2020

Where would I live in Melbourne?

Background:

Moving into a new city is difficult. Everything is different from home. Hence, it is almost essential that you find a good place to live. Home is where one feels safe and protected and is in their own space. Cities develop at quick rates with new properties propping up every single day. It is a tedious process to look for a house, checking each neighbourhood for all its amenities, evaluating each home for its price, checking proximity to essential services and much more.
Since people constantly move from one place to another this is an essential problem to solve.

Problem:

Moving into a new Melbourne is difficult. It is a big city with over 400 neighbourhoods. This study aims to evaluate the top 100 most expensive neighbourhoods and provide a list of ideal neighbourhoods for a potential buyer. The study looks at the average price for 2-bedroom houses in each neighbourhood. It looks at services such as number of medical centres, number of shopping malls, number of metro stations and much more. Through these numbers it attempts to cluster each neighbourhood and provide an insight into an ideal cluster of neighbourhoods to live in.

Data Acquisition:

The data needed for this study is:

- The list of neighbourhoods in Melbourne
- The price of each 2-bedroom house in each neighbourhood in Melbourne
- The number of Restaurants, Metro Stations, Shopping Malls, Medical Centres, Arts and Entertainment venues, Clothing Stores, Bars, Sports Venues and Grocery Stores.
- The latitude and longitude of each neighbourhood in Melbourne.

Data which was used for this analysis was found on GitHub and by the use of the Foursquare API.
The final dataset looked like this:

	Restaurants	Arts and Entertainment	Bars	Sports Venues	Medical Centers	Clothing Stores	Grocery Stores	Shopping Malls	Metro Stations	Neighborhood	Price
0	50	48	50	44	49	50	17	2	1	Albert Park	1.543793e+06
1	50	48	50	43	48	49	12	2	1	East Melbourne	1.348500e+06
2	49	36	8	30	20	12	8	1	0	Princes Hill	1.317083e+06
3	0	0	0	0	0	0	0	0	0	Ashburton	1.316429e+06
4	50	48	46	41	48	49	14	2	0	Kooyong	1.290000e+06
...	...	...	...	...	...	...	...	...	...	...	...
95	50	42	7	24	48	11	7	1	1	Clayton	7.530000e+05
96	50	43	49	47	45	47	22	3	0	Elsternwick	7.505200e+05
97	50	32	7	21	16	6	7	1	1	Kensington	7.484793e+05
98	14	0	3	2	1	4	0	0	0	Burwood East	7.460000e+05
99	50	43	49	47	45	47	22	3	0	Ripponlea	7.453889e+05

100 rows × 11 columns

Methodology:

Housing data retrieved from GitHub was processed to find the top 100 highest average priced neighbourhoods. The search query of the FourSquare API was then used to find the number of target variables in each neighbourhood. This data was then merged into a DataFrame along with the price.

