Introduction/Business Problem

A bottling company has an established role in the market of Toronto. They are the premier supplier of bottled water with over 1000 clients in Toronto's neighborhoods. His best clients consist of the following types of businesses: restaurants and bars, hotels and coffee shops. The bottled water is currently stored in a central warehouse outside of the city and is distributed to the clientele daily. The main problem with this business plan is that the distribution becomes is inefficient and becomes time consuming and costly.

Our client seeks to reduce the distribution costs by building 5 smaller warehouses in Toronto to serve as smaller hubs and distribute to his clients from those points since it'd be closer, eliminating the time and resource costs caused by the distance of the original warehouse from his clients. In order to accomplish this, our client has asked us to find the 5 best locations in the city of Toronto where the warehouses could be constructed in order to create smaller distribution clusters. The warehouses must form the centers of the clusters in order to minimize their relative distances from each client.

Data

The data required for this task are the locations of the hotels, coffee shops, bars and restaurants in Toronto. In order to gather the required data, we will gather the locations of Toronto neighborhoods from Wikipedia. With this data, we will pinpoint the locations of these neighborhoods on Foursquare. We will source the following using Foursquare:

- Neighborhood
- Neighborhood's Latitude
- Neighborhood's Longitude
- Venue Name
- Category

The data will then be filtered in order to acquire the locations of the targeted venues needed for our study. We will utilize Folium to build a map of Toronto and visualize the delivery venues. Following this, the k-means algorithm will be used to define the centroids of the five clusters which will serve as the best possible points at which distribution warehouses can be built.

Methodology

In order to define the location of each warehouse, we choose to use the k-means algorithm. By default, this algorithm minimize the distance of each point in a cluster from the centroid of the cluster. As a result, the output of the k-means algorithm is a set of clusters whose points are lying at the minimum distance from the determined centroids. As an input, we used the set of locations (lat, long) of the venues of interest. In this example the category of each venue is not required in the algorithm. All venues belonging to hotels, bars, restaurants and coffee shops are included in the list. The locations of hotels, bars, restaurants and coffee shops are shown as they were gathered from the Foursquare database in Figure 1, 2, 3, 4. A total of 258 coffee shops, 909 restaurants, 178 bars and 73 hotels were found and used in the k-means algorithm.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
8	The Beaches	43.676357	-79.293031	The Remarkable Bean	43.672801	-79.287038	Coffee Shop
11	The Beaches	43.676357	-79.293031	Buds Coffee Bar	43.669375	-79.303218	Coffee Shop
54	The Beaches	43.676357	-79.293031	Starbucks	43.668370	-79.308015	Coffee Shop
72	The Beaches	43.676357	-79.293031	Circus Coffee House	43.685483	-79.315364	Coffee Shop
75	The Beaches	43.676357	-79.293031	Starbucks	43.682446	-79.327232	Coffee Shop
121	The Danforth West, Riverdale	43.679557	-79.352188	Hailed Coffee	43.666900	-79.345432	Coffee Shop
135	The Danforth West, Riverdale	43.679557	-79.352188	Merchants of Green Coffee	43.659986	-79.354299	Coffee Shop
146	The Danforth West, Riverdale	43.679557	-79.352188	Te Aro	43.661373	-79.338577	Coffee Shop
148	The Danforth West, Riverdale	43.679557	-79.352188	Dark Horse Espresso Bar	43.658498	-79.352356	Coffee Shop
163	The Danforth West, Riverdale	43.679557	-79.352188	Sumach Espresso	43.658135	-79.359515	Coffee Shop

