EATERY ANALYSIS OF MELBOURNITES

1. INTRODUCTION

1.1 BACKGROUND

Food is the one substance that has constant variety in this world! It is not simply stuff, that people eat when hungry, food is delicious, inventive, inspiring, and much more. Unlike, other societies with a dominant agrarian history, Australians have no cuisines inherited in the traditional sense. Immigrants from all over the world brought their culinary traditions with them to Australia.

Eating out is a popular pastime in Australia and we have a huge choice of fabulous cafes, restaurants, pubs and bars in our cities and towns. Therefore, it is very advantageous for investors to set up a food business in Australia.

Any taste! Any smell! Any colour! FOOD!

1.2 PROBLEM

The main aim of this project is predicting a location for investors, to set up a restaurant in and around Melbourne Metropolitan region (Melbourne is one of the major cities in Victoria, Australia). And also, this project aims to understand what Melburnians love to eat the most.!

1.3 INTERESTS

Food is an emotion, which travels with us in every moment of our life. Investors would be interested in these predictions as it helps them to find a better place for setting up a restaurant and also for competitive advantages, to those who run food business.

2. DATA ACQUISITION AND CLEANING

2.1 DATA SOURCES

All the metropolitan restaurants data were brought in through Foursquare API and it is set into a Data frame. Restaurants were segregated based on the council areas they are located in. The boundaries of the council area are acquired from the GeoJSON file found here.

2.2 DATA CLEANING

There was a lot to do with cleaning the data.!

As I've chosen to analyse the restaurants of Melbourne Metropolitan region, there is no way, I could collect the data from Foursquare API in a single request. Because, the maximum number of venues we can pull in a single request are limited to 1000. So, I grabbed the data by sending request for every suburb and merged them into a data frame.

Depending on the radius settings, when extracting data from Foursquare there is a chance that a venue might appear multiple times. In order to overcome this issue, I decided to drop all the duplicates based on the restaurant ID.

While examining each feature, some redundancy was found. There are two features **lat** and **lng** contained the same information as **labeledLatLngs** which makes the data frame a little clumsy. So, I dropped the feature, **labeledLatLngs.**

When I saw the insights of my data frame, I found many properties in one of the features, **categories**. I made each property into a new feature and dropped some of the unnecessary features.

	categories	id	lat	lng	name
0	Dim Sum Restaurant	4b05874df964a520318a22e3	-37.812160	144.966569	Dragon Boat Restaurant
1	Dumpling Restaurant	4b1b3210f964a52074f923e3	-37.811990	144.965210	Chinatown Dumpling Restaurant
2	Chinese Restaurant	4b05874cf964a520f78922e3	-37.812027	144.966900	West Lake Restaurant
3	Chinese Restaurant	4d1417e2bb64224b6f3daa65	-37.812991	144.967075	Shanghai Dynasty Restaurant
4	Korean Restaurant	4e77090cd4c0934472942be3	-37.812742	144.960748	Yami Yami Korean & Japanese Restaurant
...	...	...	...	...	...
1156	Asian Restaurant	4f49eac2e4b093da13279669	-38.153490	145.185540	Delight Inn Malaysian & Chinese Restaurant
1157	Japanese Restaurant	5e60a54ffd0227000703987b	-38.115608	145.240721	Okami Japanese Restaurant
1158	Asian Restaurant	4bdab05d3904a5936306479e	-37.910750	145.037160	Silver Dragon
1159	Italian Restaurant	5718ba24498e88a3a859dce1	-37.849413	144.961717	Torsca italian restaurant
1160	Greek Restaurant	4be546bd910020a12cc6d214	-37.849312	144.967404	starvos restaurant

Figure 1: Data frame after removing the unnecessary features

Foursquare API dataset includes only the latitude and longitude location of restaurants. In order to segregate them, council area should be mapped based on those values.

