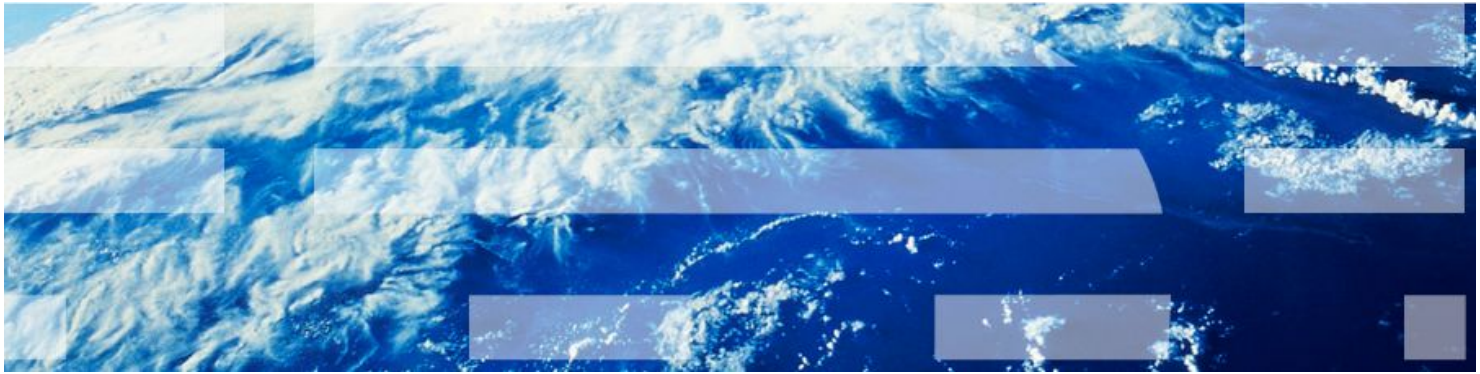
E6893 Big Data Analytics

Analysis and optimization for NYC public transportation alternatives

Project ID: 201912-42
Team Members (with UNI): Hongzhi Shi (hs3194)



Agenda

- Overall goal/tech stack/dataset with focus on the progress after the mid point presentation
- Demo of a web based app
- Some conclusion and thoughts
- Some challenges

Overall goal

- Trend and usage pattern of Taxi/FHV/Citi bike in the past couple of years. (covered by mid point presentation)

- A tool to provide insight of whether taking a taxi/uber or a citi bike is a better choice to get from one neighbourhood to another at a particular time of the day(focus of this presentation)

