

# PROJECT 2

# AMES VACATION HOUSE PRICE

*Here is how you get richer*



# AGENDA

## INTRODUCTION



Problem statement and purpose of this presentation

## EDA



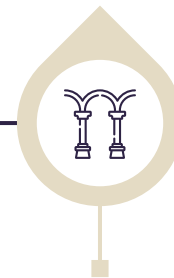
Which features matter and which do not

## MODEL



Which model was utilise

## RECOMMENDATIONS



What to do next for your vacation house

# 01

## INTRODUCTION



## PROBLEM STATEMENT

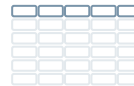
We are personal assistants to an ultrarich family. The head of the family (who shall not be named), has approached us to increase their vacation house value in Ames..





## PURPOSE

- Housing data from 2006 to 2010
- Identify features to maximise value



# SITUATION

- Chaotic housing market in 2021
- Demand of house > Supply of house



# DATA SET

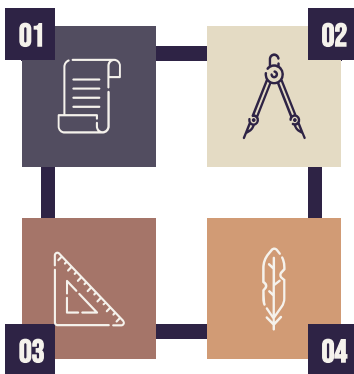
Total number of features: 81

## NOMINAL

- Alley access
- Misc features
- House style

## CONTINUOUS

- Lot frontage
