Decoding the Harmony

A Comprehensive Deep Dive into the Evolution of Music Lyrics

-for Hackathon: Cultural Evolution of our Songs

CDV0511 Analytics with Python Workshop
Instructors: Professor Sushmita and Professor Ghazal
Group Seven: Adrian An
Northeastern University Seattle Campus
College of Professional Studies, Data Analytics

Table of contents

Decoding the Harmony

- An extensive exploration into the fascinating world of music lyrics.

- Aims to unravel the complexities and understand the evolution of lyrical content in songs over the years.

- Leveraging advanced Natural Language Processing techniques and data analysis tools.

- Investigates multiple dimensions of lyrics:

  Emotional aspects   Genre-specific trends

  Linguistic patterns   Readability

- Seeks to illuminate trends and transformations that have shaped our musical journey.
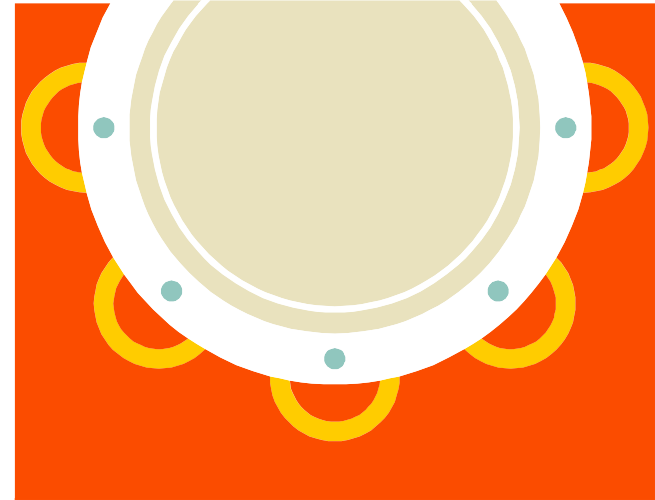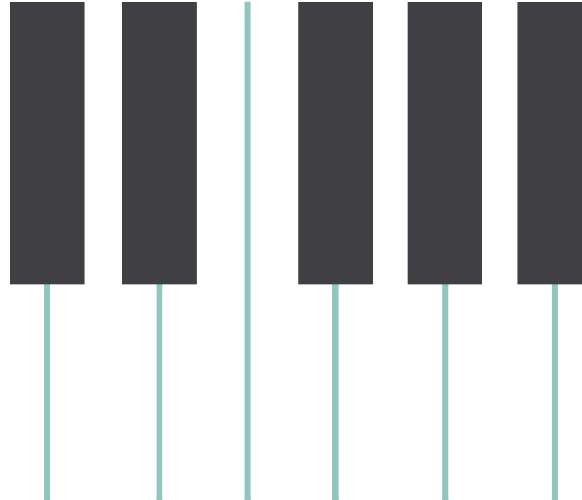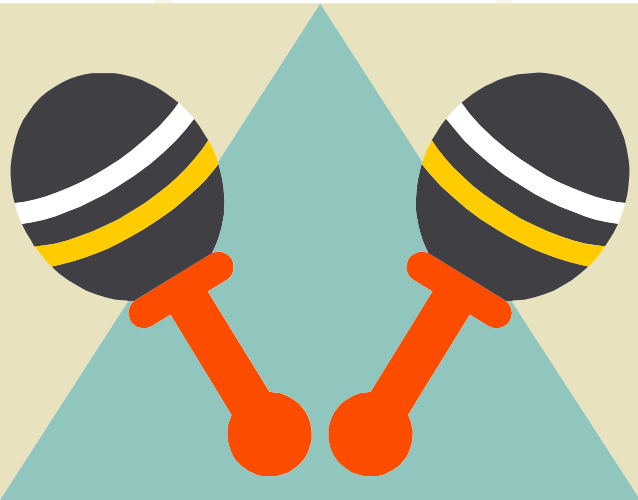
About the Data

The 'dataset_merged.csv' file is an extensive and intricate compilation of song-related data.

It represents the successful merging of three distinct datasets from Kaggle, namely 'artists-data.csv', 'lyrics-data.csv' and 'charts.csv'.

