House Price Prediction

Prathik Chukkapalli

Purpose and Problem Statement

- I am a financial analyst at a real-estate investment firm.
- Our firm has a lot of real-estate in Ames, Iowa.
- My next task is to create a mode using a dataset that was given to me to predict the sales price of the real-estate owned by the firm.
- I have a training data set that will be used to create an accurate model, and a testing data set to test my model on.

Data Dictionary

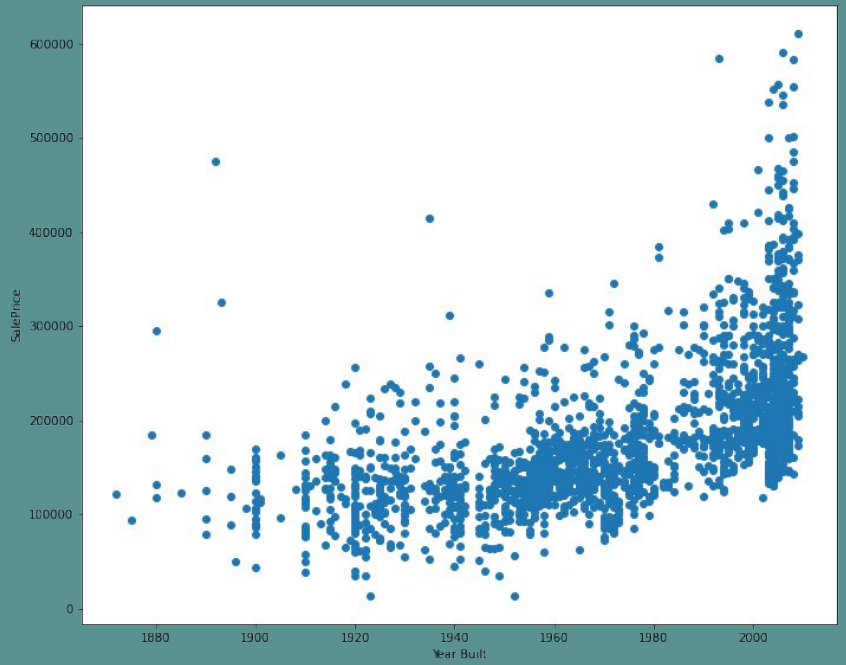
- Link to Data Dictionary

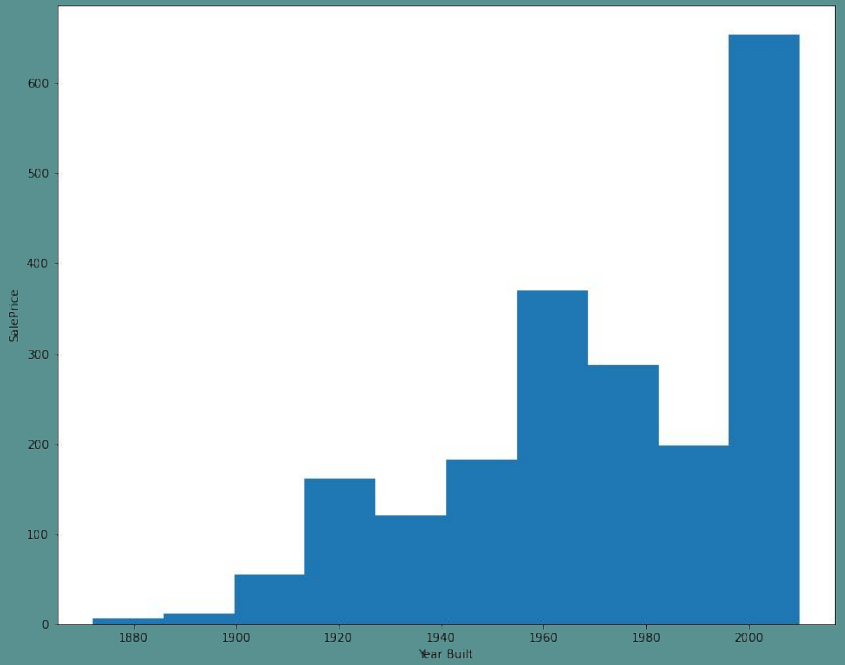
- ['Total Bsmt SF', '1st Flr SF','2nd Flr SF', 'Wood Deck SF', 'Lot Area', 'Gr Liv Area', 'Garage Area', 'Pool Area', 'Overall Qual', 'Overall Cond', 'Garage Cars', 'Year Remod/Add', 'Year Built', 'Sale Type_CWD', 'Sale Type_Con', 'Sale Type_ConLD', 'Sale Type_ConLI', 'Sale Type_ConLw', 'Sale Type_New', 'Sale Type_Oth', 'Sale Type_WD ', 'Garage Type_Attchd', 'Garage Type_Basment', 'Garage Type_BuiltIn', 'Garage Type_CarPort', 'Garage Type_Detchd', 'Kitchen Qual_Fa', 'Kitchen Qual_Gd', 'Kitchen Qual_TA']

