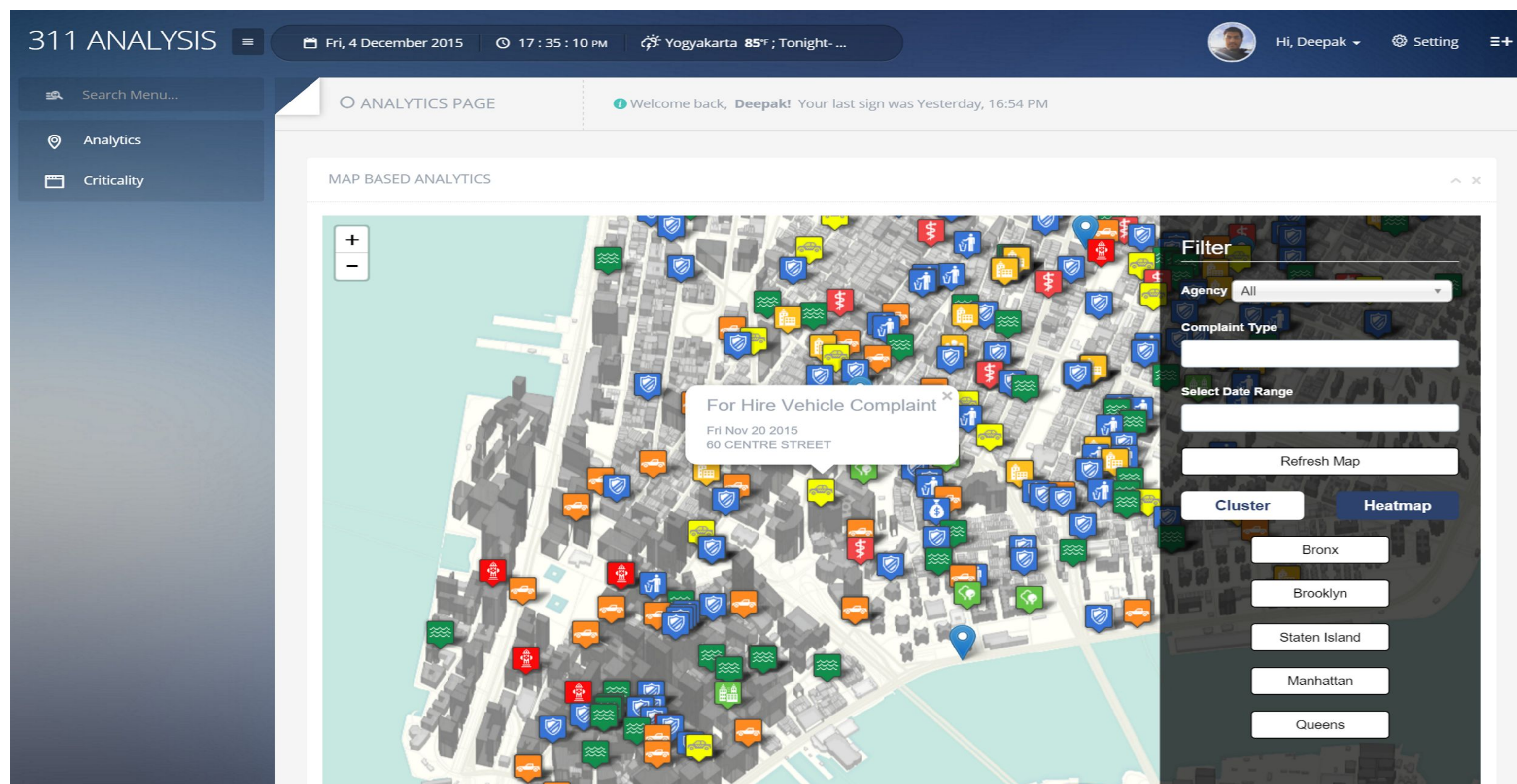# Analyzing 311 Service Requests of NYC

*Team 11 -  Aishwarya R Govindaraj, Deepak Ravindran, Karthik Kannan, Mahita Mahesh, Seema Suresh*

*CSE 6242 - Data and Visual Analytics*
*Georgia Institute of Technology*

MAIN DASHBOARD WITH ICONS REPRESENTING DIFFERENT COMPLAINT TYPES

## MOTIVATION AND METHODOLOGY

### Research questions

- Can we apply Information Visualization, Visual Analytics and Machine Learning techniques and principles to effectively explore, analyze and predict New York City's 311 service request patterns?
- Can we use NYC's historical 311 service request data to predict whether an incoming request is critical or not?
- Does a map-based visualization provide more insight into service requests patterns and trends   than traditional histograms and line charts?

