

# **The Effect of Brexit Announcement on Airbnb Prices in London: Empirical Evidence**

Semester Project Final Report

Zeyad Dahab, Matea Kocevskaa, Aigul Saiapova

Professor Xiao Liu  
Data-Driven Decision Making  
Spring 2019

*Airbnb, Inc.* entered the global lodging and real estate rental industries in its inception back in 2008 with an entirely new concept<sup>1</sup>. It was the first company of its size and kind to connect individual landlords and travelers for short-term apartment rentals for the purposes of travel accommodation, and has since grown to comfortably occupy a sizeable market share in the lodging industry. At an estimated size of \$38 billion worldwide<sup>2</sup>, Airbnb spread over many countries on almost all continents.

One of the biggest Airbnb hubs is located in London, United Kingdom's busy, touristy capital, counting more than 77 thousand listings<sup>3</sup>. This is well explained: the city welcomed an average of 17 million visitors per year since 2008<sup>4</sup>, and still remains a popular destination for local and international tourists. However, in June 2016, a major political upheaval that directly influences United Kingdom's immigration policy and border control, Brexit, has entered the equation of the country's tourism prospects. In this report, we investigate if the announcement of this radical political shift has had any effect on average prices of Airbnb listings, bringing in empirical evidence from data.

## **Part One: The Reasoning**

According to a research report<sup>5</sup>, the UK economy has shrunk by 10 million of investment dollars since Brexit was first announced. The political movement of leaving the European Union was chosen as the most desirable political strategy by the people of the United Kingdom in a referendum of 23 June 2016, a tight race between the supporters and the opposition. First of its kind and requested by one of the biggest economies in the entire political bloc, Brexit has since become synonymous to chaos, as the future of many European citizens supporting the UK economy as a source of labor has become provisional and highly dependant on the outcomes of the Brexit deal. Immigration rules are bound to become stricter, which cannot positively affect ease of travel in and out of the country (and its financial center, London). In this light, we worked under the assumption that the expected fall in demand for lodging resulted from declining inbound travel from European Union, one of the key markets of the industry, would incentivize property holders to raise prices in the months preceding the completion of the exit procedure.

---

<sup>1</sup> "Airbnb." *Wikipedia*, Wikimedia Foundation, 22 Apr. 2019, [en.wikipedia.org/wiki/Airbnb](https://en.wikipedia.org/wiki/Airbnb).

<sup>2</sup> Jaaskelainen, Liisa. "Topic: Airbnb." *Www.statista.com*, [www.statista.com/topics/2273/airbnb/](https://www.statista.com/topics/2273/airbnb/).

<sup>3</sup> "London. Adding Data to the Debate." *Inside Airbnb*, [insideairbnb.com/london/](https://insideairbnb.com/london/).

<sup>4</sup> "Inbound Visitors to London 2008-2017 | UK Statistic." *Statista*, [www.statista.com/statistics/487467/overseas-visits-to-london-united-kingdom/](https://www.statista.com/statistics/487467/overseas-visits-to-london-united-kingdom/).

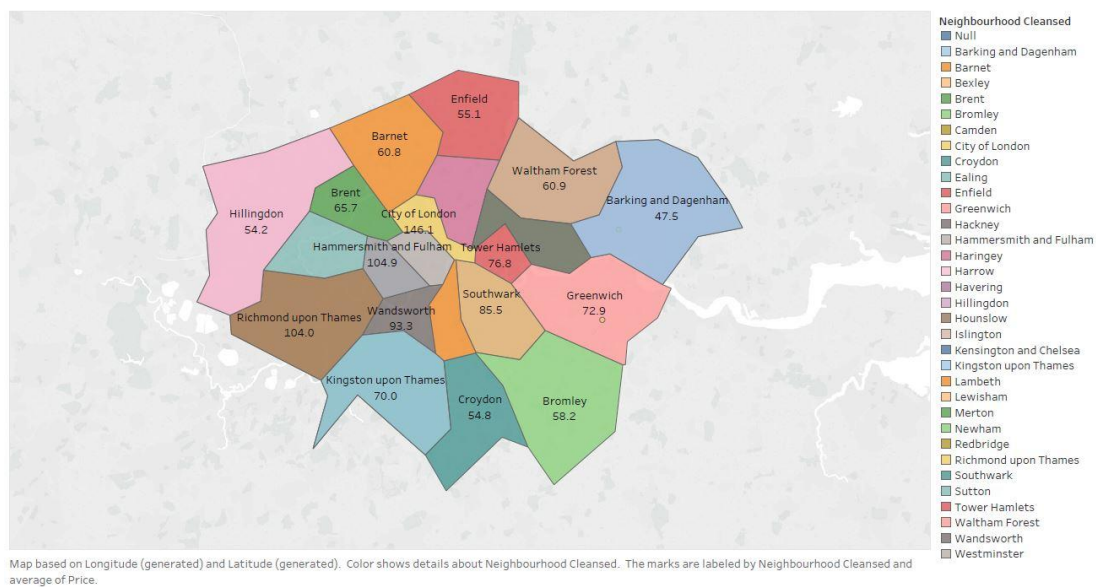
<sup>5</sup> Taylor, Chloe. "British Firms Have Diverted \$10 Billion of Investment to the EU Due to Brexit, Study Claims." *CNBC*, CNBC, 11 Feb. 2019, [www.cnbc.com/2019/02/11/uk-firms-divert-10-billion-of-investment-to-eu-over-brexit-study.html](https://www.cnbc.com/2019/02/11/uk-firms-divert-10-billion-of-investment-to-eu-over-brexit-study.html).

## Part Two: The Data

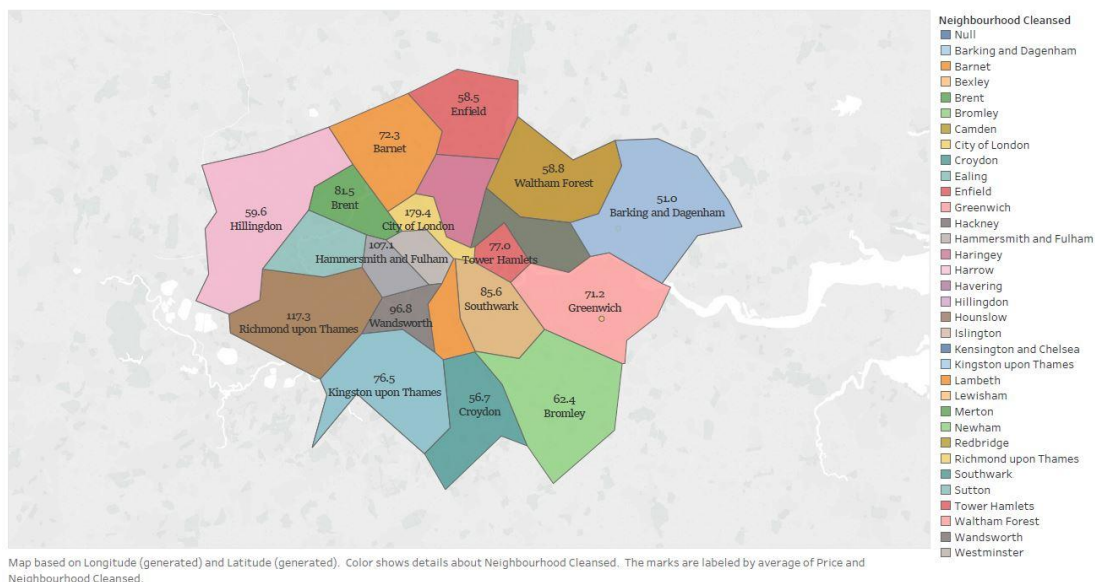
To test our hypothesis, we obtained the data from Inside Airbnb, a comprehensive web database on listings in world's major cities. We chose two periods: a few months before and after the announcement and aggregated the data to run diagnostics described below.

