**IBM DATA SCIENCE CAPSTONE**

**COURSERA CAPSTONE**

**OPENING A NEW RESTRAUNT IN NORTH YORK,TORONTO,ONTARIO**

**By**

**Piyush Wadhawan**

**INTRODUCTION**

A client has a Japanese style restaurant in Wilson Heights,Toronto which is a nice neighborhood with a traffic heavy population and as there are very few Japanese restaurants in Toronto and also because the district has a lot of restaurants due to which the restaurant attracts a large customer base and a majority of them went on to become returning customers who spread a good word creating more and more customers.So basically the location of the shop is a major factor in determining a profitable venture.

**Business Problem**

The purpose of this capstone project is to determine a suitable location for our client to open a restaurant where he will make a good profit.Using machine learning techniques like clustering to determine suitable locations for the client to expand his bussines in North York,Toronto .So the question is if someone wants to open a restaurant which place would you recommend?

Any invester looking to invest in the restaurant business could make use of this project or anyone looking to open a restaurant in general and also other owners looking to expand their franchise.

**Data used in the project**

List of neighborhoods in Toronto.

Latitude and longitude coordinates.

Venues in those neighbourhoods particularly about restaurants.

Source

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

Wikipedia

Foursquare

**METHODOLOGY**

Data consisting of a list of all the boroughs and neighbors will be read directly from Wikipedia using pandas.

Then these two dataframes were combined and data for the borough of North York will be separated into a new dataframe after which the dataframe will be further modified to contain only the data for neighborhoods along with their latitudes and longitudes.

Going ahead we get the data for the venues in under 2 km radius of the neighborhoods.

After acquiring the venues the data is further sorted and cleaned to attain the data just for the Japanese restaurants based on the customer ratings ,the quality and the daily number of customers in that particular area.

Moving ahead we apply machine learning – clustering to get a clear visualization on a map later on to make our learnings more presentable and hence providing the answer to the above question.

**RESULT**

As seen by the cluster map we have divided the borough of North York into 3 clusters based on the amount of customers,the quality of restaurants and customer reviews.The areas in green are the most suitable place as these places have the most customers per day,a good quality food and good customer ratings.

**DISCUSSION**

I noted that areas where there are a lot of people aren’t necessarily the most profitable places.Places districts with a lot of shops or a shopping mall have a significantly higher customer base and profit.

**CONCLUSION**

So the recommended places are marked in green on the cluster map like the Western Heights