PREDICTING PROPERTY PRICES (PPP)

Prepared for:
Senior Management

18 Jul 2022

Jacob, Junhong, Wende

01 **INTRODUCTION**

Background &
Problem Statement
by Jacob

02 **THE DATA**

Key Findings
by Junhong

03 **THE PREDICTION MODEL**

Model Selection &
Evaluation
by Wende

04 **THE RECOMMENDATIONS**

Recommendation &
Future Works
by Jacob

01.

INTRODUCTION

Background & Problem Statement
by Jacob

Background

Penetrating the Ames, Iowa property market

1. Buyers
2. Sellers
3. Flippers

1. Support Buyer's Budgeting

1. Help align expectations
2. Better understand their affordability
3. Maximising value of their purchase

2. Maximising Seller's Profit

IMPORTANCE OF PRICING

More buyers will compete for your property if you price it properly

3. Identifying Valuable Features

1. Investors looking to flip properties
2. Remodel key part of the properties
3. Save time and money

02.

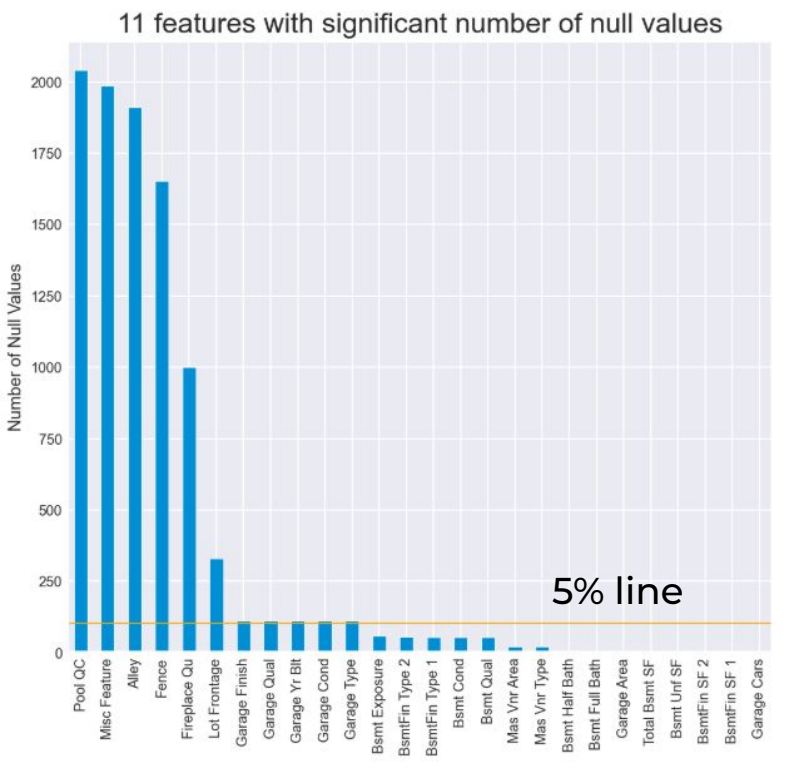
THE DATA

Key Findings
by Junhong

Feature Data

1. 80 features
2. There are no duplicated data
3. 26 columns (features) with null values.
4. `Id` & `PID` columns values are unique.

Null values



Features with many null values

1. Pool QC
2. Misc Feature
3. Alley
4. Fence
5. Fireplace Qu
6. Lot Frontage
7. Garage Finish
8. Garage Qual
9. Garage Yr Blt
10. Garage Cond
11. Garage Type

