

Team 1

The Mona Mer-lisa

Chenxi Wu, George Lindner, Qianyin Lu, Yi Mi
12/15/2019



Best Model

Before model selection, we did...

- Factorize categorical predictors
- Regroup based on price and intuition
- Remove variables with imbalance problem
- Impute missing data
- Explore potential transformation
- Use boxplot to filter out variables with obvious effect

Random Forest

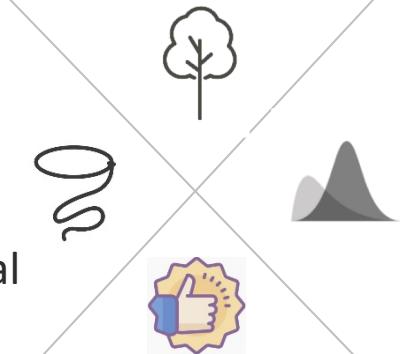
- Lack of prediction interval
- Lower coverage
- Hard to interpret

Lasso

- Lack of prediction interval
- Meaningless to normalize categorical predictors

Bayesian Model Averaging

- Higher RMSE



Final Model - Linear

- Given interval
- Easy to interpret
- High coverage and average RMSE

$$\begin{aligned} \text{logprice} = & \beta_0 + \beta_1 \text{year} + \beta_2 \text{dealer} + \beta_3 \text{expensive} + \beta_4 \text{authorstyle} + \beta_5 \text{endbuyer} \\ & + \beta_6 \text{Interm} + \beta_7 \text{Surface} + \beta_8 \text{finished} + \beta_9 \text{Irgfont} + \beta_{10} \text{diff-origin} + \epsilon \end{aligned}$$

Best Insight into Predicting Price

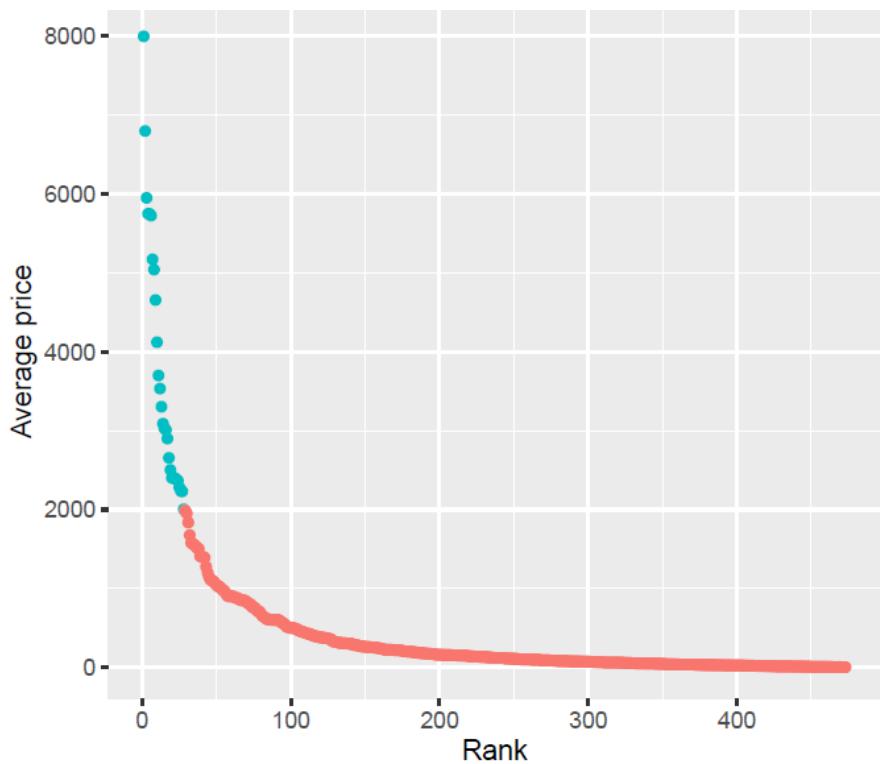


Table 4: Coefficient Summary for Final Model

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.000	12.576	-15.831	0.000	0.000	0.000
year	1.120	0.007	16.041	0.000	1.105	1.136
dealerLP	2.568	0.112	8.397	0.000	2.060	3.200
dealerR	5.356	0.103	16.260	0.000	4.374	6.558
expensive1	3.605	0.106	12.046	0.000	2.926	4.442
authorstyle1	0.388	0.148	-6.414	0.000	0.291	0.518
endbuyerD	0.788	0.101	-2.360	0.018	0.646	0.961
endbuyerEU	0.494	0.103	-6.874	0.000	0.404	0.604
Interm1	1.996	0.133	5.188	0.000	1.537	2.591
Surface	1.367	0.026	12.173	0.000	1.300	1.437
finished1	2.351	0.090	9.466	0.000	1.969	2.806
lrgfont1	2.436	0.120	7.442	0.000	1.926	3.080
diff_origin1	0.613	0.085	-5.788	0.000	0.519	0.724

