

AIRBNB: PRICE PREDICTION IN PRAGUE

Team 3: Tomáš Mikulenka, Lukáš Kuthan,
Adéla Smrčková, Jana Štolcová

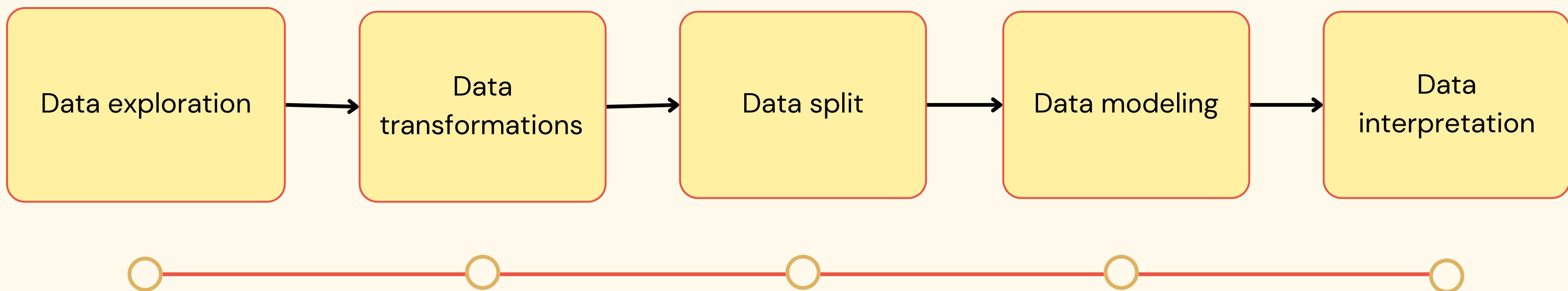


Problem & Business case

We have an apartment for 4 people in Prague center and we would like to predict the market price based on historical data.



