

Airbnb – Three Business Questions Understanding

Introduction

You must have heard of sharing economy before. Sharing economy refers to the sharing of goods or other resources by multiple people. One of the most popular sharing platforms we are enjoying in the recent years is Airbnb. Airbnb is a platform which offers you someone's home as a place to stay instead of a hotel. In this article, I am going to leverage Airbnb's listings data in Sydney, one of the most popular and international cities around the world, and the Census data to roughly get a broad view of local lodging marketplace. In addition to this, the whole project will follow up the structure including:

- Business Understanding
- Data Collecting and Understanding
- Business Questions & Answers

It is noted that this structure refers to CRISP-DM (**C**Ross **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining) methodology, while modelling, evaluation and deployment phases have been discarded and details about data preparation will not be mentioned in this paper.

Business Understanding

We cannot deny that Airbnb has completely changed how most people travel, and it brought a totally new level of comfort and intimacy that goes beyond the hotel experience. Generally speaking, most of reports prefers to provide suggestions from customer sides, such as pricing, positioning and so on in order to enjoy their travel with least cost in lodging. However, this blog mainly targets at hosts as the stakeholders. I was hoping to help hosts in the Sydney to promote their housing services, draw views on local sharing lodging market and make recommendations to potential ones who are hesitating whether join the platform.

At the same time, as the one the most popular and international cities, Sydney yearly attracts over millions of visitors from the whole world. However, as we all know, the pandemic of COVID-19 has changed the game a lot. In this article, we are going to focus on leveraging datasets acquired in the recent pandemic period in Sydney.

Data Collecting and Understanding

Regarding the utility of data, I extracted the raw datasets from 4 public resources:

1. Airbnb Listings Data in Sydney in July 2021

The raw data for Airbnb listings are originally downloaded from "Inside Airbnb" website. For the purpose of answering my business questions about the current situation, I only extracted the dataset in the Greater Sydney, in July 2021. The raw data is stored in a CSV (i.e., comma-separated value) format which contains numerous details about each house recorded, ranging from the information about the listing to the host, from reviews to price and properties, etc.

2. Sydney LGA information

LGA stands for local government area in Sydney, sorted by region. The file labelled as Geography Information for Local Government Area (LGA), in which we could get the values of LGA's code and name pair.

3. Sydney 2016 Census - 2016Census_G02_NSW_LGA.csv

The Census 2016 data is collected from Australian Bureau of Statistics. The CSV file downloaded here stands for G02 Selected Medians and Averages in which the Australia government calculates both the median and average values in rent payment and mortgage repayments, etc in an area.

4. Geospatial data for Australia - 1270055003_lga_2020_aust_shp

Geospatial data pertains to an object on Earth, which combines location, size, and shape of an object or element information. In this article, the Australian geospatial vector data is stored in a shapefile format and is zoomed into the Greater Sydney.

Business Questions & Answers

1. Which LGAs attract more rentals? Among these LGAs, which type of room attracts more rentals? (In 30 days)

For the first part of this business question, I would like to explore in which place that visitors prefer to renting a house in Sydney. I was assuming that:

1. the LGAs with more famous tourist attractions
2. the LGAs around the city
3. the LGAs in eastern side of the Greater Sydney

are more likely to get visitors.

The Percentage of Listings Having been Rent in 30 Days in LGAs [%]

