
SPOTIFY

EXPLORATORY DATA ANALYSIS

JWALA MUKHI SURESH



ABOUT THE DATA



Data Source :

Spotify daily top 200 songs with genres 2017-2021 from Kaggle

Data Background :

The dataset provides information on the daily top 200 tracks listened to by users of the Spotify digital platform around the world for a period of 4 years and 6 months from 2017-2021. The data set includes Position, Track Name, Artist, Streaming, Date and Genre(s).

Total number of records : 321046 (Prior to Quality Check)

Link : www.kaggle.com

DATA QUALITY ASSESSMENT & TRANSFORMATIONS

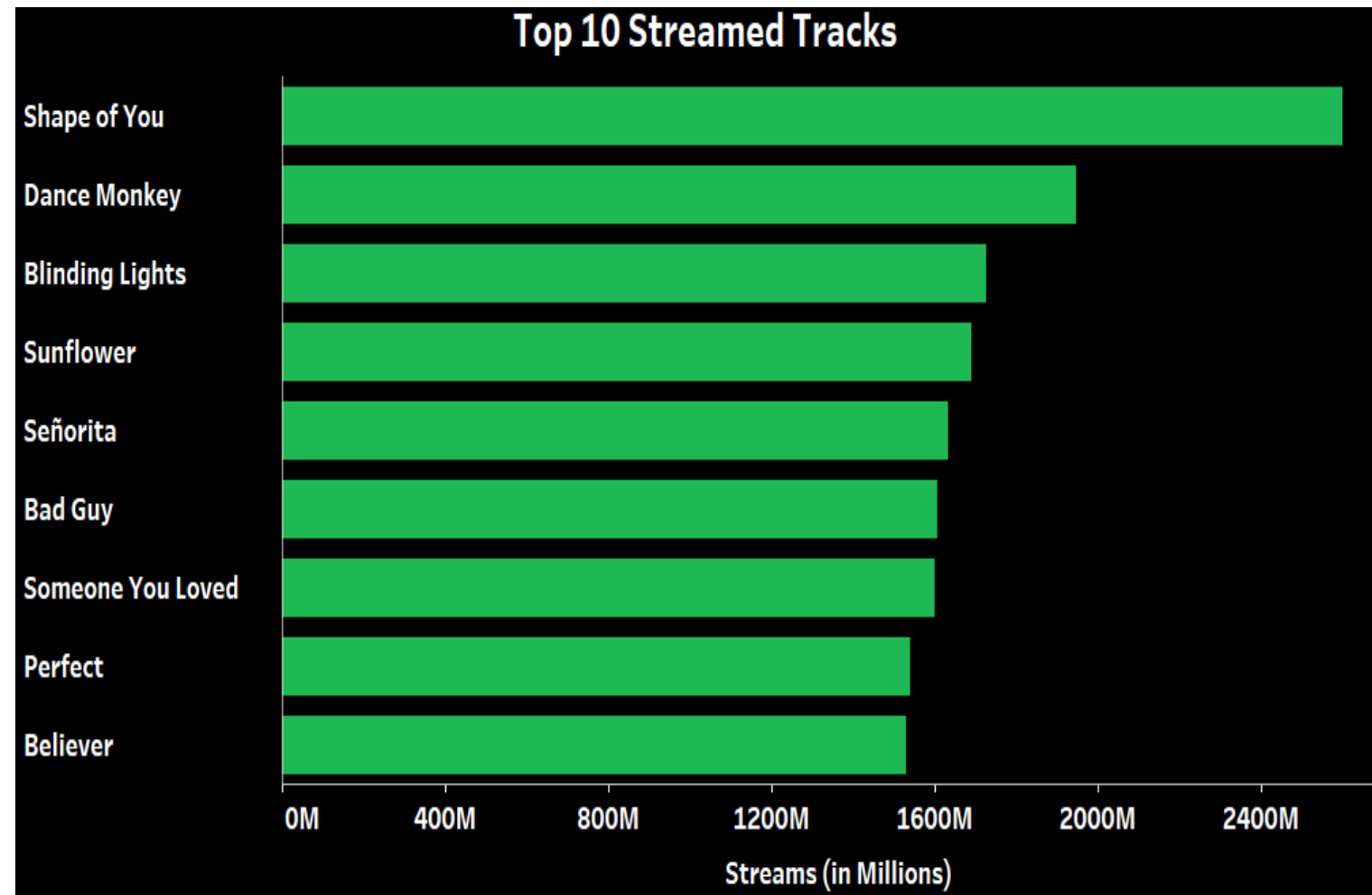
- The dataset had incomplete values for the year of 2021. So, I removed the data for the period of 2021 and did the analysis for a period of 4 years from 2017 – 2020.
Total number of records after removing 2021 data: 290048
- Outliers : There were no outliers in the dataset.
- The dataset had all values concatenated into a single column using '#'. I used Tableau prep to convert it into separate columns.
- Genre column was a multivalued column with close to 650 unique genre values. I researched and identified the main genres and finally classified them into 17 main genres. Then I created a column for each of the main genre with binary values(0 & 1) as each song could belong to multiple genres.

QUESTIONS FOR EDA

In order to explore the Spotify dataset, I chose the below questions to perform EDA.

1. What are the Top 10 streamed tracks?
2. Which tracks entered the Top 1 position during each year?
3. Which artists had the greatest number of tracks? (over each year and all years combined)
4. Top 10 artists who had the most streams and their estimated revenue from Spotify
5. What are the Top 10 streamed genres?
6. How did the streaming of the Top 5 genres change over the years?
7. Which artists had the most tracks in each of the Top 10 genres?
8. How did the overall streaming in Spotify change each year?
9. How did the position change for the songs that were Top 1 for the most time in each year?

