

Predicting Housing Prices in Ames, Iowa

Group 1: Amira, Calvin, Joel, Joseph, Joshua, Nelson

DSI-28

13 May 2022

Agenda

1. Background
2. Problem Statement
3. Methodology
4. Key Findings
5. Conclusion & Recommendations
6. Limitations & Next Steps

Background

Ames, Iowa Demographics (2019)



66,258



\$234,100 (median)



\$50,528 per capita

Mean property prices



\$236,588



\$127,269

All housing units

2 unit structures



\$248,938

Detached houses



2. Problem Statement

We: Team of data-driven property agents in Ames, Iowa.

You: Home owners in Ames, Iowa.

Why us? Provide insights backed by data instead of mere gut feel.

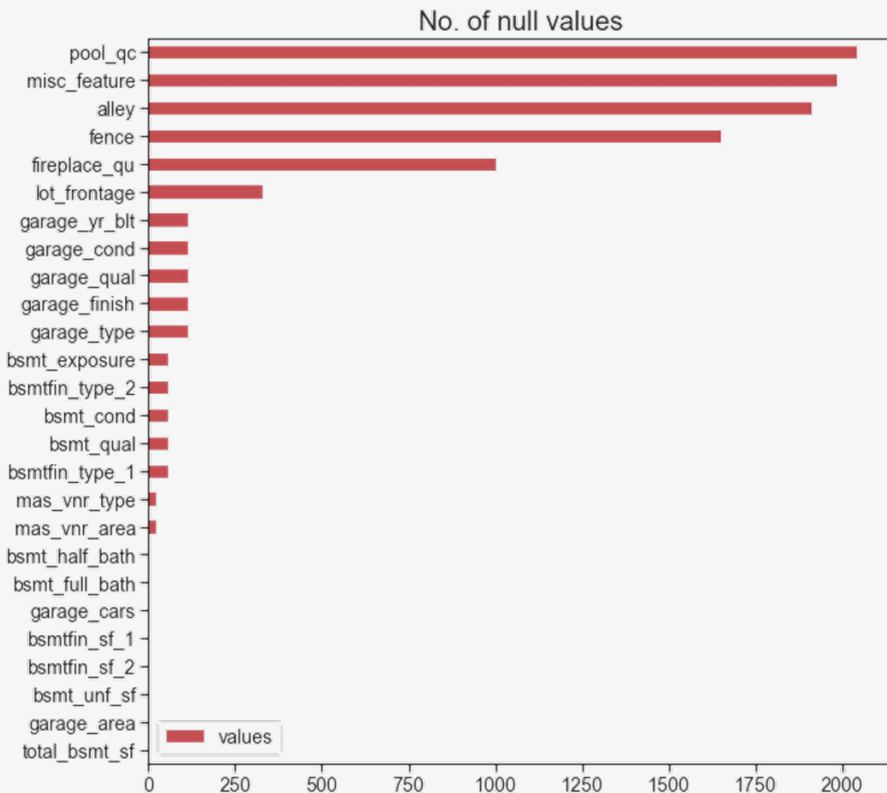
What's in it for homeowners:

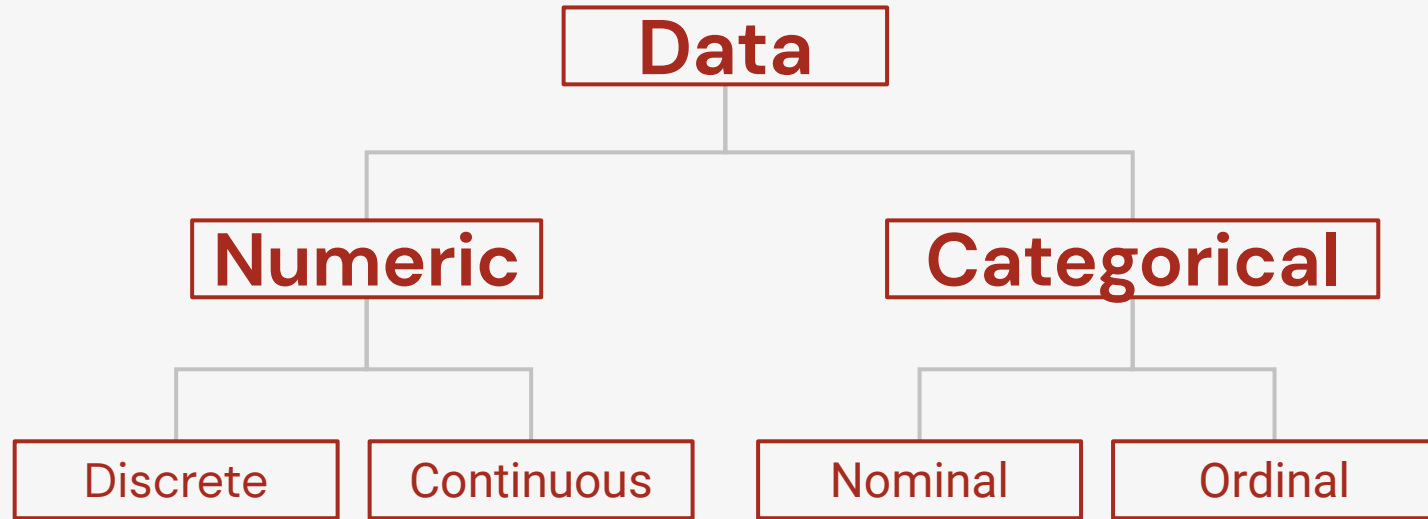
1. Optimal selling price
2. Favourable features of house &/or surroundings
3. Ways to enhance house's attractiveness to potential buyers

