1. Introduction : A description of the problem and a discussion of the background

Toronto is the capital city of the province of Ontario and the largest city in Canada by population. The economy of Toronto is the largest contributor of the Canadian economy. Many IT companies have their offices in Toronto.

A fictional IT company based in Rosedale , Toronto needs to choose a location and a possible list of hotels for their employees who will be travelling to Bangalore for few weeks to work at their client office in Bangalore, India. Bangalore is the IT capital of India, with an area of 710 km2 covering over 295 localities. The Company would like to choose only from those locations in Bangalore which are similar to Rosedale for their employees accommodation in order to make them feel more comfortable and help them adapt quickly to the new place. In this project we will attempt to help the company find those locations in Bangalore which are similar to their own locations in Rosedale, Ontario.

- **Therefore, the problem to be answered is: "Which areas of Bangalore are similar to Rosedale, Ontario ? Can we have the names and locations of Restaurants in these locations where accommodations can be arranged for their employees ?**

- **Who would be interested in this project ? In this project I attempt to help a fictional IT company in Toronto choose a locality which is similar to theirs. Anyone wishing to compare neighbourhoods similarly can check out this analysis.**

2. A description of the data and how it will be used to solve the problem

The data that we have used in this project is obtained from Wikipedia and Foursquare. We have obtained a list of localities or neighbourhoods of Toronto and Bangalore from the following Wikipedia pages and websites, as well as their geo location data like latitude and longitude.

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

https://www.mapsofindia.com/pincode/india/karnataka/bangalore/

https://www.mapdevelopers.com/batch_geocode_tool.php

These data is then scrapped using the Beautiful soup library in python, then formatted into a pandas dataframe for further processing.
Foursquare provides detailed data relating to each of the neighbourhoods. Foursquare is a local search and discovery service which provides search results for its users. Foursquare API will be used to explore the various types of venues and their categories available in each neighbourhood. The information for venues, categories, hotels, etc is extracted using foursquare to solve the business problem.
This data is used for segmenting and clustering the neighbourhoods of Bangalore. This will help to detect the similarities and dissimilarities of neighbourhoods. We will find a cluster which is most similar to Rosedale and then obtain a list of some of the hotels which are located in this cluster.

3. Methodology section to discuss and describe any exploratory data analysis done, any inferential statistical testing performed, and what machine learning was used and why.

The geolocation data that is collected from Wikipedia and www.mapdevelopers.com for Toronto and Bangalore neighbourhoods, is merged into the following format for further processing. We find the required data for Rosedale, Toronto (the original location we are interested in) and then continue our analysis with the Bangalore dataset.

```
In [16]: b.head()
```

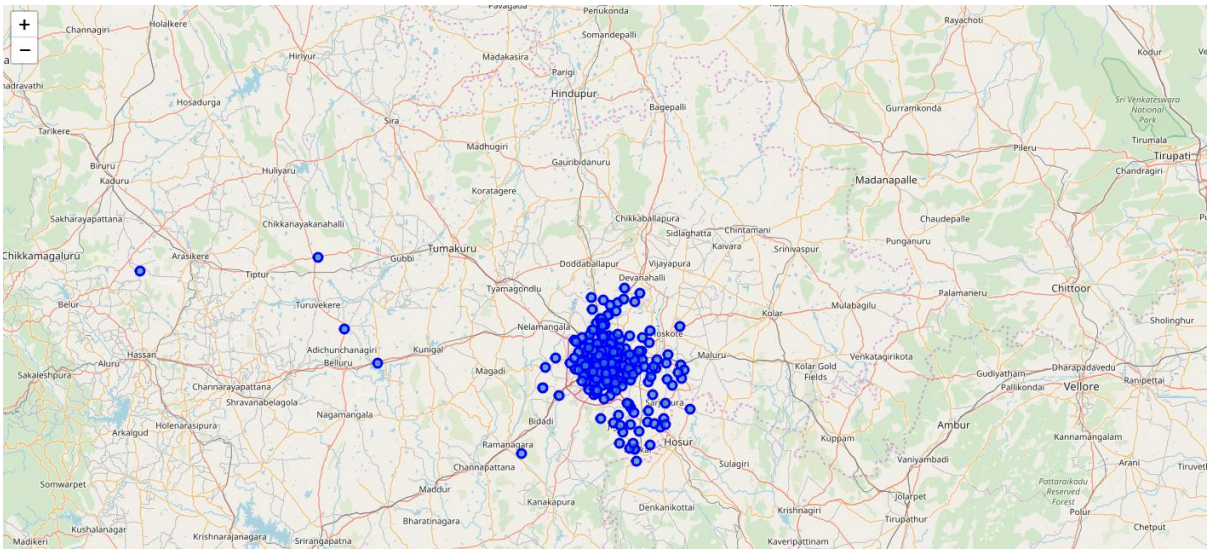
Out[16]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge ,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek ,Rouge Hill ,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood ,Morningside ,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Out[34]:

	Area	Pincode	Latitude	Longitude
0	A F station yelahanka	560063	13.132036	77.607519
1	Adugodi	560030	12.942004	77.608304
2	Agara	560034	12.923065	77.646453
3	Agram	560007	12.964357	77.621107
4	Air Force hospital	560007	12.964650	77.628136
5	Amruthahalli	560092	13.065879	77.604206
6	Anandnagar	560024	13.031328	77.591313
7	Anekal	562106	12.710526	77.691129
8	Anekalbazar	562106	12.710526	77.691129
9	Arabic College	560045	13.030375	77.621131
10	Aranya Bhavan	560093	13.010539	77.570839

Using the Folium library, we derive the following visualization of neighbourhoods in Bangalore represented by blue circles.



Then we use the Foursquare API to search/explore neighbourhoods in Rosedale and Bangalore. The explore function is used to obtain the most common venue categories within a radius of 500 metres that is extracted for each neighbourhood.

```
Rosedale
```

Out[51]:

	Area	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rosedale	43.679563	-79.377529	Rosedale Park	43.682328	-79.378934	Playground
1	Rosedale	43.679563	-79.377529	Whitney Park	43.682036	-79.373788	Park
2	Rosedale	43.679563	-79.377529	Alex Murray Parkette	43.678300	-79.382773	Park
3	Rosedale	43.679563	-79.377529	Milkman's Lane	43.676352	-79.373842	Trail

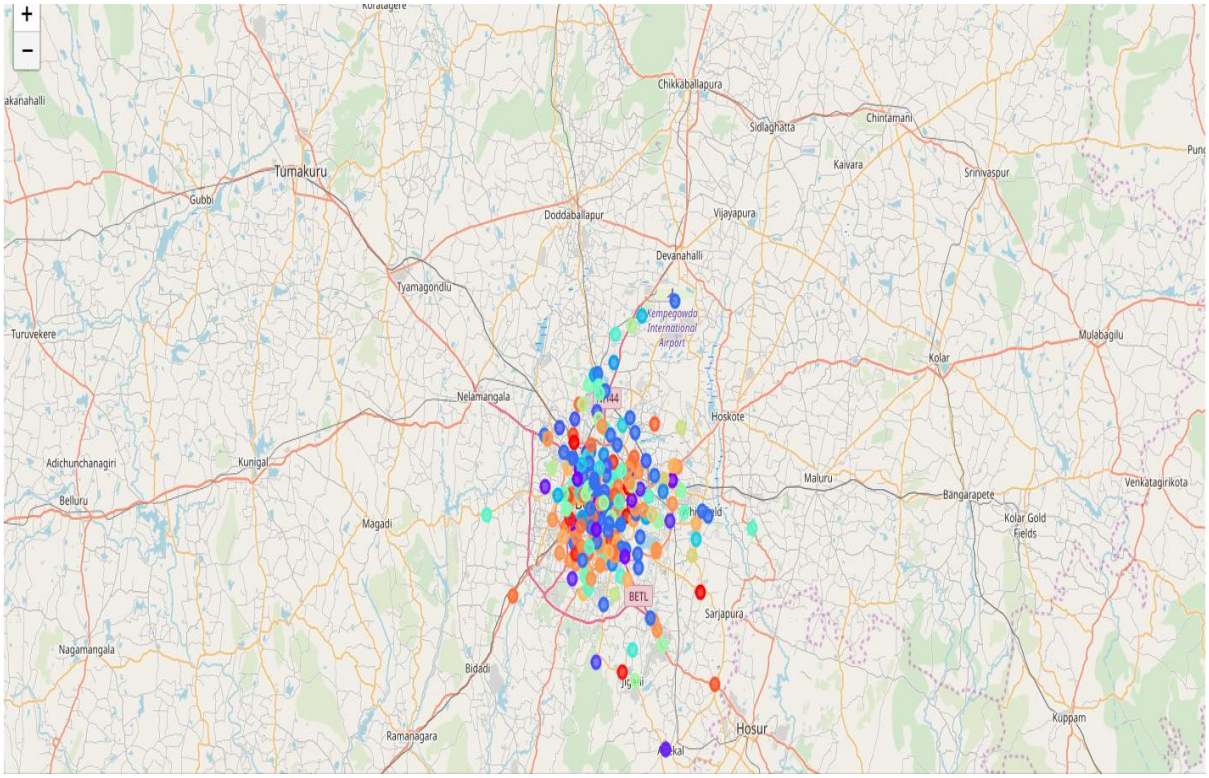
In [53]:
```
bangalore_venues.head()
```

