Which AirBnB should I visit while staying in Amsterdam?

Sharda Hinsenveld, Miranda Manschot, Nikki Tissen, Lauren Verhoeven, Lotte Verkuijlen

Introduction

Recently the municipality of Amsterdam has changed its policy concerning renting out houses. While it was previously allowed to sublet your apartment for 60 days per year, it is now only 30 days. This new legislation affects Airbnb hosts in Amsterdam financially as they can miss up to half of their income from renting out the apartment. It, therefore, made sense that hosts now have increased their prices. But whether this is actually the case and whether there is a difference between those who adhere to the new legislation and those who do not, is yet to be determined. Hence we set up the following research question: "How does the announcement of the 30-days rule impact the Airbnb housing price in Amsterdam?

First, we cleaned the dataset which was uploaded from the AirBnB website. After that, we looked at which variables still needed to be created to answer our research question. The definitions of the variables from AirBnB and those we created on our own can be found under the descriptions header.

After we cleaned the dataset and created new variables, we ran an analysis of variance (ANOVA). This was done as the independent variable was of metric origin and the dependent variables were non-metric. Afterwards we plotted the graphs to visualise the results as well. Based on these results we drew conclusions that the neighbourhood indeed influences the price, but the number of days the apartment is rented out is less significant. Those who rent it out between 30 and 60 days do ask a significantly higher price but the others do not.

Variable Descriptions

The dataset consists of the following variables:

Neighbourhood_cleansed - Neighbourhood_cleansed is a factor variable. The neighbourhood is geo coded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles. It is the neighbourhood in which the listing is located.

Room_type - Room_type is a factor variable. It is the kind of room the listing is renting, this dataset has 4 different room types.

Accommodates - Accommodates is a numeric variable and is defined by the maximum capacity of the listing.

Maximum_nights - Maximum_nights is a numeric variable and is defined by the maximum number of nights stay for the listing.

Price_1 - Price_1 is a numeric variable and is defined by the daily price in local currency.

Priceperson - Priceperson is a numeric variable and is defined by the price of the listing divided by the maximum number of people who can stay in the listing.

Classification_nights - Classification_nights is a nominal variable. Classification_nights is the classification of how many nights owners rent out their apartment. 0 indicates that they stick to the 30 day rule and do not rent it out for more than 30 days. Classification 1 means that the owners rent it out between 31 and 60 days and 3 means that the listing can be rented for more than 60 days.

Classification_neigbourhood - Classification_neigbourhood is a dummy variable. Which has the value of 0 when the accommodation is located in Amsterdam Noord, and has a value of 1 when located in Amsterdam Centrum.

Dummy30days - Dummy30days is a dummy variabele. Which has a value of 1 when the accommodation complies to the 30-days rule and rents their house for 30 days or less.

Dummy3060days - Dummy3060days is a dummy variabele. Which has a value of 1 when the accommodation complies to the 60-days rule and rents their house between 31 and 60 days.

Dummy61days - Dummy61days is a dummy variabele. Which has a value of 1 when the accommodation does not comply with the 30- and 60-days rule and rents their house for more than 60 days.