3. Methodology

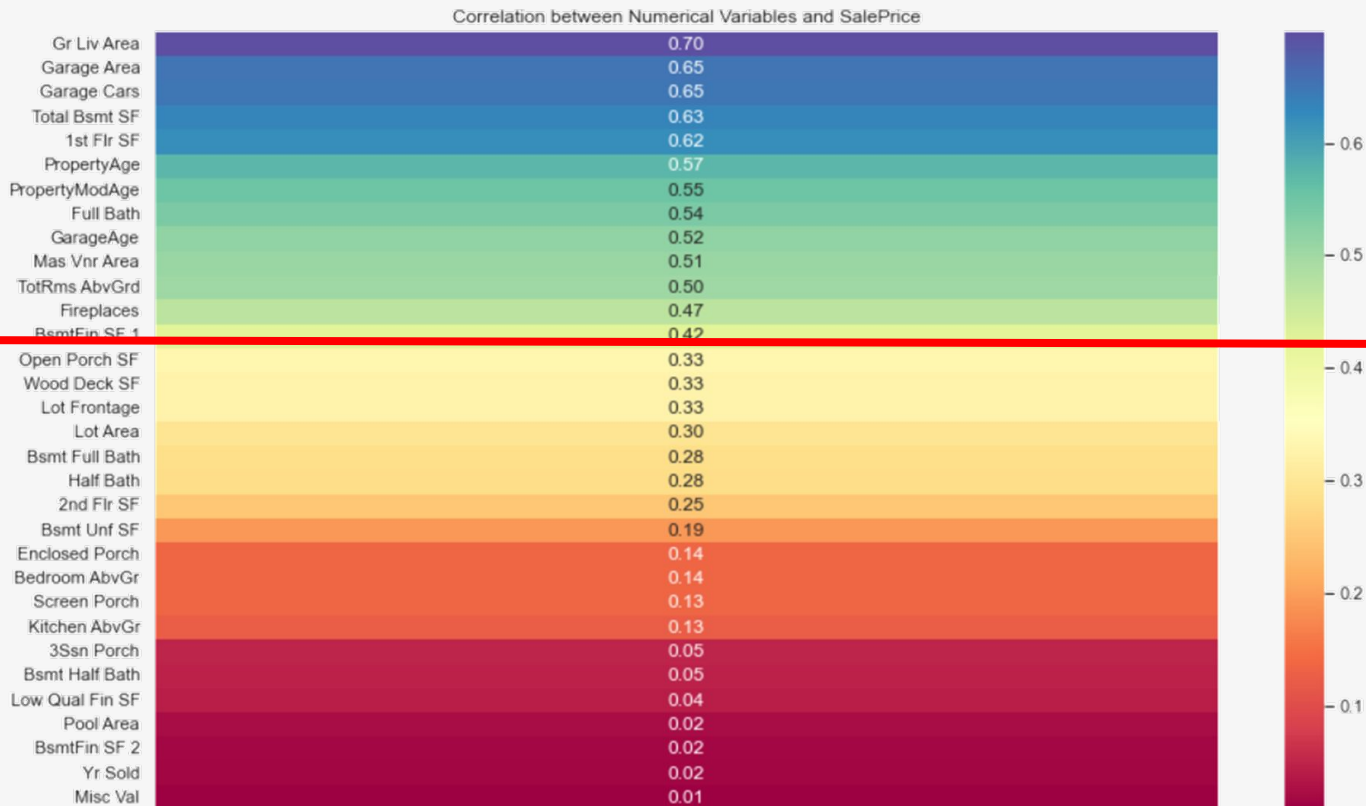
1. **Scope of data:** Housing prices between 2006 to 2010 in Ames, Iowa
2. **Exploratory Data Analysis (EDA) & Data cleaning**
3. **Data preprocessing & modelling**

A lot of our data is missing.





