

Airbnb ADE Amsterdam manuscript

Group 06

27/03/2022

This part of the report is meant to showcase the amount of reviews in the year of 2021 and give a basic analysis of the neighbourhoods in Amsterdam during the year of 2021.

First of all, an overview of the total reviews left by guests throughout 2021 was created. Herewith, a line chart per week was executed which can be found below (Figure 1). Here, one can see the number of weekly reviews left by guests in 2021. One can claim that there are several spikes in the time frame of 2021. Firstly, an increasing line of written reviews takes place around week 30, which was in the summer holiday in the Netherlands (around July and August). Secondly, a spike around week 42 till week 44 can be spotted. This is two weeks after the Amsterdam Dance Event took place.

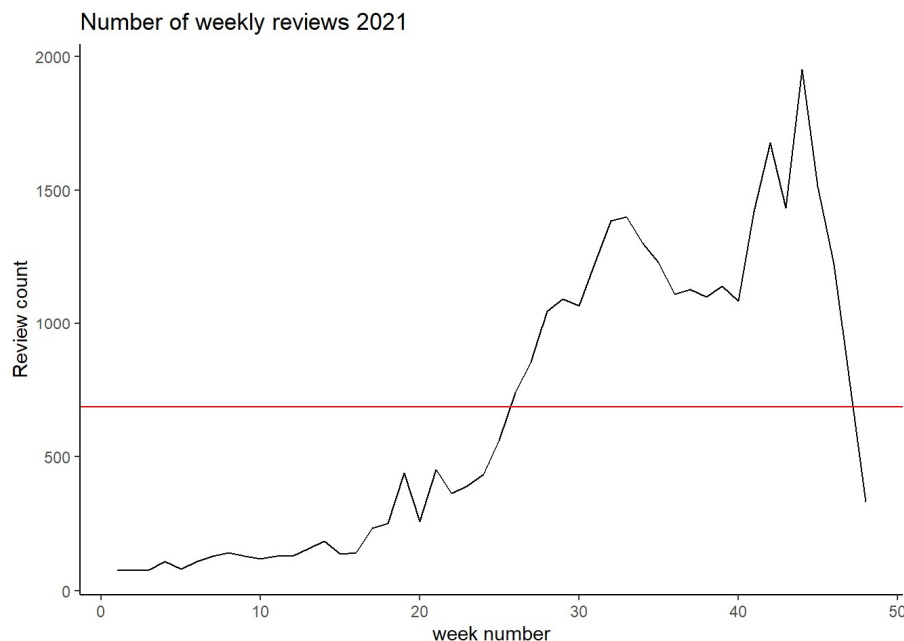


Fig 1. Weekly reviews during the ADE review period 2021-line chart

The bar chart below demonstrates the number of daily reviews during the ADE review period. Here, a spike can be spotted around October 18th, which could imply that guests left a review right after their Airbnb stay. Due to ADE ending on October 17th, this spike around October 18th is considered logic. Another spike can be seen around October 25th, which is still within the two week time frame of being able to leave a review.

