

# Ames Housing Sale Price

---

Bryan Goh  
Edmund Ho  
Melissa Ng



# PROBLEM STATEMENT

We are a team of data scientists working in a real estate consultancy firm based in Iowa to advise clients (investors) on opportunities to optimise sale price for houses in Ames.



MYHOME

REAL ESTATE



**EXPLORATORY  
DATA ANALYSIS**



**DATA CLEANING**



**FEATURE  
ENGINEERING  
AND SELECTION**



**PREPROCESSING  
AND  
MODELLING**



**RECOMMENDATION**

## It's All About Location, Location, Location



***"Real estate cannot be lost or stolen, nor can it be carried away. Purchased with common sense, paid for in full, and managed with reasonable care, it is about the safest investment in the world."** – Franklin D. Roosevelt*

You can remodel the kitchen, add on a bedroom, knock down a wall and change almost anything about a house with some tools and a paint brush ... everything, that is, but the location. Which is why veteran real estate professionals constantly remind folks of the basic rule of real estate — location, location, location...

You can remodel the kitchen, add on a bedroom, knock down a wall and change almost anything about a house with some tools and a paint brush ... everything, that is, but the location. Which is why veteran real estate professionals constantly remind folks of the basic rule of real estate — location, location, location.

It is the real estate agents' mantra: Location, location, location. You've certainly heard the phrase enough and may wonder what inspires agents to say the word three times. In a nutshell, location, location, location means identical homes can increase or decrease in value due to location. The saying is repeated three times for emphasis, and it is the number one rule in real estate, though it is often the most overlooked.

Location is number one, two and three on the list, according to many real estate experts. It's what drives the price. Are you on the beach or are you just off the beach? Even the word "waterfront" can mean different things depending on the location of the neighborhood. Waterfront might be affordable in Višnjica but it's not in "Belgrade Waterfront". Remember, you can fix up the mediocre house but you can't pick up and move the nice one.

## 8 critical factors that influence a home's value

March 27, 2019 — Written by Joe Gomez

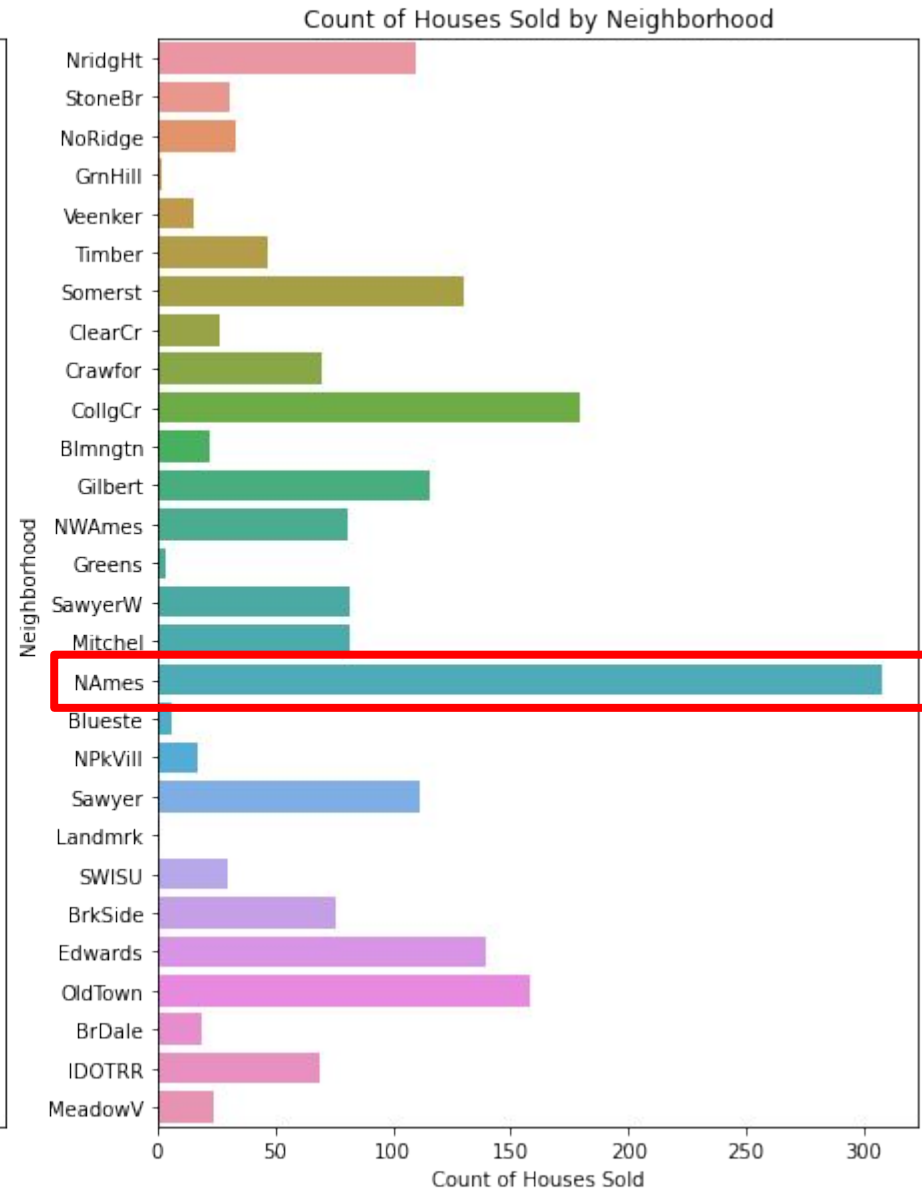
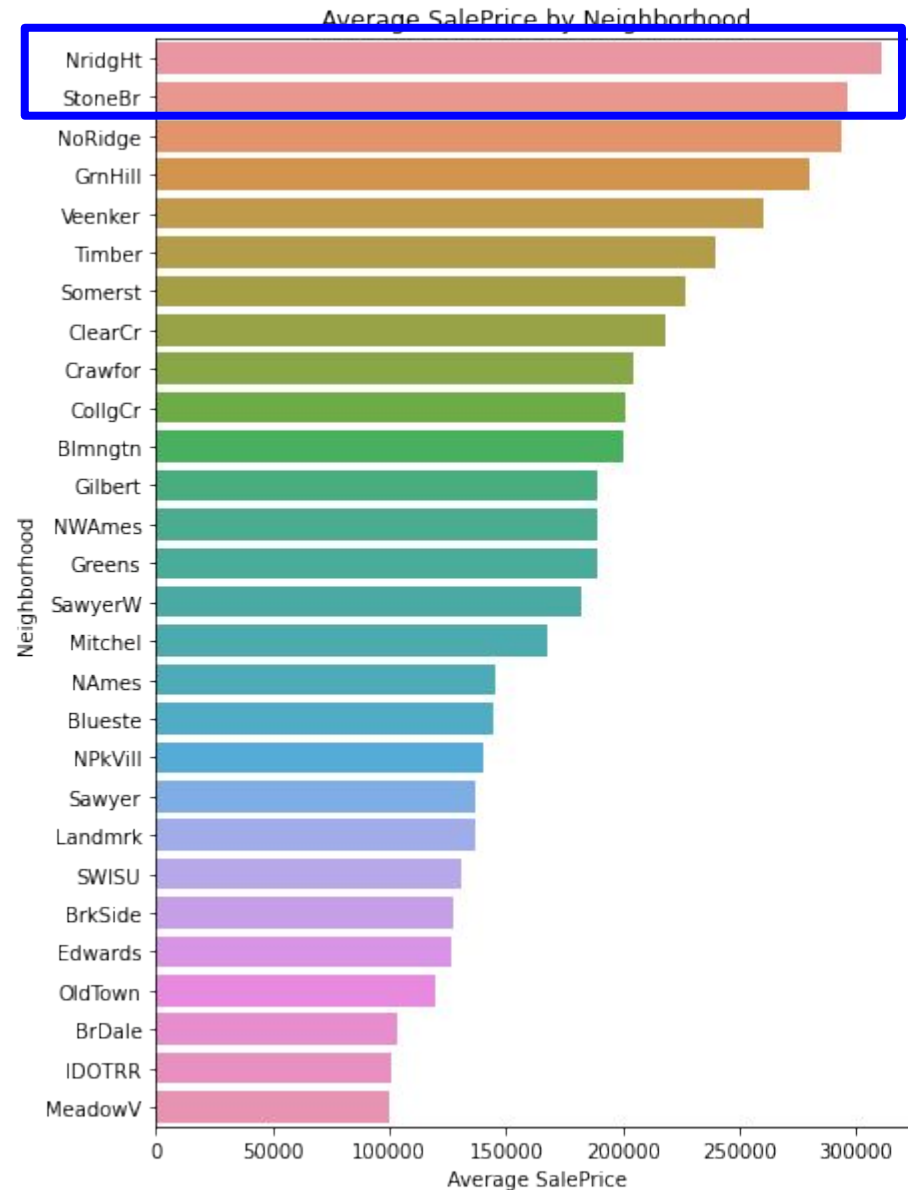
Share this article



