Predicting House Prices

A Linear Regression Project

Presentation by:
Jurgen Arias
Data Scientist

IOWA HOUSE
Ames, Iowa

Cleaning data for best predictions

Get familiar with the data

Understand what the features mean and have an idea of which ones are going to be more valuable

Think about the relationship of features to the price

Cleaning, cleaning, cleaning

Fill empty values

Combine features

Drop already combined features

Check statistics and drop outliers

Change categorical values to numerical ordered values.

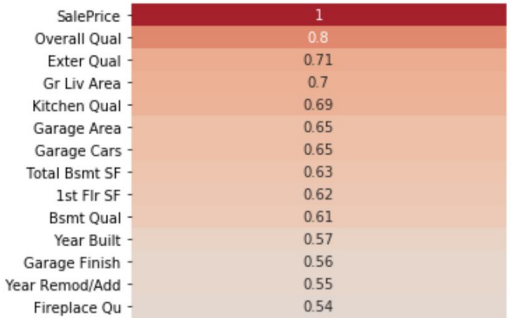
Feature Engineering

Get familiar with the correlations of the features with price

Use transformations, polynomial, log, standard scale

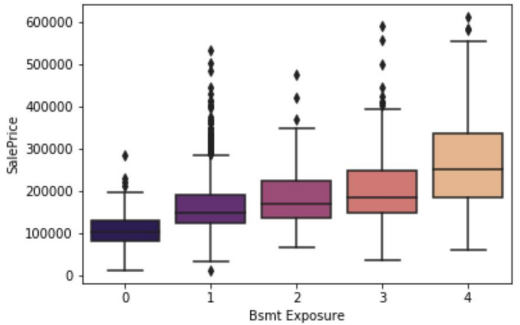
Change categorical features to dummies

Visualizations

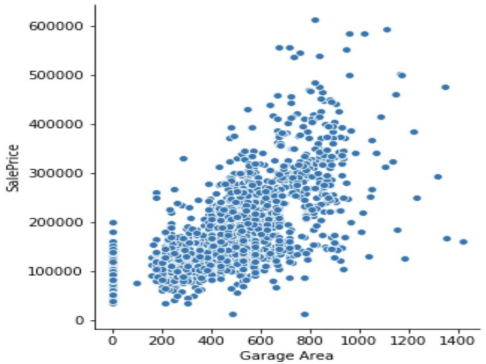


Heatmaps for correlations

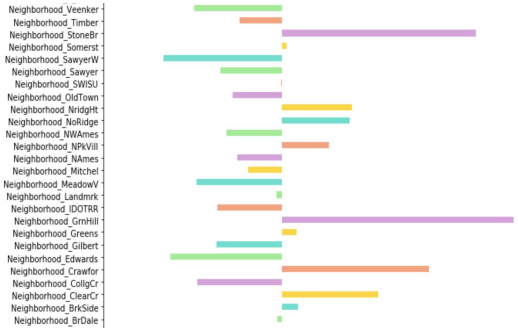
Boxplots for categorical data



Scatterplots for numerical data



Bar charts and histograms for coefficients and distributions



Modeling

Set your features dataframe and target vector

Apply train test split to evaluate your predictions

Decide what models to use and fit your training data

Score your models on your test data

Check coefficients

Decide which model works better for your data and predictions

Try Linear Regression, Lasso, LassoCV, Ridge and RidgeCV and score your models

Go back and do more cleaning, modify features, change your feature engineering, undo some of what you did before, do some new stuff

