



BUSINFO716

# PREDICTING SALES PRICE

Using Machine Learning Models



TAHA SAMEER  
SHARAN SRINIVASAN  
GLEN LASRADO  
JIAJUN LI  
CHRISTIENE CHERIAN

tsam755  
ssri440  
glas952  
jli294  
cche436





# BUSINESS PROBLEM

## GOAL

Accurately predict housing prices and determine the key factors that influence price variations to support informed decision-making

## KEY CHALLENGES

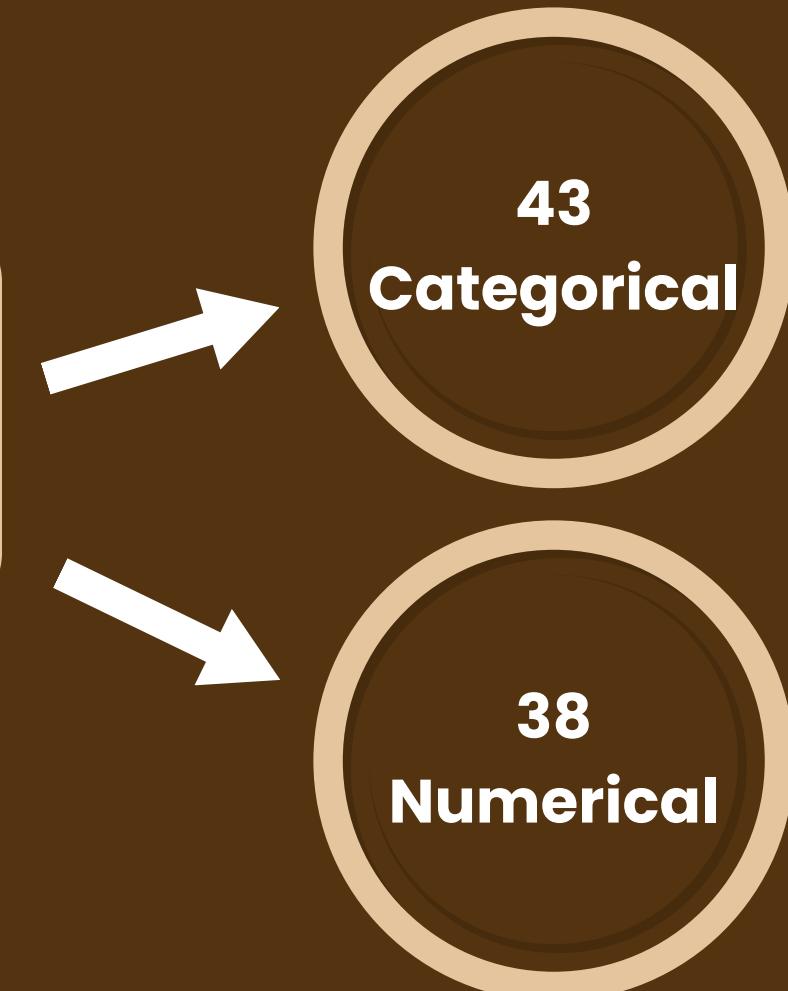
- Market Volatility
- Multiple Factors Influencing Price
- Financial Risks for stakeholders

