

206 Final Project Report

Caleigh Crossman and Ava Webster

REPO: <https://github.com/caleighc/Music-APIs>

1. The goals for your project (10 points)

The goal for our project was to grab data on the top 100 songs on Spotify, Apple Music, and Napster. We then wanted to use the data about the songs to make comparisons across the 3 platforms (Spotify, Apple Music, and Napster).

2. The goals that were achieved (10 points)

Initially, we were going to look at the top 100 songs on Spotify, Apple Music, and Napster, but we came across difficulties using the Apple Music and Napster APIs, so instead we used APIs from TheAudioDB and Genius, as well as Spotify, to look at top tracks from both artists and music charts. With TheAudioDB, I was able to look at the most popular songs for 10 different top artists. I was able to compare the popularity of artists' by calculating the average score that users give to their songs from their top 10 tracks. Additionally, I was able to see what genres were most popular and had the most songs in them. By analyzing/comparing the two charts, I was able to draw conclusions about what music types and artists are most popular. For example, I found that pop was the most popular genre but some artists whose music was in the pop genre had some of the lower averages.

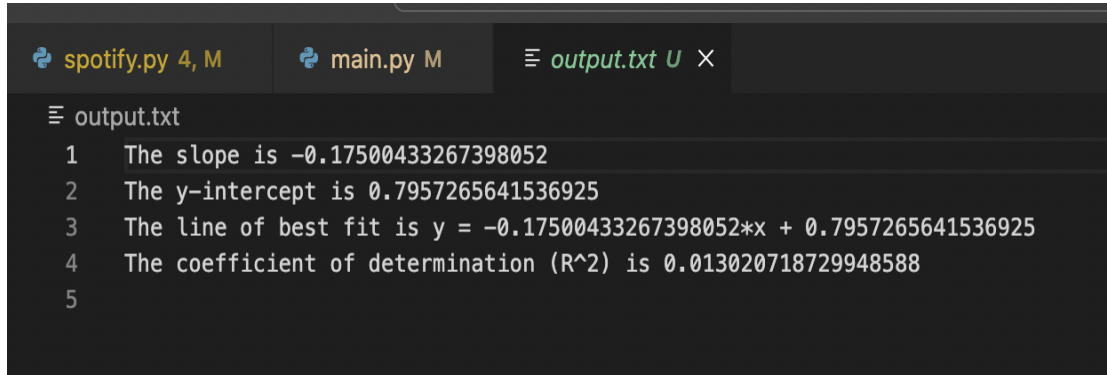
3. The problems that you faced (10 points)

A problem we faced was not being able to connect the data between our 3 APIs. Originally we had decided to use the Spotify API, Apple Music API, and Napster API. However, the Apple Music API cost \$99 to use and the Napster API didn't have the data we wanted. Therefore, we switched those APIs to the Genius API and TheAudioDB API. In the end, we were still not able to find exact data that was similar across the 3 APIs but we tried to keep our data cohesive by each looking at data on top tracks that were available on our APIs. Initially, I was going to look at the top 10 tracks by the top 10 artists (based on the Billboard Top Artists Chart), but TheAudioDB API did not have data on 10 songs for each of the top 10 artists, so I ended up looking at data for 10 artists within the top 100, but not exclusively the top 10 on the chart. Ultimately, I think this reflects better data because the Billboard charts are updated weekly, so I ended up using artists that are popular in general and not just popular for a short period of time. We also faced problems with unequal contributions to the coding portion. It made it very difficult for us to have a working cohesive project when one member wasn't pulling their weight.

4. Your file that contains the calculations from the data in the database (10 points)

The file that contains the calculations from the data for Spotify is in the database is output.txt. The coefficient of determination is 0.003 which means that there is very little

to no correction and the linear model does not fit the data. The slope of -0.079 indicates that there is a negative relationship between the danceability and energy scores of a song. Calculations will vary based on the number of songs in the database when the visualization function is run.

A screenshot of a code editor interface with a dark theme. At the top, there are three tabs: 'spotify.py 4, M' (active), 'main.py M', and 'output.txt U X'. Below the tabs, the 'output.txt' file is open, showing a list of five lines of text. The text provides linear regression results: the slope, the y-intercept, the equation of the line of best fit, and the coefficient of determination (R^2).

```
1 The slope is -0.17500433267398052
2 The y-intercept is 0.7957265641536925
3 The line of best fit is y = -0.17500433267398052*x + 0.7957265641536925
4 The coefficient of determination (R^2) is 0.013020718729948588
5
```

