

# Pricing Strategy Airbnb based on Negative Reviews

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

## Introduction

Firsly, 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.

**Name** Name is a character variable. Name of the listing.

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

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

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

**Comments** Comments is a character variable. Comments are the reviews about the listing.

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

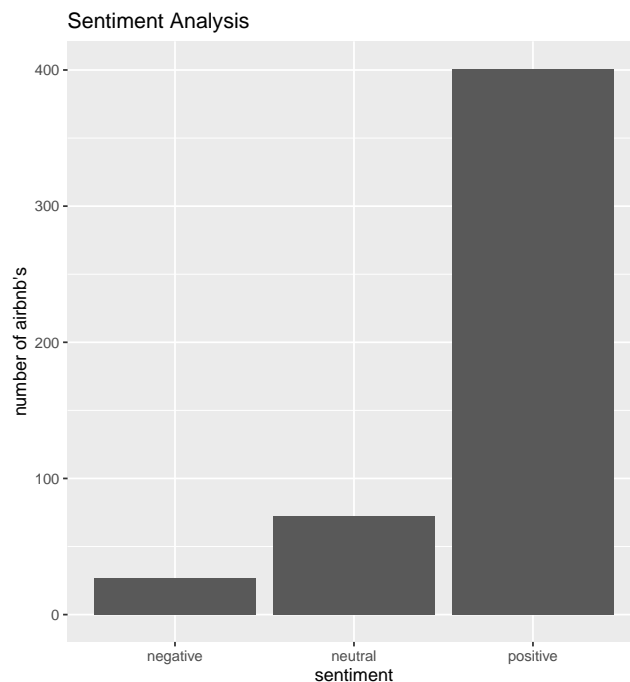
**Price** Price is a numeric variable. Price is the price in dollars per night.

```
## Rows: 226895 Columns: 8
```

```
## -- Column specification -----  
## Delimiter: ","  
## chr (4): name, neighbourhood, room_type, comments  
## dbl (4): id, accommodates, year, price  
  
##  
## 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.
```

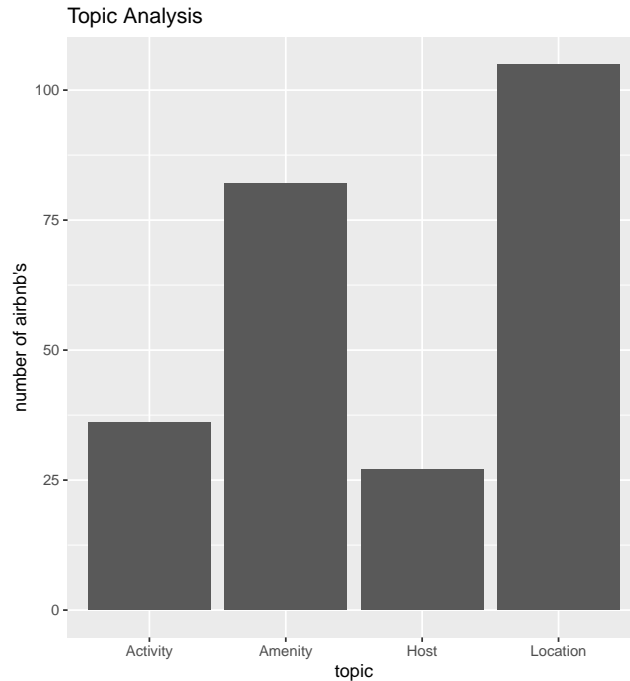
```
##           id           name      neighbourhood      room_type  
## Min.      : 2818   Length:226895   Length:226895   Length:226895  
## 1st Qu.: 7276869   Class :character   Class :character   Class :character  
## Median :17519833   Mode  :character   Mode  :character   Mode  :character  
## 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



What most standout is that most airbnb reviews are positive. A considerably small number of ratings is labeled as negative or neutral.

## Topic Analysis



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

## Model Summary

```
## Rows: 500 Columns: 10
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (5): name, neighbourhood, room_type, comments, sentiment
```

```
## dbl (5): id, accommodates, year, price, compound
```

```
##
```

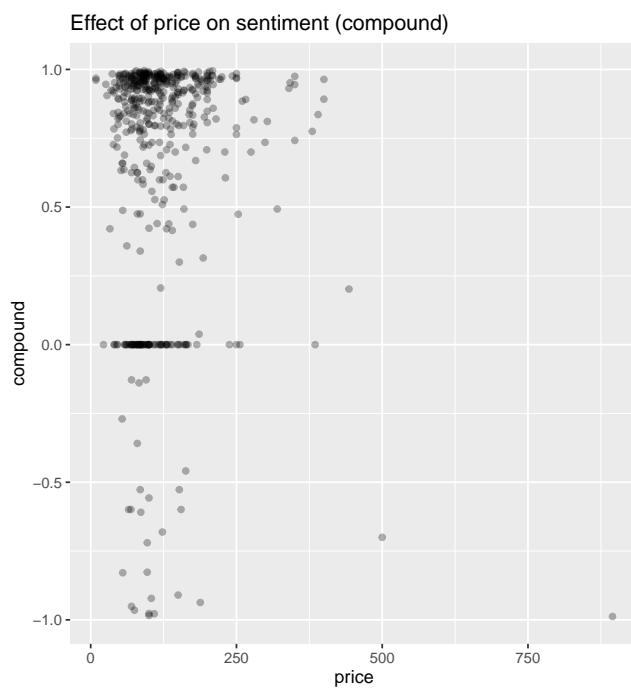
```
## i Use 'spec()' to retrieve the full column specification for this data.
```

```
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

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.

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

## 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.