FINAL REPORT AND CONCLUSION

OPENING A NEW INDIAN RESTAURANT IN TORONTO

INTRODUCTION:-

- The purpose of this project is to help my client. Here my client aims to open a indian restaurent in Torento. AS indian restaurent is a part of asian restaurents so it is very important decision to open at best location. So, I am here to design a project at which place is better for opening restaurent in Torento. So I help my client to find the best suitable location by this project.

- Business Problem:-

- The objective of this capstone project is to find the most suitable location for the client to open a new indian restaurant in Toronto, Canada. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to answer the business question:

- In Toronto, if my client wants to open a indian restaurant, where he consider opening it?

# METHODOLOGY :

- 1.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 ")

- 2.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.

- 3.To get the coordinates,I used the csv file provided by IBM team to match the coordinates of Toronto neighborhoods.

- 4.After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

- 5.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.

- 6.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.

- 7.Here, I made a justification to specifically look for "Thai restaurants". Previously, when I ran the model, I was looking for "Asian restaurants" but there are very few results (maybe due to Foursquare categorization)

- 8.Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, 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.

RESULTS:-
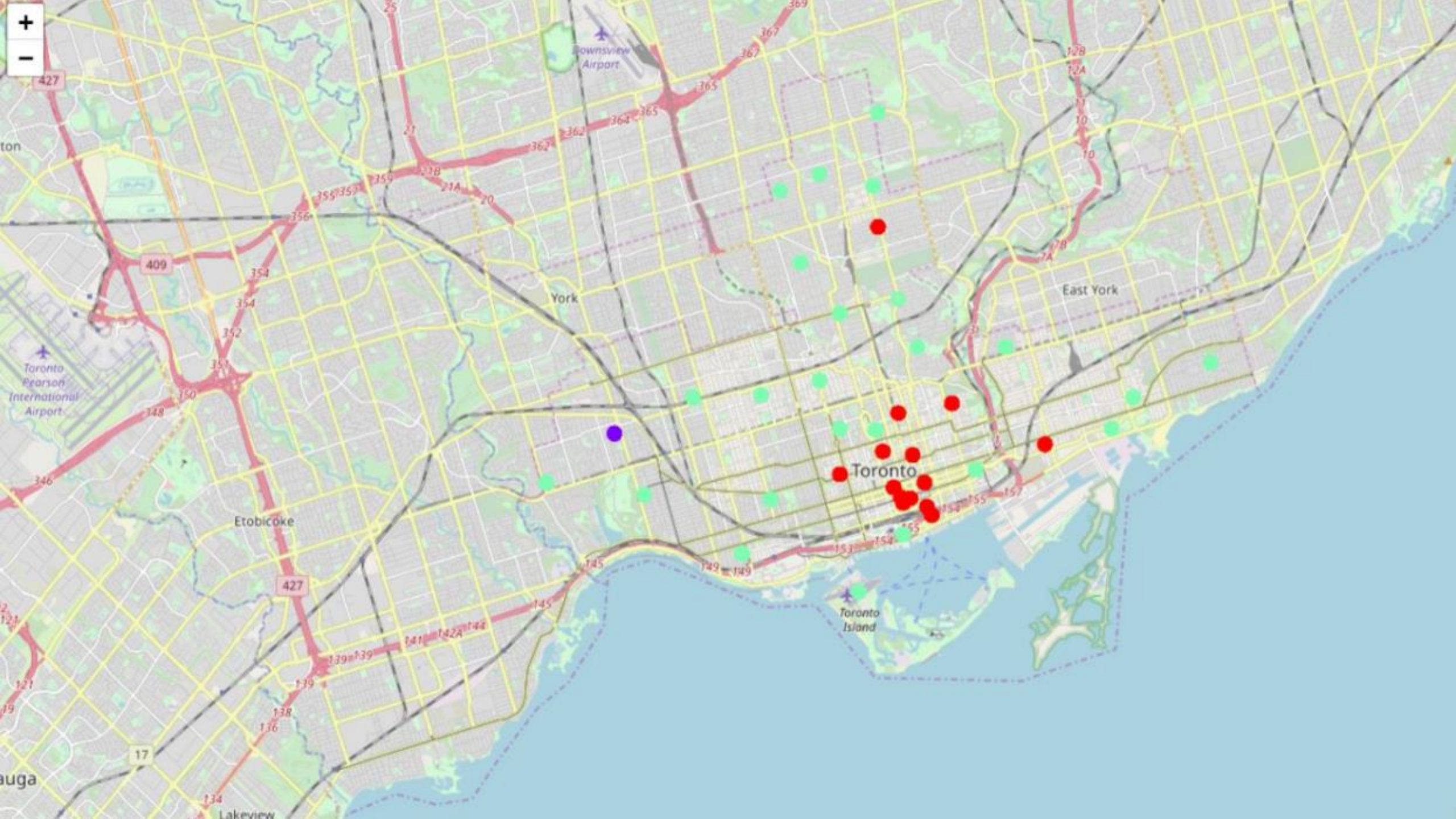
CLUSTERS

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: Neighborhoods with high number of Thai restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color and Cluster 2 in light green color.

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.