FINAL PROJECT - BILLBOARD HOT 100 ANALYSIS REPORT

Anna Ayres
DS 4023
March 8, 2024

Introduction

Popular music serves as a medium of entertainment where societal and cultural shifts throughout time are reflected. This data analysis project utilized the lyrical content of songs featured on the United States Billboard year-end "Hot 100" charts from the years 1959 to 2023. Trends of the lyrical content of the top 100 songs by year were analyzed by overall sentiment, usage of commonly offensive language, references to location, and artist information. The sentiment analysis of the lyrics shows the trend of mood in the lyrical contents throughout time. Commonly offensive words found highlights the use of explicit content in popular lyrics. Location references in lyrics were analyzed to view the locations mentioned in popular songs.

Review of Literature

The analysis of popular music lyrics to view cultural and societal trends must take into account both the popularity of individual songs that resonate amongst the population as well as shift of governmental law/corporate control over musical entertainment throughout the years of 1959 to 2023.

For the majority of the 20th century, most people's first interactions with popular music came primarily from radio broadcasts, physical copies, and live performances. Radio broadcasts served as the main medium for both music discovery and dissemination. Since the 1920s, the broadcasted content has been regulated by the government. Radio stations operating in the United States are required to adhere to Federal Communications Commission (FCC) rules that regulate content broadcasted publicly. According to the FCC, restrictions in broadcasts include topics including but not limited to indecency, obscenity, sponsorship identification, conduct of on-air contests, and hoaxes.[1]

As this dataset starts in the year 1959, it is important to acknowledge the increase in FCC regulations and restrictions that occurred in the 1960s alongside the surge in the popularity of the rock music genre. When the Rock era occurred in the early 1960s, there was a surge in FCC bans on songs mentioning "...Anti-war sentiment, class-consciousness, and racism."[2] Government bans continued with measures such as the 1990s bills introduced in 21 states that restrict the sale of records containing lyrics that are " violent, sexually explicit or 'perverse.'"[3] With the introduction of widespread internet access, the ability to download and later stream music became increasingly popular. As the songs individuals accessed on the internet were not broadcasted, they were not subject to FCC rules.

A research study conducted by the International Federation of the Phonographic Industry (IFPI) showed that in 2023, 73% of people listened to streamed music. The 2023 music engagement is broken down into 32% through audio streaming services (e.g. Spotify, Apple Music), and 31% video streaming services (e.g. YouTube, TikTok). Only 17% was through radio broadcasts and 9% through individual purchased music including physical and digital tracks.[4] Although music listened through streaming or purchased copies are not regulated by the FCC, it has not stopped censorship by corporate entities. Corporate self-censorship is done by record industry organizations for topics in response to topics such as "legal and political context and moral climate". In 1990, 85% of US record labels who are a part of the Recording Industry Association of America (RIAA) adopted the use of parental advisory labels (PAL) on records deemed to have offensive lyrics referring to topics that are sexual, overly violent, offensive, or promote drug use.[5]

Methods

The dataset used was compiled for the project using the Billboard Year-End Hot 100 singles from the years 1959 to 2023.1 The rank, title, and artist name data was compiled from Billboard Charts using BeautifulSoup and requests import tools on python.[6] The corresponding song lyrics were compiled into the dataset with the use of Genius API and the python client LyricsGenius.[7, 8] The full dataset was then compiled into a csv file named *hot100_song_data.csv*. For the analysis of the dataset, the *hot100_song_data.csv* file was read in python as a DataFrame using the Pandas Python library.

The first analysis was conducted on average sentiment within the lyrical content by year using the Natural Language Processing (NLP) tool named TextBlob.[9] The TextBlob Python library includes a sentiment analysis process which was used on the dataset. The second analysis searched through the 'Lyrics' column to match with words found in the *RadioWords.txt* file which is an compiled list of words that are commonly considered as offensive in content.[10] The figure shows the trend of offensive lyrics by year. The third analysis searched the 'Lyrics' column in the hot 100 data to match with words relating to cities in the US found in the *uscities.csv* file compiled using data from the top populated cities. A heatmap was generated of states where cities were mentioned by count of songs using the Folium *choropleth* tool and a GeoJSON file containing state boundaries information.[11] The fourth analysis compiled the top-charting artists from the count of their entries throughout the charts.

Analysis

The first analysis visualization was created with Seaborn and MatPlot library imports. 10 The Seaborn *regplot* tool was used to show the scatter plot along with the regression analysis trendline. The trendline shows a negative correlation between the sentiment of lyrics and the

years from 1959 to 2023. The trendline sloping downward suggests a decreasing average sentiment in the lyrics over the specified time period. It suggests that the overall emotional tone of popular song lyrics has shifted to less positively over the years. This could be due to factors such as societal shifts amongst the population as well as governmental law/corporate control over musical entertainment.

The second analysis visualization was created with the MatPlot library import. A scatterplot was created to show the total number of “offensive” lyrics by year. The number of offensive lyrics has increased in recent years, with a large spike happening in the 1990’s. This could be due to cultural shift as well as the increase in streaming services over FCC regulated broadcast radio.

The third analysis visualization was created with the Folium library import’s choropleth tool. The heatmap shows the highest frequency states being California, Texas, Florida, and New York. This could correlate to the fact that these states are also the highest in population of the United States, meaning more artists are more likely to have visited/be from cities in those states.

The fourth analysis visualization shows the top ten artists with bubbles where the size correlates to the amount of songs that have charted on the top 100 charts throughout time.

Conclusion

In conclusion, the analyses and visualizations show insights and trends into the lyrics of popular music from 1959 to 2023. Trends such as a decrease in lyric sentiment, increase in "offensive" lyrics, geographic hotspots, and the chart impact of top artists give an understanding of the United States music industry and society over time.

Bibliography

1. "The FCC and Speech." *Federal Communications Commission*, 31 Aug. 2022, www.fcc.gov/consumers/guides/fcc-and-speech.
2. Martin, Linda, and Kerry Segrave. *Anti-Rock: The Opposition of Rock n'Roll*, Archon Books, 1988, p. 189.
3. "Brief Timeline on Censored Music." *American Civil Liberties Union*, 26 Sept. 2005, www.aclu.org/documents/brief-timeline-censored-music.
4. "IFPI's Global Study Finds We're Listening to More Music in More Ways than Ever." *IFPI*, 11 Dec. 2023, www.ifpi.org/ifpis-global-study-finds-were-listening-to-more-music-in-more-ways-than-ever/.
5. Billboard Charts www.billboard.com/charts
6. Requests: https://pypi.org/project/requests/
7. BeautifulSoup(BS4): https://beautiful-soup-4.readthedocs.io/en/latest/
8. Genius API: docs.genius.com
9. LyricsGenius: https://lyricsgenius.readthedocs.io/en/master/
10. RadioWords.txt: edited data text file from https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words/blob/master/en
11. GeoJson Data: https://eric.clst.org/tech/usgeojson/