Analysis

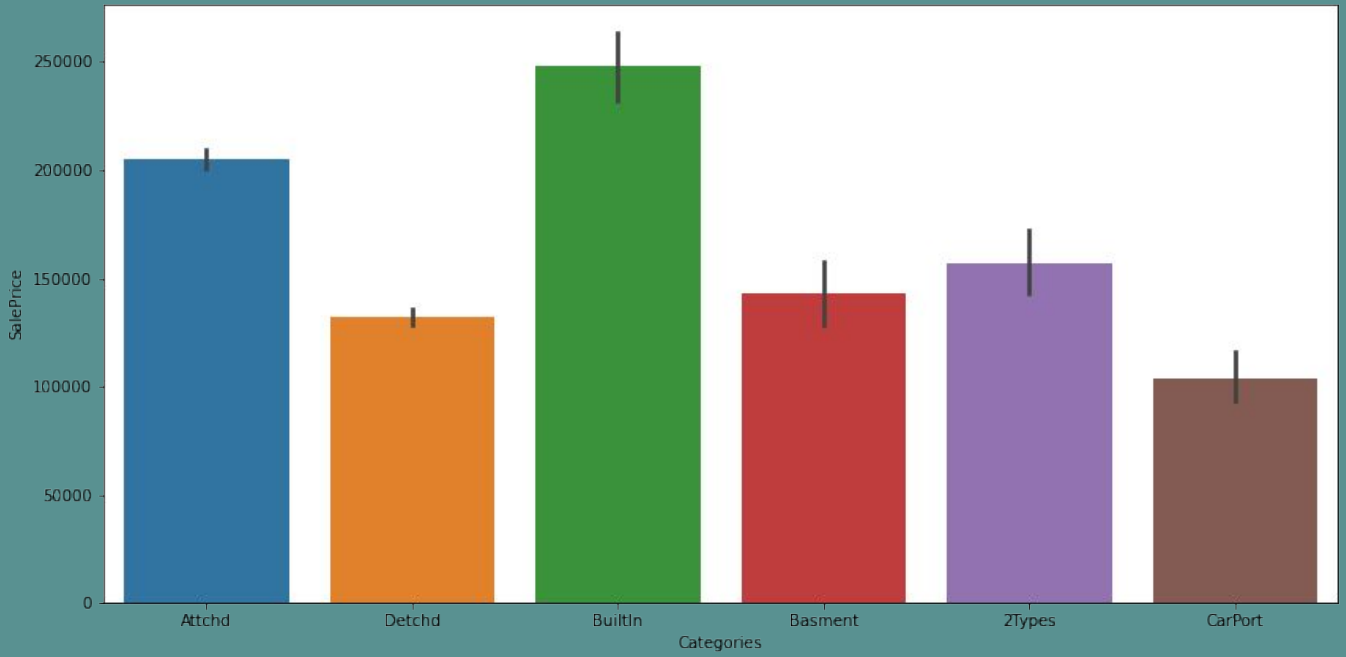
- Heat Map
  - Overall Qual
- Categorical
  - Bar Graph
- Numerical
  - Scatterplot, Histogram

Plots


SalePrice
Year Built
0
100000
200000
300000
400000
500000
600000
1880
1900
1920
1940
1960
1980
2000


SalePrice
Year Built
0
100
200
300
400
500
600
1880
1900
1920
1940
1960
1980
2000

Plots


250000
200000
150000
100000
50000
0
SalePrice
Attchd
Detchd
BuiltIn
Basment
2Types
CarPort
Categories

Models

- Linear Regression with standard scaler transformation, Ridge Models, and Lasso Models
  - r2 score was lower than other models, the model had the best scores when looking at cross_val_score, and the R2 regression score.
- Training Score = 0.824
- Testing Score = 0.834
- Cross Val Score = 0.810
- R2 regression = 0.834

Conclusion

- For the data-set given it is possible to create a good model that can predict the sales-price of houses in Ames, Iowa.
- Based on the analysis of the different types of models, the model that could best use the data given to predict sales-prices of houses in Ames, Iowa is a linear regression model that included standard_scaler.
- Make Model More Accurate