HW8: Database Joins & Matplotlib

At the end of every year, Spotify releases "Spotify Wrapped" - a personalized data story of users' listening habits throughout the year. In this homework, you'll analyze Aadarsh's Spotify listening history to create his own Wrapped-style analysis for 2024.

We have provided his listening data and starter code:
- ***2024_listening_data.db*** - a database with Aadarsh's Spotify listening data for 2024.
- ***HW8.py*** - starter code for the functions below.

Make sure you are using Anaconda ("base":conda) python for this assignment (preferred), which already has `matplotlib`. But if you have to install `matplotlib` on your own, you can use `pip3 install matplotlib` or `pip install matplotlib` or another installation method. We have also provided test cases that will pass if the functions are written correctly. You should not edit these test cases.

Note: It is okay for the extra credit test case to fail if you do not attempt the extra credit; you can also comment out those specific test cases.

Before you start: Look at the database

Check out *2024_listening_data.db* in your DB Browser for SQLite program.
1. Open DB Browser for SQLite
2. Click on "Open Database" and choose *2024_listening_data.db*.
3. Click on Browse Data
4. Take some time to familiarize yourself with the table and column names.

Table: artists

id	name
1	Uffie
2	Alex Chapman
3	Addison Rae
4	The Weeknd
5	Justice
6	Charli XCX
7	Charli xcx

Table: albums

id	name	artist_id
1	Sex Dreams and Denim Jeans	1
2	Stay The Weekend (feat. Uffie)	2
3	Aquamarine	3
4	São Paulo	4
5	Tokyo Love Hotel	1
6	† (Anniversary Edition)	5
7	Number 1 Angel	6
8	Brat and it's completely different but also still brat	7
9	Brat and it's completely different but also still brat	6

	id	name	album_id
	Filter	Filter	Filter
1	1	ADD SUV (feat. Pharrell Williams)	1
2	2	Ricky	1
3	3	Pop the Glock	1
4	4	Stay The Weekend (feat. Uffie)	2
5	5	Aquamarine	3
6	6	São Paulo (feat. Anitta)	4

Table: tracks

	id	track_id	timestamp	utc_time
	Filter	Filter	Filter	Filter
1	1	1	1731520932	13 Nov 2024, 18:02
2	2	2	1731510349	13 Nov 2024, 15:05
3	3	1	1731510145	13 Nov 2024, 15:02
4	4	3	1731509621	13 Nov 2024, 14:53
5	5	4	1731509462	13 Nov 2024, 14:51
6	6	2	1731509213	13 Nov 2024, 14:46

Table: listening_history

The database contains four tables:

1. `artists`:
    - Columns: id, name
    - Stores unique artists
    - Note: artist names are unique
2. `albums`:
    - Columns: id, name, artist_id
    - Stores unique albums
    - Foreign key: artist_id references artists(id)
    - Note: combination of (name, artist_id) must be unique
3. `tracks`:
    - Columns: id, name, album_id
    - Stores unique tracks
    - Foreign key: album_id references albums(id)
4. `listening_history`:
    - Columns: id, track_id, timestamp, utc_time
    - Stores each time a track was played
    - Foreign key: track_id references tracks(id)

Part 1: Data Cleanup

The artist "Charli xcx" recently changed how her name appears on Spotify. This means, some tracks show "Charli XCX" while others show "Charli xcx". Complete the `fix_charli_xcx_name()` function that standardizes her name to "Charli xcx" across the database. Since artists and album/artist combinations must be unique, we need to follow the following steps:
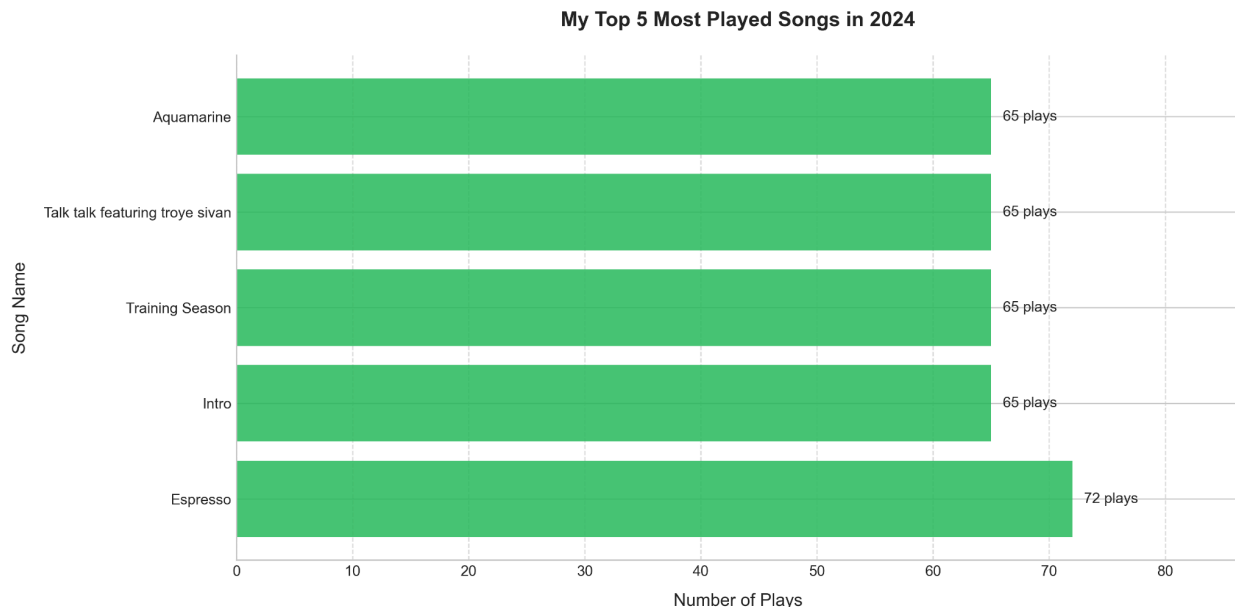
1. Fetch all the Albums for "Charli XCX"
2. For each album:
    a. Check if the album already exists for "Charli xcx"