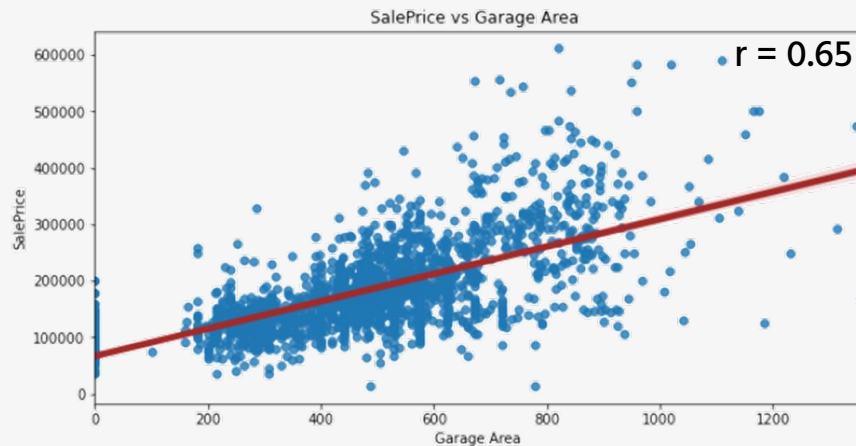
Excluded numeric data with Weak relationships* to SalePrice



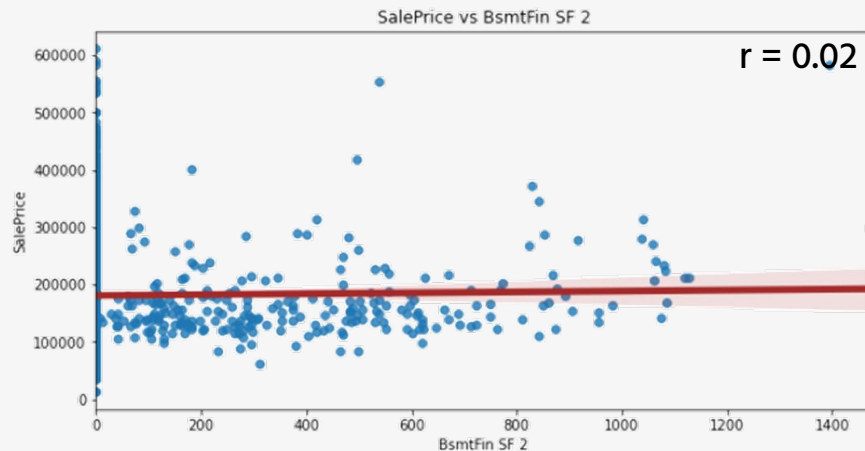
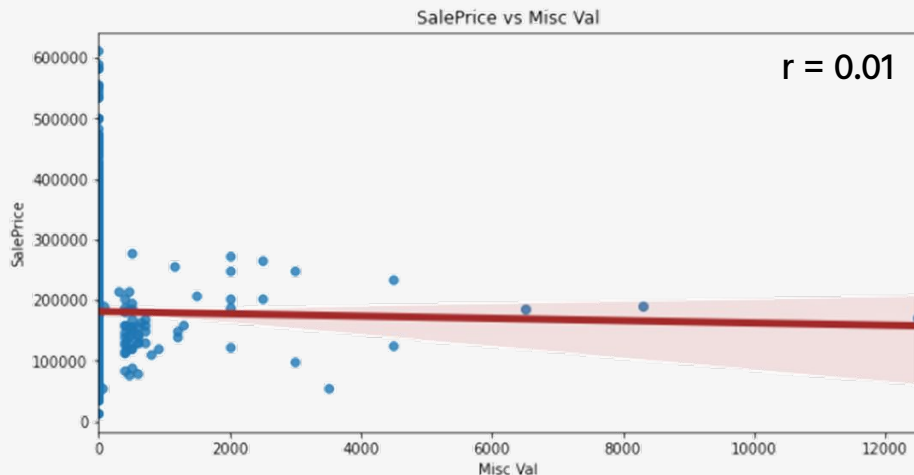
*Weak relationships are defined as those below the threshold of absolute 0.2 to 0.6 correlation coefficient.