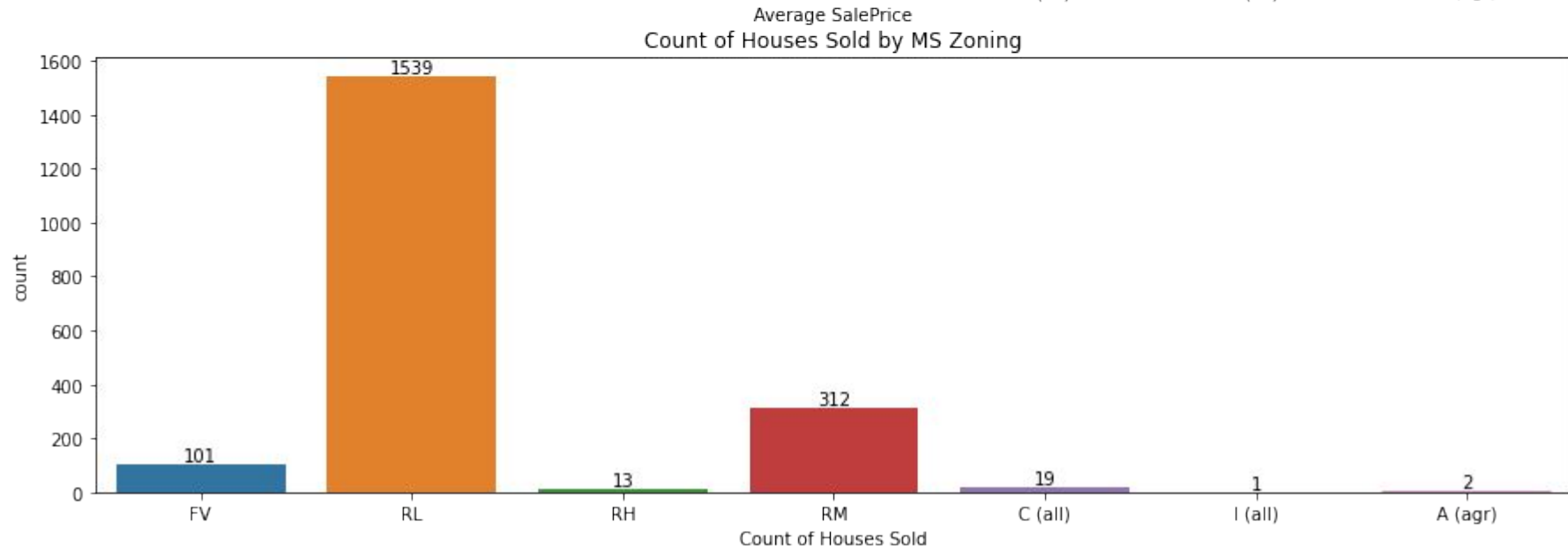
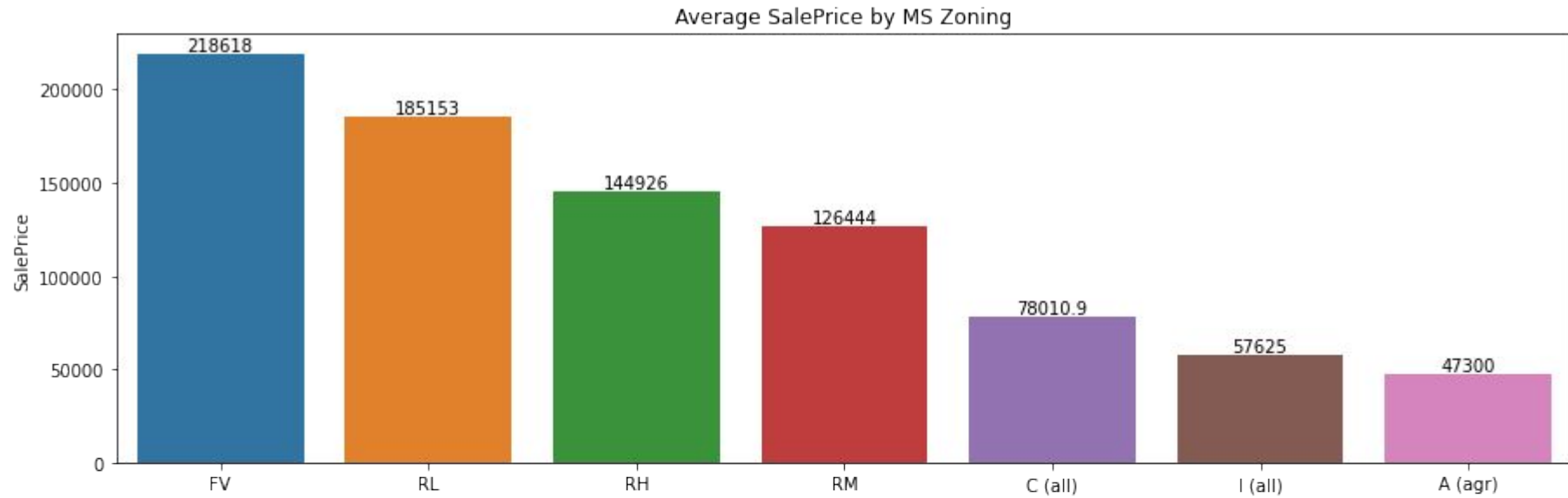
### Factors relatable to problem statement

