Decoding Popularity

A Data-Driven Exploration of Spotify Hits

By: Bennett Blanco, Gaurangi Agrawal, Olivia Peng, Yashvardhan Singh Ranawat, Lila Su, Bonny Yang

Outline

- Introduction
- Data Sources & API Documentation
- Business Problem
- Breakdown of Analysis
- Key Visualizations and Key Findings
- Recommendations

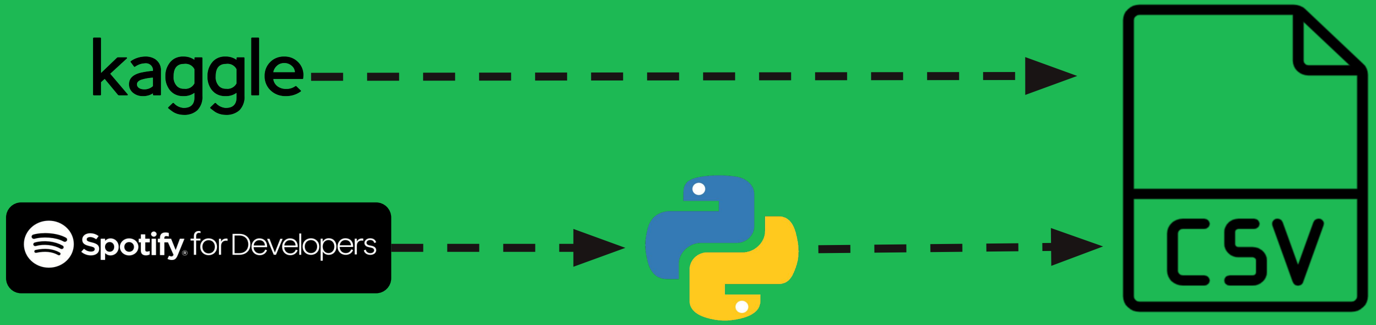
Spotify

What is it?

- Spotify is the largest music streaming service in the world.
- It has revolutionized the modern listening experience with Machine Learning applications such as NLP & Reinforcement Learning

Data Sources

- Kaggle
- Spotify API



Dataset & Variables of Interest

Summary

- ~20,000 rows of unique tracks
- Random sample of the Spotify ecosystem

Track Popularity

- Measured from 0 - 100
- Live metric that is calculated by Spotify
- Generally, songs currently being played a lot will be more popular

Audio Features

- Measured from 0 - 1
- Tracks are made up multiple features with unique values
- (ex. Danceability, Energy, Valence, Acousticness)

Business Problem Statement:

How can Track Popularity be utilized to create value for artists and Spotify stakeholders?

