DATA ANALYSIS OF NYC CAB SERVICES

Abhinandan Dubey, Raju Khanal, Teja Madiraju
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?
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

- Brooklyn: 46.13334%
- Manhattan: 32.077143%
- Queens: 21.791522%

June 2016 Tip Affluence

- Brooklyn: 59.544047%
- Manhattan: 12.832105%
- Queens: 14.610475%
- The Bronx: 3.013372%
RESULTS

May 2016 Tip Affluence

- Brooklyn: 16.230214%
- Manhattan: 23.329312%
- The Bronx: 39.282253%
- Queens: 21.158220%
How much should you tip given a ride / location?

> Recommendation System

\[
tip_{user} = \frac{\sum_{i \in U} \text{sim}(user, i) + \text{tip}_{amount}(i)}{\sum_{i \in U} \text{sim}(user, i)}
\]
How much should you tip given a ride / location?

> Recommendation System

Some sample results include:

<table>
<thead>
<tr>
<th>Users</th>
<th>$10</th>
<th>$20</th>
<th>$50</th>
<th>$60</th>
<th>$80</th>
<th>$90</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M,Q)</td>
<td>NA</td>
<td>NA</td>
<td>$8.34</td>
<td>$12.11</td>
<td>$17.44</td>
<td>$19.9</td>
</tr>
<tr>
<td>(M,M)</td>
<td>$1.98</td>
<td>$3.97</td>
<td>$11.63</td>
<td>$13.59</td>
<td>$15.54</td>
<td>$21.53</td>
</tr>
<tr>
<td>(S, M)</td>
<td>NA</td>
<td>NA</td>
<td>$9.12</td>
<td>$11.85</td>
<td>$14.06</td>
<td>$20.25</td>
</tr>
</tbody>
</table>

M = Manhattan, Q = Queens, B = Brooklyn, S = Staten Island
NA indicates that minimum fare between places is greater than the value of the item.
Do disputed payments indicate something about the neighborhood?

\[
\frac{\text{(No of disputed payments in that borough)}}{\text{(No. of rides/payments made in that borough)}}
\]
June 2016 Disputed payments

- Staten Island: 48.098841%
- Manhattan: 9.528522%
- Queens: 14.683774%
- The Bronx: 14.723614%
- Brooklyn: 12.965249%
METHODS

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

\[
\frac{\text{(Total Tip across all rides in a month)}}{\text{(Total distance covered across all rides in a month)}}
\]
RESULTS

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?

**Dec 2015 Frequent Items**

- MQ: 500000
- Others: 0
- Qo: 100000
- BrM: 0
- MElr: 0
- MM: 2500000
- OI: 400000

**Frequent Itemsets**
Do People Travel In Groups During Festive Seasons?

- Approach 1: Averaging
- Approach 2: k-passenger (k > 1) rides
Do People Travel In Groups During Festive Seasons?

Festive Time - Christmas

- 0.49%
- 3.88% higher during these holidays.
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?

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

\[ y_t = \rho y_{t-1} + u_t \]
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?

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>-0.927128</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.778907</td>
</tr>
<tr>
<td>#Lags Used</td>
<td>6.000000</td>
</tr>
<tr>
<td>Critical Value (5%)</td>
<td>-2.967882</td>
</tr>
<tr>
<td>Critical Value (1%)</td>
<td>-3.679060</td>
</tr>
<tr>
<td>Critical Value (10%)</td>
<td>-2.623158</td>
</tr>
</tbody>
</table>
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)

NYC Economic Brief - Office of the NYC Comptroller


THANK YOU