1. Location
2. Home Size and usable space
3. Age and Condition
4. Upgrades and updates

# EXPLORATORY DATA ANALYSIS - Location

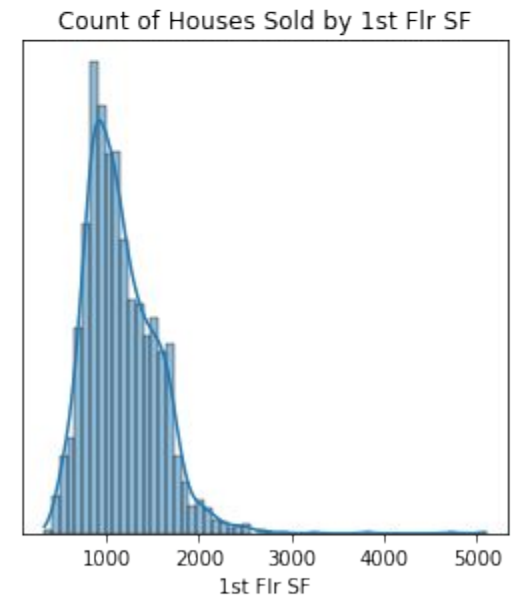
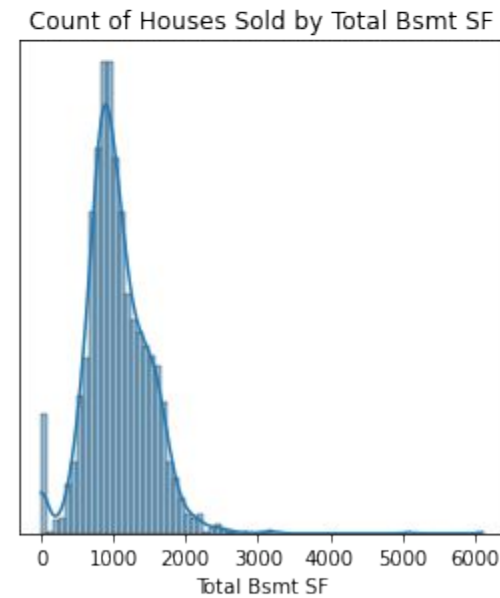
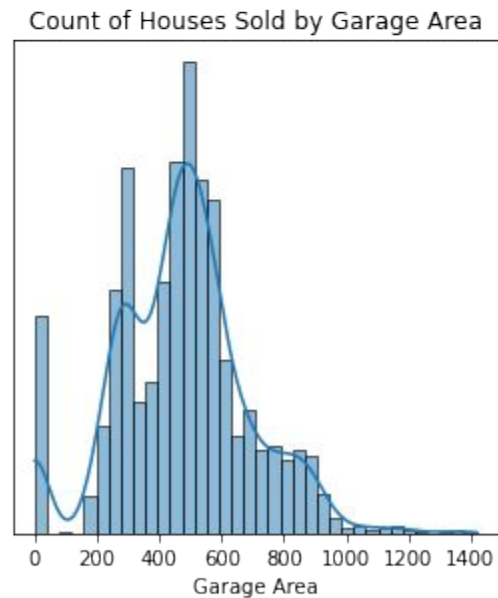
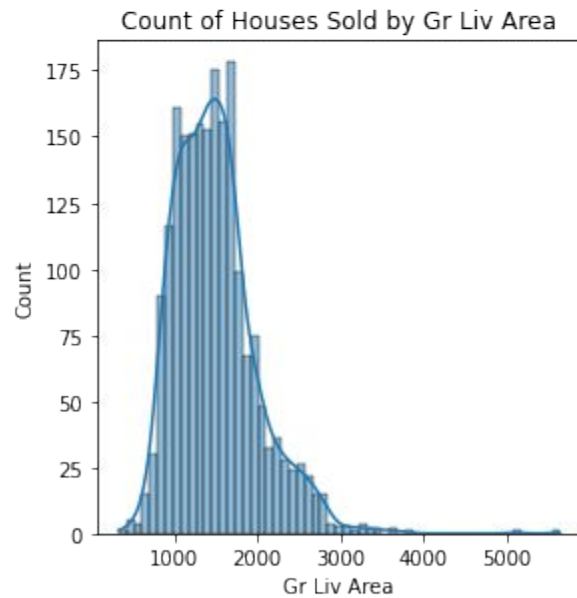
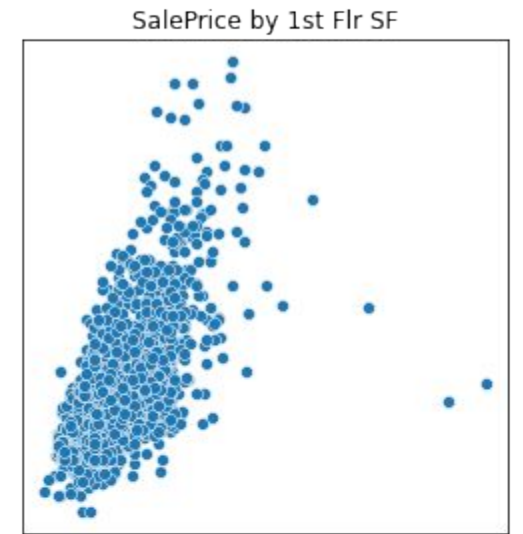
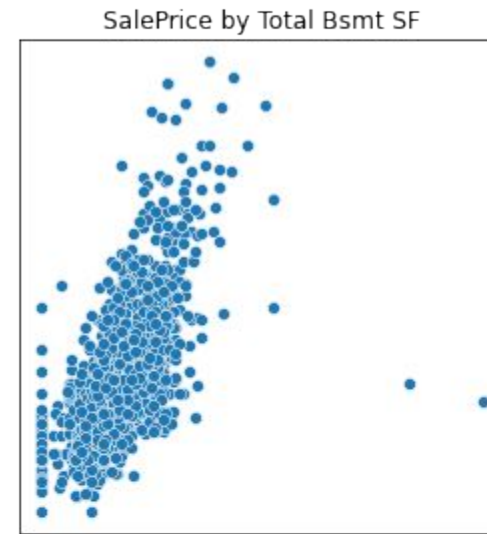
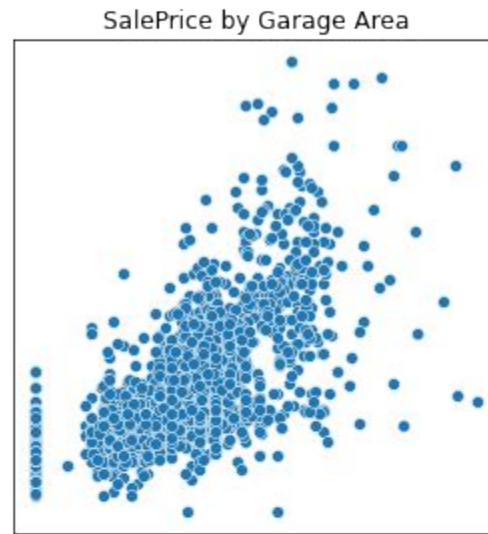
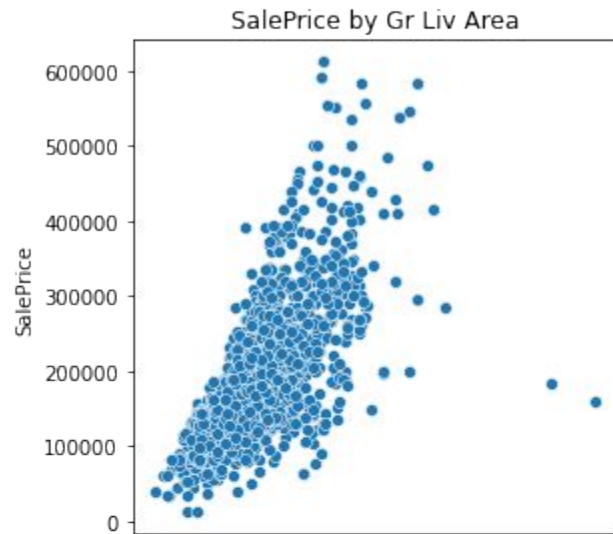