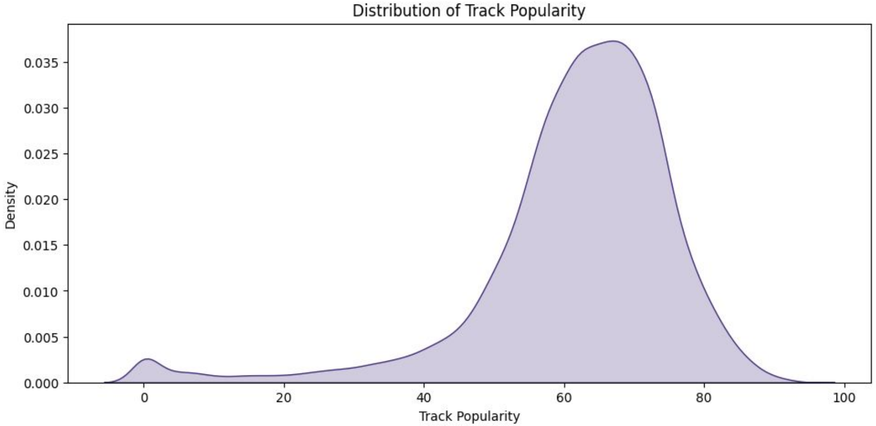
Key Findings

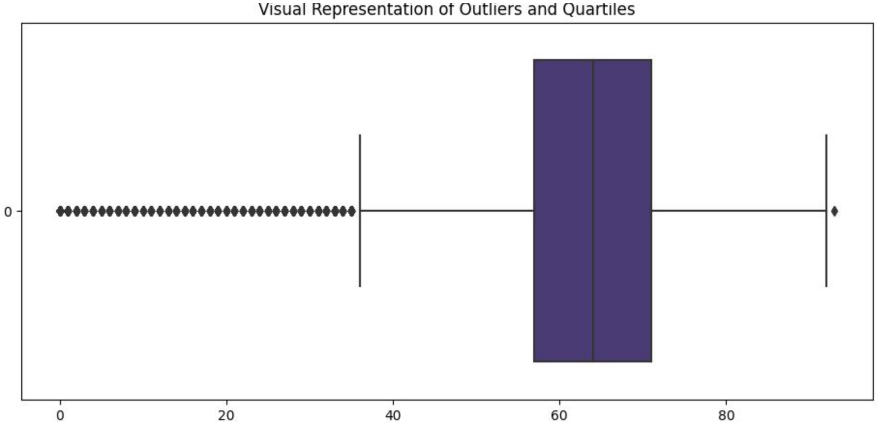
Visualizations and Analysis

Track Popularity Distribution

- Track Popularity is not normally distributed

