Ames House Prices

Problem Statement: Given a set of inputs, can we accurately predict the prices of a house?

Dataset description

- 2051 entries, 81 unique features.
- 2006 – 2010
- Lots of null, empty and repeated values.

Model

The model predicts home prices with **84% accuracy** and with an error of

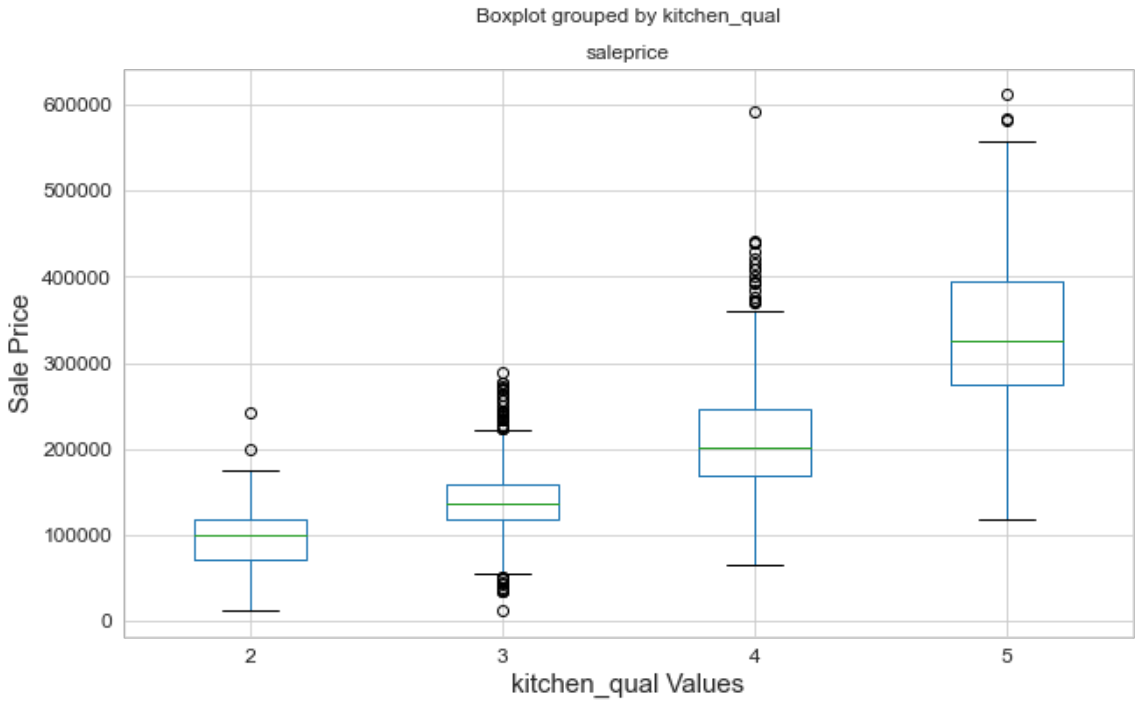
±$32,000

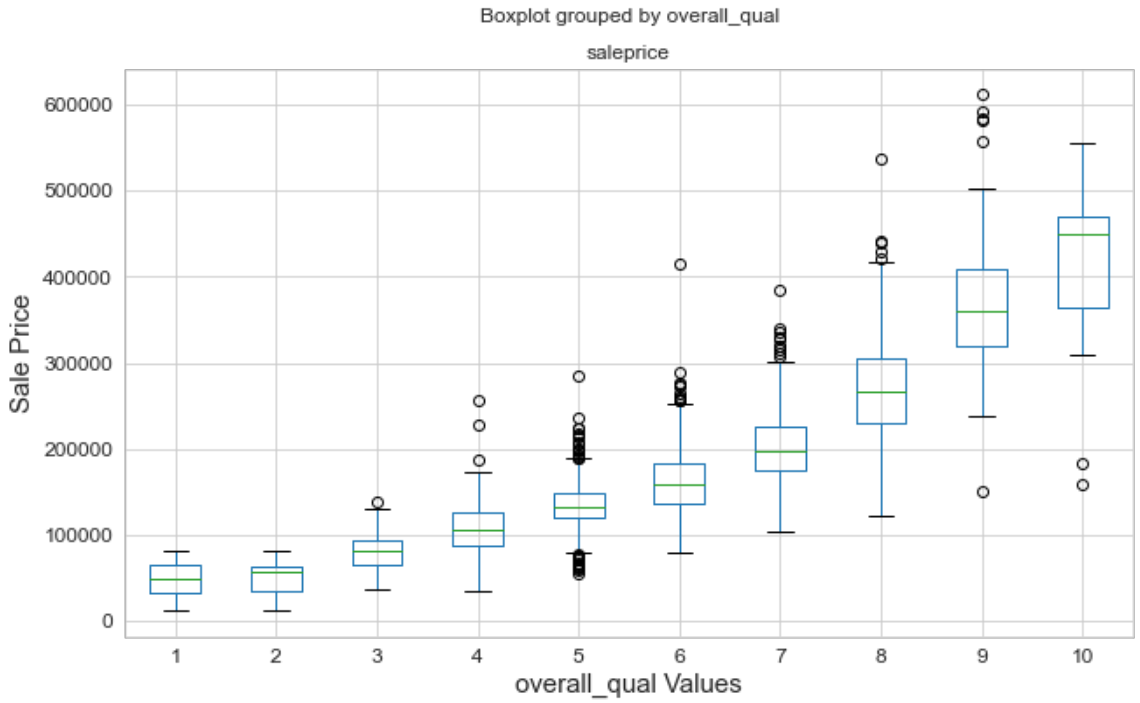
- RMSE: $ 32,000 (ave root squared error between actual model prediction)
- R^2: 0.84 (Goodness of fit between actual and model, 1.0 being a perfect model)