This consolidation has resulted in a comprehensive collection of various song attributes, offering a uniquely holistic perspective on music analysis.

Scrapped Lyrics from 6 Genres

Billboard The Hot 100 Songs

artists

lyrics

charts
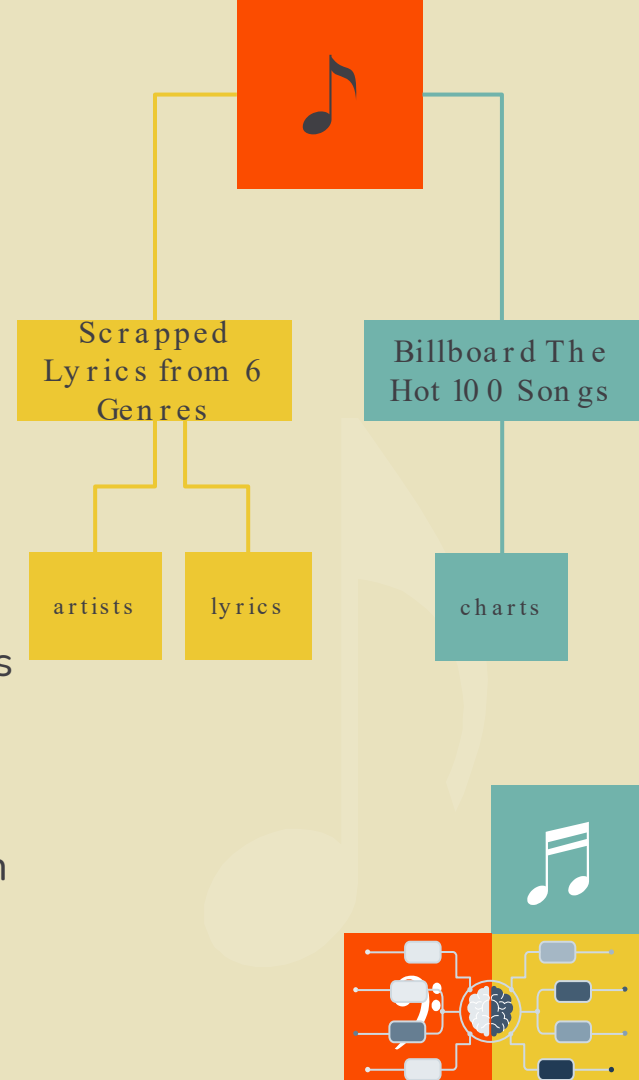
Definition of concepts

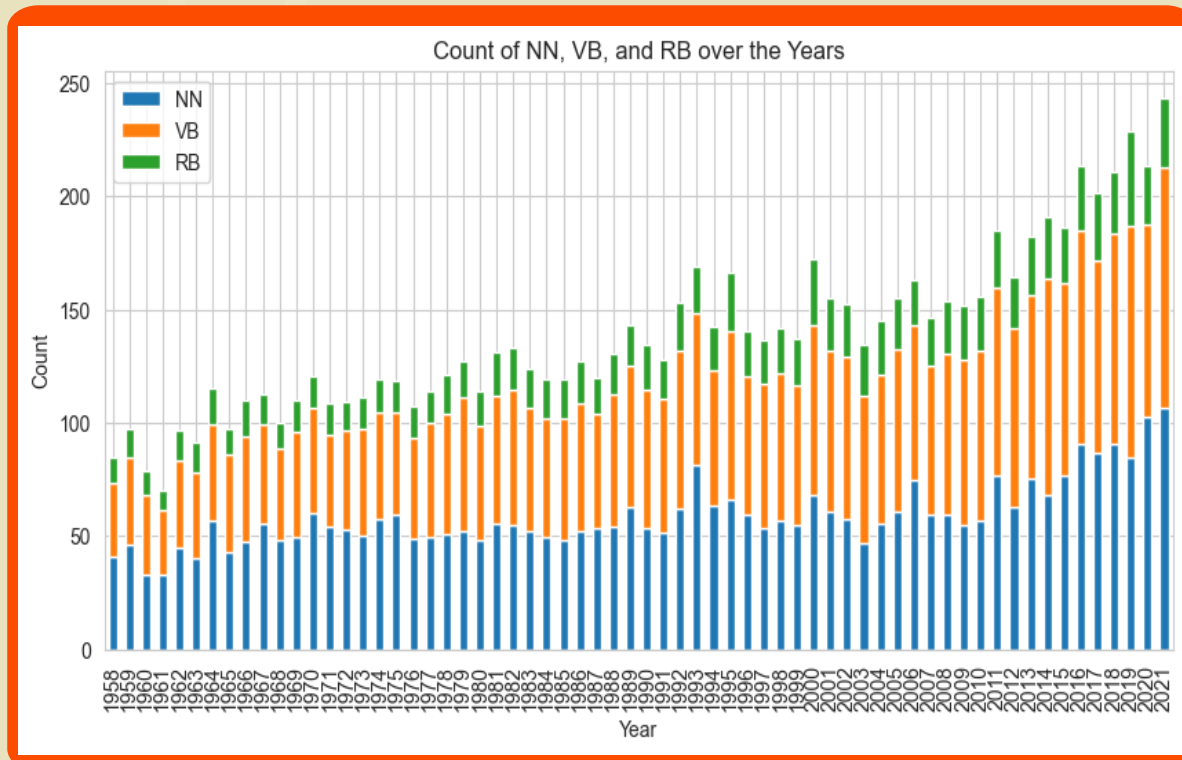| **NN**	**VB**	**RB**
