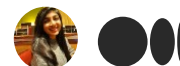


[Open in app](#)

## Aroofa Maknojia

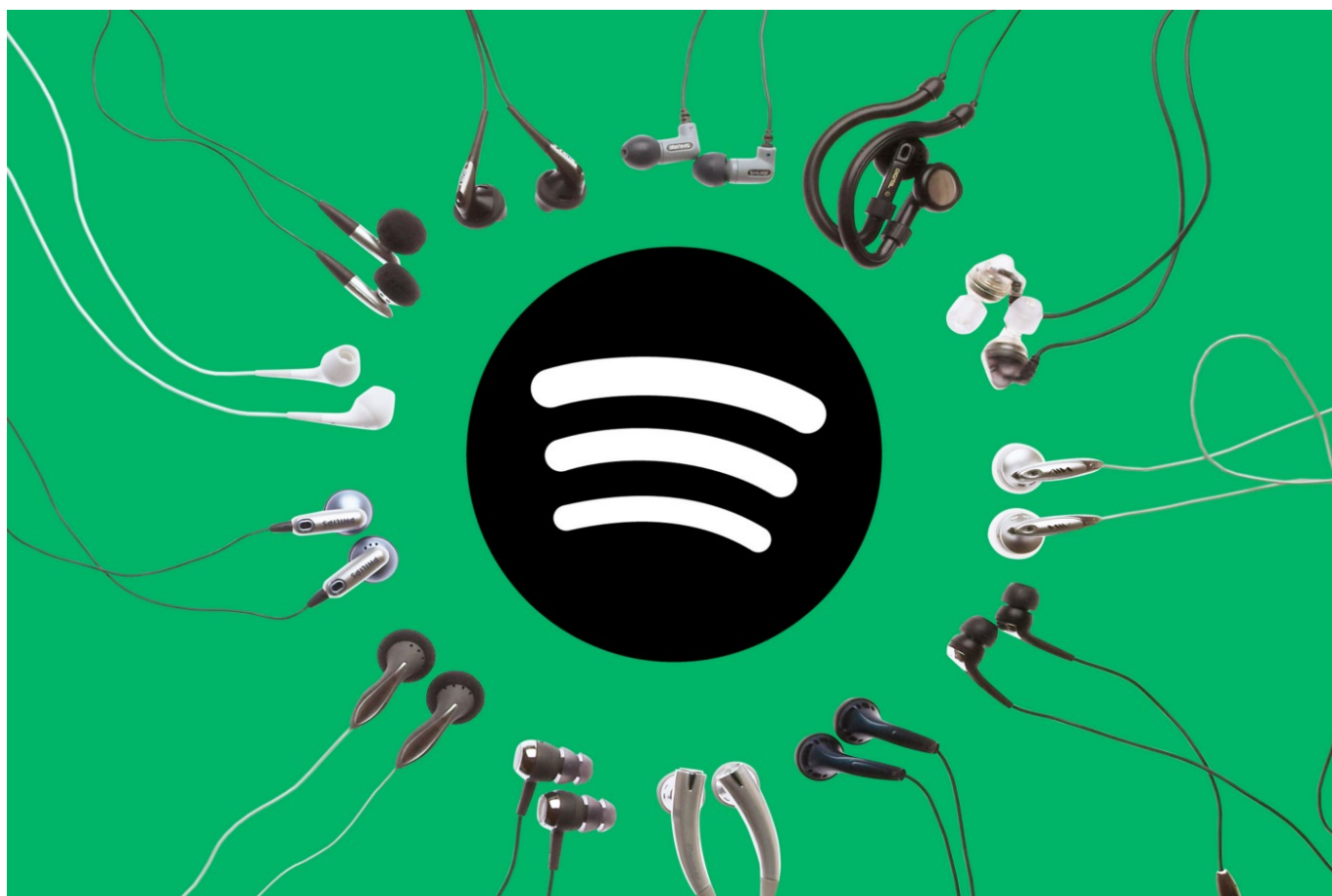
**New!** Select "Design your profile" in the menu to create a personalized space.

[Got it](#)

# An Exploratory Analysis of the Top 2000 Songs on Spotify



Aroofa Maknojia Just now · 8 min read



Spotify is one of the world's most popular music streaming platforms. From pop music to crust punk, Spotify houses a variety of genres that satisfy the preferences of their

consumers. The company makes playlists of the top songs and artists that are being streamed on their platform and displays them on their streaming service, building collections of music. Currently, they have over 50 million songs on their platform. I'll be looking into the audio characteristics of a small subset of songs that they call the top 2000 songs on Spotify.

The dataset I used contained audio statistics from the top 2000 songs. The songs were released from 1956–2019. The data included various characteristics and qualities of the songs including the title of the song, name of the artist who created it, genre, release year, tempo (measured by Beats per Minute (BPM)), and length. Additionally, other characteristics like energy, danceability, loudness (dB), liveness, valence, acousticness, speechiness, and popularity were measured based on how much the song displayed those attributes. If there was a higher value associated with one characteristic, the more likely the song exhibited that attribute. For example, if one of the songs had a score of 96 (out of a range from 0–100) for danceability, it was easier to dance to the song. However, if a song ranked 10 for danceability, it was harder to dance to the song. To clarify what some of the characteristics mean, valence relates to a positive mood for the song and speechiness means the spoken words in the song. All of these qualities and characteristics made up the columns of the dataset. Most of the values for the characteristics of the music (energy, danceability, liveness, valence, acousticness, speechiness) are judged on a scale of 0–100. Overall, there were 15 columns in the dataset including an index and 1,994 rows. All of the columns were relevant in my analysis and I used them as I explored my guiding questions. Here is the [source code](https://aroofa-maknojia.medium.com/an-exploratory-analysis-of-the-top-2000-songs-on-spotify-effcc6d24fd) for my analysis.

As I begin my exploration of the dataset, I hope to analyze my data according to the following guiding questions:

1. What characteristics of songs do Spotify users value the most?
2. How has time affected the top characteristics of popular songs?
3. What genre of songs are the most prevalent in the top 2000 songs?