I checked the council area dataset which was in GeoJSON format. Without modifying the dataset, choropleth map can be generated. But, for mapping the council areas with restaurant, I converted the GeoJSON file into a data frame.

	type	properties	geometry	poly
0	Feature	{'OBJECTID_1': 1, 'OBJECTID': 408, 'COLOR': 1,...	{'type': 'Polygon', 'coordinates': [[[141.0021...	
1	Feature	{'OBJECTID_1': 2, 'OBJECTID': 409, 'COLOR': 4,...	{'type': 'Polygon', 'coordinates': [[[142.7785...	
2	Feature	{'OBJECTID_1': 3, 'OBJECTID': 410, 'COLOR': 2,...	{'type': 'Polygon', 'coordinates': [[[142.6285...	
3	Feature	{'OBJECTID_1': 4, 'OBJECTID': 411, 'COLOR': 3,...	{'type': 'Polygon', 'coordinates': [[[142.3646...	
4	Feature	{'OBJECTID_1': 5, 'OBJECTID': 412, 'COLOR': 3,...	{'type': 'Polygon', 'coordinates': [[[143.8219...	
...	...	...	...	...
82	Feature	{'OBJECTID_1': 83, 'OBJECTID': 490, 'COLOR': 1...	{'type': 'Polygon', 'coordinates': [[[145.0105...	
83	Feature	{'OBJECTID_1': 84, 'OBJECTID': 491, 'COLOR': 3...	{'type': 'Polygon', 'coordinates': [[[144.6800...	
84	Feature	{'OBJECTID_1': 85, 'OBJECTID': 492, 'COLOR': 4...	{'type': 'Polygon', 'coordinates': [[[145.7353...	
85	Feature	{'OBJECTID_1': 86, 'OBJECTID': 493, 'COLOR': 3...	{'type': 'Polygon', 'coordinates': [[[145.3350...	
86	Feature	{'OBJECTID_1': 87, 'OBJECTID': 494, 'COLOR': 1...	{'type': 'Polygon', 'coordinates': [[[142.4223...	

Figure 2: Council area data frame.

For further analysis, I merged the polished datasets into one.

2.3 FEATURE SELECTION

After data wrangling, there were 1,164 samples in the data. As we move forward, new features might be added depending upon the requirements.

3. EXPLORATORY DATA ANALYSIS

I used Foursquare API in my study, as a database. My baseline data which has **categories, Id, name, Latitude** and **Longitude** informations of the Metropolitan Melbourne. I designed the limit as 1000 venues, radius 5000 meter for each suburb, called one-by-one and merged the data.

3.1 VISUALIZING METROPOLITAN MELBOURNE

Using **Folium** library, I visualized the geographic details of Metropolitan Melbourne restaurants. Using Latitudes and Longitudes, I created a map as shown below:

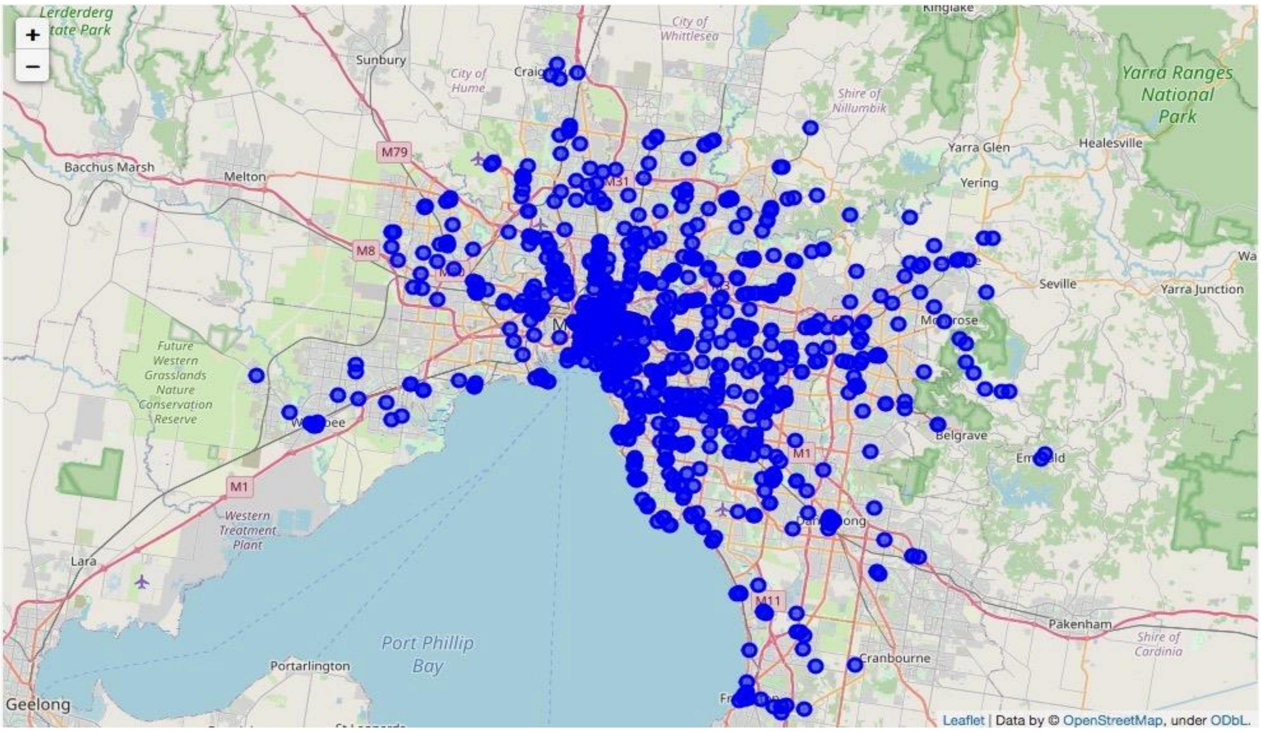

Figure 3: Restaurant locations in Metropolitan Melbourne.

3.2 CATEGORIZING RESTAURANTS

Next, I tried visualising the top 20 cuisines using bar chart. Then I learnt, some of the restaurants did not mention the cuisines they serve. So, I tried to identify the cuisines based on the restaurant names, for example the name of a restaurant includes Chinese but their cuisine is given as 'Restaurant', then I modified it to 'Chinese Cuisine'.

