Data Exploration on Melbourne Airbnb Market

1. **Use Case**

People expect tourism recovery after Covid border restrictions becoming history. When visitors’ numbers scale up as before, Airbnb will be one of the hot careers against a routine work. For those willing to be an Airbnb host or a professional host, how to set price is definitely a priority.

There was total 17,751 observations recording 74 features in Melbourne Airbnb market from 2009 to present, which I obtained from the insightsairbnb website. My aim was to figure out which factors affect Airbnb price and what is their relationship. And training a model to predict the price and recognize the important factors.

Questions were below,

Q1. Why price was higher in some part of Melbourne?

Q2. Why some room type asked for higher price?

Q3. How Airbnb price has been changed by year and region?

Q4. What was relationship between host number and price?

Q5. Need we pay more for the Airbnb room owned by a super host?

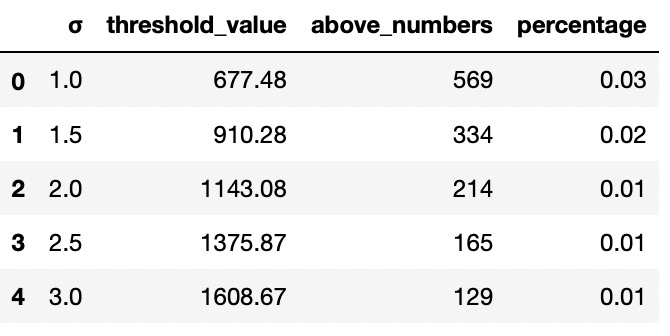
1. **Methods**

I used Python pandas to preprocess csv.gz source files, including fill missing value by median price or frequency value or reasonable value, also has done format converted and outlier definition. Then applying geopandas and seaborn packages to overview price distribution and the relations between price and regions, years, host and room types respectively. Finally training a model and further understood the top importance factors affecting price by sklearn package.

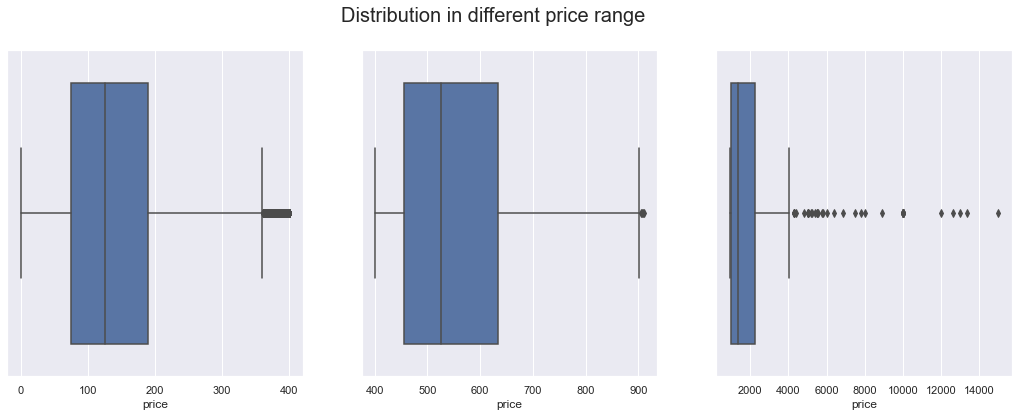
1. **Findings**
   1. Outlier definition explanation



It showed that the Melbourne Airbnb’s price was skewed to the right. And majority data with price less than $400/night.

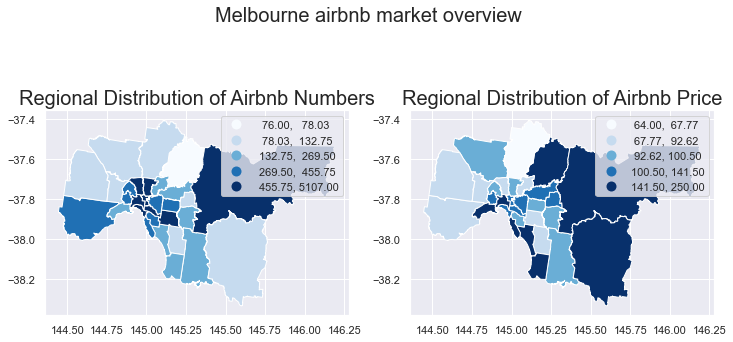


1.5 standard deviation of mean price was $910.28/night.