# EXPLORATORY DATA ANALYSIS - Location

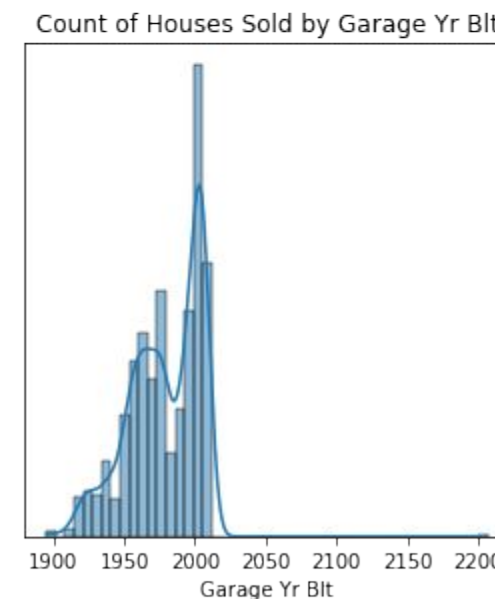
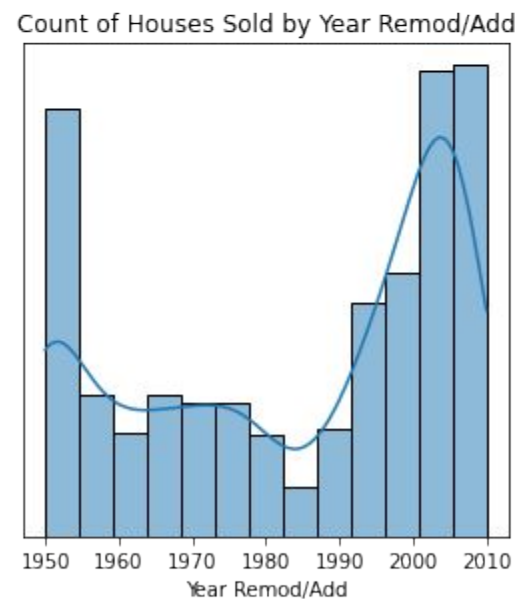
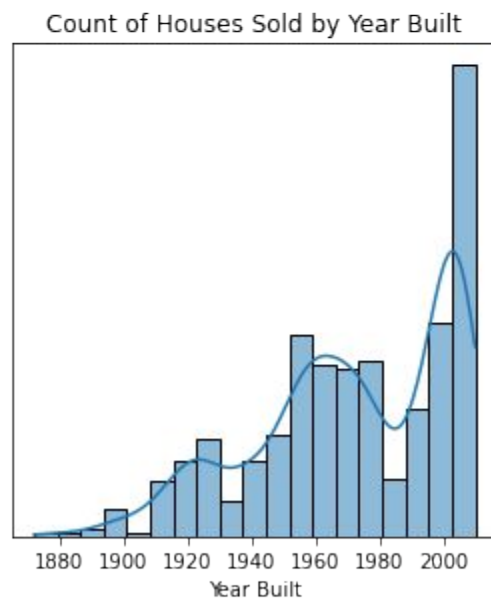
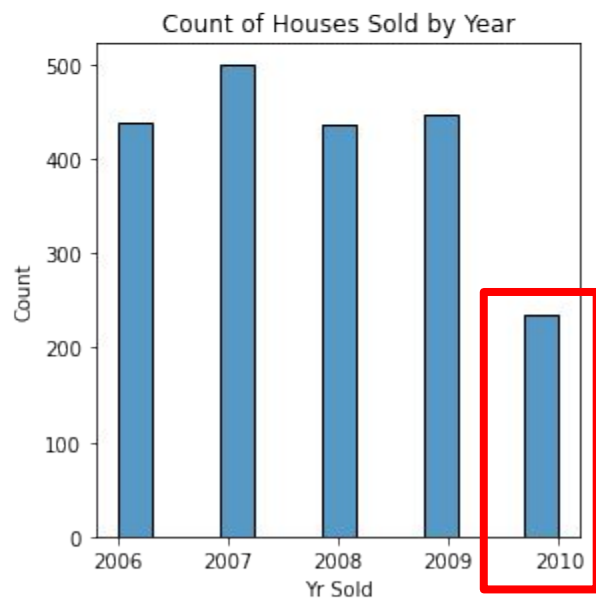
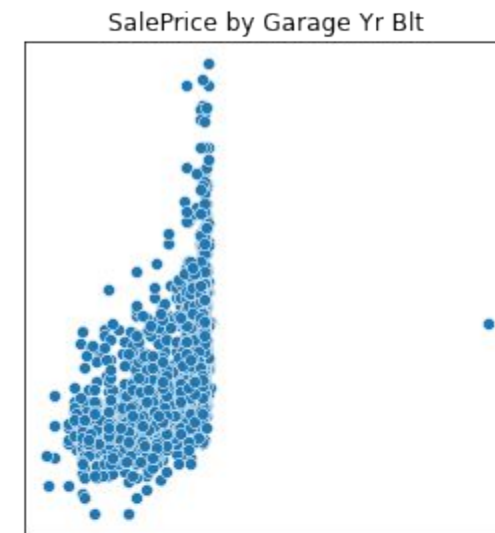
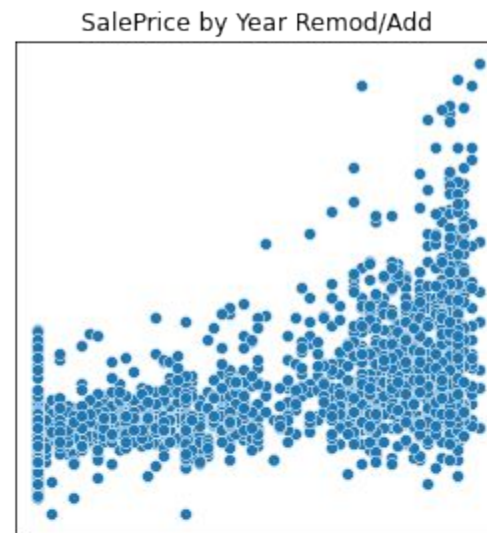
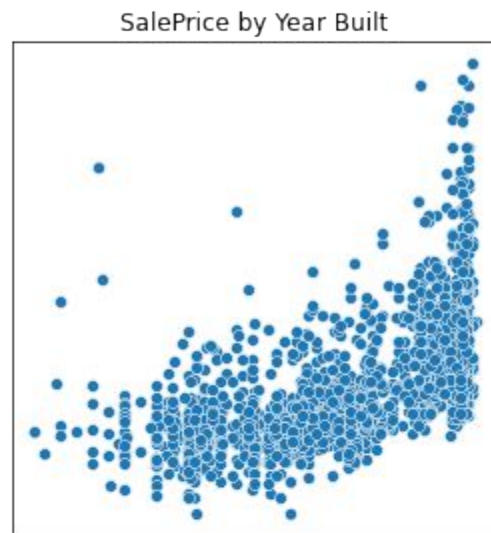
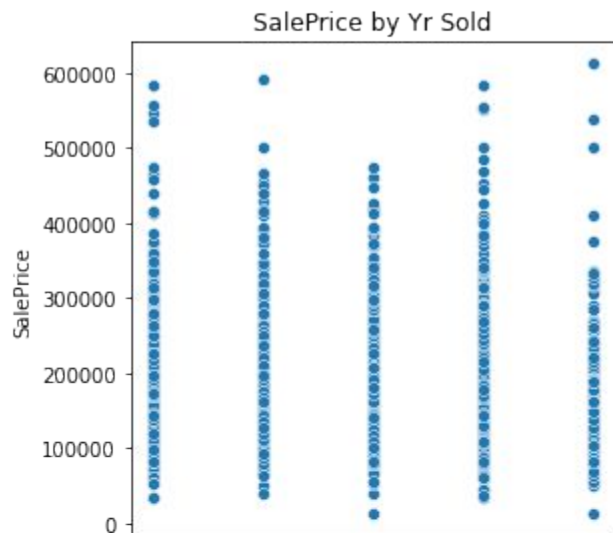




# EXPLORATORY DATA ANALYSIS - House Size

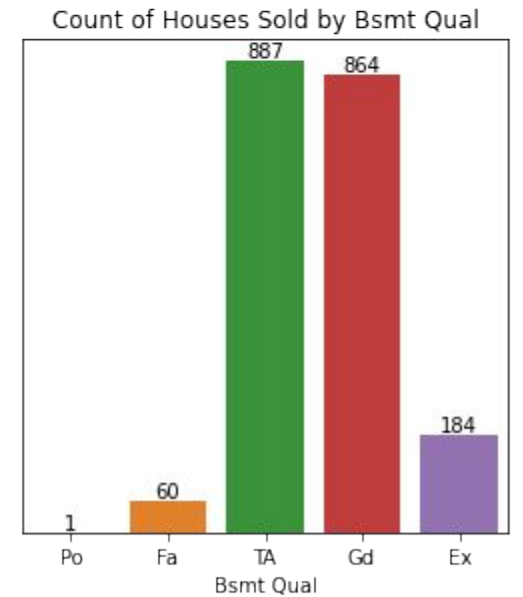
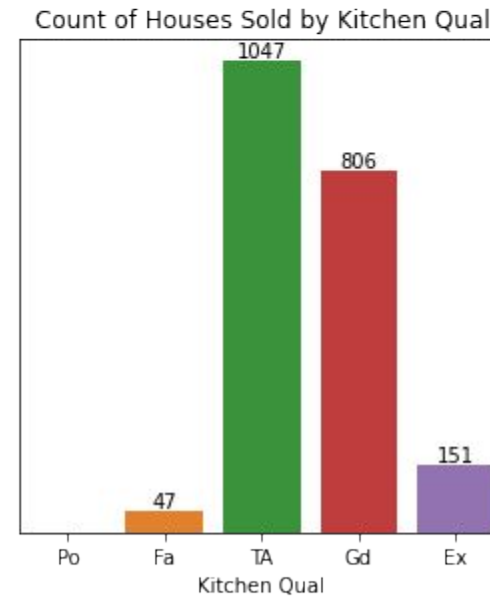
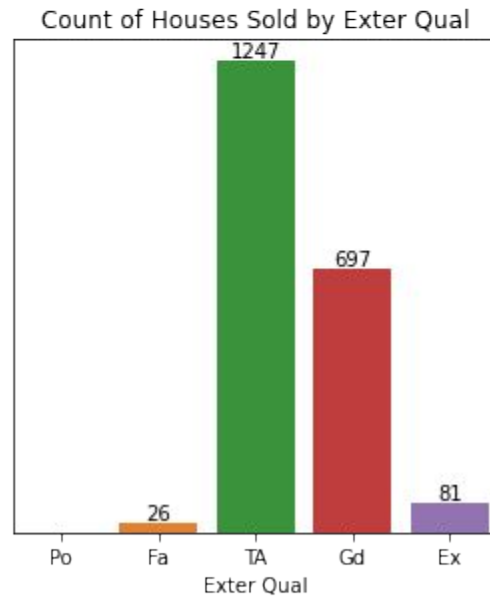
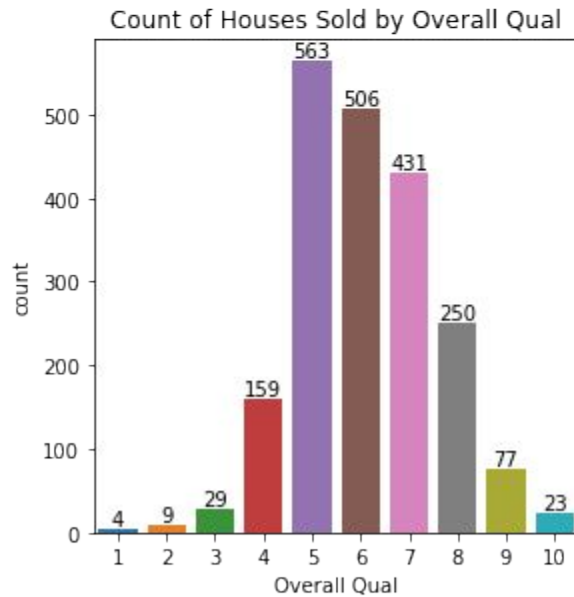
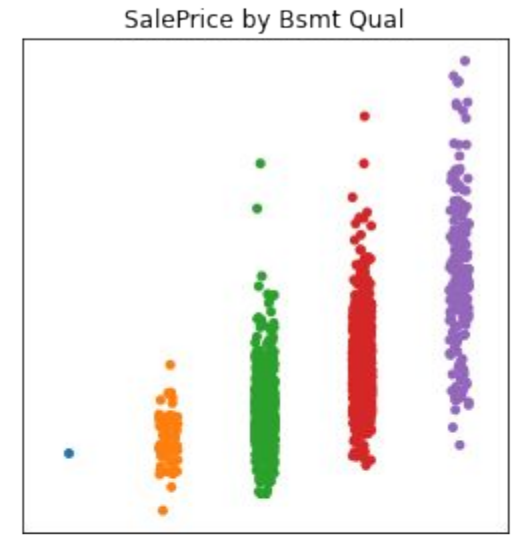
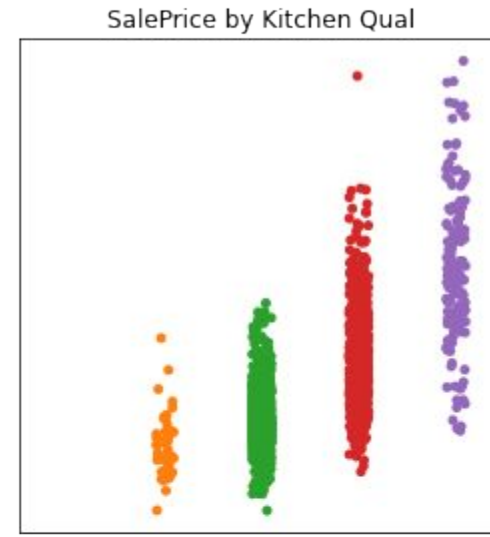
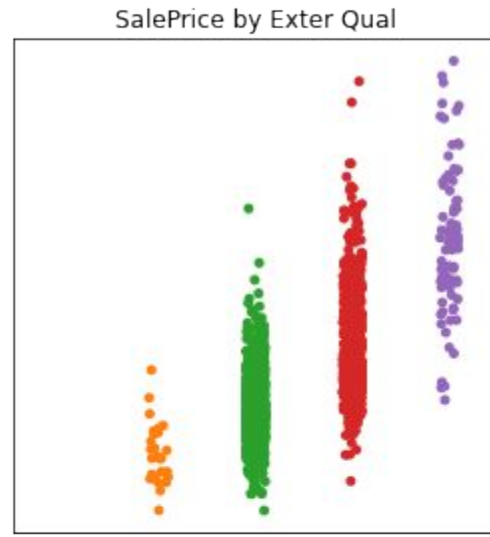
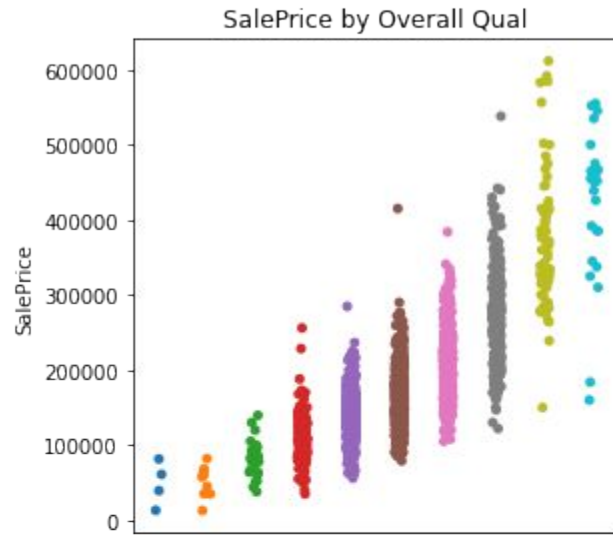


