

# EXPLORATION OF UBER PICKUPS JULY 2014

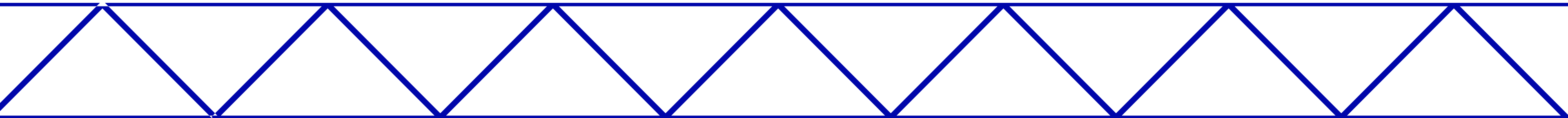
IBM DATA SCIENCE CAPSTONE PROJECT

Diego Avila  
[diego.avila.calderon@protonmail.com](mailto:diego.avila.calderon@protonmail.com)



# Usefulness of location data

- Uber uses location data to efficiently assign passengers to drivers
- Drivers don't have the aggregated data readily available
- Allocation of resources helps minimize driver's inefficiency and cut costs (fuel) by knowing where to locate passengers

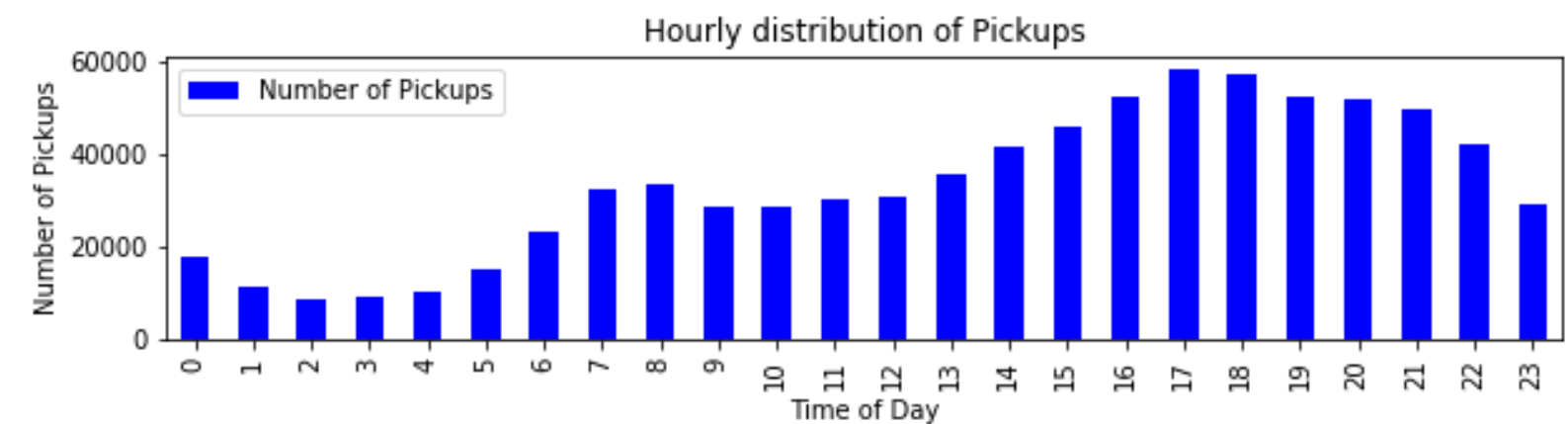
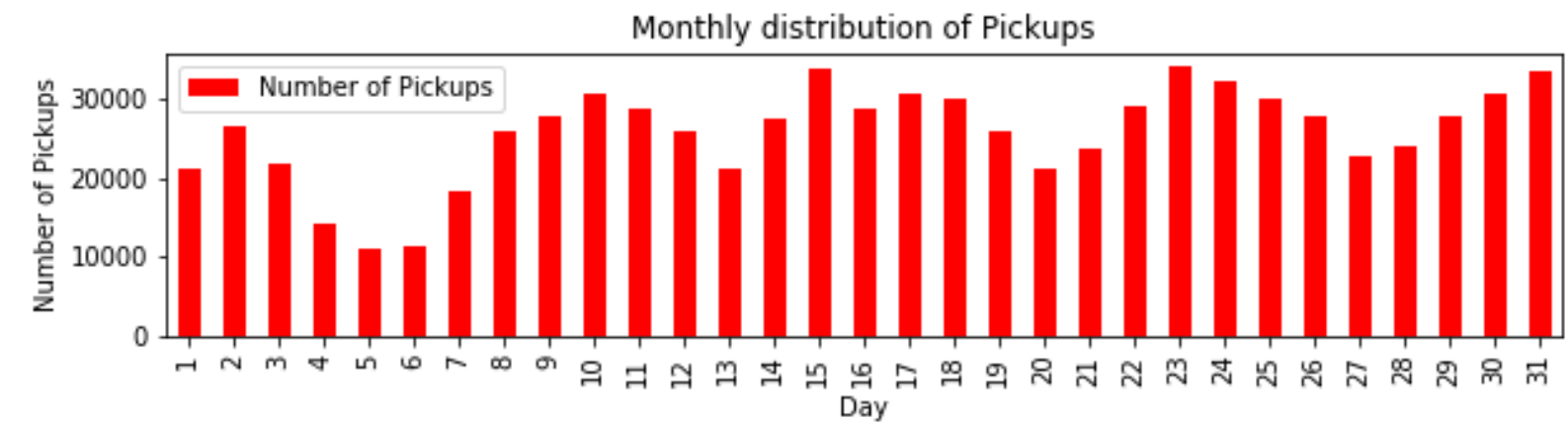
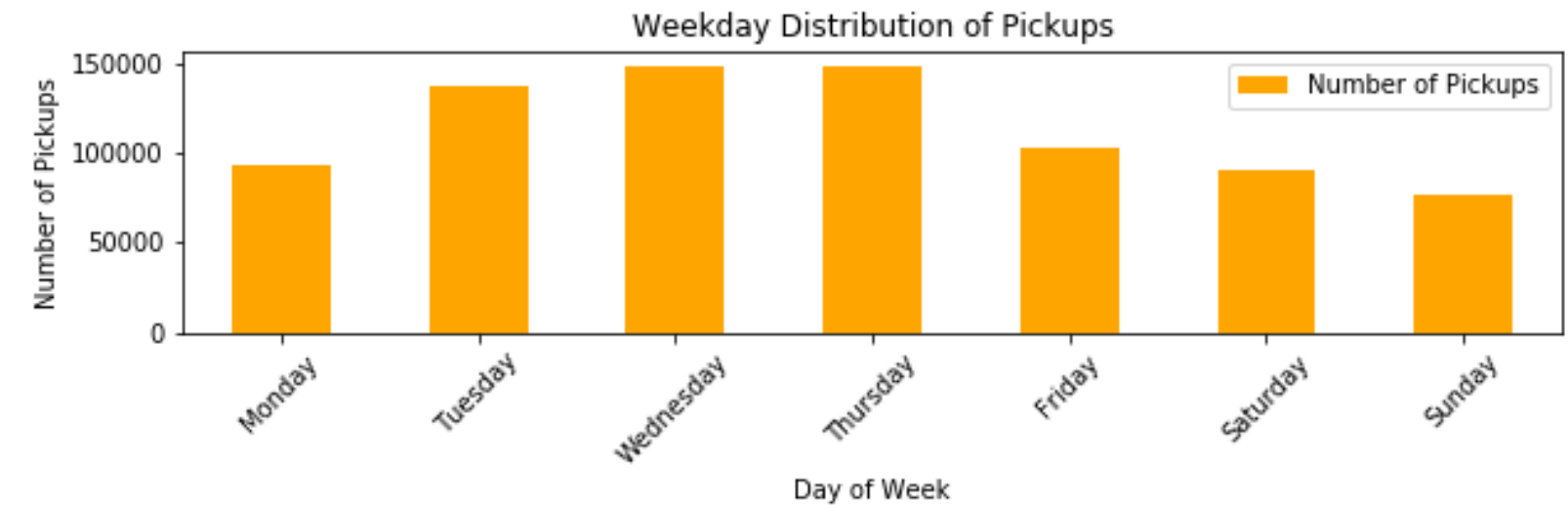


