

DATA ANALYSIS OF NYC CAB SERVICES

Abhinandan Dubey, Raju Khanal, Teja Madiraju

THE BIG PICTURE & THE PROBLEM

GOAL: Analysing commuting patterns in urban communities.

PURPOSE: Extract useful insights from NYC cab system datasets.

KEY INDICATORS: Some key factors that contribute to gaining these insights are

- Affluence of neighborhoods.
- Tipping patterns.
- Distances and routes.
- Spatio-temporal dependencies (Locations vs Times of the year)

WHY SHOULD YOU CARE?

- Key ideas involve large groups of people.
- Any solutions obtained can be mapped to other large cities.

WHY SHOULD YOU CARE?

- Key ideas involve large groups of people.
- Any solutions obtained can be mapped to other large cities.

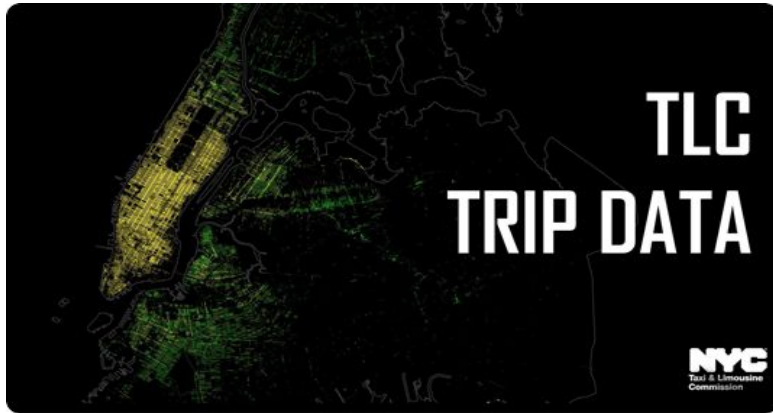
241 million | 6 hrs 18 min

Based on a ranking of states with the longest average commute-to-work times, the ACS showed that New York residents spent the most time traveling to their jobs

The US Bureau of Labor Statistics estimates that nearly 241 million passengers use NYC cabs annually.

DATA DESCRIPTION

- 19 Features : 2009 - 2016* across YellowCabs Service
- Using Datasets from 2013 - 16* : $1.8 \times 42 = 75.6$ GB



Dataset: Monthly trip data from TLC, NYC



Features from the dataset

OBSERVATIONS / EXPERIMENTS

- Do disputed payments indicate something about the neighborhood ?
- How does the tipping rate change according to location and/or peak times ?
- What are the Frequent Pick-up Drop-off points and how they relate to specific times of a year ?
- Do people travel in groups during festive seasons ?



OBSERVATIONS / EXPERIMENTS

- Do people in affluent neighborhoods really tip higher (or are they meagre ?)
- How much should you tip given a ride / location ?
- By how much will taxi fares increase in the subsequent year ?



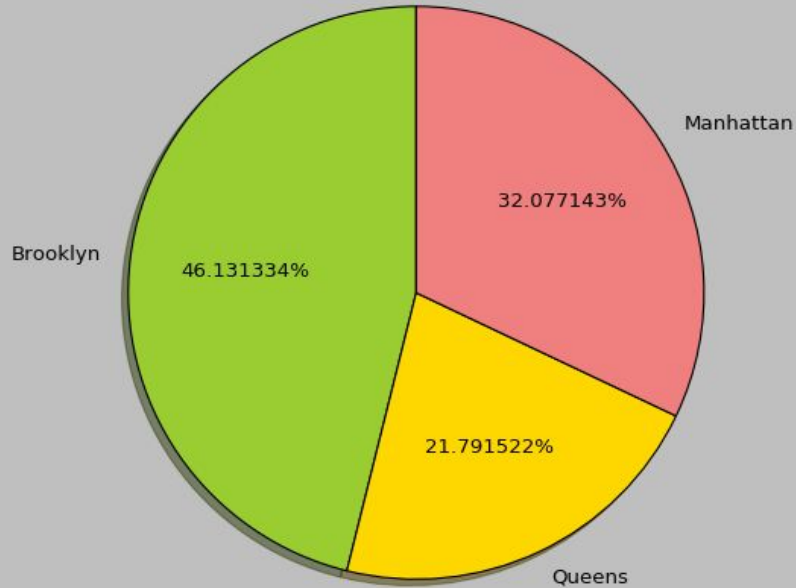
METHODS

Do people in affluent neighborhoods really tip higher ?

