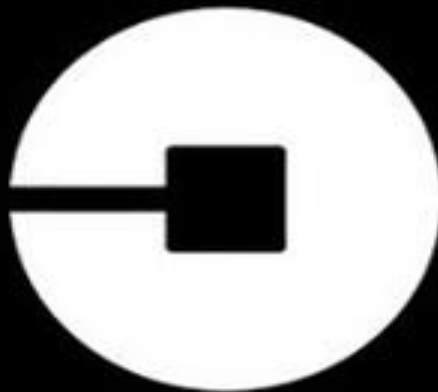


Uber Pickups



Leif Waldner, Parth Mishra

Project Description

We explored an Uber dataset that outlines Uber pickups in New York City. This dataset encompasses location, time, dispatcher, and date. We looked at trends over time of spacial and temporal uber ride distribution in New York City. We created a d3 visualization of demographic uber ride distribution in NYC.

Questions sought to Answer

This project aims to highlight some of the socioeconomic relationships when looking at pickup frequency in certain neighborhoods.

With their rise has come ethical concerns revolving around the dynamics of the driver-passenger relationship.

Dataset

<https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city>

- Cross-reference with demographic information of NYC
- 18.8 million data points in total

<https://data.cityofnewyork.us/City-Government/Demographic-Social-Economic-and-Housing-Profiles-b/kvuc-fg9b>

- Demographic, social, economic, and housing profiles by community district/puma.
-

Data Preparation

Cleaning: Ensured validity and accuracy of data set

Integration: Integrated with demographic datasets for more complex queries

Preprocessing: Re-encoded locations from Uber's location ID's to a more standardized format.

Visualization: Go through GeoJSON file and append statistics.

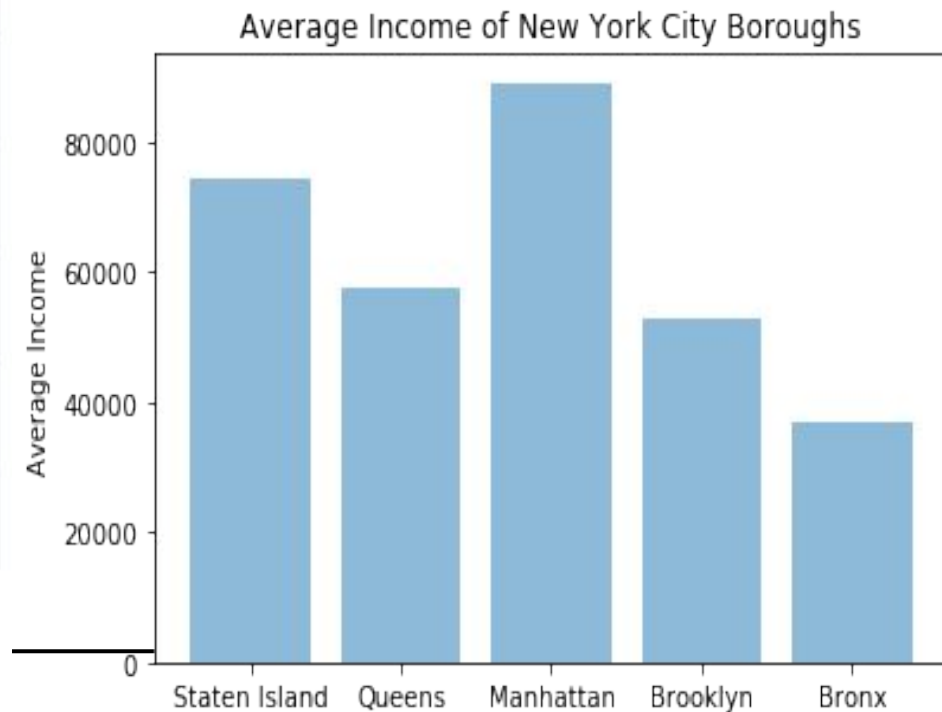
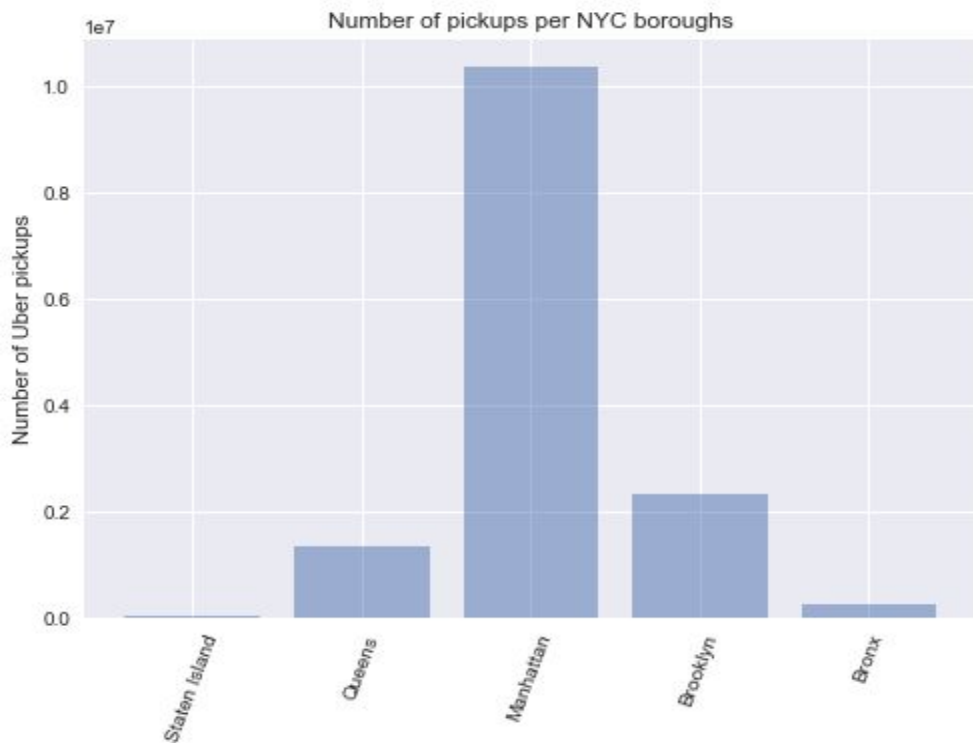
Tools

