Introduction/Business Problem

Our client, a bottled water brand, has an established role in the market of Toronto. He is the first supplier of bottled water with more than 1500 clients in the neighbourhoods of Toronto. His main clients are hotels, coffee shops, restaurants and bars. Currently, the products are stored in a big central warehouse outside Toronto and distributed in the different venues in a daily basis.

The main problem of this business plan is that the distribution of the product becomes time consuming, inefficient, and costly. Our client wants to reduces the costs of distribution by building 5 smaller warehouse in Toronto to serve locally his clients. This approach will reduce the time which is required to serve his clients, the fuel costs, and will also transform his business to a green chain. To do so, our client asked from us to find the 5 best locations in Toronto at which the warehouses must be build in order to create smaller distribution clusters. The warehouses locations must be at the centers of these clusters in order to minimize the relative distance from each of the venues.

Data

The data required for this task are the locations of the hotels, coffee shops, bars and restaurants in Toronto. To gather the data we ‘ll use the locations of all neighborhoods in Toronto gathered from Wikipedia. Based on these location we ll gather the locations of all venues in these neighborhoods from Foursquare. We will filter the data to acquire the locations of the targeted venues. In order to inspect the data, we ll use the folium library to extract the map of Toronto and visualize the locations of the venues on the map.

A k-means algorithm will be applied on the location features to define the 5 clusters of venues. The locations of the warehouse will be defined as the centroids of the clusters. Again, to visualize the map of Toronto, the 5 clusters and the locations of the warehouses we ll use the folium

library.