As ground liv area ↑
SalePrice ↑

As garage area ↑
SalePrice ↑

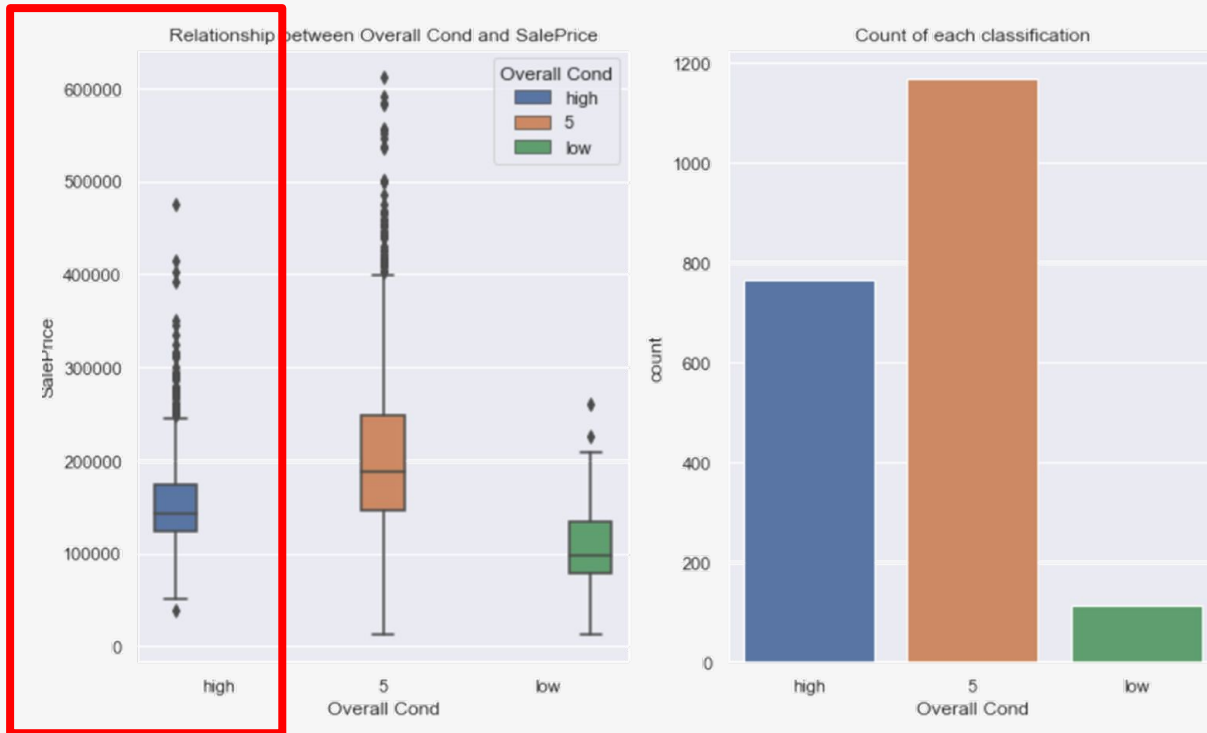


Data has insignificant relationship with SalePrice



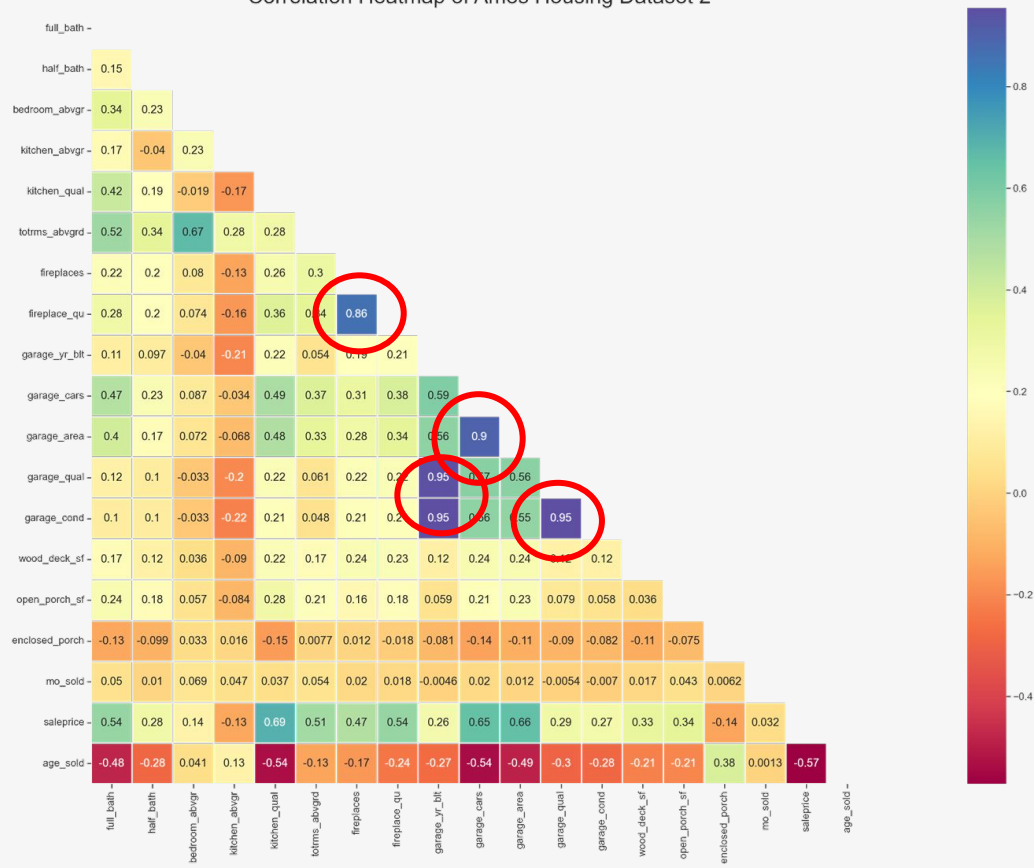
Excellent overall condition does not mean ↑ price