i. If it does exist, update the tracks album id's to point to the existing album, then delete the Charli XCX version of the album

ii. If the album doesn't exist, just update its artist_id to Charli xcx's artist id.

3. Finally, delete the Charli XCX artist entry

Part 2: Aadarsh's Top 5 Songs for 2024

Complete the `get_top_songs()` function that returns Aadarsh's most-listened-to songs and their play counts. The function should return a dictionary of his top 5 most-played songs and their play counts. The function should also create a bar chart (horizontal or vertical – figure out which one gives a better visualization) with song names along one axis and the corresponding play counts along the other axis. In the chart, the play counts should be in descending order.

Make sure to commit and push your bar chart to your github repo.

My Top 5 Most Played Songs in 2024

Part 3: Aadarsh's Top 5 Artists for 2024

Complete the `get_top_artists()` function that returns Aadarsh's most-listened-to artists and their play counts. The function should return a dictionary of his top 5 most-played artists and their play counts.

Example output:
```
{'Charli xcx': 1657, 'Beyoncé': 847, 'Tyla': 497, 'Caroline
Polachek': 362, 'Dua Lipa': 356}
```

Extra Credit: Find Featured Artists of Charli xcx

Complete the `find_featured_artists()` function that finds all of Charli xcx's songs with "feat." or "featuring" in the title and returns a list of the artists Charli xcx collaborated with. Be careful of the following cases:

1. Charli has collaborated with artists multiple times, the final list should not have duplicates.
2. Some artists are stylized with both uppercase and lowercase, for example, Troye Sivan. Troye Sivan should only be in the list once using the upper case names.
3. A.G. Cook is both in her features as "A. G. Cook" and "A.G. Cook". You should only account for one of these.
4. "Christine and the Queens" is one artist, not two different artists.

Example output:
```
['A. G. Cook', 'Abra', 'Addison Rae', 'Ariana Grande', 'Bb Trickz',
'Benga', 'Big Freedia', 'Billie Eilish', 'Bladee', 'Bon Iver',
'Brooke Candy', 'Carly Rae Jepsen', 'Caroline Polachek', 'Christine
and the Queens', 'Clairo', 'Cupcakke', 'Dorian Electra', 'Haim', 'Jay
Park', 'Jon Hopkins', 'Julian Casablancas', 'Kesha', 'Kim Petras',
'Lizzo', 'Lorde', 'Mykki Blanco', 'Pabllo Vittar', 'Rina Sawayama',
'Robyn', 'Shygirl', 'Skream', 'Sky Ferreira', 'The 1975', 'The
Japanese House', 'Tinashe', 'Tommy Cash', 'Troye Sivan', 'Uffie',
'Yaeji', 'Yung Lean']
```

Grading

Note: Test cases have been provided for you.

`fix_charli_xcx_name(db)`	15 pts
`get_top_songs(db)`	10 pts
`get_top_artists(db)`	10 pts
Submitted the bar chart image file	5 pts
Bars are ordered in descending order	5 pts
Title on the bar chart	5 pts
Informative X-axis label on bar chart	5 pts

Informative Y-axis label on bar chart	5 pts
`find_featured_artists(db)`	*6 pts extra credit*
Total	**60 pts + 6 pts extra credit**