To get a better understanding of the songs that Spotify users value, I started off by taking a look at the popularity distribution of the songs in the dataset to understand the

frequency distribution of the ratings. The histogram was left-skewed and unimodal with a mean of around 59.5. This shows that there were a large number of occurrences for the upper-value popularity ratings than there were in the lower-value ratings. The histogram set up the foundation for understanding the variation of the popularity ratings.

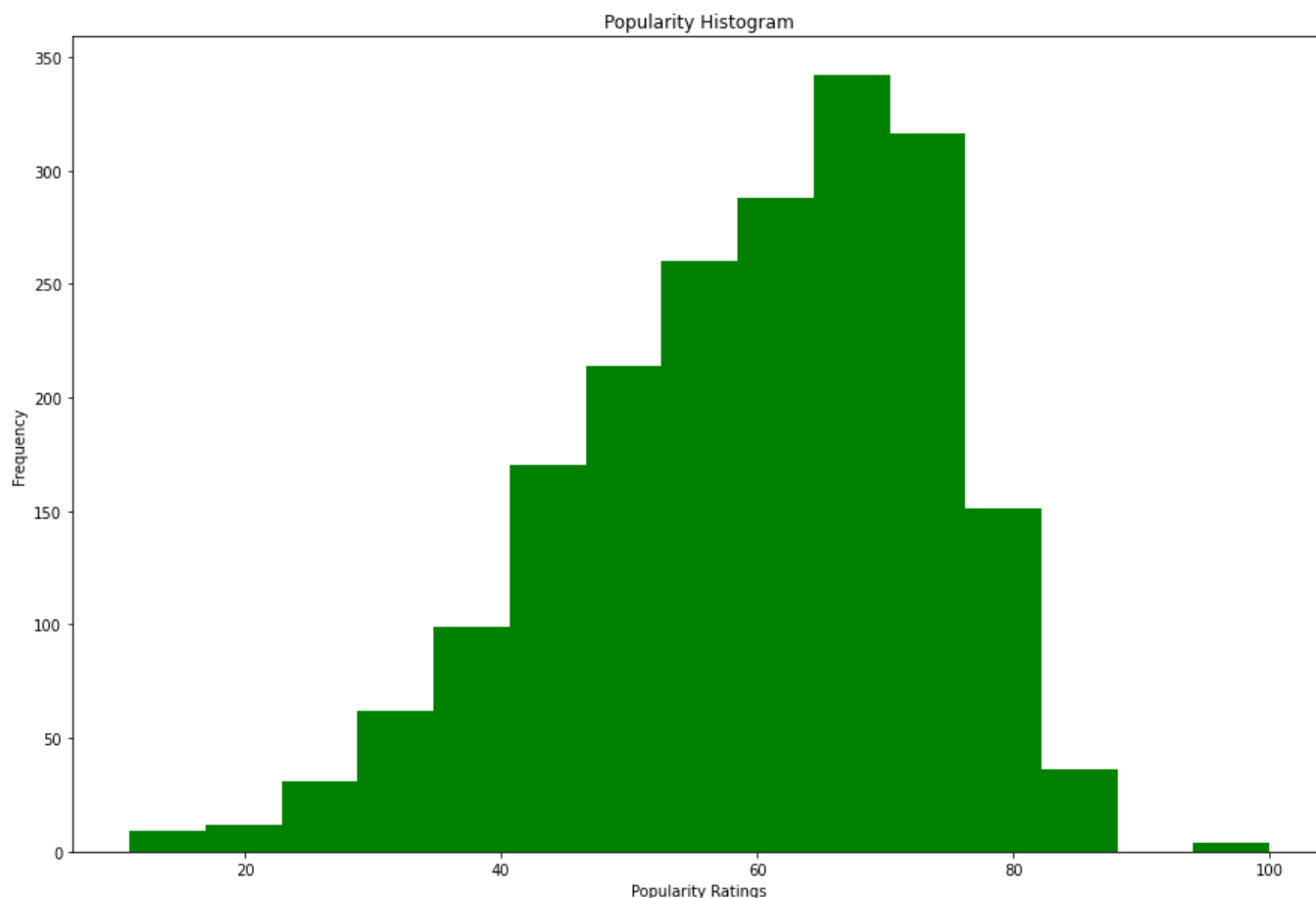


Figure 1: Histogram of the Popularity Ratings of the Songs Grouped in 15 Buckets

I wanted to get a better look at how the various song characteristics compared with the popularity ratings. Were there any trends of popular songs? Any characteristics that correlated with high popularity ratings? This would help me understand if there were any characteristics of songs that Spotify users valued the most. I started off by creating a scatter plot of the energy scores with the popularity ratings for the songs. Each dot represents one of the 1,994 tracks present in my dataset. I continued to create a scatterplot for the rest of the characteristics, creating eight scatterplots in total. Overall,

the graphs signaled that there was no correlation between the various characteristics with popularity, so the popularity of songs was largely independent of the type of characteristics it had. There is no recipe or “correct” formula in creating a hit song. Oftentimes, the popularity of a song is dictated by external factors like how well the song has been marketed, how popular the artist is, if there is significant word of mouth for the track, etc.

