Music trends

While starting this project I wasn't really sure what I wanted to look into. There is so much data around us that we could use to infer a lot of things, and answer a lot of questions. But, finding the right data and extracting useful information from it, isn't easy. For this project, I tried looking at different kinds of data for various topics. I finally wanted to look at the change in music trends across different metrics like best-selling artists, grammy winners, billboard standing, and Spotify data. I wanted to see the potential correlation or difference between this data. I wanted to find out if a specific genre of music was more popular on one of these measures. If yes, how does it differ from other ones, and how is it different currently?

To answer this question as much as I can, I decided to look at CSV files containing data from before 2020. I used three different CSV files. The first one has information about the best-selling artists from 1960 to 2019. The second one has data about the artists on billboard charts. And, the third one had data on the Grammy winners from 1999 until 2019. All of these CSVs had data until 2019. To compare these trends with more current data, I used the Spotify API to get their current numbers.

Looking at the best-selling artists' data, I wanted to see from which countries the most number of best-selling artists are from. I used value counts on the country column of the dataset, to get the total values of each country, and used Seaborn to plot it.

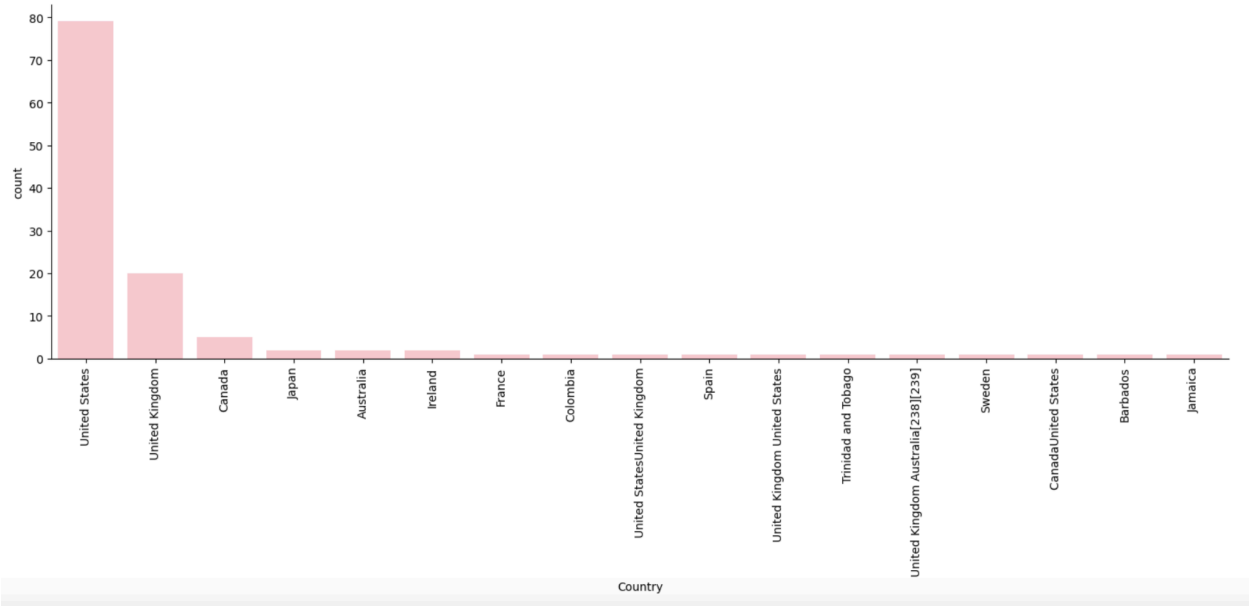

Figure 1: Best-selling artists across different countries.

From this graph, we first see that one country has the most number of artists from it. It isn't really surprising to see that it is the United States. But the surprising thing for me was how high the bar of the United States was than the other countries. It had around 80 while the next highest country was the United Kingdom, which was at 20. I don't really

know the exact reason for this skew in the result, but it might be because a lot of the data collected might have been from the Western world. Among the top 5 countries, 4 of them are English-speaking ones, so the majority of these best-selling artists are from there. Then, I wanted to look at how the distribution of the artist's countries was across the years. For this, I did a scatter plot to see the spread.