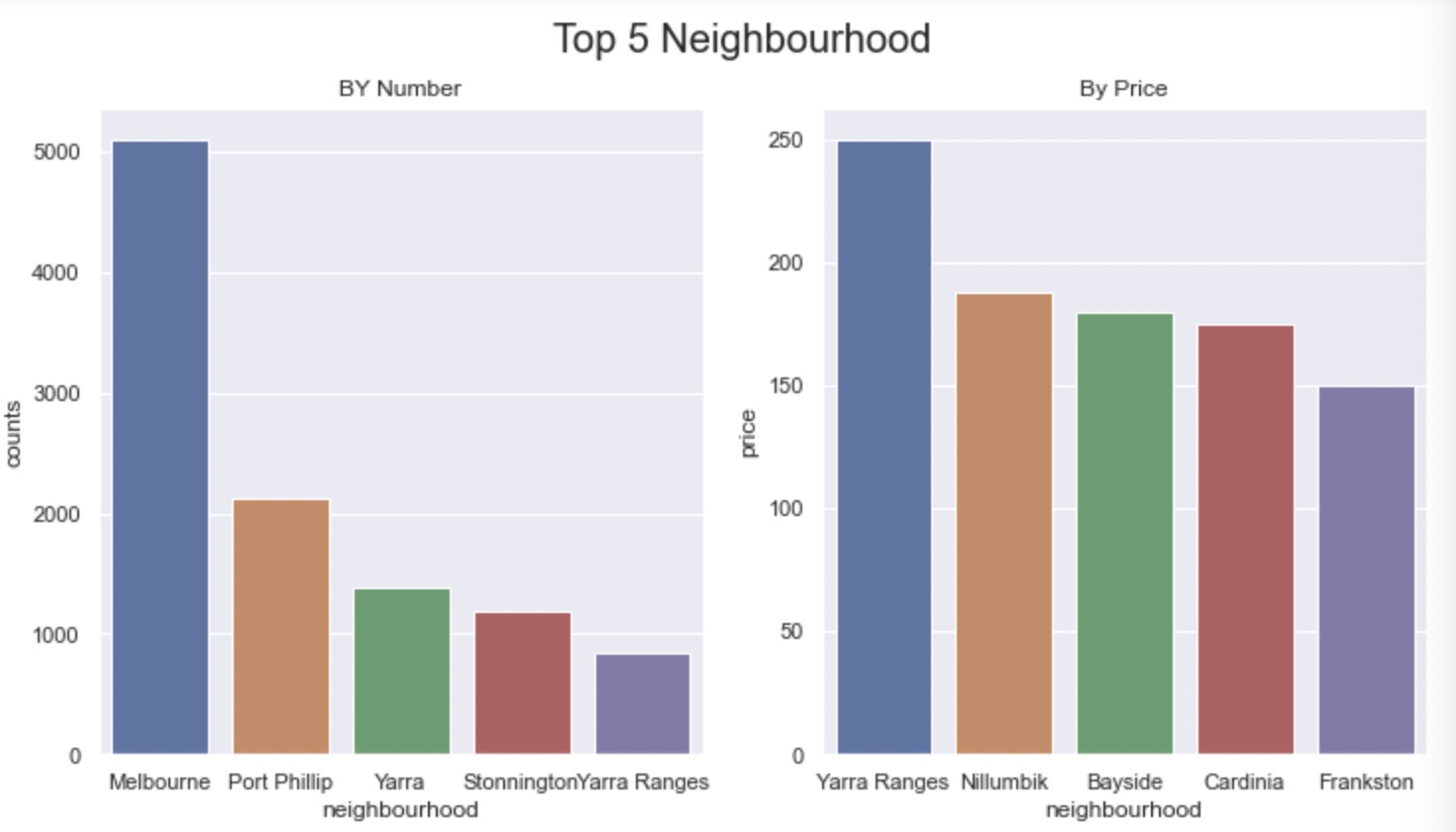


Let’s look at Airbnb’s price below $400/night, $400-$910.28/night and over $910.28/night. There were 91% Airbnb rooms under $400 per night. The median price was $ 125. 75% of them were less than $200 per night. Compared them, only 6% of the Airbnb were requested between $400 and $910 pe night, with 3 quarters were below $650. Impressively, 334 Airbnb’s were asked more than $910/night, and 44 Airbnb rooms were above $4000/night. Because only 2% of Airbnb were over $910.28 and the distribution were too wide , I regarded them as outliers.

* 1. Summary
* Distribution by region

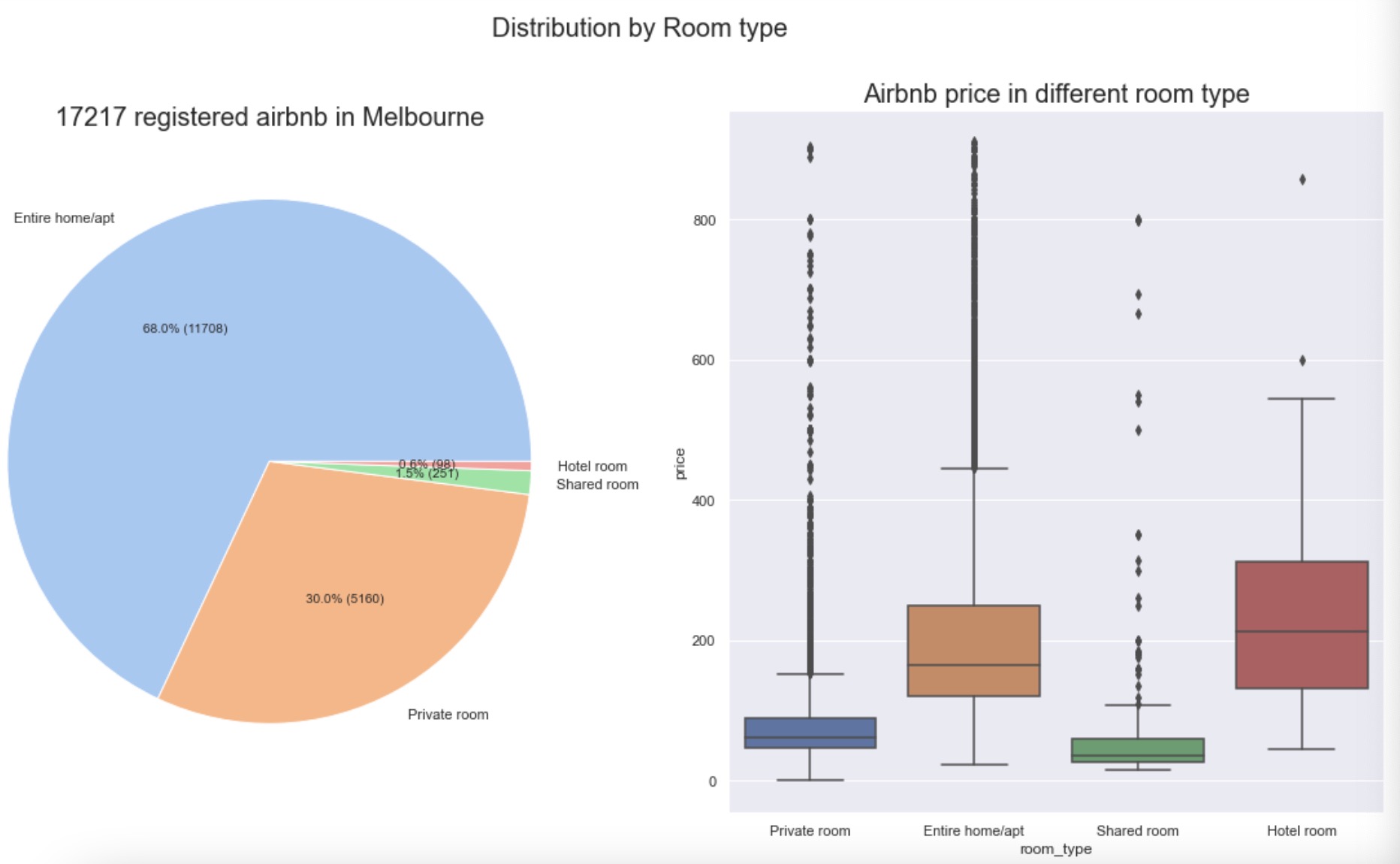


Different council areas had the different Airbnb numbers and median price. Most Airbnb located in the centre and east of Melbourne. In addition, Airbnb price in Melbourne's east were the highest. Airbnb’s along the coast were also the most expensive.