### A: Tableau.

London is a large city, resting on 607 square miles of land that are divided into nine fare zones and 32 boroughs. The farther from the city center, i.e. the City of London, the less densely populated and touristy it becomes. It made sense to us to split individual listings into boroughs and visualize them in Tableau (fig. 1&2 below).



**Figure 1.** Tableau map of listings in London by borough, February 2016.



**Figure 2.** Tableau map of listings in London by borough, October 2016.

These maps label average price for a listing in each borough, which, if examined closely, seems to have risen in the majority of the boroughs. Taking this as a positive sign, we moved on to creating a linear regression model to test our hypothesis.

### B. Data Summary

We cleaned the data set following two methods: i) we used our intuition to remove certain variables that did not have place in our testings such as HostID and ii) we computed univariate and bivariate analysis to get a better understanding of the rest of the variables. The following table shows the final variables we decided to use for our project:

Variables/Type	Continuous	Nominal/binary
Dependent	Price	
Independent	Accommodates Bathrooms Bedrooms Review Scores	Superhost Room type Neighbourhood Post-Brexit

Later in the analysis, some of these variables were further decoded to capture more detailed effects. For example, the variable neighbourhood was decoded into specific zones.

For the purpose of this report, we will show only a few important results.

#### I. Univariate analysis of our dependent variable *price*:

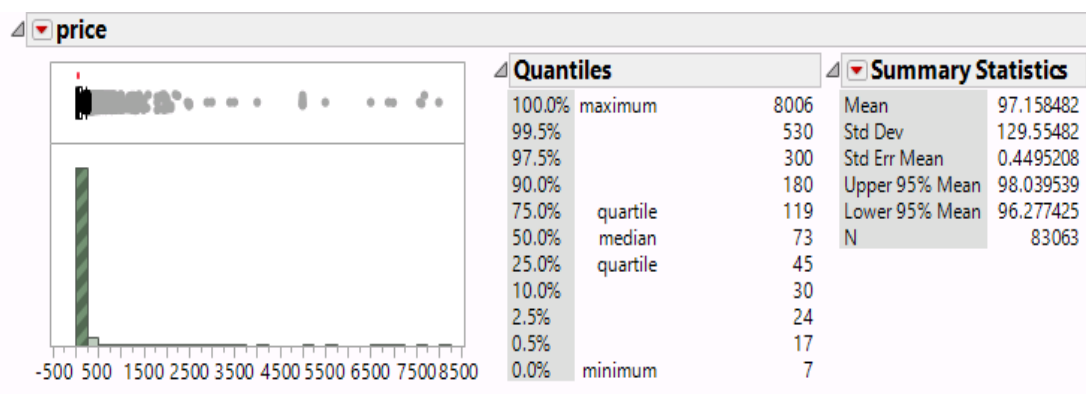
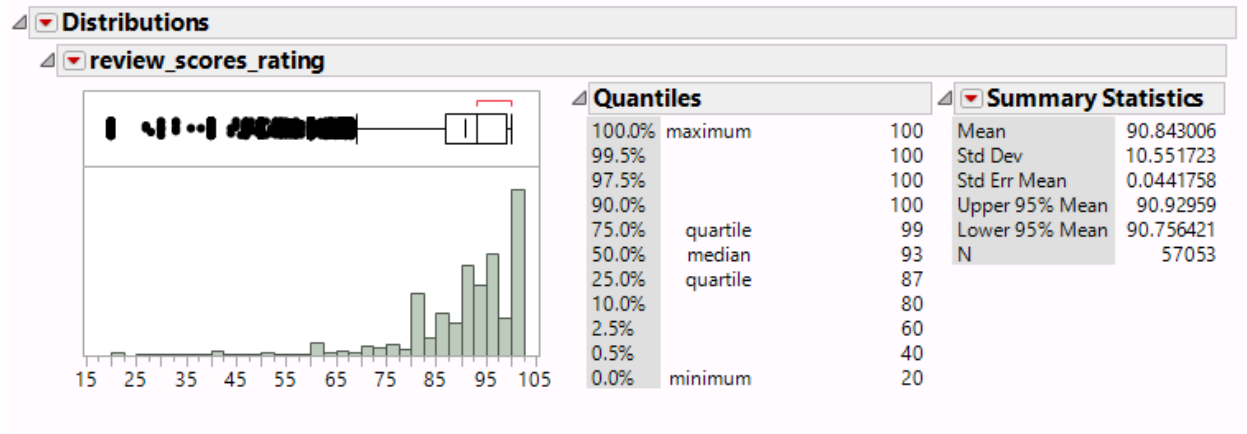


Figure 3. JMP - univariate analysis of price.

The mean price of the AirBnB booking per night showed to be around 97€ per night, however there is an enormous discrepancy between the lowest and the highest prices which is indicative that there are probably several important factors that influence the pricing strategy.

II. Univariate analysis of the variable *review scores*:

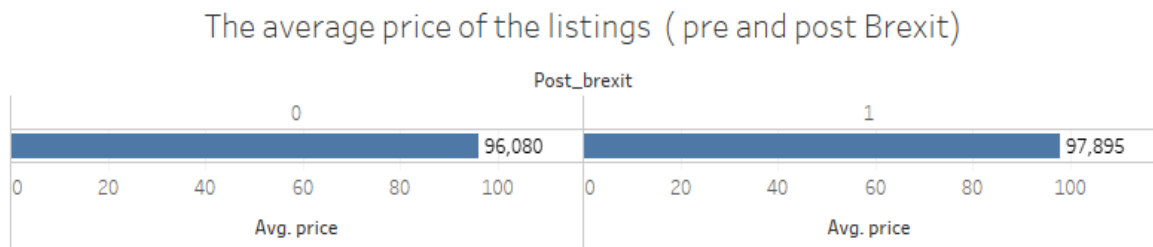


**Figure 4.** JMP - univariate analysis of review scores rating.

While there is still a great discrepancy between the maximum and minimum value, a mean review score of around 91 is indicative the customers are generally pleased with AirBnB stays in London.