Fig 1. Sample of bar locations in Toronto

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
138	The Danforth West, Riverdale	43.679557	-79.352188	The Broadview Hotel	43.659060	-79.350030	Hotel
194	The Danforth West, Riverdale	43.679557	-79.352188	The Grand Hotel & Suites Toronto	43.656449	-79.374110	Hotel
284	The Beaches West, India Bazaar	43.668999	-79.315572	The Broadview Hotel	43.659060	-79.350030	Hotel
305	Studio District	43.659526	-79.340923	The Broadview Hotel	43.659060	-79.350030	Hotel
398	Studio District	43.659526	-79.340923	The Grand Hotel & Suites Toronto	43.656449	-79.374110	Hotel
830	Moore Park, Summerhill East	43.689574	-79.383160	Four Seasons Hotel Toronto	43.671796	-79.389457	Hotel
932	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049	Four Seasons Hotel Toronto	43.671796	-79.389457	Hotel
1011	Rosedale	43.679563	-79.377529	Four Seasons Hotel Toronto	43.671796	-79.389457	Hotel
1064	Rosedale	43.679563	-79.377529	The Grand Hotel & Suites Toronto	43.656449	-79.374110	Hotel
1098	Rosedale	43.679563	-79.377529	The Broadview Hotel	43.659060	-79.350030	Hotel

Figure 2. Sample of hotel locations in Toronto

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
8	The Beaches	43.676357	-79.293031	The Remarkable Bean	43.672801	-79.287038	Coffee Shop
11	The Beaches	43.676357	-79.293031	Buds Coffee Bar	43.669375	-79.303218	Coffee Shop
54	The Beaches	43.676357	-79.293031	Starbucks	43.668370	-79.308015	Coffee Shop
72	The Beaches	43.676357	-79.293031	Circus Coffee House	43.685483	-79.315364	Coffee Shop
75	The Beaches	43.676357	-79.293031	Starbucks	43.682446	-79.327232	Coffee Shop
121	The Danforth West, Riverdale	43.679557	-79.352188	Hailed Coffee	43.666900	-79.345432	Coffee Shop
135	The Danforth West, Riverdale	43.679557	-79.352188	Merchants of Green Coffee	43.659986	-79.354299	Coffee Shop
146	The Danforth West, Riverdale	43.679557	-79.352188	Te Aro	43.661373	-79.338577	Coffee Shop
148	The Danforth West, Riverdale	43.679557	-79.352188	Dark Horse Espresso Bar	43.658498	-79.352356	Coffee Shop
163	The Danforth West, Riverdale	43.679557	-79.352188	Sumach Espresso	43.658135	-79.359515	Coffee Shop

Figure 3. Sample of coffee shop locations in Toronto

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
2	The Beaches	43.676357	-79.293031	Tori's Bakeshop	43.672114	-79.290331	Vegetarian / Vegan Restaurant
14	The Beaches	43.676357	-79.293031	Veloute Bistro	43.672267	-79.289584	French Restaurant
19	The Beaches	43.676357	-79.293031	Budapest Restaurant	43.680946	-79.310110	Hungarian Restaurant
28	The Beaches	43.676357	-79.293031	Udupi Palace	43.672480	-79.321275	Indian Restaurant
29	The Beaches	43.676357	-79.293031	Lake Inez	43.672520	-79.320712	Asian Restaurant
31	The Beaches	43.676357	-79.293031	Jatujak	43.688421	-79.270073	Thai Restaurant
40	The Beaches	43.676357	-79.293031	The Wren	43.682467	-79.328079	American Restaurant
41	The Beaches	43.676357	-79.293031	Xola	43.672603	-79.288080	Mexican Restaurant
43	The Beaches	43.676357	-79.293031	Melanie's Bistro	43.684800	-79.317167	French Restaurant
57	The Beaches	43.676357	-79.293031	Rendez-Vous Restaurant Bar & Cafe	43.682570	-79.327544	Ethiopian Restaurant

Figure 4. Sample of restaurant locations in Toronto

After I extracted the clusters, the radius of each cluster was determined using descriptive statistics and consequently, the approximate areas by where venues will be served by each warehouse. Also, the density of the venues in each cluster was calculated. Mean values were used to determine the best precise locations for the warehouses to be built.

