**Data Science Capstone Project**

**Recommendation of Neighbourhood for New Business**

**Introduction/Business Problem: Recommendation of Neighbourhood for New Business.**

An Indian business firm has many restaurants, gyms, private schools around the globe. They are now in growing state and want to expand their business. They want to open Indian Fast food and snack Centre and Fitness Club which will have most of the indoor activities in Toronto city, Canada.

However, the main hindrance is they are unable to decide in which Area/Borough/Neighbourhood they should open the restaurant and where to open the Fitness club. As the location is going to be the key factor to the success of both the types of business so considering these factors are crucial.

This problem can be solved by using Data Science and Machine Learning Methodology. Though we need some data of city to analyse and decide which place is better for which activities. Hence by using Data science and ML, we have to answer to the following questions to contractor or Business Firm. Where would you recommend to open restaurant? Similarly, where would you recommend that they setup their Fitness Club?

**Description of the Data and how it will be used to solve the problem:**

To solve this problem first we need to get the data of Toronto city. First let’s collect the list of Borough, Neighbourhoods that belongs to each Borough and longitude and latitude for each Neighbourhood. Clean the unwanted data, delete missing data lines so that it would not mislead our result.

Along with above data we need to get the nearby spots, venues for each Neighbourhood of Toronto city. We will implement this by Foursquare API. Foursquare API is best API to explore the Nearby venues of each Neighbourhood. We will implement the same and will get the Top 5 most common or favourite place by using Foursquare API's.

Once we get the most 5 common places around each Neighbourhood we will apply the ML algorithm and analyse the data. This analysis will lead us to information that which place is good for which or what kind of activities.

1. Analysis for Indian Fats food and snack centre:

We will find out Indian restaurant which are just selling Indian meals and not the Snacks and fast food. If we choose this location where Indian restaurants are present we will get Indian customers very easily and also, they can try Indian fast food and snacks which are not being sold by already present Indian restaurants.

2. Analysis for Fitness Club:

We will find out the places which are full of activities and happening area and where most of Gym and fitness centres are not present (Exclude Neighbourhoods where Gym and fitness centres are present) as competition will less and new customers who are interested in Indoor activities can be attracted. Then find the places where outdoor activities or park is nearby as these areas or spot would be great for Indoor sports too. So If he open the Fitness Club which will have all the Indoor activities like Gym, Zumba dance, Swimming Pool, Table tennis, Badminton court and Tennis Court etc will make good business at such location.

**Methodology:**

Based on business problem we should follow the Unsupervised method of machine learning algorithm to train this model. In Unsupervised method of machine learning clustering is best suitable for given problem statement.