The Count of Nouns	The Count of Verbs	The Count of Adverbs

Data Aggregation and Visualization

Count of Speech –
grouped by year

The count of Nouns (NN), Verbs (VB),
and Adverbs (RB) over the Years.

Regression Analysis

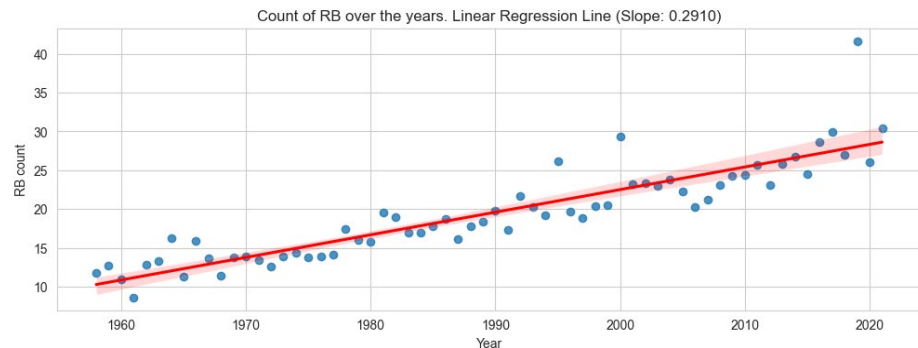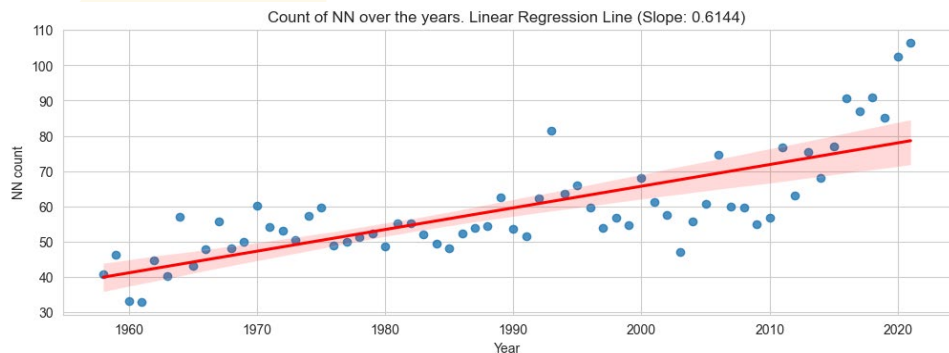Fitting a linear regression model to each part of speech count over the years.