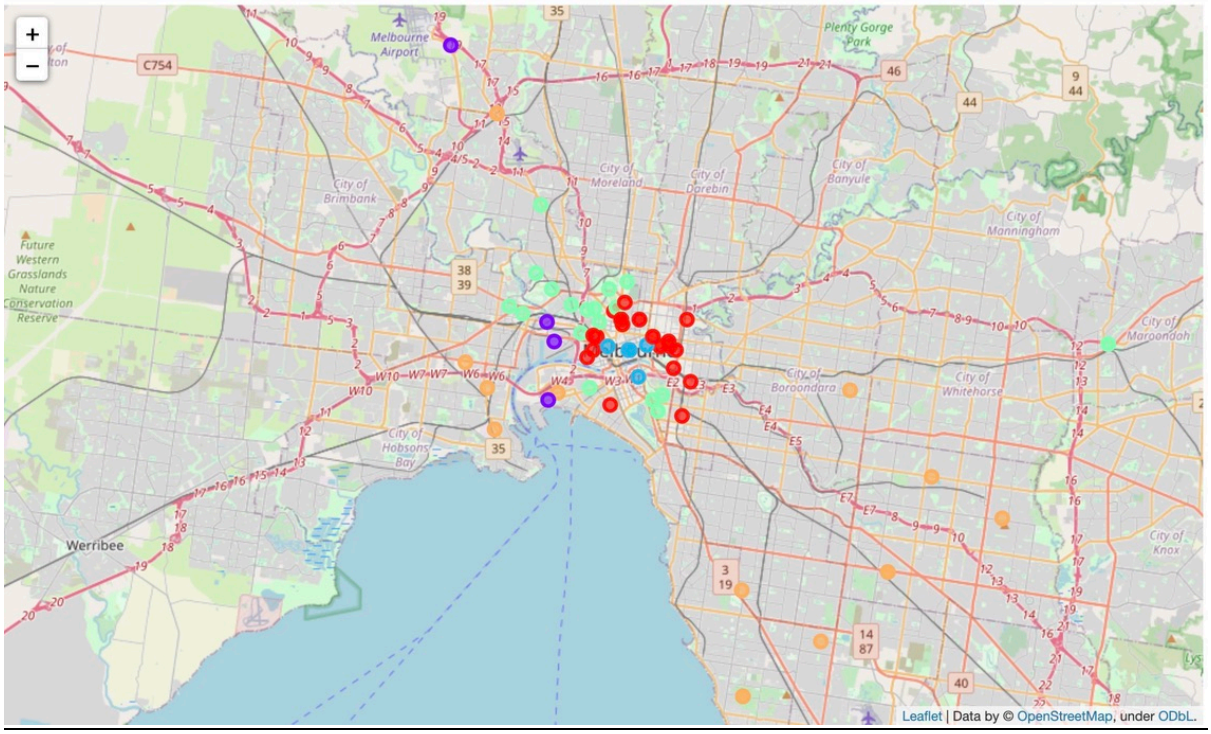
The resulting DataFrame was then clustered using K-Means Clustering. The study of the resulting clusters is essential to find the best set of neighbourhoods suited to an individual. By understanding the number of target variables in each cluster an individual would be able to make an informed decision to select a neighbourhood.

For example, an individual would want a high density of Medical venues around them in case of a medical emergency. By clustering we would be able gain neighbourhoods which have the highest concentrations of medical centres providing key information.

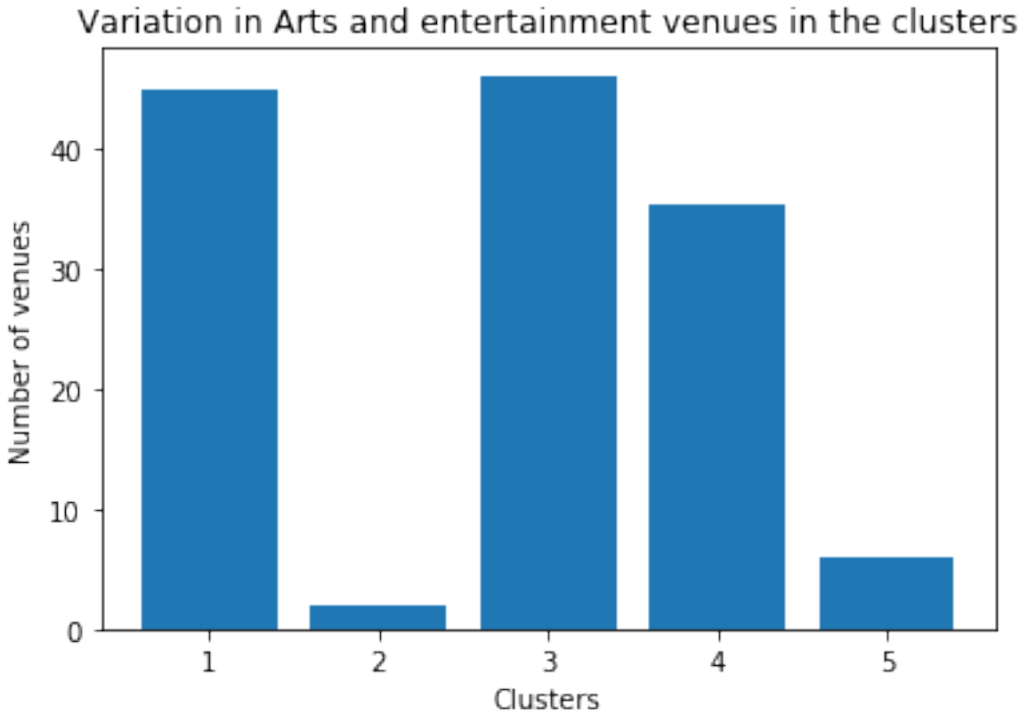
The number of clusters was set to 5 which would provide a critical insight into the neighbourhoods.

