airbnb | Melbourne

Group 6
Alexandrina Almazan - 662402746
Vigneshwaran Giri Velumani - 670664553
Priyadharshini Govindarajan - 669681953
Ashwin Kumar Mathi - 679880850
Yokesh VS - 675968182
Sahana Shrivats - 668077592

airbnb | How it works

Local destinations for a global community

Discover Amazing Places

Find hosts with extra rooms, entire homes, and unique accommodations like castles and igloos.

Book a Stay

Connect with hosts, confirm travel dates, and pay—all through Airbnb's trusted services.

Travel

Feel at home, anywhere you go in the world. Learn more about how to travel on Airbnb.

airbnb | How it works

A Community Built on Trust

Profiles and Reviews

Hosts and guests get to know each other through detailed profiles, personal reviews, and confirmed verifications.

Trusted Services

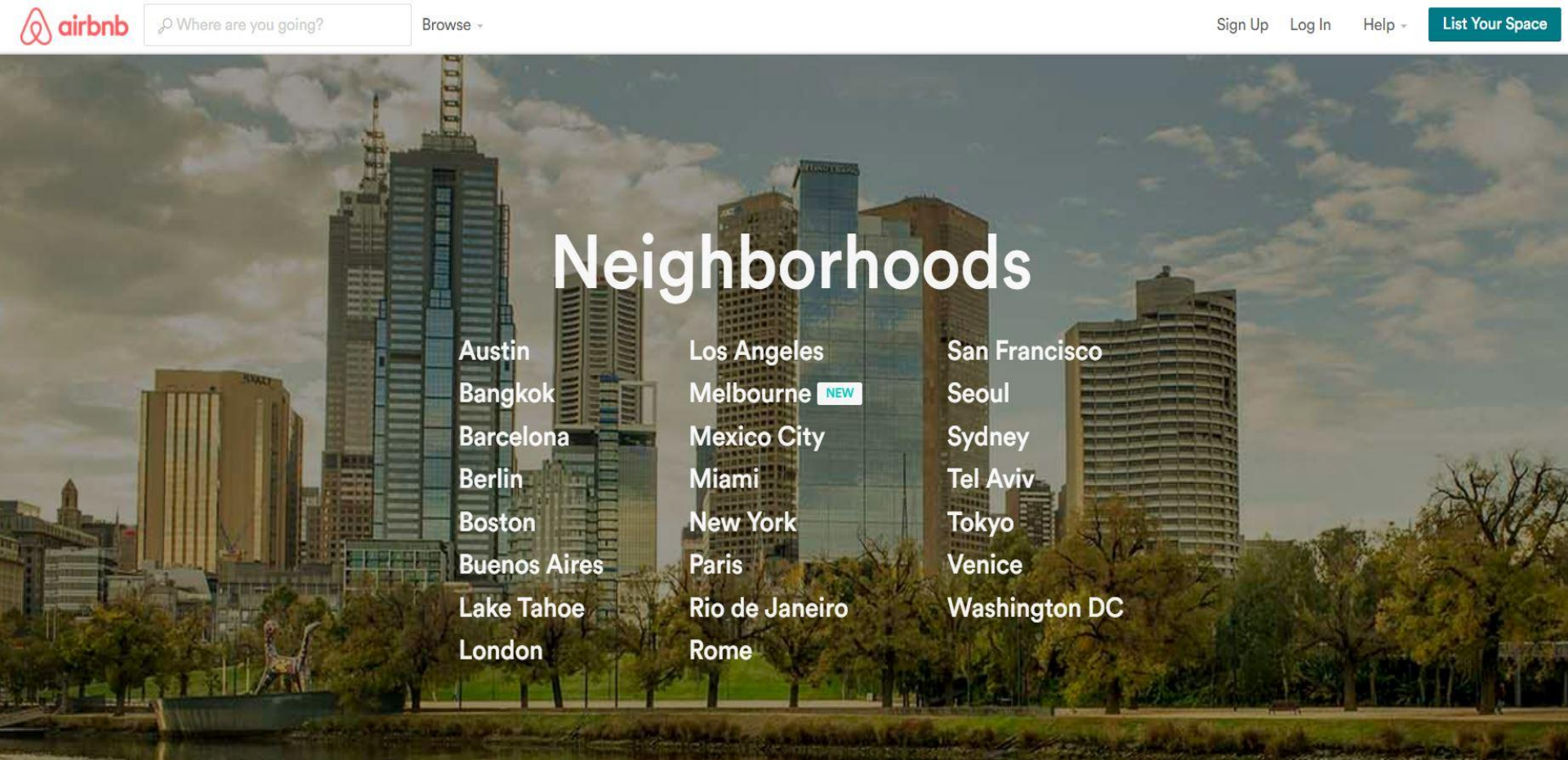
Airbnb verifies personal profiles, maintains a smart messaging system so hosts and guests can talk with confidence, and collects and transfers payments securely.