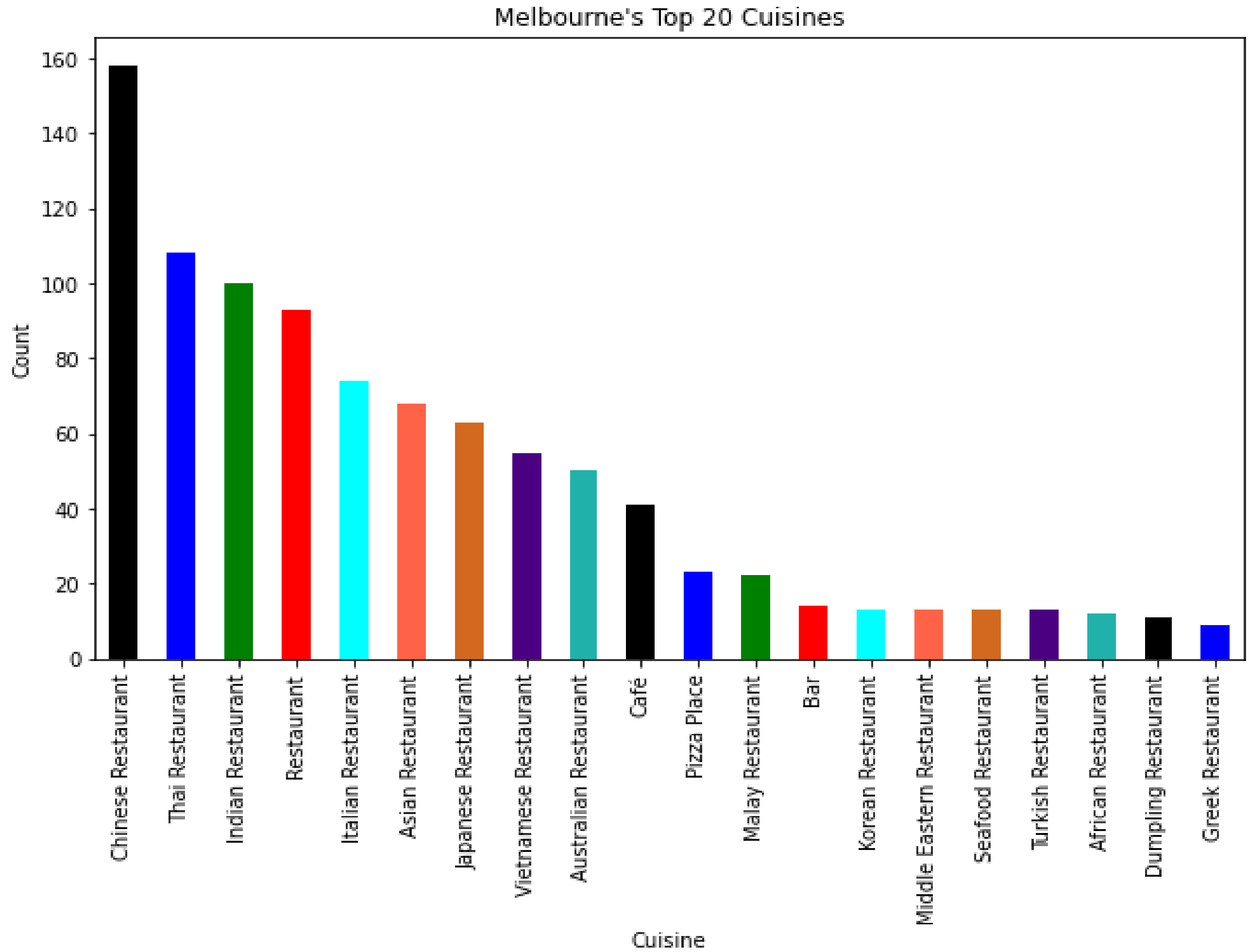

Figure 4: Melbourne's top 20 cuisines with raw data

Similarly, I observed some of the 'Bars & Restaurants' were mentioned as 'Australian Cuisines' where there is no 'Australian Cuisine'. Based on their names I modified them.

After reclassification, I observed a drop in number of unknown cuisines which is shown in the figure below:

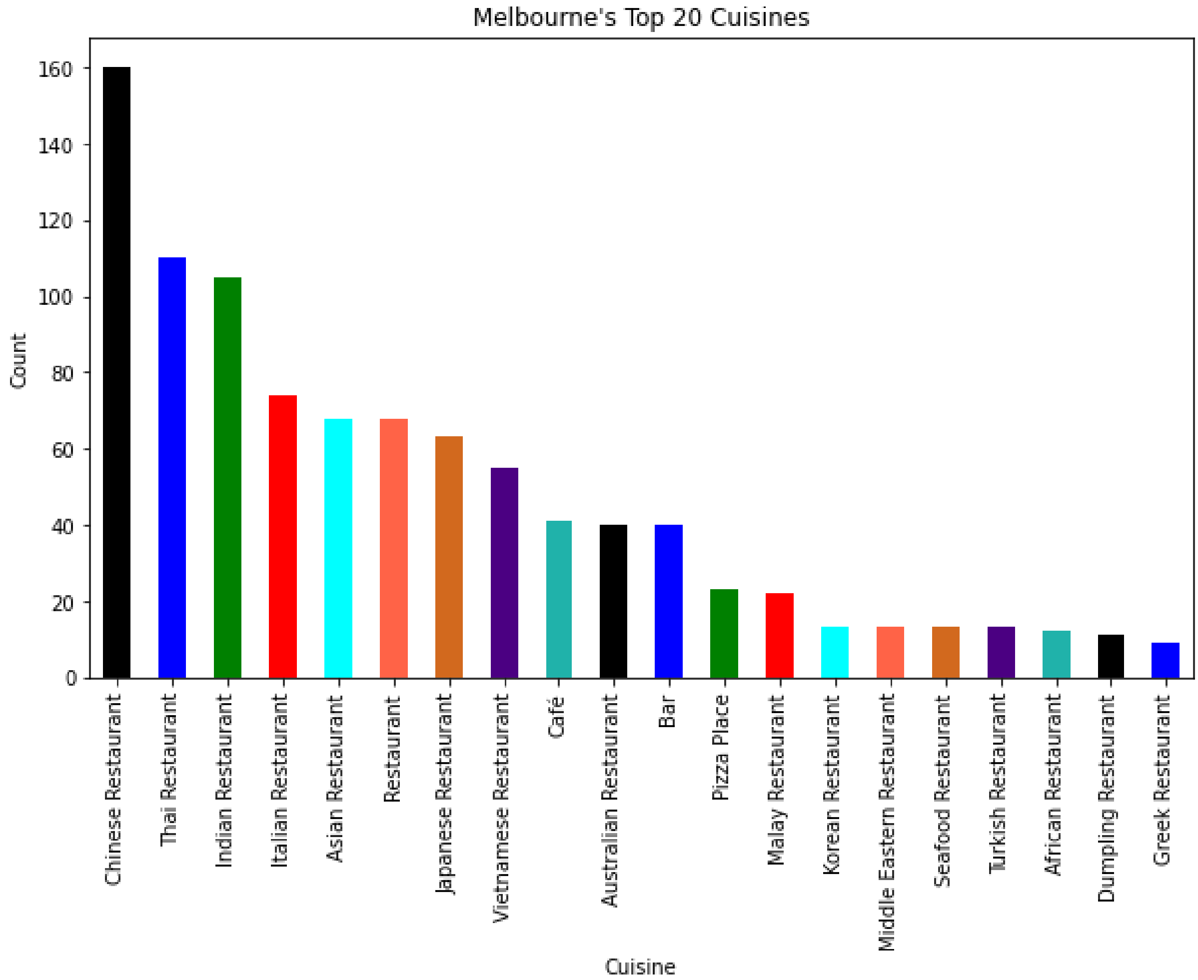

Figure 5: Melbourne's top 20 cuisines after cleaning

3.3 CUISINE ANALYSIS

Also, this project aims to understand what Melburnians are fond of! Using **categories**, a pie chart was visualized for 'cuisines' as below:

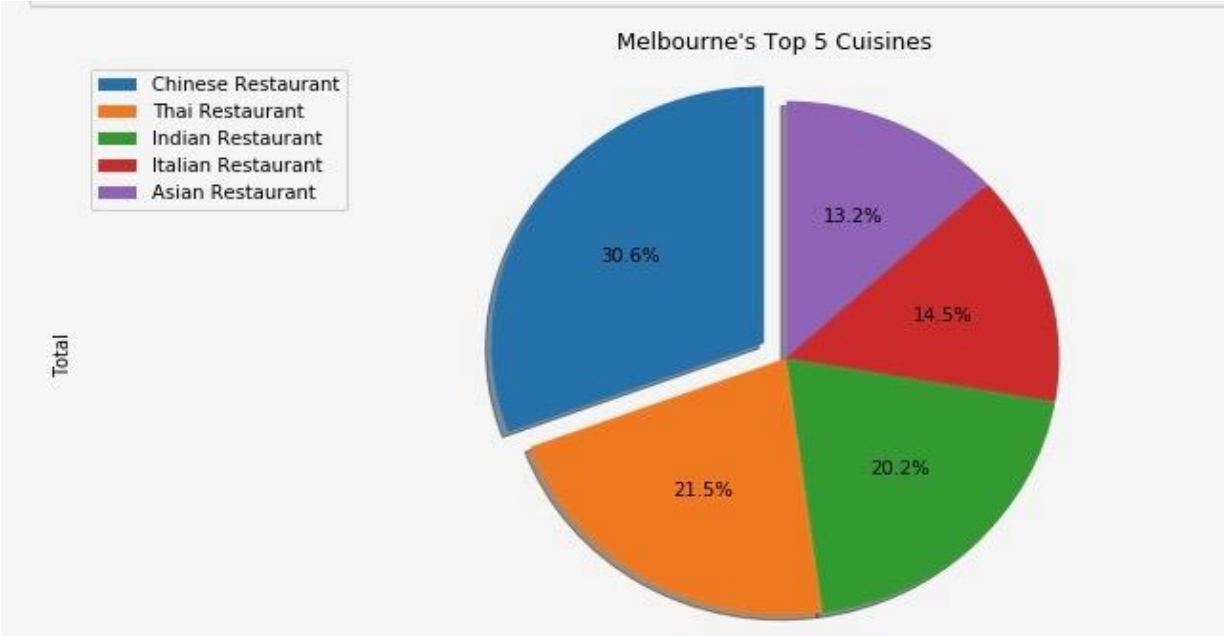

Figure 6: Highlighting top 5 cuisines.

The results states that Melburnians love to eat 'Chinese food' a lot. If we see, in the point of investors its good to start 'Chinese', 'Thai' or 'Indian' restaurants.

3.4 ACQUIRING COORDINATES

The 'Victorian council area coordinates' were brought in and formatted into a data frame from GeoJSON.

All the polygons in GeoJSON file are regenerated with the help of shapely library. Using a basic search algorithm, all restaurants are mapped to their respective council areas as shown below.

	categories	id	lat	lng	name	Councilarea
0	Dim Sum Restaurant	4b05874df964a520318a22e3	-37.812160	144.966569	Dragon Boat Restaurant	CITY OF MELBOURNE
1	Dumpling Restaurant	4b1b3210f964a52074f923e3	-37.811990	144.965210	Chinatown Dumpling Restaurant	CITY OF MELBOURNE
2	Chinese Restaurant	4b05874cf964a520f78922e3	-37.812027	144.966900	West Lake Restaurant	CITY OF MELBOURNE
3	Australian Restaurant	4c10bdba3ce120a1a37c081c	-37.815652	144.966935	Allegro Restaurant	CITY OF MELBOURNE
4	Korean Restaurant	4e77090cd4c0934472942be3	-37.812742	144.960748	Yami Yami Korean & Japanese Restaurant	CITY OF MELBOURNE
...	...	...	...	...	...	...
1159	Café	4d58675a8147f04d2801a254	-38.144340	145.121730	Nature bar	CITY OF FRANKSTON
1160	Asian Restaurant	4f49eac2e4b093da13279669	-38.153490	145.185540	Delight Inn Malaysian & Chinese Restaurant	CITY OF FRANKSTON
1161	Japanese Restaurant	5e60a54ffd0227000703987b	-38.115608	145.240721	Okami Japanese Restaurant	CITY OF CASEY
1162	Italian Restaurant	5718ba24498e88a3a859dce1	-37.849413	144.961717	Torsca italian restaurant	CITY OF PORT PHILLIP
1163	Greek Restaurant	4be546bd910020a12cc6d214	-37.849312	144.967404	starvos restaurant	CITY OF PORT PHILLIP

