

# HW 10

In this homework, you will select data from a database, process it, and create a visualization using Matplotlib. This is similar to the final steps of your pipeline for the final project.

We have provided:

- *restaurants.db* - a database with local restaurant data collected from Yelp.
- HW10.py - starter code for the functions below.

Make sure you are using Anaconda python for this assignment (preferred), or have installed Matplotlib on your own (using `pip install matplotlib` or another installation method).

[Part 1: Look at the database](#)

[Part 2: Process the data](#)

[Part 3: Visualize the data](#)

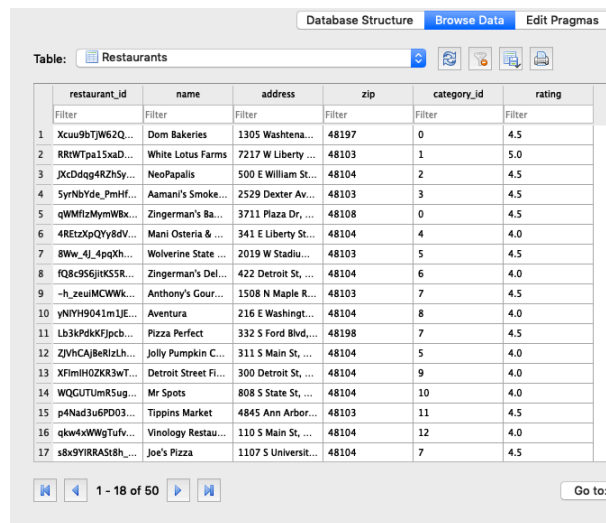
[Extra credit: Visualize other data](#)

[Grading rubric](#)

## Part 1: Look at the database

Check out *restaurants.db* in your DB Browser for SQLite program.

1. Open DB Browser for SQLite
2. Click on “Open Database” and choose *restaurants.db*.
3. Click on Browse Data
4. Take some time to familiarize yourself with the table and column names



The screenshot shows the DB Browser for SQLite application. At the top, there are tabs for 'Database Structure', 'Browse Data', and 'Edit Pragma'. The 'Browse Data' tab is active. Below the tabs, there is a section for 'Table: Restaurants' with a dropdown menu and icons for filtering, sorting, and other table actions. The main area displays a table with 6 columns: restaurant\_id, name, address, zip, category\_id, and rating. The table contains 17 rows of data, each representing a restaurant. The first row is filtered out, and the second row is highlighted. The bottom of the window shows a status bar with '1 - 18 of 50' and a 'Go to:' field.

	restaurant_id	name	address	zip	category_id	rating
Filter	Filter	Filter	Filter	Filter	Filter	
1	Xcuu9bTJW6ZQ...	Dom Bakeries	1305 Washtena...	48197	0	4.5
2	RRRWTPa1SxaD...	White Lotus Farms	7217 W Liberty ...	48103	1	5.0
3	JXcDdgg4RZh5y...	NeoPapalis	500 E William St...	48104	2	4.5
4	SyrNbYde_PmHf...	Aamani's Smoke...	2529 Dexter Av...	48103	3	4.5
5	qWMfzMyMWBx...	Zingerman's Ba...	3711 Plaza Dr, ...	48108	0	4.5
6	4REtzXpQYy8dV...	Mani Osteria & ...	341 E Liberty St...	48104	4	4.0
7	8Ww_4l_4pqKh...	Wolverine State ...	2019 W Stadiu...	48103	5	4.5
8	lQ8c956jK5SR...	Zingerman's Del...	422 Detroit St, ...	48104	6	4.0
9	-h_zeuMCWWk...	Anthony's Gour...	1508 N Maple R...	48103	7	4.5
10	yNYH9041mLJE...	Aventura	216 E Washingt...	48104	8	4.0
11	Lb3kPdkFJpcb...	Pizza Perfect	332 S Ford Blvd...	48198	7	4.5
12	ZjVhCAj8eRizLh...	Jolly Pumpkin C...	311 S Main St, ...	48104	5	4.0
13	XfImH0ZKR3wT...	Detroit Street Fl...	300 Detroit St, ...	48104	9	4.0
14	WQGUUmRSug...	Mr Spots	808 S State St, ...	48104	10	4.0
15	p4Nad3u6PD03...	Tipkins Market	4845 Ann Arbor...	48103	11	4.5
16	qkw4xWWgTufv...	Vinology Restau...	110 S Main St, ...	48104	12	4.0
17	s8x9YIRRA5t8h...	Joe's Pizza	1107 S Universit...	48104	7	4.5

