**Coursera**

IBM Applied DatabScience Capstone

Opening a New Shopping Mall in Toronto



**Introduction**

The mall culture has become a big business, as they have become multi-story structures that house a large number of shops selling various products and services. They are a collection of shops and restaurants, adjacent to the pedestrian zone or an exclusive pedestrian street.For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. Since shopping centres are the most sought-after shopping destinations, it is beneficial for a businessman to set up a store in a shopping mall. Generally, retail store owners rent shop space in a mall. Renting store space benefits the businessman in many ways. Shopping malls are usually located in prime locations which are easily accessible. If a retailer sets up a store in a mall, he can have the shop in a prime location with a minimal investment. On the other hand, if he attempts purchasing a shop in such a location, he may not be able to afford it. A rented property implies low initial investment. This enables the businessman to utilize the saved amount on his business.

**Business Probelm**

The objective is to analyse and select the best location in the New York city to open a new shopping mall. Using Data science methadology and Machine Learning techniques.

**Target Audience of this project**

The project will bbe useful for the property developers and buisnessman looking to open or invest in the new shopping malls Toronto.

**Data**

**Data required:-**

* List of neighbourhood in Toronto.
* Latitude and Longitude coordinates of neighbourhoods.
* Venue data, particularly data related to shopping malls. We will use ths data to perform clustering on teh neighbourhoods.

**Mehtod to extract the data**

We will use the Webpage to collect the data (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>), contains a list of neighbourhood in Toronto. We will use the web scrapping techniques to extract the data, with the help of Python requests and Beautifulsoup packages. Then we will collect the geographical coordinates using the geocoder package which will give us the latitude and longitude coordinates of the neighbourhood.

We will use Foursquare API to get the venue data for those neighbourhoods. This API will provide many categories of the venue data, we are particularly intrested in the shopping mall category in order to help us to solve the Business problem.

**Result**

We will get the moderate number of clusters with the highest number of shopping malls. It will represent a great oppertunity and high potential areas to open new shopping malls.