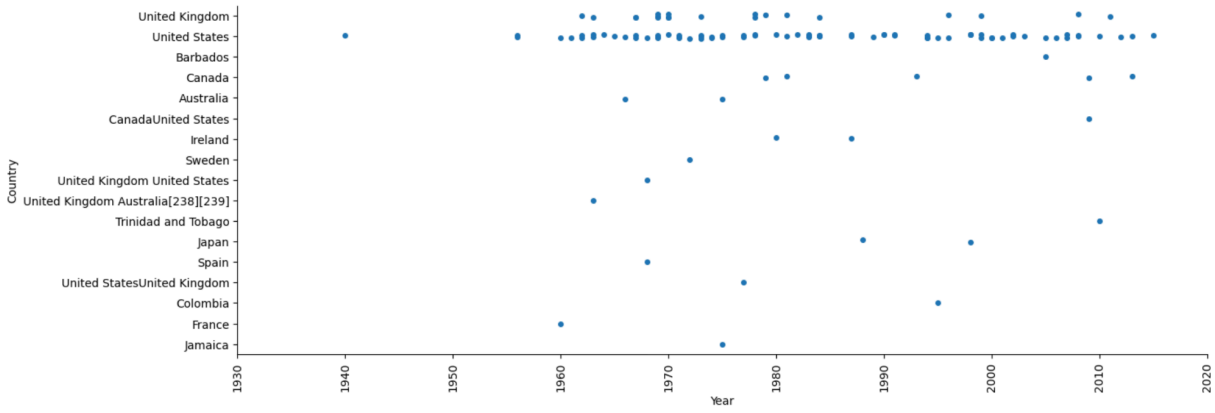

Figure 2: Countries with best-selling artists across the years

By looking at this chart, we can see that the dominance of the United States was not a recent thing but has been the same for over a couple of decades.

Then I wanted to look at wanted to look at what kinds of artists were best-selling throughout the years. As the dataset had a number of genres for each artist, I used the lambda function to split and get the artist's first genre. I used value counts to get the number for each genre.

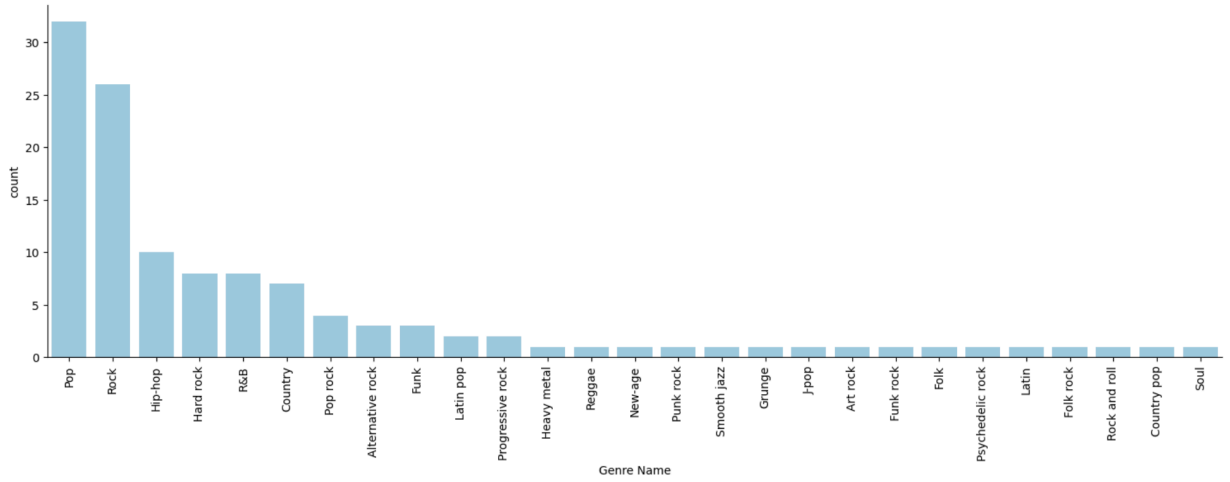

Figure 3: Genre of best-selling artists.

In this graph, we can clearly see that pop and rock are the most popular genres. Artists from pop, rock, and hip-hop take up most of the numbers. After this, I tried to see if any artists had been best-selling for a couple of years. However, in this dataset, all of the artists only had one entry, as in they were best-selling only once. This dataset did have the artists in the order of most number of sales, and unsurprisingly The Beatles, Elvis Presley, and Michael Jackson are the top ones.

