Airbnbs in the Netherlands

DS-210 Final Project

Ali Schuh

Presentation Outline

- Introduce the Data
  - Background Information
  - Research Question
- Exploratory Data Analysis & Visualizations
- Models & Statistics
  - Beginning Model & Statistics
  - Transformation Time
  - Final Model & Statistics
  - Answers to Research Questions
- Conclusion
  - Future Research

Introducing the Data

- Found data set on data.world
- Created 2 years ago
- 6,334 Airbnb locations in the Netherlands, majority in Amsterdam
- Variables included basic information about the locations
  - Some variables relating to the rating of the Airbnb
- 33 columns
- 7,833 observations
- Each row/observation represents an individual stay at an Airbnb location.

Background Information

- Airbnb: a place to stay that is rented out by the owner of the house, apartment, etc.
- Can rent out anything from a single room to the whole house/apartment
- Each property is unique
- Website is used to post locations you want to rent or to rent locations

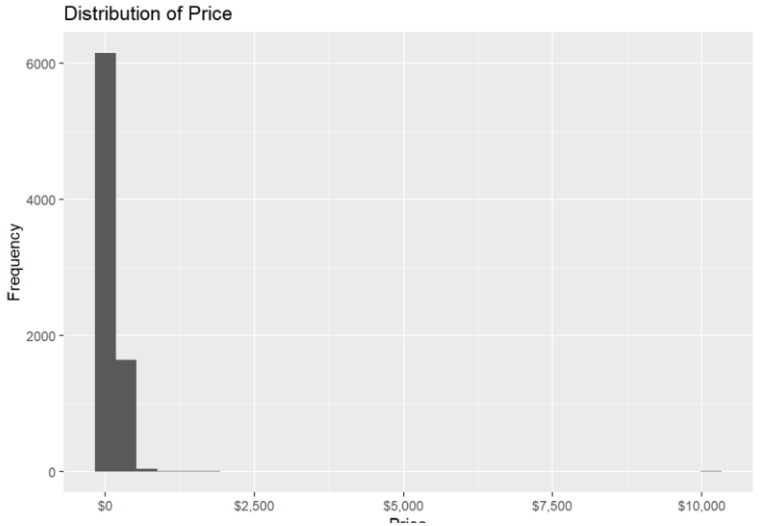
Research Question

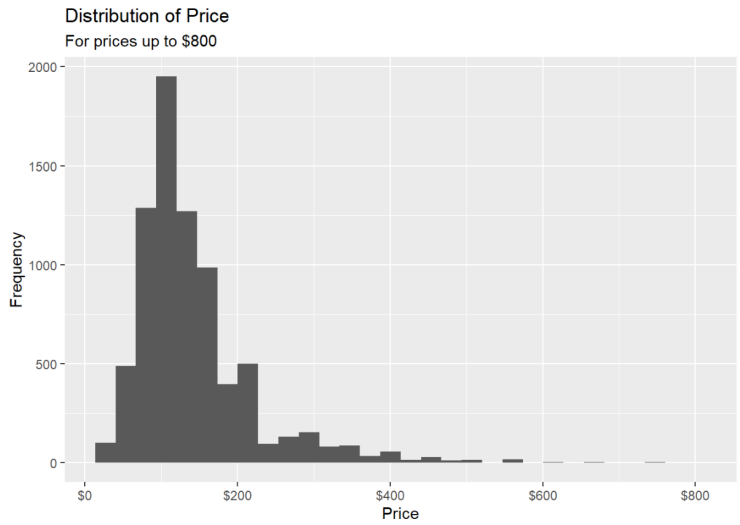
- What factors have an impact on the price of an Airbnb?
- Is there a combination of factors that have a strong effect on price?

- The factors in consideration are:
  - number of bedrooms
  - number of bathrooms
  - room type
  - minimum nights of stay
  - overall rating
  - property type
  - number of people the Airbnb accommodates.

