

Preliminary Data Exploration

To get a general idea of the layout of the data, the following general questions were asked.

- 1) Does featuring another artist on a track correlate to higher stream numbers?
Yes, there is an observed network effect in stream numbers. However, this is purely antidotal. Statistical tests are required to prove the significance of the difference in means.
- 2) What kind of growth in stream numbers has Spotify achieved over the past two years?
Between the two year period of January 2017 and December 2018, the top 200 Spotify chart daily stream numbers grew at an average of ~8700 streams. In other words, the average daily streams in a month grew by ~8700 streams.
- 3) What does the distribution of stream numbers look like for the top 200 stream list?
The distribution of stream numbers among the top 200 artist list appears to be exponential. Furthermore, we observe that the log distribution of artist stream numbers appears to be normally distributed.
- 4) When do Spotify users tend to stream the most? (day of week, month of year)
Day of Week
Highest streamed day of the week - Friday
Lowest streamed day of week - Sunday

Month of Year
Highest streamed month – December
Lowest streamed month – January
- 5) Who is the most popular artist/song in the last 2 years years?
Most popular Artist – Post Malone with 7.7 billion streams
Most popular Song – Shape of You by Ed Sheeran – with 2.1 billion streams
- 6) What types of genres tend to be on the top 200 charts?
The top 10 most popular genres among the top 200 song list is as follows:

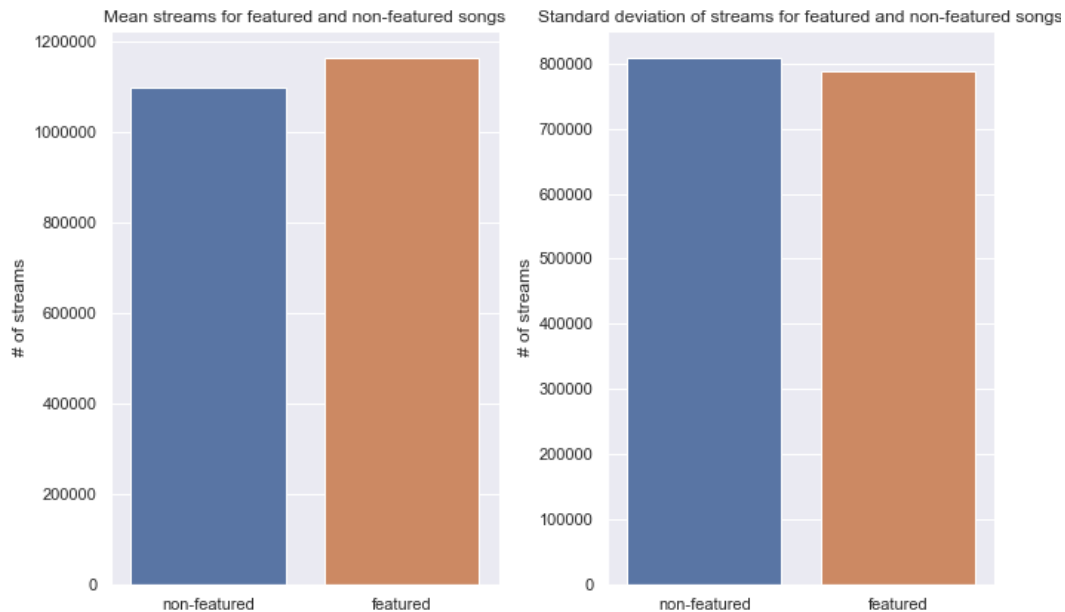
Pop - 268
Dance Pop - 154
Rap - 103
Pop Rap - 101
Post-teen Pop - 92
Trap Music - 76
Tropical house - 71
Hip Hop - 66
Electronic Dance Music - 53
Latin - 50

1) Does featuring another artist on a track correlate to higher stream numbers?

The basis of this question is to answer if there is any observed network effect across artists. In other words, will fans of a certain artist be prone to listen to another artists song if it's featuring the artist they are more familiar with.

