OPENING BEVERAGE DISTRIBUTION WAREHOUSES IN THE STATEN ISLAND

A Capstone Project - The Battle of the Neighborhoods

By IBM Applied Data Science Capstone

By

Dilip Mistry

April, 201

Table of Contents

Introduction: Business Problem

In this project, we will try to find some optimal locations for warehouse to deliver beverages to nearby restaurants, coffee shops and bars. Specifically, this report will be targeted to stakeholders interested in opening some beverage warehouses in **Staten Island** of **New York City**, USA.

The business problem of the project is that there is no central warehouse in the Staten Island and the distribution takes place from other boroughs of New York City. The distribution from other areas to the targeted areas of Staten Island became time consuming, inefficient and costly. Now, the client wants to open several warehouses to supply beverages to the restaurants, coffee shops and bars quickly and reduce the costs of the distribution of beverages. Therefore, the client wants to know which locations would be the best locations for his business in order to minimize the relative distance from each of the venues.

This project will use some data science techniques to generate neighborhoods in the Staten Island areas based on the criteria so that the client can choose the best possible five locations for warehouses and distribute beverages to the restaurants, coffee shops and bars with efficiently.

Data Sources

This project needs the following data to solve the problem:
- List of neighborhoods with coordinate information in the Staten Island
- Venue data, particularly related to restaurants, coffee shops and bars

The data in this report obtained from https://geo.nyu.edu/catalog/nyu_2451_34572 and contained boroughs and list of neighborhoods in New York City. However, the list of neighborhoods in the Staten Island have been used in this report. The data also contains the geographical coordinates of the neighborhoods.

The information regarding restaurants, coffee shops and bars in every neighborhood of Staten Island will be obtained using **Foursquare API**. The Foursquare API will provide detailed location

information, including name of location, categories for the location, geographical coordinates of the location, etc. This data will be used to determine the best 5 warehouses for distributing beverages in Staten Island, NY.

Methodology

In this project, we choose to use k-means algorithms in order to define the location of each warehouse. In first step, we have collected the required data such as location and type (category) of every restaurant, coffee shop, and bar in the Staten Island. In the second step, we will show all the locations in the map and merge all the locations of the restaurants, coffee shops and bars in a new data frame and then find the central location and the standard deviation in location. In third and final step, we will focus on optimal locations and within those create **clusters of locations** established in discussion with the client.

By default, the **k-means clustering** algorithm minimize the distance of each point in a cluster from the centroid of the cluster. As a result, the cluster points of the k-means algorithm will be at the minimum distance of the determined centroid. We used the set of neighborhood locations (Latitude, Longitude) of the venues of interest. All venues belonging to restaurants, coffee shops, and bars are included in the dataset. We will present map of all such locations and will create clusters (using **k-means clustering**) of those locations to identify general neighborhoods and search for optimal warehouse locations for the client.

There are total of 160 restaurants, 12 coffee shops and 24 bars found in the Staten Island area. The first 10 location of restaurants, coffee shops, and bars as gathered from the Foursquare database are shown in Figure 1 - 3.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	St. George	40.644982	-74.079353	Beso	40.643306	-74.076508	Tapas Restaurant
5	St. George	40.644982	-74.079353	Enoteca Maria	40.641941	-74.077320	Italian Restaurant
8	St. George	40.644982	-74.079353	The Gavel Grill	40.642157	-74.076674	American Restaurant
14	St. George	40.644982	-74.079353	The Burrito Shoppe	40.643639	-74.077919	Fast Food Restaurant
18	St. George	40.644982	-74.079353	Not Guilty Deli	40.641867	-74.077016	American Restaurant
34	Stapleton	40.626928	-74.077902	Lakruwana	40.625654	-74.075174	Sri Lankan Restaurant
36	Stapleton	40.626928	-74.077902	Vida	40.628723	-74.079802	Restaurant
37	Stapleton	40.626928	-74.077902	Bay House Bistro	40.627827	-74.076244	Asian Restaurant
39	Stapleton	40.626928	-74.077902	Campo Bello Restaurante	40.624463	-74.079652	Spanish Restaurant
43	Stapleton	40.626928	-74.077902	El Patron Restaurant & Lounge Inc.	40.629154	-74.076541	Mexican Restaurant