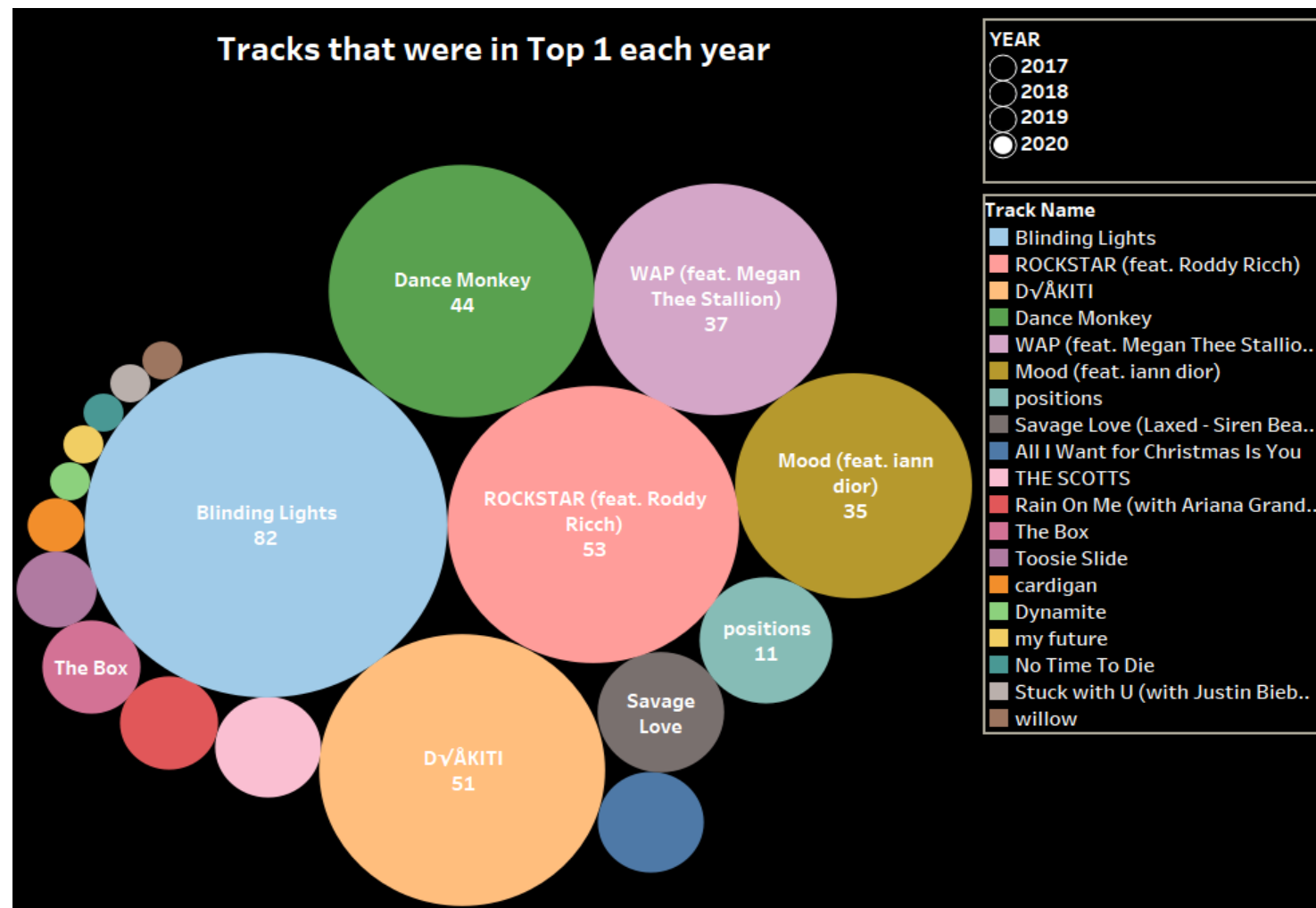
TOP 10 STREAMED TRACKS



The dataset contains the top 200 songs over the period of 4 years. To identify the top 10 most streamed tracks I have plotted the track names against their streaming. I have added a tool tip to show the Artist and Streams.