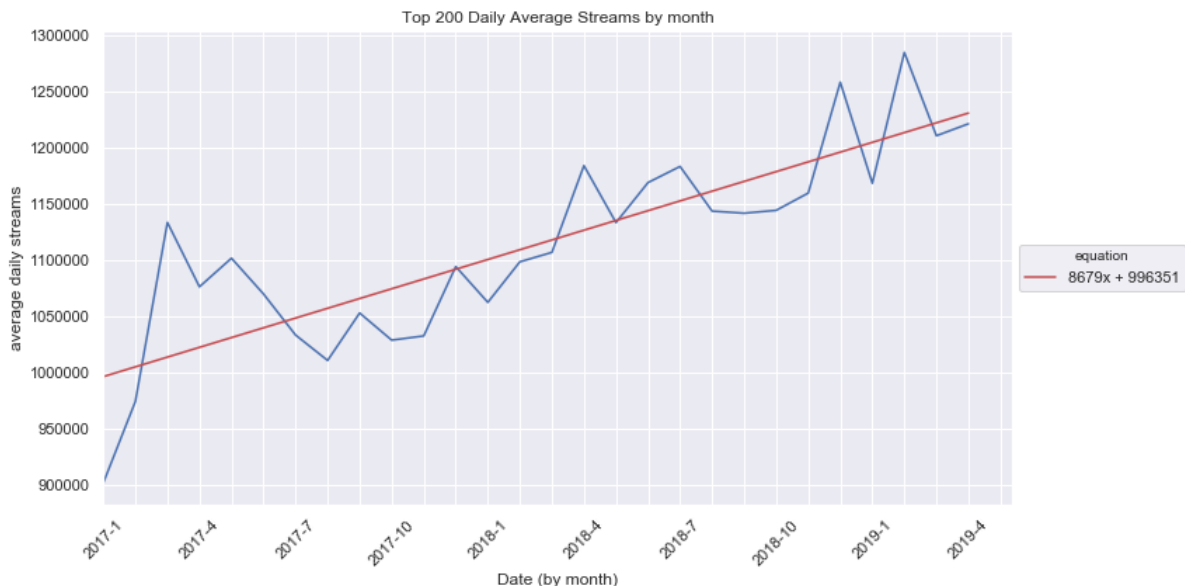
Leveraging regex, the keywords "feat" and "with" were extracted from the song name. A binary feature was then created to help with separating the two groups (1 for track featuring an artist, 0 for track not featuring anyone).

The result of mean streams by featured and non-featured stream numbers shows that featuring an artist **does** lead to higher stream numbers. However, a statistical test will be needed to confirm the validity of this hypothesis.



2 What kind of growth in stream numbers has Spotify achieved over the past two years?

Taking the compiled top 200 stream numbers and resampling by month¹, we see that the average daily stream numbers increase by approximately ~8,700 each month. That is, with each subsequent month, the top 200 daily average stream numbers are ~8,700 streams higher.

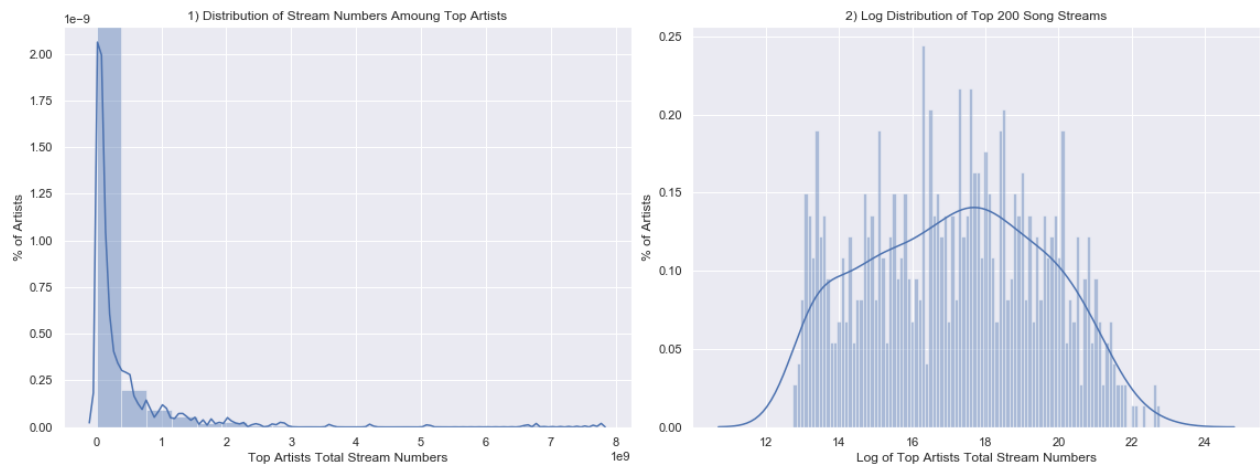


¹ Average calculated by taking the sum of all daily stream activity in a month and dividing by the number of days in the month.

3 What does the distribution of stream numbers look like for the top 200 stream list?

Understanding the distribution of top artists might provide some insight into a model that could potentially predict total Spotify stream numbers.

- 1) Plotting the distribution of artist streams, we observe (not surprisingly) that stream numbers are exponentially distributed. That is, that most artists are not superstars.
- 2) Taking the log of streams, we see a gaussian distribution of stream numbers. The fatter left hand tail coincides with the observations from the plot on the left.

