

Coursera Capstone

IBM Applied Data Science Capstone

Opening an Italian Restaurant in Melbourne, Australia

By: Matthew Burt



Introduction/Business Problem

I am thinking of opening an Italian restaurant in Melbourne but am unsure where to open it. It would be best to be outside of suburbs with a lot of existing Italian restaurants but also not too far away that people would not think to look for an Italian restaurant in that location. My business problem is to find a suburb adjacent to suburbs with a strong Italian restaurant presence. This suburb needs to have few or no Italian Restaurants and also be reasonably priced. I am planning on purchasing a residential property to convert to a restaurant to give it a more homely feel.

Data Requirements

This report will analyse the distribution of restaurant types in Melbourne's suburbs using Foursquare data. I will then find a list of Melbourne suburbs online from Wikipedia and scrape the data. I will then use geopy.geocoders to find the latitude and longitude. Using all of this data I will find the number of Italian restaurants by suburb, and also the most common restaurant type by suburb and visualize the results using Folium maps. Lastly, I will gather Melbourne's historic sales price data by suburb to analyse.

Methodology

Step 1: Gather all of the required data.

I used Pandas to scrape data from Wikipedia and also to query Foursquare.

I used geocode to convert the address of each suburb into coordinates.

I downloaded sales price data and manipulated it in Microsoft Excel before manipulating it further in Pandas.

Step 2: Cleaning and Exploring the data.

I used Folium and Pandas to explore the initial data.

I used Pandas to clean, group and merge the relevant data sources into one data frame.

Step 3: K Means Clustering

I used Sci-Kit Learns K Means Clustering algorithm to create 5 clusters and assign each suburb to one of these clusters based on Restaurants present and average house sales price.

Step 4: Visualising Results

I used Folium to geographically visualise the clusters.

I used Pandas to visualise the clusters in a tabular format.

Step 5: Conclusions

I used the Folium map and also the tabular data to decide the most appropriate location for my Italian Restaurant. I made my decision based on price and number of Italian Restaurants present in each suburb.

Results

Cluster 0: Average Italian Restaurants for each suburb in this cluster is 0.86 and average price is \$2,104,943.87.

Cluster 1: Average Italian Restaurants for each suburb in this cluster is 0.80 and average price is \$1,067,327.42.

Cluster 2: Average Italian Restaurants for each suburb in this cluster is 1.12 and average price is \$1,332,251.53.

Cluster 3: Average Italian Restaurants for each suburb in this cluster is 0.82 and average price is \$1,647,958.21.

Cluster 4: Average Italian Restaurants for each suburb in this cluster is 1.08 and average price is \$815,137.76.

Ave Italian Restaurants		Average Price	
0	0.86	0	2104943.87
1	0.80	1	1067327.42
2	1.12	2	1332251.53
3	0.82	3	1647958.21
4	1.08	4	815137.76

Discussion

Cluster 1 has the lowest average number of Italian Restaurants per suburb and is the second cheapest cluster by historic sales price.

Cluster 4 is the cheapest by historic sales price but has a relatively high average number of Italian Restaurants at 1.08.

Cluster 0 is the most expensive cluster with an average suburb sales price of \$2,104,943.87.

Cluster 3 has a relatively low number of restaurants in general and a high average sales price of \$1,647,958.21

Conclusion

Cluster 1 is the most appropriate cluster to choose the final suburb for my Italian restaurant from. This is due to the low number of existing Italian Restaurants in this cluster and its relatively low sales price.

In cluster 1, the suburb with the most Italian restaurants is Abbotsford, with 14. The two suburbs that are adjacent to Abbotsford are Richmond and Burnley. Richmond has 5 Italian Restaurants and Burnley has none.

```
In [239]: joined_df.loc[joined_df['Cluster'] == 1][['Suburb', 'Italian Restaurant', 'Price']]
```

```
Out[239]:
```

	Suburb	Italian Restaurant	Price
25	Cheltenham	0.00	969127.39
26	Gardenvale	0.00	1095312.50
169	Gardenvale	0.00	1095312.50
28	Hampton East	0.00	1096513.04
29	Highett	0.00	1083245.06
30	Moorabbin	0.00	1009232.99
162	Bentleigh East	0.00	1132489.32
164	Caulfield East	0.00	1035750.00
165	Caulfield North	0.00	958154.49
172	Murrumbeena	0.00	1061576.90
173	Ormond	0.00	1096426.63
295	South Yarra	0.00	1078120.07
296	West Melbourne	1.00	1041723.08
429	Elwood	0.00	1011153.37
439	Prahran	0.00	1149590.55
441	Windsor	0.00	1051843.75
478	Abbotsford	10.00	1040200.70
479	Burnley	0.00	1191875.62
482	Cremorne	0.00	1076294.12
486	Richmond	5.00	1072606.23

Burnley is the most appropriate suburb for my new Italian restaurant.