**Introduction/Business Problem:**

Opening a new restaurant in a new city is always a challenge as it is very important to establish new businesses in the right neighbourhood.

There were many such cases when a restaurant of particular cuisine was opened in an area but not many people visit such restaurants as the taste or cuisine preference of people in that area is very different compared to what is sold in the restaurant. For e.g. a Vietnamese restaurant is opened in an area where the majority of the population is Indian.

It's always good to go with a thorough analysis of new location-based on various parameters like how many other restaurants operate in the area, what is the most sold cuisine in the area, the distance among existing restaurants, etc.

The type of restaurant in any area is also driven by the neighbouring community and preference of the majority of people who live and work in that area.

Considering various such attributes a model is built in which restaurants of similar types are clustered into the same group. The model can be used to make a decision for any new establishment who wants to open a new restaurant in any specific area.

**Data:**

There are various datasets that are used to build the attached model in order to ensure that more accurate and reliable predictions can be made.

Below are a few keys points regarding the data & approach used in the model:

1) Canada postcodes are taken from wiki page, which is then merged with latitude and longitude coordinates for each postcode.

2) The data for different restaurant types are pulled from the ***Foursquare*** website which provides very accurate data and is used by millions of users.

3) The data is cleansed (unwanted features/ attributes) are removed from the dataset.

4 Many different transformations such as One Hot encoding, Standard Scaler, etc. are applied on model to make it more suitable for the given problem statement.

5) The data is then processed through one of the most widely used clustering algorithms like KMeans.

6) The results are visually verified by using Folium library which helps to map longitude and latitude coordinates over map.

