

FAMILY RELOCATION TO NORTH YORK

Capstone Project for IBM Data Science

By Coursera

Igor Svidritskiy
December 24th, 2019



- In The era of globalization there are many companies with subsidiaries all over the world. They have production facility in the countries with relatively cheap resources and not too far away from the markets where it's product to be sold. The headquarter or main offices of the business units are usually located in the financial centers of continents or influential global cities.
- To be competitive those companies have developed strategies, policies and procedures based on the values essential for today's business environment. Human factor is considered as the most important for a technological breakthrough and long-term development.
- One of the options to retain and develop talented employees is to offer an international assignment contract. It definitely keeps a person within the company and allows to bring new ideas/methods or even style into a different subsidiary – Diversity is a key for innovation!



PROBLEM
BACKGROUND

- An employee from Moscow sales office, Russia has been offered an international assignment for 3 years to work in Con**'s office in Toronto, North York Borough in particular, Canada. He has been working in Con** for 10 years, has a family with 2 kids of school age. Moreover, he is quite active, with healthy lifestyle and willingness to develop his children as much as possible. And they all are Christians.
- This employee has accepted the new job opportunity. Now it's time to decide what are the main preferences to consider for relocation.



PROBLEM
DESCRIPTION

- Developed neighborhood
- Gym nearby
- Sport facility for kids
- Restaurants of different cuisine in a walking distance
- Music hall or a theatre in proximity
- Also walking distance to a Catholic school

LIST OF FAMILY PREFERENCES

DATA

- Dataset 1
- City of Toronto with Postal codes, Boroughs and Neighborhoods

Out[3]:

	PostalCode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

DATA

- **Dataset 2**
- City of Toronto with Postal codes linked to their geographical coordinates.

Out[5]:

	PostalCode	Latitude	Longitude
0	M1B	43.806886	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

DATA

- **Dataset 3**
- Toronto Catholic District School Board is a governmental web site representing online resources

	SchoolName	Panel	Address
17	Bishop Allen Academy	Secondary	721 Royal York Road Etobicoke ON M8Y 2T3
19	Bishop Marrocco/Thomas Merton	Secondary	1515 Bloor Street West Toronto ON M6P 1A3
24	Brebeuf	Secondary	211 Steeles Avenue East Willowdale ON M2M 3Y6
28	St Basil-The-Great	Secondary	20 Starview Lane North York ON M9M 3B2
38	Cardinal Carter Academy	Secondary	36 Greenfield Avenue Willowdale ON M2N 3C8

Install all the required Python libraries

```
In [2]: #import libraries
import pandas as pd
import numpy as np
from sklearn.cluster import KMeans
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)

import json # library to handle JSON files
from pandas.io.json import json_normalize

import geopy
from geopy.geocoders import Nominatim

import requests
from bs4 import BeautifulSoup

import matplotlib.cm as cm
import matplotlib.colors as colors

!conda install -c conda-forge folium=0.5.0 --yes
import folium
```

Solving environment: done

All requested packages already installed.

