Where should I build my Indian or Chinese Restaurant?

Introduction

One of the things that Indians miss when they are away from home is their food and the sole reason for that is lack abundant options of Indian restaurants around them when they are abroad. And the basis of this study is to help a small group of Indians planning to open their first restaurant in Toronto. Being that Toronto is the most populated city in Canada, and continually ranks as an important global city based on a high quality of living, the choice to expand into the neighbour of the north market was an easy selection for the investing group. However, with limited knowledge of the Toronto market, the group of investors have selected us to assist in the selection of which areas of Toronto will facilitate a launch of their first restaurant.

They are interested in building in an area that meets the following criteria:

- *the habitants would like to eat in that kind of cuisines (Chinese, Indian, Indo-Chinese Fusion)*
- *the neighbourhood must be filled with a particular ethnicity*
- *they would also like to know possible location-wise opportunities if they decide to launch an Indo-Chinese Fusion restaurant.*

Business Problem Statement

1. What is the best location/ethnicity for an Indian/Chinese restaurant in Toronto?
2. In what Neighbourhood should I open an Indian/ Chinese restaurant to have the best chance of being successful?
3. What is the best location if I want to launch an Indo/Chinese restaurant?

Target Audience

Small groups of Indians or Chinese planning to open their first restaurant in Toronto

Data Collection

For this project we need the following data:

1. Toronto data that contains Borough, Neighbourhoods along with there latitudes and longitudes

* Data Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

* Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of Toronto.

2. Indian restaurants in neighbourhood of Toronto.

* Data Source: Foursquare API

* Description: By using this API we will get all the venues in a neighbourhood. We can filter these venues to get only Indian restaurants.

Methodology

Collect the Toronto data

from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

* Using Foursquare API we will get all venues for each neighbourhood.

* Filter out all venues which are Indian Restaurants.

* Data Visualization and some statistical analysis.

* Analysing using Clustering (Specially K-Means):

1. Find the best value of K

2. Visualize the neighbourhood with number of Indian, Chinese, and Indo-Chinese restaurants:

* Compare the Neighbourhoods to find the Best Place for Starting up a Restaurant

* Inference From these Results and related Conclusions

1. Exploratory Data Analysis

Get Postal code, Borough, Neighbourhood etc from

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M:

```
In [31]: # Ignore SSL certificate errors
ctx = ssl.create_default_context()
ctx.check_hostname = False
ctx.verify_mode = ssl.CERT_NONE
source = requests.get('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M').text
soup = BeautifulSoup(source, 'lxml')
table = soup.find('table',{'class':'wikitable sortable'})
table_rows = table.find_all('tr')
data = []
for row in table_rows:
    data.append([t.text.strip() for t in row.find_all('td')])

df = pd.DataFrame(data, columns=['PostalCode', 'Borough', 'Neighbourhood'])
df = df[~df['PostalCode'].isnull()]  # to filter out bad rows
```

```
In [4]: df.head(10)
```

Out[4]:

	PostalCode	Borough	Neighbourhood
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village
5	M5A	Downtown Toronto	Regent Park, Harbourfront
6	M6A	North York	Lawrence Manor, Lawrence Heights

Assign latitudes and longitudes as follows:

```
In [8]: def get_geocode(postal_code):
    # initialize your variable to None
    lat_lng_coords = None
    while(lat_lng_coords is None):
        g = geocoder.google('{}, Toronto, Ontario'.format(postal_code))
        lat_lng_coords = g.latlng
    latitude = lat_lng_coords[0]
    longitude = lat_lng_coords[1]
    return latitude,longitude
```

```
In [9]: geo_df=pd.read_csv('http://cocl.us/Geospatial_data')
```

```
In [10]: geo_df.head()
```