1164 rows × 6 columns

Figure 7: Data frame showing the council area mapping

For the purpose of visualizing map, centroids for each council area were also generated.

4. MODELING

By now the data is set and ready to perform analysis. There are some common cuisine categories in council areas, so for this reason I have chosen **‘K-Means algorithm’** for clustering the council areas.

K-Means clustering is a simple unsupervised learning algorithm, used for solving clustering problems. It partitions a set of observations into a number of clusters. Mainly it is used in statistics and also can be applied to any branch of study.

4.1 Analysing K-means:

Elbow method helps to choose the optimum value of ‘K’ by fitting the model with a range of values of ‘K’. Range of ‘K’ specifies the sum of squared distances.

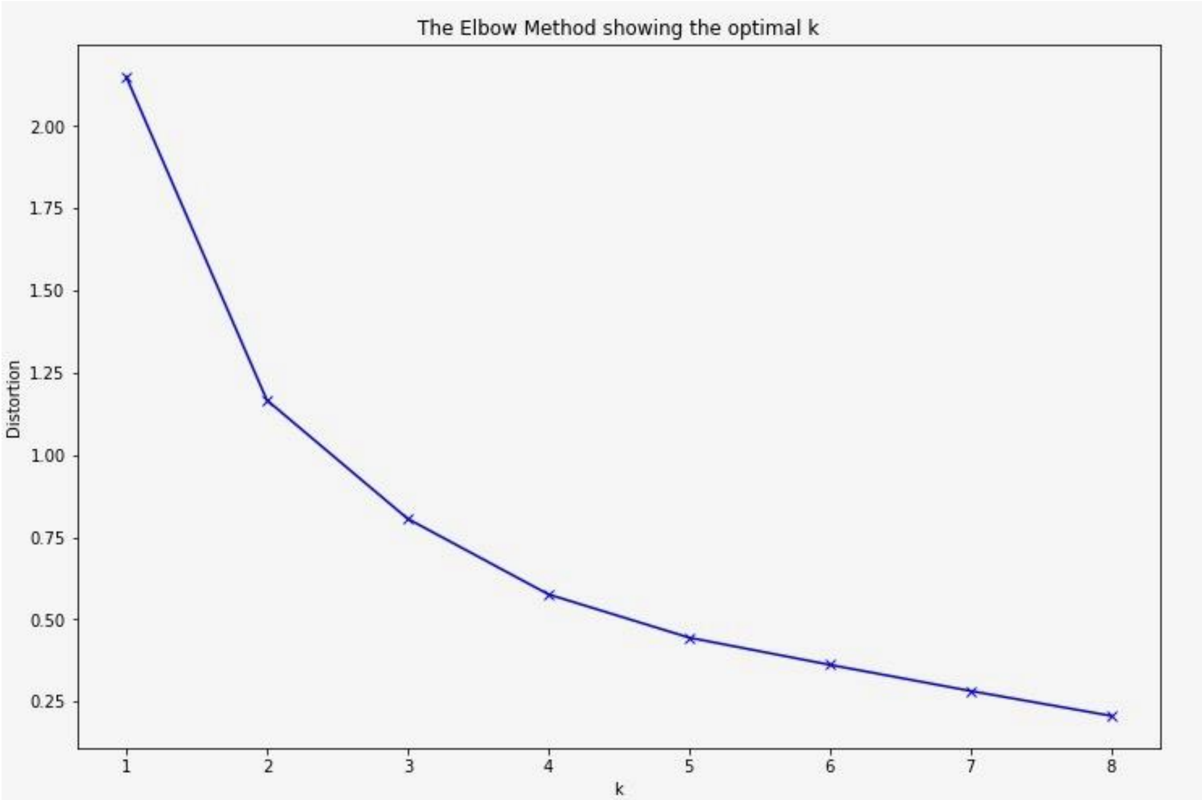

Figure 8: Finding optimum k from Elbow method

Elbow method ensures degree 4 as optimum K. Based on this council areas were clustered into 4.