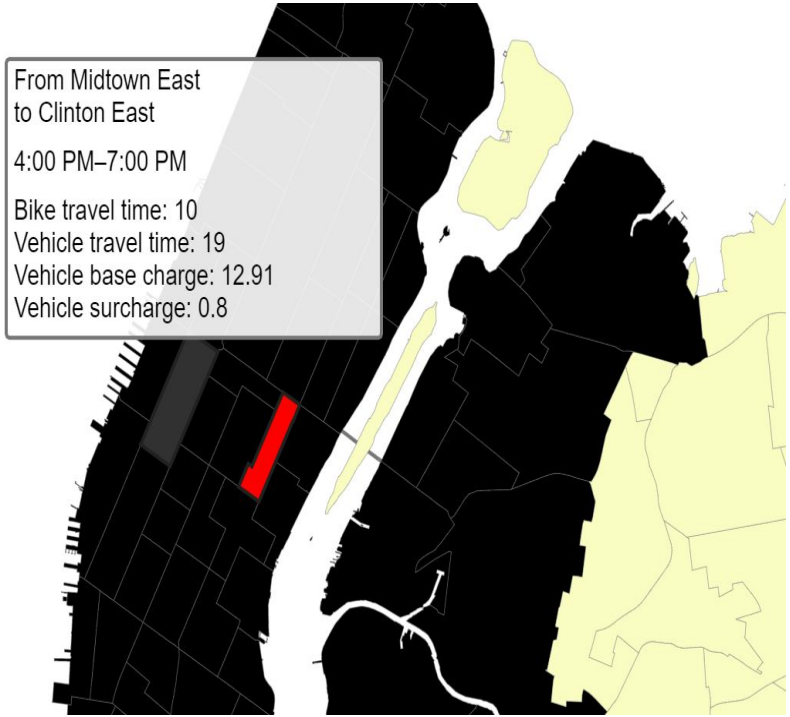
Tech stack

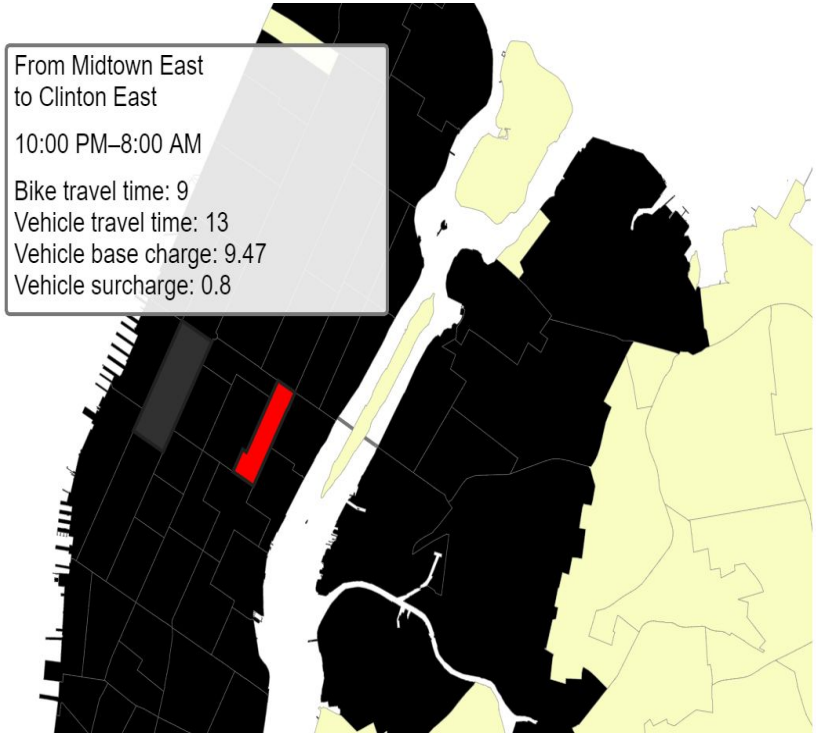
- Google cloud platform
- PySpark on dataproc for data joining/filtering/aggregation
- Cloud storage for all the data sets
- BigQuery with heavy reliance on GIS feature to do geometric computation and analysis
- Django for web app
- D3 and Vega for visualization

Dataset

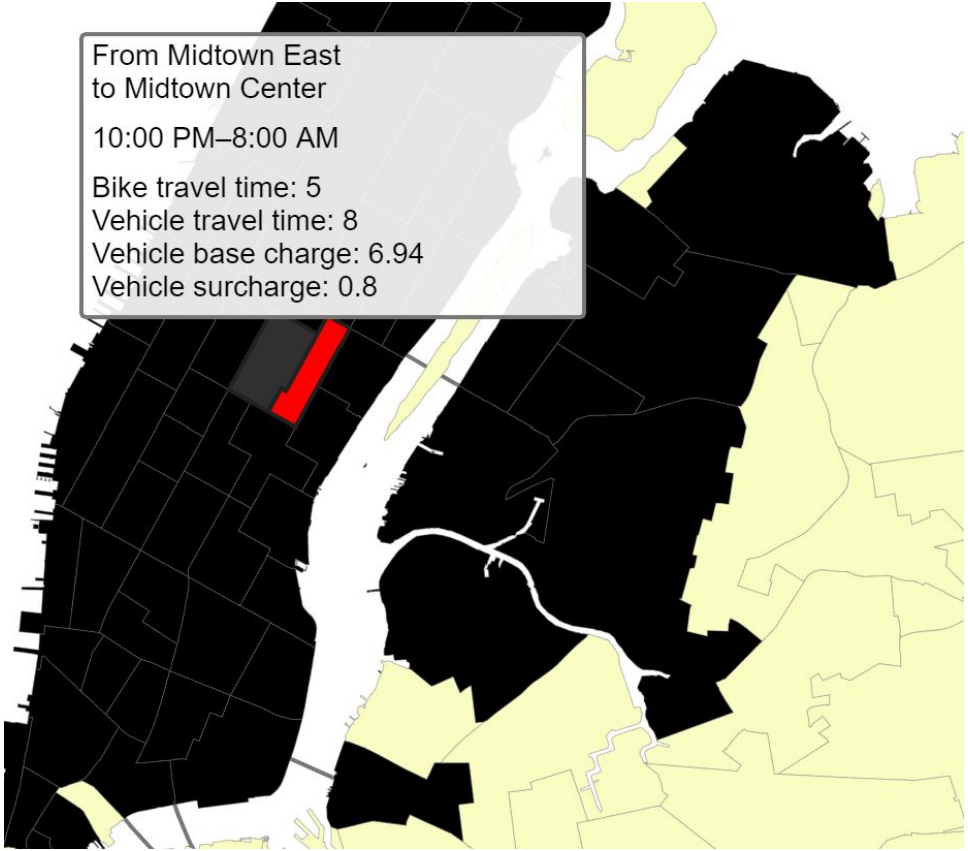
- Past 3 years of aggregate data from TLC for yellow taxi/green taxi/high volume for hire vehicles
- Past 3 years of trip data from citi bike
- Past 3 years of trip data from TLC for taxi
- NYC borough map file from NYC open data
- Taxi zone file from TLC

