Emily You and Mia Inakage
Dr. Barbara Ericson
26 April 2022

SI 206 Final Project Report: Python Girls

https://github.com/minakage/SI206PythonGirls

Original Goals for the Project

For our SI 206 project, our goal was to retrieve data using a Spotify API (Spotipy) and Wikipedia using Beautiful Soup. For Spotify, we wanted to get playlist data from one project member's Spotify account, specifically the user's top 5 playlists of 2022 with their playlist name, song names in each playlist, and the artist of each song in the playlist. Once we knew the artist of each song in the playlist, we planned to gather the number of monthly listeners for each of the artists on Spotify. Given the artist names gathered from the Spotify data, we wanted to retrieve the ages of those artists from Wikipedia. We planned to calculate a correlation between the age and monthly listeners for each respective artist.

We wanted to create 3 visualizations using matplotlib:

1. One bar graph to show the number of monthly listeners (in millions) for each artist in the user's top 5 playlists on Spotify
    a. X-axis would be artist name
    b. Y-axis would be monthly listeners (in millions)
2. One bar graph to show which age group is most common among the artists in the user's top 5 playlists on Spotify
    a. X-axis would be artist name
    b. Y-axis would be age group (grouped by 10 years)
3. One scatter plot to identify a correlation between an artist's age and their monthly listeners
    a. X-axis would be artist age
    b. Y-axis would be monthly listeners (in millions)

Our goal was to find out more information about the most commonly played artists in 2022 like which artists have the most monthly listeners and the ages of artists in the user's top 5 playlists of 2022.

New Goals and Goals that were Achieved

After reviewing our project plan, we realized our goals were a little too ambitious and that we realized that the Spotipy API would not allow us to find the number of monthly listeners that we originally intended to do. We also had difficulty with accessing a user's Spotify account and finding their top 5 playlists of 2022 while also satisfying the requirement for extracting at

least 100 data points for each API. Thus, we decided to select 10 artists and extract the popularity and duration of each song in their top 10 songs on Spotify.
10 Artists We Chose for Our Project

- Olivia Rodrigo
- Bazzi
- Billie Eilish
- Mac Miller
- Lady Gaga
- Harry Styles
- Post Malone
- Rihanna
- Elton John
- Madonna

The Spotipy API has a popularity index that rates songs on a scale of 0 to 100 and a song duration index that measures the length of each song in milliseconds. We also assigned each artist an artist id in order to avoid duplicate strings in our database. Our calculation resulted in the average popularity of the top 10 songs for each artist.

Instead of using Beautiful Soup to scrape data from Wikipedia, we decided to use a Genius API. With the same 100 songs we used from the Spotify data, we could find the page view count and annotation count for each song within the Genius API. We kept with our original plan of creating visualizations using matplotlib but changed the data that would populate in our graphs.

We ended up creating 4 visualizations using matplotlib:

4. One bar graph to show the average popularity of the top 10 songs for 10 selected artists on Spotify.
   a. X-axis is average popularity
   b. Y-axis is artist name
5. One bar graph to show the duration of each of the top 10 songs for 10 selected artists on Spotify
   a. X-axis is song title
   b. Y-axis is duration (in ms)
6. One bar graph to show the page view count of each of the 100 songs from Genius
   a. X-axis is song title and artist
   b. Y-axis is page views (in hundred thousands)
7. One bar graph to show the annotation count of each of the 100 songs from Genius
   a. X-axis is song title and artist
   b. Y-axis is annotation count

We arrived at a few new goals. For Spotify, we wanted to identify the average popularity and variation of song durations for the top ten songs for 10 artists on Spotify. For Genius, our goal was to see if there was a pattern in the number of page views or annotations from the lyrics of each song that we took from Spotify. We successfully achieved our goals because we were familiar with working with APIs in previous homework and discussion assignments. We also were reminded of how to create calculations and databases. One key lesson that we learned was how to access the Spotipy and Genius APIs and discover which indices we could find from them.

