Battle of the Neighborhoods

DATA SCIENCE CAPSTONE PROJECT

Business Problem

The scope of this project is to find the most suitable location for opening a new **Cafe Equipment Cleaning & Maintenance company** in Toronto North York. Key factors will be the number of cafes in the targeted area and the possibility of combining office location with place of residence.

Key Factors

An office location with many potential customers close by will raise revenue and minimize operational & marketing costs

It will be beneficial for the company owner to live close by to his office, as long as one of two important personal parameters are met

- The place of residence should be near parks
- There should be a variety of restaurants close by

Target Audience

Any individual who wishes to open a Cafe Equipment Cleaning & Maintenance company in North York Toronto and at the same time live in the same area.
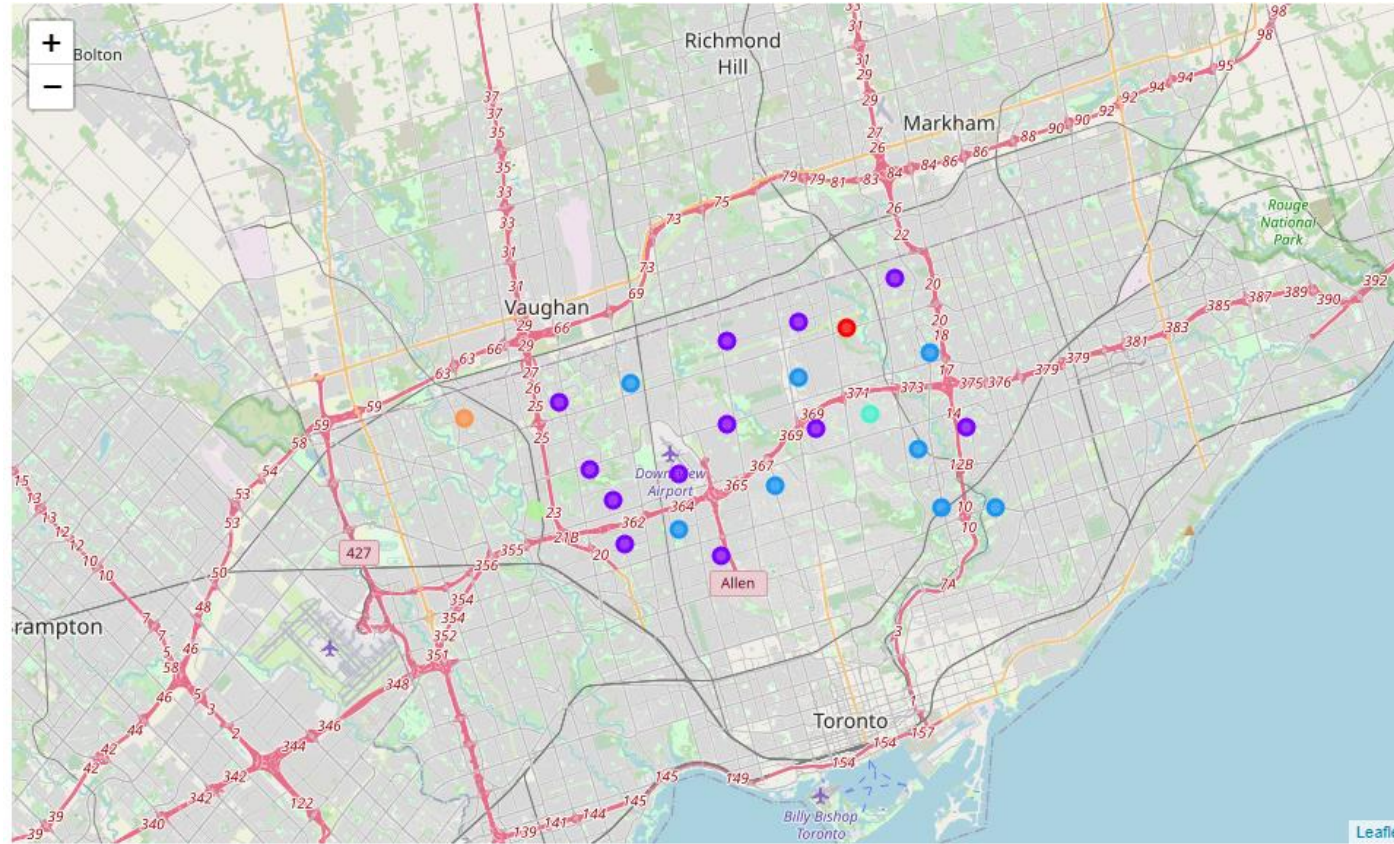
Data Sources

The necessary data and the way which will be used are:

1. Toronto neighborhoods from Wikipedia
2. Geographical coordinates of the neighborhoods with the respective Postal Codes
3. We will use Four Square API to explore North York Toronto number and type of venues