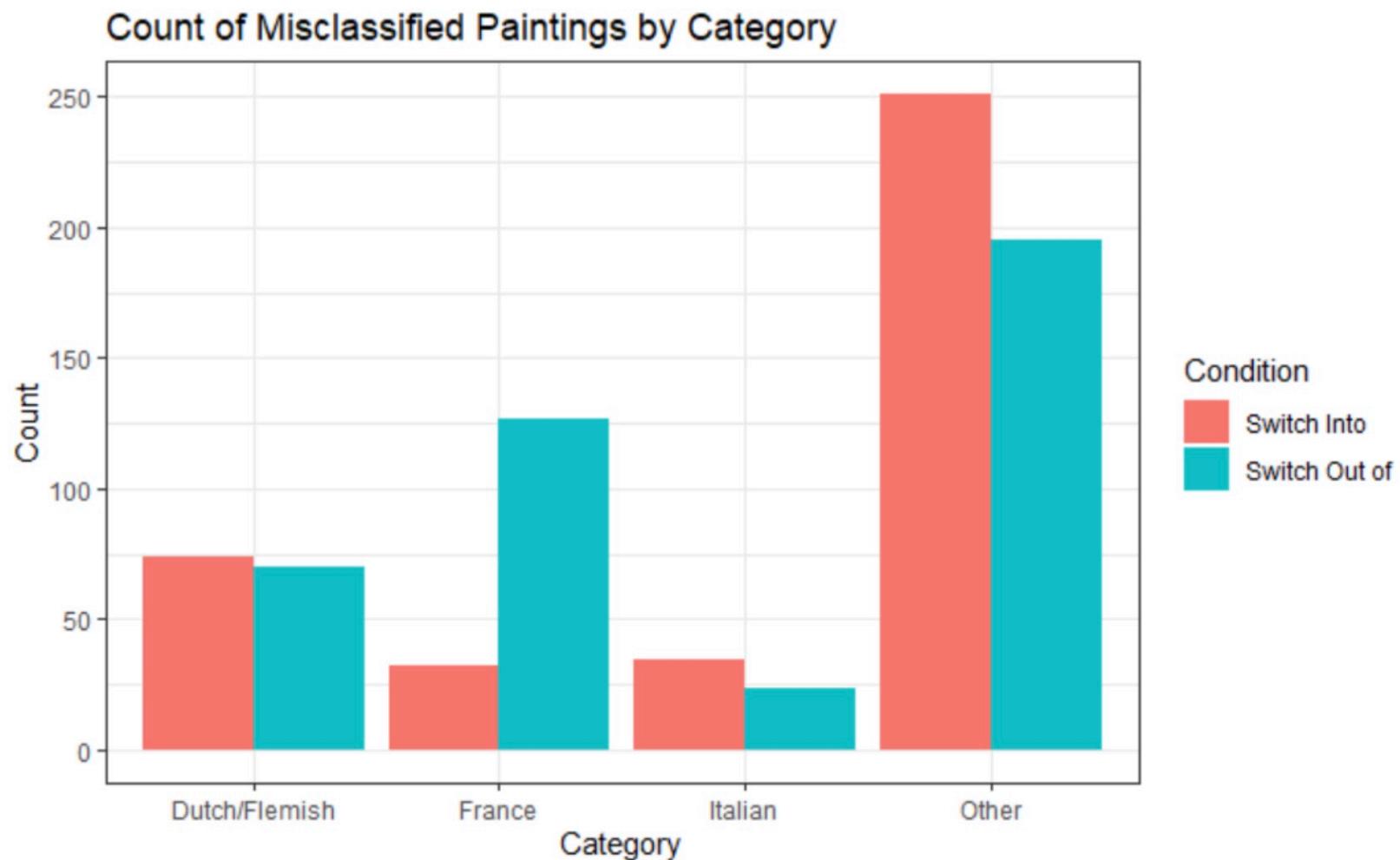
Note:

The coefficients have been exponentiated.

Most Interesting Graphic

Country	Ratio
French	3.969
Dutch/Flemish	0.946
Other	0.777
Italian	0.686

Type	Price
Other	842.18
French	514.99



3 Best Paintings to Purchase

Comparing predicted price from our best model with the actual auction price from paintings_train set, we defined undervalued paintings by:
 $\text{predicted price} - \text{auction price}$



Adoration des Rois
by Gerard de Lairesse



The Young Beggar
by Bartolomé Esteban Murillo



Le martyre de Saint Laurent
by Eustache Le Sueur