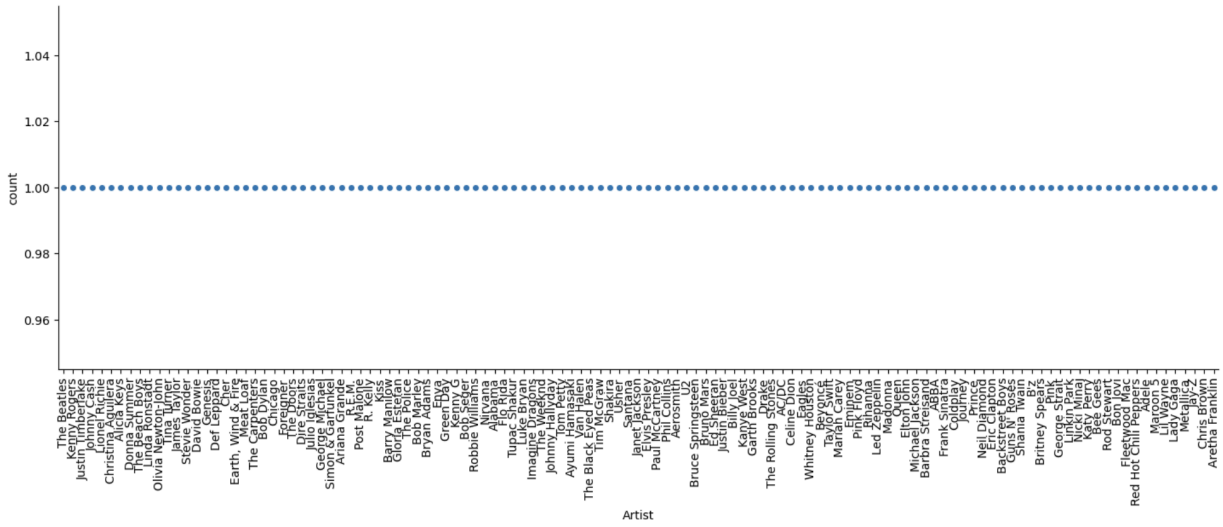

Figure 4: number of times an artist is best-selling.

Then I wanted to look at the billboard dataset. I used Groupby to get all of the entries from a single artist and then I added all of the weeks an artist was on the charts. In Figure 5, we can see that Drake and Taylor Swift are kind of close to each other at the top with around 976 and 907 weeks each. Followed by Rihanna and Kenny Chesney. It is interesting to see that these weren't the top most selling artists from the previous dataset but were on the charts for a whole lot.