24/7 Customer Support

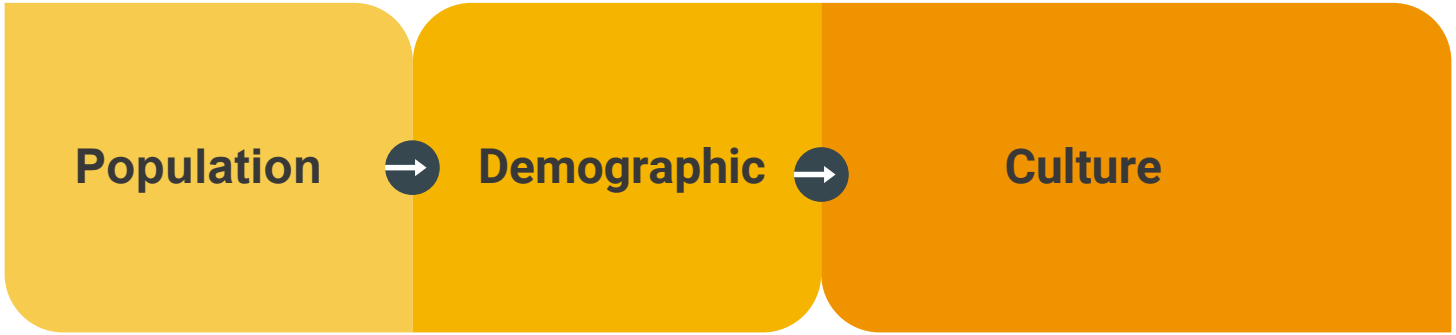
Our world-class customer support team is ready to help you around the clock, anywhere in the world.

airbnb

airbnb | Why Melbourne?

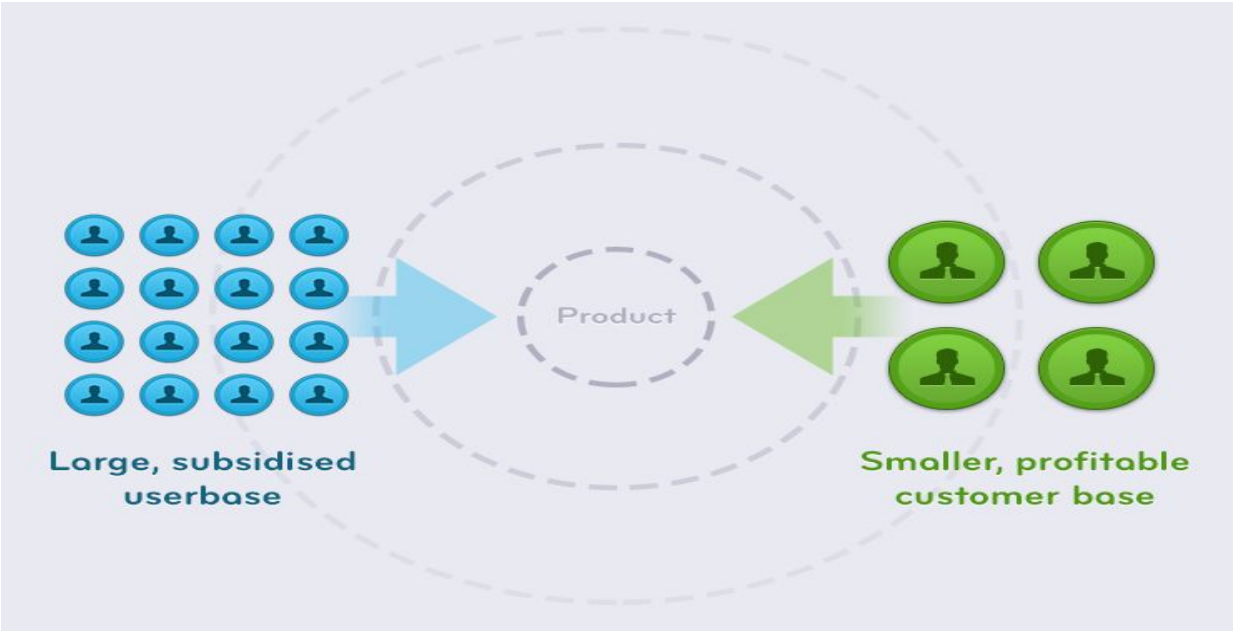


G’day Mate! | Melbourne



airbnb | Methodology

Two sided market



Airbnb Melbourne City - Preliminary Dataset Parameters

These are the variables that were provided with the preliminary dataset

• room_id	• bedrooms
• host_id	• bathrooms
• room_type	• price
• city	• minstay
• neighborhood	• latitude
• reviews	• longitude
• overall_satisfaction	• collected
• accommodates	

We have removed the highlighted variables as part of data cleaning as they do not have any significance with our analysis.

The highlighted variables are the ones on which our analysis is primarily based upon.

Additional Dataset Parameters

These are the additional dataset variables that we merged with the preliminary dataset

Postal Code Data	Crime Data	Grants Data
• postal_code	• postal_code	• postal_code
• city	• violent_crimes	• no_of_grants
	• property_crimes	• grant_value

airbnb

Airbnb Melbourne City - Final Dataset Parameters

These are the variables in the final dataset that we used for our analysis:

• room_id	• bedrooms	• grant_value
• host_id	• bathrooms	• price_cat
• room_type	• price	• cityZone
• city	• minstay	• hostCount
• neighborhood	• postal_code	• hostCategory
• reviews	• violent_crimes	• total_crimes
• overall_satisfaction	• property_crimes	
• accommodates	• no_of_grants	

