

Airbnb Berlin Price Prediction - Annex

Seong Woo AHN, Paul BÉRARD, Leila BERRADA, Tanguy BLERVACQUE

1. Introduction

This appendix is destined to gather all the graphs and data visualisations of the models and methods explored in the Airbnb Berlin Price Prediction article.

2. Data Preprocessing

2.1. Dropping irrelevant features

2.2. Handling Missing Values

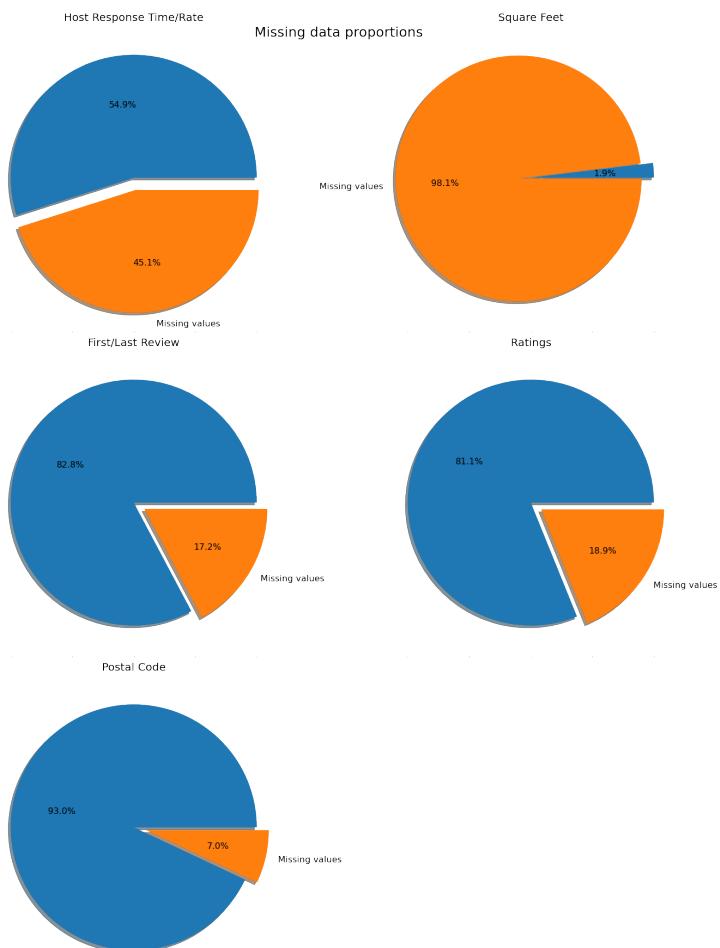