Overall Cond relationship to SalePrice and Count

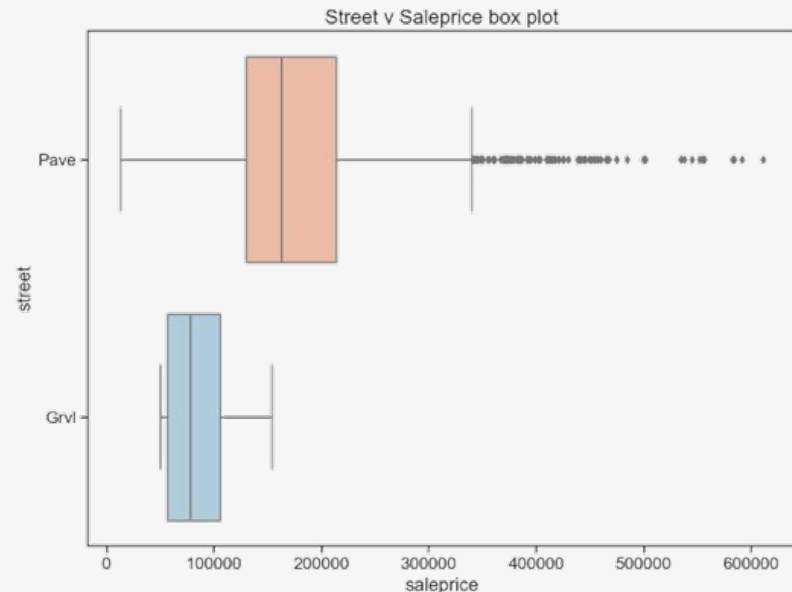
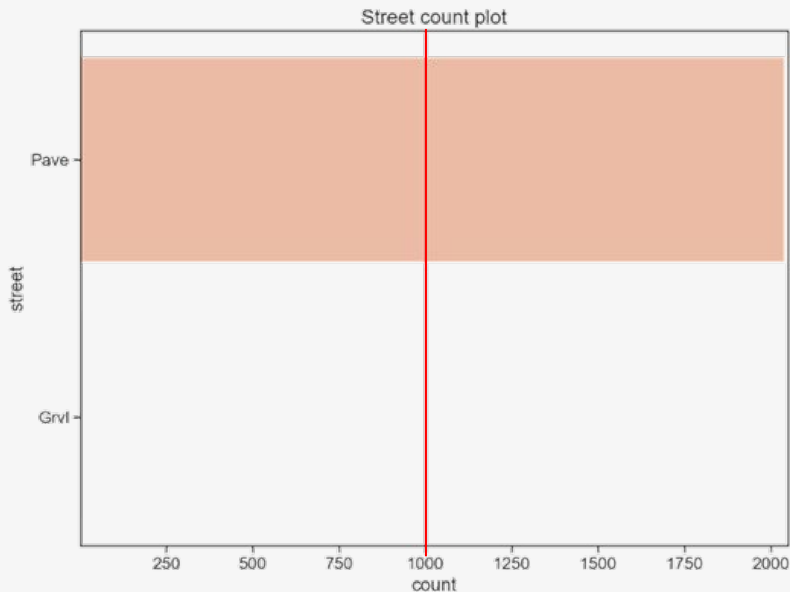


Some data points are highly related to each other

Correlation Heatmap of Ames Housing Dataset 2



Excluded categorical data that had a single, dominant category



Feature Engineering

When you clean your room so
the only trash left is you



Variables with weak correlation with Sale Price:

- | | |
|-----------------|----------------|
| - OpenPorchSF | - Bsmt_Conc |
| - EnclosedPorch | - BsmtFinType1 |
| - 3SsnPorch | - BsmtFinSF2 |
| - ScreenPorch | - BsmtFinType2 |
| - PoolArea | - WoodDeckSF |

