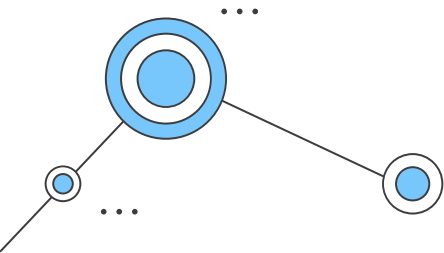
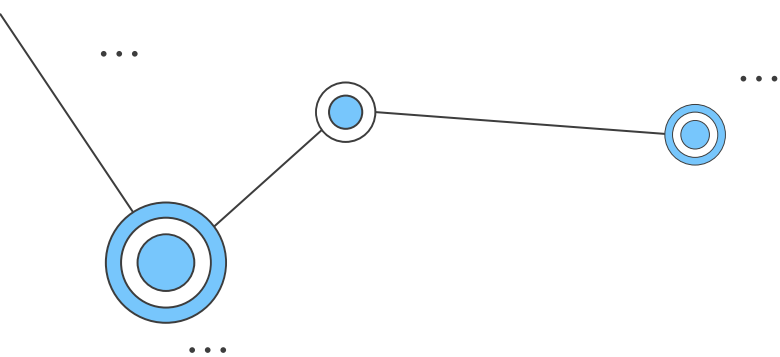


Spotify Playlist Recommendation System

Group 10

Eric Simpson, Theanie Baskevitch, Kirtan Shah, Roland Wen, Apeksha Gaonkar



Overview

- Motivation
- Dataset Description
- Methodology
- Analysis & Observation
- Summary

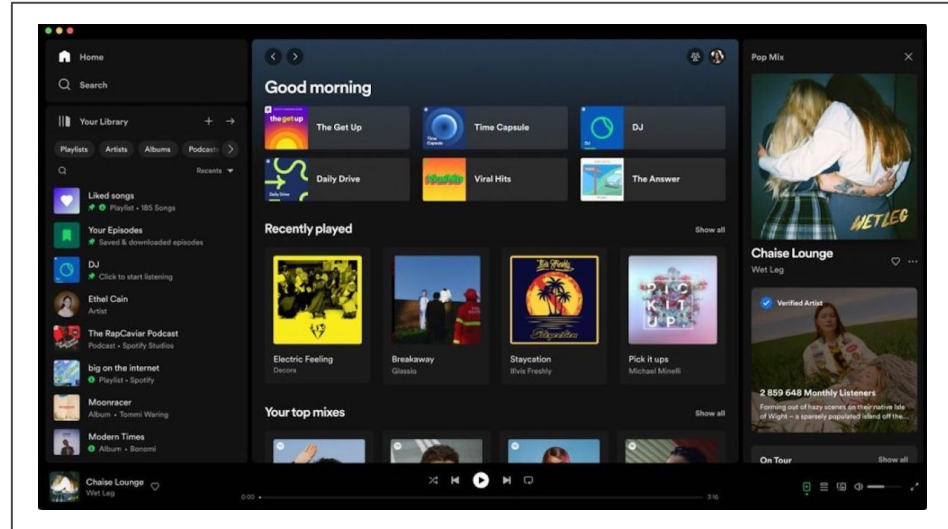
Motivation

What is Spotify?

Digital music streaming service

Why does analysis matter?

Offers valuable analytics on listener behavior, emerging trends, and overall music consumption patterns



Dataset Description



What data do we have?

- [2018 Spotify Playlist Recommendation Dataset Challenge](#)
- [Spotify Web API](#)
- **1 million** unique playlists with track, artist, album information along with various metadata (~**33 GB**)
- **22,622,292** unique tracks
- **295,860** unique artists
- **734,684** unique albums

How was the data prepared?

Separate into 3 dataframes:

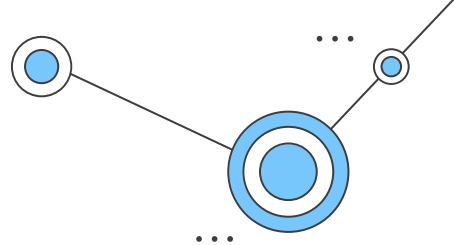
- **Tracks:** Track information + unique track id
- **Playlists:** Playlist information + a list of track ids
- **Playlists & Tracks Mapping:** Expand track ids for fast access

Compressed to **1.7 GB**

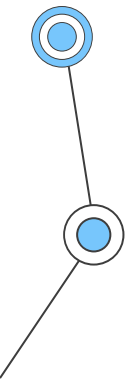


Analysis & Observation

Analysis & Observations



- Understanding the Data
 - Tracks - insights about unique tracks across **all** playlists.
 - Playlists - insights about unique playlists.
 - High level basic “feel” for various features of the dataset.
 - Use as insight into asking more pointed questions.
- Clustering Analysis & Track Recommendations
 - K-means clustering to recommend N tracks based on audio characteristics of given track or playlist.





