How Covid-19 infected the Airbnb prices in Amsterdam, The Netherlands

Bram van den Bemt, Doga Bayraktar,Demi de Groot, Ellen van Berlo, Sam Messaoudi

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Introduction

For the first step, the data extracted from the Airbnb site has been cleaned. Based on this clean data-set,the variables have been further reduced and a new variable has been added creating the data-set only_price_district. The descriptions of the variables can be found below. Before running the analysis, summary statistics have been obtained. This can be found in the summary stats table below. Later, the data-frame has been visualized with a mean plot, which plots districts by price difference. Results of the mean plot are presented in meanplot_pricedif_district. After accounting for the outliers, a box plot has been plotted that can be found in boxplot_pricedif_district_filtered. Thereafter, the assumptions for the ANOVA (analysis of variance) have been tested to see if this analysis fits the dataset well. First, the Levene's test was applied to check the homogeneity of the variances across the treatment group. Ultimately, as all the points fall approximately along this reference line, we have assumed normality. Lastly, ANOVA has been applied to see if prices significantly differ across the districts that can be found in only_price_disrict.aov below. Using the filtered dataset and checking the assumptions of the ANOVA analysis, Tukey HSD (Tukey Honest Significant Differences, R function: TukeyHSD()) has been ran to perform multiple pairwise-comparison between the means of groups. The results of the Tukey HSD have been presented in a plot and table. These can be found in ANOVA Tukey Plot Table and ANOVA Tukey Plot Graph accordingly. To conclude, partial eta-squared has been applied to see the effect of the ANOVA with a confidence interval of 95%. The results of this can be found in the partial eta-squared table. The results of all the analysis and tables will also be briefly discussed below.

Variable Descriptions

The cleaned data-set "../gen/temp/complete_price_comparison.csv" is composed of the following variables:

ID ID variable is an integer variable. Represents the unique id of every listing.

Host ID Host ID is an integer variable. Represents the unique id of the host that is holding the listing.

Neighbourhood Cleansed Neighbourhood cleansed is a character variable. Represents the district that the listing is located. It has been reduced from 22 neighbourhoods of Amsterdam to 7 districts of Amsterdam.

Room type Room type is a character variable. Represents the accommondation type of a listing. There is 4 room types in total.

Price 12.20 Price 12.20 is numeric variable. Represents the price of the accommondation per night in dollars in the month of December 2020.

Price 01.21 Price 01.21 is numeric variable. Represents the price of the accommondation per night in dollars in the month of January 2021.

Price 02.21 Price 02.21 is numeric variable. Represents the price of the accommondation per night in dollars in the month of February 2021.

Price 03.21 Price 03.21 is numeric variable. Represents the price of the accommondation per night in dollars in the month of March 2021.

Price 04.21 Price 04.21 is numeric variable. Represents the price of the accommondation per night in dollars in the month of April 2021.

Price 05.21 Price 05.21 is numeric variable. Represents the price of the accommondation per night in dollars in the month of May 2021.

With further coercion of the variables and an additional variable the dataframe has been reduced further into "../gen/temp/only_price_district.csv".