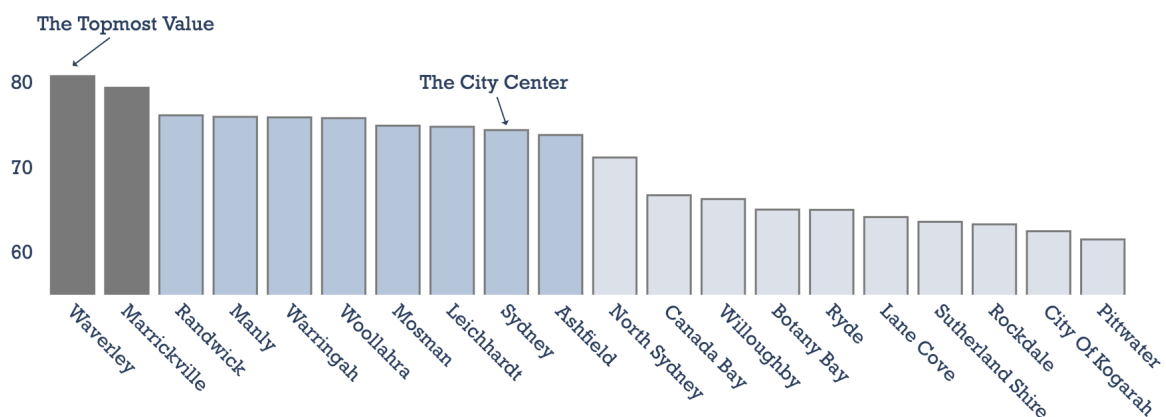


Figure 1. The percentage of listings having been rent in 30 days in LGAs

To verify my assumption, a bar chart has been drawn to expose the percentage of Listings Having Been Rent in 30 days. The values have been sorted from the highest to lowest but in top 20, and then I split them into 3 group:

1. Group 1: 75% – 80%
2. Group 2: 70% – 75%
3. Group 3: lower than 70%

It is obvious that Waverley is leading beyond the others with about 80% of listings have been rent out. It is noted that the Waverley Council local government area is made up of four wards: Hunter, Bondi, Waverley and Lawson. One of the most famous beaches – the Bondi beach is located in the Bondi suburb. According to the survey from Destination NSW in 2019, There were nearly 1.7 million international visitors and 1 million domestic visitors who visited Bondi in YE March 2019.

In contrast, about 74% of houses is in the center of city (i.e., Sydney). Places like the Sydney Opera House, the Harbour Bridge, the Rocks and the Darling Harbour are the most appealing ones that attract millions of travelers visiting the city each year. However, there are various reasons caused the ratio :

1. The global pandemic of COVID-19
2. The variety and volume of house options in the center of Sydney

As we all know, the whole world has suffered enormously from the global pandemic of COVID-19 virus. Here below is a line chart (i.e., figure 2) that displays a significant decrease in international tourism due to the effects of the coronavirus crisis from 2020 to 2021, comparing with figures before 2019. Additionally, travel restrictions and travel ban policies implemented by the Australia government led a negative impact on both the number of international arrivals and domestic travels. Since the tourism industry has been damaged so much, there is no wonder the rental housing industry in the center of Sydney would be heavily hit accordingly.

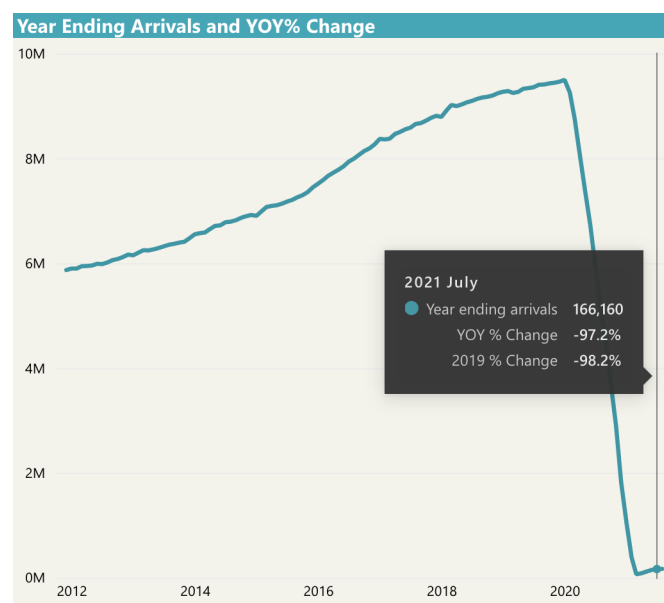


Figure 2. International market performance in tourism (from Tourism Australia)

The second possible reason is caused by the variety and volume of house options in Sydney. As the denominator in calculating the ratio, the sum of listings in Sydney is much higher than the other LGAs (referring to the figure). Noted that both the number of entire home/apt and private room in Sydney is about 1.5 times larger than numbers in Waverley.