Figure 1: Locations of Restaurants in Staten Island

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
90	West Brighton	40.631879	-74.107182	Fab Cup	40.630129	-74.109486	Coffee Shop
91	West Brighton	40.631879	-74.107182	Starbucks	40.630320	-74.106087	Coffee Shop
94	West Brighton	40.631879	-74.107182	Beans & Leaves	40.630718	-74.103376	Coffee Shop
175	New Springville	40.594252	-74.164960	Barnes & Noble Cafe	40.592385	-74.162901	Coffee Shop
264	Eltingville	40.542231	-74.164331	Country Donuts	40.545532	-74.160893	Coffee Shop
306	Woodrow	40.541968	-74.205246	Starbucks	40.542153	-74.207067	Coffee Shop
367	Park Hill	40.609190	-74.080157	Starbucks	40.606393	-74.078682	Coffee Shop
383	Arlington	40.635325	-74.165104	Unique Coffee Roasters	40.637370	-74.162424	Coffee Shop
589	Charleston	40.530531	-74.232158	Starbucks	40.528280	-74.233315	Coffee Shop
614	Arden Heights	40.549286	-74.185887	APlus at Sunoco	40.551233	-74.184030	Coffee Shop

Figure 2: Locations of Coffee Shops in Staten Island

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
6	St. George	40.644982	-74.079353	Steiny's Pub	40.642185	-74.076599	Bar
9	St. George	40.644982	-74.079353	Ruddy & Dean	40.644074	-74.076683	Bar
38	Stapleton	40.626928	-74.077902	The Hop Shoppe	40.629034	-74.079758	Beer Bar
40	Stapleton	40.626928	-74.077902	Vinum Wine Bar & Cafe	40.624853	-74.074890	Bar
100	West Brighton	40.631879	-74.107182	Jody's Club Forest	40.630894	-74.101912	Bar
107	West Brighton	40.631879	-74.107182	Better Gourmet Health Kitchen	40.630881	-74.102813	Juice Bar
109	West Brighton	40.631879	-74.107182	Liberty Tavern	40.630929	-74.102130	Bar
152	Port Ivory	40.639683	-74.174645	Jonesys Tavern	40.639640	-74.171252	Bar
165	Castleton Corners	40.613336	-74.119181	Danny Boy's	40.613212	-74.123846	Bar
178	New Springville	40.594252	-74.164960	1001 Nights Cafe	40.598126	-74.162204	Hookah Bar

Figure 3: Locations of Bars in Staten Island

In order to determine the radius of each cluster, descriptive statistics were used and thus, the area in which the venues of interest will be served by the specific warehouse. Moreover, in order to examine the result, the density of the venues in each area was calculated. Mean values

were used to determine the exact location of each warehouse. This project created 5 zones representing their centers and addresses will be the final result. Although zones are shown on map, their shape is actually very irregular and their centers should be considered only as a starting point for exploring area neighborhoods in search for potential warehouse locations.

