Restaurant prediction and recommendation system

Content:

1. Introduction

This project has been created in the interest of IBM Data Science professional Certification: capstone project.

Here, in this project we can figure out for the businessman in Toronto, who will require opening a restaurant. We will figure out which location will suit the businessman to open the restaurant, specifically a cafe. In this project, we can assume the bet location, and the place of interest for the people who will be interested for the Café near their location.

2. Data Selection and Source of Data:

To analyze and predict the location of the café, we will use the Toronto location data with pin code, from it's Wikipedia page(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). The data is scrapped from the Wikipedia page using Beautifulsoup api and dataframe is obtained using python.

A preview of the dataframe obtained from the data set is given below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Po

We're going to predict two things with the dataset and the neighborhood location,

- First is to predict the cafe location in Toronto, where there isn't a café nearby. We will group 50 venues with in 1km radius and cluster them, to check if there is a café with in the cluster.
- Then our object will be to select the location, where a café can be opened.

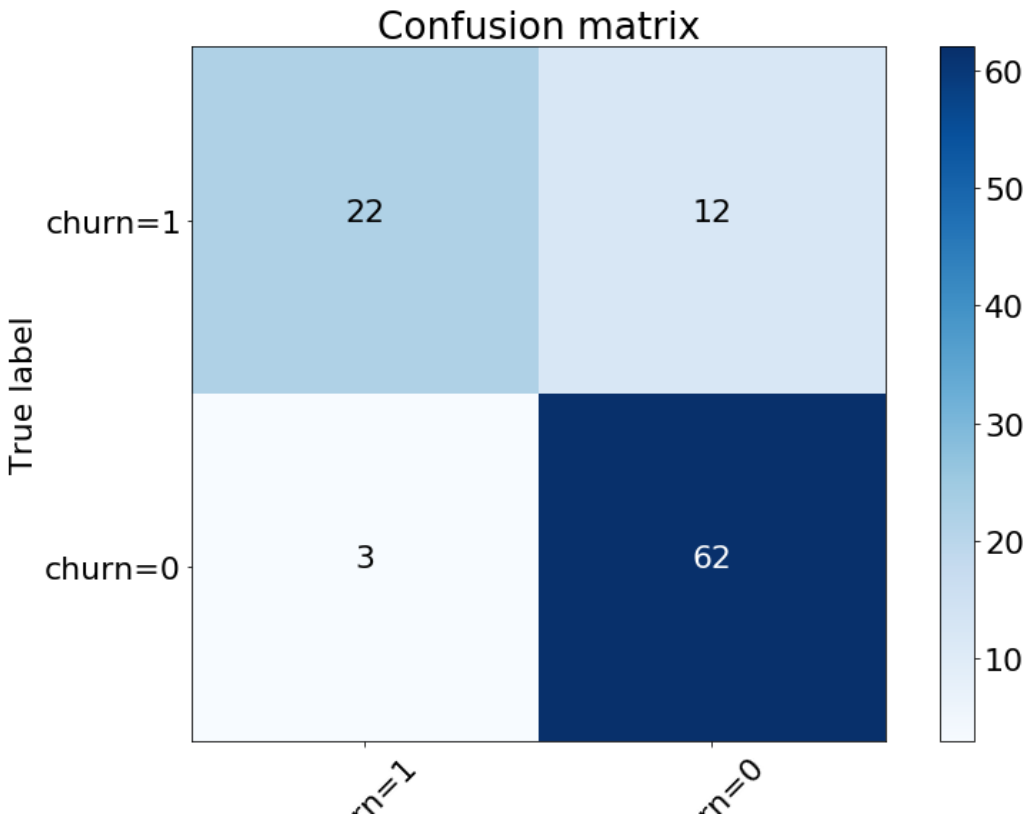
3. Result

The first step is to transform the data frame in to a binary code. It will be substituted by 1 if the venue is present in the neighborhood else it will be 0.