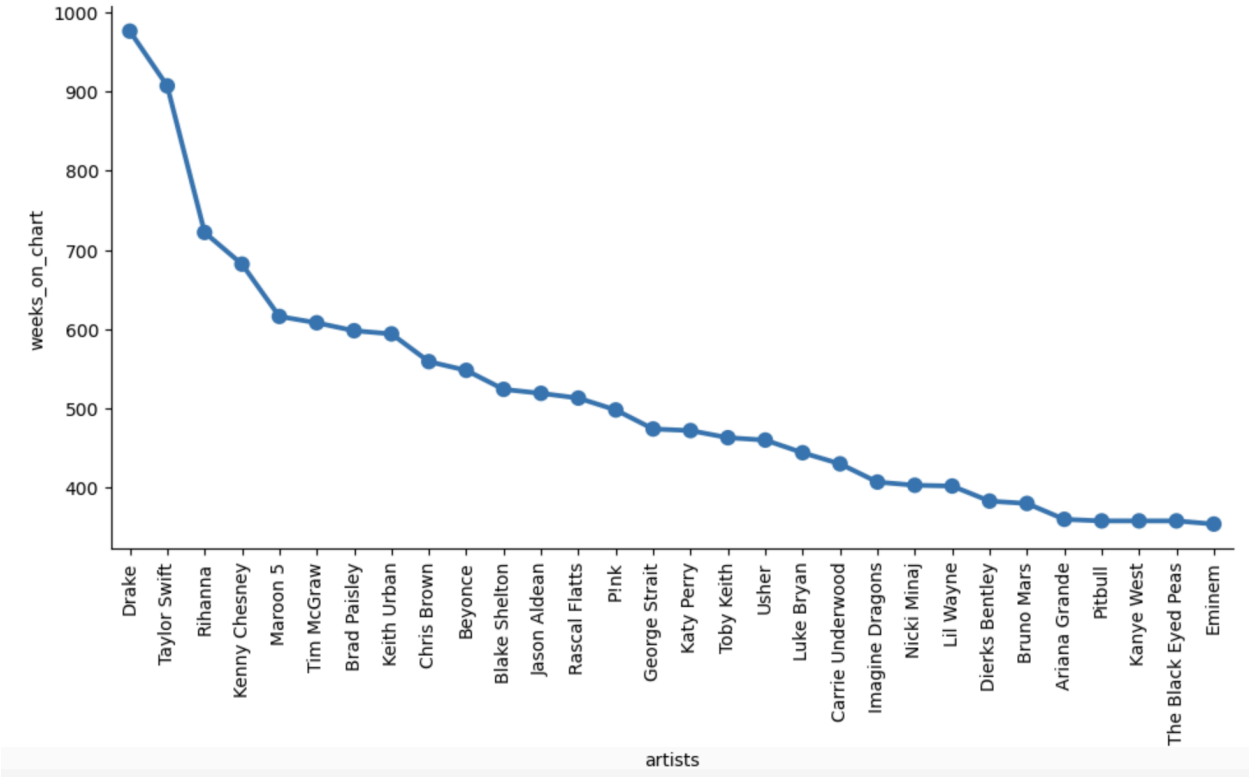

Figure 5: Highest weeks on the chart

Then I wanted to look at the most popular genre on the charts, to compare it with Figure 3. To do this I used a lambda function to get the primary genre and then added the number of weeks after using Groupby on the genre. In Figure 6, we can see a significant difference between the spread across the genres. Country was very high on the chart with almost 5000 weeks more than it's follower pop. Country, pop, and rap are clearly very higher than the other genres. If we compare this to Figure 3, we can see a difference that pop, rock, and hip-hop are the most-selling genres but country, pop, and rap are more favored when it comes to the billboard data. This data combines factors such as sales as well as streams.

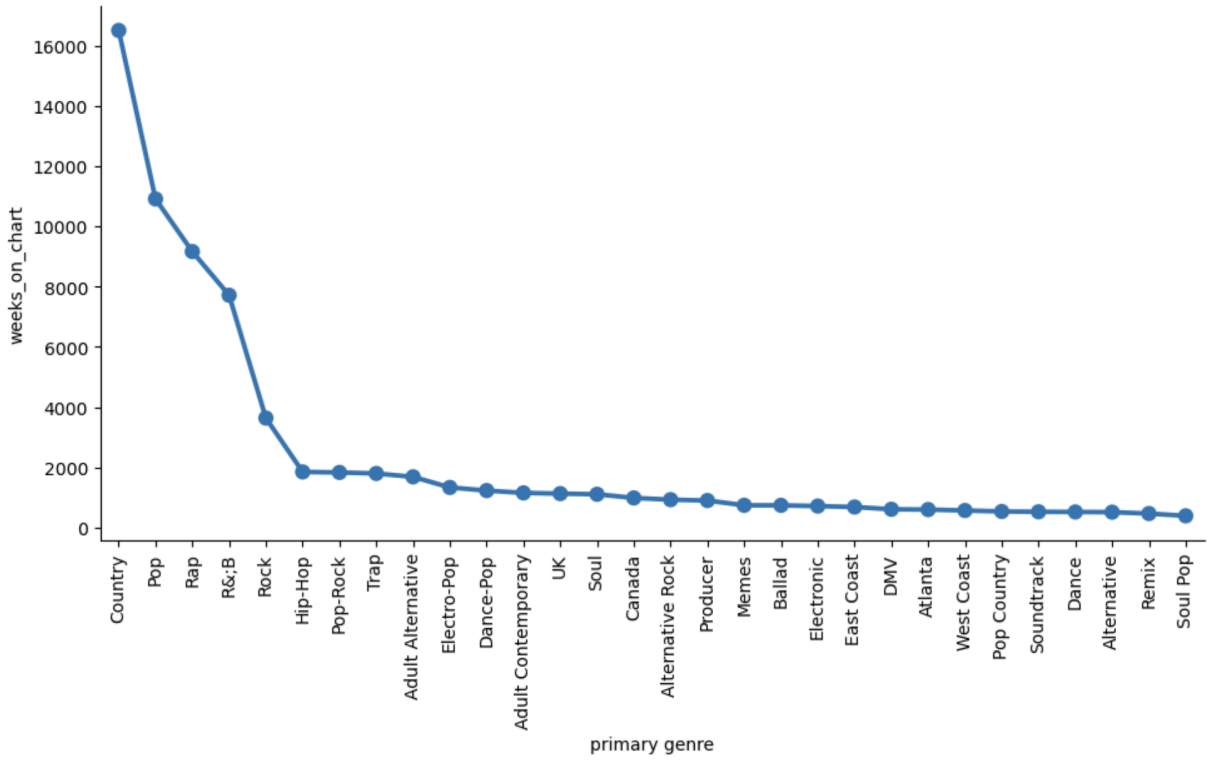

Figure 6: Genre on the charts

Now I wanted to see the Grammy data, to see if being on the charts or being the best-selling would affect the awards. Firstly, I wanted to see the artists with the most number of Grammys from 1999 to 2019. To do this I used value counts on the artist column of the dataframe. Then I used Seaborn to plot the top 30 artists with the most number of Grammys.

In Figure 7, we can see that Adele, U2, Alicia Keys, and Carrie Underwood were the ones with the most number of Grammys from 1999 to 2019. Figure 5, weeks on the chart doesn't really have any correlation with Grammy awards as we see different artists at the top.