Problems Faced

We faced multiple problems while trying to achieve our new goals, one of which was having duplicate strings in our Spotipy and Spotipy 2 database tables. Specifically, since we gathered the top 10 songs from 10 artists, we would have the same artist for 10 rows in the 'artist' column of our Spotipy and Spotipy 2 database tables. For example, an artist (i.e. 'Olivia Rodrigo') would appear 10 times in the rows with a song name and popularity or song duration value. In order to eliminate the duplicate strings, we modified our tables so that each artist was given an artist_id (integer). Each table then had duplicate id values instead of duplicate strings. Another problem we faced was that the artist_id value could not be the primary key shared between the Spotipy and Spotipy 2 tables because artist_id had duplicate values. In order to ensure that both tables shared a primary key, we made the song title the primary key between the two tables and VARCHAR NOT NULL, which forces the data in the song title column to always contain a value. The song titles are unique to each row. We also had a problem with readjusting the size of the figure of our visualizations because the data was difficult to view. The titles and bars on the bar graph were often either too cramped together, or the titles were out of the frame and unable to be read. We found a solution to this problem by changing the figure size of the plot using fig = plt.figure(figsize=(#,#)) and adjust the spacing of the bottom using fig.subplots_adjust(bottom=spacing) to show the x-axis.

Git Commits

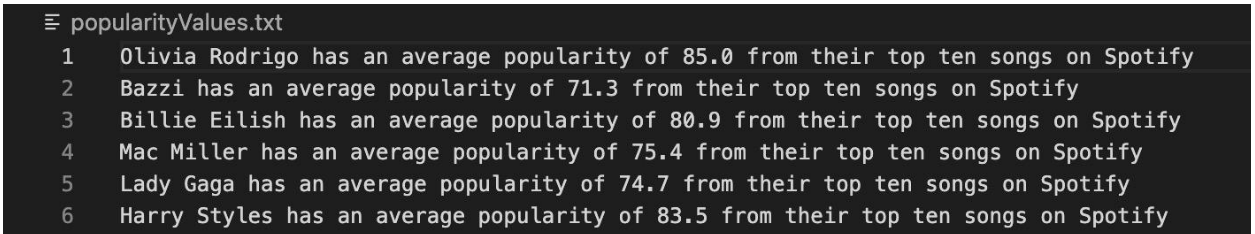
My partner and I decided not to create a collaborative Github to avoid running into the problem of git merge conflicts. Since we almost always worked on the project together in-person, we had Mia commit m

No collaborative Github → My partner and I work almost always on the project in person so one partner typically commits

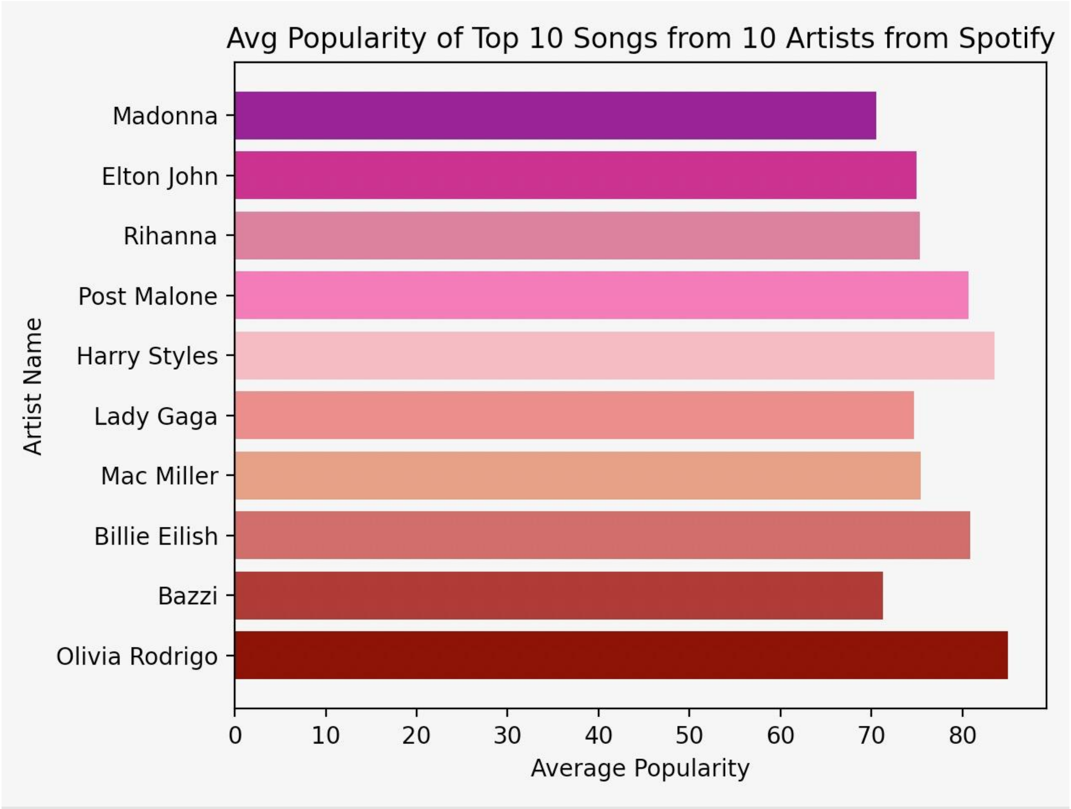
Explain that in your report. Otherwise only one of you will get the 15 points for the github commits.