Try your models again

Repeat over and over again

Ridge

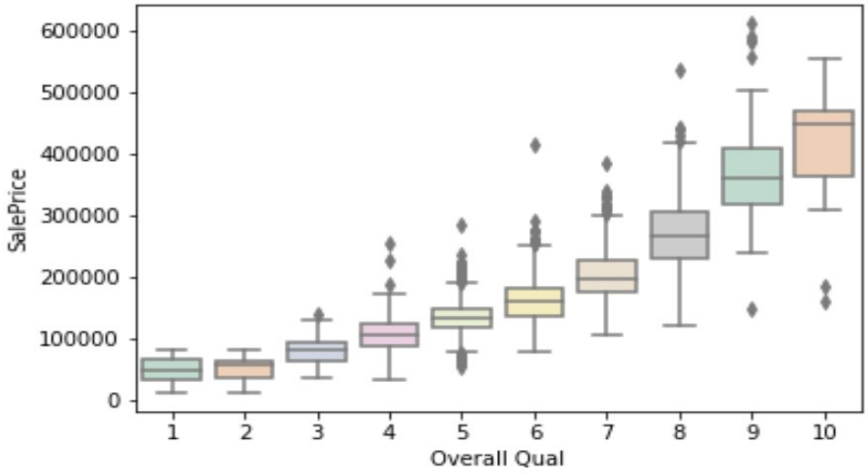
Ridge scores

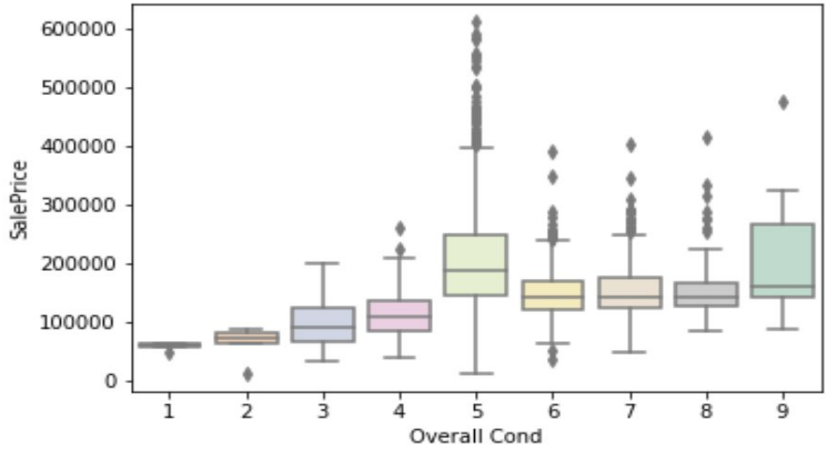
Ridge score on train data: **0.952**
Ridge score on test data: **0.942**
Ridge cross val score: **0.918**

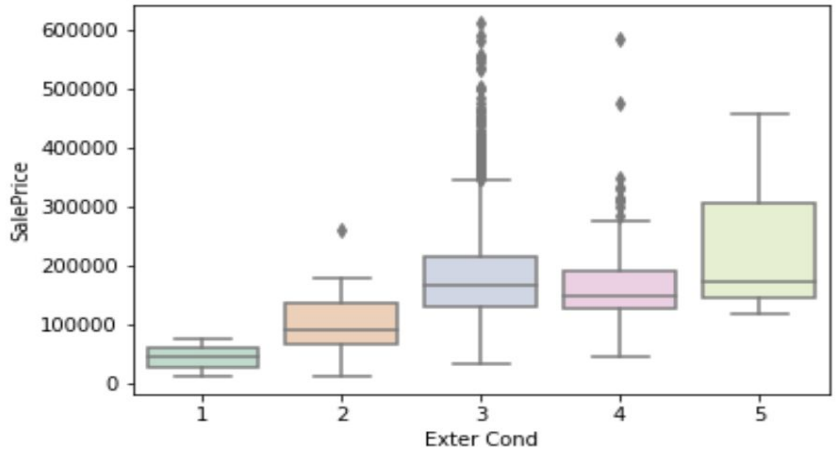
Ridge Root Mean Squared Error: **18879.9**

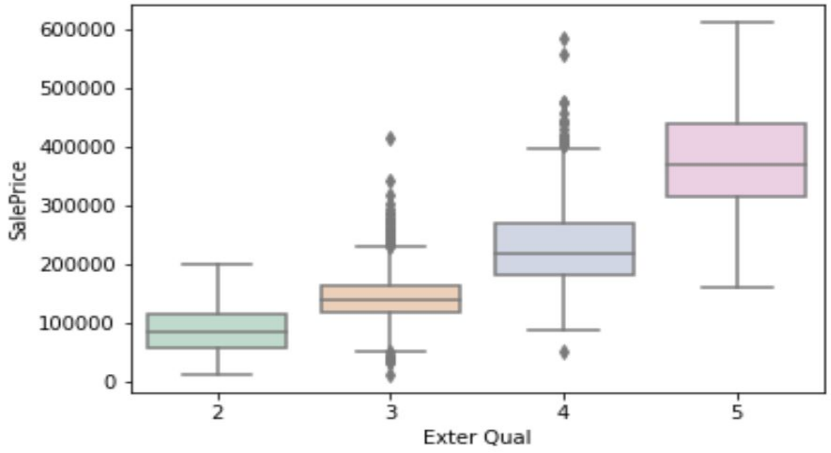
Coefficient of determination R^2
(proportion of the variance in the dependent variable that is predictable from the independent variable)
https://en.wikipedia.org/wiki/Coefficient_of_determination