Figure 1. Missing data proportion

2.3. Handling outliers

2.4. Feature scaling

2.5. Handling categorical data

2.6. Feature engineering

2.7. Visualisation

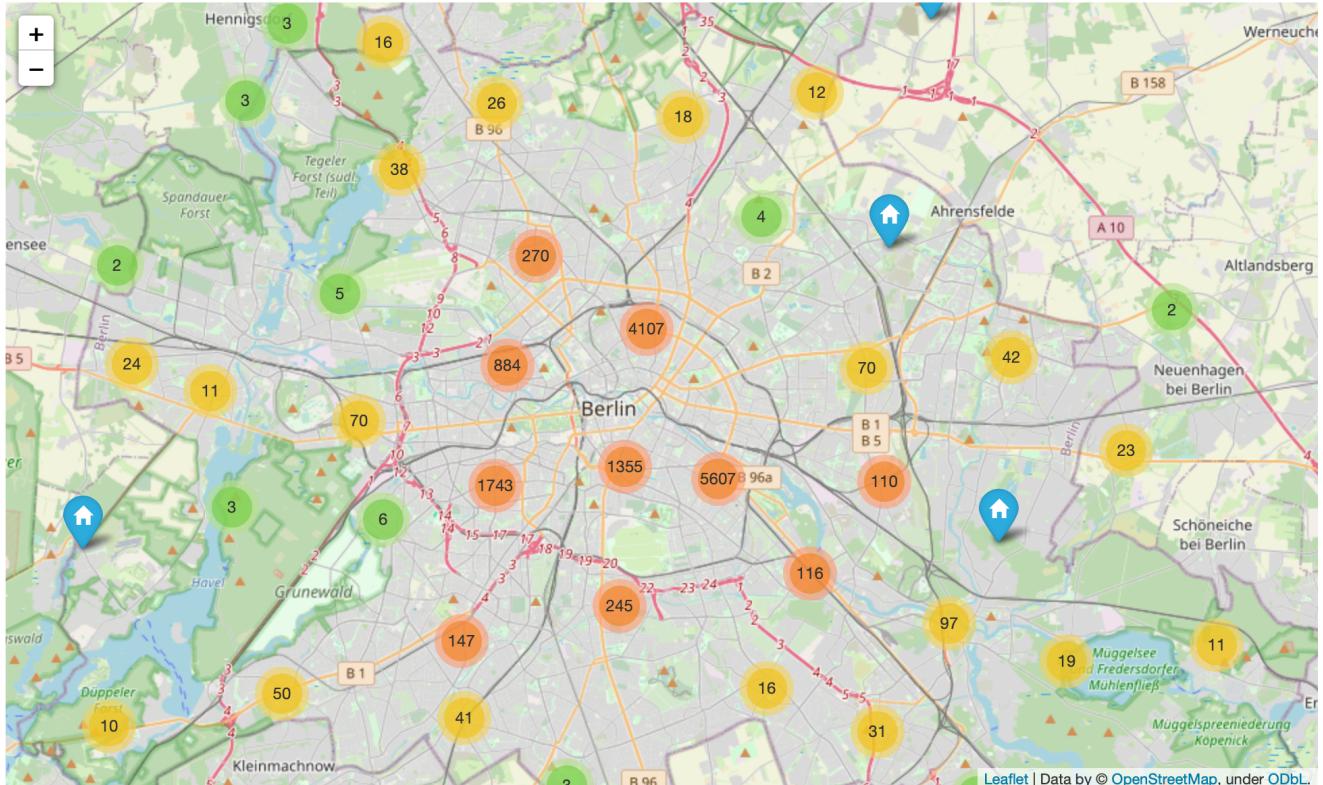


Figure 2. Visualisation of our Airbnb Listings on the Map of Berlin

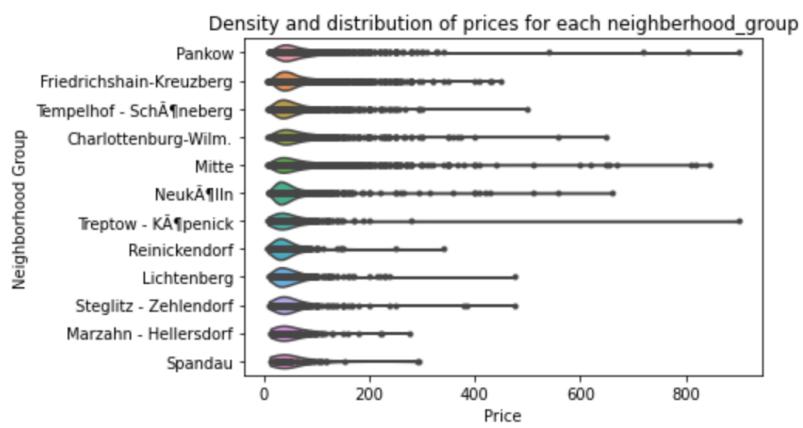


Figure 3. Repartition of prices in the different Neighborhood Groups

3. Price Prediction Models

3.1. Linear Regression