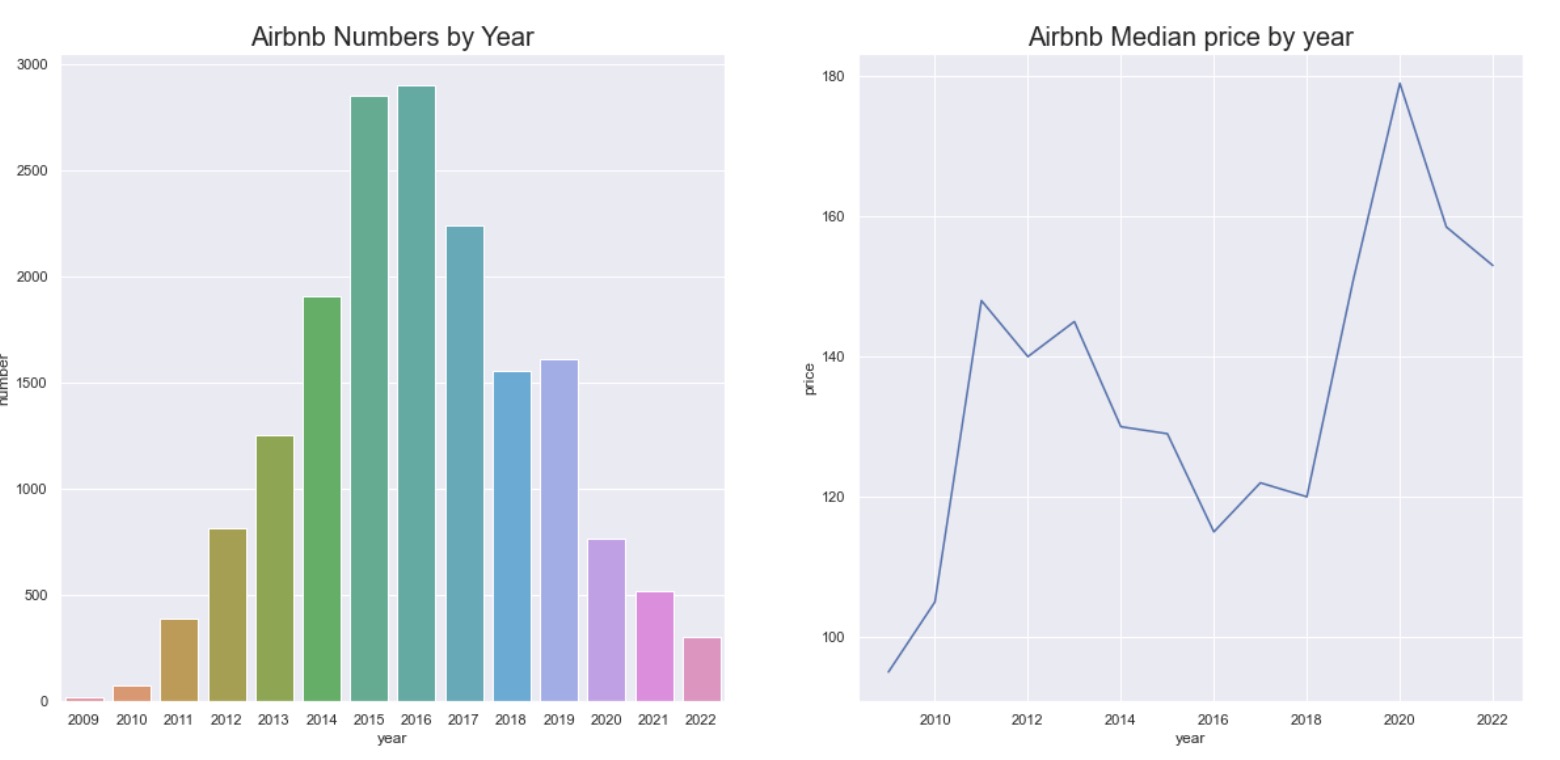
Among them, the number of Airbnb in the Melbourne City Council District were much higher than that of other City Council districts. In terms of specific council areas, the Yarra Ranges had the highest median prices, with Bayside a close second.