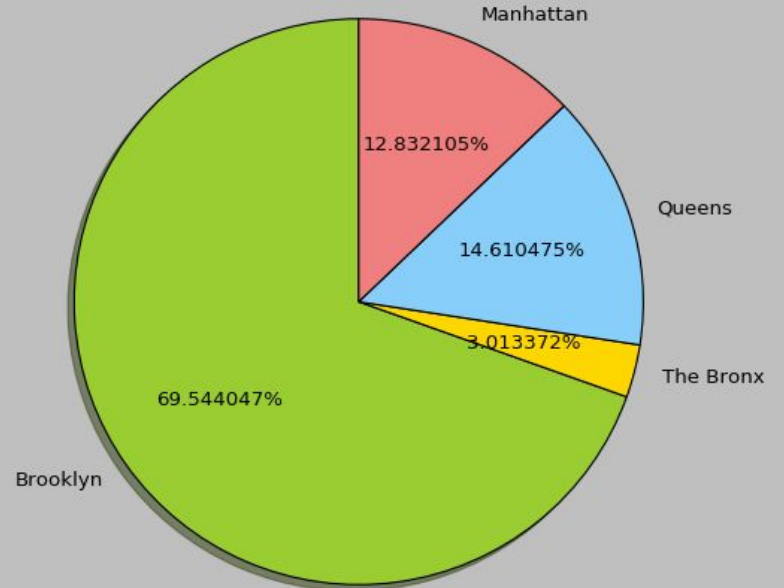
- Filtered rides with fares > \$150 (approx.)
- Computed tip per mile grouped by boroughs
- Compared with affluency of neighborhoods.

RESULTS

Dec 2015 Tip Affluence

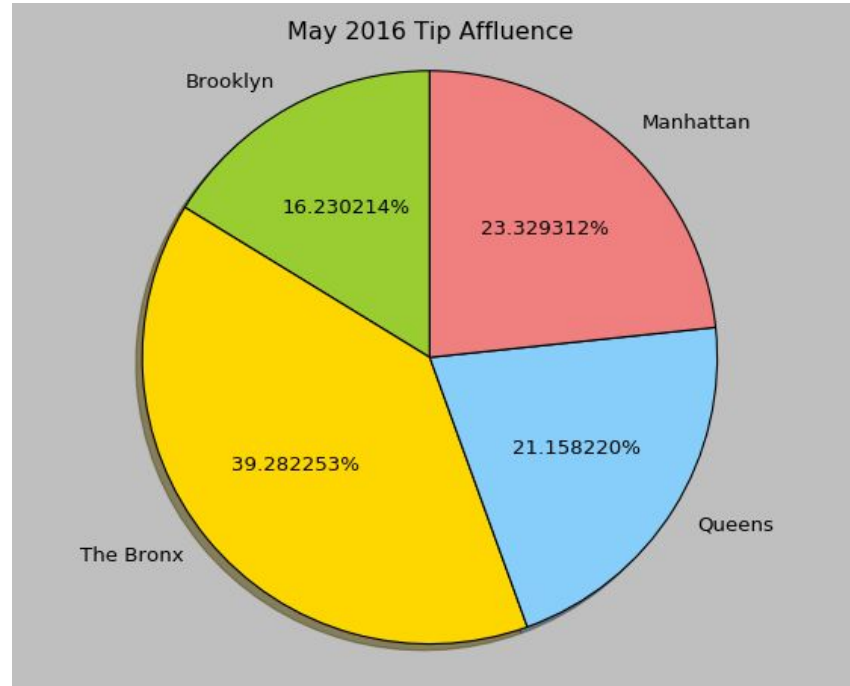


June 2016 Tip Affluence



RESULTS

-



METHODS

How much should you tip given a ride / location ?

> Recommendation System

$$tip_{user} = \frac{\sum_{i \in U} sim(user, i) + tip_{amount}(i)}{\sum_{i \in U} sim(user, i)}$$

RESULTS

How much should you tip given a ride / location ?