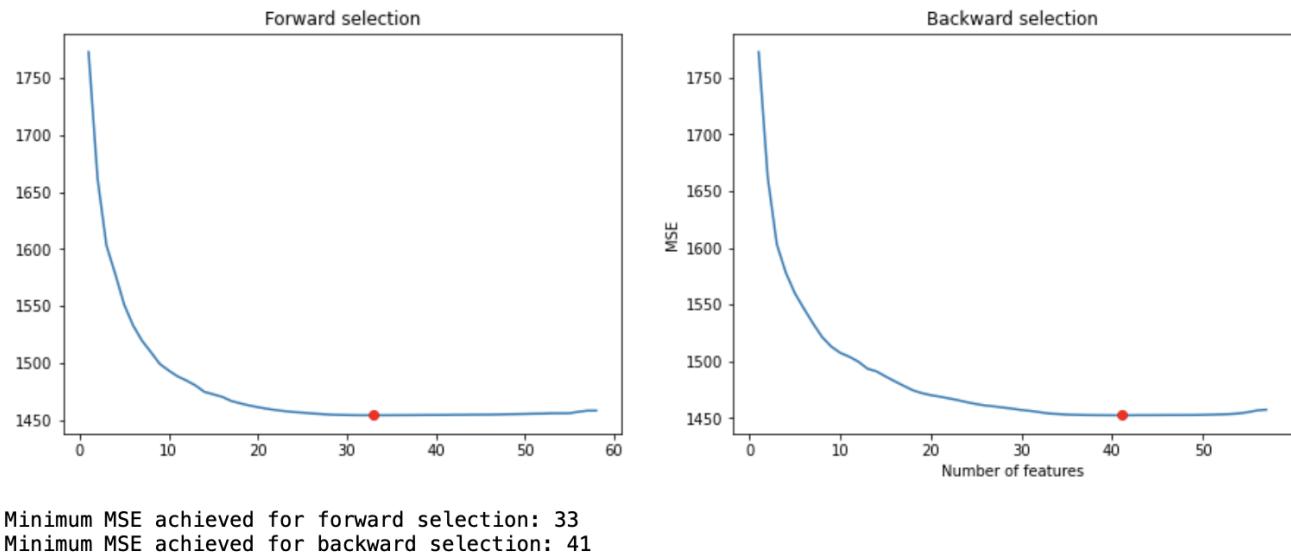


Figure 4. Linear Regression - subset selection

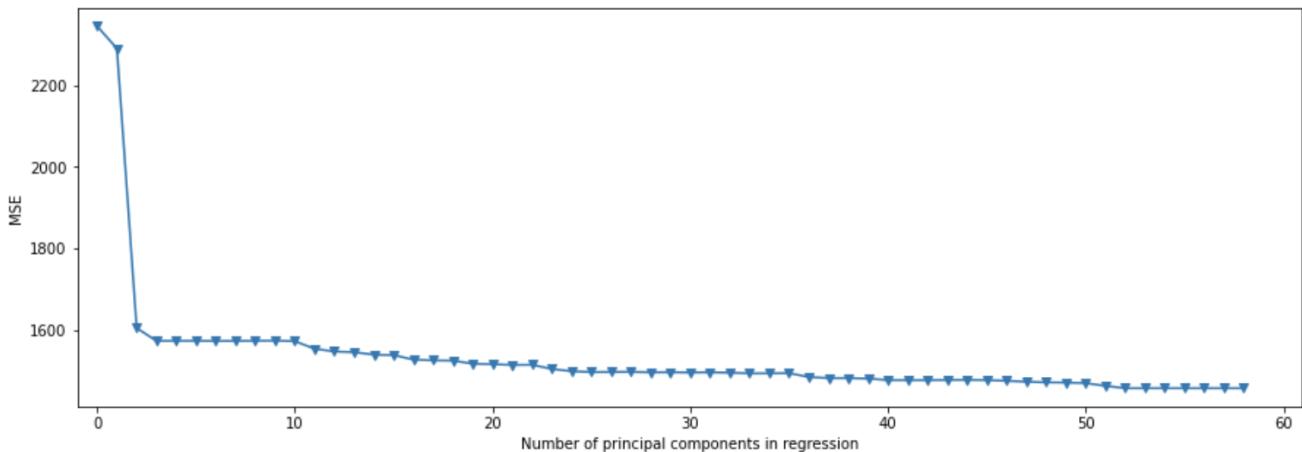


Figure 5. Linear Regression - PCA



Figure 6. Linear Regression - Visualisation of Shapley Analysis

3.2. Decision Tree Regression

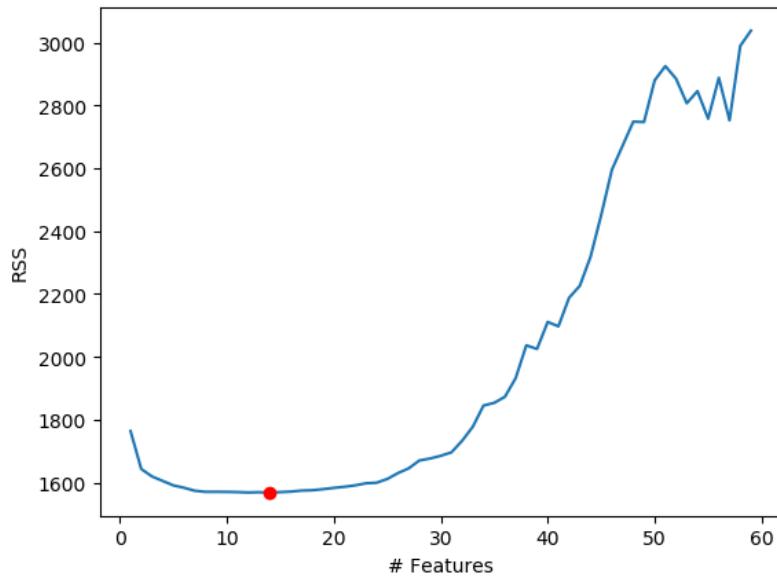


Figure 7. Evolution of MSE for forward feature selection

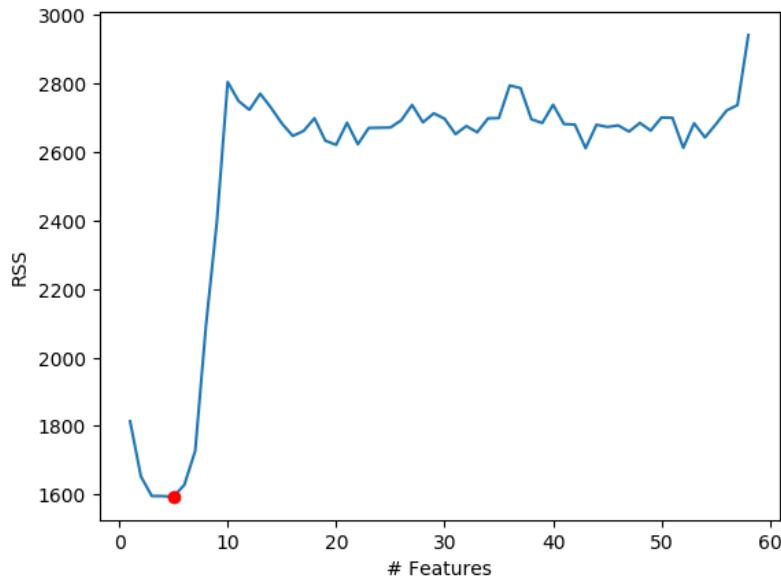


