Predicting Housing Category

Jhih-Rou Huang

Goal

Build a classification model to predict the category of the houses

- Predicting the category of the house in the real estate company is basic or luxury
- Customers can find their ideal houses easily and efficiency based on their needs
- The sales of houses can increase

Tools

Pandas

Numpy

Matplotlib

Seaborn

Sklearn

xgboost

Data

ParisHousingClass from Kaggle
10,000 houses with 18 features

- squareMeters
- numberOfRooms
- floors - number of floors

Size

- cityPartRange- the higher the range, the more exclusive the neighborhood is
- numPrevOwners- number of previous owners
- Made- year

Age

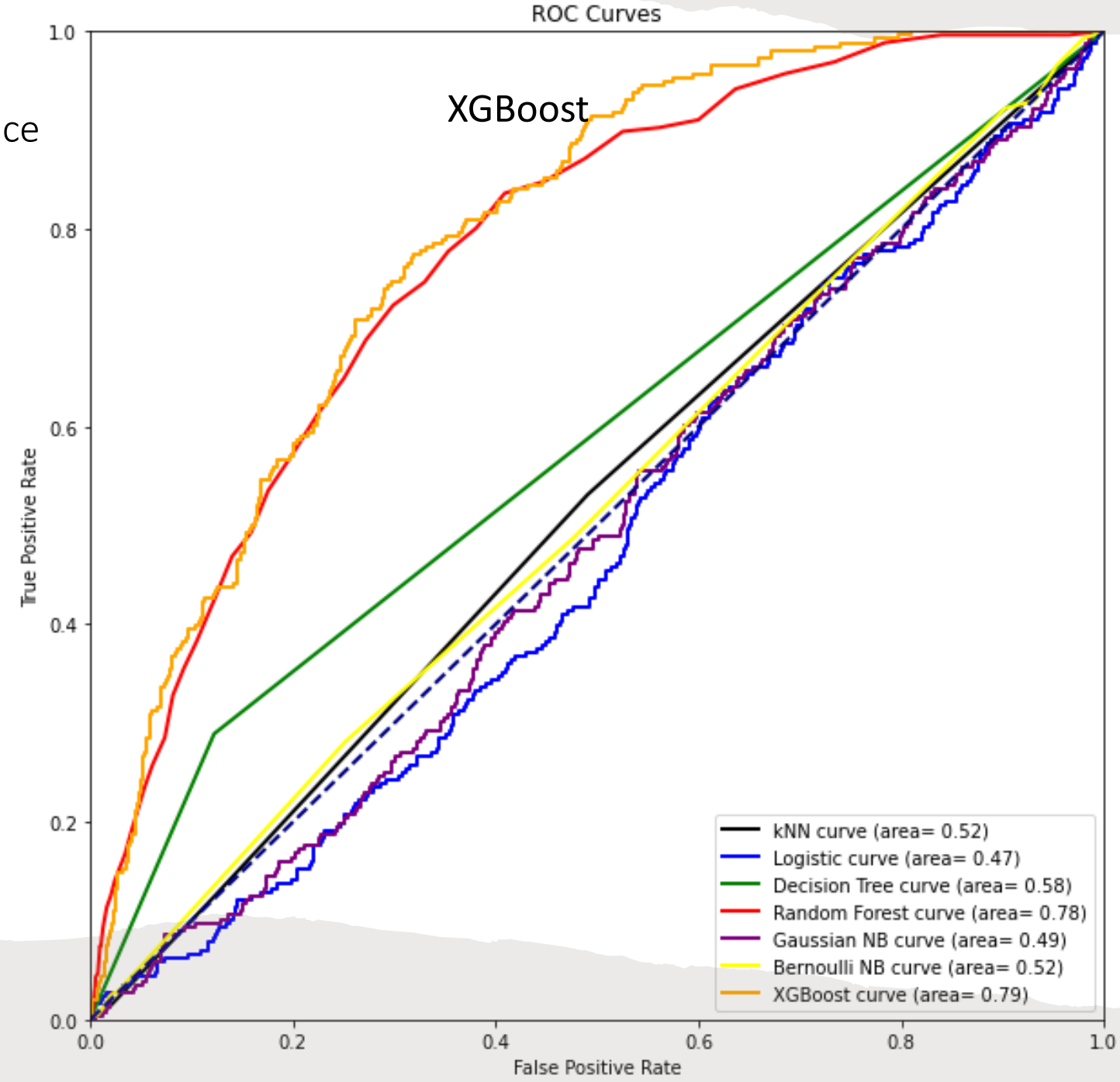
- hasStormProtector
- Basement- basement square meters
- Attic- attic square meters
- Garage- garage size
- hasStorageRoom
- hasGuestRoom- number of guest rooms
- Price- the price of a house

