Recommendation of location for an Indian restaurant in Melbourne

Yasmeen Fathima

10th April 2021

1. Introduction

1.1 Background:

Finding a suitable location for food business in a metropolitan city like Melbourne can be challenging. Looking at the variety and number of Indian restaurants in the city, it is vital that aspects such as location, cuisine and parking are considered carefully. Similarly, it is important to analyse the type of restaurants and population in that area. For example, considering the owner is interested in opening an Indian restaurant, the location should be selected where there are less competition and more scope for the business. This project aims only in the northern suburbs of Melbourne.

1.2Problem

Data that might be contributing to determining the best recommendations for a restaurant is based on the number of restaurants in the area, type of restaurants and venues. This project aims to suggest the most suitable location for opening a restaurant in Melbourne Northern suburbs.

1.3Interest

Restaurant owners and partners would be undoubtedly interested in finding the apt site for opening their restaurant in Melbourne.

2. Data acquisition and cleaning

2.1Data sources

To begin with, location data is based on the suburbs in Melbourne and postcodes. After carefully reviewing the number of websites, the website https://postcodes-australia.com/state-postcodes/vic seemed right for scrapping.

In addition to Victorian suburbs and postcodes, the Indian population in the suburb locations was also needed to determine the ideal location for a restaurant. This data was hard to find, finally, I extracted a survey report published in 2016 https://www.multicultural.vic.gov.au/images/2016_Census/CommunityProfiles2016/India-Community-Profile-2016-Census.pdf, which had a list of top locations where most Indians are located

2.2 Data cleaning

Data obtained from postcodes-australia.com is a list of 100s of locations. The list is scrapped from the website and stored in tabular form or a Data frame. Multiple locations were registered for a single postcode, for example, for postcode 3036 there were two locations. For simplicity purposes, it was reduced to one location.

For demographic details for Indians in suburbs, multicultural.vic.gov.au is scrapped. It is a pdf file. To scrape table from pdf file “tabula” is used to convert it to “.csv” file and the table was read using “read_csv”. This .csv file has one table extracted from the pdf, which is then converted into the data frame. We drop the unnecessary columns from the table and replace commas and strip the white spaces. The single suburb has multiple locations in the

2.3 Feature selection

As the scope of this project only focuses on the Northern suburbs of Melbourne, locations that are based in postcodes between 3000 and 3100 are only considered. This alone will fetch us 89 different locations to study and further analyse.

```
df = pd.DataFrame(sublist)
df.shape
```

```
Out[3]: (89, 2)
```

Demographic data obtained from the table also had multiple locations under one suburb. From this table, the top 20 locations with the most Indian population is extracted. The goal of the project is to now predict the location with no or least number of Indian restaurants with maximum Indian population in the vicinity.

	Place
0	Tarneit
1	PointCook
2	Truganina
3	HoppersCrossing
4	WyndhamVale
5	HamptonPark
6	CranbourneEast
7	Berwick
8	NarreWarrenSouth
9	CranbourneNorth
10	NoblePark
11	Dandenong
12	Springvale(Vic.)
13	Keysborough
14	DandenongNorth
15	Epping(Vic.)
16	Wollert
17	SouthMorang
18	Lalor
19	Thomastown

