

Exploring Uber/taxi data

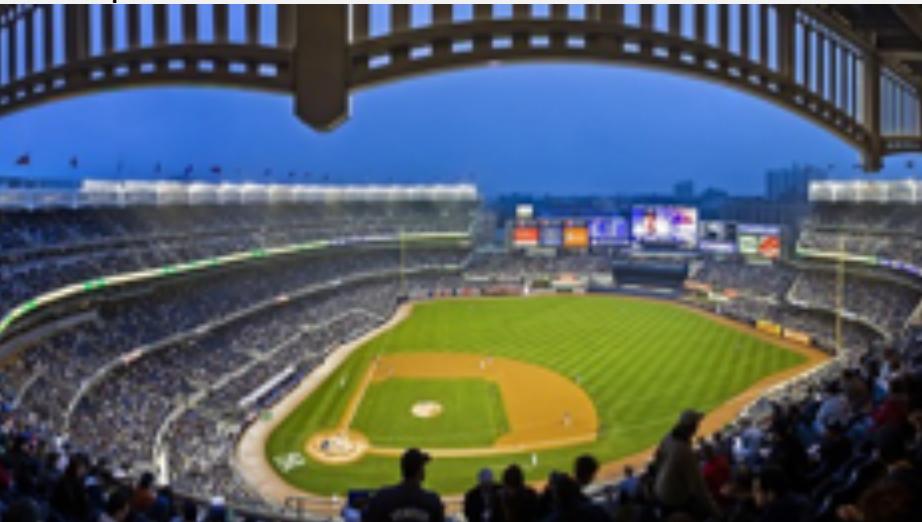
Case 3 Uber Case

PRESENTED BY Group 5

WPI



Background & Intro



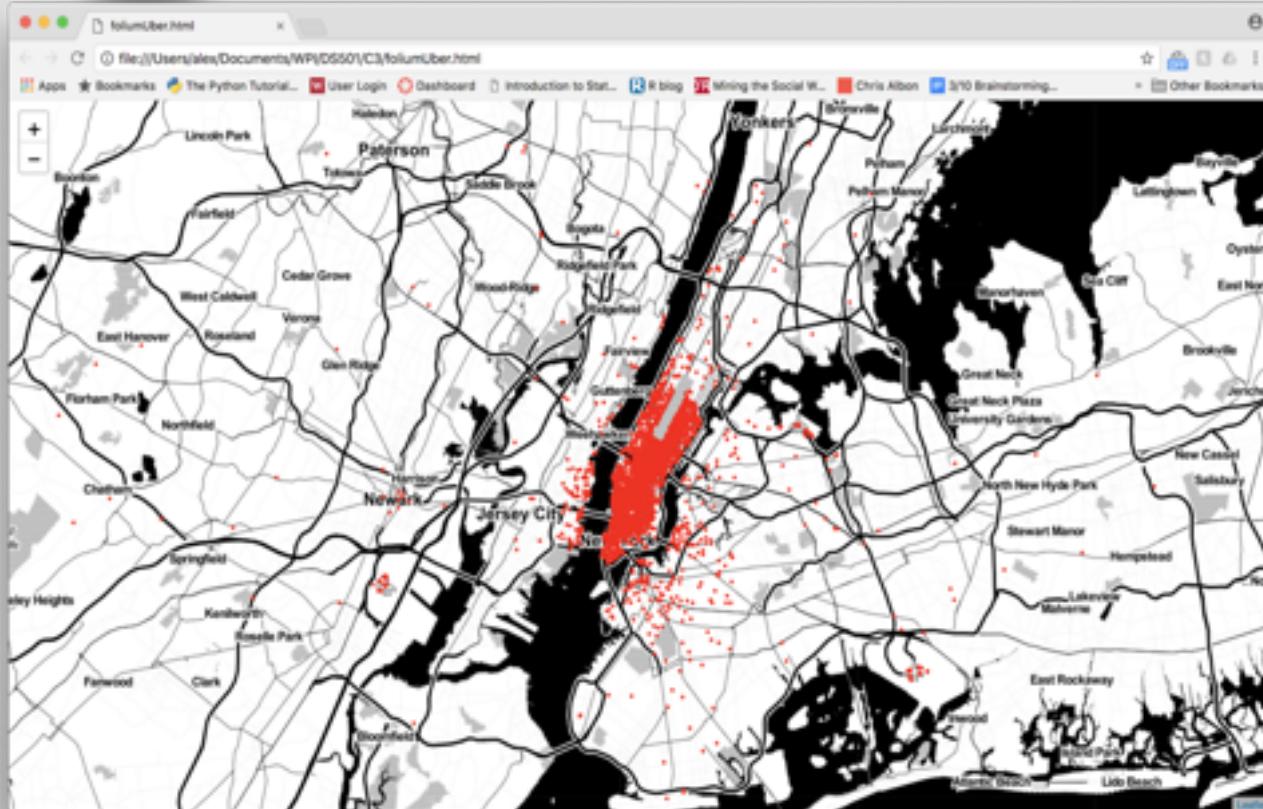
Sports Stadiums in New York

Plenty Taxi Data





Spatial distribution of 5000 Uber pickups in Apr, 2014



Data source: FiveThirtyEight



Game Day Data Exploration

- Where do fans come and go?
- Where do fans work and live?
- Tipping pattern on different game results?



Methodology

Database
Setup



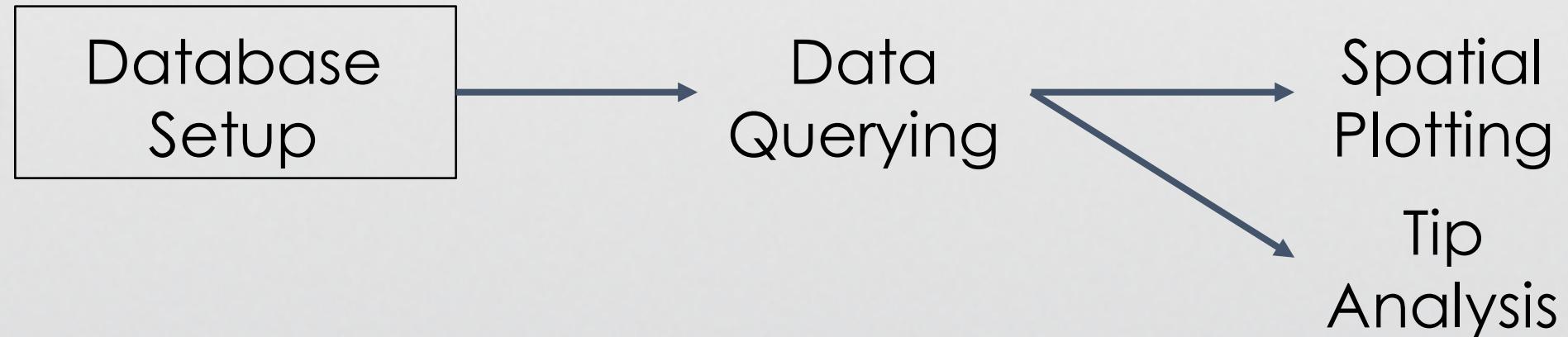
Data
Querying



Spatial
Plotting

Tip
Analysis





Google Cloud Compute

- single instance
- 16gb ram
- 300gb disk



300M taxi GPS tuples

- 2015 only
- yellow/green only



PostgreSQL

- simple, quick



Database
Setup



Single, obnoxious query (kind-of)