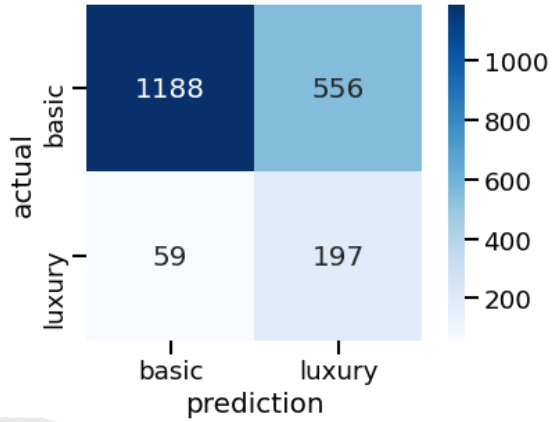
Price

Target
Category: Luxury or Basic

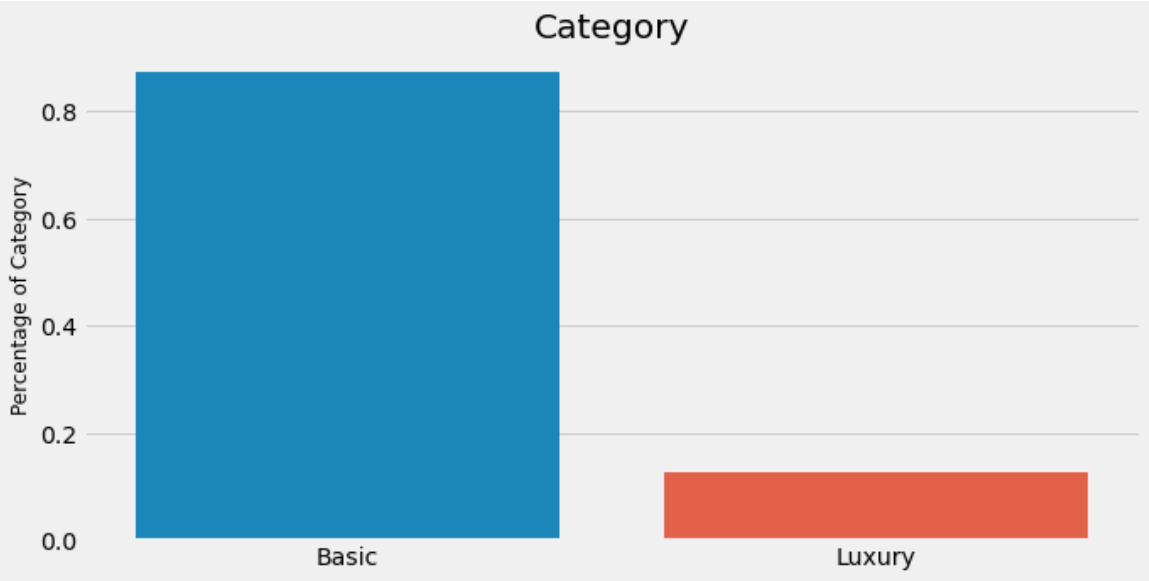
Price per square meters=price/ square meters
Age =2021-made