The **highlighted** variables are derived variables

Assumptions

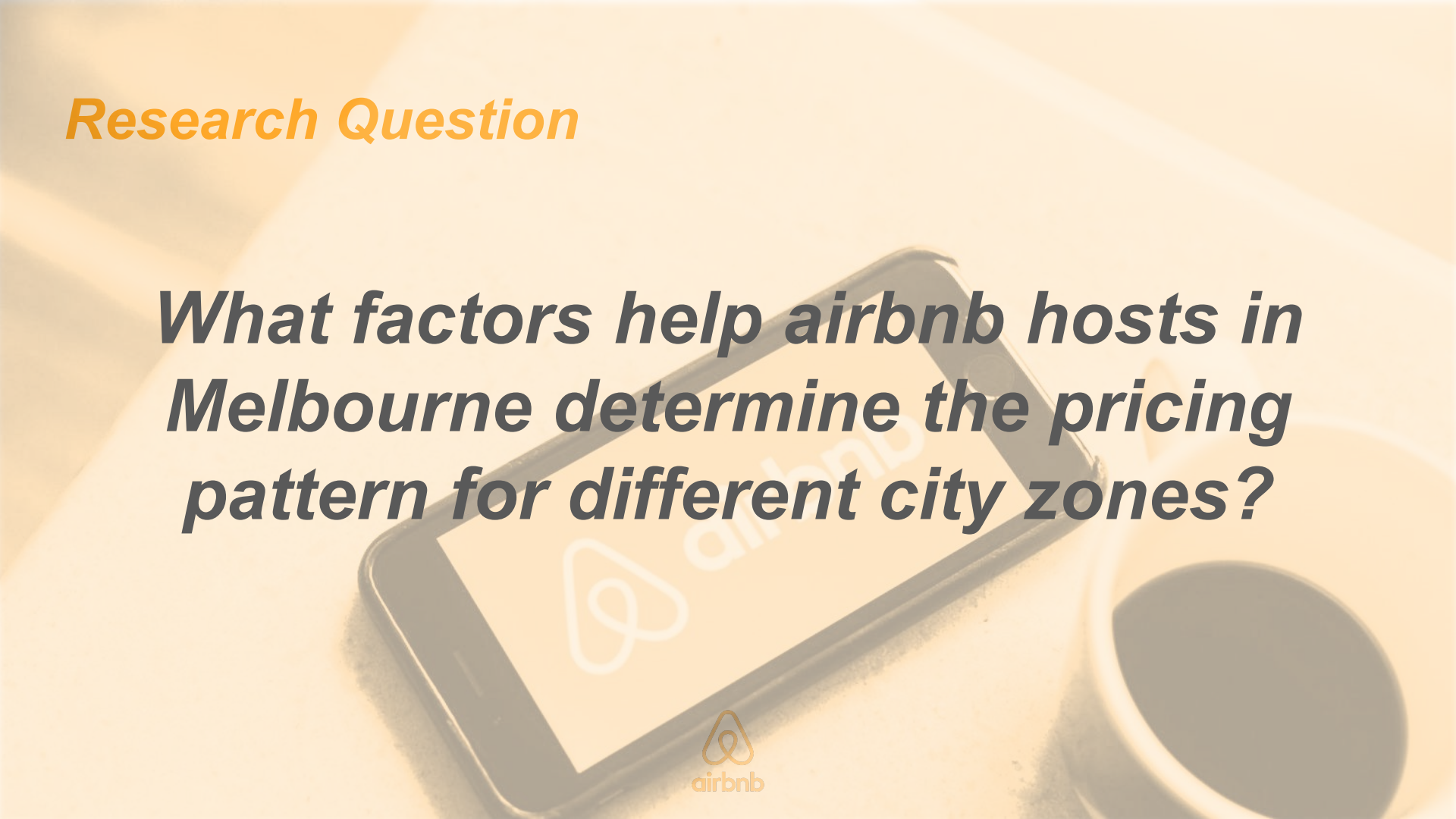
- The city limits are less than 5 km from the center of Melbourne
- Suburbs are more than 5 km from the center of Melbourne
- Grants restrict first time home buyers from renting out their units
- Utilization of airbnb listings in both city zones are the same

Challenges

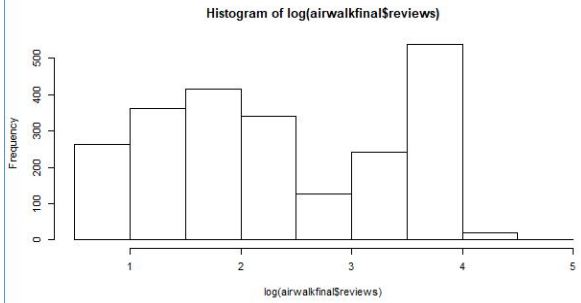
- Deciding and coming up with the city limits of Melbourne
- The ‘crime data’ and ‘grants data’ were not explicitly available
- The ‘city data’ required data cleaning
- Unavailability of utilization and time-series data

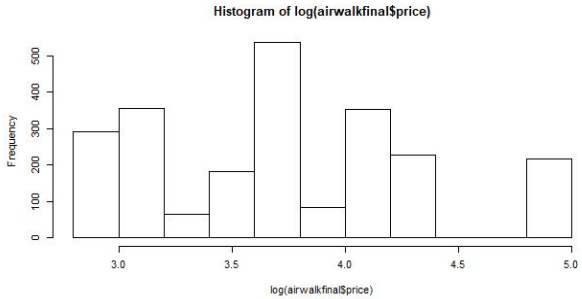
Research Question

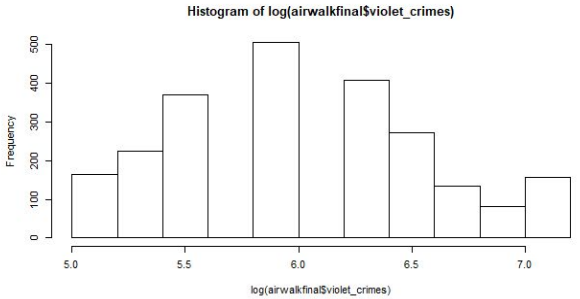
What factors help airbnb hosts in Melbourne determine the pricing pattern for different city zones?