Characteristics vs. Popularity

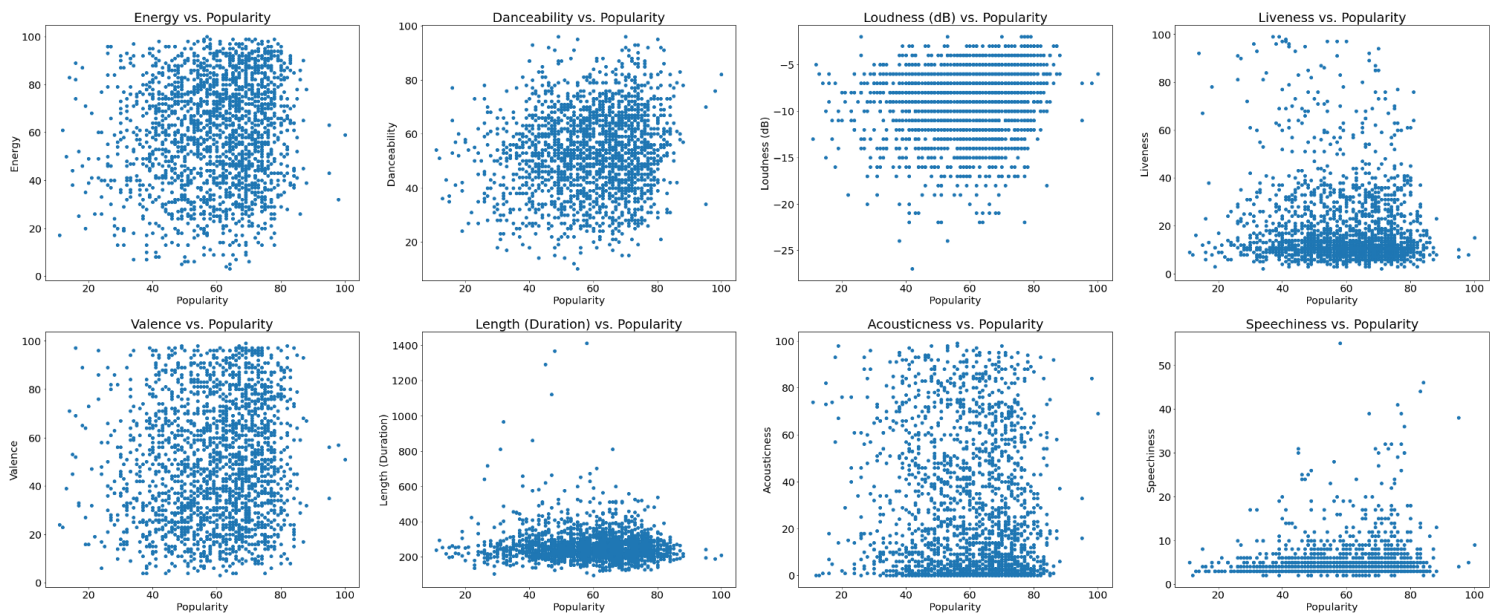


Figure 2: Scatter Plots for the Various Characteristics and the Popularity Ratings for the Songs

If popularity and the characteristics weren't correlated, were there any characteristics that were highly correlated with one another? To examine this, I created a heatmap that would display the correlation coefficients between the characteristics. The heatmap confirmed my assertion that the song characteristics and popularity are not correlated. Here are some of my takeaways from the graph:

- Energy and Acousticness have a strong negative correlation. High energy songs have a lower acousticness value.
- Valence and Danceability have a moderate positive correlation as well as Valence and Energy. This shows that positive songs tend to be easier to dance to and have higher energy values.

- Loudness and energy have a strong positive correlation which means loud songs tend to have high energy values.
- Acousticness and Loudness have a moderate negative correlation which means acoustic songs are not loud.
- The rest of the correlation values between the characteristics are weak which means there isn't a linear relationship between them.

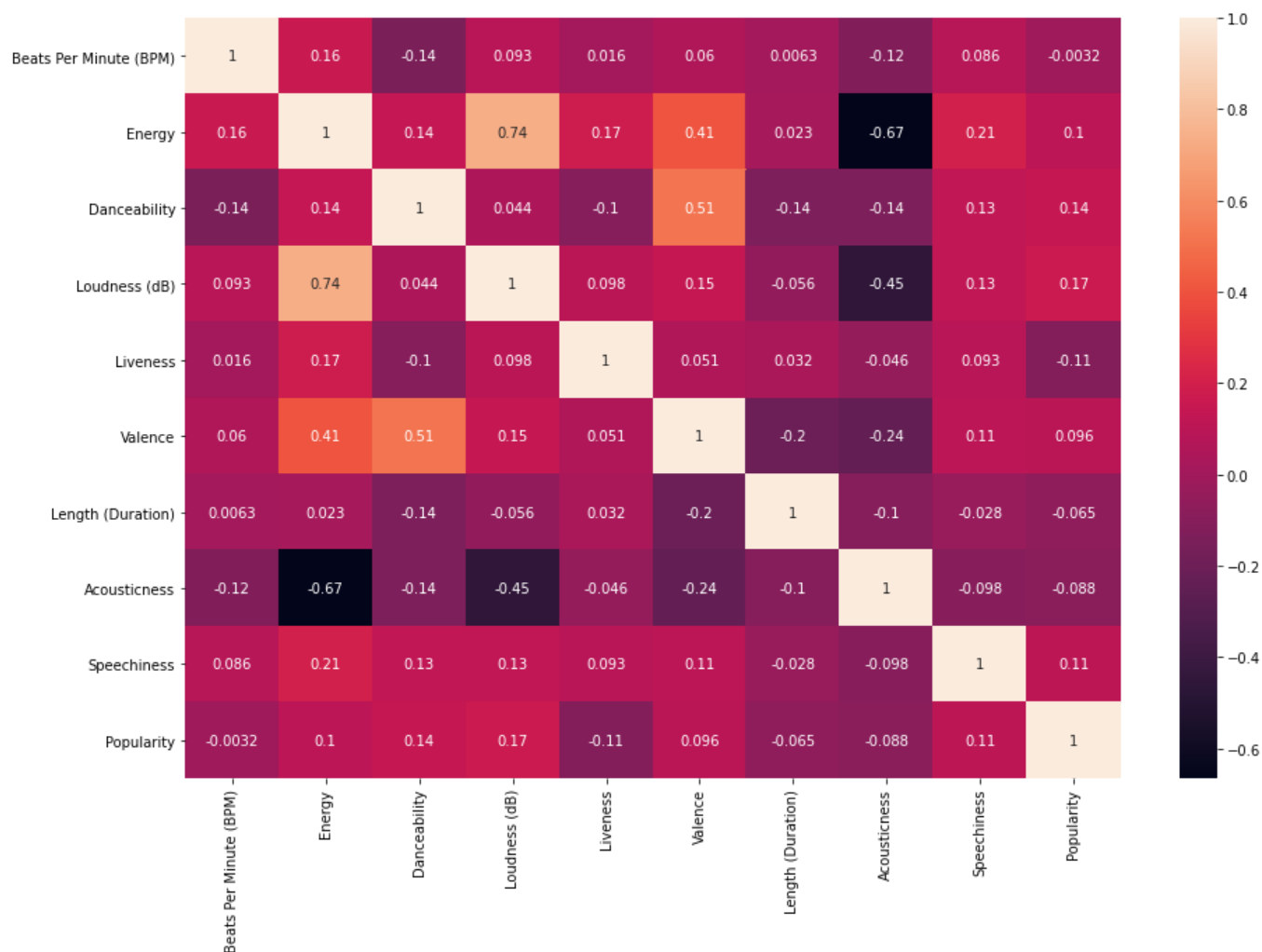


Figure 3: Heat Map of the Various Characteristics of Songs Displaying the Correlation Coefficients

