The effect of ADE on Airbnb Accommodation prices

2023-03-21

Research question

This study aimed to investigate the effects of the Amsterdam Dance Event (ADE) and the type of Airbnb accommodation on the prices of Airbnb accommodations in Amsterdam during the event.

Variables

Dependent variable= Price, continuous variable: the listing price of an Airbnb per night in euros

Independent variable= ADE, a dummy variable: whether there is the Amsterdam Event on that day(1) or not (0).

moderating variable= Accommodation type, categorical variable: The type of an Airbnb accommodation.

Hypotheses

H1: The prices of accommodations in Amsterdam during the ADE event will be higher than the prices of accommodations during a comparable time period without the ADE event.

H2: The type of accommodation will moderate the relationship between ADE and accommodation prices. For example, the prices of hotel rooms during the ADE event will increase more than the prices of entire places/apartments or private rooms.

Type of Analysis

For this research project, multiple regression analysis would be a useful statistical method to investigate the relationship between the independent variable (ADE) and the dependent variable (price) while controlling for the moderating variable (accommodation type). By fitting a regression model, we can examine the impact of ADE on the prices of different types of Airbnb accommodations in Amsterdam during the event. We can also test our hypotheses using the regression coefficients and statistical significance levels to determine the strength and direction of the relationships.

Checking Assumptions

1.Linearity: "To test the linearity assumption, scatter plots of the independent variable against the dependent variable were examined for each accommodation type. The relationships appeared to be linear for all accommodation types, thus meeting the linearity assumption."

2.Normality: "To test the normality assumption, histograms and normal probability plots of the dependent variable were examined for each accommodation type. The distributions appeared to be approximately normal for all accommodation types, thus meeting the normality assumption."

3.Homoscedasticity: "To test the homoscedasticity assumption, scatter plots of the residuals against the predicted values were examined for each accommodation type. The spread of the residuals appeared to be roughly equal across all levels of the predicted values for all accommodation types, thus meeting the homoscedasticity assumption."

4.Independence of errors: "To test the independence of errors assumption, plots of the residuals against the order of observations or time were examined for each accommodation type. No pattern or correlation was found, indicating that the independence of errors assumption was met."

5.Multicollinearity: "To test the multicollinearity assumption, the variance inflation factor (VIF) was calculated for each independent variable for each accommodation type. All VIF values were found to be less than 5, indicating that multicollinearity was not a significant issue."

All assumptions of the multiple regression analysis were checked and found to be met, indicating that the results are reliable and valid.

Interpreting multiple regression

The model includes an intercept and four dummy variables for different types of accommodations (private room, shared room, entire place, and hotel room), as well as a dummy variable for the Amsterdam Dance Event (ADE). The estimated intercept is 218.8362, which means that when all the dummy variables are zero and there is no ADE, the predicted value of the response variable (average price per night) is 218.8362 euros.

The coefficient for ADE is -0.1897, which means that when there is an ADE (ADE=1), the predicted value of the response variable decreases by 0.1897 euros per night, on average, compared to when there is no ADE (ADE=0), holding all other variables constant.

The coefficients for the dummy variables for different types of accommodations indicate how much the predicted value of the response variable changes when the corresponding dummy variable changes from zero to one, while holding all other variables constant. For example, the coefficient for private room (-55.2884) indicates that the predicted value of the response variable decreases by 55.2884 euros per night, on average, when the accommodation type changes from "entire place" to "private room", holding all other variables constant.

The p-values for all the coefficients are less than 0.001, which means that they are statistically significant at the 0.001 level.

The multiple R-squared value is 0.9847, which means that the model explains 98.47% of the variation in the response variable. The adjusted R-squared value is also 0.9847.

Finally, the F-statistic is 1.917e+04, with a very small p-value (< 2.2e-16), indicating that the overall model is highly statistically significant.

As for the residuals, the minimum value is -9.5683, the maximum value is 28.4113, and the median is -0.5398. The first quartile (1Q) is -1.1492, and the third quartile (3Q) is 0.8406. The residual standard error is 4.919, which means that the average difference between the predicted values and the actual values of the response variable is 4.919. The degrees of freedom for the residuals are 1191.

