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 to 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 traveling to Melbourne for few weeks to work at their client office in Melbourne, Australia. Melbourne is one of the largest cities in Australia, with an area of 9,992.5 km2 covering 31 municipalities and over 345 localities. The company would like to choose only from those locations in Melbourne which are similar to Rosedale for their employees' accommodation in order to make them feel more comfortable and help them to adapt quickly to the new place. In this project we will attempt to help the company find those locations in Melbourne which are similar to their own location in Rosedale, Ontario.

- **Therefore, the problem to be answered is: "Which areas of Melbourne are similar to Rosedale, Ontario? Can we have the names and locations of hotels in these locations where accommodation 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 the list of localities or neighbourhoods of Toronto and Melbourne from the following Wikipedia pages and website, as well as their geo-location data like latitude and longitude.

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

https://en.wikipedia.org/wiki/Postal_district_numbers_of_Melbourne

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

These data is then scraped using the Beautiful Soup library in python, and then formatted into a pandas dataframe for further processing.

:

	Area	Latitude	Longitude
0	(Carnegie North)	-37.884829	145.058806
1	(Hawthorn East)	-37.831000	145.050000
2	(Williamstown West)	-37.808000	144.927000
3	Abbotsford	-37.803000	145.002000
4	Aerodrome Essendon	-37.726410	144.901488
5	Albert Park	-37.842000	144.950000

Foursquare provides detailed data relating to each of these neighbourhoods. Foursquare is a local search-and-discovery service mobile app 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 neighborhood. The information for venues, categories, hotels, etc is extracted using foursquare to solve the business problem.

This data is to be used for segmenting and clustering the 345 localities of Melbourne. This will help to detect the similarities and dissimilarities of neighborhoods. We find the cluster which is most similar to Rosedale and then obtain a list of some of the hotels which are located in this cluster.

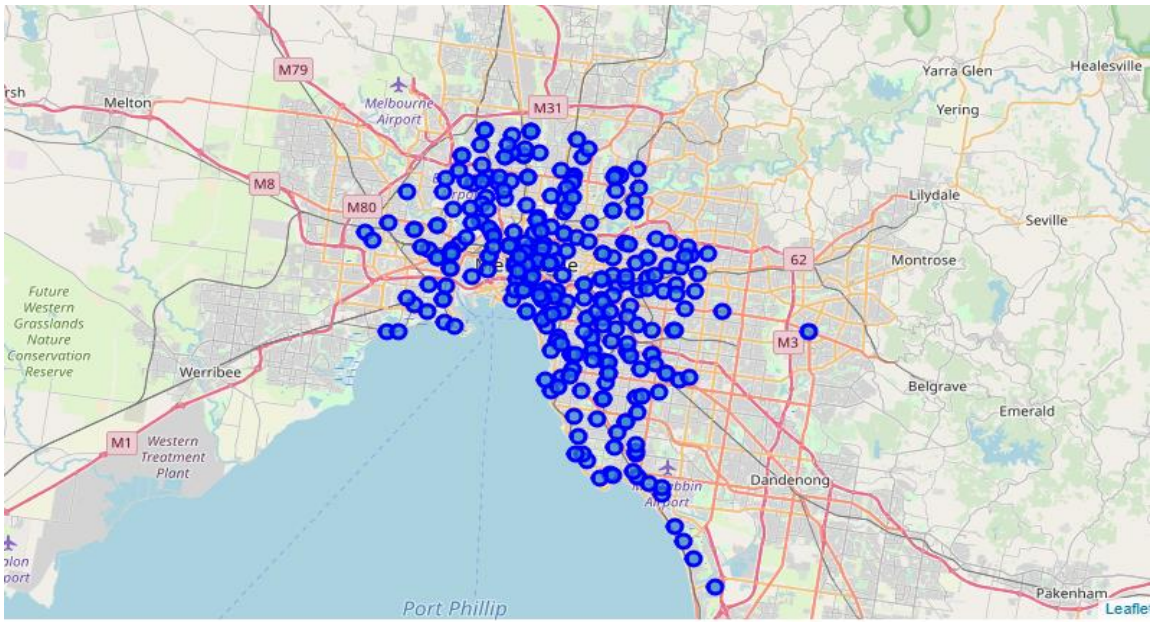
3. Methodology section to discuss and describe any exploratory data analysis done, any inferential statistical testing performed, and what machine learning was used and why.