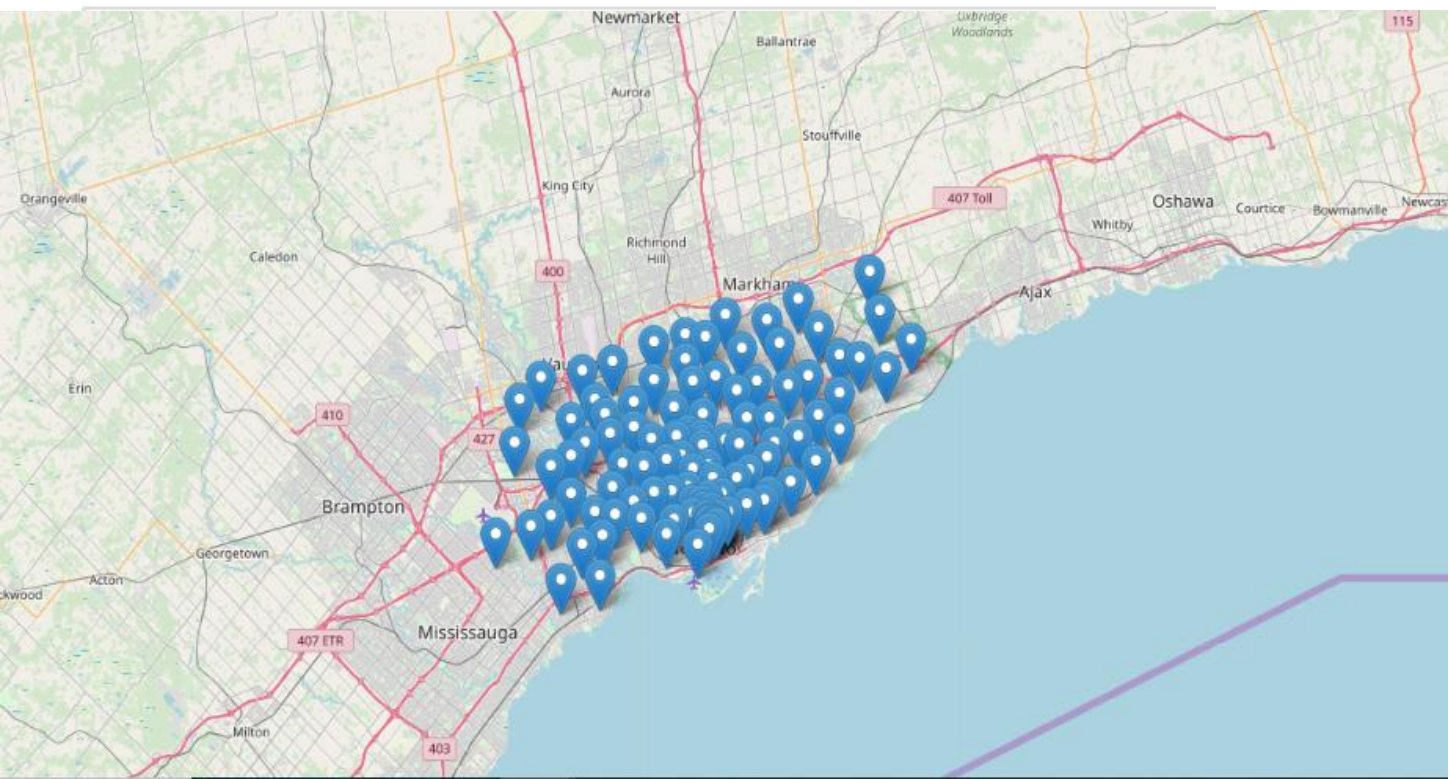
ANALYSIS

We parse Wikipedia web page to get 103 postal codes of Toronto with Boroughs and Neighborhoods they belong to.

Then we merge this info with geographical coordinates of each postal code

Out[6]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806888	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476



ANALYSIS

Let's select now North York since that is a designated point of the family to move to eventually.

And a step before showing the map of North York let's use the Foursquare API to get Venues or Points of Interest (POI) information of the area. As you can see below – there are 243 POIs in the selected area of 23 postal codes.

(243, 9)

Out[11]:

	PostalCode	Borough	Neighborhood	Lat	Lon	VenueName	VenueLat	VenueLon	VenueCategory
84	M2H	North York	Hillcrest Village	43.803762	-79.363452	Eagle's Nest Golf Club	43.805455	-79.364186	Golf Course
85	M2H	North York	Hillcrest Village	43.803762	-79.363452	New York Fries	43.803684	-79.363905	Fast Food Restaurant
86	M2H	North York	Hillcrest Village	43.803762	-79.363452	AY Jackson Pool	43.804515	-79.366138	Pool
87	M2H	North York	Hillcrest Village	43.803762	-79.363452	Villa Madina	43.801685	-79.363938	Mediterranean Restaurant
88	M2H	North York	Hillcrest Village	43.803762	-79.363452	Duncan Creek Park	43.805539	-79.360695	Dog Run



ANALYSIS

Machine Learning clustering with “KMeans” algorithm gives us the following map.