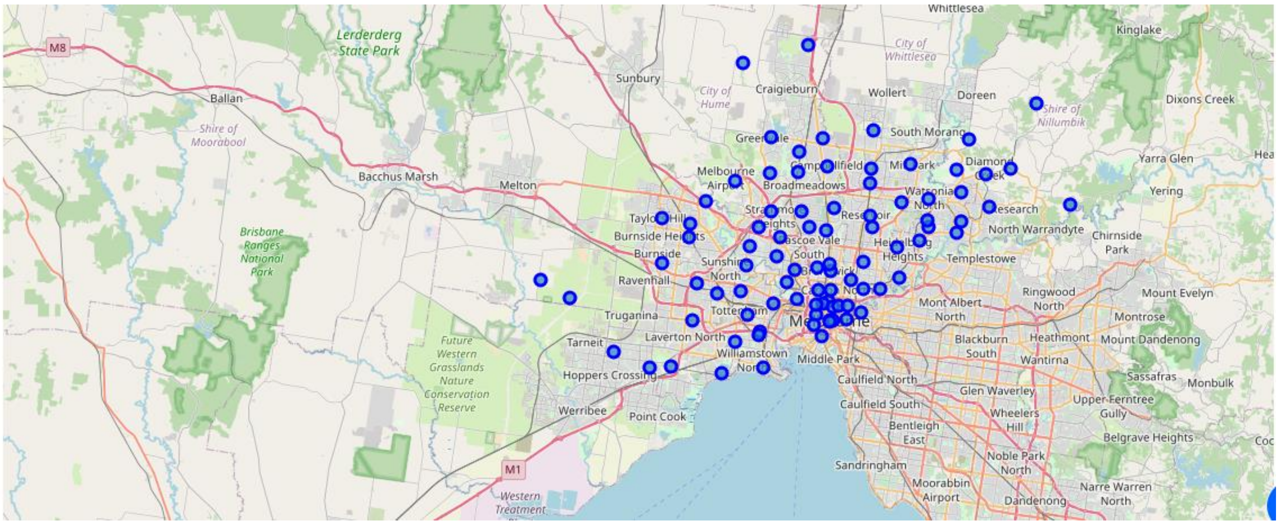
3. Methodology

3.1 Exploratory Data Analysis

Using BeautifulSoup class, data is scrapped through the website. After a thorough cleaning and formatting of data, I get a data set of lists of suburbs and their postcodes. Using pgeocode package, the latitudes and longitudes of the respective postcodes for all the 89 locations are calculated. This will help in further analysing the data.

	Place	Postcode	Latitude	Longitude
0	Melbourne	3001	-37.8140	144.963300
1	East Melbourne	3002	-37.8133	144.983900
2	West Melbourne	3003	-37.8088	144.947300
3	Melbourne	3004	-37.8140	144.967600
4	World Trade Centre	3005	-37.8152	144.963200

Using Folium, the map is generated which plots all the suburb locations ranging postcode from 3000 to 3100 on an Australian map. As it can be seen from the below map, locations are widely spread over the map.



Now, I need the information related to the venues i.e., restaurant details in these locations. Therefore, I use "Four Square" API to get these details. This is done by generating an URL and calling it using "requests" package. I need the following details to generate an URL:

```
CLIENT_ID,
CLIENT_SECRET,
VERSION,
neighborhood_latitude,
neighborhood_longitude,
radius,
LIMIT)
```

CLIENT ID, CLIENT SECRET, VERSION is obtained from Four Square developer's account. The latitude and longitude are taken from the table for the location with a Radius of 2000 meters and LIMIT of 100. With these details, it will display all the venues in these locations, which is overwhelming amounts of data, but I am only interested in restaurants that too specifically restaurants that serve cuisine that is not Indian. For this reason, I have modified the URL by adding another attribute called "category_id". Category IDs for places can be obtained from https://developer.foursquare.com/docs/build-with-foursquare/categories/. Therefore, the final URL looks like: https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&categoryId=4d4b7105d754a06374d81259&categoryId!=4bf58dd8d48988d10f941735&ll={},{}&radius={}&limit={}

Hence, the venue details with locations that do not have any Indian restaurants is obtained.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Melbourne	-37.8140	144.963300	Tipo 00	-37.813527	144.961978	Italian Restaurant
1	Melbourne	-37.8140	144.963300	B'cos Brazil	-37.815486	144.963085	Brazilian Restaurant
2	Melbourne	-37.8140	144.963300	Trattoria Emilia	-37.815220	144.962636	Italian Restaurant
3	Melbourne	-37.8140	144.963300	Calia	-37.812724	144.963930	Japanese Restaurant
4	Melbourne	-37.8140	144.963300	Nosh	-37.815396	144.962999	Asian Restaurant
...	...	...	...	...	...	...	...
948	Eltham	-37.7043	145.158267	House Of Salad	-37.704270	145.160196	Pizza Place
949	Eltham	-37.7043	145.158267	Wild Bean Cafe	-37.704172	145.160553	Café
950	Eltham	-37.7043	145.158267	Wild Bean Cafe	-37.704151	145.160565	Café
951	Eltham	-37.7043	145.158267	Nanna G's	-37.702307	145.155719	Australian Restaurant
952	Wattle Glen	-37.6672	145.184900	Wattle Glen General Store	-37.666571	145.182908	Café

953 rows × 7 columns

The list of locations

```
Melbourne
East Melbourne
West Melbourne
Melbourne
World Trade Centre
Southbank
Docklands
University Of Melbourne
Footscray
Brooklyn
Yarraville
Newport
Williamstown
Altona
Braybrook
Albion
Albanvale
Ardeer
Burnside
```

It shows that there are 953 locations of food places in the given postcodes. Further analysis by grouping data and calculating clusters will aid me to reduce the dataset with more feasible data sets.

