PROJECT 2
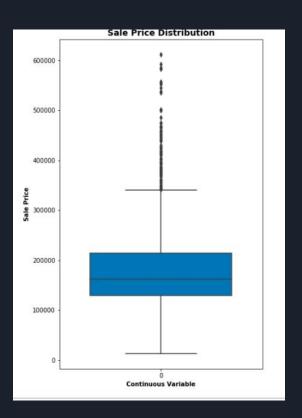
Predict price of homes - acmes iowa

PROBLEM STATEMENT

As a data scientist in Property Inc, our company is looking to develop residential homes in Aimes. It is important to forecast where and what kind of details to put into our development. The company is not just looking to build quality hoems, they also want to find out the variablss that make the house profitable. In order to do that, we have the housing data of Aimes Iowa.
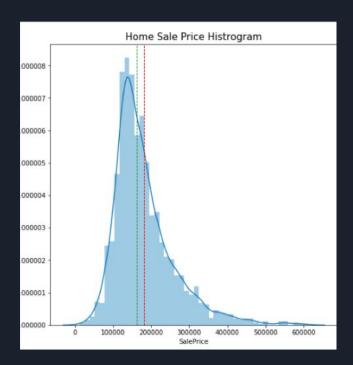
SALE PRICE ANALYSIS

We can see most house prices are priced below 4 00 000. There are quite a number of houses priced outside the main price range.

Sale Price Distribution

SALE PRICE ANALYSIS

Home prices seem to be right skewed with most house prices below 200000.
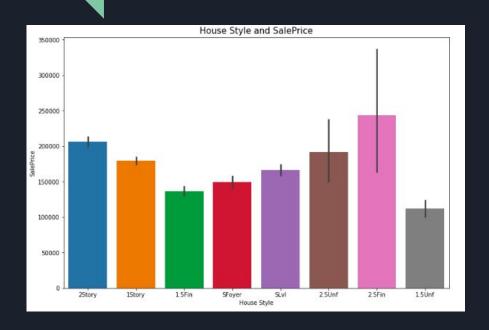
Features that have close relationship with sale price

Overall Qual	0.800975
Exter Qual	0.712880
Gr Liv Area	0.698046
Kitchen Qual	0.692190
Garage Area	0.650000
Garage Cars	0.648103
Total Bsmt SF	0.631507
1st Flr SF	0.621181
Bsmt Qual	0.617419
Year Built	0.572256
Year Remod/Add	0.550055
Garage Finish	0.549081
Full Bath	0.537369
Mas Vnr Area	0.504025
TotRms AbvGrd	0.502909

The numbers show how related the features are to the Sale Price. 1 means totally related.
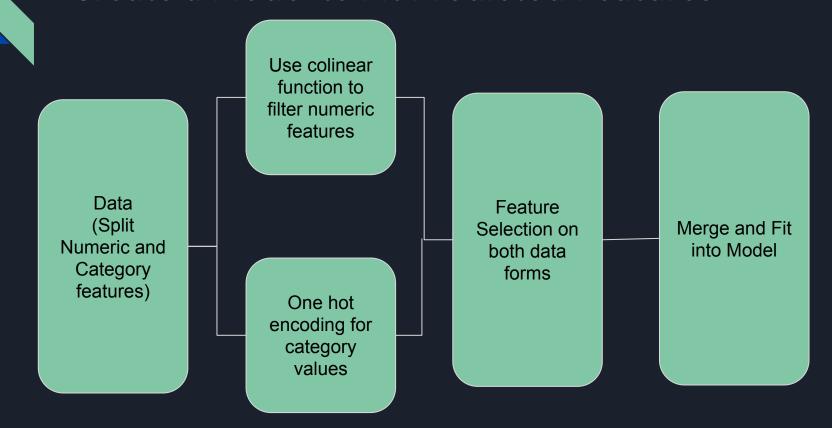
Category Features that Affect Sale Price

From the plots we can see these features affect Sale Price.

< House Style, Neighbourhood, Exterior, Foundation

Create a Model to Fit Predicted Features

```
┌─────────────┐          ┌─────────────────┐          ┌─────────────┐          ┌─────────────┐
│    Data     │     ┌────│  Use colinear   │────┐     │   Feature   │          │ Merge and   │
│   (Split    │     │    │  function to    │    │     │ Selection on│          │ Fit into    │
│  Numeric and│─────┤    │ filter numeric  │    ├─────│  both data  │──────────│   Model     │
│  Category   │     │    │    features     │    │     │    forms    │          │             │
│  features)  │     │    └─────────────────┘    │     │             │          │             │
│             │     │    ┌─────────────────┐    │     │             │          │             │
│             │     │    │    One hot      │    │     │             │          │             │
│             │     └────│  encoding for   │────┘     │             │          │             │
│             │          │   category      │          │             │          │             │
│             │          │    values       │          │             │          │             │
└─────────────┘          └─────────────────┘          └─────────────┘          └─────────────┘
```

Features Extracted From Ridge Regularization

Neighbourhoods and Quality of homes are many of the key features extracted.

Conclusion

After analysis, we can see that location or neighbourhood makes a significant increase in Sale Price. Not only does it affect price positively, some neighbourhood affects negatively. After location, quality of the house and other qualities like exterior and basement quality also affects the price. Garage also is a factor.

To do well, finding the right neighbourhood is crucial with a strong focus on the quality of the house will be a key factor. Another thing to consider is also the kitchen quality.