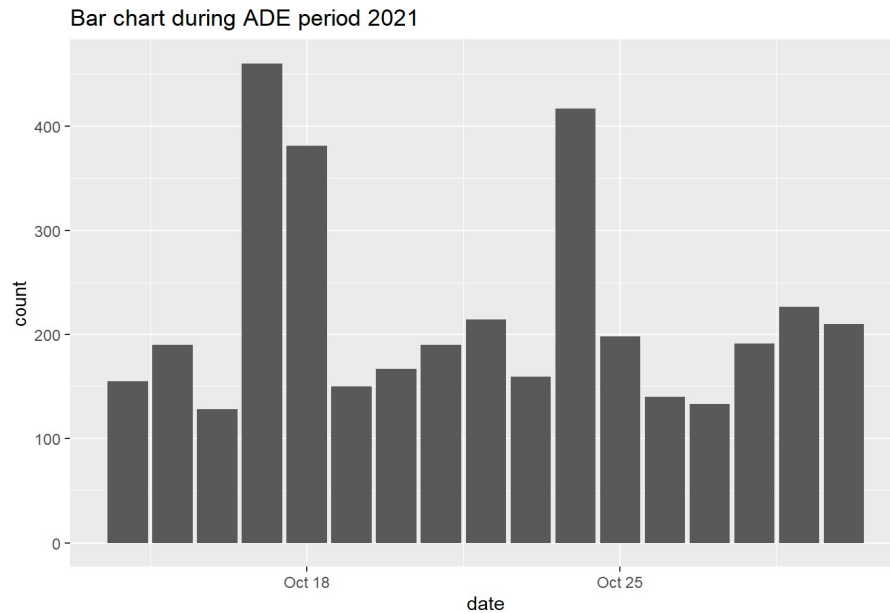


Fig 2. Daily reviews during ADE review period 2021-bar chart

In the corresponding bar chart below, an overview of the daily reviews during 2021 can be found. Here, an increase can be noticed starting in July, which remains stable until the end of October. At the end of October 2021, another increase in the number of reviews left by guests can be found. This corresponds to the previous models, where the number of reviews written increases after ADE has taken place. The highest value can be spotted at the end of October, giving a value close to 600 reviews on a daily basis.

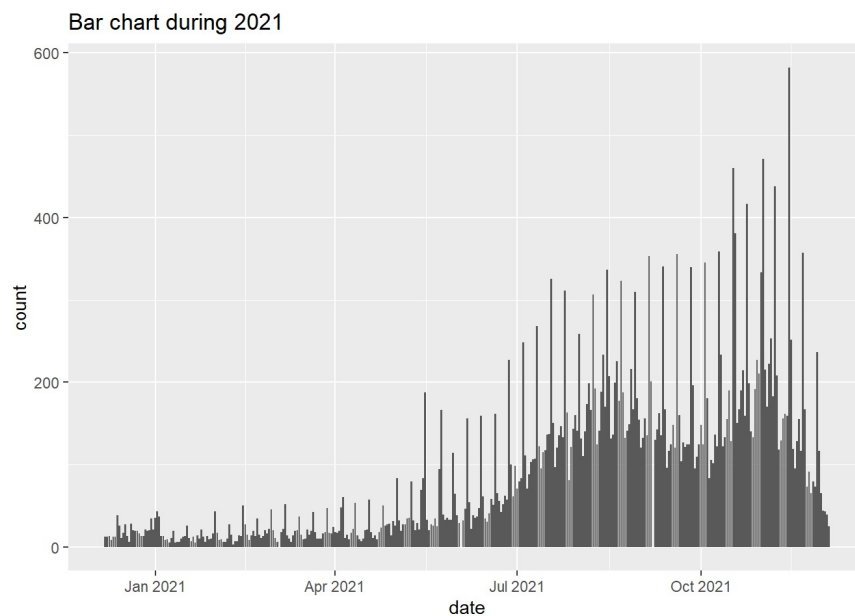


Fig 3. Daily reviews during 2021-bar chart

Below, a line chart demonstrates the total reviews left by guests throughout the year 2021. The line chart exhibits a similar view as the above mentioned bar chart. An increase starting around July continues to another increase at the end of October 2021. Herewith, another model confirms the spotted increase in number of reviews left by guests after ADE has taken place.