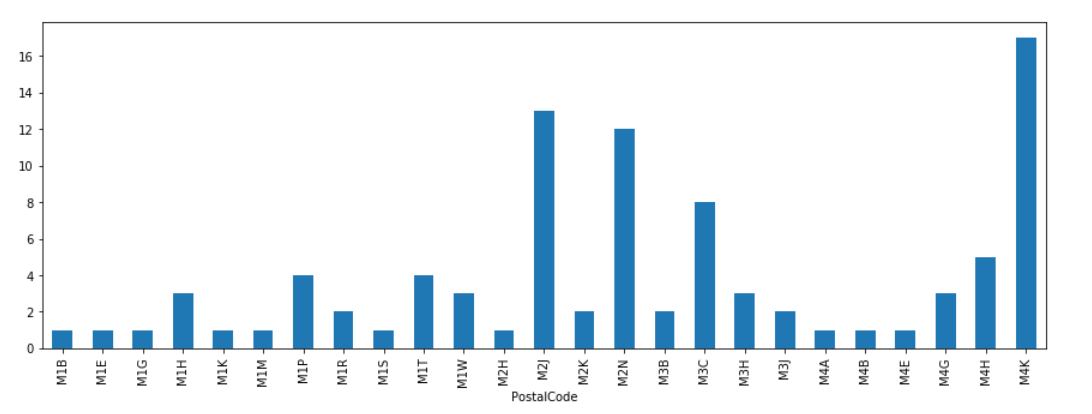
**Methodology section:**

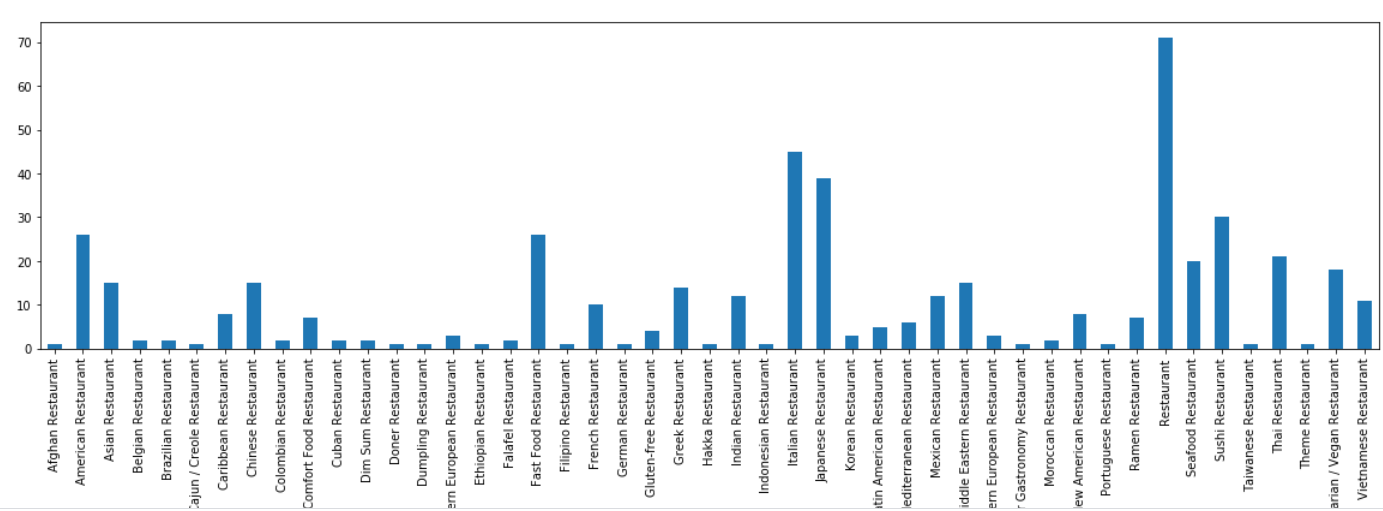
The data for the use case is pulled using various sources like wiki pages and Foursquare website.

The data attributes pulled from Foursquare website is explored and below attributes are chosen for the use case.

* Postcode
* Borough
* Neighborhood
* Latitude
* Longitude
* Venue name
* Venue Latitude,
* Venue Longitude
* Venue Category

The dataset is further refined for the use case by picking data relevant to restaurants only. The restaurant data is analysed across postcode as below to understand how many restaurants operate in each postcode.

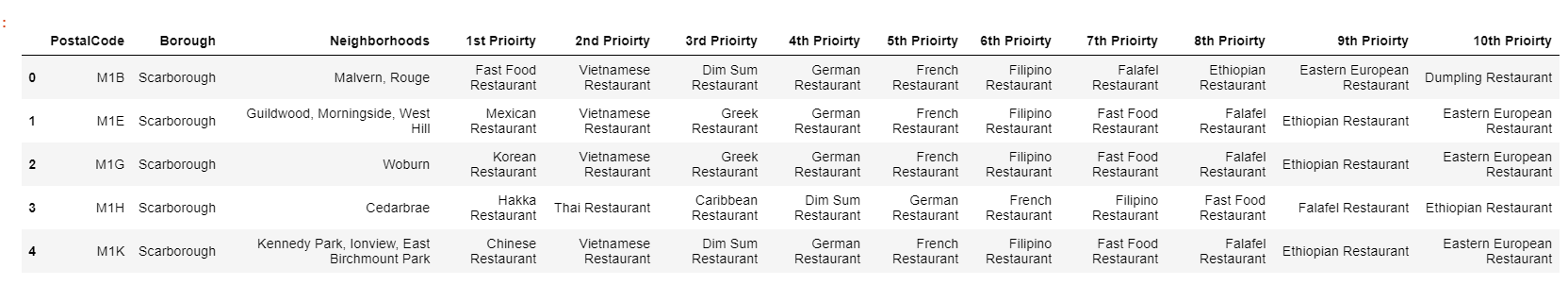


The data is further explored to understand number of restaurant by cuisine type.

Using Onehot encoding the data is transformed from categorical to numeric format so that it can be processed by various clustering algorithms.



The data is further processed to identify top 10 restaurant preference on each post code. Priority 1 is given to restaurant type which has maximum outlets.