### Data

- The dataset was taken from the NYC Open data website
- The dataset consists of over 1 million service requests across over 40 agencies, 110 complaint types,  and with requests originating from 5 boroughs of NYC
- The requests were tagged as critical if there were multiple service requests with the same complaint, less than 12 hours apart, with similar incident locations.
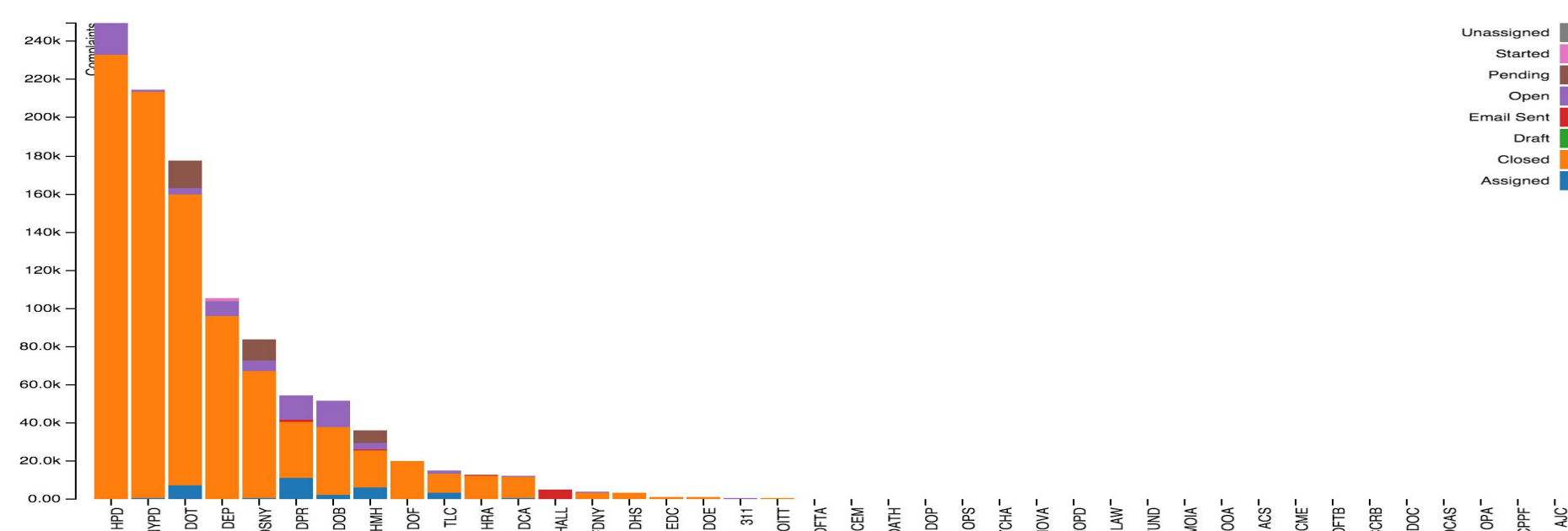


