Capstone Project - The Battle of Neighborhoods

Mostafa Naemi

June 18 – 2019

1.Introduction

1.1 background

Melbourne is a large city of the Victoria state in Australia with a population of more than 5 million. Due to the diversity of people living in Australia and especially in Melbourne, there are various types of cuisine available. There is a high competition between restaurants in the suburbs with a high concentration of food stores and restaurants. Therefore, it is crucial to investigate the suitability of different suburbs for opening a new restaurant to reduce the risk of investment failure and financial loss.

1.2 Problem

Recently, a client who is going to open a restaurant in Melbourne asked us to introduce the best place for opening his restaurant. For this purpose, we need to analyse the number of restaurants in each desired suburb to find the best one. This project therefore aims to find the best place for a restaurant in Melbourne, Australia. Since the restaurant is going to be open until late night, the client has concerns about the crime in the suburb in which restaurants is located. As there is a trade-off between the crime rate and the suitability of a suburb for a restaurant, at the end of this project the followings are delivered to provide an insight into different options for our client:

- Our recommended list of top 3 suburbs for opening the restaurant.
- Crime data and number of restaurants in these suburbs.
- A geographical map of these suburbs.

2. Data

For this project, we need to first obtain the geographical data of all suburbs along with their postcode. Also, we require the total number of restaurants in each suburb. Furthermore, the historical crime data of Melbourne broken by suburbs is needed. The following table summarises the data required for this project:

Table 1: Required data and sources.

Data description	Source
List of Melbourne suburbs and their geographical location information	https://www.matthewproctor.com/full_australian_postcodes_vic
Number of restaurants within certain radius of each suburb	Foursquare API
Crime data of Melbourne suburbs	https://www.crimestatistics.vic.gov.au/crime-statistics/latest-crime-data/download-data

The following figure shows the clean dataframe of the 20 desired suburbs with their crime, geographical and number of restaurants data.

	Suburb	Latitude	Longitude	Postcode	No Restaurants	Incidents Recorded
0	ALBANVALE KEALBA KINGS_PARKST_ALBANS	-37.743361	144.796693	3021	12	148273
1	ALBION GLENGALA SUNSHINE SUNSHINE_NORTHSU...	-37.777204	144.829945	3020	12	7572
2	ALTONA SEAHOLME	-37.861720	144.812736	3018	5	6812
3	ARDEER DEER_PARK_EAST	-37.795937	144.793961	3022	1	7453
4	BRAYBROOK BRAYBROOK_NORTHROBINSON	-37.784683	144.854014	3019	3	671
5	DOCKLANDS	-37.814719	144.948039	3008	61	24358
6	EAST_MELBOURNE	-37.818517	144.982207	3002	14	15308
7	FOOTSCRAY SEDDON SEDDON_WEST	-37.807101	144.907953	3011	4	880
8	MELBOURNE	-37.817403	144.956776	3000	70	25150
9	MELBOURNE ST_KILDA_ROAD_CENTRAL	-37.844246	144.970161	3004	4	16537
10	NEWPORT SOUTH_KINGSVILLESPOTSWOOD	-37.838242	144.880556	3015	6	7049
11	SOUTH_WHARFSOUTHBANK	-37.824272	144.959000	3006	45	9118
12	UNIVERSITY_OF_MELBOURNE	-37.796152	144.961351	3010	24	9665
13	WEST_MELBOURNE	-37.810871	144.949592	3003	21	7096
14	WILLIAMSTOWNWILLIAMSTOWN_NORTH	-37.863743	144.888461	3016	3	9498
15	WORLD_TRADE_CENTRE	-37.824608	144.950858	3005	33	47594
16	YARRAVILLEYARRAVILLE_WEST	-37.814183	144.888686	3013	32	37309
17	ARDEER DEER_PARK_EAST	-37.795937	144.793961	3022	0	3682
18	BURNSIDE BURNSIDE_HEIGHTSCAIRNLEA CAROLINE_S...	-37.775501	144.768500	3023	0	31644
19	FIELDSTONEMAMBOURIN MANOR_LAKESMOUNT_COTTRELLW...	-37.863403	144.580165	3024	0	10275

Figure 1: Data required for this project

3. Methodology and Analysis

Now, we need to first visualise the data to derive some values from it and select the best method to analyse the data.