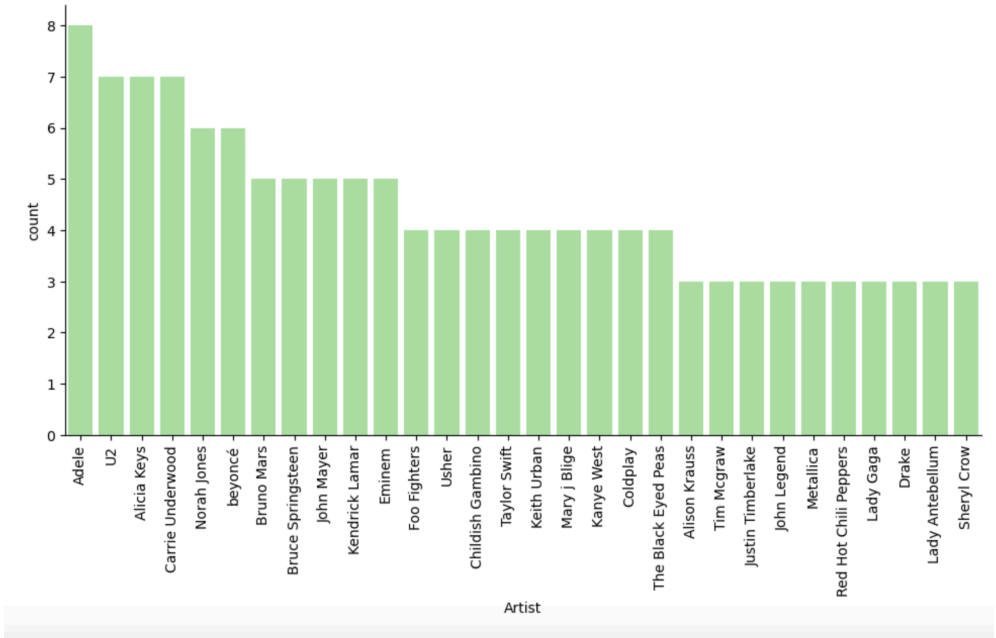

Figure 7: artists with the most number of Grammys(1999-2019)

Then I did the graph with the highest genres that got the award. To do this, I used value counts on the Genres. In Figure 8, we can see that rock, and R&B are tied with each other, and pop is pretty close to them. Pop is the genre that was consistently present in the top 3 for all of the graphs that I previously did. Even though rock is the most popular among the Grammys and was high on the best-selling, it was significantly lower on the graph with weeks on the chart.

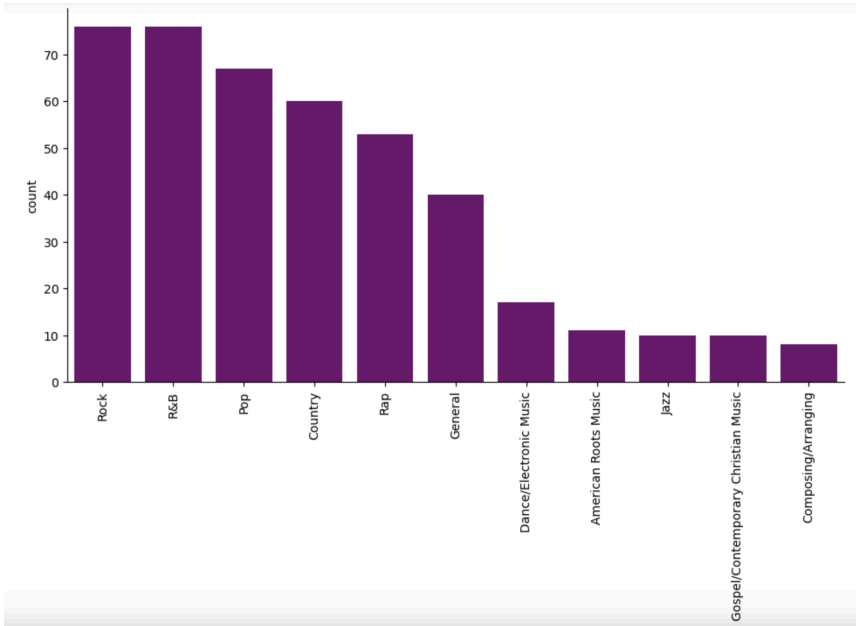

Figure 8: Genres with most Grammys.

Then I looked into the songs with the most number of Grammys and what genres they were from. To do this, after finding the songs with the most number of Grammys with value count, I created a list with these songs. I used this list to extract the columns with these songs from the Grammy dataset. Then I did the value counts for the genre of these selected songs and plotted a graph to show the distribution. In these figures, we can see that general had the most songs followed by rap. However, the most Grammys overall were for neither of these from Figure 8.