Files Containing Calculations from Data

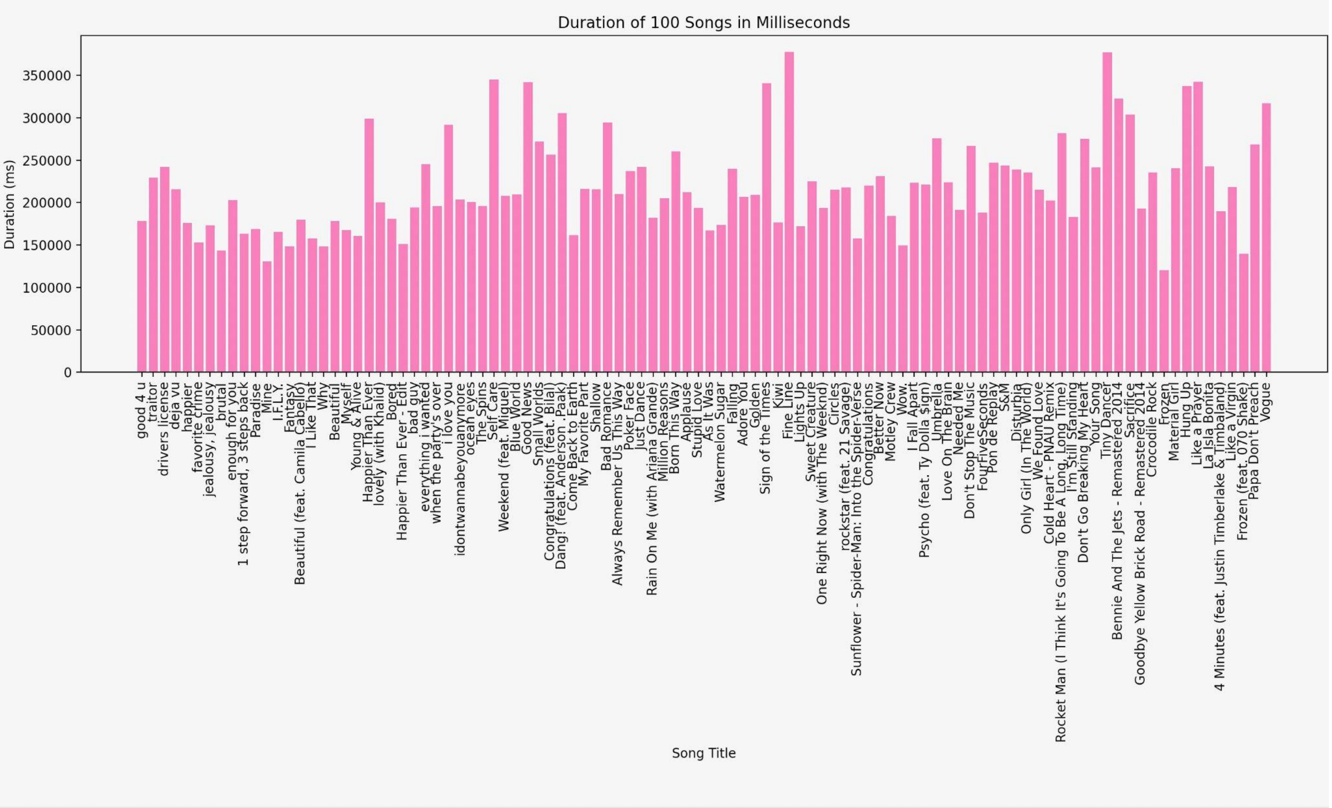
File containing our calculations from "spotipyPop.py" file: appears in popularityValues.txt