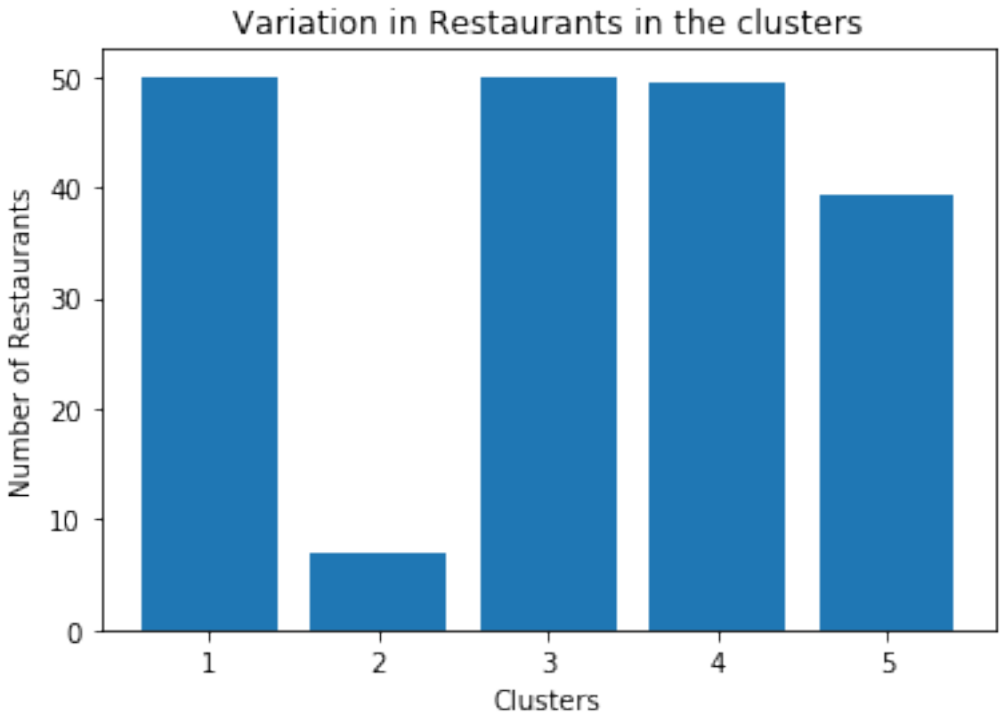
Results:

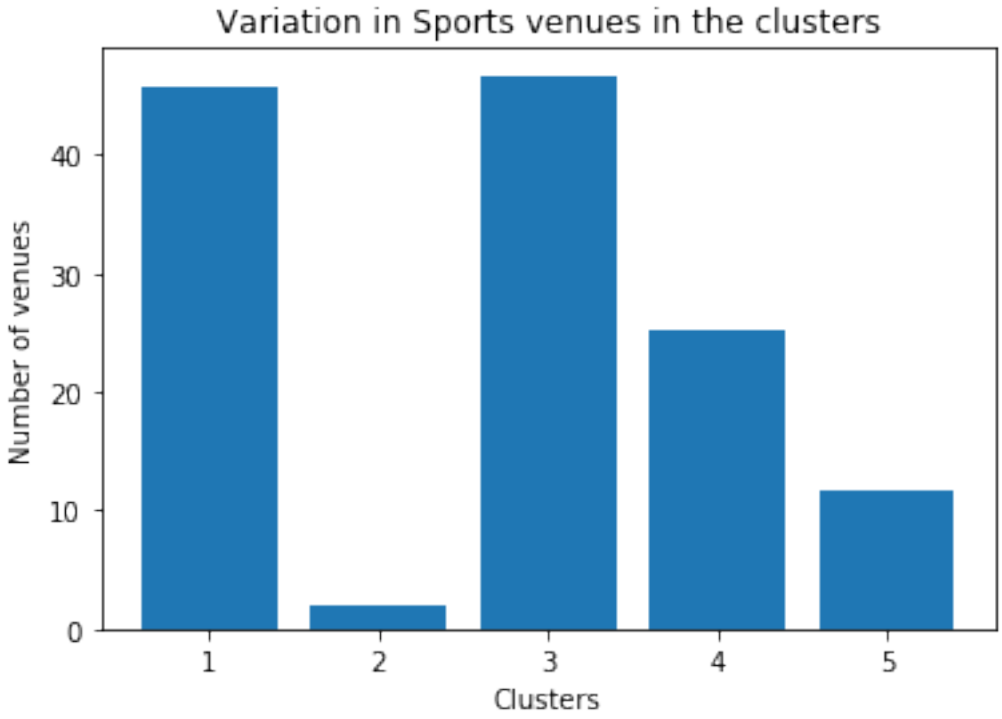
From clustering the neighbourhoods, the following results were obtained.



