CAPSTONE PROJECT

ANALYSIS TO FIND BEST LOCATION FOR OPENING A RESTAURANT

CONTENTS

- INTRODUCTION
- BUSINESS PROBLEM
- DATA SECTION
- METHODOLOGY
- RESULTS
- CONCLUSION

INTRODUCTION

TORONTO IS THE CAPITAL CITY OF THE CANADIAN PROVINCE OF ONTARIO. WITH A RECORDED POPULATION OF 2,731,571, IT IS THE MOST POPULOUS CITY IN CANADA AND THE FOURTH MOST POPULOUS CITY IN NORTH AMERICA.

ACCORDING TO THE UNITED NATIONS DEVELOPMENT PROGRAMME, TORONTO HAS THE SECOND-HIGHEST PERCENTAGE OF CONSTANT FOREIGN-BORN POPULATION AMONG WORLD CITIES, AFTER MIAMI, FLORIDA.

IN THIS PROJECT WE WILL TRY TO ANALYSE ALL THE **INDIAN RESTAURANTS** CURRENTLY PRESENT IN TORONTO'S DIFFERENT NEIGHBOURHOOD AND FIND TOP INDIAN RESTAURANT'S BASED ON POPULARITY FROM ITS LIKES, RATING, TIPS ETC AND WILL ABLE TO PROVIDE AN FINAL RESULT/ANALYSIS AT THE END OF THIS PROJECT.

BUSINESS PROBLEM

THE OBJECTIVE IS TO FIND A SUITABLE LOCATION(S) TO OPEN AN INDIAN RESTAURANT IN TORONTO CITY, ONTARIO, CANADA. THIS PROJECT MAKES USE OF VARIOUS DATA SCIENCE AND MACHINE LEARNING METHODOLOGIES (K-MEANS CLUSTERING) TO PROVIDE A SOLUTION TO THE CLIENT.

THE PROJECT AIMS TO PROVIDE A SOLUTION TO THE QUESTION: **‘WHERE SHOULD YOU CONSIDER OPENING AN INDIAN RESTAURANT IN TORONTO CITY?’**

DATA SECTION

- TORONTO NEIGHBOURHOOD DATA THAT CONTAINS BOROUGH, NEIGHBOURHOODS ALONG WITH THEIR LATITUDES AND LONGITUDES

  DATA SOURCE: *GEOSPATIAL CO-ORDINATES CSV FILE*

- TORONTO POSTAL CODE DATA THAT CONTAINS POSTAL CODES FOR NEIGHBOURHOOD PRESENT WITHING TORONTO CITY.

  DATA SOURCE: *HTTPS://EN.WIKIPEDIA.ORG/WIKI/LIST_OF_POSTAL_CODES_OF_CANADA:_M*

- INDIAN RESTAURANTS TORONTO CITY.

  DATA SOURCE: FOURSQUARE API

METHODOLOGY- WEB SCRAPING

WE WILL FIRST USE, THE DATA **BEAUTIFUL SOUP PACKAGE** TO SCRAPE THE POSTAL CODE INFORMATION OF TORONTO NEIGHBOURHOOD FROM THE WIKIPEDIA PAGE, AND STORE THE SAME INFORMATION WITHIN **PANDAS** DATAFRAME.

In [6]:
```
tor_zip_url = requests.get('https://en.wikipedia.
soup = BeautifulSoup(tor_zip_url, 'xml')

table=soup.find('table')
column_names = ['Postal code','Borough','Neighborh
df = pd.DataFrame(columns = column_names)
```

In [7]:
```
for tr_cell in table.find_all('tr'):
    row_data=[]
    for td_cell in tr_cell.find_all('td'):
        row_data.append(td_cell.text.strip())
    if len(row_data)==3:
        df.loc[len(df)] = row_data
```

In [8]:
```
df.head()
```

Out[8]:

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

METHODOLOGY- DATA CLEANSING

DATA WILL BE CLEANED TO REMOVE UNASSIGNED VALUES PRESENT IN EXTRACTED DATAFRAME.