- Sale prices
- Square feet



## ORDINAL

- External Quality
- External Conditions
- Overall condition

## DISCRETE

- Number of bathrooms
- Number of fireplaces
- Garage cars



# 02

## EDA





# DATA TREATMENT

## FILLING THE NULL

Impute with Mean, Mode, 0

## TREATING THE ORDINAL

With Ratings

- Quality
- Conditions
- Yes No

## FEATURE ENGINEERING

Combining the SQFT together  
Create the age feature

## DROPPING COLUMNS / OUTLIER

PID - Property ID  
Dropping outlier

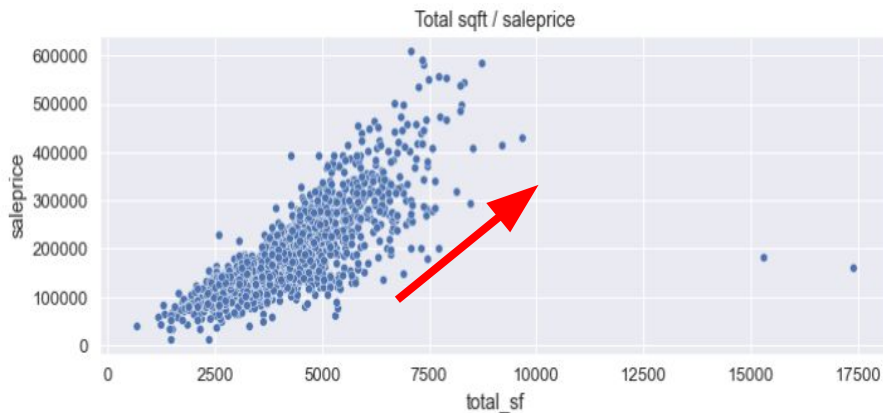
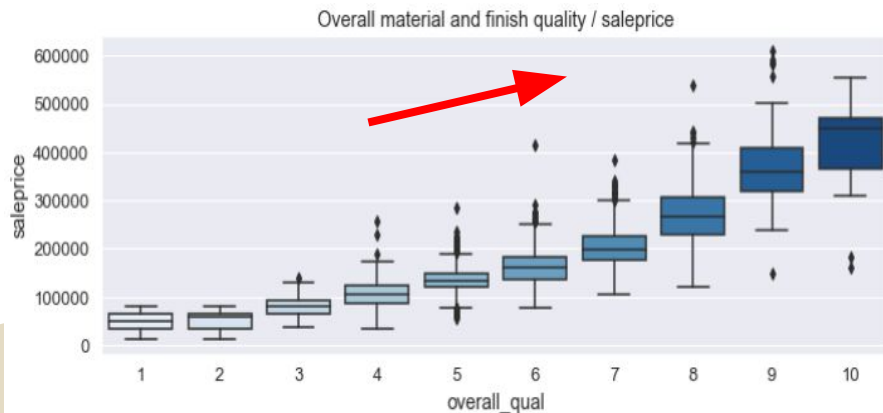
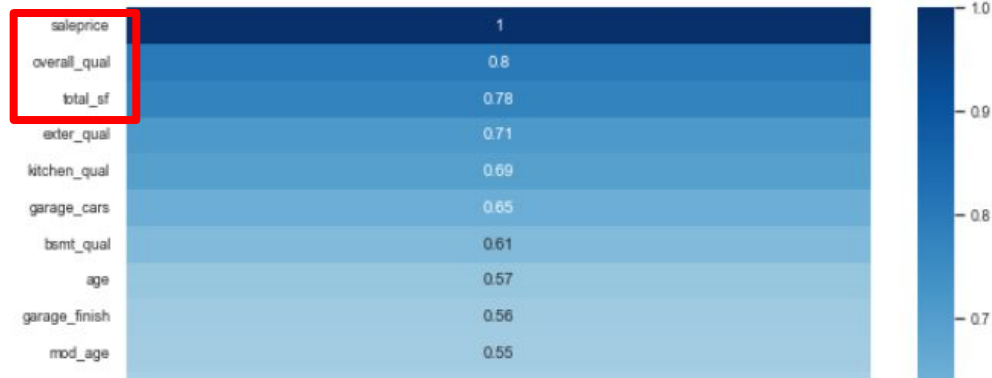
	Count	Percent	Dtype
Pool QC	2916	99.556163	object
Misc Feature	2823	96.381017	object
Fireplace Qu	1422	48.548993	object
Lot Frontage	490	16.729259	float64
		5.428474	object

27 Features

```
{ 'Ex' : 5,
  'Gd' : 4,
  'TA' : 3,
  'Fa' : 2,
  'Po' : 1,
  'NA' : 0 }
{ 'Gd' : 4,
  'Av' : 3,
  'Mn' : 2,
  'No' : 1,
  'NA' : 0 }
```