- Outliers are likely not true artists, so it doesn't make sense to include them

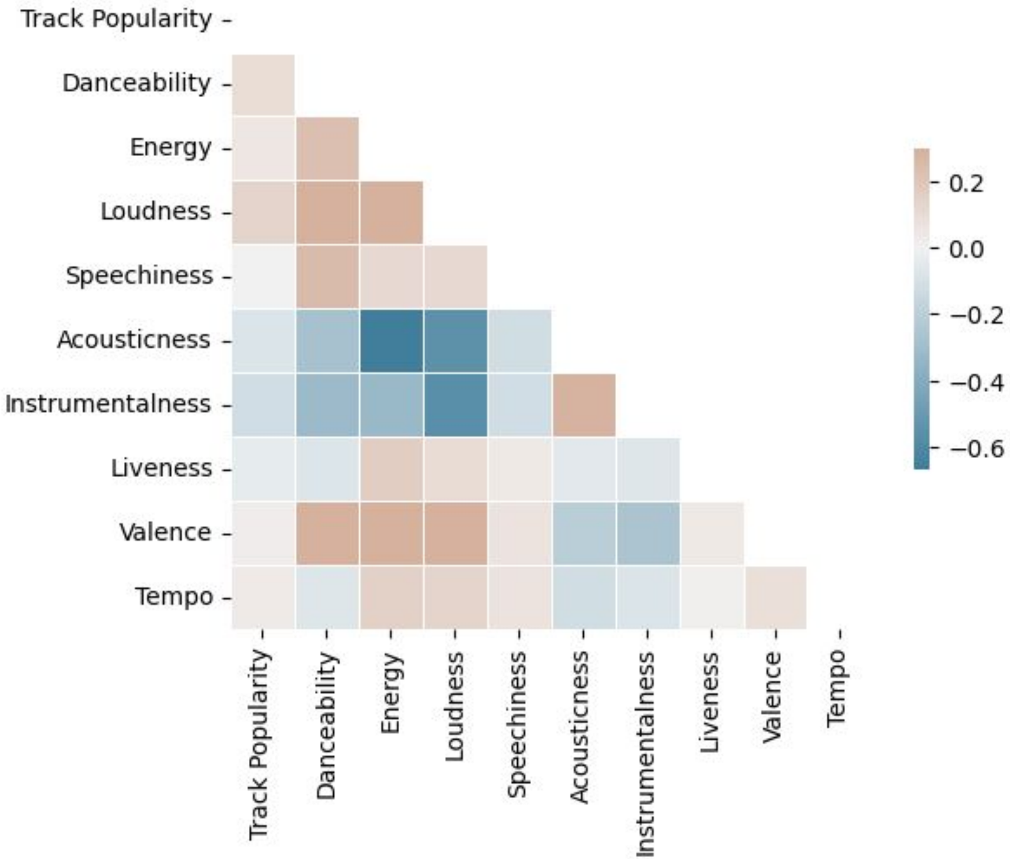




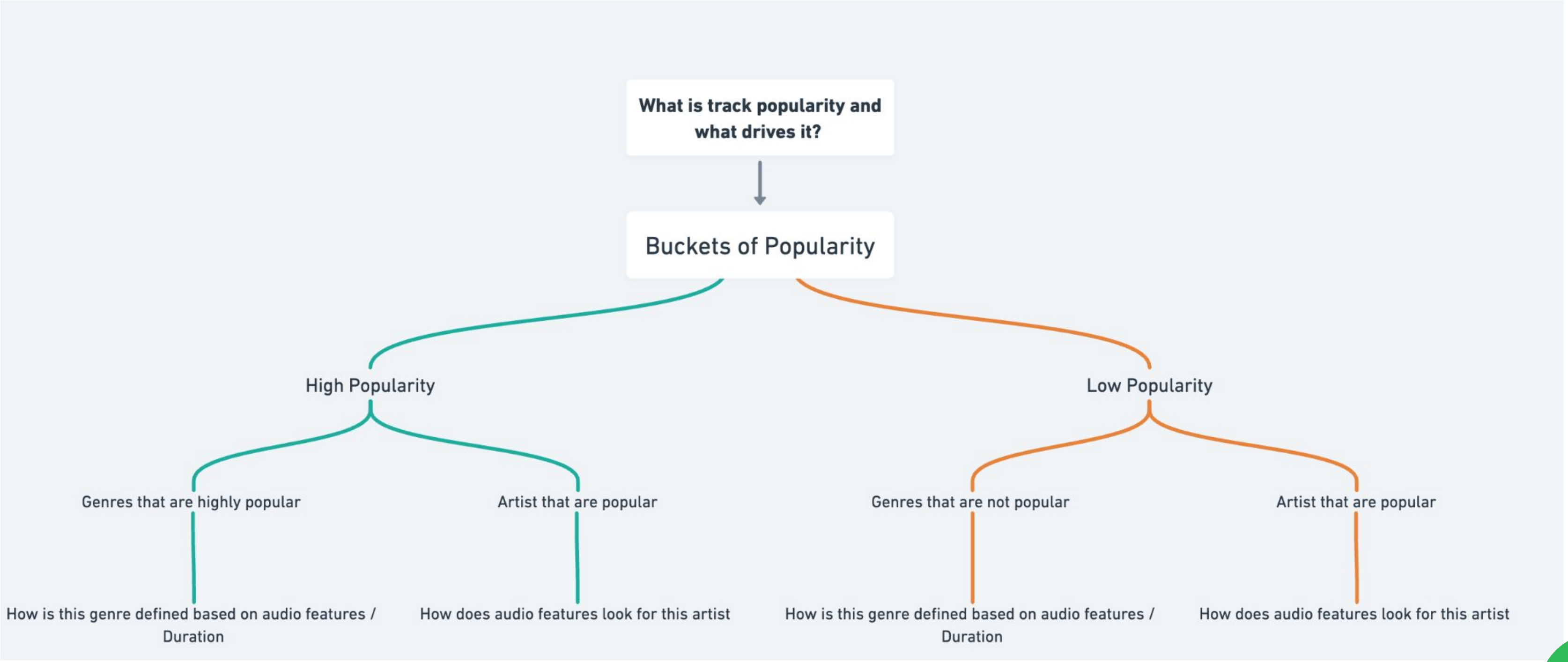
Do Individual Features Drive Track Popularity?