Overall and Exterior Quality and Condition

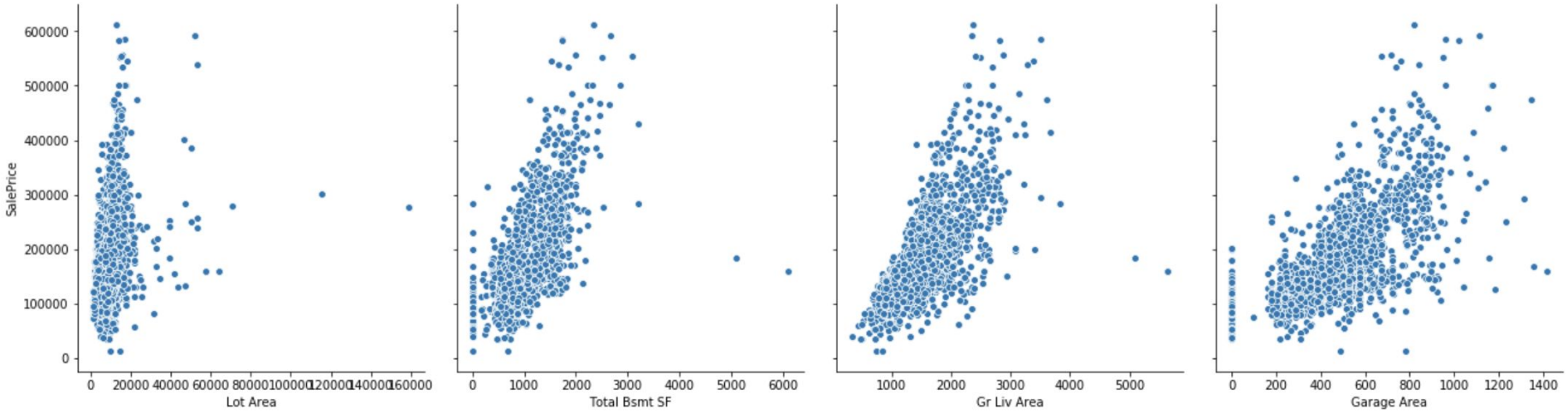




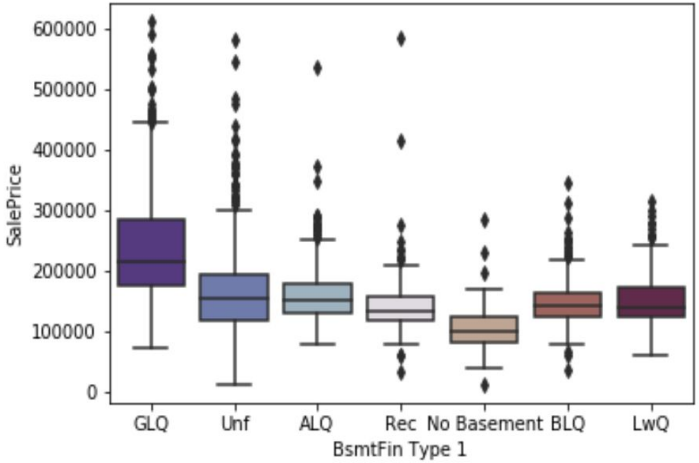


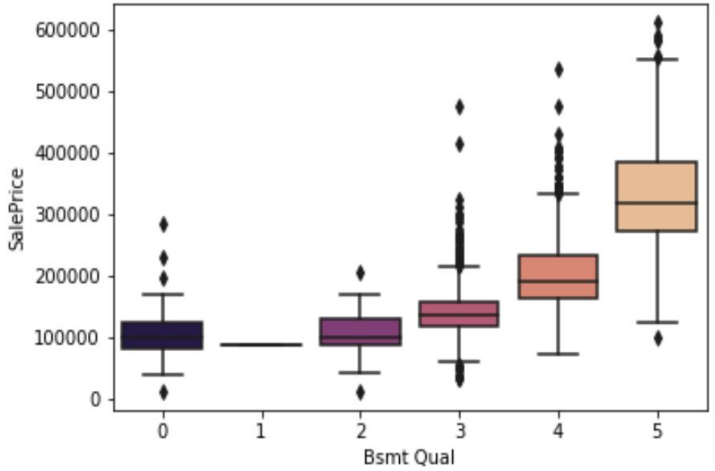


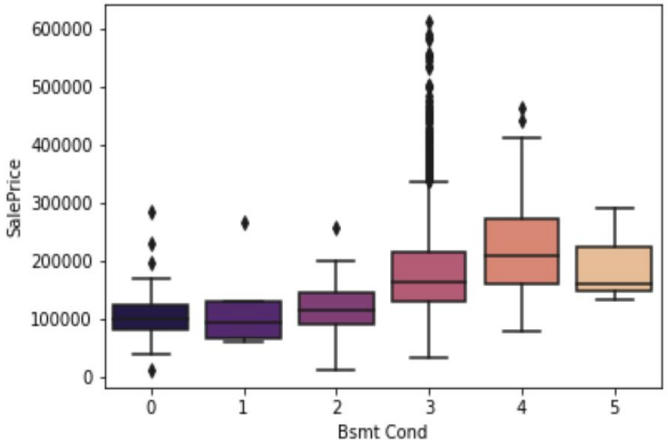
