

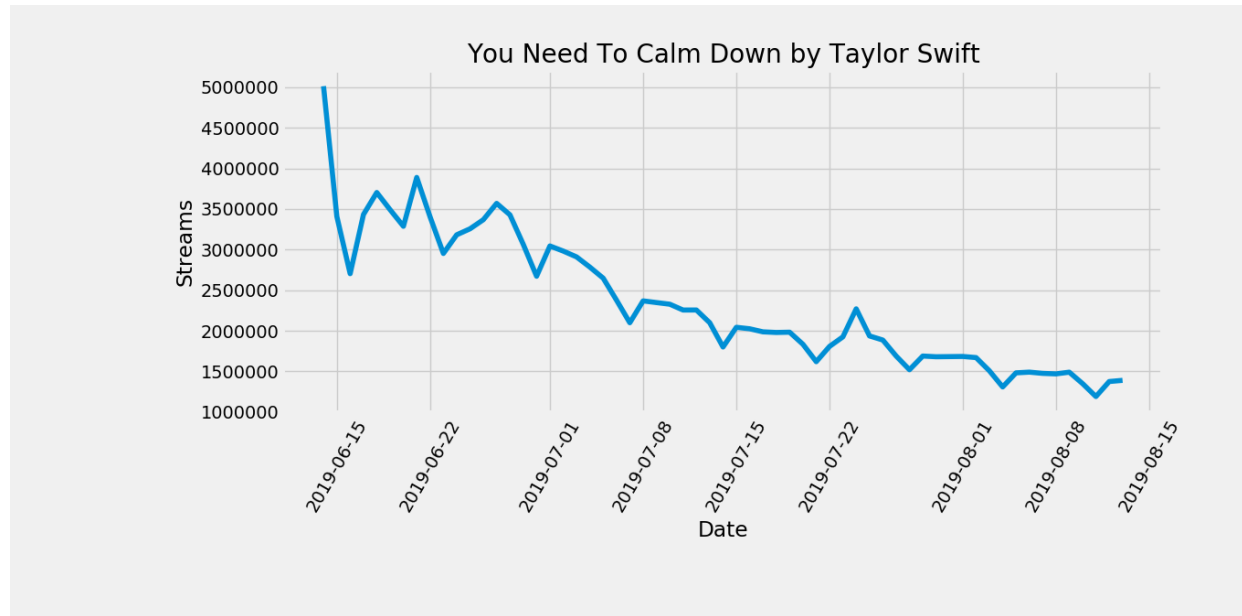
Personal projects

1. Spotify streams count

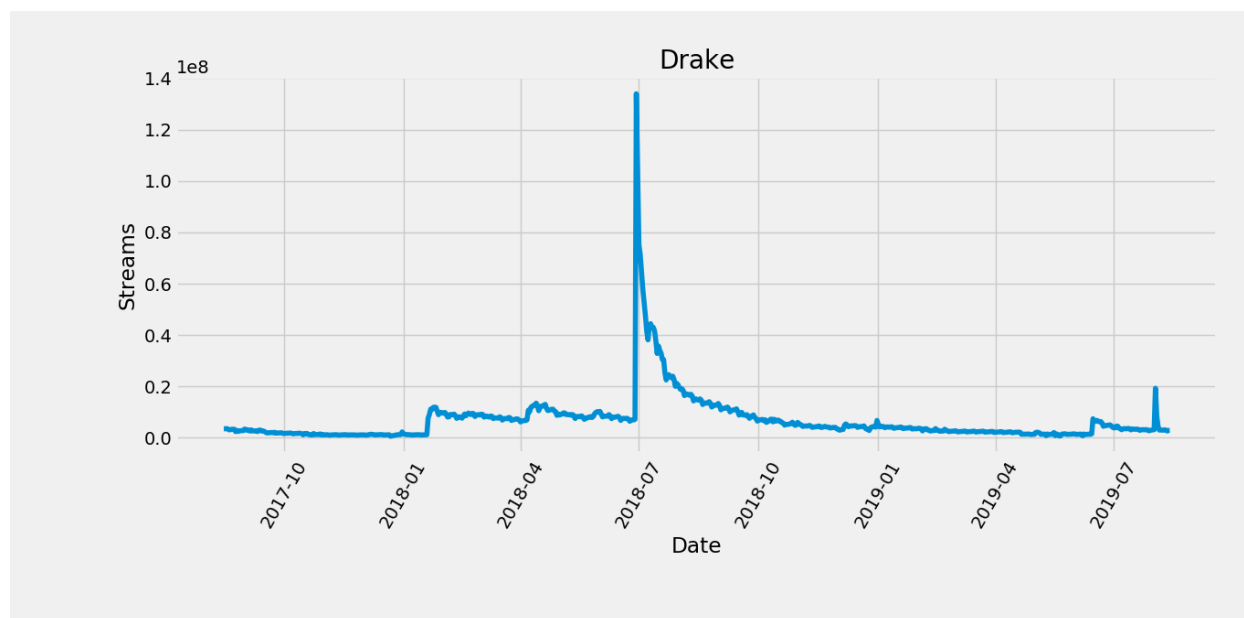
- The intent of the code is to get daily Spotify streaming number for any particular song by any particular artist (with function: `get_data_track(track, artist)`) or by any particular artist (with function: `get_data_artist(artist)`).
- It will pull daily stream number for 2 years (730 days) and plot the streams vs time.
- The number is obtained from this official Spotify chart: <https://spotifycharts.com/regional/global/daily/latest>
- The DataFrame obtained is in form of this format. It keeps appending the daily Top 200 chart and the final result of the DataFrame is shown below.