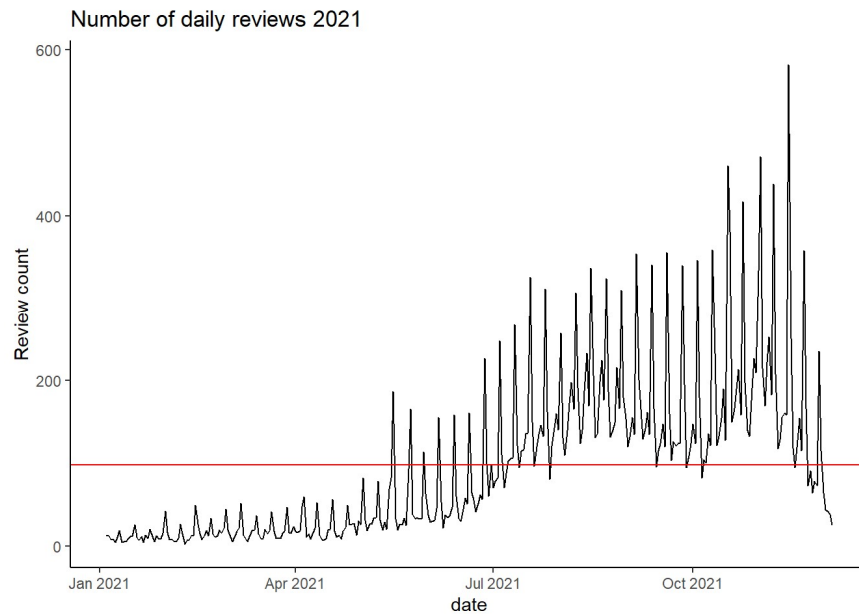


Fig 4. Daily reviews during 2021-bar chart

A following plot contains the weekly number of reviews written, where the week 42 to 44 are marked red in order to clearly state the weeks where the Amsterdam Dance Event might have left an impact. Guests have up to two weeks after their stay to leave their review which is taken into account in this plot. As the Amsterdam Dance Event took place in week 41 of 2021, guests were able to leave reviews up to the latest week 44. One could claim that indeed in those weeks after the Amsterdam Dance Event, the number of written reviews were more than the other weeks of 2021. In the weeks 42 to 44, the most reviews were left by guests in Amsterdam.

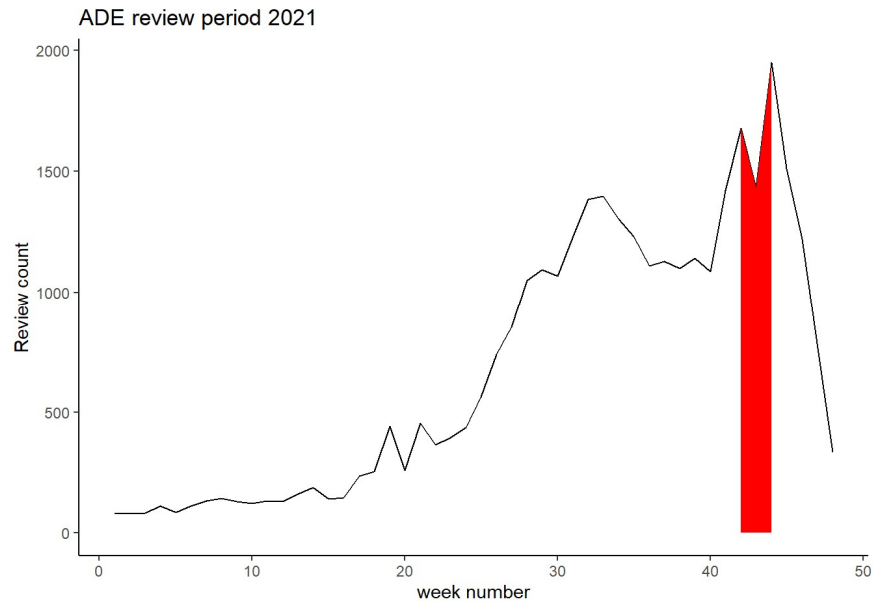


Fig 5: ADE review period 2021-line chart

The above mentioned plot was also recreated for previous years. Here, 2018, 2019 and 2020 were taken into account. Starting with 2020, as this is an interesting year to look at, the weekly number of reviews after ADE were tremendously low. This can be explained by the fact that ADE 2020 did not take place physically. The organization made some type of online ADE available where visitors could watch shows online without being physically present in Amsterdam. Due to the Coronavirus, such an enormous event was cancelled by means of high chances of an infection outbreak. Due to these circumstances, no additional stream of tourism was initiated as there was no physical presence of the event. This possibly led to less bookings of Airbnb's and subsequently, less reviews were written.