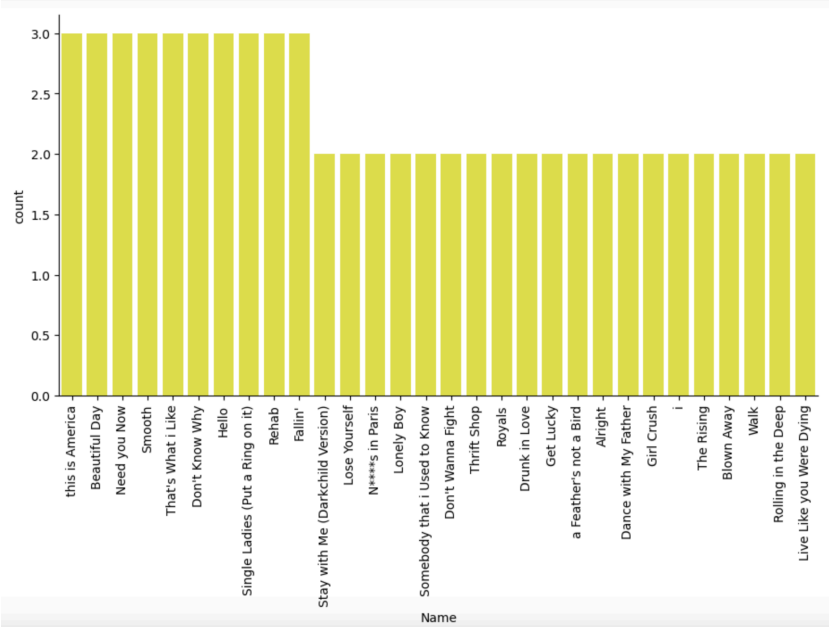

Figure 9: songs with the most number of Grammys (1999-2019)

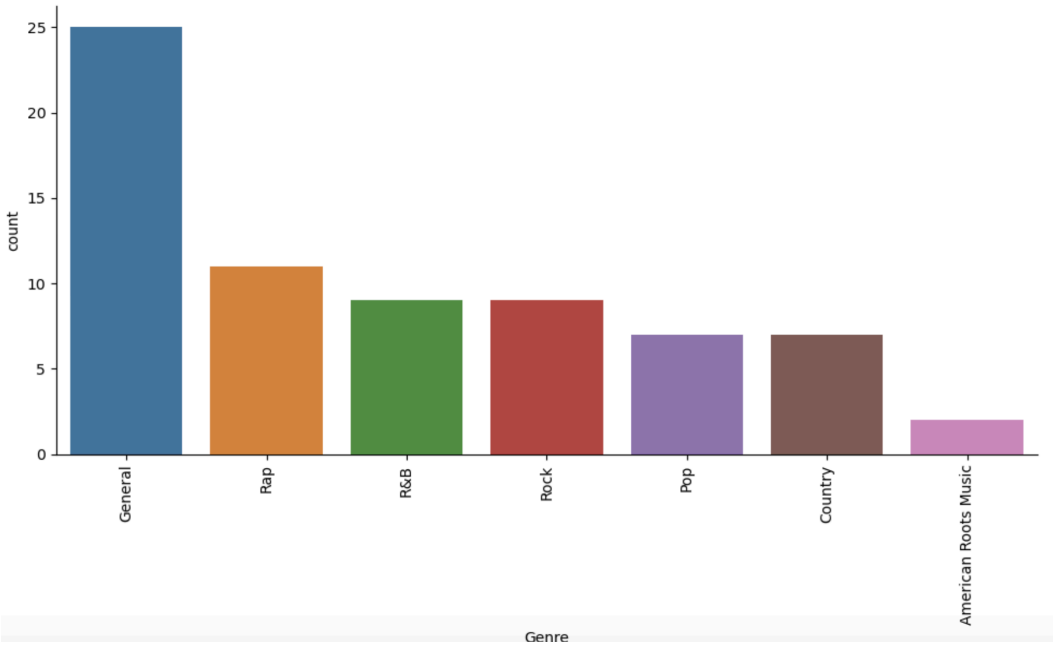

Figure 10: genres of songs with the most Grammys

Now, to compare this with the current data, I used the Spotify AP to get the popularity and followers for the most Grammy winners and most weeks on the charts. To do this I created two separate lists of the most Grammy winners and artists with the most weeks on the chart, by using .to_list() on the columns with these data. I then passed that list into the API request to get information about artists.