Figure 8. Evolution of MSE for backward feature selection

3.3. Bayesian Ridge Regression

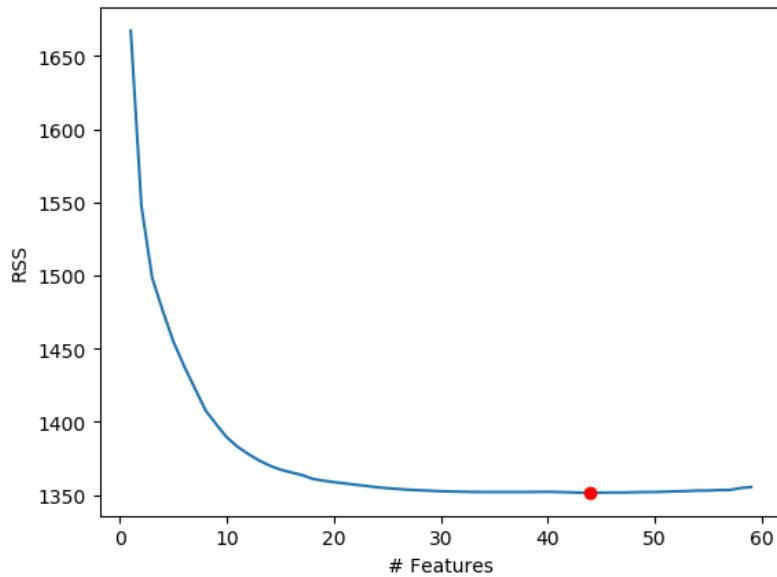


Figure 9. Evolution of MSE for forward feature selection

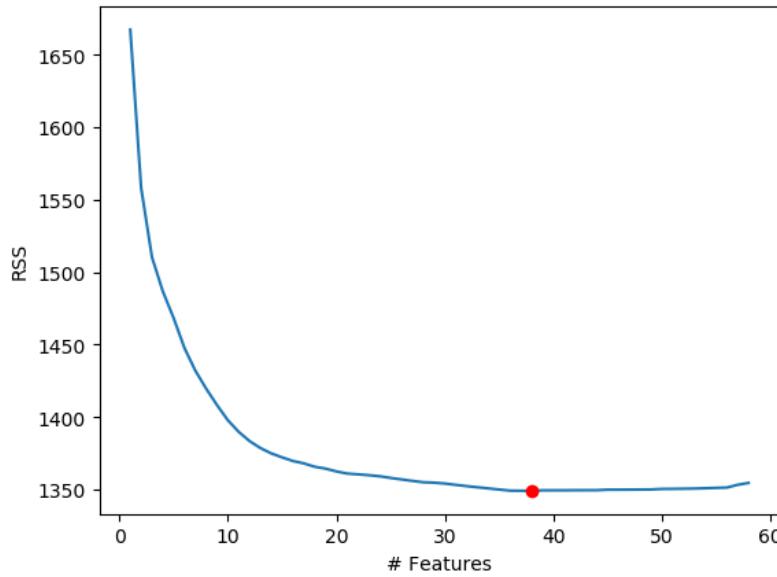


Figure 10. Evolution of MSE for backward feature selection

3.4. KNN Regression

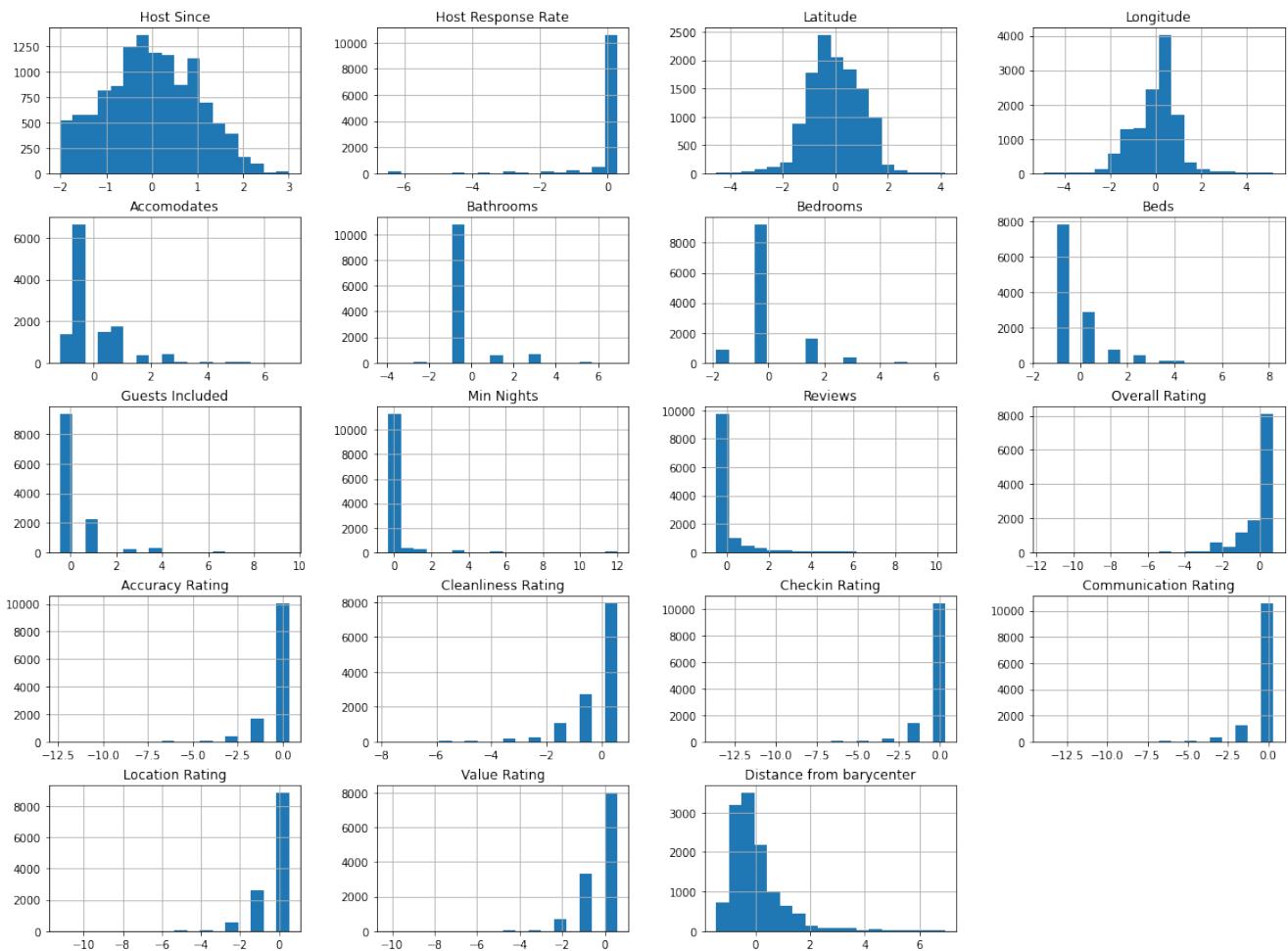


Figure 11. Distribution of data for train set (imputed and scaled)