# EXPLORATORY DATA ANALYSIS - Age





# EXPLORATORY DATA ANALYSIS - Condition



# DATA CLEANING

## Missing Values

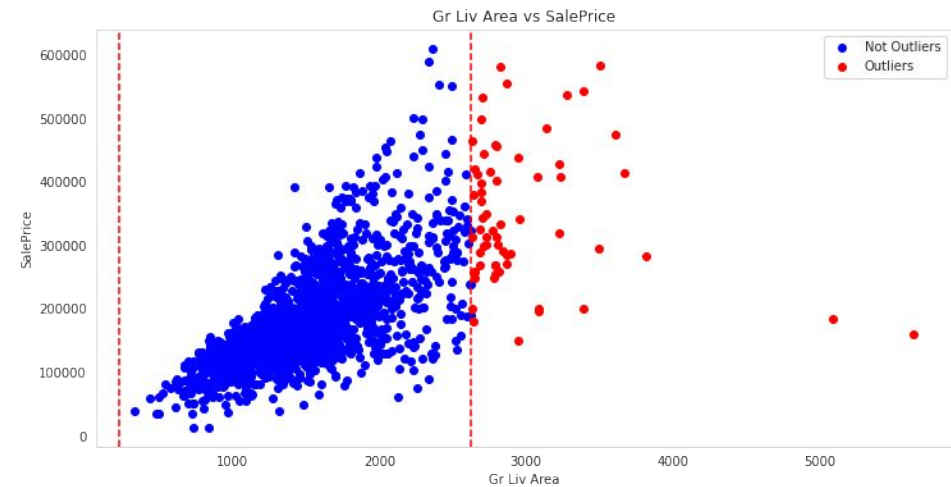
- Impute true missing values with mean for Lot Frontage
- Replace NaN with 0 or "NA" for the rest

|                | missing_count | missing_percent |
|----------------|---------------|-----------------|
| Pool QC        | 2042          | 99.561190       |
| Misc Feature   | 1986          | 96.830814       |
| Alley          | 1911          | 93.174061       |
| Fence          | 1651          | 80.497318       |
| Fireplace Qu   | 1000          | 48.756704       |
| Lot Frontage   | 330           | 16.089712       |
| Garage Finish  | 114           | 5.558264        |
| Garage Qual    | 114           | 5.558264        |
| Garage Yr Blt  | 114           | 5.558264        |
| Garage Cond    | 114           | 5.558264        |
| Garage Type    | 113           | 5.509508        |
| Bsmt Exposure  | 58            | 2.827889        |
| BsmtFin Type 2 | 56            | 2.730375        |

|                |    |          |
|----------------|----|----------|
| BsmtFin Type 1 | 55 | 2.681619 |
| Bsmt Cond      | 55 | 2.681619 |
| Bsmt Qual      | 55 | 2.681619 |
| Mas Vnr Area   | 22 | 1.072647 |
| Mas Vnr Type   | 22 | 1.072647 |
| Bsmt Half Bath | 2  | 0.097513 |
| Bsmt Full Bath | 2  | 0.097513 |
| Garage Area    | 1  | 0.048757 |
| Total Bsmt SF  | 1  | 0.048757 |
| Bsmt Unf SF    | 1  | 0.048757 |
| BsmtFin SF 2   | 1  | 0.048757 |
| BsmtFin SF 1   | 1  | 0.048757 |
| Garage Cars    | 1  | 0.048757 |

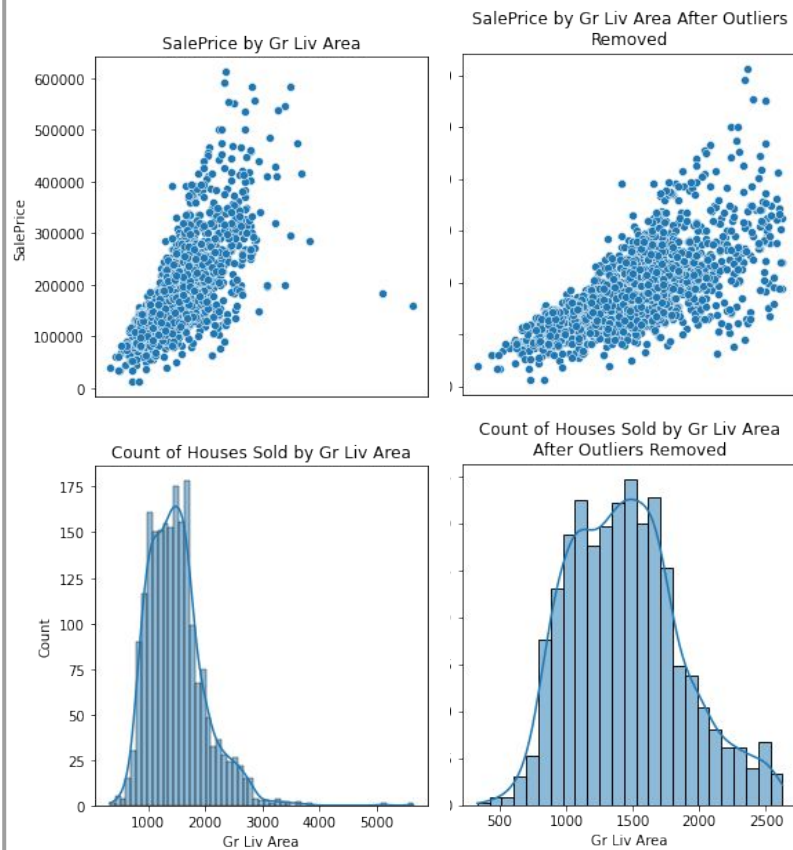
## Outliers

- Removed according to Inter-Quartile Range (IQR)