Area and Square Footage

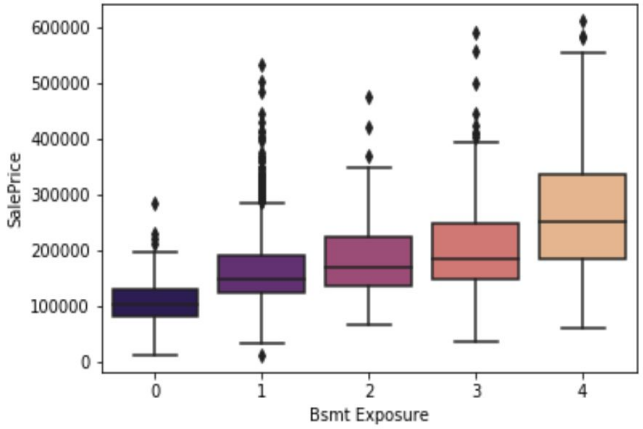


Basement

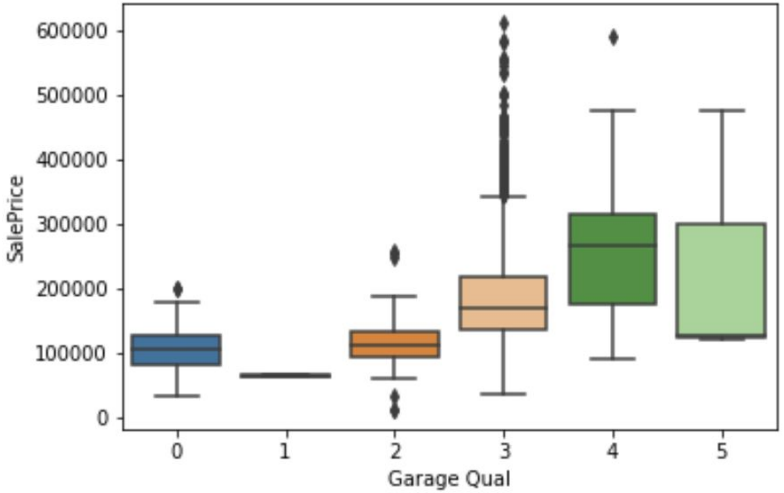

SalePrice
600000
500000
400000
300000
200000
100000
0
GLQ
Unf
ALQ
Rec
No Basement
BLQ
LwQ
BsmtFin Type 1