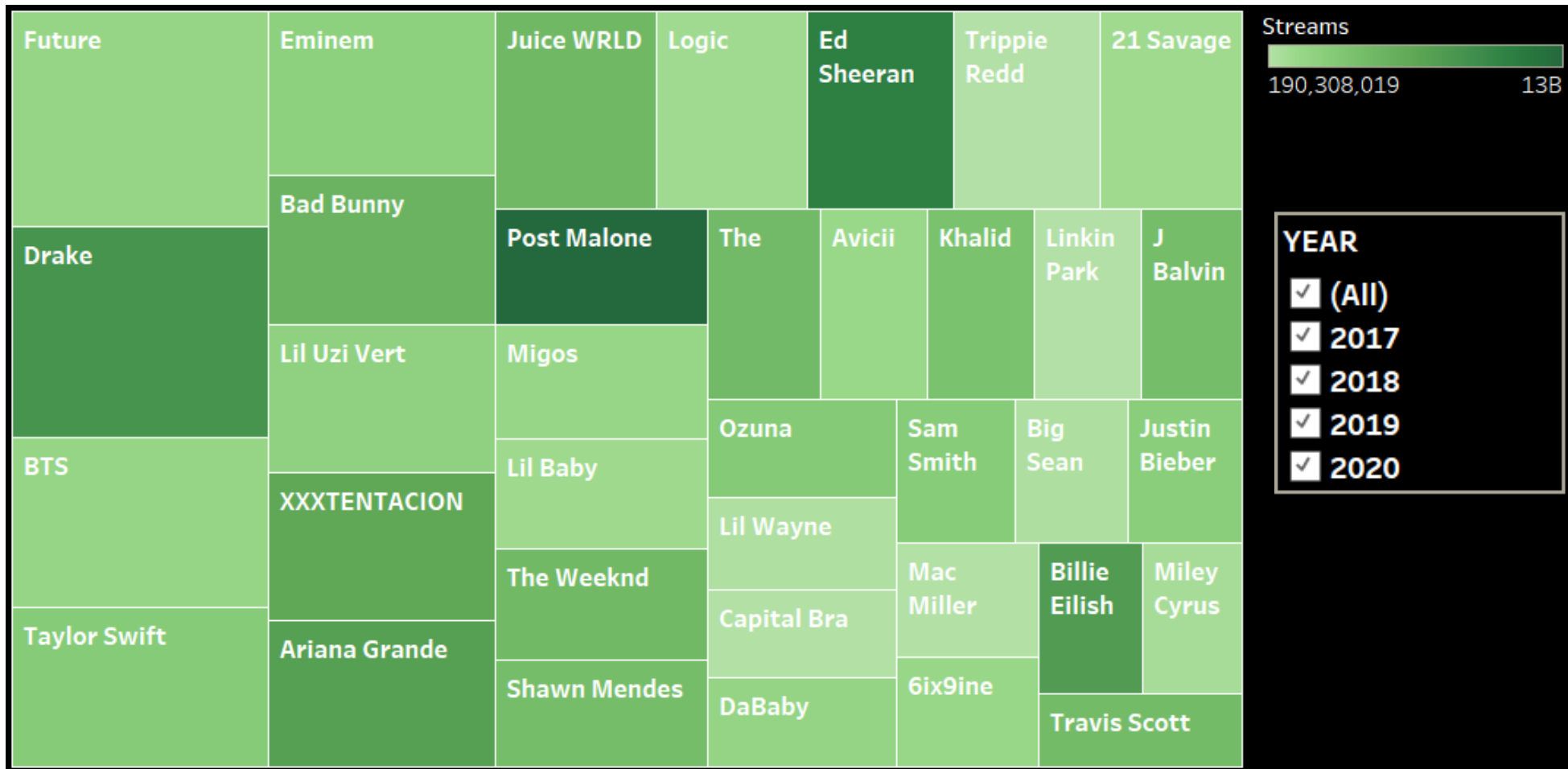
The bar graph shows that “Shape of You” was the most streamed song with a streaming close to 2.6 Billion. The second most popular song was “Dance Monkey” with a streaming close to 1.9 Billion but interestingly 0.6 Billion less than the first. The top 10 chart includes two songs by “Ed Sheeran” – Shape of You & Perfect.

TRACKS THAT ENTERED TOP 1 EACH YEAR



To understand the popularity of tracks, I created a bubble chart showing the tracks that grabbed the top 1 position during each year. I have used a single value filter to choose the year. The legend shows the list of tracks with a color assigned to each track. The size of the bubble represents the number of times a track entered the top 1 spot. I have also used label for the same. As these details are not visible for smaller bubbles, I have included a tool tip for easy interpretation.

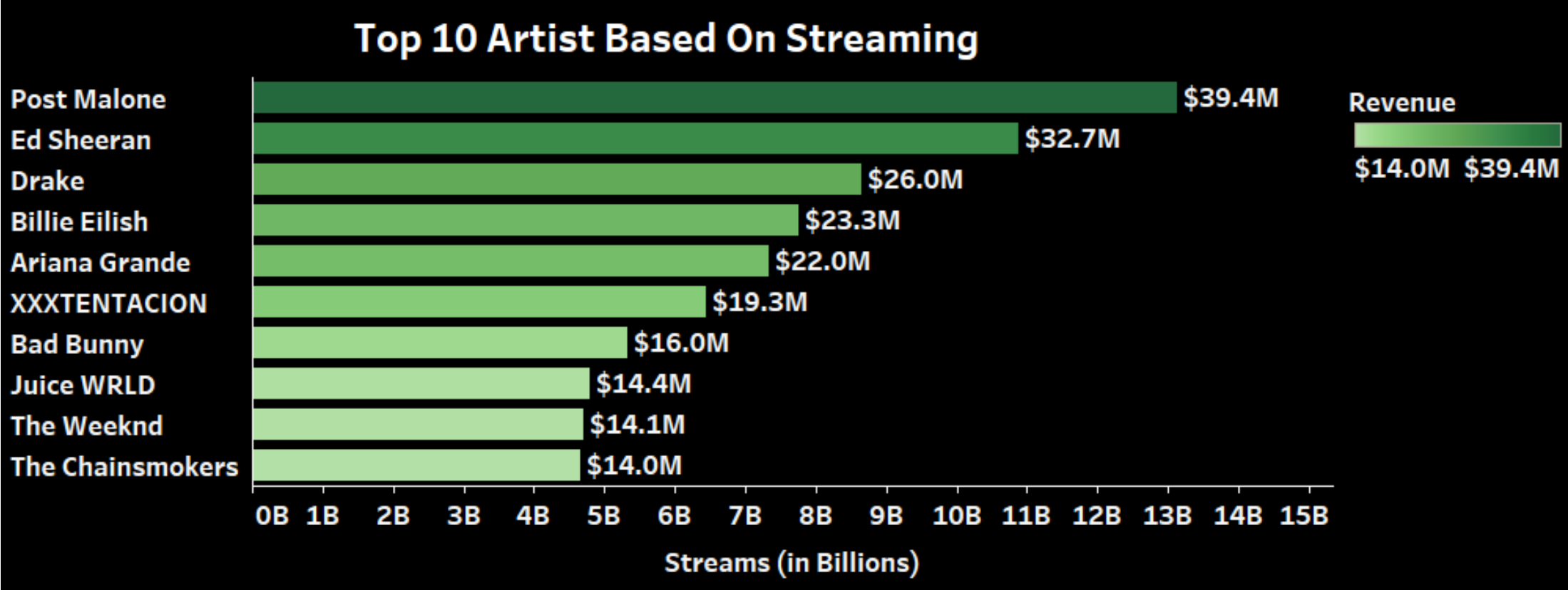
The image added here just represents the data for 2020. From the visualization we can see that “Blinding Lights” grabbed the Top 1 position 82 times in 2020. The song “Willow” was able to grab the top 1 position only once. By using the filter similar insights can be made about the song popularity in other three years.



