Chicago Taxi Rides 2016

Team 10:

Noah Leuthaeuser | Joseph Alsko | Heesuk Jang | Rei Isobe

Questions Sought to Answer

- How do certain factors affect pickup frequency?
  - Weather: Temperature/Wind/Snow Depth/Precipitation
  - Date: Month/Day/Hour
    - Weekends vs. Weekdays
    - Holidays vs. Non-Holidays
  - Per Capita Income
- Which community areas give the most generous tips?
- Which exact locations describe the most frequency of pickup?
- Which routes describe the most frequency between pickup and dropoff community area?

Data Preparation Work

- **Data Cleaning:**
  - Removed non-applicable *columns*: Taxi ID, taxi company, tolls, etc.
  - Removed *rows* with null values for: Fare, trip seconds or miles, etc.
- **Data Reduction:**
  - Dimensionality & Numerosity Reduction:
    - *i.e.* Selecting/sampling only trip timestamp data on 3-5 top community areas
- **Data transformation:**
  - Parse/construct month/day/hour separate features from timestamp
- **Data integration**
  - Merged primary dataset with community area income data and weather conditions

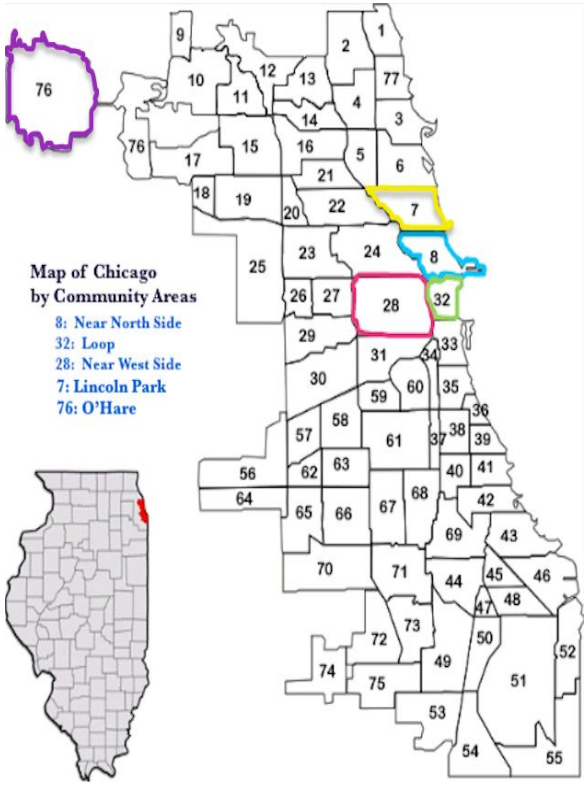
Tools Used

- **Python3/ iPython**
  - Pandas/Numpy/Scipy
  - Parse *from* dateutil.parser
  - Shapefile
  - KMeans *from* sklearn.cluster
  - Pdist and Squareform *from* scipy.spatial.distance
  - Itemmining *from* pymining
  - Plotly/Matplotlib
- **Git/Github/Slack**

Relim (Recursive Elimination) **Algorithm:** Frequent Itemset

Five most frequently traveled community areas for pickup and dropoff

Area 1	Area 2	Count
Near North Side	Loop	2281451
Near North Side	Near West Side	899440
Near West Side	Loop	814120
Lincoln Park	Near North Side	475912
Near North Side	O'Hare	453597