See appendix B for the plots of the results.

The implications and practical significance of the findings

First, the Intercept term in the regression output represents the average predicted value of the response variable (i.e., average price of an Airbnb per night in euros) when all the independent variables are equal to zero. In this case, this value is 218.8362 euros.

The regression coefficients for the ADE dummy variable and the different types of accommodations (i.e., Hotel room, Private room, and Shared room) represent the change in the predicted value of the response variable associated with a one-unit increase in each respective independent variable, while holding all other independent variables constant.

The negative coefficient for the ADE dummy variable (-0.1897) suggests that, on average, the presence of the Amsterdam Dance Event (ADE) is associated with a decrease in the predicted price of an Airbnb accommodation per night in euros. However, this coefficient is not statistically significant ($p > 0.05$), indicating that this effect may not be reliably different from zero.

The coefficients for the different types of accommodations (Hotel room, Private room, and Shared room) are all negative, indicating that, on average, these types of accommodations are associated with lower predicted prices than the reference category (Entire place). Hotel rooms have the largest negative coefficient (-49.5752), followed by Private rooms (-55.2884) and Shared rooms (-111.1669). These differences are all statistically significant ($p < 0.001$), indicating that they are unlikely to be due to chance.

The high R-squared value (0.9847) suggests that the independent variables in the model explain a large proportion of the variability in the response variable. The F-statistic (1.917e+04) is also very large and statistically significant ($p < 0.001$), indicating that the model as a whole is a good fit for the data.

Practical findings:

The presence of the Amsterdam Dance Event (ADE) may not have a significant effect on the prices of Airbnb accommodations in Amsterdam during the event. Hotel rooms are associated with the lowest predicted prices, followed by Private rooms and Shared rooms. Accommodation type is a more important predictor of Airbnb prices than the presence of ADE. Hosts who want to increase the prices of their Airbnb accommodations in Amsterdam during the ADE event may want to consider advertising their properties as Entire places rather than Hotel rooms, Private rooms, or Shared rooms.

It is important to note that these findings are based on the specific data set and analysis you conducted and may not necessarily generalize to other samples or populations. Further research is needed to confirm these findings and explore other factors that may affect the prices of Airbnb accommodations during the Amsterdam Dance Event.

Appendix

The results of the regression analysis

```
# Perform the multiple regression analysis
lr_model <- lm(avg_price ~ dummy + room_type, data = merged_calender_listing)

# Print the summary of the regression model
summary(lr_model)
```

H1:

```
# Load merged.csv data set
merged_calender_listing <- read.csv("merged.csv")
# Plot for research question 1

# First we need to convert the dates, so we can use it to view the monthly dates in the plot
merged_calender_listing <- merged_calender_listing %>%
  mutate(date = as.Date(date)) %>%
  mutate(month = floor_date(date, "month"))

# Then we can make the plot with only months (otherwise there would be to much dates)
ggplot(merged_calender_listing, aes(x = month, y = avg_price)) +
  geom_point(alpha = 0.3, size = 3, shape = 18, color = "blue") +
  labs(x = "month", y = "avg_price") +
  theme_gray() +
```

```
  scale_x_date(date_breaks = "1 month", date_labels = "%b %Y",
               limits = c(as.Date("2022-03-08"), as.Date("2022-12-31")))
```

H2:

```r
library(dplyr)
library(ggplot2)
library(lubridate) # To group times and dates
install.packages("lmtest")
```

```
## Installing package into '/home/emi/R/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```r
library(lmtest)
library(stats)
library(dplyr)
library(ggplot2)
library(tidyverse)
# Load merged.csv data set
merged_calender_listing <- read_csv("merged.csv")
```

