Citibike and its Impact in NYC

Maikol Cerda (Postdoc, Yale University)

Data Incubator Interview

May 14th, 2020

Goals and Databases

1. Understand the Demand's Dynamics and Demand's Forecasting
   - Final User: Lyft.
   - Optimization of Resources and Better Competitive Position
2. Quantify the impact of bike-paths on City's Development (2020 expansion to the Bronx)
   - Final user: Government (NYC) and Lyft
   - Urban economics and strategic expansion

- Time period: 2015-2019

Dataset	Freq	Size	Features
1. Citibike Trips	Daily	13.3GB	Trips duration, start-stop times/stations
2. Taxis and ridesharing trips	Monthly	0.8GB	Trips per-day, vehicles per-day, avg minute per-trip
3. Weather conditions	Daily	1.5GB	Temperature, wind speed, precipitation, humidity
4. Tourists-passengers	Monthly	40MB	Number of domestic and international visitors
5. Pollution and rental prices	Daily-Monthly	400MB	Main pollutants, overal AQI

1.a Demand's Dynamics and Demand's Forecasting

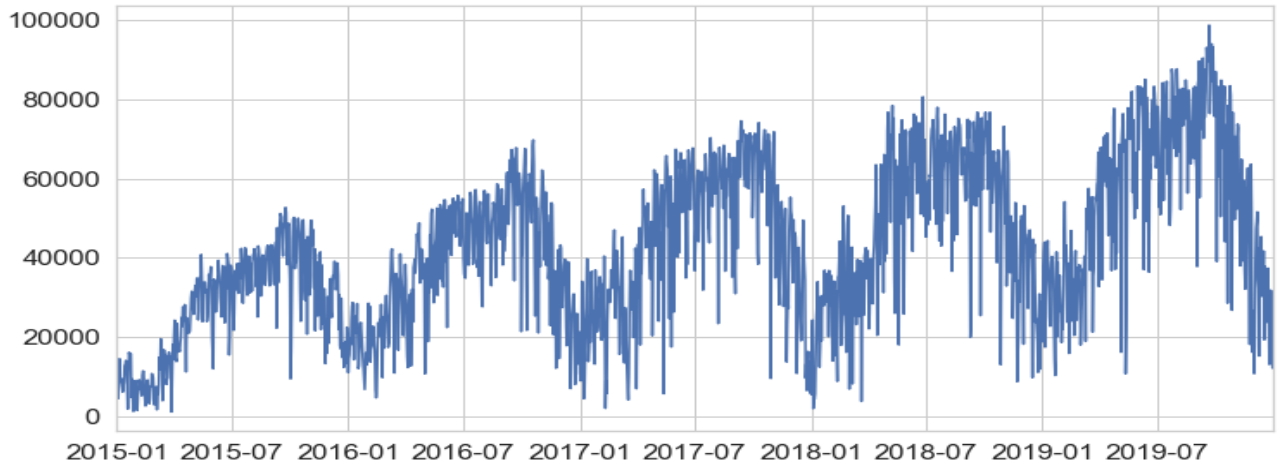


Figure 1: Number of trips per-hour, 2015-2019

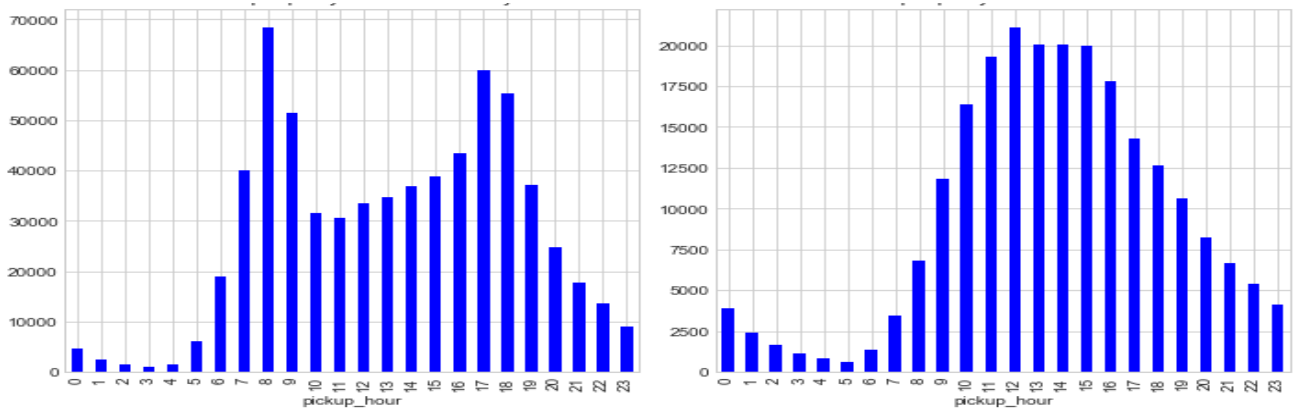


Figure 2: Daily Seasonality: Trips per-hour week day (left) vs weekend (right)