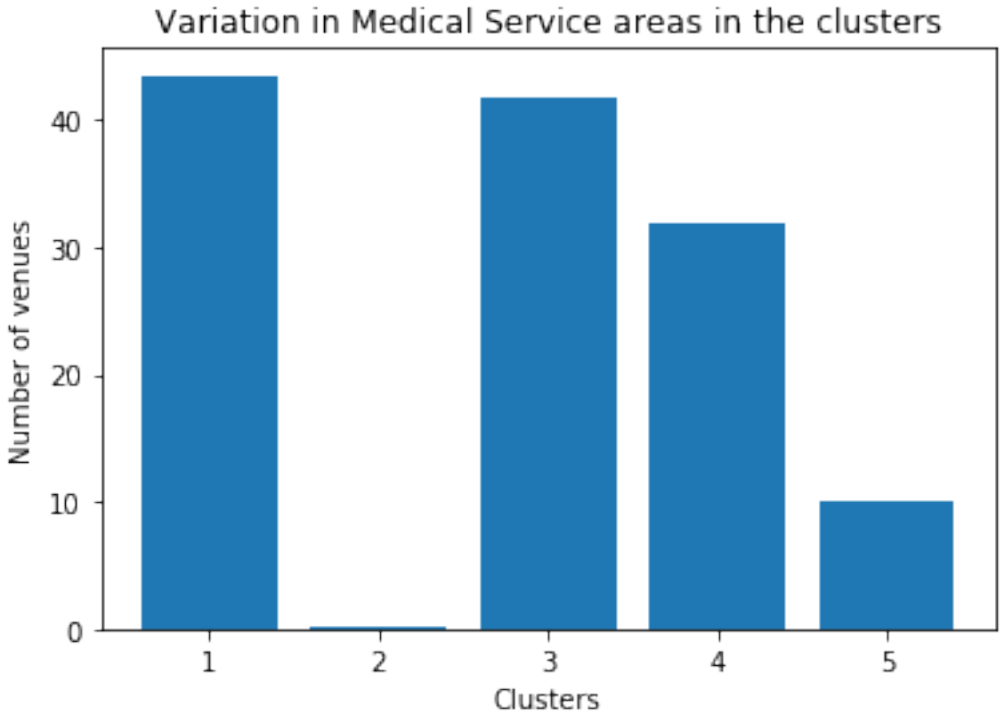
The average of the number of target variables from each cluster was calculated and plotted on a bar chart to see how they varied.

