**Battle of Neighborhood**

**Opening a new Indian Restaurant in Toronto**

Sarika Sastha Karunan

**1.Introduction**

* 1. Background:

As of 2020, there are one million Indian Citizens in Canada on permanent or temporary status. According to Forbes, the number of Indians who became permanent residents in Canada increased from 39,340 in 2016 to 80,685 in 2019.Toronto has the largest Indian population. Almost 51% of the entire Indian Canadian community resides in the Greater Toronto Area. Since the Indian population is increasing in Toronto year by year, opening a new Indian restaurant in Toronto is a good business idea for anyone who wants to run their own business.

* 1. Problem:

We can use the Foursquare API to analyse location data of Toronto. This project aims to find the best place in Toronto to open an Indian Restaurant with minimum competition with other Indian Restaurants.

* 1. Interest:

This project is beneficial for anyone who wants to open a restaurant in Toronto.

It can be either a person who wants to start a new business, or a hotel group who wants to open their new restaurant outlet in Toronto.

**2.Data Acquisition and Cleaning**

2.1 Data Sources

The data, including postal codes, Boroughs and Neighborhoods scrapped from the Wikipedia page: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

The latitude and longitude values of different Neighborhoods of Toronto was available from: <https://cocl.us/Geospatial_data>

The location data was accessed from Foursquare location data.

2.2 Data Cleaning

* The Toronto data including Postal code, Borough and Neighborhood was scrapped from Wikipedia and changed into a pandas Data Frame.
* As part of Data Cleaning, first removed all rows in which Borough= “Not Assigned”.
* More than one Neighborhood can exist in one Postal Code. Combined these rows into one row.
* If Neighborhood= “Not Assigned”, then, value of corresponding Borough is given to Neighborhood.
* Downloaded the Geospatial data of Toronto and merged this data to the existing Data Frame.
* Our final dataframe is shown below:

Table

Description automatically generated

The final dataframe contains 103 rows and 5 columns.