Out[10]:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

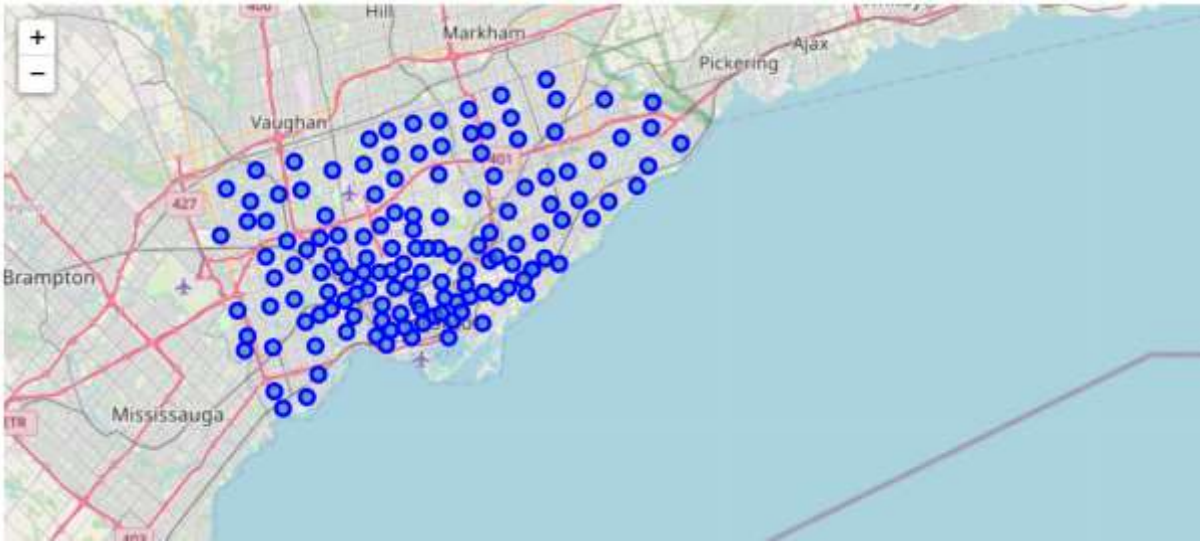
Foursquare data is very comprehensive, and it powers location data for Apple, Uber etc. For this business problem I have used, as a part of the assignment, the Foursquare API to retrieve information about the Venue, Venue category with there longitudes and latitudes. The call returns a JSON file and we need to turn that into a data-frame. Here I've chosen 100 popular spots for each neighbourhoods a radius of 500 meters. Below is the data-frame obtained from the JSON file that was returned by Foursquare:

```
In [19]: toronto_venues.groupby('Neighborhood').count()
```

Out[19]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	55	55	55	55	55	55
Brockton, Parkdale Village, Exhibition Place	23	23	23	23	23	23
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	16	16	16	16	16	16
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16	16	16	16	16	16
Central Bay Street	68	68	68	68	68	68
Christie	16	16	16	16	16	16
Church and Wellesley	75	75	75	75	75	75
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	33	33	33	33	33	33
Davisville North	9	9	9	9	9	9
Dufferin, Dovercourt Village	13	13	13	13	13	13
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West, Forest Hill Road Park	4	4	4	4	4	4
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
High Park, The Junction South	25	25	25	25	25	25
India Bazaar, The Beaches West	19	19	19	19	19	19

The data is plotted on Toronto Map using folium package as follows:



It clearly shows that our dataset covers all of city of Toronto.

Find out the top common ethnic origins by neighbourhood:

Agincourt North

Origin	Count
Chinese	16950.0
Sri Lankan	2230.0
East Indian	2090.0
Filipino	1465.0
Canadian	1295.0

Alderwood

Origin	Count
English	2320.0
Canadian	2245.0
Irish	1900.0
Scottish	1720.0
Italian	1275.0

