

Spotify Streaming History Analysis

Dataset

For this project, I have used the spotify streaming history of my account. On top of the data provided by spotify, I have downloaded audio features for all songs and lyrics for all the songs which had lyrics available on the Genius API.

The initial exploration of the dataset involved the process of validating the dataset to understand whether or not I have downloaded the right data for lyrics and audio features. I tried to remove all the rows or columns which did not add value to the data analysis I wanted to perform.

I started with 17627 rows, this included songs and podcasts. Upon removing podcasts from this list of songs, I was left with 9610 unique songs.

After the initial exploration, the findings are shown in the explanatory_analysis jupyter notebook. This has the results of the exploration phase, all the valid visualisations and findings.

Findings

The average tempo of all the songs streamed on spotify, over the course of the last 365 days is moderate - 120bpm. 120 bpm translates to around 2 beats per second, which is similar to the heartbeat or the average walking stride. The average valence, energy and danceability of the songs is also moderate.

It was also observed that Thursdays and Tuesdays had the least number of streams and Wednesdays had the highest streams. And, the songs were played during the hours of 12 - 18.

Most of the songs were streamed during the winter and summer seasons and the overall theme of all the songs is Love.

Resources

1. developer.spotify.com
2. genius.com/developers
3. stackoverflow.com
4. [kaggle.com](https://www.kaggle.com)
5. Pandas, Seaborn, Matplotlib documentation.