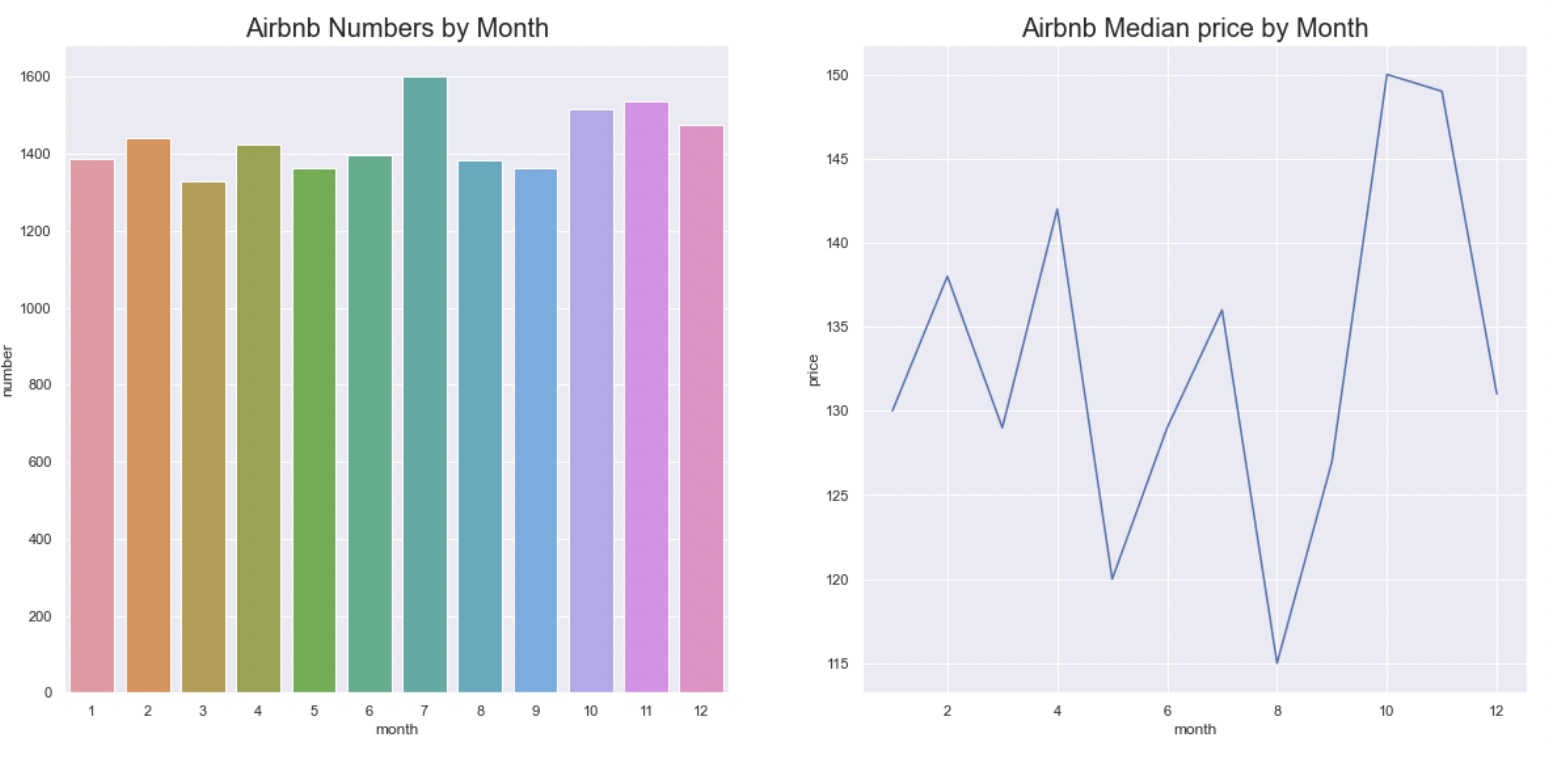
* Distribution by Room Type



The median price varied in terms of Airbnb room types. Entire houses or apartments accounted for the most room types on Airbnb, followed by private rooms. Hotel rooms were minimal. But the median price of a hotel room was the most expensive. Shared rooms were the cheapest.

* Distribution by Year



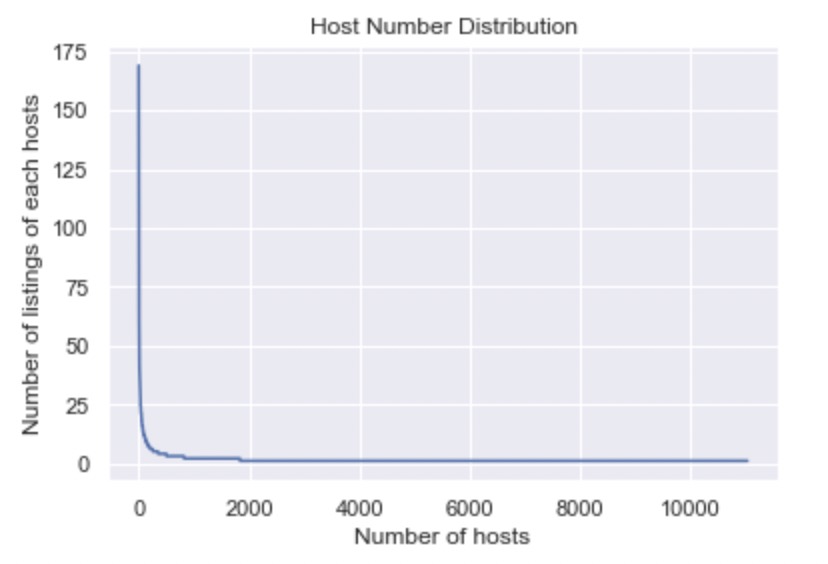


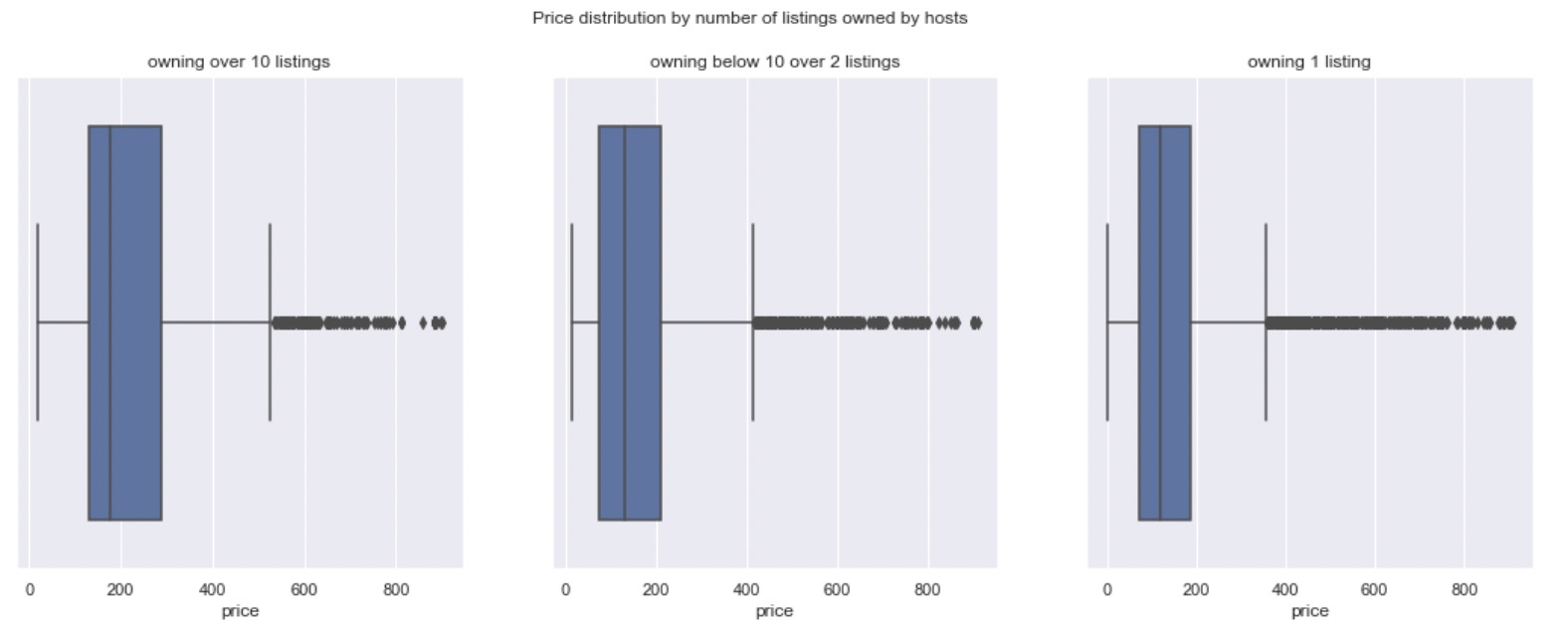
If the registered number of Airbnb was rising, the price tended to fall. The number of Airbnb increased substantially starting in 2011 and peaked in 2016. However, half pass 2022, the number of Airbnb has dropped to pre-2011 levels.