4.2 COMMON VENUES

After choosing the optimum 'K' all the cluster labels were assigned to their respective councils. Also, the 'most common venues' were estimated for every cluster.

Bar chart was drawn for finding the proper labels for each cluster.

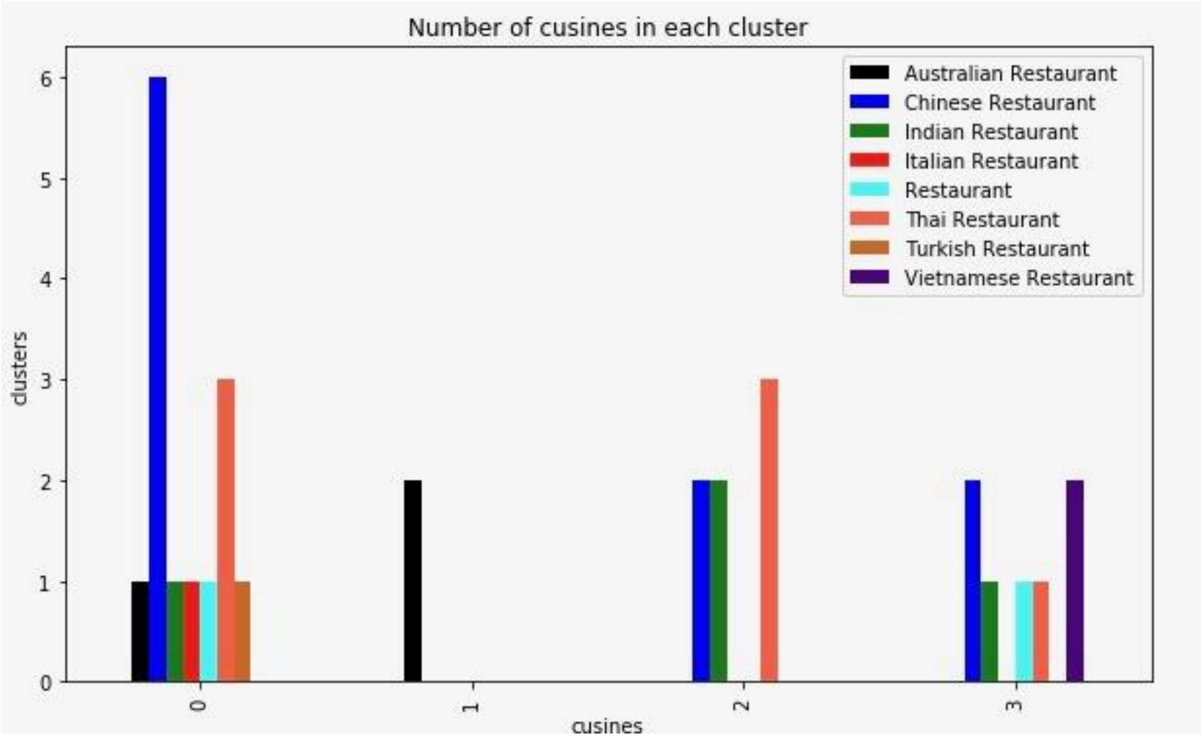

Figure 9: Restaurants in each cluster.

Scrutinizing the above graph, I labelled each cluster as follows for reference:

- Cluster 0: "Chinese Restaurants"
- Cluster 1: "Australian Restaurants"
- Cluster 2: "Thai Restaurants"
- Cluster 3: "Multicultural Restaurants"

As there are more Chinese restaurants in the first cluster, I labelled 'cluster 0' as 'Chinese Restaurants'.

While examining the second cluster, I understood that there were not many restaurants and I labelled 'cluster 1 as 'Australian Restaurants'.

Third cluster as 'cluster 2' and labelled it to 'Thai Restaurants'.

Fourth cluster as 'cluster 3' and labelled it as 'Multicultural Restaurants', because there were many cuisines resided into this cluster.

A keen observation on the ‘Bar graph’ suggests the investors, may find the second cluster (Cluster 1) as a best place for setting up a business because there were only few Restaurants.

4.3 CHOROPLETH MAP

Choropleth map, a thematic map where the distribution on property is shown using different colours. It helps in easy visualizing on how measurements vary across geographic areas.