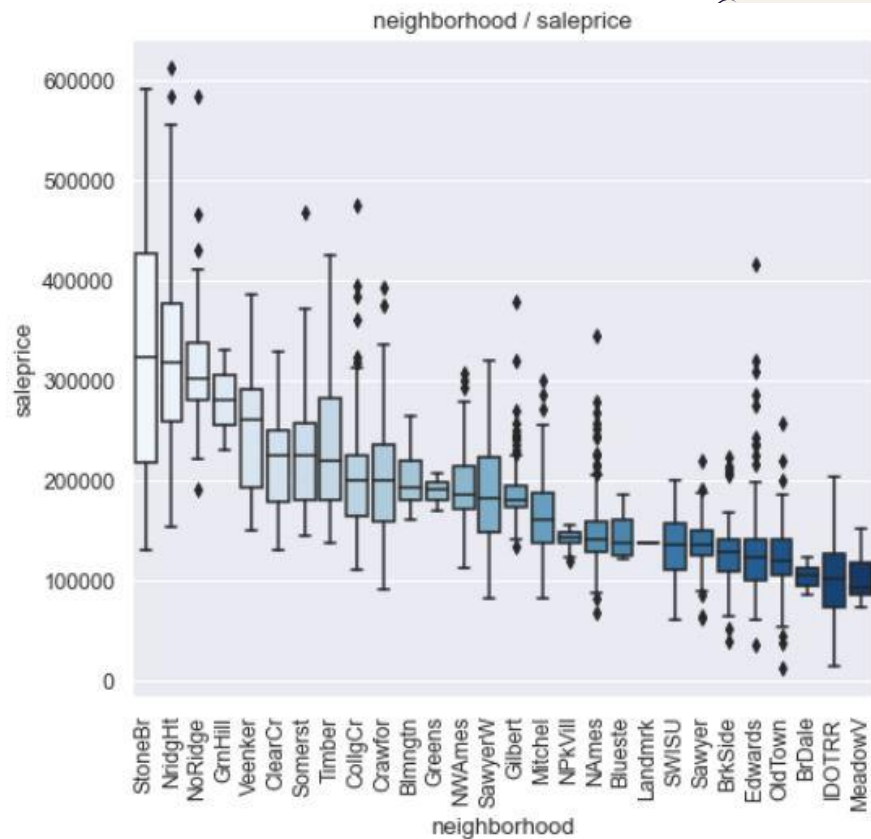
# FEATURE SELECTION

- Overall Quality
- Total Sqft
- Neighborhood

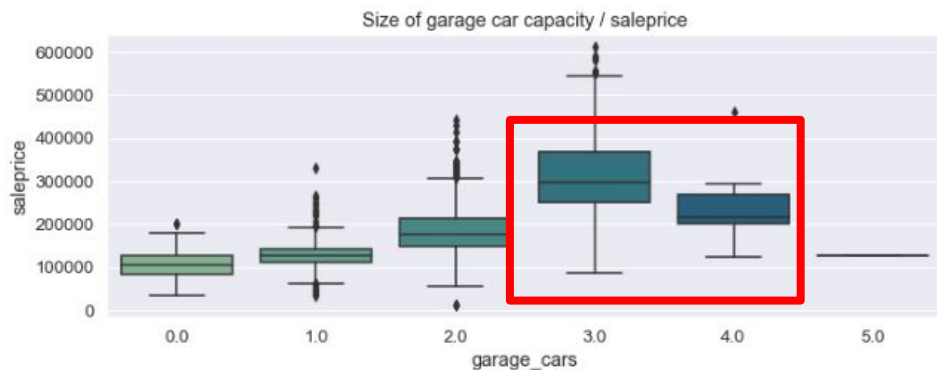


# FEATURE SELECTION

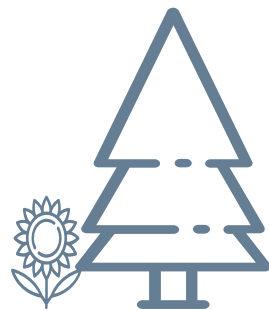
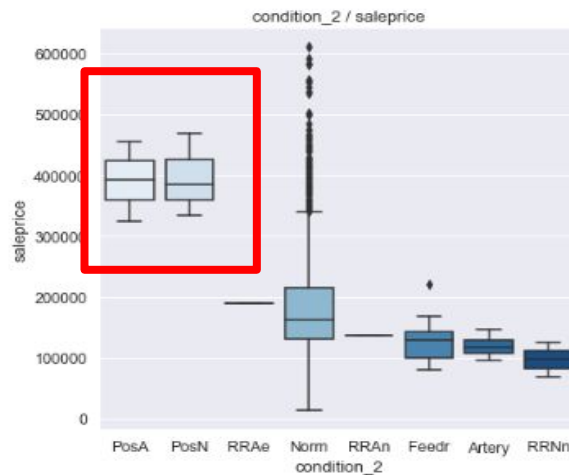
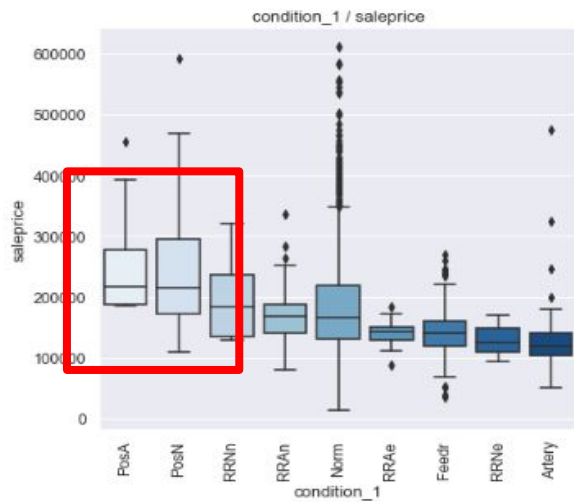
The Top 3 expensive  
neighbourhood price vs the  
Cheapest neighbourhood