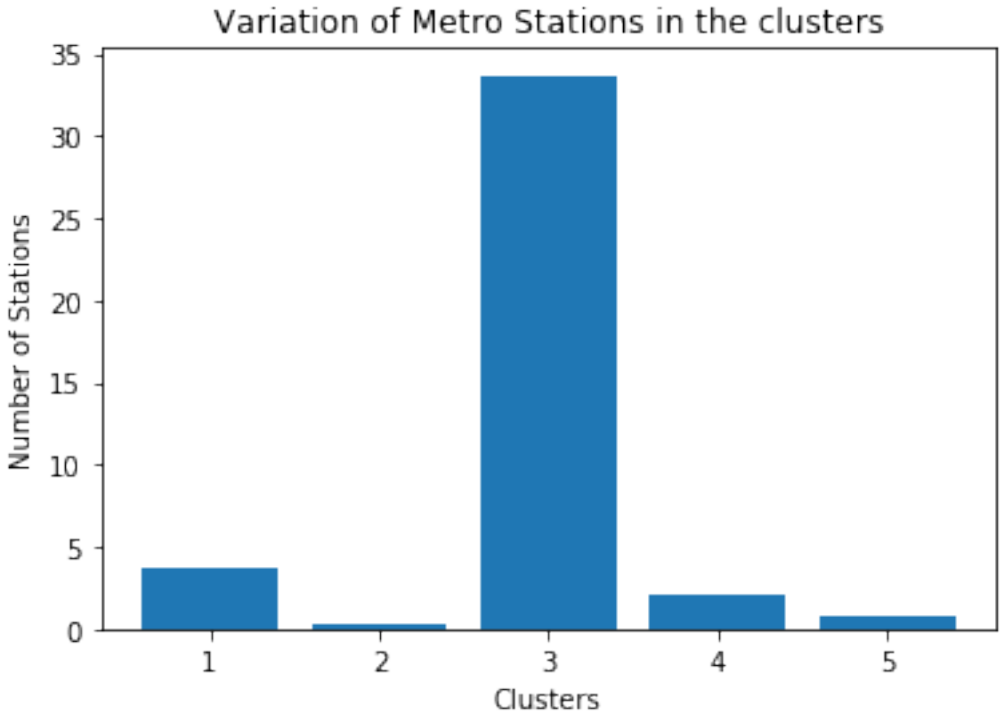

Variation in Arts and entertainment venues in the clusters
Number of venues
0
10
20
30
40
1
2
3
4
5
Clusters

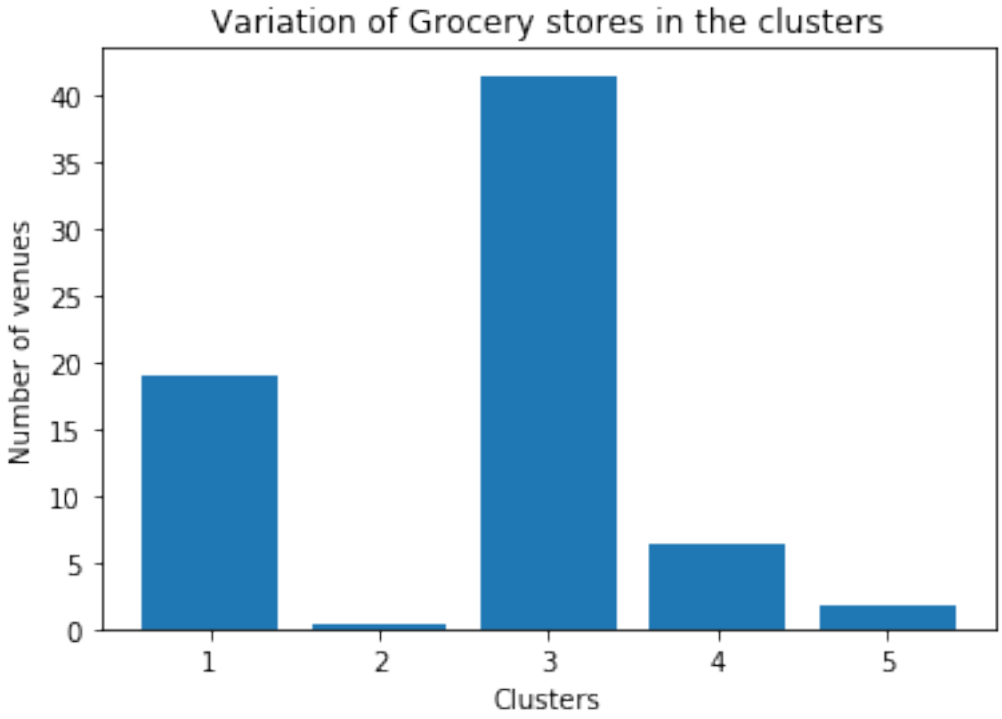

Variation in Restaurants in the clusters
Number of Restaurants
0
10
20
30
40
50
1
2
3
4
5
Clusters

