Capstone 1 Proposal – Top 200 Spotify Streaming Data

Overview

Since January 1st, 2017, Spotify has published a daily csv file that details the top 200 most popular songs on Spotify by number of streams. The top 200 csv data includes rank (out of 200), track name, artist and number of streams.

Analyzing the types of songs and artists that tend to be the most popular could provide some insights into music trends among Spotify users. Furthermore, a predictive model could shed some insight on what types of songs and artists will be the most popular in the coming months/years.

Problem Statements

- What types of songs or artists have the highest streaming numbers?
- Do certain song qualities impact the number of streams a song will receive?
- Is there a manner in which the prediction of a songs popularity can be determined by the features of a song?
- Do the top 200 streaming numbers correlate with Spotify stock price?

Data

Spotify Top 200 streaming figures – https://spotifycharts.com/regional
- CSV web scrape using Python requests module

Spotify API – https://developer.spotify.com/documentation/web-api/
- Extract song features for songs in top 200 chart

musiXmatch API – https://developer.musixmatch.com/
- Song meta data (genre, lyrics, MusixMatch AI lyric text analysis)

Quandl API – https://docs.quandl.com/docs/in-depth-usage
- Spotify stock price data

Methodology

1. Extract data by web scraping and referencing public APIs
2. Perform basic analysis (Most popular artist/song/genre etc.)
3. Perform corollary analysis (Correlation between stream numbers and data features)
4. Generate predictive model
5. Test accuracy of model