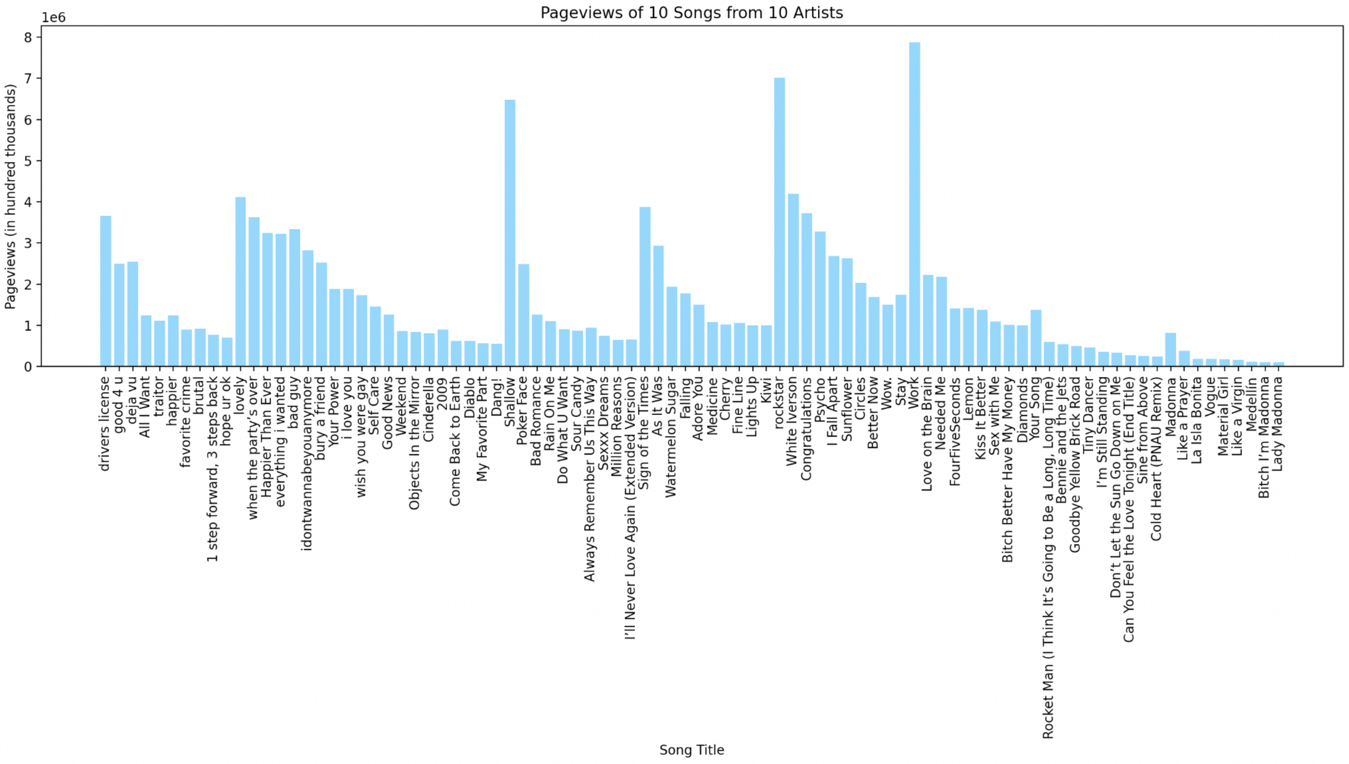


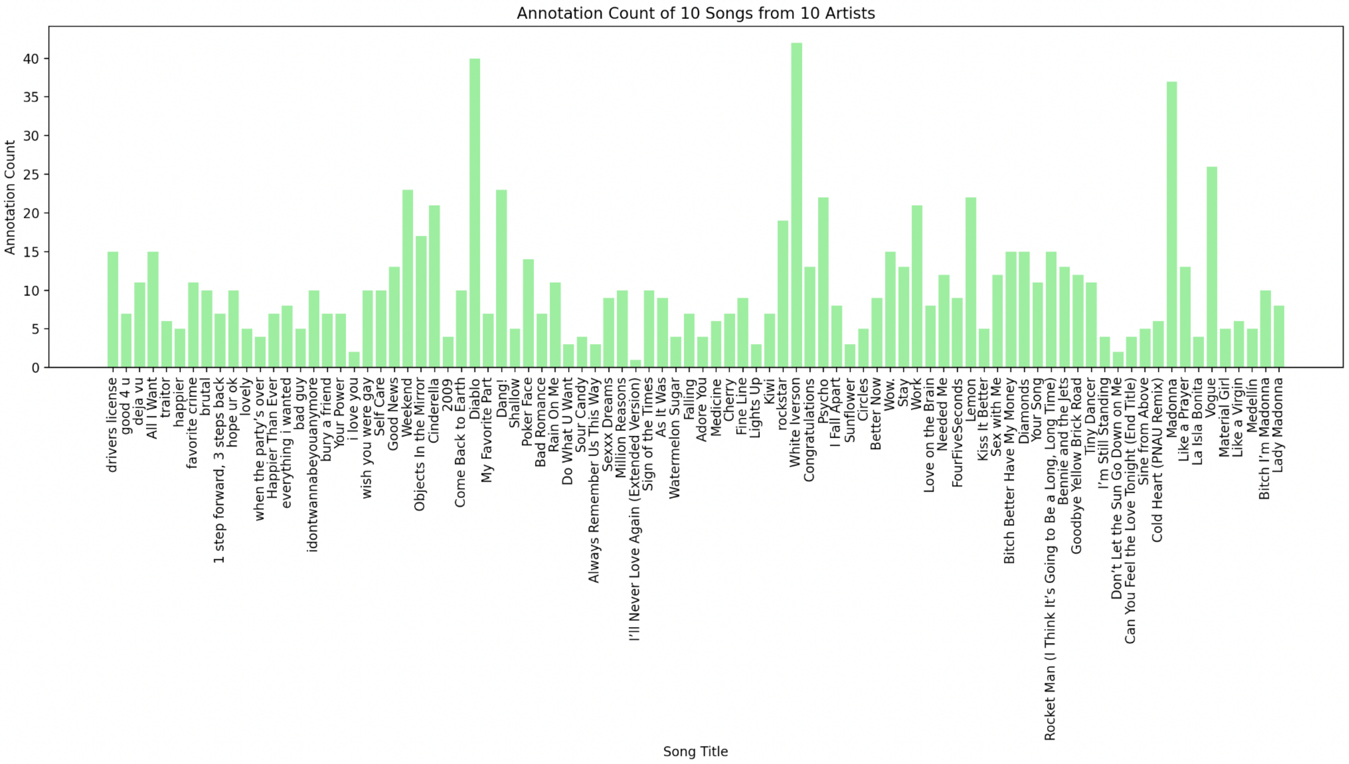
```
Olivia Rodrigo has an average popularity of 85.0 from their top ten songs on Spotify
Bazzi has an average popularity of 71.3 from their top ten songs on Spotify
Billie Eilish has an average popularity of 80.9 from their top ten songs on Spotify
Mac Miller has an average popularity of 75.4 from their top ten songs on Spotify
Lady Gaga has an average popularity of 74.7 from their top ten songs on Spotify
Harry Styles has an average popularity of 83.5 from their top ten songs on Spotify
```



Visualizations (4)