To analyze which artist had the greatest number of tracks in the top 200 Spotify charts I created this visualization with a treemap. The visualization uses hue to show the streaming trend and the size shows the number of songs. I have added a multivalued filter, so that the user can view a particular year or for all the four years. I have added a filter to include only artists that had a minimum of 30 songs.

The visualization shows the artists with the greatest number of tracks and streams for all 4 years combined. It can be seen that “Future” and “Drake” had the greatest number of songs but songs by “Drake” was more streamed than “songs by “Future”. Even though “Post Malone” had a smaller number of songs, his songs were the most streamed compared to all other artist. Similarly, for “Ed Sheeran” and “Billie Eilish” the artists were able to gain popularity just by making a smaller number of songs which is interesting as this shows the numbers don’t matter but quality of songs is most important.

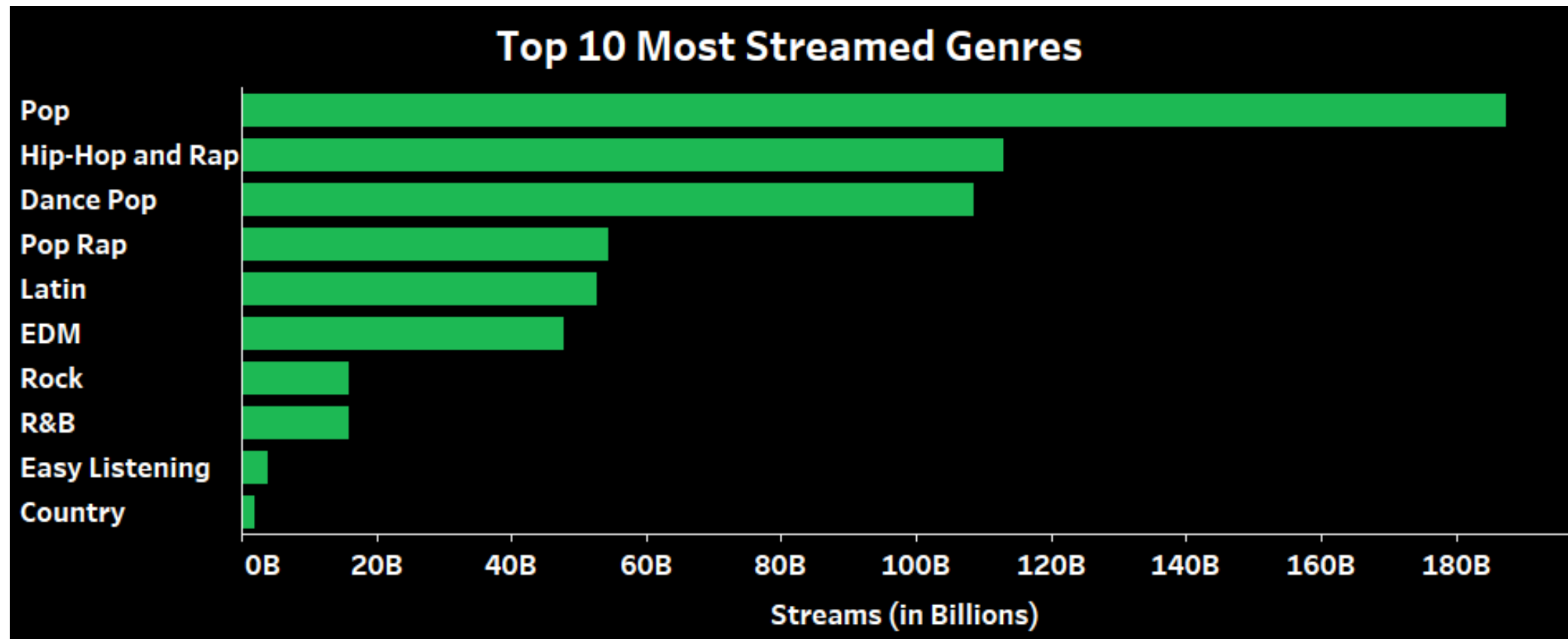
TOP 10 STREAMED ARTISTS AND THEIR REVENUE



This analysis shows the top 10 most streamed artists for the period of 4 years. I was curious to understand how the artists benefited from their songs being on Spotify. I found the interesting fact that each artist gets paid a \$0.003 per streaming. To show this in my visualization I have plotted Artist against their total streaming. The revenue earned is shown with label and hue.

The visualization shows that “Post Malone” was the most popular artist with just 41 songs (from previous treemap) , he made a streaming of over 13 Billion and he earned a total of \$39.4 Million from Spotify. Whereas Drake who had a total of 90 songs (from previous treemap) was able to make only \$26.0M from the streaming which is less than what Post Malone & Ed Sheeran made from Spotify with fewer songs.