North York Toronto
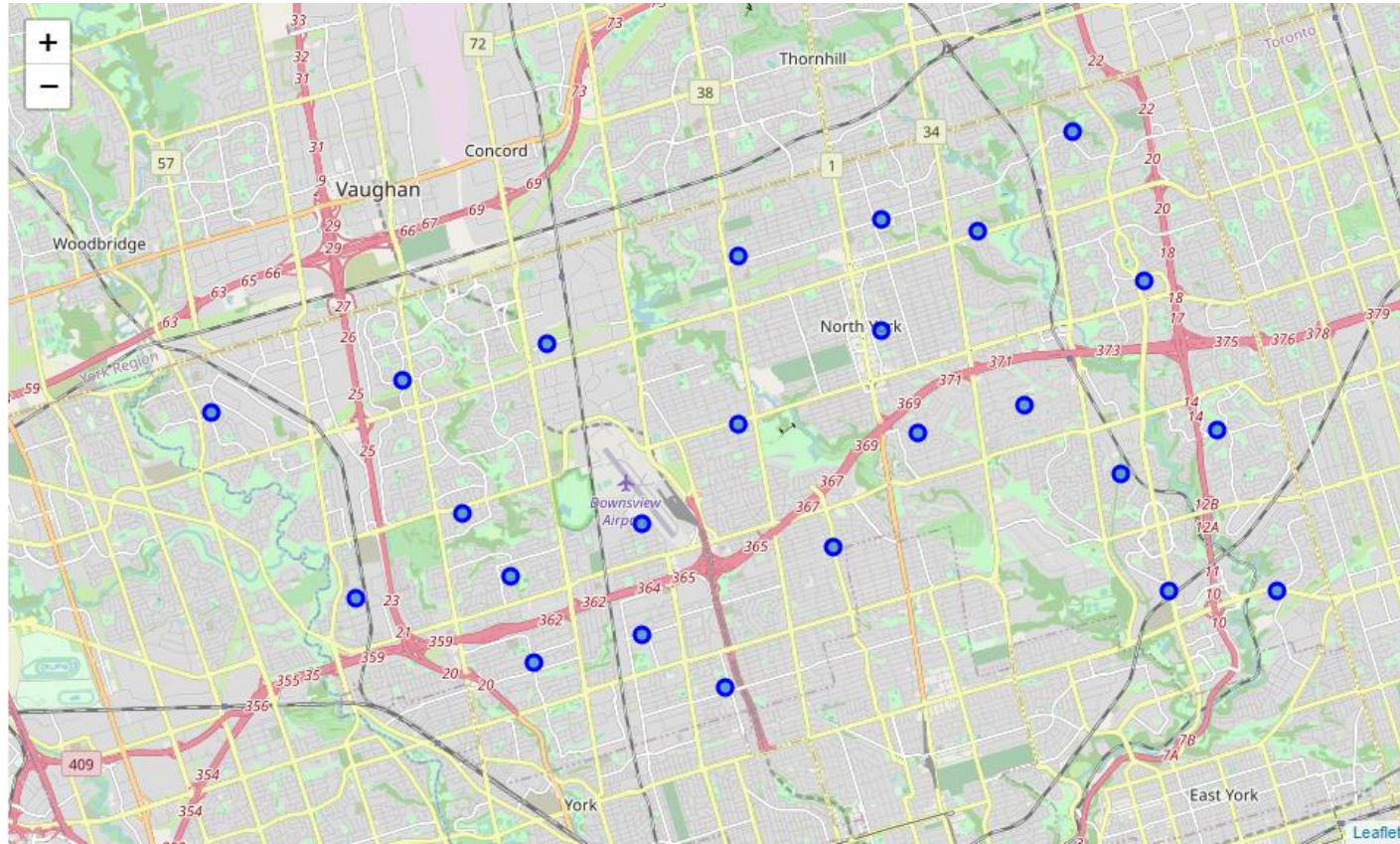
Cluster North York Neighborhoods based on common venue types

Cluster 1

Cluster 1 contains only Bayview Village neighborhood. There are a lot of Gas stations, banks & grocery stores in the area. Cafes are the 6th most common venue. Based on our desired parameters, Bayview Village is not a desired area to open a new office for our company or as a place of residence.

Cluster 2

Cluster 2 contains a bigger number of neighborhoods. We can see that we have either parks or cafes in top common venue types and there are a lot of restaurants. Based on our desired parameters, these neighborhoods are worth consideration as a desired area to open a new office for our company and as a place of residence.

Cluster 3

Cluster 3 contains also a big number of neighborhoods. We can see that Coffee shops are the most common venue by far. There are a lot of restaurants but not many parks. Based on our desired parameters, these neighborhoods are worth consideration as a desired area to open a new office for our company and as a place of residence.

Cluster 4

Cluster 4 contains only 2 neighborhoods (York Mills & Silver Hills). We can see that Parks are the most common venue. There are some restaurants nearby but not Coffee shops. Based on our desired parameters, we cannot consider these neighborhoods for opening a new office, but we can consider them as a place of residence.

Cluster 5

Cluster 5 contains only Humberlea, Emery neighborhood. We can see that Discount Stores and Convenience Stores are the most common venue. There are Parks nearby but not Coffee shops. Based on our desired parameters, we cannot consider these neighborhoods for opening a new office, but we can consider them as a place of residence.

Cluster 6

Cluster 6 contains only Humber Summit neighborhood. We can see that Electronic Stores and Banks are the most common venue. There are Parks nearby and some Restaurants but not Coffee shops. Based on our desired parameters, we cannot consider these neighborhoods for opening a new office, but we can consider them as a place of residence.

Final Results

Based on our overall results we will drop Clusters 1, 4, 5 & 6 with Neighborhoods as potential candidates for a new office location or a place of residence. Cluster 2 and Cluster 3 Neighborhoods seem more promising.

By examining more closely our data and after more investigation we decide not to consider Downsview as a candidate especially as a residential area because of the close proximity to the airport.

It seems that our best bets are North Park, Maple Leaf Park & Upwood Park from Cluster 2 and all Neighborhoods of Cluster 3.

Recommendations

Our efforts would be more precise if we could take under consideration:

- General competition in the area

- Coffee shop types in order to determine the possible new equipment or maintenance budget

- Import other parameters for our residential candidates such as crime rates, median income etc.