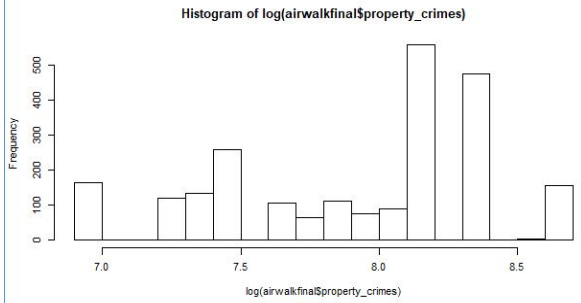


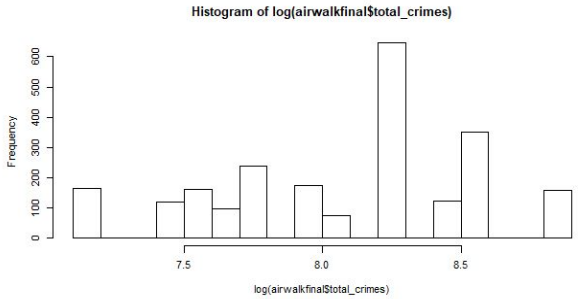
Histograms of Log Transformed Numeric Variables

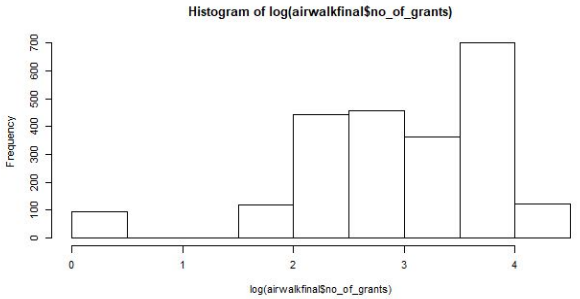


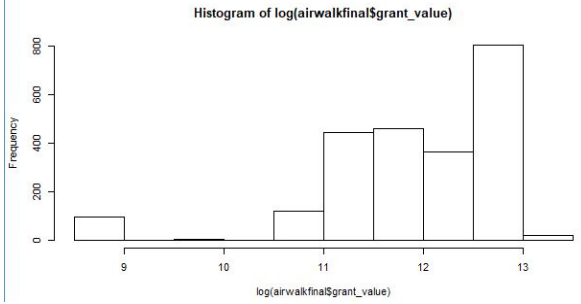


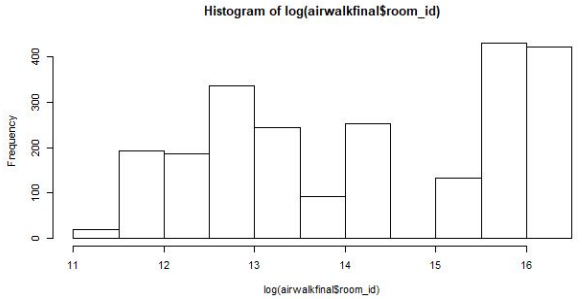


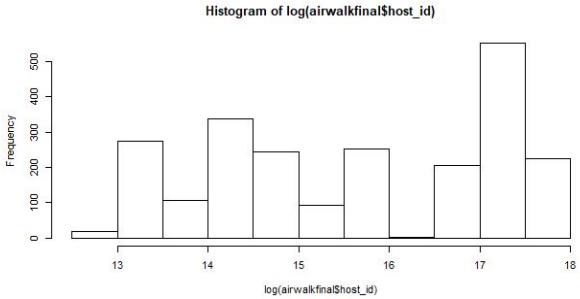








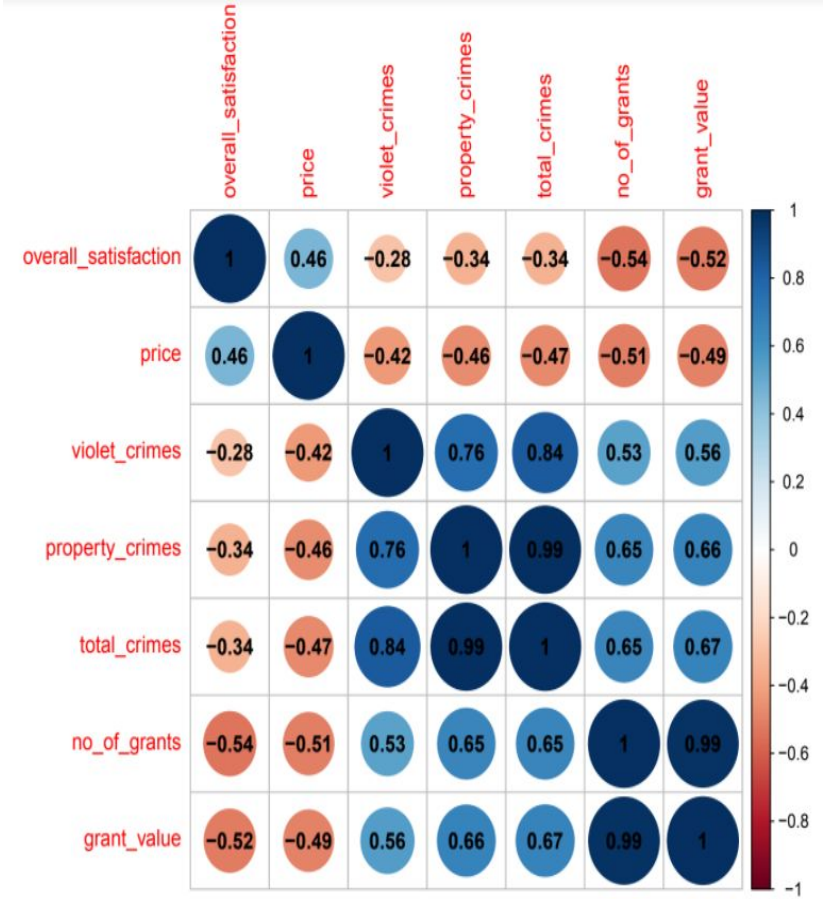




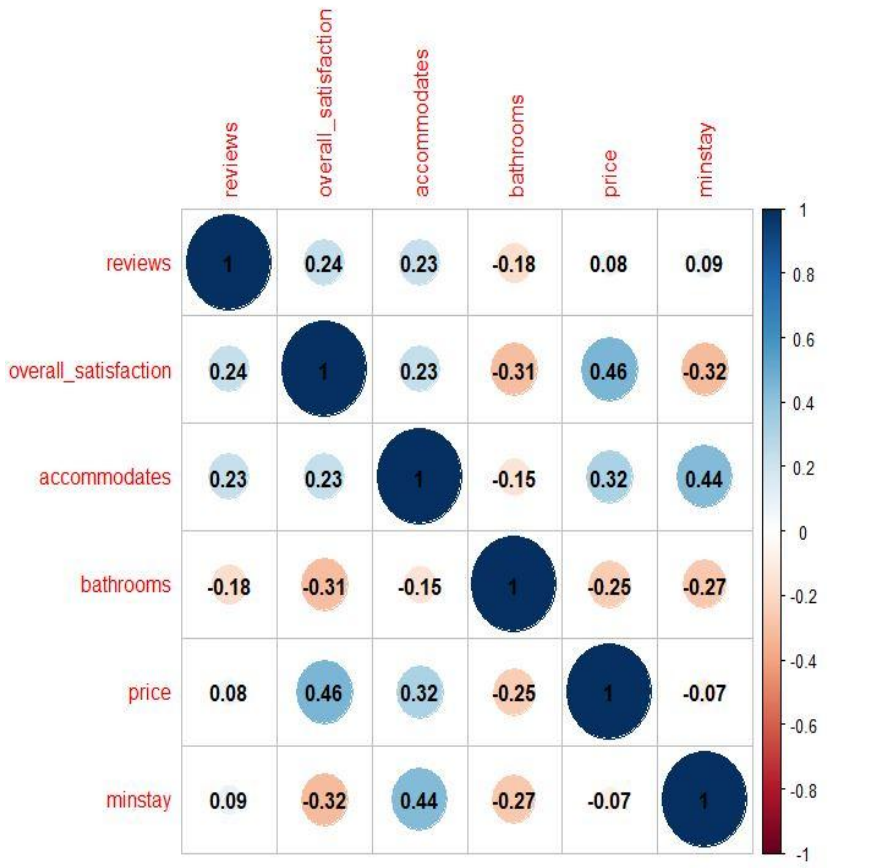
Hypotheses

- The price of an airbnb listing in Melbourne for the year 2016 is higher in a low crime zone