III. We also conducted a visual comparisons to get an idea of what our data tells us, before we continue towards testing the significance of those results:

a) Difference in average price pre and post Brexit announcement

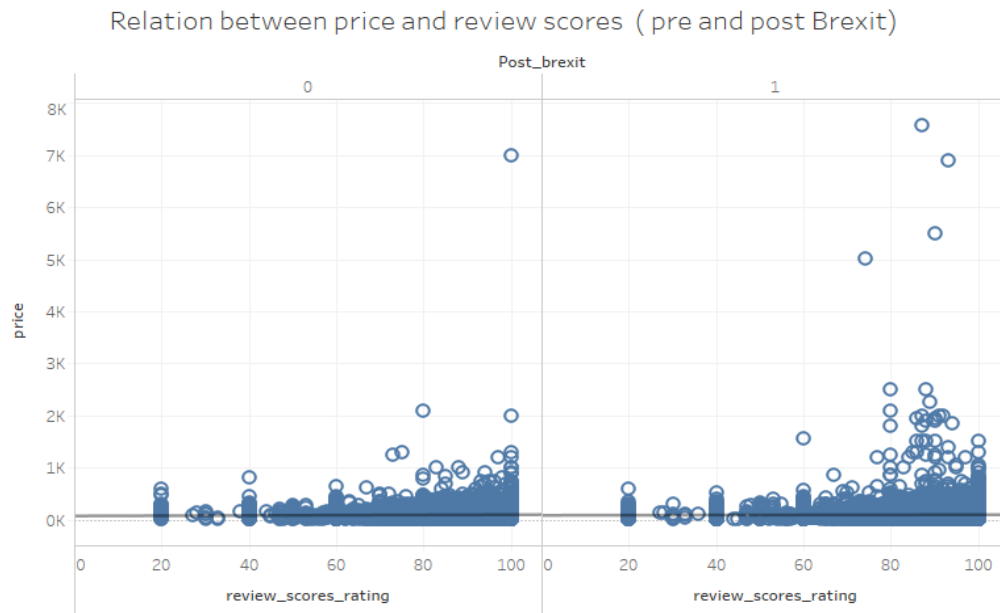


**Figure 5.** JMP - visuals of univariate analysis of price.

The visual shows that the mean price per a nights stay is higher by approximately 2£ in the post Brexit announcement month . However, there are several issues with this. First, we suspect that Brexit announcement is not the only variable impacting the prices: there might be other variables, interaction terms etc. Second, we are unsure if this difference is statistically significant. Therefore, we proceed our report by conducting further analysis.

IV. Bivariate analysis- scatterplots

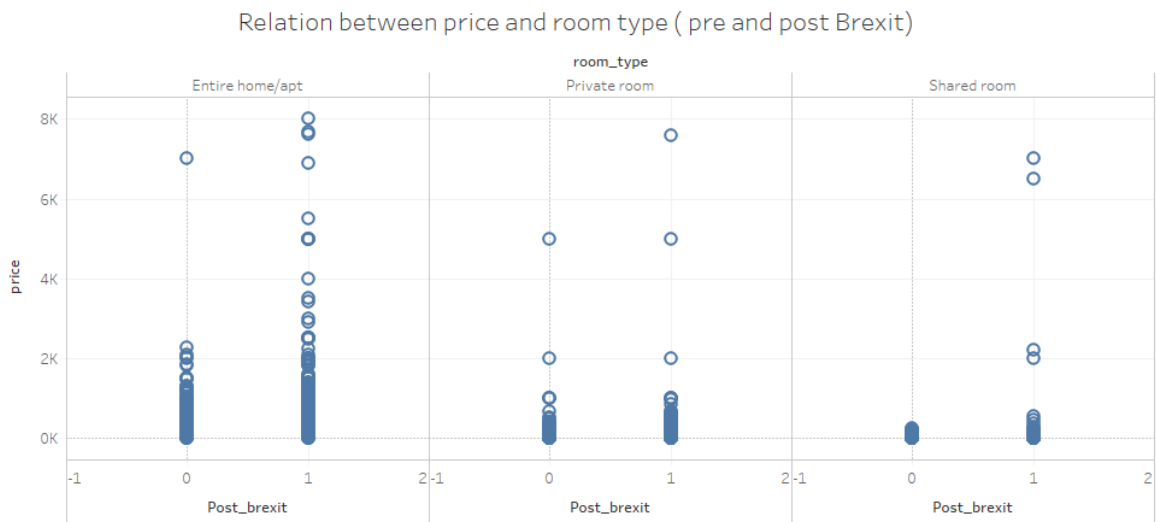
a)



**Figure 6.** JMP - bivariate analysis of review scores rating and price.

There does not seem to be much difference in the correlation between price and review scores pre- and post- Brexit announcement.

b)



**Figure 7.** JMP - bivariate analysis of room type and price.

The biggest difference seems to be the prices for an entire apartment.

### C: Linear Regression.

To simplify the regression model, we grouped the existing boroughs into five fare zones and modeled the prices using the following model (fig.3).

Summary of Fit				
RSquare		0.29207		
RSquare Adj		0.291857		
Root Mean Square Error		85.46992		
Mean of Response		91.90316		
Observations (or Sum Wgts)		56637		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	17	170641534	10037737	1374.071
Error	56619	413607848	7305.1069	Prob > F
C. Total	56636	584249382		<.0001*
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-42.73224	5.784327	-7.39	<.0001*
Post_brexit[0]	-0.448528	0.374486	-1.20	0.2310
accommodates	7.3006733	0.325931	22.40	<.0001*
bathrooms	23.844858	0.761561	31.31	<.0001*
bedrooms	24.519559	0.728142	33.67	<.0001*
business[0]	-7.026762	0.965863	-7.28	<.0001*
romantic[0]	-4.21867	1.445826	-2.92	0.0035*
family[0]	-0.675681	1.108428	-0.61	0.5421
social[0]	0.4440089	1.063162	0.42	0.6762
Entire home/apt[0]	-23.67549	0.451171	-52.48	<.0001*
Shared room[0]	19.511102	1.644539	11.86	<.0001*
f[0]	2.3148108	0.649434	3.56	0.0004*
review_scores_rating	0.183044	0.035103	5.21	<.0001*
zone_1[0]	-5.972561	0.880526	-6.78	<.0001*
zone_2[0]	3.399985	0.91563	3.71	0.0002*
zone_3[0]	5.4887469	0.935108	5.87	<.0001*
zone_4[0]	9.0428818	1.37862	6.56	<.0001*
zone_5[0]	8.0494146	2.650318	3.04	0.0024*

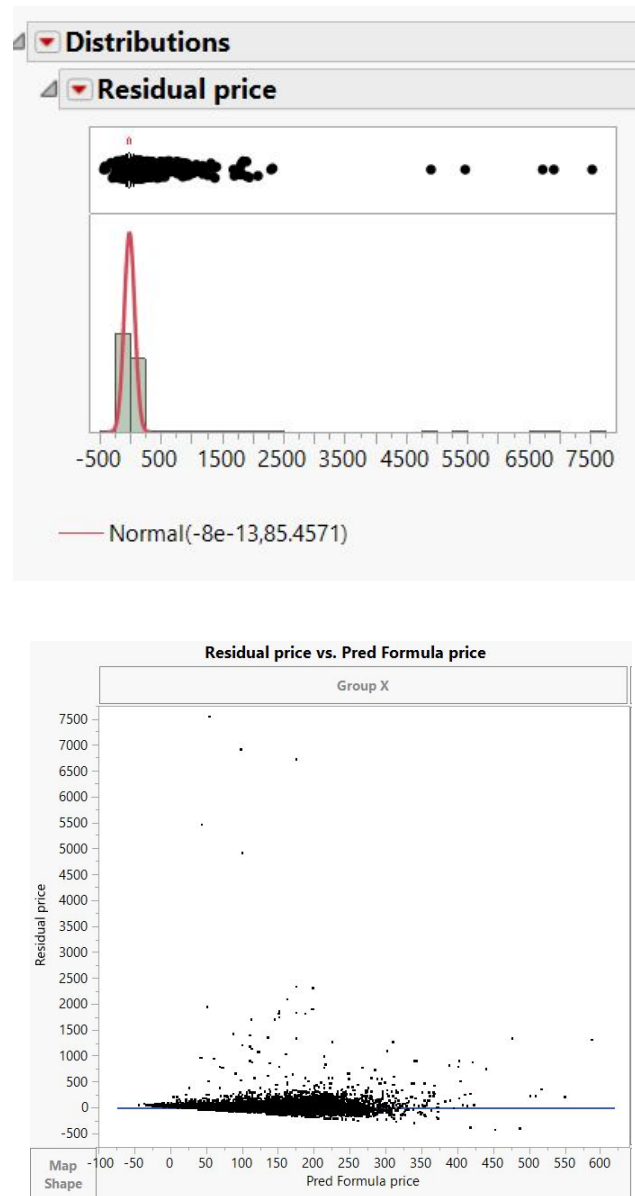
**Figure 3.** Linear regression of price on chosen variables in JMP.

