1. Introduction

This report explores the idea of openning a new Thai restaurant in Melbourne, a multi-culture costal city with over 5 million population. This report will select a suburb so that the new restaurant will have competitive advantage. First part of the report will discuss data used for analysis and then cluster suburbs by similarity. Among the suburb clusters that favour Thai restaurants, select one with least number of Thai restaurant. This suburb should provide the new restaurant a good chance to succeed.

2. Data

This report will mostly utilise FourSquare data API and public data on the internet. Venue information are obtained via FourSquare API based on coordiates provided by http://www.corra.com.au/ (http://www.corra.com.au/). Suburb information such as postcodes are scrapped off Wikipedia using BeautifulSoup

3. Methodology

After data are cleaned and parsed, there are 246 suburbs with all necessary data. Then data is normalised by calculating venue frequency in each suburb as the table below.

	Suburb	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Antique Shop	...	Vineyard	Volleyball Court	Warehouse Store	Whisky Bar	Wine Bar
0	Abbotsford	0.0	0.010204	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
1	Altona North	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
2	Ardeer	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
3	Armadale	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
4	Arthurs Seat	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
5	Ashwood	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
6	Avondale Heights	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
7	Balaclava	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
8	Balwyn North	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0
9	Bangholme	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.00	...	0.0	0.0	0.00	0.0	0.0

k-clustering is appled to group the suburbs since the aim is to find a group of suburbs that would favor Thai food. k=50 is used to keep the number of suburbs in each cluster low. The list below shows clusters with high frequency of Thai restaurants. The list is sorted by median frequency of Thai restaurants in each cluster. For example, Cluster 1 has median 4.6% of venues are Thai restaurants.

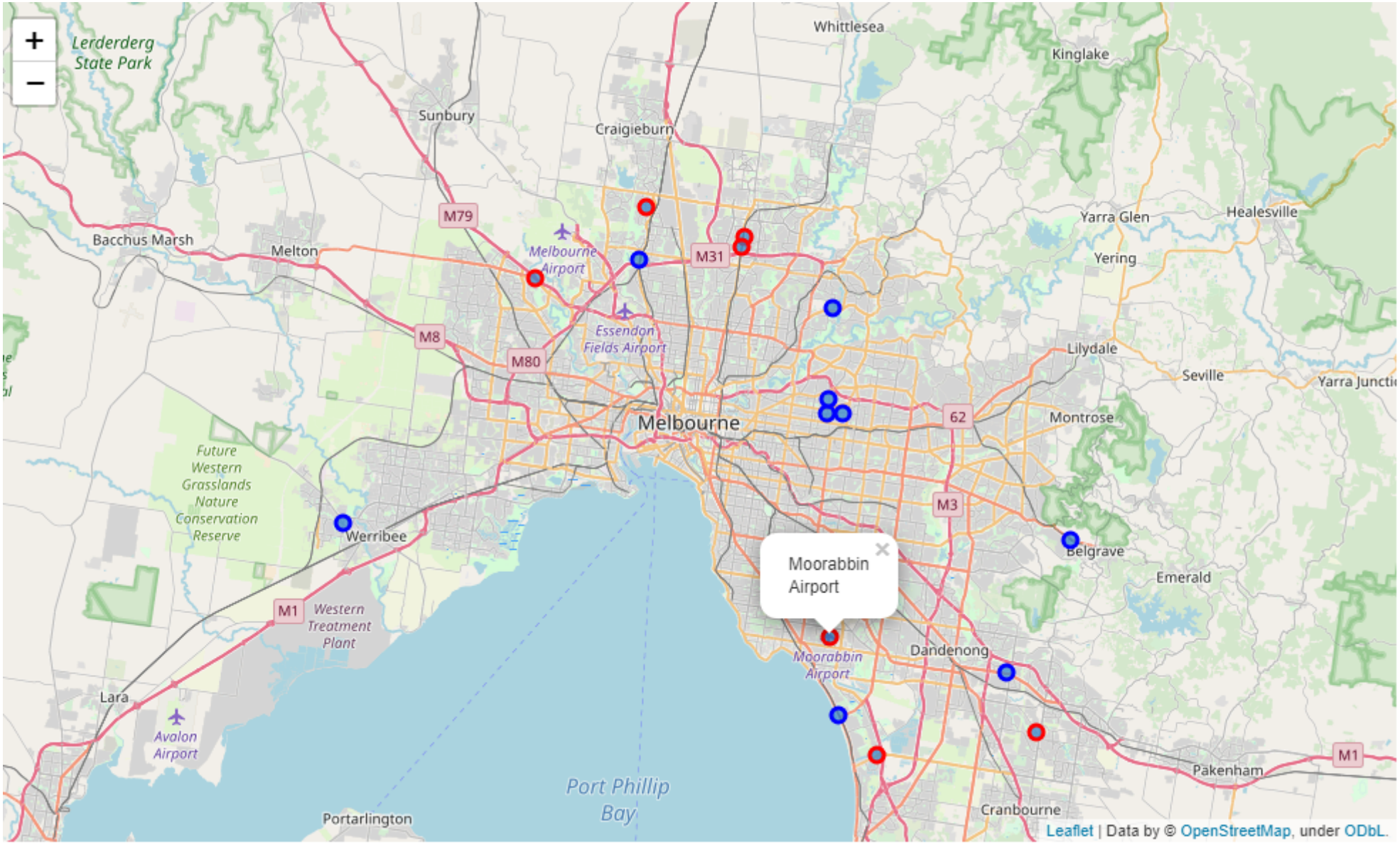
```
mel_grouped.groupby(['Cluster Labels'])['Thai Restaurant'].median().sort_values(ascending=False).head()
```

```
Cluster Labels
1     0.046296
25    0.005000
12    0.000000
22    0.000000
21    0.000000
Name: Thai Restaurant, dtype: float64
```