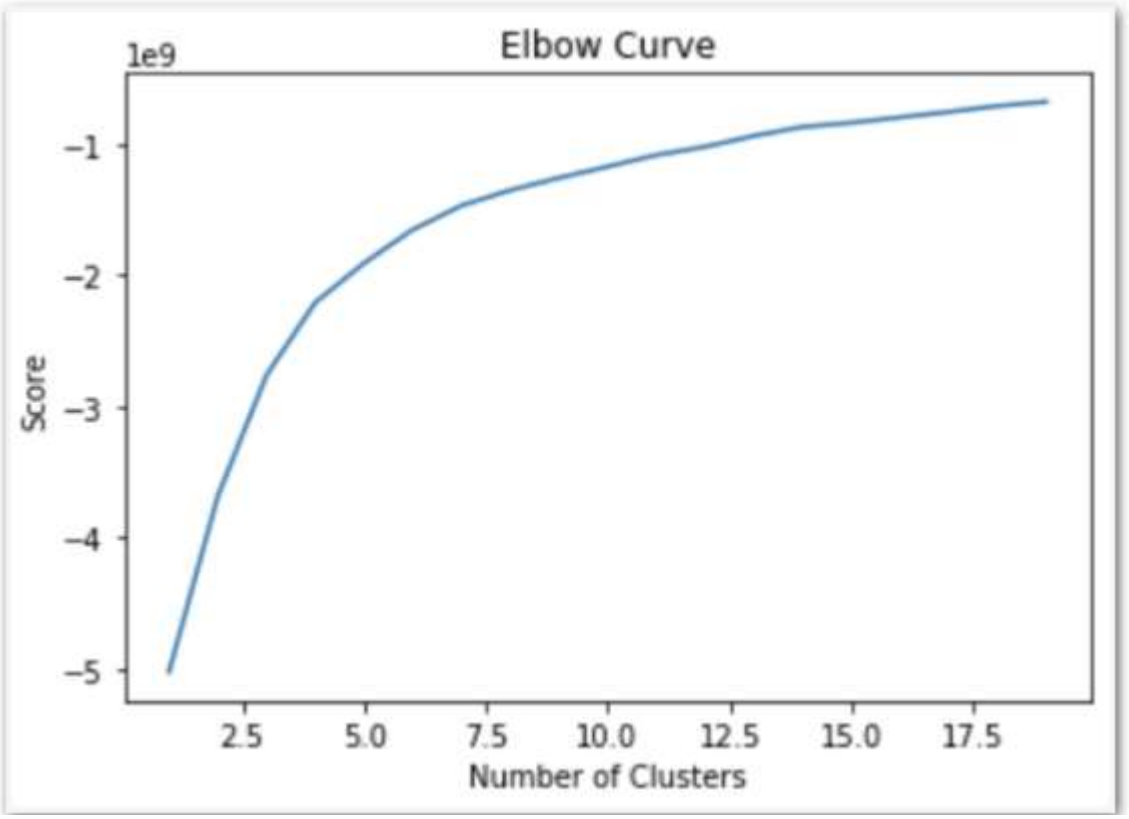
2. **Machine Learning Cluster Algorithm used**

K-Means Clustering

- k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
- It is widely used for unsupervised learning and clustering
- It is simple and less time consuming than other more complex clustering algorithms

Number of Clusters (Elbow Method):

We will extract Indian restaurant data from above table and fit this into the code for finding best value of K:



The elbow line is visible for k = 5 and hence we take k as five:

Pass the data through K-Means algorithms:

```
kclusters = 5

toronto_areas_clustering = df_demographic.drop('City_Area', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(data)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([3, 3, 0, 4, 1, 0, 1, 1, 1, 0])
```

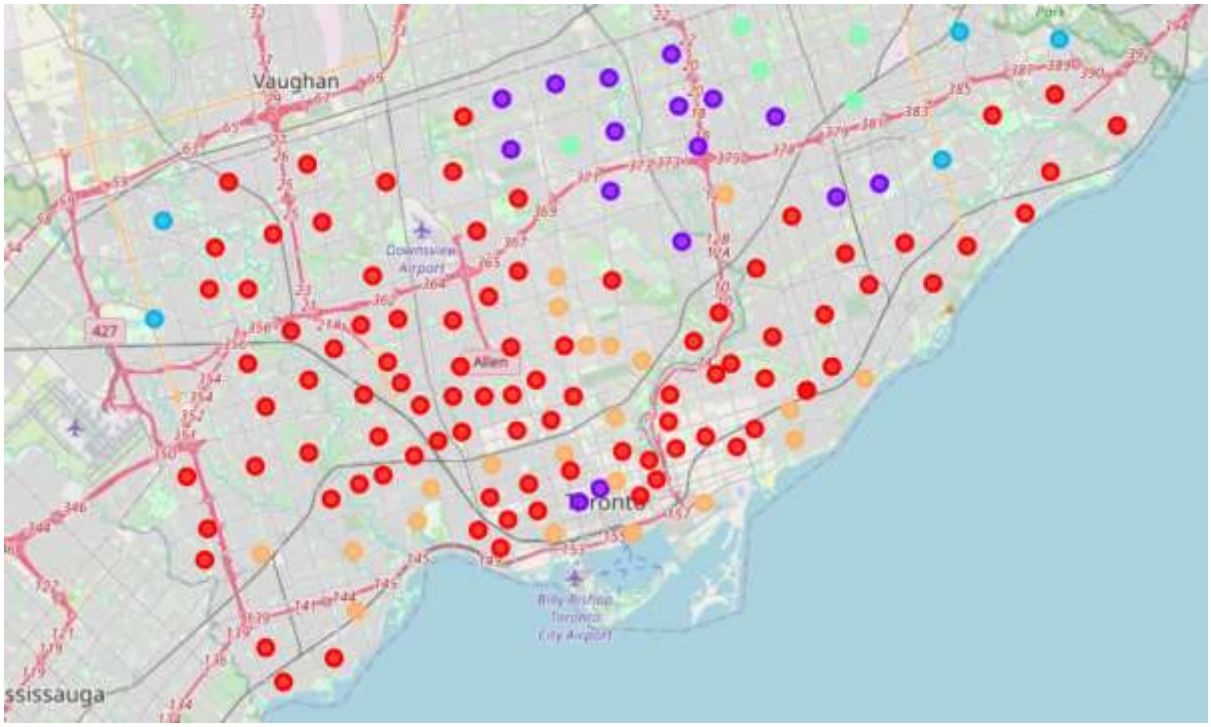
```
# add clustering labels
areas_ethn_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
```

```
toronto_merged = df_cityAreas

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
toronto_merged = toronto_merged.join(areas_ethn_sorted.set_index(['City_Area']), ons['City_Area'])

toronto_merged.head(100) # check the last columns!
```

	CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Most Common Origin	5th Most Common Origin	7th Most Common Origin	8th Most Common Origin
0	129	Agincourt North	43.80930	-79.267070	3	Chinese	Sri Lankan	East Indian	Filipino	Canadian	English	Tamil	Jamaican
1	128	Agincourt South-Malvern West	43.78736	-79.269350	3	Chinese	East Indian	Filipino	Sri Lankan	Canadian	English	Scottish	Jamaican

Results

Here is the clustering output is as follows:



Clusters formed:

Red–European and Canadian

Light blue – Indian

Green – Chinese

Yellow – Irish, Scottish, and English

Purple – Asian

Cluster 0 : European and Canadian people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor O
33	Alderwood	43.6049 60	-79.541160	0	English	Canadia n	Irish	Scottish	Ita
97	Broadview North	43.6893 70	-79.354298	0	English	Irish	Scottish	Greek	Ca n
11	Briggate-Centennial-West Deane	43.6979 10	-79.577390	0	Canadia n	English	Italian	Irish	Sc

Cluster 1 : Asian people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor O
127	Bendale	43.7596 3	-79.25739	1	Chinese	East Indian	Filipino	Canadia n	En
47	Don Valley Village	43.7855 2	-79.35017	1	Chinese	Filipino	East Indian	Iranian	En
126	Dorset Park	43.7553 3	-79.27748	1	Filipino	East Indian	Chinese	Canadia n	Sri Lan

Cluster 2 : Indian people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor O
132	Malvern	43.8097 1	-79.22004	2	East Indian	Sri Lankan	Filipino	Chinese	Ja
2	Mount Olive-Silverstone-Jamestown	43.7473 1	-79.33826	2	East Indian	Kore	Jamaican	Canadia n	Sri
133	Rouge	43.8079 6	-79.17855	2	East Indian	Sri Lankan	Canadia n	Filipino	Ja

Cluster 3 : Chinese people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor O
129	Agincourt North	43.8001 3	-79.26759	3	Chinese	Sri Lankan	East Indian	Filipino	Ca n
128	Agincourt South-Malvern West	43.7973 5	-79.26941	3	Chinese	East Indian	Filipino	Sri Lankan	Ca n
131	L'Amoreaux	43.7972 6	-79.31320	3	Chinese	East Indian	Canadia n	Sri Lankan	Fil

Cluster 4 : Irish, Scottish and English people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor C
35	Annex	43.6681 6	-79.40281	4	English	Irish	Scottish	Canadia n	G
122	Birchcliffe-Cliffside	43.6941 2	-79.26465	4	English	Irish	Canadia n	Scottish	Fr
73	Church-Yonge Corridor	43.6692 4	-79.37868	4	English	Irish	Scottish	Chinese	C n

Discussion

We are more interested in looking into the clusters that comprise of Indians and Chinese as follows:

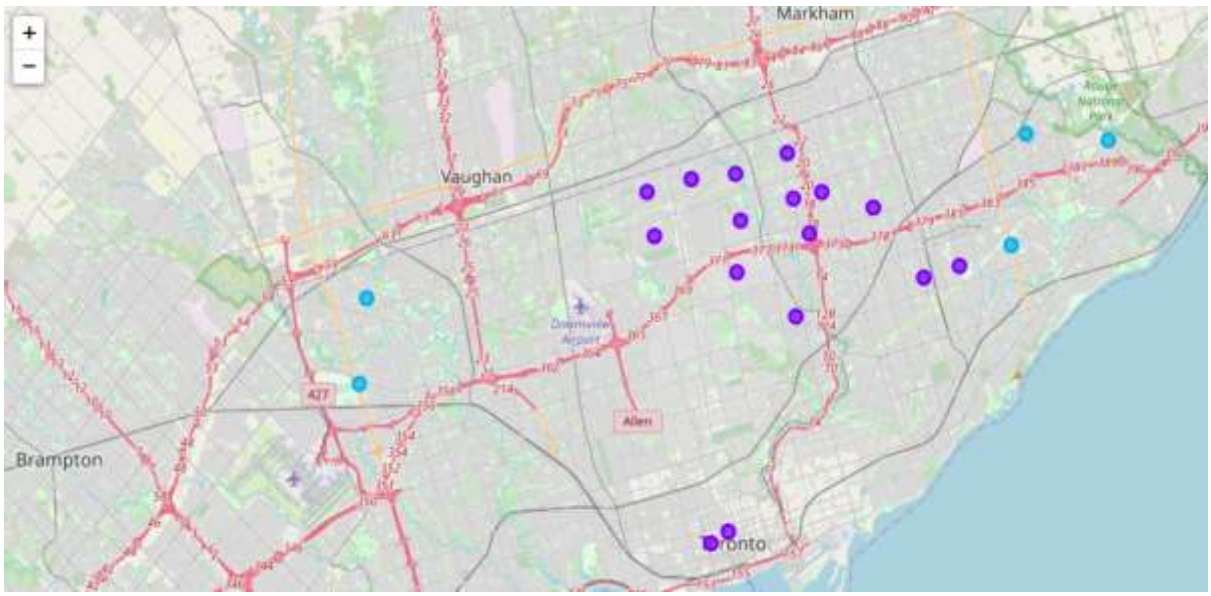
Indian:

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor Or
132	Malvern	43.80977	-79.22084	2	East Indian	Sri Lankan	Filipino	Chinese	Jar
2	Mount Olive-Silverstone-Jamestown	43.74721	-79.58826	2	East Indian	Iraqi	Jamaican	Canadian	So
131	Rouge	43.80766	-79.17405	2	East Indian	Sri Lankan	Canadian	Filipino	Jar

Chinese:

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor Or
129	Agincourt North	43.80930	-79.26707	3	Chinese	Sri Lankan	East Indian	Filipino	Can
128	Agincourt South-Malvern West	43.78735	-79.26941	3	Chinese	East Indian	Filipino	Sri Lankan	Can
117	L'Amoreaux	43.79726	-79.31220	3	Chinese	East Indian	Canadian	Sri Lankan	Fili

Isolating clusters that comprise of only Indians and Chinese as follows:



Conclusion

The following areas are the ideal places for setting up an Indian or Chinese or Indo-Chinese restaurants as these neighbourhoods are frequently visited and are inhabited by a lot of Chinese, Indians and Asians as well who will for sure would love to try out Chinese, Indian or fusion of both the cuisines:

In conclusion, to end off this project, we had an opportunity on a business problem, and it was tackled in a way that it was like how a genuine data scientist would do. We utilized numerous Python libraries to fetch the information, control the content and break down and visualize those datasets. We have utilized Foursquare API to investigate the settings in neighbourhoods of Toronto, get a great measure of data from Wikipedia which we scraped with the Beautifulsoup Web scraping Library. We also visualized utilizing different plots present in seaborn and Matplotlib libraries. Similarly, we applied AI strategy to anticipate the error given the information and utilized Folium to picture it on a map.

Places that have room for improvement or certain drawbacks give us that this project can be additionally improved with the assistance of more information and distinctive Machine Learning strategies. Additionally, we can utilize this venture to investigate any situation, for example, opening an alternate cuisine or opening of a Movie Theater and so forth. Ideally, this task acts as an initial direction to tackle more complex real-life problems using data science.