L I N E A R

Count of NN over the years. Linear Regression Line (Slope: 0.6144)

Count of RB over the years. Linear Regression Line (Slope: 0.2910)

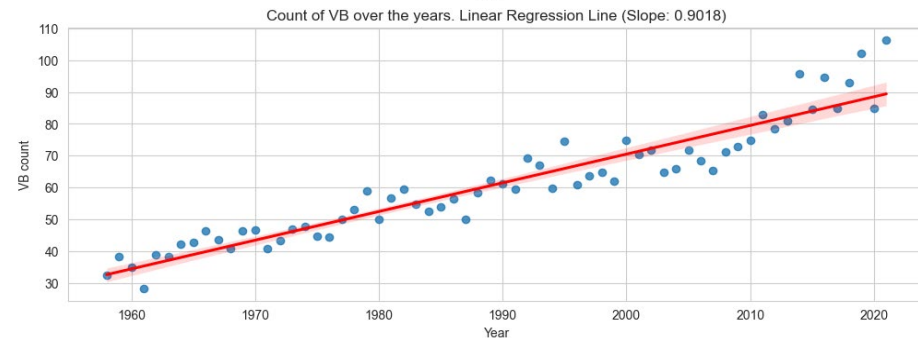Count of VB over the years. Linear Regression Line (Slope: 0.9018)
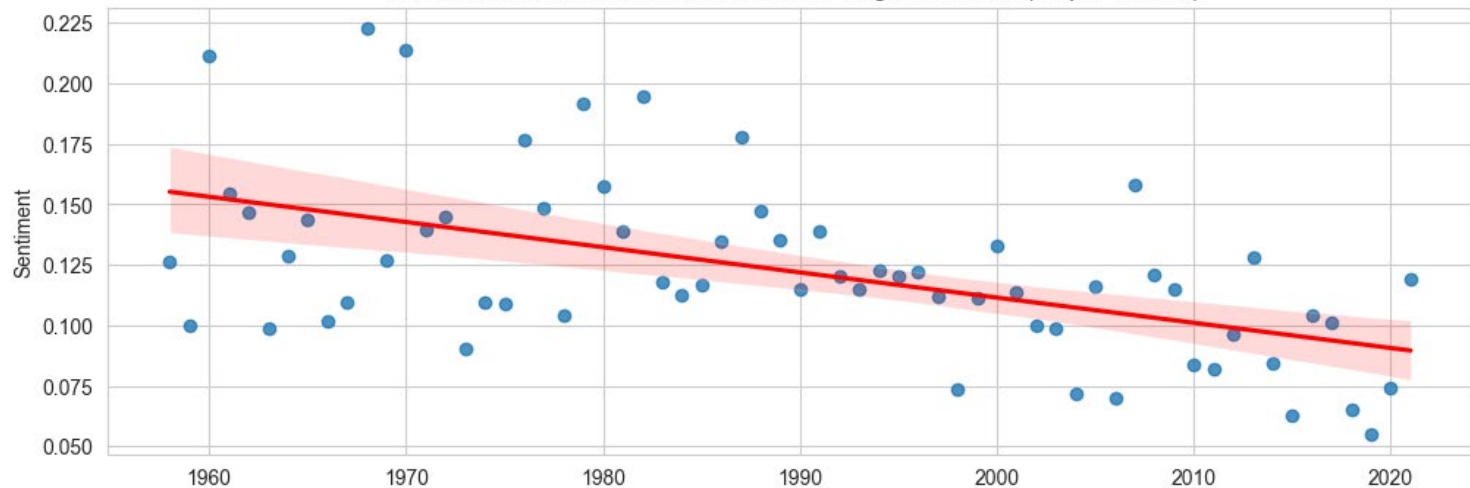
Plots of word count over time, separated by type of speech (NN, VB, RB).

The positive slope of the regression line, suggesting that songs have been using more unique words over time.

TextBlob Sentiment over the Years (Slope: -0.0010)

Emotions
Exploration

By measuring the sentiment of song lyrics and analyze its trend over the years, the slope of the linear regression line can suggest the sentiment of songs is becoming more negative over time.