> Recommendation System

Some sample results include :

	Fare (items)					
Users	\$10	\$20	\$50	\$60	\$80	\$90
(M,Q)	NA	NA	\$8.34	\$12.11	\$17.44	\$19.9
(M,M)	\$1.98	\$3.97	\$11.63	\$13.59	\$15.54	\$21.53
(S, M)	NA	NA	\$9.12	\$11.85	\$14.06	\$20.25

M = Manhattan, Q = Queens, B = Brooklyn, S = Staten Island

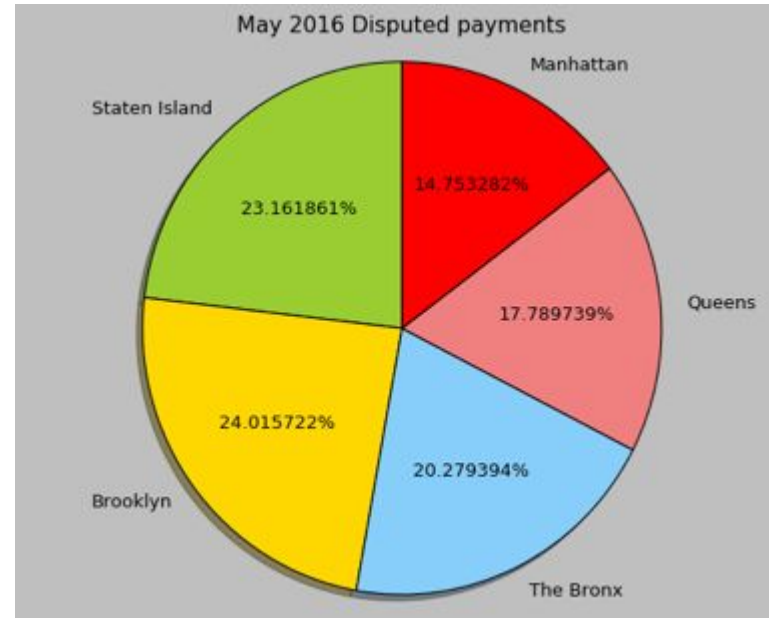
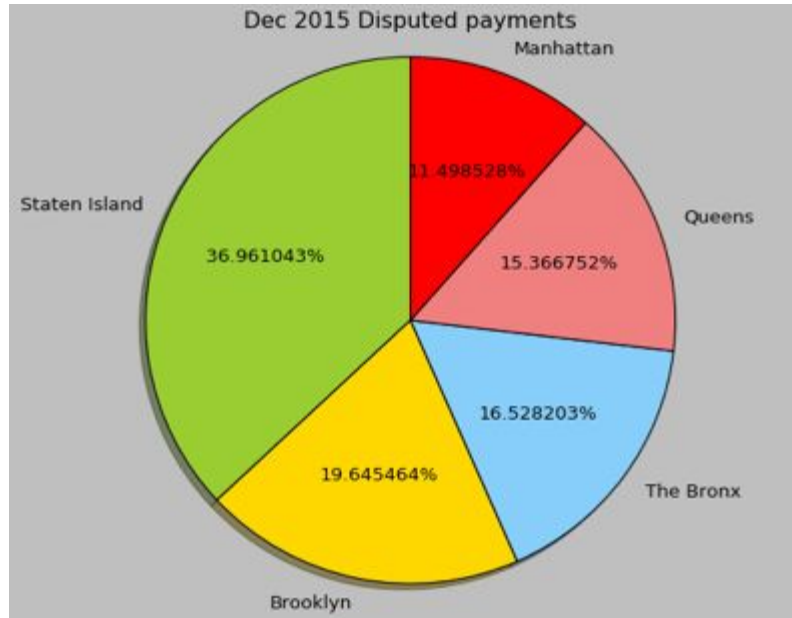
NA indicates that minimum fare between places is greater than the value of the item.