3.1 Exploratory Data Analysis

First, it is useful to compare the crime rate and the number of restaurants of various suburbs to develop a better understanding of the problem. Figure 2 shows the histogram of the crime data for all studied

suburbs. It can be seen that suburbs with postcode 3000, 3020 and 3021 have the highest number of crimes among all of studied suburbs.

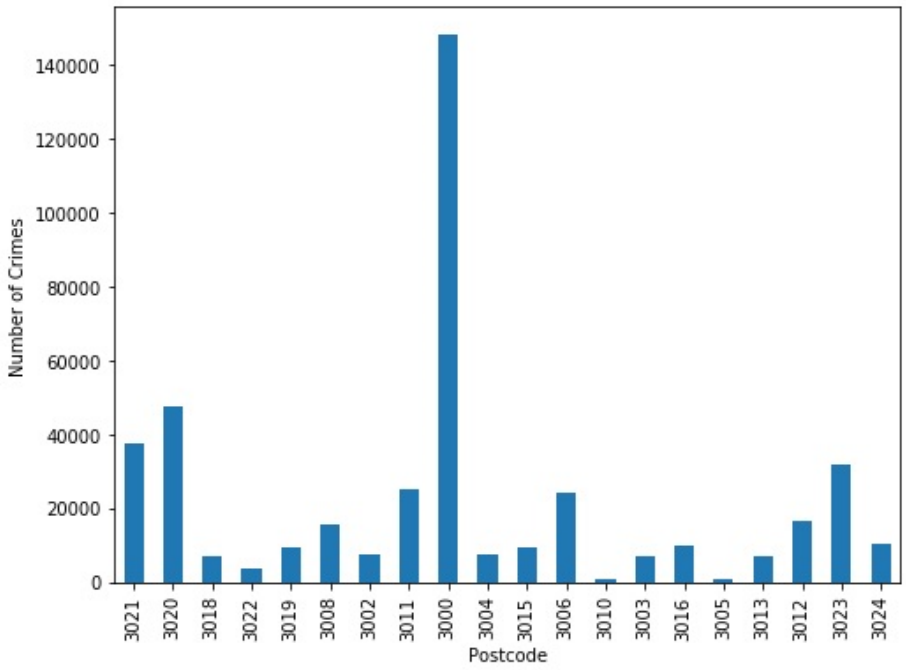

Figure 2: Histogram of the number of crimes for all studied suburbs.

Figure 3 shows the histogram of the number of restaurants for all suburbs. It can be seen that suburbs with postcode 3000, 3008 and 3006 have the highest number of restaurants. In other words, the number of competitors in these suburbs is higher.

