Determining What Elements of a Song Make it Popular Based On Trend Studies and Clustering Methods

Logan Wrinkle and Mahim Mathur

Introduction

- Music is an almost universal passion which has major differences across cultures and time periods. Our goal is to try and explore some of that history using data.
- Our dataset contains 2000 of the most popular songs on Spotify over 20 years (100 for each year), which tracks not only genre but elements like energy, danceability, upbeatness and acousticness for each song.

Motivation

Our group wants to explore two questions
- What trends can we spot over the 20 years that the dataset covers?
- What elements, or combination of elements, is associated with popular songs?

Using Python libraries and data mining tools, we can visualize and analyze Spotify data to understand what listeners haved like the most this century.

Methodology

- First we worked to understand the dataset, answering questions such as:
  - How many songs are there per year?
  - What track elements are there?
- Once we understood the dataset more, we visualized the data with Pandas and Matplotlib libraries
  - Average popularity over time
  - Number of songs in each genre per year
  - % contribution of each genre per year
- Then, we analyzed which track elements work best together using a heatmap of correlation coefficients
- We applied K-means clustering with Sklearn to try identifying interesting relationships between elements and popularity
  - Used the elbow-method to find the optimal number of clusters (4)
  - Tried different methods of visualizing the clusters

Results

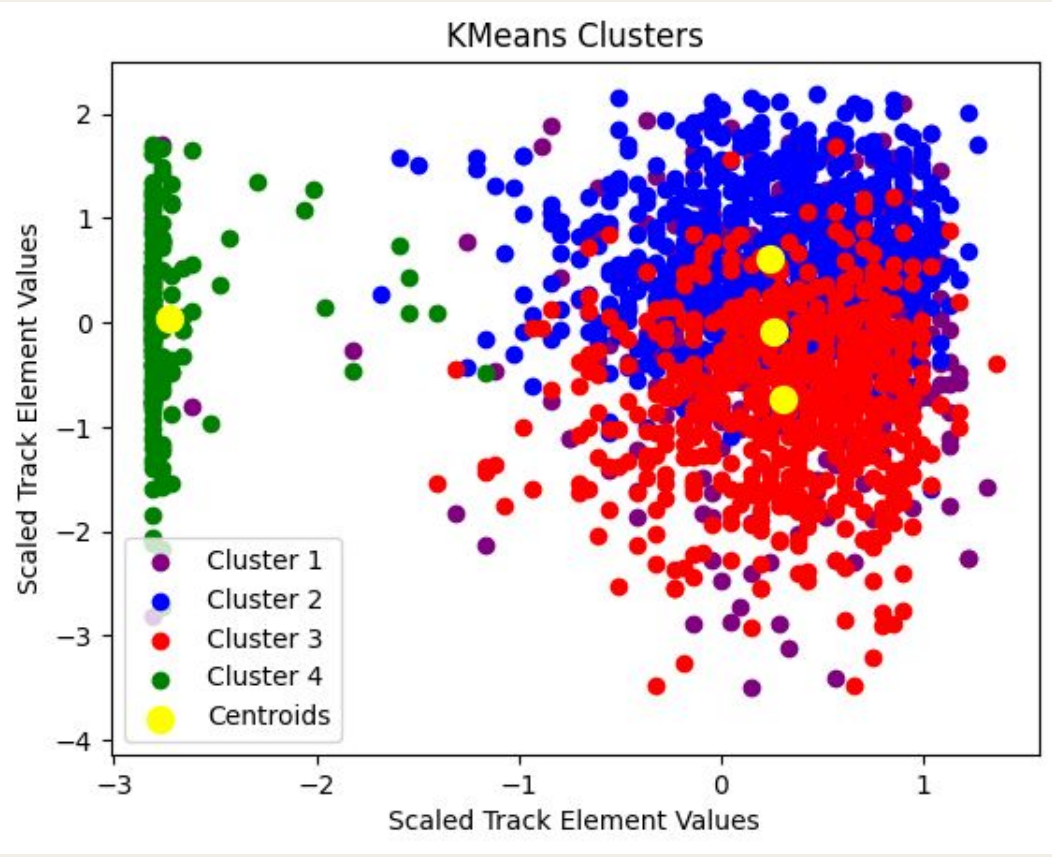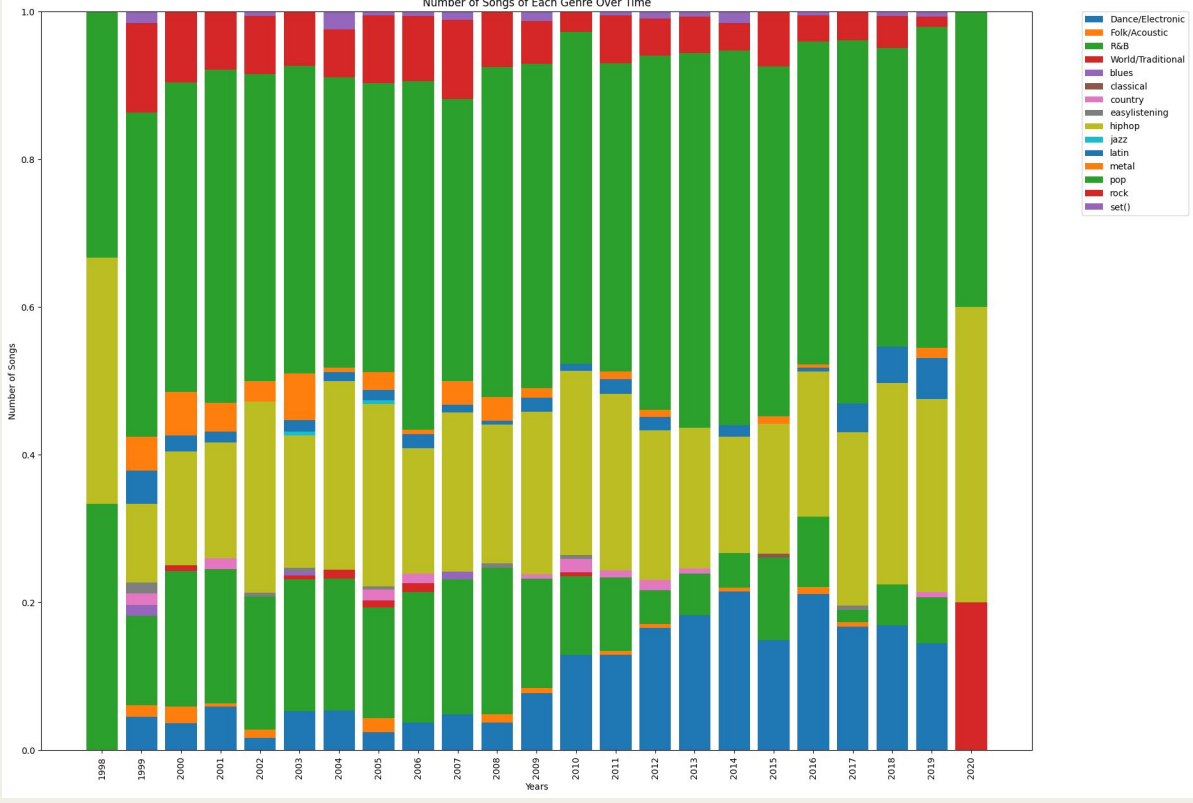The most popular songs this century came from 2018.
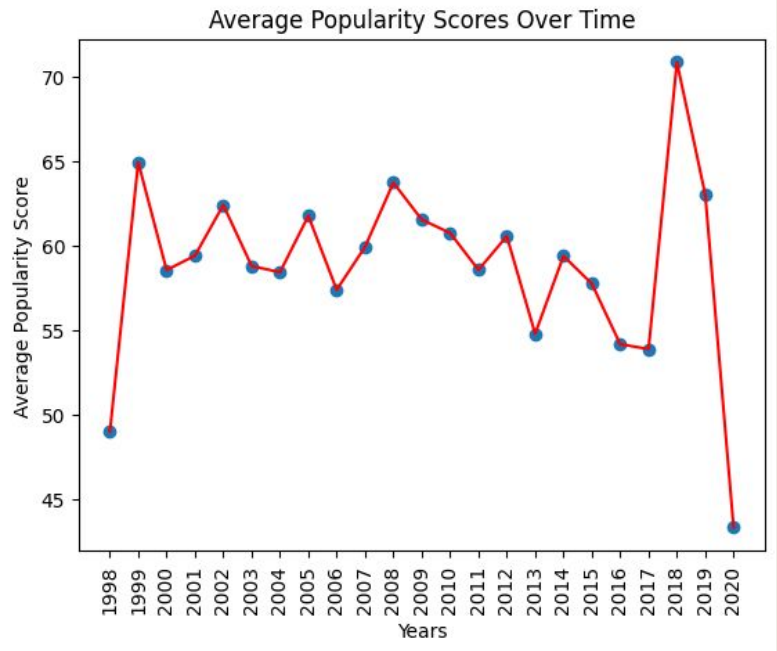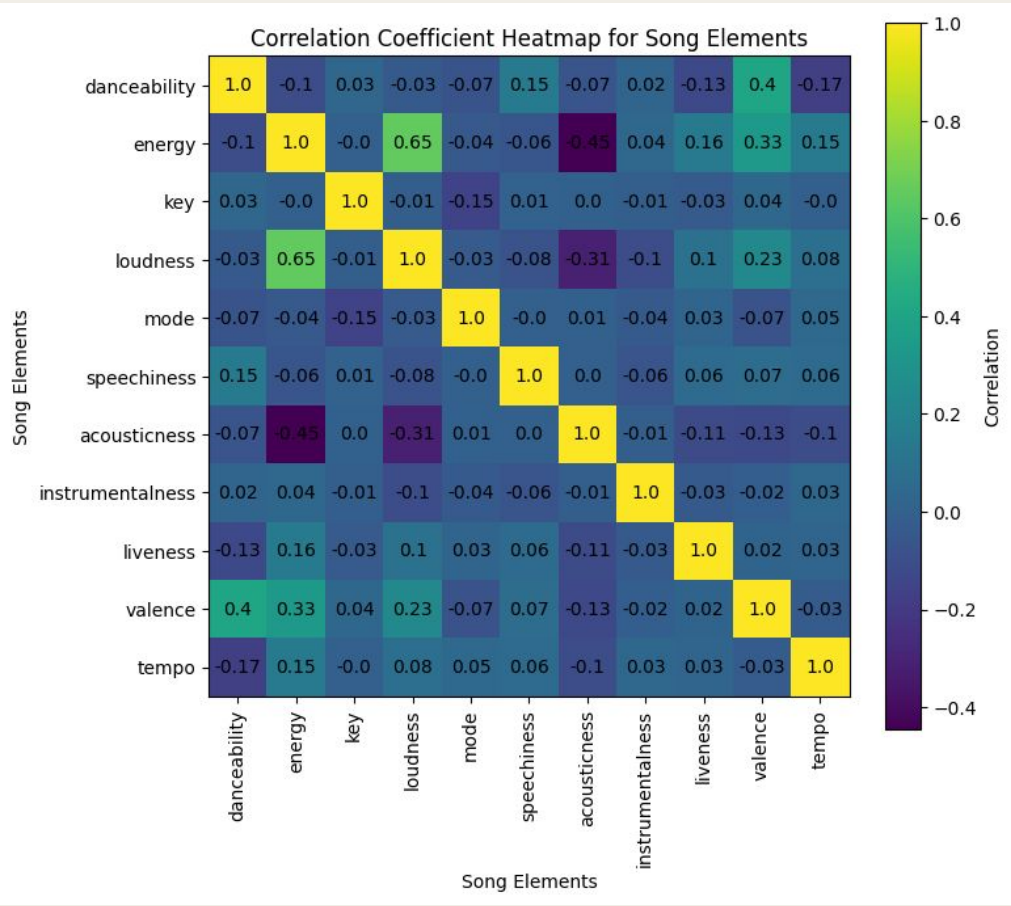
Loudness and Energy are track elements that work together best.
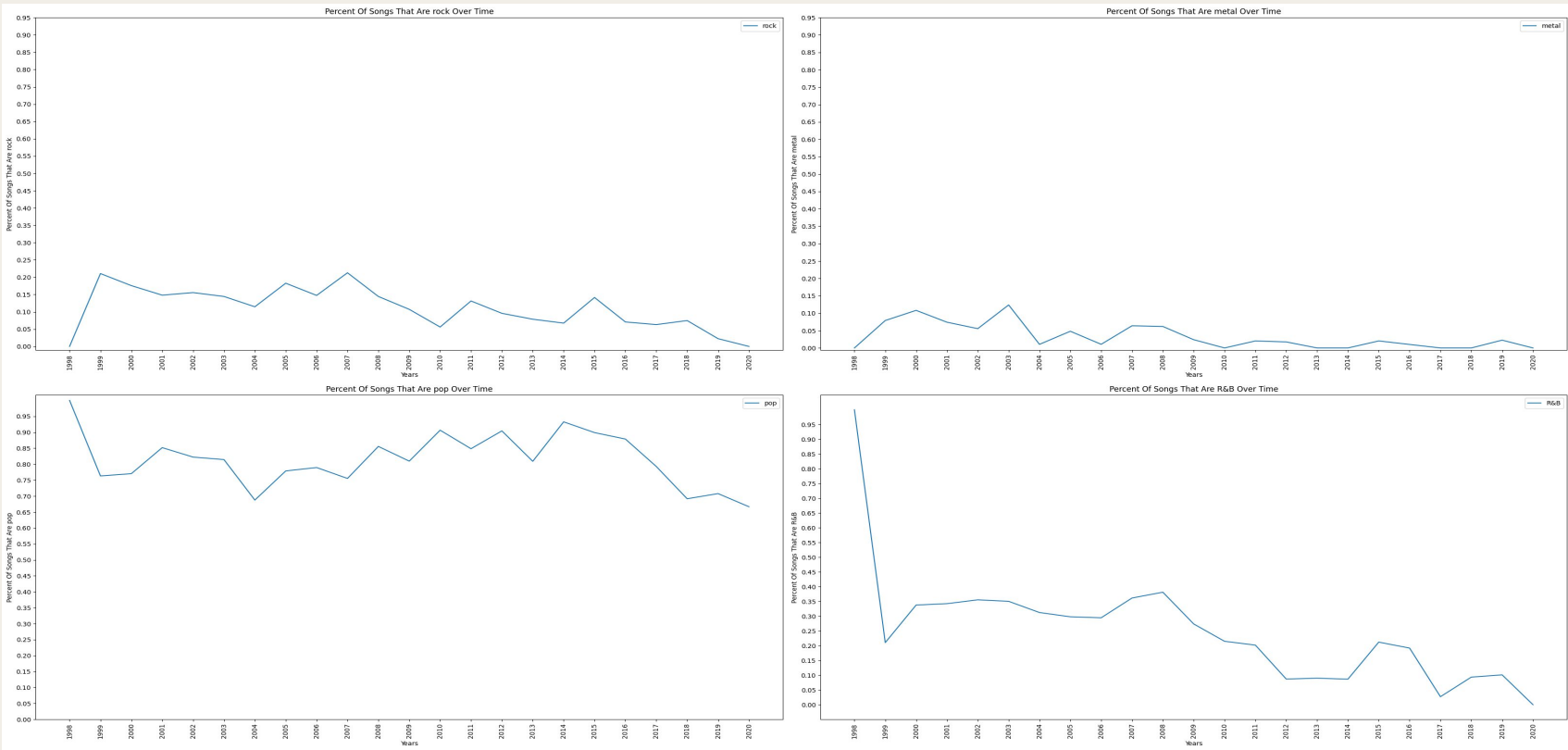
We can see that Rock, Metal and R&B are in a decline, while Pop is in a steady position.