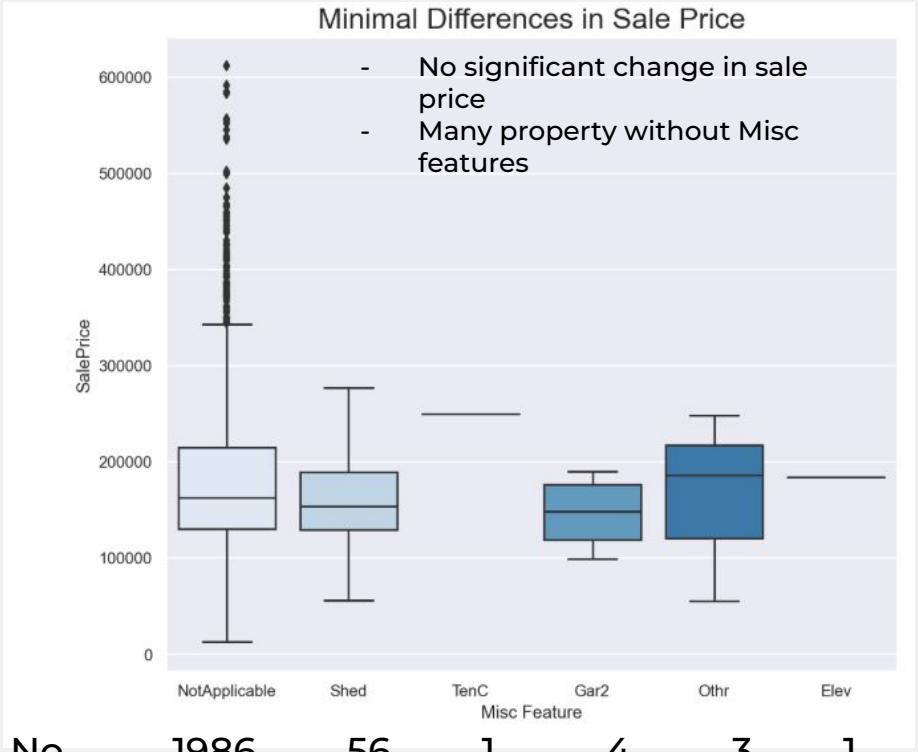
Pool features

	Pool QC	Pool Area	SalePrice
52	Fa	519	250000
657	Gd	576	171000
761	Gd	800	260000
952	Ex	228	315000
960	Gd	480	160000
1130	Fa	648	181000
1249	Gd	738	274970
1635	TA	368	153000
1875	TA	561	130000