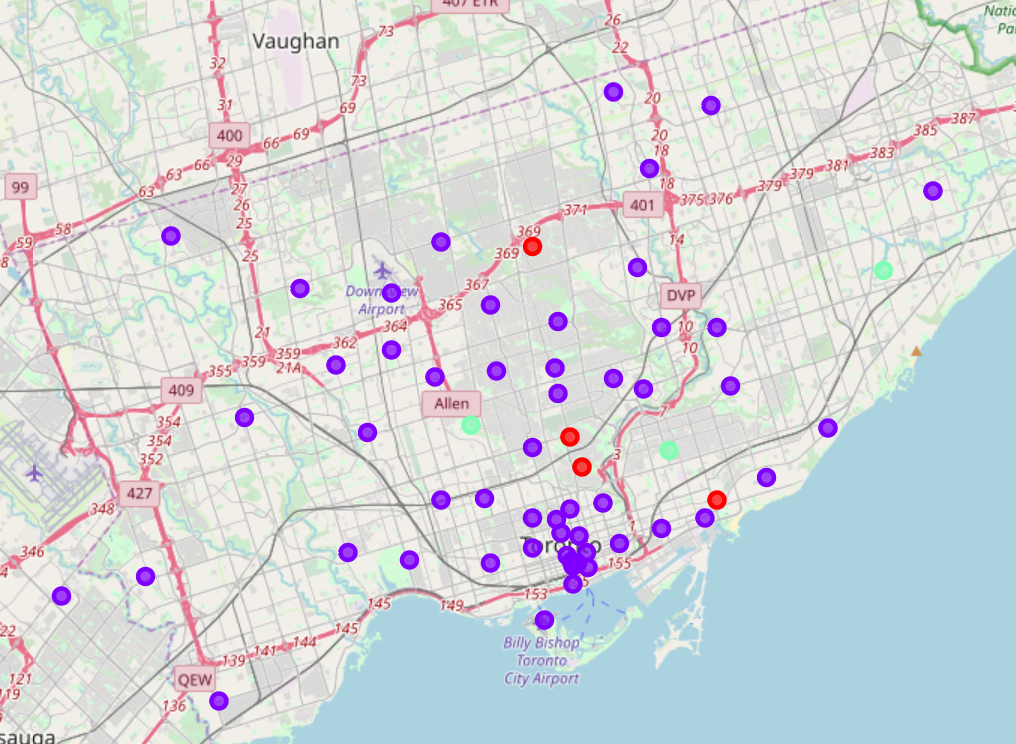
The *K*-means clustering algorithm is used to find groups which have not been explicitly labelled in the data. This can be used to confirm business assumptions about what types of groups exist or to identify unknown groups in complex data sets. Once the algorithm has been run and the groups are defined, any new data can be easily assigned to the correct group.

After applying Kmeans algorithm on collected dataset we have got the three major clusters plotted on Toronto map as below.

Cluster 1 : There are many restaurants, Cafe's and eateries in this region

Cluster 2 : This is famous for Women Health & Beauty Service and Women's Store

Cluster 3 : There are parks and Dessert Shop and Farmer's Market



This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome.

**Evaluation:**

We have tried the best possible iteration of centroid update and best possible number of clusters by deciding the elbow point k=3.

**Results:**

We have got the 3 major clusters and based on these three clusters we have identified below points:

1. Result for Indian Fats food and snack centre: We have found out Indian restaurant which are just selling Indian meals and not the Snacks and fast food.

2. Result for Fitness Club: We have found out the places which are full of activities and happening area and where most of Gym and fitness centres are not present (By Excluding Neighbourhoods where Gym and fitness centres are present). Also found the places where outdoor activities or park is nearby as these areas or spot would be great for Indoor sports too.

**Discussion:**

1. Analysis for Indian Fats food and snack centre:

If we choose these location where Indian restaurants are present we will get Indian customers very easily and also, they can try Indian fast food and snacks which are not being sold by already present Indian restaurants.

1. Analysis for Fitness Club:

If we open the Fitness Club which will have all the Indoor activities like Gym, Zumba dance, Swimming Pool, Table tennis, Badminton court and Tennis Court etc will make good business at such location.

**Conclusion:**

* **Recommended location for Indian Fast food and snacks canter:**

In Cabbagetown, St. James Town, Thorncliffe Park Neighbourhood there are two Indian Restaurant.Our Contractor wants to open new restaurant for Indian Snacks and Fast Food.

So If he open here he can lure the Customer which are visiting Indian restaurants present at these locations and as these restaurants are not offering/selling any Indian fast food and snacks(They are just selling Indian Meals). So opening new Indian Fast food and snacks center he can make the good start by attracting Indian Customers already vising here. Hence Cabbagetown, St. James Town, Thorncliffe Park Neighbourhood is recommended location for our Contractor.

* **Recommended location for Fitness Club:**

In The Beaches ,East Toronto seems to be the best place for Fitness Club. This area is having park and outdoor activities but not the Indoor activities. Also all the activities are not at one place e.g Dance studio is seperate , plaground is at different spot,park is somewhere else so user or customer has to visit different places if he has to do all the sport for fitness activities.

So If he open the Fitness Club which will have all the Indoor activities like Gym, Zumba dance , Swimming Pool, Table tennis, badminton and tennis Court etc will make good business here.