Relim (Recursive Elimination) **Algorithm:** Frequent Itemset

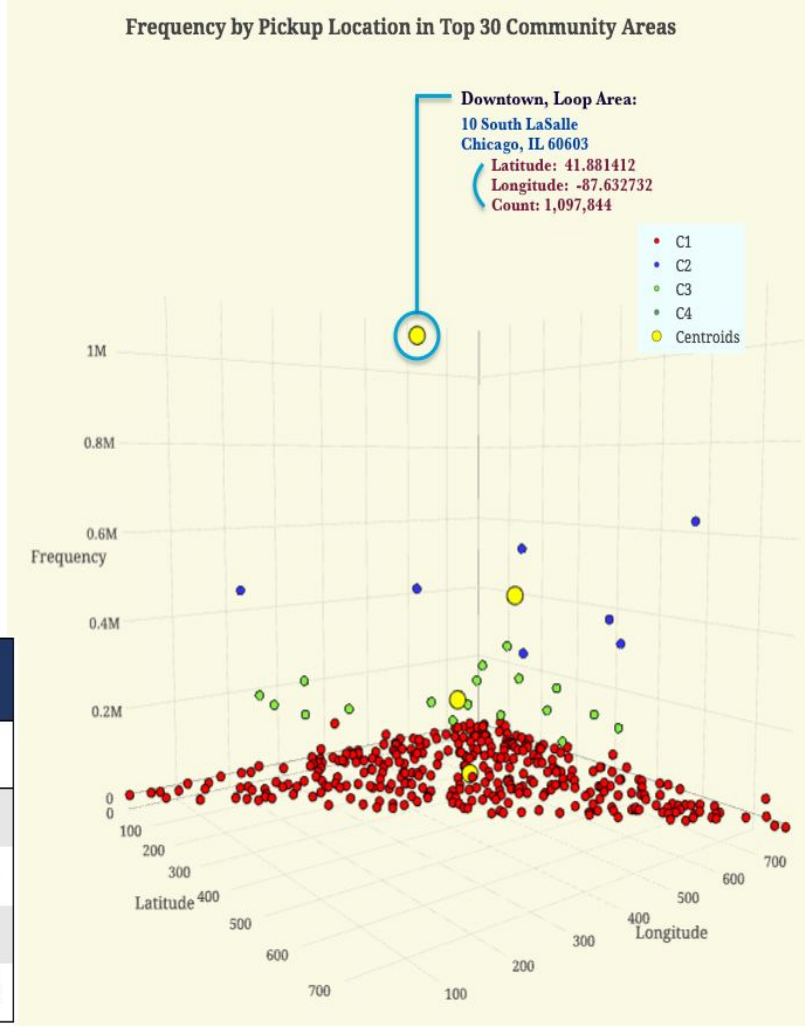
The Most Frequent Three Pickup Locations in Top 3 Community Areas				
Area	**Latitude**	**Longitude**	**Count**	**Approximate Location**
Near North	**41.9129**	**-87.7614**	**555,344**	**1759 N Lotus Ave. Chicago, IL 60639**
(8)	41.8925	-87.6262	456,092	43 E Ohio St., Chicago, IL 60639
	41.7889	-87.6197	378,455	5806 S Prairie Ave., Chicago, IL 60637
Loop	**41.8810**	**-87.6327**	**1,097,844**	**10 S LaSalle St., Chicago, IL 60603**
(32)	41.9788	-87.6538	628,970	5301 N Sheridan Rd, Chicago, IL 60640
	41.8789	-87.6252	306,818	Theodore Thomas Orchestra Hall, Chicago, IL 60640
Near West	**41.8793**	**-87.6426**	**328,605**	**621-601 Historic U.S. 66, Chicago, IL 60661**
(28)	41.8853	-87.6428	199,683	177-199 N Jefferson St., Chicago, IL 60661
	41.8740	-87.6635	176,340	1445 W Harrison St., Chicago, IL 60607

Note: *An approximate pickup location is defined by its latitude and longitude.*

K-Means: Clustering Pickup Location with its Frequency

- **Elbow Method:** the optimal k = 4
- **Final Centroids** are located in four different community areas.
- **Cluster 4** contains a single data point with an abnormally high frequency