Understanding the Data (Tracks)

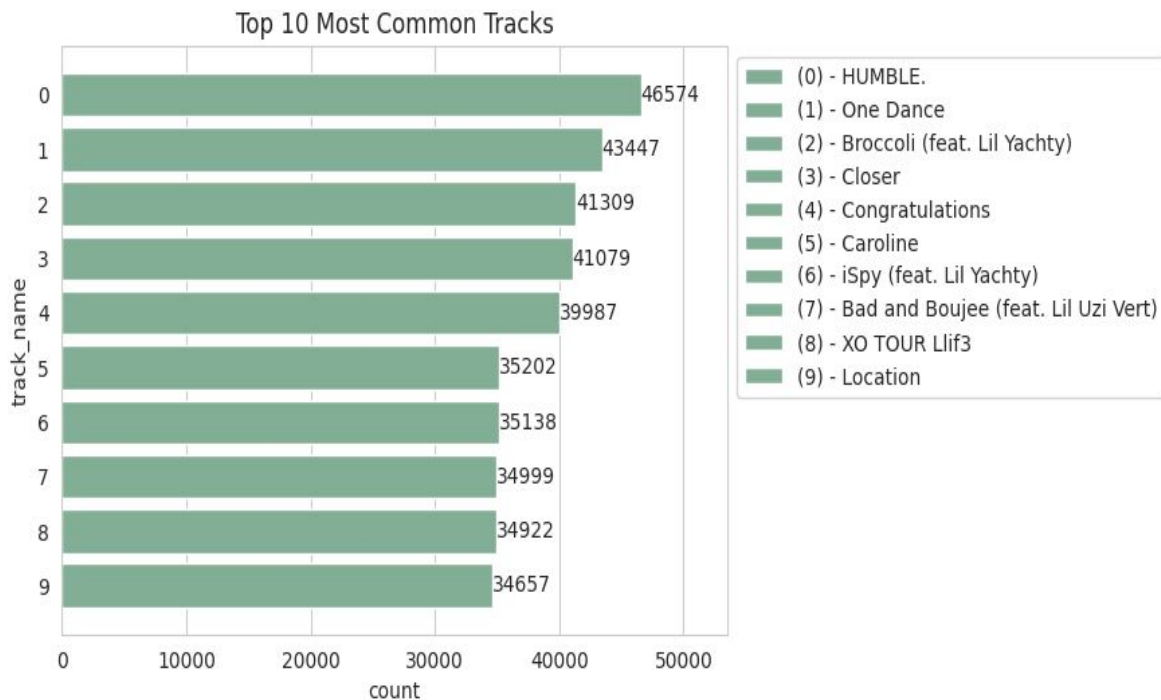
What kind of tracks, artists, or albums do we expect to be the most popular?

...