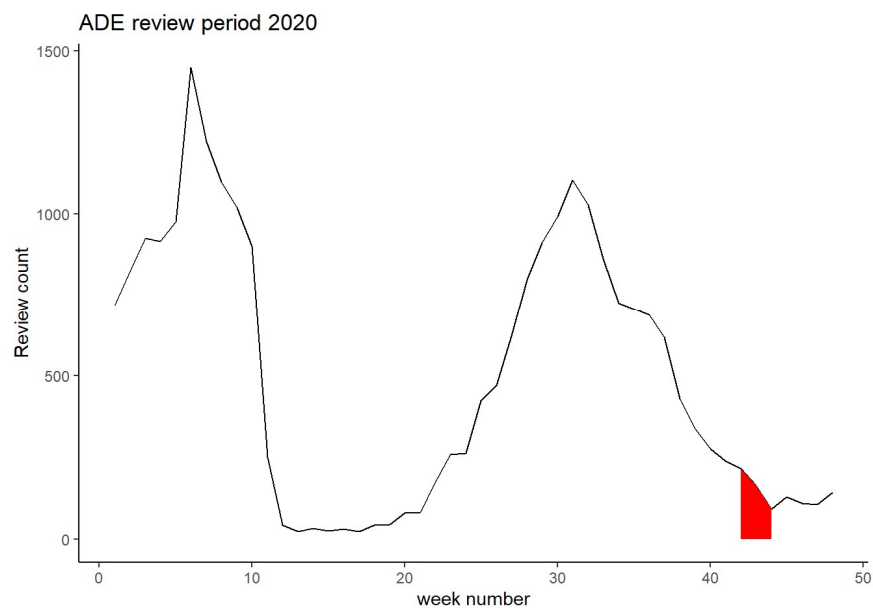


Fig 6: ADE review period 2020-line chart

Interestingly, the number of reviews left in 2019 around the same time period, was quite high. This can be explained by the fact that ADE did take place physically in 2019. Overall in 2019, there are quite some spikes that can be noticed. This may be due to 2019 not being impacted by the pandemic at all. More tourism, and therefore also more listings and reviews, in the spring and summer can explain the spikes starting from week 15. Additionally, a spike around ADE can be seen. Due to the review count being quite high already on an average level, the spikes can be perceived as less invasive. In the year 2021, which was partially affected by the pandemic, the spikes are larger and more spontaneous than in 2019. This can be dedicated to the pandemic which has limited the tourism and number of listings throughout the year.

Part 2 of analysis: Regression analysis (based on the regression building block)

In this model “the reviews measured in a day in the city of Amsterdam” is regressed as an outcome variable on the predictor variables “room type”, which indicates the type of accommodation listed on Airbnb, the “neighbourhoods, referring to one of the wider scale neighbourhoods in Amsterdam, and lastly”ade” which means whether the period is within the review time period of two weeks. There is also an interaction added between these two last variables to best understand the effects. Adaptions of this regression model used in the analysis a logarithmic transformation of the outcome variable, and a model that removes the outliers detected by the screening process in the analysis.

Outcome variable	Predictor variable 1	Predictor variable 2	Predictor variable 3
reviews	room type	neighbourhood	ade
per day	1-Private room 2-Entire home/apt 3-Hotel Room 4-Shared Room	1-centre 2-north 3-east 4-south 5-west 6-new_west 7-southeast	*whether ade is going yes/no 1/0

Fig 8: Variables and levels used in the regression model

The formulation of the model is as followed:

```
regression_ade <-  
lm(reviews~as.factor(neighbourhood_cleansed)+as.factor(room_type)+ade+ade*a  
s.factor(neighbourhood_cleansed)
```

1. Evaluate linear model assumptions

Firstly, the model assumptions can be estimated through different plots. As can be seen, the model has a poor fit initially.

