THE BATTLE OF NEIGHBORHOODS

Choosing accommodation in Melbourne

Data Science Capstone Project

MAY 2020

Ruben de las Heras

This page was left blank intentionally

INDEX

1. INTRODUCTION

Australia is a popular destination for travelers around the world. According to Austrade, the country recorded more than **8.5 million visitors** in 2019[1].

It is easy to understand why: unique wildlife and some of the most beautiful landscapes you will ever see. Not to mention some of the most extraordinary places in the world like the Great Barrier Reef.

For some people, spending a couple of weeks or a month is not enough. There is some much to see, to do, to experience that in 15-20 days you definitely cannot see everything. In that regard, a good way to enjoy the *'full Aussie experience'* is by living in the country. A very handy and common choice is the **Working Holiday Visa**, which allows you to live and travel for a year, with a potential extension of an additional one.

As the Department of Home Affairs states, the number of Working Holiday Visa makers in 2019 was **198,293**[2]. Out of that figure, the Australian Bureau of Statistics indicates that 17% chooses Melbourne as a destination or starting point[3], making a good **33,701** people.

Why Melbourne? It has been selected for several years in a row as the most liveable city in the world. Clean, vibrant, dynamic, multicultural, and many other good characteristics make Melbourne a great place to live.

Moving to another country is such a great experience but like everything in life, it has its difficult part. One of the biggest concerns is finding accommodation. The aim of this report of this study is **to help WHM makers, or anyone interested, providing information about rental prices and venues** in the main suburbs in Melbourne.

2. DATA

In this study, I used data from different sources, as indicated below:

- **Melbourne:** I found a Melbourne geojson on GitHub which covered a wide area. I modified it reducing the number of suburbs to leave just the ones close to the city center. From there I extracted the list of suburbs.[4][5]

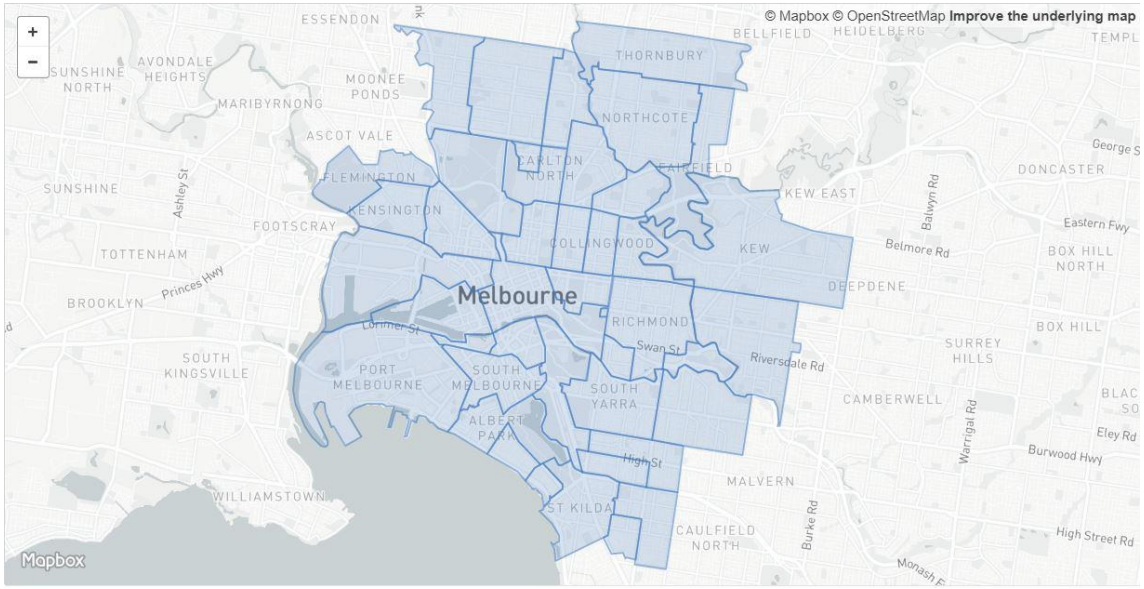
- **Rental prices**: from the Victorian Department of Health and Human Services I got a rental report in 2019[6] in which among other figures I was able to extract the average price per room in every suburb of my concern.

- **Coordinates:** used Geopy library.

- **Venues:** utilized FourSquare to get the most common venues in every suburb.

