Project Presentation

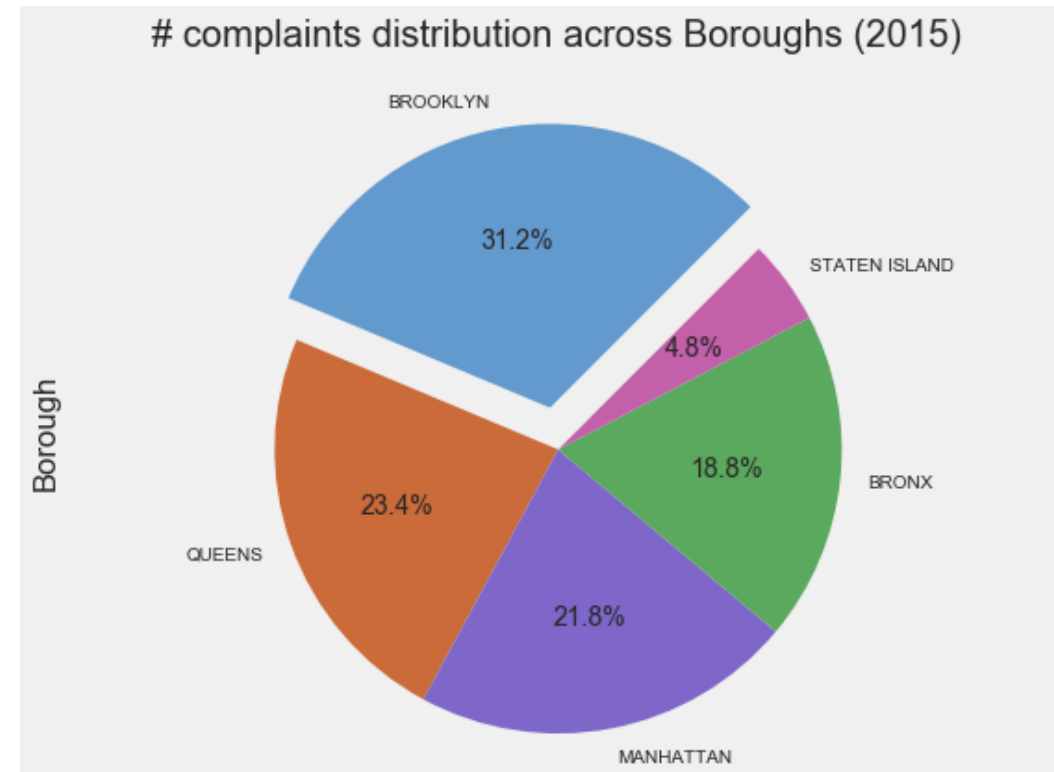EXPLORING NYC 311 DATA

BY:

AYUSH SETHI
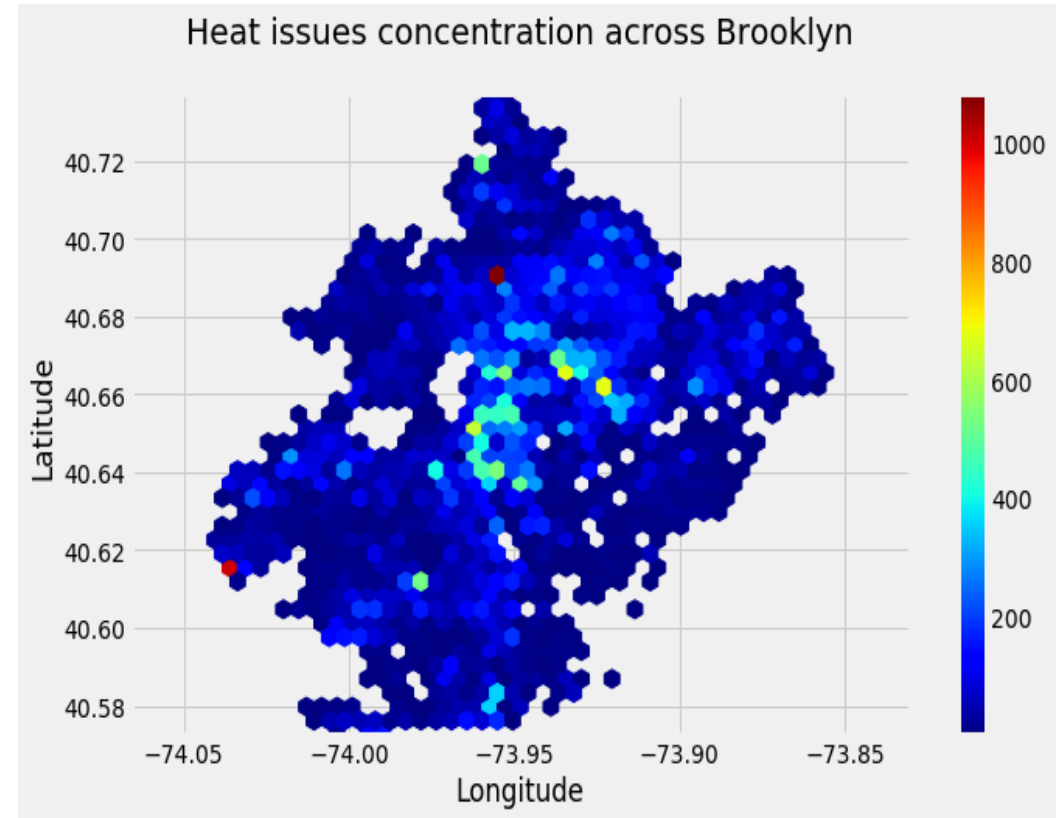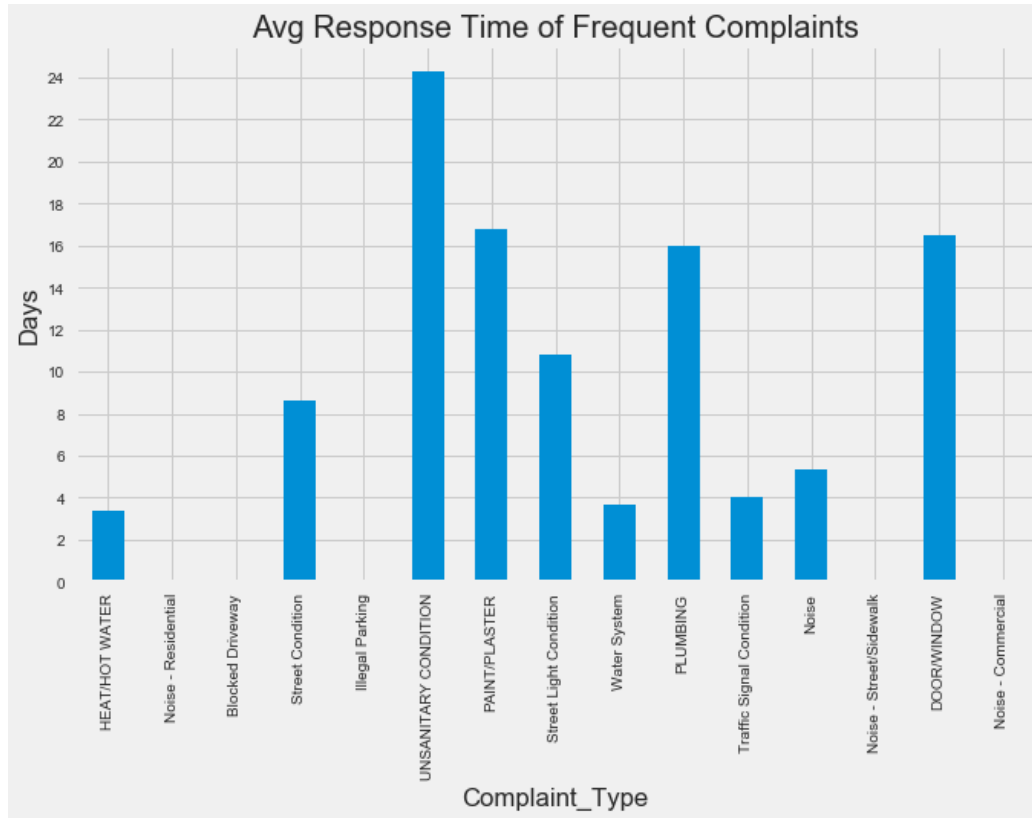
ARNAV SHAH

Exploring NYC 311 Data

Overview:
- Records of non-emergency complaints raised by New York City residents.
- Evaluating 2015 data.

