**Spotify Recommendations Model Project - CSCIE109a**

**Milestone 2**

**Project Statement**

Spotify users can select individual songs, albums or artists to listen to. This project will allow users to create playlists based on their preferences.

This project will allow the user to “prime” a playlist with a “context”. This context will drive the nature of the playlist and will be considered for all playlist recommendations. The context can include any combination of one, none or many of the following song aspects:

* Genre
* Mood
* Artist
* Song

Once the user “primes” the playlist with the context, the model will generate a list of 10 songs. The user accepts a song by listening to the song and rejects the song by skipping it or deleting it. Once 5 songs remain un-judged, the recommender will generate another 10 songs based on the songs that were accepted, if any. The recommended will always use the most recent songs to generate subsequent recommendations. A maximum of 10 most recent accepted songs will be used.

The final call to generate a playlist will look as follows:

|  |
| --- |
| get\_playlist( context:dict = { genre:list = **None**,  mood:list = **None**,  artist:list = **None**,  song:list = **None** },  recent\_songs:list = **None** ) |
|  |

This function can be used to “prime” the list with a context of one or several contextual markers with no recent songs. Once songs have been confirmed, the function can be called again with the original context as well as a list of recent songs.

**Preliminary EDA**

**Million Songs Listing**

The primary data for this project will come from list of all available playlists and songs in the supplied dataset. The list of CSV files has been loaded into an SQLite database for easier retrieval. The code base uses SQLalchemy as the database’s ORM wrapper.

The current dataset includes the following:

Playlist id

Song position in playlist

Track Name

Track URI

Artist Name

Artist URI

Album Name

Album URI

Song Duration in ms

After analysis of the database, the following statistics have been gathered:

Total number of song entries: 66,346,428

Number of unique songs: 1,483,760

Total Number of Playlists: 999,001

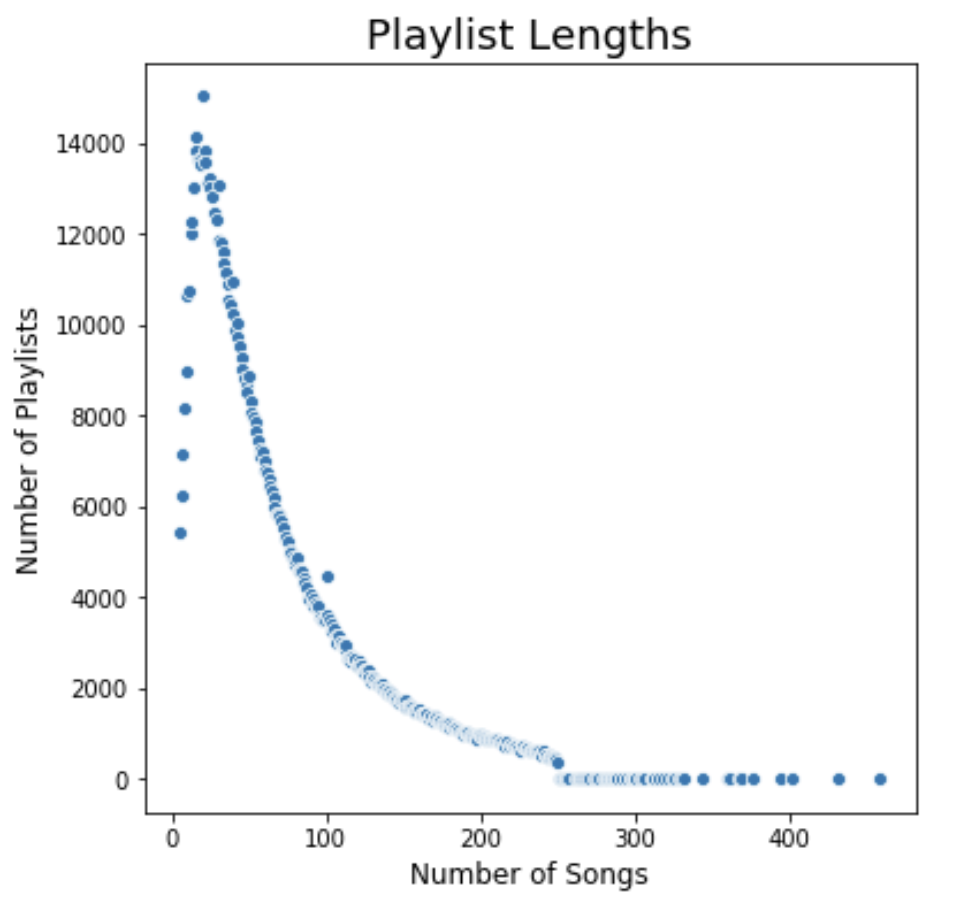
Number of Unique Artists: 287,740

**Distribution of Playlist Lengths**

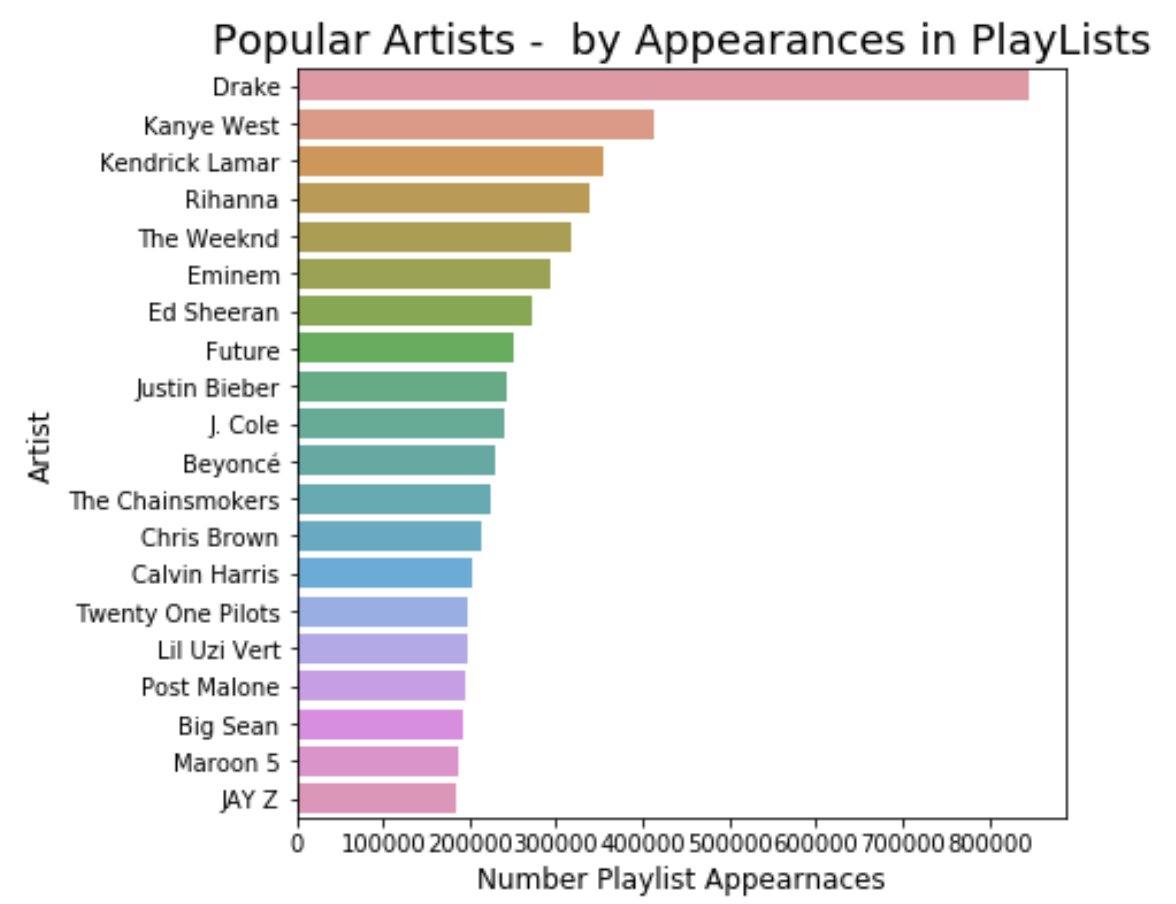
Most common playlist length: 20 songs

~90% of playlists have under 150 songs

~50% of playlists have under 50 songs



**Artist Popularity**



**Spotify API**

A link to the Spotify API has been established. Using the API, additional fields are available; however, many are already collected in the million song database. Most noteworthy is the following field:

From track\_uri

Popularity

From artist\_uri

Followers

Genres

Popularity

From album\_uri

Popularity

Release Date

These can help to identify the genre as well as popularity both of which can be useful when creating a recommended playlist.

For the sake of keeping the project’s scope reasonable and achievable within the timeline, we will restrict our data to these sources.

Additional EDA Efforts:

Group Efforts:

* Identification of data necessary to develop model
* Identification of sources for data
* Definition of data structures
* Assignment of tasks to team members
* Finalization of dataset

Individual Member Efforts

* Collecting data from data sources and consolidating into single database
* Removal of redundant and corrupted entries from each source
* Exploring efforts to speed retrieval of stored data
* Individual review/audit of final data set by each member

**Project Milestones**

The first effort will focus on finalizing our dataset and defining our use case. Once this is done, we will each create our own proposal for modeling a solution.

During the development, we may rule out some of these models and consolidate fforts.

Our object is to present various ways of solving the problem and presenting the pros and cons of each option and the preferred model.

|  |  |
| --- | --- |
| Task | Deadline |
| **Milestone3: Proposal** | **11/25** |
| Develop Data APIs | 11/2 |
| Finalize Dataset/Use Case | 11/2 |
| Finalize Base-Line Model | 11/9 |
| Model Proposals – 1 per member | 11/16 |
|  |  |
| **Milestone3: EDA & Revised Statement** | **11/20** |
| Build models – test and train #1 | 11/23 |
| Build models – test and train #1 | 11/30 |
| Models Complete | 12/7 |
| **Milestone4: Completion** | **12/11** |