3. METHODOLOGY

Firstly, I worked on the original Melbourne geojson. The area that covered was a really good one actually but as I mentioned before, I wanted the suburbs close to the city centre. Those suburbs bring more options regarding bars, shops, and entertainment in general, which I believe is an important factor for travellers. Besides, the city has a vast amount of suburbs, which could lead to work on a lot of information which after will not be very useful at all. Therefore, I also reduced the size of the sample as it was pretty big. This is the modified one in which I have based my study on:



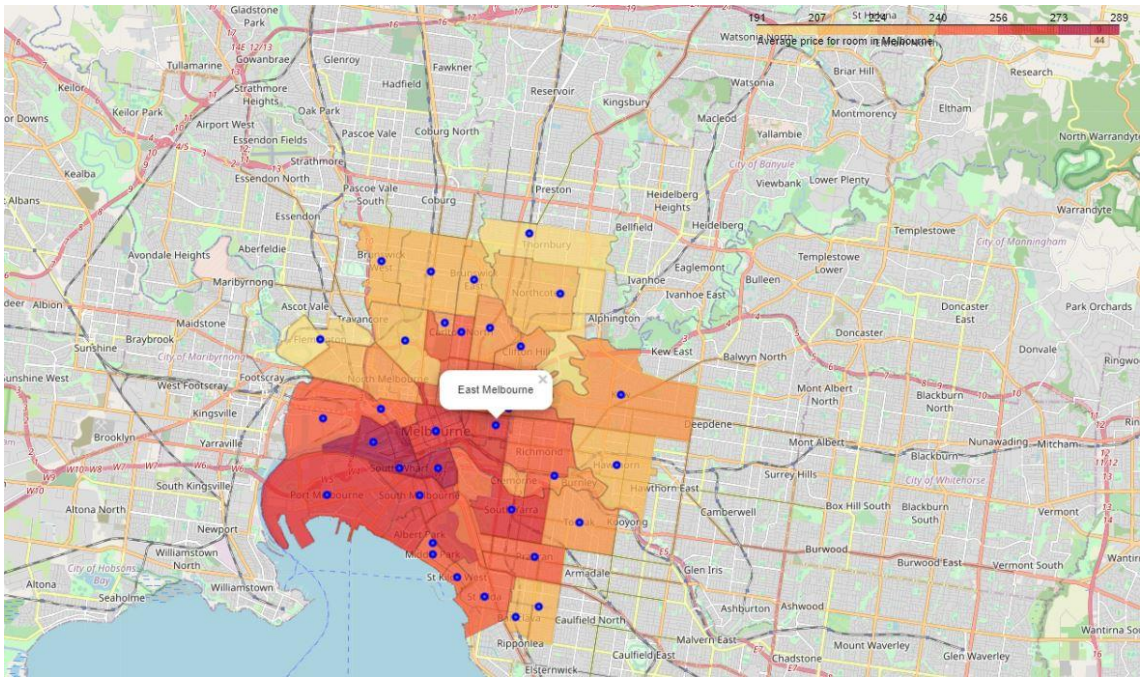
Secondly, I obtained the values of rental prices. In the Table from Rental Report, I considered 3-bed houses, as it is the more common way to share accommodation. Dividing the full price between 3 I got the average rental price per room per week. Then, I created the dataframe on the notebook:

	name	price
0	Balaclava	216
1	Windsor	244
2	Fairfield	197
3	Cremorne	238
4	Fitzroy North	238
5	Hawthorn	216

Having the name of the suburbs, I created a loop to make a call in Geopy to obtain the geographical coordinates of each suburb:

	suburb	price	lat	long
0	Balaclava	216	-37.869921	144.993428
1	Windsor	244	-33.605534	150.821953
2	Fairfield	197	-33.867440	150.889310
3	Cremorne	238	-33.826008	151.225433
4	Fitzroy North	238	-37.783332	144.983707
5	Hawthorn	216	-37.824425	145.031721

With all that data I was able to have the first glance on a map. I used a choropleth map as it provides a more visual way to see the difference in prices between suburbs. Every suburb is marked with a blue circle.



Then I started using the FourSquare API to make different calls to obtain the venues on each suburb, the most common ones, and group them into categories.

As I had common categories in every suburb, I used the K-means algorithm to cluster them. K-means is a widely utilized method to cluster. In my case, I ran 5 clusters.

With K-means, the dataframe looked more organized and easy to understand:

	suburb	price	lat	long	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Balaclava	216	-37.869921	144.993428	0	Café	Coffee Shop	Bar
1	Windsor	244	-33.605534	150.821953	3	Café	Shopping Mall	Grocery Store
2	Fairfield	197	-33.867440	150.889310	0	Grocery Store	Park	Pub
3	Cremorne	238	-33.826008	151.225433	3	Café	Cocktail Bar	Bar
4	Fitzroy North	238	-37.783332	144.983707	3	Coffee Shop	Pub	Park

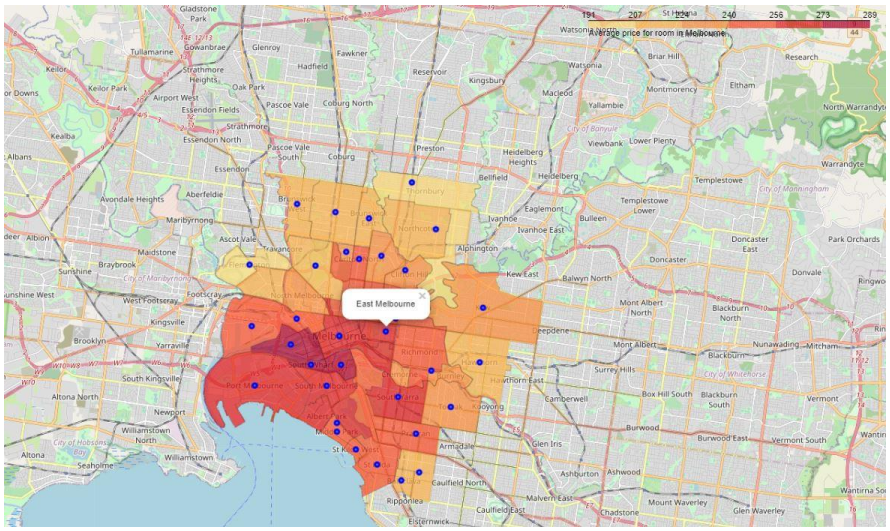
From there, it was just a matter of analyzing each cluster individually to, later on, create a map to see the results.