ANOVA

Analyis of Variance (ANOVA) with Price per person as dependent variable and classification_nights & classification_neighbourhood as independent variable + their interaction effect. From the output below we can conclude that both the neighbourhood and the number of nights the apartment is rented out have a significant effect on the price. Note that classification_nights is only marginally significant. This implies that there is an effect, however, we must be careful when drawing conclusions based on it. The neighbourhood has a greater effect than the number of nights. The interaction between these two variables is not significant.

```
                                  Df   Sum Sq Mean Sq F value Pr(>F)
classification_nights              2     3564    1782   2.323 0.0983
.
classification_neigbourhood        1   210832  210832 274.847 <2e-16
***
class_nights:class_neigbourhood 2      967     967   0.630 0.5327
Residuals                    1637  1255726     767
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

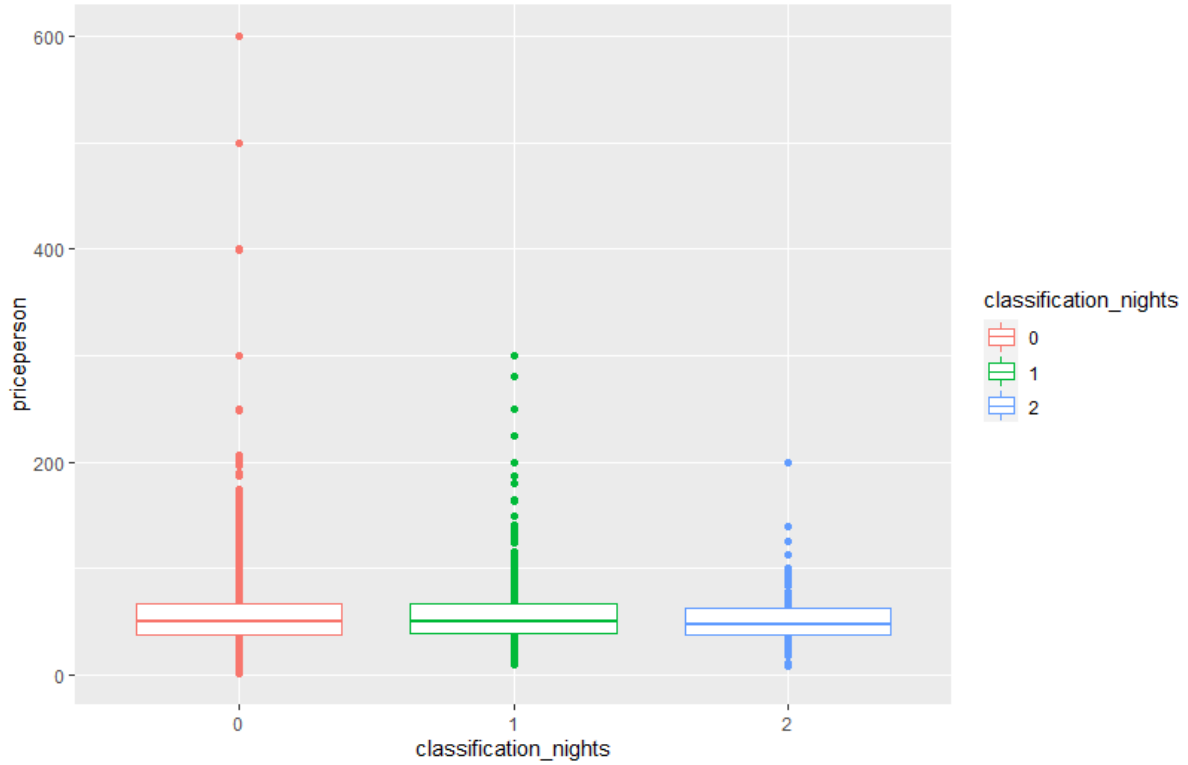
Analysis of Variance (ANOVA) with Price per person as dependent variable and and the dummies of those who adhere to the rules (dummy30days) and those who do not (dummy3060days and dummy61days) as independent variables. Here we can see that those who do not stick to the rules, but do not rent out their apartment for more than 60 days, have a significant effect on the price. People who rent out their apartment for more than 60 days or stick to the rules, do not have a significant effect on the price.

```
                  Df   Sum Sq Mean Sq F value Pr(>F)
dummy30days        1     3242    3242   3.592 0.0581 .
dummy3060days      1     3514    3514   3.894 0.0485 *
Residuals       6981  6300401     903
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
                  Df   Sum Sq Mean Sq F value Pr(>F)
dummy3060days      1     5817    5817   6.446 0.0111 *
dummy61days        1      939     939   1.040 0.3078
Residuals       6981  6300401     903
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