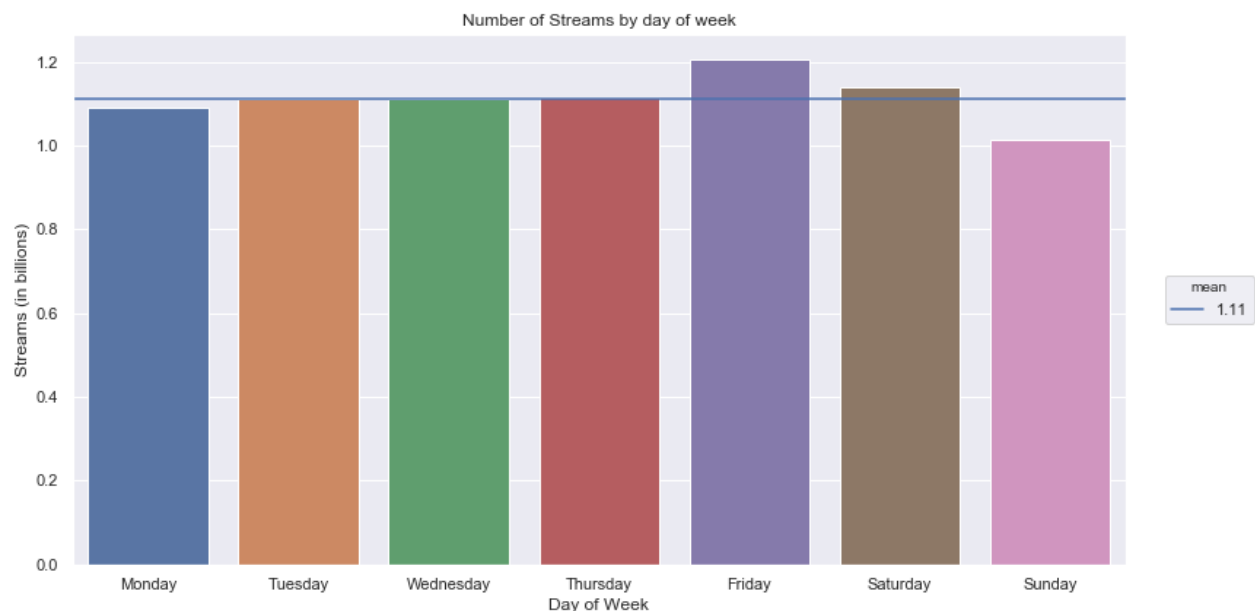


4 When do Spotify users tend to stream the most (day of week, time of year, etc.)?

To test the hypothesis that leisure time correlates to higher stream numbers, date time features were created using pandas and datetime methods.