Comparison of Model Performance

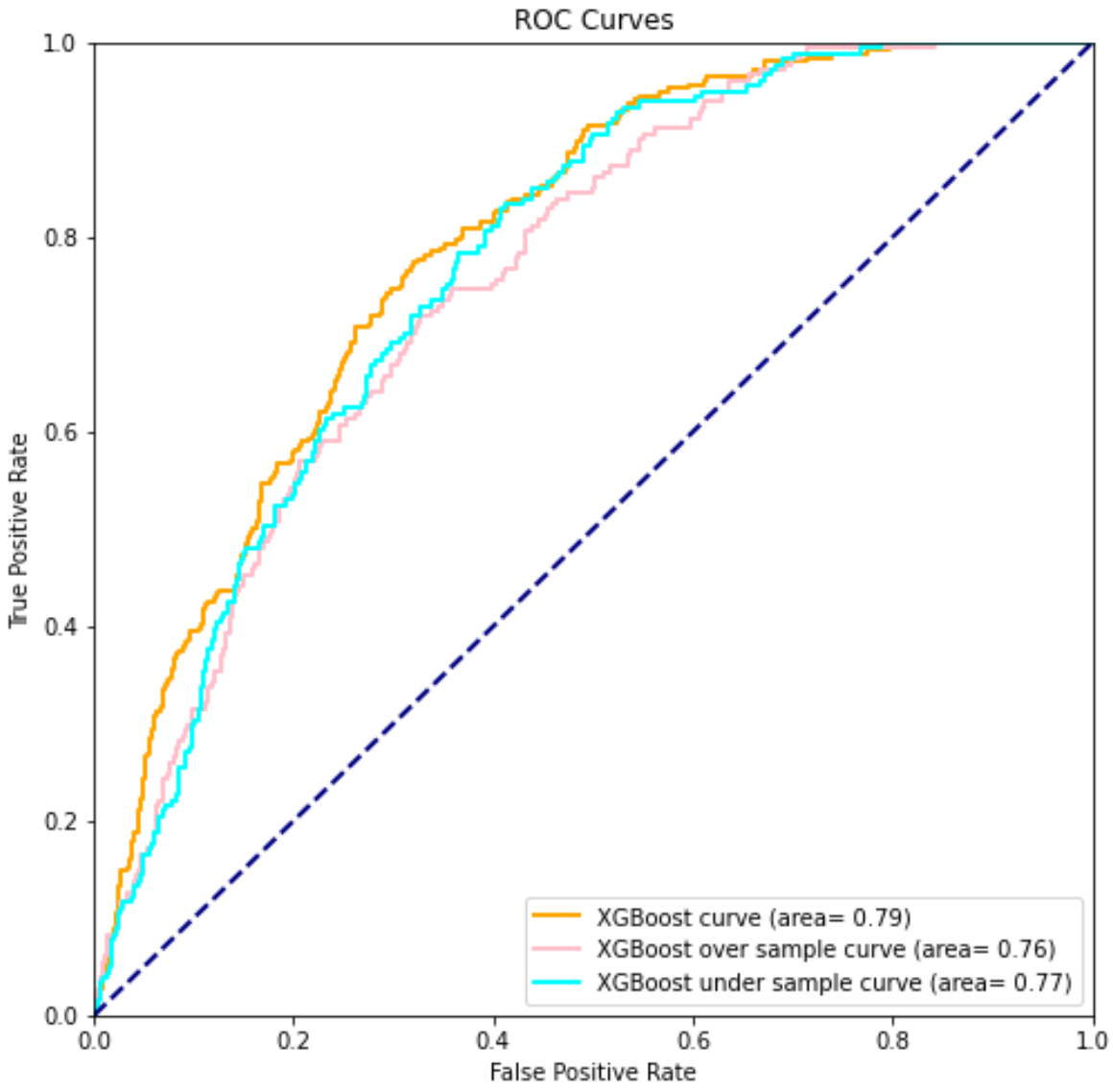




Resample



Model	accuracy	precision	recall	f1	AUC
KNN	0.858	0.088	0.012	0.021	0.518
Logistic	0.872	0.000	0.000	0.000	0.473
Decision Tree	0.803	0.258	0.289	0.273	0.583
Random Forest	0.872	0.500	0.004	0.008	0.778
Gaussian	0.872	0.000	0.000	0.000	0.487
Bernoulli	0.872	0.000	0.000	0.000	0.516
XGBoost	**0.863**	**0.402**	**0.152**	**0.221**	**0.792**
XGBoost over sample	**0.839**	**0.289**	**0.287**	**0.288**	**0.758**
XGBoost under sample	**0.684**	**0.223**	**0.718**	**0.340**	**0.771**