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson

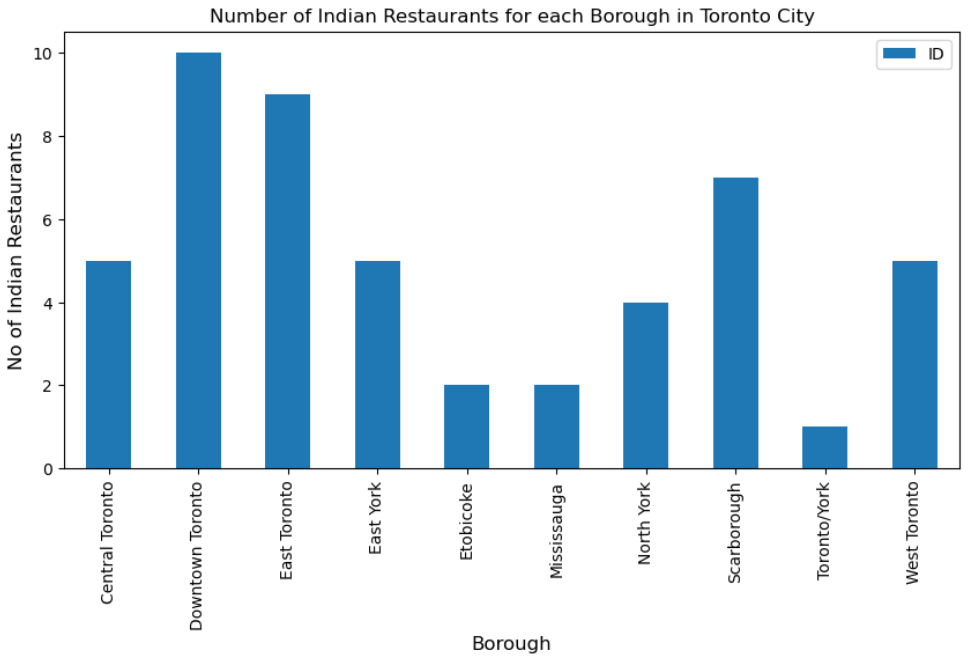
	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

METHODOLOGY- DATA WRANGLING

- IMPORT THE CONTENTS OF GEOSPATIAL CO-ORDINATES CSV FILE INTO THE DATAFRAME. CREATE A SINGLE DATAFRAME FOR FURTHER ANALYSIS