METHODS

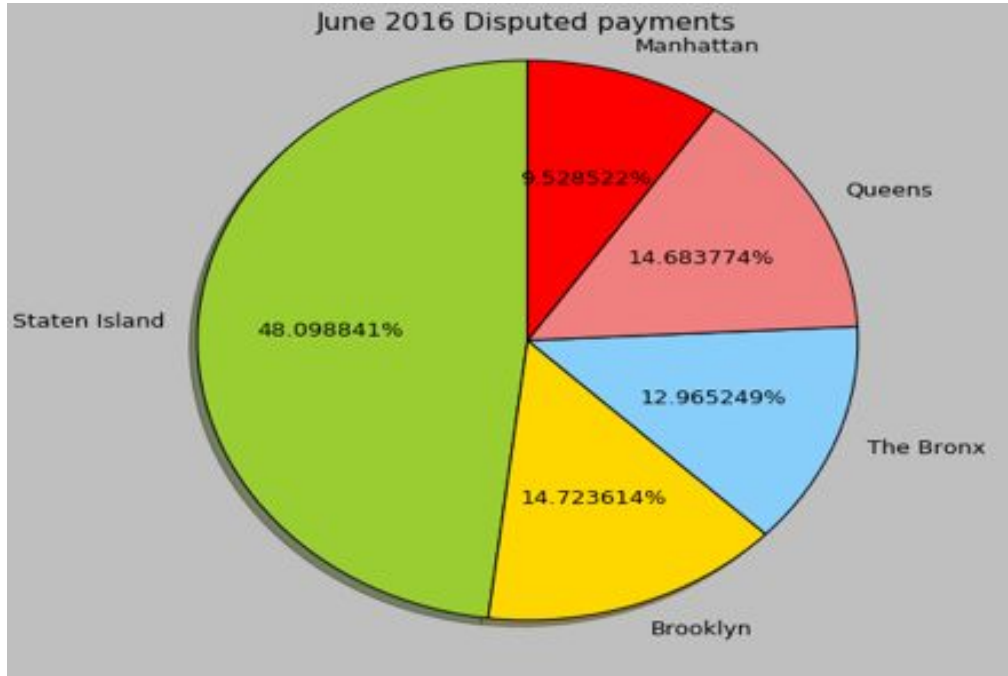
Do disputed payments indicate something about the neighborhood?

$$\frac{\textit{(No of disputed payments in that borough)}}{\textit{(No.of rides/payments made in that borough)}}$$

RESULTS



RESULTS



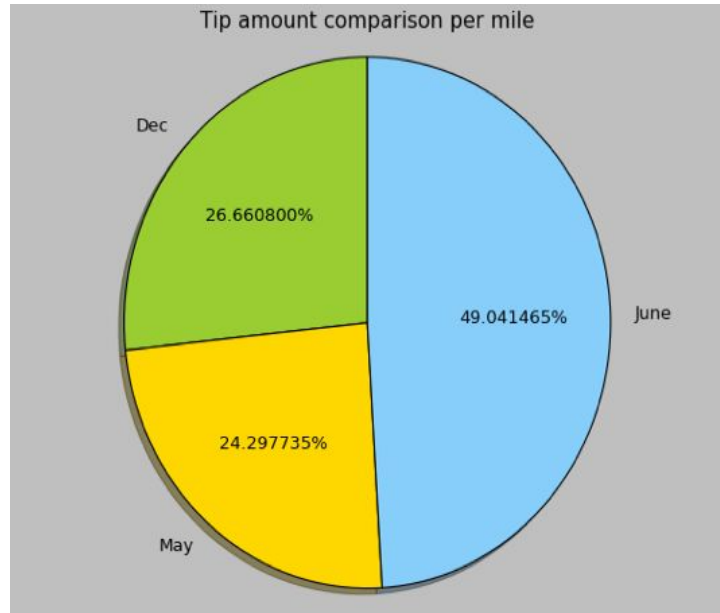
METHODS

How does the tipping rate change according to location and/or peak times ?

$$\frac{\textit{(Total Tip across all rides in a month)}}{\textit{(Total distance covered across all rides in a month)}}$$