	Position	Track Name	Artist	Streams	URL	Date
0	1	Señorita	Shawn Mendes	6916437	https://open.spotify.com/track/6v3KW9xbzN5yKlt9YKDYA2	8/15/2019
1	2	China	Anuel AA	4428958	https://open.spotify.com/track/2ksOAxTlxY8yEIEWw8RhGK	8/15/2019
2	3	Ransom	Lil Tecca	3781152	https://open.spotify.com/track/1kBgxAUc4HQv7cgIaylDH	8/15/2019
3	4	Goodbyes (Feat. Young Thug)	Post Malone	3534394	https://open.spotify.com/track/6vBdBcoOhKHlYDDOcorfNo	8/15/2019
4	5	Beautiful People (feat. Khalid)	Ed Sheeran	3415344	https://open.spotify.com/track/70eFcWOvIMObDhURTqT4Fv	8/15/2019
5	6	I Don't Care (with Justin Bieber)	Ed Sheeran	3353944	https://open.spotify.com/track/0hVXuCcRiWRGvwMV1r5Yn9	8/15/2019
6	7	boyfriend (with Social House)	Ariana Grande	3081637	https://open.spotify.com/track/0Ryd8975WihbObpp5cPW1t	8/15/2019
7	8	How Do You Sleep?	Sam Smith	2997104	https://open.spotify.com/track/6b2RcmUt1g9N9mQ3CbjX2Y	8/15/2019
8	9	bad guy	Billie Eilish	2841885	https://open.spotify.com/track/2FxmHks0bxGSBdJ92vM42m	8/15/2019
145993	195	Wonderwall - Remastered	Oasis	436035	https://open.spotify.com/track/5qqabll2vWzo9ApSC317sa	8/16/2017
145994	196	Stressed Out	Twenty One Pilots	430843	https://open.spotify.com/track/3CRDbSiZ4r5MsZ0YwxuEkn	8/16/2017
145995	197	Báilame - Remix	Nacho	429646	https://open.spotify.com/track/3VCLLWqNn80Llv0PJOsKgr	8/16/2017
145996	198	Loca (feat. Cali y El Dandee)	Maité Perroni	428814	https://open.spotify.com/track/05ZXxyeZFhZtA6DAP3Aw7x	8/16/2017
145997	199	Heavy (feat. Kiiara)	Linkin Park	428695	https://open.spotify.com/track/104buTcnP2AsxqB7U1FIZ4	8/16/2017

- `get_data_track('You Need To Calm Down', 'Taylor Swift')` will get you this:



- `get_data_artist('Drake')` will get you this (notice the spike because of his new release in July 2018):



2. Evaluate whether a song appearance in Spotify Global Top 50 tracklist will help boost the performance of the song.

On Spotify, 50 songs with the highest streams daily will be featured in Global Top 50 Chart Playlist. Spotify users may check out new entries in Top 50 and play the song (there's a mark for the song which just climbed to Top 50). Does the appearance of the song in the chart help boost its performance?

To evaluate this, I am going to create 3 different groups of data:

Day 1 is the day before day 2, day 2 is the day before day 3.

