**Locate an Optimal Food Service Delivery Area**

# Business Problem

Considering multiple factors such as time of day, traffic patterns, clustering of food providers, and population density help optimize the delivery area by providing maximally efficient routes and service coverage areas, to reduce driver costs and time and shorten customer delivery times.  
The audience for this data report is investors, existing or planned businesses, and logistics providers.

### Data Requirements and Description[¶](http://127.0.0.1:8888/notebooks/datasciencecert/capstone/Coursera_Capstone/capstone-week3-v2.ipynb#Data-Requirements-and-Description)

#### The following datasets are available:

#### NYC Neighborhood Data:

This dataset contains data about 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood is needed.  
The NY dataset exists for free on the web and will be downloaded from this link: <https://geo.nyu.edu/catalog/nyu_2451_34572>

#### Foursquare Data:

Location based venue information. This data will be provided by the free Foursquare API. Combined with the NYC Data, various regression and clustering analytics will be used to drive the analysis and tune models.

# Data Acquisition and Cleansing

Data was acquired for the NY dataset over the internet, and the raw JSON file was parsed and loaded into a pandas frame. The frame structure was refined to correlate with Foursquare API venue data to simplify downstream analysis.

The features selected as key indicators were the number of venues, and within a fixed radius of a neighborhood. These features provide an overall relative view of neighborhoods with maximum delivery potential and minimum cost.

The primary final cleansed and corrected data is shown below:



Figure 1. Cleansed and corrected dataset used for analysis.

# Exploratory Data Analysis

In order to calibrate the data and prepare for deeper analysis, an overall view of the geographic distribution of locations was created. This shows that a local analysis of the top ranking venues by location is likely the best approach to start, given the targeted focus of maximizing ROI.

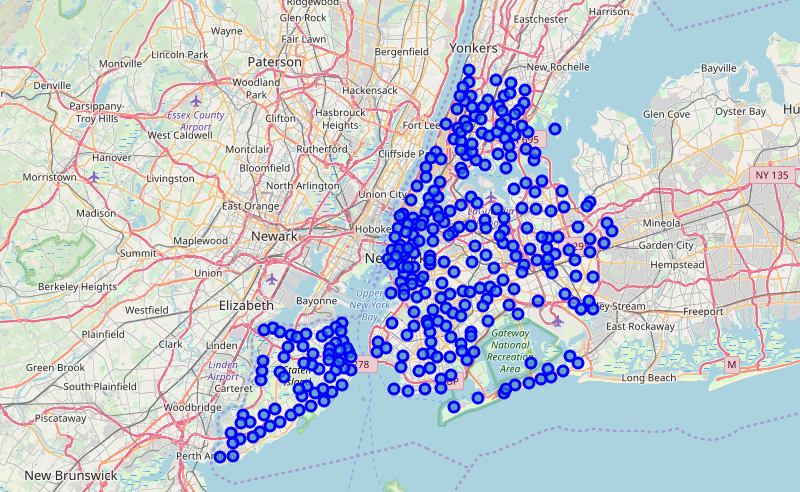


Figure 2. Distribution of neighborhoods

The neighborhoods with the most number of densities in the radius area was then identified and the sample data identified for the next steps of the analysis.

