Data Science Term Project Report

Title: Analyzing Spotify Listening Behavior during Exam and Non-Exam Periods

Introduction:

The objective of this data science project was to investigate whether there are discernible patterns in the songs I listen to during exam periods compared to non-exam periods. Spotify streaming history data spanning multiple years was utilized for this analysis.

Data Collection and Preprocessing:

1. **Data Sources:**
   - Spotify streaming history data was collected for the years 2021-2023.
   - The data included information such as timestamp, track details, and other relevant metadata.

2. **Data Cleaning and Feature Selection:**
   - Columns of interest, including timestamp, track name, artist, and Spotify track URI, were selected.
   - Timestamps were converted to datetime objects for temporal analysis.
   - Monthly listening trends were explored to identify patterns.

Hypothesis:

The hypothesis driving this project was that my music preferences vary between exam and non-exam periods. Specifically, I expected to observe changes in the genres, energy levels, and specific songs listened to during these distinct time frames.

Exploratory Data Analysis:

1. **Monthly Listening Trends:**
   - Monthly listening time was visualized to identify trends over time.

- Examining listening behavior during different periods revealed potential patterns.

2. **Daily Average Listening Time:**

- Daily average listening times were calculated and compared between exam and non-exam periods.

- Visualization of daily listening time variations provided insights into my music consumption habits.

Playlist Creation and Genre Analysis:

1. **Playlist Creation:**

- Playlists were created based on songs listened to during exam and non-exam periods.

- The Spotipy API was utilized for playlist creation and management.

2. **Genre Distribution:**

- Genres of songs in a selected playlist were analyzed to understand the diversity of music genres during different periods.

Song Comparison and Jaccard Similarity:

1. **Common Songs:**

- A comparison of songs during exam and non-exam periods identified common songs.

- Jaccard similarity was calculated to quantify the degree of overlap in song choices.

2. **Venn Diagram:**

- A Venn diagram visually represented the overlap of songs between exam and non-exam periods.

Linear Regression for Future Prediction:

1. **Linear Regression Model:**

- A linear regression model was trained to predict listening behavior during future exam periods.

- The model's performance was evaluated using mean squared error.

2. **Actual vs. Predicted Values:**

- The model's predictions were compared with actual listening times, providing insights into potential trends.

Conclusion:

The analysis provided valuable insights into my music consumption patterns during exam and non-exam periods. The results suggest that there are discernible differences in the songs I choose to listen to during these distinct phases. The hypothesis was partially supported by the observed variations in genres, energy levels, and specific songs.

Recommendations and Future Work:

1. **Refinement of Hypothesis:**

- Further exploration and refinement of the hypothesis could involve considering additional factors such as mood, external events, or specific exam subjects.

2. **Feature Engineering:**

- Exploration of additional features, such as song sentiment analysis or lyrics, could enhance the predictive power of the model.

3. **User-Specific Recommendations:**

- Customized recommendations for music during exam periods based on the observed patterns could be implemented.

Acknowledgments:

I would like to express my gratitude to the Spotify API for providing access to streaming history data, and to the Spotipy library for simplifying interactions with the API.

References:

1. Spotify for Developers. (https://developer.spotify.com/)

2. Spotipy Documentation. (https://spotipy.readthedocs.io/)