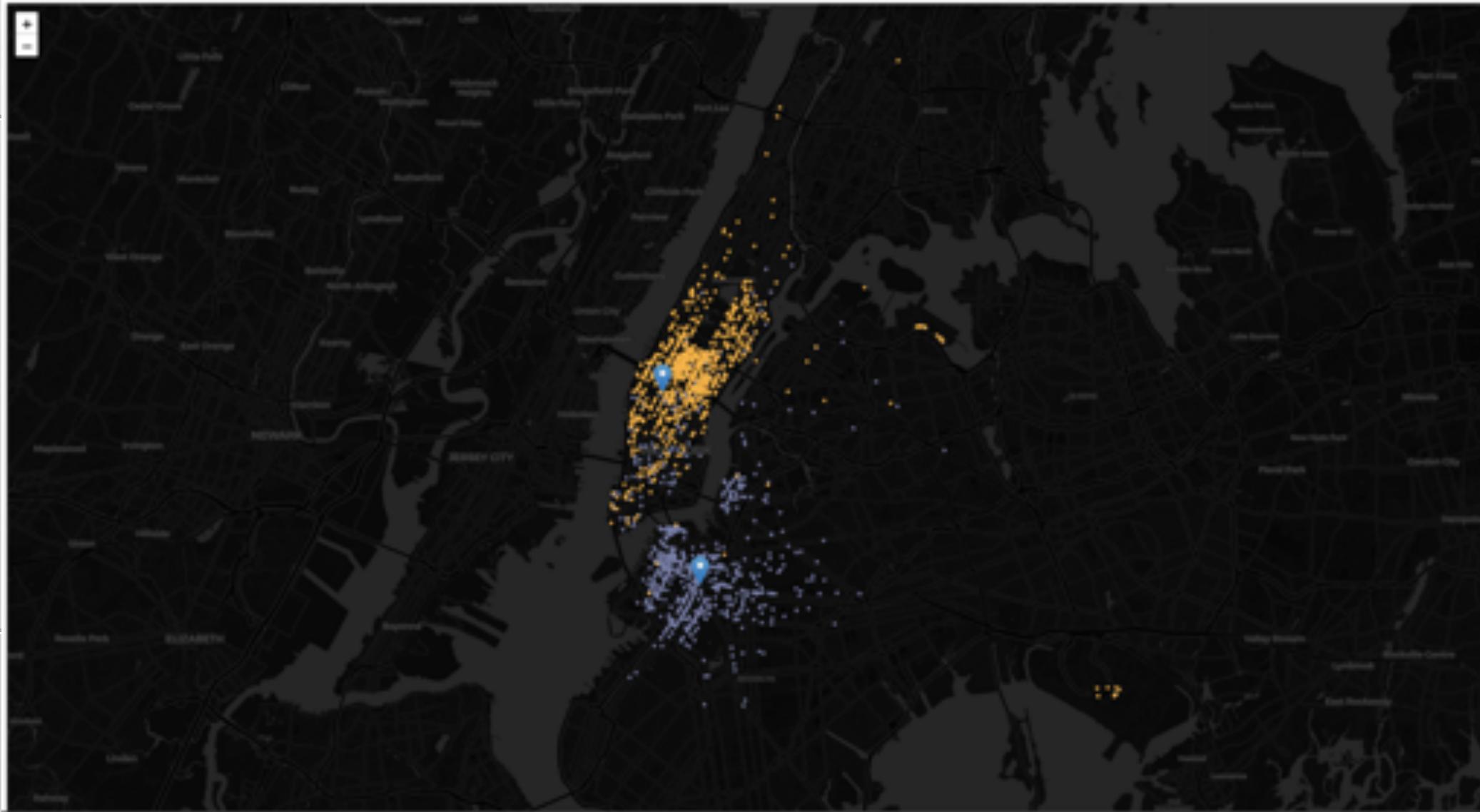
```
SELECT * FROM rides where
    (GPS_LAT between LAT_MIN and LAT_MAX) and
    (GPS_LONG between LONG_MIN and LONG_MAX) and
    ((DROP_TIME between GAME_1_START and GAME_1_END) or
     (DROP_TIME between GAME_2_START and GAME_2_END) or
     ...
     (DROP_TIME between GAME_N_START and GAME_N_END))
```



- Visualize pre-game pickups and post-game dropoffs