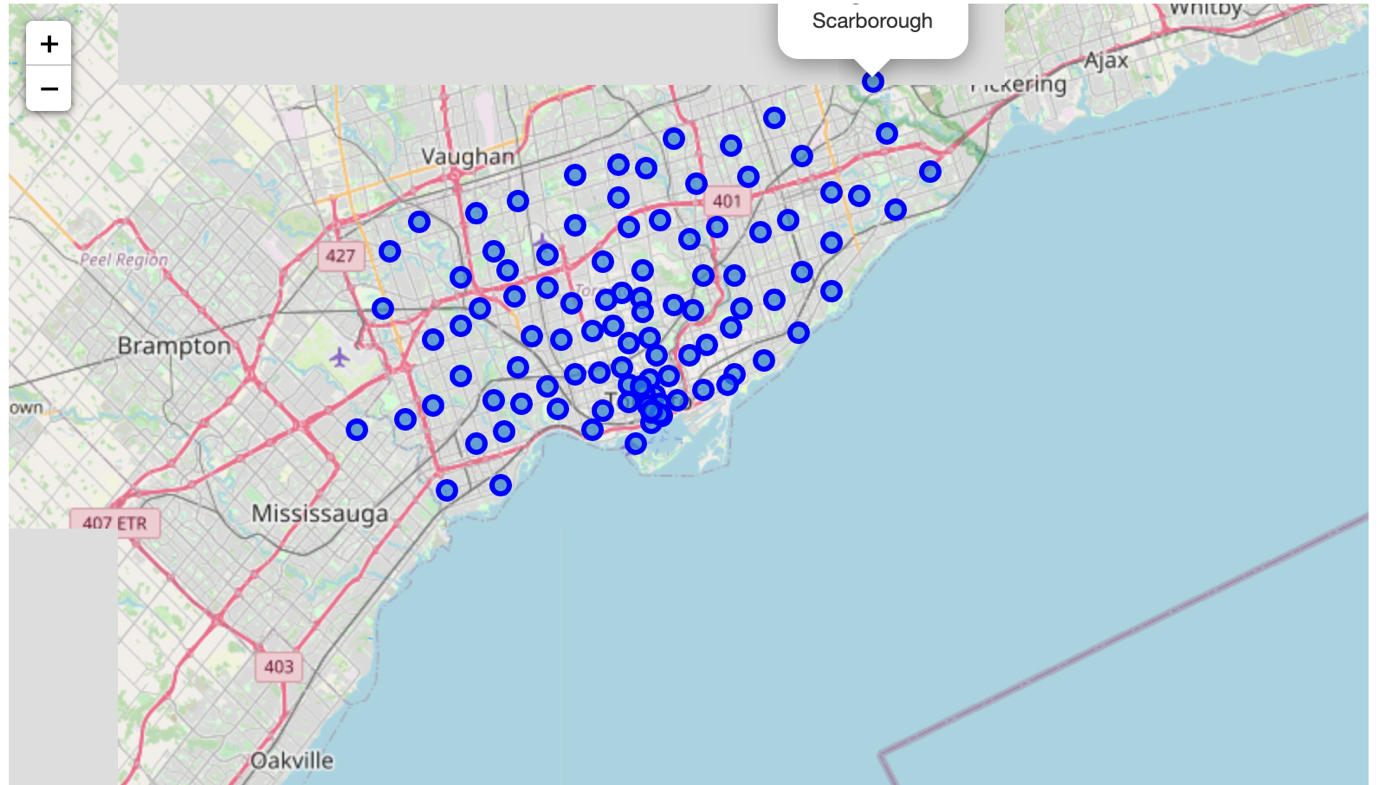
2.3 Location Analysis

Accessed the location data of Toronto including nearby venues using Foursquare API. Collected the information regarding Indian restaurants in different Neighborhoods.

