

CS 181, Spring 2021

Group 9: Grace Le & Tristan Luu

Date: May 18th, 2021

FINAL PROJECT

Introduction

In this final project, our group is interested in investigating the entertainment field, especially the music industry. Therefore, we have decided to narrow our focus down to two popular music platforms: Spotify and Billboard. Using the data acquired, we start to concentrate on the three questions below in order to gain insights into our chosen topic:

1. Based on the data from our sample of Spotify's users, artists from which genres tend to have the highest followers and popularity?
2. Is there a relationship between our artists' number of followers and popularity rate?
3. Does a song's rank on the Billboard's "Hot 100" have any influence on its sales' rank on the "Digital Song Sales" chart?

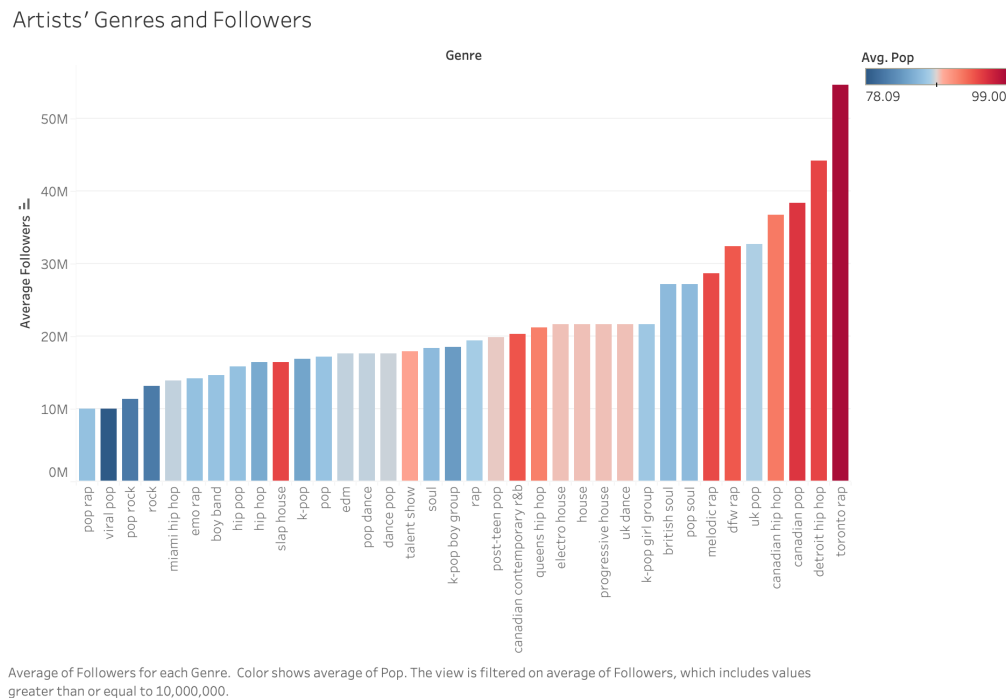
Background

Firstly, we use Spotify as our Web API provider in order to build requests and get data from seven users, who are all Vietnamese high school and college students. The big chunk of data includes their public playlists, top tracks, followed artists, streaming devices, and more. In this project, we focus on interpreting the data about their following artists and those artists' related artists. Each artist has information corresponding to five variables: ID, name, followers, popularity, and genre. The popularity rate is an invented one by Spotify to keep track of the artist's popularity. In addition, for the second data source, we utilize a special Python package to acquire data from a second source: the Billboard's website. We mainly get some popular chart's data such as the "Hot 100", "Billboard 200" and "Digital Song Sales", which serves as complements to our data from Spotify's users.

Visualization and Interpretation

Question 1: Artists' Followers by Genres and Popularity

In order to tackle this question about artist's followers by genres, we start off by asking for consent and getting the data from several Vietnamese Spotify's users. Then, we manage to select more than 500 related artists based on their song streaming data and build two tidy data frames: one with artist's ID, name, followers, and popularity, and another with artists' ID and their genres.