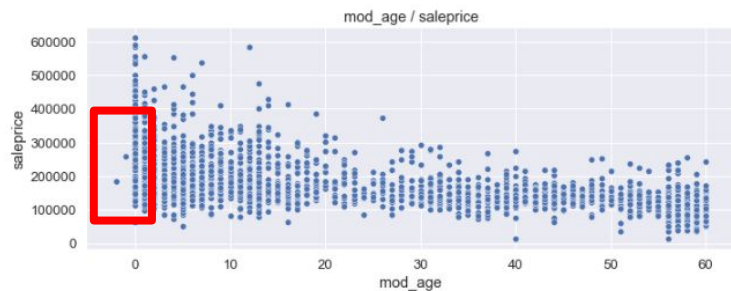
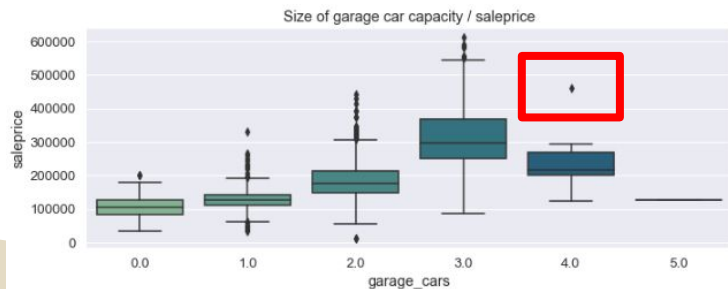
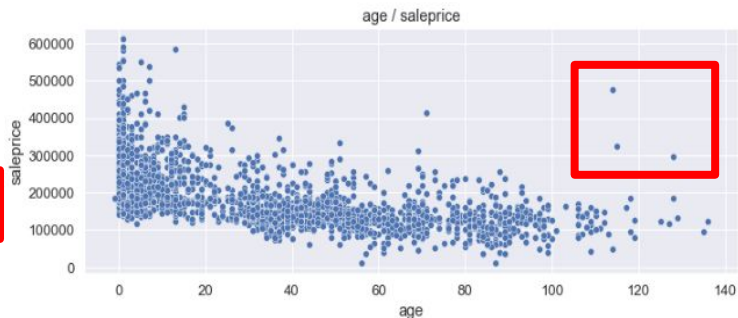
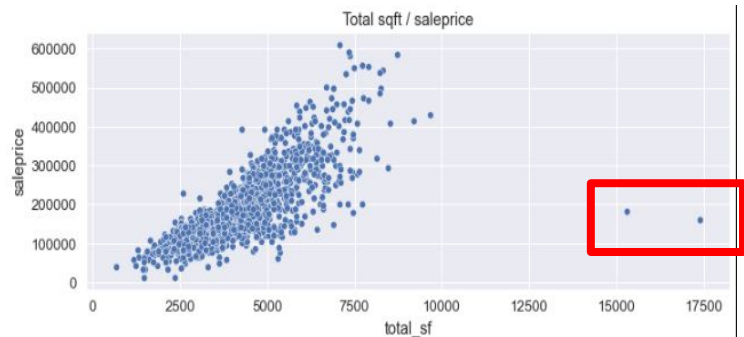
# INTERESTING FACT