The Confidence Interval

Expected Slope Mean: The slope of the regression line is -0.0010. This means that, on average, there is a very slight decrease in sentiment per year. Essentially, song sentiment appears to be becoming marginally more negative over time, but the change is minor.

Expected Slope Standard Deviation: The standard deviation of the slope, which measures the variability in the year-to-year changes in sentiment, is 513.6830. This value is substantial, indicating a high degree of variability in the sentiment change from year to year.

95% Confidence Interval: The confidence interval for the slope is (-1026.8387, 1026.8366). This means that we can be 95% confident that the true slope of the population (if we were to have data for all songs, not just our sample) lies between these two values.

```
Expected Slope Mean: -0.0010
Expected Slope Standard Deviation: 513.6830
95% Confidence Interval: (-1026.8387, 1026.8366)
```

Explore the Emotion in the Top 6 Genres

List of Tuples
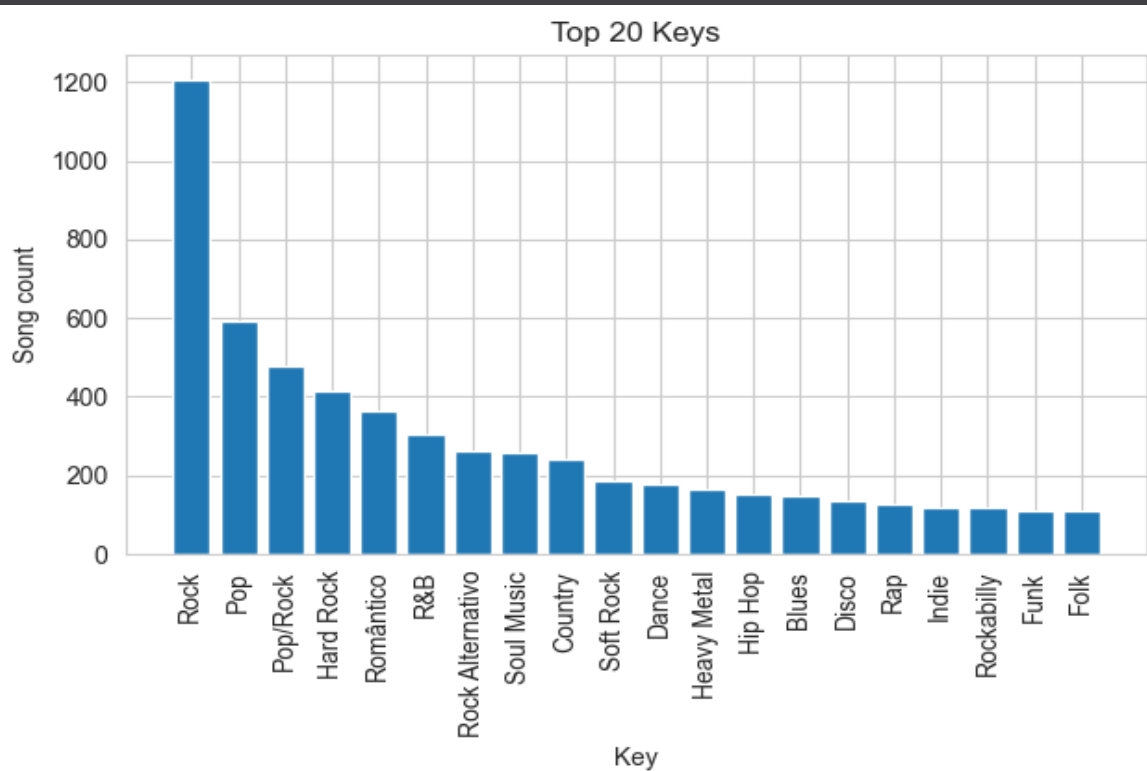
Explore the Emotion in the Top 6 Genres

Rock Alternativo

This genre is represented by 275 songs in the dataset, making it the sixth most common genre in the data.

Romântico

There are 405 songs in this genre, ranking it fifth.

Hard Rock

With 414 songs, this genre is the fourth most common in the dataset.