Feature Engineering

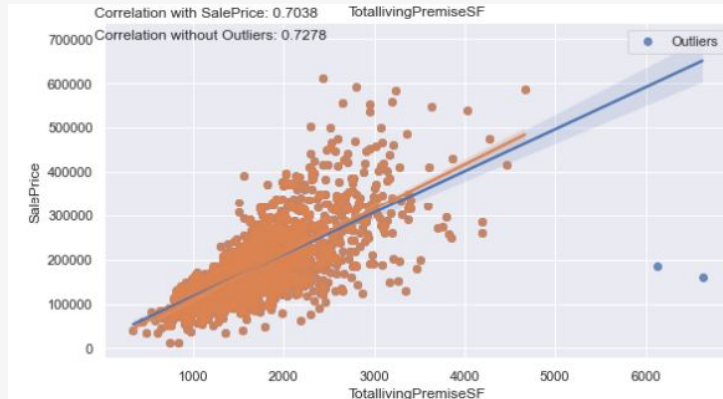
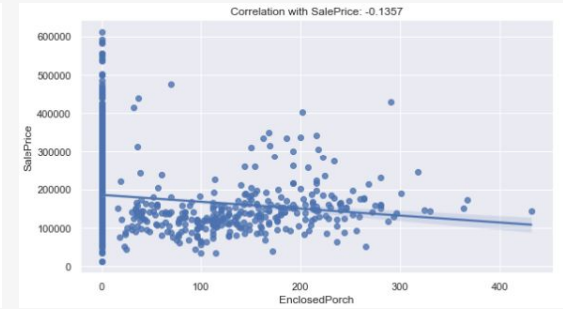
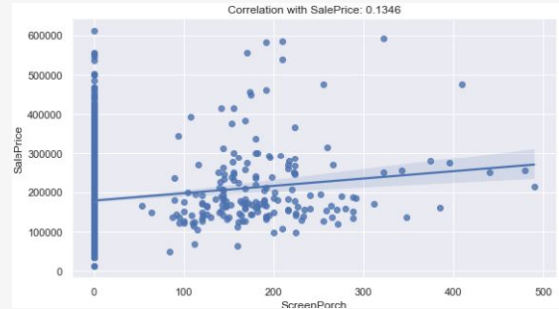
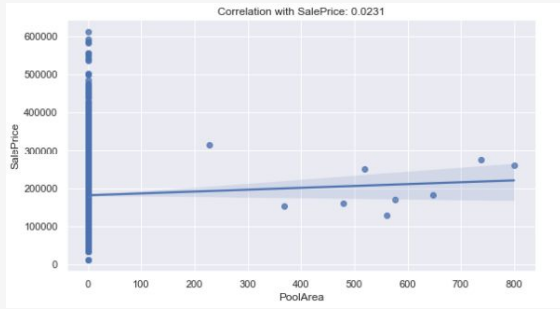
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea



TotallivingPremiseSF

OpenPorchSF + EnclosedPorch + 3SsnPorch + ScreenPorch + PoolArea + GrLivArea

Feature Engineering



Feature Engineering

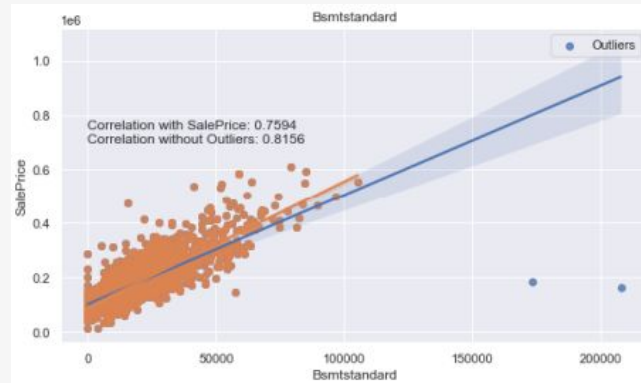
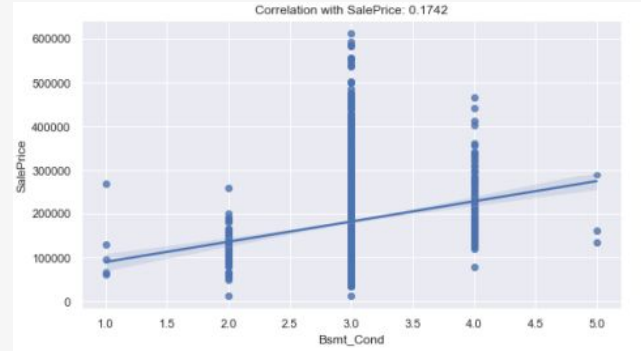
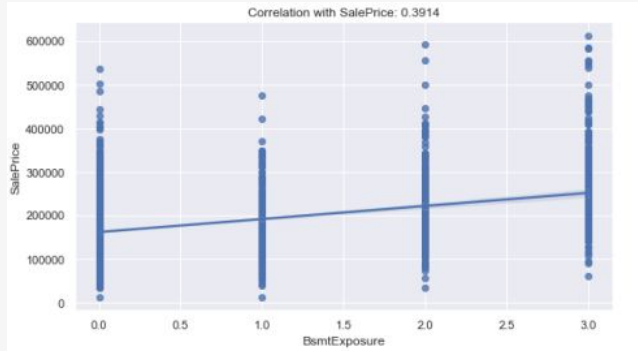
TotalBsmtSF
Bsmt_Qual
Bsmt_Cond
BsmtExposure