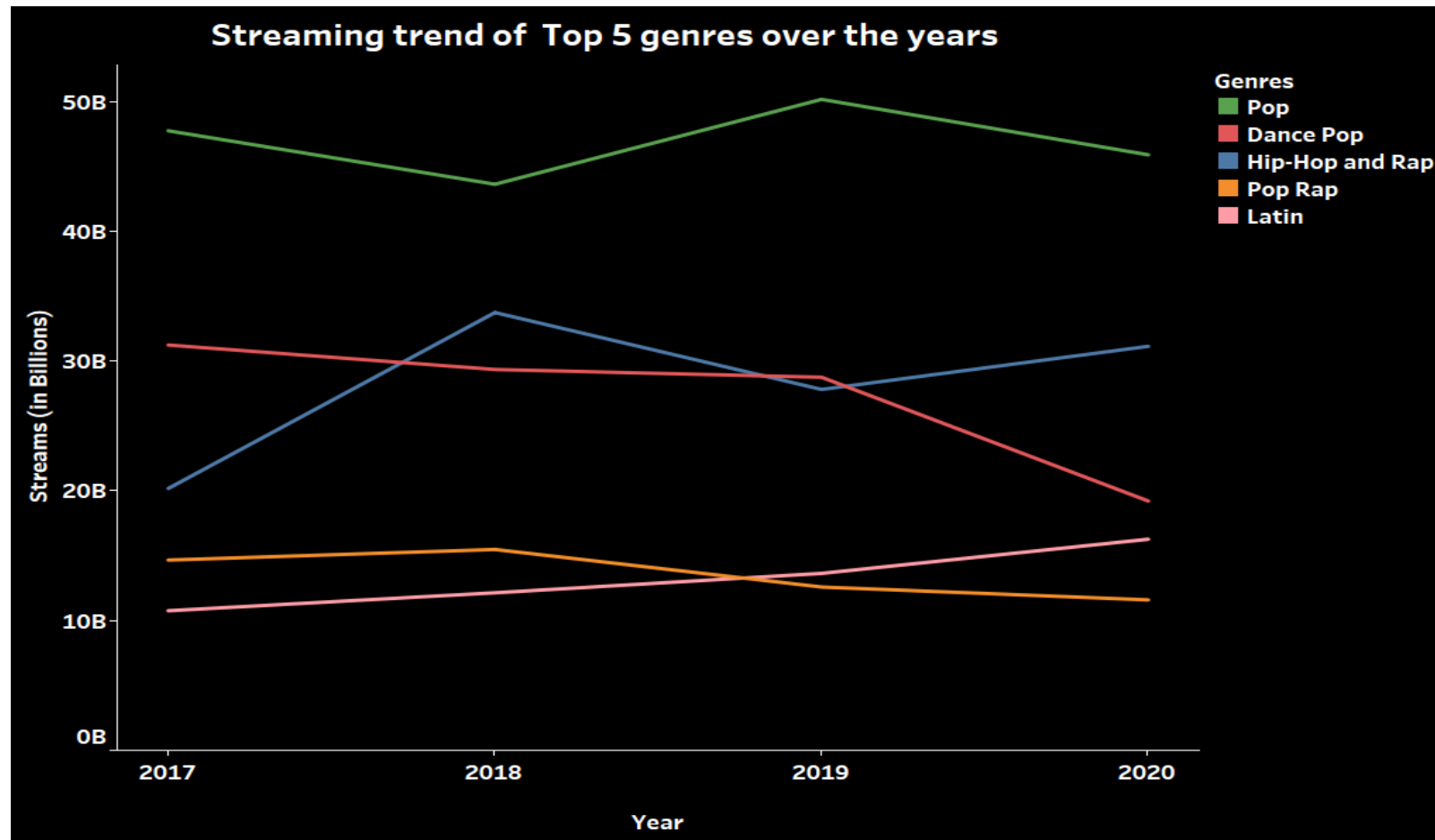
TOP 10 STREAMED GENRES



As I have classified 650 genres into 17 main genres, I was interested in finding out the top 10 most streamed main genres. To show this in my visualization I created a calculated field for each genre to find the total streaming of each genre. I used a bar chart for the visualization with the genres plotted against the streams.

The visualization shows that over a period of 4 years, “Pop” is the most popular genre with a streaming of over 180B. From the chart we can clearly notice that even though the songs in other genres did enter the top 200 Spotify ranking they weren’t able to gain as much popularity as the Pop songs.