Visualizing a **Choropleth map** with the below information:

- **Cluster reference.**
- **Top 3 Cuisines.**
- **Number of restaurants**

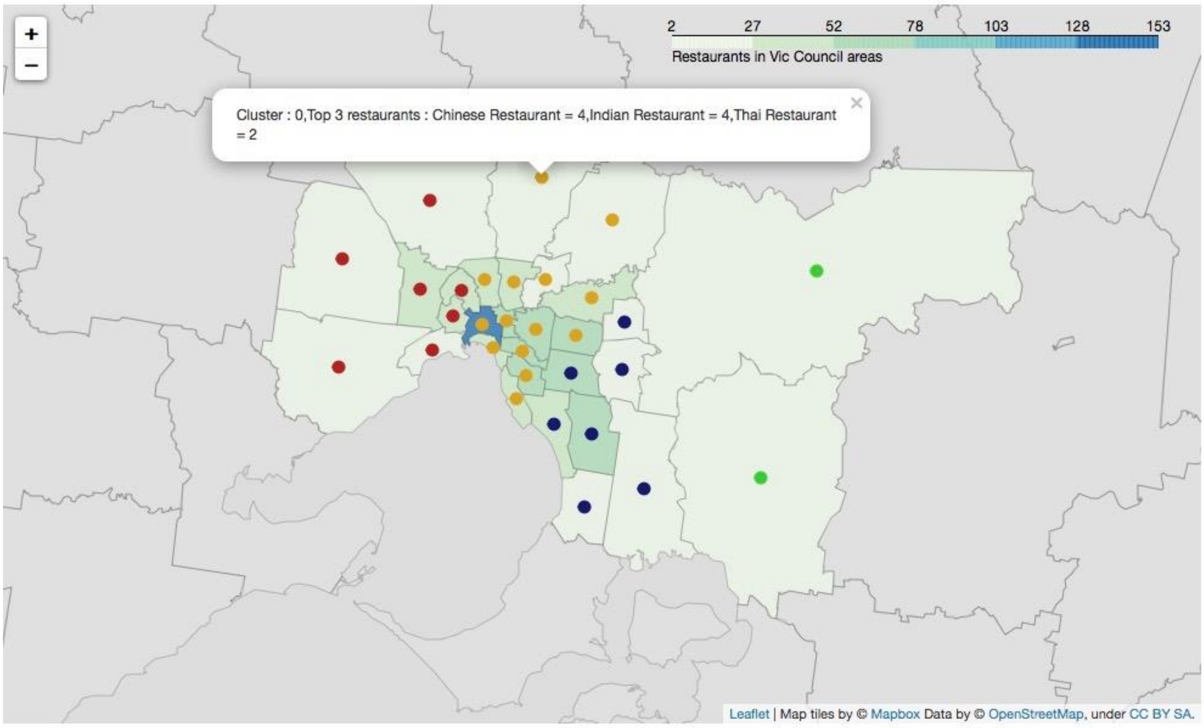

Figure 10: Choropleth map highlighting the clusters and top restaurants.

Looking good! The above **choropleth map** helps the users to understand how many restaurants are there in each **council area.** Also, it indicates what are the top **3 cuisines.** The cluster label indicates the type of cuisine that is more common in that region.

5. RESULTS AND CONCLUSIONS:

This case study was done to find the proper and best locations for investors to set up a new business. All the restaurants data were brought in through **Foursquare API** and restaurants were segregated by using **Victorian council boundaries**.

A lot of wrangling was done on Master data and a clean data frame was set for prediction. On pie charting the data, we analysed the top 5 cuisines. This also helps us to understand which cuisines Melbournites enamour the most.

Using **Elbow** method, we divided the restaurants(categories) into 4 clusters for easy understanding and analysing the top 3 restaurants for every cluster was displayed in the choropleth map. **Folium** was used for visualizing the map, from these results we can say there are many restaurants in down town when compared.

The second cluster says it all! Because there were very few restaurants located which helps the investors to choose a location for their business.

6. FUTURE DIRECTIONS

As a note, we would like to tell the audience! Since we called the data from **Foursquare API,** we got very few restaurants data which is approximately around 1,100. In future if more data is available this case study may provide better and accurate predictions.

Generally, population has a linear relation with business. More people! More restaurants!

So, it is also a one of the major factors to be considered. Substantially, multiculturalism also plays a vital role for FOOD BUSINESS.