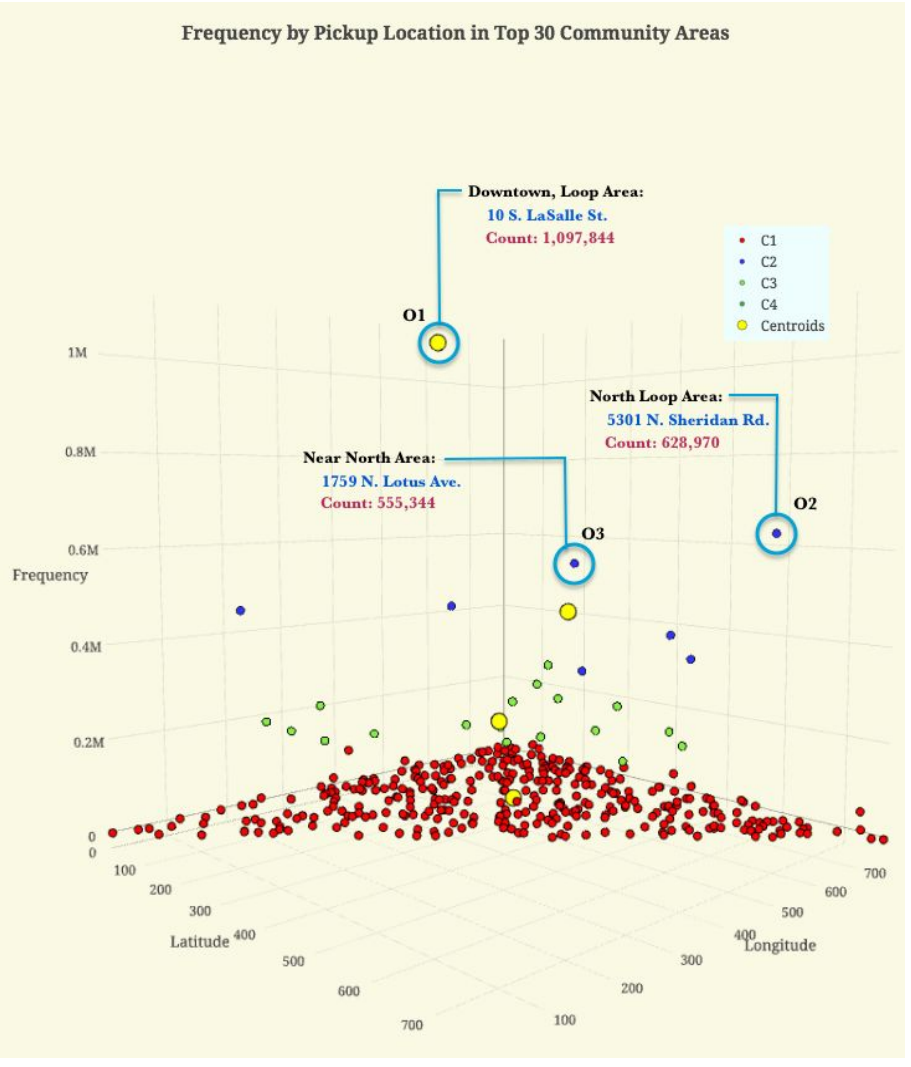
Centroids : K-Means Clustering				
	Latitude	Longitude	Count	Approximate Location
Cluster 1	41.77643	-87.70121	4,909	6426 S. Albany Ave, Chicago IL 60629
Cluster 2	41.95340	-87.66419	449,089	3958-3932 N Southport Ave, Chicago IL 60613
Cluster 3	41.90838	-87.66128	186,301	1460 N Elston Ave, Chicago IL 60642
Cluster 4	41.88099	-87.63275	1,097,844	10 S. LaSalle St, Chicago IL 60603 *(Downtown)*



