CS 181 - Final Project: Web Service API's

Authors: Amna Khalid and Loc Tran
Date: 12/16/2019

Tables on SQL database:

1. SPOTIFYCHARTS: data acquired from https://spotifycharts.com/regional as a .csv file to see monthly popular songs on spotify
2. TOPTRACKS: data acquired via OAuth2 in JSON format from spotify's API to see top tracks in our classmates spotify accounts
3. followingArtists: data acquired via OAuth2 from spotify's API in JSON format to see the artists followed by classmates on their spotify accounts
4. savedTracks: data acquired via OAuth2 from spotify's API in JSON format to see the tracks that have been saved by our classmates to their spotify accounts

SQL Query Explanation:

Query # 1: Number of streams of tracks on Monthly Spotify Top Charts that are also in Saved Tracks

```
SELECT SPOTIFYCHARTS.Track_Name, SPOTIFYCHARTS.Artist, SPOTIFYCHARTS.Streams
FROM savedTracks inner join SPOTIFYCHARTS on SPOTIFYCHARTS.Track_Name =
savedTracks.Song_name;
```

In this query we are trying to find what tracks are on the top monthly spotify charts and are also in the saved tracks data of our classmates and how many times these tracks have been streamed monthly on spotify. We can get the required data by inner joining the SPOTIFYCHARTS and savedTracks tables and displaying the track name, artist name and number of streams.

Query # 2: Artists with top popularity on Spotify that are also on the Monthly Top Spotify Charts

```
SELECT DISTINCT SPOTIFYCHARTS.Artist, followingArtists.popularity FROM followingArtists inner
join SPOTIFYCHARTS on SPOTIFYCHARTS.Artist = followingArtists.Artist
ORDER BY followingArtists.popularity DESC;
```

In this query we are trying to find what artists are on the monthly top spotify charts and are also on the followed by users in out class. After that we want to see the top ten artists that have the highest popularity on the spotify followed artists data. We get the required data by inner joining our SPOTIFYCHARTS and followingArtists tables. We use DISTINCT to avoid getting repeated values. Additionally, we use we use ORDER BY to get the values in descending order since we're looking at the popularity in order.

Query # 3: Popular Genres in users Following Artists

```
SELECT *
```

```
FROM (SELECT  count(Genre) as count FROM followingArtists WHERE Genre LIKE '%pop%') as pop
UNION
SELECT *
FROM (SELECT  count(Genre) as count1 FROM followingArtists WHERE Genre LIKE '%rap%') as rap
UNION
SELECT *
FROM (SELECT  count(Genre) as count2 FROM followingArtists WHERE Genre LIKE '%house%') as edm
UNION
SELECT *
FROM (SELECT count(Genre) as count2 FROM followingArtists WHERE Genre LIKE '%rock%') as rock;
```

In this query we are trying to find what genres are popular among our classmates by looking at the followingArtists data set. We do this by using the COUNT function combined with WHERE and LIKE to see the number of occurrences of pop, roc, rap and edm in our genre column. We are using nested queries for the count acquisition of each genre separately. UNION is further used to combine all these subqueries into a single query that can display our results in a single query.

Code Explanation:

The program is using OAuth2 to access users information (like their top tracks etc.) on spotify. OAuth2 is used to access information without having to know users passwords for using their information.

Using OAuth2, in our token acquisition notebook, we are creating an authorization url for spotify and using a get request with our .json file containing client id, redirect uri, response type, state and scope to get an access token. Once this access token is received, it is used to update the token by sending a post request to get a refresh token so we have authorization to the dataset without having to acquire tokens over and over again. All the information about the access token, refresh token, token type, scope and expiration time is stored in our “spotify.json” file for all users whose data we are accessing. Do note that not all providers have refresh tokens; spotify is one of many which has it available and thus is a requirement for our API; making our work easier.

For our data acquisition notebook, we are using the scopes: user-follow-read (read access to the list of artists and other users that the user follows), user-top-read (read access to user’s top artists and tracks), user-read-recently-played (read access to a user’s recently played tracks) and user-library-read (read access to a user's "Your Music" library). After receiving data from all these scopes from spotify, we are putting them into .json files (each scope in a different file). To clean up the data, we are iterating through the .json dictionary using enumerate and picking specific parts of the data such as artist name, track name, popularity etc. to keep and dropping the columns such as external_uri, href, id etc. that will not be useful to us in exploring our data. Further, the columns are also renamed and indexes have been set to match the confines of tidy data.