Factors that affected the model

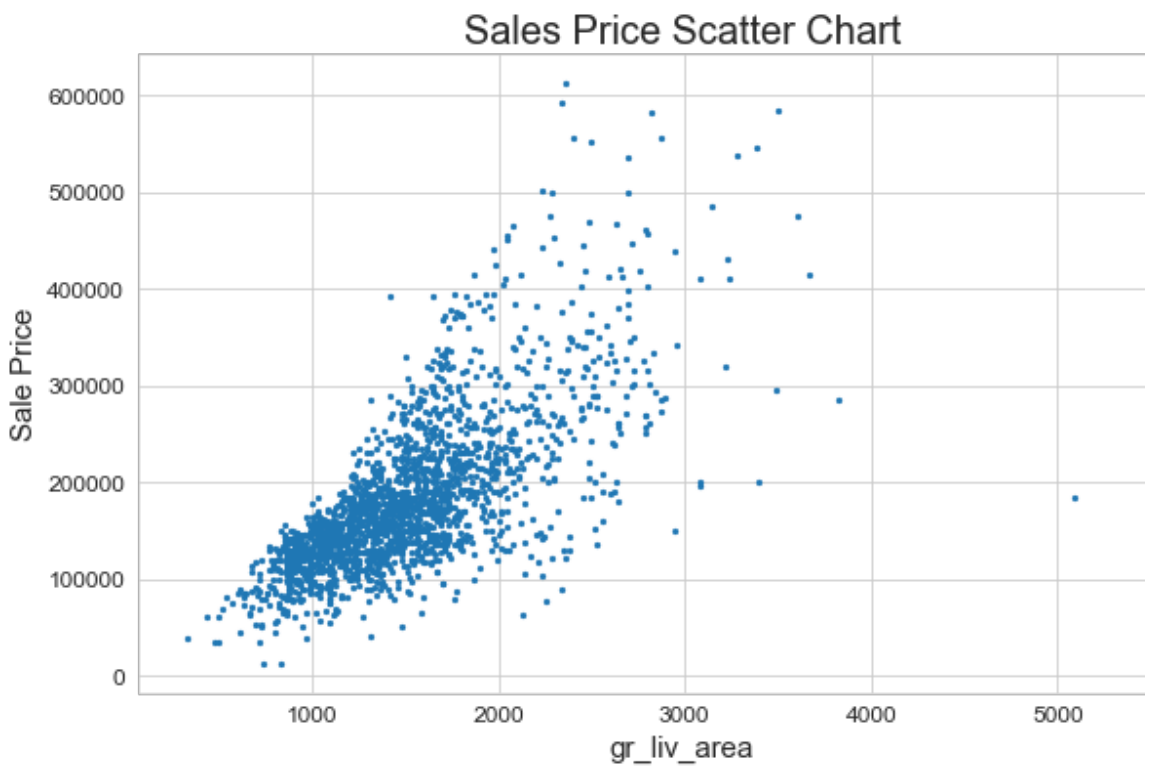
- Quality
- Size
- Location
- Age

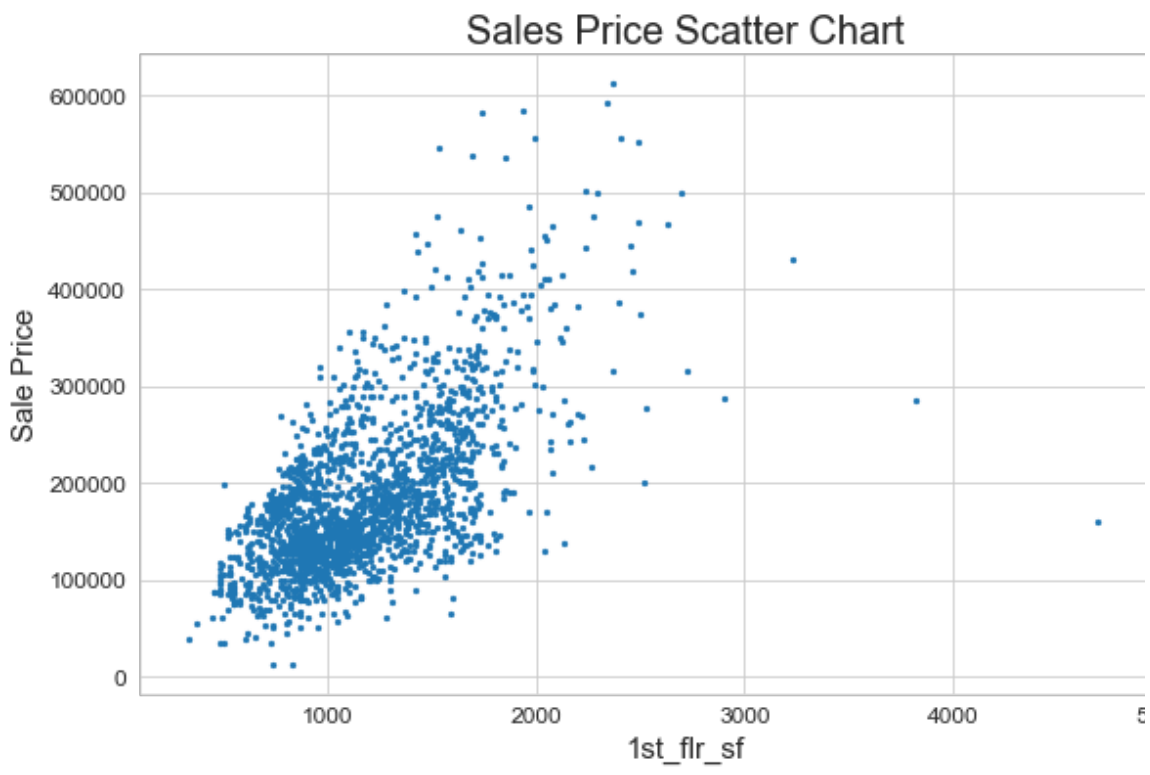
Quality