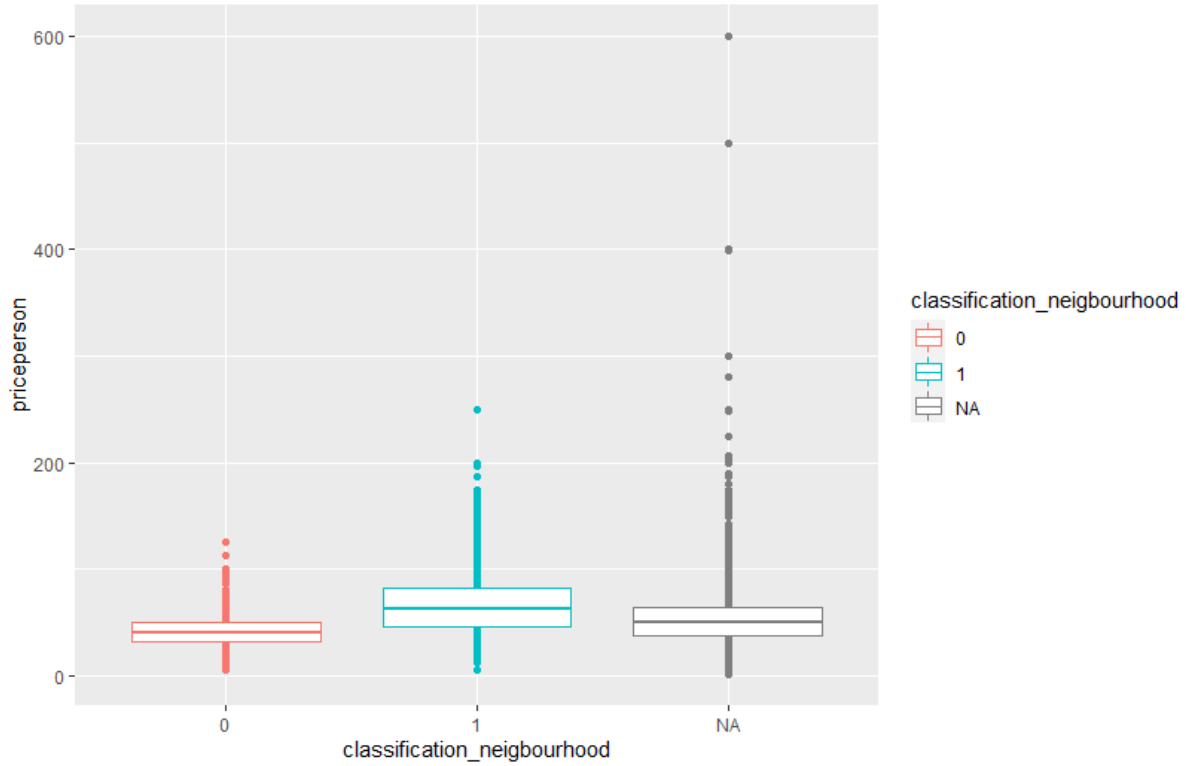
Graph 1

In the first graph we see the classification_nights on the X-axis and Priceperson on the Y-axis. Here we can distinguish the classification_nights between 0, 1 and 2 where 0 represents the group that does stick to the 30-day rule, 1 people who do not but do not rent out their property for more than 60 days and 2 that represents people who rent out their property for more than 60 days. Priceperson is the Price per person, calculated by the price per night divided by the number of people the location accommodates. As supported by the ANOVA analysis, we can see that classification 1 has a slightly higher average than classification 0 and 2.

