

A stylized, light-colored illustration of a house with a gabled roof and a central chimney, serving as a background for the text.

Housing Prices in Boston

Predicting Housing Prices

Audience

- Realtors
- Potential Buyer and Seller



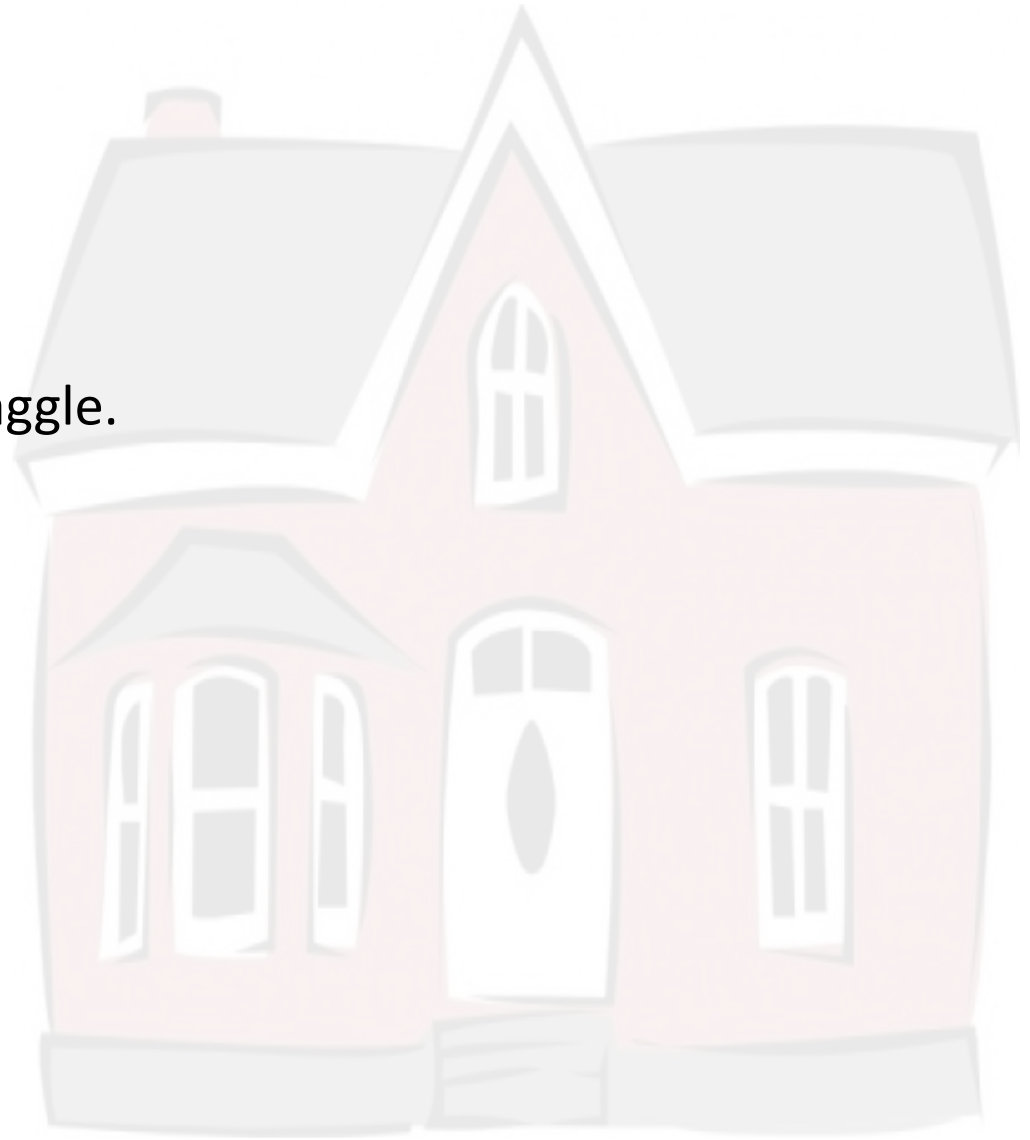
Problem Statement

- Unsure of prices in the housing market.
- Able to predict prices accurately to attract buyer
- Buyers have a good idea what type of house they get with the specific range.



