The dataset for this exercise has been derived from the `Indeed Data Scientist/Analyst/Engineer` [dataset](https://www.kaggle.com/elroyggj/indeed-dataset-data-scientistanalystengineer) on kaggle.com.

#### Provide the SQL queries and answers for the following questions/tasks using the data\_analyst\_jobs table you have created in PostgreSQL:

1. How many rows are in the data\_analyst\_jobs table? **1793**

**SELECT count (\*)**

**FROM indeed\_data\_analyst\_jobs**

2. Write a query to look at just the first 10 rows. What company is associated with the job posting on the 10th row? **ExxonMobil**

**SELECT company**

**FROM indeed\_data\_analyst\_jobs**

**LIMIT 10**

3. How many postings are in Tennessee? **21**

**SELECT COUNT (location)**

**FROM indeed\_data\_analyst\_jobs**

**WHERE location = 'TN'**

How many are there in either Tennessee or Kentucky? **27**

**SELECT count (title)**

**FROM indeed\_data\_analyst\_jobs**

**WHERE location IN ('TN', 'KY')**

4. How many postings in Tennessee have a star rating above 4? **3**

**SELECT count (\*)**

**FROM indeed\_data\_analyst\_jobs**

**WHERE location = 'TN'**

**AND star\_rating > 4;**

5. How many postings in the dataset have a review count between 500 and 1000? **151**

**SELECT count (\*)**

**FROM indeed\_data\_analyst\_jobs**

**WHERE review\_count BETWEEN 500 AND 1000**

6. Show the average star rating for companies in each state. The output should show the state as `state` and the average rating for the state as `avg\_rating`. Which state shows the highest average rating? **NE**

**SELECT AVG(star\_rating) AS avg\_rating, location AS state**

**FROM indeed\_data\_analyst\_jobs**

**WHERE location IS NOT null**

**AND star\_rating IS NOT null**

**GROUP BY location**

**ORDER BY AVG(star\_rating) DESC**

7. Select unique job titles from the data\_analyst\_jobs table. How many are there? **881**

**SELECT COUNT (DISTINCT title)**

**FROM indeed\_data\_analyst\_jobs**

8. How many unique job titles are there for California companies? **230**

**SELECT COUNT (DISTINCT title)**

**FROM indeed\_data\_analyst\_jobs**

**WHERE location = 'CA';**

9. Find the name of each company and its average star rating for all companies that have more than 5000 reviews across all locations. How many companies are there with more than 5000 reviews across all locations?

**SELECT (company), AVG(star\_rating)**

**FROM indeed\_data\_analyst\_jobs**

**WHERE review\_count > 5000**

**AND company IS NOT NULL**

**GROUP BY company;**

**"company" "star\_rating"**

**"JPMorgan Chase" 3.900000095**

**"Lockheed Martin Corporation" 4.099999905**

**"Kroger" 3.599999905**

**"Siemens" 4.099999905**

**"PwC" 4**

**4**

**"Nike" 4.199999809**

**"Unilever" 4.199999809**

**"Dell" 4**

**"Staples" 3.599999905**

**"Goodwill Industries International, Inc. (GII)" 3.599999905**

**"Capital One" 4**

**"The Home Depot" 3.799999952**

**"American Express" 4.199999809**

10. Add the code to order the query in #9 from highest to lowest average star rating. Which company with more than 5000 reviews across all locations in the dataset has the highest star rating? What is that rating? **American Express** **4.199999809**

**SELECT DISTINCT (company), star\_rating**

**FROM indeed\_data\_analyst\_jobs**

**WHERE review\_count > 5000**

**ORDER BY star\_rating DESC**

11. Find all the job titles that contain the word ‘Analyst’. How many different job titles are there? **754**

**SELECT DISTINCT title**

**FROM indeed\_data\_analyst\_jobs**

**WHERE LOWER (title) LIKE '%analyst%';**

12. How many different job titles do not contain either the word **‘Analyst’ or the word ‘Analytics’? 4**

**SELECT \***

**FROM indeed\_data\_analyst\_jobs**

**WHERE UPPER (title) NOT LIKE '%ANALYST%'**

**AND UPPER (title) NOT LIKE '%ANALYTICS%'**

What word do these positions have in common?

**Tableau**

\*\*BONUS:\*\*

You want to understand which jobs requiring SQL are hard to fill. Find the number of jobs by industry (domain) that require SQL and have been posted longer than 3 weeks.

- Disregard any postings where the domain is NULL.

- Order your results so that the domain with the greatest number of `hard to fill` jobs is at the top.

- Which three industries are in the top 4 on this list? How many jobs have been listed for more than 3 weeks for each of the top 4?