

Pricing Strategy Airbnb based on Negative Reviews

Rashel Mol, Noa Beekmans, Astrid Raaijmakers, Pien Korteweg, Amber Vermeer

Introduction

Firstly, we cleaned the raw dataset. The descriptions of the variables are mentioned below. Based on this cleaned dataset, we executed a text analysis. The results of this analysis are shown in a sentiment plot. Furthermore, we did a topic analysis to check what topics are spoken of the most. These results are shown in the sentiment topic plot below. To conclude, we performed a regression analysis. The results are shown in model summary. Also, we briefly describe the results in the section below. To conclude, we visually checked the correlation between price and sentiment (compound).

Note: our results are based on a prototype sample. The results can differ when conducting it on the whole dataset.

Variable Descriptions

The cleaned dataset “gen/temp/airbnb.csv”, consists of the following variables.

ID

ID is a numeric variable. Every listing has an unique ID.

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##      2818   7276869 17519833 17400701 24732648 51316529
```

Name

Name is a character variable. It is the name of the listing. Below is the summary and an example of a name:

```
##      Length      Class      Mode
##      226895 character character

## [1] "Quiet Garden View Room & Super Fast WiFi"
```

Neighbourhood

Neighbourhood is a factor variable. The neighbourhood in which the listing is located. There are 22 classified neighbourhoods.

##		
##	Bijlmer-Centrum	Bijlmer-Oost
##	2110	1113
##	Bos en Lommer	Buitenveldert - Zuidas
##	8963	2277
##	Centrum-Oost	Centrum-West
##	29775	43206
##	De Aker - Nieuw Sloten	De Baarsjes - Oud-West
##	2235	32383
##	De Pijp - Rivierenbuurt	Gaasperdam - Driemond
##	22998	2049
##	Geuzenveld - Slotermeer	IJburg - Zeeburgereiland
##	2691	4123
##	Noord-Oost	Noord-West
##	3418	4645
##	Oostelijk Havengebied - Indische Buurt	Osdorp
##	8382	1747
##	Oud-Noord	Oud-Oost
##	8992	11238
##	Slotervaart	Watergraafsmeer
##	4499	4219
##	Westerpark	Zuid
##	13033	12799

Room Type

Room type is a factor variable. There are 4 possible room types.

##				
##	Entire home/apt	Hotel room	Private room	Shared room
##	100546	4025	120366	1958

Accommodates

Accommodates is a numeric variable. Accommodates is the number of guests that can stay in the listing.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.000	2.000	2.000	2.692	4.000	16.000

Comments

Comments is a character variable. Comments are the reviews about the listing. Below are some examples of comments.

```
## [1] "A model host. A model apartment. Airbnb private rooms just don't get any better than this!"
## [2] "Come and enjoy"
## [3] "<U+8377><U+5170><U+7684><U+623F><U+4EF7><U+548C><U+7269><U+4EF7><U+666E><U+904D><U+6709><U+70B9>"
## [4] "To me Daniel is the definition of Superhost. His place is very well organized and spotlessly clean."
## [5] "Daniel was a great host! He gave us great tips for transportation and local shops and restaurants."
## [6] "I had to cancel my trip to Amsterdam so unfortunately, I didn't get to stay with Daniel. I got very disappointed."
```

Year

Year is a numeric variable. Year is the year the review is written.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2018    2018    2019    2019    2019    2021
```

Price

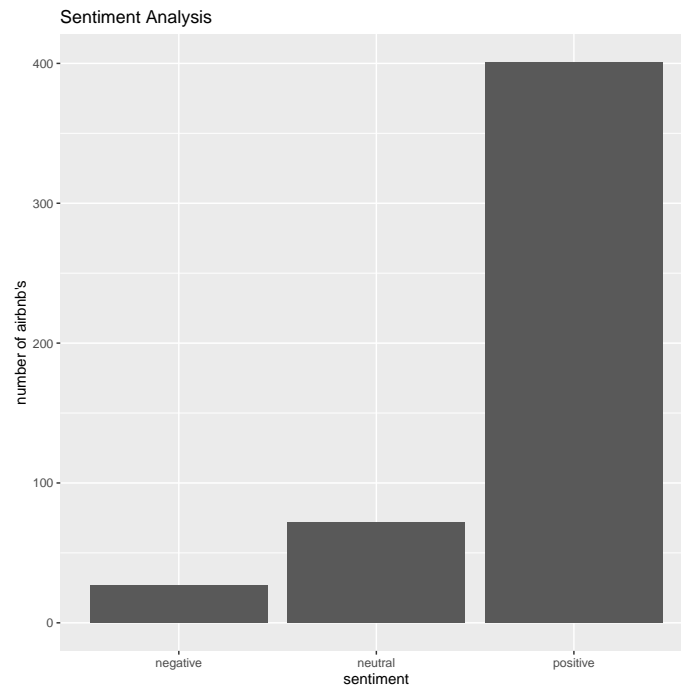
Price is a numeric variable. Price is the price in dollars per night.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       4.0    79.0   105.0   128.8   150.0   7999.0
```

Total overview of variables

