

Assessing House Prices Using Linear Regression Model

DSI – SF

Emily Lin



Agenda

- Problem Statement
- Modeling Process
- Conclusions
- Recommendations and Limitation
- Questions and Comments



Problem Statement

- To find the most accurate linear regression model that assess Ames housing price by identifying the appropriate indicators.



Modeling Process



Data Cleaning



Exploratory
Data Analysis



Featuring
Engineering



Modeling



Metrics

Data Cleaning

Categorizing Ordinal Data

Exterior Quality

Exter Qual (Ordinal): Evaluates the quality of the material on the exterior Scale in dataframe		
Ex	Excellent	4
Gd	Good	3
TA	Average/Typical	2
Fa	Fair	1
Po	Poor	0

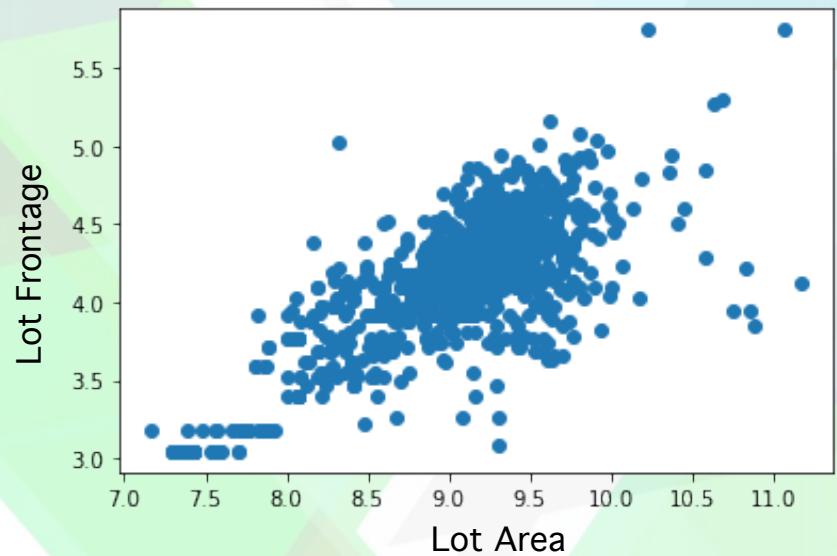
Filling in null values

BsmtFin Type 2 [Ordinal] 1