- Goal is to identify a specific driver of Track Popularity

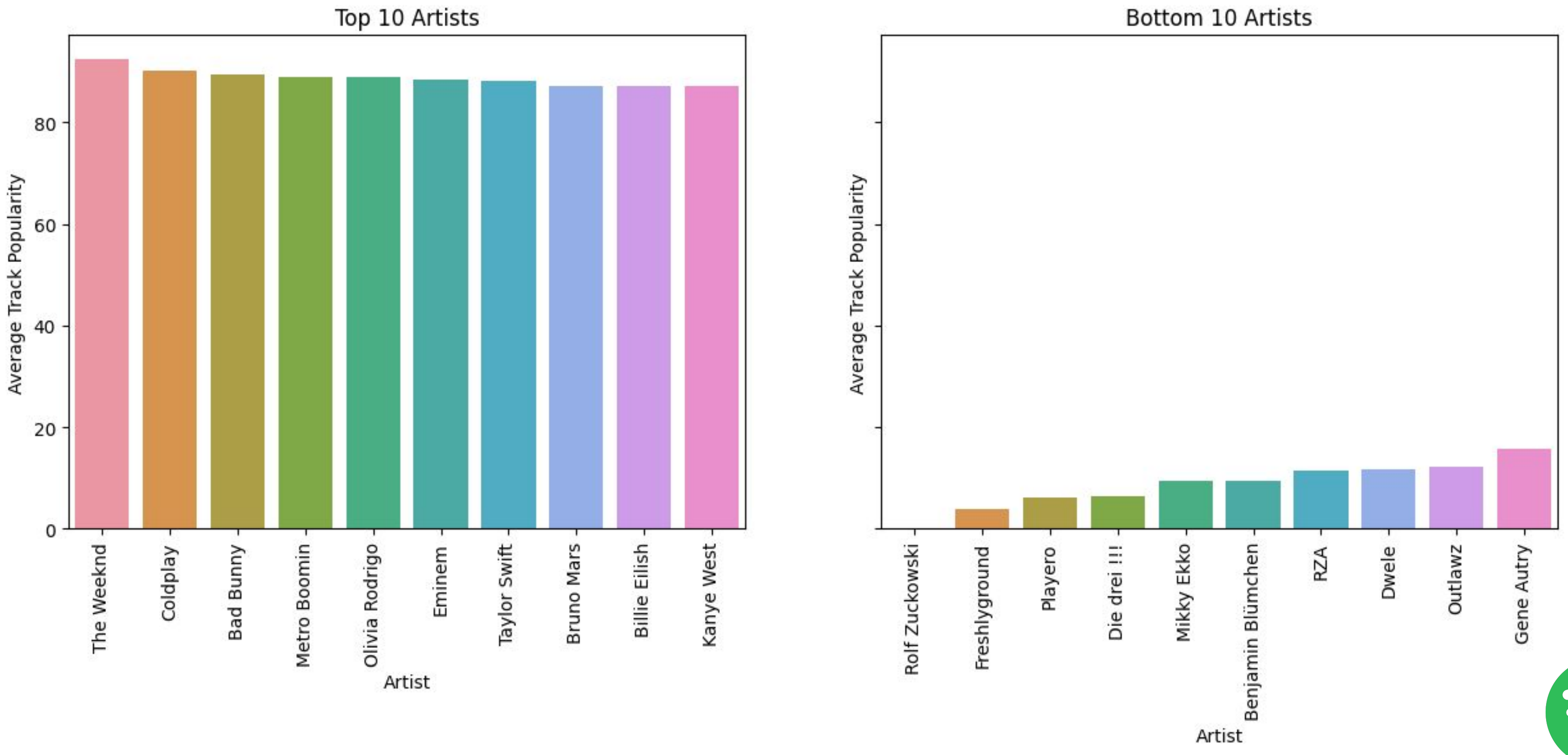
- Heatmap shows us that this is actually not the case



