Maria Scricco
MTH 2300
Midterm Project
March 21, 2024

Introduction

Music Festivals and concert lineups are on the rise once again following the COVID-19 pandemic (Brecht, 2023a). Consequently, festivals like Coachella, Governors Ball, Lollapalooza, Rolling Loud, and dozens of others are experiencing an influx of event-goers and attendees with music-fanatics in their in-person listening concert experiences. Events such as these combine several days of music performances from a variety of artists and genres. According to recent reports, "Revenue in the music events market is projected to reach U.S.$ 30.14 billion in 2023, with a projected market revenue of 36.71 billion dollars by 2027 (Marx, 2023). With such a growing presence, it is vital for music event marketers, organizers, and managers to understand the relationships between music genres in curating effective and appropriate lineups for music festivals.

This project seeks to connect the dots between the 6 most popular music genres and see how they are related to one another in order to assist festival management in blending associated genres with one another to best suit the customer experience for the better.

Methods & Study of Data

About the Dataset

The present dataset was a selection of Spotify data from a TidyTuesday competition, which outlines top songs from Spotify over the past century, alongside a variety of characteristics of the songs, each given their own value (i.e.: each song's tempo, key, danceability, valence, etc). This dataset was downloaded from the original TidyTuesday post linked: here.

Data Cleaning & Management

In terms of the methods of this project, the initial data management steps required a great deal of cleaning and data removal in order to be subsetted for an accurate sized dataset. I completed the main data cleaning steps within an Excel file, utilizing the table formats to search through all of the data points. First, I started off by removing unnecessary data columns from the original Excel dataset of Spotify songs. This required me to remove columns for danceability, liveliness, loudness, track album name, subgenre, speechiness, release date, instrumentalness, and valence. Consequently, the columns of interest that remained were tempo, duration, key, mode, and popularity. Next, I removed any blank or extraneous values (i.e.: values that were nonsense or clearly typos).

Moving on to data management, this original dataset from Spotify was over 25,000 data points, which would be extremely difficult and redundant to utilize for the present project. Consequently, I decided to subset the data by genre, since that was the variable of interest I wanted to see connections across. So, I decided to select the top 10 songs from each genre listed in the dataset by their popularity ranking. This required me to sort the values by popularity (largest to smallest) and then subset 10 songs for each genre listed. The existing genres were: Pop, EDM, Rap, R&B, Latin, and Rock.

Data Analysis & Studying

In terms of data analysis, I decided to utilize a Principal Component Analysis to see which genres were closest to one another, and which genre was the closest to the "Pop" genre. As we commonly know, "Pop" or "Popular" music is characterized by trends within music listening, and investigating which genres are the closest to "Pop" music can help assist concertgoers and event planners in associating genres for the best audience turnout and retention.