```
##      id              name      neighbourhood      room_type
##  Min.   :    2818  Length:226895  Length:226895  Length:226895
## 1st Qu.: 7276869  Class :character  Class :character  Class :character
## Median :17519833  Mode  :character  Mode  :character  Mode  :character
## Mean   :17400701
## 3rd Qu.:24732648
## Max.   :51316529
##  accommodates  comments              year      price
##  Min.   : 1.000  Length:226895  Min.   :2018  Min.   : 4.0
## 1st Qu.: 2.000  Class :character 1st Qu.:2018 1st Qu.: 79.0
## Median : 2.000  Mode  :character Median :2019 Median : 105.0
## Mean   : 2.692              Mean   :2019 Mean   : 128.8
## 3rd Qu.: 4.000              3rd Qu.:2019 3rd Qu.: 150.0
## Max.   :16.000              Max.   :2021 Max.   :7999.0
```

Sentiment Analysis



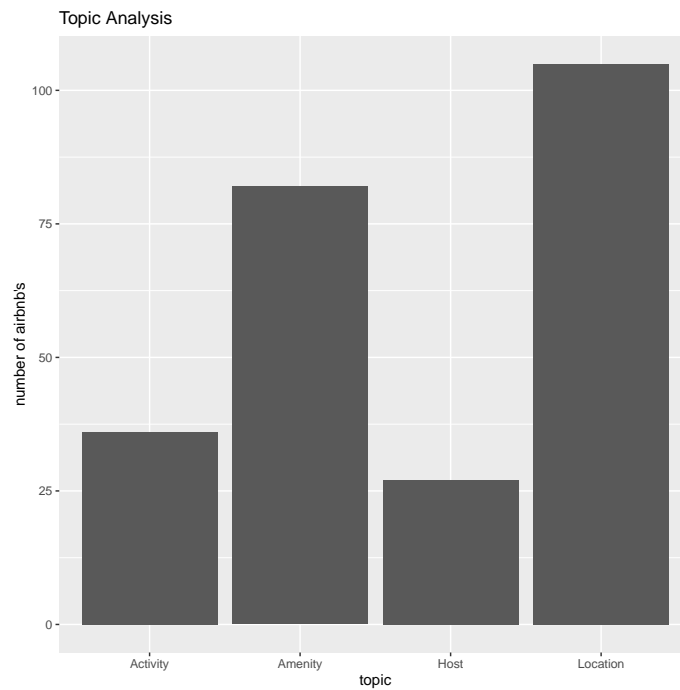
Based on the plot above, which is generated in the `text_analysis.R` script, we can see the following:

- The vast majority of reviews in the Airbnb dataset is defined positive.
- Only a very small part of the reviews in this same dataset is considered negative.

Therefore, we can conclude that the majority of reviews created by Airbnb guests has a positive nature.

Model 1	
(Intercept)	0.687 (0.040)
price	0.000 (0.000)
Num.Obs.	500
R2	0.002
R2 Adj.	0.000
AIC	649.2
BIC	661.8
Log.Lik.	−321.595
F	0.798

Topic Analysis



Noticeable is that most reviews are about Location and Amenity. A relatively small number is about Activity and Host.

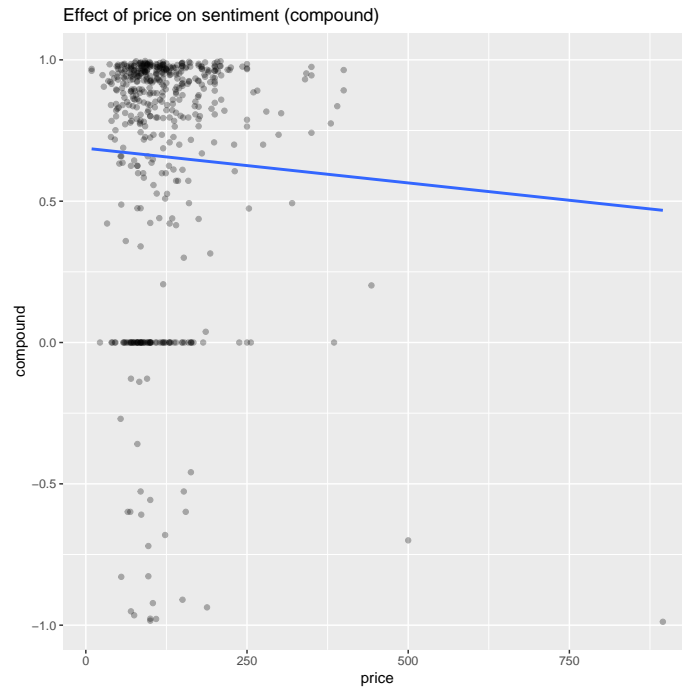
The plot above shows the topics most often mentioned in the reviews written by Airbnb guests. Remarkable is that most reviews are about Location and Amenity. A relatively small number is about Activity and Host.

Regression Analysis

Model Summary

The effect of price on compound (sentiment) is non significant, with a p-value of .372. There is very little variation explained by the model, resulting in a R square of .002.

Plot Price and Compound



As we expected based on the regression results, there is visually no correlation to be seen between price and compound (sentiment). Also, as earlier mentioned in one of the intermediate plots, most reviews are labeled as positive.

However, since only a limited sample size has been used in this analysis, one should be careful about rejecting this hypothesis. A significant relationship could still be identified when analyzing the full dataset.