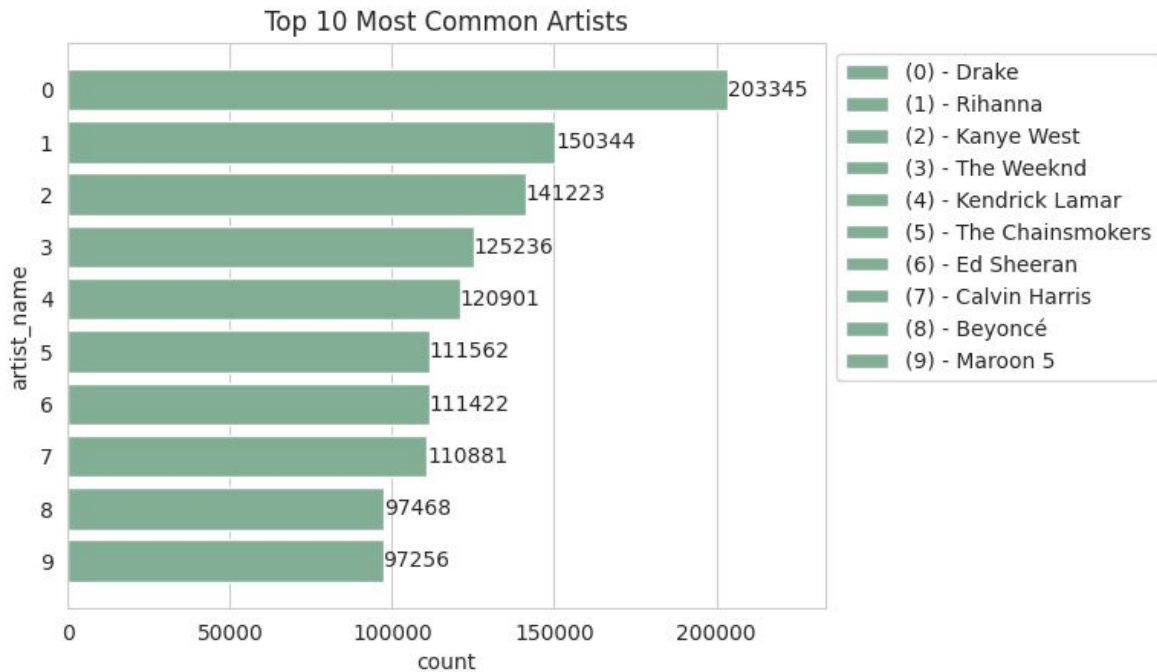
Top Tracks

Which tracks are most popular across all playlists? **Pop/Hip-Hop**



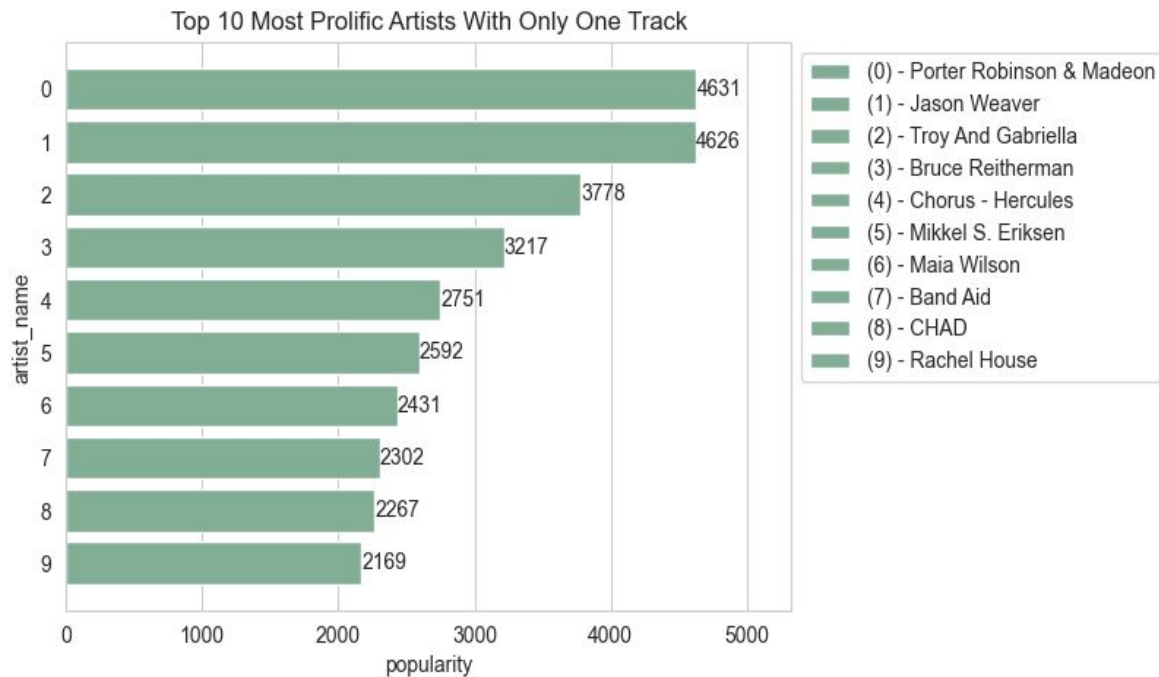
Top Artists