Bsmtstandard

$\text{TotalBsmtSF} + (\text{Bsmt_Qual}^2 + \text{Bsmt_Cond} + \text{BsmtExposure}^2)$

Feature Engineering



Feature Engineering

GarageArea
GarageQual

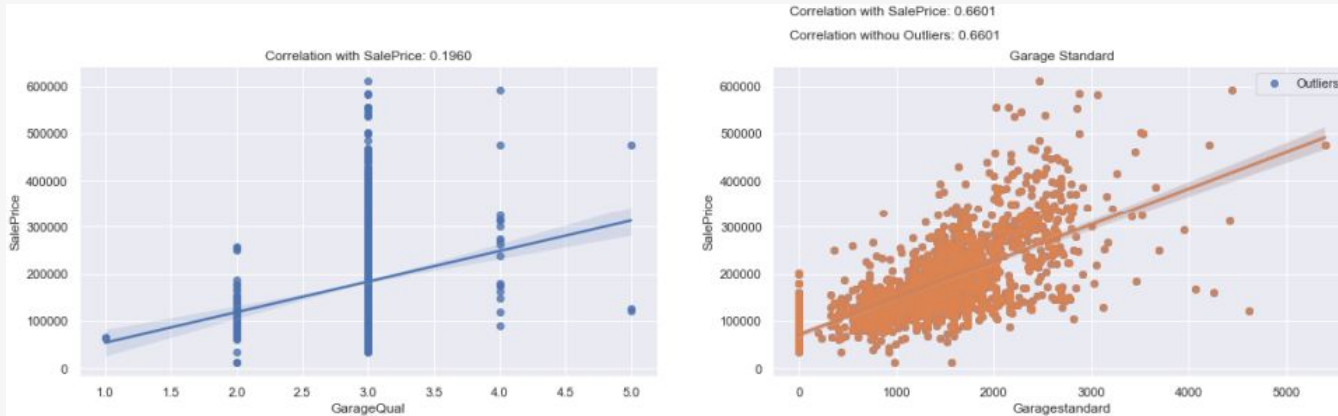


GarageArea x GarageQual



Garagestandard

Feature Engineering



List Of Regressions Used

- Linear Regression
- Lasso Regression
- Ridge Regression



Agent 1's Model – Ridge Regression

- Ridge Regression best, based on Root Mean Square Error (RMSE)
- Imputed continuous, numerical variables using average, and categorical variables using the most observed categories.

Agent 2's Model – Lasso Regression

- Rank features based on the strength of their influence
→ the stronger, the better
- Imputed continuous, numerical variables using iterative imputer

H2H Model Performance

Model	Metric (RMSE)	Stability
Model 1 - Agent 1's Ridge	26, 364	Less stability
Model 2 - Agent 2's Lasso	25, 530	More stability

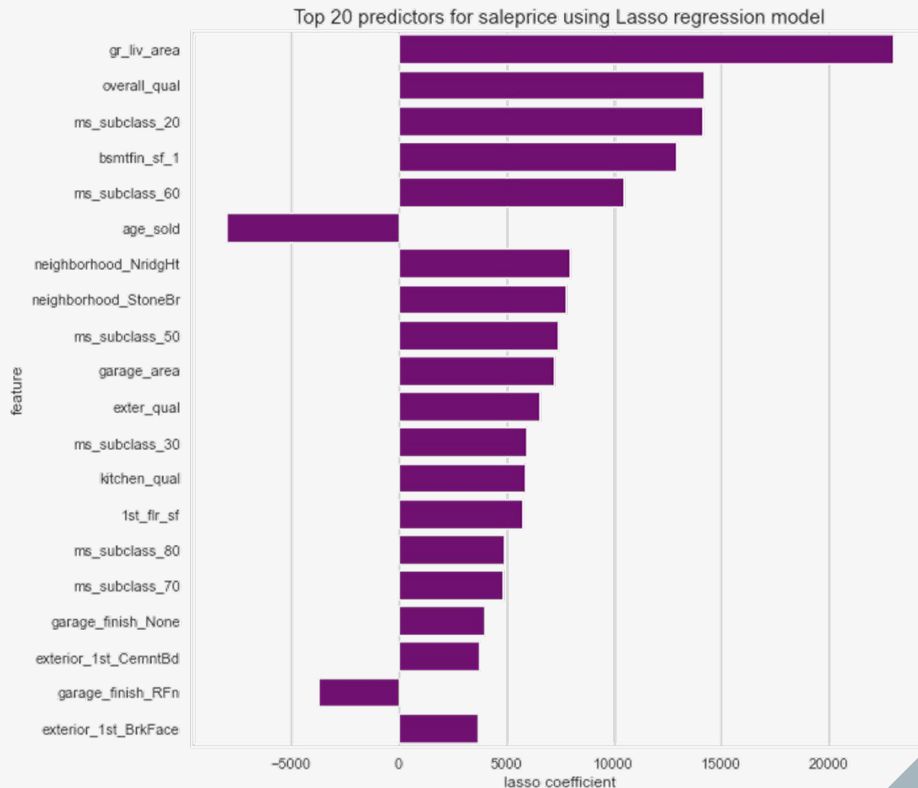
Final number of features: Model 1 - 44 features (16 categorical and 28 numerical features)