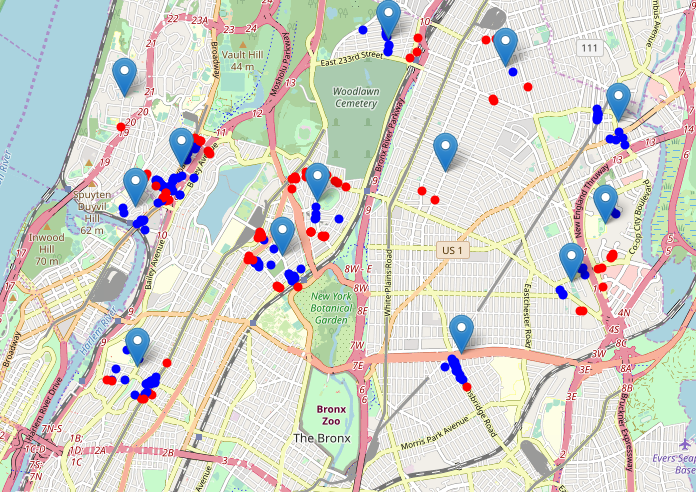


Figure 3. Maximal targeted dataset to proceed with analysis

# K-Clusters and Heatmaps

In the next phase of the analysis, the identified sample dataset was used to represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones with top location candidates.

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

The visualization below could be used in future analysis as a way to spot key locations among a larger dataset:

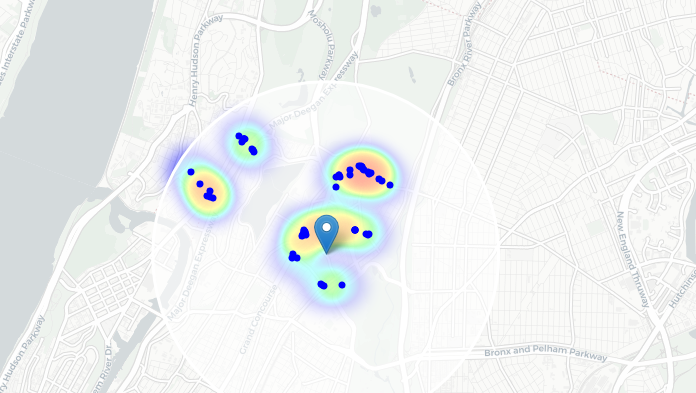


Figure 4. Heatmap shows grouping strength of the target variables

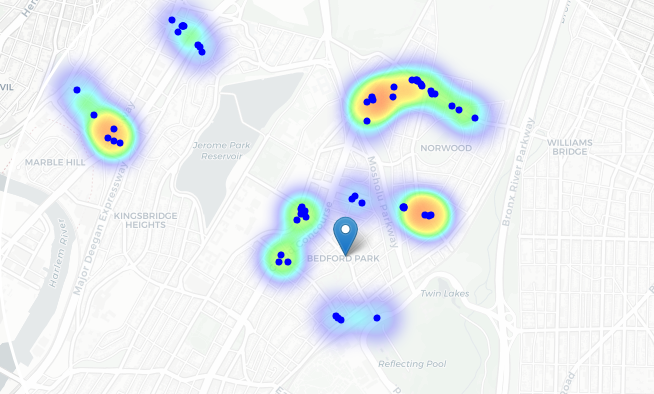


Figure 5. Heatmap with k-cluster k=15

# Conclusion

The shaded areas below indicate the cluster strength. The upper right quadrant clearly shows the strongest relationship between the features selected toward the business goal of high efficiency and low cost delivery services. High density venue neighborhoods with close proximity are indicted by the clustering, and relative relationship between the neighborhoods sampled is shown by the heat map and blue circles.

From this analysis, a broader review can be done city-wide, and provide the groundwork for prioritizing delivery services with maximum ROI potential.

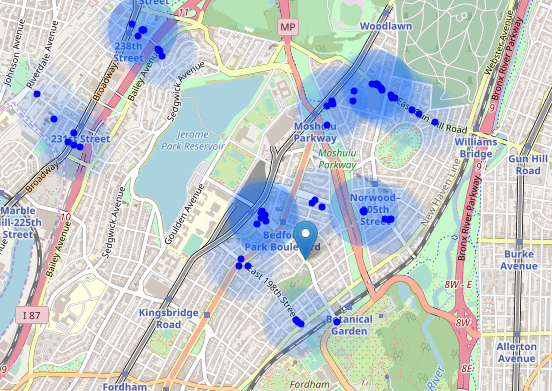


Figure 6. blue circles show cluster strength and maximal targets for the business model