Which artists are most popular across all playlists? **Pop/Hip-Hop**



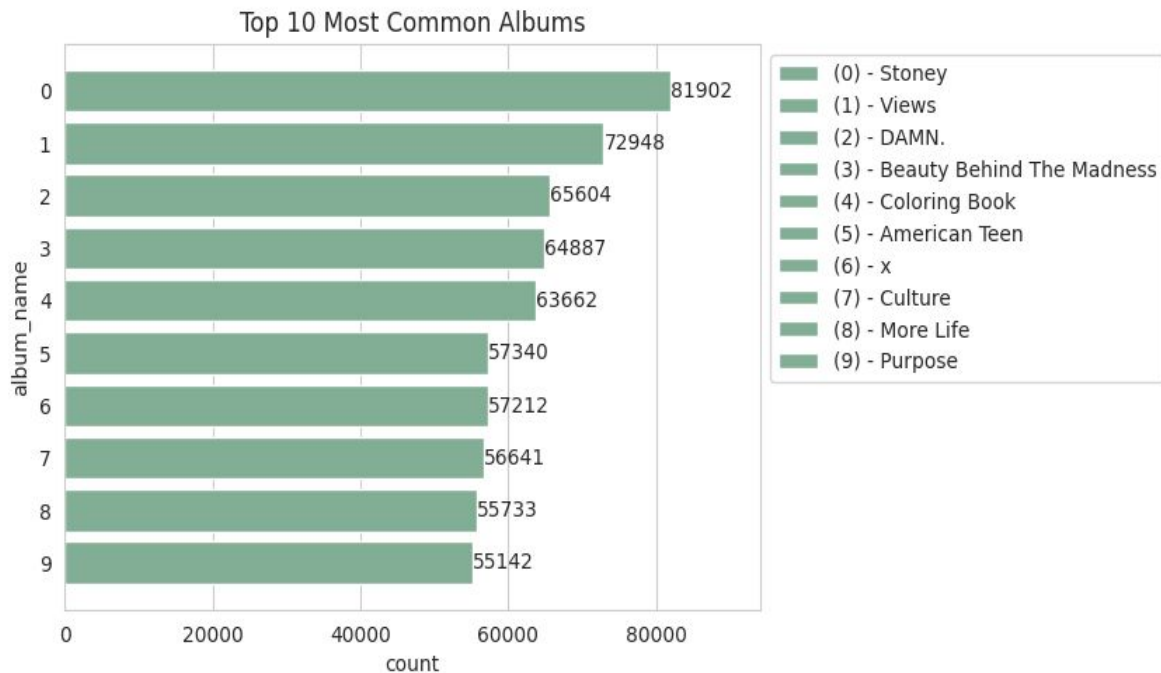
Top Artists with One Track

What artists with only one track are the most popular?