Out[53]:

	Area	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A F station yelahanka	13.132036	77.607519	AERO INDIA 2017	13.132061	77.608366	Hot Spring
1	Adugodi	12.942004	77.608304	Bharathi Refreshments(South Indian Food) - Adu...	12.943388	77.607840	Fast Food Restaurant
2	Adugodi	12.942004	77.608304	Stoneart	12.941271	77.608701	Design Studio
3	Adugodi	12.942004	77.608304	Celio Ascendas	12.942748	77.609074	Men's Store
4	Adugodi	12.942004	77.608304	audugodi	12.942543	77.607353	Bus Station

Now the the Original neighbourhood and the Bangalore neighbourhoods are grouped into cluster on the basis of their shared characteristics and features. The number of cluster is chosen as 70. The k-means clustering algorithm is employed to perform this task. The sklearn.cluster package from the sci-kit library is used for performing the clustering.

Finally the Folium library is used to visualize the neighbourhoods of Bangalore and their clusters.



4. Result section to discuss the result

The analysis described above was performed and the following cluster of neighbourhoods in Bangalore was obtained that is predicted to show similar features to the original neighbourhoods of Rosedale, Toronto. Then a list of restaurants and cafe's was obtained using the Foursquare API.

ourpreferredcluster

Out[90]:

	Area	Latitude	Longitude	Pincode	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
2	Agara	12.9231	77.6465	560034.0	11	Indie Movie Theater	Chinese Restaurant	Italian Restaurant	Hyderabadi Restaurant	Fried Chicken Joint	Pizza Place
10	Aranya Bhavan	13.0105	77.5708	560003.0	11	Indian Restaurant	Playground	Dessert Shop	Chinese Restaurant	Bakery	Tea Room
11	Ashoknagar	12.9709	77.6048	560050.0	11	Indian Restaurant	Pub	Restaurant	Café	Lounge	Brewery
14	Austin Town	12.9588	77.616	560047.0	11	Indie Movie Theater	Department Store	Tibetan Restaurant	Electronics Store	Food Truck	Food Court

5. Discussion section to discuss any observations noted and any recommendations to make based on the results.

The clustering analysis done above was useful at classifying the neighbourhoods with similar features; however it should be noted that it is not a perfect solution. There are a large number of other factors that can be considered for clustering and segmenting purpose. Other data sources, demographic and socio-economic features data, detailed information related to the cafe's and restaurants can be added to the analysis for further refinement, and no algorithm can predict an accurate analysis of the neighbourhoods in consideration as an actual visit and stay in these places.

6. Conclusion section to conclude the report

To conclude, the neighbourhoods in Bangalore that are similar to Rosedale, Toronto are found, along with a list of cafe's and restaurants which the IT company can recommend to their employees who are travelling to India. This project can be replicated for different neighbourhoods and cities in order to find similar nerighbourhoods, and this can be expanded further using different datasets and taking more factors into consideration to provide more accurate analysis.