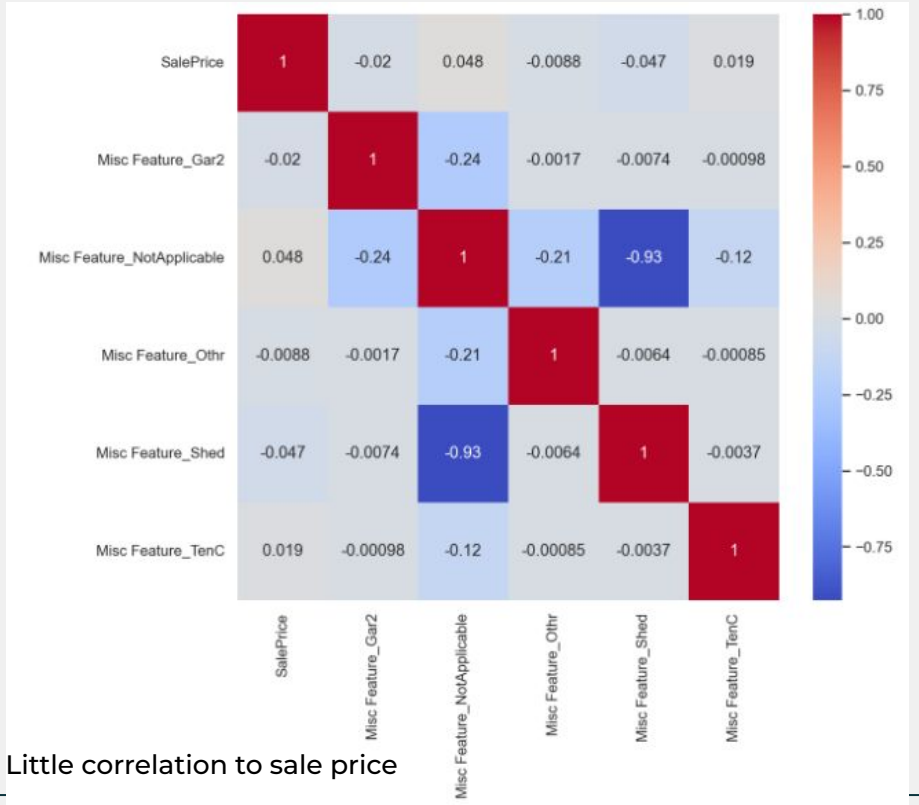
9 Properties have a pool

2042 Properties does not have a pool

Misc features

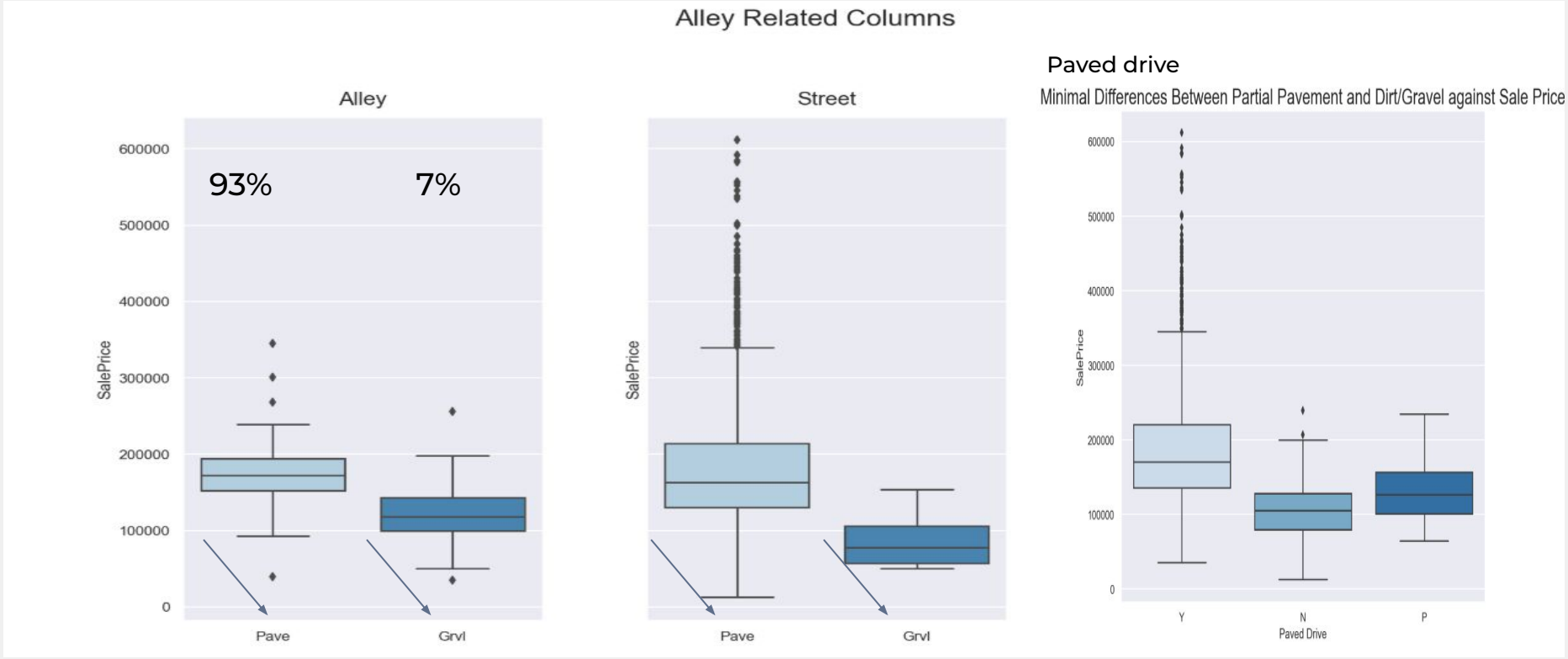


No.	1986	56	1	4	3	1

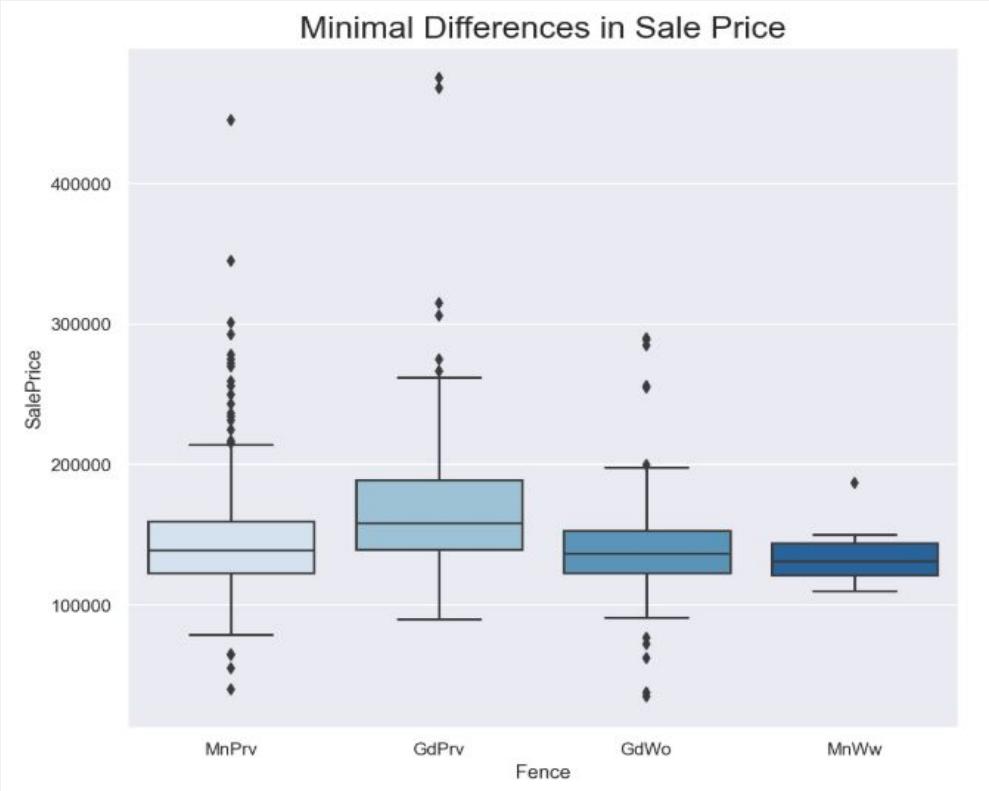


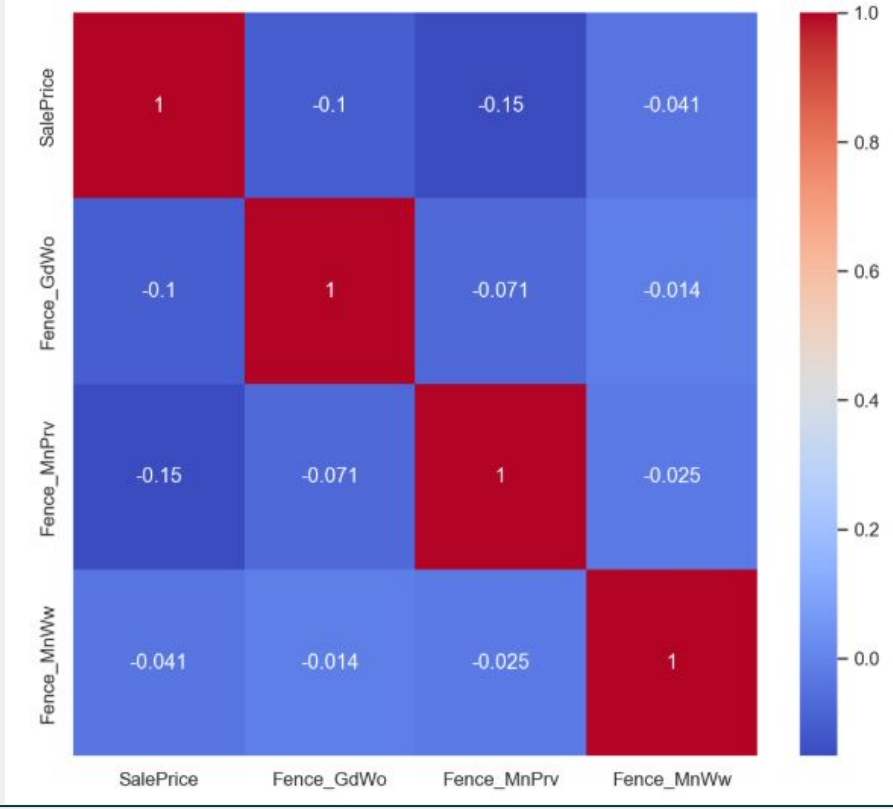
Little correlation to sale price

Alley Features



Fence Feature





Features Summary

1. Misc Feature related and fence are dropped
   - null values
   - low correlation to sale price
