

# 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 2025 thus far.

We have provided his listening data and starter code:

- **2025\_listening\_data.db** - a database with Aadarsh's Spotify listening data for 2025.
- **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 `2025_listening_data.db` in your DB Browser for SQLite program.

1. Open DB Browser for SQLite
2. Click on "Open Database" and choose `2025_listening_data.db`.
3. Click on Browse Data
4. Take some time to familiarize yourself with the table and column names.

	<u>id</u>	name	<u>album_id</u>
	Fil...	Filter	Filter
1	1	Wifey Riddim	1
2	2	HOTBOX	2
3	3	Room of Fools	3
4	4	Drums of Death	3
5	5	Perfect Stranger	3
6	6	Physical (feat. Troye Sivan)	4
7	7	Dolphin (feat. Shygirl)	5
8	8	Mosquito	6

	<u>id</u>	name
	Fil...	Filter
1	1	Shygirl
2	2	Lil Nas X
3	3	FKA twigs
4	4	Dua Lipa
5	5	Earthater
6	6	PinkPantheress

	<u>id</u>	name	artist_id
	Fil...	Filter	Filter
1	1	Wifey Riddim	1
2	2	Hotbox	2
3	3	EUSEXUA	3
4	4	Physical (feat. Troye Sivan)	4
5	5	Shark Brain / Dolphin	5
6	6	Heaven knows	6
7	7	HIT ME HARD AND SOFT	7

	<u>id</u>	<u>track_id</u>	timestamp	utc_time
	Filter	Filter	Filter	Filter
1	1	1	1743428130	31 Mar 2025, 13:35
2	2	2	1743427951	31 Mar 2025, 13:32
3	3	3	1743427689	31 Mar 2025, 13:28
4	4	4	1743427497	31 Mar 2025, 13:24
5	5	5	1743427300	31 Mar 2025, 13:21
6	6	6	1743427110	31 Mar 2025, 13:18

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

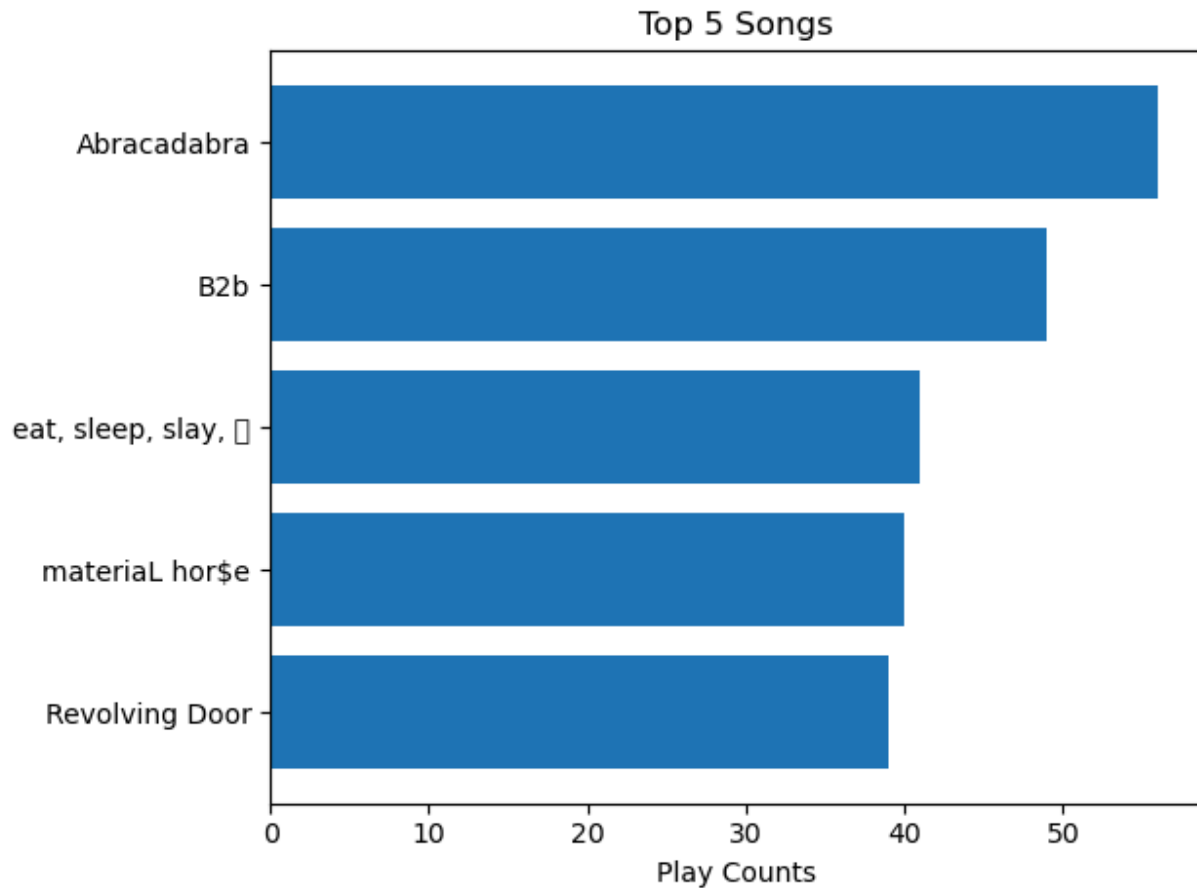
## Part 2: Aadarsh's Top 5 Songs for 2025

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.

**Group the data by both track name and artist to ensure accurate play counts.** This step accounts for different album versions of a song and distinguishes identical titles by different artists.

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

\*If you get a UserWarning: Glyph when trying to output your bar chart, don't worry about it. If the test case fails, it won't be because of this warning.



### Part 3: Aadarsh's Top 5 Artists for 2025

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:**

```
{  
    'Bad Bunny': 236,  
    'Lady Gaga': 225,  
    'Tate McRae': 224,  
    'FKA twigs': 186,  
    'Charli xcx': 181  
}
```

## 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.

### Example output:

```
['A. G. Cook', 'Addison Rae', 'Ariana Grande', 'Bb Trickz', 'Big  
Freedia', 'Billie Eilish', 'Bladee', 'Bon Iver', 'Brooke Candy',  
'Caroline Polachek', 'Clairo', 'Cupcakke', 'Haim', 'Jon Hopkins',  
'Julian Casablancas', 'Kesha', 'Kim Petras', 'Lizzo', 'Lorde',  
'Pablllo Vittar', 'Robyn', 'Shygirl', 'Sky Ferreira', 'The 1975', 'The  
Japanese House', 'Tinashe', 'Tommy Cash', 'Troye Sivan', 'Yaeji',  
'Yung Lean']
```

## Grading

Note: Test cases have been provided for you.

<code>fix_charli_xcx_name(db)</code>	15 pts
<code>get_top_songs(db)</code>	10 pts
<code>get_top_artists(db)</code>	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
<code>find_featured_artists(db)</code>	6 pts extra credit
<b>Total</b>	<b>60 pts + 6 pts extra credit</b>