| sum_listings | | | | | |
|------------------------|-----------------|------------|--------------|-------------|----------------------------------|
| room_type | Entire home/apt | Hotel room | Private room | Shared room | |
| neighbourhood_cleansed | | | | | |
| Ashfield | 119.0 | 0.0 | 114.0 | 10.0 | Lane Cove144.015.0103.04.0 |
| Auburn | 180.0 | 5.0 | 143.0 | 7.0 | Leichhardt517.00.0209.01.0 |
| Bankstown | 65.0 | 0.0 | 157.0 | 5.0 | Liverpool56.00.069.03.0 |
| Blacktown | 65.0 | 0.0 | 190.0 | 0.0 | Manly1213.03.0338.05.0 |
| Botany Bay | 214.0 | 1.0 | 366.0 | 12.0 | Marrickville542.04.0492.012.0 |
| Burwood | 74.0 | 4.0 | 123.0 | 11.0 | Mosman335.013.074.02.0 |
| Camden | 21.0 | 0.0 | 33.0 | 1.0 | North Sydney776.021.0336.06.0 |
| Campbelltown | 44.0 | 0.0 | 60.0 | 3.0 | Parramatta193.05.0283.015.0 |
| Canada Bay | 187.0 | 0.0 | 177.0 | 4.0 | Penrith103.00.063.02.0 |
| Canterbury | 107.0 | 0.0 | 200.0 | 10.0 | Pittwater1035.00.080.01.0 |
| City Of Kogarah | 57.0 | 0.0 | 121.0 | 4.0 | Randwick1631.034.01230.057.0 |
| Fairfield | 41.0 | 0.0 | 53.0 | 0.0 | Rockdale276.00.0454.027.0 |
| Holroyd | 23.0 | 0.0 | 51.0 | 2.0 | Ryde214.00.0288.018.0 |
| Hornsby | 226.0 | 0.0 | 170.0 | 3.0 | Strathfield50.00.0114.06.0 |
| Hunters Hill | 30.0 | 0.0 | 15.0 | 1.0 | Sutherland Shire350.03.0129.01.0 |
| Hurstville | 67.0 | 0.0 | 102.0 | 8.0 | Sydney4461.098.02727.0223.0 |
| Ku-Ring-Gai | 149.0 | 2.0 | 137.0 | 2.0 | The Hills Shire116.00.0156.04.0 |
| | | | | | Warringah1256.00.0356.03.0 |
| | | | | | Waverley2976.04.01451.082.0 |
| | | | | | Willoughby243.01.0169.06.0 |
| | | | | | Woollahra886.06.0379.011.0 |

Table 1. Sum of listings for each room type by LGA

Moreover, from the table 1, we can find that apart from the two LGAs mentioned previously, almost all the LGAs listed in the bar chart are located at the eastern or north-eastern side of the Greater Sydney, and all the top 7 LGAs such as Randwick and Manly are located at the coastline. The result of this do verify our third hypothesis at the beginning of this section.

2. In which LGA, hosts are likely to get more revenue? And which type of room could make more profits? (In 30 days)

Revenue stands for the total amount of income generated by the sales of goods or services. For a sharing economy, revenue must be one of the most important elements that each participant (i.e., host) concerns a lot. In the plot below, we could find that the bars in Light Salmon colour indicate the average number of ideal revenues in thirty days among twenty LGAs. Mosman is at the first place with over AU\$15,000 per host per month. From the Sydney map, we can see that the Mosman Council is located at the northern suburbs of Sydney adjoining Sydney Harbour. Manly Council, which is on the second place, is nearly AU\$4,000 lower than Mosman. Both of these LGAs are at the east costal line of the Greater Sydney. It is noted that Manly, Pittwater and Warringah LGAs has been amalgamated into the Northern Beaches Council.