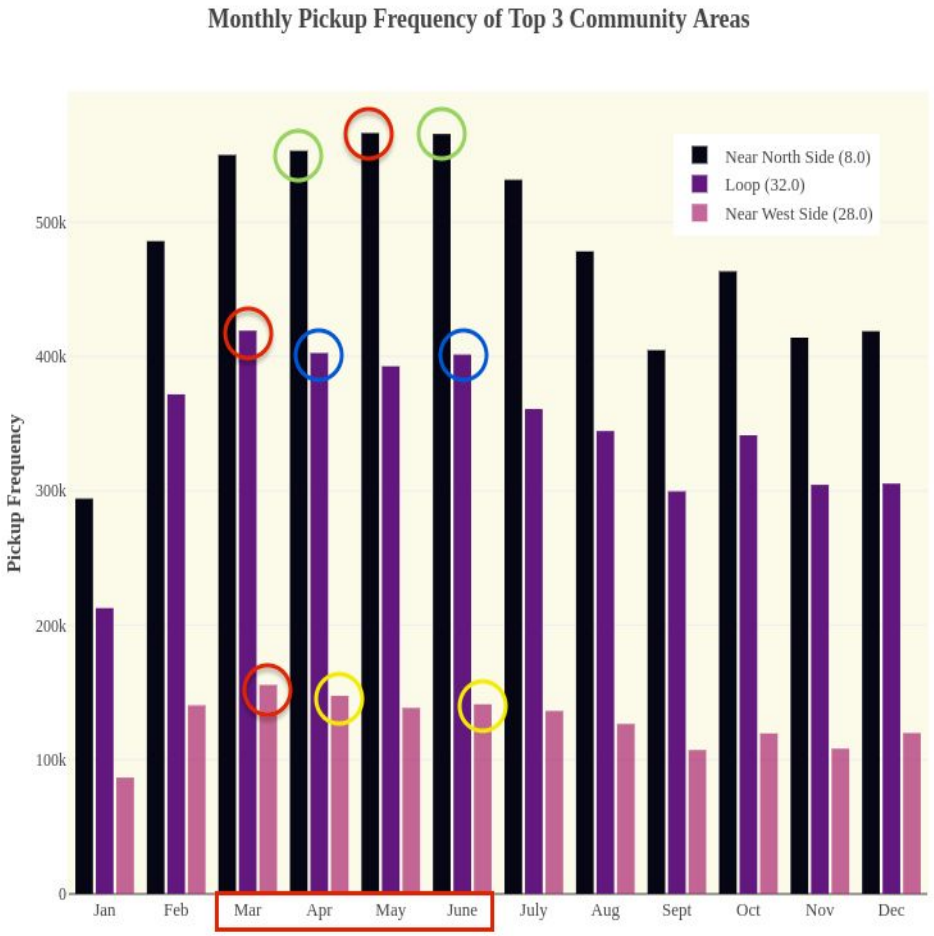
K-Means: Detecting Outlier

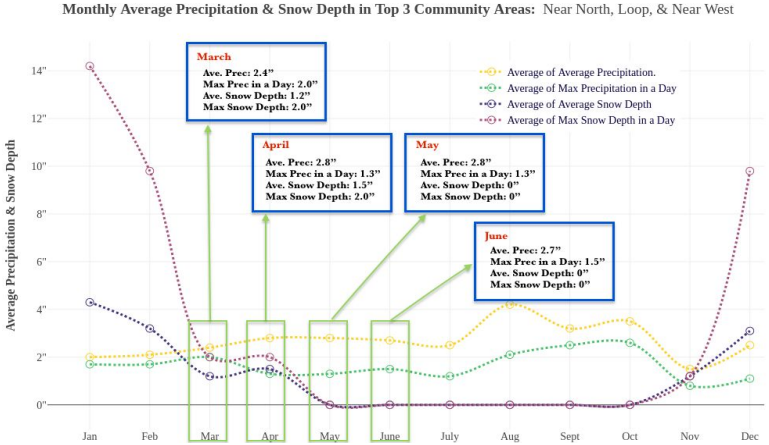
- **Euclidean Distance Approach:**
  Threshold: (min dist+max dist)/2
  Distance > Threshold: Outlier

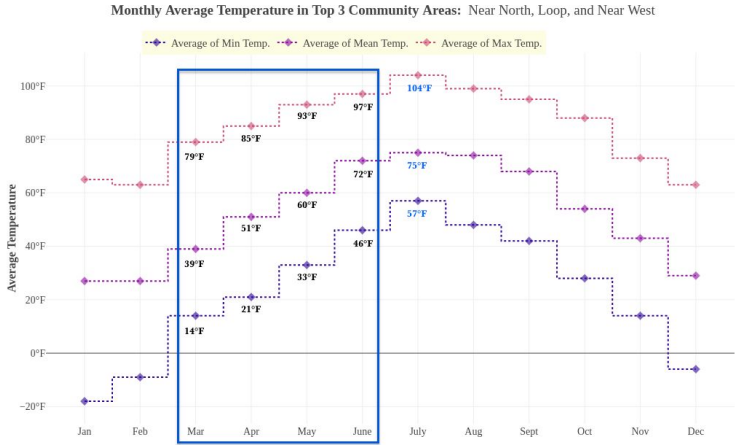
- **Detection of Three Outliers:**
  O1: Loop, Downtown
  O2: North Loop
  O3: Near North