DISTRIBUTION OF COMPLAINTS ACROSS AGENCIES BY STATUS
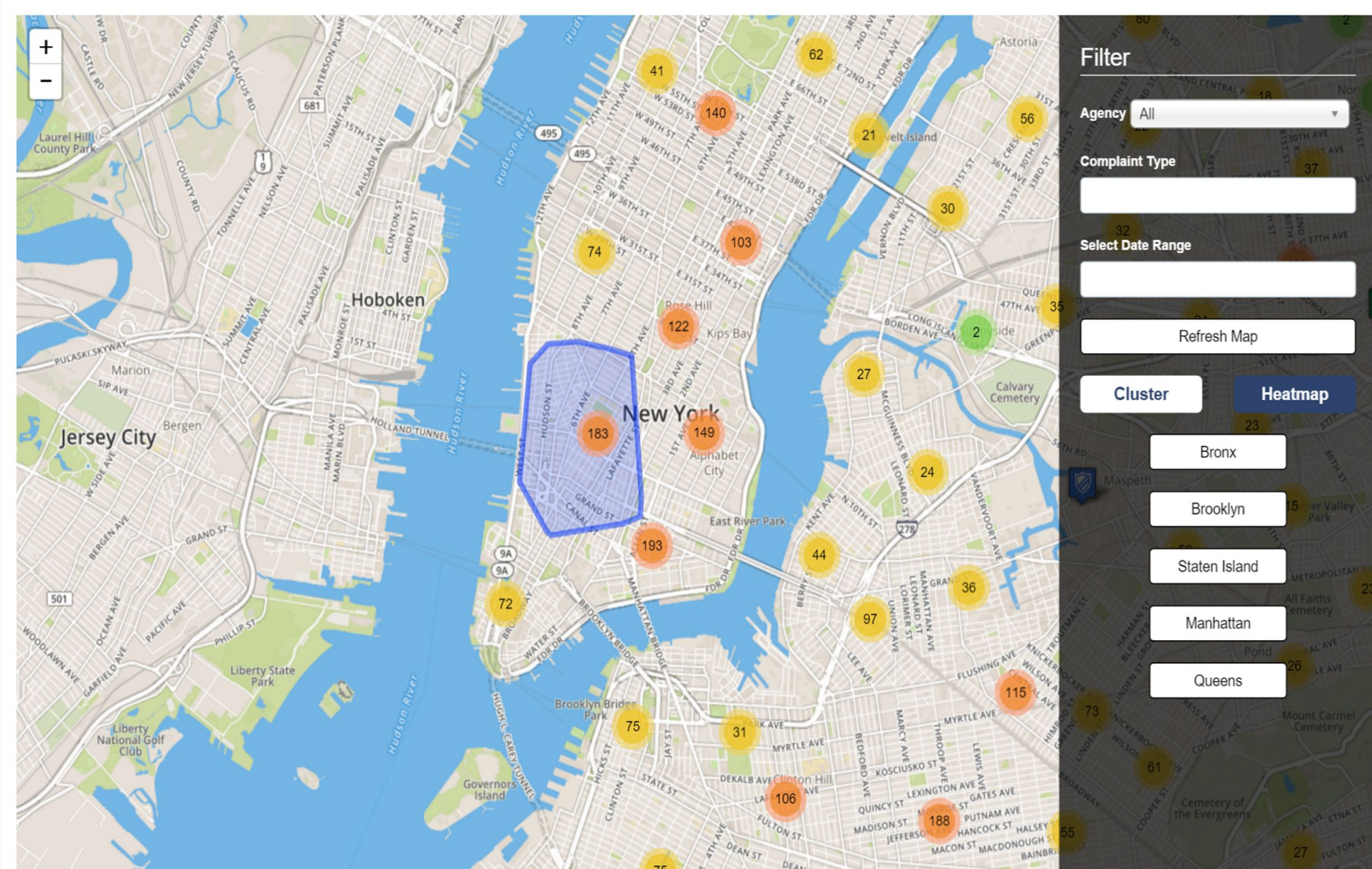


CLUSTERED REPRESENTATION OF COMPLAINTS

### Map Visualization

- Consists of two major parts - Clustering and Heat map
- The dynamic clustering combines the data into groups based on the zoom level of the map. Analytics can be done at different scales.
- The heat map gives a distribution of the large number of data over the region and its concentration
- The map can be filtered by department, problem type and by date range.
- Clicking on map icons gives the details about the complaint