Average Revenue in 30 Days Among LGAs

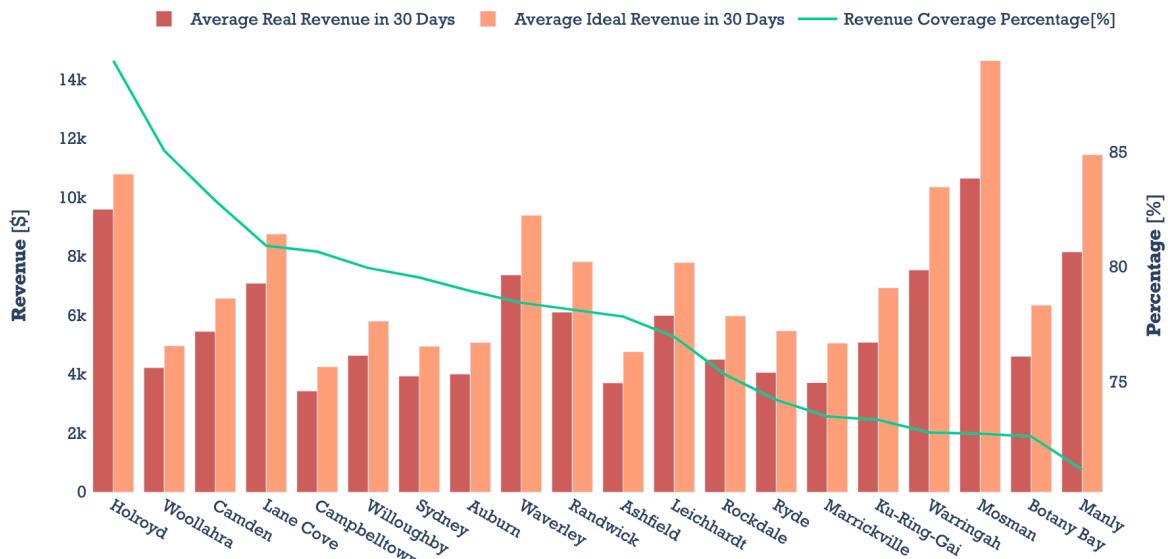


Figure 3. Average revenue (real and ideal) in 30 days by LGA

From the visual, it seems the hosts in Sydney LGA are less likely to get a higher average number in both ideal and real revenue in July, even though the Sydney LGA has the topmost value in the volume of listings. One of the possible reasons is that the total number of private rooms is much higher than some other LGAs (referring to table 2). It is obvious that the price of a single room must be lower than an entire apartment or house. However, the proportion of room type in entire home/apartment is relatively higher in LGAs like Manly, Mosman and Warringah.

Furthermore, in this figure, I also calculate the coverage of real revenue out of ideal revenue in 30 days. The result has been drawn in the green line. LGAs like Sydney, Auburn (Noted: parts of cities of Auburn, Holroyd and Parramatta has been merged to Cumberland Council in 2016) and Randwick performs better, which means hosts in these LGAs are more likely to get an ideal revenue from their sharing houses and they are less affected by the COVID-19 pandemic in Australia.

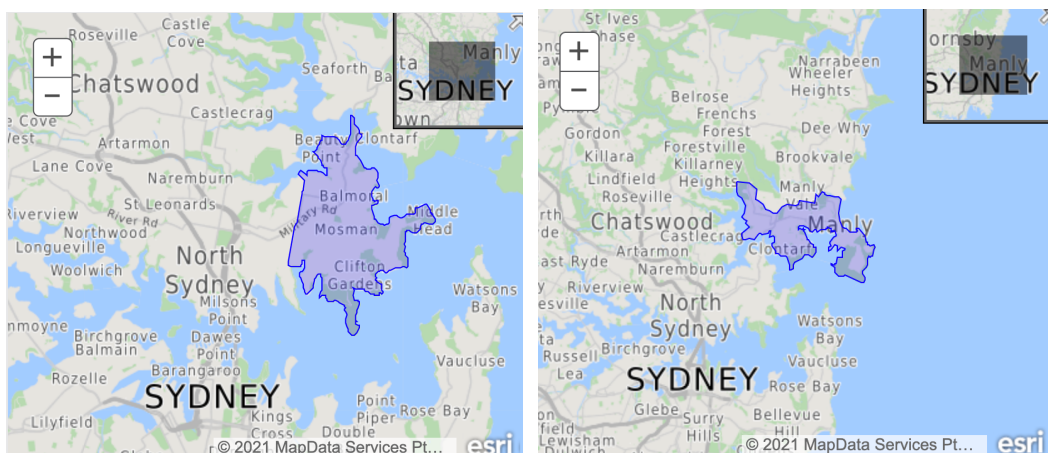


Figure 4. Geographic map for Mosman (left) and Manly (right) LGA (from Esri maps)

