Assignment Week 5 Report

This report presents as a briefly report to explain since the defining problem in the business sector, which focused in Ontario, Canada. There are six components, which are introduction, data, methodology, result, discussion, and finally conclusion.

**Introduction**

In the business perspective, before launching any new businesses, to be able to know suitable location before launching the businesses is quite important, especially in the competitive market. From week 3 assignment, we can see that in Toronto, Ontario, there are many businesses within the cities. In fact, when it comes to the first tier cities, i.e., London, New York, Beijing, Tokyo, to be able to launch new businesses, it is a hard problem as there are many choices for customer. Therefore, in this assignment, we will perform a task to observe the exist businesses with ranking in top 10 venues in three cities in Ontario, Canada, where the zipcode is started with M.

In Canada, there are many provinces, which Ontario is one of the provinces. From Wikipedia website, Ontario is the most populous province in Canada, where people live in Ontario about 40%. So, it means that in Ontario, the businesses, education, and others should be developed further than other provinces. However, when it comes to 2020, which the world is running with competitive market, it is undeniable that more popular city tends to have more competitive market within the city, resulting in harder to launch the new business successfully. In addition, when it comes to lower tier cities, for instance, in UK case, the first tier city is London, while the second tier maybe Kent, it is still not easy to launch the new business as well as we need to evaluate the consumption and behavior of the people who live in that cities. In this assignment, we will use the basic model to observe the top 10 businesses within Toronto, and other two cities within Ontario province, which are Scarborough, and York. We will use the zipcode as an indicator to link to those cities.

**Data**

Data plays an important role in every segments. In this assignment, there are three data sources, which we need to gather them in order to be able to complete this work. Firstly, as mentioned earlier, we will use the zipcode as an indicator to link with three different cities. Therefore, we need a data that contains zipcode, and the area within that zipcode, for example, in Dundee city, Scotland, UK, there are four main area which the zipcode or postal code start with the city name, i.e., DD, followed by the number and area number, like DD1 5HA. After we have the zipcode data, the second data that is needed is the location of each cities that we want to explore, i.e., latitude and longitude. Furthermore, as we have both data, the most important things that we need to do is to link them together. After the linking is done, the final data that we need is the names of the places, which in this case is the venues. In order to get the venues data, we need to use an API, which called Foursquare API, which we can get the venues in the selected areas. The first data and second data are listed below with the links.

1. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
2. http://cocl.us/Geospatial\_data

**Methodology**

As this is the brief report, the procedure will be described in short, which are the main processes to get the data we want.

1. Data Wrangling

In this process, we need to clean the data to make sure it is usable. As there are two sets of data, which come from the first data and second data, hence we need to clean both of them. After cleaning process is done, then we need to combine them to be a single table likes below.

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

1. Extract the data from each cities that we want to explore with the same format as shown in the first step.
2. Use the geopy library to create the map in order to see how does the explored cities look like (optional).
3. Use Foursquare API to obtain the venues in each cities, within the limit and radius. In this case, we assume that people in the cities will travel by foot. Therefore, the radius should not be too far.
4. Rank the venues in each cities in Top 10 venues for each neighborhoods in each cities.

**Result**

A picture containing cabinet, station, train, computer

Description automatically generatedToronto

York

Table

Description automatically generated

Scarborough

Table

Description automatically generated

**Discussion**

From the result section, in can be seen that in each neighborhoods for each cities, the popularity of the businesses in different areas is also a good indicator that reflects to the lifestyle of the people in the each neighborhoods, and this data is the most important data for the business launching. To be able to success in launching the new business, we need to know and understand the culture in each areas. To be able to see the usefulness of the table, let’s say we want to launch a business in Central Bay Street, Toronto. As we can see, most of the shops are western style, like café, burger shop, sandwich shop, with the eastern style at the end of the list, Thai restaurant , and Japanese restaurant. So, if we want to launch new business, as least we knows the trend that people in this area prefer western style of food than eastern style.

**Conclusion**

From this assignment, we have seen that data is an important tool for evaluate the risk and successfulness of launching new business. However, there are many factors that are needed to be taken into account before launching new business. Based from this work, it can be improved much better in many aspects, for example, use machine learning to predict the future outcome if we know which business we are going to launch, but it also means we need more information as well. Therefore, this work can be think of as the starting point.