The calculations from the data gathered from TheAudioDB are in the calculations1.csv and calculations2.csv files. Users are able to assign a score to a song, so the calculations1.csv file shows the average score of the top 10 songs from 10 top artists. Additionally, I did a second calculation for a second visualization for extra credit where I calculated the genres of the 100 songs from my data and got the count for how many songs were in each genre. This is shown in the calculations2.csv file.

```

calculations1.csv
1  Artist Name,Average Score
2  Ariana Grande,9.63
3  Drake,10.0
4  Ed Sheeran,9.72
5  Justin Bieber,10.0
6  Kendrick Lamar,10.0
7  Metallica,8.99
8  Michael Jackson,8.95
9  Rihanna,8.62
10 Taylor Swift,8.64
11 The Weeknd,9.96
12 |

```

```

calculations2.csv
1  Genre,Number of Songs in Genre
2  Country,8
3  Hip-Hop,20
4  Pop,52
5  R&B,10
6  Thrash Metal,10
7

```

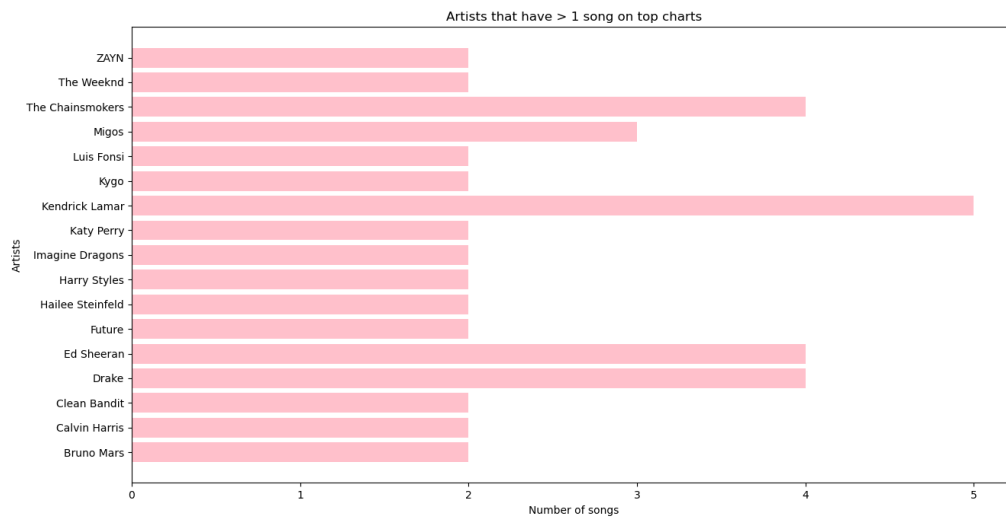
5. The visualization that you created (i.e. screenshot or image file) (10 points)

This bar chart shows the artists who have more than one song on the top charts on the y-axis and the number of songs they have on the top charts on the x-axis. I made this visualization in order to see if artists had multiple songs on the top charts and how many songs they had. I was curious to see if popular artists (artists on the top charts) typically had one top hit or multiple popular songs. I was able to get the correct data from my tables using the query