- Group 1: Position in day 1 is between 51 and 55, day 2 between 46 and 50 (Global Top 50 exposure on day 2, not day 1)
- Group 2: Position in day 1 is between 56 and 60, day 2 between 51 and 55 (no Global Top 50 exposure on day 1 & 2)
- Group 3: Position in day 1 is between 46 and 50, day 2 between 45 and 41 (Global Top 50 exposure on day 1 & 2)

Then, I am computing the change in position from day 2 to day 3, relative to the change from day 1 to day 2. I'll compare this among the 3 groups to see the Top 50 chart exposure impact.

But beforehand, I need to reformat the DataFrame into the format below for me to easily compute the relative change in position. This is a sample of the 2 year data.

Track	Artist	Date1	Position1	Stream1	Date2	Position2	Stream2	Date3	Position3	Stream3
Señorita	Shawn Mendes	8/13/2019	1	6978889	8/14/2019	1	7042248	8/15/2019	1	6916437
China	Anuel AA	8/13/2019	2	4461167	8/14/2019	2	4537298	8/15/2019	2	4428958
Ransom	Lil Tecca	8/13/2019	3	3782345	8/14/2019	3	3803651	8/15/2019	3	3781152
Goodbyes (Feat. Young Thug)	Post Malone	8/13/2019	4	3557126	8/14/2019	4	3548999	8/15/2019	4	3534394
Beautiful People (feat. Khalid)	Ed Sheeran	8/13/2019	5	3482740	8/14/2019	5	3462900	8/15/2019	5	3415344
I Don't Care (with Justin Bieber)	Ed Sheeran	8/13/2019	6	3440285	8/14/2019	6	3425717	8/15/2019	6	3353944

boyfriend (with Social House)	Ariana Grande	8/13/2019	7	3240749	8/14/2019	7	3137073	8/15/2019	7	3081637
How Do You Sleep?	Sam Smith	8/13/2019	8	3036674	8/14/2019	8	2999446	8/15/2019	8	2997104
Loco Contigo (feat. J. Balvin & Tyga)	DJ Snake	8/13/2019	11	2832140	8/14/2019	9	2885674	8/15/2019	11	2815213

Afterwards, I calculate the relative position change and relative streams change as follows:

- $\text{spot['rel_pos_change']} = -((\text{spot['Position3']} - \text{spot['Position2']}) - (\text{spot['Position2']} - \text{spot['Position1']}))$ (there's a negative sign in front, because for ranking, the lower is better)
- $\text{spot['rel_stream_change']} = (\text{spot['Stream3']} - \text{spot['Stream2']}) - (\text{spot['Stream2']} - \text{spot['Stream1']})$

For analysis, even though I compute both relative position and streams change, I only use the relative position change. There is a pattern in Spotify streaming number, such as Friday being the peak number for streamings and Sunday the lowest (the difference between Friday and Sunday can reach 20%) and over the year, around March is the peak month of streaming. To remove that seasonality, I use relative position change.