1.a Demand's Dynamics and Demand's Forecasting

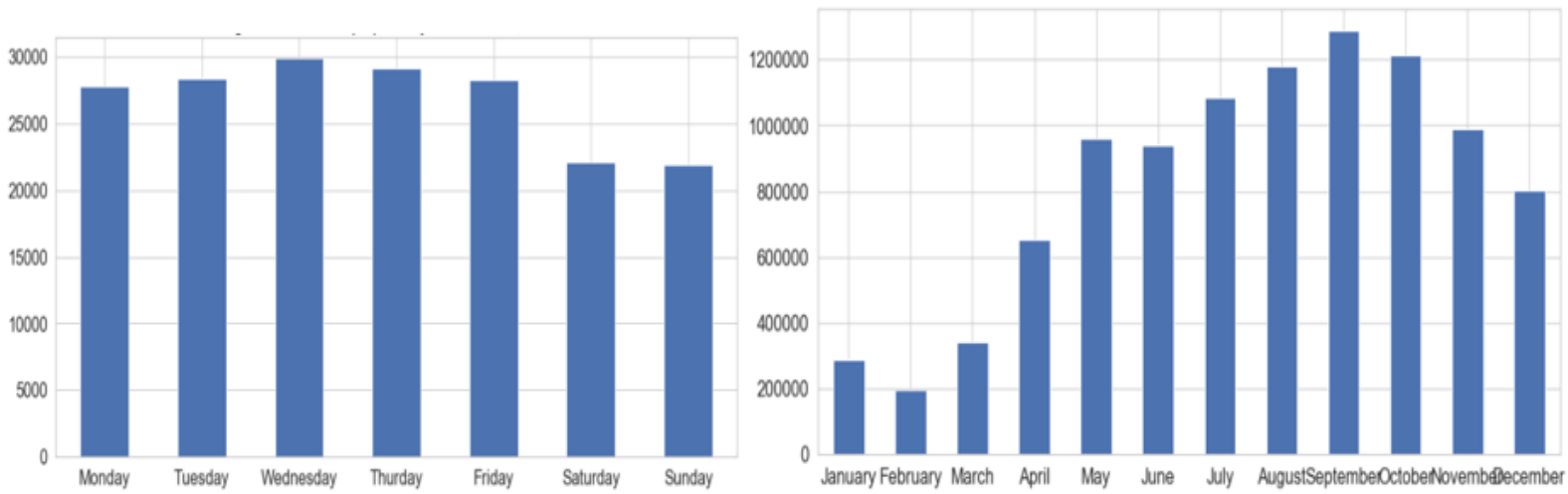


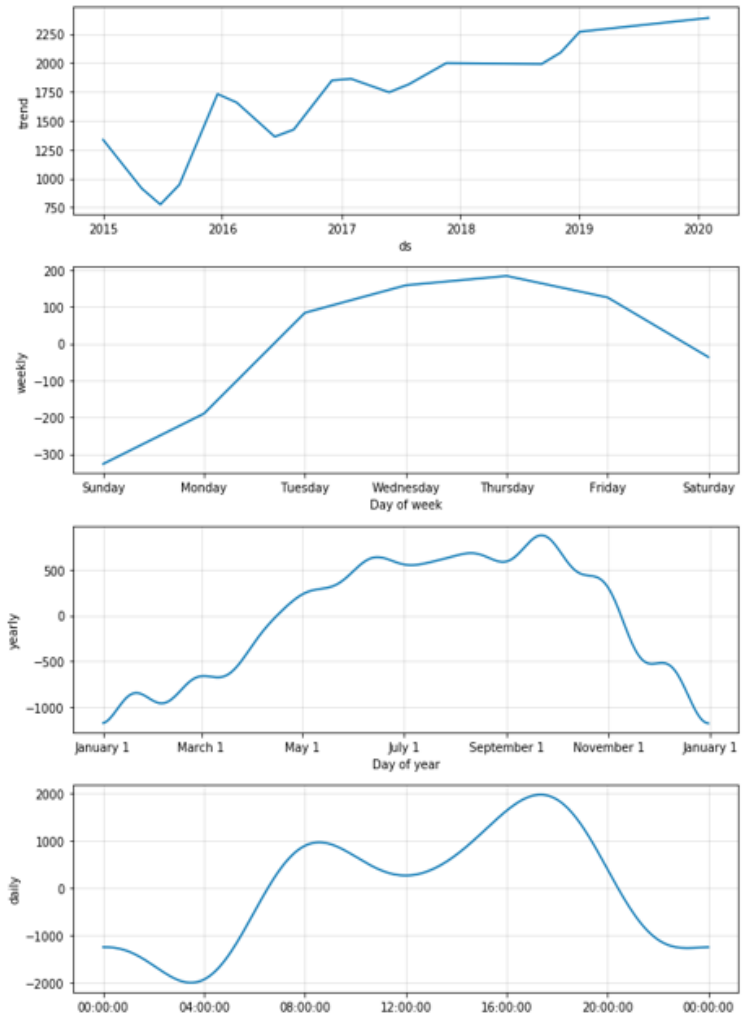
Figure 3: Monthly (left) and Weekly (right) Seasonalities: average number of trips

- Prophet, Time Series Forecasting:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- Includes:
  - $g(\cdot)$: aggregate long-term trend
  - $s(\cdot)$: seasonalities with Fourier series
  - $h(\cdot)$: impact of holidays

1.a Demand's Dynamics and Demand's Forecasting



Model	Specification	RMSE errors
SARIMAX	(1,2,1)x(4,0,2,12)	9.1565
LSTM	(2,2)	8.6743
Prophet		8.2320

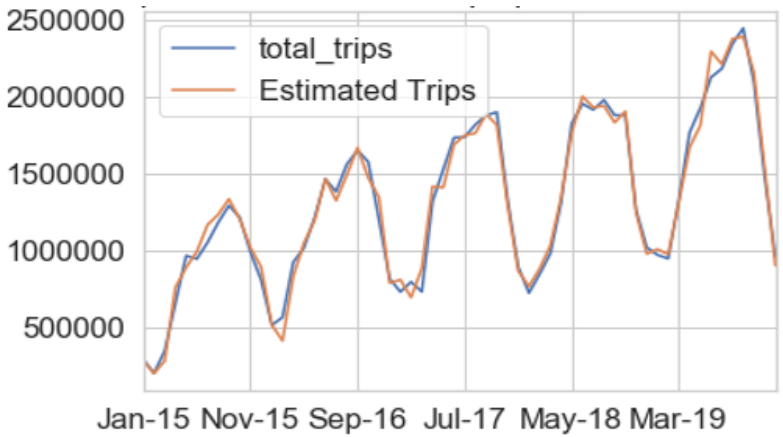
Figure 4: Prophet's Forescasting (left) and Model evaluation (right)

1.b Explanatory Analysis

- Model for the number of trips:

$$y = f\left(\overrightarrow{x}\right) = f\left(\#\text{ visitors, T, rain, substitutes, season, gender}\right)$$

- Approximation: f is quadratic



total_trips	Coef.	Robust Std. Err.	t
dom_passengers	.1790183	.1414179	1.27
int_passengers	-.1756811	.1350174	-1.30
t_avg	.2654983	.0942052	2.82
rain	-.0510522	.0241002	-2.12
taxi_vehicles_per_day	-.2307219	.1378688	-1.67
_cons	.2542904	.0292383	8.70

Figure 5: Estimation fitting (left) and results (right)

- Gradient boosting decision tree
  - Combine several simple tree models to achieve optimized prediction performance
  - Interpretation of the model results by identifying the key explanatory variables