After looking at how the characteristics and popularity correlated with one another, I got curious at how the characteristics of the top 2000 songs have fared over the years. How do modern-day songs compare to those in the past? I started my analysis by taking the average value for the song characteristic for every year and I graphed these values in a line chart. I repeated this for every song characteristic to get a total of 10 charts. There tends to be a spike at the start of every line graph. This happened because there were

fewer songs in my dataset from the 1960s compared to the other years. Overall, modern-day songs have more energy, are louder, and less acoustic. However, what was the most interesting to see was that songs in the past had higher popularity ratings than those of the present.

Song Characteristics Over Time

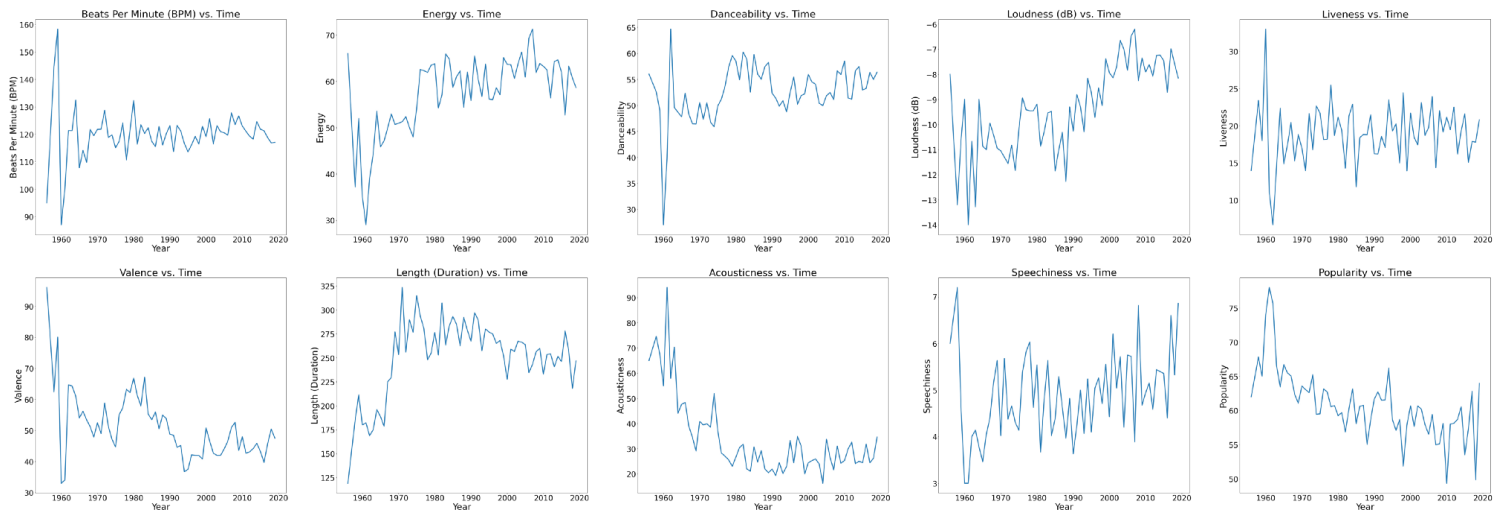
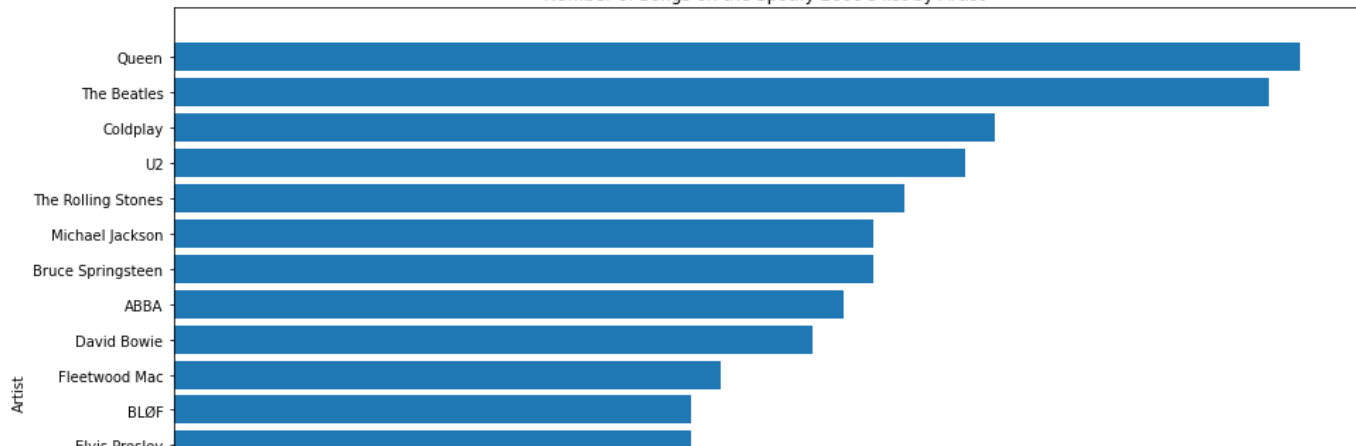


Figure 4: Line Chart of the Various Characteristics with Time

After looking at how the song characteristics have fared over the years, I wondered which artist had the most songs on the top 2000s list. I created a horizontal bar chart that would display the number of songs that every artist created that landed on the chart. However, there were hundreds of artists with many having only one or two songs land on the list, so I picked the top 20 artists that had the most songs. The rock band, Queen, had the most number of songs with 37 songs in total followed by the Beatles that had 36 and Coldplay that had 27. It was interesting to see that the top three artists with the most frequent songs were all British rock bands.

