Prediction with Machine Learning for Economists 2021/22 Fall

Assignment 2 Summary Report

Target: Help company set to price their new apartments (small and mid-size **apartments** hosting **2-6 guests**) not on the market in **Berlin**

Data: Airbnb Berlin data, more than 10K

Target variable: **Airbnb prices** for small and mid-size apartments hosting 2-6 guests in Berlin (in Berlin, types of room mainly divide to Entire room/apt and Private room)

Predictors: Size, type, location, amenities and others

Model selection: OLS, LASSO, CART, GBM

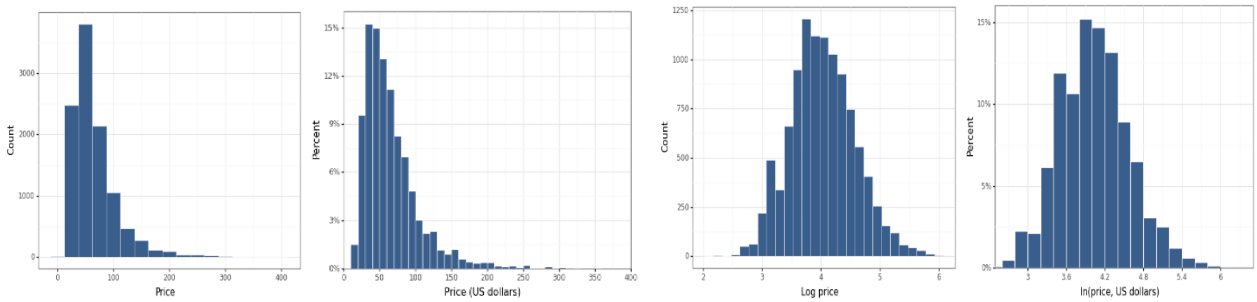


Figure1 Distribution of Airbnb Price in Berlin (before and after logarithm)

Most airbnb apartments hosting 2-6 guests price in Berlin locate in 30-40/40-50 Euros (around 15%), and 50-60/60-70 Euros also take up more than 10%. The price is almost normal distribution after logarithm.

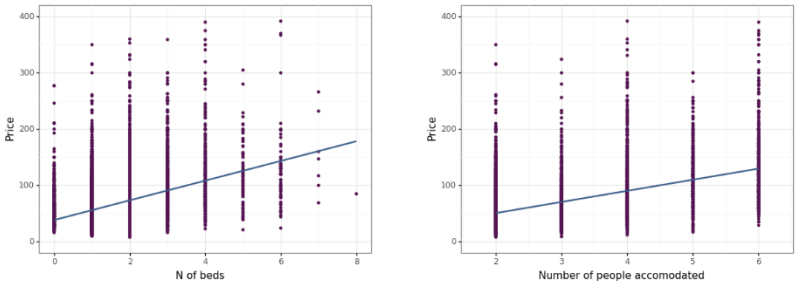


Table2 Relation between Beds/People Accomodated and Price

Beds/People Accomodated is significantly correlated with the price. Most of our samples have 1 or 2 beds and average price of 1 bed is 53 Euros and 73 Euros for 2 beds. The average spending on 2 people accomodated is 52 Euros, and 65 for 3.

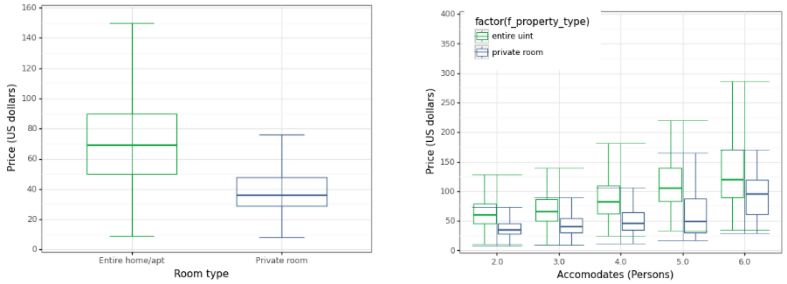


Figure3 Relation between Beds/People Accomodated and Price

Room type is used for further analysis with Box-plot. The price of Entire home/apt is much higher (around 70 Euros) than Private room (Around 38) which makes sense. The pattern is also found by interacted with the people accomodated. The lower quarter of entire unit is almost the same with the upper quarter of the single room.