The geolocation data that is collected from Wikipedia and www.mapdevelopers.com for Toronto and Melbourne neighborhoods, 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 Melbourne dataset.

	Postal Code	Bourough	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

	Area	Suburb	Latitude	Longitude
0	(Carnegie North)	(Carnegie North)	-37.884829	145.058806
1	(Hawthorn East)	(Hawthorn East)	-37.831000	145.050000
2	(Williamstown West)	(Williamstown West)	-37.808000	144.927000
3	Abbotsford	Abbotsford	-37.803000	145.002000
4	Aerodrome Essendon	Essendon	-37.726410	144.901488

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



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

Rosedale

	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

Melbourne:

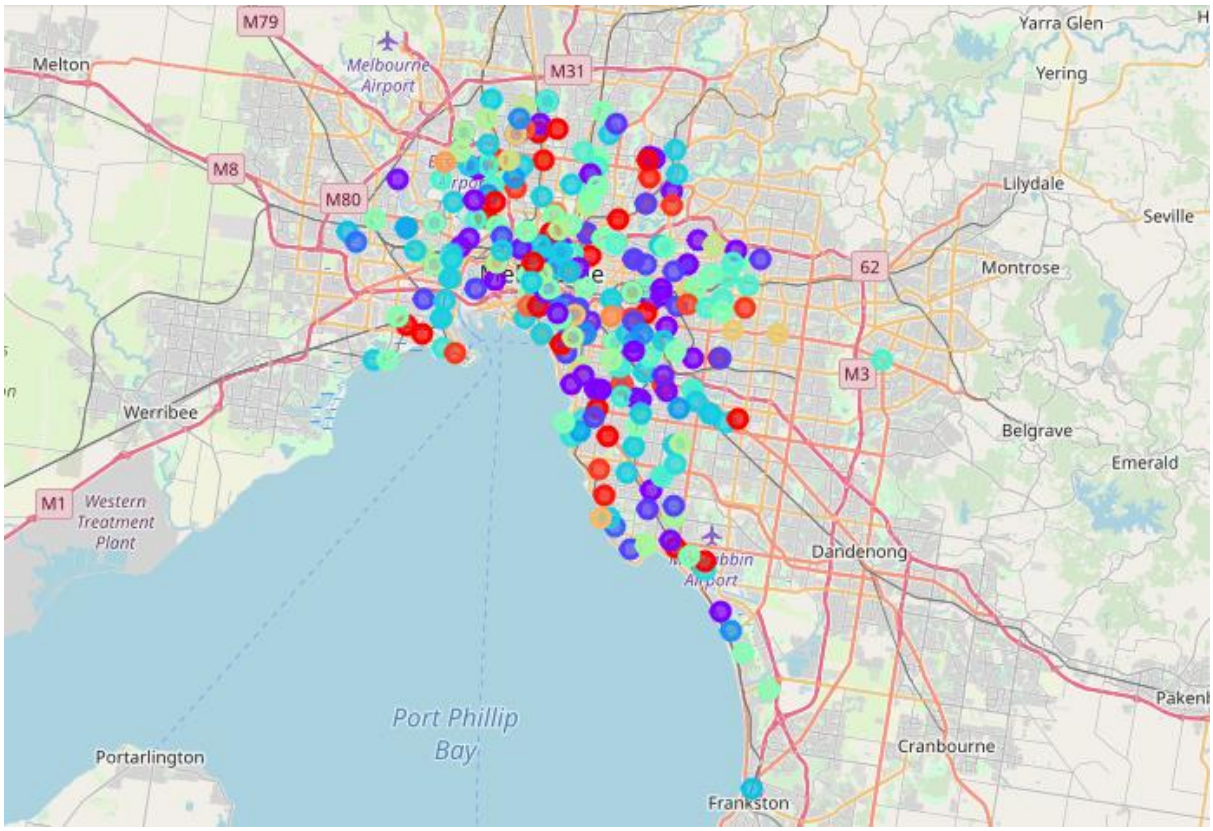
	Area	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	(Carnegie North)	-37.884829	145.058806	Grill'd	-37.887390	145.057494	Burger Joint
1	(Carnegie North)	-37.884829	145.058806	Shyun Japanese Restaurant	-37.888466	145.057225	Japanese Restaurant
2	(Carnegie North)	-37.884829	145.058806	Huff Bagelry	-37.888168	145.057269	Bagel Shop
3	(Carnegie North)	-37.884829	145.058806	Mrs Kim's Grill	-37.888670	145.057160	Korean Restaurant
4	(Carnegie North)	-37.884829	145.058806	Platform 3	-37.885513	145.057592	Breakfast Spot