Number of Songs on the Spotify 2000's list by Artist



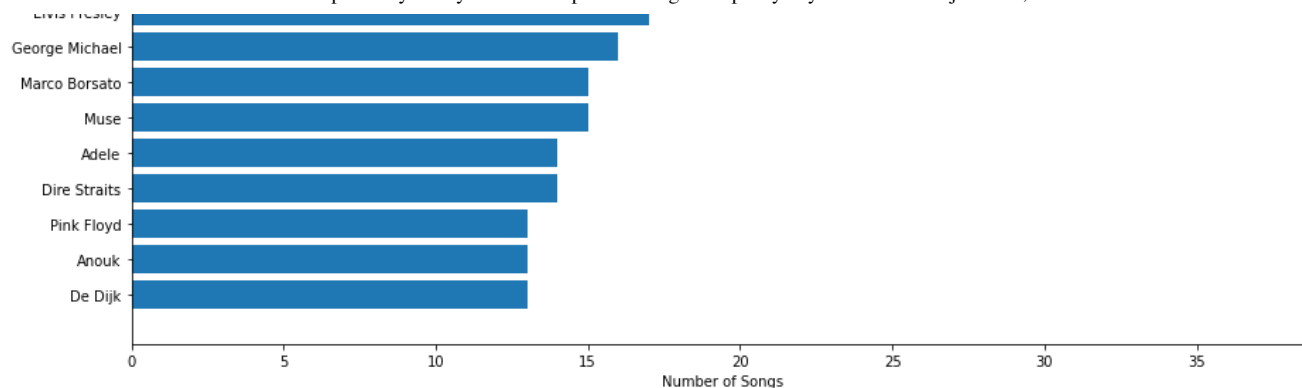


Figure 5: Horizontal Bar Chart Displaying the Number of Songs Released by 20 of the Artists on the List

After seeing that Queen released the most songs, it made me curious what years these songs were released. I created a pie chart that would show the distribution of how many of their songs were released per year. Although 1991 is a big release year for songs that landed on the list for Queen, the majority of songs that landed on the list were released in the 70s or 80s. It was interesting to see that after 23 years of their last modern release that landed on the charts in 1995, they were able to get another song on the list in 2018 called “I Want to Break Free”.

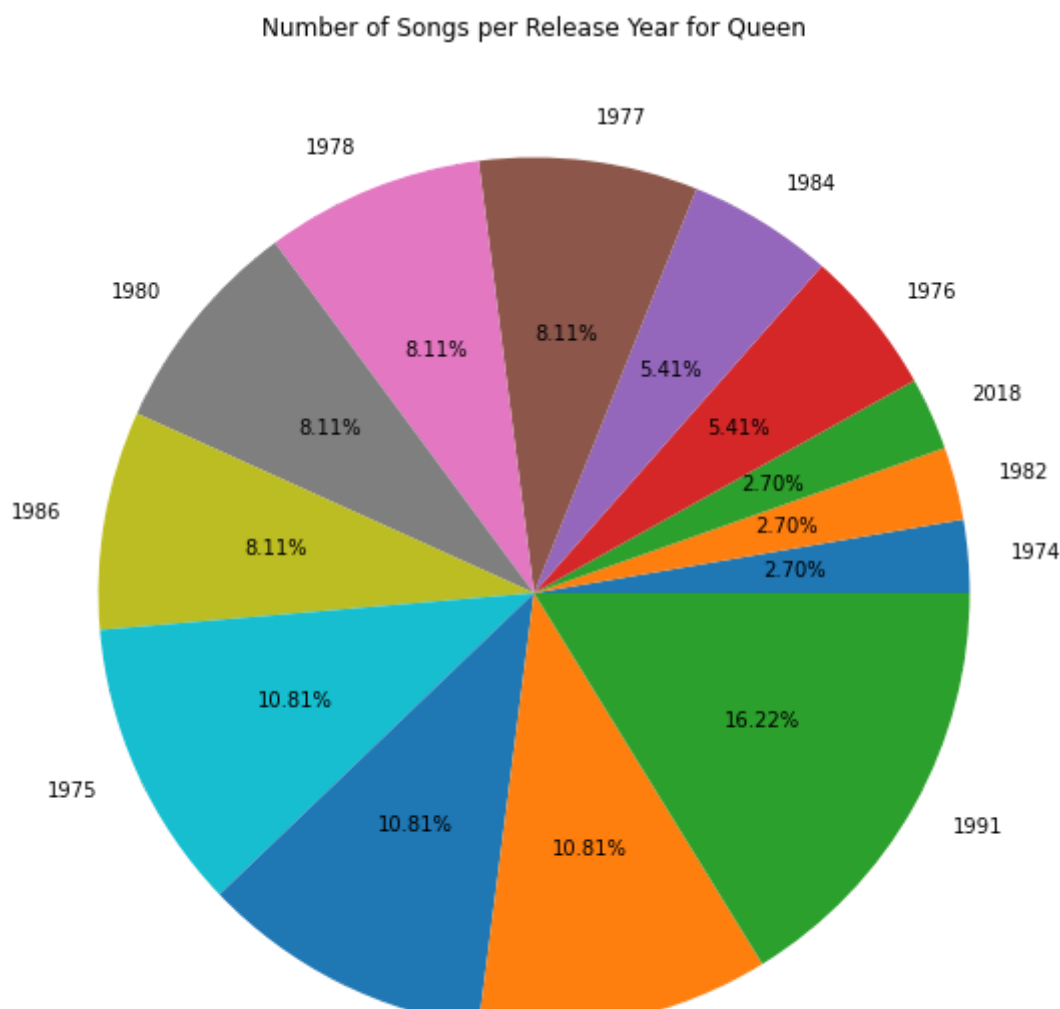


Figure 6: Pie Chart Displaying the Percentage of Songs Queen Released per Year that Landed on the Chart

I went back and expanded my focus on the top 20 artists with the most tracks. I hypothesized that since these artists had the greatest number of tracks on the list, they were extremely popular for the Spotify users which meant that their songs overall would have less lower popularity ratings and more higher popularity ratings. I created a bar chart that would cover all of the popularity ratings for the tracks of all of the top 20 artists that had the most songs on the list. I found that as I hypothesized, there were fewer songs that had lower popularity ratings and more songs that had higher popularity ratings.