## Part 2: Process the data

Complete the `get_category_dict(..)` function that accepts the filename of the database as a parameter, and returns a dictionary with the number of restaurants that are in each category. The keys should be the category names and the values should be the number of restaurants in that category. The dictionary should look like:

```
{'American (New)': 1, 'American (Traditional)': 1, 'Bakeries': 2...}
```

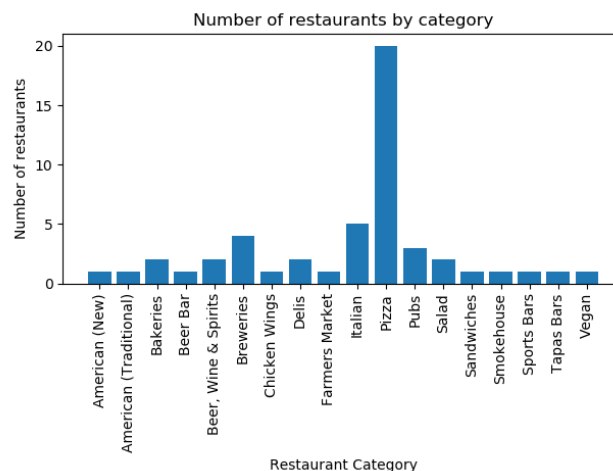
(HINT: If you did the extra credit for HW9, you can modify and re-use that code.)

Your function must pass all the unit tests to get full credit.

## Part 3: Visualize the data

Complete the function `barchart_restaurants_by_cat(..)`, which takes in a dictionary created by the function in Part 1, and uses matplotlib functions to draw a bar chart with restaurant categories on the x axis, and the number of restaurants in that category on the y axis. The chart must have appropriate axis labels and a title.

Sort the X axis alphabetically from left-to-right. Your chart should look like this:



Finally, this function should sort the dictionary items alphabetically and return the resulting list of tuples. Your list should look like:

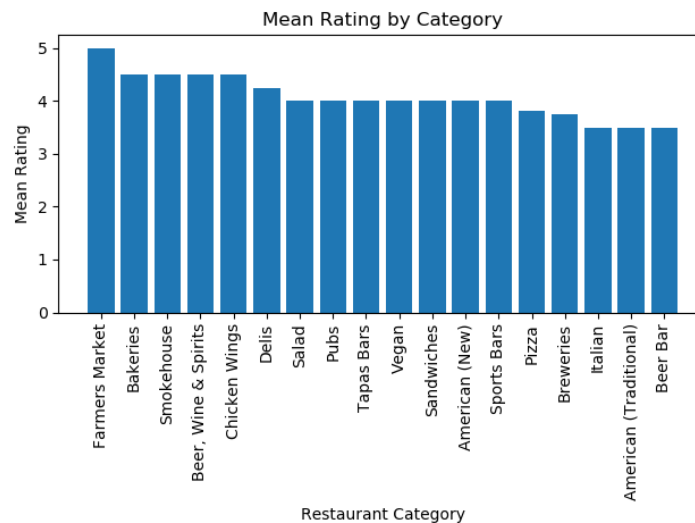
```
[('American (New)', 1), ('American (Traditional)', 1), ('Bakeries', 2) ...]
```

Submit an image file of your bar chart to Canvas, along with your repository link.

## Extra credit: Visualize more data

Do different restaurant categories tend to get higher ratings? Let's make another chart and see!

Complete function `mean_rating_by_category(..)` to plot a barchart of the mean rating for each restaurant category. Put the restaurant on the x axis and the average rating on the y axis. Sort the x axis in descending order from left-to-right. The chart must have appropriate axis labels and a title. Your chart should look like this:



Finally, this function should return a dictionary where the keys are restaurant categories and the values are the mean ratings for the corresponding category. Your dictionary should look like this:

```
{'American (New)': 4.0, 'American (Traditional)': 3.5, 'Bakeries': 4.5...}
```

Submit an image file of your chart to Canvas.

## Grading

Code passes all unit tests	14 pts (2 pts per unit test with 7 tests)
Submission of bar chart image file	5 pts
Created a bar chart from the data	26 pts
Title on bar chart	5 pts

Informative X-axis label on bar chart	5 pts
Informative Y-axis label on bar chart	5 pts
<i>Correct code and image file for mean ratings</i>	<i>3 pts extra credit</i>
<b>Total</b>	<b>60 pts + 3 pts extra credit</b>