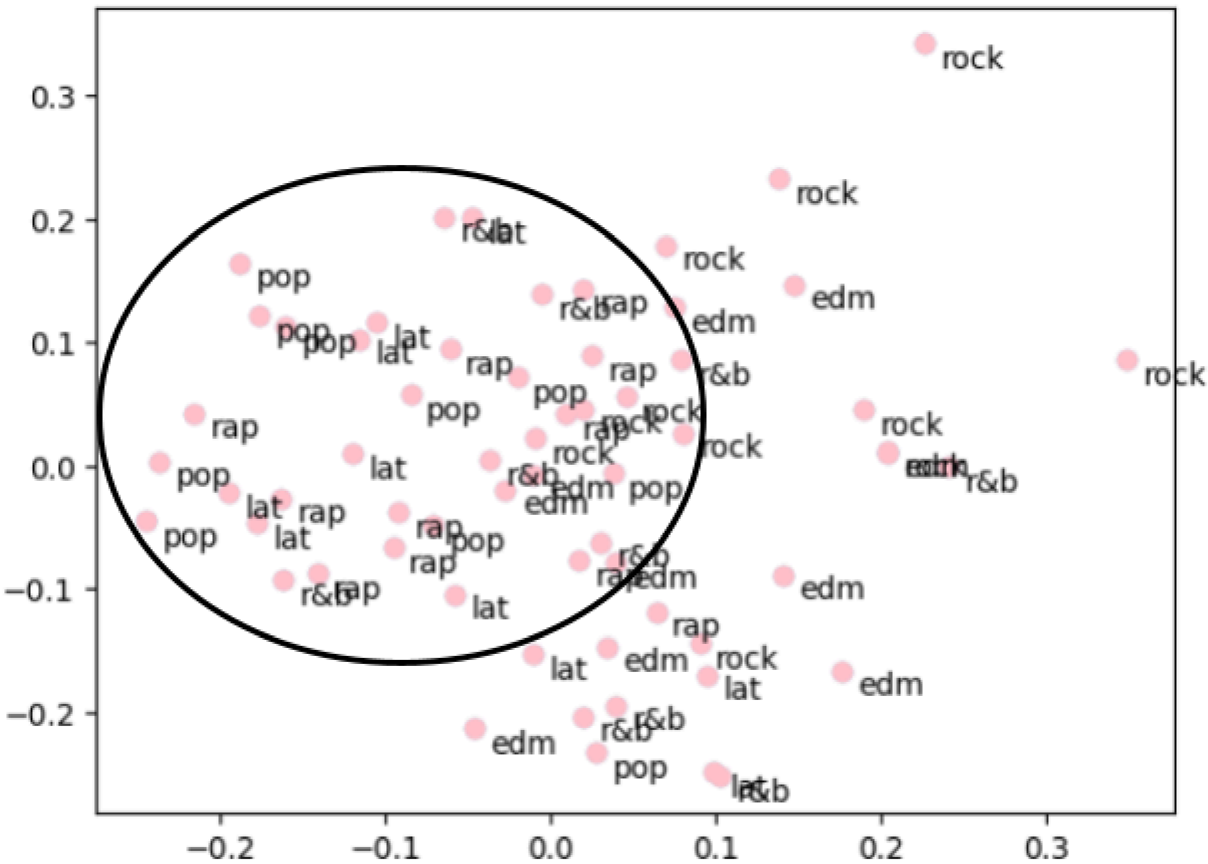
This first Principal Component Analysis showed a scatterplot of all 60 songs (top 10 songs for each genre) plotted, alongside the respective Singular Value Decomposition results. These results can be seen in *Figure 1.*

Moving on, in order to compare genres to "Pop", I decided to take an average of the top 10 songs of each genre within the dataset to better compare. This required the creation of a second dataset, which averaged the standardized popularity, key, mode, tempo, and durations of the songs by each genre. This new dataset was uploaded for a second Principal Component Analysis in the second section of my Python file.

This second Principal Component Analysis showed a scatterplot of the 6 genres in relation to one another, alongside the respective Singular Value Decomposition results. These results can be viewed in *Figure 2.* On top of this, although the scatterplot showed a visual representation of the closeness between genres, a for loop was utilized to calculate the actual numeric values associated with each genre in order to assist in comprehending which was closest to Pop.