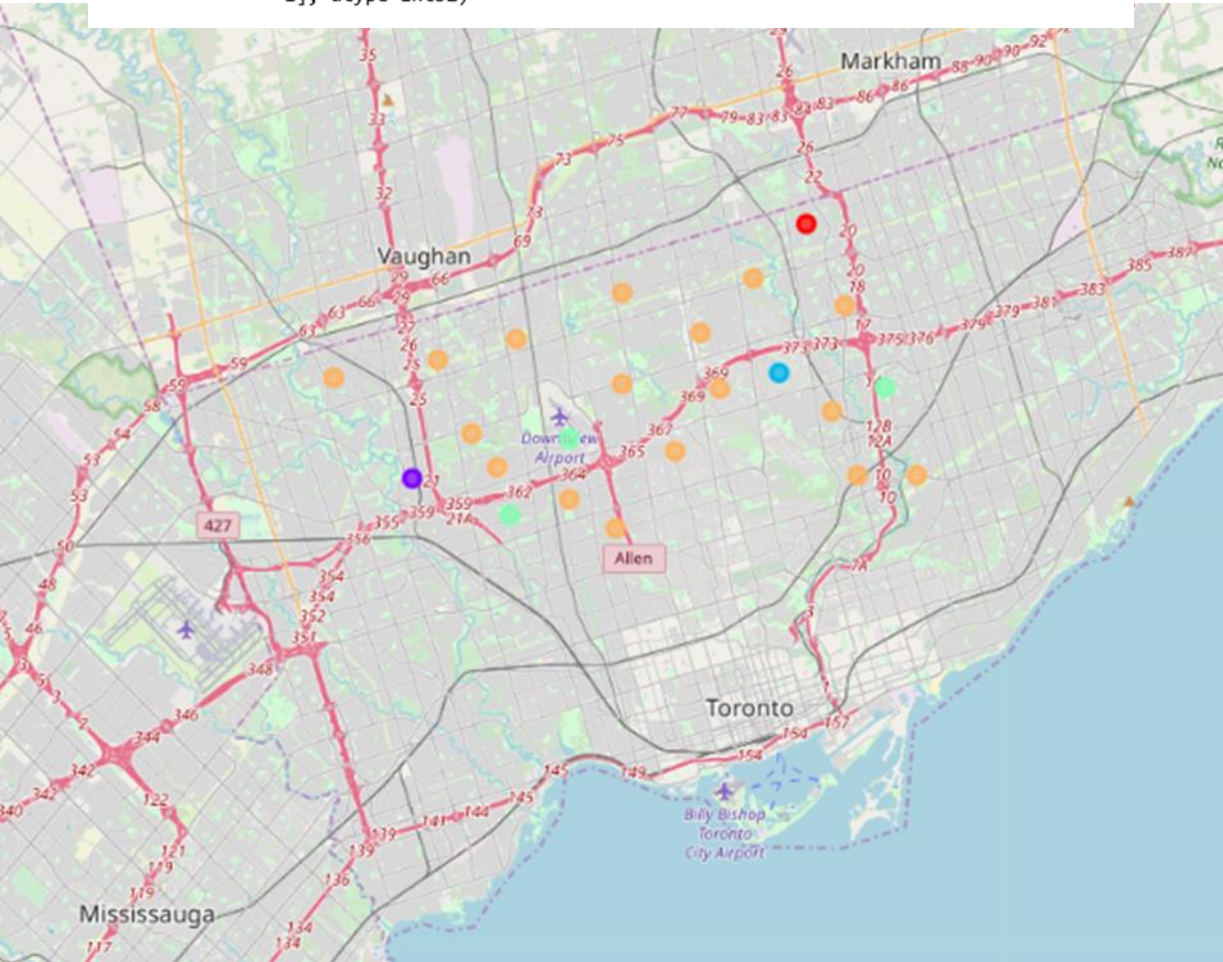
We select 5 clusters partitioning resulting into the array:

```
In [19]: #cluster the Area
kclusters= 5

NY_clustering= NY_grouped.drop(["PostalCode", "Borough", "Neighborhoods"], 1)
kmeans= KMeans(n_clusters= kclusters, random_state= 0).fit(NY_clustering)

kmeans.labels_[0:24]

Out[19]: array([0, 4, 4, 2, 4, 4, 4, 3, 4, 4, 4, 4, 3, 4, 4, 4, 4, 4, 4, 3, 4,
1], dtype=int32)
```