- The price of an airbnb listing in Melbourne for the year 2016 is higher when higher number of reviews
- The minimum stay for an Airbnb listing affects the price fixed by the hosts in Melbourne for the year 2016

Relation of Price with Crime and Grant variables



Relation in listing's attributes



airbnb

Hypothesis testing

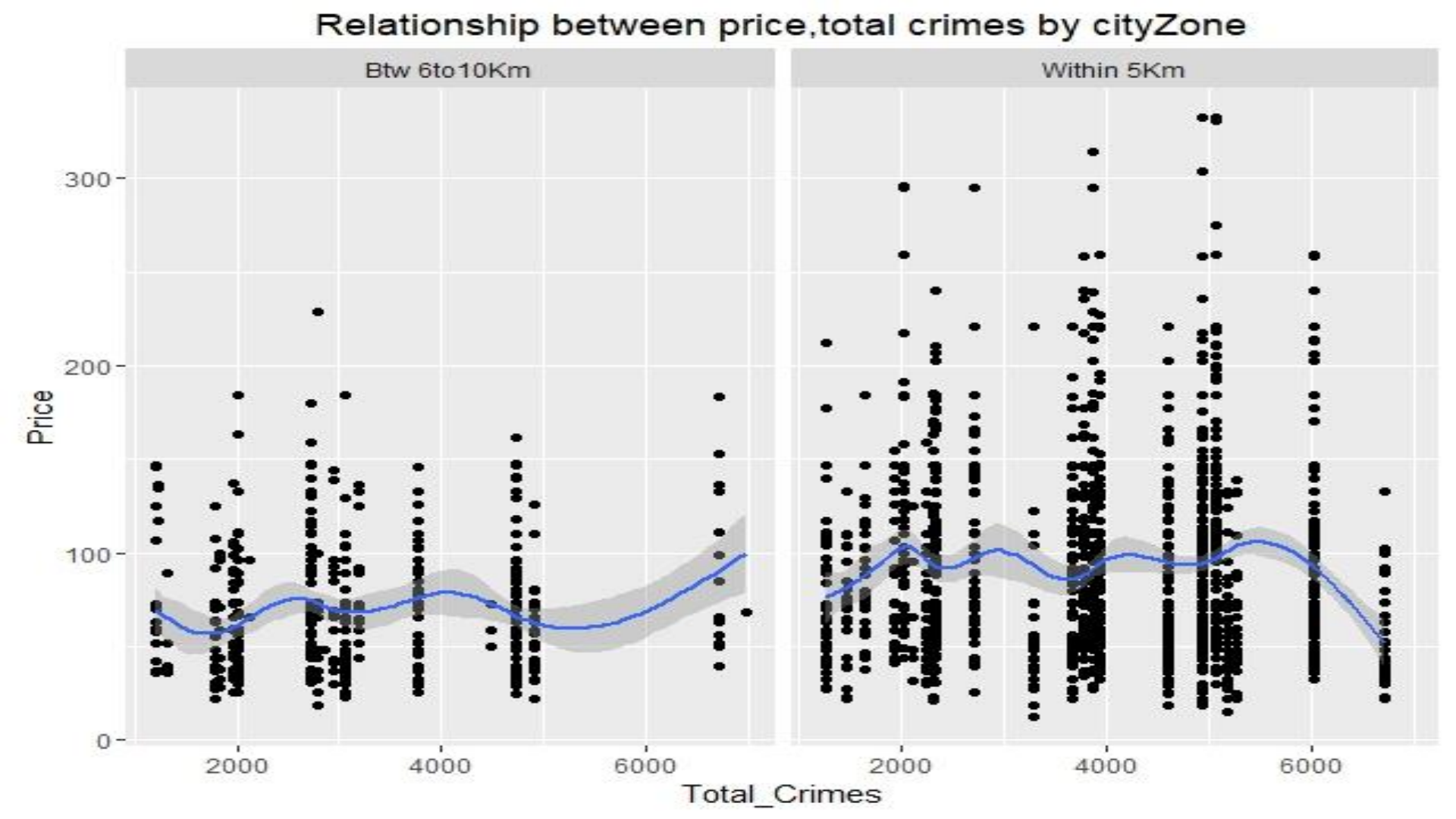
```
        One Sample t-test

data:  zone_6to10$price
t = -52.488, df = 130, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50.79
95 percent confidence interval:
 20.51086 22.71052
sample estimates:
mean of x
 21.61069
```