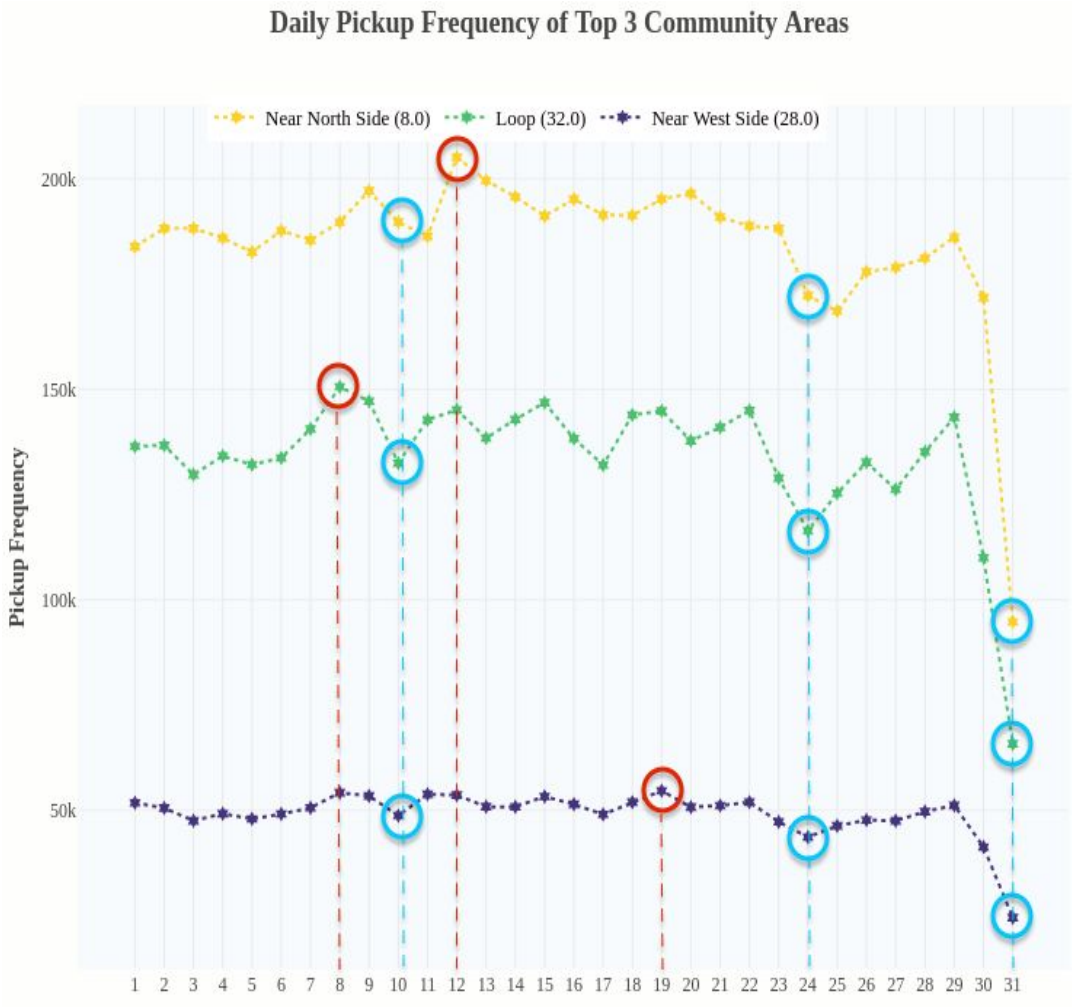


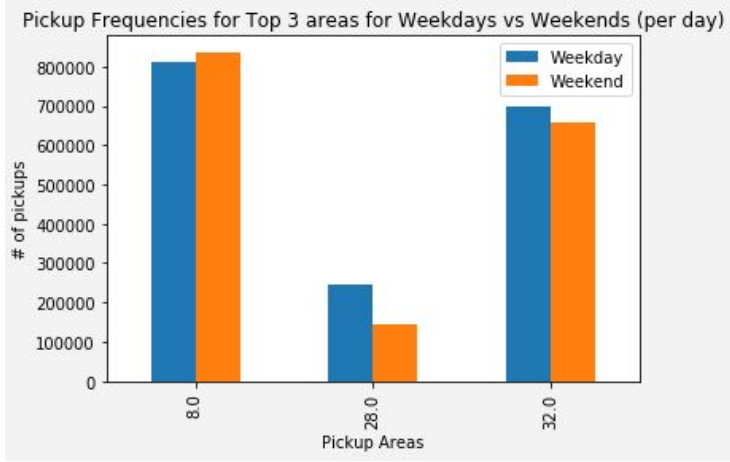
Correlation: Weather vs. Monthly Pickup Frequency