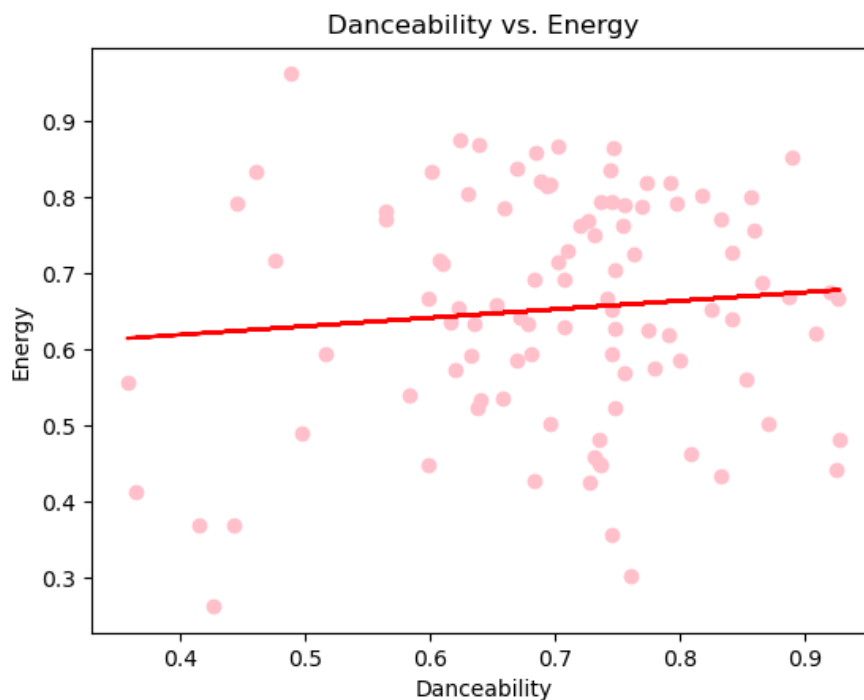
```

SELECT artists.artist, COUNT(*)
FROM songs
JOIN artists ON artists.artist_id = songs.artist_id
GROUP BY artist
HAVING COUNT(*) > 1

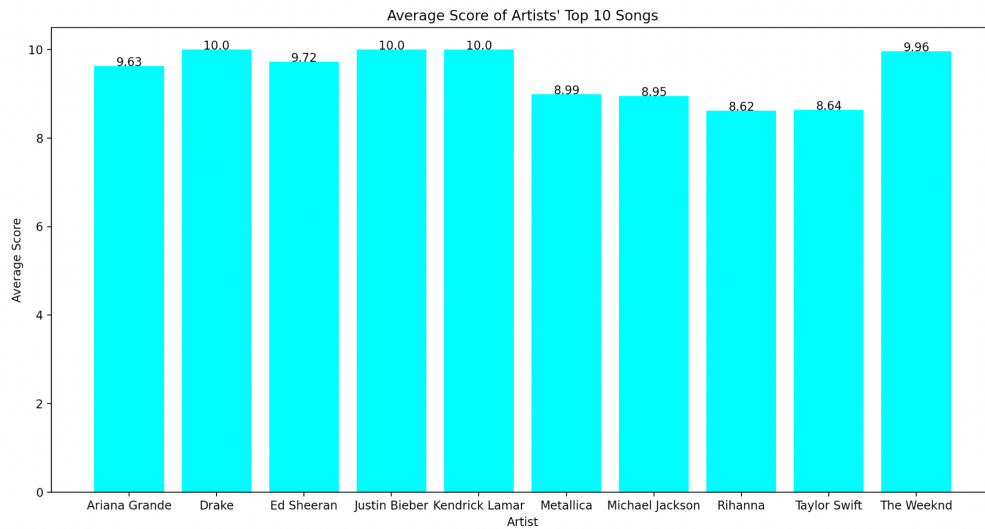
```



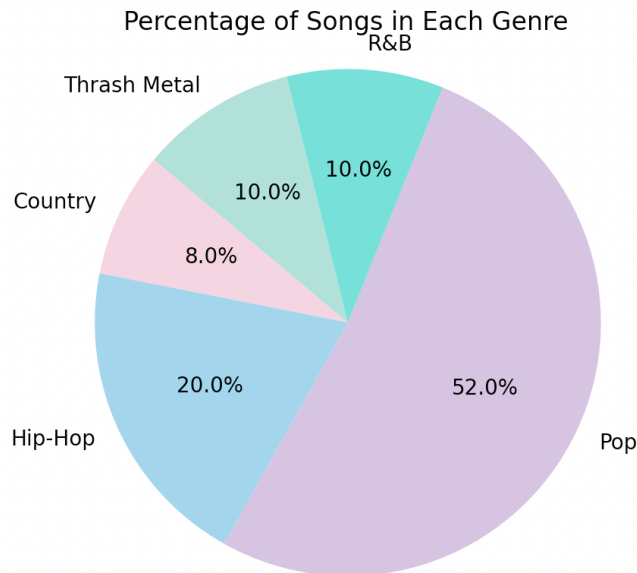
This scatter plot shows the energy score on the y-axis and the danceability score on the x-axis for each of the songs in my table. I made this scatter plot to analyze the relationship between danceability and energy for a song. In addition, I calculated and graphed the line of best fit for this relationship. As seen, there is not a linear relationship between danceability and energy. I was able to get the data from my table to create the visualization below using the query
 SELECT danceability,energy FROM songs



The first visualization I created was a bar graph that displays the average score for the top 10 tracks for each of the 10 artists, which represents not only an artists' average rating, but also shows how their scores compare to each other.



The second visualization I created is a pie chart that represents the percentage of the 100 songs that belong in each genre.



6. Instructions for running your code (10 points)

- Run main.py 4 times and then 2 spotify visualizations will show up behind one another
- Then in main.py in main() function comment out run_spotify(cur,conn), uncomment run_audio(cur,conn), and run main.py 4 times to get all 100 rows of data in (25 rows inserted each time code runs). By the end of the 4 times, both completed visualizations should pop up