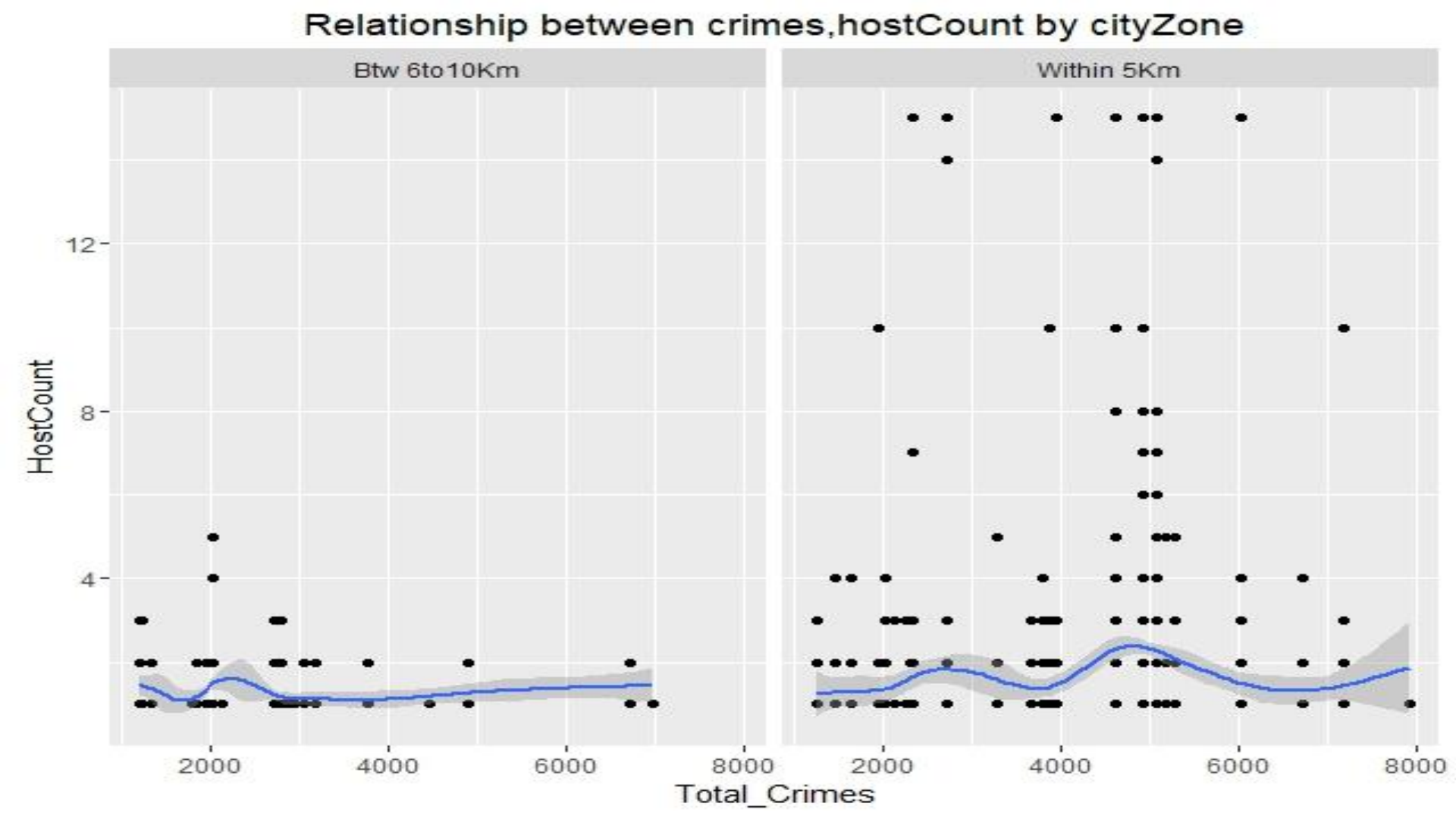
Price is higher for Zone 1 as compared to Zone 2

Crime is higher for Zone 1 as compared to Zone 2

```
        One Sample t-test

data:  zone_6to10$total_crimes
t = -7.611, df = 130, p-value = 4.892e-12
alternative hypothesis: true mean is not equal to 3528.2
95 percent confidence interval:
 2950.676 3188.973
sample estimates:
mean of x
 3069.824
```