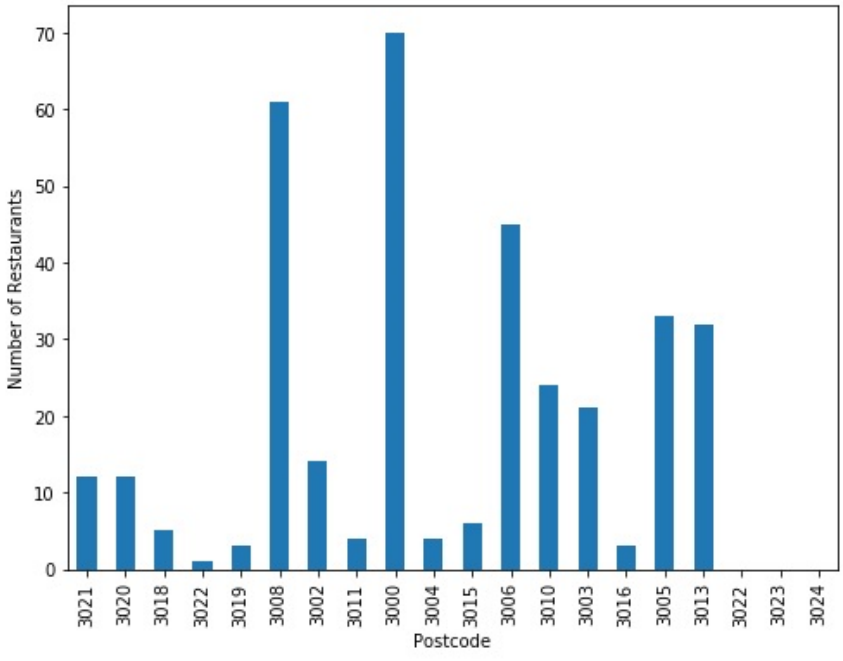

Figure 3:: Histogram of the number of restaurants for all studied suburbs

Figure 4 demonstrates the summary of the last two figures and presents the percentage of the number of restaurants and crime for each suburb. According to this figure, we can see that the suburb with postcode=3000 has the highest number of crimes and restaurants. Therefore, the first interpretation that we can get from this visualization is that this suburb is probably not a good candidate for opening a new restaurant.

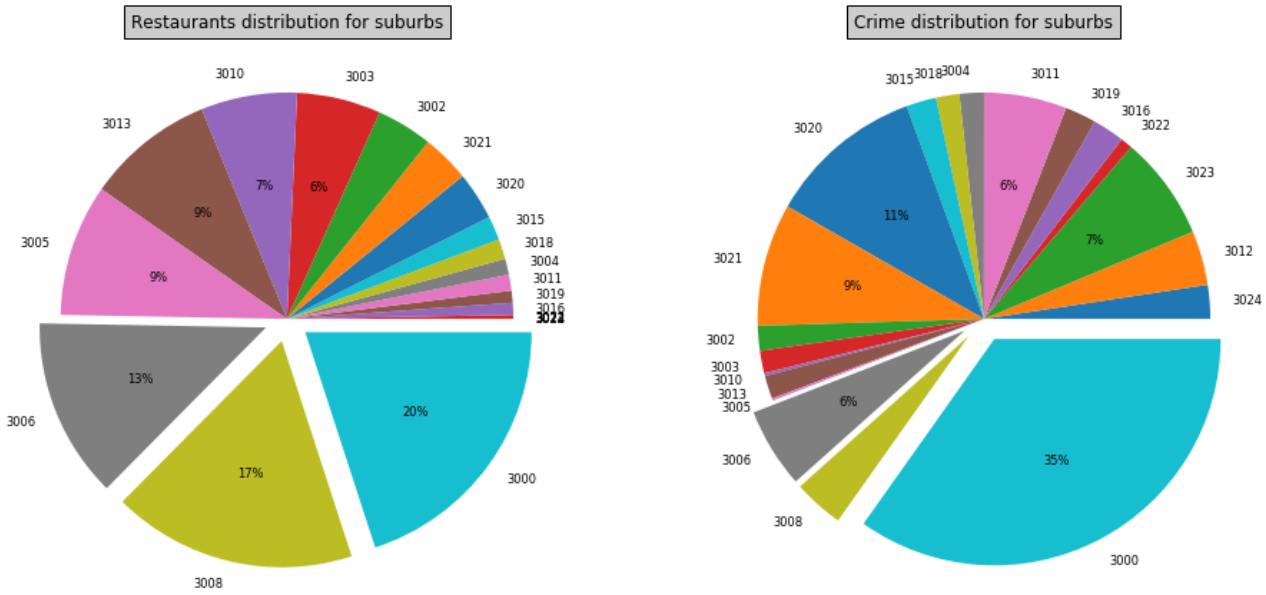

Figure 4: Pie charts of number of restaurants and crime for all studied suburbs.