The median Airbnb price in 2016 was the lowest from 2009 to the present. From 2011-2017 prices fell when the number of Airbnb rose, and prices rose when the number of Airbnb registrations fell.

In addition, Airbnb always had the most listings in July, and the median price fluctuated up and down seasonally.

* Distribution by Host

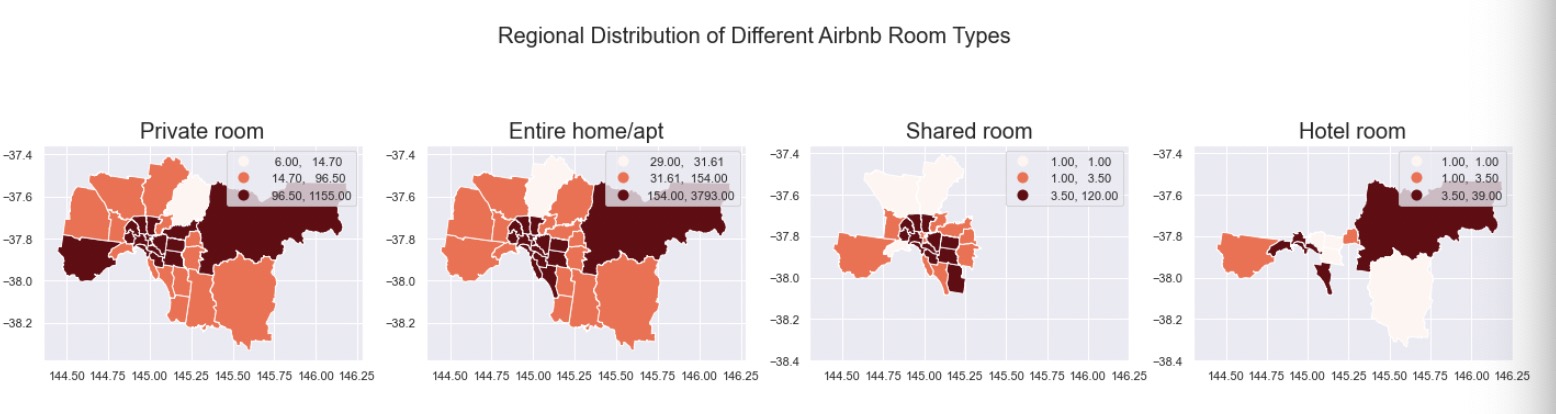




About 10,000 hosts had only 1 Airbnb home, and median price of 75% of them were less than 200. But 117 hosts had more than 10 listings, accounting for 1%, price of more than 50% of them were over 200 per night.

