Coursera Capstone

IBM Applied Data Science Capstone Project

Finding the most suitable place to open a restaurant in Tororntо

By: Ajay Mathew

July 2020

Introduction

Whenever someone wants to establish a new business, that person would want it to be successful. For the business to be successful, there should be minimum competition for customers in its surrounding area. The aim of this project is to identify places within Toronto that have increased chances of profitability if a restaurant is opened there.

The target audience: The target audience of this project might be a single person or a business group hoping to move to a new city or start their first ever restaurant. For the businesses like restaurants, the less the competition, the better. So, we would be trying to find out places in Toronto that has less number of restaurants.

Data

To solve this problem, we need the following data:

1, The list of neighborhoods and boroughs in Toronto

2, The coordinates of all the neighborhoods

3, Venue data

Sources of Data:

For the data of the list of neighborhoods in Toronto, we could simply scrape the Wikipedia page using the pandas.read_html() method. The resulting dataframe after cleaning it is:

	Postal Code	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

The coordinates for all the nwighborhoods can be obtained from the link : " http://cocl.us/Geospatial_data" . The corresponding dataframe is :

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-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

Combing both these dataframes, we get:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

This DataFrame would be the basis of all our data analysis.

Geocoder library will also be used to retrieve coordinates of needed locations as per demand.

We would be utilizing the Foursquare API to get venue data for the selected neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by more then 125,000 developers. The resulting Json file from the query would be converted into DataFrames for futher use in our program. The corresponding DataFrame is:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790