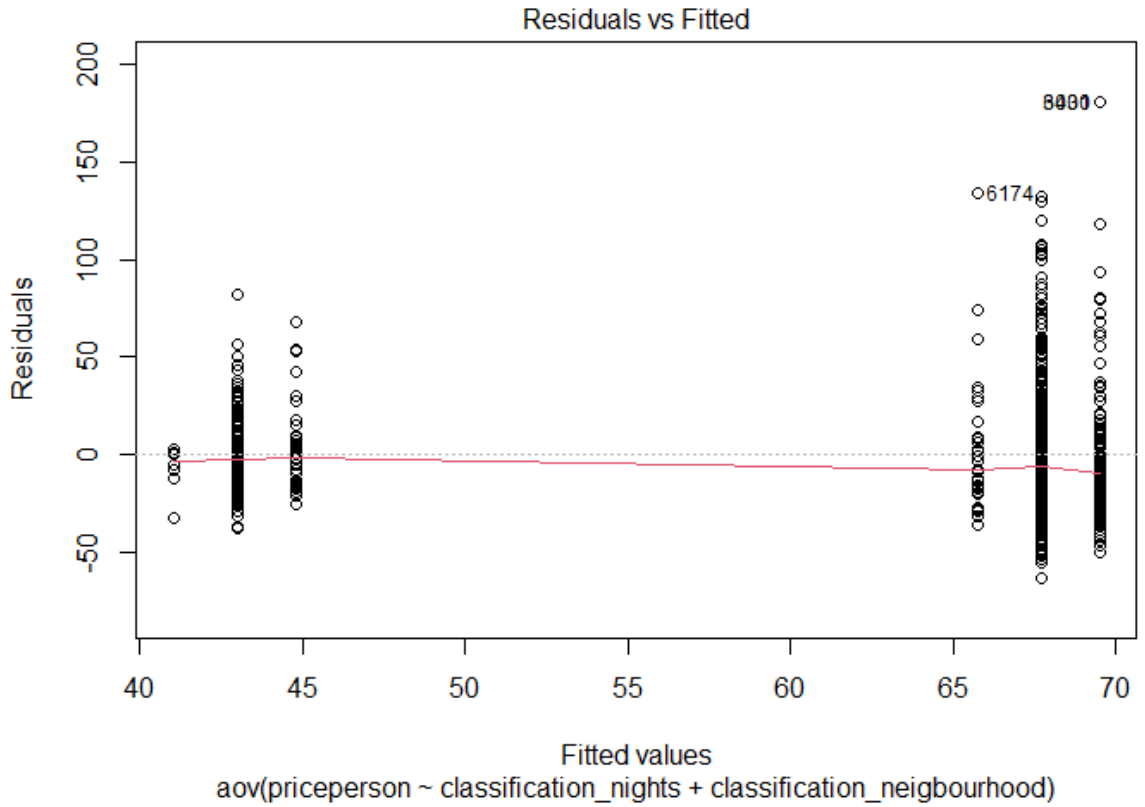


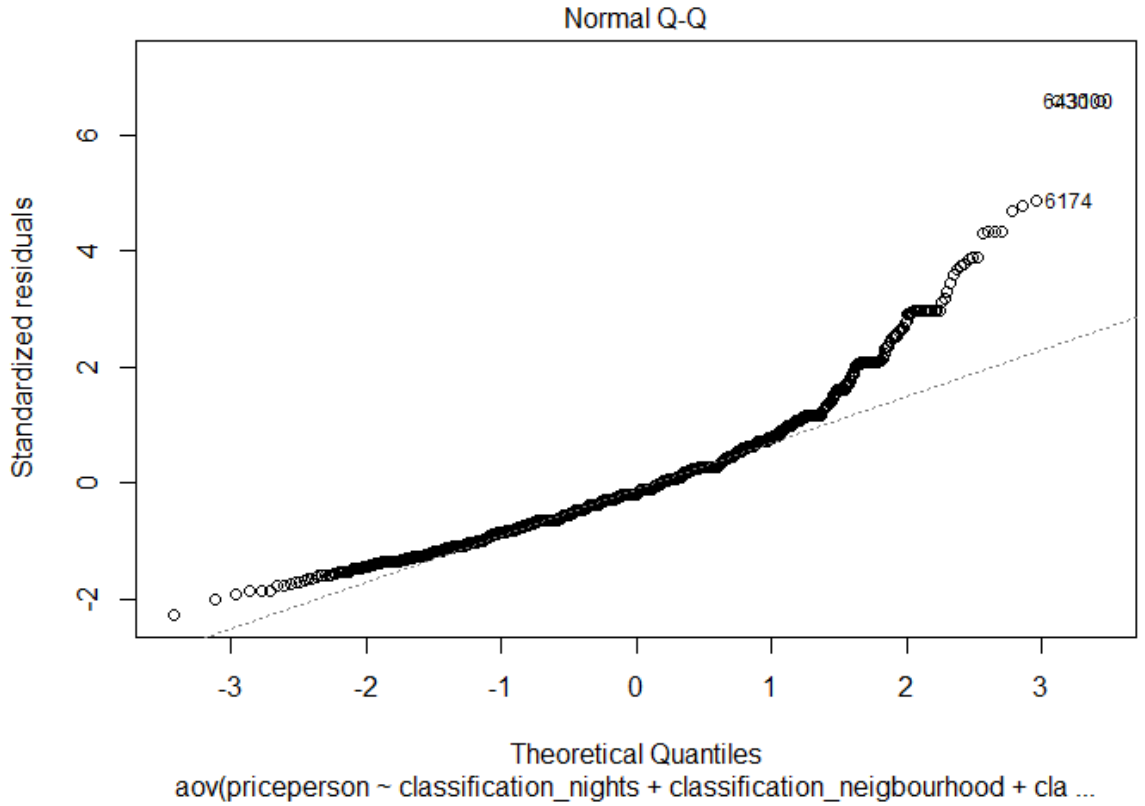
Graph 2

In the second graph we see the classification_neighbourhood on the X-axis and Priceperson on the Y-axis. The classification_neighbourhood can be distinguished in 3 groups: 0 which represents Amsterdam Noord, 1 which represents Amsterdam Centrum and the nameless group which represents the rest of Amsterdam. Priceperson is the Price per person, calculated by the price per night divided by the number of people the location accommodates. As supported by the ANOVA analysis we can clearly see that the neighbourhood has a significant effect on the price, where an Airbnb in the centre, on average, demands a higher price per person than the other neighbourhoods.



Assumptions


Residuals vs Fitted
Residuals
200
150
100
50
0
-50
6174
40
45
50
55
60
65
70
Fitted values
aov(priceperson ~ classification_nights + classification_neigbourhood)


Normal Q-Q
Standardized residuals
6
4
2
0
-2
6174
-3
-2
-1
0
1
2
3
Theoretical Quantiles
aov(priceperson ~ classification_nights + classification_neigbourhood + cla ...

Levene's Test for Homogeneity of Variance (center = median)

```
        Df F value    Pr(>F)
group    5  18.359 < 2.2e-16 ***
      1637