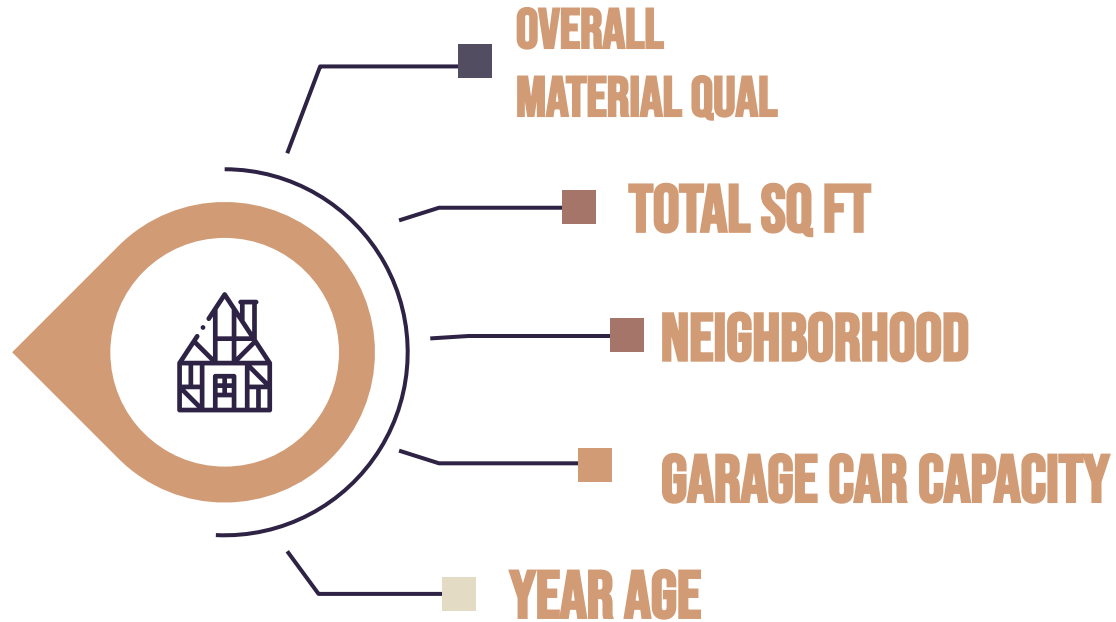
# INTERESTING FACT



# THE OUTLIERS



# IMPORTANT FEATURES



# \$611,657

The most expensive house

# 7,058 SQFT

Total Living Area