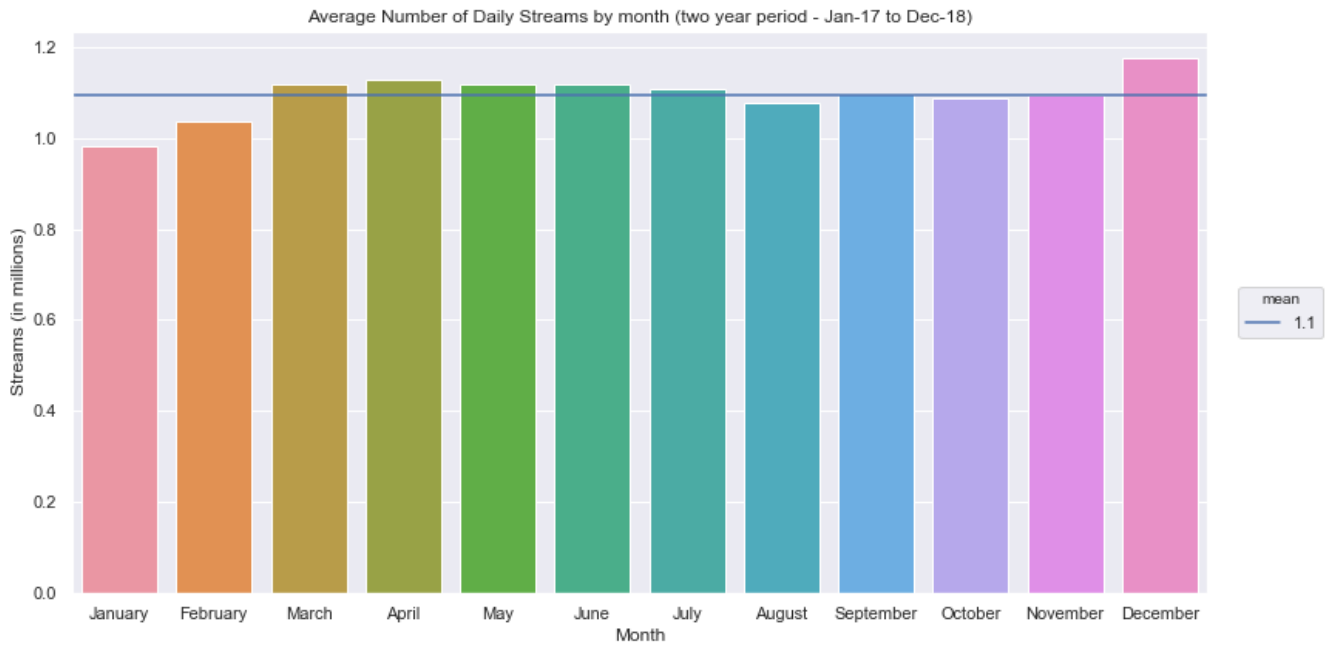
Day of week

The result was the plot below that shows the observed difference in days of the week over the 2.33 year sample. An interesting observation is that despite Sunday being a 'weekend' day, we see significantly lower stream numbers when compared to other days of the week.



Monthly

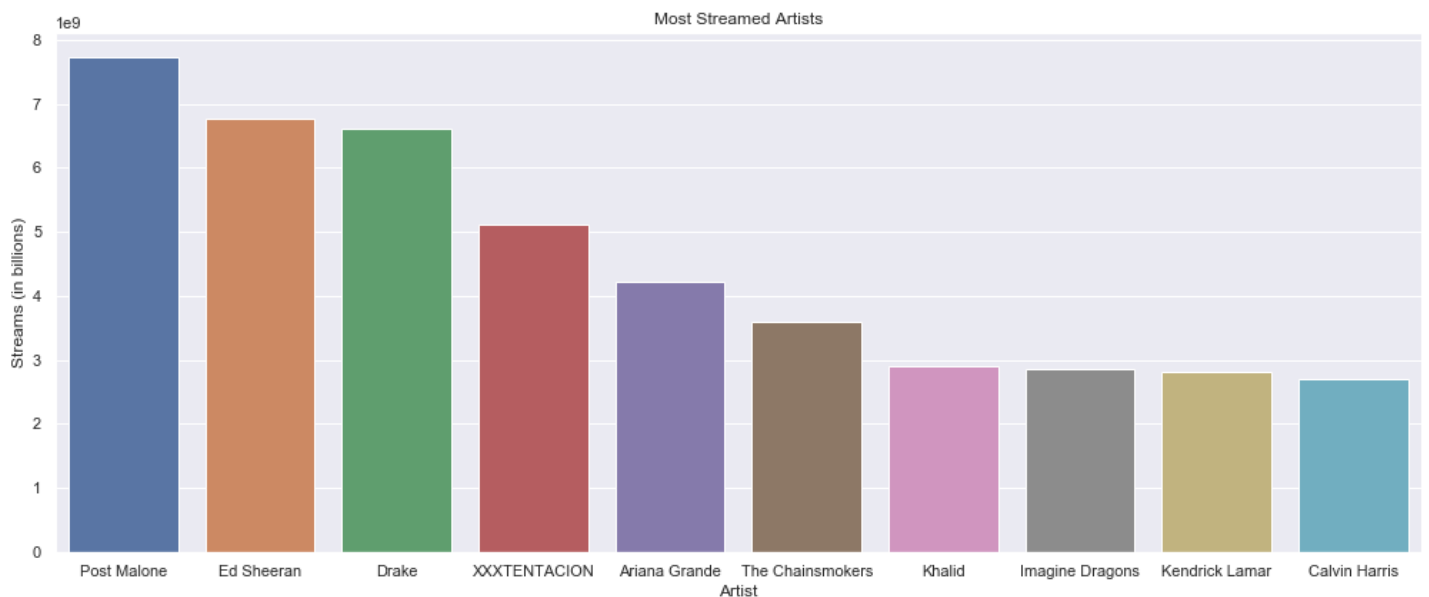
The same methodology was applied to months. The below plot shows the observed difference in daily streams by month. We see that December is by far the highest streamed month with January as the lowest streamed month. January being the lowest streamed month might also refute the hypothesis that streams correlate to user leisure time.



5 Who is the most popular artist/song in the last 2.33 years?

Artist

“Post Malone” is the most streamed artist from the past two years.



Song

Shape of You” by “Ed Sheeran” is the most streamed song from the past two years.