Exploratory Data Analysis & Visualizations

- Data Wrangling
  - Converting price (euros) to dollars
  - Filtering outliers for price variable
  - Changed variable types for important variables
  - Filter out "NA" values for some variables

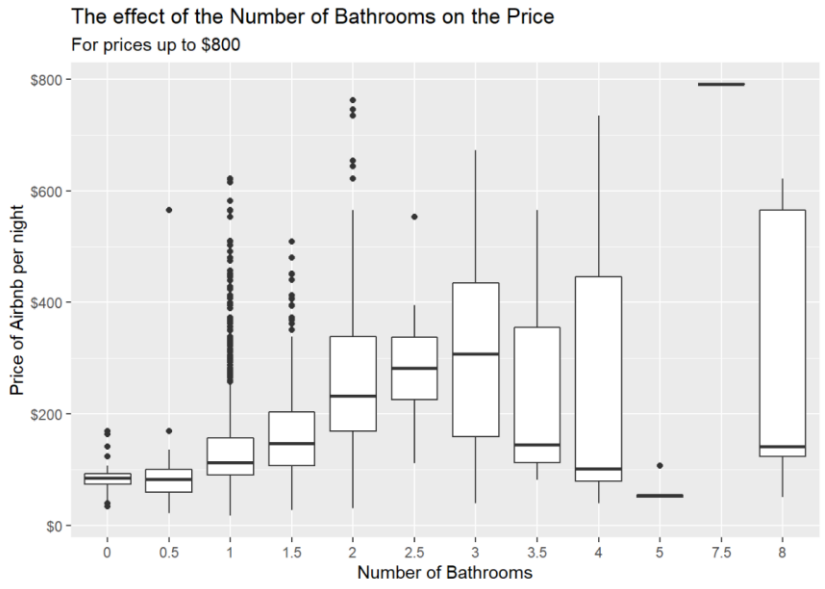


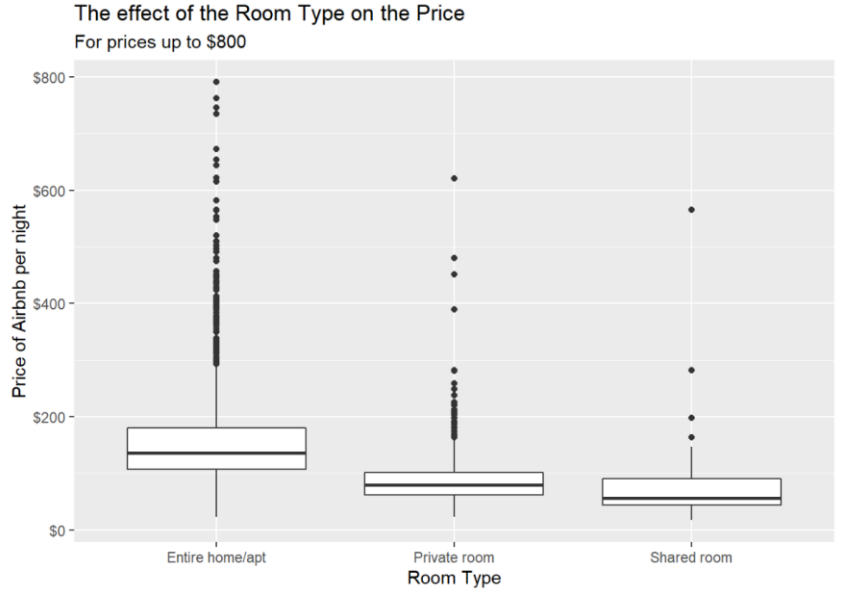


Exploratory Data Analysis & Visualizations

- EDA
  - Created boxplots for each individual variable
  - Produced some linear regression models for some variables
    - Noted the r-squared values