Linear regression equation:

$$\text{Price} = -42.7 - 0.449 \cdot \text{post\_brexit} + 7.3 \cdot \text{Accommodates} + 23.8 \cdot \text{bathrooms} + 24.5 \cdot \text{bedrooms} - 7.02 \cdot \text{business} - 4.22 \cdot \text{romantic} - 0.676 \cdot \text{family} + 0.44 \cdot \text{social} - 23.7 \cdot \text{entire home/apt} + 19.5 \cdot \text{shared room} + 2.31 \cdot \text{f} + 0.18 \cdot \text{review\_Score} - 5.97 \cdot \text{Zone\_1} + 3.4 \cdot \text{Zone\_2} + 5.49 \cdot \text{Zone\_3} + 9.04 \cdot \text{Zone\_4} + 8.05 \cdot \text{Zone\_5}$$

This linear regression shows that Brexit has had no significant effect on prices. The effects of other variables are as expected. we can see that a listing being in Zone 1 has a significantly positive effect on price, which makes sense because it is the most expensive zone. It also makes sense that listings that accomodate more people or have higher review

scores have higher prices. However, we can see that this model has a very low RSquare value (0.29). This means that our model only explains 29% of the variation in our data. Therefore, we decided to run the following model diagnostics:



**Figure 4.** *Residual Diagnostics for linear regression.*

We can see that both the assumptions of residual normality and linearity are violated. This means that we should use a log transformation in order to have a better performing model.

### *Semi-Log Regression Model*

Moving on we decided to run a semi-log regression model. We applied a log transformation on our dependent variable (Price) and we got the following results:



Summary of Fit				
RSquare		0.684003		
RSquare Adj		0.683908		
Root Mean Square Error		0.370811		
Mean of Response		4.282546		
Observations (or Sum Wgts)		56637		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	17	16851.670	991.275	7209.232
Error	56619	7785.154	0.138	Prob > F
C. Total	56636	24636.823		<.0001*
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	2.9275275	0.025095	116.66	<.0001*
Post_brexit[0]	0.0095537	0.001625	5.88	<.0001*
accommodates	0.063389	0.001414	44.83	<.0001*
bathrooms	0.0680007	0.003304	20.58	<.0001*
bedrooms	0.155175	0.003159	49.12	<.0001*
business[0]	-0.059757	0.00419	-14.26	<.0001*
family[0]	-0.018134	0.004809	-3.77	0.0002*
romantic[0]	-0.040328	0.006273	-6.43	<.0001*
social[0]	-0.006603	0.004613	-1.43	0.1523
f[0]	0.0373646	0.002818	13.26	<.0001*
Entire home/apt[0]	-0.347873	0.001957	-177.7	<.0001*
Shared room[0]	0.3199691	0.007135	44.85	<.0001*
review_scores_rating	0.0035537	0.000152	23.33	<.0001*
zone_1[0]	-0.049113	0.00382	-12.86	<.0001*
zone_2[0]	0.044818	0.003972	11.28	<.0001*
zone_3[0]	0.0805985	0.004057	19.87	<.0001*
zone_4[0]	0.1283643	0.005981	21.46	<.0001*
zone_5[0]	0.1268051	0.011498	11.03	<.0001*

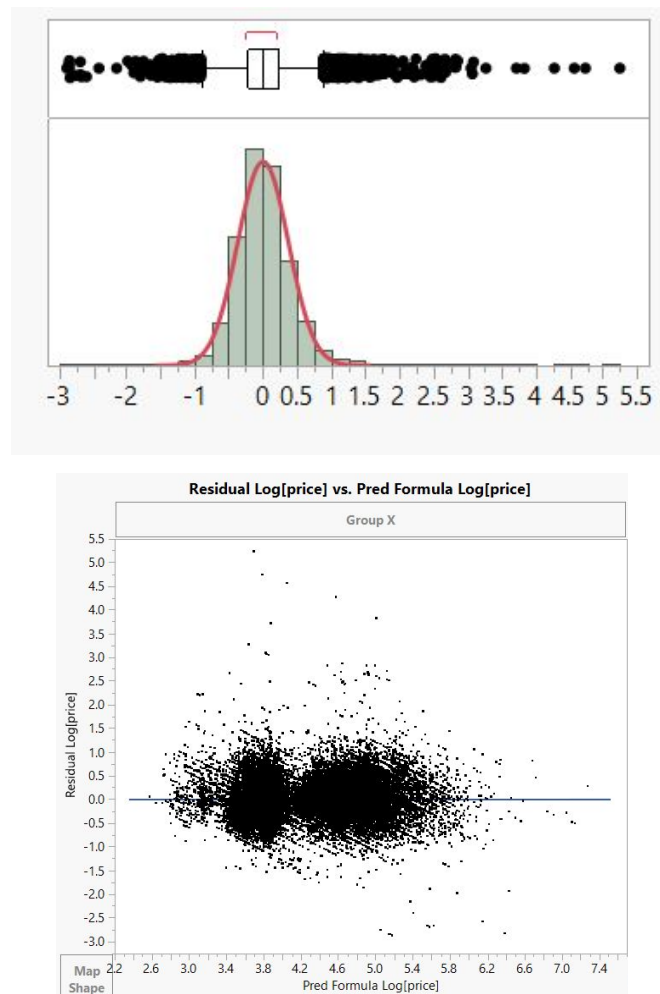
**Figure 5.** Semi-log regression of price on chosen variables in JMP.

This gives us the following regression equation:

$$\ln(\text{Price}) = 2.93 + 0.0096 \cdot \text{post\_brexit} + 0.063 \cdot \text{Accommodates} + 0.068 \cdot \text{bathrooms} + 0.155 \cdot \text{bedrooms} - 0.059 \cdot \text{business} - 0.018 \cdot \text{romantic} - 0.040 \cdot \text{family} - 0.0066 \cdot \text{social} - 0.35 \cdot \text{entire home/apt} + 0.32 \cdot \text{shared room} + 0.037 \cdot \text{f} + 0.0036 \cdot \text{review\_Score} - 0.049 \cdot \text{Zone}_1 + 0.045 \cdot \text{Zone}_2 + 0.081 \cdot \text{Zone}_3 + 0.128 \cdot \text{Zone}_4 + 0.127 \cdot \text{Zone}_5$$

Interestingly our New model shows that Brexit did indeed have a significantly negative effect on AirBnB prices in London. This model also has a significantly higher Rsquare value (0.68) compared to our previous model. This means that the semi\_log regression model explains 68% of the variation in our data.