# Data Collection and Cleaning

- Uber trip data New York City 2014 (Kaggle): geocoordinates of pickups, time & date information (converted to datetime)
- Over 795,000 entries
- Foursquare API (venues, venue category)
- NYC Neighborhoods GeoJson: coordinates, borough name, neighborhood name
- NYC Boundary limits of postal codes GeoJson (Github)
- For Choropleth Maps, a random sample of 600 entries was taken

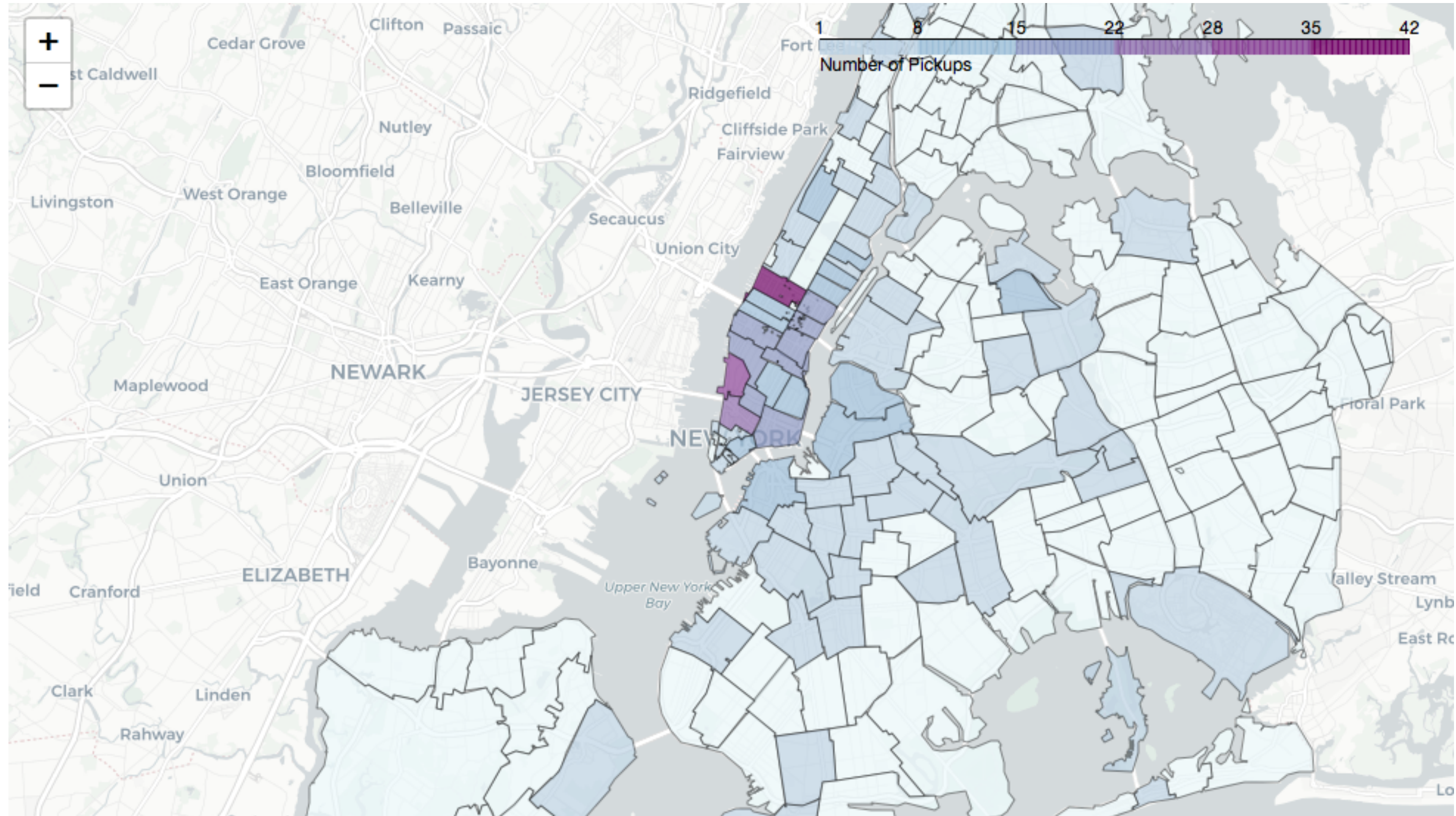
# Datetime Distribution

- Weekdays are busier than weekends and Monday
- Sinusoidal distribution throughout the month
- Holidays and long weekend inhibit less activity (Ex: 4th of July)
- Pickups show greater activity at rush hour times (7-8 am and 6pm).