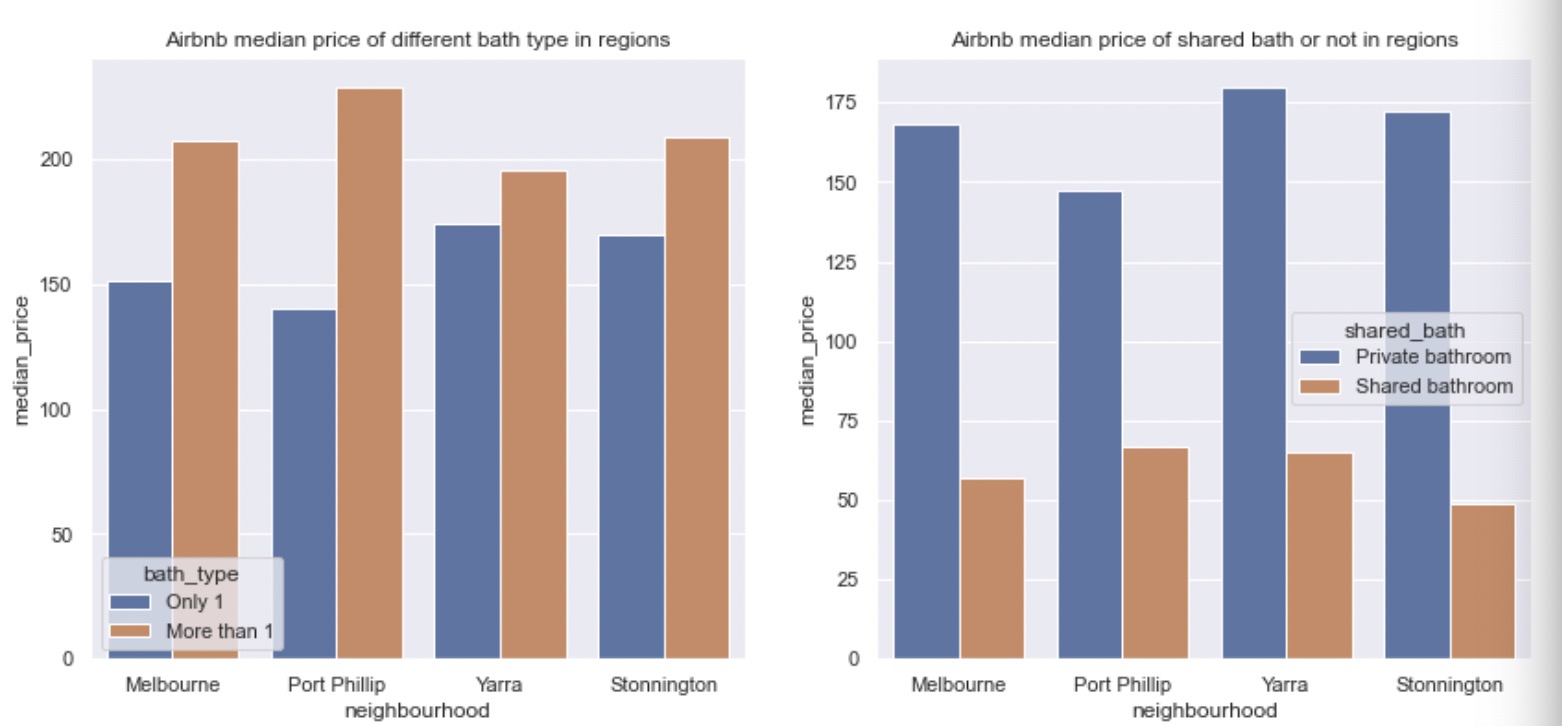
* 1. Questions

Q1. Why price was higher in some part of Melbourne?



From the previous pictures, we knew that the median price of a hotel room was the most expensive. And most hotel rooms were only available in Melbourne's north-east, which may be one reason that median price of Airbnb in this areas were higher than other places. Another reason was that the most entire Airbnb rooms and private rooms were also located in northeast Melbourne.

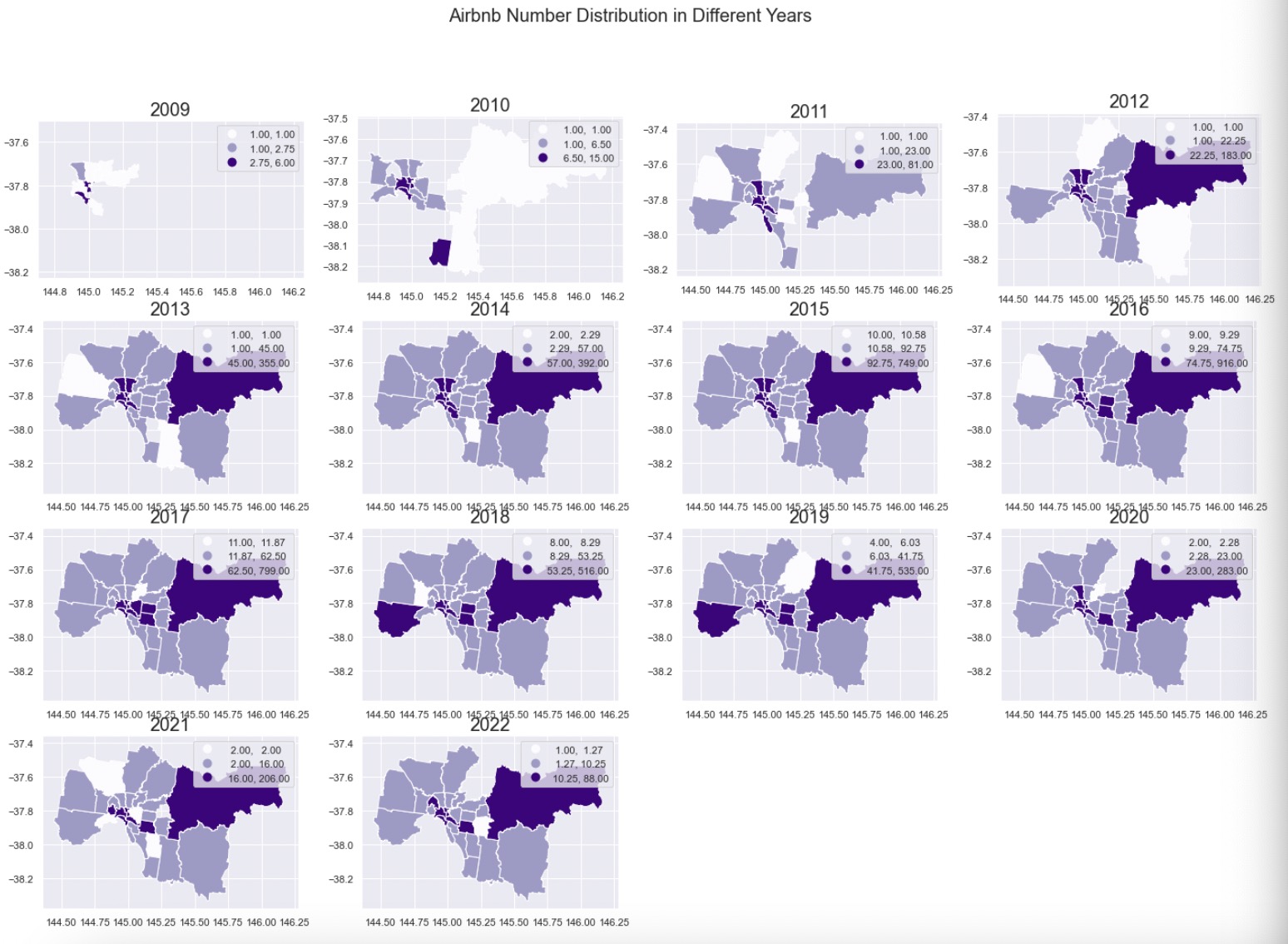
Q2. Why some room type asked for higher price?