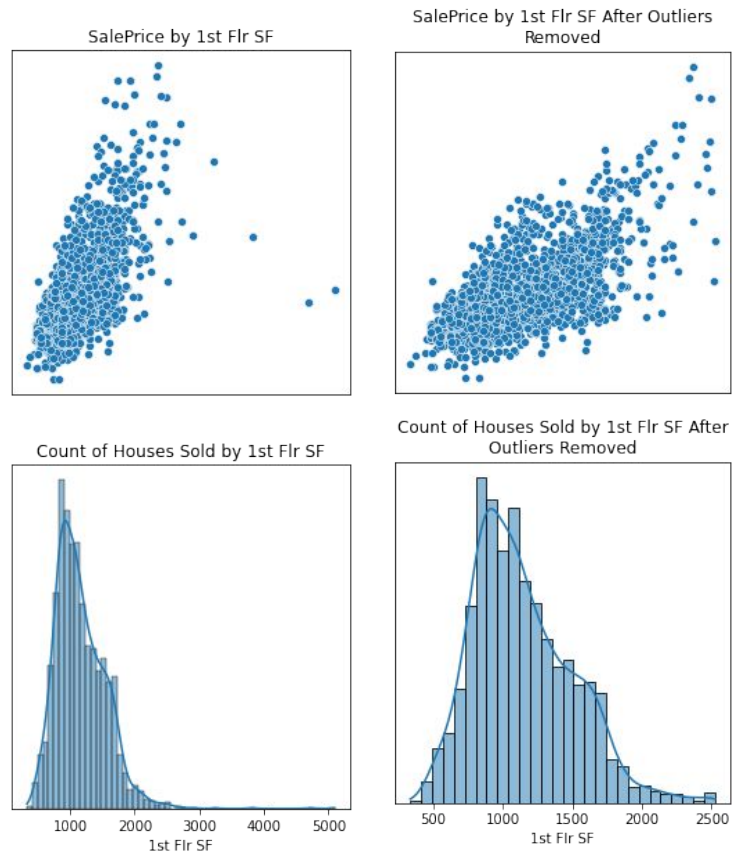


# DATA CLEANING

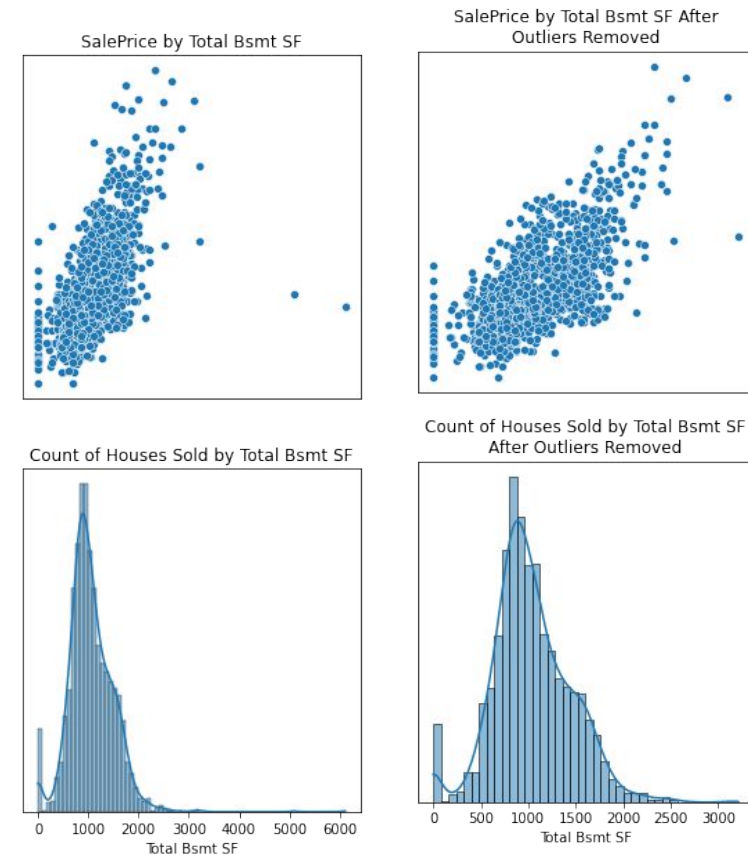
## Gr Liv Area



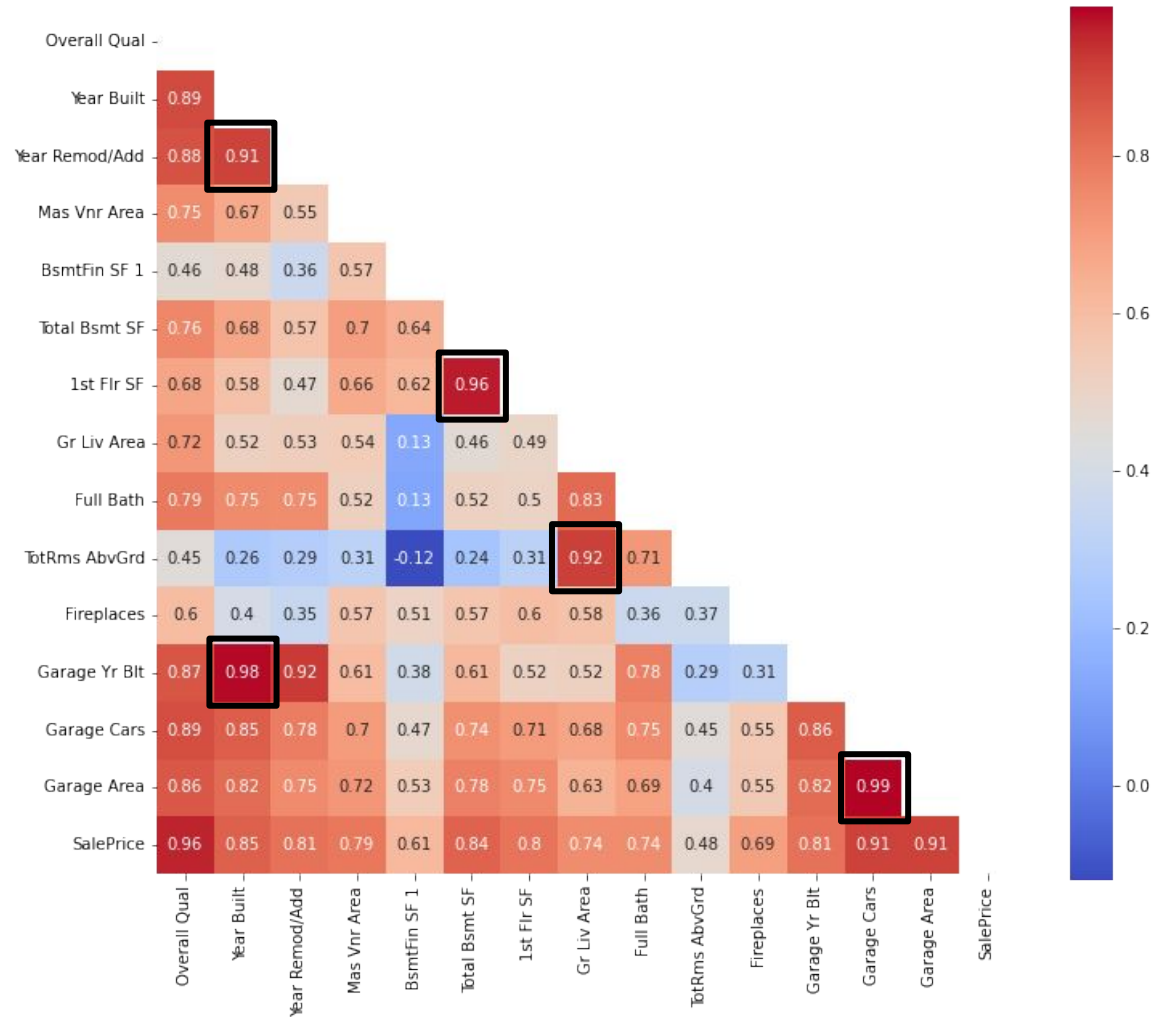
## 1st Flr SF



## Total Bsmt SF

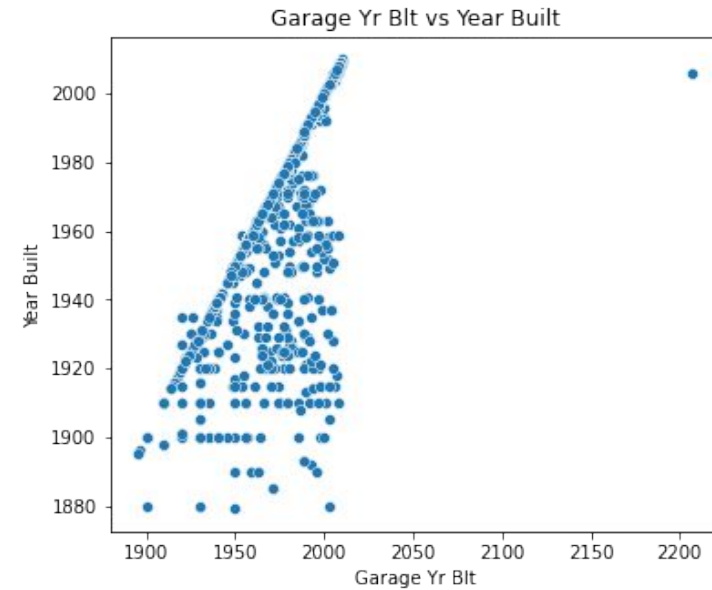
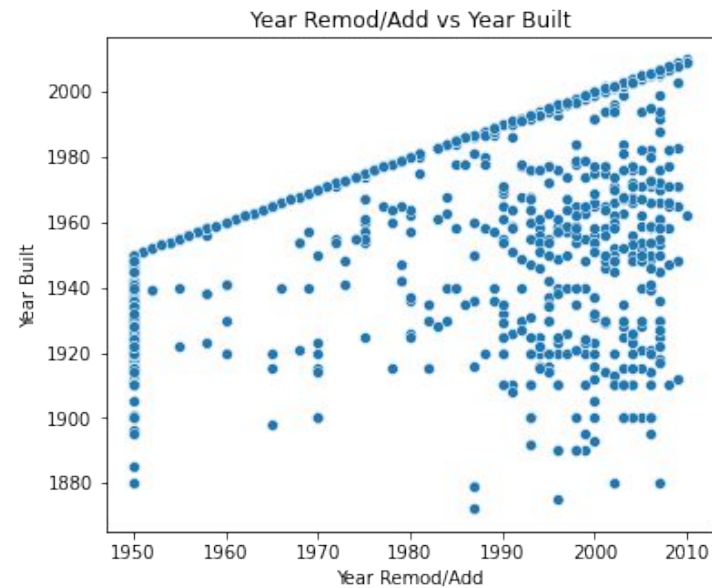
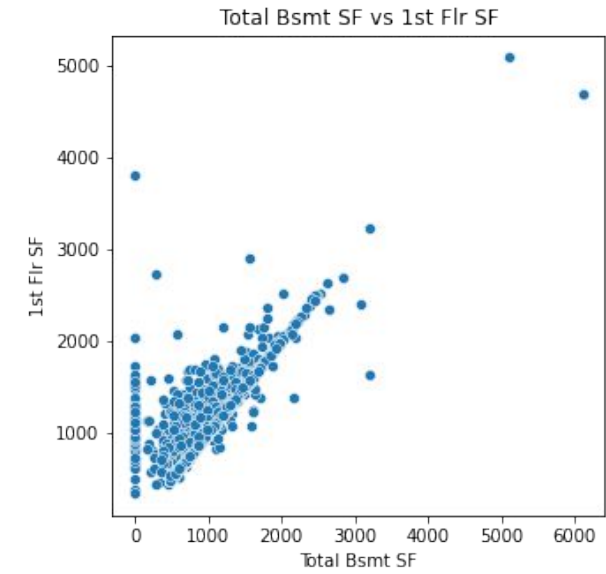
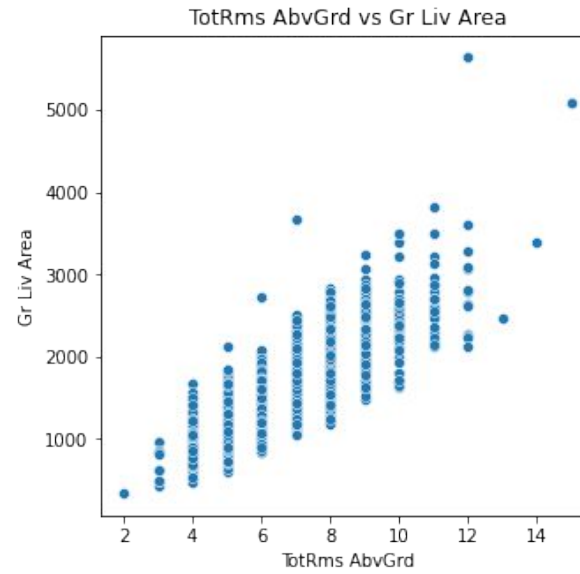
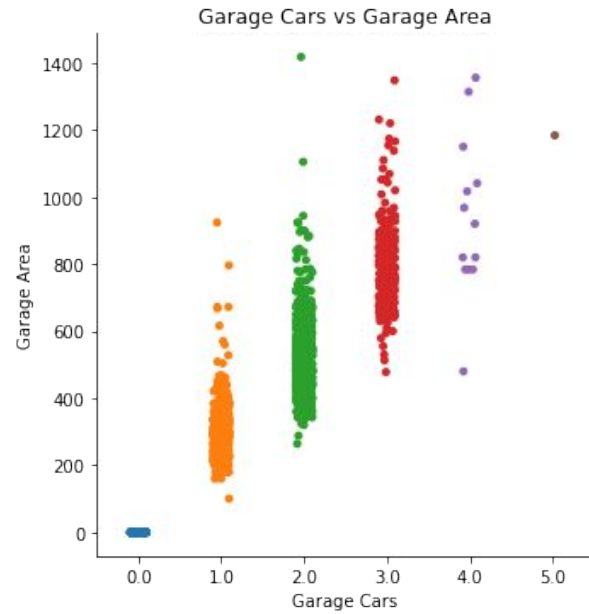


# FEATURE ENGINEERING AND SELECTION - MULTICOLLINEARITY



- Garage Cars vs Garage Area
- Total Bsmt SF vs 1st Flr SF
- Gr Liv Area vs TotRms AbvGrd
- Year Built vs Year Remod/Add
- Year Built vs Garage Yr Blt

# FEATURE ENGINEERING AND SELECTION - MULTICOLLINEARITY



## FEATURE ENGINEERING AND SELECTION

- Transform ordinal features to numerical features
- Create new features
- Drop features that add up to existing features

| New Features Created/<br>Existing features | Features Dropped   |
|--|--|
