a limited known song count of 11, challenges the assumption of a direct relationship between track count and artist ranking. This suggests that my affinity for an artist may not necessarily be dictated by the number of tracks I know from them. His unexpected ranking in my top 15 artists highlights the dynamic and multifaceted nature of listening preferences. It underscores the role of personal connection, emotional resonance, and individual track impact in shaping my overall appreciation for the artist.

(2) Quality over Quantity

Again, this just proves the significance of the quality and personal resonance of the tracks rather than the sheer quantity of songs known. Thus is why (though it can be assumed) there is no direct relationship between track count and artist ranking



(1) Listening Patterns Over Time:

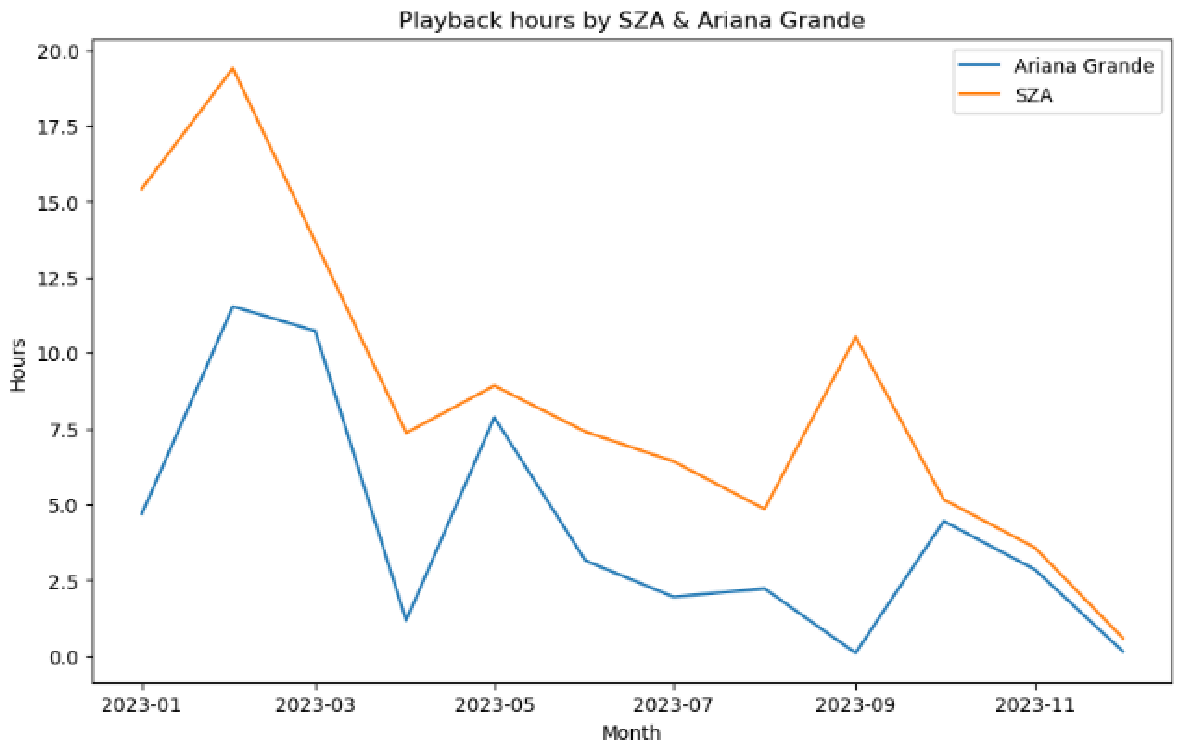
My top two favorite artists, SZA and Ariana Grande, have consistently been prominent in my listening habits throughout the year, reflecting a continuous appreciation for their music.