It is determined that the Melbourne dataset contains 313 unique venue categories. The 10 most common venue categories are found for each of the 345 neighbourhoods in Melbourne.

	Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	
0	(Carnegie North)	Malay Restaurant	Thai Restaurant	Dumpling Restaurant	Supermarket	Korean Restaurant	Asian Restaurant	Japanese Restaurant	
1	(Hawthorn East)	Malay Restaurant	Asian Restaurant	Park	Liquor Store	Multiplex	Movie Theater	Café	
2	(St Kilda Road North)	Café	Hotel	Indonesian Restaurant	Australian Restaurant	Sushi Restaurant	Mobile Phone Shop	Park	
3	(St Kilda Road South)	Café	Hotel	Indonesian Restaurant	Australian Restaurant	Sushi Restaurant	Mobile Phone Shop	Park	
4	(Williamstown West)	Pier	Flea Market	Zoo Exhibit	Exhibit	Farm	Farmers Market	Fast Food Restaurant	
5	Abbotsford	Café	Pub	Farmers Market	Japanese Restaurant	Vegetarian / Vegan Restaurant	Convenience Store	Coffee Shop	Farm

Now, the original neighbourhood (Rosedale, Toronto) and the 345 Melbourne neighborhoods are grouped into clusters on the basis of their shared characteristics and features. The number of clusters 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 utilized to visualize the neighborhoods in Melbourne and their clusters.



4. Results section to discuss the results.

The analysis described above was performed and the following cluster of neighbourhoods in Melbourne was obtained that is predicted to show similar features to the original neighbourhood of Rosedale, Toronto. Then a list of five hotels in the selected cluster was obtained using the Foursquare API.

```
ourpreferredcluster
```

	Area	Latitude	Longitude	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
32	Beaumaris South	-37.9865	145.034	7	Café	Grocery Store	Chinese Restaurant	Shopping Mall	Breakfast Spot	Seafood Restaurant
63	Camberwell	-37.835	145.071	7	Train Station	Park	Café	Light Rail Station	Thai Restaurant	Tennis Court
71	Carlton South	-37.8001	144.967	7	Italian Restaurant	Café	Coffee Shop	Ice Cream Shop	Wine Bar	Burger Joint
73	Carnegie South	-37.835	144.96	7	Café	Pub	Coffee Shop	Bar	Bakery	Gastropub
89	Cheltenham North	-37.9593	145.067	7	Café	Playground	Athletics & Sports	Flower Shop	Farmers Market	Fast Food Restaurant
92	Clifton Hill East	-37.789	144.999	7	Café	Gym / Fitness Center	Gas Station	Pub	Park	Zoo Exhibit
100	Collins Street	-37.8183	144.957	7	Café	Coffee Shop	Japanese Restaurant	Hotel	Korean Restaurant	Bar
102	Cotham	-37.8086	145.046	7	Light Rail Station	Zoo Exhibit	Exhibit	Farm	Farmers Market	Fast Food Restaurant
108	Darling	-37.8385	144.992	7	Café	Italian Restaurant	Hotel	Japanese Restaurant	Dessert Shop	Convenience Store
110	Deepdene	-37.8131	145.068	7	Café	Gym	Zoo Exhibit	Farmers Market	Fast Food Restaurant	Filipino Restaurant
111	Degraves Street	-37.8173	144.966	7	Bar	Café	Coffee Shop	Italian Restaurant	Hotel	Asian Restaurant
114	Domain Road	-37.8341	144.983	7	Light Rail Station	French Restaurant	Australian Restaurant	Japanese Restaurant	Bakery	Hotel
116	Draytonville	-37.8136	144.963	7	Café	Bar	Coffee Shop	Cocktail Bar	Shopping Mall	Sushi Restaurant
172	Hawthorn	-37.8136	144.963	7	Café	Bar	Coffee	Cocktail	Shopping	Sushi