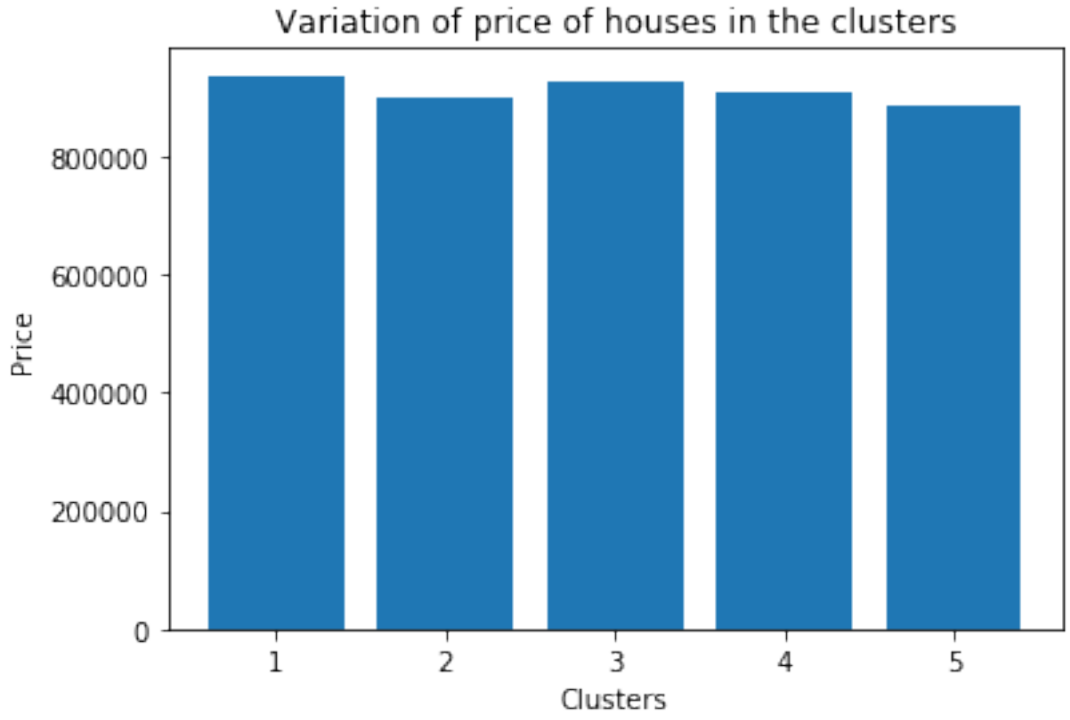

Variation in Sports venues in the clusters
Number of venues
0
10
20
30
40
1
2
3
4
5
Clusters


Variation of Clothing stores in the clusters
Number of venues
0
10
20
30
40
50
1
2
3
4
5
Clusters


Variation in Medical Service areas in the clusters
Number of venues
0
10
20
30
40
1
2
3
4
5
Clusters


Variation of Metro Stations in the clusters
Number of Stations
0
5
10
15
20
25
30
35
1
2
3
4
5
Clusters


Variation of Grocery stores in the clusters
Number of venues
0
5
10
15
20
25
30
35
40
1
2
3
4
5
Clusters


Variation of price of houses in the clusters
Price
0
200000
400000
600000
800000
1
2
3
4
5
Clusters

Results Analysis:

It is interesting to note that the average price of each neighbourhood is almost equal with little variation. However, the average count of all the target variables in the clusters show distinct variation.

Cluster 2 has a low average of all the target variables making it less ideal for a prospective buyer. However, cluster 3 has a high average number of all the target variables making it suitable to move into.

Clusters 1,4 and 5 follow a pattern where cluster 4 leads the average number of target variables followed by clusters 1 and 5.

From the variations in the graphs above, it can be seen that the clusters have the similar number of certain target variables such as Arts and Entertainment venues and Restaurants which suggests that these venues are spread out and are not limited.

However, other features such as variation of Sports venues are more dissimilar and makes each cluster different.

Where would I choose to live?

I would choose to live in any neighbourhoods in cluster 4 as they have the right proportion of features needed for a smooth life.
To explain further, Cluster 3 has a high average number of every feature; this suggests a lot of buildings in close proximity and a cluttered sky.
This can be seen on the map and cluster 3 is located in the centre of Melbourne.

Cluster 2 on the other hand, has a very low number of features and it wouldn't be ideal to live there.

Hence, only clusters 1,4 and 5 are left. Personally, I would choose a higher number of feature variables near me, since cluster 4 provides that and the average price across clusters doesn't vary, I would choose cluster 4.

However, clusters 1 and 5 are suitable alternatives to people with different opinions of what they would like in their neighbourhood.

How can this project be improved?

Instead of taking the absolute number of feature variables in each area, if we took the density of each variable in neighbourhoods it would be a better metric. This is because even though two neighbourhoods might have the same number of restaurants, they might differ in size making one neighbourhood denser with restaurants than another. The buyer would be able to better gauge which is the most dominant feature variable in each cluster.

The rent prices used in this study are from 2018, if more up-to-date prices were used the study would be more accurate as the top 100 most expensive neighbourhoods would have changed, providing different results.

Conclusion:

The project was immensely helpful in applying the data science skills I had learnt through this professional certificate. The project allowed me to explore my curiosity in finding a new home in Melbourne.
From the coding difficulties to the difficulties of finding data it has been a challenging and great experience!