Ideal VS. Real Revenue in 30 Days in Australian Suburbs

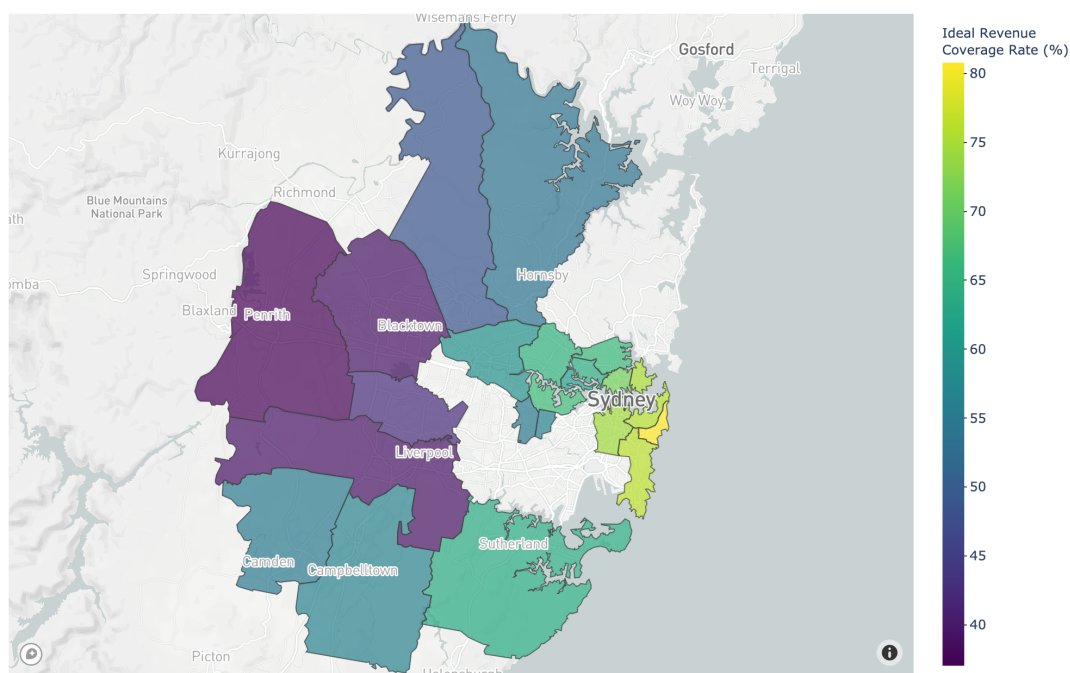


Figure 5. Spatial map for revenue coverage ratio by LGA

| The Distribution of Room Types in LGAs [%] | | | | |
|--|---------------------------|------------|--------------|-------------|
| | Entire home/ Apartment | Hotel room | Private room | Shared room |
| Blacktown | 25.5% | 0.0% | 74.5% | 0.0% |
| Bankstown | 28.6% | 0.0% | 69.2% | 2.2% |
| Strathfield | 29.4% | 0.0% | 67.1% | 3.5% |
| City Of Kogarah | 31.3% | 0.0% | 66.5% | 2.2% |
| Canterbury | 33.8% | 0.0% | 63.1% | 3.2% |
| Botany Bay | 36.1% | 0.2% | 61.7% | 2.0% |
| Rockdale | 36.5% | 0.0% | 60.0% | 3.6% |
| Burwood | 34.9% | 1.9% | 58.0% | 5.2% |
| Hurstville | 37.9% | 0.0% | 57.6% | 4.5% |
| Parramatta | 38.9% | 1.0% | 57.1% | 3.0% |

Figure 6. The percentage of room types by LGA

| Neighbourhood | Room Type | Percentage |
|---------------|-----------------|------------|
| Randwick | Entire home/apt | 83.99 |
| Waverley | Entire home/apt | 82.67 |
| Ashfield | Entire home/apt | 80.98 |
| Marrickville | Private room | 79.93 |
| Warringah | Entire home/apt | 79.22 |
| Leichhardt | Entire home/apt | 79.06 |
| Mosman | Entire home/apt | 78.69 |
| Manly | Entire home/apt | 76.72 |
| Woollahra | Entire home/apt | 76.25 |
| Sydney | Private room | 76.02 |