Results

Figure 5 represents the map of Staten Island. The locations of the venues of interest are marked on the map with different colors according to their category.
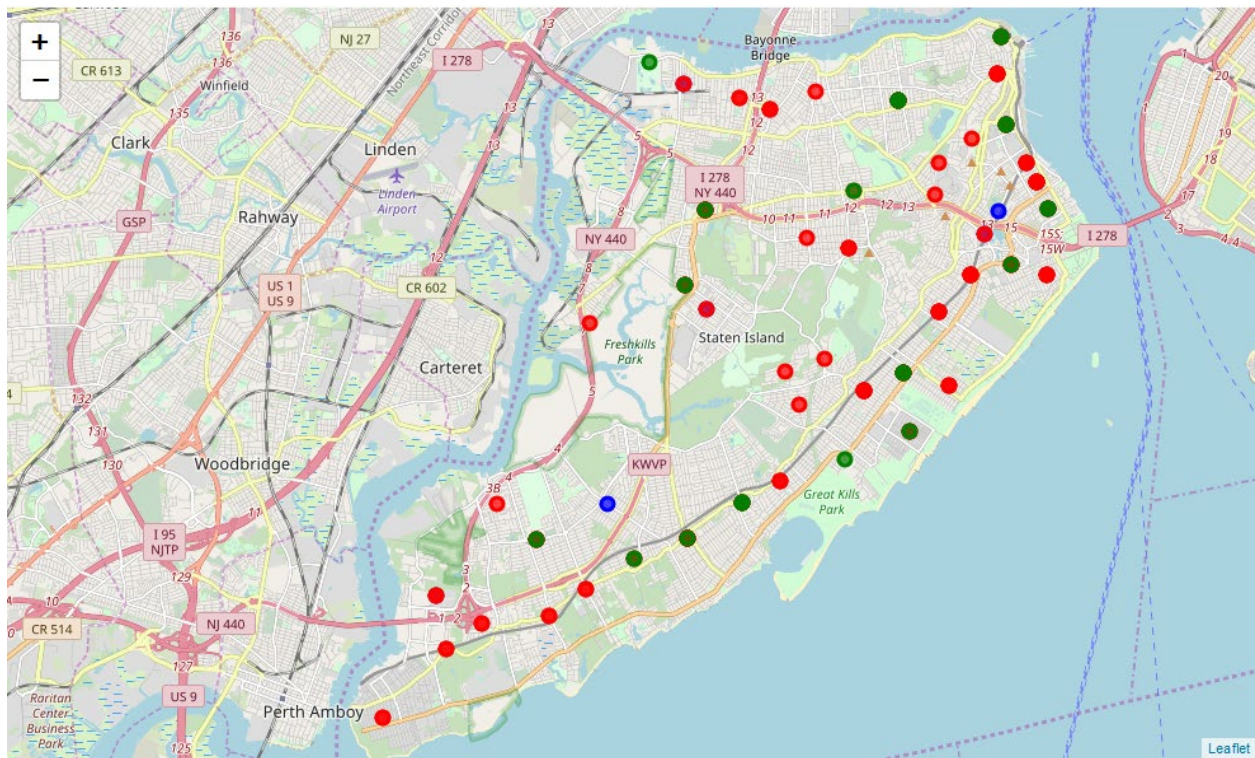
Figure 5: Locations of Restaurants (red), Coffee Shops (blue) and Bars (green) in the Staten Island

