

Proposal 9 -- Analyzing Trends of Music Lyrics and Features Throughout the Decades

Problem:

Analyze the most popular songs of the last decades and understand how the songs change throughout the years in terms of genres, features, generations, lyrics, etc.

Dataset:

Python libraries/ APIs: LyricsGenius, PyLyrics Module, billboard.py, Spotify Web API

- 1) LyricsGenius is a library that wraps the genius.com API. This library provides the lyrics, title, artist, release date, etc. PyLyrics can be used as an alternative as well.
- 2) Billboard.py provides different charts of hot songs and their artists for a specific timeframe. The most popular songs of the last decades can be found by using this library.
- 3) Spotify Web API can be used for analyzing tracks (acoustics, liveness, danceability, tempo, etc) and for analyzing artists (genres, popularity). The Python library, Spotipy can be used for this cause too.

Links to the data sources:

1-a. LyricsGenius: <https://github.com/johnwmillr/LyricsGenius>

1-b. PyLyrics <https://pypi.org/project/PyLyrics/>

2-a. Billboard.py <https://github.com/guoguo12/billboard-charts>

3-a. Spotify Web Api <https://developer.spotify.com/documentation/web-api/>

3-b. Spotipy <https://spotipy.readthedocs.io/en/2.7.1/>

Proposed Solution and Real-world Application:

The main goal of this project is to understand and analyze how the most popular songs change throughout the years. Starting from the 1960s, we will focus on each year's the most popular songs and try to answer questions such as:

- How do the most frequent words in the songs change with respect to genres, releasing years, etc.?
- How do popular genres change throughout the years?
- For specific artists, how do their styles change throughout the years?
- Can we explain these changes by the sociopolitical status of the related era? How are these changes related to real life?

The real-world application of this solution is the insights of our analysis, which can help us understand the music trends and the changes in genres better. Also, it can provide some insights for other people who are in need to do feature extraction, model training, and recommendation systems in this field, etc.

In conclusion, this project looks for relations between genres, intertextual connections, acoustic phenomena (track features), and the time.

Project steps:

Step	Estimated completion time	Person(s) in charge
1. Data extraction	One week	Yuepeng, Ismail
2. Data cleaning	One week	Yucheng, Ismail
3. Data Processing, Transformation, and Analysis	One week	Yucheng, Yuepeng
4. Data visualization	One week	Siyi Wang