Table 2. The highest percentage of room types for each LGA

3. With current revenue, whether it could cover the mortgage payment in LGAs?

For the last question, I would like to know whether hosts are able to pay back their mortgage. The amount of median monthly mortgage repayment has been grabbed from the Census 2016 which is collected by the NSW government. To calculate the real revenue for each host in different LGAs, the formula of “price * the number of days the house having been rent” has been implemented. Then, the host who could undertake the mortgage will be labelled as “True” and the others will be labelled as “False”. From the table below, we can see that the city of Sydney gets the topmost value both in “True” and “False”. One of the reasons is that Sydney have significantly more listings have been recorded by Airbnb. Therefore, it is relatively difficult to make comparison on the question until I also get the result of percentage of mortgage having “True”s and sort them in a ascending order. As a consequence, Mosman gets 62.5% in a highest place which means around 60% of hosts in this LGA are able to pay their monthly mortgage with housing revenue. LGAs from the second to the fifth place can make a score around 50% respectively. However, we can find that all of these top 5 LGAs are located at or next to the eastern coastline of Sydney.

2021 July Mortgage Coverage

| | neighbourhood_cleansed | cover_mortgage_f | cover_mortgage_t | sum_listings | cover_mortgage_t_percent |
|----|------------------------|------------------|------------------|--------------|--------------------------|
| 0 | Mosman | 159 | 265 | 424 | 62.50 |
| 1 | North Sydney | 549 | 590 | 1139 | 51.80 |
| 2 | Waverley | 2185 | 2328 | 4513 | 51.58 |
| 3 | Sutherland Shire | 241 | 242 | 483 | 50.10 |
| 4 | Sydney | 3804 | 3705 | 7509 | 49.34 |
| 5 | Woollahra | 663 | 619 | 1282 | 48.28 |
| 6 | Randwick | 1571 | 1381 | 2952 | 46.78 |
| 7 | Hunters Hill | 25 | 21 | 46 | 45.65 |
| 8 | Lane Cove | 148 | 118 | 266 | 44.36 |
| 9 | Canada Bay | 218 | 150 | 368 | 40.76 |
| 10 | Willoughby | 262 | 157 | 419 | 37.47 |
| 11 | Botany Bay | 393 | 200 | 593 | 33.73 |
| 12 | Ryde | 355 | 165 | 520 | 31.73 |
| 13 | Camden | 38 | 17 | 55 | 30.91 |
| 14 | Rockdale | 524 | 233 | 757 | 30.78 |
| 15 | Campbelltown | 75 | 32 | 107 | 29.91 |
| 16 | Hornsby | 283 | 116 | 399 | 29.07 |
| 17 | Penrith | 120 | 48 | 168 | 28.57 |
| 18 | Parramatta | 363 | 133 | 496 | 26.81 |
| 19 | Strathfield | 126 | 44 | 170 | 25.88 |
| 20 | Fairfield | 70 | 24 | 94 | 25.53 |
| 21 | Burwood | 168 | 44 | 212 | 20.75 |
| 22 | The Hills Shire | 221 | 55 | 276 | 19.93 |
| 23 | Liverpool | 105 | 23 | 128 | 17.97 |
| 24 | Blacktown | 216 | 39 | 255 | 15.29 |

Table 3. The percentage of mortgage coverage ratio by LGA

Conclusion

In this report, we have covered three main questions regarding the listings' status in Sydney on the Airbnb platform. Here are some takeaways:

1. In the Greater Sydney, houses located at the eastern costal line of the New South Wales, Australia and located at the place with more tourism attractions are more likely to get tenants on Airbnb.
2. Although the global pandemic of COVID-19 has a negative impact on sharing houses industry, listings which are located at eastern places are more likely to meet their ideal revenue and cover their monthly mortgage in Sydney.
3. Entire home/apartment is the room type that exists most on Airbnb, while tenants are more preferable to private rooms.
4. Listings labelled as private room are less likely to be affected by crisis and hosts can get a lower loss from house sharing.