

FINAL DATA ANALYSIS PROJECT

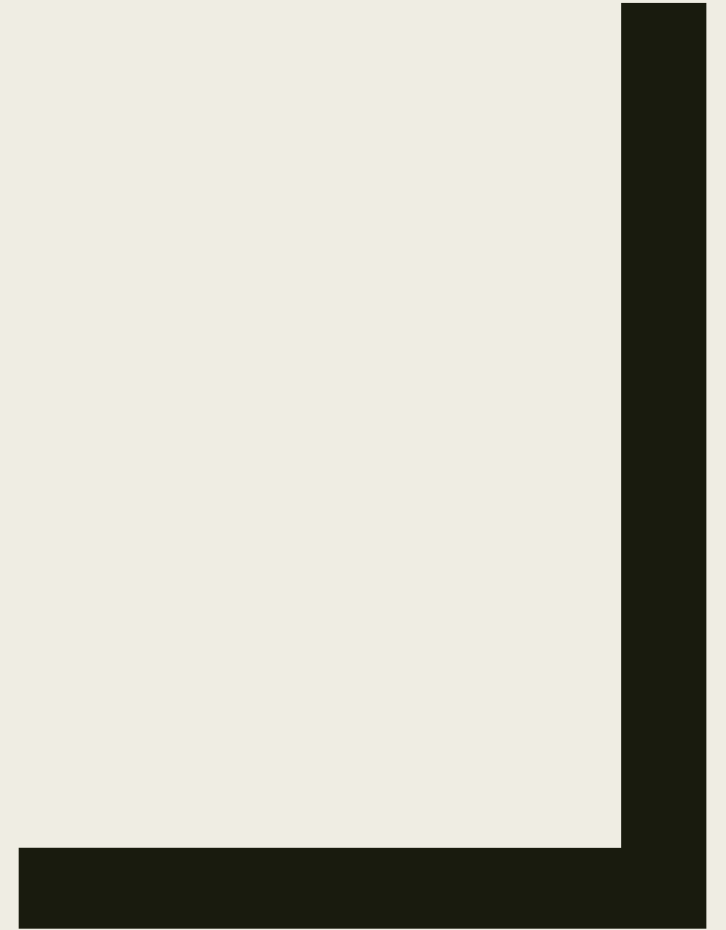
Team 3

Mingxuan Yang

Jiawei Chen

Machao Deng

Jishen Yin



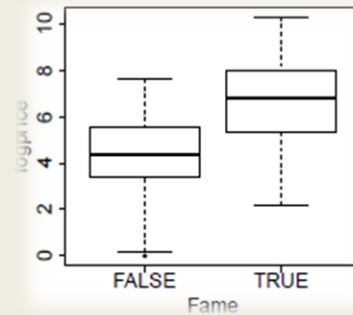
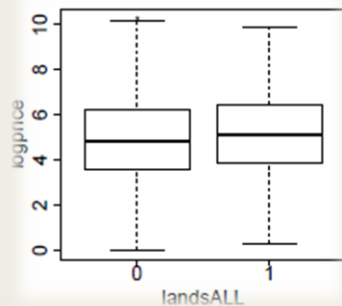
• EDA

• Variable Investigation

- *General*
- *Variable to impute*
 - Impute with existed value | impute with `n/a`
- *Variable to drop*
 - Def | Too hard to handle | Correlated
- *Variable to manipulate*
 - Correction | Creation