Pop/Rock

This hybrid genre is represented by 507 songs, placing it third.

Pop

As one of the most popular genres in the dataset, it is represented by 616 songs.

Rock

With 1160 songs, Rock is the most represented genre in the dataset.
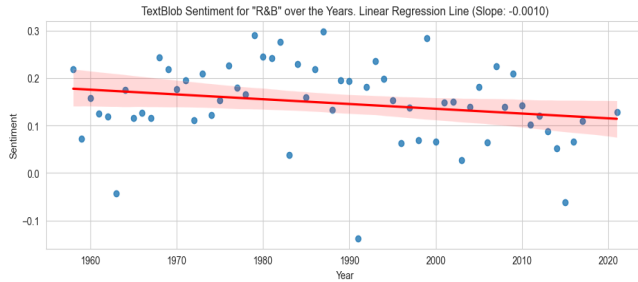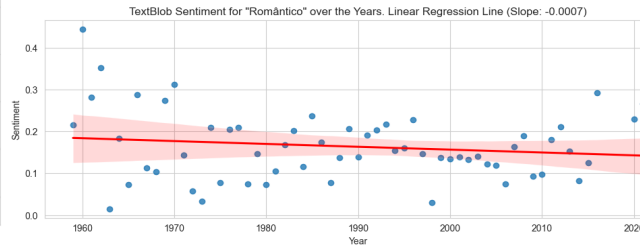
Sentiment Analysis
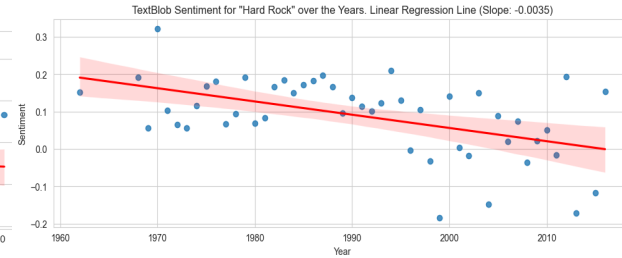
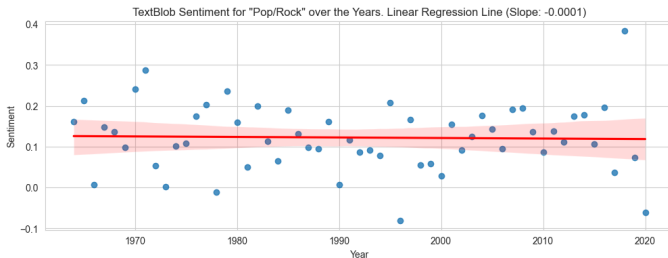- Grouping by year and visualizing the sentiment trend using linear regression
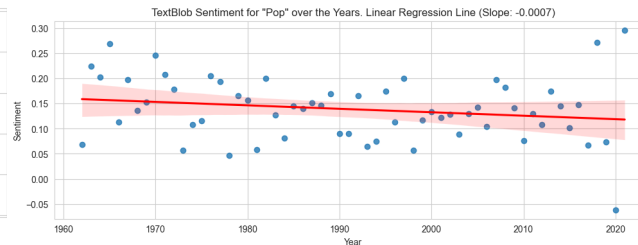