Duration of 100 Songs in Milliseconds
Duration (ms)
0
50000
100000
150000
200000
250000
300000
350000
Song Title
good 4 u
traitor
drivers license
deja vu
happier
favorite crime
jealousy, jealousy
brutal
enough for you
1 step forward, 3 steps back
Paradise
Mine
I.F.L.Y.
Fantasy
Beautiful (feat. Camila Cabello)
I Like That
Why
Beautiful
Myself
Young & Alive
Happier Than Ever
lovely (with Khalid)
Bored
Happier Than Ever - Edit
bad guy
everything i wanted
when the party's over
i love you
idontwannabeyouanymore
ocean eyes
The Spins
Self Care
Weekend (feat. Miguel)
Blue World
Good News
Small Worlds
Congratulations (feat. Bilal)
Dang! (feat. Anderson .Paak)
Come Back to Earth
My Favorite Part
Shallow
Bad Romance
Always Remember Us This Way
Poker Face
Just Dance
Rain On Me (with Ariana Grande)
Million Reasons
Born This Way
Applause
Stupid Love
As It Was
Watermelon Sugar
Falling
Adore You
Golden
Sign of the Times
Kiwi
Fine Line
Lights Up
Sweet Creature
One Right Now (with The Weeknd)
Circles
rockstar (feat. 21 Savage)
Sunflower - Spider-Man: Into the Spider-Verse
Congratulations
Better Now
Motley Crew
Wow.
I Fall Apart
Psycho (feat. Ty Dolla $ign)
Umbrella
Love On The Brain
Needed Me
Don't Stop The Music
FourFiveSeconds
Pon de Replay
S&M
Disturbia
Only Girl (In The World)
We Found Love
Cold Heart - PNAU Remix
Rocket Man (I Think It's Going To Be A Long, Long Time)
I'm Still Standing
Don't Go Breaking My Heart
Your Song
Tiny Dancer
Bennie And The Jets - Remastered 2014
Sacrifice
Goodbye Yellow Brick Road - Remastered 2014
Crocodile Rock
Frozen
Material Girl
Hung Up
Like a Prayer
La Isla Bonita
4 Minutes (feat. Justin Timberlake & Timbaland)
Like a Virgin
Frozen (feat. 070 Shake)
Papa Don't Preach
Vogue


Pageviews of 10 Songs from 10 Artists
1e6
Pageviews (in hundred thousands)
0
1
2
3
4
5
6
7
8
Song Title
drivers license
good 4 u
deja vu
All I Want
traitor
happier
favorite crime
brutal
1 step forward, 3 steps back
hope ur ok
lovely
when the party's over
Happier Than Ever
everything i wanted
bad guy
idontwannabeyouanymore
bury a friend
Your Power
i love you
wish you were gay
Self Care
Good News
Weekend
Objects In the Mirror
Cinderella
2009
Come Back to Earth
Diablo
My Favorite Part
Dang!
Shallow
Poker Face
Bad Romance
Rain On Me
Do What U Want
Sour Candy
Always Remember Us This Way
Sexxx Dreams
Million Reasons
I'll Never Love Again (Extended Version)
Sign of the Times
As It Was
Watermelon Sugar
Falling
Adore You
Medicine
Cherry
Fine Line
Lights Up
Kiwi
rockstar
White Iverson
Congratulations
Psycho
I Fall Apart
Sunflower
Circles
Better Now
Wow.
Stay
Work
Love on the Brain
Needed Me
FourFiveSeconds
Lemon
Kiss It Better
Sex with Me
Bitch Better Have My Money
Diamonds
Your Song
Rocket Man (I Think It's Going to Be a Long, Long Time)
Bennie and the Jets
Goodbye Yellow Brick Road
Tiny Dancer
I'm Still Standing
Don't Let the Sun Go Down on Me
Can You Feel the Love Tonight (End Title)
Sine from Above
Cold Heart (PNAU Remix)
Madonna
Like a Prayer
La Isla Bonita
Vogue
Material Girl
Like a Virgin
Medellín
Bitch I'm Madonna
Lady Madonna


Annotation Count of 10 Songs from 10 Artists
Annotation Count
0
5
10
15
20
25
30
35
40
Song Title
drivers license
good 4 u
deja vu
All I Want
traitor
happier
favorite crime
brutal
1 step forward, 3 steps back
hope ur ok
lovely
when the party's over
Happier Than Ever
everything i wanted
bad guy
idontwannabeyouanymore
bury a friend
Your Power
i love you
wish you were gay
Self Care
Good News
Weekend
Objects In the Mirror
Cinderella
2009
Come Back to Earth
Diablo
My Favorite Part
Dang!
Shallow
Poker Face
Bad Romance
Rain On Me
Do What U Want
Sour Candy
Always Remember Us This Way
Sexxx Dreams
Million Reasons
I'll Never Love Again (Extended Version)
Sign of the Times
As It Was
Watermelon Sugar
Falling
Adore You
Medicine
Cherry
Fine Line
Lights Up
Kiwi
rockstar
White Iverson
Congratulations
Psycho
I Fall Apart
Sunflower
Circles
Better Now
Wow.
Stay
Work
Love on the Brain
Needed Me
FourFiveSeconds
Lemon
Kiss It Better
Sex with Me
Bitch Better Have My Money
Diamonds
Your Song
Rocket Man (I Think It's Going to Be a Long, Long Time)
Bennie and the Jets
Goodbye Yellow Brick Road
Tiny Dancer
I'm Still Standing
Don't Let the Sun Go Down on Me
Can You Feel the Love Tonight (End Title)
Sine from Above
Cold Heart (PNAU Remix)
Madonna
Like a Prayer
La Isla Bonita
Vogue
Material Girl
Like a Virgin
Medellín
Bitch I'm Madonna
Lady Madonna