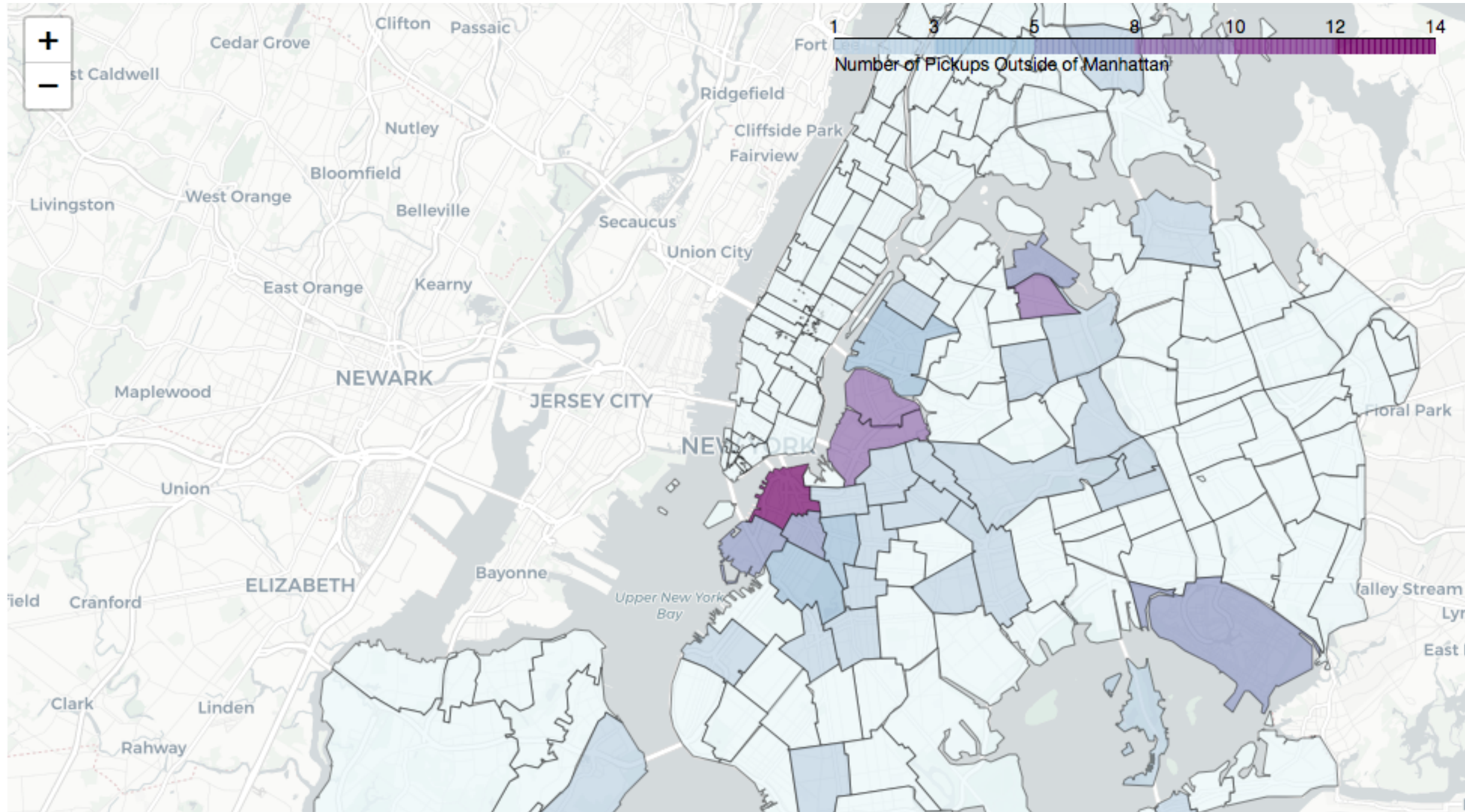


# Location Distribution





# Distribution without Manhattan



- Besides Manhattan, Brooklyn inhibits most pickup activity
- Areas near the bridges to Manhattan are important pickup areas