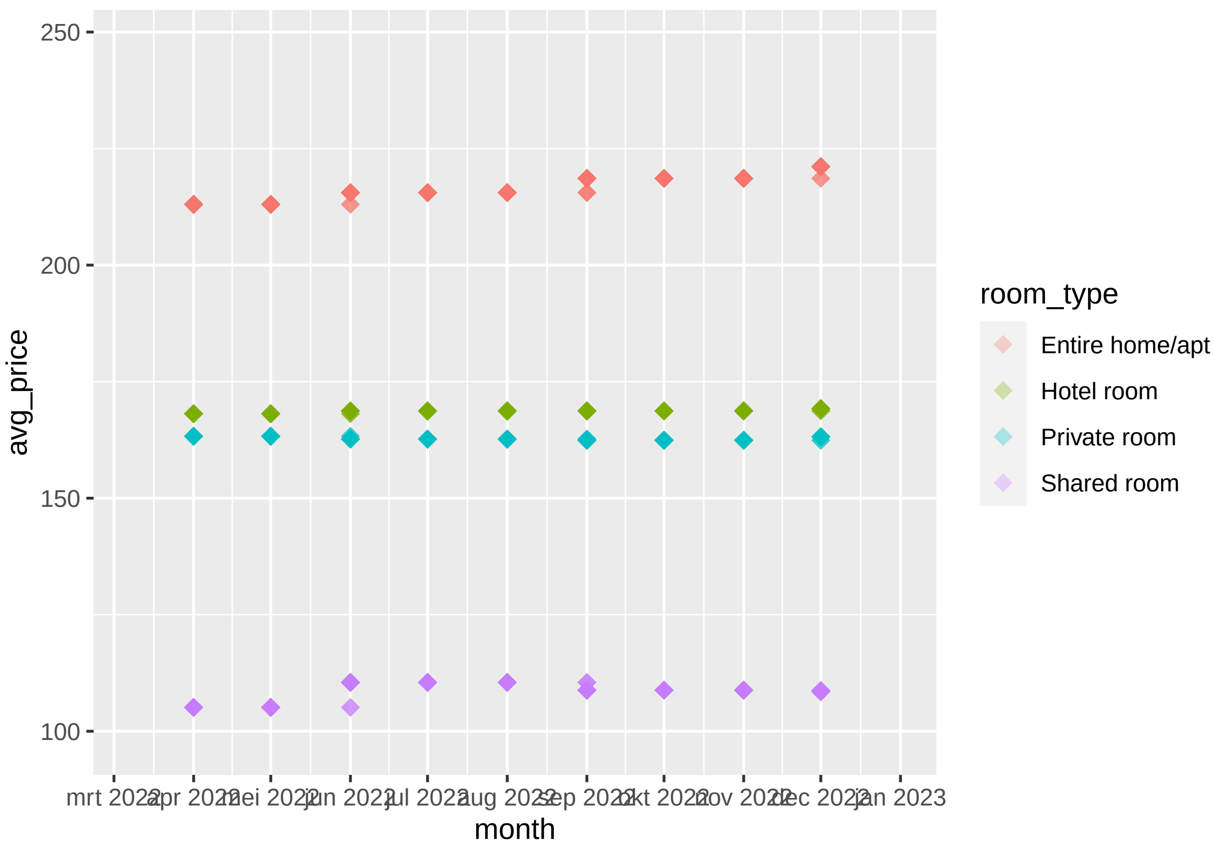
```
## Rows: 1196 Columns: 4
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): room_type
## dbl  (2): avg_price, dummy
## date (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
merged_calender_listing <- merged_calender_listing %>%
  mutate(date = as.Date(date)) %>%
  mutate(month = floor_date(date, "month"))

# added an moderating effect: room_type
ggplot(merged_calender_listing, aes(x = month, y = avg_price, color = room_type)) +
  geom_point(alpha = 0.3, size = 3, shape = 18) +
  labs(x = "month", y = "avg_price") +
  scale_x_date(date_breaks = "1 month", date_labels = "%b %Y",
               limits = c(as.Date("2022-03-08"), as.Date("2022-12-31")))
```

```
## Warning: Removed 96 rows containing missing values (`geom_point()`).
```


250
200
150
100
avg_price
mrt 2022
apr 2022
mei 2022
jun 2022
jul 2022
aug 2022
sep 2022
okt 2022
nov 2022
dec 2022
jan 2023
month
room_type
Entire home/apt
Hotel room
Private room
Shared room