Figure 5 shows the scatter plot of the number of crimes versus the number of restaurants. This provides an insight into the required number of clusters. We can see that most of suburbs are located in the bottom left part of the figure, meaning that there is low number of crimes and restaurants. This region of the figure represents the best locations for opening a new restaurant.

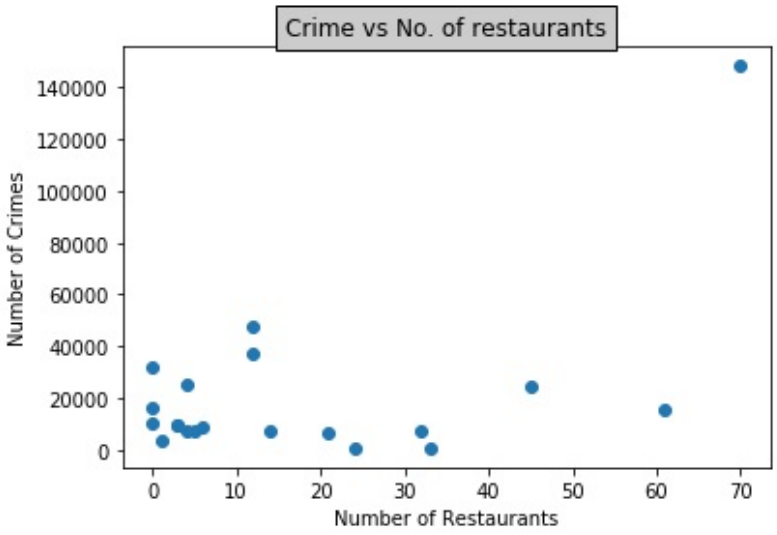

Figure 5: Scatter plot of number of crimes vs number of restaurants for all studied suburbs.

3.2 Clustering

After visualising the data of different suburbs, we need to cluster them in order to put similar suburbs in a same group. This enables us to recommend a list of suburbs which are suitable for the new restaurant. For clustering, we use K-means method with k determined according to the following figure which shows the distance between the cluster centres and all the other datapoint. According to this figure and using the elbow method, the optimal number of clusters is 5. Therefore, we divide the suburbs into 5 separate groups.

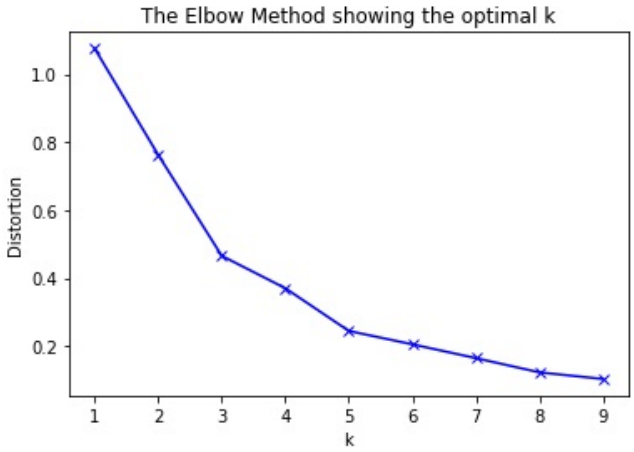

Figure 6: optimal number of clusters (k) required for this problem using distance between cluster centres and other data points.

4. Results and discussion

Table shows all suburbs data along with their cluster index. By looking at this table we can draw some insightful conclusion which can be described as:

- Cluster 0: Suburbs with relatively high number of crimes and high number of competitors
- Cluster 1: Suburbs with relatively high number of crimes and low-medium competitors
- Cluster 2: Suburbs with extremely high number of crimes and high competitors
- Cluster 3: Suburbs with relatively high number of restaurants and competitors
- Cluster 4: Suburbs with low-medium number of crimes and low number of competitors

Therefore, we should look for the best suburb in cluster 4. But we need to first investigate the geographical location of these suburbs on the map.