# N\_RIDG\_HT

Town





# 03

## MODEL

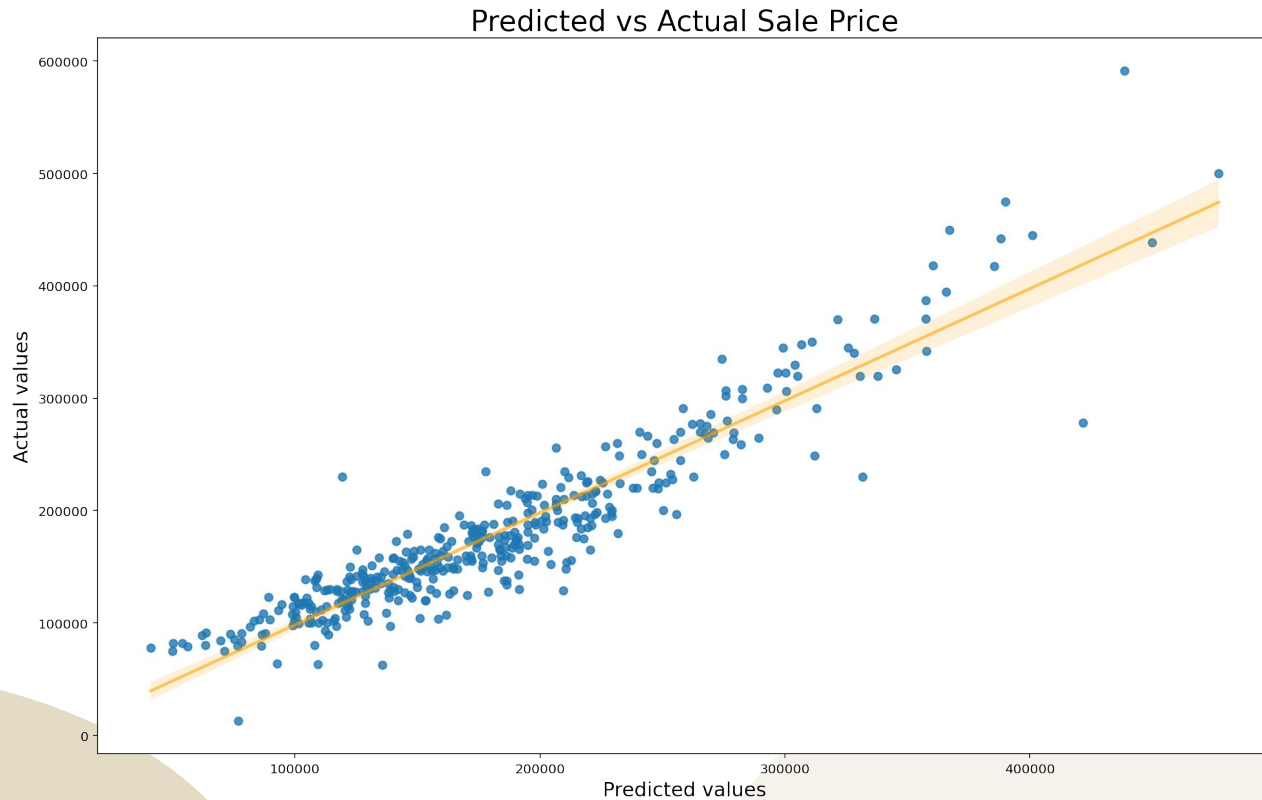


# LASSO MODEL

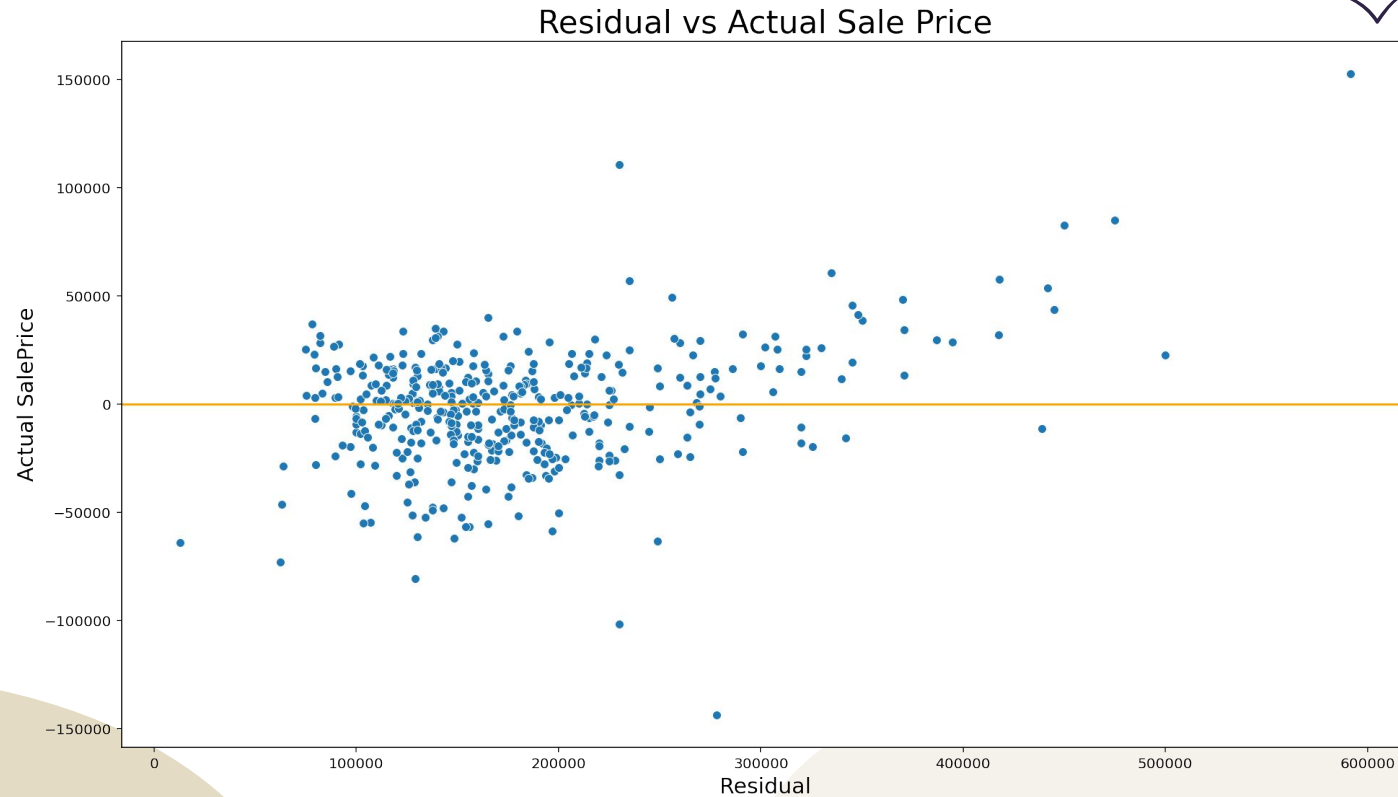
Based on the highest  $R^2$  score and  
lowest RMSE score