Project timeline



Data preparation

Metrics

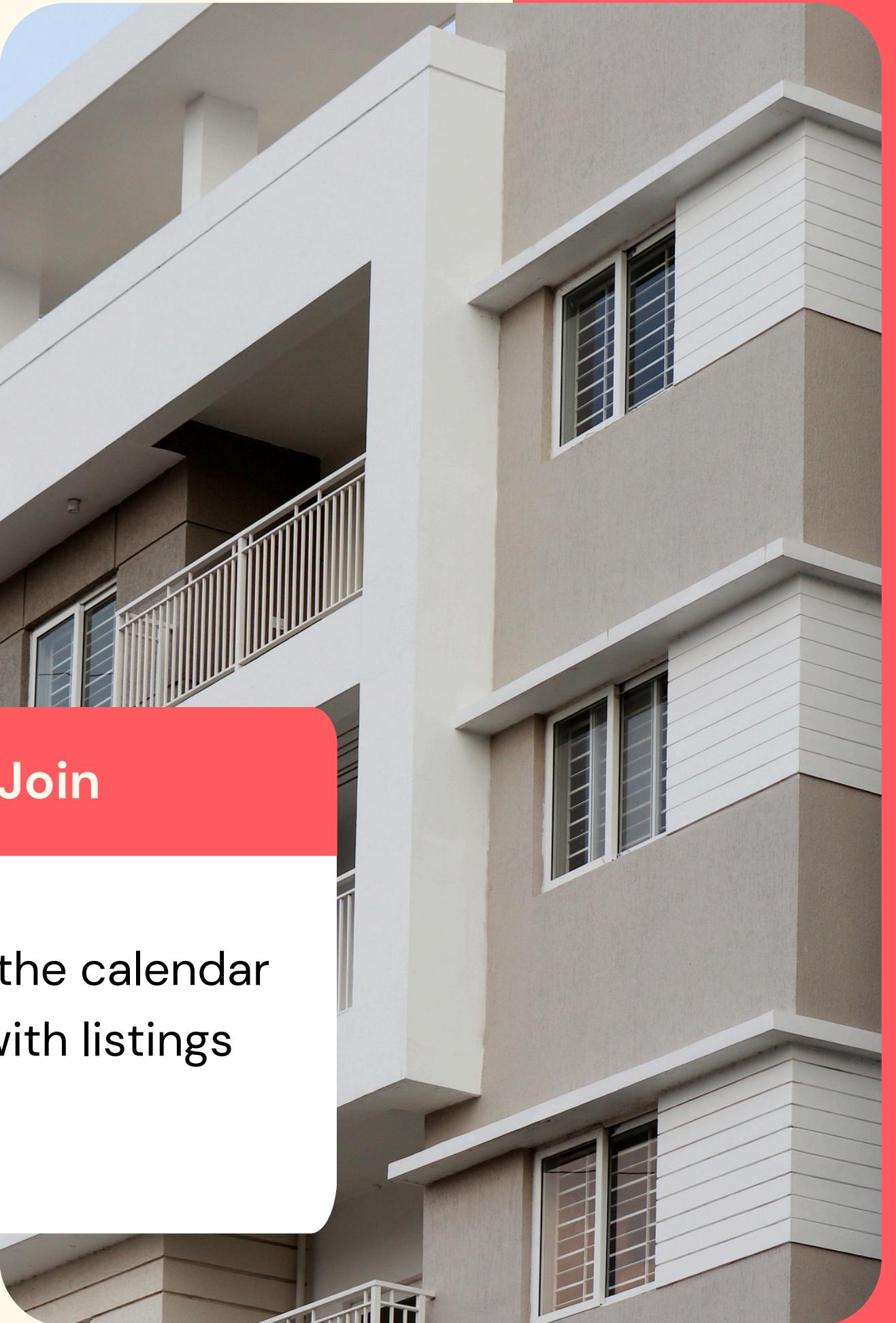
Selection of relevant metrics

Transformations

Data binning
Conversion to numeric
Selecting outliers

Join

Joining the calendar data with listings



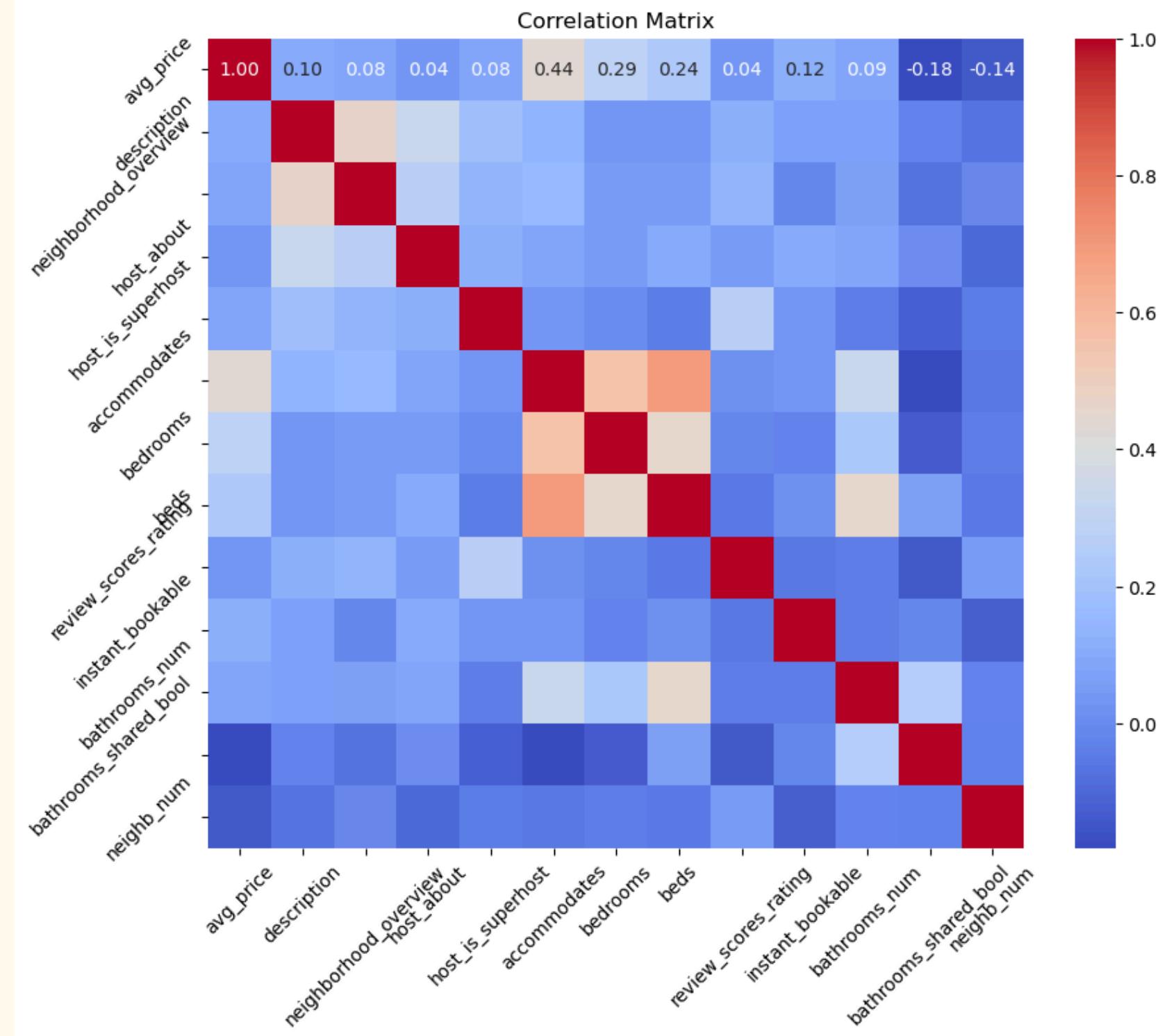
Parameters used

Column	Keep	Change	Reason
description	Yes, but Count number of signs		Could be relevant
neighborhood_overview	Yes, but Is or not - boolean		Could be relevant
host_since	Yes, but Difference from today and in bins (less than year, more than year)		Could be relevant
host_about	Yes, but Is or not - boolean		Could be relevant
host_is_superhost	Yes, but 1/0 boolean		Could be relevant
host_has_profile_pic	Yes, but 1/0 boolean		Could be relevant
host_identity_verified	Yes, but 1/0 boolean		Could be relevant
neighbourhood_cleansed	Yes, but Make bins		Relevant
room_type	Yes		Relevant
accommodates	Yes		Relevant
bathrooms_text	Yes, but Convert to number and create a new column boolean shared_bathroom		Relevant
bedrooms	Yes, but Replace blanks with 1		Relevant
beds	Yes, but Replace blanks with 1		Relevant
number_of_reviews	Yes		Could be relevant
review_scores_rating	Yes		Could be relevant
instant_bookable	Yes, but 1/0 boolean		Could be relevant

Legend
adjust
keep

Column	Keep	Change	Reason
date	Yes, but Change the resolution to month and year only		Relevant
price	Yes, but Aggregate price as average for given month and year		Relevant