RESULTS

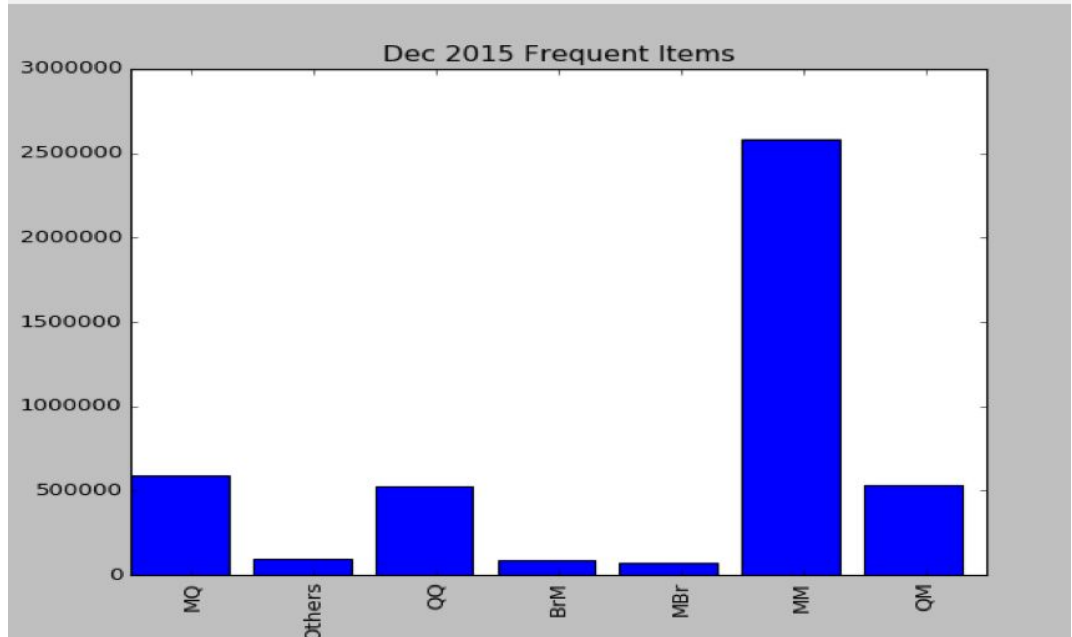
How does the tipping rate change according to location and/or peak times ?



METHODS / RESULTS

What are the Frequent Pick-up Drop-off points and how they relate to specific times of a year ?

FREQUENT ITEMSETS



METHODS

Do People Travel In Groups During Festive Seasons ?

- Approach 1 : Averaging
- Approach 2 : k-passenger ($k > 1$) rides

RESULTS

Do People Travel In Groups During Festive Seasons ?

Festive Time - Christmas

- 0.49%



- 3.88 % higher during these holidays.

METHODS

How much do you end up paying more for the fare because of traffic, and does it have a specific pattern of increasing over the years?

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

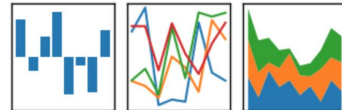
$$y_t = \rho y_{t-1} + u_t$$

cloudera

spark-ts

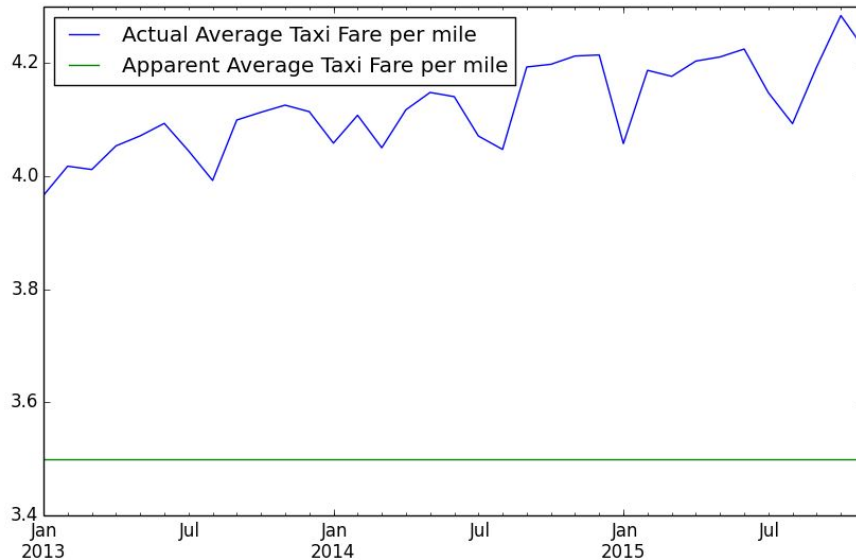
Maven[™]

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



RESULTS

How much do you end up paying more for the fare because of traffic, and does it have a specific pattern of increasing over the years?



Test Statistic	-0.927128
p-value	0.778907
#Lags Used	6.000000
Critical Value (5%)	-2.967882
Critical Value (1%)	-3.679060
Critical Value (10%)	-2.623158

DISCUSSION

- Urban communities and their commuting patterns.
- Increasing average trip times
- Tipping Patterns



CONCLUSION

Some benefits include

- PASSENGERS: How much to tip for a ride?
- CAB DRIVERS: Locality selection with frequent itemsets.
- CITY TRANSPORT AUTHORITY: Infrastructure planning, (Increasing Trend for fares due to traffic)

http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

[NYC Economic Brief - Office of the NYC Comptroller](#)

US Bureau of Labor Statistics - NYC Cab Statistics

THANK YOU