Blake Zurman
Scripting for Data Analysis Project
12/12/2024

Song Sentiment Analysis and Musical Elements

Describe the Data:
The data that was used for this project are of CSV format and JSON from an API.

- 3000_spotify_songs.csv (Retrieved from Kaggle)
- spotify_data_2017_2021.csv (Retrieved from Kaggle)
- JSON Lyric data (Actively pulls from the Genius API)

For preprocessing, the CSV data was clean. The only work done to the CSV data was taking the average rank by week to determine the songs popularity. This allowed me to rank an artist's songs in the DF based on the amount of weekly hits they got.

For preprocessing the API data, my code searches the Genius API for the song and artist that is inputted. If the song is found, it extracts the details into a dictionary. I then take that dictionary and transform it into a pandas df for further processing.

Method of Analysis:
My idea of this project was to compare the musical elements of popular songs to the lyrical sentiment of the song. Ideally, I wanted to discover any trends or patterns between musical elements, and the message that the lyrics give. So, in summary, the questions that I wanted answered from the project were the following

> Do popular songs have a relationship between sentiment and musicality?
>
> Do the certain artists have a signature combo of musicality and sentiment that work for them?
>
> Does sentiment analysis even work on modern music?

A possible limitation that I predicted was that the sentiment analyzer would get tricked by the lyrics. I was concerned with the nature of how popular songs are written, in that they often contain metaphors, figures of speech, or slang, which confuses the sentiment analyzer. Alas, I pushed forward to find some cool insights.

Description of the Program:

I described earlier how I ranked the songs, and the purpose of this was to limit my research to `Relevant` music. I did this because I figured that the trends in these popular songs may be similar.

The code has a bit a preprocessing in the beginning (as described up top), and has a function block that pulls data from the API on command.

I ranked the popularity of Drake, Coldplay, and Post Malone songs because I like their music. I also chose these artists to test the robustness of sentiment analysis using Vader.

> Coldplay songs have sad lyrics and are mellow, so I was expecting a relationship between musical elements and sentiment.
>
> Post Malone songs are upbeat but have lyrics that would most likely trick the sentiment analysis.
>
> Drake songs are upbeat and sometimes aggressive, but the lyrics are filled with slang.

Next, I created functions that ingested a song and artists, pulled the musical data from the CSVs, pulled lyrical data from the API.

Lastly, I did sentiment analysis with Vader and displayed the results (and a word cloud because they're cool).

Results:

These were the results of the song 'Orphans' by Colplay

{'music_elements': {'danceability': 0.503,
'energy': 0.808,
'key': 11.0,
'loudness': -5.051,
'mode': 0.0,
'speechiness': 0.045,
'acousticness': 0.0254,
'instrumentalness': 0.00162,
'liveness': 0.557,
'valence': 0.283,
'tempo': 107.975},
'sentiment_analysis': 'The sentiment of the song is positive with a compound score of 0.98.'}

I was expecting this song to be positive because it is upbeat, high energy, and moderately danceable. The tempo is often a good indicator of the mood of the song. Slow songs are either sappy and about love, or sad. This song is 108 bpm which is a moderately fast song.

These were the results of ‘Yellow’ by Coldplay… a notoriously sad song.

{'music_elements': {'danceability': 0.429,
'energy': 0.661,
'key': 11.0,
'loudness': -7.227,
'mode': 1.0,
'speechiness': 0.0281,
'acousticness': 0.00239,
'instrumentalness': 0.00013,
'liveness': 0.234,
'valence': 0.285,
'tempo': 173.365},
'sentiment_analysis': 'The sentiment of the song is positive with a compound score of 0.97.'}

Don’t let the tempo fool you, 174 / 2 is the real speed so it’s a slow song at 67 bpm.
The lyrics of this song are certainly melancholy, so Vader was tricked. The energy being high here can be attributed to the load orchestral (sad) music.

Here is ‘Circles’ by Post Malone

{'music_elements': {'danceability': 0.704,
'energy': 0.758,
'key': 0.0,
'loudness': -3.537,
'mode': 1.0,
'speechiness': 0.0402,
'acousticness': 0.233,
'instrumentalness': 0.00144,
'liveness': 0.0924,
'valence': 0.534,
'tempo': 120.003},
'sentiment_analysis': 'The sentiment of the song is positive with a compound score of 0.96.'}

This song talks about how being in love feels like running in circles and doesn’t necessarily have a negative or positive tone. However, the danceability and tempo resemble that of House Music, so positivity could be the interpreted mood of the song. Vader says it’s positive, and I’m not going to argue that.

God’s Plan Drake

{'music_elements': {'danceability': 0.905,
'energy': 0.617,
'key': 2.0,
'loudness': -8.039,
'mode': 1.0,
'speechiness': 0.0596,
'acousticness': 0.00287,
'instrumentalness': 0.00044,
'liveness': 0.0484,
'valence': 0.572,
'tempo': 134.972},
'sentiment_analysis': 'The sentiment of the song is positive with a compound score of 0.99.'}

No surprises here, the happiest and most danceable song of the lot – nice job vader!

In conclusion, this project was a fun exercise of combining multiple sources of data, using different formats of storage in python, and gaining some insights through sentiment analysis, some ranking calculations, and word clouds.