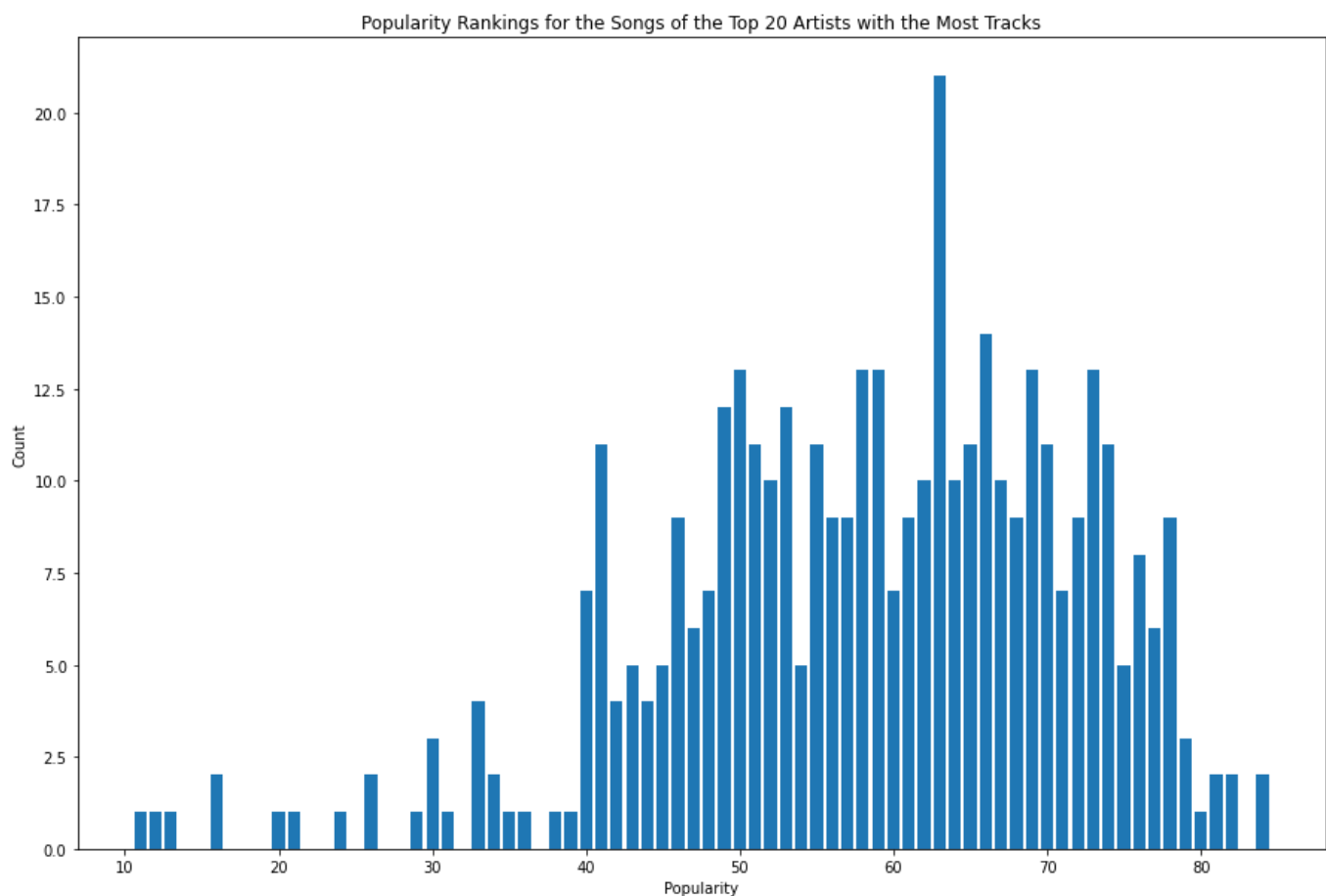


Figure 7: Bar Chart Displaying the Distribution of Popularity Ratings for the Songs of the Top 20 Artists with the Most Tracks

I continued to expand my analysis on the top 20 artists with the most tracks by looking into what genre was the most popular for the songs of these artists. I found that Album



Rock is the most popular genre and British Soul was the least popular.