All the above points were used in the k-means clustering algorithm. The resulted clusters are shown in Figure 6.
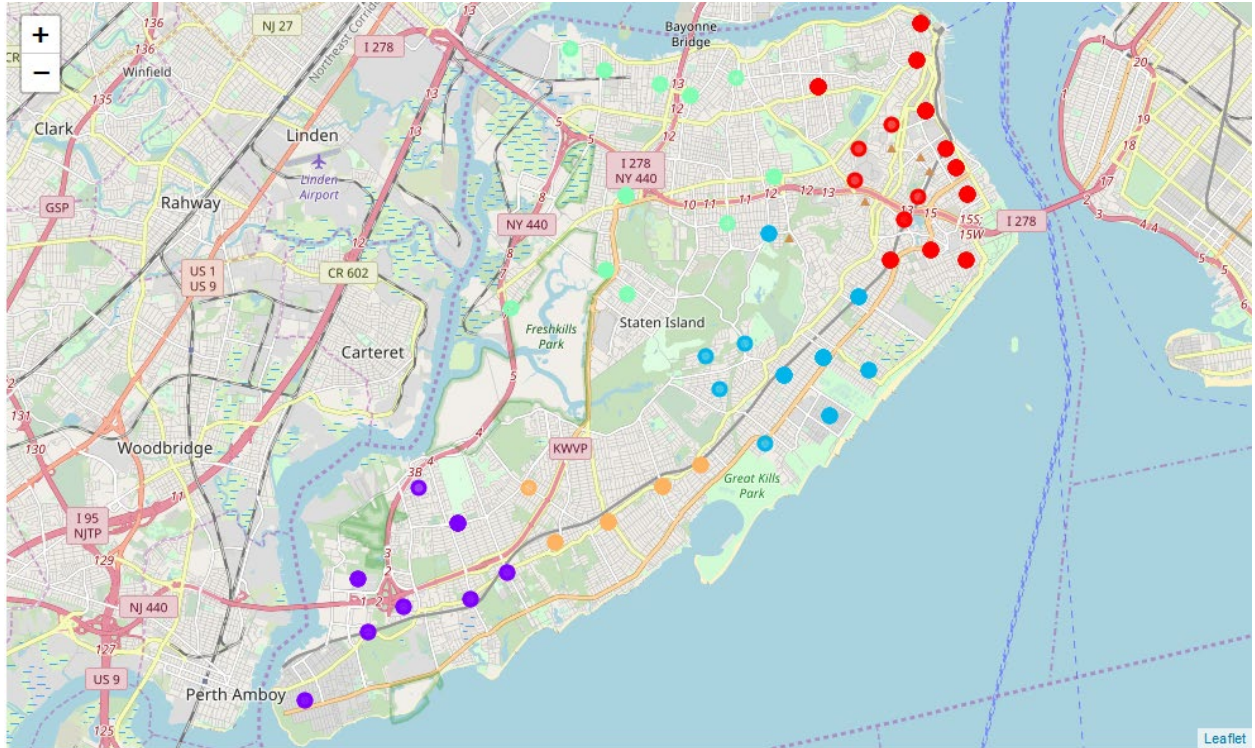
Figure 6: The 5 clusters of the venues indicated by different colors

The location of the warehouse was identified by the centroids of the clusters calculated by the k-means clustering algorithm. The addresses can be extracted from the location's latitude and longitude values. The area of each cluster was measured by calculating the mean radius of each cluster. Finally, the density of each cluster was calculated from the number of venues based on the total area of each cluster. The results are shown in the table below.

Table 1: The detail location information of 5 warehouses for Beverages in the Staten Island

Cluster Number	Color of Cluster	Location Center		Mean Radius (KM)	Area (KM²)	Number of Venues	Density
		Latitude	Longitude				
Warehouse 1	Red	40.6199	-74.0823	2.07	121.31	77	0.63
Warehouse 2	Purple	40.5304	-74.2161	1.96	108.91	20	0.18
Warehouse 3	Blue	40.5768	-74.1099	1.38	53.50	38	0.71
Warehouse 4	Green	40.6114	-74.1557	1.98	110.68	32	0.29
Warehouse 5	Orange	40.5455	-74.1592	1.20	40.96	29	0.71