Python - Pandas, SKlearn for any model building

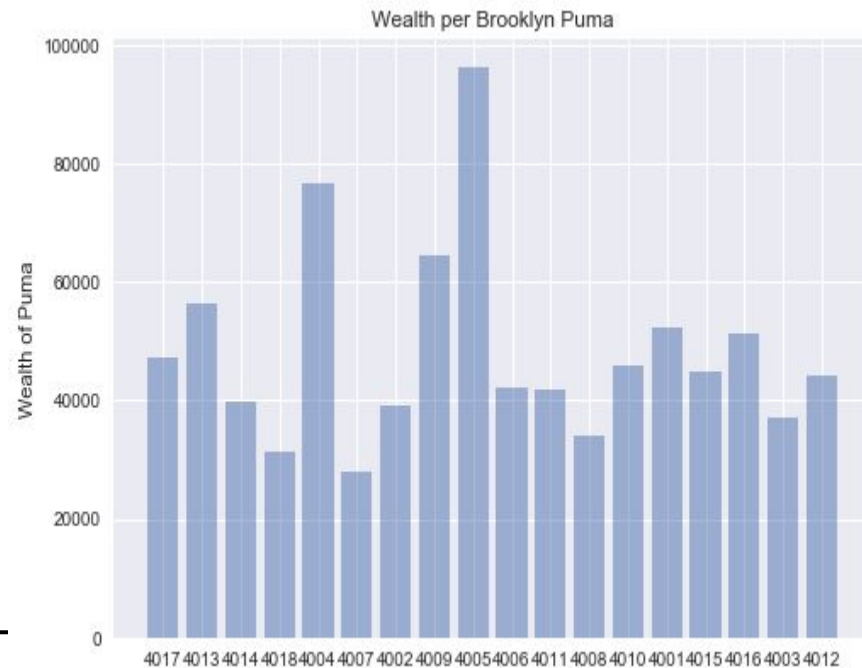
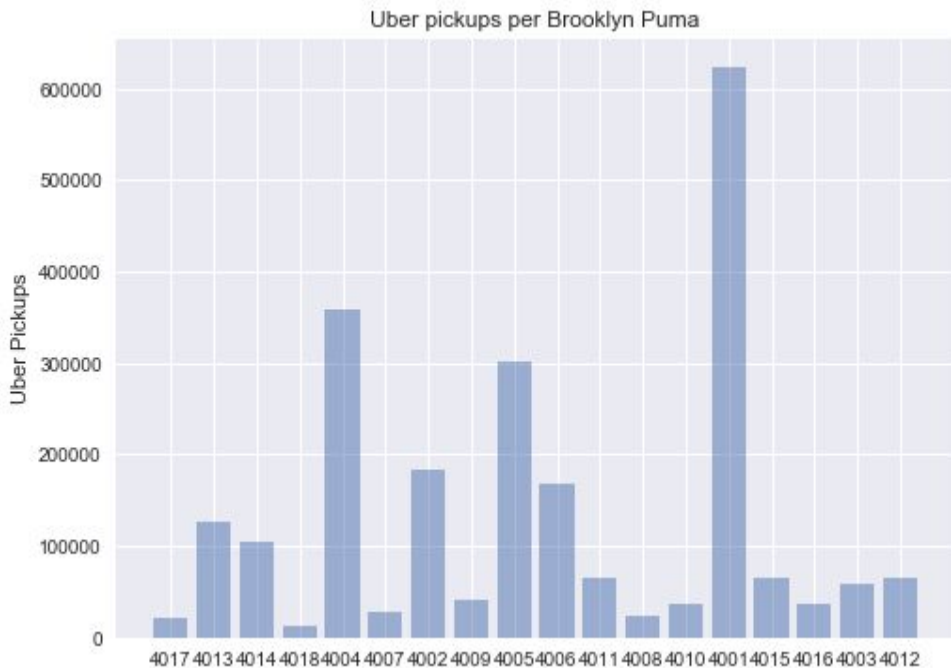
R - Exploratory data analysis

D3, Matplotlib - Visualization

Knowledge Gained



Brooklyn Puma Information



Visualization

DEMO!

Limitations

- Correlation/Causation due to data integration
 - Does not account for non-demographic causal factors such as geography(e.g. Staten Island).
 - Uber pickup data does not necessarily mean rider is from that location
 - Assuming that most Uber rides are within a PUMA distance
-