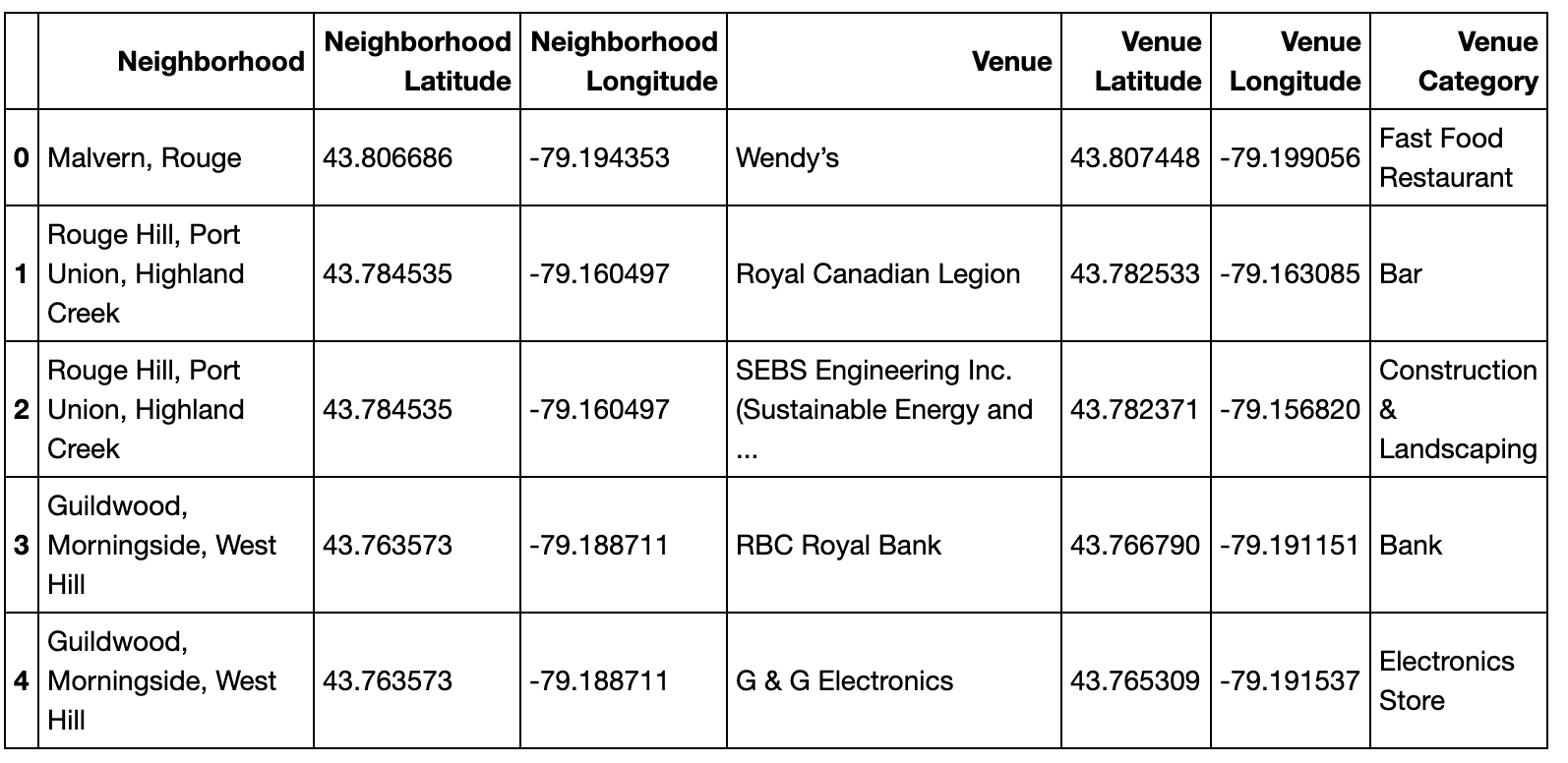
**3.Methodology**

3.1 Data Exploration

* The next step is Data Exploration. As part of this, created the map of Toronto using folium. Different neighborhoods are marked as blue circles on the map.



* Next, I used Foursquare API to get a list of venues in Toronto. The resulting dataframe (toronto\_venues) is consists of 2139 venues.



* From the above list, I extracted the Indian restaurants. There are only 13 Indian restaurants in Toronto.

Table

Description automatically generated

Map

Description automatically generated

From the above map, we can see that all the Indian restaurants are located in a specific area.

3.2 Machine Learning

Before creating machine learning model, we need to perform **one hot encoding.** It allows the representation of categorical data to be more expressive. Many machine learning models cannot work with categorical data directly. The categories must be converted into numbers. That’s why we are performing one hot encoding.

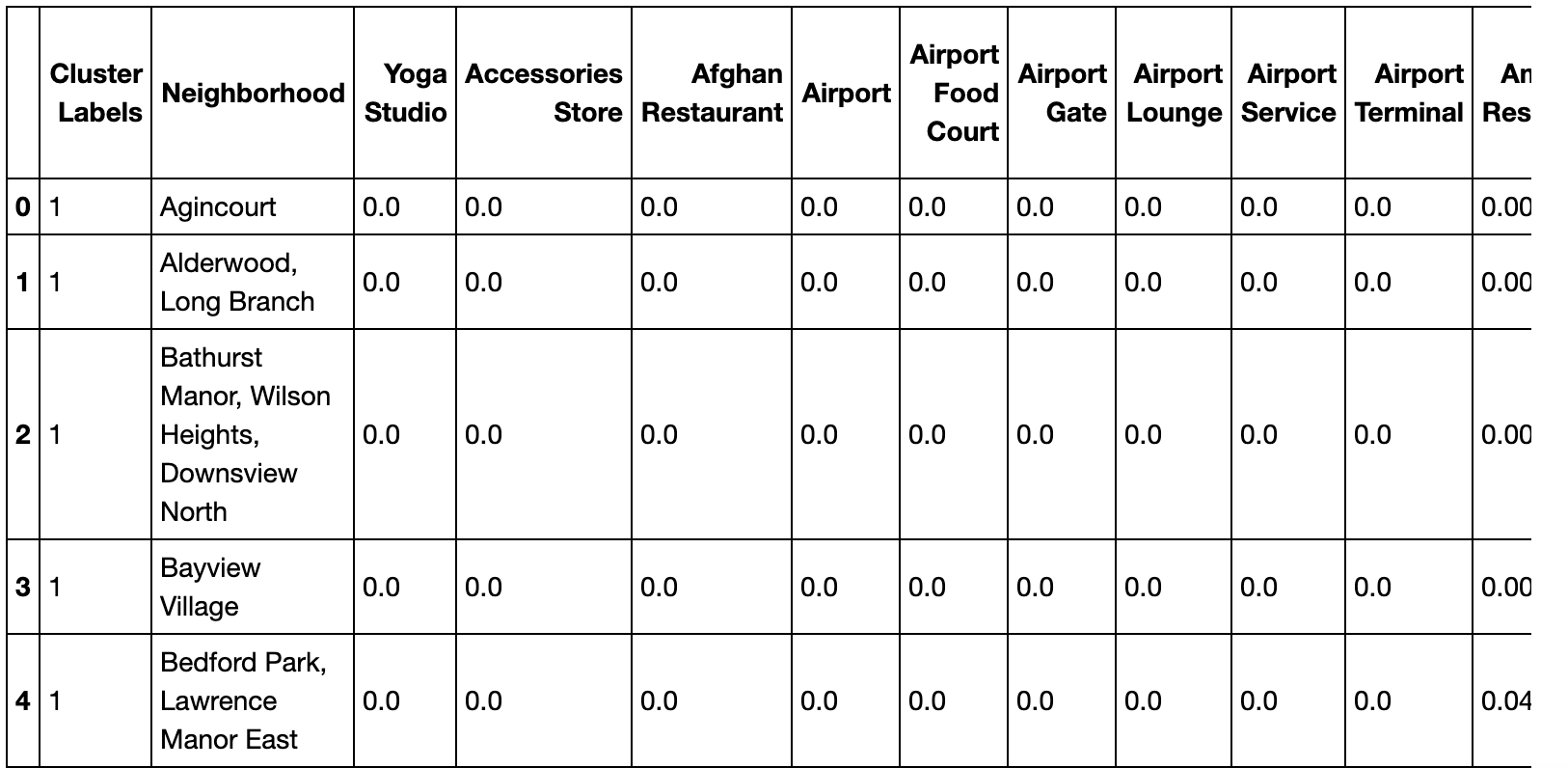
After performing one hot encoding, I grouped the rows by ‘Neighborhood’ and by taking the average of the frequency of occurrence of each venue category.