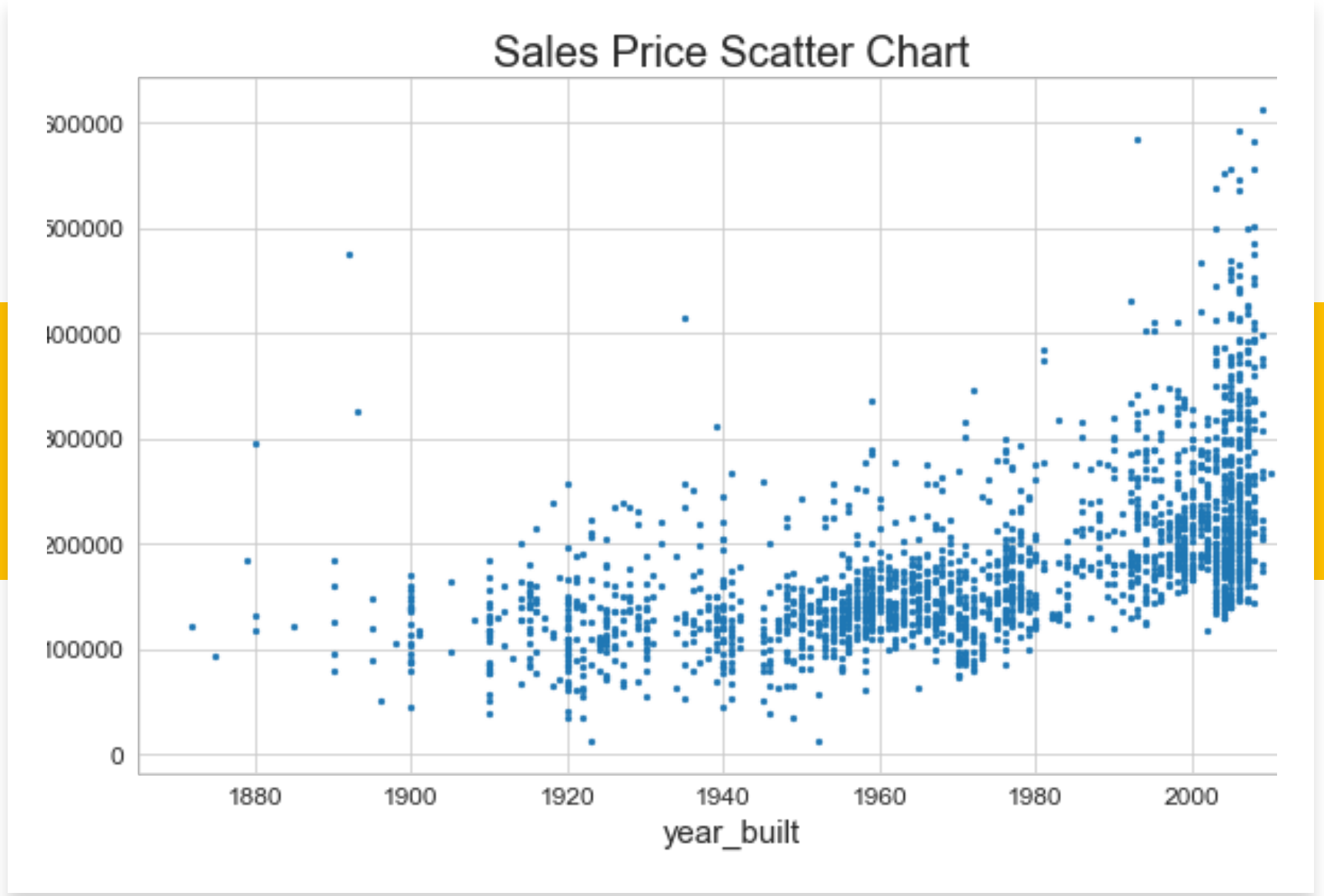

Boxplot grouped by kitchen_qual
saleprice
Sale Price
600000
500000
400000
300000
200000
100000
0
2
3
4
5
kitchen_qual Values


Boxplot grouped by overall_qual
saleprice
Sale Price
600000
500000
400000
300000
200000
100000
0
1
2
3
4
5
6
7
8
9
10
overall_qual Values