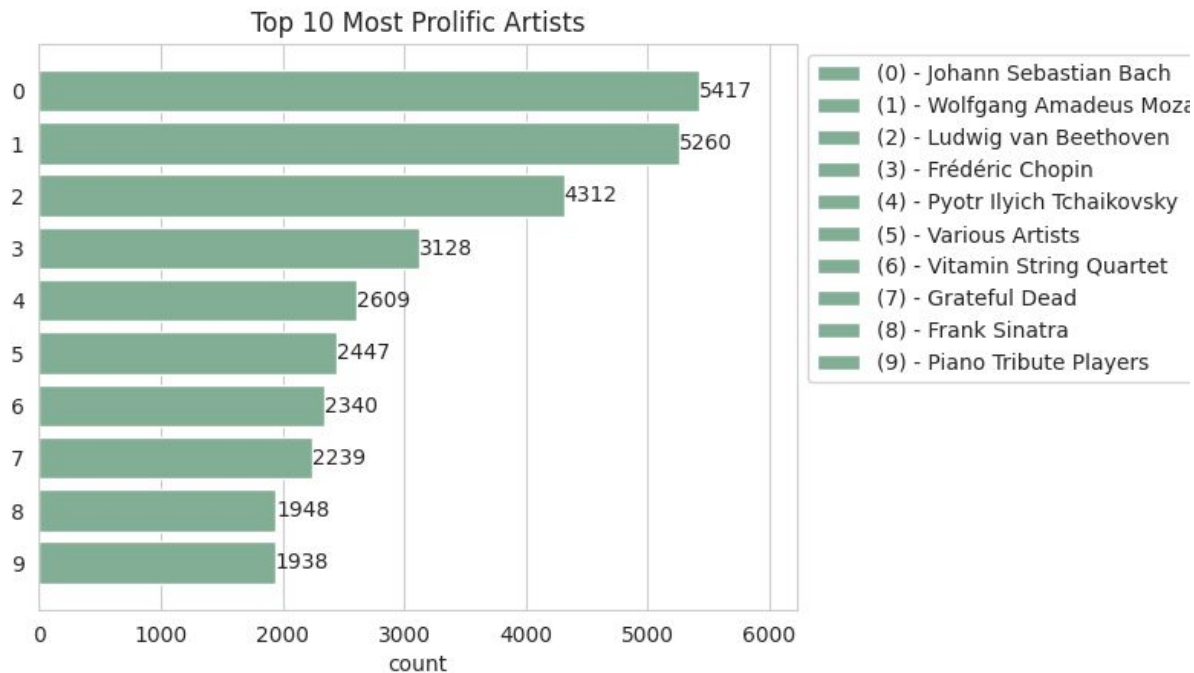
Top Albums

Which albums are most popular across all playlists? **Pop/Hip-Hop**



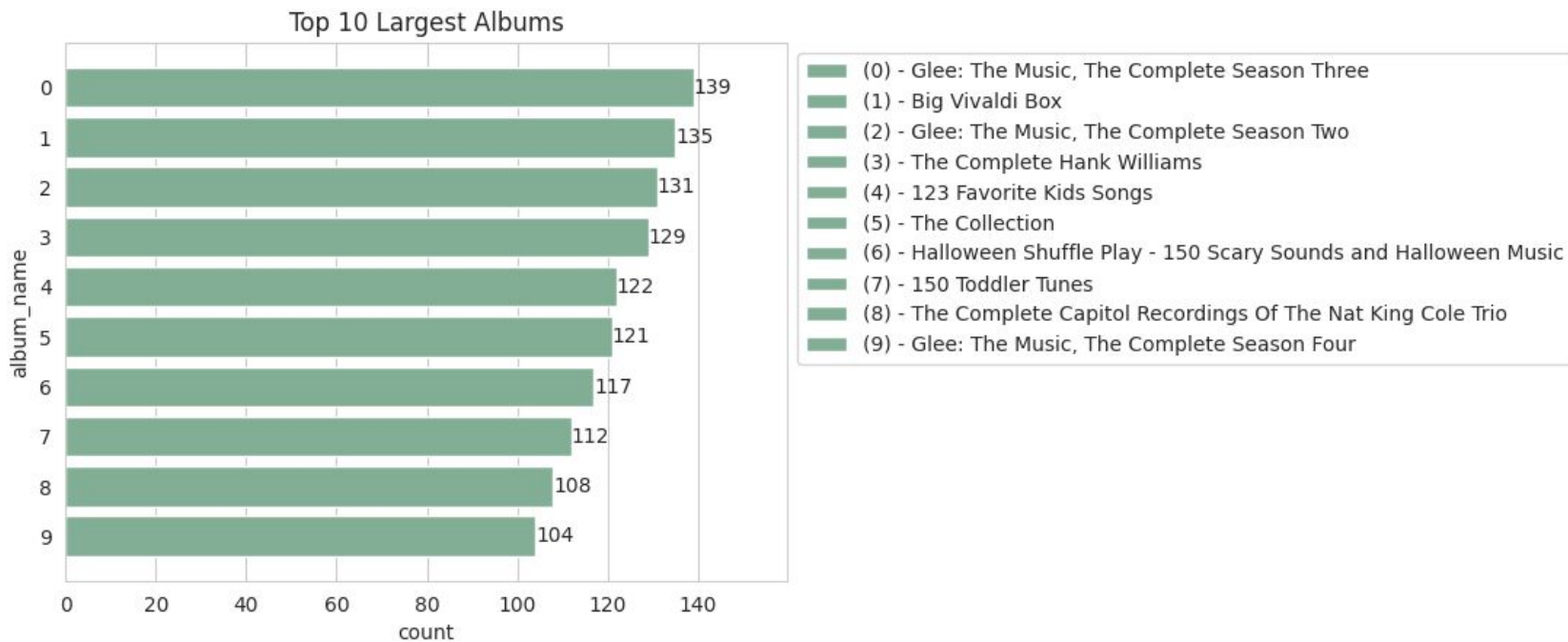
Most Prolific Artists

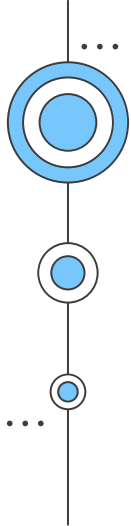
Which artists are the most prolific across all playlists? **Classical**



Largest Albums

Which albums are the largest across all playlists? **TV/Seasonal/Collection**





Understanding the Data (Playlists)

What factors matter to listeners when deciding how to
curate a playlist?