We then decided to examine the residual diagnostics for this model as well:



**Figure 6.** *Semi-log regression diagnostics.*

We can see here that the assumption of residual normality is not violated. The semi log model seem to be more homoscedastic than the linear Model.

#### *D: Principal Component Analysis.*

In an aim to reduce the great volume of our dataset, or more specifically, to reduce the number of columns we conducted a *Principal Component Analysis*. By considering the magnitude of the principal components through looking at their eigenvalues we decided to retain 9 factors as shown in the following table:

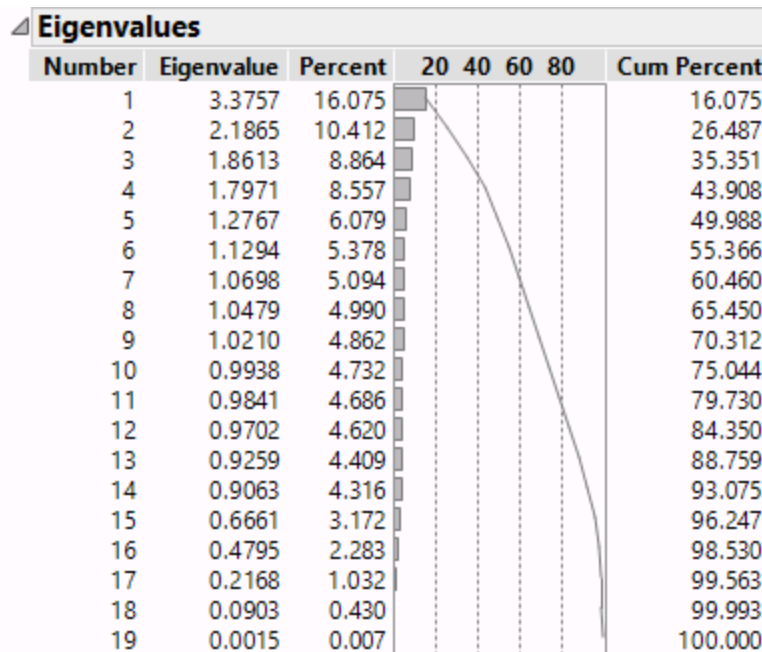


Figure 7. JMP - Eigenvalues

By using 9 Factors, where the eigenvalue is greater than 1, we capture about 70% of total information.

Furthermore, we conducted a *Factor Analysis* and by identifying high values in the correlation matrix between our factors and independent variables in the set we assigned names to our factors accordingly.

**Rotated Factor Loading**

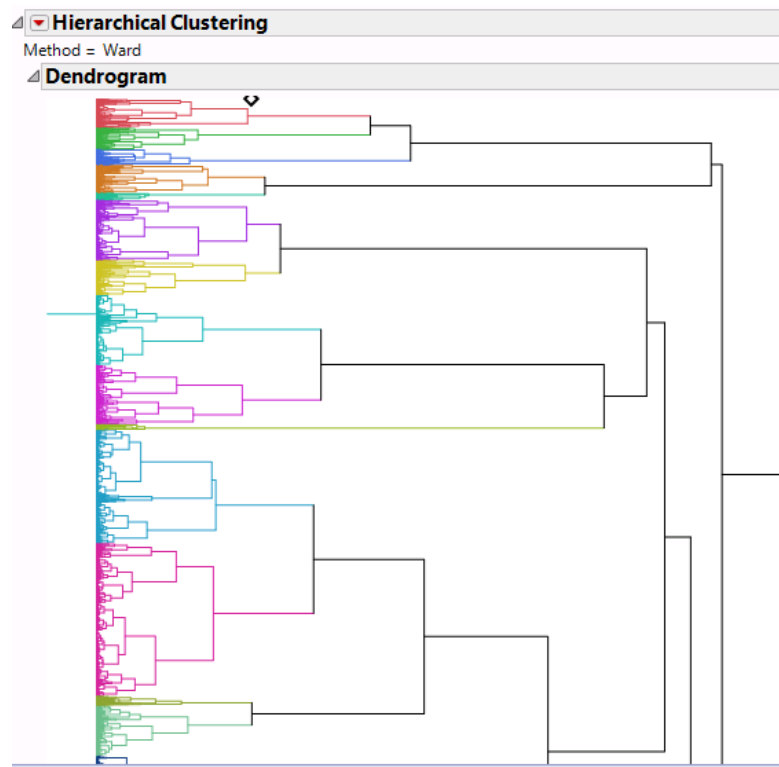
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
accommodates	0.770160	0.415676	-0.032130	-0.008331	-0.022498	0.002166	-0.068644	-0.006898	-0.033938
bathrooms	0.787269	-0.153624	0.001994	0.041950	0.008075	0.000249	-0.013470	0.003822	0.134856
bedrooms	0.833795	0.214534	-0.017264	0.042335	-0.032924	0.055288	0.024877	0.018691	-0.084787
business	-0.024935	-0.010535	-0.050918	0.002106	<b>0.962416</b>	0.003976	0.084473	-0.018852	0.032604
Post_brexit	0.028565	-0.005438	0.078780	-0.051415	-0.050598	0.099967	0.026631	<b>0.661768</b>	-0.014038
Entire home/apt	0.290288	<b>0.912260</b>	0.008540	-0.053201	-0.001075	-0.061501	-0.004999	-0.016140	-0.100436
f	-0.026091	0.030601	0.072454	0.014071	-0.093802	0.029333	<b>-0.718495</b>	0.018428	-0.034928
family	0.235629	0.025751	<b>-0.475601</b>	0.016624	0.136059	0.099141	-0.277039	-0.004702	<b>-0.342422</b>
none	-0.036289	-0.048462	<b>0.840174</b>	0.002516	<b>-0.528826</b>	-0.009803	-0.007716	0.036105	0.068257
price	<b>0.588407</b>	0.232527	0.034107	-0.079955	0.019086	-0.095901	0.064474	-0.010035	-0.055813
Private room	-0.294008	<b>-0.916192</b>	0.000271	0.050430	-0.008167	0.048479	0.030586	0.015440	-0.116286
review_scores_rating	0.004467	0.018525	0.001909	0.032966	-0.019126	0.074316	<b>0.739270</b>	0.029561	-0.117317
romantic	-0.202753	0.324160	<b>-0.417140</b>	0.037976	-0.015348	-0.032101	0.044824	-0.011327	-0.019215
Shared room	0.013973	0.011217	-0.037111	0.011935	0.038950	0.054939	-0.109292	0.002990	<b>0.911332</b>
social	0.012237	-0.160359	<b>-0.738994</b>	-0.051087	-0.194988	-0.059903	0.160194	-0.033692	0.186898
zone_1	0.011919	0.053234	-0.005951	<b>-0.838911</b>	0.013694	<b>-0.471530</b>	-0.029603	-0.129189	-0.015269
zone_2	0.007138	-0.042774	0.011192	<b>0.945595</b>	0.012514	-0.192936	0.000676	-0.090442	-0.000870
zone_3	-0.035038	-0.001358	-0.027396	0.149679	-0.086280	<b>0.773285</b>	0.051136	0.070961	0.045465
zone_4	0.001803	-0.066631	0.039254	-0.082680	0.073039	<b>0.595382</b>	-0.009931	-0.071356	-0.007029
zone_5	-0.027181	-0.018944	-0.041257	0.057796	0.035461	-0.124134	-0.019867	<b>0.779546</b>	0.018906

Figure 8. JMP - Factor Analysis

Therefore, Factor1= Accommodates, Factor2= Size, Factor3= None, Factor4= Zone 2, Factor5= Business, Factor6= Mid Market, Factor7= Reviews, Factor8= Brexit and Factor9= Shared Room.

