Data Science Capstone Project Report

French gourmet restaurant grand opening

By: Thomas Chanier

Date: 9/11/2019

Introduction

We want to open a new French gourmet restaurant in Toronto. We are looking to open in a rich neighborhood where a very few restaurants are implanted. For that purpose, we need to classify neighborhood in terms of income. We will use 3 neighborhood classifications: 1: Low Income, 2: Medium Income, 3:High Income. We explore the High Income classification and find the neighborhood with the lowest number of restaurants. We will open a restaurant in this neighborhood.

Data

We will take Toronto postal code data and geolocalization data for each neighborhood and demographic data of Toronto from Wikipedia. We will use kMeans to classify neighborhoods by income. We will obtain three classifications which will be analyzed statistically to determine the socio-economic status of each neighborhood. We will then use Foursquare to determine the number of restaurants in each neighborhood. We decide to open a restaurant in a high income neighborhood with the lowest number of restaurants.

Data:
Canada postal code: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Geospatial data: http://cocl.us/Geospatial_data
Population data: https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods
Foursquare API: number of restaurants in a given neighborhood

Methodology

We use the kMeans algorithm to classify Toronto neighborhoods per income. kMeans is an unsupervised machine learning algorithm which will group the data by Income. We will print

these results on a map. We determine the number of restaurants for each neighborhood using Foursquare API and select the richest neighborhood (results of kMeans).

Results

1. **Toronto neighborhoods per household income**

The results of the kMeans algorithm are presented on Table 1. The low income category contains 67 neighborhoods for an average income of $ 34396, the middle income contains 10 neighborhoods for an average income of $ 88471 and high income contains 3 neighborhoods for an average income of $ 194292. The results are plotted on Figure 1.

LABEL	Latitude	Longitude	Population	Income	Density
0	43.689903	-79.412407	12670.300000	88471.500000	6150.600000
1	43.709180	-79.382418	15862.208955	34395.865672	6928.582090
2	43.699052	-79.383160	6266.333333	194292.000000	2869.333333

Table 1: kMeans results: The data are split into 3 categories: 0: medium income, 1: low income and 2: high income.
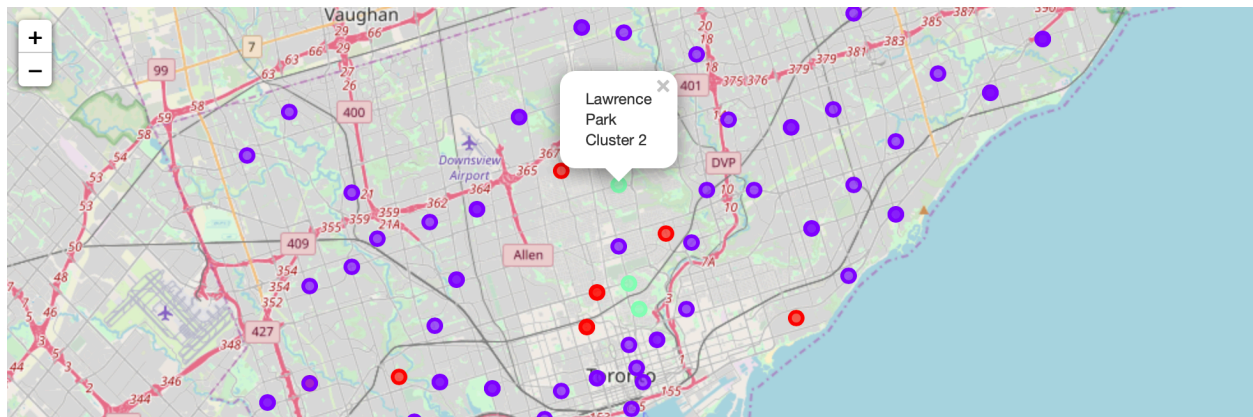

Fig 1: Toronto neighborhoods per household income. High Income are in green, Middle Income are in Red and Low Income are in Purple.

2. Toronto neighborhoods restaurants

We use the Foursquare API to determine the number of restaurants for each neighborhood. Table 2 is the head of the obtained dataframe:

	Postcode	Latitude	Longitude	Borough	Neighborhood	Population	Income	Density	LABEL	restaurant
0	M1B	43.806686	-79.194353	Scarborough	Rouge	22724.0	29230.0	791.0	1	18.0
1	M1B	43.806686	-79.194353	Scarborough	Malvern	44324.0	25677.0	5003.0	1	18.0
2	M1C	43.784535	-79.160497	Scarborough	Highland Creek	12853.0	33640.0	2505.0	1	5.0
3	M1C	43.784535	-79.160497	Scarborough	Rouge Hill	11167.0	32858.0	2878.0	1	5.0
4	M1C	43.784535	-79.160497	Scarborough	Port Union	12450.0	48117.0	2310.0	1	5.0

Table 2: Dataframe obtained after Foursquare API call. The column restaurant contained the number of restaurant per neighborhood.

We then choose the neighborhood with the highest income (Table 1). The results are presented on Table 3. We then decide to implant our new French gourmet restaurant in Lawrence Park neighborhood which has the lowest number of restaurant.

	Postcode	Latitude	Longitude	Borough	Neighborhood	Population	Income	Density	LABEL	restaurant
35	M4N	43.728020	-79.388790	Central Toronto	Lawrence Park	6653.0	214110.0	1828.0	2	9.0
37	M4T	43.689574	-79.383160	Central Toronto	Moore Park	4474.0	154825.0	3959.0	2	59.0
40	M4W	43.679563	-79.377529	Downtown Toronto	Rosedale	7672.0	213941.0	2821.0	2	27.0

Table 3: Number of restaurants for Toronto richest neighborhoods.

Discussion

The target or attractiveness was high income neighborhood with the lowest number of restaurants. The real estate cost is very expensive in these areas. We can look at medium income area too where we can find also a good tradeoff between real estate cost and attractiveness. Among these neighborhoods, the York Mills neighborhood seems also to be a good choice because it has a middle-high household income of $ 92099 and only 4 restaurants.

Conclusion

To maximize the attractiveness of our new French gourmet restaurant, it must be located in a rich area with a few other restaurants. Based on a kMeans algorithm analysis, the socio-economic status of each Toronto neighborhoods has been studied. We have determined the number of restaurants per neighborhood by the Foursquare API. Based on the results, we will

choose to implant our new restaurant in Lawrence Park neighborhood, Toronto. York Mills neighborhood will be considered too.