# DATASET OVERVIEW



2,908  
**Observations**

81  
**Features**

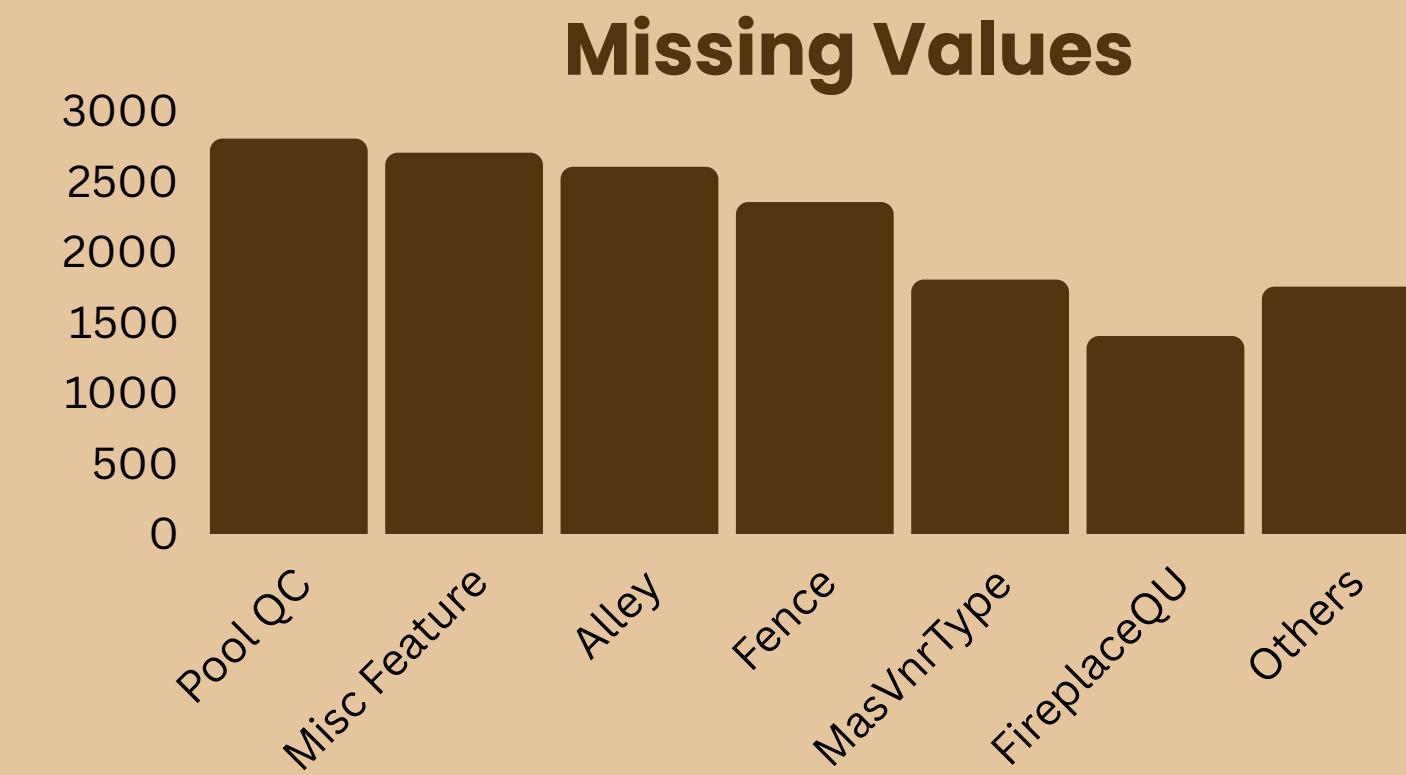


## LIMITATIONS

01. Missing Values

02. Outliers

03. Small Dataset



# DATA CLEANING

## Missing values Imputation

Feature Name	Description	Imputed Value
GarageType	Garage location	No Garage
PoolQC	Pool quality	Not Applicable
FireplaceQu	Fireplace quality	No Fireplace

## Variable Transformation

Feature Name	Original Value	Imputed Value
	0 - 4	below average
Overall Quality	5	average
	6 - 10	above average



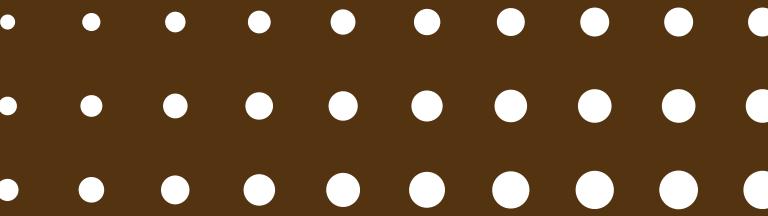
# FEATURE SELECTION

**Drivers for Variable Selection**

**Model 1**  
**Statistical Approach**

**Model 2**  
**All features used**

# MODEL RESULTS 1

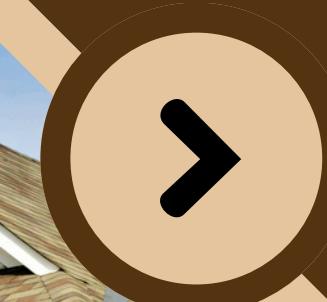


## ASSUMPTIONS

Features were selected using statistical approach

Model	Performance on Unseen Data
Linear Regression	Difference of \$25,893 from actual value
Ridge Regression	Difference of \$18,600 from actual value
Lasso Regression	Difference of \$45,444 from actual value

# VARIABLES DROPPED

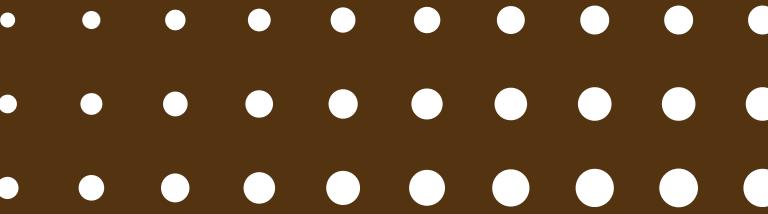


**Open porch area (sq. ft)**

**Lot size (sq. ft)**

**Number of kitchens and Bedroom**

# MODEL RESULTS 2



## ASSUMPTIONS

All features are significant and impact the sale price.

Model	Performance on Unseen Data
Linear Regression	Difference of \$18,600 from actual value
Ridge Regression	Difference of \$18,600 from actual value
Lasso Regression	Difference of \$33,479 from actual value

# RESULT COMPARISON



	Model Results 1	Model Results 2
Linear Regression	Difference of \$25,893 from actual value	Difference of \$18,600 from actual value
Ridge Regression	Difference of \$18,600 from actual value	Difference of \$18,600 from actual value
Lasso Regression	Difference of \$45,444 from actual value	Difference of \$33,479 from actual value

**RIDGE IS THE BEST MODEL**

# BUSINESS TAKEAWAY

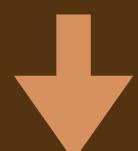
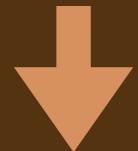
## KEY FEATURES

**Overall Condition**  
**Below Average Rating**

**Overall Quality**  
**Below Average Rating**

**Home  
Functionality**

**Sales Price**



## LIMITATION OF THE ANALYSIS

- Training Data sample size
- Regional specific
- Macro economic influence not take account into our analysis



BUSINFO716

# THANK YOU

Q&A?