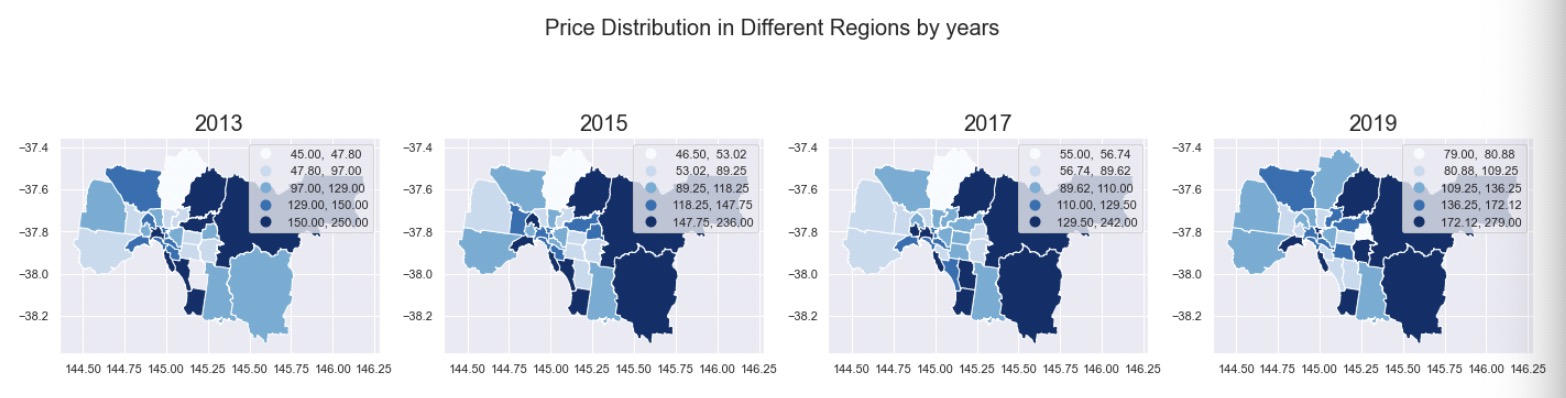
There were many features related to room details, but they had a great deal of missing values expect bath types and bath numbers. So, I picked these two features. In the same year and same room condition, we can see rooms with more than 1 bathroom were more expensive than rooms with only 1 bathroom. A room with a private bathroom was more expensive than a room with a shared bathroom.

75% of the entire house/apartment had more than 1 bathroom, and the entire house/apartment or hotel room offered a private bathroom rather than a shared bathroom, resulting in their higher median prices than private and shared rooms. That was the reason why room type may affect the price.

Q3. How Airbnb price has been changed by year and region?

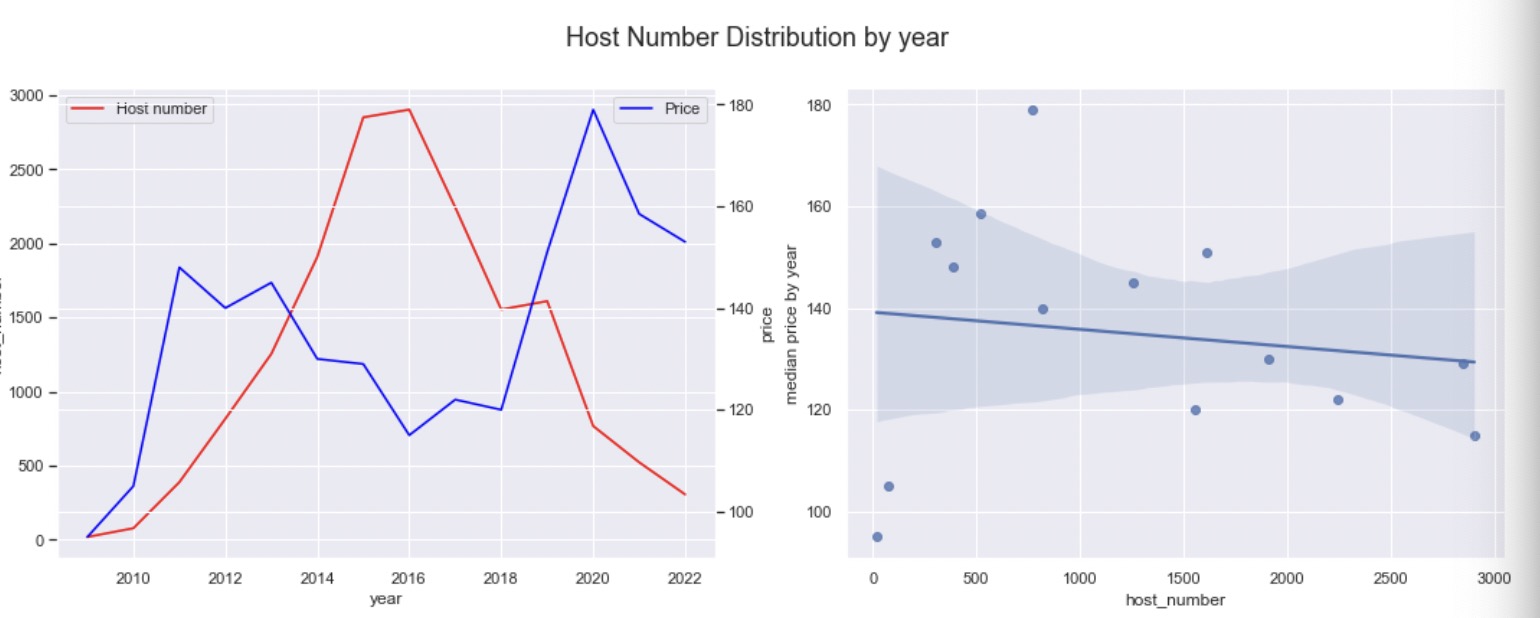


In 2009, Airbnb began to appear in the city centre and southern coastal areas. A year later, more than ten Airbnb appeared in the eastern and southern coastal areas. Since 2013, they have developed to all Melbourne, with the most in the eastern area for nearly 10 years except 2019.