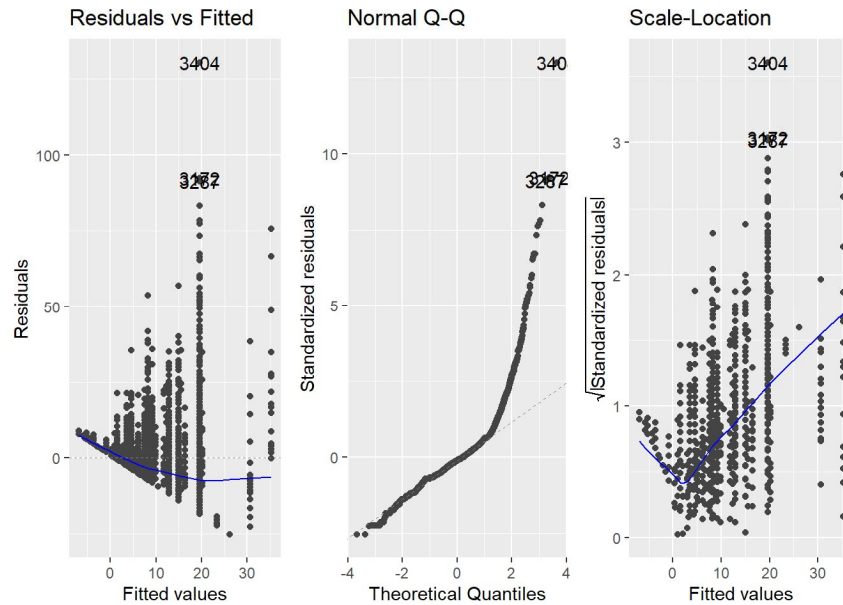


Fig 9: initial model for regression

Making a log transformation for the outcome variable in the model.

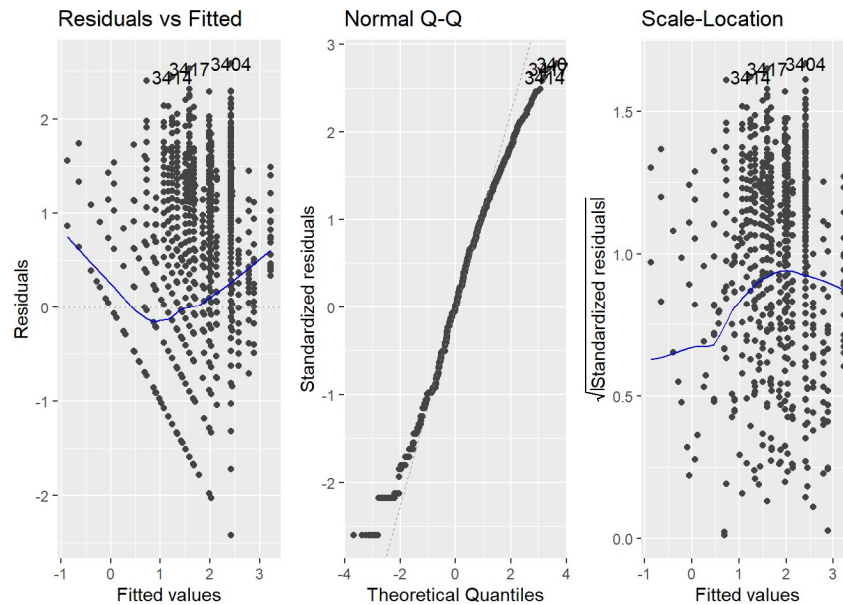


Fig 9: Log transformation on outcome variable

2. Outlier screening

The screening of outliers was done through the measures of leverage and influence. As can be seen, the model plotted against the number of observations showcases some relatively peaks in the model towards Cook's distance, which is plotted against the observations.

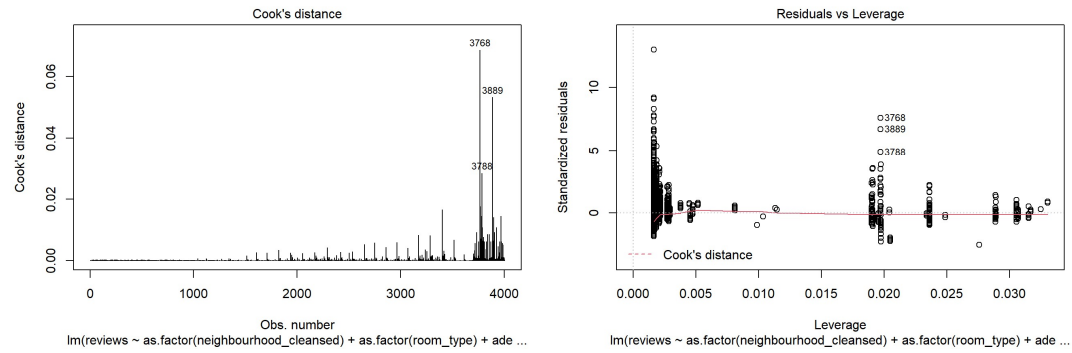


Fig 10 and 11: Outlier measurement before

The outliers were detected based on the leverage and cooks distance, measuring the greatest deviating effects in the model.