	Cluster Labels	Postcode	Suburb	Latitude	Longitude	No Restaurants	Incidents Recorded
0	1	3021	ALBANVALE KEALBA KINGS_PARKST_ALBANS	-37.743361	144.796693	12	37309
1	1	3020	ALBION GLENGALA SUNSHINE SUNSHINE_NORTHSU...	-37.777204	144.829945	12	47594
2	4	3018	ALTONA SEAHOLME	-37.861720	144.812736	5	7096
3	4	3022	ARDEER DEER_PARK_EAST	-37.795937	144.793961	1	3682
4	4	3019	BRAYBROOK BRAYBROOK_NORTHROBINSON	-37.784683	144.854014	3	9498
5	0	3008	DOCKLANDS	-37.814719	144.948039	61	15308
6	4	3002	EAST_MELBOURNE	-37.818517	144.982207	14	7572
7	1	3011	FOOTSCRAY SEDDON SEDDON_WEST	-37.807101	144.907953	4	25150
8	2	3000	MELBOURNE	-37.817403	144.956776	70	148273
9	4	3004	MELBOURNE ST_KILDA_ROAD_CENTRAL	-37.844246	144.970161	4	7453
10	4	3015	NEWPORT SOUTH_KINGSVILLESPOTSWOOD	-37.838242	144.880556	6	9118
11	0	3006	SOUTH_WHARFSOUTHBANK	-37.824272	144.959000	45	24358
12	3	3010	UNIVERSITY_OF_MELBOURNE	-37.796152	144.961351	24	880
13	3	3003	WEST_MELBOURNE	-37.810871	144.949592	21	6812
14	4	3016	WILLIAMSTOWNWILLIAMSTOWN_NORTH	-37.863743	144.888461	3	9665
15	3	3005	WORLD_TRADE_CENTRE	-37.824608	144.950858	33	671
16	3	3013	YARRAVILLEYARRAVILLE_WEST	-37.814183	144.888686	32	7049
17	4	3012	BROOKLYN KINGSVILLEKINGSVILLE_WESTMAIDSTONE T...	-37.807135	144.861162	0	16537
18	1	3023	BURNSIDE BURNSIDE_HEIGHTSCAIRNLEA CAROLINE_S...	-37.775501	144.768500	0	31644
19	4	3024	FIELDSTONEMAMBOURIN MANOR_LAKESMOUNT_COTTRELLW...	-37.863403	144.580165	0	10275

Figure 7: Dataframe of all suburbs data along with their cluster index.

We now visualise the clusters on the geographical map of Melbourne to see where suburbs are located and how they are clustered. Figure 8 shows the Melbourne suburbs clusters as circles. Each colour represents one cluster and the radius of circles is approximately proportional to the number of restaurants. We can clearly see that the cluster 1 (black) and 0 (purple) are close to each other which have high number of restaurants. Cluster 3 (blue) has relatively high number of restaurants. Cluster 2 (red) and 4 (yellow) have lower number of restaurants.