Price difference Price difference is a numeric variable. Represents the price difference between December 2020 and May 2021 per night in dollars.

```
## New names:
## * `` -> ...1

## Rows: 16289 Columns: 11

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): neighbourhood_cleansed, room_type
## dbl (9): ...1, id, host_id, price_12.20, price_01.21, price_02.21, price_03....

##
## 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.

##       ...1             id              host_id          neighbourhood_cleansed
##  Min.   :    1   Min.   :    2818   Min.   :     3159   Length:16289
##  1st Qu.: 4073   1st Qu.: 9922340   1st Qu.:  9442056   Class :character
##  Median : 8145   Median :18746775   Median : 27776233   Mode  :character
```

```
##  Mean   : 8145   Mean   :20126011   Mean   : 65095999
##  3rd Qu.:12217   3rd Qu.:29384140   3rd Qu.: 84449589
##  Max.   :16289   Max.   :46953753   Max.   :379498141
##   room_type          price_12.20      price_01.21    price_02.21
##  Length:16289       Min.   :   0.0   Min.   :   0   Min.   :   0.0
##  Class :character   1st Qu.:  94.0   1st Qu.:  95   1st Qu.:  95.0
##  Mode  :character   Median : 125.0   Median : 125   Median : 127.0
##                     Mean   : 151.9   Mean   : 152   Mean   : 152.9
##                     3rd Qu.: 179.0   3rd Qu.: 179   3rd Qu.: 180.0
##                     Max.   :8000.0   Max.   :8000   Max.   :8000.0
##   price_03.21    price_04.21      price_05.21
##  Min.   :   0   Min.   :   0.0   Min.   :   0
##  1st Qu.:  95   1st Qu.:  95.0   1st Qu.:  95
##  Median : 128   Median : 129.0   Median : 129
##  Mean   : 153   Mean   : 153.7   Mean   : 154
##  3rd Qu.: 180   3rd Qu.: 180.0   3rd Qu.: 180
##  Max.   :8000   Max.   :8000.0   Max.   :8000

## New names:
## * `` -> ...1

## Rows: 16289 Columns: 9

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): neighbourhood_cleansed
## dbl (8): ...1, price_12.20, price_01.21, price_02.21, price_03.21, price_04....

##
## 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.

##       ...1       neighbourhood_cleansed  price_12.20      price_01.21
##  Min.   :    1   Length:16289           Min.   :   0.0   Min.   :   0
##  1st Qu.: 4073   Class :character       1st Qu.:  94.0   1st Qu.:  95
##  Median : 8145   Mode  :character       Median : 125.0   Median : 125
##  Mean   : 8145                          Mean   : 151.9   Mean   : 152
##  3rd Qu.:12217                          3rd Qu.: 179.0   3rd Qu.: 179
##  Max.   :16289                          Max.   :8000.0   Max.   :8000
##   price_02.21      price_03.21    price_04.21      price_05.21
##  Min.   :   0.0   Min.   :   0   Min.   :   0.0   Min.   :   0
##  1st Qu.:  95.0   1st Qu.:  95   1st Qu.:  95.0   1st Qu.:  95
##  Median : 127.0   Median : 128   Median : 129.0   Median : 129
##  Mean   : 152.9   Mean   : 153   Mean   : 153.7   Mean   : 154
##  3rd Qu.: 180.0   3rd Qu.: 180   3rd Qu.: 180.0   3rd Qu.: 180
##  Max.   :8000.0   Max.   :8000   Max.   :8000.0   Max.   :8000
##  price_difference
##  Min.   :-987.000
##  1st Qu.:   0.000
##  Median :   0.000
##  Mean   :   2.039
##  3rd Qu.:   0.000
##  Max.   :7200.000
```

Summary Statistics

```
## New names:
## * `` -> ...1
```

```
## Rows: 7 Columns: 5
```