The table below shows the suburb in cluster 1 along with most common venues in each suburb. It is obvious that this cluster has strong flavour for Asian food.

	Suburb	lat	lon	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	...	92th Most Common Venue	93th Most Common Venue	94th Mos Commo Venu
0	Deepdene	-37.807312	145.096698	1.0	Thai Restaurant	Wine Shop	Japanese Restaurant	Korean Restaurant	Gym / Fitness Center	Gym	...	Deli / Bodega	Dance Studio	Bookstor
1	Edithvale	-38.037415	145.107846	1.0	Beach	Paper / Office Supplies Store	Thai Restaurant	Harbor / Marina	Café	Train Station	...	Bakery	Bar	Baseba Fiel
2	Mambourin	-37.891641	144.629467	1.0	Sandwich Place	Playground	Gym	Thai Restaurant	Pizza Place	Asian Restaurant	...	Bar	Baseball Field	Basketba Cou
3	Jacana	-37.690301	144.915729	1.0	Shopping Mall	Sandwich Place	Grocery Store	Thai Restaurant	Middle Eastern Restaurant	Café	...	Badminton Court	Bagel Shop	Baker
4	Upwey	-37.903672	145.331310	1.0	Convenience Store	Bakery	Restaurant	Thai Restaurant	Supermarket	Train Station	...	Bagel Shop	Bar	Baseba Fiel
5	Yallambie	-37.727482	145.102309	1.0	Grocery Store	Pub	Golf Course	Thai Restaurant	Park	Fish Market	...	Baseball Field	Basketball Court	Basketba Stadiur
6	Hallam	-38.004302	145.269261	1.0	Pub	Restaurant	Fried Chicken Joint	Fast Food Restaurant	Thai Restaurant	Shopping Mall	...	Baby Store	Badminton Court	Bage Sho
7	Box Hill North	-37.807247	145.112050	1.0	Indian Restaurant	Track Stadium	Fish & Chips Shop	Park	Café	Thai Restaurant	...	Badminton Court	Bagel Shop	Baker
8	Fingal	-38.355330	144.905978	1.0	Beach	Fast Food Restaurant	Indian Restaurant	Burger Joint	Thrift / Vintage Store	Bakery	...	History Museum	Food	Deli Bodeg
9	Balwyn North	-37.796947	145.098434	1.0	Asian Restaurant	Shopping Mall	Japanese Restaurant	Grocery Store	Food Court	Fish & Chips Shop	...	Dance Studio	Cupcake Shop	Cultura Cente
10	Lalor	-37.673141	145.016634	1.0	Grocery Store	Supermarket	Asian Restaurant	Fried Chicken Joint	Fish & Chips Shop	Middle Eastern Restaurant	...	Bagel Shop	Bakery	Ba

Next step is to select suburbs in cluster 1 that do not have Thai restaurants so that these suburbs share the same taste as other suburbs in cluster 1 and provide this new restaurant a head start.



The map above marks suburbs in cluster 1. Blue dots indicate suburbs that already have Thai restaurant whereas red dots are suburbs in cluster 1 but have not had Thai restaurant.

4. Result

The cluster 1 are all mid-ring suburbs which are geographically similar so K-Clustering is proved to be an appropriate techinque to be used in this task. It is also reasoable that these suburbs would share similar taste for Thai food. Among the candidate suburbs (red dots on the map), Morrobin Airport provides dense population and location close to CBD compared to other candidate suburbs.

5. Conclusion

In summary, K-clustering is applied on Melbourne suburban data acquired from various sources. After a series of analysis, Moorabbin Airport appears to be a very promissing suburb for a new Thai restaurant to succeed.