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 Interest

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.

2. Data acquisition and cleaning

2.1 Data sources

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

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

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

https://cocl.us/Geospatial_data

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.

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