

Introduction-to-SQL Lab (through CARTO)

Huy T. Vo



NYU

Center for Urban
Science + Progress

Objectives

- Manipulate (spatio-temporal) data through SQL
 - Citibike ridership data
 - **<https://serv.cusp.nyu.edu/~hvo/files/citibike.csv>**
 - No Python or programming involved
 - Interactive SQL queries using Carto (previously known as CartoDB)



Task 1 — Familiarize with SQL Clauses

- Sort data by start_station_id, tripduration
 - Only checking trips with duration \leq 3 hours
- Only show the top/last 10 records (aka **head** and **tail** in SQL)
- List all unique start_station_id values
- Aggregation functions:
 - Count the number of trips (aka **wc -l** in SQL)
 - Find the average/min/max trip duration

Task 2 — Working with date/time

- Selecting trips started on Feb-02-2015 only
- Selecting trips started on the weekends
 - What are average trip duration during weekends?
- Can we do the same for weekday?



Task 3 — Working with Space

- Showing the list of start station locations
 - Using GROUP BY
- Showing the number of trips started per station
- ... but only for stations within 500m of Time Square!
 - The coordinates of Time Square is (40.7577,-73.9857)



Task 4 — Putting it all together

- Find the station that had the longest average trip duration during weekends and within 500m of TimeSquare!
- Extra: **create lines** for trips started from stations within 500m of Times Squares and **lasted less than 2 hours**. The number of trips per each pair of stations are output as attributes of these lines.