1.c Microanalysis and Optimization of Resources

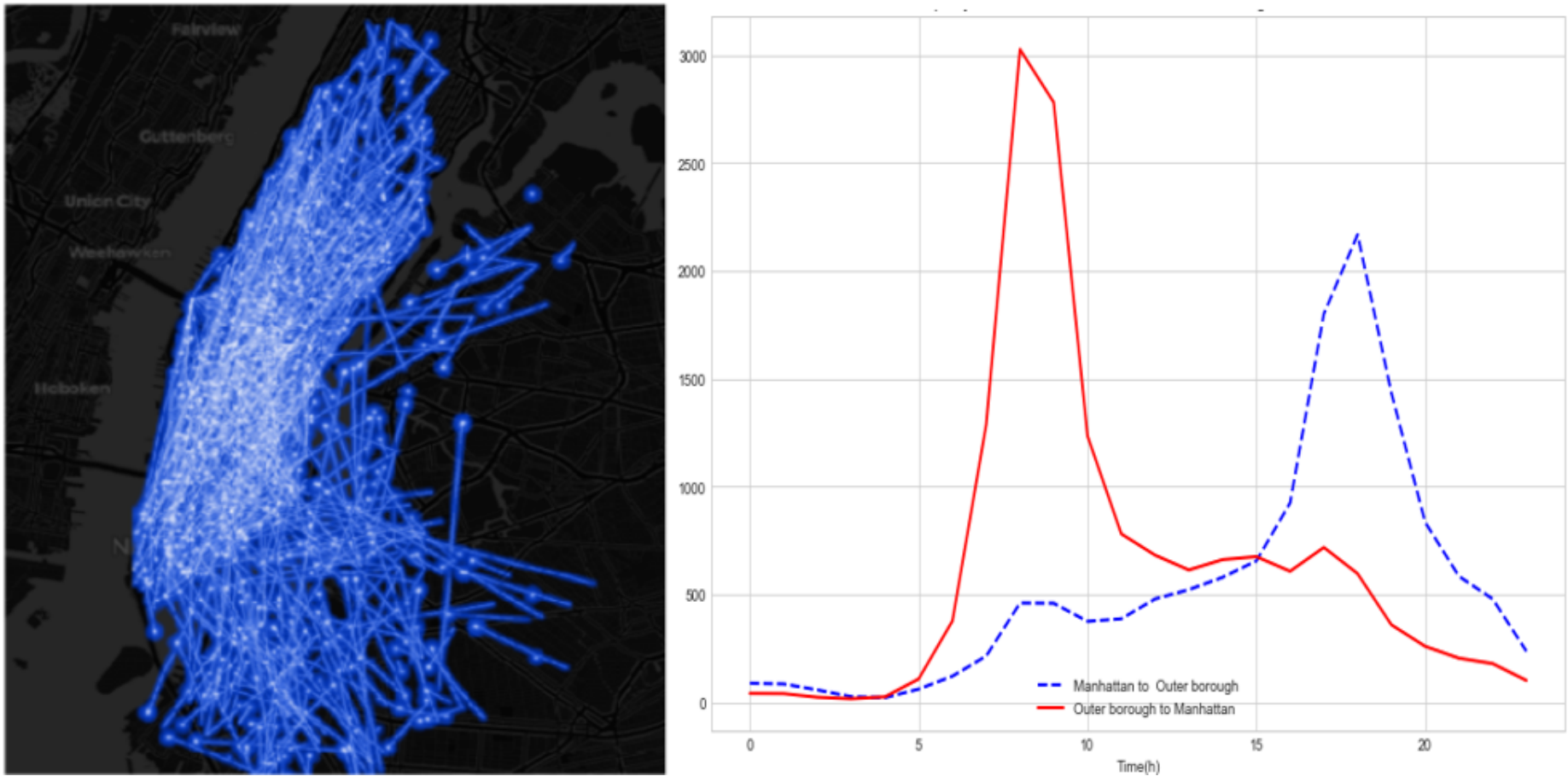


Figure 6: Relevance of routes (left) and trips among borough (right)