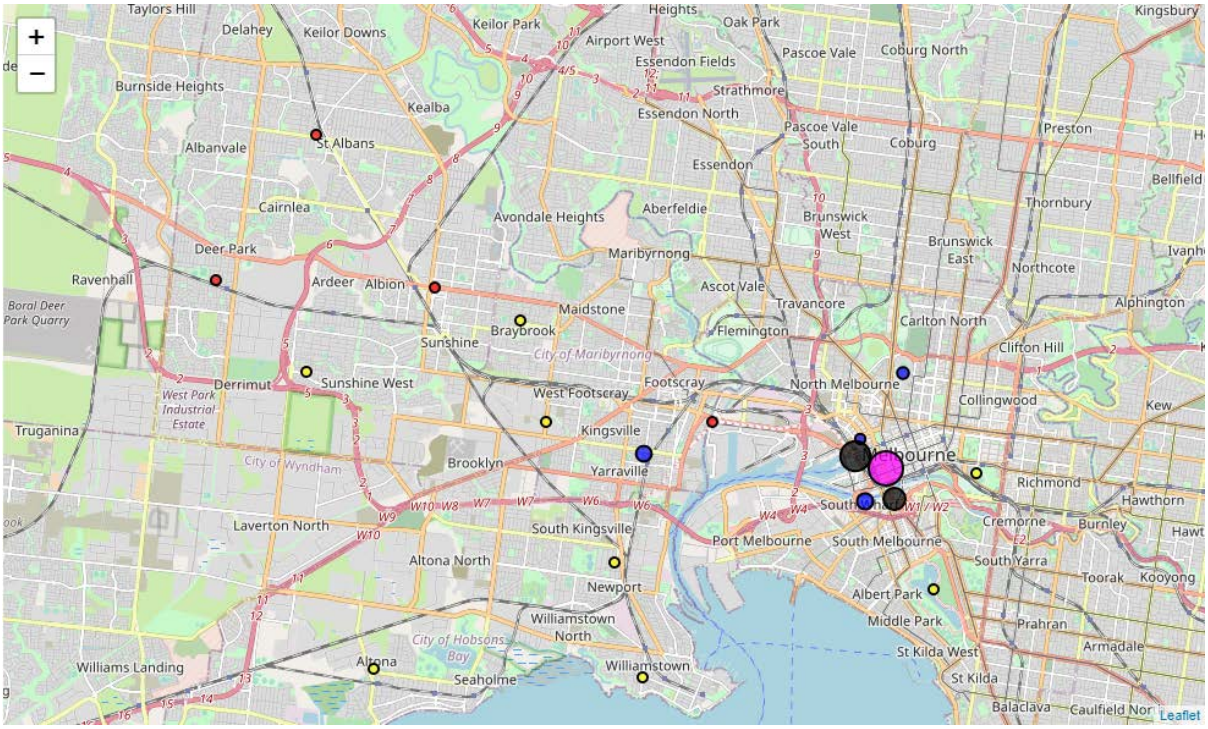

Figure 8: Map of Melbourne suburbs clustering. The radius of circles is approximately proportional to number of restaurants. Each color represents one cluster

Cluster 4 which is the best cluster for opening the new restaurant is shown in yellow which are scattered form east to west of the map. Now, let's look at the Figure 9 which shows the Melbourne suburbs clustering. In this figure the radius of circles is approximately proportional to their number of crimes. Now, the difference between cluster 2 (red) and 4 (yellow), having low number of restaurants, is more distinct. Comparison of these clusters show that the cluster 2 (red) has higher crime rate. Cluster 3 (blue) also has lower crime rate but Figure 8 indicated that this cluster has a high number of restaurants which makes the competition for the new restaurant more difficult.