A picture containing table

Description automatically generated

**K-Means Clustering**

It divides data into non-overlapping subsets without any cluster internal structure. In this project, I am using k-means clustering algorithm to divide Toronto neighborhoods into different clusters. I am selecting k as 5. The resulting dataset with cluster labels is shown below:



From the above dataframe, I am creating a new dataframe (Indian\_restaurants)

Table

Description automatically generated

3.4 Data Analysis

We have 5 clusters. Now, we are going to analyze these clusters.

First, we can visualize the number of neighborhoods in each cluster.

Chart, waterfall chart

Description automatically generated

From the above graph, we can see that the maximum number of neighborhoods are in cluster 2 (cluster label=1). Cluster 1 contains 9 neighborhoods. Other clusters having only one neighborhood in each cluster.

Now, let’s visualize the Indian restaurant in each cluster.

Chart, waterfall chart

Description automatically generated

From the above graph, we can see that all the Indian restaurants are located in cluster 2 only.

Let’s see the neighborhoods in other clusters

Cluster1

Table

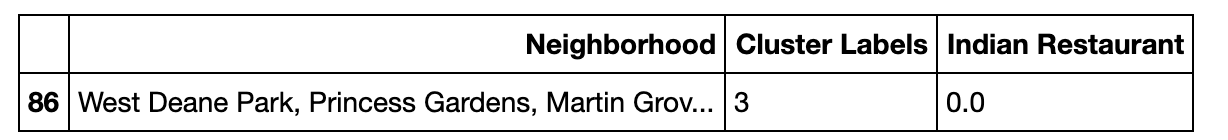
Description automatically generated

Cluster3

Table

Description automatically generated

Cluster4



Cluster5

A picture containing text, table

Description automatically generated

**4. Results**

From the data analysis part, we can clearly see that all the Indian restaurants are located in cluster 2. Also, Cluster2 has the maximum number of neighborhoods (more than 90). Other clusters do not have any Indian restaurants. Among that cluster 1 has 9 neighborhood and other three clusters have one neighborhood each. So, we can say that for opening a restaurant with minimum competition, we can choose a location from these four clusters.

**5. Discussion**

In the beginning of this project, we mentioned that there are around one million Indian citizens in Canada. But we could see that the number of Indian restaurants in Canada as compared to the Indian population is very less. We could see that all the Indian restaurants are located in a specific area. Cluster 2 contains all the Indian restaurants. However, the percentage of Indian restaurants in cluster 2 is around 0.009 as it contains maximum number of neighborhoods. Clusters 1, 3, 4 and 5 are having no Indian restaurants. After cluster 2, cluster 1 contains a greater number of neighborhoods. Based on our analysis, we can suggest starting an Indian restaurant in cluster1 as it’s the second largest cluster. Also, we can consider cluster 3,4 and 5 too as it contains no Indian restaurants.

**6. Conclusion**

In this study, I analyzed the location data of Toronto using Foursquare API and various Python packages. As a result, I identified that there are very few Indian restaurants in Toronto and are located in a specific area. As Toronto has large Indian population and is going to increase in coming years as well, starting a new Indian restaurant is a good business idea for those who wants to start a business in Toronto. In this project, I could find the areas with minimum number of Indian restaurants so that we can avoid competition.