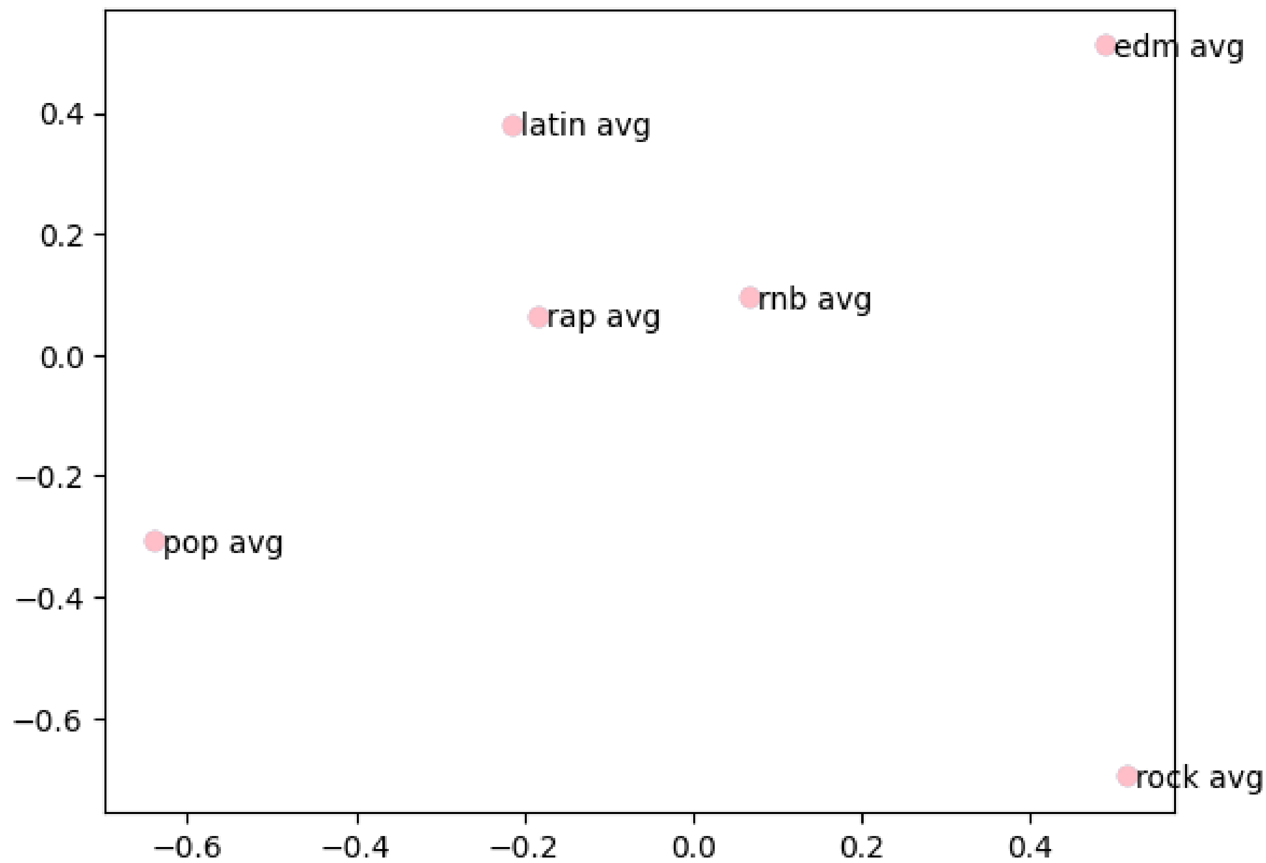
Results / Graphs

Figure 1. Songs by Genre PCA (Top 10)



Caption. Scatterplot showing groupings of related top 10 songs by genre from the Spotify dataset. As we can see, there is a large group of pop songs and rap songs in a similar area, denoted with a circle.

Figure 2. Closest Genre PCA (Average)



Caption. Scatterplot indicating the grouping/relatedness of song genres (by averages). Visually, it seems that latin or rap music seem to lie closest to the Pop genre.

Finally, the following results were incorporated to calculate the distances to indicate which genre actually was the closest.

```
(POP)  0 0.0
(RNB)  1 0.8146075347327164
(ROCK) 2 1.219639004630895
(EDM)  3 1.3948972855968993
(LATI) 4 0.8092239905478674
(RAP)  5 0.5890311307157639
```

Conclusions & Future Studies

Conclusions

This information can be used to fuel festival lineups and comprehend the manner in which performance lineups should be curated. From what we learned, it seems that the Rap genre is in closest relation to the Pop genre, indicating that this genre should likely be sought after when curating pop music lineups for large music events, festivals, or concerts.

Future Ideas

In terms of future ideas for studying, it may be of interest to investigate different aspects of subgenres that were originally available in this Spotify dataset and even further see how to group genres or how related the subgenres are to one another. In addition, future research could investigate the relationships between fans' listening habits and specific related artists to one another, or specific songs, and cluster/group based on that.

References

Brecht, Robert M. “Selecting Music Genres for Live Events.” *TSE Entertainment*, TSEE, 11 Apr. 2023, tseentertainment.com/selecting-music-genres-for-live-events/.

Brecht, Robert M. “Music Festival Economics.” *TSE Entertainment*, TSEE, 6 Sept. 2023, tseentertainment.com/music-festival-economics/#:~:text=Festivals%20have%20become%20another%20venue,economy%20is%20seriously%20big%20business.

Marx, Madison. “Genre Trends at America’s Biggest Music Festivals.” *Vivid Seats*, Vivid Seats, 19 May 2023, www.vividseats.com/blog/music-festival-genre-trends.

“Spotify Songs.” *GitHub*, GitHub , 21 Jan. 2020, github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-01-21/readme.md.