Instructions for Running the Code

Spotify:

- **Install spotipy by typing this in terminal** `pip install spotipy --upgrade`
- **Browse to https://developer.spotify.com/dashboard/applications.**
- **Log in with your Spotify account.**
- **Click on ‘Create an app’.**
- **Pick an ‘App name’ and ‘App description’ of your choice and mark the checkboxes.**
- **After creation, you see your ‘Client Id’ and you can click on ‘Show client secret` to unhide your ’Client secret’.**
- **Use your ‘Client id’ and ‘Client secret’ to retrieve a token from the Spotify API. The function get_spotify_api_token() performs all necessary steps with your ‘Client id’ and ‘Client secret’ to retrieve a token. Consider to assign this character string to a variable named ‘my_token’, which is the default value for all spotipy functions that are in need of a token.**
- **For the redirected uri go to the settings of the app you created an enter in** `http://localhost:8888/callback`
- **Go to spotipy.py and update your username, your client id, and your client secret**

Genius:

- **Go to this website https://genius.com/signup_or_login and sign up**
- **Once you’re signed in, you should be taken to https://genius.com/api-clients, where you need to click the button that says “New API Client.”**
- **After clicking “New API Client,” you’ll be prompted to fill out a short form about the “App” that you need the Genius API for. You only need to fill out “App Name” and “App Website URL.”**
- **You can simply put “Song Lyrics Project” for the “App Name” and the URL for our course website “https://melaniewalsh.github.io/Intro-Cultural-Analytics/” for the “App Website URL.”**
- **When you click “Save,” you’ll be given a series of API Keys: a “Client ID” and a “Client Secret.” To generate your “Client Access Token,” which is the API key that we’ll be using in this notebook, you need to click “Generate Access Token”.**
- **copy and paste your “Client Access Token” into genius.py**

Documentation: explain what each function does including its input and output

Function 1

- total_lists_songs_popularity
    - Doesnt take input
    - This function

```
total_list_songs_duration_ms()
   -   Doesnt take input
```

Resources

Date	**Issue Description**	**Location of Resource**	**Result**
04/13/22	Trouble with the Spotipy API	https://spotipy.readthedocs.io/en/2.19.0/	We were able to successfully use the Spotipy API and create functions to access top songs of artists we chose
04/14/22	Find Spotipy Client ID and Client Secret	https://cran.r-project.org/web/packages/spotidy/vignettes/Connecting-with-the-Spotify-API.html	We were able to sign up for Spotify Developer and input the Client Secret
04/15/22	Trouble with the Genius API	https://lyricsgenius.readthedocs.io/en/master/ https://melaniewalsh.github.io/Intro-Cultural-Analytics/04-Data-Collection/07-Genius-API.html https://docs.genius.com/#/getting-started-h1	We were able to successfully use the Genius API and get help with accessing the Client Access Token
04/17/22	Search for Spotify feature to scrape from Spotipy API	https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-tracks https://towardsdatascience.com/extracting-song-data-from-the-spotify-api-using-python-b	We were able to find two indices to scrape from Spotipy API: song popularity and song duration (in ms)

		1e79388d50	
04/18/22	How to create databases	file:///Users/emilyyou/Downloads/Databases-v7%20(1).pdf file:///Users/emilyyou/Downloads/DatabaseNormAndJoin-v4.pdf	We were able to follow examples from slides in order to get information into our databases for Spotify and Genius data
04/20/22	How to visualize data using Matplotlib	file:///Users/emilyyou/Downloads/Matplotlib-v7.pdf	We successfully plotted visualizations using matplotlib in Python files
04/24/22	How to add and change colors of matplotlib bar chart	https://matplotlib.org/3.5.0/gallery/color/named_colors.html	Customized colors using this website for each bar in all bar graphs