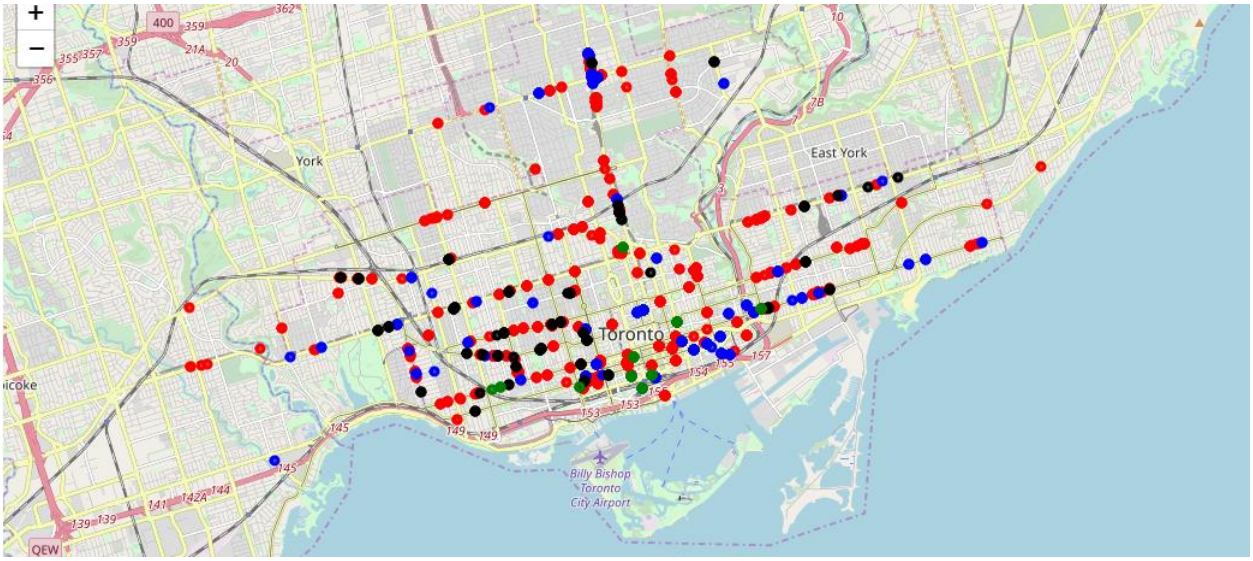

Figure 5. Bars(black), Hotels(green), Restaurants(red), Coffee Shops(blue)

The above points were all used in the k-Means algorithm with the resultant clusters shown in Fig 6.

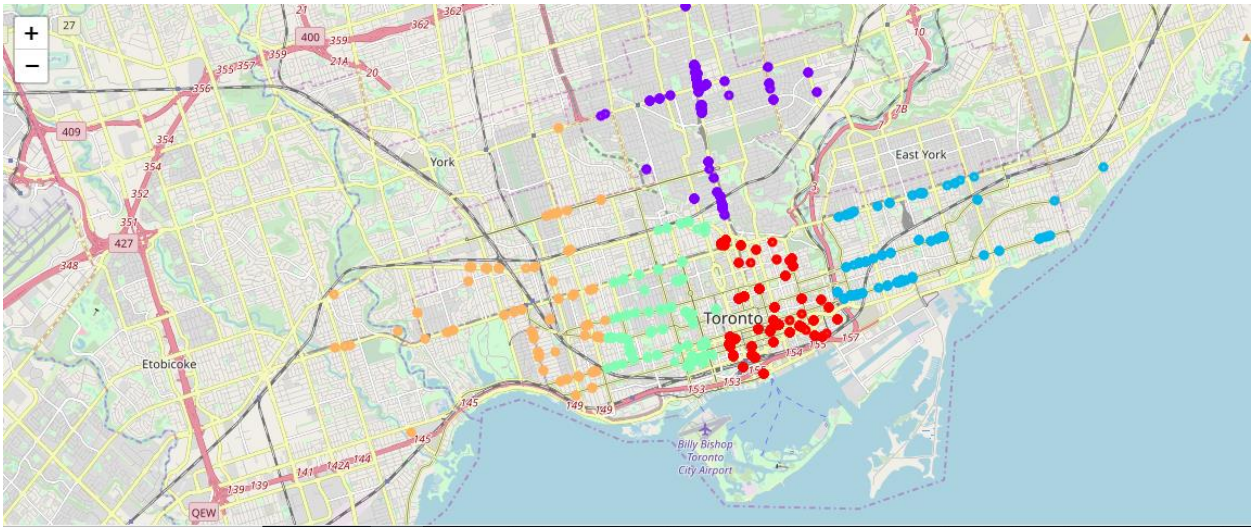

Figure 6: The venue clusters are indicated by different colors, assigned with colormap. The clusters were created using the k-Means algorithm.