1.c Microanalysis and Optimization of Resources

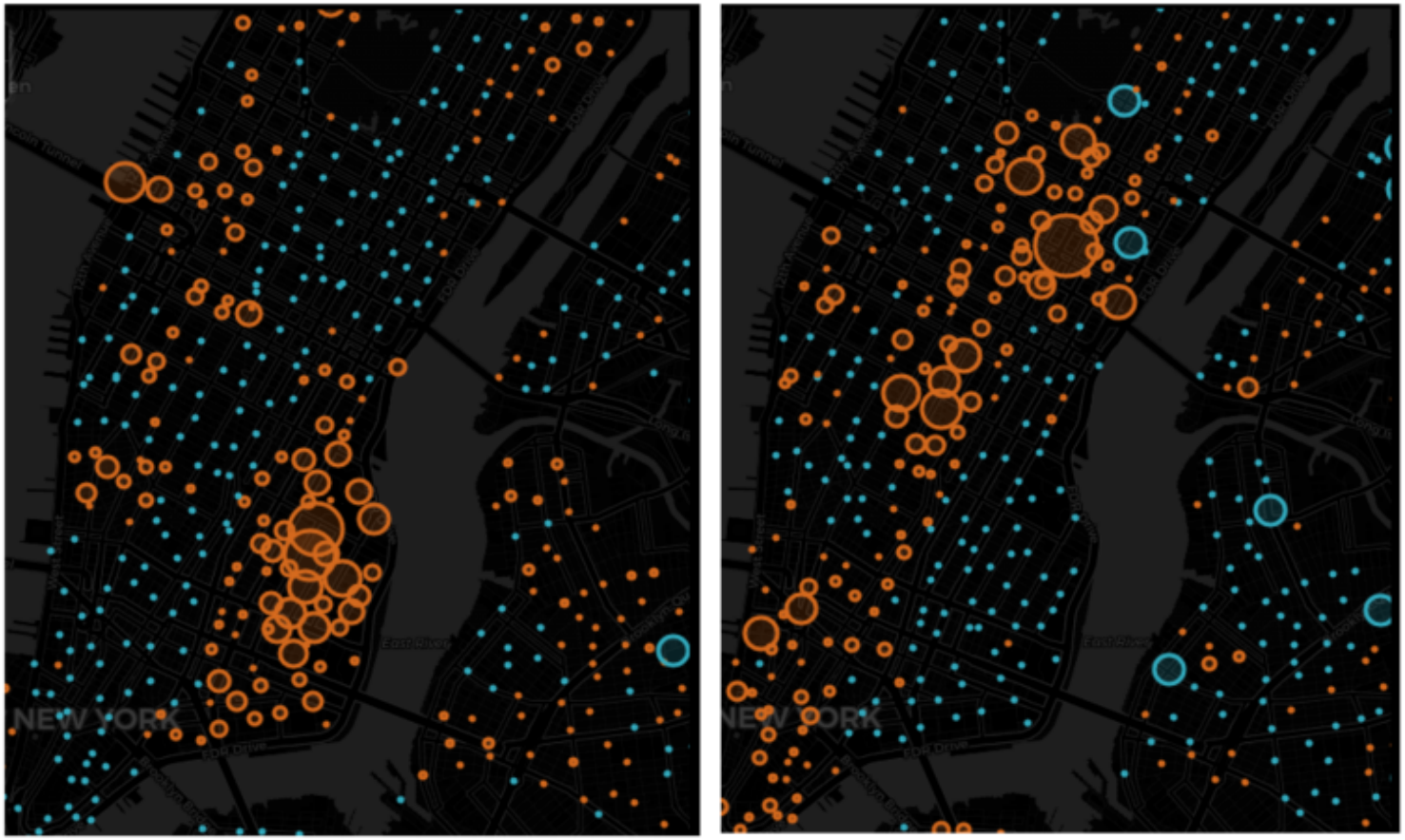


Figure 7: Net Departures by station, 9am (left) vs. 6pm (right). (Orange: departures>arrivals ; Blue: otherwise)

2.a Citibike and City's Development

- Citibike and its impact on other variables of interest:

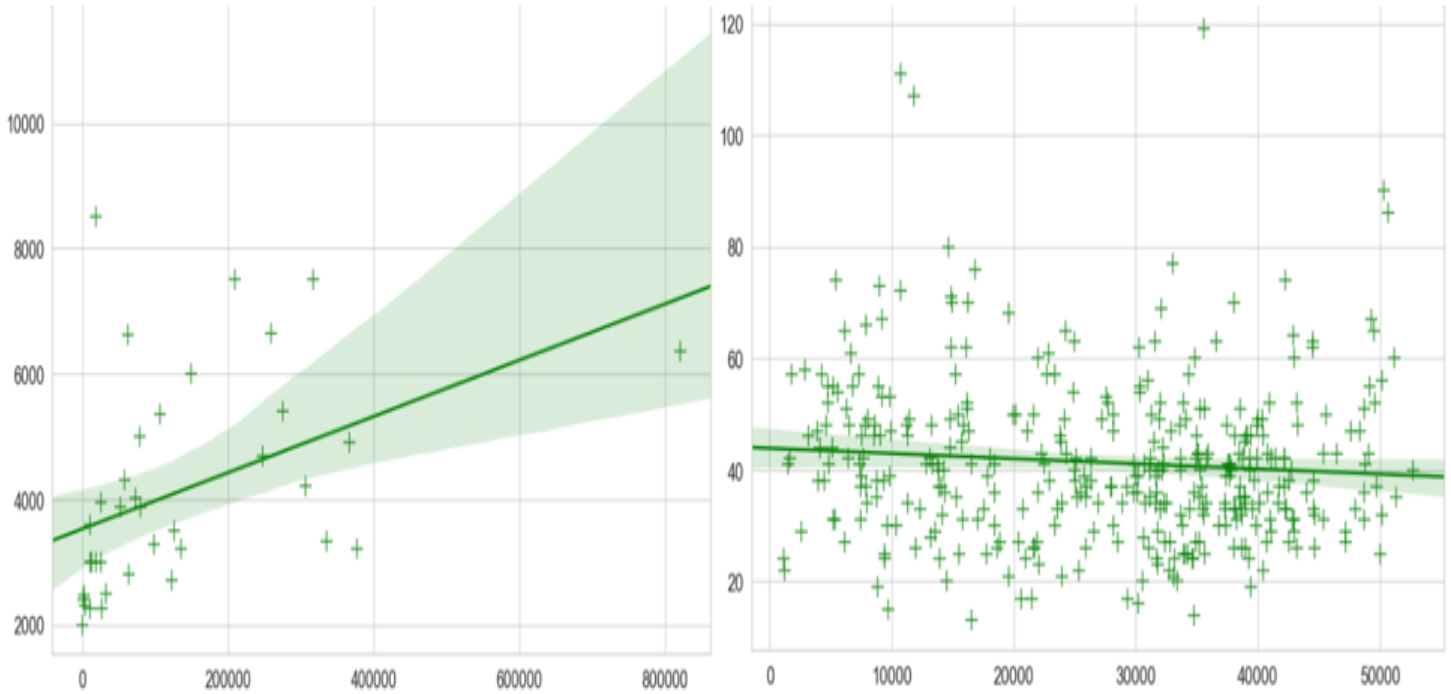


Figure 8: Avg. rental prices by neighborhood (left) and daily NO2 pollutant level as a function of the number of trips

2.b Citibike and Expansion's Plan

- Lyft vs. Taxis rides:

	service_zone	Zone	Borough
PULocationID			
168	1368	Mott Haven/Port Morris	Bronx
247	891	West Concourse	Bronx
69	668	East Concourse/Concourse Village	Bronx
213	639	Soundview/Castle Hill	Bronx
159	636	Melrose South	Bronx

- The planned expansion through 2023 will include the following neighborhoods in the Bronx:
  - Mott Haven, Melrose, Port Morris, Highbridge, Claremont, Morrisania, Longwood, Concourse, and Mt. Eden
  - Source: https://www.citibikenyc.com/blog/major-citi-bike-expansion-map-revealed

2.b Citibike and Expansion's Plan

- Search Problem:

$$\{y_i\}_{i=1}^{N} = \arg\min \sum_{i=1}^{N} n_i \sum_{j=1}^{3} \omega_{ij} d_{ij}$$

$$\text{s.t.} \sum_{j=1}^{3} \omega_{ij} = 1 \; ; \sum_{i=1}^{N} y_i \leq S \; ; \omega_{i1} > \omega_{i2} > \omega_{i3}$$

- where: N=neighborhoods ; $y_i \in \{0, 1\}$; S=max of stations