2. Alley is dropped
   - little values
   - multi-collinearity.
3. Lot Frontage is dropped
   - multi-collinearity
4. Id and PID are dropped
   - unique, not useful in predicting SalePrice

03.

THE PREDICTION MODEL

Model Selection & Evaluation
by Wende

MODELLING PROCESS

START

Linear Regression Model (Baseline) — Determine time frame

Make Adjustments (Remove/Engineer Features)

Create, Run, Evaluate Models (Linear, Ridge, Lasso)

Iterative Process

Train with New Data

Compare Models — More Evaluation — The Best PPP Model

Overview of Model Development

1

Linear Model (Baseline)

75 Features
(Complex)

2

Regularisation 1: Ridge

75 Features
(Better accuracy but still complex)

3

Regularisation 2: Lasso

75 Features
(Better Accuracy but still Complex)

4

Ridge Model (Remove anomaly and columns)

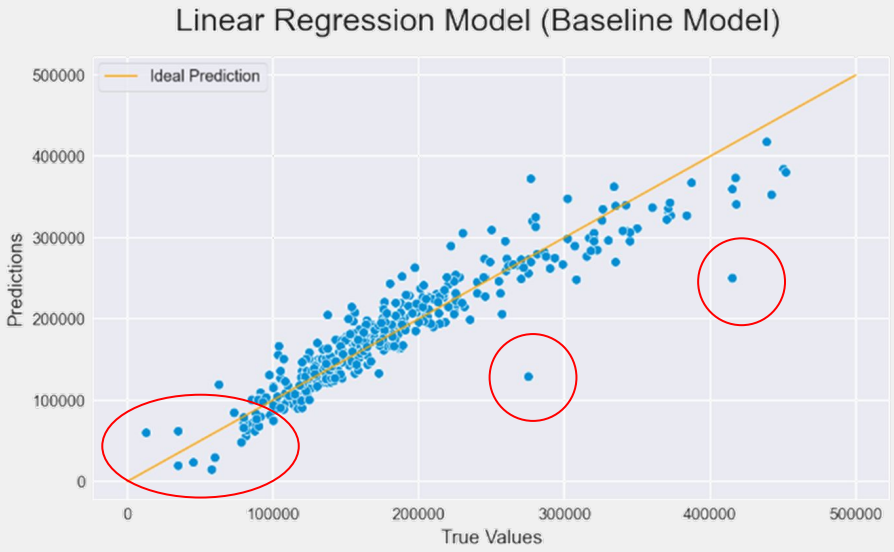
56 Features
(Better accuracy, not so complex)