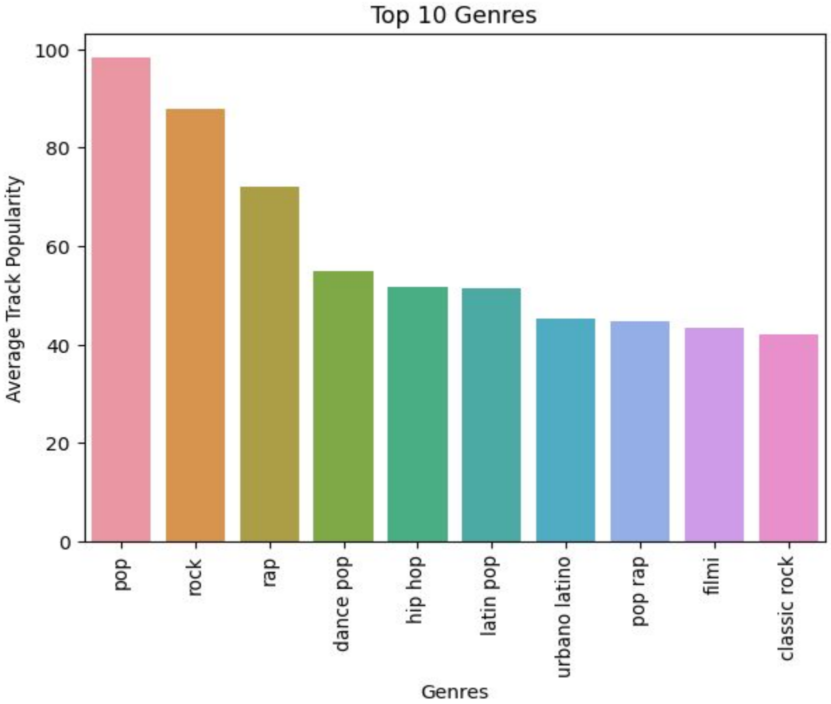
Rethinking our Analysis

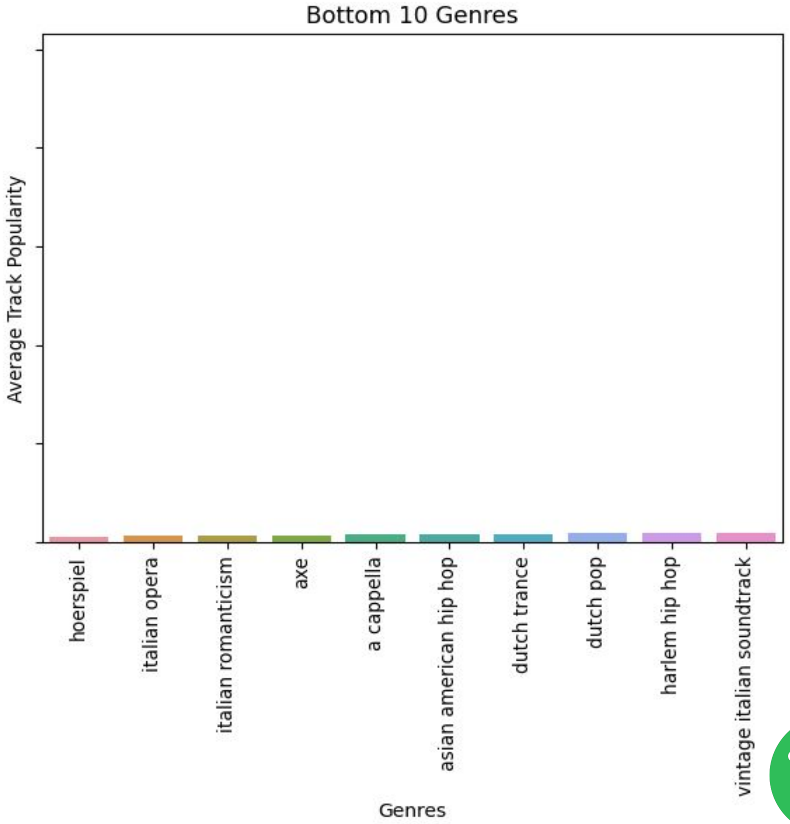


What's Popular Now? (Trending Artists)



What's Popular Now? (Trending Genres)