Correlation matrix



Data split

60 % Train

20 % Valid

20 % Test

2023

2024

Sorted by date

Used models

Linear
Reggresion

Lasso
Regression

Random Forrest

XGBoost

RMSE as key measure for evaluation & comparison

Linear Regression

Validation
RMSE

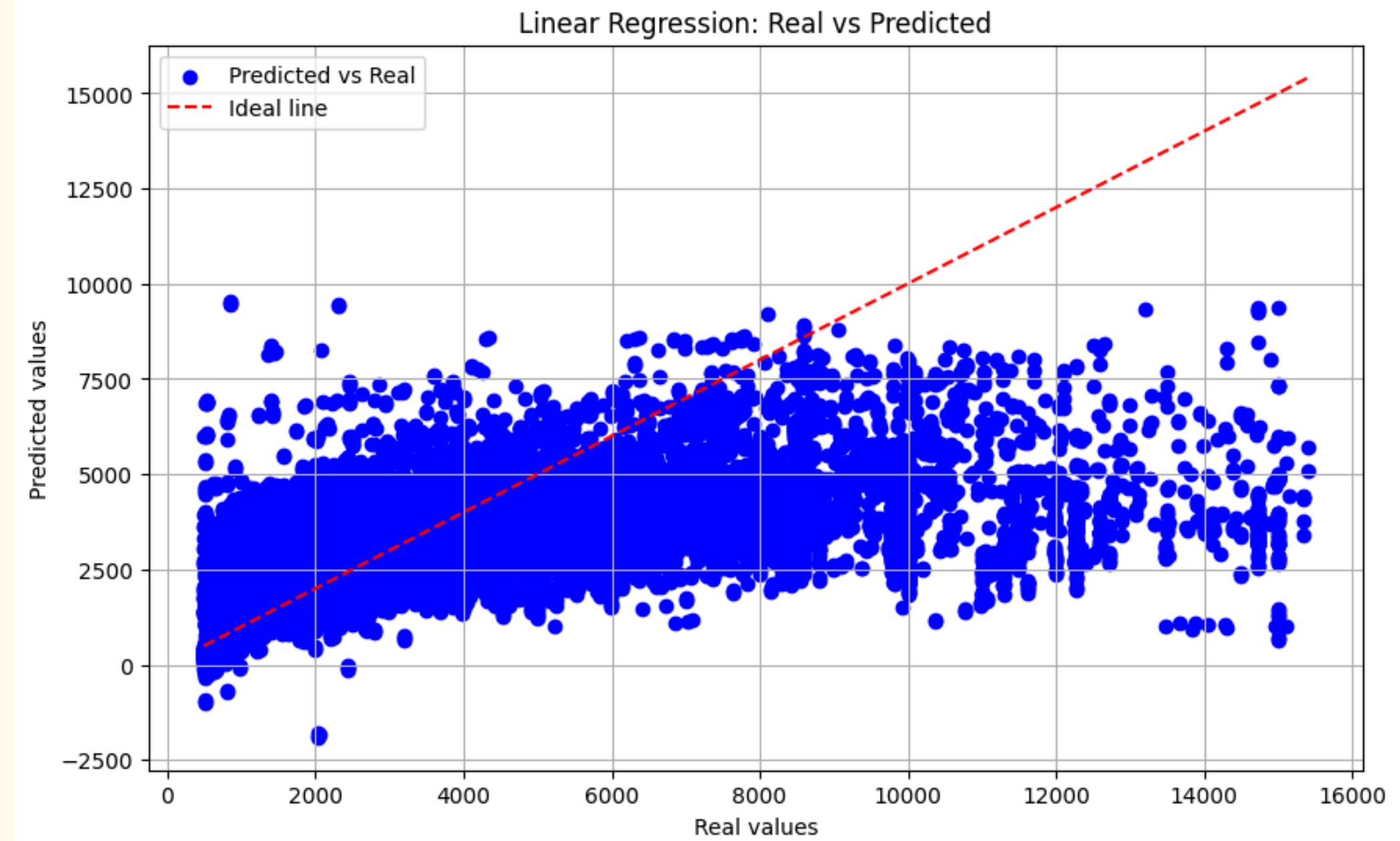
2403.74

Test
RMSE

2481.63

Coefficients

```
[ 5.70560978 22.73911238  
-130.36032637 -55.00277619  
199.31731746 429.44269482  
-918.90714623 403.68461469  
209.79935586 -79.76090382  
-2.39249414 203.46521552  
245.94407167 -130.36032637  
62.2399841 -503.21439886  
297.31264558 38.99819051  
-195.82174176 676.85574994  
-481.03400817 56.06359818  
639.56760892 108.49498473  
-804.12619182]
```