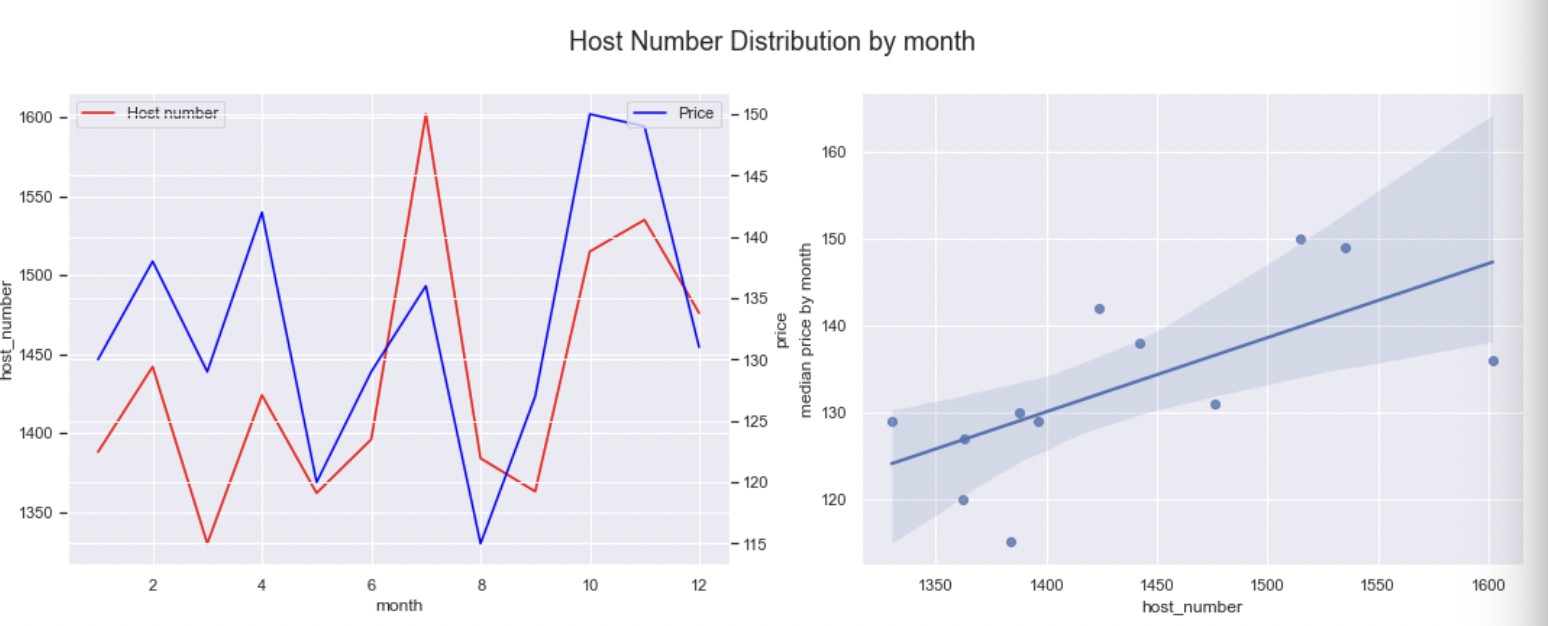


Since Airbnb has been available in every council area since 2013, I've selected data from 2013, 2015, 2017 and 2019 to see how prices were distributed in the area. Median Airbnb prices in the east were consistently higher than others but the southern coast and remained unchanged from 2013 to 2019.

Q4. What was relationship between host number and price?

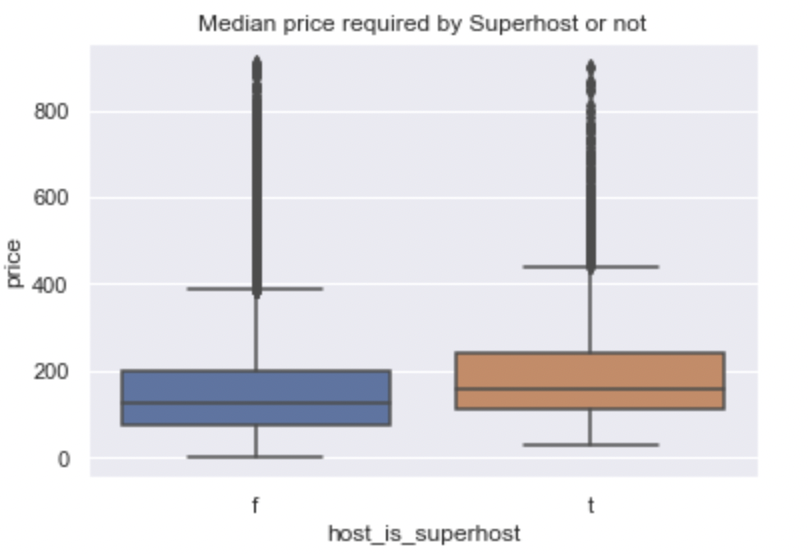


One possible reason for the lowest price in 2016 was due to the highest number of host in the same year.



The number of Airbnb host and the median price of Airbnb had the similar distribution trends by month. Possible reason for the lower price in March, May and August was due to the lower number of host in the same months. School holidays and pubic holidays do not happen in these months.

Q5. Need we pay more for the Airbnb room owned by a super host?



Most features related with host had many missing values except 'host\_is\_superhost'. We can see that the median price of Airbnb owned by super hosts were more expensive than those people who were not super host.

1. **Modeling**

I created a model to predict price and further understanding the relations between features and price. The R square of training data indicated 91% variance of room price can be explained by the features I chose. But the r square of testing data was not good, which means the model might be overfitting and the model needs to be do more hyperparameter tuning next step.

1. **Conclusion**

**First, I overviewed the distribution of Airbnb price over four dimensions including region, year, room type, and host via different forms of charts. All those Airbnb were less than $910.28/night. In addition, I used maps to show the evolution of Airbnb over time, so that Melbourne residents and investors who are unfamiliar with Melbourne can better understand the factors that affect Airbnb prices.**

**After deep analysis, I chose 10 features out of 74 total features to train a model for predicting Airbnb price which were less than $400/night. And 91% variance in Airbnb room price can be explained by the features I chose through the model. But the model is a bit overfitting, I will better it in the future research.**