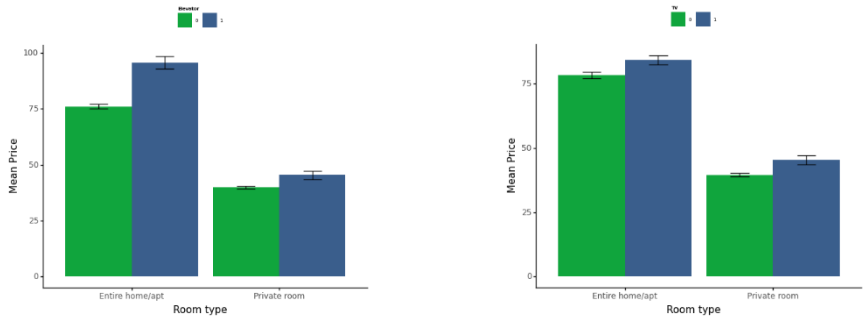
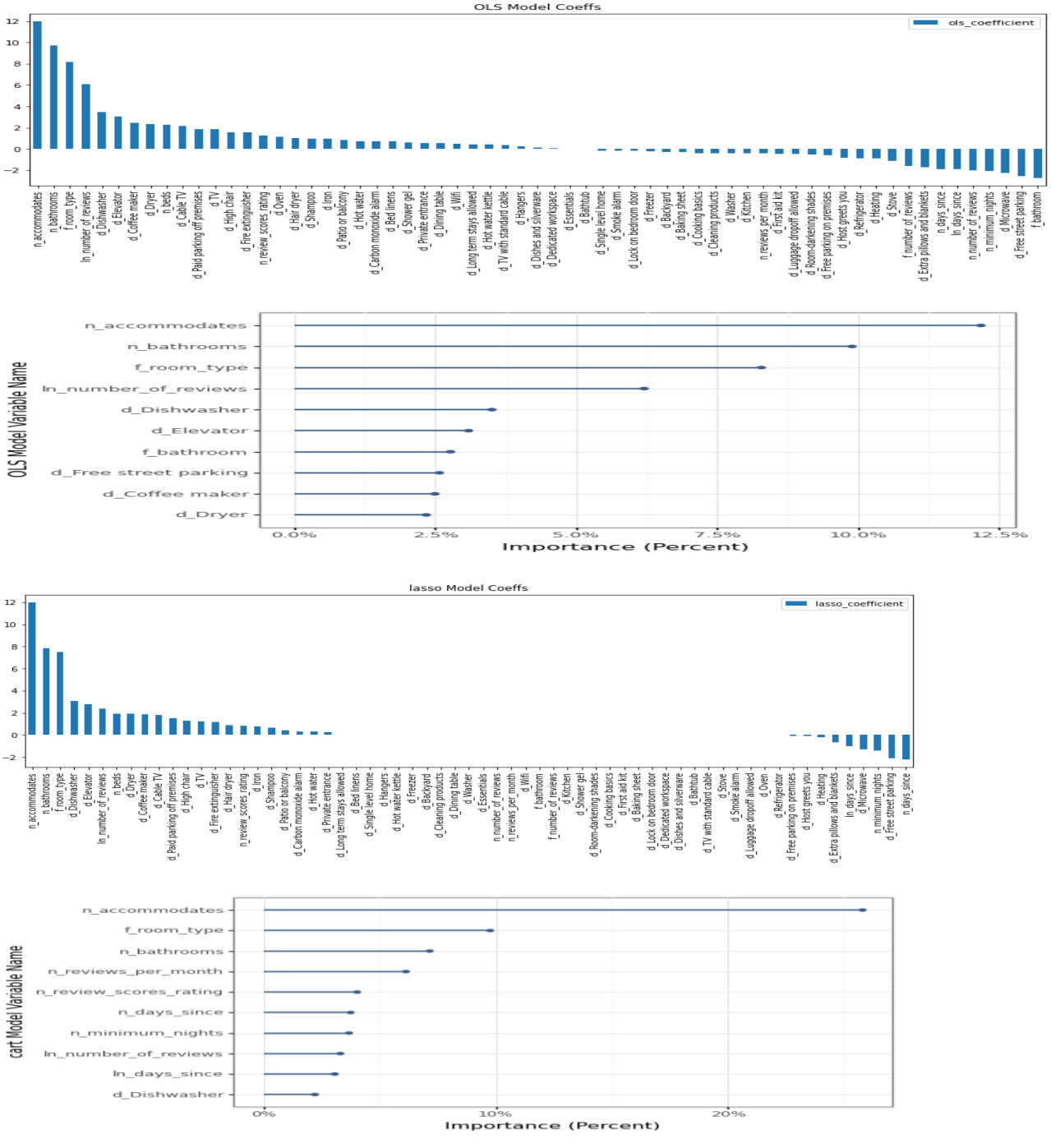