*E: Clustering.*

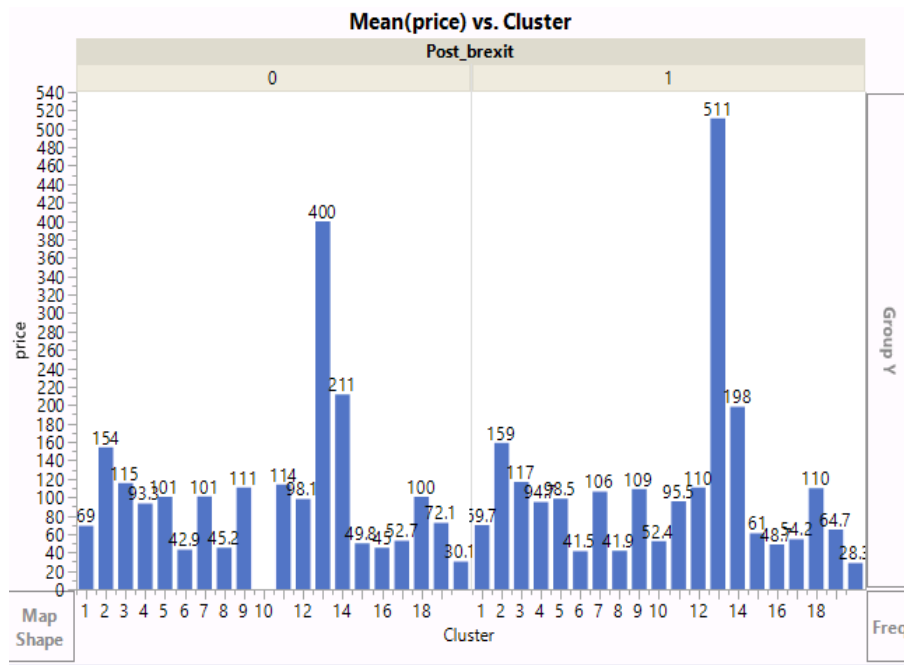
Next, we increased the number of rows grouping a set of objects in such a way that objects in the same group are more similar to each other - *Hierarchical Cluster Analysis*.



**Figure 9.** *JMP - Clusters.*

The analysis resulted in 20 clusters.

Then, to relate the analysis to our project hypothesis about the effect the brexit announcement had on AirBnB prices in London, we were interested to see if our dependent variable-price varies by cluster when comparing pre and post Brexit announcement values



**Figure 10.** JMP - A bar chart of price variations by clusters.

In conclusion, we were able to identify small variations sometimes by increases in price such as cluster 12 and sometimes by decreases such as cluster 19 and 20. However, there is not a clear linear trend.

### *Linear Regression with PCA:*

Now that we have reduced the number of factors in our data using PCA, we decided to rerun the linear regression model using the rotated components that we saved to our data set.

Summary of Fit				
RSquare		0.438223		
RSquare Adj		0.438104		
Root Mean Square Error		76.65417		
Mean of Response		92.29261		
Observations (or Sum Wgts)		42456		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	194554081	21617120	3678.970
Error	42446	249406853	5875.8623	Prob > F
C. Total	42455	443960934		<.0001*
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	90.873794	0.374653	242.55	<.0001*
Accommodates	60.816638	0.387072	157.12	<.0001*
Size	30.874279	0.364083	84.80	<.0001*
None	1.3908661	0.32837	4.24	<.0001*
Zone 2	-9.824617	0.377618	-26.02	<.0001*
Business	2.2406554	0.323715	6.92	<.0001*
Mid Market	-10.88462	0.382804	-28.43	<.0001*
Reviews	5.6763183	0.343544	16.52	<.0001*
Brexit	-1.28538	0.374809	-3.43	0.0006*
Shared Room	-8.075077	0.383865	-21.04	<.0001*

**Figure 11.** Linear regression of price on rotated components.

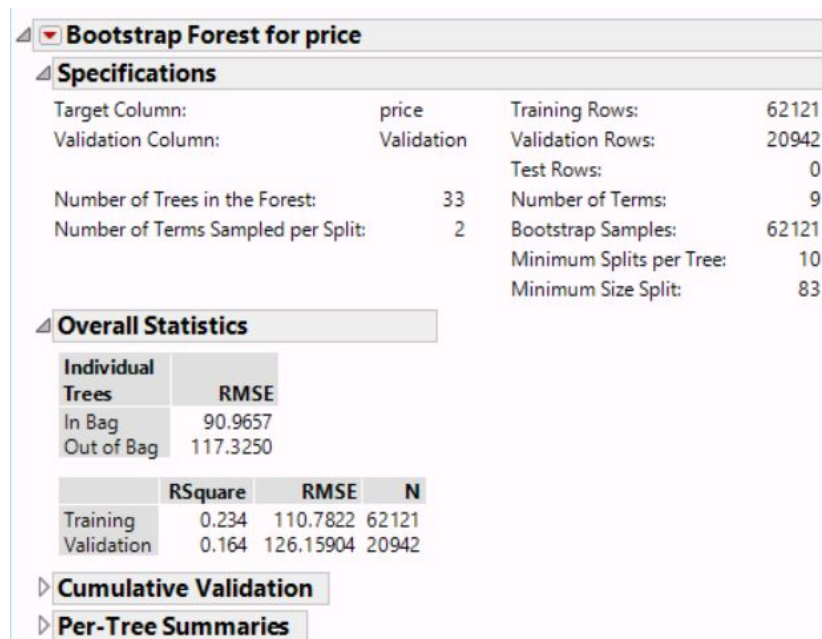
Again, we can see that Brexit has a statistically significant negative effect on prices. The adjusted RSquare value in this case is 0.43. What we did next was that we saved the predicted values for prices to our data set so that we can include this linear regression in our model comparison, once we start running machine learning models.

#### *G: Machine Learning.*

Moving on, we are now going to run four different machine train four different machine learning models in order to see which model can better predict AirBnB prices in London. Given that our dependant variable (Price) is a continuous variable, the metric that we will be looking at in our case will be RSquare rather than misclassification rate. This is important because we will be able to choose the model that best fits our data and, therefore, gives a more accurate prediction.

Before running any of the models, we created a validation column, which split our data into a training set and a validation set. Our training set includes 75% percent of the data, While the validation set only includes 25%

#### *Bootstrap Forest:*



The screenshot shows the configuration and performance metrics for a Bootstrap Forest model. The 'Specifications' section lists the target column as 'price', validation column as 'Validation', 33 trees, 2 terms sampled per split, and 62121 bootstrap samples. The 'Overall Statistics' section provides RMSE values for in-bag and out-of-bag predictions, and a table of RSquare, RMSE, and N for training and validation sets. The validation RSquare is notably low at 0.164.

