



citi bike

Lorena De La Parra Landa

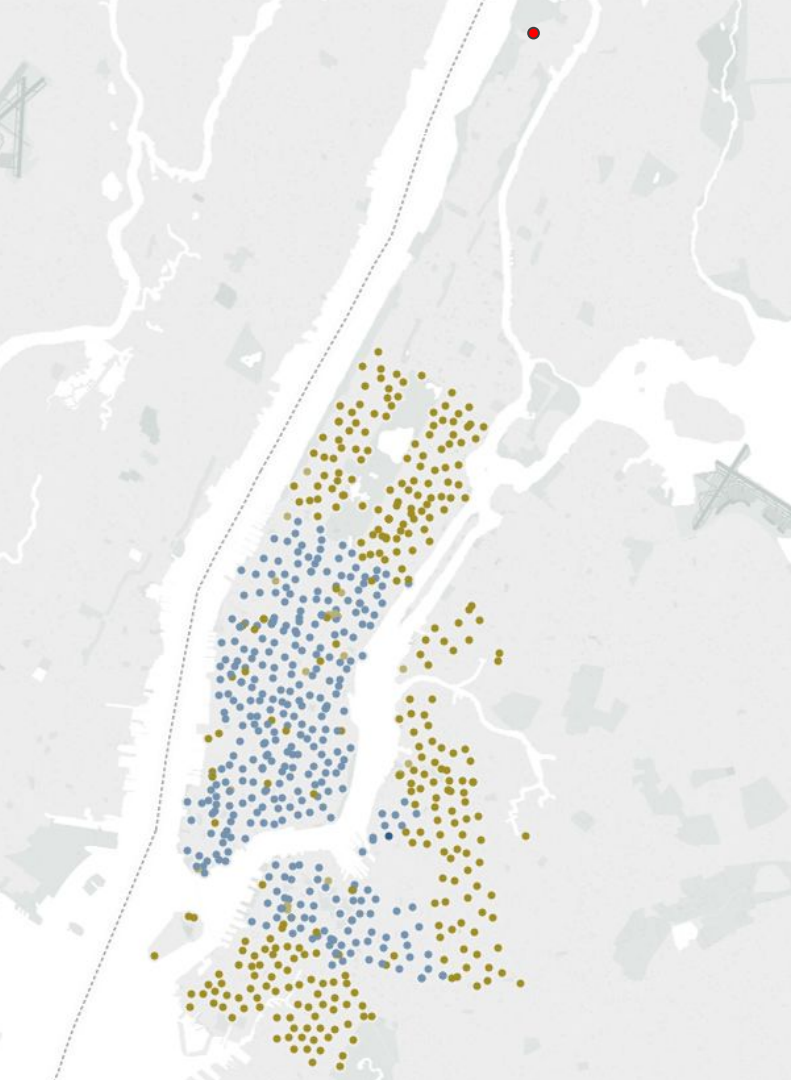
est. 2013



36,902,025



~700 stations



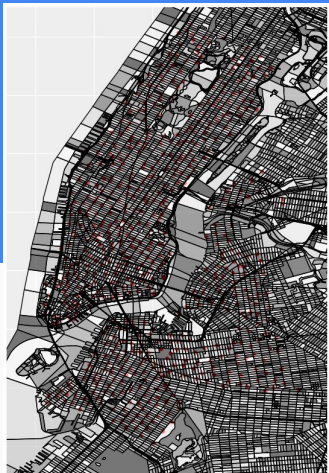
**Will I have a bike station
on my block?**

**Can I predict if a block
has a station?**

Data



Block Features



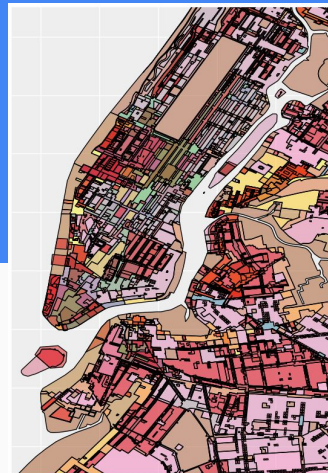
Stations

Location, age, distance, monthly average, and total counts on the three closest Citibike stations.



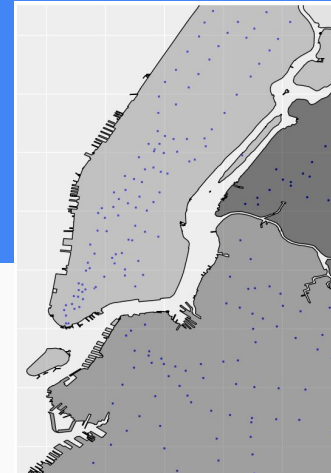
Zipcodes

Block zipcode, density, housing units, and income.



Districts/Areas

Block (assembly, electoral, school, fire...) districts, and commercial zone type.



MTA

Distance to the three closest MTA stations.

Project Process

Data tools

- Pandas
- Geopandas
- PostgreSQL

Balance library

- Imbalanced-learn: SMOTE

Clean Data

Merge Data

Categorize
/ Normalize

Run Models

Filter /
Oversample

Random Forest

Python libraries

- Patsy
- Sklearn

Sklearn models

- KNN
- Logistic Regression
- SVM
- Naive Bayes
- Decision Tree

Model Scores

Manhattan

Blocks: 8,117
Stations: 435 (~5%)



First models

Accuracy: 0.95
Precision: 0.80
Recall: 0.09



Final model

Accuracy: 0.93
Precision: 0.42
Recall: 0.51



Important Features

- Block size (shape area, length)
- Distance to Citibike stations
- Bike/dock demand of closest stations
- Latitude
- Distance to MTA stations





Future Work

- More data (extra features, expansion...)
- Qualitative data
- Real estate application

Thank you!