- Folium library (Python + Leaflet.js)



Pre-game spatial plotting of Knicks and Nets fans





Post-game spatial plotting of Knicks and Nets fans





Pre/Post-game comparison on Nets fans





→ Tip Ratio

- ◆ defined as: $\text{tip_amount} / \text{total_amount}$ (other than tips)

→ Focusing on trips with card payment

- ◆ cash: 0 tips

→ Capture the game result

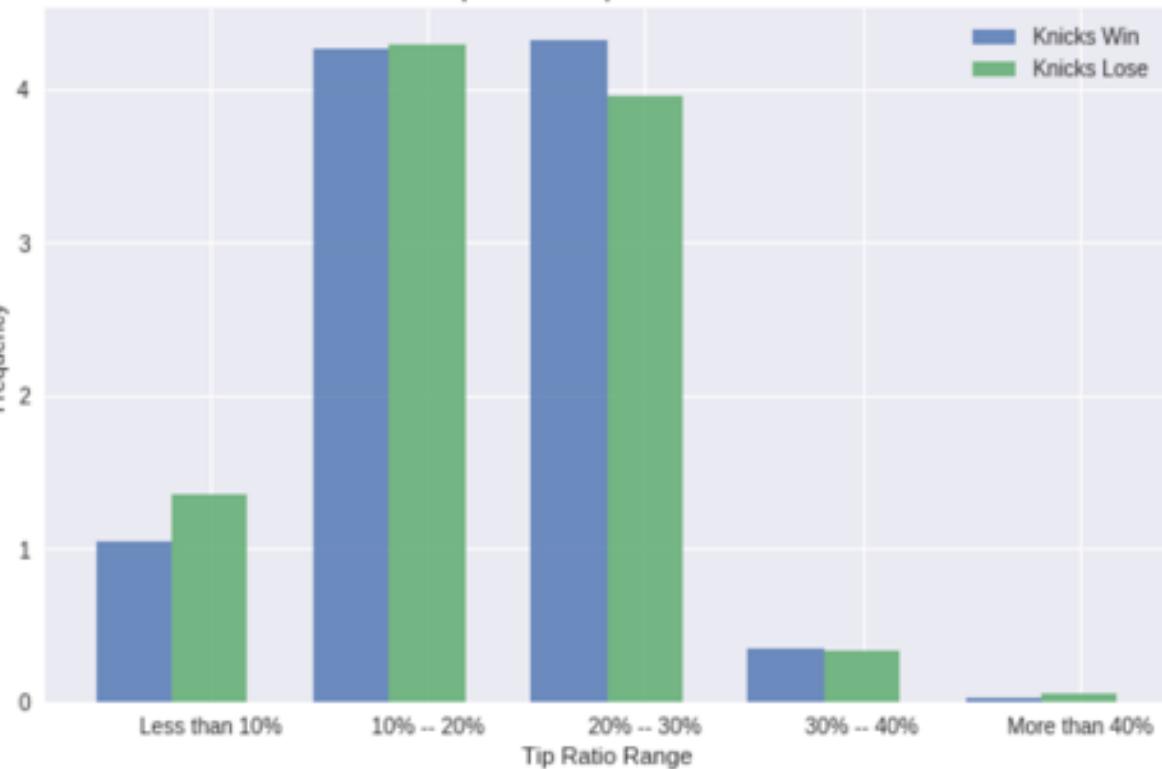
- ◆ (Win or Lose)