							Shop	Bar	Mall	Restaurant
173	Hawthorn North	-37.8136	144.963	7	Café	Bar	Coffee Shop	Cocktail Bar	Shopping Mall	Sushi Restaurant
189	Ivanhoe North	-37.7703	145.046	7	Japanese Restaurant	Café	Sandwich Place	Park	Shopping Mall	Pizza Place
191	Kensington	-37.791	144.931	7	Café	Pub	Pizza Place	Japanese Restaurant	Fast Food Restaurant	Bakery
195	Kew North	-37.8053	145.036	7	Café	Italian Restaurant	Fast Food Restaurant	Japanese Restaurant	Fish Market	Supermarket
197	Kingsville	-37.809	144.878	7	Café	Italian Restaurant	Fish & Chips Shop	Thai Restaurant	Coffee Shop	Zoo Exhibit
198	Kingsville South	-37.8301	144.87	7	Pizza Place	Indian Restaurant	Bakery	Playground	Zoo Exhibit	Fish & Chips Shop
204	Maidstone	-37.783	144.878	7	Gym	Sports Club	Zoo Exhibit	Flea Market	Farm	Farmers Market
251	North Brighton	-37.9036	145.005	7	Coffee Shop	Café	Movie Theater	Malay Restaurant	Fast Food Restaurant	Lounge
272	Port Melbourne	-37.8239	144.911	7	Bus Stop	Australian Restaurant	IT Services	Harbor / Marina	Zoo Exhibit	Flea Market
274	Prahran	-37.852	144.998	7	Café	Bar	Burger Joint	Italian Restaurant	Dumpling Restaurant	Vietnamese Restaurant
282	Regent West	28.0717	-80.6534	7	Athletics & Sports	Zoo Exhibit	Exhibit	Farm	Farmers Market	Fast Food Restaurant
307	St Kilda	-37.864	144.982	7	Café	Japanese Restaurant	Thai Restaurant	Hotel	Pizza Place	Hostel
308	St Kilda Junction	-37.8552	144.983	7	Café	Hotel	Light Rail Station	Gym	Pub	Vietnamese Restaurant
317	Stocksville	-37.8676	145.105	7	Park	Pizza Place	Athletics & Sports	Hockey Field	Zoo Exhibit	Farm
334	Toorak Road	-37.8443	145.035	7	Yoga Studio	Hockey Field	Dog Run	Park	Zoo Exhibit	Fish Market
346	Rosedale	43.6796	-79.3775	7	Park	Trail	Playground	Zoo Exhibit	Farm	Farmers Market

	name	location
0	Adina Apartment Hotel	{'address': '157 Fitzroy St', 'crossStreet': '...
1	Quest Windsor Hotel	{'address': '111 Punt Rd', 'lat': -37.85107890...
2	Park Regis Hotel	{'address': '604 St. Kilda Rd.', 'lat': -37.85...
3	Windsor Castle Hotel	{'address': '89 Albert St.', 'crossStreet': 'a...
4	The Elephant & Wheelbarrow Hotel	{'address': '169 Fitzroy St', 'crossStreet': '...

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 neighborhoods 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 segmentation purpose. Other data sources, demographic and socio-economic features data, detailed information related to the hotels, 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 at the places.

6. Conclusion section to conclude the report.

To conclude, the neighbourhoods in Melbourne that are similar to Rosedale, Toronto are found, along with a list of 5 hotels in these areas, which the IT company can use to arrange for accommodation for their employees. This project can be replicated for different neighbourhoods and cities in order to find similar neighbourhoods, and this can be expanded further using different datasets and taking more factors into consideration to provide more accurate analysis.