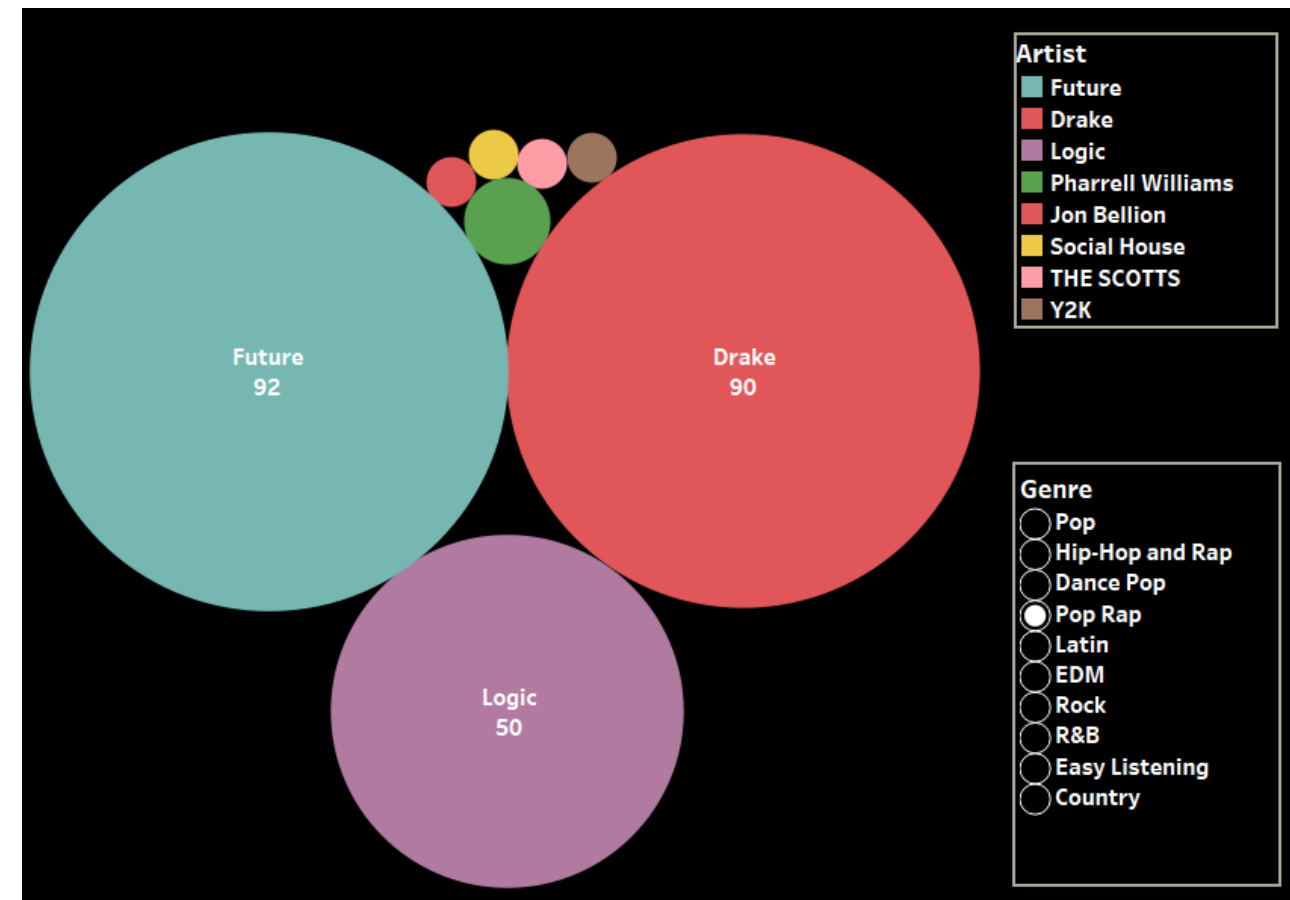
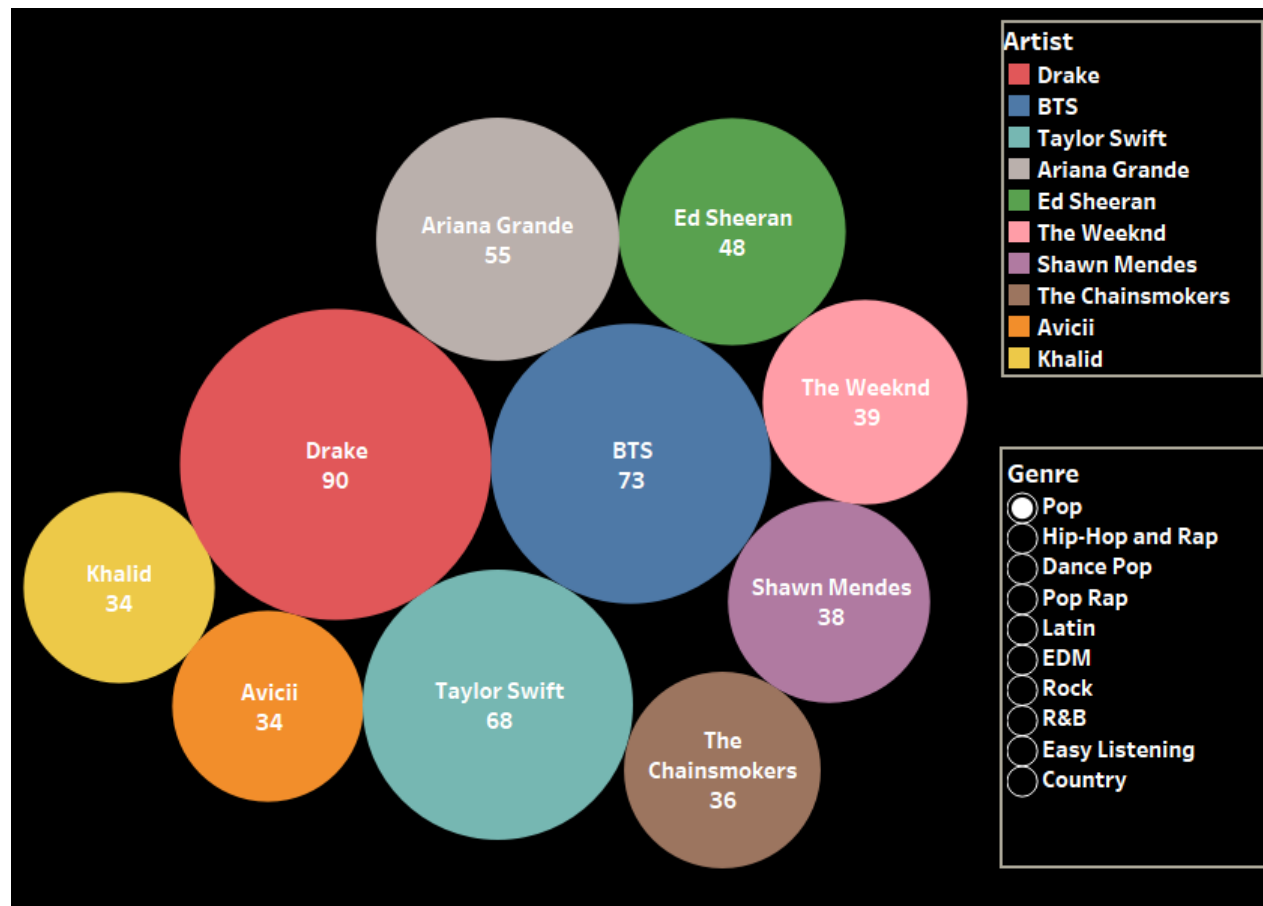
STREAMING TREND OF TOP 5 GENRES



As I was able to understand the most streamed genres, I was curious to see how the trend changed for the top 5 genres over the 4 years. Was the streaming pattern same for all years or was it because any immediate change in trend during a particular year. To understand the trend, I created this visualization showing the streaming pattern of the top 5 genres over the 4 years. I used a different color to represent the 5 genres to distinguish them easily.

