Hyo Sung (Angelica) Kim and Chin Yee Lee

MUSA 620 Final Project Proposal - Taxi and Uber Data

Topic Choice//Overview of Concept:

Increased popularity of transportation network companies (TNC) such as Uber and Lyft in recent years has raised discussion on their competition against the long established taxicab industry. We are interested in using taxi trip data and Uber data in order to further explore and compare how these businesses perform in terms of passenger pick-ups. A spatial visualisation of where and when these pick-ups occur is expected to reveal spatial-temporal similarities or differences in competitiveness and popularity.

We are also interested to see whether there are observable relations between the volume of pick-ups and traffic volume over the study duration.

Technical Implementation:

We want to create a web based, interactive dashboard that compares taxi data against Uber data in NYC between 2014 and 2015. The NYC taxi data comes from data.cityofnewyork.us and contains over 180M rows and 165M rows of data for 2014 and 2015 respectively. This data can be queried using an API call, which this project will do. The Uber data comes from Kaggle and contains over 14,000,000 rows of data for 2015 alone (data from 2014 contains over 4,500,000 rows of data). As the Uber data was obtained via a data request made by Kaggle, it is presented in the CSV format, and is restricted to April - September 2014, and Jan - July 2015. To aid comparison, we will only use NYC taxi data during this period of time as well.
https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city#other-Lyft_B02510.csv

Historic traffic volume data is available for the period of 2012 - 2013, and is delineated by hours of the day. There are 5945 observations recorded for this data set, accessible via an API. For streets with missing data, we will impute the traffic volume of the nearest adjacent street.
https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts-2012-2013-/p424-amsu

We plan to use (Jupyter Notebook/Python) and (Observable/JavaScript;D3) to create our visualizations and embed them into our final web page.

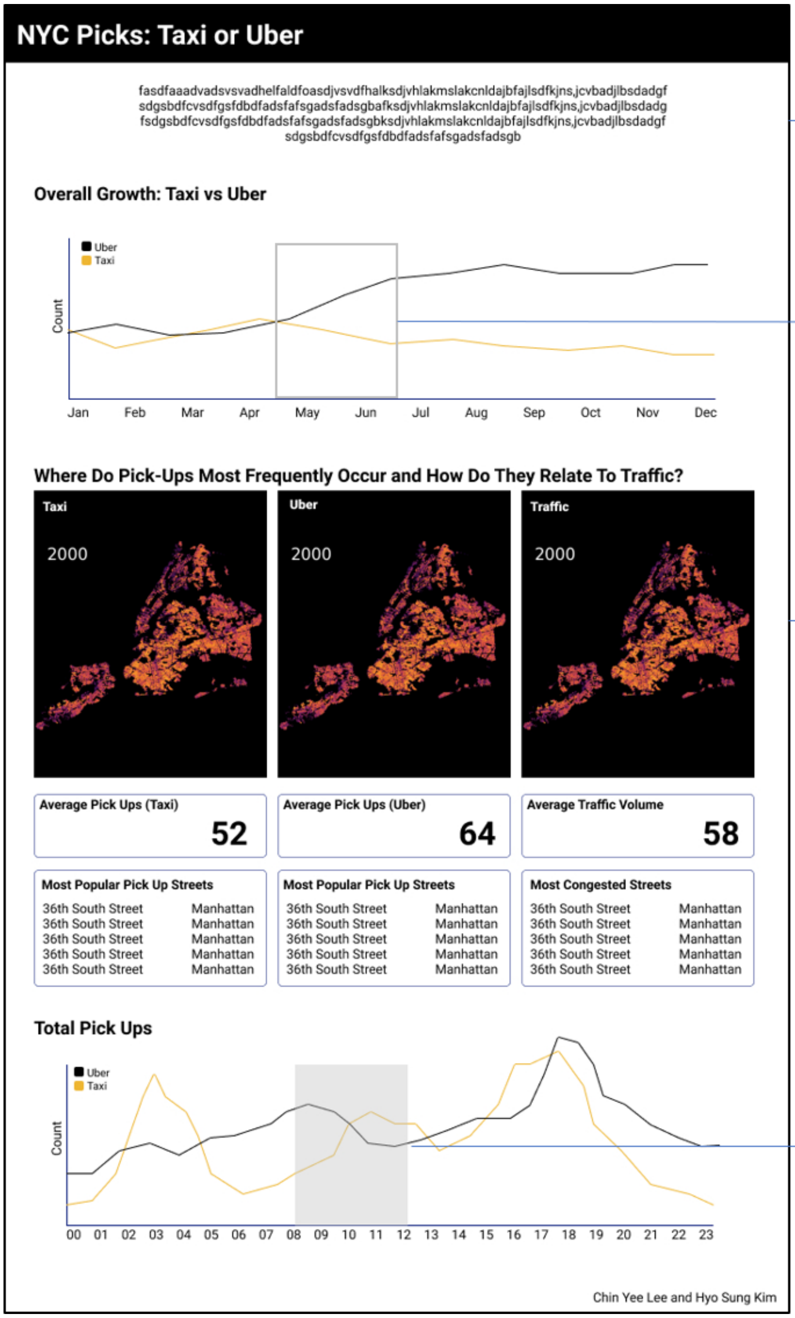
From Guidelines Checklist

✔ Data is collected through a means more sophisticated than downloading (e.g. scraping, API).
✔ At least one of the datasets contains more than 1,000,000 rows.
✔ It combines data collected from 3 or more different sources.
✔ The analysis of the data is reasonably complex, involving multiple steps (geospatial joins/operations, data shaping, data frame operations, etc).
✔ You use one of the analysis techniques for urban street networks (e.g., osmnx, pandana) or

clustering (e.g., scikit-learn)
✔ The webpage includes a significant interactive component (cross-filtering, interactive widgets, etc)

Design Details:

We want to include interactive dashboards where users will be able to visualize pick-up locations and counts based on user-selected filters. Please refer to the attached document/file for a draft visual of what we plan our dashboard web page to look like.



NYC Picks: Taxi or Uber

fasdfaaadvadsvsvadhelfaldfoasdjvsvdfhalksdjvhlakmslakcnldajbfajlsdfkjns,jcvbadjlbsdadgf
sdgsbdfcvsdfgsfdbdfadsfafsgadsfadsgbafksdjvhlakmslakcnldajbfajlsdfkjns,jcvbadjlbsdadg
fsdgsbdfcvsdfgsfdbdfadsfafsgadsfadsgbksdjvhlakmslakcnldajbfajlsdfkjns,jcvbadjlbsdadgf
sdgsbdfcvsdfgsfdbdfadsfafsgadsfadsgb

Overall Growth: Taxi vs Uber

Where Do Pick-Ups Most Frequently Occur and How Do They Relate To Traffic?

Average Pick Ups (Taxi) 52

Average Pick Ups (Uber) 64

Average Traffic Volume 58

Most Popular Pick Up Streets	
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan

Most Popular Pick Up Streets	
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan

Most Congested Streets	
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan
36th South Street	Manhattan

Total Pick Ups

Chin Yee Lee and Hyo Sung Kim

Introduction to the application site and key question

Interactive feature:
Selection Box
- The average values displayed in the maps and plot below will be derived from the values within this window.

Map 1 and 2 will display where the respective pick-ups occur.
Map 3 will display where traffic volume is high.

Interactive feature:
Selection Box
- By selecting a time period, the maps above will change to reflect the average value derived from this time period.