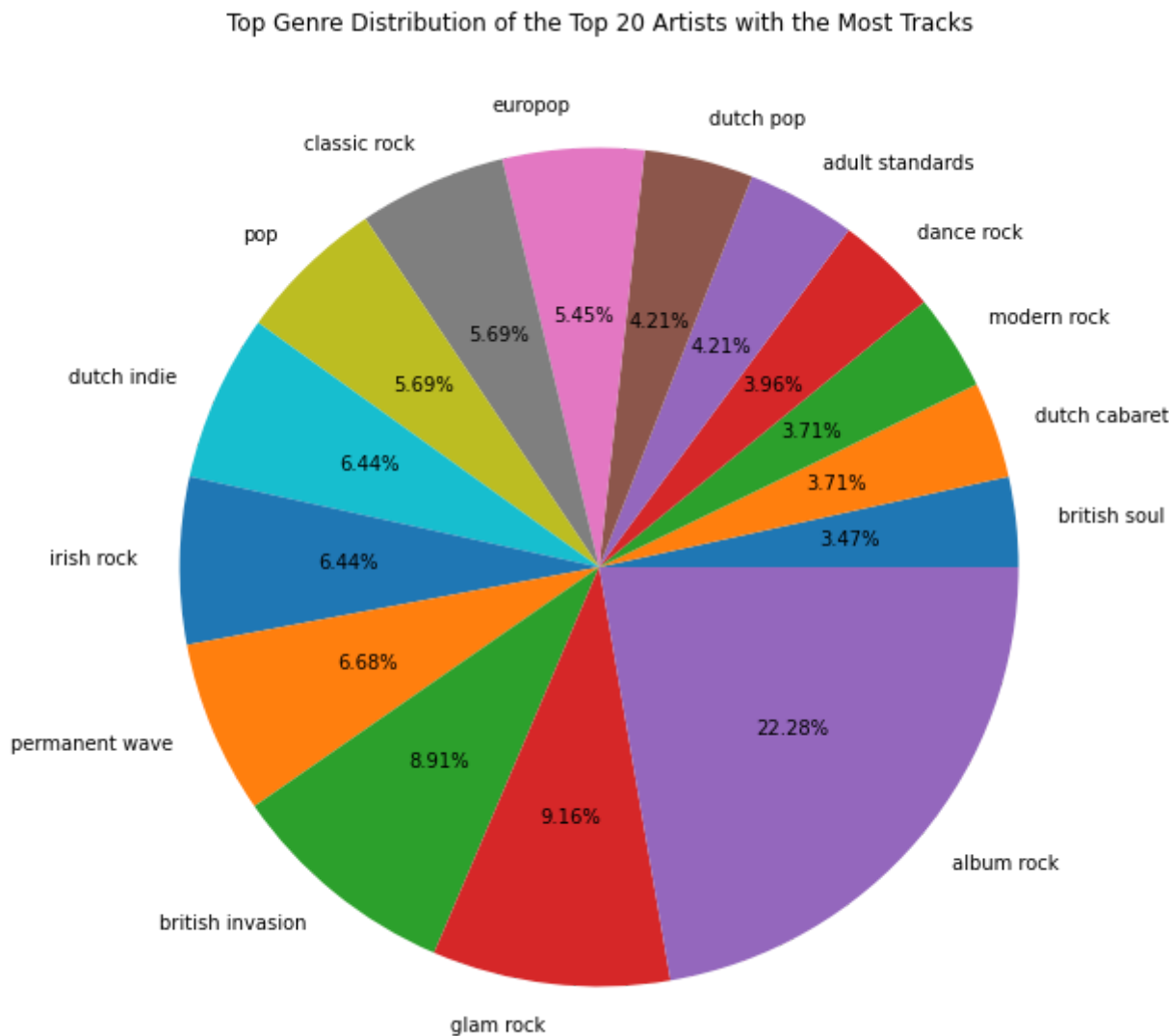


Figure 8: Pie Chart Displaying the Percentage of Songs that Fell into Each Genre Category for the Top 20 Artists with the Most Songs on the List

After looking into the distribution of genres for the songs of the Top 20 artists, I expanded my outlook to look into the top genres of all of the songs on the top 2000s list. Similar to the artists, there were hundreds of genres available, so I limited my search to the top 35 genres. Album Rock ranked first on the list as the most popular genre which is similar to the genre distribution of the top 20 artists. There were 413 songs that fell into this category. Adult Standards was second with 123 songs.