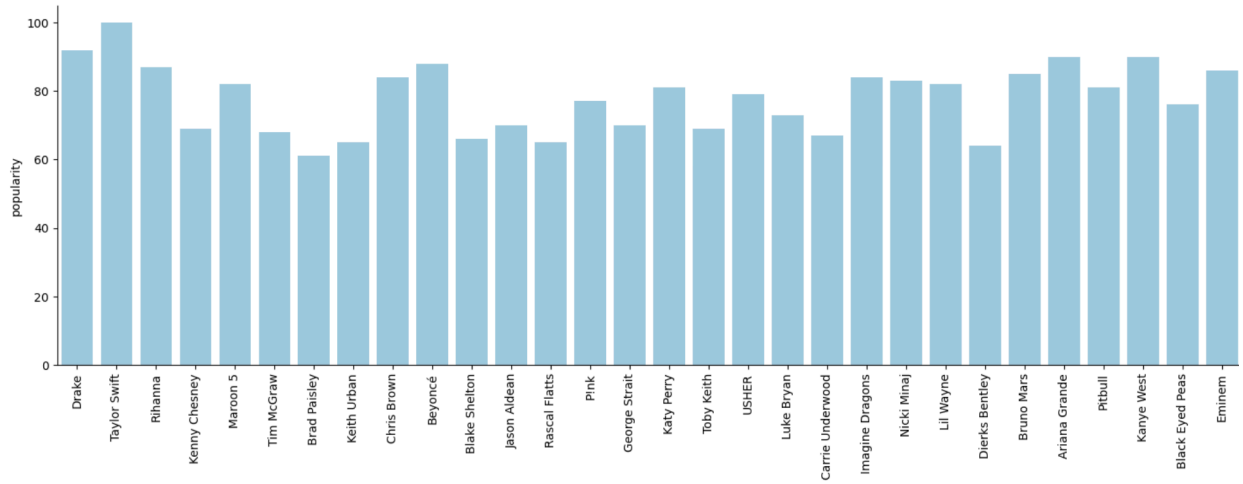

Figure 11: Popularity of artists in weeks on charts on Spotify

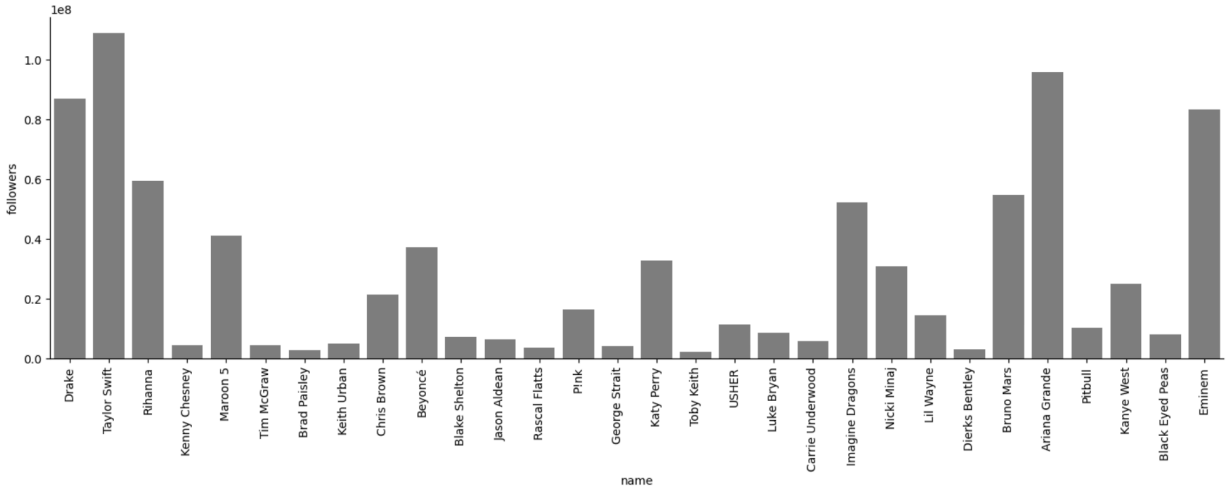

Figure 12: followers of artists in weeks on charts on Spotify

In Figure 11, we can see that the popularity of the artists is more evenly distributed compared to the weeks on charts from before 2019, which was very skewed. We also see that Taylor Swift has gained more popularity than Drake. In Figure 12, we see a lot of change across the artists, Taylor is leading Drake by a significant amount, and we can see that Ariana Grande and Eminem significantly shot up compared to their weeks on the charts.