4. RESULTS

With the first piece of information, I created a horizontal bar chart to check how the rental prices were distributed:

As we can see, the average price is around **$240** while the distribution is pretty equal. It is easy to find prices both over and under the average.

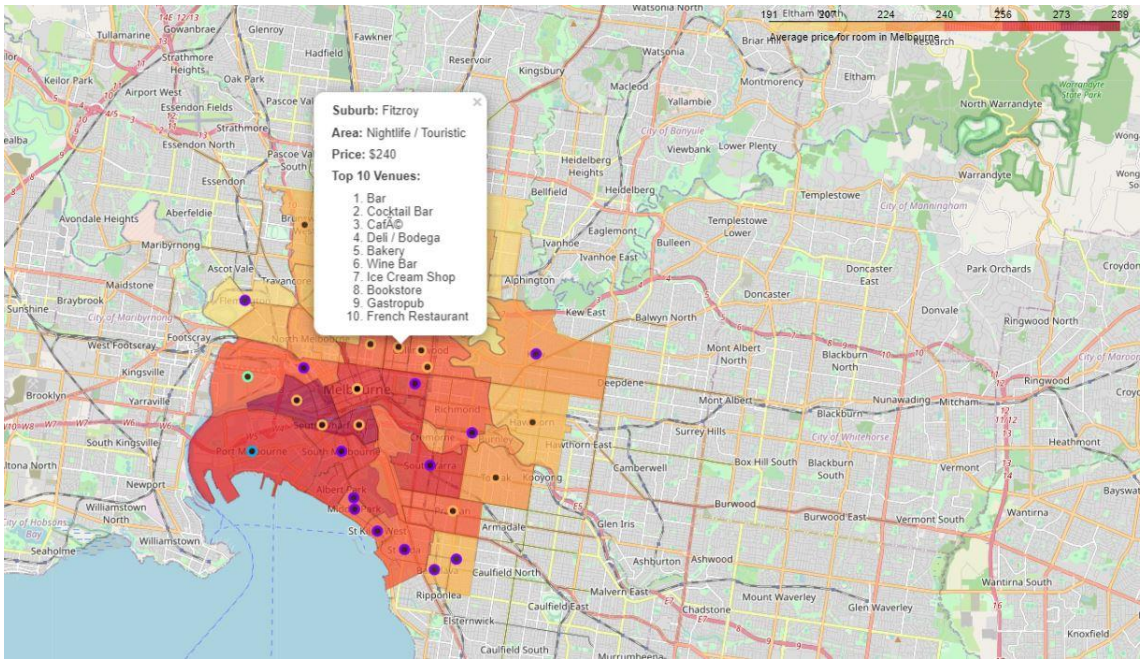
Transferring that information to a more visual way, I created the next choropleth map, so it is easier to see each suburb's prices.



FourSquare returned a total of **200 unique categories** with **1066 results**.

The final result is shown in the below map. It displays all the information I thought it would be useful at the beginning of this study, being that:

- Suburb
- Type of area: nightlife, residential, commercial...
- Price: average rental price
- Top 10 venues



5. CONCLUSION

Melbourne is a big city, way bigger than the area used for this study. That means that rental prices usually are cheaper the further you get from the city center. Public transportation is pretty good which helps to live in areas that are not close to downtown.

However, I thought it would be a better idea to focus the study on the suburbs around the CBD as they have a wider variety of venues and they also have a higher number of travelers / young people, which I think is important while settling in a new country.

From there, I worked on the geojson file which allowed me to understand how to use and modify that format. It also was the platform base from where I started creating the different dataframes. The geojson file can be found on my GitHub profile as well as the original one (check references).

Regarding the K-means algorithm, I decided to go for 5 clusters as I wanted to have a wider variety of areas to label on at the end. Perhaps it is too much and I could have ended up using 3, but I preferred to go higher than shorter to give a more depth insight.

I used data visualization tools throughout the entire process as they are very useful tools to understand the data. The final result can be seen in the latest map.

6. REFERENCES

[1] Latest International Visitors Survey - Austrade

[2] Working Holiday Report 2019 - DHA

[3] Insights from the Australian Census and Temporary Entrants - ABoS

[4] Original Melbourne geojson - GitHub

[5] Modified Melbourne geojson - GitHub

[6] Tables from Rental Report - VDHHS