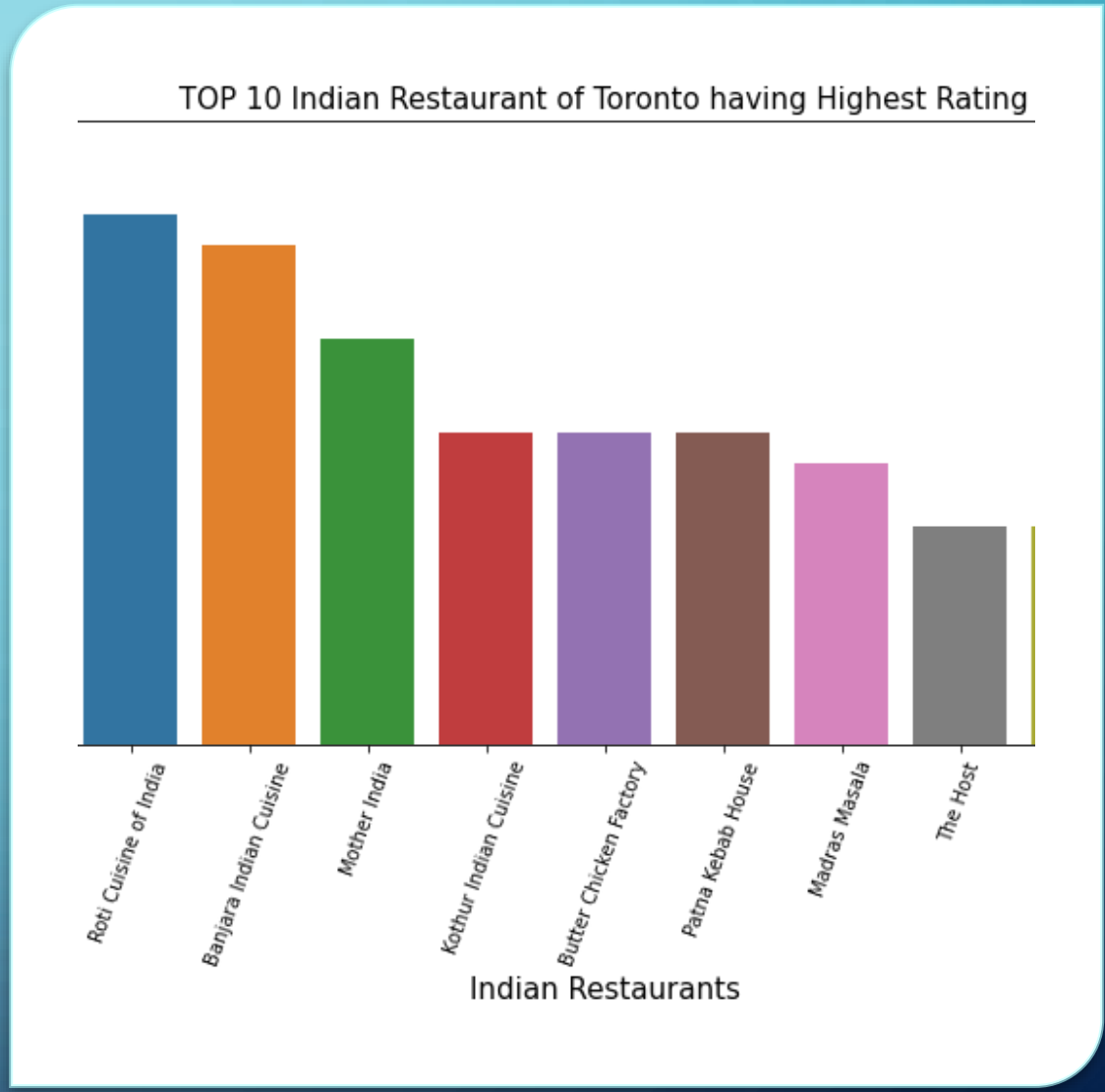
METHODOLOGY- EXPLORATORY ANALYSIS

USE FOURSQUARE API TO FETCH INDIAN RESTAURANTS PRESENT WITHIN DIFFERENT BOROUGHS IN TORONTO