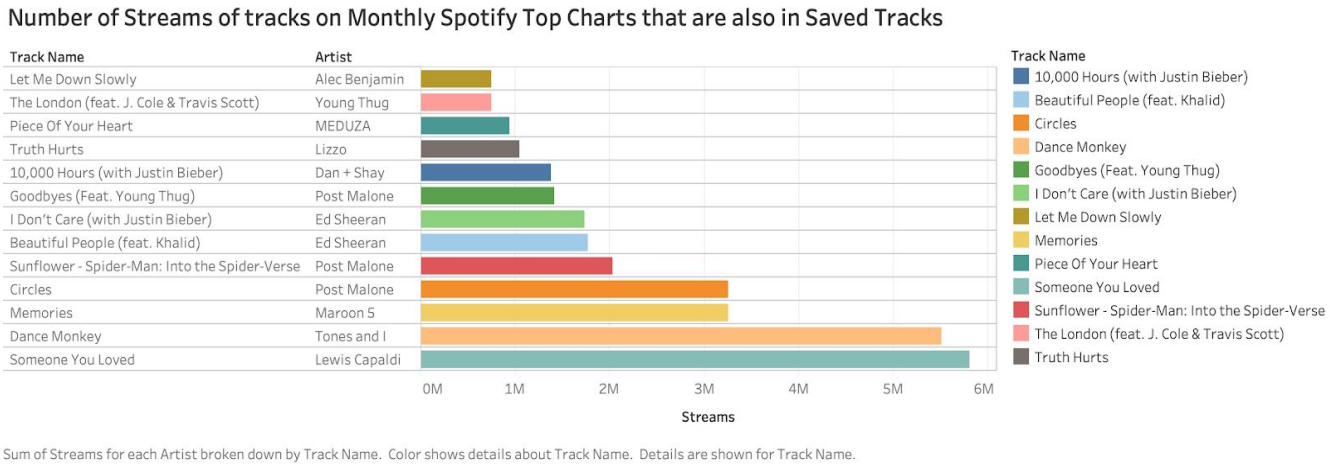
As our secondary data source, we downloaded the .csv file of the monthly top spotify charts from https://spotifycharts.com/regional. We put this data set into a pandas dataframe and performed operations to make it tidy.

Proceeding into the program, we created an sqlalchemy connection string, engine, and connection. On this database, we create four different tables (SPOTIFYCHARTS, TOPTRACKS, followingArtists, savedTracks) and then put our pandas dataframes into the respective tables using INSERT. Once that is done, we are making different queries to answer our data exploration questions and exporting them to .csv files to further explore the data via graphing in tableau.

Data Exploration:

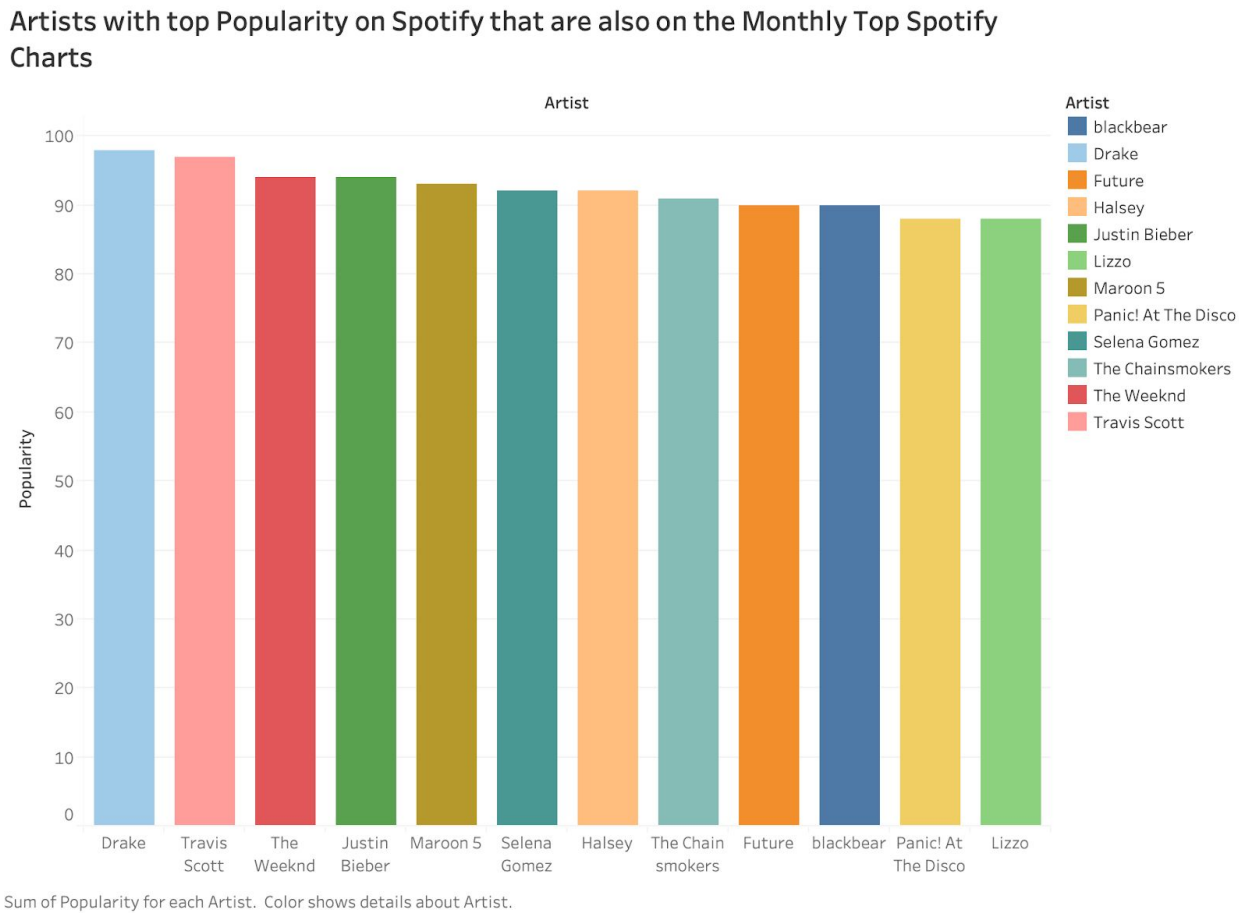
Q1: Are there similar tracks in saved tracks and in the monthly top spotify charts? If so, how many times have those tracks been streamed?

We explore the number of streams of similar tracks on the monthly top spotify charts and on the saved tracks data by representing the statistics on a bar graph since our data is discrete.
We find that the song that there are only fourteen songs that are common in both datasets. The song "Someone You Loved" by Lewis Capaldi has the most streams of 5,801,086 while the song "Let Me Down" by Alec Benjamin has 742,528 streams.
The song by Capaldi and Benjamin are both of pop genre. What is interesting to see here is that all the songs on our query are either of pop or hip-hop genre. Looking into this further, research shows (AAA Music) that the top genres of music in 2019 in the U.S. are of rap, pop and rock which is consistent with what we find from our data. The data also shows that most of the tracks are collabs with more than one artist which can result in higher streams. Other than that, Post Malone alone has three songs in the data followed by Ed Sheeran who has two which shows that right now, Post Malone and Ed Sheeran are doing well in the music industry.



Q2: Which artists that are followed by users are also on the top monthly spotify charts? And what is their popularity on spotify?

We explore how many artists are followed are also on the monthly top spotify charts and what is there popularity on spotify by representing this data on a bar graph.
Popularity of an artist is measured on a scale of 0-100; 0 being the least and 100 being the most popular among listeners. Popularity is measured by averaging over the number of streams that artist had collectively (of all their songs) in comparison to other artists.
We find that there are only twelve artists in common in both datasets. Drake has the most popularity of 98 while the least, 88, is held by Lizzo. This result makes sense since the artists would not have been on the monthly top spotify charts if they did not have a high popularity to begin with.
Furthermore, all of the artists listed are more prominent for rap, rock or pop music. It is also interesting to see that there are only 3 bands in this resultant query: Maroon 5, The Chainsmokers and Panic! At The Disco. Other than that, there are only 3 female artists with high popularity.



Q3: What are the most popular genres of artists that are followed by our classmates?

We wanted to explore what genres of music is popular in followed artists by the users in our dataset. We explore this by displaying the data on a bar graph.
We find that the most common genre is that of pop (includes k-pop, dance pop, hip pop, deep pop etc.) that has a count of 65. Pop is followed by the genre of rap and hip hop with a count of 24. Which is further followed by rock music with a count 20. Like in our first query, the data we find is consistent with the top listened to genres (rap, pop and rock) in the U.S. in 2019 (AAA Music) showing that are music the user's in our data listen is very common music. The data falls into normative of trending music genres.