Outliers in model summary				
as.factor(room_type)	as.factor(neighbourhood_cleansed)	ade	leverage	cooks_dist
1	1	1	0.019708633	0.06859528
1	1	1	0.019708633	0.05326130
1	1	1	0.019708633	0.02840515
2	1	1	0.019726212	0.01761478
1	1	0	0.001659189	0.01650366
2	5	1	0.019085661	0.01452678

Fig 10: Outlier measurement after logarithmic transformation and outlier

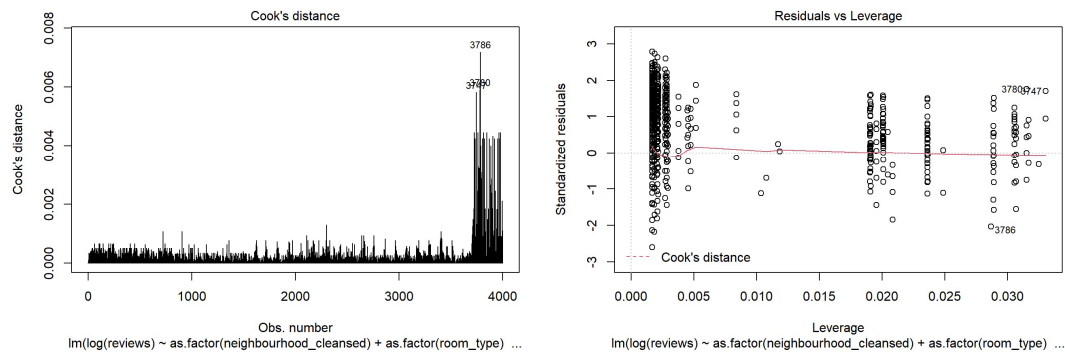


Fig 11 and 12: Outlier measurement after logarithmic transformation and outlier

3. model reporting

A summary is given for the models on fit statistics for the models displayed. Parameters such as the R-squared, and the adjusted R-squared show the model improves after the alteration of the variables and taking the outliers out of the model.

Review count per neighbourhood and ade active			
	Frequency of reviews		
	log(reviews)		
	Full model (1)	Transformation (2)	Outliers excluded (3)
Neighbourhood/price	-13.406*** (0.614)	-1.255*** (0.057)	-1.259*** (0.057)
as.factor(neighbourhood_cleansed)3	-11.392*** (0.556)	-0.907*** (0.052)	-0.919*** (0.052)
as.factor(neighbourhood_cleansed)4	-10.394*** (0.558)	-0.740*** (0.052)	-0.744*** (0.052)
as.factor(neighbourhood_cleansed)5	-6.742*** (0.525)	-0.389*** (0.049)	-0.393*** (0.049)
as.factor(neighbourhood_cleansed)6	-12.036*** (0.630)	-1.072*** (0.058)	-1.077*** (0.058)
as.factor(neighbourhood_cleansed)7	-12.972*** (0.741)	-1.502*** (0.069)	-1.506*** (0.069)
as.factor(room_type)2	-4.625*** (0.348)	-0.435*** (0.032)	-0.435*** (0.032)
as.factor(room_type)3	-11.873*** (0.511)	-1.556*** (0.047)	-1.558*** (0.047)
as.factor(room_type)4	-9.101*** (0.965)	-1.440*** (0.090)	-1.388*** (0.091)
ade1	15.656*** (1.439)	0.803*** (0.133)	0.833*** (0.135)
as.factor(neighbourhood_cleansed)2:ade1	-12.925*** (2.280)	-0.086 (0.211)	-0.116 (0.212)
as.factor(neighbourhood_cleansed)3:ade1	-11.529*** (2.018)	-0.180 (0.187)	-0.215 (0.188)
as.factor(neighbourhood_cleansed)4:ade1	-8.442*** (2.143)	0.095 (0.199)	0.065 (0.200)
as.factor(neighbourhood_cleansed)5:ade1	-8.337*** (2.021)	0.058 (0.187)	0.028 (0.188)
as.factor(neighbourhood_cleansed)6:ade1	-13.360*** (2.322)	-0.294 (0.215)	-0.324 (0.216)
as.factor(neighbourhood_cleansed)7:ade1	-17.354*** (2.378)	-1.025*** (0.220)	-1.055*** (0.221)
Constant	19.606*** (0.409)	2.417*** (0.038)	2.422*** (0.038)
Observations	4,005	4,005	3,999
R ²	0.293	0.366	0.366
Adjusted R ²	0.290	0.364	0.364
Residual Std. Error	10.044 (df = 3988)	0.931 (df = 3988)	0.931 (df = 3982)
F Statistic	103.437*** (df = 16; 3988)	144.115*** (df = 16; 3988)	143.955*** (df = 16; 3982)
Significance levels	* p<0.1; ** p<0.05; *** p<0.01		