Variable	R-squared
Bedrooms	0.347
Bathrooms	0.168
Overall Rating	0.008

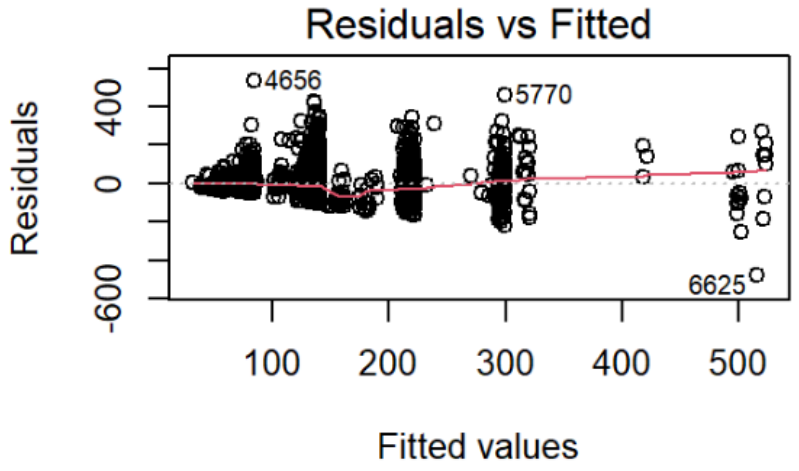


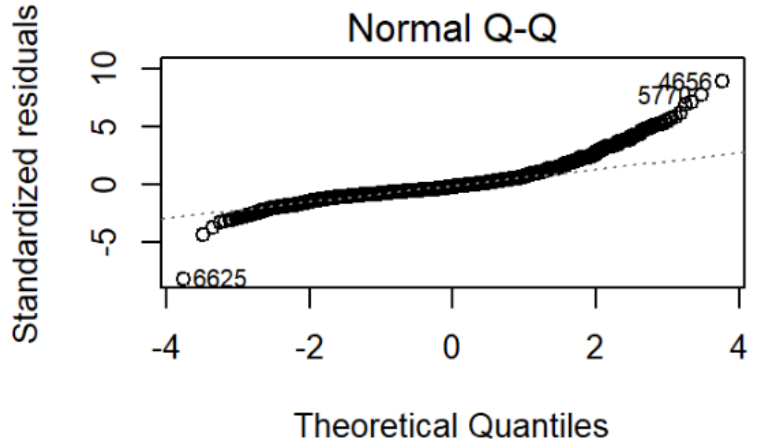


Beginning Model & Statistics

- Model #1
  - Create levels within variables
  - 5 variables included
  - Backwards Elimination
  - Adjusted R-squared = 0.379
  - Problem with variance and normality

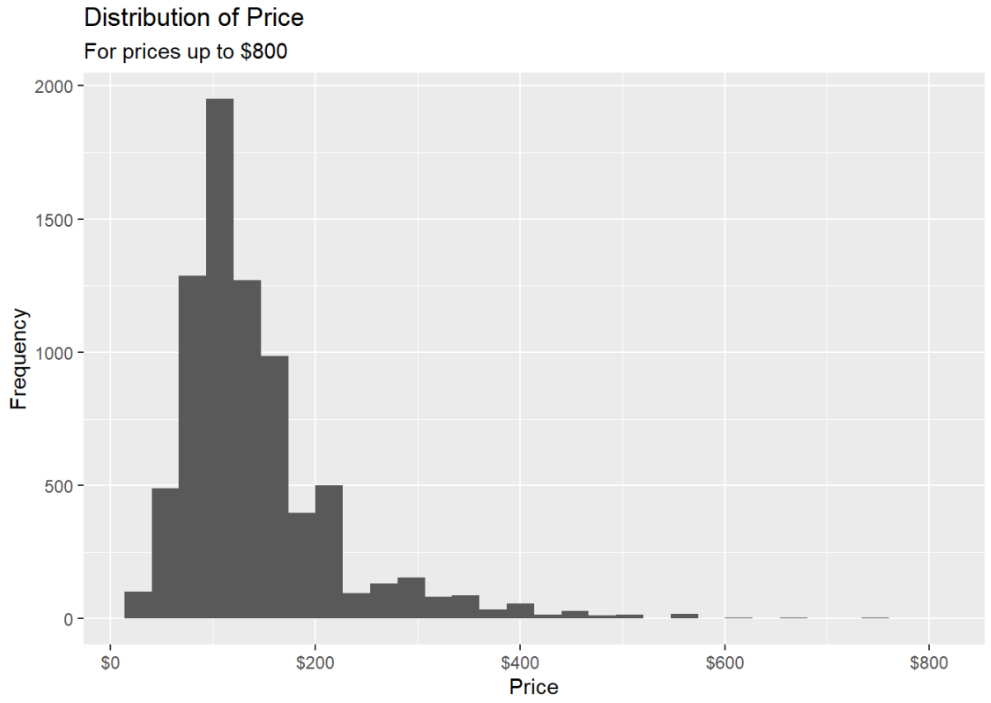
```
## # A tibble: 11 x 3
##    term                    estimate   p.value
##    <chr>                      <dbl>     <dbl>
##  1 (Intercept)               92.8    1.67e- 14
##  2 room_typePrivate room    -56.1    1.40e-152
##  3 room_typeShared room     -78.6    7.63e- 15
##  4 bathrooms1-1.5            15.3    3.42e-  2
##  5 bathrooms2-2.5            92.2    2.11e- 29
##  6 bathrooms3+              114.     1.12e- 19
##  7 bedrooms3-5               81.3    5.40e-143
##  8 bedrooms6+               284.     4.70e- 94
##  9 minimum_nights4-7         -2.21   3.56e-  1
## 10 minimum_nights8+         -32.2    1.90e-  5
## 11 review_scores_rating       0.327  1.66e-  3
```