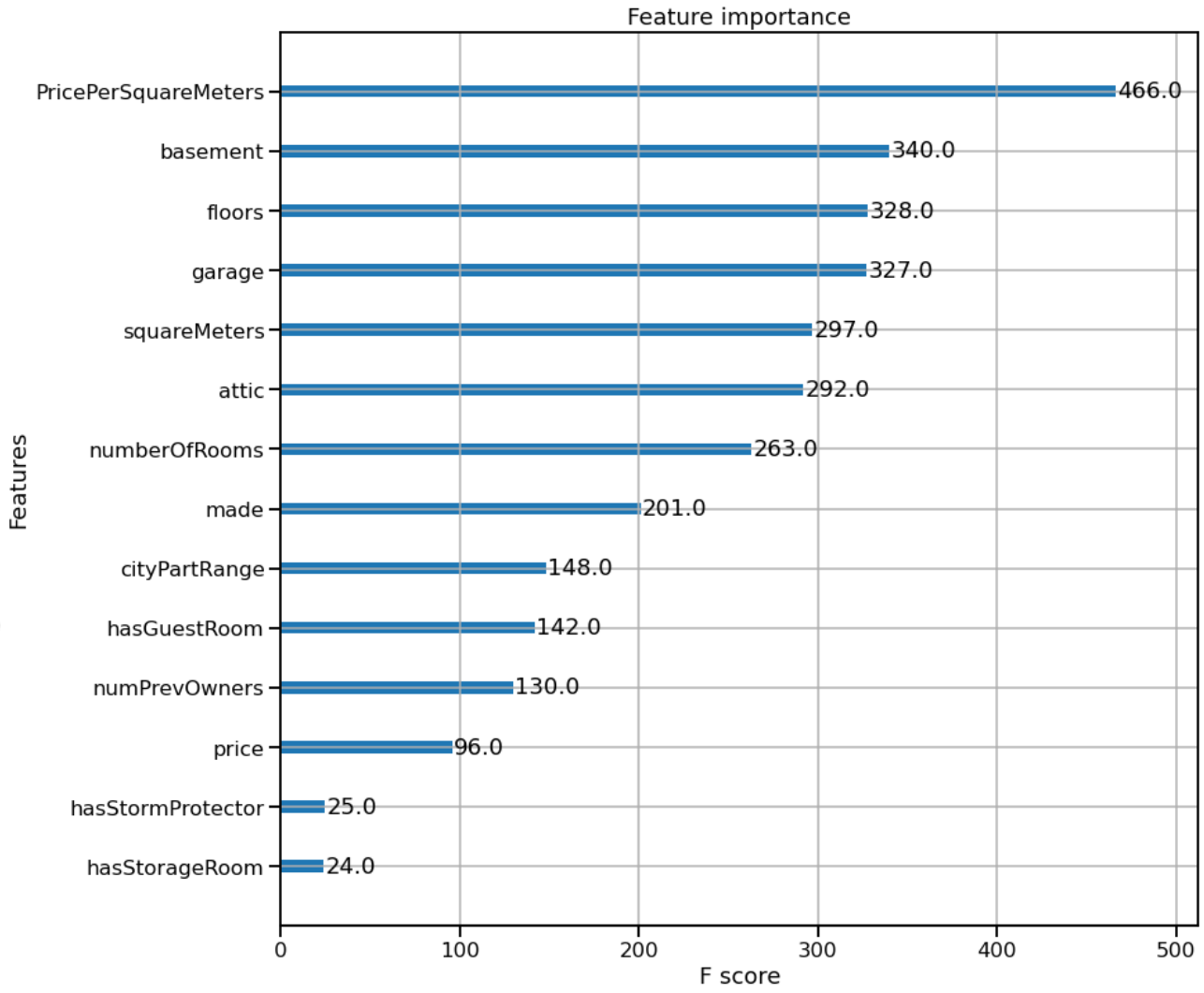


Insights and Recommendations

- Price per square meters is the most important feature
- The size of the house also plays an important role

Future Improvements

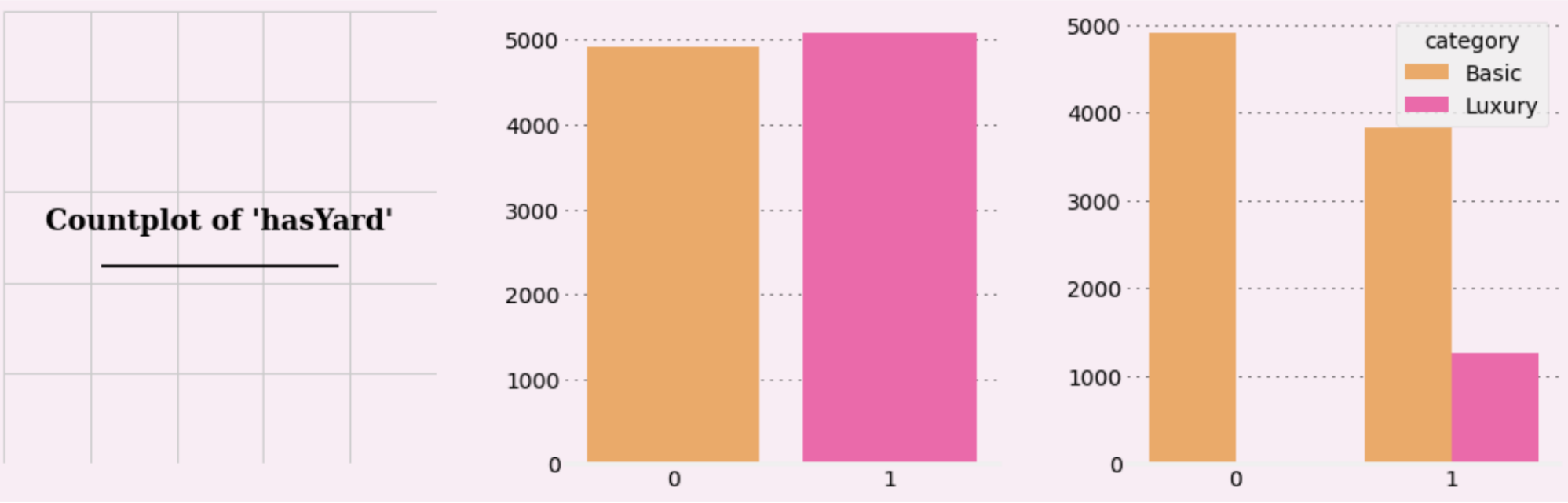
Build more robust model (more data and features)

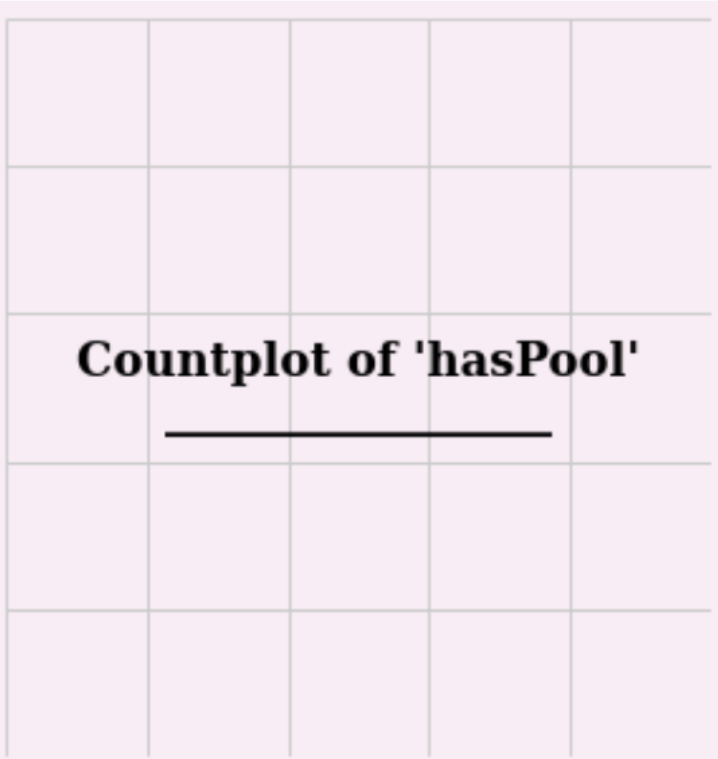
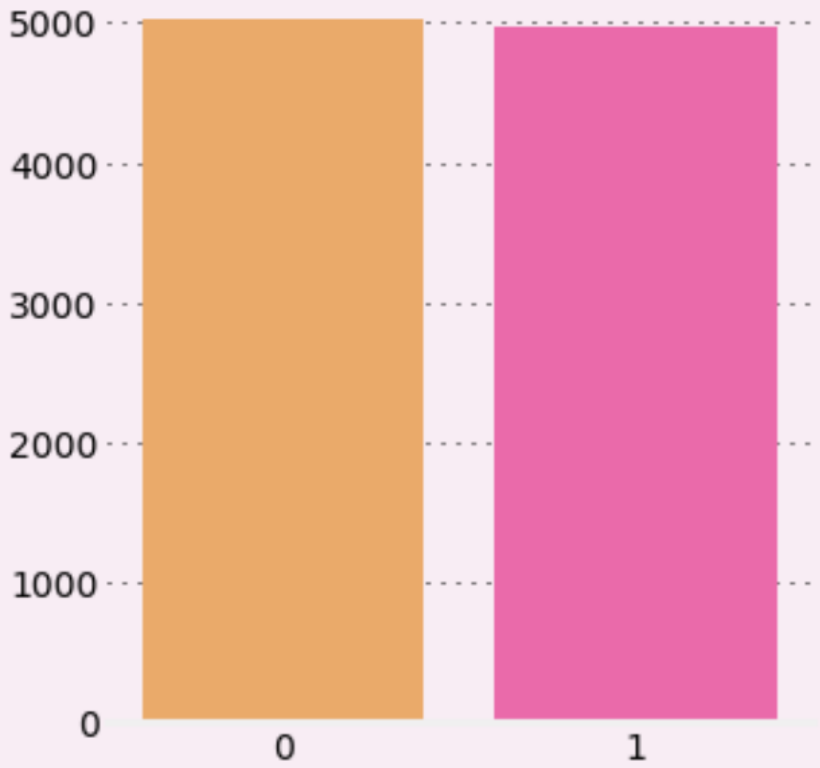
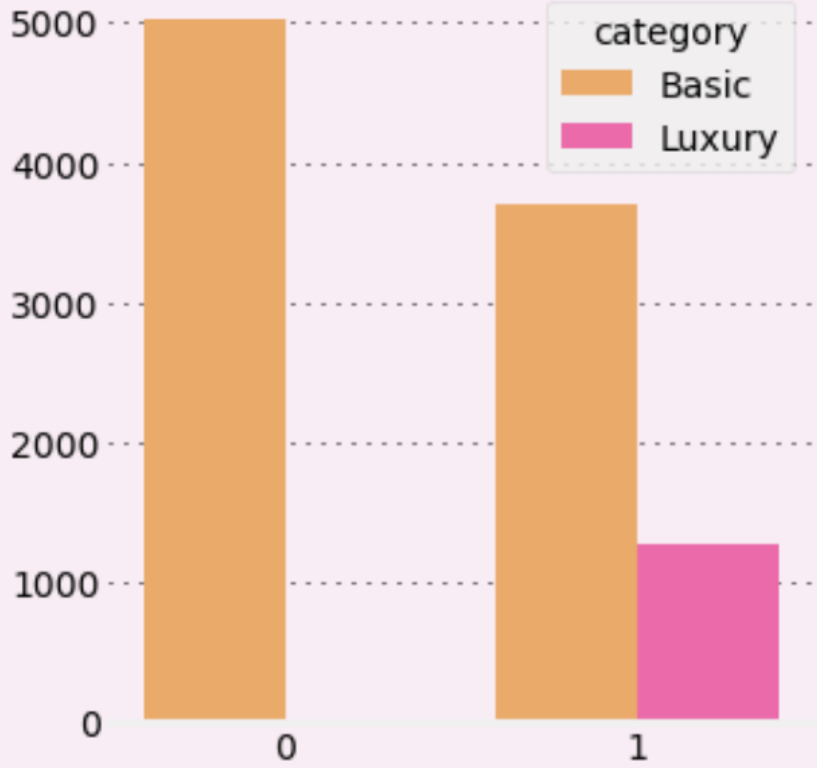


Appendix

‘has Yard’

hasYard	0	1	All
category			
Basic	4913	3822	8735
Luxury	0	1265	1265
All	4913	5087	10000



‘has Pool’

hasPool	0	1	All
category			
Basic	5032	3703	8735
Luxury	0	1265	1265
All	5032	4968	10000