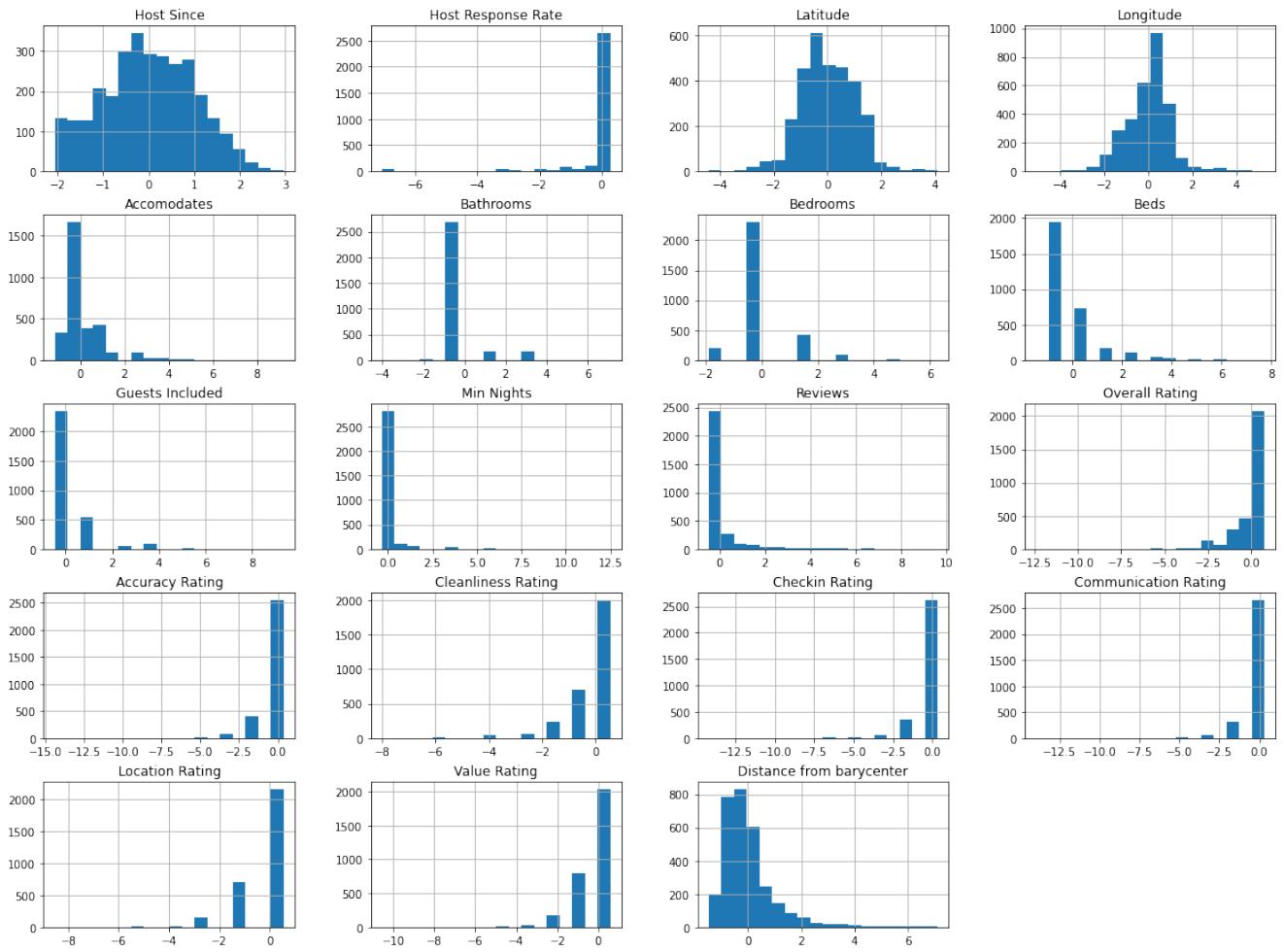


Figure 12. Distribution of data for test set (imputed and scaled)