→ Plot Bar Chart to compare the result

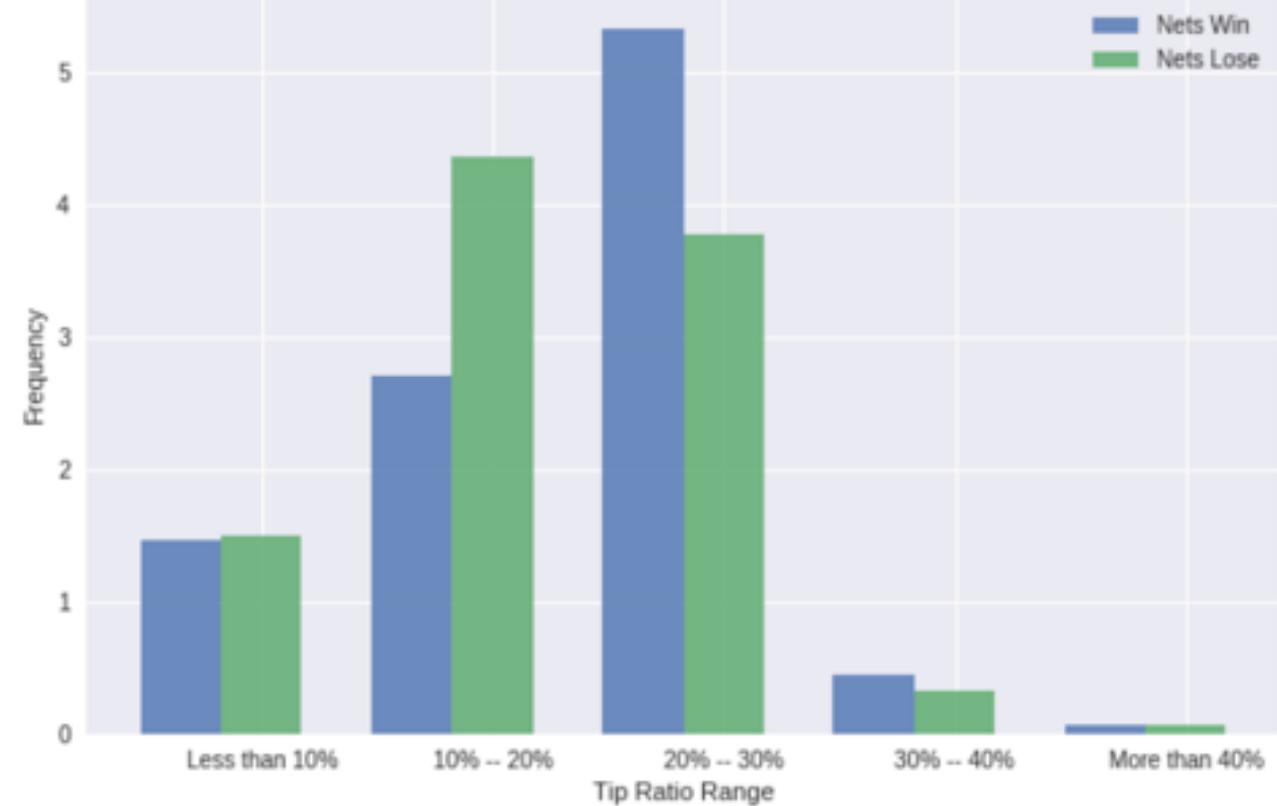


Tip Analysis

Tip Ratio Comparison Knicks

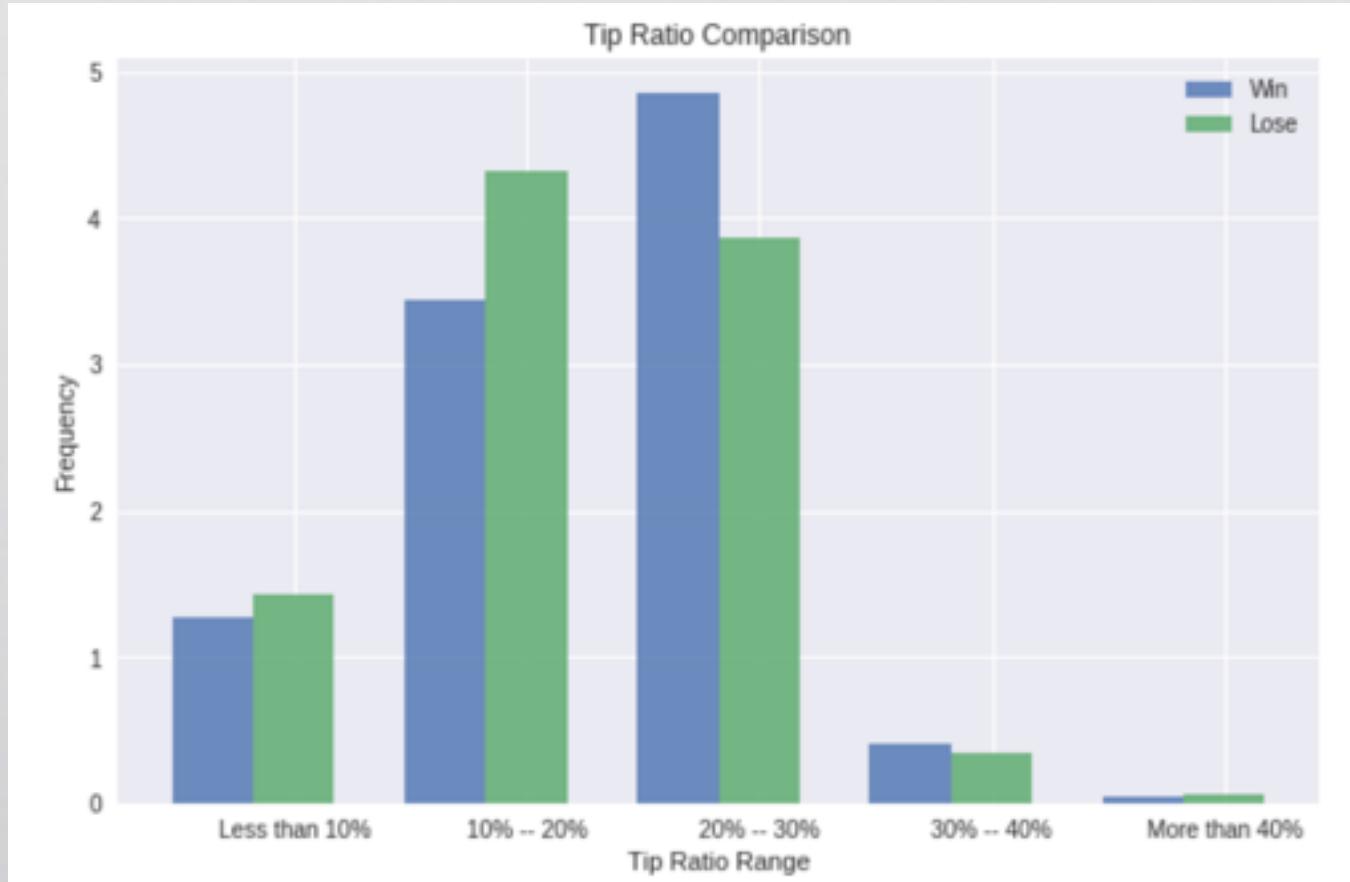


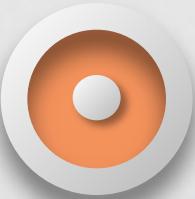
Tip Ratio Comparison Nets





Tip Analysis





Conclusions

- Billions of NYC Taxi/Uber data enable data exploration in all aspects
- Spatial plotting of pre-game GPS_pickups and post-game GPS_dropoffs enable visualization of fans habitants/advertising targets
- New Yorkers tend to tip more when their home team win



THANKS