Problem:
- Predict time required in terms of days to resolve a specific complaint in a specific borough.
- Find trends in various complaints when a storm hits the city like seeing spikes in volume of calls of a particular complaint during a storm.

complaints distribution across Boroughs (2015)

BROOKLYN 31.2%
STATEN ISLAND 4.8%
BRONX 18.8%
MANHATTAN 21.8%
QUEENS 23.4%

Progress so far…

Avg Response Time of Frequent Complaints

Heat issues concentration across Brooklyn

Progress so far…

Feature selection for prediction problem:

- Day of week
- Day of month
- Month
- Incident Zip
- Descriptor

Target Variable: Difference between created date and closed date (days)

Divided target variable into 3 classes: < 2 days, 2-6 days, > a week

Model applied: Decision Tree Classifier (accuracy = 61.5%)

- ❖ Descriptor variable is text, so converted it into dummy variables.

- ❖ Scaled Incident Zip feature so that it doesn't dominate other features.
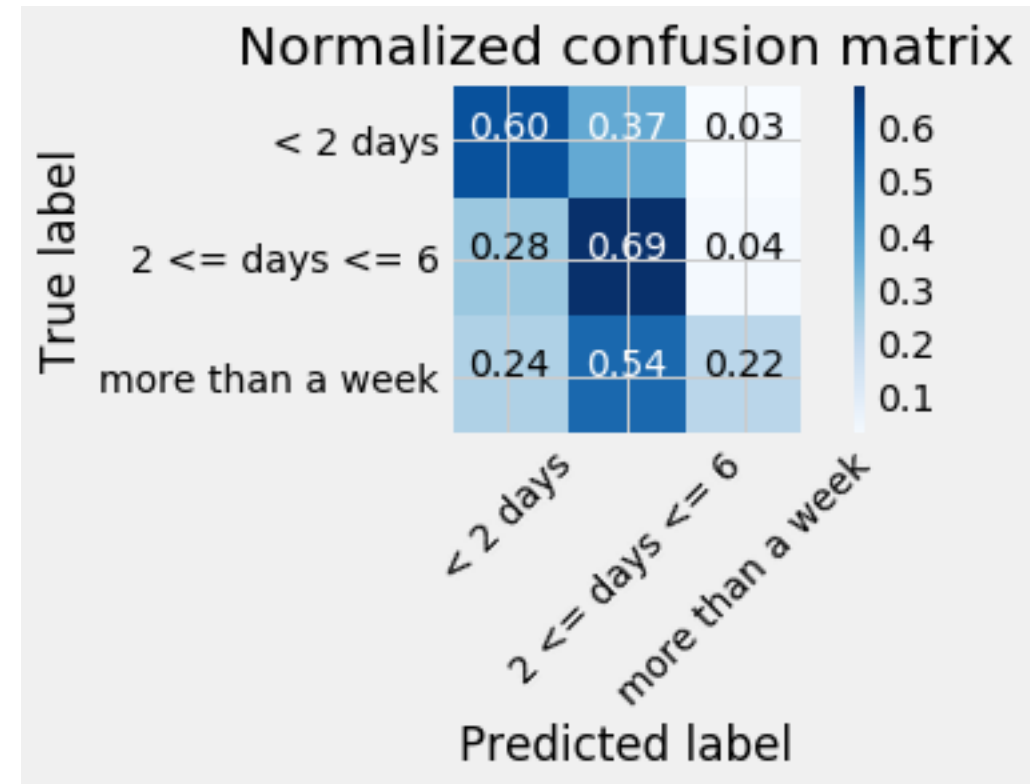
What is left…

- Improving the performance of the current model.
- Combining Storms dataset with the 311 dataset.
- To study how the average response time for a complaint changes during and immediately after the storm.