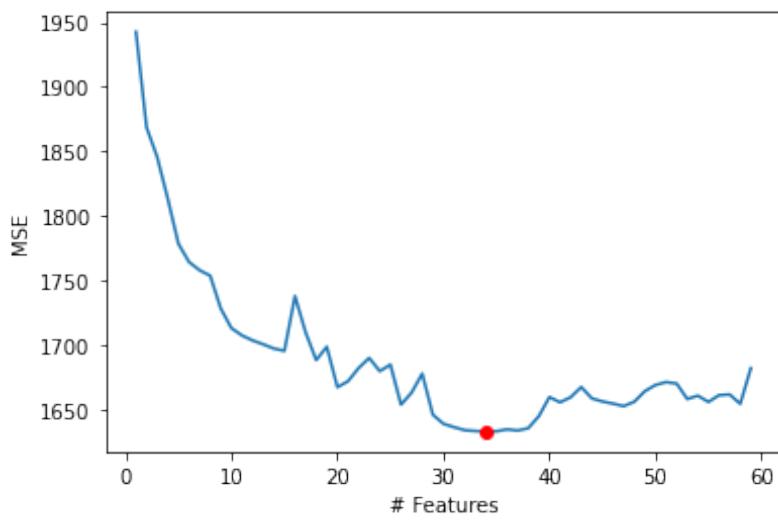


Figure 13. Evolution of MSE for forward feature selection

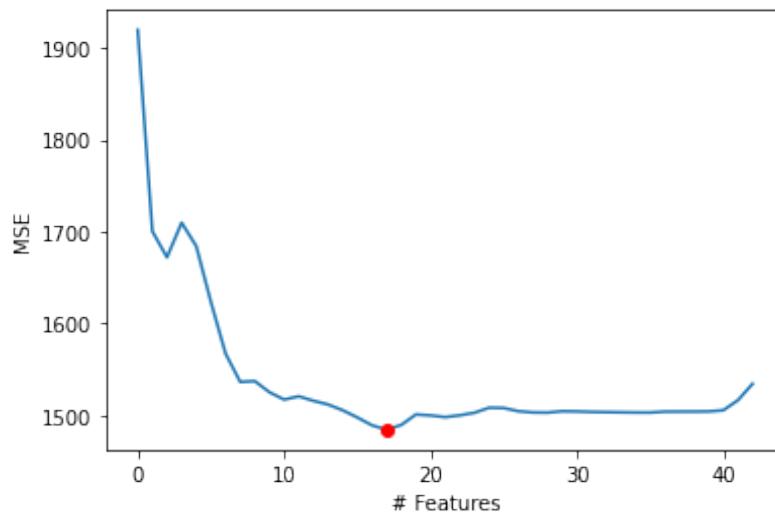


Figure 14. Evolution of MSE for backward feature selection

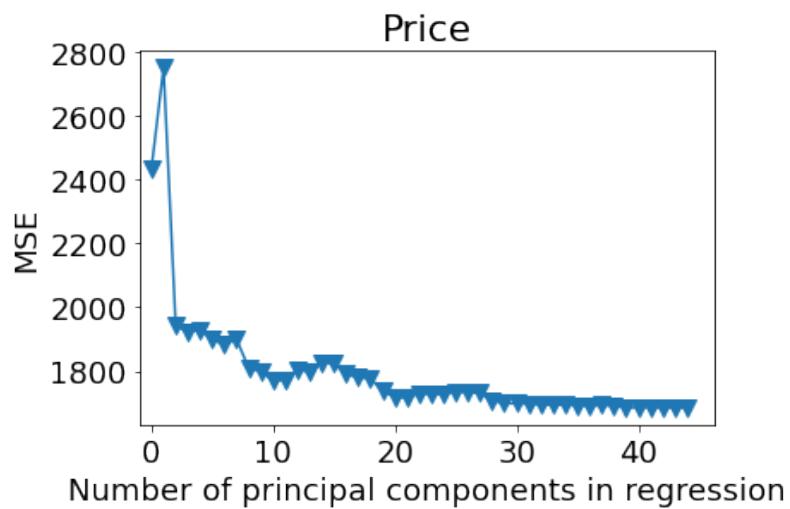


Figure 15. Evolution of MSE for PCA

3.5. Random Forest Regression

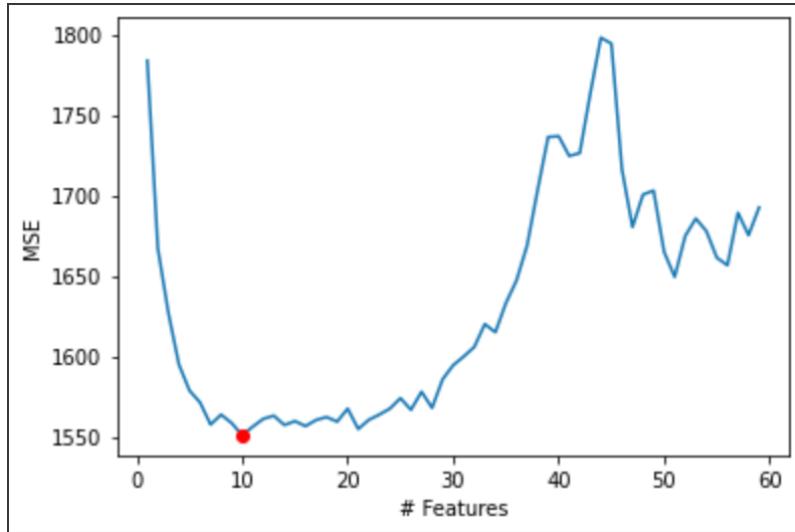


Figure 16. Random Forest - Graph forward feature selection

```
Variable: Accomodates           Importance: 0.11
Variable: Bedrooms               Importance: 0.1
Variable: Room Type_Entire home/apt Importance: 0.08
Variable: Room Type_Private room Importance: 0.08
Variable: Beds                   Importance: 0.07
Variable: Bathrooms              Importance: 0.06
Variable: Distance from barycenter Importance: 0.06
Variable: Latitude                Importance: 0.05
Variable: Guests Included        Importance: 0.05
Variable: Host Since              Importance: 0.04
Variable: Longitude               Importance: 0.04
Variable: Reviews                 Importance: 0.03
Variable: Min Nights              Importance: 0.02
Variable: Overall Rating          Importance: 0.02
Variable: Property Type_Loft       Importance: 0.02
Variable: Host Response Rate      Importance: 0.01
Variable: Cleanliness Rating      Importance: 0.01
Variable: Location Rating         Importance: 0.01
Variable: Value Rating             Importance: 0.01
Variable: Host Response Time_unknown Importance: 0.01
Variable: Host Response Time_within an hour Importance: 0.01
Variable: Property Type_Apartment Importance: 0.01
Variable: Property Type_Hotel      Importance: 0.01
Variable: Property Type_Serviced apartment Importance: 0.01
Variable: Neighborhood Group_Mitte Importance: 0.01
Variable: Boy or Girl?_Boy        Importance: 0.01
Variable: Boy or Girl?_Girl       Importance: 0.01
Variable: Is Superhost            Importance: 0.0
Variable: Is Exact Location        Importance: 0.0
Variable: Accuracy Rating          Importance: 0.0
Variable: Checkin Rating           Importance: 0.0
Variable: Communication Rating    Importance: 0.0
Variable: Instant Bookable         Importance: 0.0
```