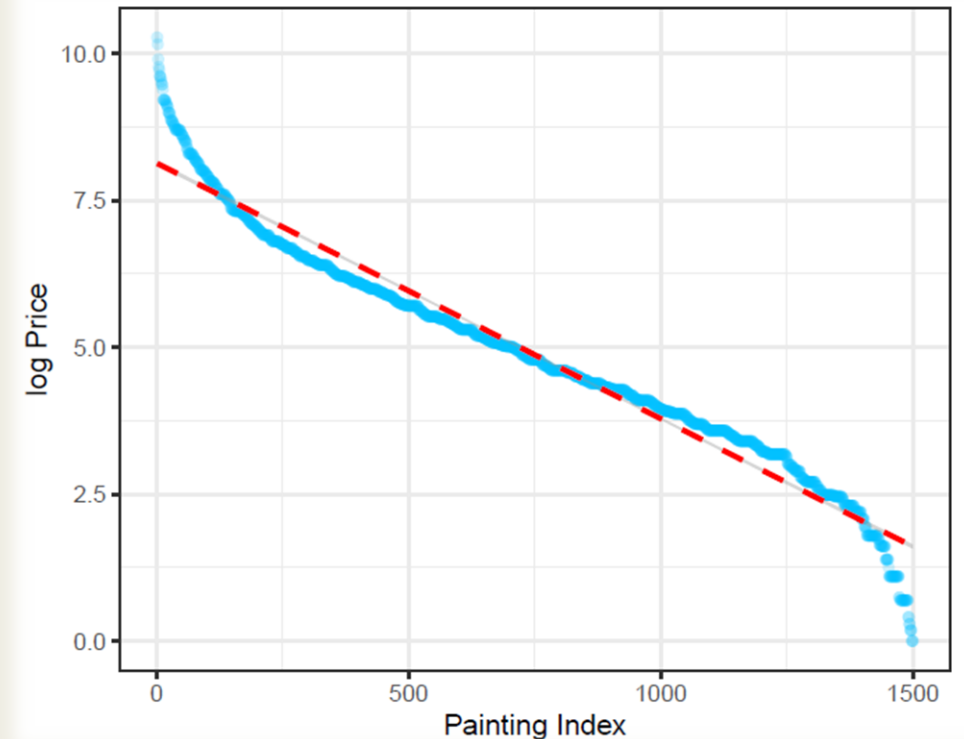
• Important predictor identification

dealer, year, origin_author, finished,
winningbiddertype, endbuyer,
type_intermed, lrgfont, Fame



Freq		type_intermed				
		B	D	E	EB	n/a
Interm	0	0	0	0	0	960
	1	11	94	39	1	0

"squ_rect" "oval" "octagon" "round" "miniatur"



Linear Model Variable

dealer

year

origin_author

endbuyer

log_Surface

finished

lrgfont

winningbiddertype

year:winningbiddertype

Linear Model

Bias 206.171

Coverage 0.952

maxDeviation 13558.248

MeanAbsDeviation 541.419

RMSE 1487.499

PRELIMINARY
MODEL

Random Forest Variable

year

log_Surface

dealer

lrgfont

endbuyer

origin_author

winningbiddertype

finished

type_intermed

diff_origin

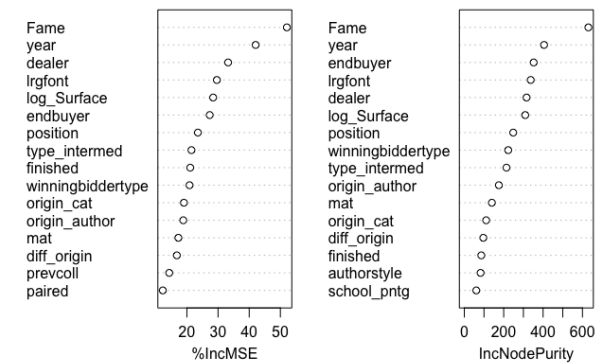
precvoll

paired

Fame

mat

	Linear Model	Random Forest
Bias	206.171	153.389
Coverage	0.952	0.935
maxDeviation	13558.248	8422.370
MeanAbsDeviation	541.419	313.295
RMSE	1487.499	886.098



FINAL MODEL

Table 6: Top 10 valued paintings

author	year	dealer	surface	endbuyer
Federico Barocci	1768	J	832.00	
Robert Tournières	1768	J	1394.00	
Flemish	1768	J	1386.00	
Johann Wilhelm Baur	1768	J	945.00	U
Flemish	1768	J	30.25	C
Guido Reni	1768	J	180.00	U
Johann Wilhelm Baur	1768	J	154.00	E
Johann Wilhelm Baur	1764	L	192.00	
Anonymous	1768	J	48.00	D
Antoine Dieu	1768	J	48.00	D

Table 7: Variable Importance Plot

	%IncMSE
Fame	94.331
dealer	83.071
year	80.737
log_Surface	44.684
endbuyer	40.299

CONCLUSION