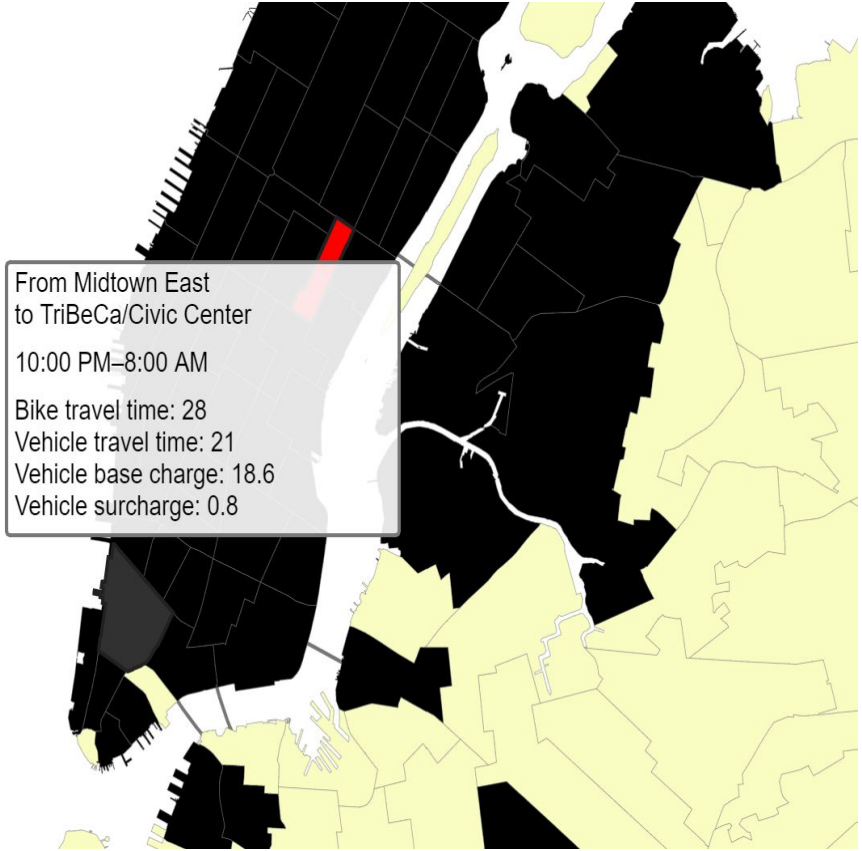
Demo : Cross town rush hour impact on travel time