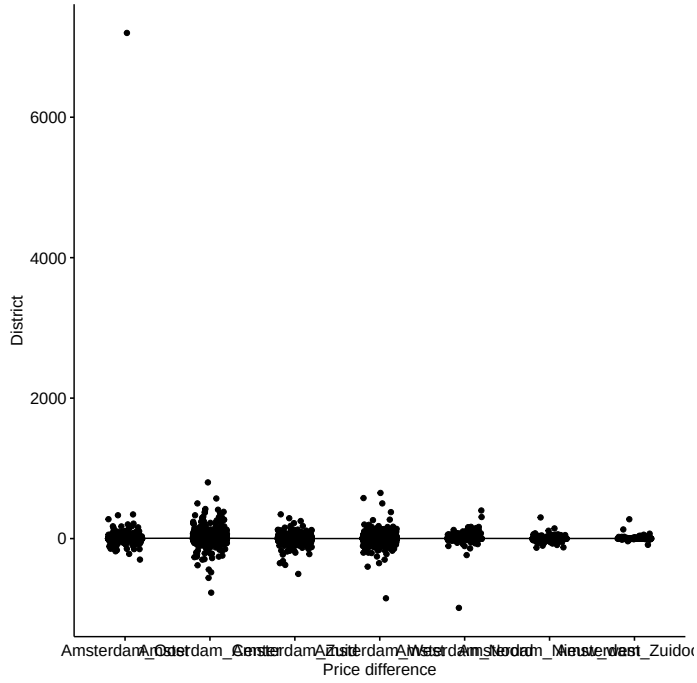
```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): neighbourhood_cleansed
## dbl (4): ...1, count, mean, sd
```

```
##
## 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.
```

```
##       ...1     neighbourhood_cleansed     count             mean
##  Min.   :1.0   Length:7               Min.   : 289.0   Min.   :0.3295
##  1st Qu.:2.5   Class :character       1st Qu.: 928.5   1st Qu.:0.8048
##  Median :4.0   Mode  :character       Median :2706.0   Median :2.1488
##  Mean   :4.0                          Mean   :2327.0   Mean   :2.1924
##  3rd Qu.:5.5                          3rd Qu.:3251.5   3rd Qu.:2.9474
##  Max.   :7.0                          Max.   :4934.0   Max.   :5.3645
##        sd
##  Min.   : 18.54
##  1st Qu.: 22.76
##  Median : 29.37
##  Mean   : 46.24
##  3rd Qu.: 45.04
##  Max.   :140.18
```

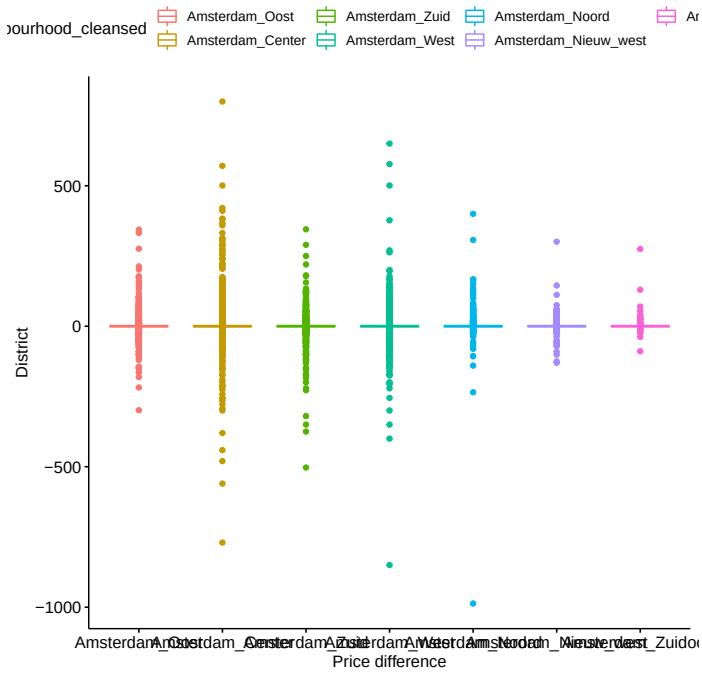
Summary of the variables neighbourhood cleansed and price difference. Count of the accommodations with price difference according to districts (neigbourhood cleansed), their mean and standard deviation.

Mean Plots



Plot of district by price difference by means, added error bars. An outlier is seen for price difference as it is larger than 6000.

Box Plots



Plot of district by price difference and color by district after filtering out the outlier > 6000.