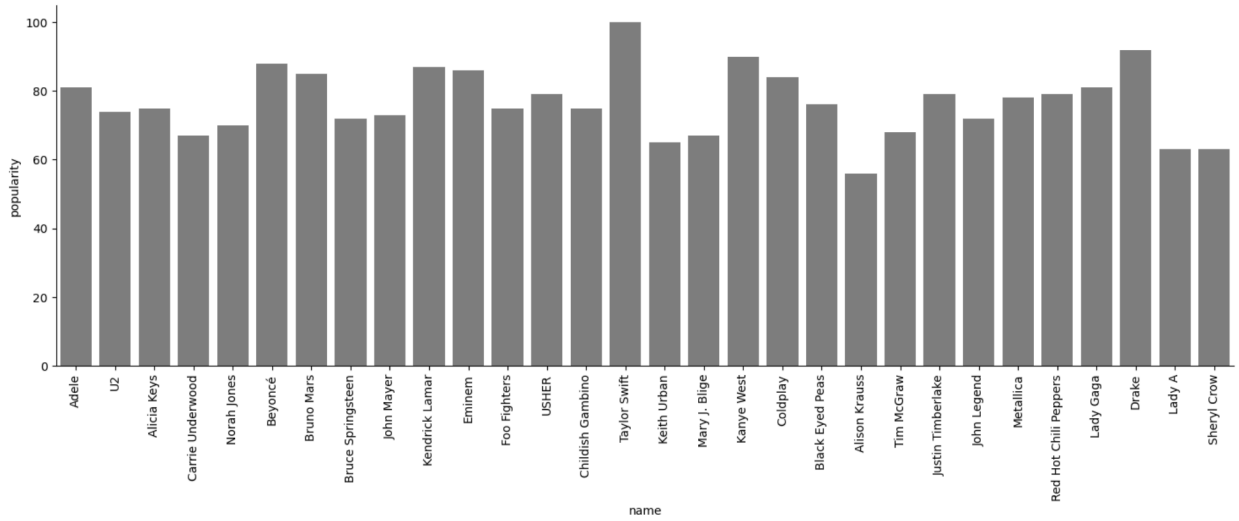

Figure 13: Popularity of the artist with the most Grammys(1999-2019)

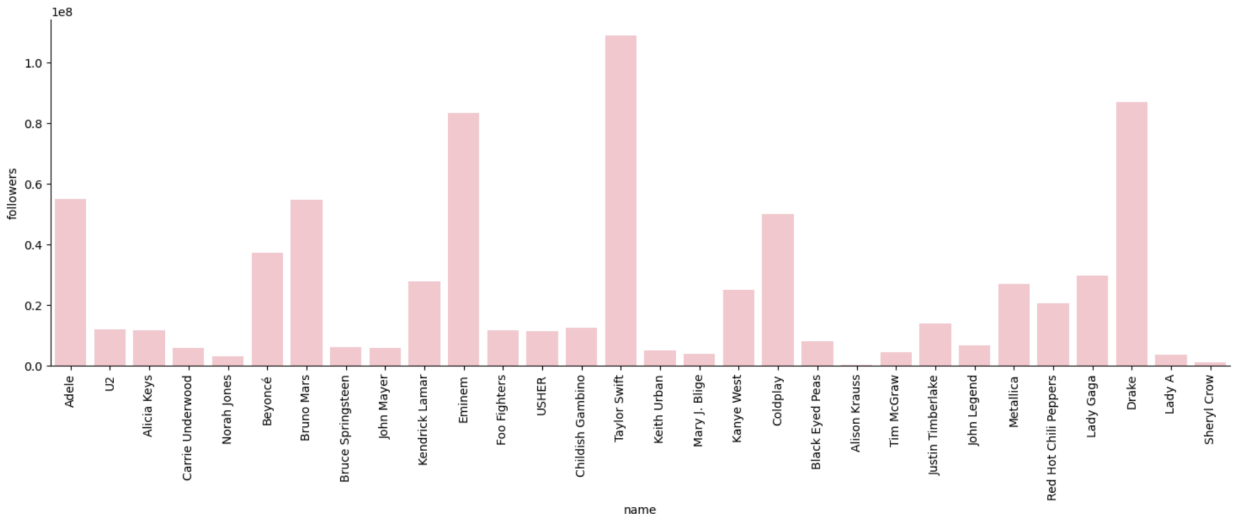

Figure 14: followers of the artist with the most Grammys(1999-2019)

In Figure 12, we can see that the popularity of these artists is somewhat distributed evenly, with Taylor, Drake, Kanye West, and Beyoncé somewhat standing up. In Figure 13, we can see that Taylor and Drake have more followers currently, despite being lower on the rankings with Grammy awards.

All of these graphs and plots show how the trends in music constantly evolve. There are so many factors that contribute to the standings of these artists on various metrics. After all of this analysis, we can see that monetization (best-selling artists) doesn't really mean recognition(Grammys) and vice versa.

If I were to further continue working on this topic, I'd try to find out what is the reason for having the United States as the country with the most best-selling artists. I'd look into how these patterns in music and the popularity of the artists vary across various countries. I'd try to figure out if the language the song is made in factors in determining its popularity. I would have to use more recent and proper data to get to proper conclusions about more things regarding this topic. I could also use data from other streaming services and look for any differences or parallels between several data points. Another area for further research would be to see how different generations of people might influence the popularity of different artists or genres of music.