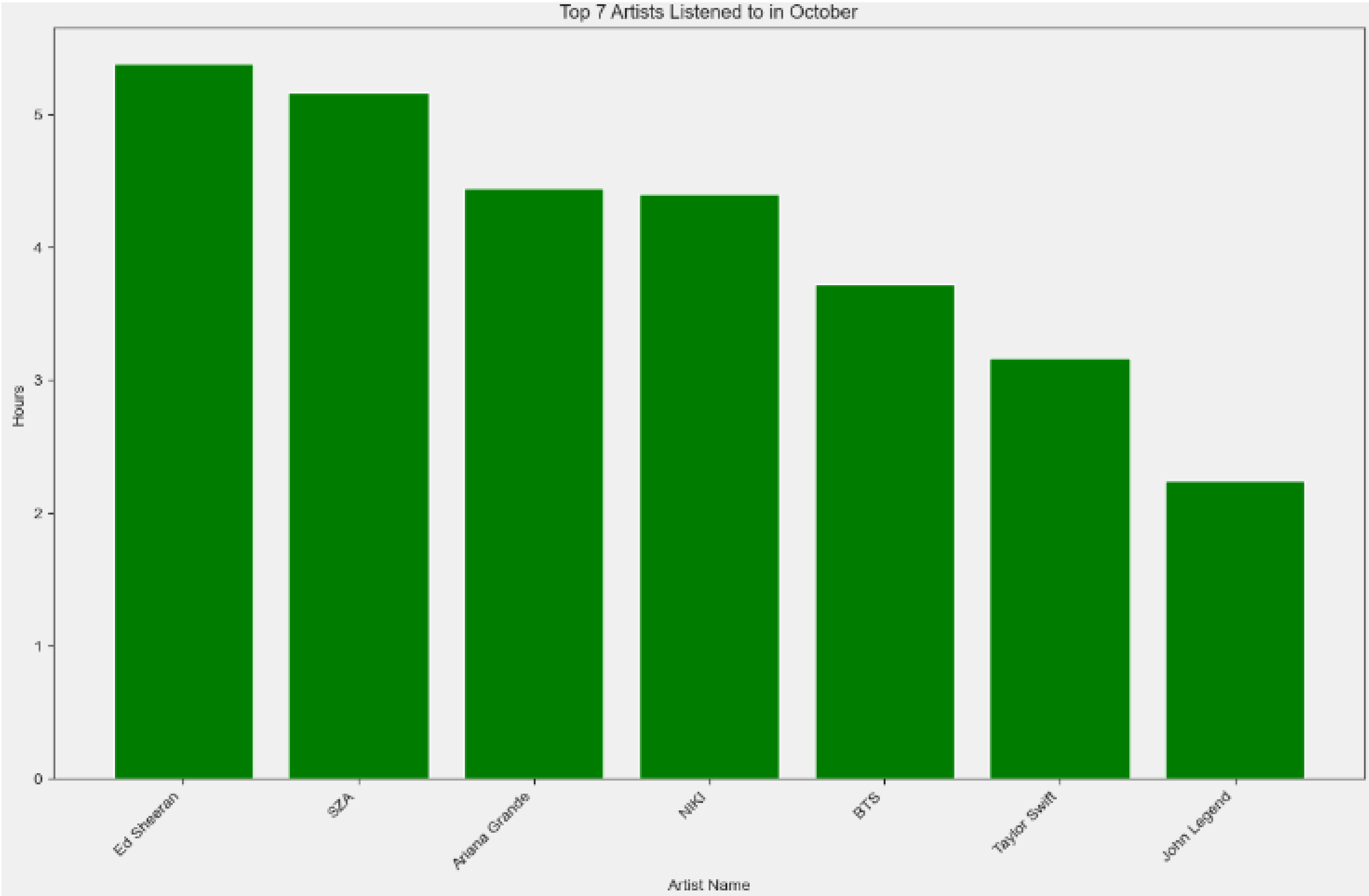
(2) Dip in October:

The investigation reveals a noticeable dip in the listening frequency of both SZA and Ariana Grande in October. Upon closer examination, this dip can be attributed to my heightened engagement with Ed Sheeran during that month.

(3) Ed Sheeran's Album Release:

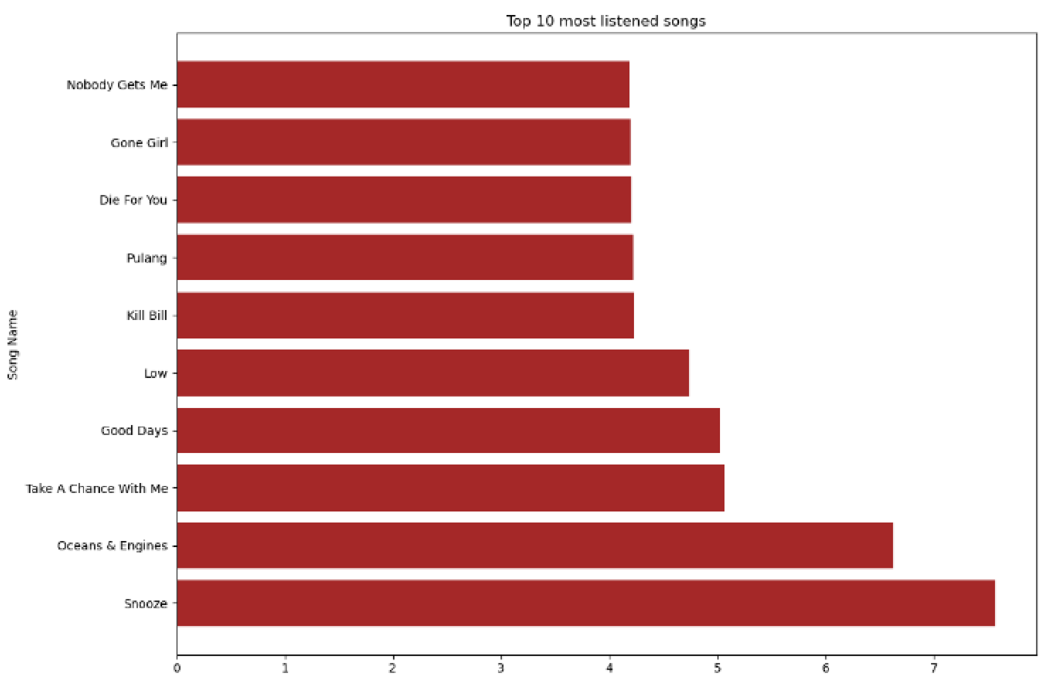
The release of Ed Sheeran's album "Autumn Variations" in late September aligns with the observed dip in listening to my top two favorite artists in October. The surge in Ed Sheeran's music consumption during that period indicates a significant focus on exploring and enjoying the new album in October 2023.





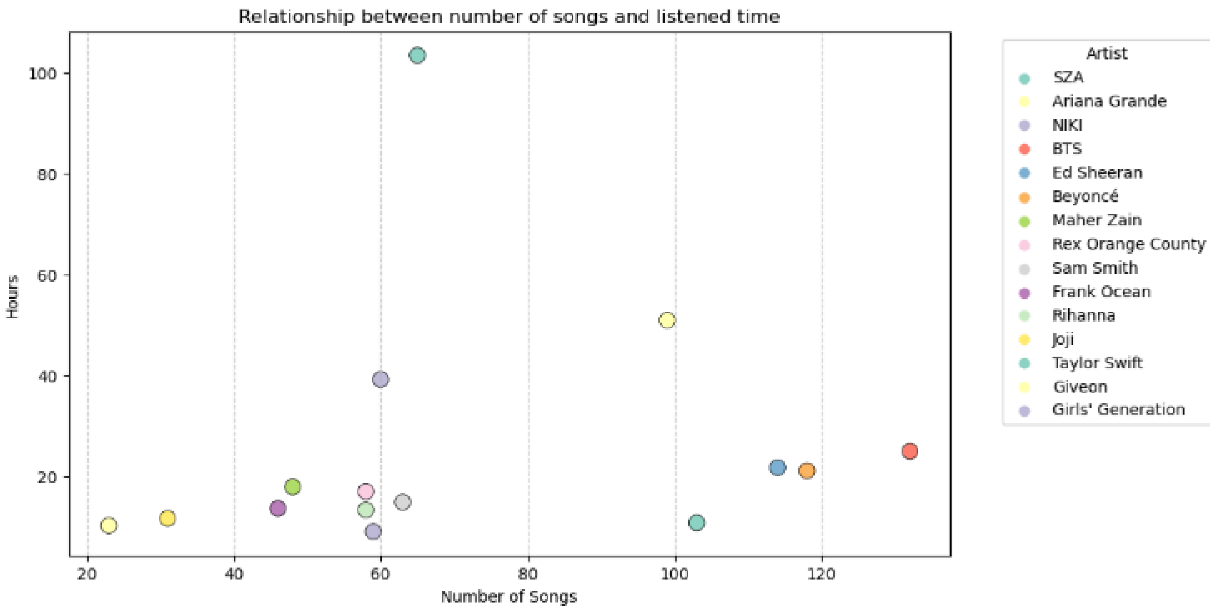
Surprising Lack of Familiarity:

Despite the frequent listening to "Snooze," it is noted that I still haven't memorized all the lyrics to the song. This discrepancy between high listening frequency and a lack of full memorization raises questions about the factors influencing song familiarity.



