

## Sources

Spotify API - JSON - <https://developer.spotify.com/documentation/>

Coachella 2019 Data - CSV-

[https://www.reddit.com/r/Coachella/comments/bc6b63/coachella\\_2019\\_dataset\\_data\\_nerds\\_rejoice\\_a/](https://www.reddit.com/r/Coachella/comments/bc6b63/coachella_2019_dataset_data_nerds_rejoice_a/)

Coachella 2018 Data - CSV - Handmade -

<https://www.stereogum.com/1977133/coachella-2018-lineup/news/>

Coachella 2017 Data - CSV - Handmade

<https://pitchfork.com/news/70100-coachella-2017-lineup-radiohead-beyonce-kendrick-lamar/>

## Overview/Design

My project used the spotify api to explore a wide variety of music data. Through spotify, I was able to collect relevant information on top artists, genres, and playlists, and effectively store this information in data frames that I would later use for analysis. I took data from past coachella lineups, and enhanced that data with pandas, using the Spotify api to populate columns with relevant information for each coachella artist. I then compiled this information into new CSVs, which I used for my final analysis of the data.

## Process

### *Data Cleaning*

#### *Artists - What are the current top Artists on spotify?*

I spent the majority of this project cleaning my data and getting it into the format that I wanted for analysis. I started with getting information on the top artists according to spotify. I did this using spotify's *popularity score* which is a data point they keep for each artist and track. I compiled a dataframe of spotify's top artists and populated it with relevant information on each artist.

(NOTE - The only way I was able to iterate through each artist on spotify was to do a query for every letter in the alphabet, which is what I did)

#### *Genres - What are the current top genres on spotify?*

Next, I looked at the top genres on spotify, by populating a data frame with each genre associated with an artist's name. I then grouped that dataframe by genre and saved relevant information on each genre.

#### *Playlists - What are the best Playlists on Spotify? What makes them the best?*

I spent probably 10 hours writing a function that classifies a playlist by the genres it is composed of. This proved to be more difficult than I anticipated, because the only way to do this

was to analyze each artist on each track on the playlist, and return information about those artists' genres.

*Coachella Data - What artists Performed when at coachella? Can we get relevant information on each of these artists from the spotify api? What genres dominate coachella?*

I was inspired by a coachella csv file I found through kaggle on reddit, but I only found 2019 data, so I used images of coachella's 2017 and 2018 lineups to populate new csv files I would use in analysis. I imported these csv's and added to them using information from the spotify api. I ended up with three data frames filled with relevant information on artists who performed at coachella in 2017, 2018, and 2019.

### *Data Analysis*

After compiling 6 relevant dataframes, I exported those dataframes to new csv's, so I could analyze my findings in a fresh environment. I used mainly seaborn to create visualizations and discover trends in my datasets.

### **Results**

I had a real hard time taking a screenshot of my graphs on my microsoft surface, and I'm also having difficulty exporting my graphs to this document. Please reference the visualizations in my analysis ipnyb.

#### *Artists*

- I visualized the top artists on spotify, by followers for the first graph, and popularity score for the second.
- I noticed popularity score and followers were positively correlated

#### *Genres*

- Returned the most popular genres in a neat dataframe. It was interesting for me to scroll through this dataframe and see how certain genres are doing.
- Visualized artist genre popularity by both followers and popularity score.
- Visualized a distribution of popularity score sum and mean, noticing a 'sweet spot' where popularity score(sum) spiked even though mean was not at its max.

#### *Playlists*

- I was curious if the number of genres present in a playlist were related to the play playlist popularity
- I found that these variables were not positively correlated and that there was a sweet spot for a diversity score around 70

#### *Coachella Analysis*

- Found a strong correlation between coachella rank and year.
- Found that more popular artists tend to perform on sunday
- Found a correlation between recency and artist popularity
- Top Genres at coachella historically -- alternative r&b, alternative dance, and dance pop

**Future Analysis and Implications**

If I were to continue to work on this project, I would look more in depth at how each genre interacts with other genres. I think this would be super interesting and also possible to do using the spotify api as a base.

Thanks for a great semester!