7. Documentation for each function that you wrote. This includes the input and output for each function (20 points)

Name	Function	Description	Input	Output
Caleigh	set_up_db(db_name)	This function in main takes the name of the database and creates it	Name of the database (music)	returns cur, conn
Caleigh	run_spotify(cur,conn)	This function runs the functions in spotify.py	cur,conn	None
Caleigh	main()	Main function that calls set_up_db, run_spotify, run_audio, and run_genius	None	None
Caleigh	make_requests(playlist_id)	Takes in a spotify playlist ID and makes a request to the API to return the playlist JSON data	Playlist_id (the list of playlist ids are in run_spotify)	playlist which is a dictionary that contains the data about the tracks
Caleigh	create_artists_table(cur,conn)	Creates the Artists table in the music database	cur,conn	None
Caleigh	create_songs_table(cur,conn)	Creates the Songs table in the music database	cur,conn	None
Caleigh	add_artists_id(data, cur,conn)	Adds the artist names and artist_ids to the Artists table in the database	data,cur,conn (data is the dictionary that is returned from make_requests)	None
Caleigh	add_songs(data,cur, conn)	Adds each songs and song data to the Songs table in the database	data,cur,conn (data is the dictionary of song data that is returned from make_requests)	None
Caleigh	danceability_vs_energy_plot(cur,conn)	Uses a query to the database to create a	cur,conn	Visualization

		scatter plot of the danceability vs energy scores of each song in the Songs table		
Caleigh	artists_visualization(cur,conn)	Uses a query to the database to create a horizontal bar plot of the number of songs each artists has on the top charts	cur,conn	Visualization
Ava	request_data(artist_ids)	Requests data from API and parses through JSON object using regex to get the desired data	artist_ids - list of artist ids to use in url	None
Ava	create_table(cur, conn)	Creates the SongsAudiodb and ArtistsAudiodb tables	cur, conn	None
Ava	insert_data(cur, conn)	Inserts the data into the SongsAudiodb table (25 rows each time program runs) and into the ArtistsAudiodb table (only has 10 items)	cur, conn	None
Ava	calculate1(cur, conn)	Does the first calculation which calculates the average score of each artist based on the scores for each of their songs	cur, conn	Returns result, which is a list of tuples with the artist and their average score
Ava	calculate2(cur, conn)	Does the second calculation which calculates the number of songs in each genre	cur, conn	Returns result, which is a list of tuples with the genre and number of songs in that genre
Ava	write_to_csv1(result)	Writes the results of the first calculation to a CSV file	result (from calculate1)	None
Ava	write_to_csv2(result)	Writes the results of the second calculation to a CSV file	result (from calculate2)	None
Ava	make_chart1(result)	Makes first visualization (bar graph) of first	Result (from calculate1)	visualization

		calculation		
Ava	addLabels(x, y)	Adds labels to bar graph to be used to show average score in first visualization	x, y - x and y axis used in graph	None
Ava	make_chart2()	Makes second visualization (pie chart) of second calculation	None	visualization
Ava	run_audio	In main.py, this function runs all of the above functions from my audio.py file	cur.conn	None (visualizations/r results from functions within this function are output)

8. You must also clearly document all resources you used. The documentation should be of the following form (20 points)

Name	Date	Issue Description	Location of Resource	Result (did it solve the issue?)
Caleigh	11/29/2022	Having trouble calling functions in main to create table in database	Homework 6 & 7	Yes, I was able to figure out how to do it properly
Caleigh	12/4/2022	Was having trouble using matplotlib	Discussion 12	Yes, I was able to see how to properly use matplotlib to create visualizations
Caleigh	12/4/2022	Client credentials flow	https://github.com/spotify/web-api-auth-examples/blob/master/client_credentials/app.js	Yes, I was able to figure out the flow for the client credentials
Caleigh	12/9/2022	Calculations	https://numpy.org/	Yes, I was able to import numpy and use it to help me make calculations
Caleigh	12/3/2022	Horizontal bar plot and scale on x and y axis	https://matplotlib.org/stable/api/pyplot_summary.html	Yes, I was able to figure out how to make the plot and change the scale
Ava	12/08/2022	Accessing data in the	Regex	Yes, I was able to use

		json object was not easy to do via indexing; found best solution was regex	summary/glossary on Runestone	regex expressions to extract the data i wanted from the json object
Ava	12/08/2022	Formatting the API request	https://www.theaudio.db.com/api_guide.php	Yes, this webpage for my API provided documentation get requests and how to request specific data
Ava	12/08/2022	Couldn't find artist ID to use in url request	https://musicbrainz.org/doc/MusicBrainz_API	Yes, i was able to search this website to get the artist id that would work in the url request
Ava	12/11/2022	Writing calculations to csv file	https://tutorial.eyehunters.com/python/python-on-file-modes-open-write-append-r-r-w-w-x-etc/ https://www.pythontutorial.net/python-basics/python-write-csv-file/	Yes, using these resources I was able to write my calculations to a CSV file
Ava	12/10/2022	Adding labels to bar graph	https://www.geeksforgeeks.org/adding-value-labels-on-a-matplotlib-bar-chart/	Yes, I was able to create an addlabels function to add labels to my bar graph