Rock Alternativo

Romântico
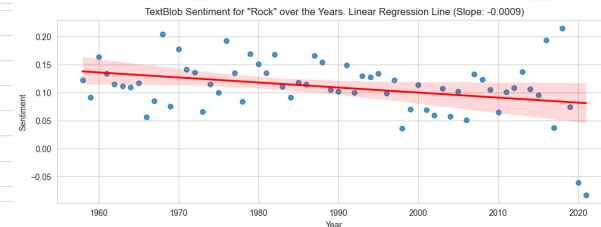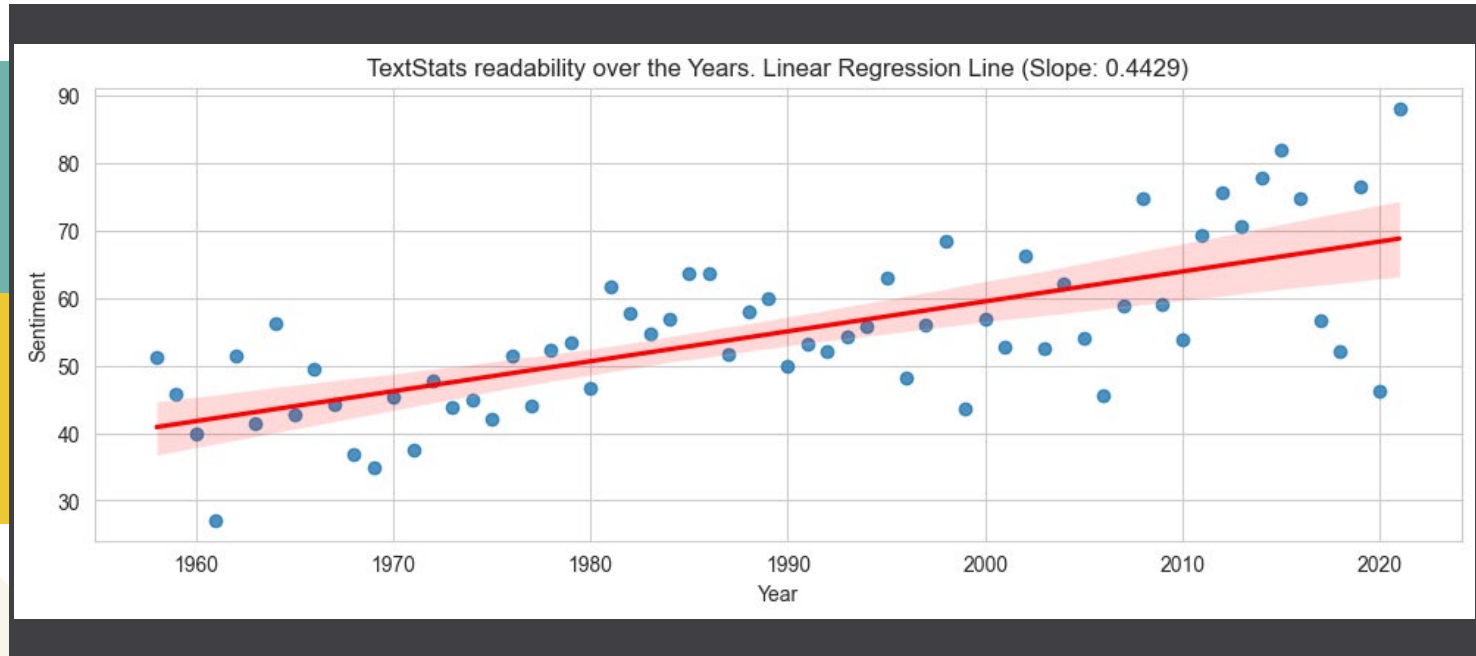
Hard Rock

Pop/Rock

Pop

Rock

Show sentiment analysis plots for the top 6 genres ("R&B", "Romântico", "Hard Rock", "Pop/Rock", "Pop", "Rock").
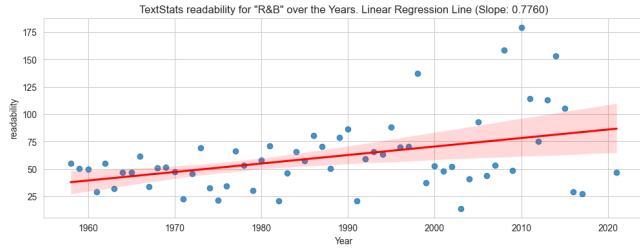
Song Readability

The slope of this line (0.4429) shows that the average readability score increases by about 0.4429 points for each additional year. This means that, on average, songs are becoming more complex (or harder to read) over time.
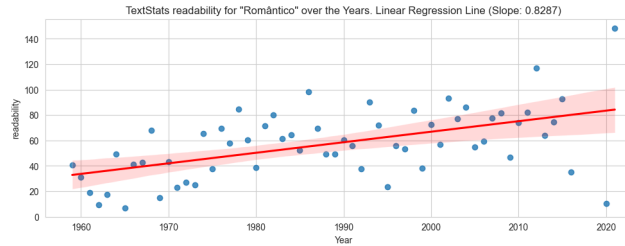
Genre Analysis