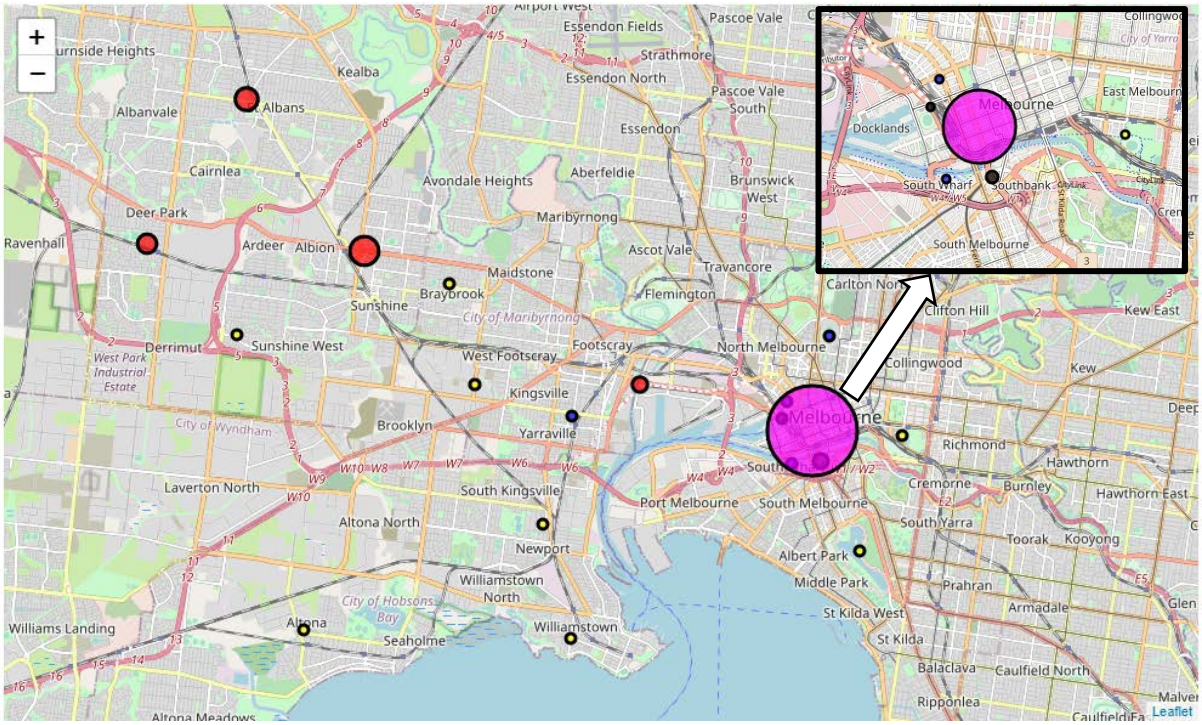

Figure 9: Map of Melbourne suburbs clustering. The radius of circles is approximately proportional to number of crimes. Each color represents one cluster.

The following figure shows all suburbs in cluster 4. Our desired suburb is in this cluster so, we need to first compare these suburbs. Considering our client concerns about the crime and the number of competitors, the best recommended suburbs are listed in the following table.

	Cluster Labels	Postcode	Suburb	Latitude	Longitude	No Restaurants	Incidents Recorded
19	4	3024	FIELDSTONEMAMBOURIN MANOR_LAKESMOUNT_COTTRELLW...	-37.863403	144.580165	0	10275
17	4	3012	BROOKLYN KINGSVILLEKINGSVILLE_WESTMAIDSTONE T...	-37.807135	144.861162	0	16537
3	4	3022	ARDEER DEER_PARK_EAST	-37.795937	144.793961	1	3682
14	4	3016	WILLIAMSTOWNWILLIAMSTOWN_NORTH	-37.863743	144.888461	3	9665
4	4	3019	BRAYBROOK BRAYBROOK_NORTHROBINSON	-37.784683	144.854014	3	9498
9	4	3004	MELBOURNE ST_KILDA_ROAD_CENTRAL	-37.844246	144.970161	4	7453
2	4	3018	ALTONA SEAHOLME	-37.861720	144.812736	5	7096
10	4	3015	NEWPORT SOUTH_KINGSVILLESPOTSWOOD	-37.838242	144.880556	6	9118
6	4	3002	EAST_MELBOURNE	-37.818517	144.982207	14	7572

Figure 10: Dataframe of all suburbs in cluster 4.

Table 2: Top 3 suburbs recommended for the new restaurant.

Priority	Postcode	Number of restaurants	Number of crimes
1	3022	1	3682
2	3004	4	7453
3	3018	5	7096

Figure 11 shows the location of these three suburbs on the Melbourne map. This provides a better insight for the client to decide on the location of the restaurant. As it can be seen, two recommended suburbs are located in the western side of the Melbourne, while one of them (postcode=3004) is located in the centre of Melbourne. Considering the proximity of this suburb (postcode=3004) to city centre, this can be a plus point because more people are commuting in the neighbourhood of Melbourne CBD (Commercial Business District). However, as shown in the above table this suburb has approximately two times more crimes compared with postcode=3022. Therefore, this depends on the client preference to choose one suburb out of these three recommended ones.