3.2 Inferential Statistical Testing

The number of unique categories in the dataset is 82. After converting categorical data into indicator variables using data manipulation, get_dummies, here is how the table looks like:

	Neighborhood	Afghan Restaurant	African Restaurant	Argentinian Restaurant	Armenian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	...	Spanish Restaurant	Steakhouse	Sushi Restaur
0	Melbourne	0	0	0	0	0	0	0	0	0	...	0	0	0
1	Melbourne	0	0	0	0	0	0	0	0	0	...	0	0	0
2	Melbourne	0	0	0	0	0	0	0	0	0	...	0	0	0
3	Melbourne	0	0	0	0	0	0	0	0	0	...	0	0	0
4	Melbourne	0	0	0	0	1	0	0	0	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Grouping the neighbourhood data using groupby operation along with calculating mean values of each location will result in following

	Neighborhood	Afghan Restaurant	African Restaurant	Argentinian Restaurant	Armenian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	...	Spanish Restaurant	Steakhouse	Sushi Restaura
0	Abbotsford	0.0	0.0	0.00	0.0	0.083333	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
1	Aberfeldie	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
2	Airport West	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
3	Albanvale	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	1.00	0.00	...	0.00	0.0	0.00
4	Albion	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
60	West Melbourne	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
61	Williamstown	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
62	World Trade Centre	0.0	0.0	0.01	0.0	0.020000	0.01	0.0	0.01	0.01	...	0.02	0.0	0.05
63	Yarrambat	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00
64	Yarraville	0.0	0.0	0.00	0.0	0.000000	0.00	0.0	0.00	0.00	...	0.00	0.0	0.00

Now, calculating the top five venues at each of the locations. This analysis will help us determine the competition in each location for any restaurant. The frequency of food places at each location is calculated and it is displayed in ascending order. The following image shows the top 3 places with the highest number of food places in a given location. The same is calculated for all 64 neighbourhood locations.

```
----Abbotsford----
                             venue  freq
0                             Café  0.50
1              Japanese Restaurant  0.08
2   Vegetarian / Vegan Restaurant  0.08
3                 Asian Restaurant  0.08
4               Italian Restaurant  0.08


----Aberfeldie----
                 venue  freq
0                 Café   0.4
1      Thai Restaurant   0.2
2               Bakery   0.2
3           Food Court   0.2
4    Afghan Restaurant   0.0


----Airport West----
                       venue  freq
0         Italian Restaurant  0.33
1            Fish & Chips Shop  0.33
2                     Bakery  0.33
3   Middle Eastern Restaurant  0.00
4           Ramen Restaurant  0.00
```

A further deep analysis into the data to look at the 10 most common venues at each location. Observing this data set I can see that there are no Indian restaurants in these common venues. That means these locations would make a good choice for an Indian food business, given that Indian cuisine is popular. However, to determine the popularity of a particular food, it's important that there is a population of that ethnicity.

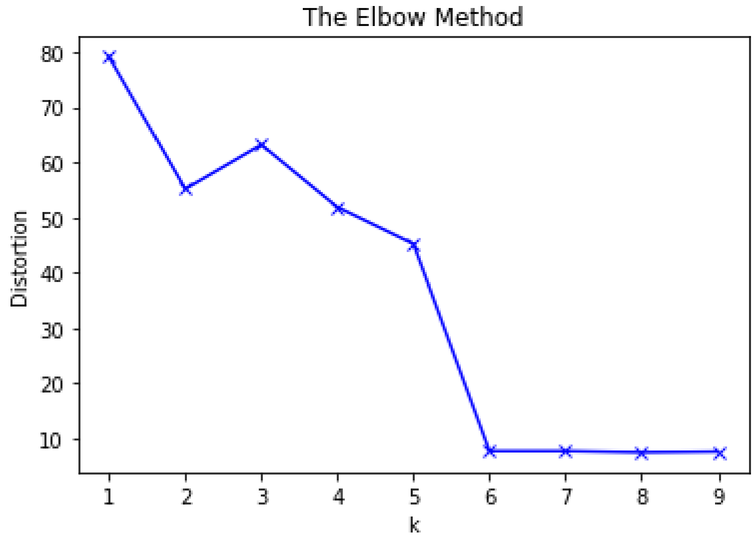
Here is the table with the 10 most common venues at each location.