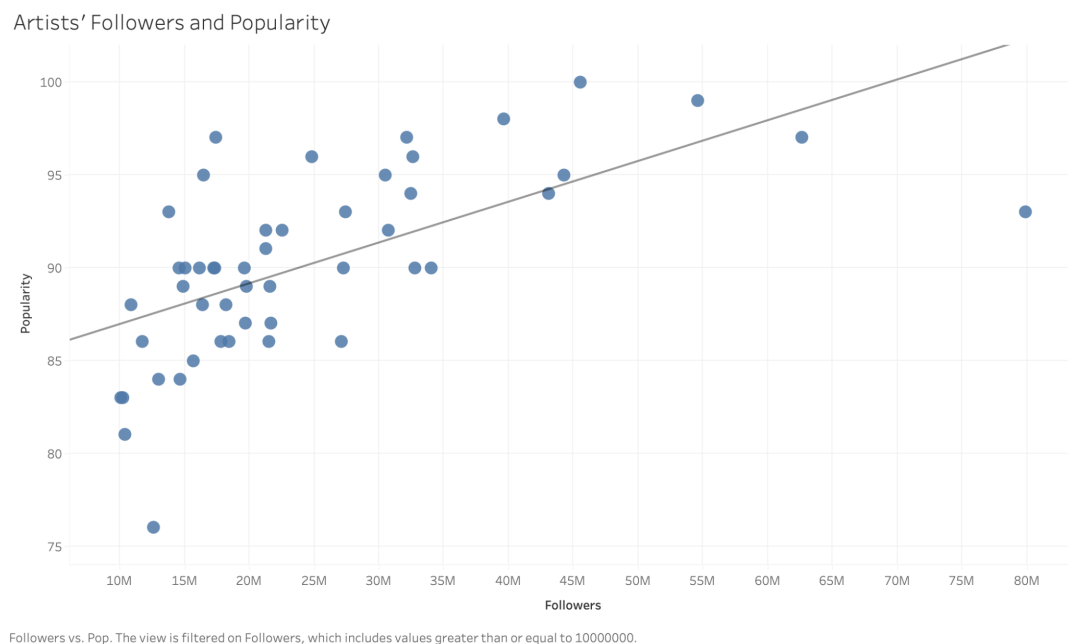
To convey our observations, we utilize Tableau Desktop to construct a bar chart using artists' genres on the x-axis and artists' average followers on the y-axis, colored by artists' average popularity. The followers variable is filtered to contain only values larger than 10,000,000. The height of each bar represents the average number of followers according to artists' genres. The graph is organized in an increasing order of height from left to right, which illustrates artists' genres with lower number of followers on the left hand side than those on the right. In this graph, it is apparent that artists from genres such as Canadian pop, Canadian hip hop and Toronto rap have the highest number of followers, of 35 million and above, while rock artists tend to not have as many followers, with only around 10 million.

This result implies that Canadian artists of multiple genres are very popular globally, meaning that Canada must have a very developed music industry and many talented artists.

Additionally, artists' popularity is also represented on the chart by the color of the bars. Red indicates a very well-known artist, whereas blue means an artist's popularity rating is not very high. The graph shows that most of the higher bars on the right are in red, while the majority of the lower ones on the left are blue. Since a higher number of followers is a factor associated with an artists' popularity, the result turns out as expected that artists' genres with a higher number of followers also have higher popularity ratings.

Question 2:

How significant is the relationship between the artists' number of followers and popularity rate?