Lasso Regression

Coefficients

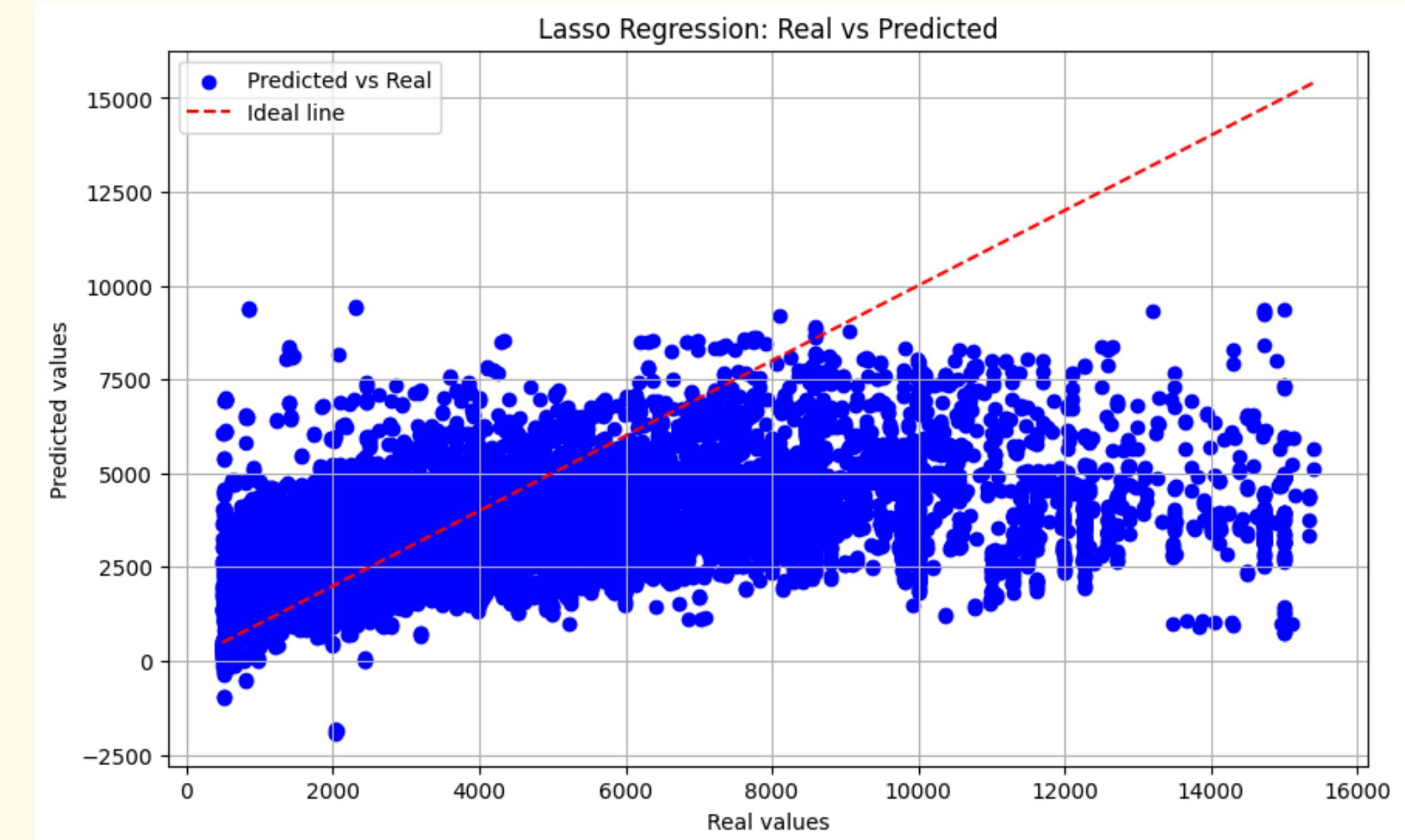
```
[ 6.51754046 18.64938677  
-58.20277747 -50.45990542  
193.25813266 371.24951444  
-889.94211536 406.56873007  
209.73652472 -82.97424289  
-2.39099193 199.64321291  
240.51406609 -148.89703873  
56.10230389 -486.96076918  
274.38286681 36.46507818 -0.  
869.51694696 -259.65078537  
-12.52204476 514.6849589  
23.81181889 -777.3042681 ]
```