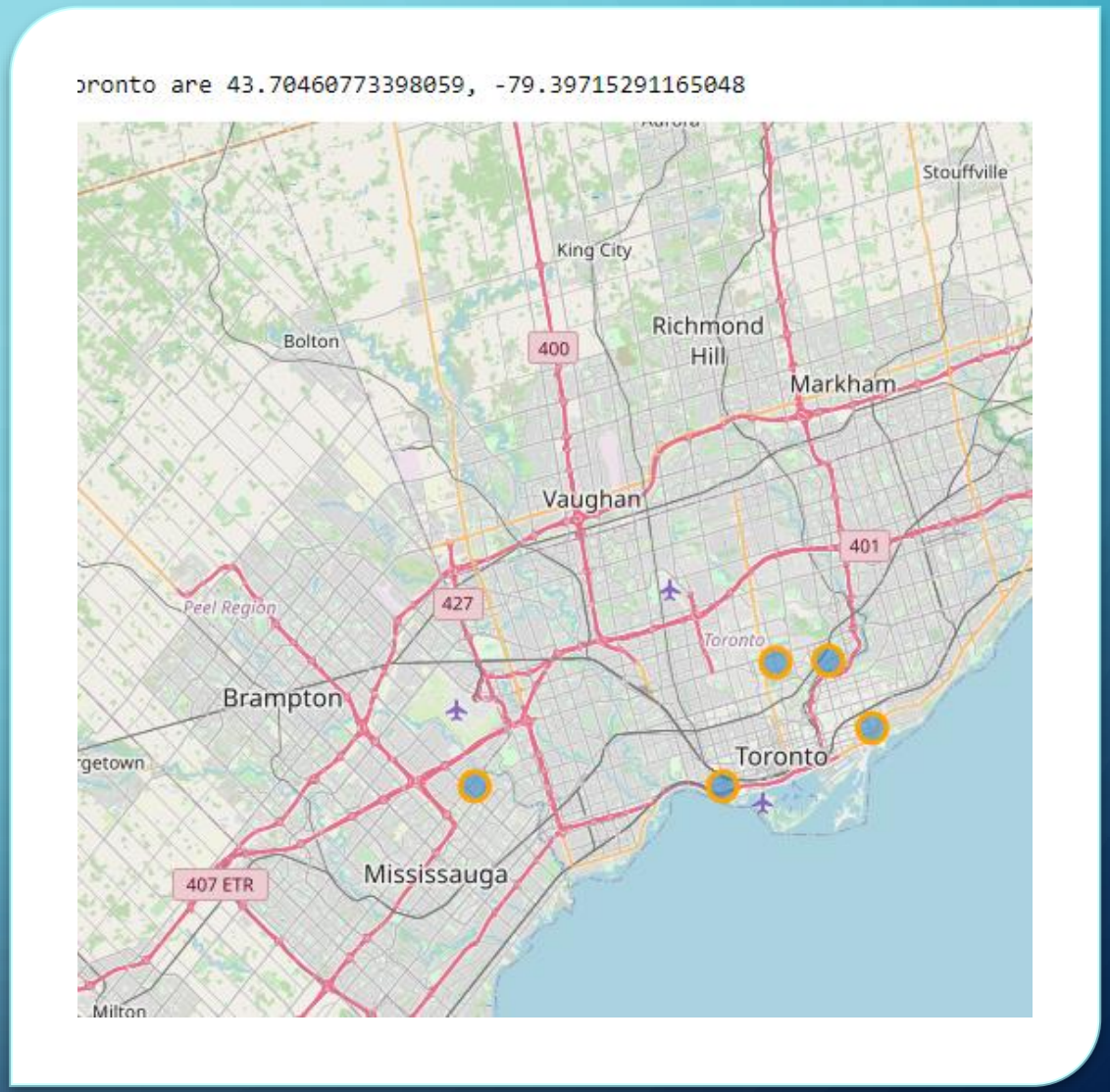
METHODOLOGY-ANALYSIS

USE FOURSQUARE API AND DATA WRANGLING TO FETCH TOP 10 RESTAURANTS BASED ON RATING, TIPS AND LIKES.