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbotsford	Café	Vegetarian / Vegan Restaurant	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Japanese Restaurant	Burger Joint	Yunnan Restaurant	Fast Food Restaurant	Donut Shop
1	Aberfeldie	Café	Bakery	Thai Restaurant	Food Court	Fish & Chips Shop	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant
2	Airport West	Bakery	Italian Restaurant	Fish & Chips Shop	Food	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant
3	Albanvale	BBQ Joint	Yunnan Restaurant	Food	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant	Fast Food Restaurant
4	Albion	Chinese Restaurant	Gastropub	Pizza Place	Café	Bakery	Yunnan Restaurant	Fast Food Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant

3.3 Machine Learning

K Means algorithm is one of the most common cluster methods of unsupervised machine learning. I have some common venue categories in the neighbourhoods; therefore, K means clustering algorithm is used.

To find the optimal k value for the K Means algorithm, I used the elbow method. It is shown below.



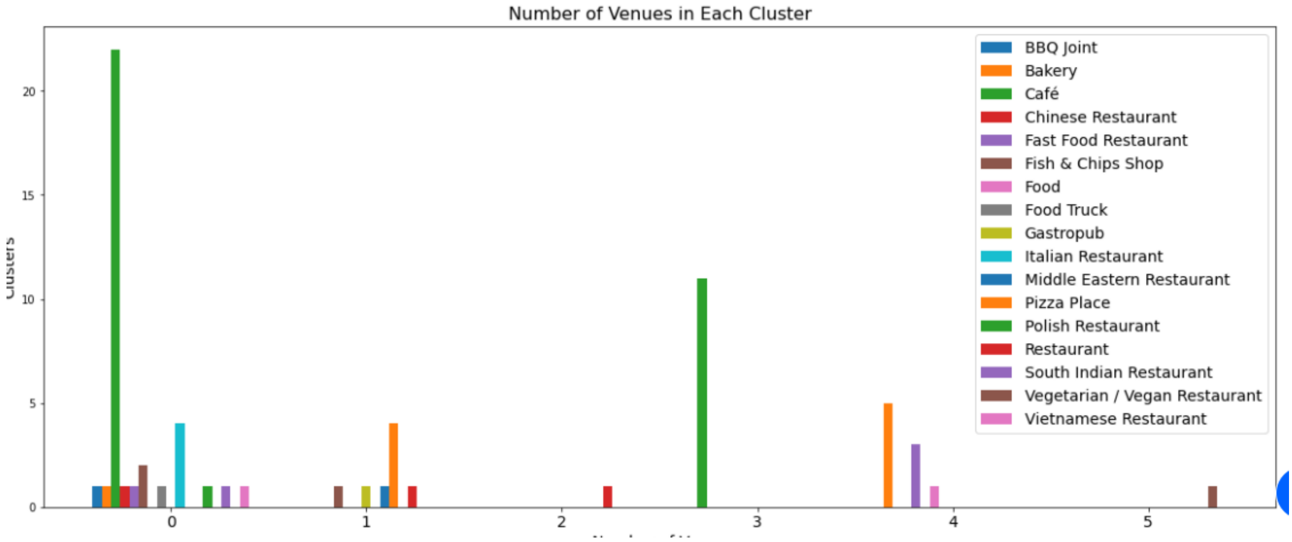
Merging the tables with neighbourhood venue details and original data set, we can see all the information in one table. I can use this merged table naming melb_merged, for generating clusters on the map. The merged table looks as follows:

	Place	Postcode	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Melbourne	3001	-37.8140	144.963300	0.0	Café	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Korean Restaurant	Asian Restaurant	Burger Joint	Ramen Restaurant	Dumpling Restaurant	Indonesian Restaurant
1	East Melbourne	3002	-37.8133	144.983900	0.0	Café	Australian Restaurant	Sushi Restaurant	Restaurant	Pizza Place	Indian Restaurant	Italian Restaurant	Fish & Chips Shop	Falafel Restaurant	Diner
2	West Melbourne	3003	-37.8088	144.947300	0.0	Café	Pizza Place	Sandwich Place	Italian Restaurant	Food Truck	Bakery	Noodle House	Mexican Restaurant	Bistro	Himalayan Restaurant
3	Melbourne	3004	-37.8140	144.967600	0.0	Café	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Korean Restaurant	Asian Restaurant	Burger Joint	Ramen Restaurant	Dumpling Restaurant	Indonesian Restaurant
4	World Trade Centre	3005	-37.8152	144.963200	0.0	Café	Italian Restaurant	Korean Restaurant	Mexican Restaurant	Sushi Restaurant	Japanese Restaurant	Sandwich Place	Dumpling Restaurant	Burger Joint	Bakery
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
80	Diamond Creek	3089	-37.6728	145.153700	0.0	Café	Fish & Chips Shop	Malay Restaurant	Fast Food Restaurant	Yunnan Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant
82	Yarrambat	3091	-37.6390	145.133400	3.0	Café	Yunnan Restaurant	Fish & Chips Shop	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant	Fast Food Restaurant
84	Montmorency	3094	-37.7188	145.124500	0.0	Café	Chinese Restaurant	Thai Restaurant	Burger Joint	Fast Food Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant

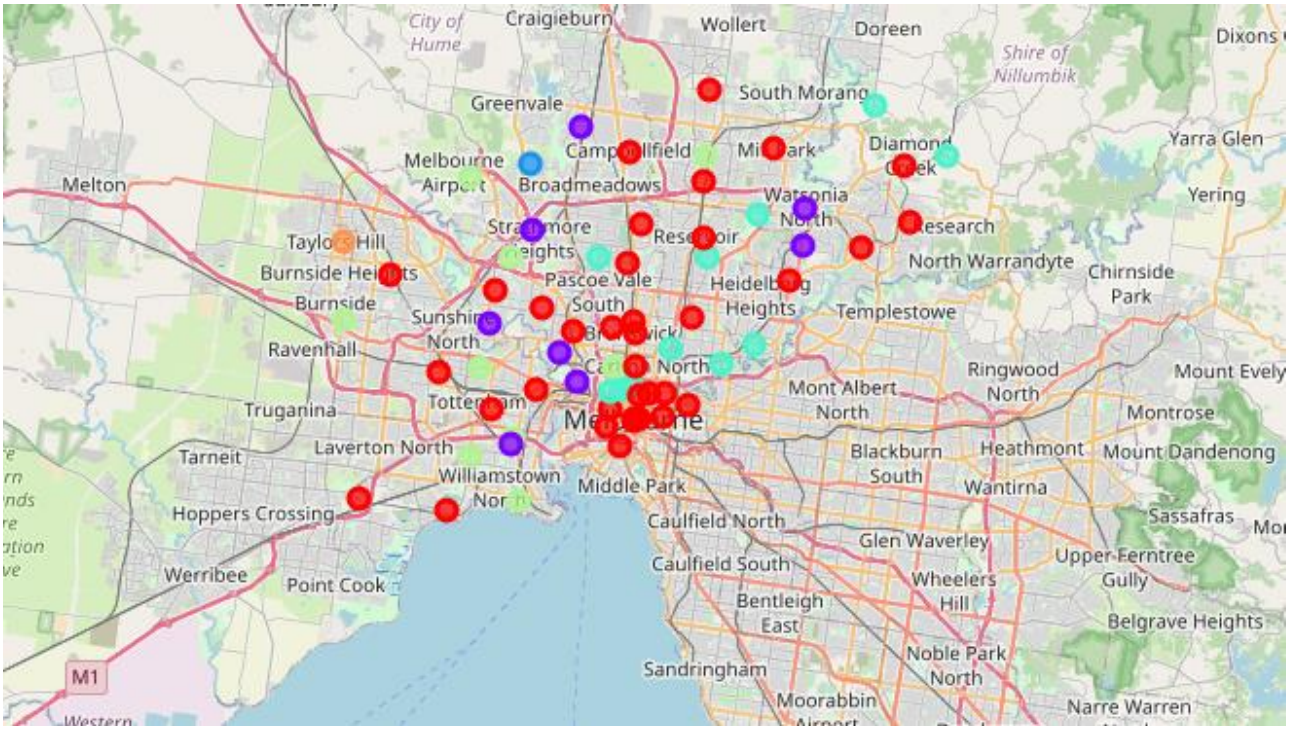
After estimating the number of 1st Most common venue in each cluster, we can create a bar chart which may help us to find proper label names for each cluster