Low Proportionate Listening Frequency:

BTS has the most number of saved songs in my library, indicating a considerable collection of their tracks in my Spotify library. Despite having a large number of saved songs, the listening frequency for BTS is not proportionately high. This suggests that, while I have a significant collection of their songs, Ido not actively listen to BTS as frequently as the number of saved songs might imply

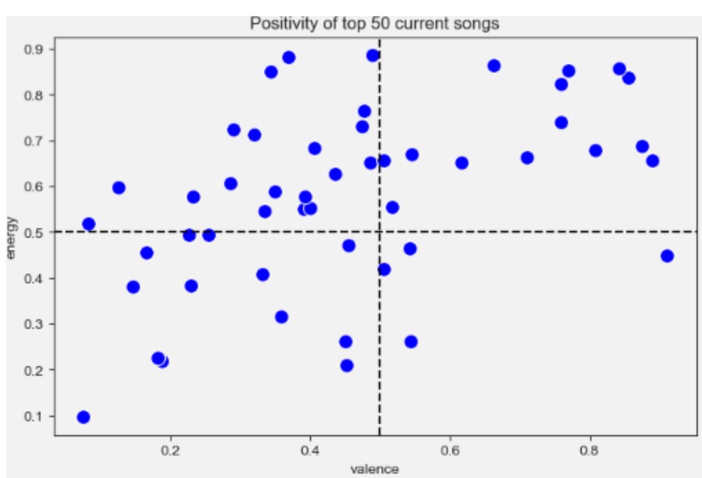


(1) High-Energy Preference:

My preference for high-energy music, characterized by fast tempo, loud volume, and upbeat qualities, is evident. This suggests a tendency to gravitate towards tracks that are dynamic, lively, and energizing.

(2) Valence Dynamics:

It's interesting to note that despite having an equal number of high and low valence tracks, all tracks share high-energy attributes. This indicates a unique and dynamic combination of positive (happy, cheerful, euphoric) and negative (sad, depressed, angry) emotions within the context of high-energy music.