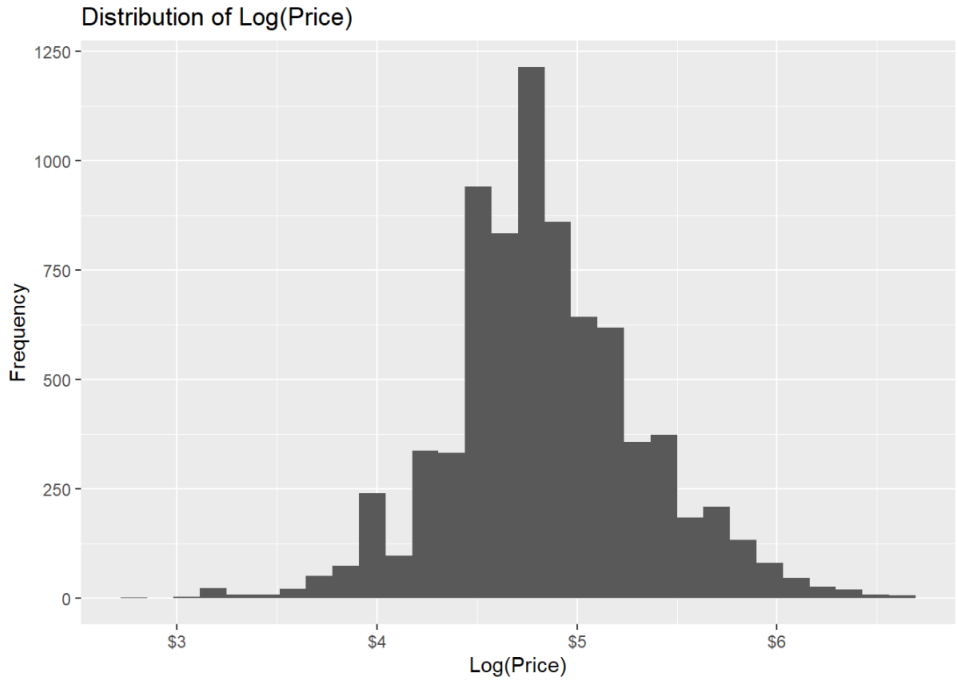




Transformation Time

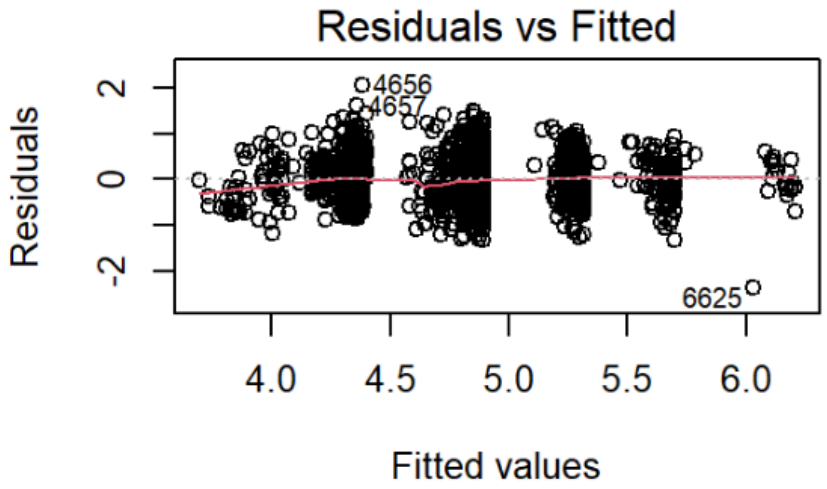
- Price data = Right-skewed -> Log transformation = Not skewed
- New adjusted r-squared = 0.393