Levene's Test of Homogenity

```
## New names:
## * `` -> ...1
```

```
## Rows: 2 Columns: 4
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ...1
## dbl (3): Df, F value, Pr(>F)
```

```
##
## 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.
```

```
##      ...1                 Df           F value          Pr(>F)
##  Length:2           Min.   :    6   Min.   :63.85   Min.   :0
##  Class :character   1st Qu.: 4075   1st Qu.:63.85   1st Qu.:0
##  Mode  :character   Median : 8144   Median :63.85   Median :0
##                     Mean   : 8144   Mean   :63.85   Mean   :0
##                     3rd Qu.:12212   3rd Qu.:63.85   3rd Qu.:0
##                     Max.   :16281   Max.   :63.85   Max.   :0
##                                     NA's   :1       NA's   :1
```

The p-value is significant: Pr(<F) 2.2e-16 < 0.05. Concluding that there is no evidence to suggest that the variance across groups is significantly different. Therefore, we can assume the homogeneity of variances in the different treatment groups.

Analysis of Variance (ANOVA)

```
## New names:
## * `` -> ...1
```

```
## Rows: 1000 Columns: 8
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (8): ...1, X.Intercept., neighbourhood_cleansedAmsterdam_Nieuw_west, nei...
```

```
##
## 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.
```

```
##       ...1         X.Intercept.   neighbourhood_cleansedAmsterdam_Nieuw_west
##  Min.   :   1.0   Min.   :2.417   Min.   :-7.668
##  1st Qu.: 250.8   1st Qu.:4.816   1st Qu.:-4.917
##  Median : 500.5   Median :5.373   Median :-4.215
##  Mean   : 500.5   Mean   :5.380   Mean   :-4.210
##  3rd Qu.: 750.2   3rd Qu.:5.945   3rd Qu.:-3.458
##  Max.   :1000.0   Max.   :8.116   Max.   :-1.209
```

```
##  neighbourhood_cleansedAmsterdam_Noord neighbourhood_cleansedAmsterdam_Oost
##  Min.   :-7.613                        Min.   :-7.789
##  1st Qu.:-3.873                        1st Qu.:-5.327
##  Median :-2.880                        Median :-4.609
##  Mean   :-2.900                        Mean   :-4.646
##  3rd Qu.:-1.820                        3rd Qu.:-3.977
##  Max.   : 1.064                        Max.   :-1.125
##  neighbourhood_cleansedAmsterdam_West neighbourhood_cleansedAmsterdam_Zuid
##  Min.   :-7.836                       Min.   :-8.858
##  1st Qu.:-5.552                       1st Qu.:-5.633
##  Median :-4.944                       Median :-5.018
##  Mean   :-4.961                       Mean   :-5.056
##  3rd Qu.:-4.293                       3rd Qu.:-4.423
##  Max.   :-1.439                       Max.   :-1.567
##  neighbourhood_cleansedAmsterdam_Zuidoost
##  Min.   :-7.654
##  1st Qu.:-4.235
##  Median :-3.267
##  Mean   :-3.212
##  3rd Qu.:-2.236
##  Max.   : 2.833
```

The p- value (7.46e-10) is less then the significance level of 0.05. Therefore we can conclude that there is a significance difference between the different districts.

Partial Eta-Squared

```
## New names:
## * `` -> ...1
```

```
## Rows: 1 Columns: 6
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Parameter
## dbl (5): ...1, Eta2, CI, CI_low, CI_high
```

```
##
## 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.
```

```
##       ...1    Parameter             Eta2               CI
##  Min.   :1   Length:1           Min.   :0.003309   Min.   :0.95
##  1st Qu.:1   Class :character   1st Qu.:0.003309   1st Qu.:0.95
##  Median :1   Mode  :character   Median :0.003309   Median :0.95
##  Mean   :1                      Mean   :0.003309   Mean   :0.95
##  3rd Qu.:1                      3rd Qu.:0.003309   3rd Qu.:0.95
##  Max.   :1                      Max.   :0.003309   Max.   :0.95
##      CI_low            CI_high
##  Min.   :0.001726   Min.   :1
##  1st Qu.:0.001726   1st Qu.:1
##  Median :0.001726   Median :1
```

```
##  Mean   :0.001726   Mean   :1
##  3rd Qu.:0.001726   3rd Qu.:1
##  Max.   :0.001726   Max.   :1
```

Confidence intervals being perfectly zero is very rare, but it is most likely due to the fact that the lockdown was a rare circumstance in the corresponding time frame.