Figure4 Relation between Amenities and Price

The left figure shows the mean price by room type with/without elevator, and the price differs a lot within Entire/apt. The right figure is divided by with/without TV, which suggests no huge gap within the room type.



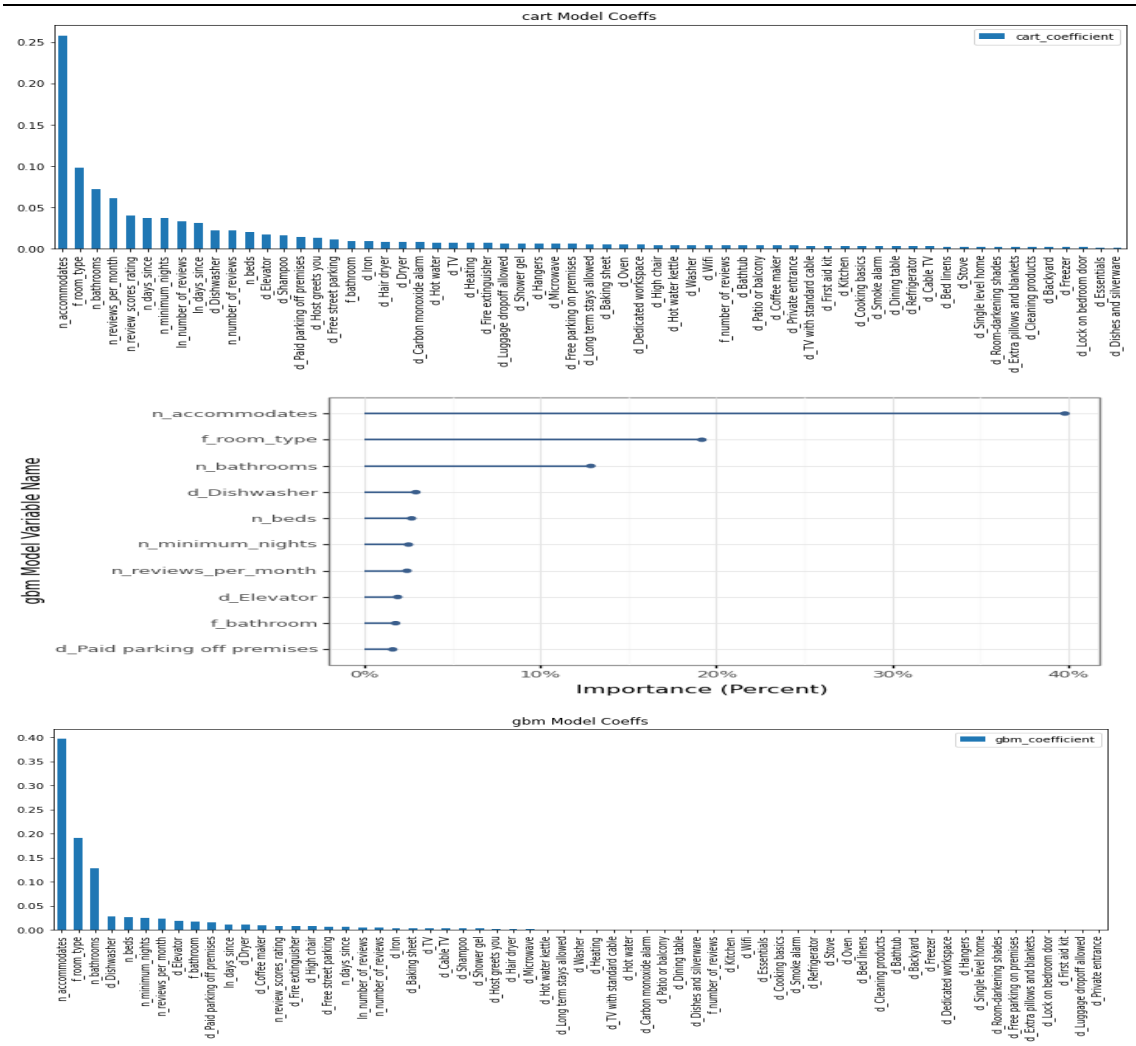
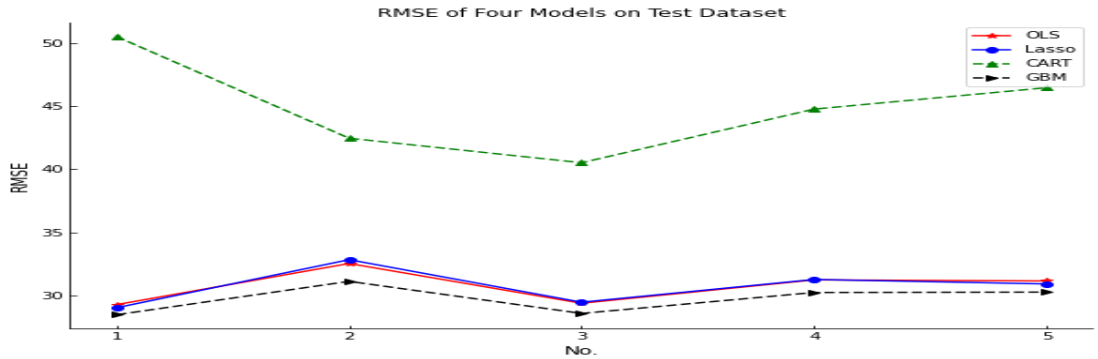


Figure5 Coefficients and Importance of OLS, LASSO, CART and GBM Predication

For OLS regression, poeple accommodate is the most imortant predictor, one extral person will increase the price by 12 Euros. The importance of bathroom, room type, number of reviews and dishwasher follow the accommodates. The LASSO regression result is almost the same but number of reviews is shrunk. Both predication with CART and GMB confirm the importance of people accommadates, room type and number of bathroom, but the reviews per month/ score rank 4/5 in CART and dishwasher/number of beds rank 4/5 in GBM.



	OLS	LASSO	CART	GBM
Mean CV RMSE	30.72	30.71	44.94	29.74

Figure5 RMSE for four Predication Models

The predication performance of OLS and LASSO are almost the same, CART plays the worst result, and GBM has the lowest RMSE. Therefore, GBM has the best predication result among the four models.