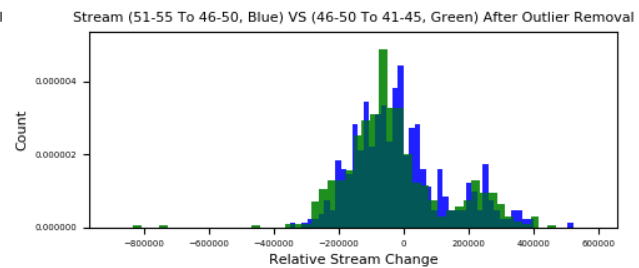
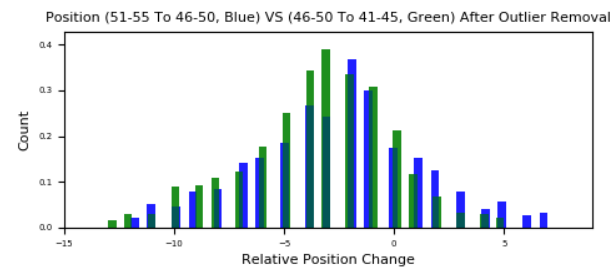
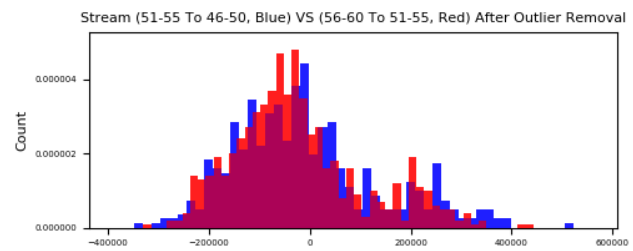
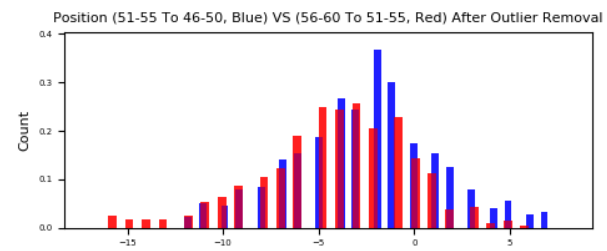
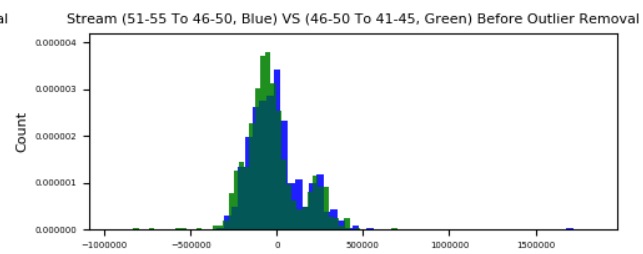
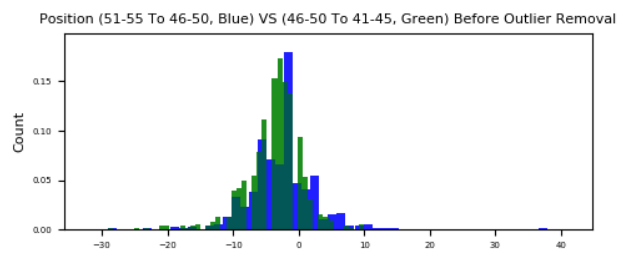
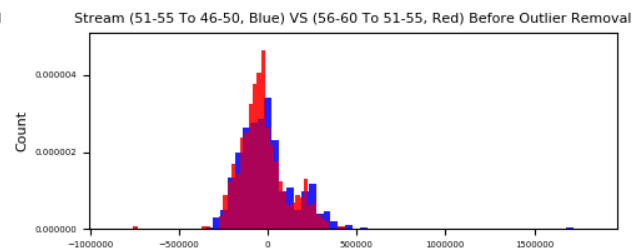
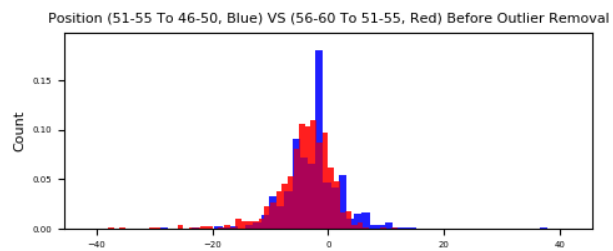
Thereafter, I created 3 groups of relative position change data (group 1, group 2 and group 3) based on the classification earlier. I perform exploratory data analysis to see the distribution of the data and realize that there are a few outliers (for example, Christmas period is over and Christmas songs tumbled down or new music video is released, causing big surge). I use interquartile method to remove these outliers.

The plot below is the distribution of the data. Each group has more than 400 data points.

- Group 1 (Day 1: 51-55 To Day 2: 46-50) in blue
- Group 2 (Day 1: 56-60 To Day 2: 51-55) in red
- Group 3 (Day 1: 46-50 To Day 2: 41-45) in green

I compare blue vs red and blue vs green for both relative position and stream change. From here, it's quite obvious that the average of the relative position change for group 1 is higher than group 2 and group 3. Calculating the mean

difference: group 1 vs group 2 (+1.55) and group 1 vs group 3 (+0.86). For my further analysis, I only use the plots on the left (relative position change).



To validate this finding, I perform bootstrap resampling analysis. This is also to measure how confident we are that the above observation is accurate. The null hypothesis is below.

- Null hypothesis_1: the mean of group 1 and group 2 relative position change are the same.
- Null hypothesis_2: the mean of group 1 and group 3 relative position change are the same.

$p_value_1 = 0$ and $p_value_2 = 5e^{-5}$ (both are close to 0, hence the null hypothesis can be rejected even with 99.50% confidence interval)

Conclusion: Spotify Global Top 50 Chart exposure can boost the track performance. By having the exposure, the average of position change is about 0.86 to 1.55 more (in ranking) compared to position change without the exposure.

One interesting note: the mean of the calculated relative position is less than 0. So most of the time, after the song increases in position, it generally will not maintain the same increase. After thinking about it, this quite makes sense because when I calculate the relative position, I compare it against day 1 -> day 2 performance (which is always positive in my case). Hence, it'll be difficult to gain more than this. Even when a song is gaining traction, the daily position will be hovering up and down (long term wise, the trend should increase).