ANOVA Tukey Plot Table

```
## New names:
## * `` -> ...1
```

```
## Rows: 1000 Columns: 8
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (8): ...1, X.Intercept., neighbourhood_cleansedAmsterdam_Nieuw_west, nei...
```

```
##
## 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.
```

```
##       ...1         X.Intercept.   neighbourhood_cleansedAmsterdam_Nieuw_west
##  Min.   :   1.0   Min.   :2.750   Min.   :-7.1597
##  1st Qu.: 250.8   1st Qu.:4.769   1st Qu.:-4.9839
##  Median : 500.5   Median :5.361   Median :-4.2026
##  Mean   : 500.5   Mean   :5.386   Mean   :-4.2344
##  3rd Qu.: 750.2   3rd Qu.:5.967   3rd Qu.:-3.4934
##  Max.   :1000.0   Max.   :8.268   Max.   :-0.7196
##  neighbourhood_cleansedAmsterdam_Noord neighbourhood_cleansedAmsterdam_Oost
##  Min.   :-7.917                        Min.   :-7.734
##  1st Qu.:-3.938                        1st Qu.:-5.297
##  Median :-2.868                        Median :-4.612
##  Mean   :-2.922                        Mean   :-4.647
##  3rd Qu.:-1.905                        3rd Qu.:-3.974
##  Max.   : 1.040                        Max.   :-1.728
##  neighbourhood_cleansedAmsterdam_West neighbourhood_cleansedAmsterdam_Zuid
##  Min.   :-8.222                       Min.   :-7.971
##  1st Qu.:-5.622                       1st Qu.:-5.750
##  Median :-4.947                       Median :-5.019
##  Mean   :-4.952                       Mean   :-5.049
##  3rd Qu.:-4.297                       3rd Qu.:-4.335
##  Max.   :-2.029                       Max.   :-2.288
##  neighbourhood_cleansedAmsterdam_Zuidoost
##  Min.   :-7.129
##  1st Qu.:-4.279
##  Median :-3.282
##  Mean   :-3.220
##  3rd Qu.:-2.284
##  Max.   : 3.567
```

The plot which shows differences between Amsterdam- Center and all the other districts gives a significant p-value below 0.05. For all the other differences between districts, there is no significant difference.

ANOVA Tukey Plot Graph

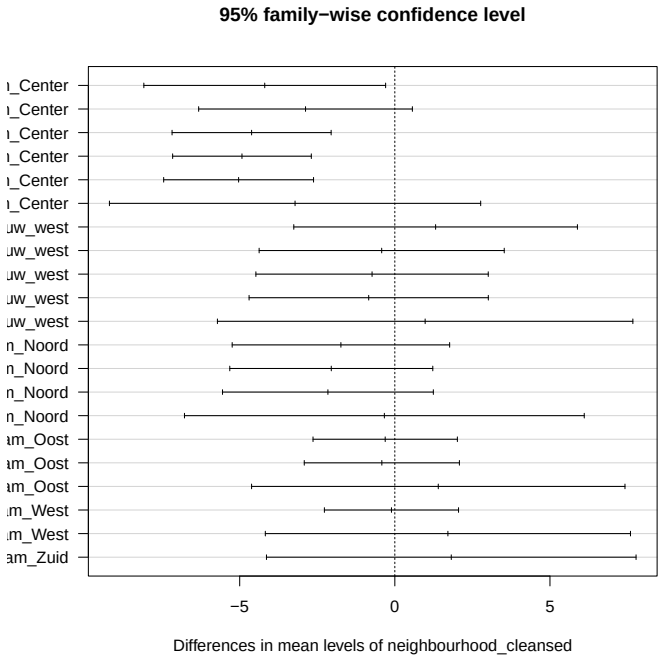


Table for the Tukey HSD test.