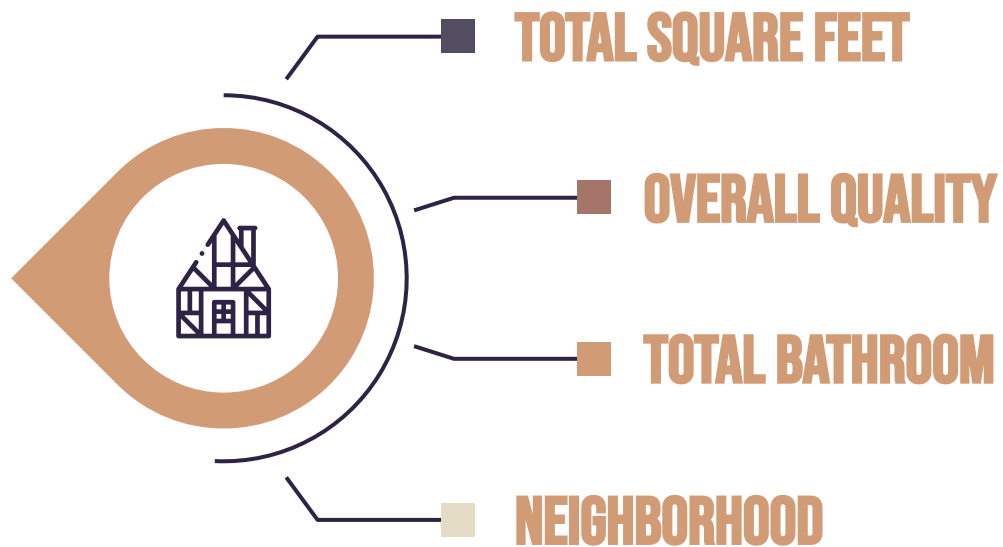
# PREDICTED VS ACTUAL PRICE



# RESIDUAL PLOT



# TOP PREDICTORS



# PREDICTION OUTCOME

	variable	coef	abs_coef
68	total_sq_feet	28648.833230	28648.833230
39	overall_qual	13957.349454	13957.349454
50	total_bathroom	10927.784046	10927.784046
5	neighborhood	10593.938058	10593.938058
26	kitchen_qual	9550.569083	9550.569083