The mean radius was calculated for each cluster and the locations of the warehouses defined by the clusters' centroids which resulted from the k-means algorithm. The number of venues in each cluster has been extracted and used in order to define each area's density. The results are shown in the following table, and in Figure 7, the area of each cluster is presented.

Cluster	Location Center (latitude)	Location Center (longitude)	Mean Radius	Area(km²)	Number of Venues	Density(km²)
Red	43.65390051	-79.37688846	1.15	37.79	431	11.4
Purple	43.70337104	-79.39690812	1.49	62.96	231	3.66
Blue	43.66966925	-79.33465418	1.7	81.98	174	2.12
Green	43.653691	-79.40603255	1.3	47.86	408	8.52
Orange	43.65630314	-79.44835502	1.92	104.75	172	1.64

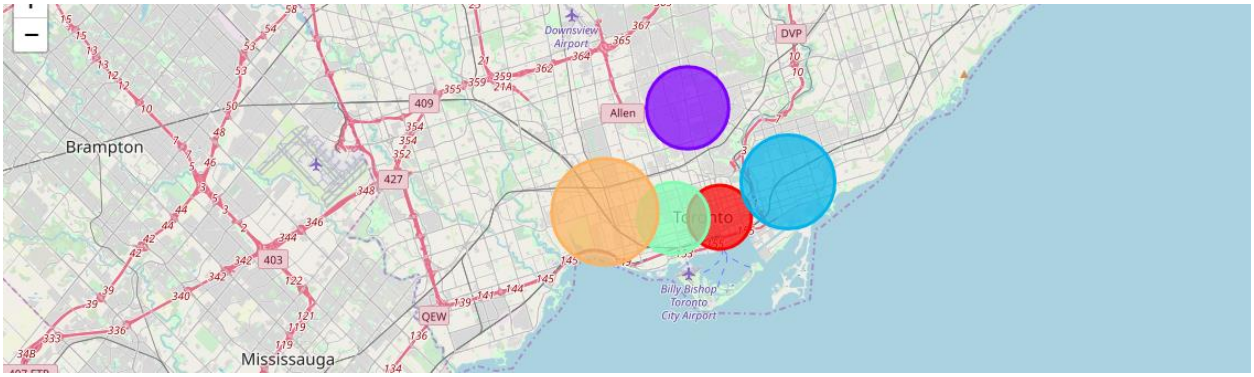

Fig. 7: Area of the Clusters. Each warehouse will be responsible for distribution to the venues within the areas highlighted by the circles.

Discussion

From our results we can see the variability in the distribution of different venues in Toronto. As can be seen, the two central clusters(green and red) in the city center contain at least twice as many venues as the remaining three clusters. Based on this observation, I suggest that the warehouses to be built in the green and red areas be much larger than the remaining three warehouses, as they will need the extra storage space to serve the bulk of venues. On that same thread, there should likewise be far more people employed at the warehouses serving red and green areas to help meet demand in a productive fashion. It also follows that the warehouses in those two areas should also have more delivery trucks than those from the remaining three clusters.

Conclusion

We have solved the problem of defining the locations for the five distribution warehouses to be built for the purpose of improving bottled water distribution in Toronto's most populated city area. The approach was to divide the city into five subregions via Clustering. We used the distribution, ie, the density of the venues in the city to accomplish this. The centers of these subregion clusters were generated by using the k-means algorithm. The standard deviation of the clusters were used to generate and define the areas which contained the venues of interest to which the warehouses must distribute its products. Based on our results, we have successfully defined the warehouse locations, their size, numbers to be employed and number of distribution trucks required.