Bootstrap Forest for price			
Specifications			
Target Column:	price	Training Rows:	62121
Validation Column:	Validation	Validation Rows:	20942
		Test Rows:	0
Number of Trees in the Forest:	33	Number of Terms:	9
Number of Terms Sampled per Split:	2	Bootstrap Samples:	62121
		Minimum Splits per Tree:	10
		Minimum Size Split:	83
Overall Statistics			
Individual Trees	RMSE		
In Bag	90.9657		
Out of Bag	117.3250		
	RSquare	RMSE	N
Training	0.234	110.7822	62121
Validation	0.164	126.15904	20942
Cumulative Validation			
Per-Tree Summaries			

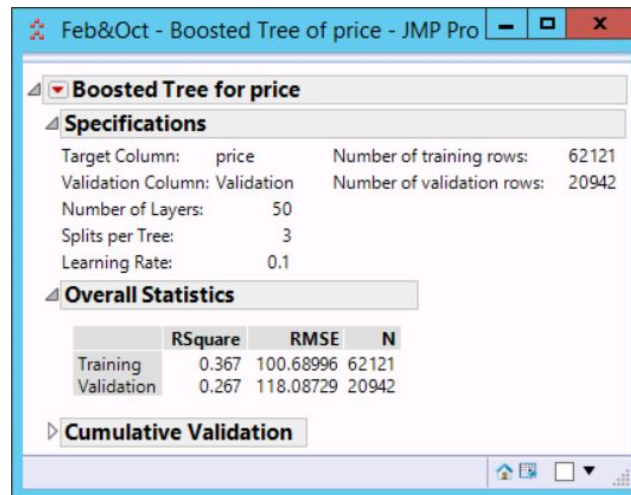
**Figure 12.** *Bootstrap Forest Model.*

In order to this model, we are interested in seeing the RSquare on the validation set. We can see here that this model has a very low RSquare value of 16.4%, which is much lower than



our linear regression model.

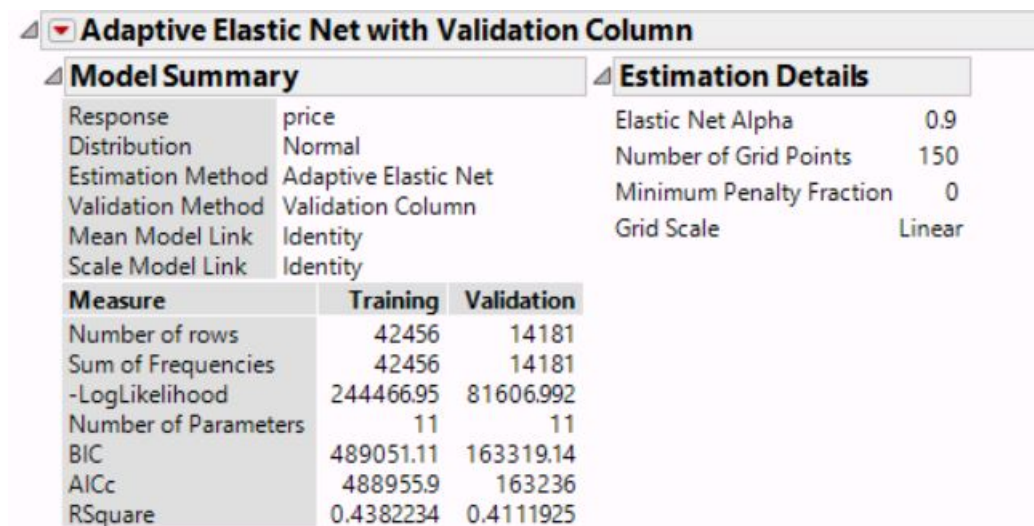
*Boosted Tree Model:*



**Figure 13.** *Boosted Tree Model.*

We can see that our RSquared value on the validation set is 0.267.

*Elastic Net:*



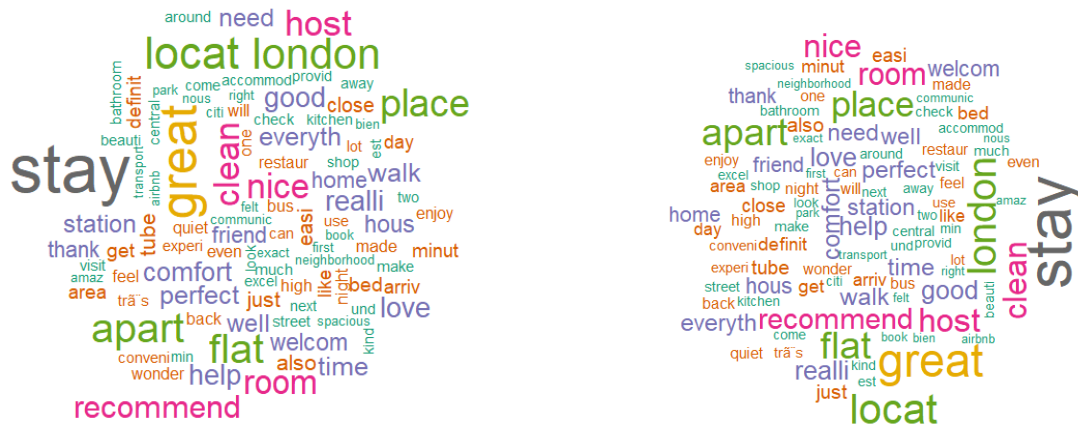
**Figure 14.** *Elastic Net Model.*

Conversely, our RSquared is 0.411.



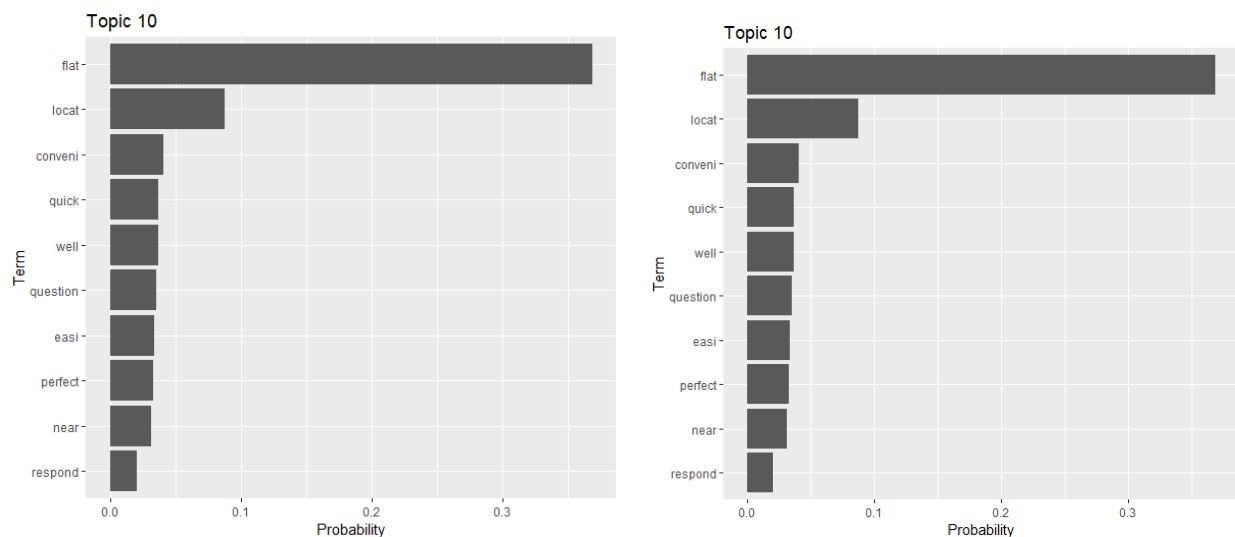


from period 1 to 2. We made a word cloud of the most frequently used terms in 3500 randomly chosen reviews and obtained the following result (fig.17).