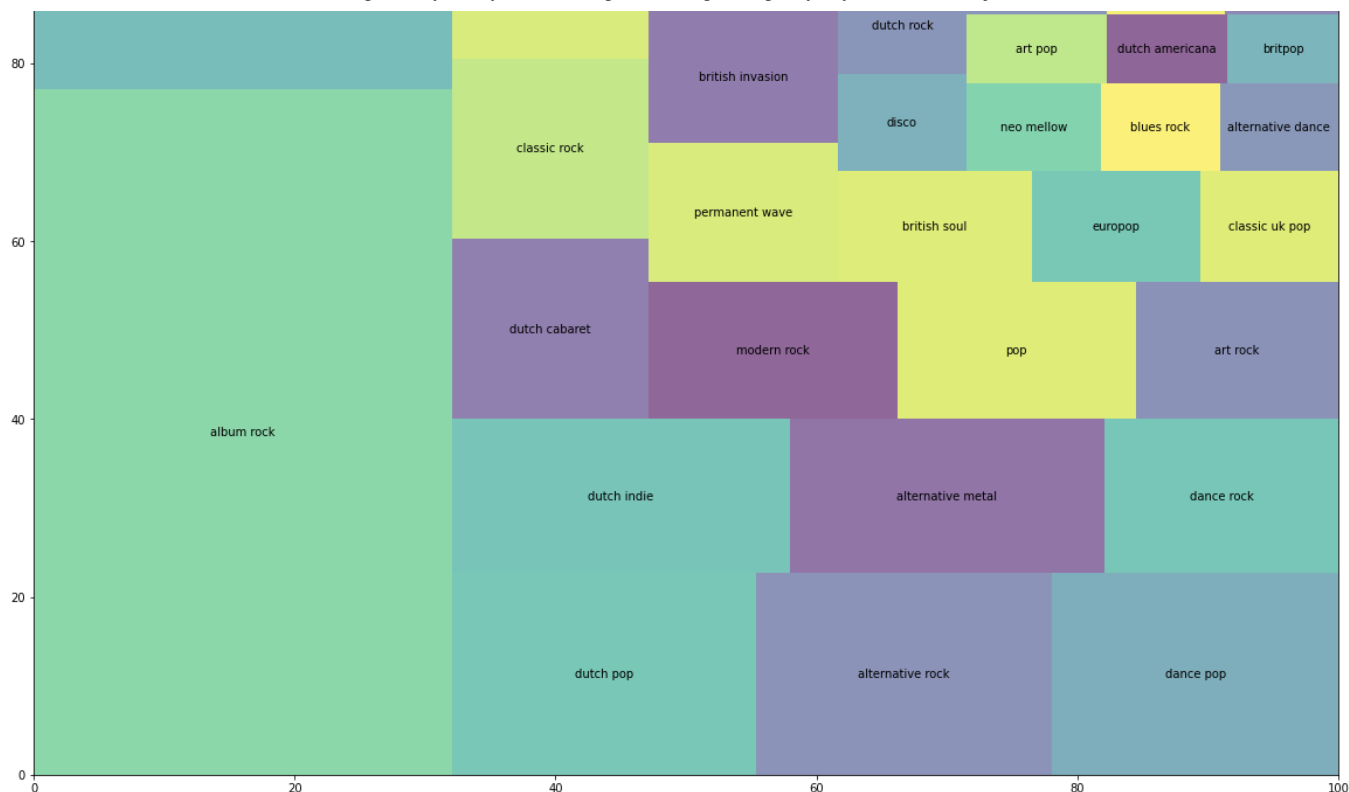
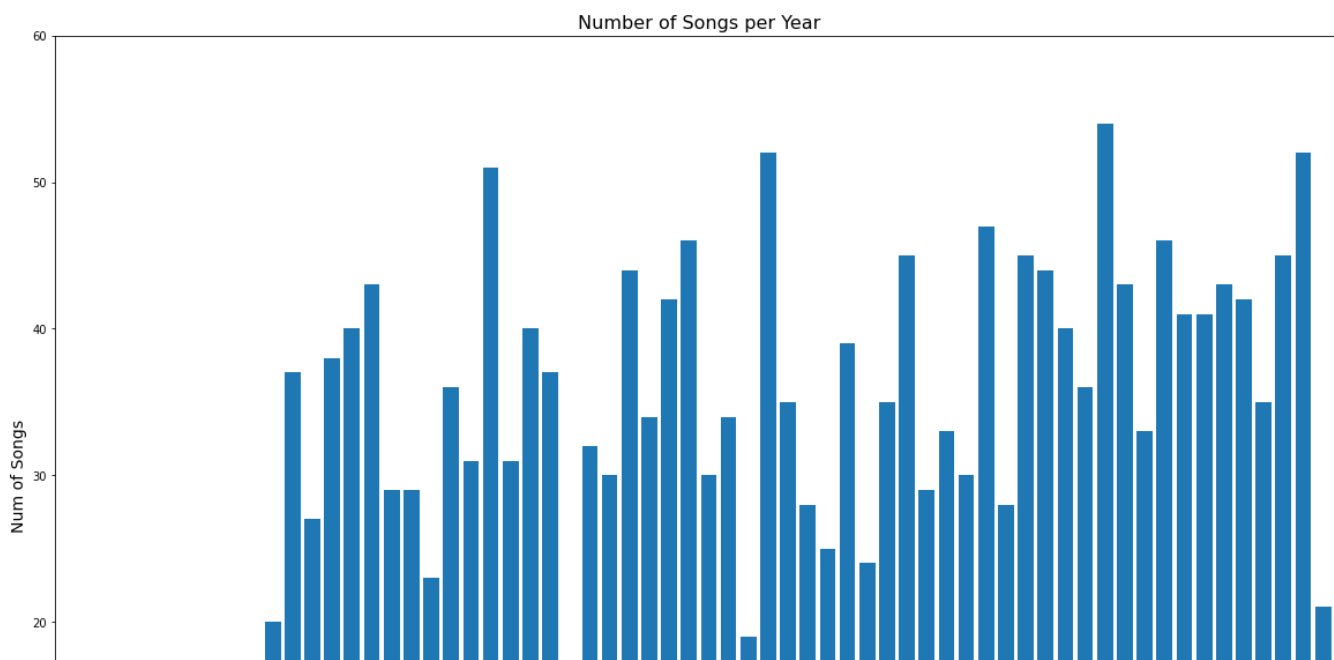


Figure 9: Treemap of the Top Genres on the Spotify 2000s List

I wondered if the dataset had any decades that it had more songs in than other decades. This would show if a decade had an impact on the top genre quantities. I created a bar chart that would categorize all of the songs by the release year. The songs were fairly evenly distributed over all of the decades except for the 60s which didn't have as many songs as the other decades. The top genre of Album Rock was not influenced by a particular decade.



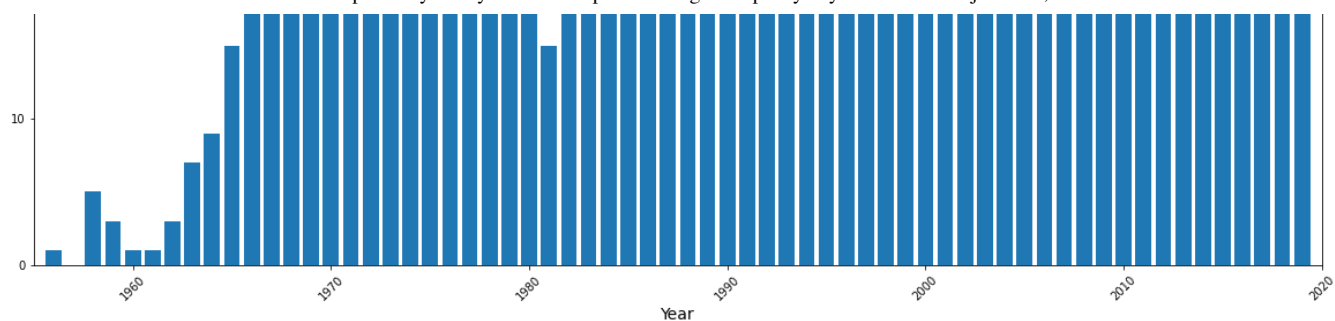


Figure 10: Bar Chart Displaying the Number of Songs Released per Year

Here are some of my takeaways after doing the exploratory data analysis on the Spotify Top 2000s list:

- 1. Popular songs don't have common song characteristics.**
- 2. Modern-day songs have more energy, are louder, and less acoustic.**
- 3. Album Rock is the most popular genre on the Top 2000s list.**

It was exciting to dive deeper into this dataset and extract meaningful insights from the raw data. I hope you got a better picture of the Spotify 2000s list from this analysis!

Data Science

Exploratory Data Analysis

Spotify

Pandas

Visualization

[About](#) [Help](#) [Legal](#)

Get the Medium app

