Lebanese restaurant in Toronto

1- Introduction

1.1 Background

Lebanese cuisine is an ancient one and part of the cuisine of the Levant. Many dishes in Lebanese cuisine can be traced back thousands of years and were influenced by different foreign civilizations that held power each in an era where the most recent one was the French after World War 1, therefore it is loved and intended by a lot of people from different cultures and nationalities.

Toronto is a multinational city which has a lot of Lebanese people residing in it, therefore an idea of a Lebanese restaurant can be very appealing and profitable for any businessman with an investing mind.

1.2 Problem

The challenge here is to find the most suitable location for this restaurant to be successful and profitable. A bad location choice can be crucial and lead to the restaurant closing, therefore I am going to use machine learning and data science and study different neighborhoods and venues in order to get the best location possible.

1.3 Interest

Any entrepreneur who wants to make more money would be really interested in this project.ZX????

2- Data acquisition and Cleaning

`

2.1 Data Sources

For the data, first I need all the information to explore and cluster the neighborhoods in Toronto which can be found in a Wikipedia page here. Then I need a csv file here that has the geographical coordinates of each postal code and last I can utilize the Foursquare location data.

2.2 Data Cleaning

The data from the Wikipedia page is returned in a form of array of tables, but the info I need are in the first table which made it easy for me to get. However the received table is full of empty values and boroughs that are not assigned. I removed all the unassigned boroughs and filtered the data on boroughs that contains the word ‘Toronto’ since I’m only interested in this city and not all of Canada.

Then I download the CSV file which contains the longitude and latitude linked to the Postal Code. The column ‘Postal Code’ can be found in both datasets and contain the same data, therefore I can merge them into one dataset which contains all the needed data with their coordinates. Only one small change needs to be done which is to rename the column ‘Postal Code’ in the first dataset because it is not written in the same case as the second set and the merge is case sensitive on the column name. Now I’m ready to work with the foursquare API and explore the neighborhoods using the explore function to get the locations of restaurants and then use this feature to group the neighborhoods into clusters and decide which is the best location for our restaurant to be in.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
2	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
19	M4E	East Toronto	The Beaches	43.676357	-79.293031