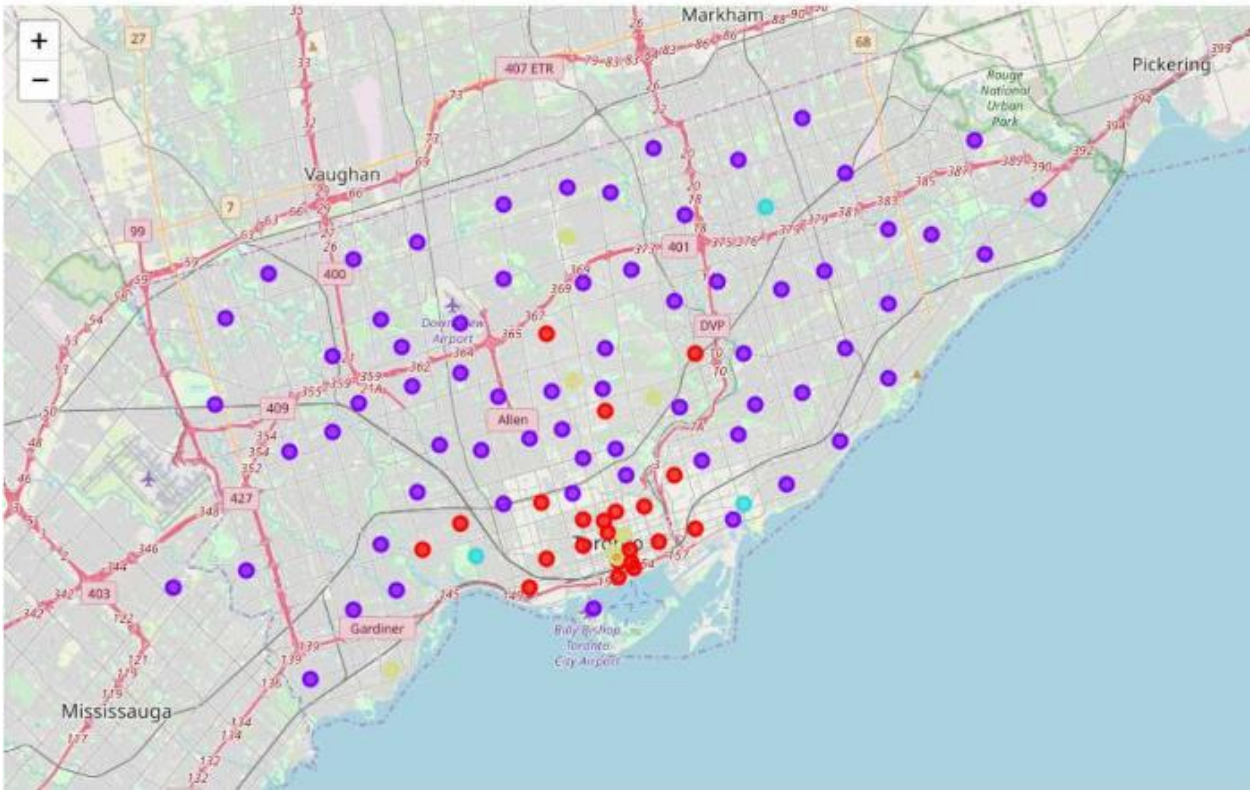
The below image shows the result of the dataframe.

	Yoga Studio	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Baker
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.

A model is predicted of existence if café in different neighborhoods. To obtain so the data is splitted in to test dataset and train dataset. The confusion data matrix obtained from the dataset is presented below.

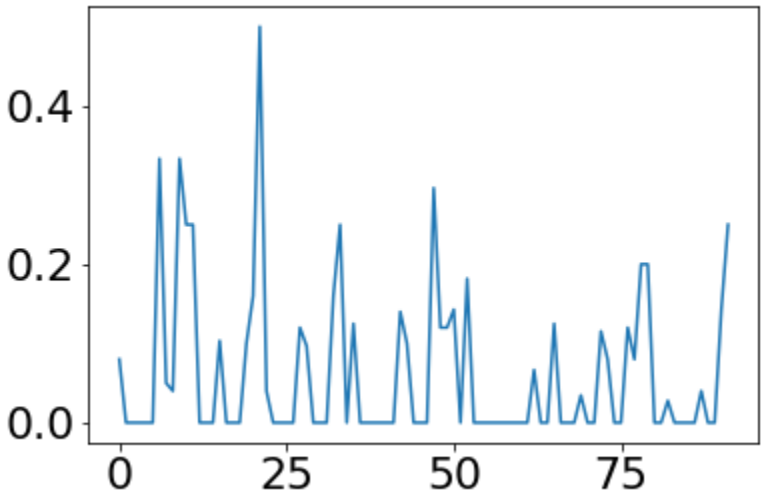


As from the figure it can be seen that 3 places are there where Café could be there, but it's not there in the city. This can be presented as a folium map also. Refer the below image for the map obtained for the dataset.



The neighborhoods where a café is not present are the best places to start with for the project.

To select the best place for a café, linear regression model can be used. Due to the vast number of features 3 most suitable places are selected. The co-relation of the model is used where the absolute peak in the diagram, are the best suitable places to start with.



Conclusion:

In the study, we obtained the 3 most suitable place to start with for opening a café , with the basis of the location and taking the neighborhood in to consideration. There is a lot of scope of work in the subject and a lot of factors can also be taken in to consideration to provide a better output.

But, for the scope of the project 3 places are obtained and hence it concludes the project with a proper output and a model.