The visualization shows that “Pop” has clearly been listeners favorite in all 4 years. But the trend line shows that there is a growth and dip in number of listeners every other year. “Dance Pop” had a streaming of above 30B in 2017 but had a huge drop in popularity by 2020 to a 20B streaming. “Hip-Hop and Rap” started with a 20B streaming and crossed the streaming rate of “Dance pop” in 2018 and 2020 and grabbed the 2nd position. “Pop Rap” and “Latin” had a lower streaming compared to other 3 but “Latin” gained more popularity than “Pop Rap” in 2019 and 2020.

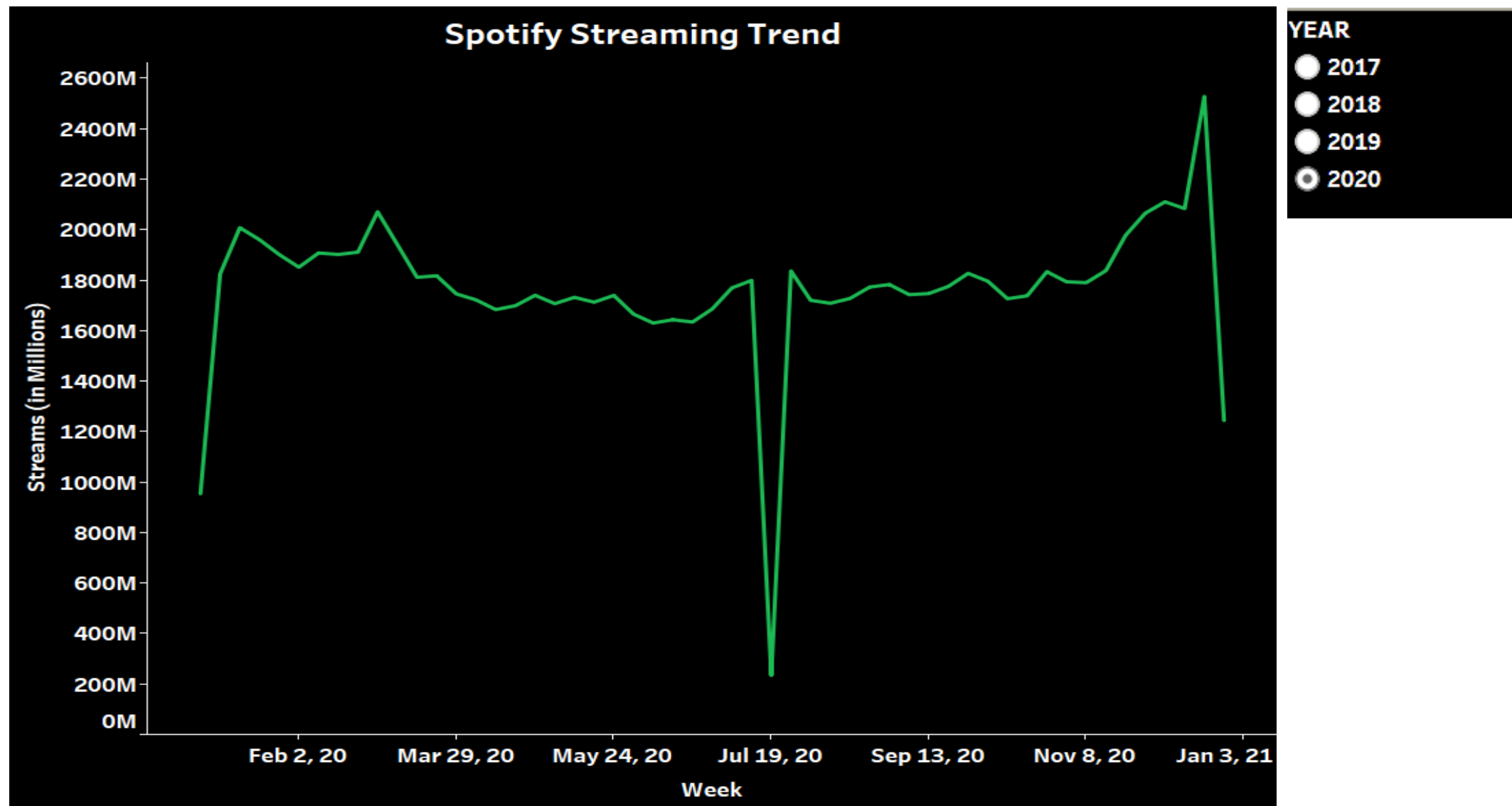
ARTISTS WITH MOST TRACKS IN THE TOP 10 GENRES



To analyze which artists had the greatest number of tracks in each of the top 10 genres, I created a bubble chart with a filter to choose the genre. The size of the bubble indicates the count of distinct songs each artist had in a particular genre. I have included a label to easily identify the artist and the number of tracks. The tool tip also provides these details for easy interpretation of smaller bubbles.

The visualization shows for the genre “Pop”, Drake had the greatest number of tracks. In “Pop Rap” 3 artists had the greatest number of songs. It can be understood from the chart that it was the songs by these 3 artists that made the genre “Pop Rap” the 4th most streamed genre.

