**Opening a new Grocery store in Singapore**

**Problem Background**

Singapore is the well developed city , which has lots of business environments, IT Businesses and friendly environment that attracts lot of population. However, that also means the market is highly competitive and as well as developed city, the cost of doing business is also high interms of investment. So for this reason, one should be very carefull in establishing any new business in the Toronto should examine the neighorhood places such that they should have high amounts of profit in return.

**Problem Description**

In this city called Toronta, XYZ grocery market is supposed planning to establish in the Singapore. As a startup company, even though well funded, they have to choose the location wisely based on highly populated areas in neighborhoods of Singapore.

**Target Audience**

As a Data Scientist, I should recommend the best location for the XYZ Grocery store. I should be able to examine the statistical analysis based on the machine learning algorithms, to locate and recommend the best place for their first store to be located.

**Success Criteria**

The success criteria for this project is to recommend the best neighborhood choice for the XYZ grocery store based on highly populated area and lack of grocery stores in that particular neighborhood

**Data Description**

The data that is available from the website is not structured and therefore we have to scrap it through an existing website that has all the information that is required to explore and cluster the neighborhoods in Toronto.We have to obtain the coordinates of the neighborhood such that it would be easy to work with the data. The data should be preprocessed and should be in strutured format.

**Data Features**

We will be leveraging the features in a reliable location information provider such as the FourSquare.com to explore the neighborhood venue places. We will explore neighborhood within a radius of 200 and should have following variable to solve this:

Neighborhood

Neighborhood Lat

Neighborhood Long

Venue name

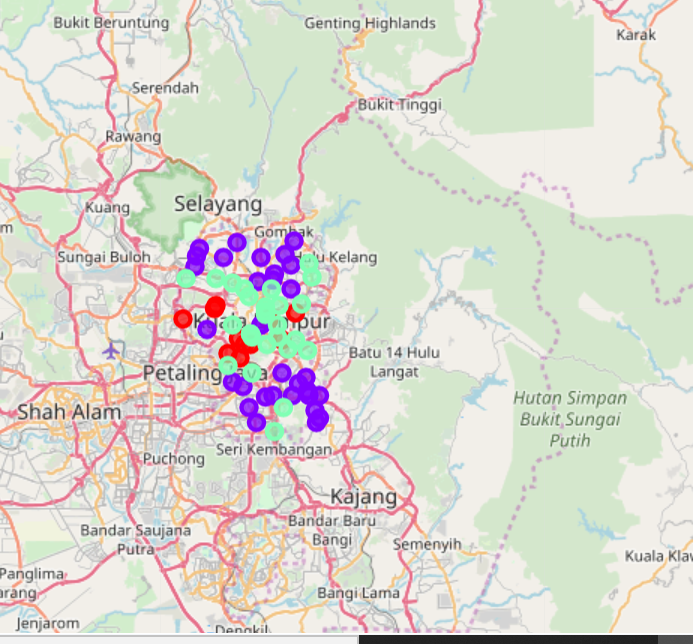
Venue Lat

Venue Long

Venue Category

**Methodolgy**

Foursquare API gets the top 100 venues that are within 200metre radius. once we got the structured data, we used Kmeans clustering for segmentation and to find the best location for the grocery store.



**Results:**

Aftre Kmeans, we got 3 clusters

Cluster 1 : Neighborhoods with moderate number of stores

Cluster 2 : Neighborhoods with low to no existence number of stores

Cluster 3 : Neighborhoods with high number of stores

**Conclusion:**

* Cluster 1 are the most preferred location to open a new grocery store.
* This will provide better understanding to stakeholders to decide.