KMeans Cluster algorithm is used to identify the best clusters based on above datasets. Below are parameters set for KMeans algo

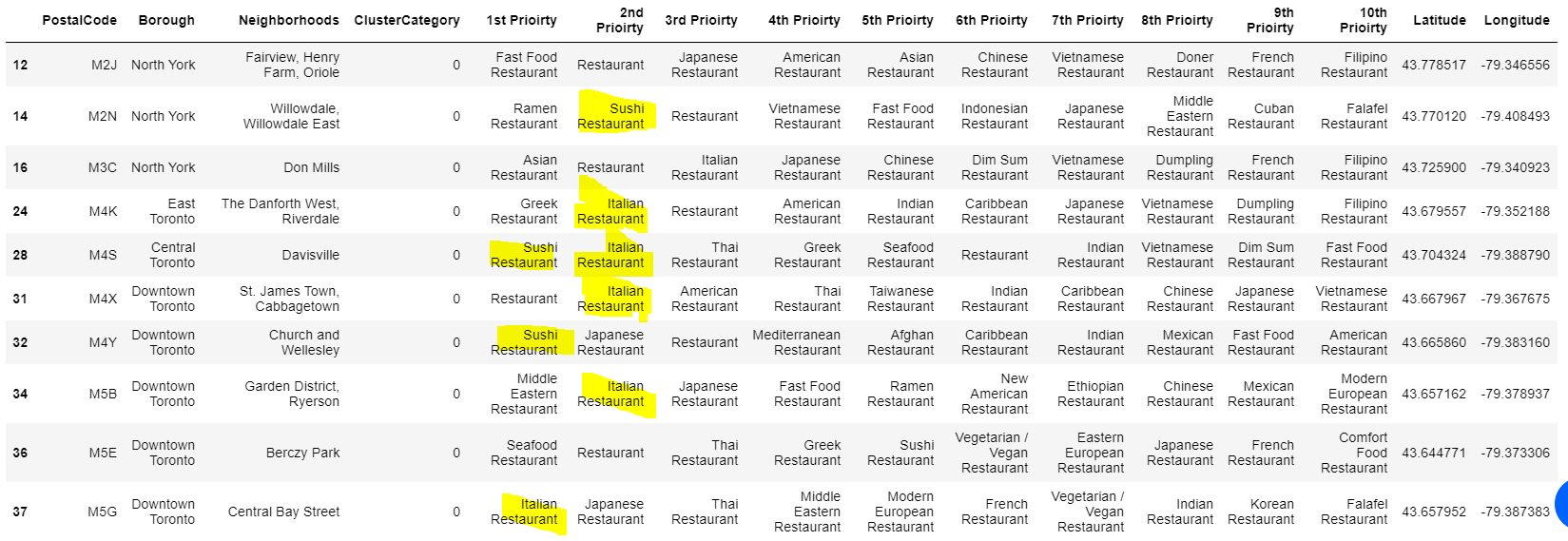
*KMeans(n\_clusters=cluster\_no,init='k-means++', n\_init=10)*

*Where cluster\_no is set to 5*

**Result section:**

The algorithm helps to identify correct cluster type or location to open a new restaurant based on restaurants’ cuisine type.

As per the results, cluster 0 has top priority of Italian & Sushi restaurants.



Similarly, cluster 1 mainly consists of Vietnamese, Indian and Chinese restaurants and cluster 2 consists of American restaurants.

**Discussion section:**

Given it’s an unsupervised learning algorithm where the result can’t be compared against, I believe the algorithm can generate more accurate results if the sample size is more. The given use case is built against a very small sample dataset and I believe for unsupervised learning algorithm to produce better results its important that there should be sufficient volume of data so that it can learn the patterns.

**Conclusion section:**

The solution nicely outputs the priority of restaurant by cuisine type across each postal code which can help any business person who wish to setup a restaurant in the new area.