Summary

Introduction

Spotify is one of the largest online platforms for streaming music and podcasts. In this module, we use Spotify's podcast data to construct informative metrics for podcasts like music-specific metrics developed by Spotify.

Podcast Data

1. We use Spotify's API to download podcast data. Our dataset includes a total of 5787 podcast data from seven categories with a time range of 2015-2024, for each category we choose 50 shows.
2. Data cleaning. First, remove entries in the description column containing non-English words, such as emojis. Next, we processed the description column by eliminating punctuation marks, converting all text to lowercase, and removing URLs.
3. We do word counting, and find two metrics.

Informative Metrics

1. Special Metrics
   a. We do word count for every word in the podcast name and use TF to show the results. "episode", "show", "season" are some common words. We can use IDF to show if the word is rare or not. N in the equation is the number of total podcasts, df(w) is the number of podcasts that include that word. When we calculate these TF-IDE scores, the higher score we get, this podcast is rarer.

$$IDF(w) = log(N/(1 + df(w)))$$

   b. We calculate Cosine correlation between that podcast name and other names to show the similarity.
   c. We combine two above to define special scores. When this name is special, the special score is close to 1.
2. Emotion Metrics

We counted the emotional words in the description based on the Bing emotion dictionary. Positive words were assigned a score of 1, and negative words were assigned a score of -1. The scores were then normalized to a range between 0 and 1. Neutral words and NA values were assigned a score of 0.5.

Shiny App

https://cjiang232.shinyapps.io/shinyapp

Limitations

The drawbacks can include insufficient data volume, the ability to retrieve only from the existing dataset, and the fact that the Shiny app cannot interact with the internet.

Contributions

Contributions	Chenyu Jiang	Siyu Wang
Summary	Provide feedback	Responsible summary
Code	Responsible for code	Provide feedback
Shiny App	Responsible for Shiny app	Provide feedback