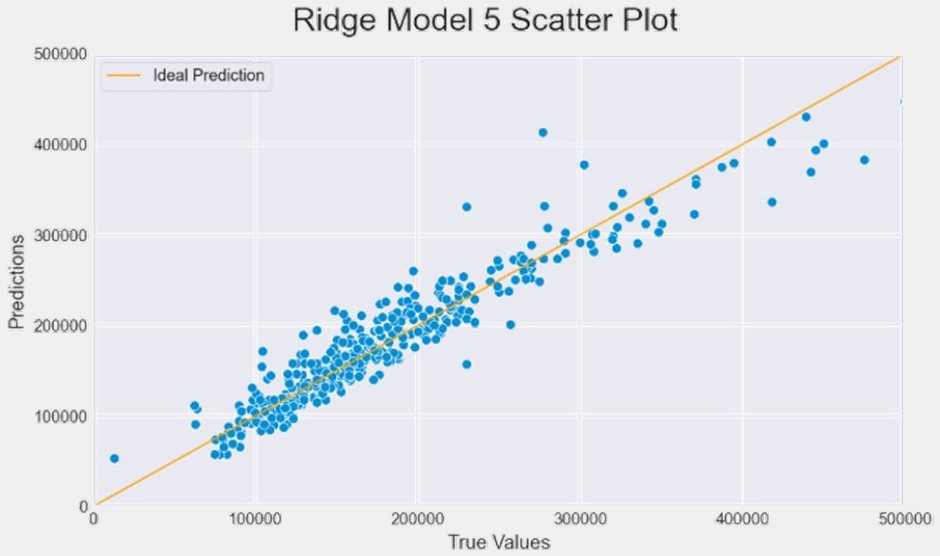
5

Ridge Model 5 (Remove columns, Engineer Features)

50 Features
(Best accuracy, not so complex)

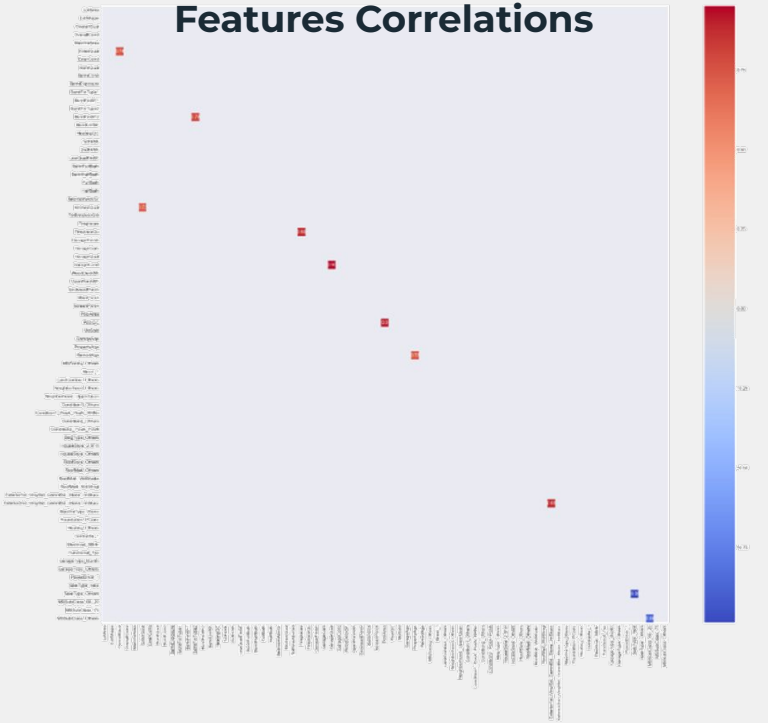
Best PPP Model Closer to Ideal Prediction