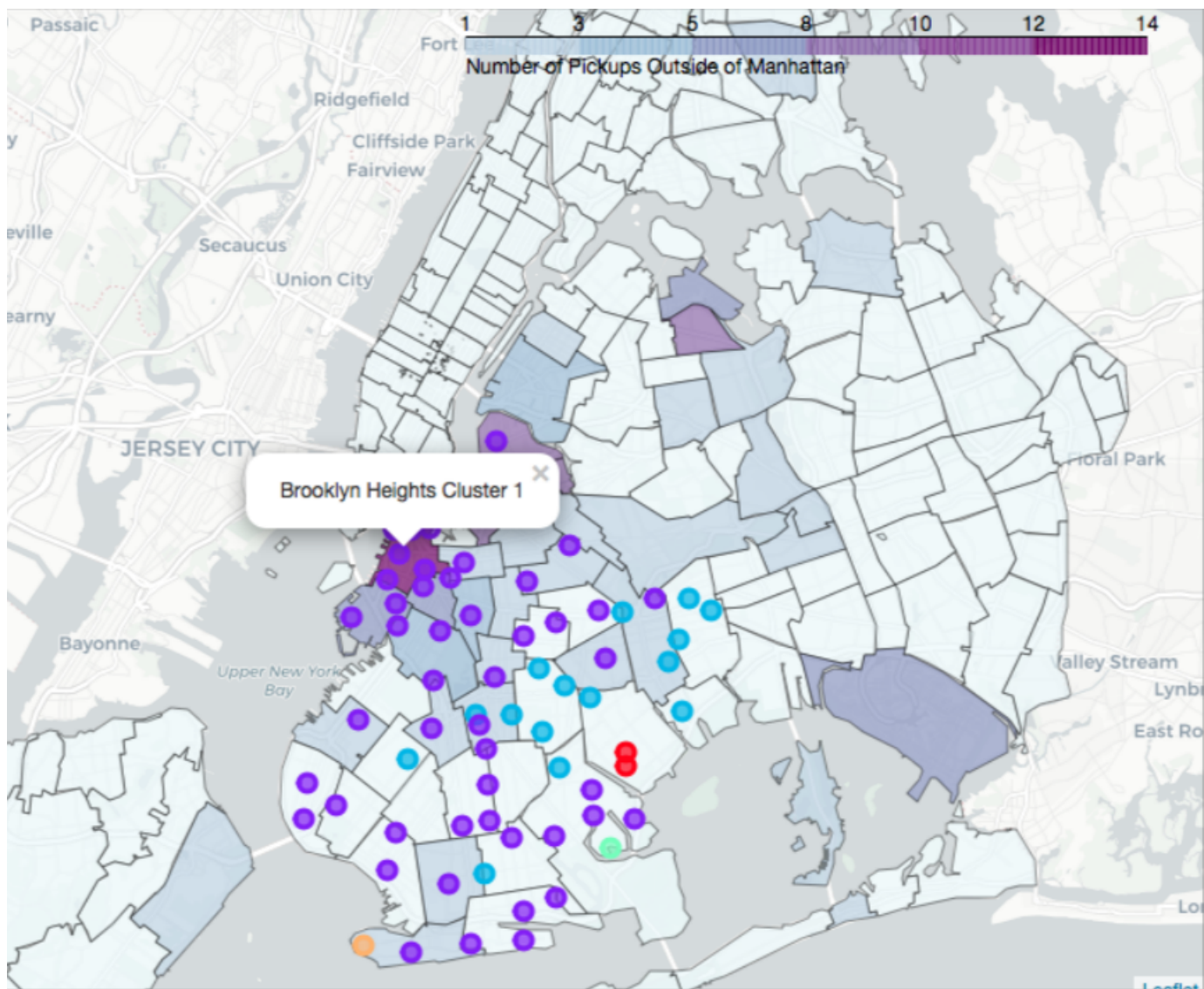
# K-Means Clustering the Neighborhoods

- K-means clustering is an unsupervised machine learning algorithm that groups data entries into groups, known as clusters, through the calculation of distance between cluster centroids
- Neighborhoods were grouped by similarities between their 10 most common venues
- Allows us to see what venue types are commonly found in areas of interest

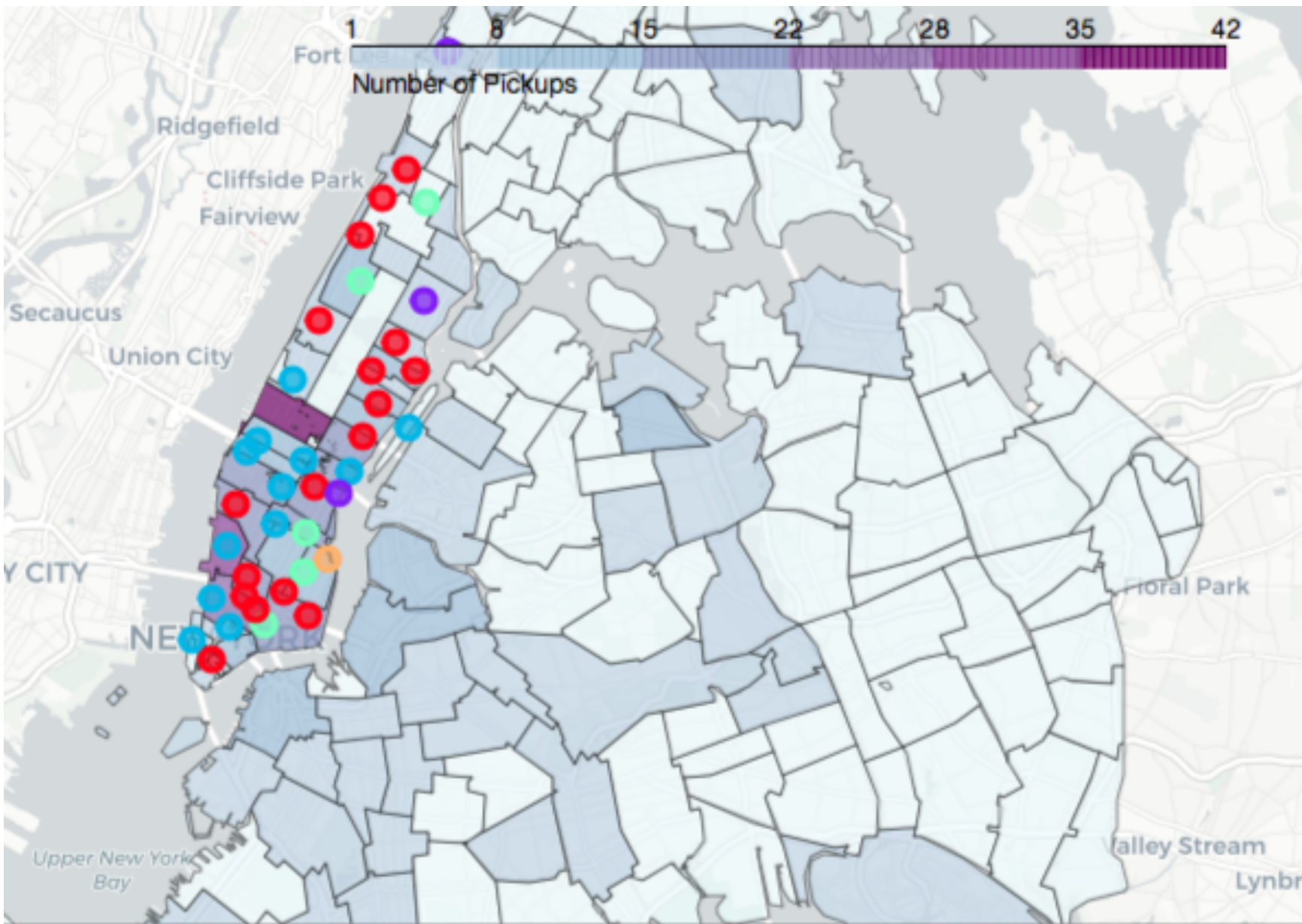


# Clustered Neighborhoods Results

Brooklyn:



Manhattan:





# Venue Types

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Battery Park City	2	Park	Hotel	Coffee Shop	Gym	Memorial Site	Boat or Ferry
5	Civic Center	2	Coffee Shop	Hotel	Gym / Fitness Center	Cocktail Bar	French Restaurant	Yoga Studio
6	Clinton	2	Theater	Italian Restaurant	Gym / Fitness Center	Wine Shop	American Restaurant	Coffee Shop
10	Flatiron	2	Gym / Fitness Center	Vegetarian / Vegan Restaurant	Cycle Studio	Mediterranean Restaurant	Italian Restaurant	New American Restaurant
14	Hudson Yards	2	American Restaurant	Gym / Fitness Center	Hotel	Italian Restaurant	Café	Coffee Shop

Two different cluster types were identified as important

- Cluster 1: Near Hotels with venues for targeted to locals (Gym, parks)

- Cluster 2: Neighborhoods big on the cuisine, lots of choice for eating
- Coffee shops are abundant

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Carnegie Hill	0	Coffee Shop	Café	Italian Restaurant	Pizza Place	Yoga Studio
3	Chelsea	0	Coffee Shop	Art Gallery	American Restaurant	French Restaurant	Bakery
9	Financial District	0	Coffee Shop	Cocktail Bar	Bar	American Restaurant	Pizza Place
12	Greenwich Village	0	Italian Restaurant	Sushi Restaurant	Café	Clothing Store	Indian Restaurant
13	Hamilton Heights	0	Pizza Place	Café	Coffee Shop	Deli / Bodega	Mexican Restaurant

# Conclusion

- Weekdays inhibit more activity than other days of the week
  - Afternoons, near rush hour are the time of day with most pickup requests
- 
- Besides Manhattan, Brooklyn and areas closer to Manhattan show the most activity
  - Neighborhoods with hotels
  - People get taxis in areas with lots of food-related venues