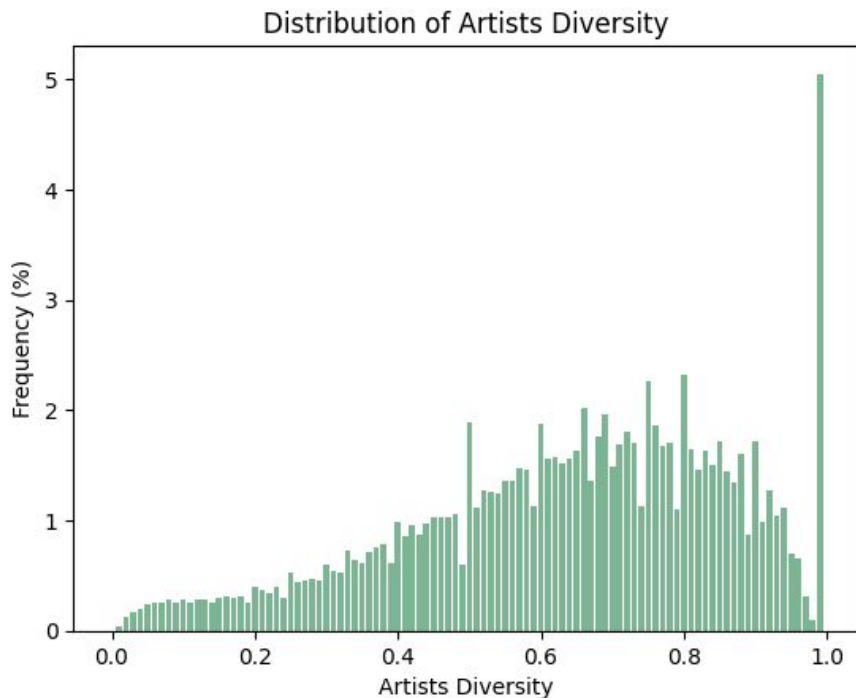
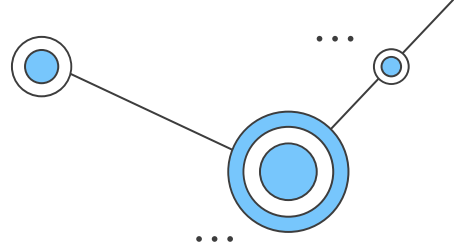
...





Artist Diversity

What is the diversity of artists like in the playlists? **Highly diverse**



Artists diversity =
 $\frac{\text{\# of unique artists}}{\text{\# of tracks}}$

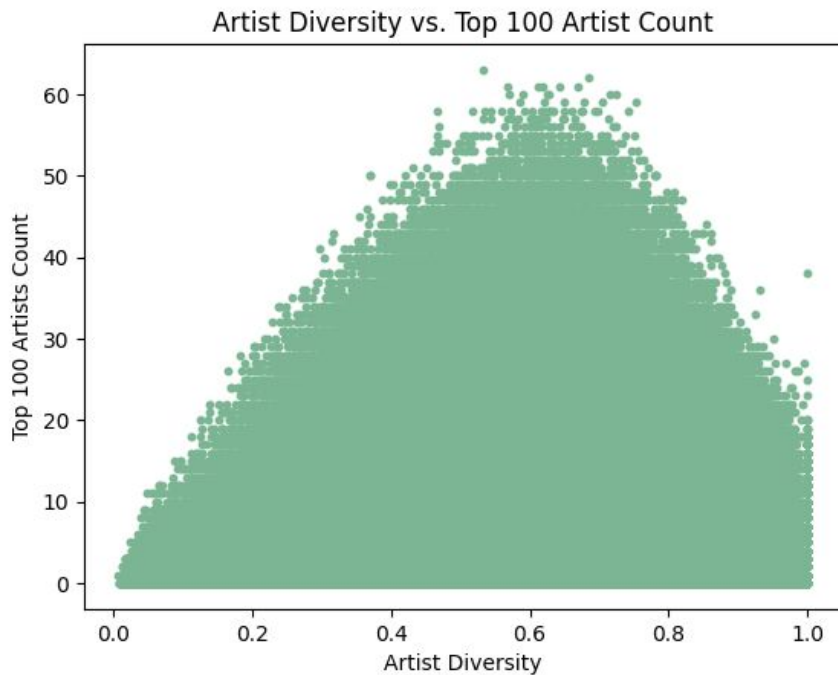
Hypothesis

Playlists with higher artist diversity are not as interested in popular artists.

...

Artist Diversity

Playlists with **more artist diversity** are **less interested in popular artists**



Hypothesis

Popular tracks are popular because they share common characteristics that differentiate them from more “average” tracks.

...