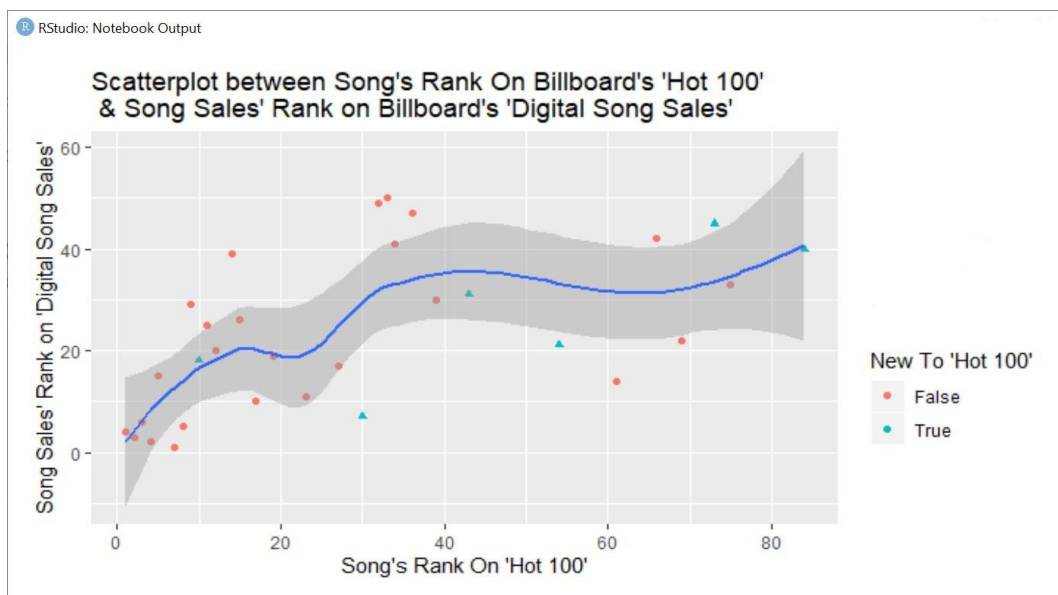


From the data exploration step, we recognize that the distributions of the two variables “followers” and “popularity” are quite similar, suggesting that there might be a correlation between the two. Therefore, in this part, we want to determine if there is any significant relationship between the artists' number of followers and popularity rate.

From the tidy data set containing information about artists' ID, name, followers, and popularity, we construct a scatter plot that plots the artists' number of followers on the x-axis and the popularity rate on the y-axis. The data is filtered to only contain artists with more than 10,000,000 followers to eliminate insignificant observations from the test. As we can see, for artists having between 10,000,000 followers and 70,000,000 followers, there is a relatively strong positive correlation between their number of followers and popularity. The more followers they have, the more popular they are. The slope of the best fitted line is not 1 but close to 1. This indicates that the number of followers influences the popularity score positively, but it does not have the entire power on the artists' popularity. Spotify maintains other factors that contribute to the popularity rate as a whole to keep it objective. Generally, artists with more followers stand a higher chance of having a high popularity score.

Question 3: Effect of Song's Rank on Song Sales' Rank

Lastly, using data from the Billboard website, we decide to investigate factors that may have an effect on song sales' rank on the "Digital Song Sales" chart. After getting data from the Billboard's "Hot 100" and "Digital Song Sales" charts, we begin to put them into tables and clean as part of our data wrangling process. In order to prepare for the visualization, we take a closer look at our two chosen variables: song's rank from the "Hot 100" table and song sales' rank from the "Digital Song Sales" table. With the aid of RStudio, we successfully produce our visualization to show the correlation between the song's rank on the "Hot 100" and their sales' rank on the "Digital Song Sales", colored by whether the song is new to the "Hot 100" chart or not.



As we initially intend to investigate the correlation between the two variables, our method is to use a scatter plot with a curved line of best fit to graph the result. Our x-axis is the song's rank on the "Hot 100", and the song sales' rank on the "Digital Song Sales" is on the y-axis. According to the figure, the dots form an upward trend overall: they concentrate on the lower left hand corner of the graph and then move upward from left to right. This indicates that there might be a correlation between our two chosen variables: the higher a song's rank in the "Hot 100", the higher its sales' rank. In addition, the color of the dots illustrates whether a song is new to the "Hot 100" chart or not. It is apparent that the majority of them are red, which implies that most of the songs on the graph have been on the "Hot 100" for a while. This result shows that newly released songs that have just been featured on the "Hot 100" tend not to rank high on "Digital Song Sales" chart. This is not a big surprise because it is quite self-explanatory. It is true that most of the time, a newly released song may not be popular enough to the audience to achieve a significant number of digital sales. It usually takes some time for a song to gain its popularity and have a lot of people buy its digital version.

Moreover, we also graph a curved best fit line for a better fitting visualization than a straight one. From the graph, the line generally goes upwards with a positive slope, which reaffirms our claim above that a song's rank on the "Hot 100" has a positive impact on its sales' rank on the "Digital Song Sales" chart. However, the dark shadow representing the fluctuation of the correlation is not very small, meaning that this line may not fit very closely with our data. One reason for this could be that our sample is not big, so it is not very likely that our result is accurately representative of all songs.

Conclusion

After investigating more closely, we have come to the conclusion that artists of genres such as Canadian pop, Canadian hip hop, and Toronto rap generally tend to have more followers, which leads to a higher popularity rate, than artists of different genres. In addition, there are also chances that a correlation exists between a song's rank on the "Hot 100" and its sales' rank on the "Digital Song Sales" chart. However, our results above may only be true for our sample of Spotify's users and might not be very accurate if generated to a larger population because our sample size is quite small.

References

“Authorization Guide.” *Spotify for Developers*,
developer.spotify.com/documentation/general/guides/authorization-guide/.

“Charts.” *Billboard*, www.billboard.com/charts.