Final number of features: Model 2 - 57 features (32 categorical and 25 numerical features)

Top 3 Tips to enhancing your house's value

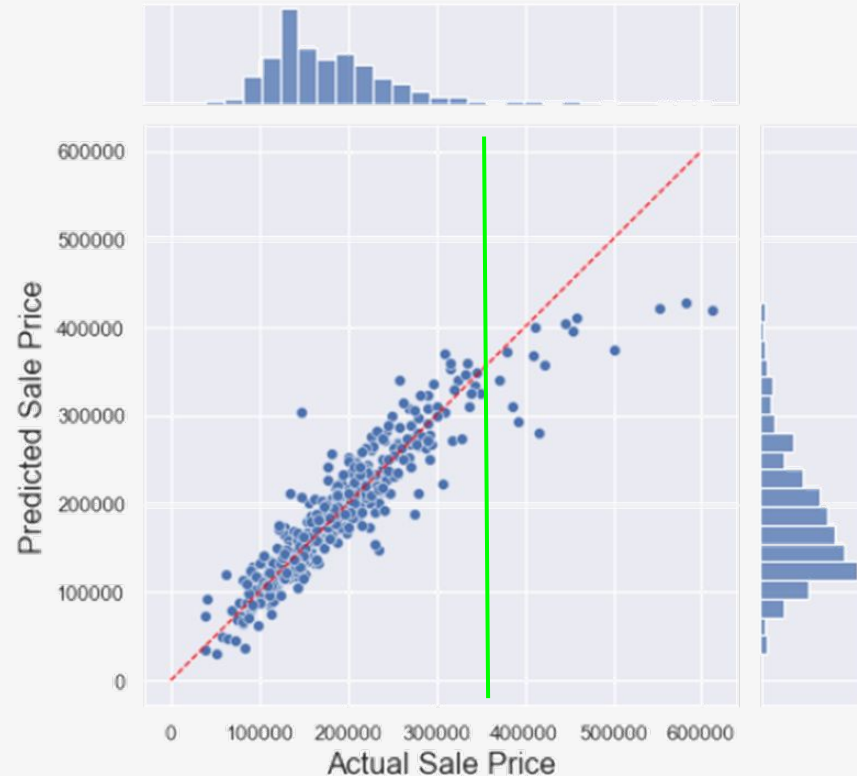
Want better offers for your house?

1. Maintain good / excellent quality of house thru regular maintenance and upkeep of interior / exterior features
2. Have good / excellent quality kitchen to attract better prices
3. Consider remodelling esp if house is old, to reinstate it to its former glory, i.e. Uplift overall quality of house



Technical Limitations and Next Steps

1. Our pricing model will underprice properties with high expected Sale Price
2. Looking at the increasing size of error as Sale Price gets large, can experiment with non-linear forms

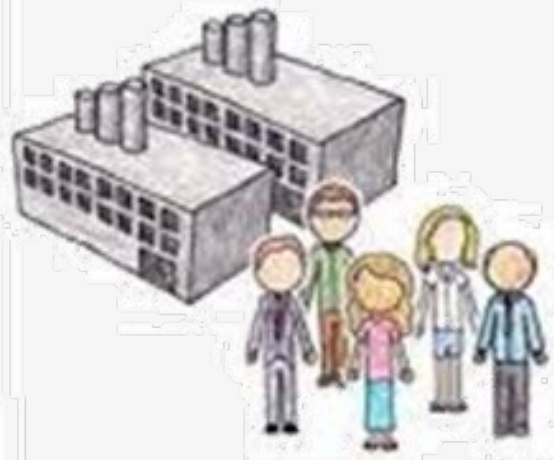


General Limitations

MACRO



MICRO





MICRO	MACRO
Lacking further data points	Bank interest mortgages
Ambiguity in given dataset (overall qual? Bsmt qual? How is it determined?)	Unemployment rates
Dataset is specific to Ames only	Crime rates





**THANK
YOU**