Figure 17. Random Forest - Important features from RF

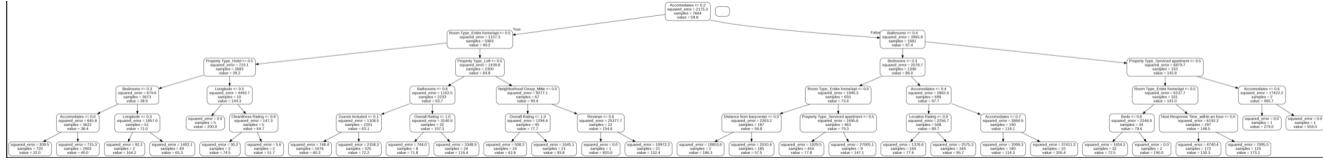


Figure 18. Random Forest - One tree of our forest



Figure 19. Random Forest - Visualisation of Shapley Analysis

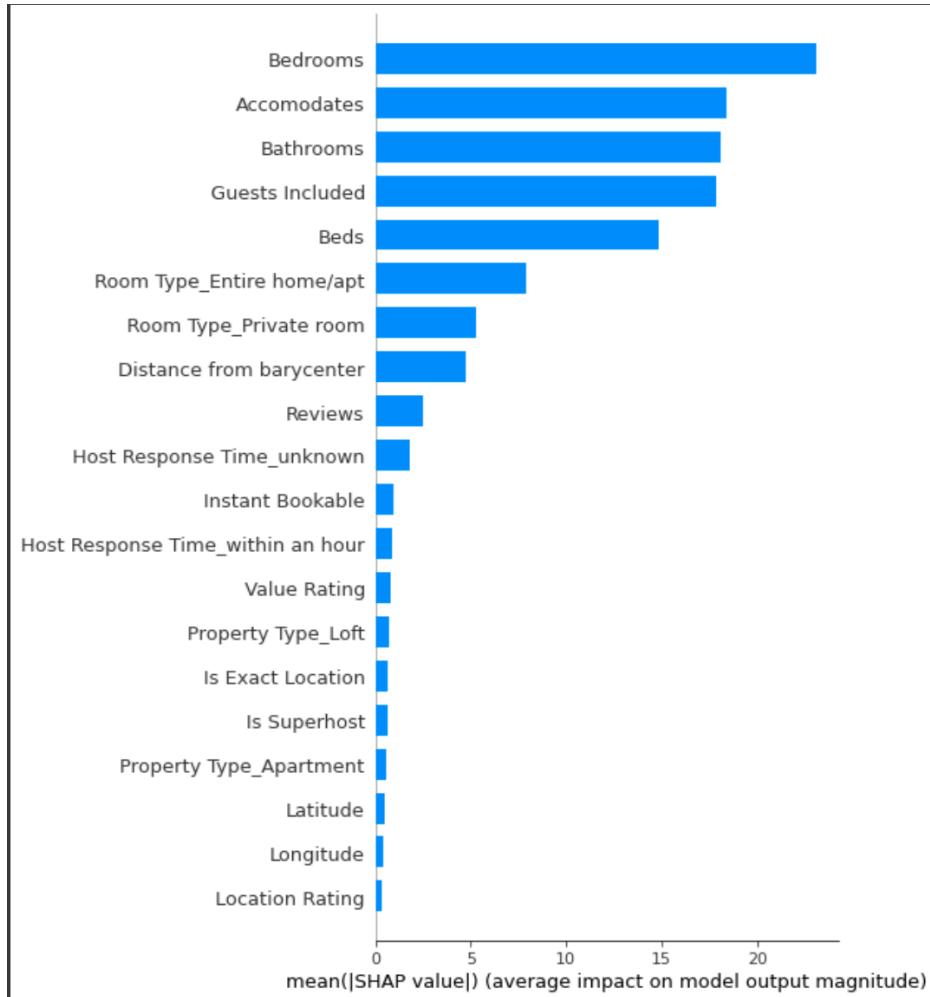


Figure 20. Random Forest - Visualisation of Shapley Analysis

3.6. SVR

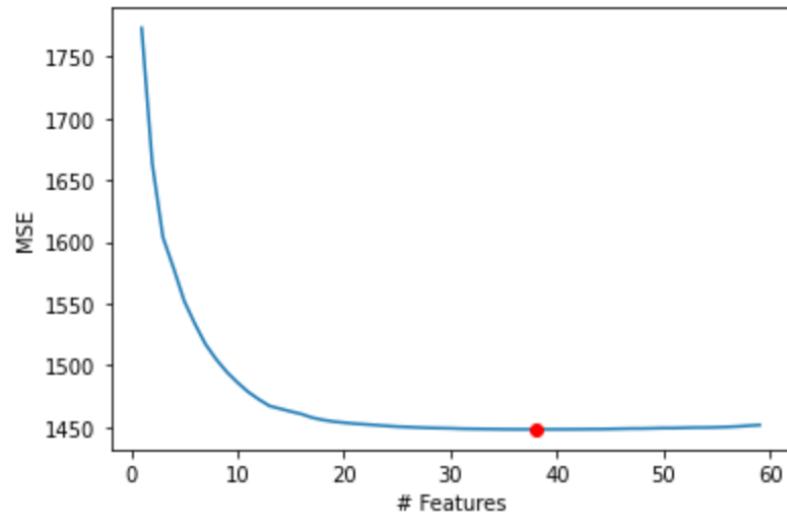


Figure 21. Linear SVR - Forward Feature Selection



