**Introduction**

The entrepreneur is thinking of opening this restaurant in locations where Asian food is. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur.

**Business Problem**

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Burmese restaurant in Toronto, Canada. By using machine learning methods such as clustering, this project aims to provide solutions to answer the business question: In Toronto, if an entrepreneur wants to open a Burmese restaurant, where should they consider opening it?

I will need below data:

● List of neighborhoods in Toronto, Canada.

● Latitude and Longitude of these neighborhoods.

● Venue data related to Asian restaurants. This will help us find the neighborhoods that are most suitable to open a Burmese restaurant. Extracting the data

● Scrapping of Toronto neighborhoods via Wikipedia

● Using Foursquare API to get venue data related to these neighborhoods.

**Methodology**

First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia page (“https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M”) I did the web scraping by utilizing pandas html table scraping method as it is easier and more convenient to pull tabular data directly from a web page into dataframe.

However, it is only a list of neighborhood names and postal codes. I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later. Here, I made a justification to specifically look for “Thai restaurants”.

Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Thai food”. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

**The results** from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Thai restaurants are in each neighborhood:

● Cluster 0:

Neighborhoods with little or no Thai restaurants

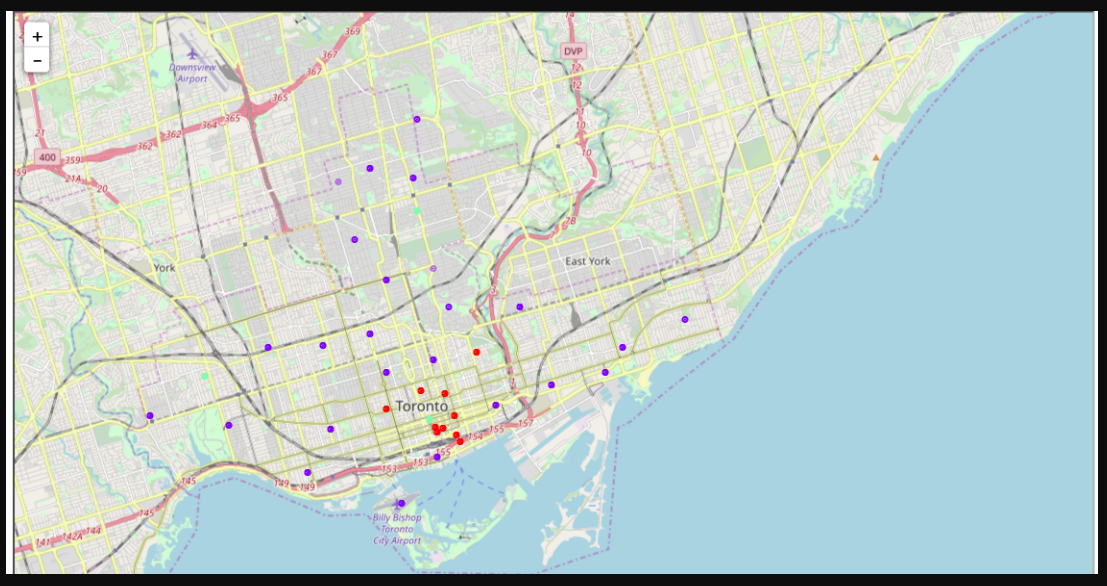
● Cluster 1:

Neighborhoods with no Thai restaurants

● Cluster 2:

Recommendations Most of Thai restaurants are in Cluster 2 which is around Adelaide, King, Richmond areas and lowest (close to zero) in Cluster 1 areas which are North Toronto West and Parkdale areas. Also, there are good opportunities to open near Chinatown, St James town as the competition seems to be low. Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Asian restaurants in these areas.

Therefore, this project recommends the entrepreneur to open an authentic Burmese restaurant in these locations with little to no competition, I only take into consideration of one factor: the occurrence / existence of Thai restaurants in each neighborhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new restaurant.



**Conclusion** In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

**References** List of neighborhoods in Toronto: https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M Foursquare Developer Documentation: <https://developer.foursquare.com/docs>