ANALYSIS

The Cluster 5 is the most developed and, the most important, -
suits the purpose of this Project the most.

```
In [28]: #Cluster 5
NY_new.loc[NY_new['Cluster Labels'] == 4, NY_new.columns[[0] + list(range(2, NY_new.shape[1]))]]
```

Out[28]:

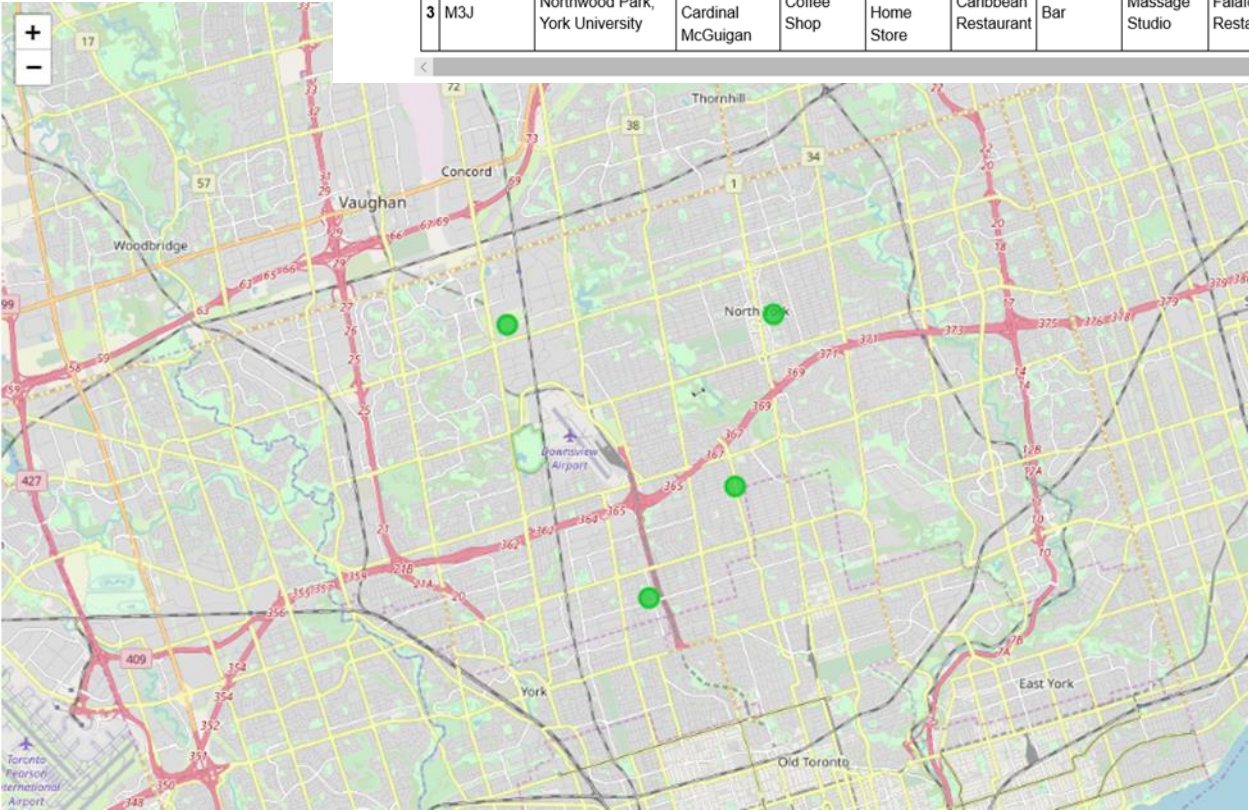
	PostalCode	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
19	M6B	Glencairn	Pizza Place	Bakery	Pub	Japanese Restaurant	Women's Store	Diner	Concert Hall	Construction & Landscaping	Convenience Store
18	M6A	Lawrence Heights, Lawrence Manor	Clothing Store	Accessories Store	Sporting Goods Shop	Miscellaneous Shop	Coffee Shop	Furniture / Home Store	Boutique	Gift Shop	Vietnamese Restaurant
17	M5M	Bedford Park, Lawrence Manor East	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Sandwich Place	Restaurant	Liquor Store	Juice Bar	Café	Pharmacy
16	M4A	Victoria Village	Intersection	Pizza Place	Portuguese Restaurant	Financial or Legal Service	Hockey Arena	Coffee Shop	Diner	Concert Hall	Construction & Landscaping
15	M3N	Downsview Northwest	Grocery Store	Gym	Gym / Fitness Center	Discount Store	Women's Store	Dog Run	Construction & Landscaping	Convenience Store	Cosmetics Shop
14	M3M	Downsview Central	Home Service	Food Truck	Baseball Field	Women's Store	Dog Run	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega
13	M3L	Downsview West	Grocery Store	Bank	Hotel	Park	Shopping Mall	Women's Store	Discount Store	Concert Hall	Construction & Landscaping