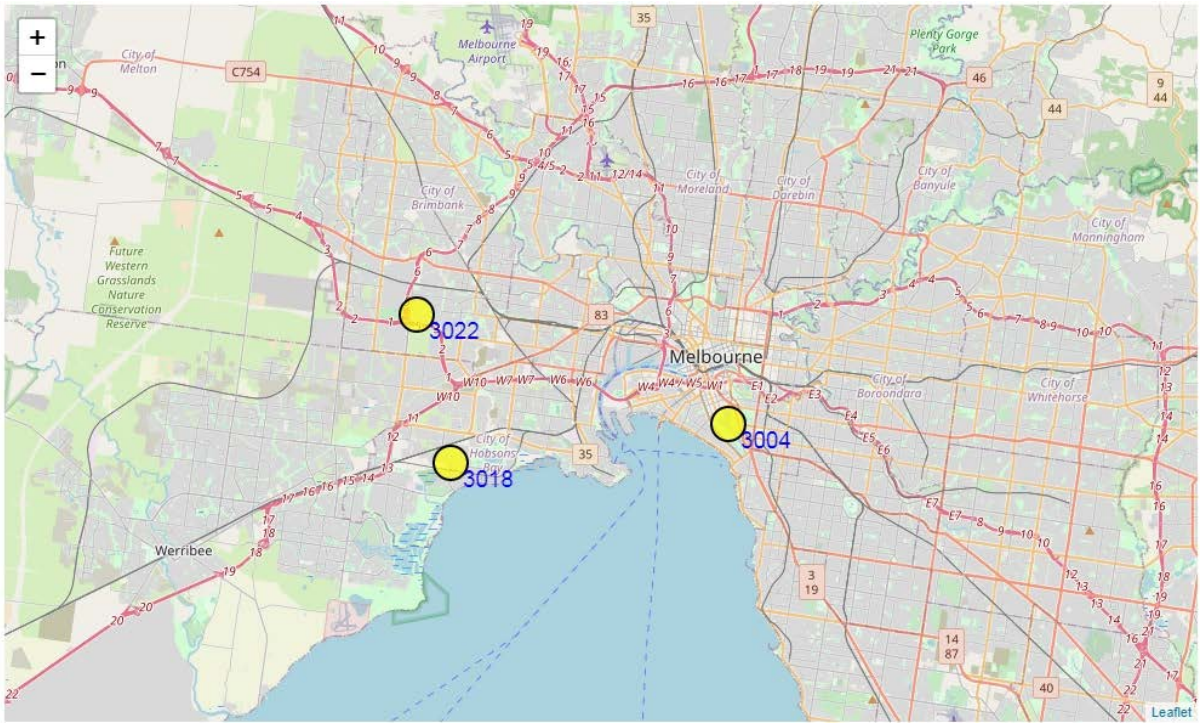

Figure 11: location of top 3 suburbs for the new restaurants (numbers are the postcodes).

5. Conclusion

This project aimed to find the top 3 suburbs for opening a new restaurant considering the number of restaurants in the suburbs and their crime rate. We first obtained the data from various sources and clean them to prepare the final dataset required for this project. After analysing various suburbs, we cluster them based on their crime data and their number of restaurants into 5 different clusters. We found that one of the clusters has lower number of competitors and also have a relatively low number of crimes. Various suburbs belonged to this cluster, among which we selected the top 3 of them to the client. These suburbs include postcode 3022, 3004 and 3018. We summarised the crime data and number of

restaurants for these suburbs that showed some pros and cons in their relative suitability for the new restaurant. Finally, we plotted these top 3 suburbs on Melbourne's geographical map. this provides information about the location of these suburbs and help the client to choose the best one.