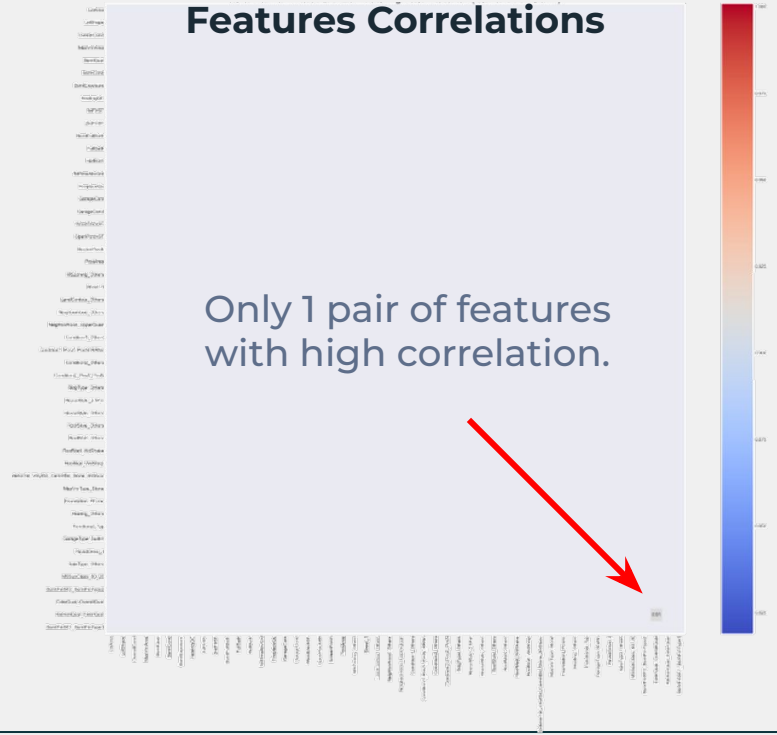


Significantly Reduced Multicollinearity

Baseline Model
Features Correlations



Ridge Model 5
Features Correlations



$33,517

RMSE of Baseline model

$24,840

RMSE of Best PPP model

26%

Reduction in RMSE[1]

[1] RMSE is interpreted as the average difference in Sale Price between the predicted values from the model and the actual values in the dataset.

90%

Of variability in Sale Price can be explained by the features in Ridge Model 5.

05.

THE RECOMMENDATIONS

Recommendation & Future Works
by Jacob

Propose to Adopt
Ridge Model 5

1. **Accurate**
   (Low RMSE, difference in the testing and training RMSE is about $800.)

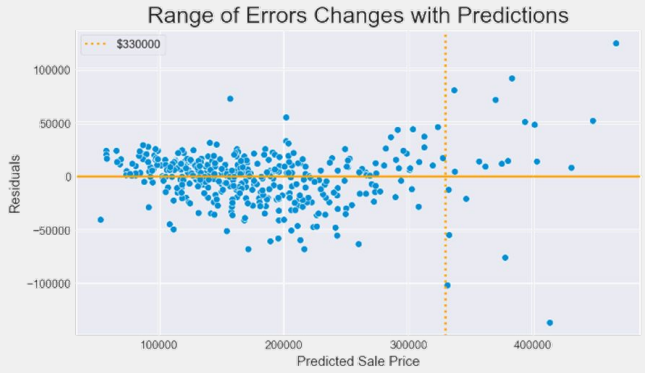
2. **Not complex**
   (10 features with significant coefficient despite having 50 features in the model.)

3. **High Confidence**
   (Explains up to 90% (R^2) of the variability in Sale Price.)

Top Five Features

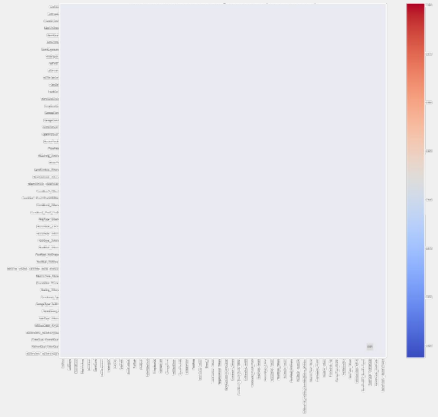
1. 1stFlrSF
2. ExterQual_OverallQual
3. 2ndFlrSF
4. BsmtFinSF1_BsmtFinType1
5. Neighborhood_UpperQuan

Future Work



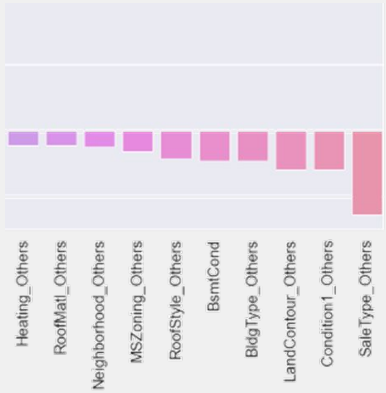
More Data

Build up database for sale price above $330k



Less Multi-colinearity

Address the pair of features with significant correlation



Review Nominal Sub-category

Improve resolution of the ‘others’ feature.

THANKS!

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik