Investment in Popular Venues of Neighborhoods

Manoj Banisetty

May 20, 2020

1. Introduction

1.1 Background

Rich people are all across the globe. They earn very much. The main aim of everybody is to have money and should be able to double what we have. So , that's where the concept of investment came into play. Everybody likes to invest what they have in something they like to earn more from that investment. So, investment is important in everyone's life. Investment can be done in anything now. Be it a Corporate Company , Multi National Company , or it could even be in restaurants, shops,etc. Anything a person believes they can gain from, they will invest in it.

1.2 Problem

Data that might help investors who are looking to invest in anything which is popular around a particular city/neighborhoods – in our case, we will be checking the neighborhoods of Toronto- to understand the people living in that neighborhood. We can simply understand it by the derived data by estimating what business is more prevailing in that city/neighborhood so that the investors can invest in that business.

1.3 Target Audience

The advantage in this problem is that we don't need to search for a particular person who is looking to invest in a particular stream. Instead, any person who wants to invest in any business can look up at this data to understand and decide where they will invest their investments. Any person who is looking to invest in anything can refer to these conclusive datasets and come up with a plan to setup a business at a particular neighborhood where there is more demand for a particular product.

2. Data acquisition and cleaning

2.1 Data sources

I have obtained the required data from the following web pages:

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

This link consists of all the Neighborhoods, Boroughs along with their postal codes.

	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

This CSV file consists of all the geographical coordinates of the Neighborhoods, Boroughs, etc.

2.2 Data cleaning

1) Firstly, we will remove all the rows which contain the 'Not assigned' or 'NaN' values from the postal codes dataset.

2) Secondly, we will do the same with the dataset containing the geographical coordinates of the Neighborhoods and all the venues present in their particular neighborhoods.

3) Lastly, we can merge the 2 data sets into one by using the common column in both the data sets which is the Postal Code column which in future will help us locate the coordinates on the map.

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn

2.3 Feature selection

The features required in the selection from the data is we need all the coordinates of the different venues present in different neighborhoods.

We have to particularly select a neighborhood and go through all of the various venue categories in order to arrive to a particular decision and mapping those venues