On the right is the break down of genres per year with Pop, Hiphop, and dance/electronic music being the largest groups.

K-means fails to find meaningful clusters. We hypothesize that K-means is too simple of an algorithm or music is too diverse to properly cluster

Conclusion

- Contributions
  - Identified clear trends with genres over time
  - Found the correlation between song elements and popularity
  - Showed that K-means was insufficient to cluster our music dataset
- Future Work / Issues
  - Find a larger and more wholistic dataset. This dataset only contained the top 100 songs on Spotify for 20 years, which is relatively small
  - Perform research into why the trends we see are occurring. (ex. Why is metal not as popular?)
  - Research alternative clustering methods
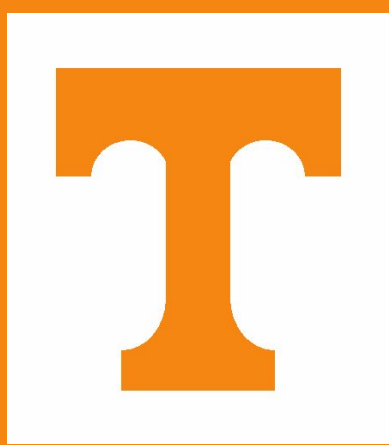
THE UNIVERSITY OF TENNESSEE KNOXVILLE