Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Toronto

Gabriele Minep

May 2020

Introduction

Aa today, shopping mall are plenty of shoppers, grocery shopping, restaurants and cinemas, etc.
A shopping mall could be a really interesting investment for real estate investor, but it will be really important the chosen location. This criterium will determine whether the mall will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Toronto, to open a new shopping mall. Using data science methodology and the techniques of clustering, this project aims to provide solutions to answer the business question if In the city of Toronto, a real estate developer should open a new shopping mall.

Target Audience of this project

This project is particularly useful to real estate developers and investors looking to open or invest in new shopping malls in Toronto.

Data

To solve the problem, we will need the following data:

- List of neighborhoods in Toronto.

- Latitude and longitude coordinates of above-mentioned neighborhoods. This is required in order to plot the map and also to get the venue data.

- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

The Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, contains a list of neighborhoods of Toronto, with a total of many neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. In the next section, will be presented the Methodology section where it will be discussed the steps taken, the data analysis taken.