ANALYSIS

At the last stage of the Analysis we have found that out of 31 Catholic Schools located in North York Borough, only 4 meet the criteria to be in Cluster 5.

Out[36]:

	PostalCode	Neighborhoods_x	SchoolName	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	M6B	Glencairn	Dante Alighieri	Pizza Place	Bakery	Pub	Japanese Restaurant	Women's Store	Diner	Concert Hall	Construction & Landscaping	Convenience Store
1	M5M	Bedford Park, Lawrence Manor East	Loretto Abbey	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Sandwich Place	Restaurant	Liquor Store	Juice Bar	Café	Pharmacy
2	M2N	Willowdale South	Cardinal Carter Academy	Ramen Restaurant	Coffee Shop	Sushi Restaurant	Café	Pizza Place	Sandwich Place	Juice Bar	Japanese Restaurant	Lounge
3	M3J	Northwood Park, York University	James Cardinal McGuigan	Coffee Shop	Furniture / Home Store	Caribbean Restaurant	Bar	Massage Studio	Falafel Restaurant	Discount Store	Construction & Landscaping	Convenience Store



ANALYSIS



RESULT

Cluster 5 of North York Are is the most developed. It has a lot of restaurants and cafés, small shops, department or grocery stores. There different sport facilities available like Gyms, a Pool, a Baseball Field and even a Hockey Arena. This Cluster also has a Concert Hall and number of parks.

However, there are only 4 Catholic Schools located in this Cluster, and there might be a compromise decision to cover only some of the Family preferences and being close to one of four Schools.

A hand holding a silver pen points towards a financial candlestick chart displayed on a screen. The chart features green and red candlesticks, with various numerical values and percentages visible. The background is dark blue with white and yellow text. The pen is positioned diagonally across the frame, pointing towards the lower right.

RECOMMENDATION

Out of the initial list of preferences we meet the following ones:

- Developed neighborhood
- Restaurants of different cuisine in a walking distance
- Music hall or a theatre in proximity
- Also walking distance to a Catholic school

Gym and sport facility for kids are located not within 500 m around the dedicated postal codes. So, the Family has to decide if 4 out of 6 main criteria are suffice.

If the answer is YES, then the parents should foresee School's gym facility for the kids and 1-1.5 km distance to the adult gym for themselves which is not a bad option for newly settlers.

- The scope of this analysis based on a limited data. There only 3 datasets are used, and the primary selection of Borough North York assumes that the relocating Employee walks to the office.
- However, if we consider a short commute to the office of, say, 10-15 minutes of drive, then the searching area could be extended to the Boroughs of Vaughan, Markham and Toronto Downtown. That can bring a lot more option for the Family to decide where to live for 3 years.



CONCLUSION