Validation
RMSE

2406.16

Test
RMSE

2484.01



Random Forrest

Validation
RMSE

1534.02

Test
RMSE

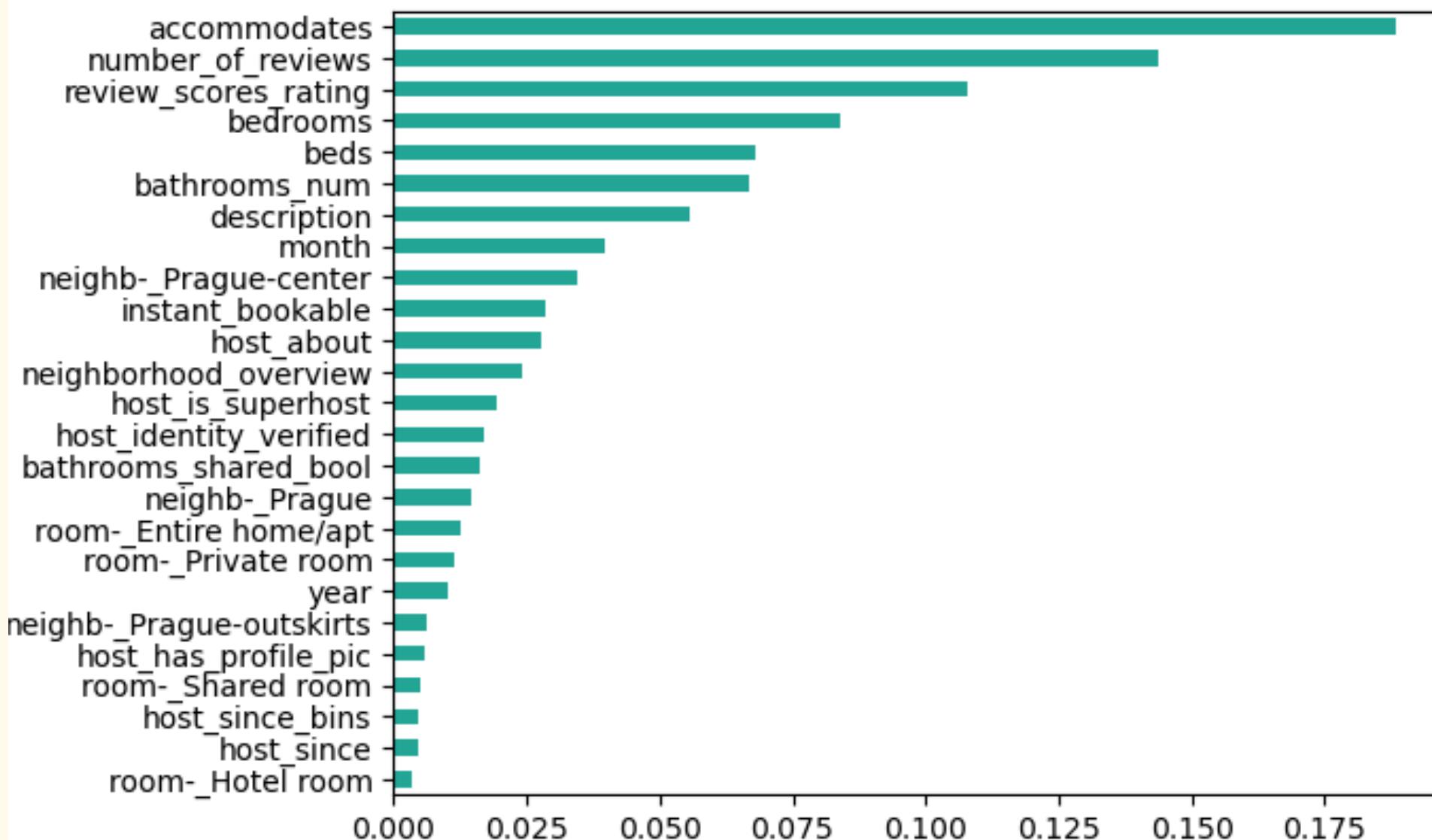
1692.49

Grid Search

```
n_estimators:  
[4000,5000,5500],  
max_features:  
[5,9,12],  
min_samples_split:  
[10, 20],  
max_depth:  
[None, 5, 10]
```

Winning parameters

```
n_estimators:  
[4000],  
max_features:  
[12],  
min_samples_split:  
[10],  
max_depth:  
[None]
```



XGBoost

Validation
RMSE

1468.51

Test
RMSE

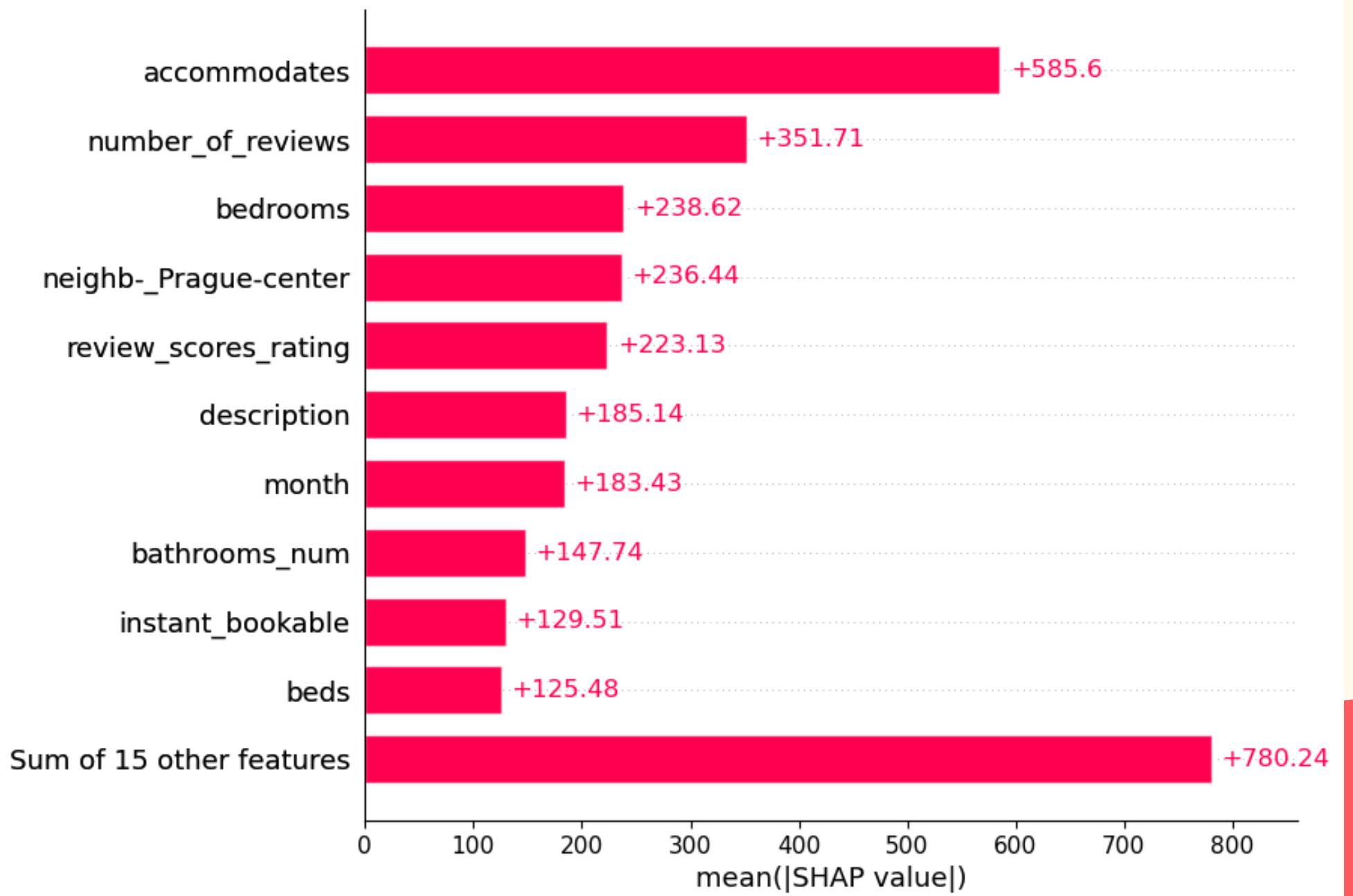
1614.74

Grid Search

```
learning_rate: [0.1, 0.5, 0.9],  
n_estimators: [20, 50, 100],  
subsample: [0.5, 0.8, 1],  
max_depth: [5, 8, 11],  
colsample_bytree: [0.5, 1]
```

Winning parameters

```
learning_rate: [0.5],  
n_estimators: [100],  
subsample: [1],  
max_depth: [11],  
colsample_bytree: [0.5]
```