| House_Age, Remod_Age                       | Yr Sold, Year Remod/Add, Garage Yr Blt                   |
| Total_BathRooms                            | Bsmt Half Bath, Bsmt Full Bath, Full Bath, Half Bath     |
| Total_PorchSF                              | Open Porch SF, Enclosed Porch, 3-Ssn Porch, Screen Porch |
| Gr Liv Area                                | 1st Flr SF, 2nd Flr SF, Low Qual Fin SF                  |
| Total Bsmt SF                              | Bsmt Fin SF 1, Bsmt Fin SF 2, Bsmt Unf SF                |
| Garage Area                                | Garage Cars  |



# PREPROCESSING AND MODELLING

## Features Transformation

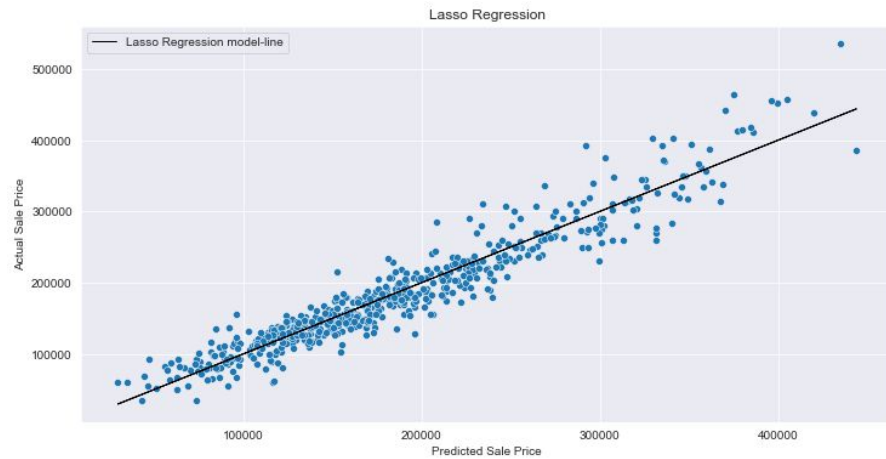
|                     |                                |
|---------------------|--------------------------------|
| Categorical Feature | One-hot encoding               |
| Numerical Feature   | StandardScaler vs MinMaxScaler |