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As the p-value of the Levene’s test is lower than 0.05, the variance across the treatment groups is not equal. This means that the null hypothesis of homoscedasticity is rejected. This violates the assumption of the Anova. However, the sample sizes of the treatment groups are equal. Moreover, a covariate has been added to the research (neighbourhood as it only affects the dependent variable). This means that the rejection of homoscedasticity is not a large problem for this case.

Conclusion

In conclusion, this project investigated the effect of the neighbourhood and the effect of the implemented 30-days rule on the Airbnb prices in Amsterdam. The neighbourhood affects the price per person the greatest, as the prices in Amsterdam Centrum are significantly higher than in Amsterdam Noord. Whether hosts stick to the 30-days rule or not does not have a significant effect on the Airbnb prices in Amsterdam. However, when looking at the groups separately, one can see that those who rent out their apartment between 30 and 60 days, do ask a significantly higher price than those who stick to the 30-day rule and those who rent it out for more than 60 days.

It should be noted that when running the assumptions of the ANOVA are met however at the beginning and end, the values deviate more. The homoscedasticity could be more equally spread. This was also seen in the Levene’s which had a p-value <0.05, which indicates that the variance across the treatment groups is not equal. This implies that further investigation is needed as some assumptions of ANOVA concerning the classifications_nights are not met.