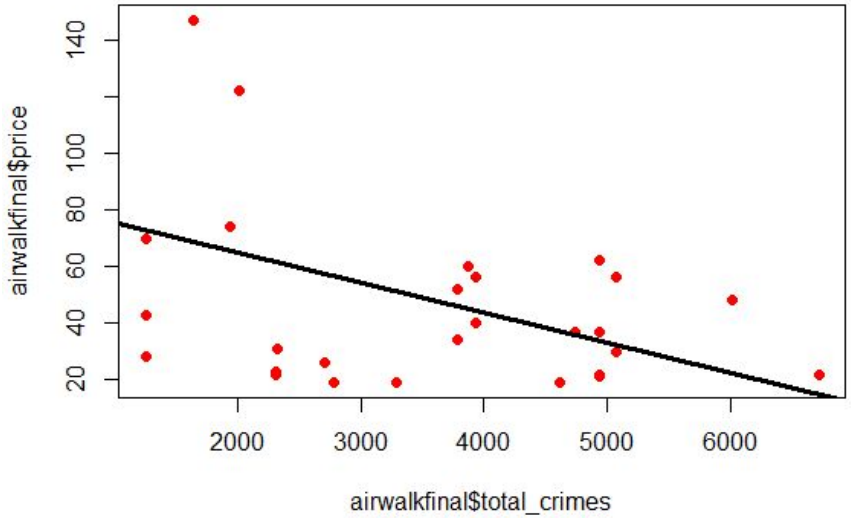
airbnb


Relationship between price,total crimes by cityZone
Btw 6to10Km
Within 5Km
Price
300
200
100
0
2000
4000
6000
Total_Crimes


Relationship between crimes,hostCount by cityZone
Btw 6to10Km
Within 5Km
HostCount
12
8
4
2000
4000
6000
8000
Total_Crimes

Bivariate Analysis

1.Price vs. Total Crimes



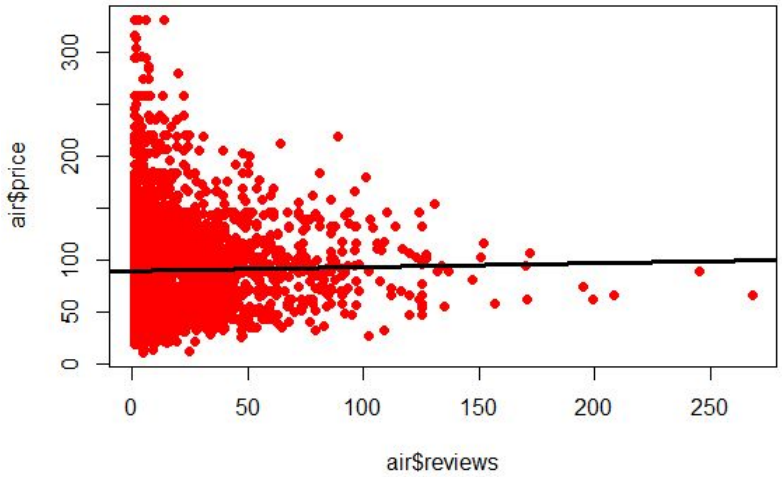

```
> cor.test(log(airbnb$price), log(airbnb$total_crimes))

        Pearson's product-moment correlation

data:  log(airbnb$price) and log(airbnb$total_crimes)
t = -25.58, df = 2308, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5011661 -0.4375933
sample estimates:
      cor
-0.469989
```

Summary: 47% Negative Correlation

airbnb

Bivariate Analysis

2.Price vs. Number of Reviews



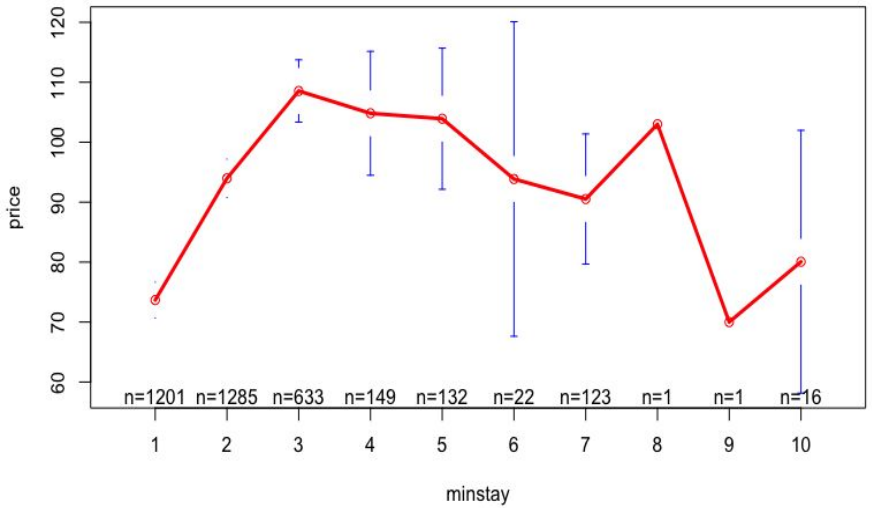

```
> cor.test(airbnb$price, airbnb$reviews)

        Pearson's product-moment correlation

data:  airbnb$price and airbnb$reviews
t = 3.6236, df = 2308, p-value = 0.0002968
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.03453493 0.11564106
sample estimates:
       cor
0.07521239
```

Summary: 7.5% Positive Correlation

airbnb

Bivariate Analysis

3.Price vs. Minstays



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
minstay	9	618744	68749	34.34	<0.0000000000000002 ***
Residuals	3553	7112538	2002		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

