Coursera Capstone Project

Introduction/Business Problem

Our client, a bottled water distributor, has an established role in the market of Toronto. He is the top supplier of bottled water with more than 1500 clients in the Toronto neighborhood. His main clients are hotels, coffee shops, restaurants, and bars. Currently, the products are stored in a big central warehouse outside of Toronto and distributed to different venues daily. The main problem with this is that the distribution of the product becomes increasingly time-consuming and costly. Our client wants to increase efficiency and reduce the cost by building 5 smaller warehouses in Toronto to serve his clients locally. This approach will reduce the time to spend on roads, fuel cost and become more environmentally friendly.

To do so our client asked us to find the best 5 locations I Toronto at which he can build warehouses in order to create smaller distribution clusters. After this, he will try to build his warehouses at the center of those clusters to minimize the distance to each venue.

Data

The data required are the locations of hotels, coffee shops, bars and restaurants in Toronto. To gather the data, we will use the locations of all neighborhoods in Toronto gathered from Wikipedia. Based on these locations we will 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 will 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 to the locations features to define the 5 clusters of venue. 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 will use the folium library.