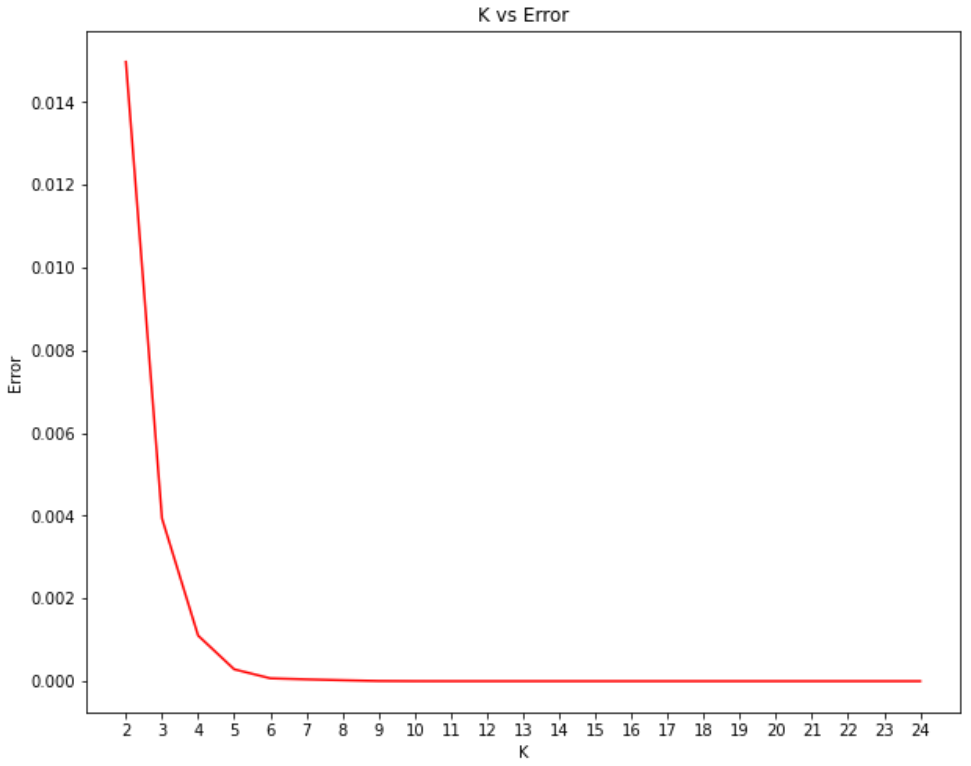
METHODOLOGY –GEOSPATIAL ANALYSIS

VISUALIZE THE DATA OF TOP NEIGHBOURHOOD BASED ON THE VENUE CATEGORIES IN A TORONTO CITY MAP USING *FOLIUM PACKAGE*.



METHODOLOGY – CLUSTERING

CLUSTER THESE TOP NEIGHBOURHOOD BASED ON THE VENUE CATEGORIES AND USE K-MEANS CLUSTERING MACHINE LEARNING TECHNIQUE TO BUILD A MODEL USING ELBOW POINT METHOD.



METHODOLOGY – CLUSTERING

FINDING THE CLUSTER WITH BEST INDIAN RESTAURANTS FOR OUR ANALYSIS.

Cluster 2 has most of the **Indian Restaurant** followed by **Cluster 5** and **Cluster 3**

RESULTS

- ***DOWNTOWN TORONTO*** BOROUGH HAS MAX NO OF INDIAN RESTAURANT IN TORONTO CITY
- ***DOWNTOWN TORONTO*** BOROUGH HAS ***10*** INDIAN RESTAURANT, WHERE AS, ***YORK*** BOROUGH HAS LEAST INDIAN RESTAURANT PRESENT ITS COUNT IS ONLY **1** IN TORONTO CITY RESPECTIVELY.
- ***INDIA BAZAAR, THE BEACHES WEST*** NEIGHBOURHOOD HAS MAXIMUM NO OF INDIAN RESTAURANT WITH A COUNT OF ***6***.
- WE CAN OBSERVE ***ROTI CUISINE OF INDIA*** OF INDIAN RESTAURANT GOT THE MAXIMUM RATING. IT BELONGS TO ***THE ANNEX, NORTH MIDTOWN, YORKVILLE*** NEIGHBOURHOOD AND OF ***CENTRAL TORONTO*** BOROUGH.
- ***CLUSTER 2*** HAS MOST OF THE ***INDIAN RESTAURANT*** FOLLOWED BY ***CLUSTER 5*** AND ***CLUSTER 3***.

CONCLUSION

IN THE END, I REACHED THE GOAL THAT I DECLARED IN THE FIRST SECTION. I THINK WITH THE MAP ON THE RESULTS SECTION A TOURIST CAN SEE A SIMPLE GUIDE ABOUT RESTAURANTS IN DIFFERENT DISTRICTS.

BUT THERE IS SCOPE IN FUTURE TO FETCH MORE DEMOGRAPHIC AND FINANCIAL DATA TO GET MORE CONCRETE ANALYSIS. IN TOTAL, I HOPE YOU ENJOYED THESE RESULTS OF THE CAPSTONE PROJECT.