Please **read this entire assignment**, **every word**, before you start working on the code. Several of the questions build off of previous questions.

**Turn in this document with your SQL code and screenshots. It can be either a MS Word file (doc or docx) or a pdf file. Edit your copy of this assignment and place your answers into your copy. Submit your completed copy into Canvas.**

**DUE: October 19, 2023.**

This assignment makes use of data generously provided by Yelp about donut shops located in the United States. Yelp provided a sample of their data for use to use. Do not use this data outside of this class.

**All questions are worth 5 points**. Questions 1 and 2 a free 5 points each.

**This is an individual assignment.**

1. Run the SQL script f23-Lab1-tables.sql. This will create a schema and a table within that schema against which we will run some queries. You’ll find this file as part of the assignment in Canvas.
2. Run the SQL script f23-Lab1-data.sql. This will load 5,000 rows of data into the table you created in step 1. This may take up to a minute to complete. It took about 50 seconds to load on the PSU server from off-campus for me. You’ll find this file as part of the assignment in Canvas.
3. Validate that the data correctly loaded into your table. Run the following SQL query. The count() aggregate function is very commonly used in cases like this. Add a screenshot into your document showing the result of the query.

select count(\*) from yelp\_donut.donut\_data;

1. Write a SQL query that shows **only** the first 10 rows of data from the table. Use the wildcard character to select all columns. Use the LIMIT clause. Do not order the data. Place your query and a screenshot of the 10 rows of data below.
2. Modify the query from question 4, but order the data by city. What seems odd about the first 2 rows of the result set? Answer that and place your SQL and a screenshot of the first 10 rows of the result set below.
3. Modify the query from question 5, but using the OFFSET clause, jump 1000 rows ahead in the data. Place your query and a screenshot of the 10 rows of data below.

**In the following queries, only select the following columns:**

**name, city, county, state, country, rating, review\_count, price\_indicator**

1. Write a SQL statement that shows only the donut shops that are in the state of Oregon. Place your SQL and the first 5 rows of data (as a screenshot) below.
2. Write a SQL query to answer the following question: Which donut shops in Oregon has the most number of reviews? Sort by review\_count. Place your SQL and the first 5 rows of data (as a screenshot) below.
3. Modify your query from 8 to show the donut shops in Oregon with the fewest number of reviews. Sort by review\_count. Place your SQL and the first 5 rows of data (as a screenshot) below.
4. Modify your query from 8 to show the donut shops in Oregon with at least 100 reviews. Sort by review\_count. Place your SQL and the first 5 rows of data (as a screenshot) below.
5. Modify your query from 10 to show the donut shops in Oregon with the highest rating. Place your SQL and the first 5 rows of data (as a screenshot) below.
6. Write a SQL query that only shows donut shops in a city named Portland. Don’t order the data. Place your SQL and the first 5 rows of data (as a screenshot) below.
7. Write a SQL query that only shows donut shops in Portland, Oregon. Don’t order the data. Place your SQL and the first 5 rows of data (as a screenshot) below.

**You’ll need to change the returned columns for the following questions 14 and 15.**

1. Write a SQL query that shows the distinct states that have a donut shop in this data. Order the result by state. Place your SQL and the first 5 rows of data (as a screenshot) below.
2. Modify the query from 14 the show the distinct states that have a donut shop in a city named Portland. Place your SQL and the rows of data (as a screenshot) below.

**You’ll need to change the returned columns for the following question.**

1. Write a SQL query that shows the name, city, and liked\_by\_vegans columns for only the the donut shops in Portland, Oregon that are liked by vegans. Place your SQL and the rows of data (as a screenshot) below.

**Return to showing the following columns:**

**name, city, county, state, country, rating, review\_count, price\_indicator**

1. Write a SQL query that answers the following question: Which donut shop in Oregon has the highest rating? Place your SQL and the first 5 rows of data (as a screenshot) below.
2. Modify the query from 17 to show all donut shops that have a rating of 5. Order the data by state and city. Place your SQL and the first 5 rows of data (as a screenshot) below.
3. Modifying the query from 18 and using the count() aggregate function (described in question 3), write a SQL query that shows how many donut shops have a rating of 5. Place your SQL and the rows of data (as a screenshot) below.
4. Modify your SQL from question 17, but show only the donut shops in **California** which have a name that begins with a ‘Co’. Place your SQL and the first 5 rows of data (as a screenshot) below.
5. Modify your query from question 20, but return donut shops in **California** that have a name that begins with a ‘co’. Place your SQL and the rows of data (as a screenshot) below.
6. Modify your query from question 21, but return donut shops in **California** that have a name that begins with a ‘co’ with any capitalization. Place your SQL and the first 5 rows of data (as a screenshot) below.
7. Write a SQL query that answers the following question: What donut shops in Oregon have the word ‘donut’ (with any capitalization) anywhere in the name? Place your SQL and the rows of data (as a screenshot) below.