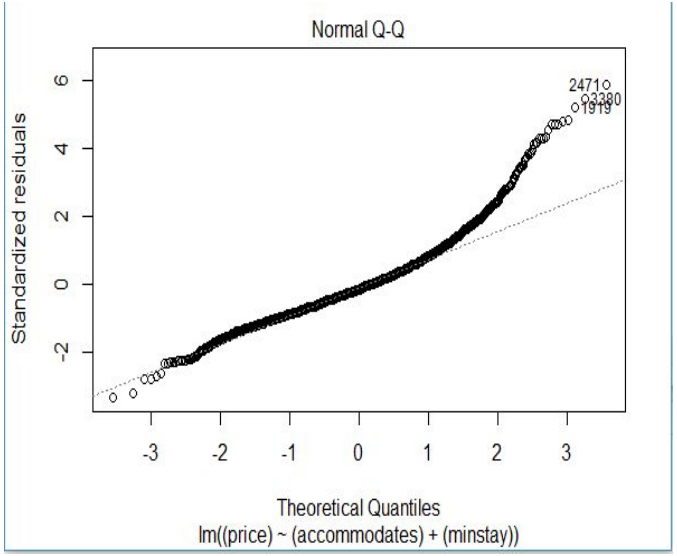
Summary: Correlation is High

airbnb

Linear Regression

Model 1: Price with Accommodates and Minstay in Zone1 (Within 5 Kms)

```
Call:
lm(formula = (price) ~ (accommodates) + (minstay), data = zone1)

Residuals:
    Min      1Q  Median      3Q     Max
-118.54  -23.16   -4.91   16.78  209.53

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    37.526      2.755  13.622  < 2e-16 ***
accommodates2  29.697      2.828  10.502  < 2e-16 ***
accommodates3  48.962      3.641  13.447  < 2e-16 ***
accommodates4  83.942      3.040  27.612  < 2e-16 ***
accommodates5 103.842      3.913  26.535  < 2e-16 ***
minstay2       11.687      1.688   6.922 5.57e-12 ***
minstay3       22.071      2.049  10.772  < 2e-16 ***
minstay4       17.235      3.474   4.961 7.47e-07 ***
minstay5       25.337      3.966   6.388 1.97e-10 ***
minstay6       17.690      9.651   1.833 0.066935 .
minstay7       13.188      3.942   3.345 0.000833 ***
minstay9        2.777     35.817   0.078 0.938200
minstay10      10.676      9.647   1.107 0.268566
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.79 on 2649 degrees of freedom
Multiple R-squared:  0.4389,    Adjusted R-squared:  0.4364
F-statistic: 172.7 on 12 and 2649 DF,  p-value: < 2.2e-16
```

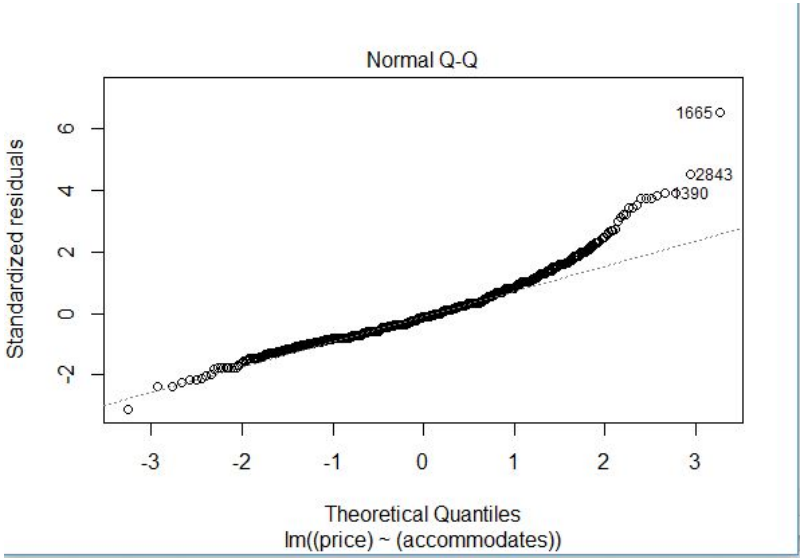


Adjusted R-squared: 43.64%

Linear Regression

Model 2: Price with Accommodates in Zone2 (Between 6-10 Kms)

```
Call:
lm(formula = (price) ~ (accommodates), data = zone2)

Residuals:
    Min      1Q  Median      3Q     Max
-98.176 -20.998  -3.998  13.812 203.462

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    38.188      3.940   9.693  < 2e-16 ***
accommodates2  24.811      4.171   5.948 3.89e-09 ***
accommodates3  46.276      5.470   8.460  < 2e-16 ***
accommodates4  73.989      4.564  16.210  < 2e-16 ***
accommodates5  90.351      5.884  15.354  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.52 on 896 degrees of freedom
Multiple R-squared:  0.3918,    Adjusted R-squared:  0.3891
F-statistic: 144.3 on 4 and 896 DF,  p-value: < 2.2e-16
```



Adjusted R-squared: 38.91%

It's True Mate, Fair Dinkum! | Results

- In Zone 1, Airbnb hosts set the price of listings based on the number of accommodation the property can hold and its corresponding minimum stay requirement.
- There is a mild correlation between the property's price and its minimum stay requirement
- In Zone 2, Airbnb hosts set the price of the listings only based on the size of the accommodation
- Hosts don't follow a pattern based on property's utilization or its locality and instead consider the allowed number of accommodation.

It's Crickey | Research Findings

- Crime rate is comparable across two city zones
- Negative correlation exists between price and total crimes
- Average price of listings in Zone 1(Less than 5kms from Melbourne CBD) is higher than that in Zone 2 (between 6 to 10kms from Melbourne CBD)

Recommendations

- Hosts should price the listings based on its utilization rather than just the unit size or minimum stay values in Zone 1.
- Hosts should follow a similar trend in Zone 2

Thank You!