Motivation:

My motivation arose from deep interest in understanding my own music preferences and how it changes over time. By analyzing my own Spotify data, I expected to gain insights into my musical taste, my favorite tracks, and how it varies over different seasons. It is also a plus that while I am exploring my own data, I get to learn data science techniques in a practical context.

Data Source:

The data I have used during this project was directly sourced by Spotify. This dataset includes detailed .json files covering a year of my Spotify usage from December 2022 to December 2023. Data encompasses information such as track names, artist names, minutes played, and the end time of each listening session. Additionally, I used CSV files from specific playlists to further enrich the dataset. I have sorted out genres of specified playlists from CSV files for further investigation.

Data Analysis:

Data analysis consist of 5 parts:

Data Preparation:

- First, I have imported necessary libraries. Throughout project, I have utilized python libraries such as pandas for data manipulation, json for parsing JSON files, numpy for numerical computations, and matplotlib and seaborn for visualization.
- Imported my sources, .json and CSV files, into pandas dataframes for easier handling.
- Handled missing values.
- Standardized data formats
- Filtered irrelevant information to focus on key metrices like track names, artist names, and listening durations.

EDA:

- Basic statistics such as mean, median, mode are computed.
- Time of the day and frequency of listening are examined for correlations.
- Listening durations, frequencies of tracks and artists are examined for distribution analysis.

Seasonal Analysis:

- Data is divided into different seasons to spot the seasonal differences in music preferences.
- Analyzed the genres of music preferred in different seasons.
- Most frequently played tracks and preferred artists are identified.

Time Analysis:

- Analyzed total listening time per month to observe any monthly trends or anomalies.
- Peak hours of listening are identified.

Visualization:

- Bar charts, pie charts, line charts for visual representation of the findings

Findings:

- Seasonal variations in music preferences are noted. I preferred calm and slow music in winter, more energetic and house music in spring, more rhythmic and latin music in summer and jazz and blues in fall.
- Discovered that total listening time peaked in November. There is no patterns spotted other than dramatic increase in total listening time in the autumn months.
- Observed diverse range of genres and soundtracks are my consistent preference.
- Found that I listened to music more in evenings. This might suggest that I have a routine at the end of the day.
- Weekends showed a more distributed pattern of listening throughout the day, with a notable increase in morning sessions compared to weekdays.
- Analyzed specific playlists such as turkce xxl, #, **** and S. Apparently, playlist S has much more tracks than all other playlists combined.

Limitations and Future Work:

This project, although it is thorough, faced limitations such as lack of hyperparameter tuning, and re-training the model. The predictive accuracy could potentially be enhanced by experimenting with various hyperparameters and conducting a re-training process with a more diverse dataset. Also, the analysis do not taking account of changing music preferences over time. Future work could include a study to track changes in listening habits. The application of more sophisticated machine learning models like deep learning algorithms could be beneficial in the further investigations.