Figure 22. Linear SVR - Visualisation of Shapley Analysis

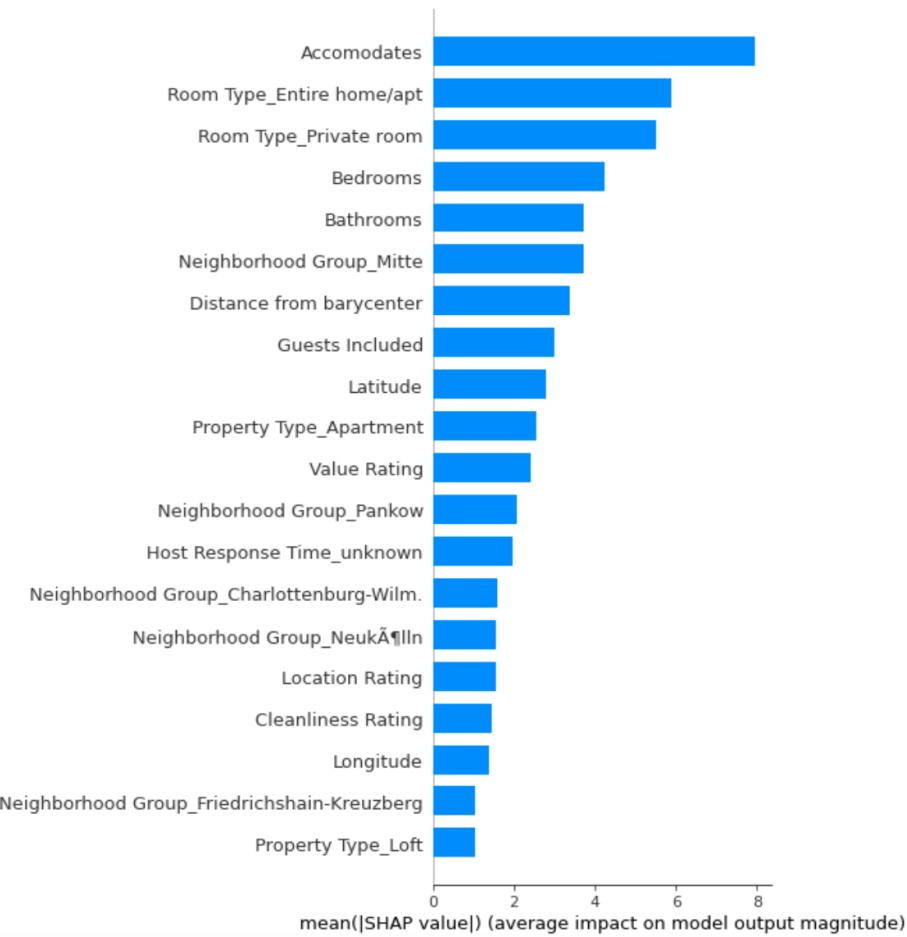


Figure 23. Linear SVR - Visualisation of Shapley Analysis

3.7. AdaBoost

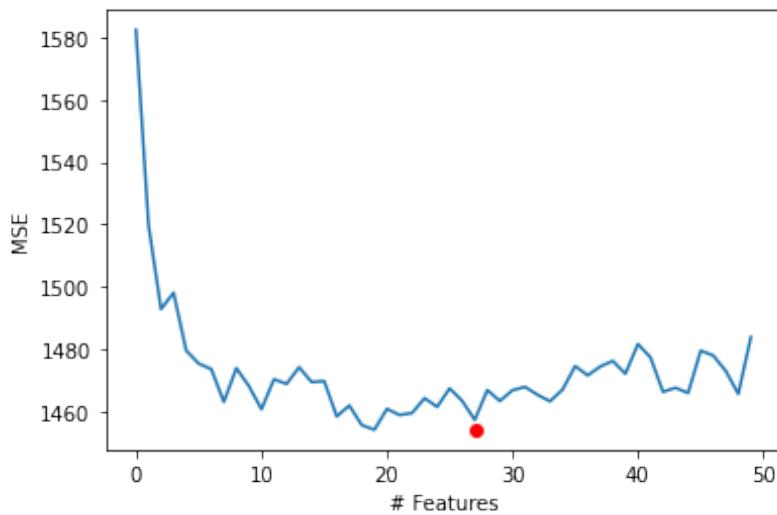


Figure 24. Evolution of MSE for backward feature selection

3.8. XGBoost

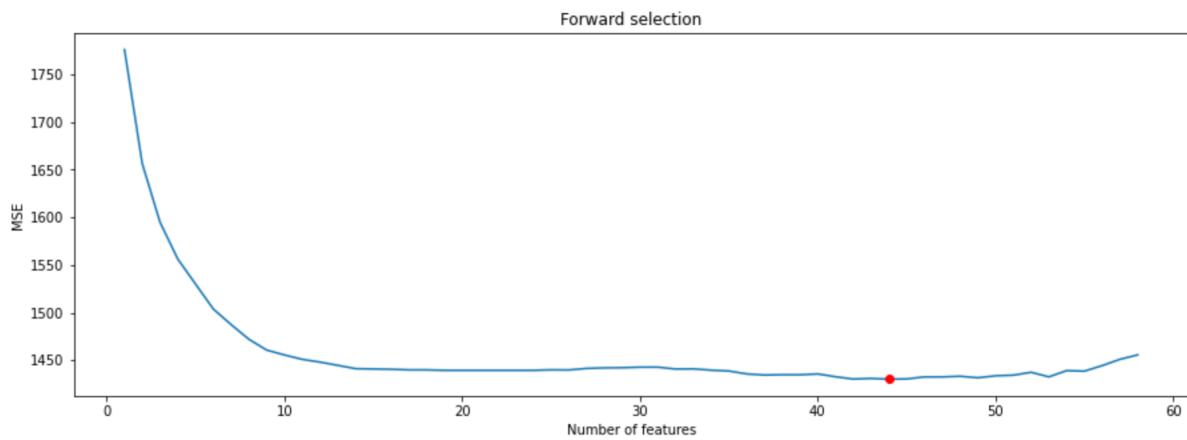


Figure 25. XGBoost - Forward feature selection

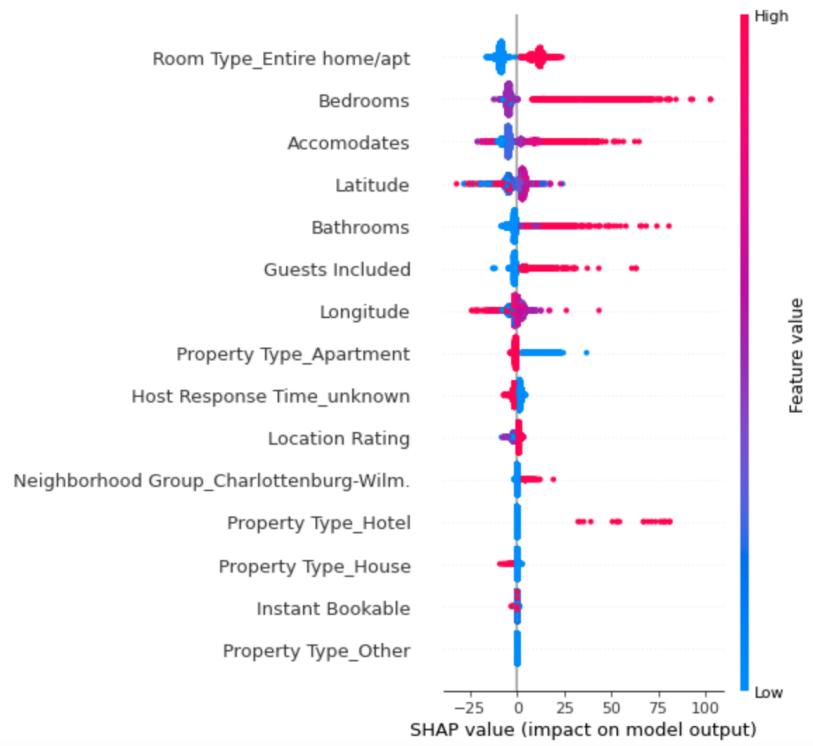


Figure 26. XGBoost - Visualisation of Shapley Analysis