Artist Archetypes

- Consistency among top artists in audio feature makeup

- Lack of consistency among unpopular artists

Bottom Artists

Top Artists

Genre Archetypes

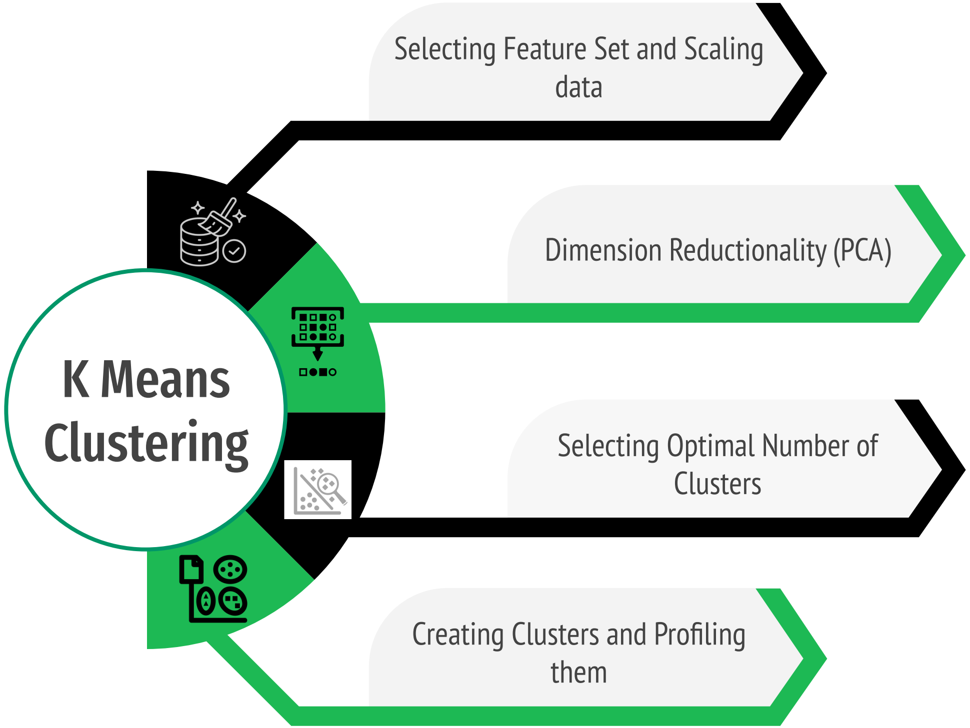
- Consistency among top genres in audio feature makeup

- Lack of consistency among unpopular genres

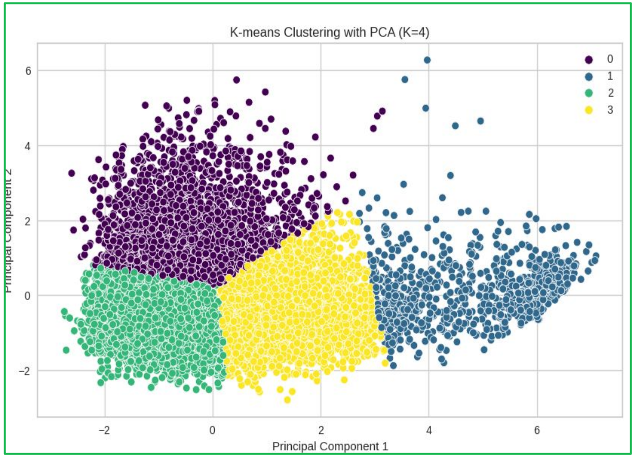
Bottom Genres

Top Genres

Clustering


K Means Clustering
Selecting Feature Set and Scaling data
Dimension Reductionality (PCA)
Selecting Optimal Number of Clusters
Creating Clusters and Profiling them

Clustering



Feature	Metric	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Track Popularity	Mean	60	57	66	63
Danceability		0.55	0.32	0.72	0.56
Energy		0.78	0.16	0.71	0.46
Acousticness		0.13	0.87	0.18	0.51
Instrumentalness		0.03	0.63	0.01	0.05
Liveness		0.36	0.13	0.14	0.14
Valence		0.49	0.15	0.66	0.39

Recommendations

- For Spotify
  - Utilize clusters with highest feature to pursue users on platforms that correspond with that feature
  - Utilize cluster with highest track popularity for top-funnel marketing and on audiences with little known information
  - Utilize clusters to create new playlists

- For Upcoming Artists
  - Utilize audio feature radar charts and produce tracks within those parameters for higher growth

Appendix

- Notebook & Data Dictionary:
  https://github.com/jbblancojr/spotify/blob/main/A01_Decoding_the_Secret_to_Popularity_Spotify.ipynb

Elbow Plot for Clustering