COURSERA IBM DATA SCIENCE CAPSTONE PROJECT

Opening a new restaurant in Toronto, Canada

Publication by Muzamal Azam

April 27th, 2020

Introduction

For this project, given a hypothetical scenario where a businessman would like to open a restaurant and he/she is interested to learn more about the city by having an insight of where he can develop his business and the type of customer he may attract given depending where he might open it

Business Problem

Mr Bellucci would like to open an Italian pizza restaurant and he needs to understand better each neighbourhood in Toronto City, where there might be a good idea to open a restaurant and which neighbourhoods are popular

Audience

Italian Restaurant owner

Data

In order to make a rough analysis the current market and the points of interests in Toronto we will use opendata provided by Toronto City. Please note that the data was scrapped before importing on python in order to simplify the processing.

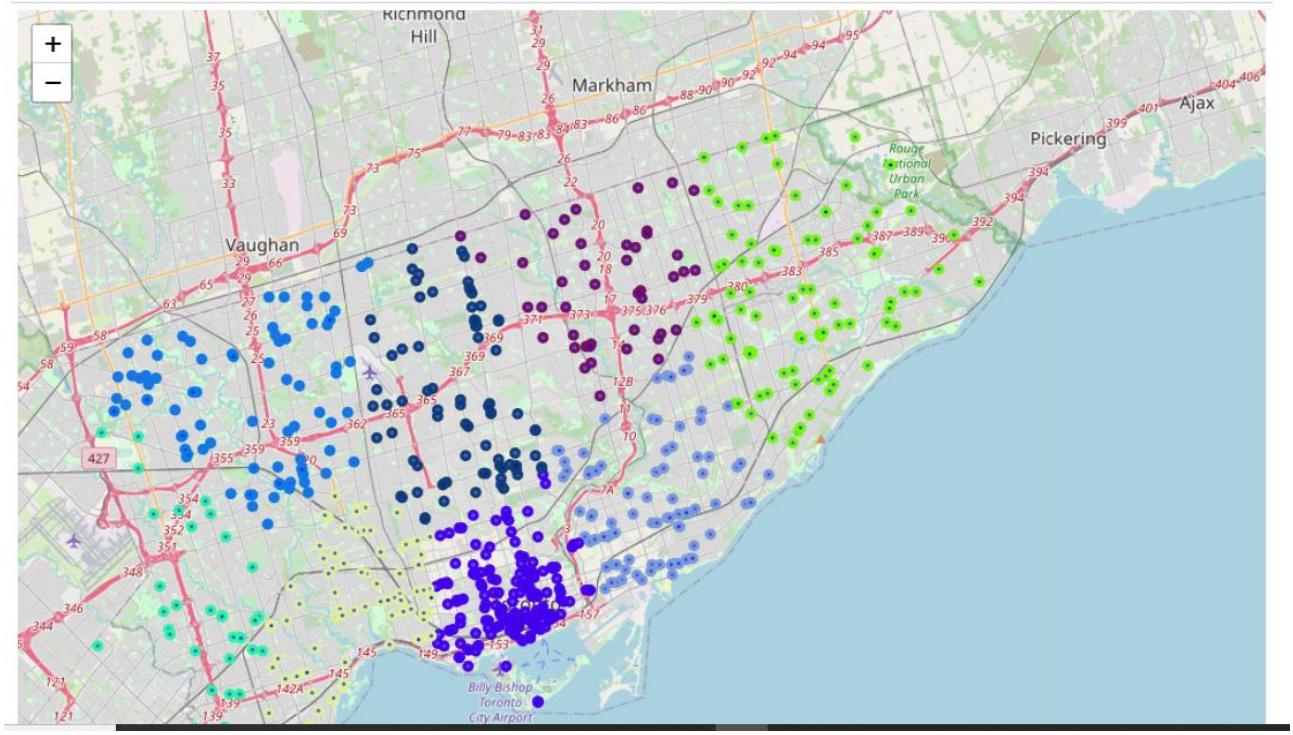
- Dataset with Postal Codes associated to their bougher and neighbourhood
- Dinesafe provided by the department of health of Toronto
- Toronto points of interest

Methodology

Given a set of postal codes of Canada associated to their neighbourhood and a set of coordinates for each postal code I'll make clusters of the city with K-mean and draw a map of the neighbourhood associated to their cluster.

Then I'll be assigning the list of restaurants to the closest cluster and understand where there is a high density of pizza shops.

Results



As we can see the cluster 5 is the most denser in terms of pizza shops as it is part of the old town of Toronto and it's a residential area.

Cluster number	Pizza shops #
0	70
1	65
2	71
3	80
4	77
5	144
6	27
7	35

But if we take into account the point of interests of Toronto these are the results which shows that the cluster 3 has most of the points of interest which makes it perfect for tourism attraction and it's a lot better in terms of business competition as there are fewer businesses than cluster 5

Cluster number	Points of interests
0	8
1	0
2	25
3	102
4	64
5	0
6	54
7	1

Considerations

The cluster number 3 has been identified as one of the possible locations to open the business, however it would be ideal to have the data regarding the rents and make a fair comparison and also the data of Defaults within the entire city to assess the risks of the business. We haven't used forsquare as it is not the right tool to extract the entire list of restaurants of the entire city due to its pricing limitation.

Conclusions

In this project we have identified a business problem and provided a possible solution with the limited data accessible , using clustering and geographic drawing libraries.