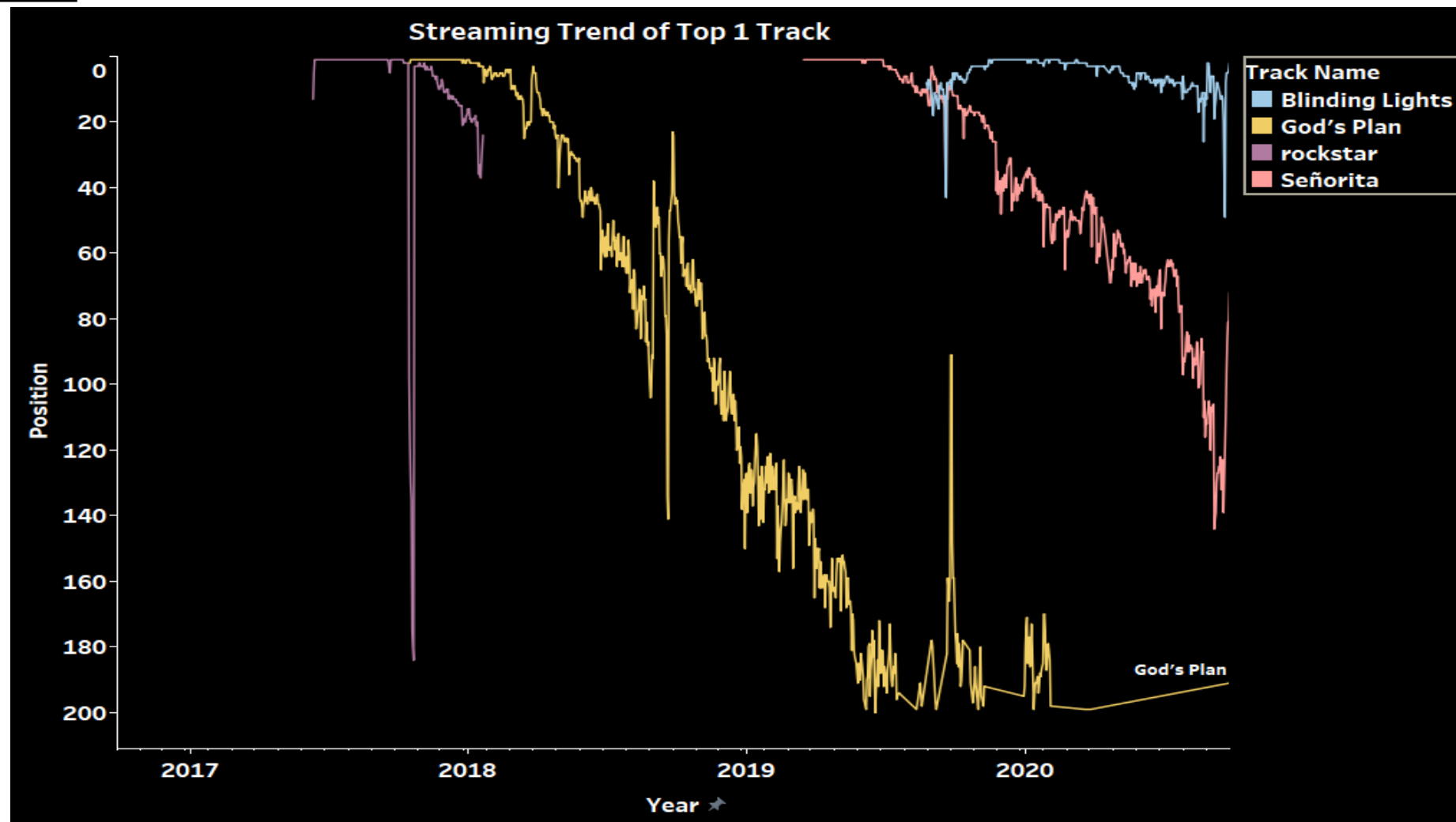
SPOTIFY STREAMING TREND



To understand the popularity of Spotify during each of the 4 years, I believe a good way to look at that would be to plot the overall streaming of the top 200 songs in this duration. I have plotted the streaming against weeks in each year. The single valued filter is provided to choose the year.

The visualization shows that for the year 2020 there was a huge drop in the number of listeners for a week in the month of July, but later streaming increased to 1800M. By the end of 2020 we can see that the streaming increased further to almost 2500M which is the most streaming Spotify had for the year of 2020.

STREAMING TREND OF TOP 1 TRACK



To analyze how the trend of the Top 1 streamed track in each year changed, I have plotted the position of the tracks against year. The y-axis has been reversed as the longer a song stayed at a lower position means the more popular the song was.

The visualization shows that “God’s Plan” entered the Top charts in 2018 and continued to stay in the Top 200 charts even in 2020. The song “Rockstar” that entered the top charts in 2017, and towards the end of 2017 it dropped to position 184 and then suddenly gained popularity by going back to 2nd position. “Blinding Lights” that entered the Top 1 in November 2019 was able to stay within the Top 50 position till the end of 2020.

KEY TAKEAWAYS & FUTURE SCOPE

- Over the course of 4 years from 2017-2020, Spotify had 4922 unique songs in their Daily Top 200 chart with an overall streaming of ~324 Billion.
- There were 1019 unique artists whose songs made into the Daily Top 200 chart. The most streamed artists were Post Malone, Ed Sheeran and Drake. Post Malone and Ed Sheeran had a fewer songs compared to other top artists, but their songs were more streamed and made the highest revenue from Spotify.
- Pop, Hip-Hop and Rap & Dance Pop were the most popular genres based on streaming. Pop had 2312 songs with a streaming of over 187 Billion. At the beginning of 2017 Dance Pop was more popular than Hip-Hop and Rap but there was a sudden increase in the streaming of Hip-Hop and Rap making the genre 2nd most popular.
- “God’s plan” that was released in 2018 made grabbed the Top 1 spot in the same year and continued to stay in the Top 200 chart even by the end of 2020.

In future it would be interesting to add locations to this dataset to understand the trends all over the globe. If a particular user’s information was available, then it would help to see the person’s streaming pattern and how Spotify makes suggestions to users.



TABLEAU PUBLIC

Link to Packaged Workbook : [Spotify Exploratory Data Analysis](#)

Link to Dashboard : [Spotify Trends \(2017-2020\)](#)