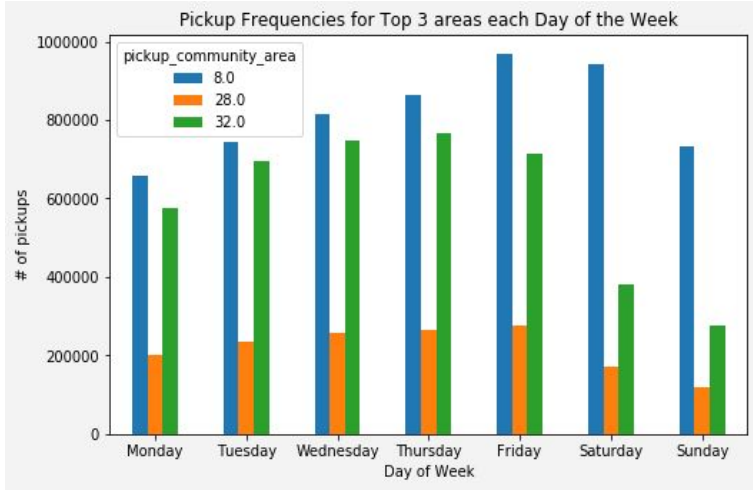


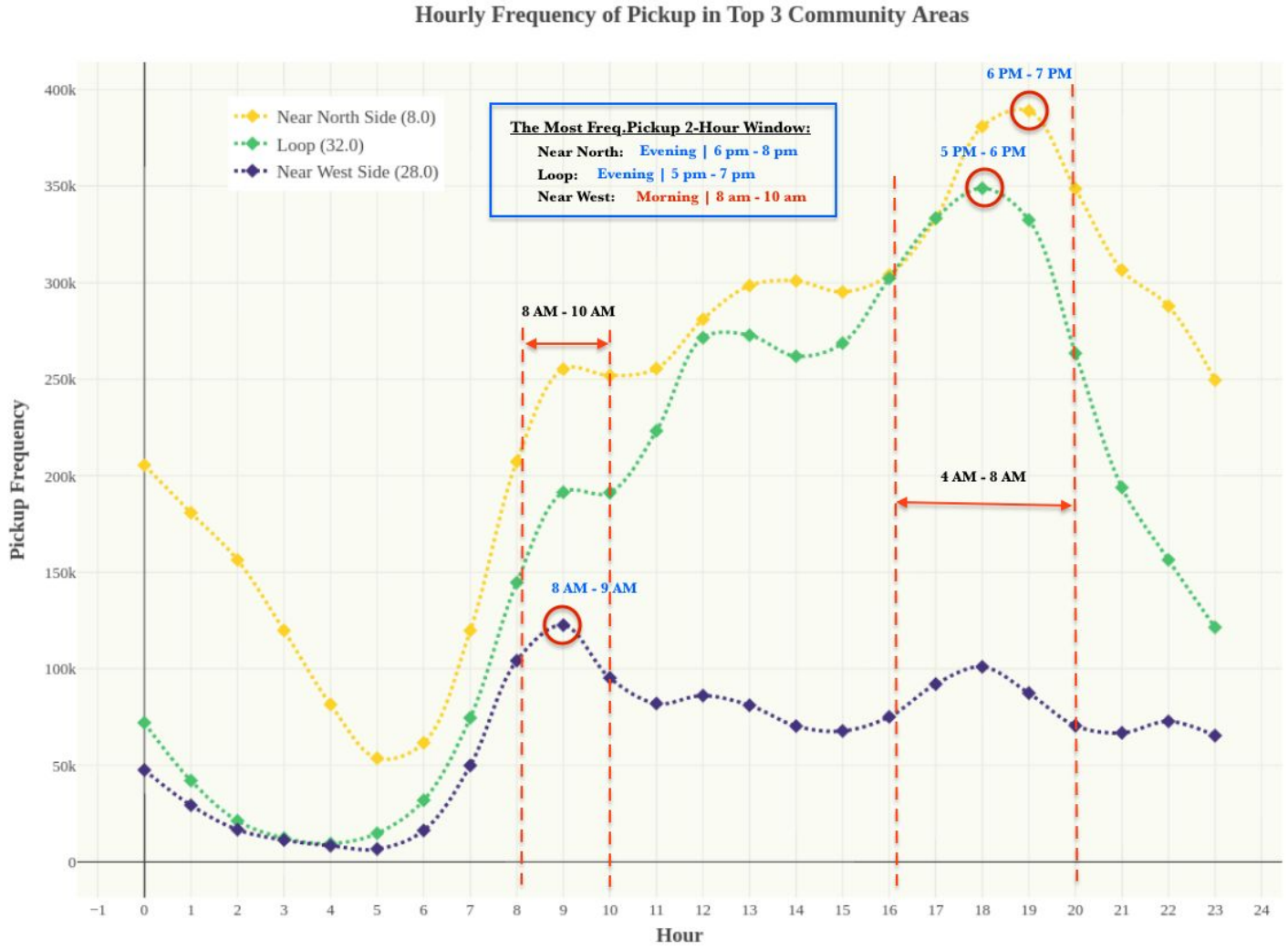





Daily Pickup Frequency of Top 3 Community Areas
Near North Side (8.0)
Loop (32.0)
Near West Side (28.0)
Pickup Frequency
200k
150k
100k
50k
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31


Pickup Frequencies for Top 3 areas for Weekdays vs Weekends (per day)
Weekday
Weekend
# of pickups
800000
700000
600000
500000
400000
300000
200000
100000
0
8.0
28.0
32.0
Pickup Areas


Pickup Frequencies for Top 3 areas each Day of the Week
pickup_community_area
8.0
28.0
32.0
# of pickups
1000000
800000
600000
400000
200000
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week


Hourly Frequency of Pickup in Top 3 Community Areas
Near North Side (8.0)
Loop (32.0)
Near West Side (28.0)
The Most Freq.Pickup 2-Hour Window:
Near North: Evening | 6 pm - 8 pm
Loop: Evening | 5 pm - 7 pm
Near West: Morning | 8 am - 10 am
6 PM - 7 PM
5 PM - 6 PM
8 AM - 10 AM
4 AM - 8 AM
8 AM - 9 AM
Pickup Frequency
Hour

Knowledge Gained

- Best tips by community area
- Frequently traveled routes
- Pickup frequency by location
- Time series trends in pickup frequency
- Factors that correlate with pickup frequency
    - Area income, weather, weekends/weekdays, time

Application of knowledge

Taxi drivers can know:

- The best community areas to target based on:
  - Tips, pickup/dropoff frequency, safety, predicted drop-off
- The best pickup locations within community areas
- The best dates/times to work community areas

Best tips: **The Loop**

Highest ride frequency:
Near North Side

Month with most rides:
March