

Principles of Urban Informatics

Homework #7 – SQL Basics

Instructor: Huy T. Vo – huy.vo@nyu.edu

Overview and Preparation

In this homework, you will be tasked with writing SQL snippets to extract data from a database given specific constraints. All assignments in this homework can be accomplished through CartoDB using the same data set that we have interrogated in Lab 6. The solutions for Lab 6, which has been posted on NYU Class under [Resources/Lab6_Keys.pdf](#), should be served as a good resource for this homework.

CartoDB setup: this is a reminder that as a NYU student, you are entitled to a CartoDB “Coronelli” equivalent account, i.e. with up to 500MB of storage. Please be sure to login to CartoDB using your NetID and password at <https://nyu.cartodb.com/>. More information can be found at: <http://guides.nyu.edu/cartodb>.

Data set: we are reusing the data set that was used in Lab 6, CitiBike trip records for 1 week of February 2015. If you have not already ingested this data to your CartoDB account (as part of the lab), you can do so by creating a new data set on CartoDB pointing to:
<https://serv.cusp.nyu.edu/files/hvo/citibike.csv>

Submission Info:

We do not require github for this assignment. Please submit your solutions through NYU Classes website only. For **each** of the assignment below, please turn in only **a single file** naming: **PUI2015_<NetID>_hw7_<AssignmentNumber>.sql**, where **<AssignmentNumber>** should be replaced with the assignment number below. For example, I would turn in 2 files:

PUI2015_hvt210_hw7_1.sql

PUI2015_hvt210_hw7_2.sql

... if I am to complete all 2 assignments of this homework. Each of the file would contain **a valid SQL command** that can be used to query and prepare the data to answer the questions presented in the assignments. A SQL command is considered “valid” in this case if it could be directly in the CartoDB’s SQL box against the citibike table to retrieve your answers. Thus, to facilitate the grading process and minimizing typos and mistakes, please be sure to:

1. Name your table in CartoDB as **citibike**.
2. Answer the “sanity” check questions inline in the submission text online.

Assignment 1

We would like to know the IDs and names of the top 5 origin-destination pairs that have the most rides in the data set. Please write a SQL command to extract that information from the data. The result table should have 5 columns: **start_station_id**, **start_station_name**, **end_station_id**, **end_station_name**, and **trip_count**, where **trip_count** is a new column created by your query containing the total number of trips started from the corresponding start and end station.

TABLE OUTPUT COLUMNS:

```
start_station_id
start_station_name
end_station_id
end_station_name
trip_count
```

SANITY CHECK QUESTION: *what is the top result, aka the station pair with the most riders?*
Please place the answer inline in your submission text online.

SANITY CHECK ANSWER 1: *(sample only, not actual answer)*

```
1234,Jay St & Myrtle,4321,Smith St & Atlantic, 9999
```

Assignment 2

We would like to plot where riders are coming to Washington Square Park from on weekends. Please write a SQL command to retrieve the start station IDs, names, and locations of all trips ending within 1000 meters of NYU, sorted in descending order by the number of trips from the start station.

TABLE OUTPUT COLUMNS:

```
start_station_id
start_station_name
start_station_location
trip_count
```

Similar to the previous assignment, both **start_station_location** and **trip_count** are new columns created by your query to specify the location of the start station and the total number of trips started from that station.

SANITY CHECK QUESTION: *what was the station ID and name with the most trip count to the Washington Square Park on the weekends? And what was the count?* Please place the answer inline in your submission text online.

SANITY CHECK ANSWER 2: *(sample only, not actual answer)*

1234,Jay St & Myrtle, 9999

NOTE:

- The latitude and longitude of the Washington Square Park are **40.7307602, -73.9974086**.
- The start station location (**start_station_location**) must be in the web-mercator projection so that we could visualize it on the map.