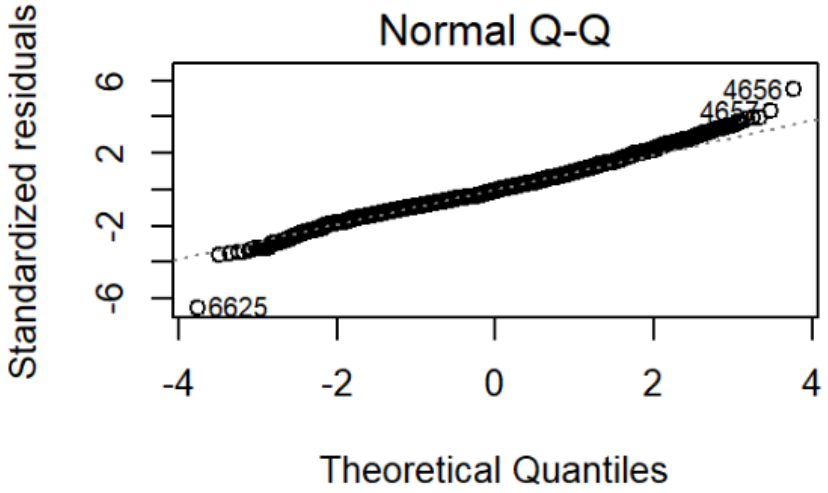




Transformation Time

- Residual problem = fixed
- Normality problem = fixed

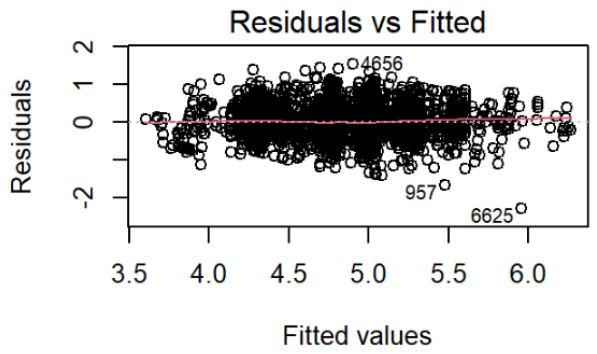


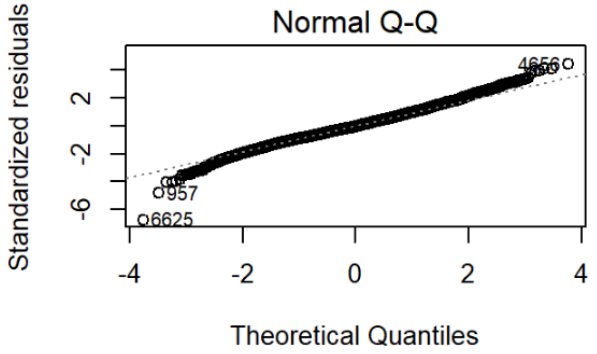


Final Model & Statistics

- Can we produce a stronger model?
  - Property type & accommodates added
- Model #2 & Final Model
  - Create levels within variables
  - 7 total variables
  - Backwards Elimination
  - Adjusted r-squared = 0.472

```
## # A tibble: 20 x 4
##    term                          estimate   p.value estimate_exp
##    <chr>                            <dbl>     <dbl>        <dbl>
##  1 (Intercept)                    4.20     0              66.9
##  2 room_typePrivate room         -0.468    3.62e-249       0.626
##  3 room_typeShared room          -0.814    3.16e- 43       0.443
##  4 bathrooms1-1.5                 0.120    4.41e-  3       1.13
##  5 bathrooms2-2.5                 0.395    1.05e- 16       1.49
##  6 bathrooms3+                    0.192    9.24e-  3       1.21
##  7 bedrooms3-5                    0.233    5.17e- 32       1.26
##  8 bedrooms6+                     0.514    8.06e- 10       1.67
##  9 minimum_nights4-7             -0.00208  8.80e-  1       0.998
## 10 minimum_nights8+              -0.419    8.56e- 22       0.658
## 11 review_scores_rating           0.00468  8.46e- 15       1.00
## 12 accommodates4-6                0.250    1.32e-120       1.28
## 13 accommodates7+                 0.524    8.46e- 54       1.69
## 14 property_typeHouse             0.0561   1.10e-  3       1.06
## 15 property_typeBed & Breakfast   0.162    2.82e- 12       1.18
## 16 property_typeBoat              0.162    1.44e- 12       1.18
## 17 property_typeLoft              0.0319   4.94e-  1       1.03
## 18 property_typeCabin            -0.288    1.31e-  2       0.750
## 19 property_typeCamper/RV        -1.07     5.82e- 18       0.344
## 20 property_typeOther             0.0179   7.76e-  1       1.02
```





Answers to Research Question

- What factors have an impact the price of an Airbnb?
  - All 7 factors researched
    - Final model's backward elimination results

- Is there a combination of factors that have a strong effect on price?
  - 7 variables instead of 5 variables
    - Adjusted r-squared values: 0.393 (5) vs. 0.472 (7)

Conclusion

- Introduction of Data
- EDA & Visualizations
- Models & Statistics

- Key Takeaways
  - Multiple variables are better than just 1!
  - Ways to fix residuals
    - Transformation and additional variables
  - Every part of the process is important!

Future Research

- Explore the variable "city"
  - Would require lots of data wrangling
- Geographic location vs. Price
  - Possibility of map plot using latitude and longitude
  - May have to filter by town or neighborhood
- Transform overall rating variable
  - Left-skewedness creates difficulties


Thank you!