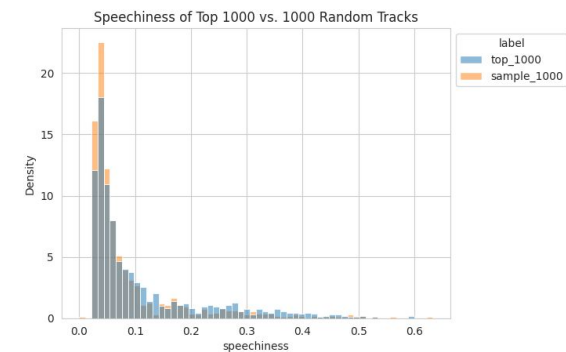
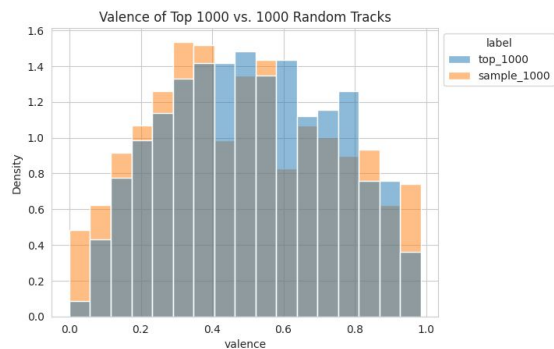
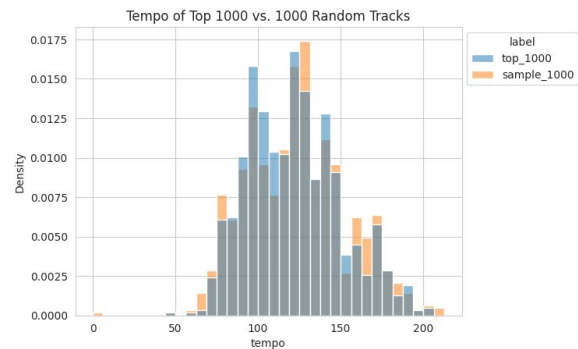
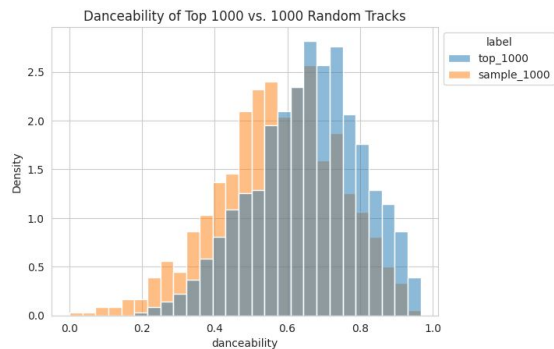


Audio Characteristics

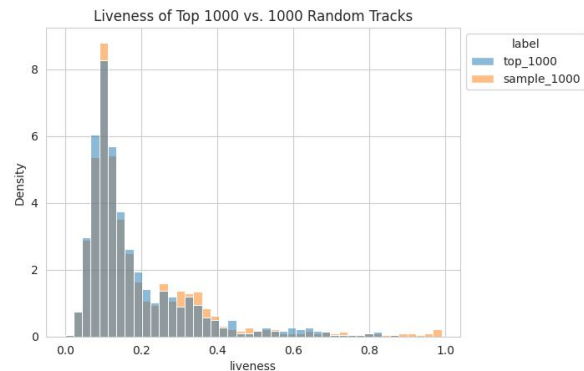
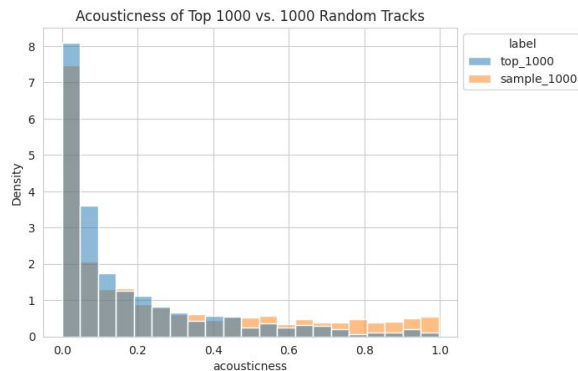
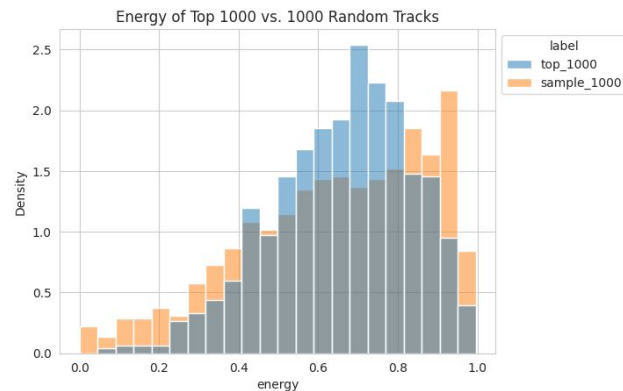
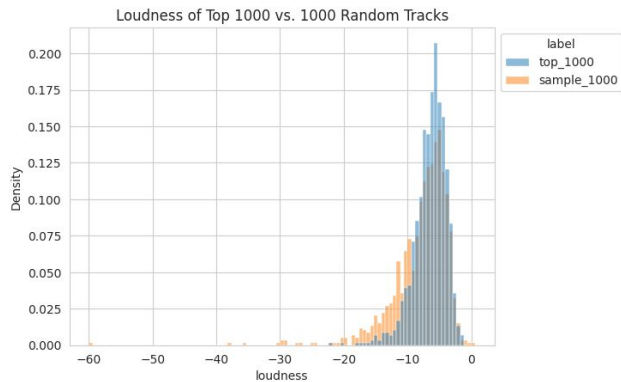