Figure 7 below shows the area of each cluster overlay on the map of Staten Island. Each warehouse will be responsible for the venues lying at the area indicated by the corresponding circle.
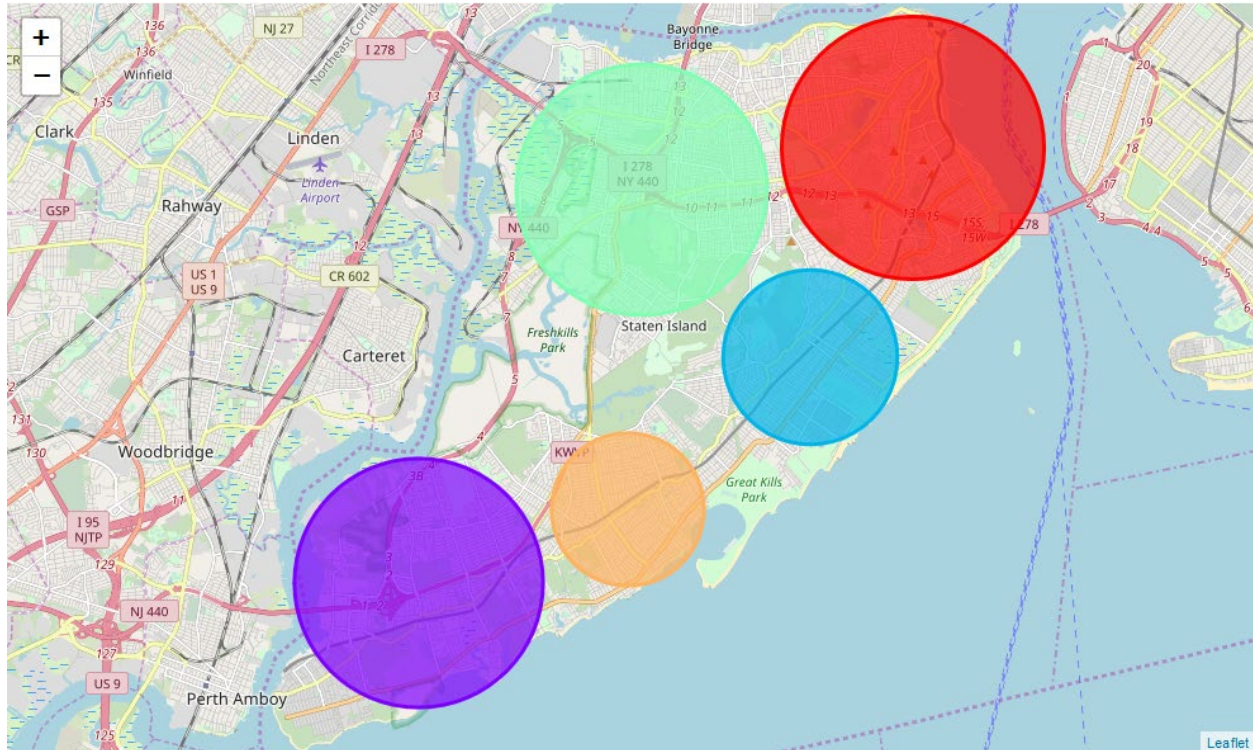
Figure 7: Overlay of Cluster Area in the Staten Island Map

Discussion

The clustering map shows the distribution of restaurants, coffee shops, and bars in the Staten Island area in five clustering zones. The red, purple and green clusters hold the larger area ranging from 109~121 sq. km. Whereas, the blue and orange clusters hold the smaller area ranging from 41~54 sq. km. The red cluster contains the largest number of restaurants, coffee shops, and bars than each of other four clusters. Even though, green and purple clusters hold the larger area as red cluster, they contain relatively less number of restaurants, coffee shops, and bars than the red cluster.

Based on the observation, we suggest the following:

- The number of trucks, employees, and drivers should be chosen based on the number of venues in each cluster
- The size of the warehouse in the red clustering zone would be bigger than any other warehouses
- The number of trucks, employees, and drivers in red clustering zone would be at least double than the number of any warehouses

Conclusion

The purpose of this project was to locate five warehouses for the client in the Staten Island area in order to supply beverages to the restaurants, coffee shops and bars more efficiently and cost effectively. The approach was to divide the Staten Island area into five sub regions based on the criteria of the distribution of the restaurants, coffee shops and bars in the entire Staten Island area. The centroid of each cluster was identified by running the k-means clustering algorithm. In order to estimate the clustering area containing the restaurants, coffee shops, and bars to which the warehouse would provide beverages, a standard deviation of each cluster was used. Based on the results, this project was able to locate the location of the warehouses, their size, and the number of employees and truck drivers in each warehouse needed.

Even though, this project recommended the optimal locations for warehouses, final decision on optimal location will be made by the client based on specific characteristics of neighborhoods taking into consideration additional factors like proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.