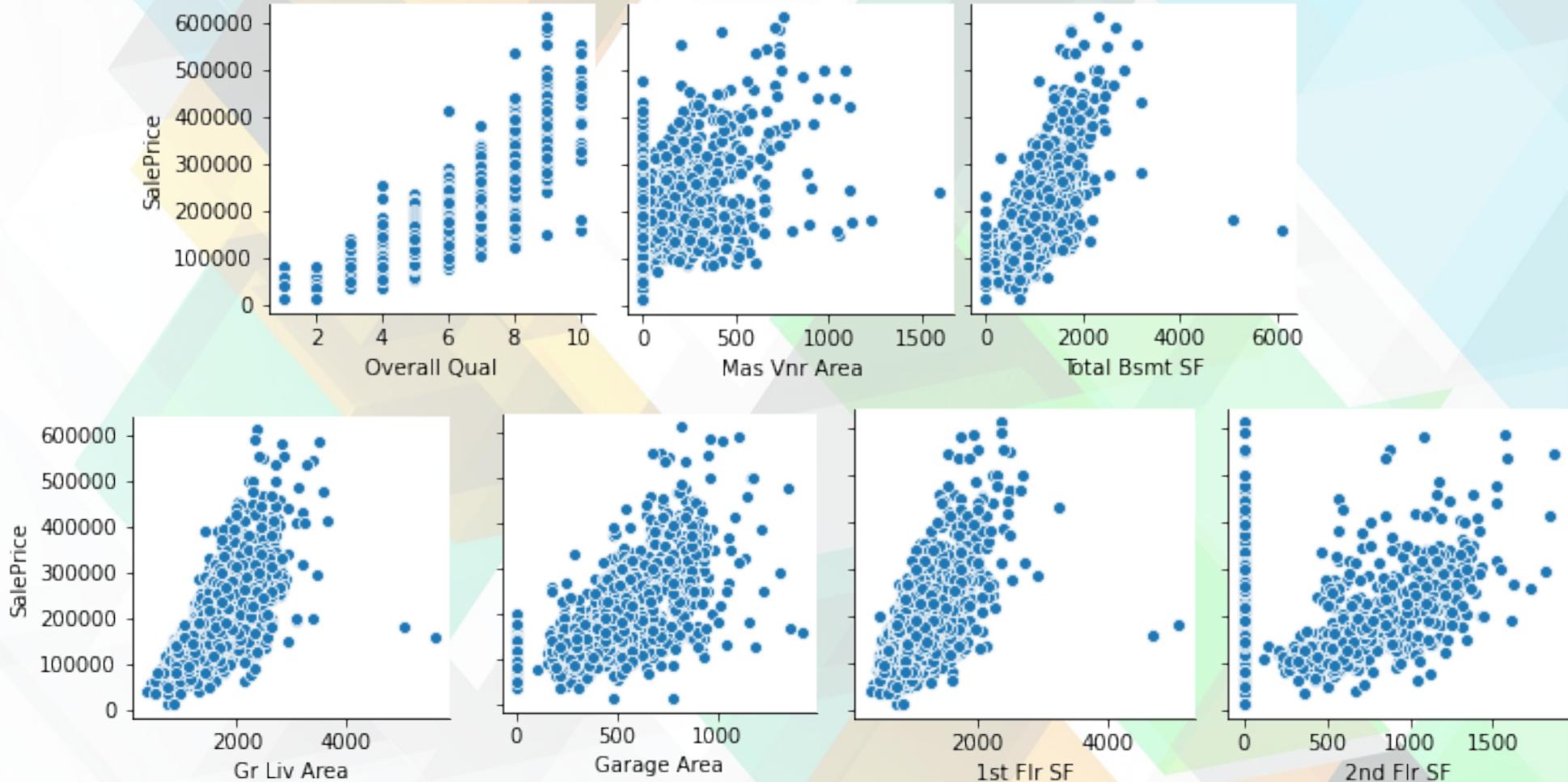
```
test_data['BsmtFin Type 2'].fillna(0, inplace=True)
```

Goal: Don't lose too much data by filling in the null values as much as possible.

Lot Frontage



Exploratory Data Analysis



Linear Regression Model

Predictors

- Overall Quality
- Mason Veneer Area
- Total square feet of basement area
- Above grade (ground) living area square feet
- Size of garage in square feet
- First Floor square feet
- Second floor square feet

TRAIN METRICS:

24473.45

Mean Absolute Error

37720.22

Root Mean Squared
Error

0.77

R2

TEST METRICS:

24328.06

Mean Absolute Error

34369.98

Root Mean Squared
Error

0.83

R2

After attempting log transformation:

TRAIN METRICS:

23134.07

Mean Absolute Error

56981.66

Root Mean Squared
Error

0.47

R2

TEST METRICS:

21031.67

Mean Absolute Error

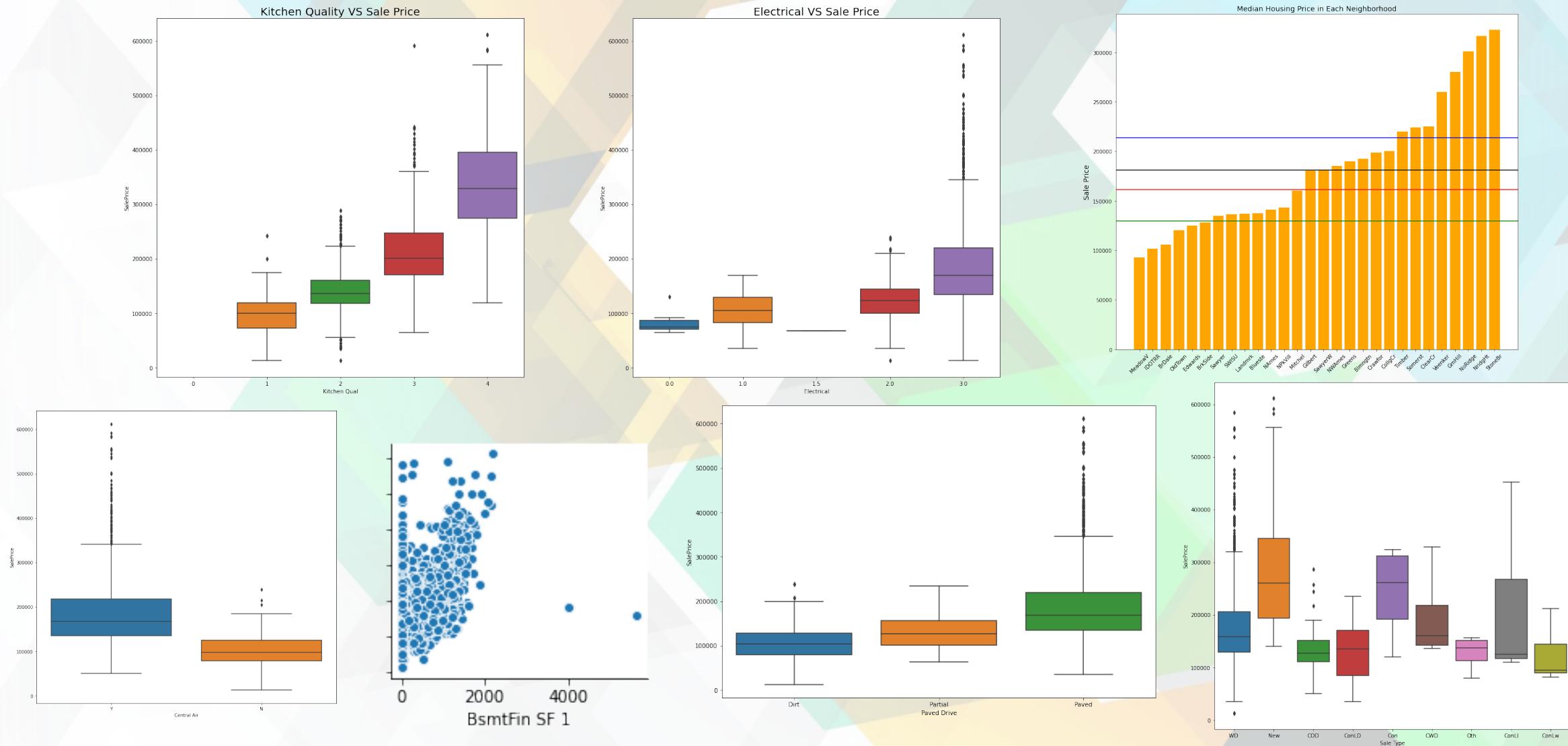
29358.19

Root Mean Squared
Error

0.87

R2

More EDA



Linear Regression Model

Predictors

- Overall Quality
- Mason Veneer Area
- Total square feet of basement area
- Above grade (ground) living area square feet
- Size of garage in square feet
- First Floor square feet
- Second floor square feet
- Paved Drive [Engineered]
- Sale Type (New or Con) [Engineered]
- Type 1 finished square feet
- Central Air
- Electrical
- Kitchen Quality
- Neighborhood Codes [Engineered]
 - (1 for neighborhoods below average housing price, 1.5 for right at average or under 75th percentile, 2 for above 75th percentile)

After more data cleaning and feature engineering:

TRAIN METRICS:

17137.33

Mean Absolute Error

24234.81

Root Mean Squared Error

0.90

R2

TEST METRICS:

16582.7

Mean Absolute Error

23463.17

Root Mean Squared Error

0.92

R2

Conclusions

Predictors	% Change in Sale Price if Predictor is Increased by 1 Unit
Central Air	16.66%
Neighborhood Codes	15.15%
Paved	11.4%
Overall Quality	8.75%
Sale Type (New or Con)	7.39%
Kitchen Quality	6.25%
Electrical	2.29%
1st Floor Square Foot	0.028%
2nd Floor SF	0.023%
Garage Area	0.014%
Type 1 finished square feet	0.012%
Total Basement Square Foot	0.008%
Mason Veneer Area	-0.0026%
General Living Area	0.0025%



Sale Price
\$500,000



General Living Area
(all else held equal)



\$501,250

Constant = 27473.7

Recommendations and Limitation

- **Overall Quality, Kitchen Quality**
- All agents are familiar with guidelines and rubrics.
- Have more than one agent to evaluate a house.

Questions, Comments, Concerns?