- Spotify API
 - **Danceability** – How suitable a track is for dancing.
 - **Energy** – Represents a perceptual measure of intensity and activity.
 - **Loudness** – The overall loudness of a track in decibels (dB).
 - **Speechiness** – A measure of the presence of spoken words in a track.
 - **Acousticness** – A confidence measure of whether the track is acoustic.
 - **Liveness** – Detects the presence of an audience in the recording.
 - **Valence** – Describes the musical positiveness conveyed by a track.
 - **Tempo** – The overall estimated tempo of a track in beats per minute (BPM).

Audio Characteristics



Audio Characteristics



Audio Characteristics Takeaways

- For some of the audio features, there is a distinguishable amount of separation between the distributions of average tracks vs. the most popular.
- Most notably, the audio features of **danceability, loudness, & energy** show signs of separation while the remaining features are relatively indistinguishable.
- Popular tracks have large variance when it comes to **valence** (positivity) and moreover do not differ much from average tracks.
- We consider these audio features as a motivating factor for building the recommendation model.

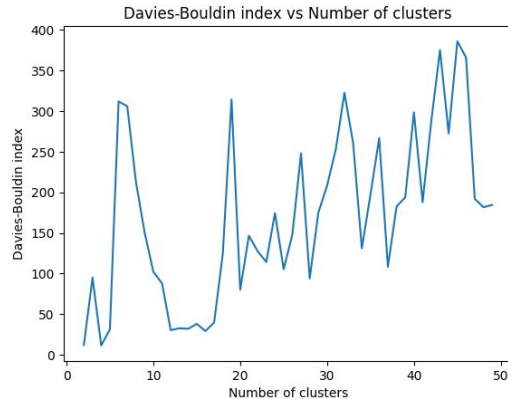
Cluster Analysis

Performed K means clustering

Selected number of clusters based on below metrics :

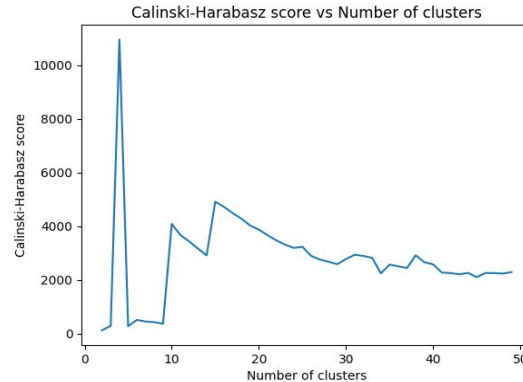
1. Davies Bouldin Index value,
2. Calinski harabasz score.

K = 18 gave best value for both these metrics



Calculated Cosine similarity

- High similarity value was observed for tracks within a cluster.
- Low similarity value was observed for Tracks in different cluster.



Recommended N tracks

1. Select tracks from same k-means cluster as current song
2. Select tracks from give playlist with high cosine similarity with current song.

```
python recommend_track.py 2jQiSYrwJehQAcuaaQrXnS --playlist_id 248269
```

Recommended songs for Despacito

24K Magic

Celebration

Get Down On It - Single Version

The Only One For Me

Dancing On The Ceiling

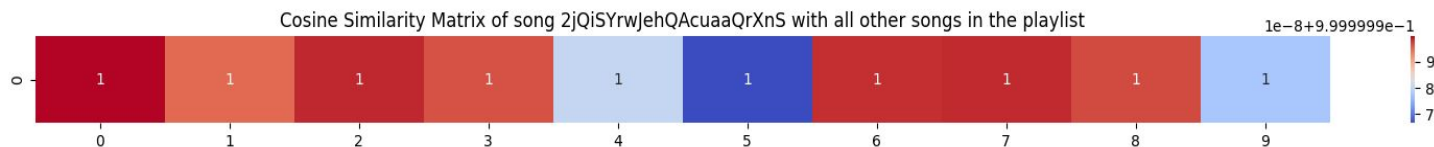
End Of The Road

Poison

Get Up Offa That Thing

U Got It Bad

Case Of The Ex (Whatcha Gonna Do)



Thank you!

Any questions?