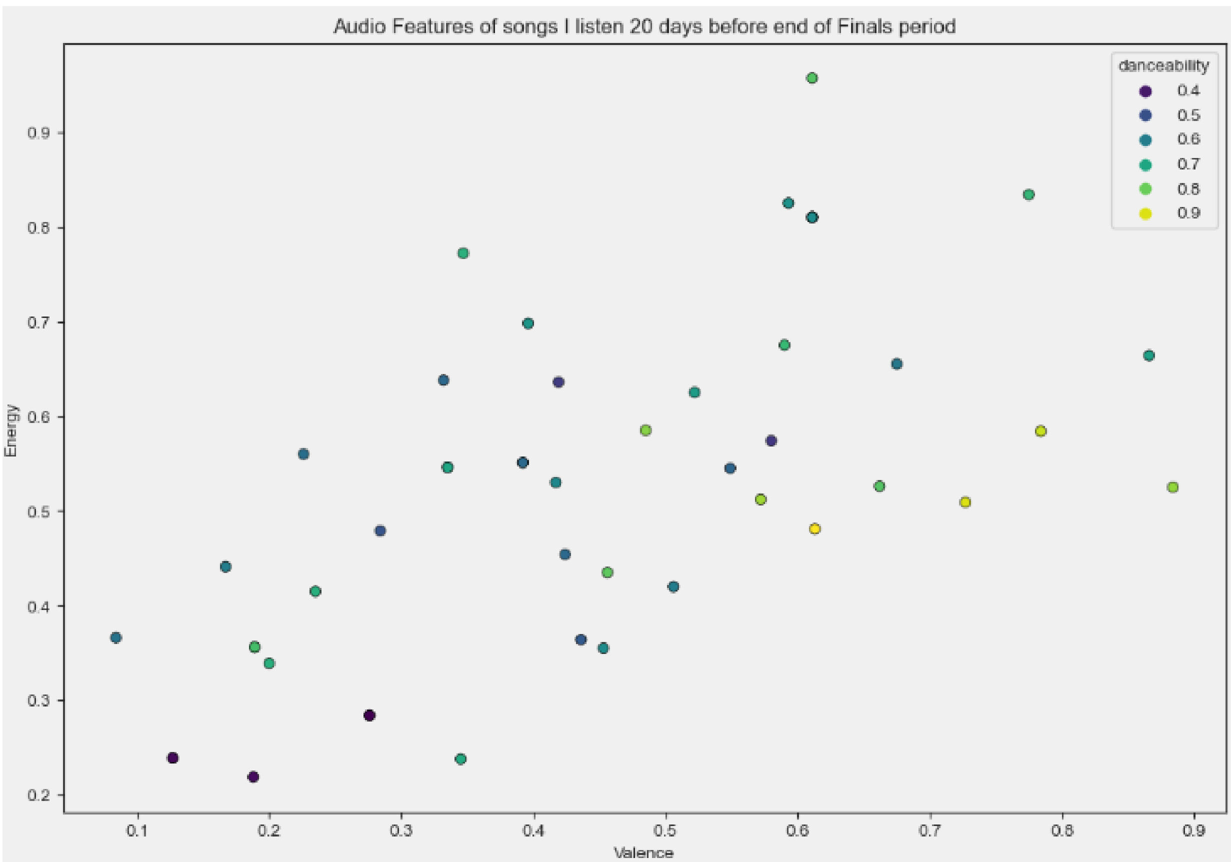


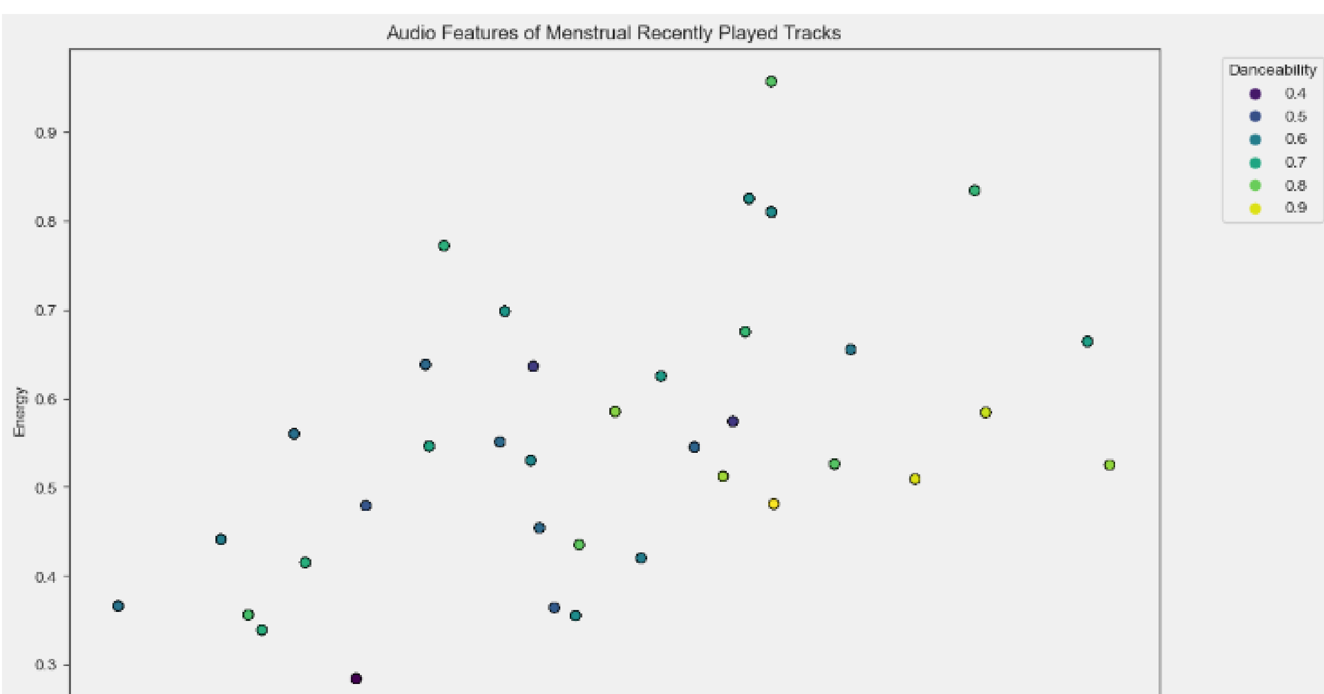
(1) No Discernible Patterns in External Factors:

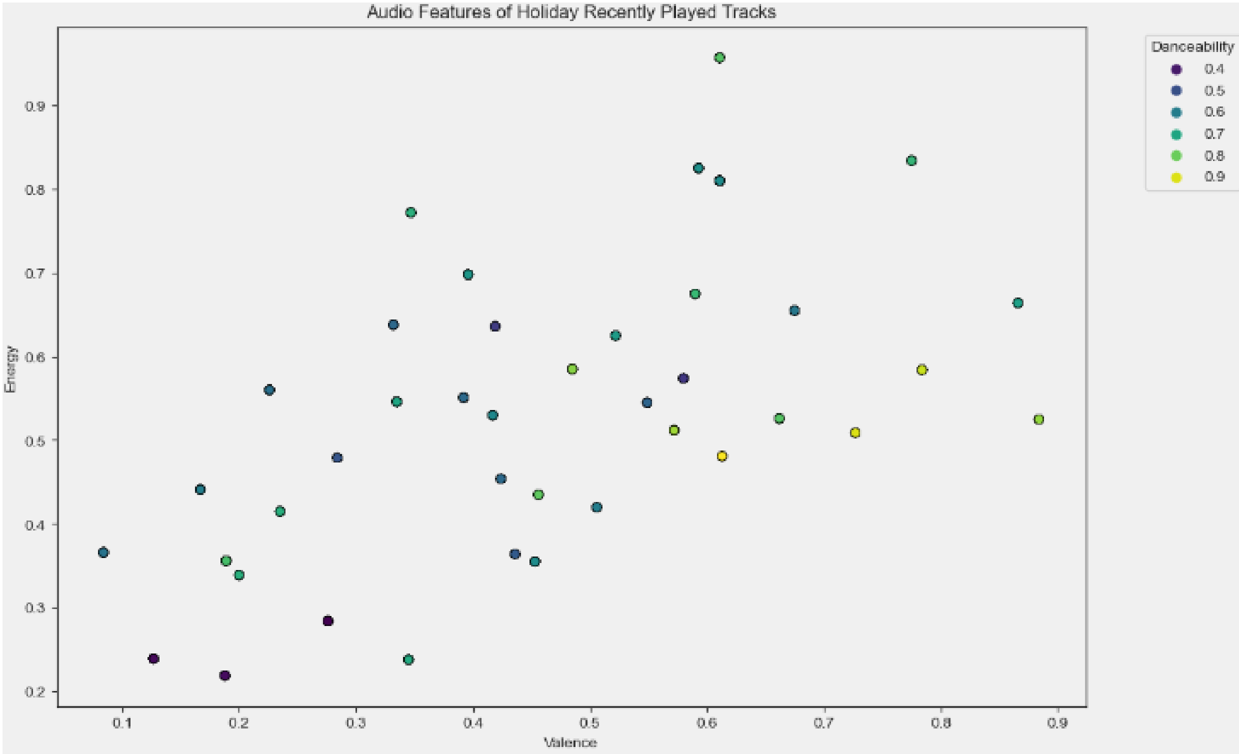
Despite the initial exploration, no clear patterns between my mood and external factors such as finals season, menstrual cycles, or public holidays were identified. The influence of these factors on my overall mood or music preferences appears to be complex and not easily discernible from the available data.

(2) Direct Relationship Between Energy and Valence:

However, a significant finding is the direct relationship between the level of energy and valence in my music preferences. This suggests that my go-to songs likely exhibit both high energy and high valence, indicating a preference for tracks that are not only dynamic and lively but also convey positive emotions.







Findings: What you learned about yourself?

I was really surprised to find out that I do not listen to songs of my mother tongue as much as I want to. I also did not realise that I barely listen to music when I'm out. I guess I tend to use other social media platforms or maybe be in the moment and enjoy ambience/ scenery. I learned that I am a big fan of SZA and Ariana Grande and will be looking forward to their concerts when they come to Singapore. I also learned that I like happy, fast songs all day all night.

Limitations and future work: What could be done better? Do you have any future plans about your project?

Such limitations like not being able to compare my top albums was a real downer as I wanted to do a comparison on my two favorite albums. There are also other limitations with the data given to me by Spotify and the data generated from API (as they only provide 50 results). For example, I may have thousands of data from my own data but require additional information from API but API will only generate the first 50 for me. I believe more conclusions can be derived if they increase the number of output from API. On top of that, due to the limited number of requests from Spotify API, I can only do XX number of analysis per day thus, it delayed my project by a lot.

Additionally, I wish I spent more time in making the visualizations interactive as it will definitely help discovering hidden patterns and enhance my ability to tell a compelling data story.

I think the biggest downside is my competency. I am a year 4 accounting undergraduate with 2 years of basic accounting in high school and 3 years diploma. I am no coder so it was really really tough for me to complete this coding project but I am thankful I managed to pull through. Thus, I wish I sought help more instead of drowning myself with my own doubts and insecurities.

As for the future plans about my project, I have none and I will leave it as it is as of now as I am very proud of its current status. However, when I have improved my coding skills, I do wish to come back and do a comparison of 2023 and maybe 2025/2026's spotify data(?).