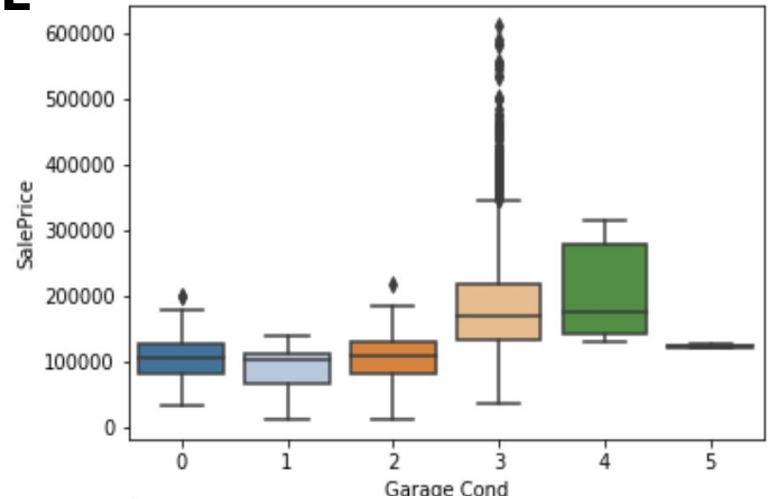

SalePrice
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
Bsmt Qual

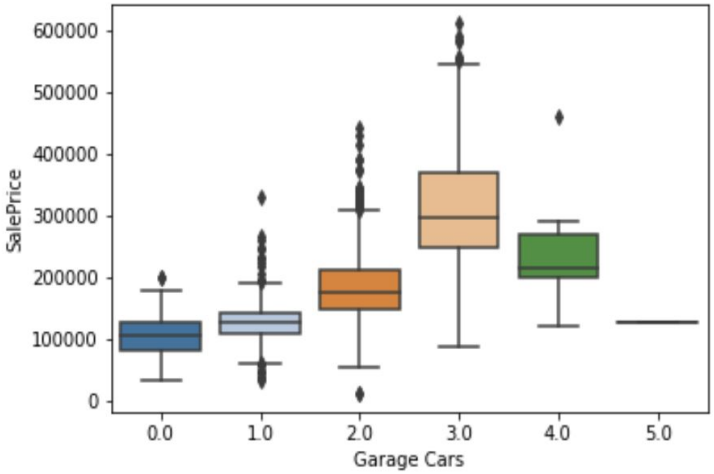

SalePrice
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
Bsmt Cond

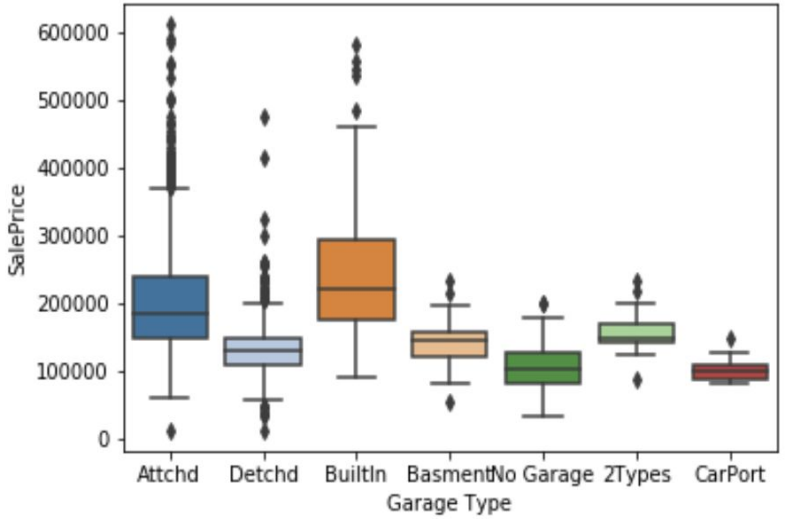

SalePrice
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