- Extracted songs of each genre, calculated the yearly average readability, and fitted linear regression models

Rock Alternativo

TextStats readability for "R&B" over the Years. Linear Regression Line (Slope: 0.7760)

Romântico

TextStats readability for "Romântico" over the Years. Linear Regression Line (Slope: 0.8287)

Hard Rock

TextStats readability for "Hard Rock" over the Years. Linear Regression Line (Slope: 0.8120)

Pop/Rock

TextStats readability for "Pop/Rock" over the Years. Linear Regression Line (Slope: 0.5485)

Pop

TextStats readability for "Pop" over the Years. Linear Regression Line (Slope: 0.4623)

Rock

TextStats readability for "Rock" over the Years. Linear Regression Line (Slope: 0.4852)

Show plots of TextStats readability for each of the top 6 genres.

Summary

Emotion Analysis

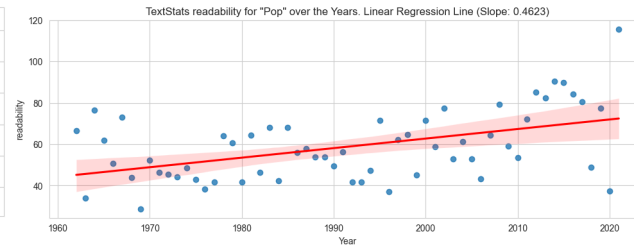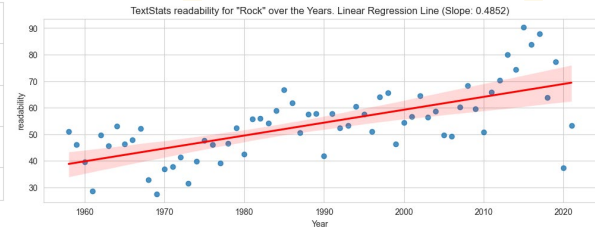A comprehensive sentiment analysis, aided by TextBlob, revealed a slight negative trend in the emotionality of song lyrics over the years. This insight was further substantiated by genre-specific analysis, where 'Hard Rock' emerged as the genre with lyrics becoming more emotionally negative over time.

Part-of-Speech Analysis

Unraveling linguistic patterns in lyrics showcased a consistent presence of nouns, verbs, and adjectives in songs, indicating a rich tapestry of descriptive language and action words in lyrics, underpinning the narrative storytelling aspect of music.

Genre Analysis

The project unveiled 'Rock', 'Pop', and 'Pop/Rock' as the top three genres with the highest number of songs, highlighting the popularity and influence of these genres in the music industry.

Readability Analysis

An interesting trend of decreasing song readability over time was noted, implying that song lyrics have, on average, become slightly less complex. This trend varied across genres, offering a unique view into the lyrical complexity within different music genres.

Conclusion

Sentiment Analysis

A slight decrease in sentiment scores across all genres over the years.

Part-of-Speech Analysis

An increase in the usage of unique words over the years, as suggested by the positive slopes in the regression line for nouns (NN), verbs (VB), and adverbs (RB).

Genre Analysis

'Rock', 'Pop', and 'Pop/Rock' have the most song entries, indicating their dominance and influence in the industry.

Readability Analysis

A decrease in readability scores over time, with a slope of 0.4429. This suggests that lyrics have become less complex and more accessible.
The genre-specific analysis showed varied results. For instance, 'R&B' showed the highest increase in readability over the years, with a slope of 0.7760.

This project offers a unique perspective on the intricate relationship between music and language, capturing the dynamic evolution of lyrics over the years.
By highlighting these patterns and trends, we hope to foster further discussions and research in the fascinating realm of music lyrics.

Decoding the Harmony

Thanks!

Does anyone have any questions?
an.te@northeastern.edu| Northeastern University - Seattle | 225 Terry Ave N