## RESULTS
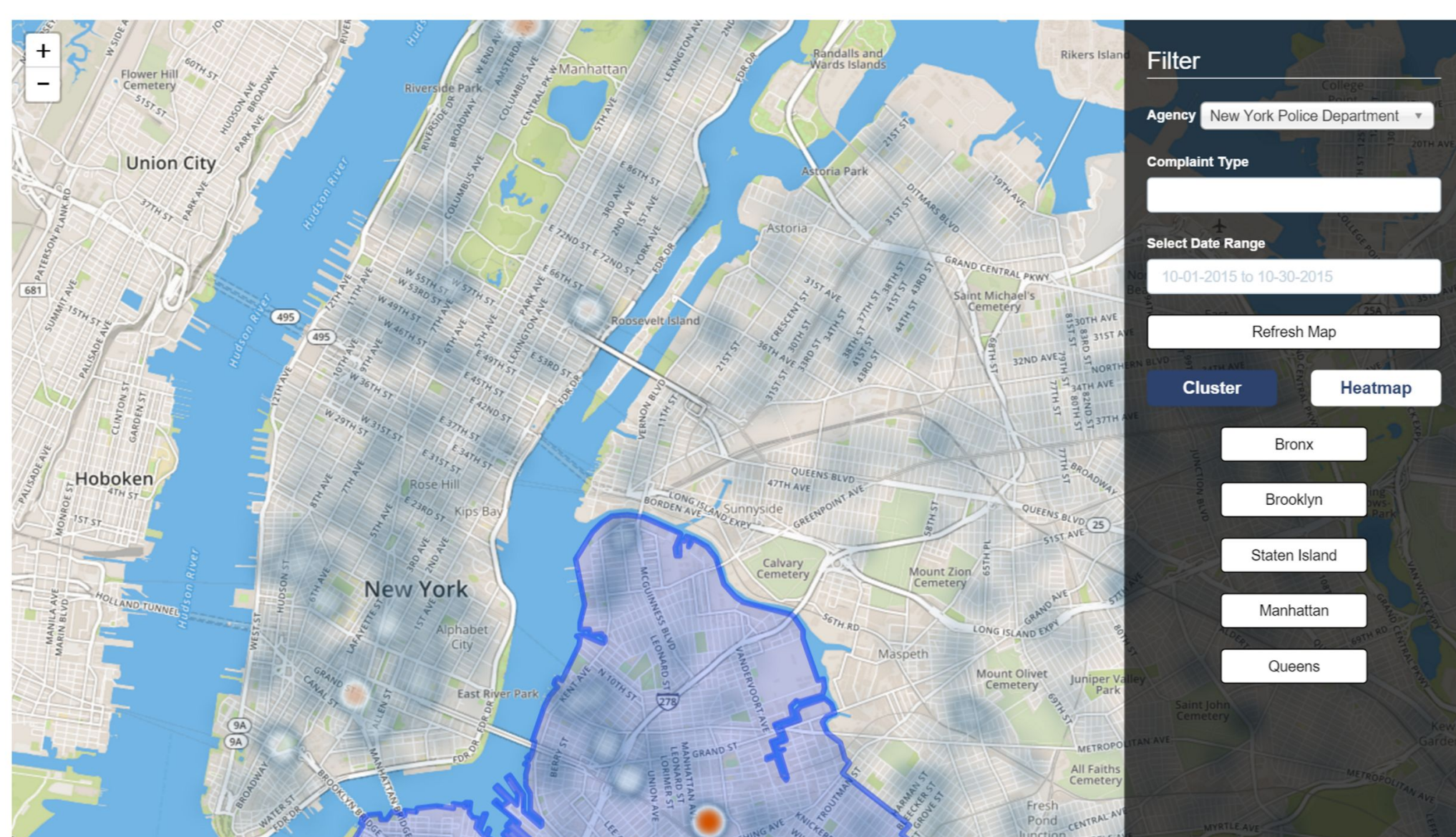


HEAT MAP REPRESENTATION OF COMPLAINTS

### Experiment Results:

- Random Forest classifier model used to predict the criticality of an incoming service request. Features include agency, complaint type, location type, incident zip, address type, city, facility type, status, borough and (x,y) coordinate of location.
- 82.47% accuracy achieved as an average of 5 runs with 70-30 random split of dataset into train and test.
- User Interface incorporates a form to fill in service request data and receive a prediction on whether the request is potentially critical or not.
- Scalability achieved using Apache Spark as the cluster computing framework and MLlib as the machine learning library
- Map visualization can retrieve and display data from any month in 2010-2015.
- Clustering the map data vastly improves readability