Fig 13: Outlier measurement after logarithmic transformation and outlier

4. Visualize variables in the model

The variable “neighbourhood” can also be set off against time to display the effect throughout the year. It can be seen that the review count per neighbourhood moves relatively evenly between the different neighbourhood.

The daily reviews per neighborhood show that the city centre has the largest number of reviews throughout the year. There is a clear trend recognizable around July and again around October. This is in line with the previously reviewed plots, which also shows a trend of an increase in the number of reviews in July and at the end of October. The city center has the highest amount of reviews per day on average.

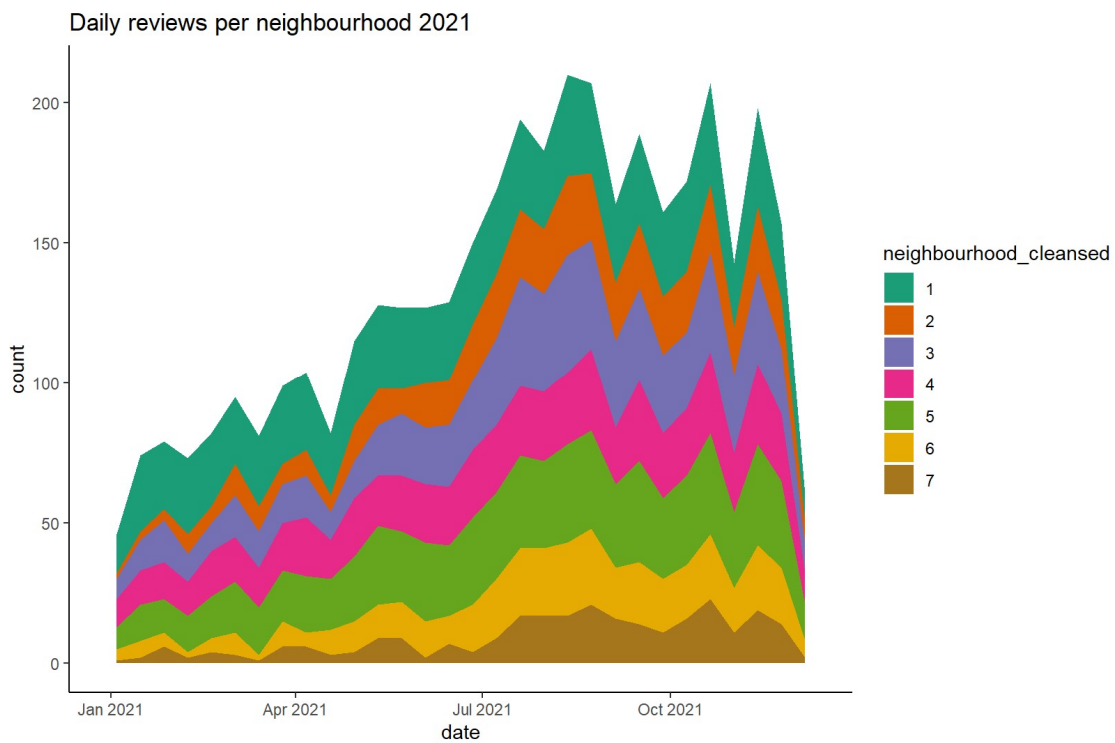


Fig 14: Daily reviews per neighbourhood- graph

**Neighbourhoods 1-centre 2-north 3-east 4-south 5-west 6-new_west 7-southeast*