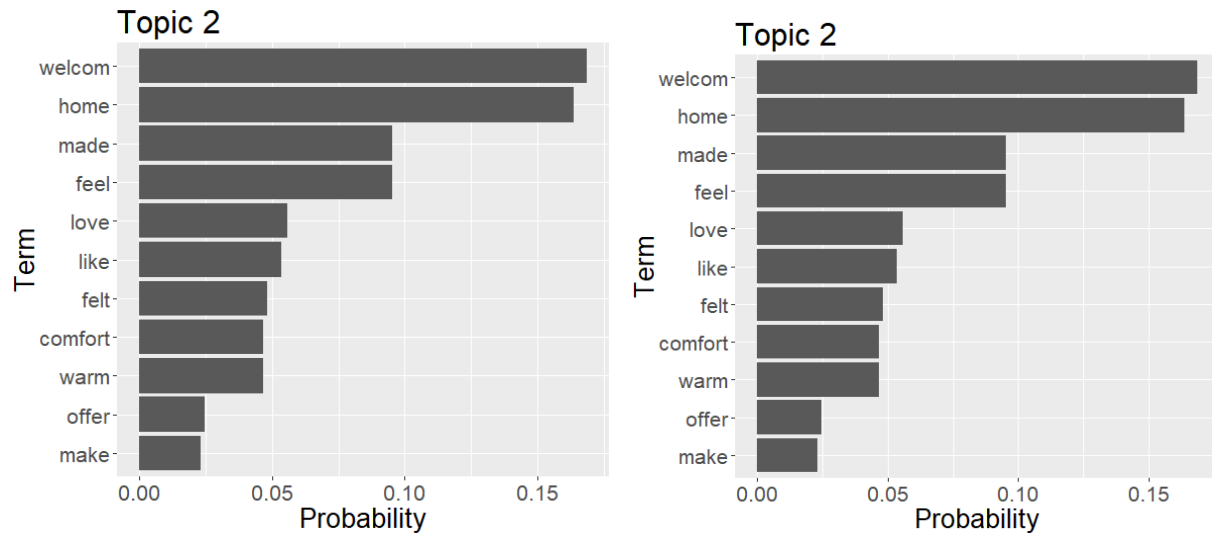


**Figure 17.** Review word clouds from February, 2016 and October, 2016, respectively.

Although this is not very informative, a conclusion can be largely made that the overall perception of Airbnb listings in London has not changed significantly after the announcement. Moving on, we modeled sentiments by topic and saw the results reflected in Figure 18.

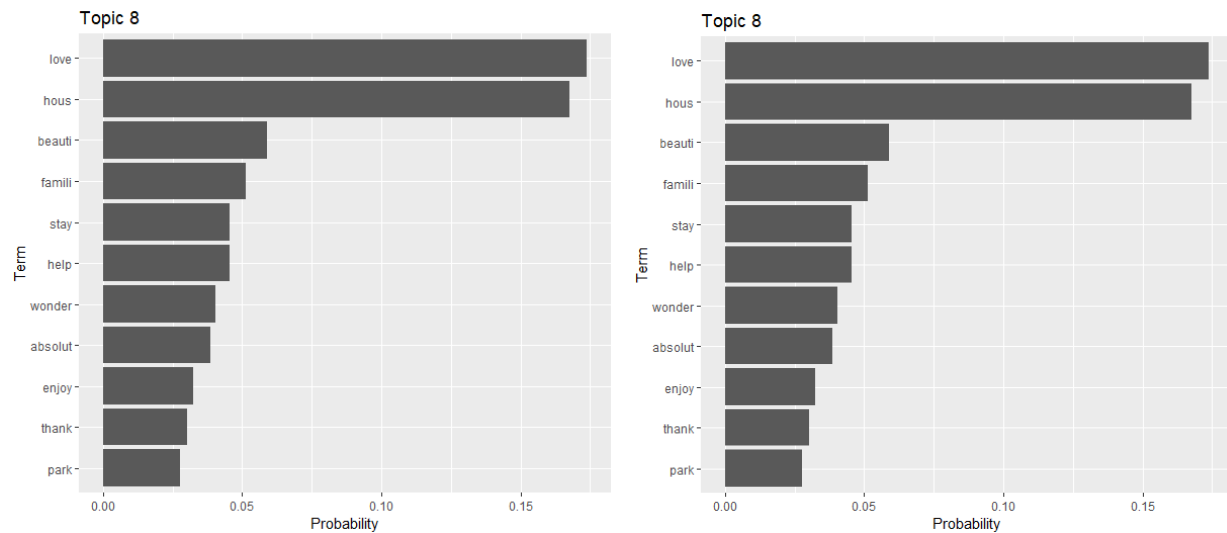


**Figure 18.** Topic modeling (topic: convenience) of reviews in February and October, 2016. Respectively.



**Figure 19.** *Topic modeling (topic: host) of reviews in February and October, 2016, respectively.*

As we can observe, there has not been any noticeable change in the sentiment of reviews after the announcement. Further testing only proved the conclusion, as shown in Figure N+3.



**Figure 20..** *Topic modeling (topic: impressions) of reviews in February and October, 2016, respectively.*

Therefore, as in other parts of our analysis, we concluded that there has not been any meaningful change in the sentiment of the reviews since the announcement, which can be explained by the relative novelty of the event and its larger irrelevance towards the quality of rented apartments in the city.

### Part Three: Limitations and Conclusion.

We contemplated extending the timeframe of our research model, but concluded that in the current environment of political uncertainty in the country, this would lie beyond the scope

of our project. We do, conversely, acknowledge the limitations.

### *Limitations.*

The first and foremost limitation is the time horizon chosen for the project. We acknowledge that the discovered changes in prices, as indicated by our linear regression model, might have been seasonal. Therefore, a more thorough study would include data from multiple seasons to account for that.

Secondly, we cannot use London data as a generalizable sample to make a prediction for other parts of the UK, as this is not a representative sample.

Thirdly, a more thorough text mining analysis can be done in order to evaluate the change in specific sentiments across multiple timeframes.

### *Conclusion.*

Having conducted thorough analysis using the acquired data, we can firmly conclude that as of October, 2016, the announcement of Brexit has had a negative impact on the price, yet has not significantly impacted the quality of listings or the sentiment carried by the majority of the reviews. Therefore, the findings of our project are neutral for the time being, but we acknowledge that that can change as the terms of the Brexit deal solidify and begin to take place in reality. Until then, we advise everyone to visit the wonderful city of London and take advantage of its numerous lodging opportunities.