# 04

## RECOMMENDATION



# THINGS HOMEOWNERS COULD DO TO INCREASE HOUSE PRICE

**REGULAR MAINTENANCE OF THE HOUSE**

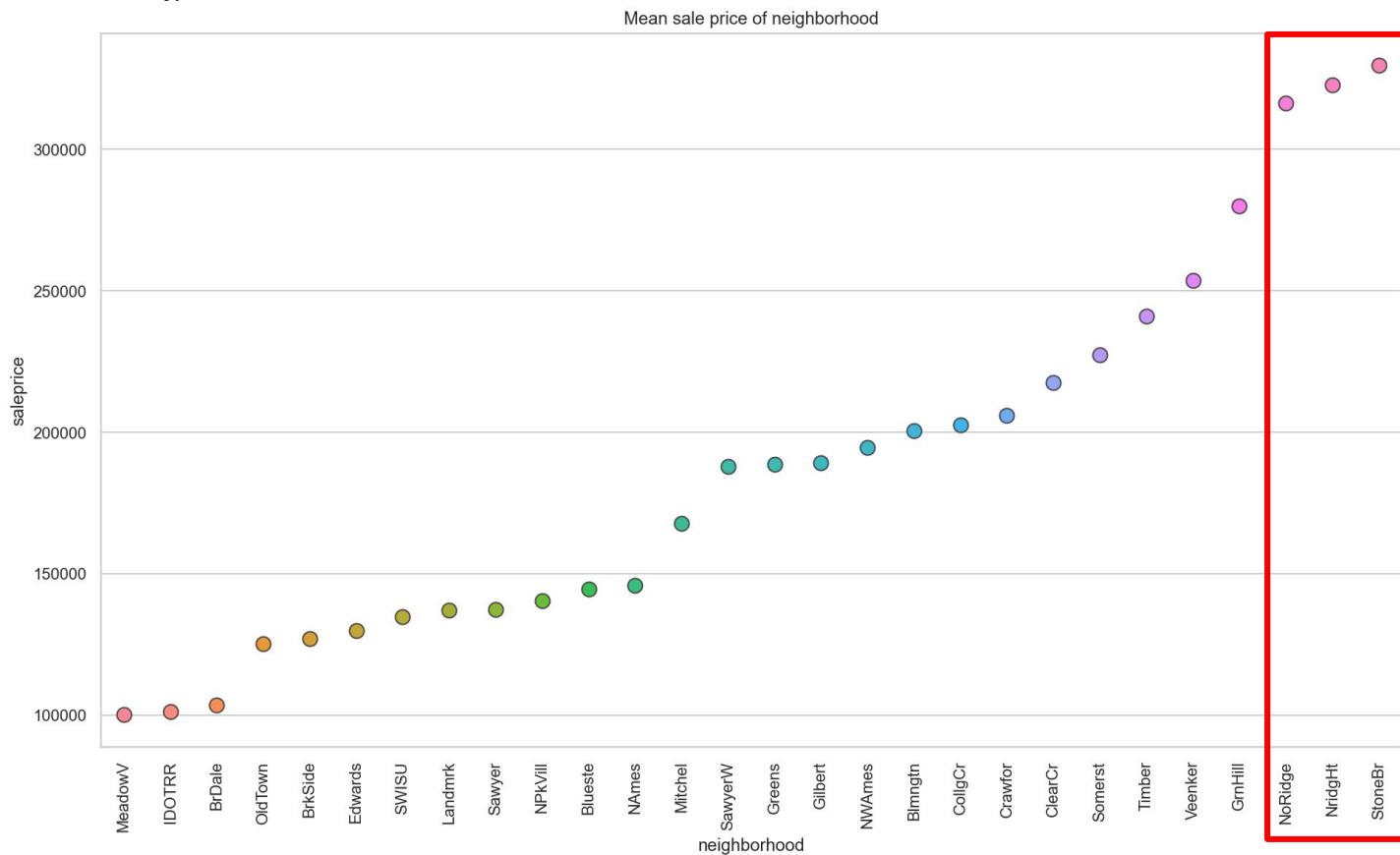
**REMODELING OF THE HOUSE**

**IMPROVE QUALITY OF KITCHEN**





# NEIGHBORHOODS THAT SEEM TO BE A GOOD INVESTMENT



# FURTHER EXPLORATION

**EXPLORE RELATIONSHIP  
BETWEEN FEATURES**

**BETTER IMPUTATION  
METHOD**

**USING AUTOMATED  
FEATURE SELECTION**

**USE MACHINE LEARNING  
TOOLS**



# THANKS!

Do you have any questions?

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