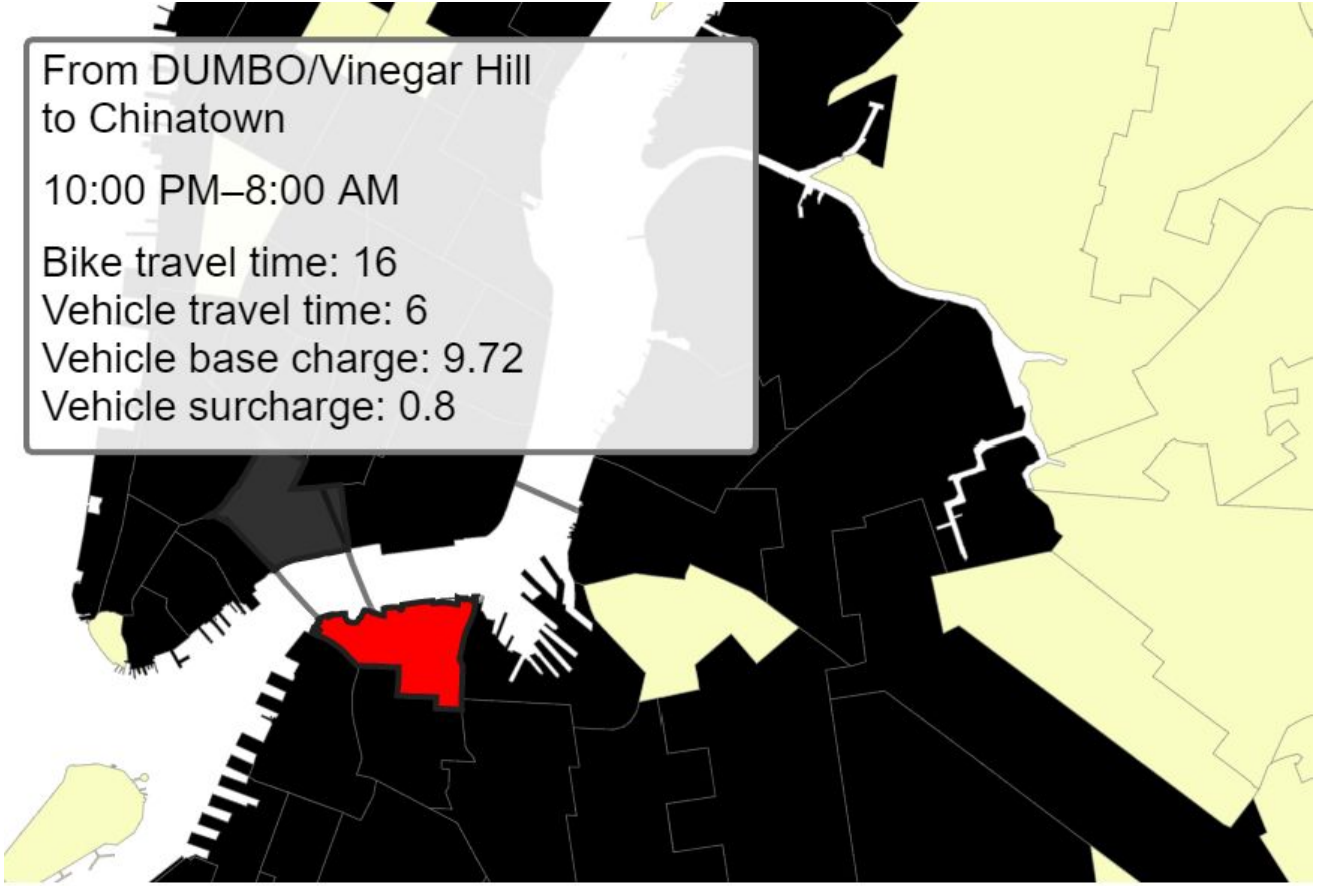


Demo : distance impact on travel time

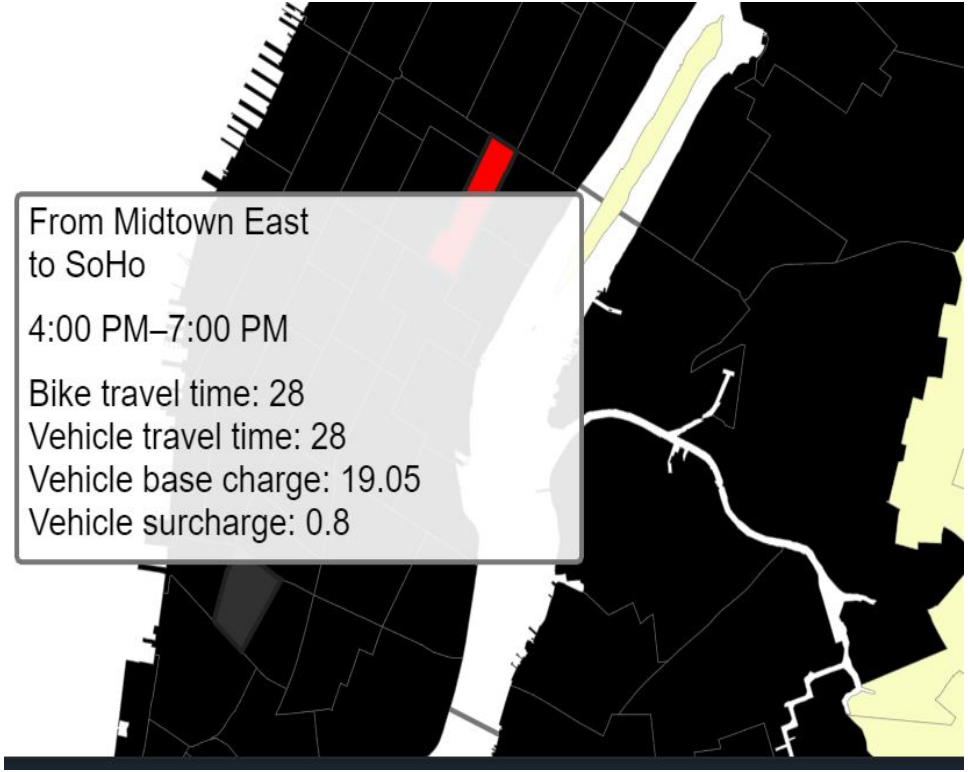


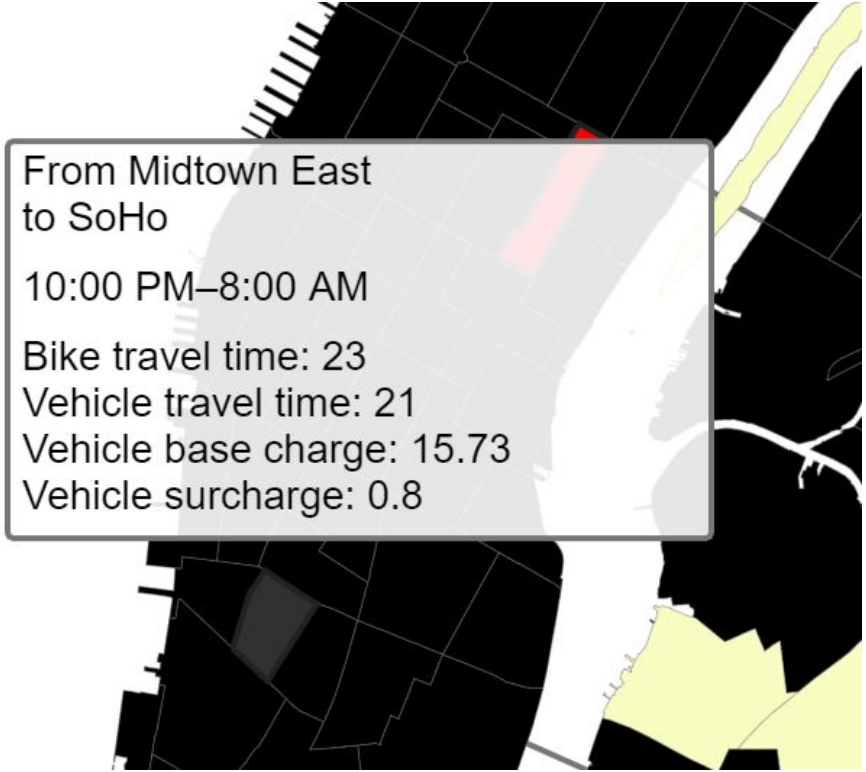


Demo : cross bridge impact on travel time



Demo : rush hour impact on fare





Conclusion

- Each plays a different role with FHV and Citi bike taking more shares from the Taxi service
- Citi bike is both faster and more budget friendly for shorter trips and rush hour trips
- Taxi and FHV is still a better choice for most of the long distance trips and cross bridge trips.

Challenges

- Different data format provided by different organizations
  - TLC goes by taxi zone VS citibike goes by specific station id with latitude and longitude
  - Solution: use BigQuery GIS feature to join the bike station location to the taxi zone, which is in the format of a polygon specified by a geojson file

- Lack of data for congestion and rush hour surcharge from FHV to provide better comparison between the FHV and Taxi