1st Most Common Venue	BBQ Joint	Bakery	Café	Chinese Restaurant	Fast Food Restaurant	Fish & Chips Shop	Food	Food Truck	Gastropub	Italian Restaurant	Middle Eastern Restaurant	Pizza Place	Polish Restaurant	Restaurant	South Indian Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant
0	1	1	22	1	1	2	0	1	0	4	0	0	1	0	1	0	1
1	0	0	0	0	0	1	0	0	1	0	1	4	0	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
3	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	5	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Here is the bar chart displaying the clusters



And the map with clusters with optimal k=6, as shown below



Now I have the optimal list of clusters with food joints across Melbourne suburbs with postcodes 3000 and 3100. Based on the demographic data of most Indians residing in the suburban areas, comparing the tables, we get the top 3 locations that can be considered ideal to open an Indian restaurant.

4. Results

After cleaning the data in the table i.e., striping the extra spaces and commas, I get the list of the top 20 locations of Melbourne where there are most Indians. Now, merge this table with melb_merged, with common Place names. This gives the final table with three locations.

	Place	index	Postcode	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Epping	69	3076	-37.6307	145.0171	0.0	Italian Restaurant	Café	Yunnan Restaurant	Fish & Chips Shop	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant
1	Lalor	68	3075	-37.6673	145.0145	4.0	Bakery	Food	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant	Fast Food Restaurant	Fish & Chips Shop
2	Thomastown	67	3074	-37.6821	145.0126	0.0	Fish & Chips Shop	Fried Chicken Joint	Café	Middle Eastern Restaurant	Halal Restaurant	Gastropub	French Restaurant	Food Truck	Food Court	Food

Epping, Lalor and Thomastown are three locations common in both the tables. In these places, the frequency of Indian restaurants is less compared to other restaurants.

If interested, here are the top 8 locations with a similar search of locations.

					-	Labels	Venue	Venue	Venue	Venue	Venue	Venue	Venue	Venue	Venue	Venue
0	Epping	69	3076	-37.6307	145.01710	0.0	Italian Restaurant	Café	Yunnan Restaurant	Fish & Chips Shop	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant
1	Lalor	68	3075	-37.6673	145.01450	4.0	Bakery	Food	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	Falafel Restaurant	Fast Food Restaurant	Fish & Chips Shop
2	Thomastown	67	3074	-37.6821	145.01260	0.0	Fish & Chips Shop	Fried Chicken Joint	Café	Middle Eastern Restaurant	Halal Restaurant	Gastropub	French Restaurant	Food Truck	Food Court	Food
3	Melbourne	0	3001	-37.8140	144.96330	0.0	Café	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Korean Restaurant	Asian Restaurant	Burger Joint	Ramen Restaurant	Dumpling Restaurant	Indonesian Restaurant
4	Melbourne	3	3004	-37.8140	144.96760	0.0	Café	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Korean Restaurant	Asian Restaurant	Burger Joint	Ramen Restaurant	Dumpling Restaurant	Indonesian Restaurant
5	Docklands	6	3008	-37.8183	144.94450	0.0	Café	Restaurant	Italian Restaurant	Chinese Restaurant	Pizza Place	Seafood Restaurant	Sandwich Place	Middle Eastern Restaurant	Steakhouse	Australian Restaurant
6	Southbank	5	3006	-37.8291	144.95305	0.0	Café	Asian Restaurant	Restaurant	Burger Joint	Fast Food Restaurant	Sandwich Place	Caucasian Restaurant	Mexican Restaurant	Paella Restaurant	Pizza Place
7	Carlton	47	3053	-37.8007	144.96710	0.0	Italian Restaurant	Café	Thai Restaurant	Pizza Place	Asian Restaurant	Vegetarian / Vegan Restaurant	Malay Restaurant	Fish & Chips Shop	Burger Joint	Japanese Restaurant

There is also scope in these locations

5. Discussion

The datasets used in this project are sourced from different websites. I used BeautifulSoup to scrape this data.

Considering the demographic data obtained is from the year 2016, there can be a conflict with the result and current business situation. Therefore, it is vital that we obtain an updated list of locations.

I used the K means algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 6.

As there are a smaller number of restaurants compared to the Indian population in these areas, it can be inferred that these locations are the ideal locations to open an Indian restaurant.

6. Conclusion

Businesses that focus on the food industry, therefore, choose the location based on the population density in the area. Moreover, considering the ethnicity of the population in the selected location can have a significant impact on business.