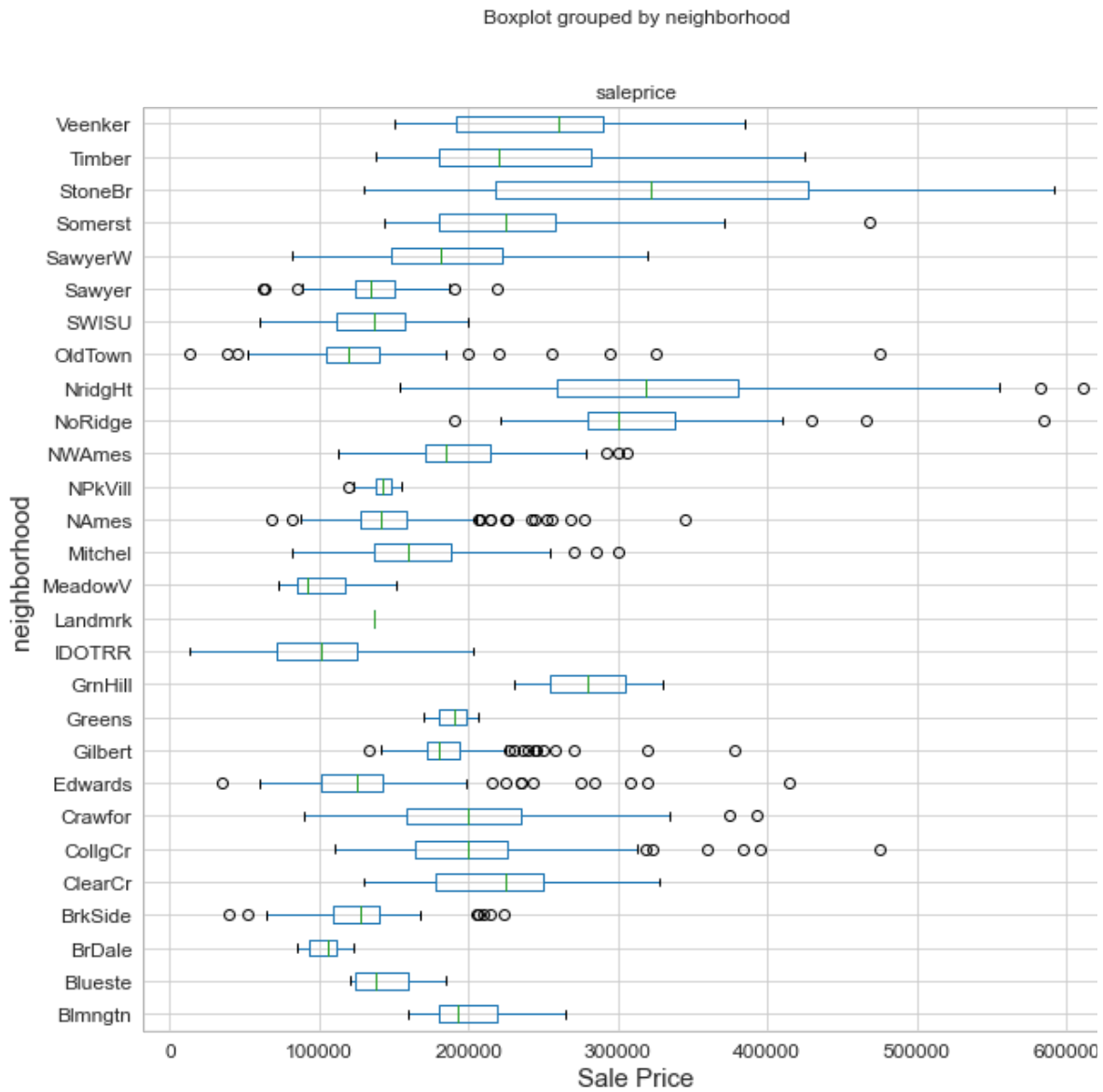
Size


Sales Price Scatter Chart
Sale Price
600000
500000
400000
300000
200000
100000
0
1000
2000
3000
4000
5000
gr_liv_area


Sales Price Scatter Chart
Sale Price
600000
500000
400000
300000
200000
100000
0
1000
2000
3000
4000
1st_flr_sf