## Comparing Mean Squared Error Generalisation

|                            | StandardScaler | MinMaxScaler |
|----------------------------|----------------|--------------|
| <b>Linear</b>              | -1.46e+25%     | -1.52e+22%   |
| <b>Ridge (alpha = 100)</b> | 10.53%         | 13.63%       |
| <b>Lasso (alpha = 100)</b> | 5.16%          | 2.44%        |
| <b>Enet</b>                | -6.93%         | -8.23%       |

# Model Performance

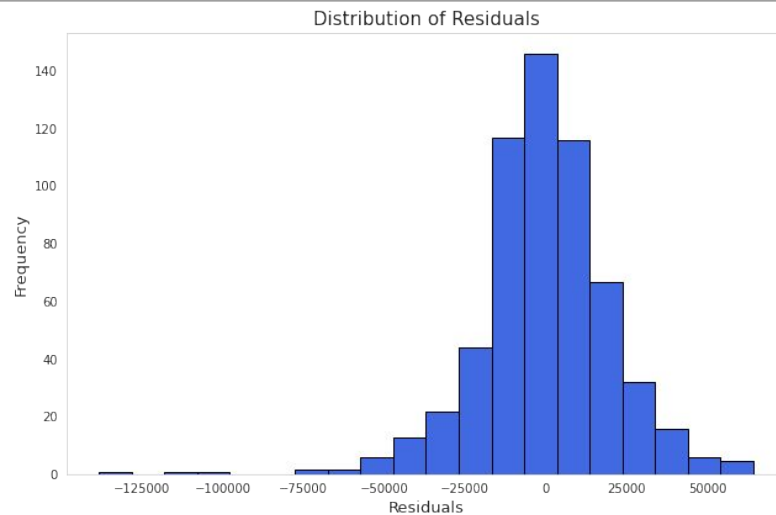
## Linearity



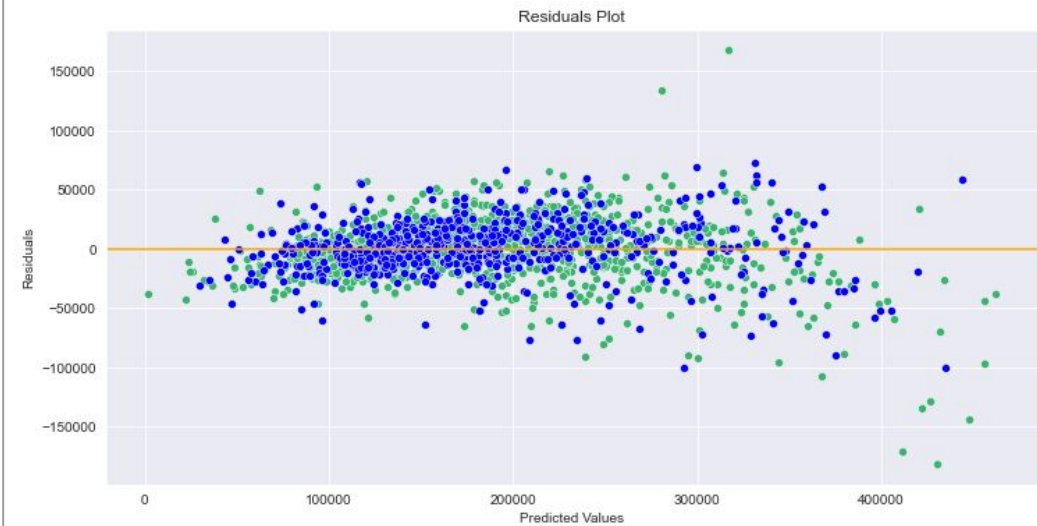
## Independence

Observations are independent of one another

## Normality

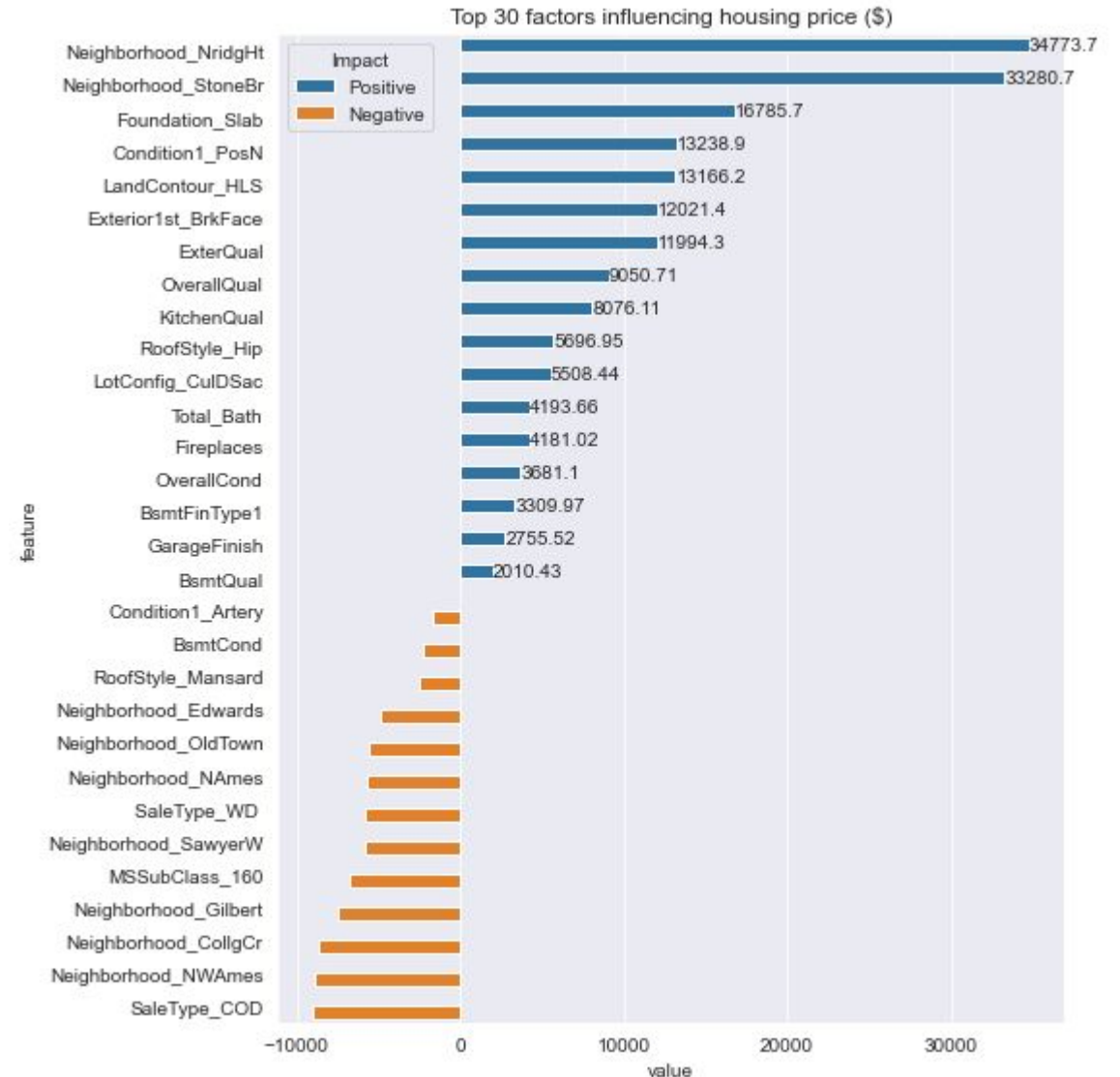


## Equal Variance/Homoscedasticity



# RECOMMENDATIONS

| Recommendations            |   |
|----------------------------|---|
| Neighbourhoods             | • Stone Brooke • Northridge Heights   |
| Home improvement projects  | <ul style="list-style-type: none"> <li>• Improve the Overall quality, External quality and Kitchen quality (+ \$ 8000/unit)</li> <li>• Increase the no. of bathrooms (+ \$ 4100/unit)</li> <li>• Installation of Fireplace if not available (+ \$ 4100/unit)</li> <li>• Improve the finishing of the garage (+ \$ 2700/unit)</li> </ul> |
| Limitation                 |   |
| Multicollinearity          | Collinearity between features are not eliminated totally (ex. Exter Qual, Kitchen Qual, Overall Qual)   |
| Limited timeframe          | Inability to capture changes in house price due to external factors such as the current economy   |
| Predicted Sale Price range | Predict up to \$300,000, which accuracy will decrease thereafter  |
| Historical Data            | Retraining of model required for better performance of current house prices due to inflation  |



**Thank you**