Bsmt Exposure

GARAGE

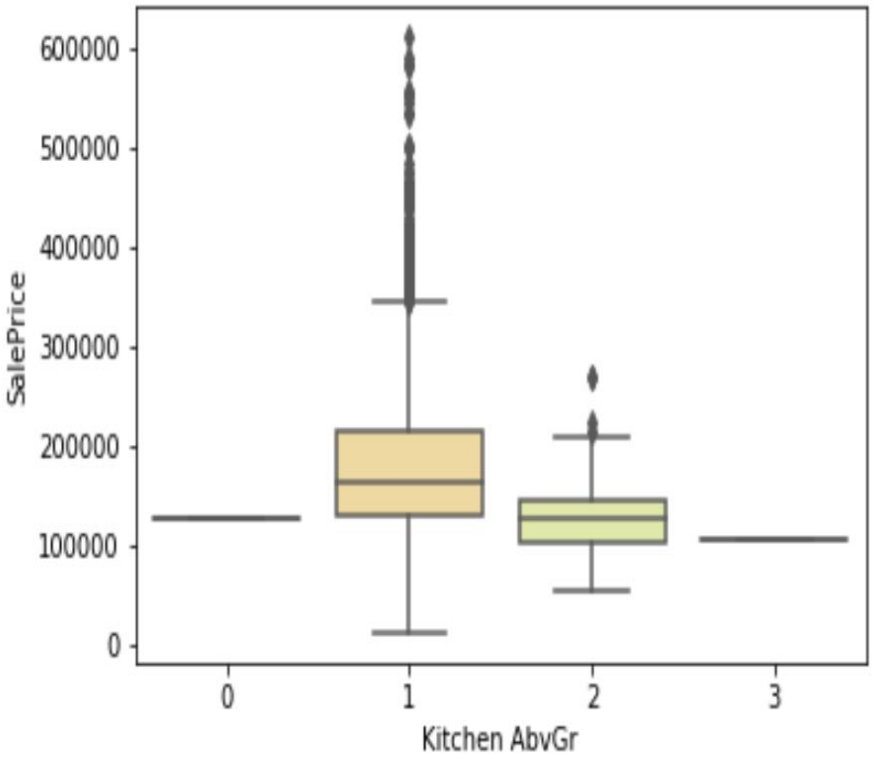

SalePrice
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
Garage Qual

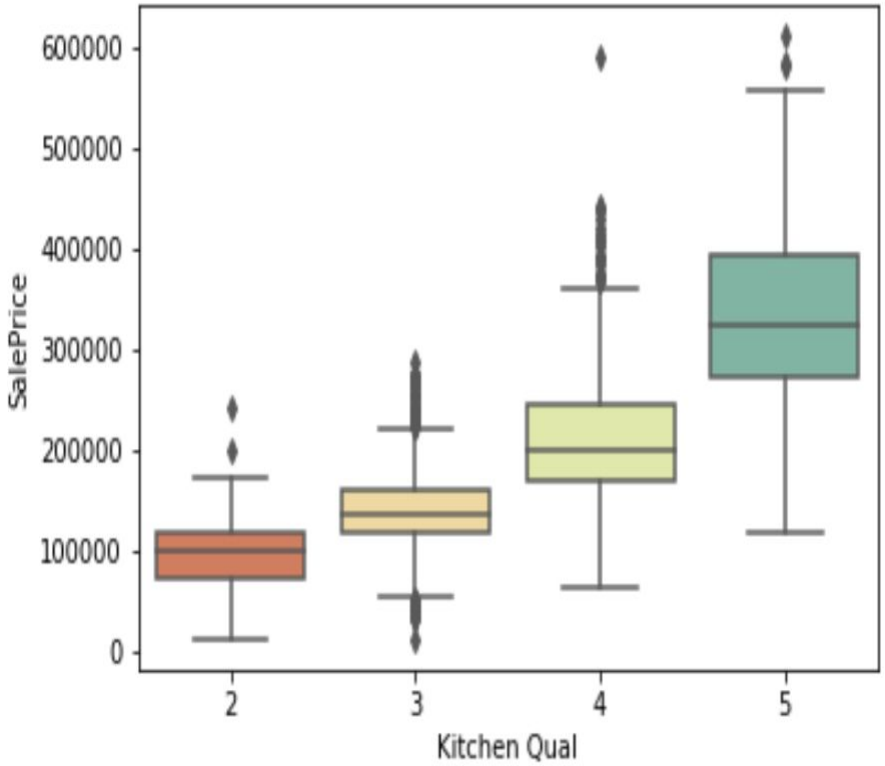

SalePrice
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
Garage Cond


SalePrice
600000
500000
400000
300000
200000
100000
0
0.0
1.0
2.0
3.0
4.0
5.0
Garage Cars