Comparasion

Linear
Reggresion

Lasso
Regression

Random
Forrest

XGBoost

Validation
RMSE

2403.74

2406.16

1534.02

1468.51

Test
RMSE

2481.63

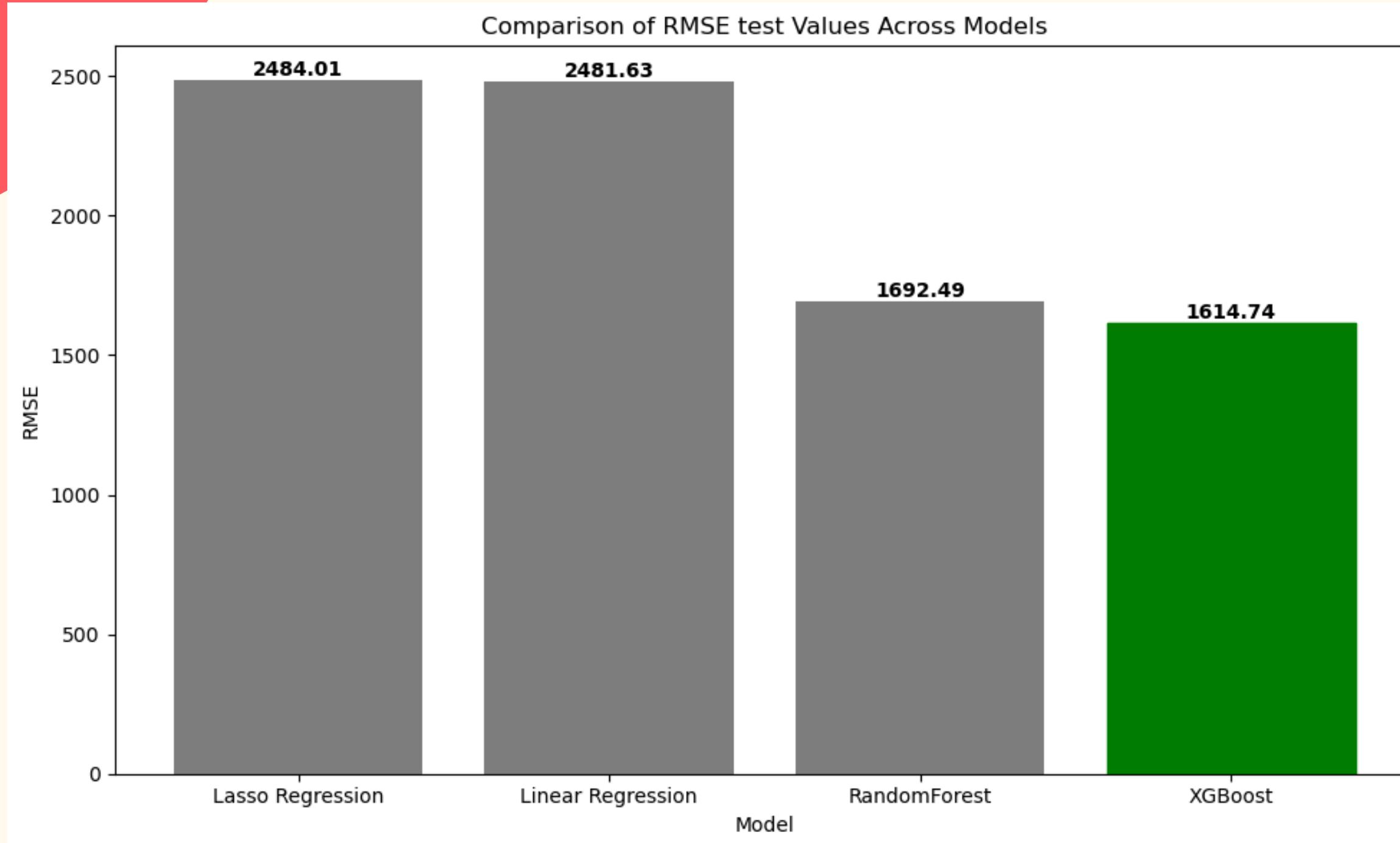
2484.01

1692.49

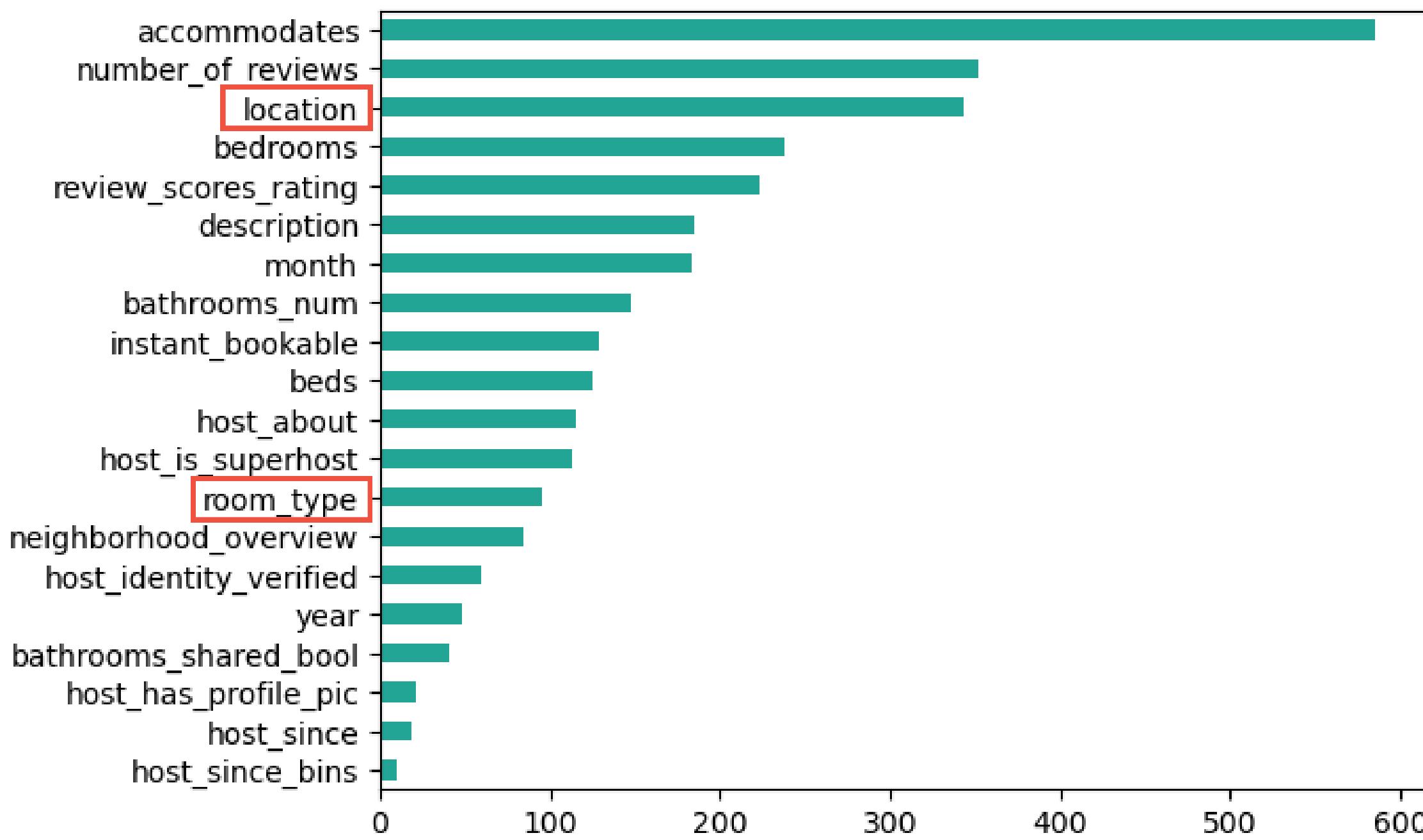
1614.74

Winner

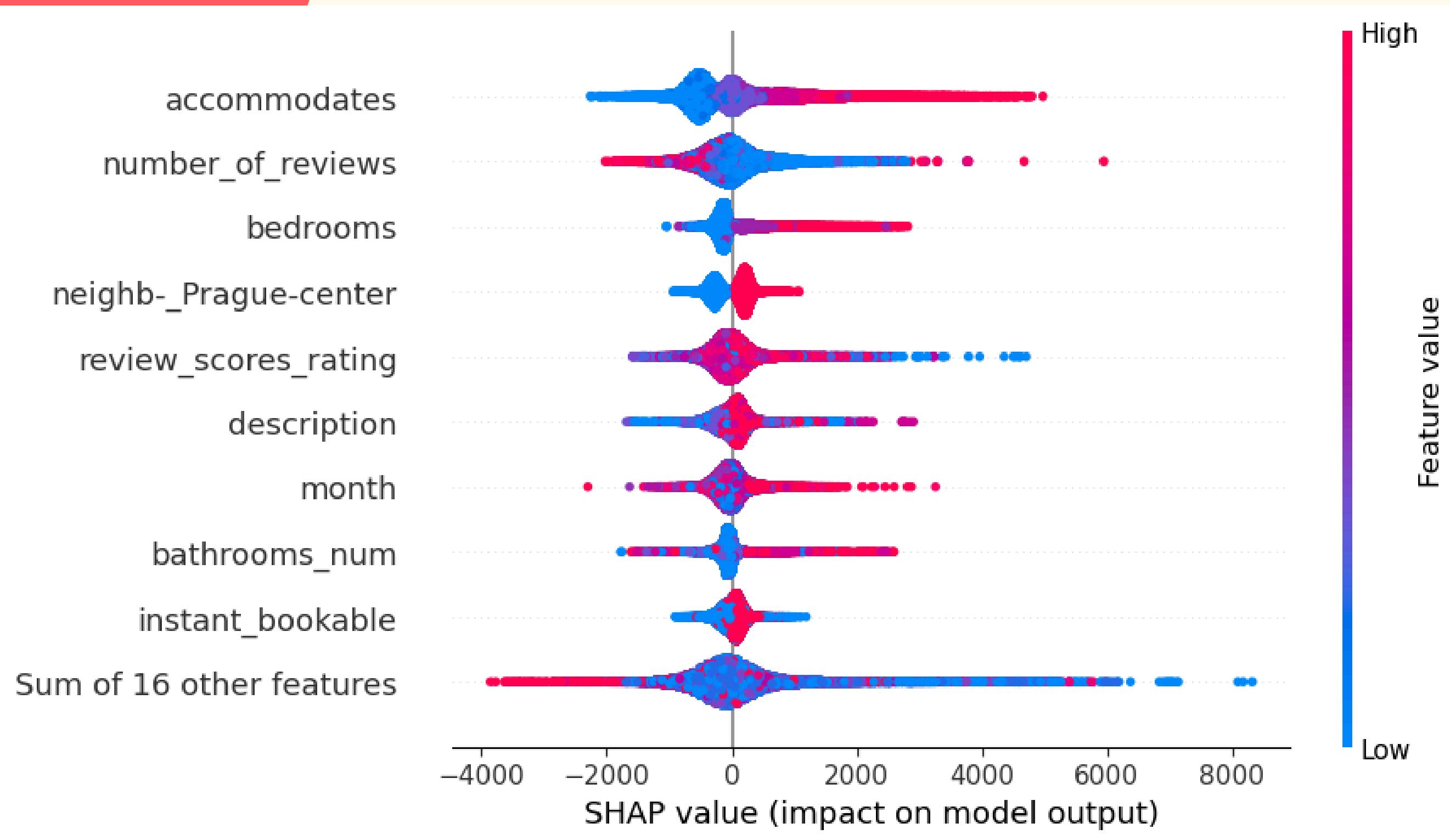
XGBoost



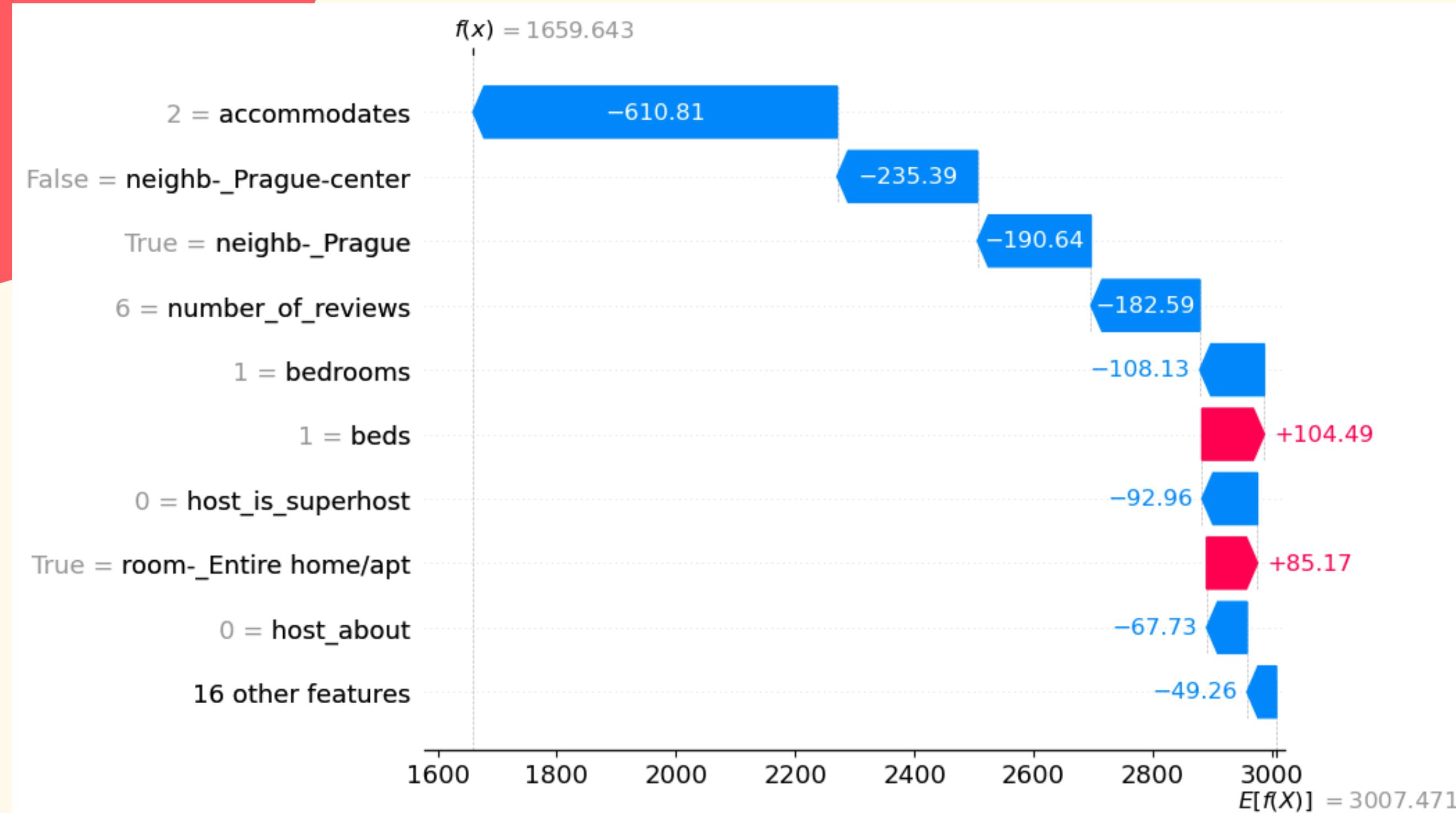
XGBoost SHAP



XGBoost SHAP



XGBoost SHAP



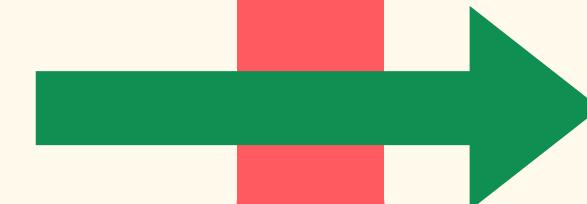
Solution to our problem

Prague center apartment

Starting Airbnb business as new user with lack of experience

- 4 accommodates
- 2 bedrooms
- 2 beds
- 1 bathroom
- expected rating 4.6 (Ø)
- 0 reviews
- poor description (bin 2)
- not superhost

1 814 CZK



Prague center apartment

More skilled user and host after couple of months at Christmas

- 4 accommodates
- 2 bedrooms
- 2 beds
- 1 bathroom
- expected rating 4.7
- 25 reviews
- improved description (bin 4)
- superhost ✓

2 969 CZK



Suggestions for improvement

Sentiment analysis

Conduct sentiment analysis on the standalone reviews dataset and append the results to the listings.

Amenities

Every Airbnb listing boasts its own range of amenities. Evaluate the significance of these amenities and integrate them into the model.

Geojson

Utilize the GeoJSON dataset to leverage location polygons to have more precise geographical data.

Discussion