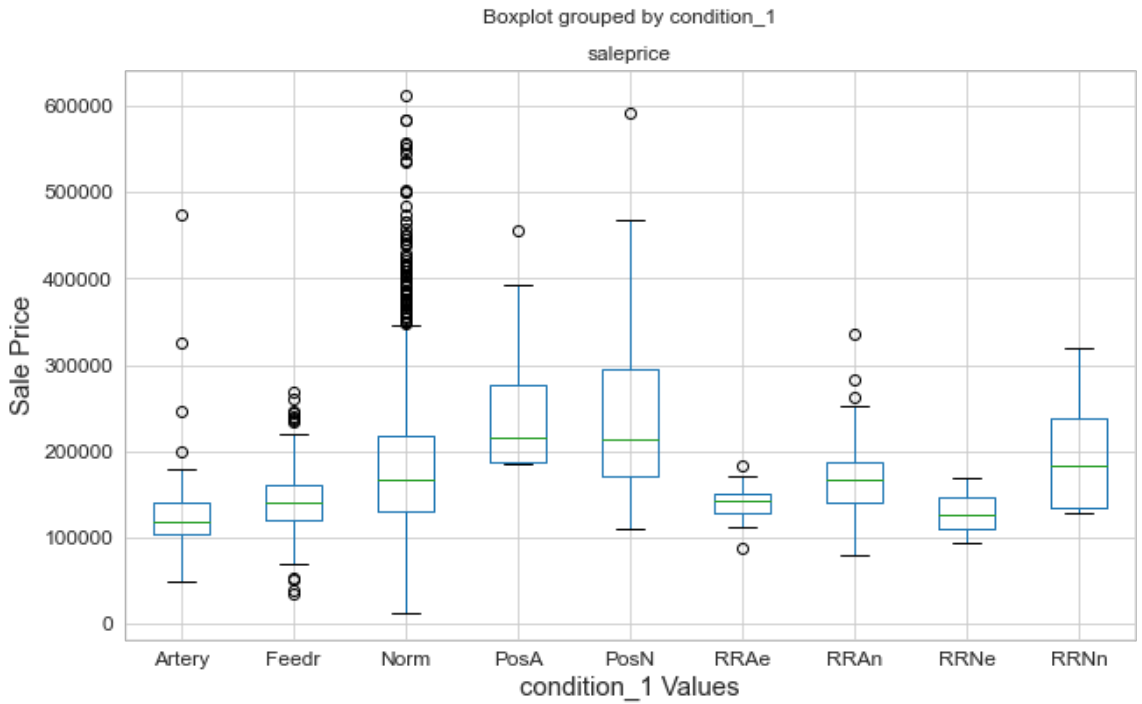

Sales Price Scatter Chart
600000
500000
400000
300000
200000
100000
0
1880
1900
1920
1940
1960
1980
2000
year_built

Age


Boxplot grouped by neighborhood
saleprice
Veenker
Timber
StoneBr
Somerst
SawyerW
Sawyer
SWISU
OldTown
NridgHt
NoRidge
NWAmes
NPkVill
NAmes
Mitchel
MeadowV
Landmrk
IDOTRR
GrnHill
Greens
Gilbert
Edwards
Crawfor
CollgCr
ClearCr
BrkSide
BrDale
Blueste
Blmngtn
neighborhood
0
100000
200000
300000
400000
500000
600000
Sale Price

Location

Location



Artery	Adjacent to arterial street
Feedr	Adjacent to feeder street
Norm	Normal
RRNn	Within 200' of North-South Railroad
RRAn	Adjacent to North-South Railroad
PosN	Near positive off-site feature--park, greenbelt, etc.
PosA	Adjacent to postive off-site feature
RRNe	Within 200' of East-West Railroad
RRAe	Adjacent to East-West Railroad

Who will find the model useful?

- Buyers : Will be able to figure out if they are overpaying or underpaying
- Sellers: Will know what to do to improve the value of their home
- Investors: Will be able to know what locations to invest in

Limitations of model

- Data is only 4 years long and out of date
- Factors such as inflation may not have been taken into account for over the years
- Model is only applicable to Ames, Iowa and cannot be generalized to other regions.

Thank you very much