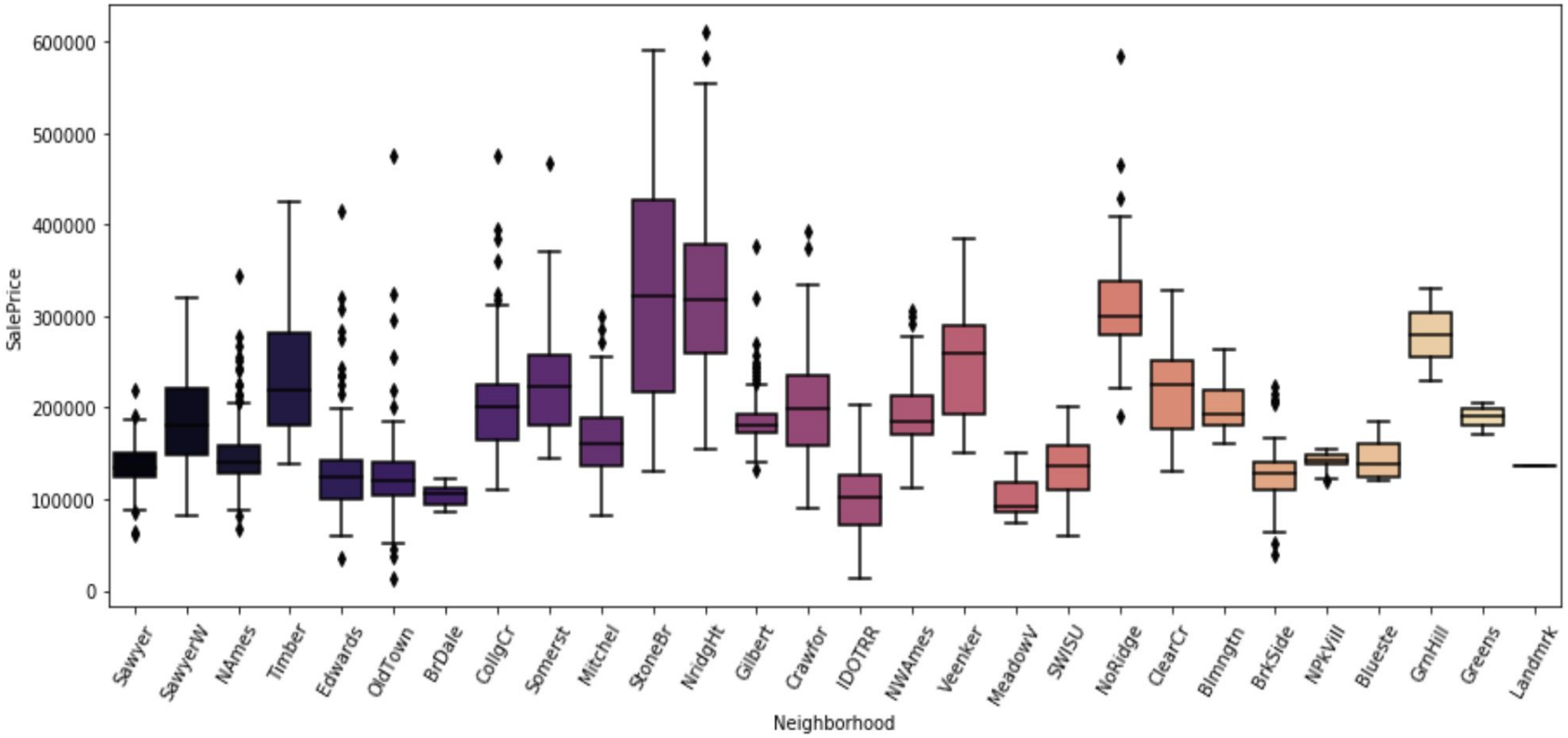

SalePrice
600000
500000
400000
300000
200000
100000
0
Attchd
Detchd
BuiltIn
Basment
No Garage
2Types
CarPort
Garage Type

Kitchen


SalePrice
600000
500000
400000
300000
200000
100000
0
0
1
2
3
Kitchen AbvGr


SalePrice
600000
500000
400000
300000
200000
100000
0
2
3
4
5
Kitchen Qual

Neighborhoods



Thanks