CIS 9440

Section Number: UWA 32009

Homework #1

**Student name:** Jason Jiang

**Student ID:** 23435668

**Date:** 9/18/2022

**1. Follow the instructions on Blackboard to create a new project in Google BigQuery**

**Once complete, take a screenshot of the results of any query you write.**

**ANSWER:**

**Graphical user interface, application

Description automatically generated**

**2. Take a screenshot of your results and post them as the answer to this question.**

**ANSWER:**

Graphical user interface, application

Description automatically generated

**3. Continuing to use the “new\_york\_trees” dataset, write a query to find the top 5 most common trees in the “tree\_census\_2015” table. More specifically, you are looking for the top 5 most common “spc\_common” in the table.**

**ANSWER:**

**Table

Description automatically generated with medium confidence**

**4. Continuing to use the “new\_york\_trees” dataset, write a query to find the average tree diameter of trees in “Good” health by Borough in the “tree\_census\_2015” table.**

**For more details, in the “tree\_census\_2015” table the tree diameter is in column “tree\_dbh”, tree health is in column “health”, and Boroughs are in column “boroname”.**

**ANSWER:**

**Graphical user interface, text, application

Description automatically generated**

**5. Continuing to use the “new\_york\_trees” dataset and the “tree\_census\_2015” table, write a query to find the common name of the tree with the largest tree diameter in the Borough of “Brooklyn”.**

**For more details, in the “tree\_census\_2015” table the tree diameter is in column “tree\_dbh”, tree common name is in column “spc\_common”, and Boroughs are in column “boroname”.**

**ANSWER:**

**Graphical user interface, text, application

Description automatically generated**

**6. Continuing to use the “new\_york\_trees” dataset and the “tree\_census\_2015” table, write a query to determine which “curb\_loc” has the largest average “tree\_dbh”.**

**For more details, in the “tree\_census\_2015” table the tree diameter is in column “tree\_dbh”, tree common name is in column “spc\_common”, and Boroughs are in column “boroname”.**

**ANSWER:**

**Graphical user interface, text, application, email

Description automatically generated**

**7. Continuing to use the “new\_york\_trees” dataset and the “tree\_census\_2015” table, write a query to determine the zip code with the most trees in “Good” health.**

**For more details, in the “tree\_census\_2015” table the tree diameter is in column “tree\_dbh”, tree common name is in column “spc\_common”, and Boroughs are in column “boroname”.**

**ANSWER:**

**Graphical user interface, text, application

Description automatically generated**

**8. Continuing to use the “new\_york\_trees” dataset and the “tree\_census\_2015” table, write a query to help determine if “guards” improve the trees’ “health”. Be creative with this query, and explain your answer.**

**For more details, in the “tree\_census\_2015” table the tree diameter is in column “tree\_dbh”, tree common name is in column “spc\_common”, and Boroughs are in column “boroname”.**

**ANSWER:**

**Table

Description automatically generated**

Below is to help with the analysis in question #8 to make sure the counts add up

**Graphical user interface, text, application

Description automatically generated with medium confidence**

**Explanation to Question #8**

Guards **do not** help improve the tree’s health because the majority - about 87% ) of trees with good health do not have guards at all.

**9. Continuing to use the “new\_york\_trees” dataset and the “tree\_census\_2015” table, write a query to determine the most common “user\_type” for trees that are “London planetree”.**

**For more details, in the “tree\_census\_2015” table the tree diameter is in column “tree\_dbh”, tree common name is in column “spc\_common”, and Boroughs are in column “boroname”.**

**ANSWER:**

**Graphical user interface, text, application, email

Description automatically generated**

**10. Now, we will move onto a different Google Public Dataset. Find the “chicago\_taxi\_trips” dataset in the “bigquery-public-data” project.**

**In the table “taxi\_trips”, write a query to find the average tip left by taxi riders that paid with a “Credit Card” for rides that were longer than 15 minutes.**

**For more details, the tip is in the “tips” column, the payment type is in the “payment\_type” column, and the trip duration is in the “trip\_seconds” column.**

**ANSWER:**

**Graphical user interface, text, application, email

Description automatically generated**

**11. Still using the “taxi\_trips” table in the “chicago\_taxi\_trips” dataset,**

**Write a query to find the payment type that resulted in the largest average tip for rides that were longer than 10 minutes and between 5 and 10 miles.**

**For more details, the tip is in the “tips” column, the payment type is in the “payment\_type” column, the trip duration is in the “trip\_seconds” column, and the trip distance is in the “trip\_miles” column.**

**ANSWER:**

**Graphical user interface, text, application, email

Description automatically generated**

**12. Still using the “taxi\_trips” table in the “chicago\_taxi\_trips” dataset,**

**Use a SQL query to find the most expensive taxi “company”. You choose how to define “expensive”. Please paste a screenshot of your query, your query results, and your explanation of “expensive” as your answer.**

**ANSWER:**

**Graphical user interface, text, application

Description automatically generated**

**Explanation to Question #12**

Metro Jet Taxi A. is the most expensive taxi company because it has an average fare cost of about $415.15. I decided to use “fares” instead of the other metrics (tips, tolls, extras, and trip\_total) because these other metrics are influenced by other factors outside of the company’s control. Fares gives us just the amount charged by the company.

**13. Now, we will move onto a different Google Public Dataset. Find the “stackoverflow” dataset in the “bigquery-public-data” project.**

**Write a SQL query to find the “title”, “tags”, “view\_count”, and “score” of the 5 posts with the highest “favorite\_count” in the table “stackoverflow\_posts”.**

**ANSWER:**

**Table

Description automatically generated with medium confidence**

**14. Continue in the “stackoverflow” dataset in the “bigquery-public-data” project.**

**Write a SQL query to find the 10 most viewed posts regarding BigQuery.**

**Hint: leverage the LIKE operator.**

**ANSWER:**

**Graphical user interface, table

Description automatically generated with medium confidence**

**15. Continue in the “stackoverflow” dataset in the “bigquery-public-data” project.**

**Write a SQL query that joins the “stackoverflow\_posts” and “users” tables to return the “title”, “view\_count”, “owner\_display\_name”, and “reputation” of the 10 titles with the most comments (“comment\_count”). Please add a WHERE clause to filter out NULLs in the “title” column.**

**Hint: “reputation” is from the “users” table.**

**ANSWER:**

**Table

Description automatically generated**

**Question 15 ANSWER (for those whose “owner\_display\_name” is not null):**

**Table

Description automatically generated**

**16. Paste your ERD from Lucidchart as your answer.**

**ANSWER:**

**Table

Description automatically generated**

### Academic Integrity Statement - This needs to be signed by you.

*The work in this assignment is my own. I have not used outside help when answering the questions and have not used sites to procure answers (e.g., Chegg.com). Any outside sources have been properly cited.*



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Jason Jiang

(Signature)