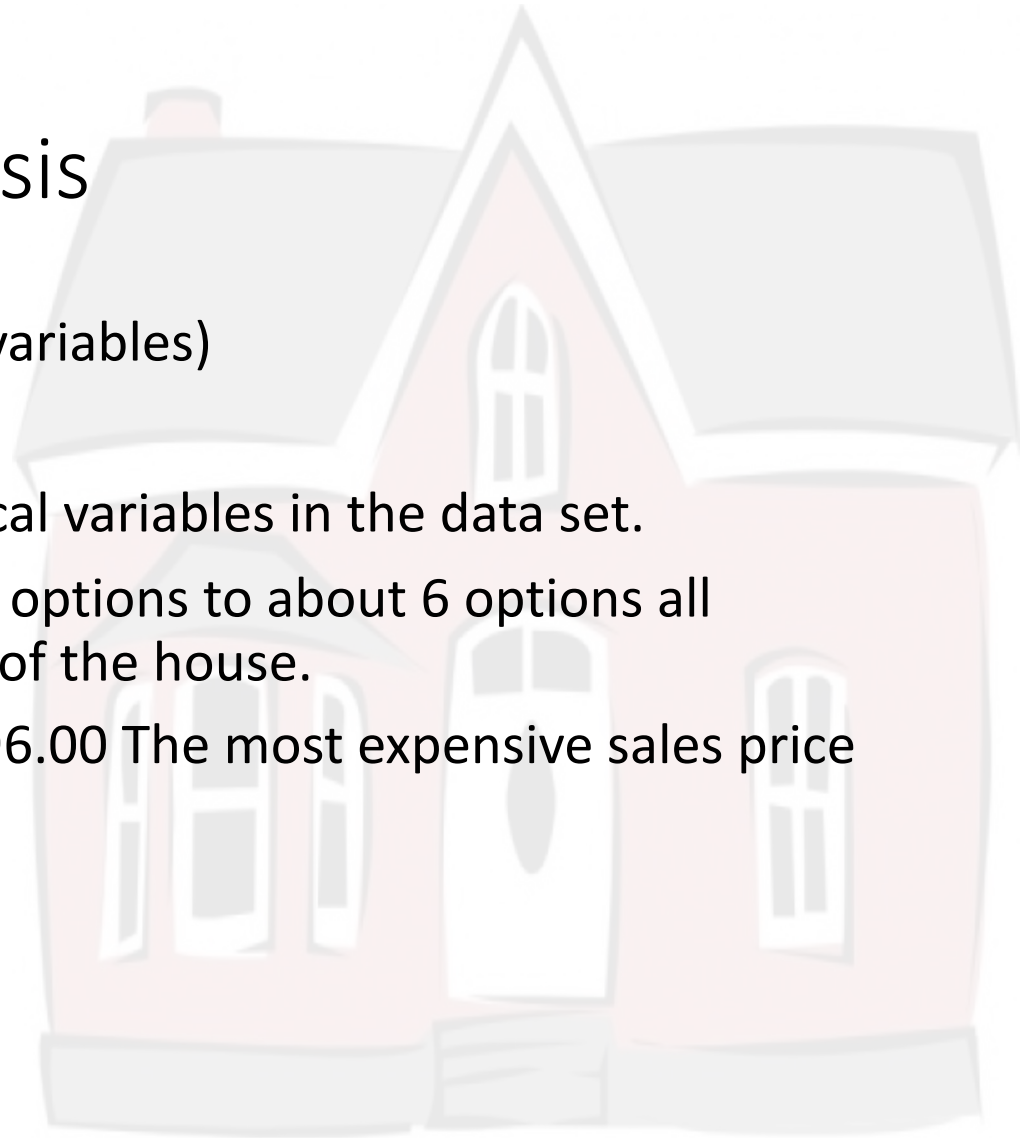
Data

- Boston Housing Data
Housing set was taken from Kaggle.



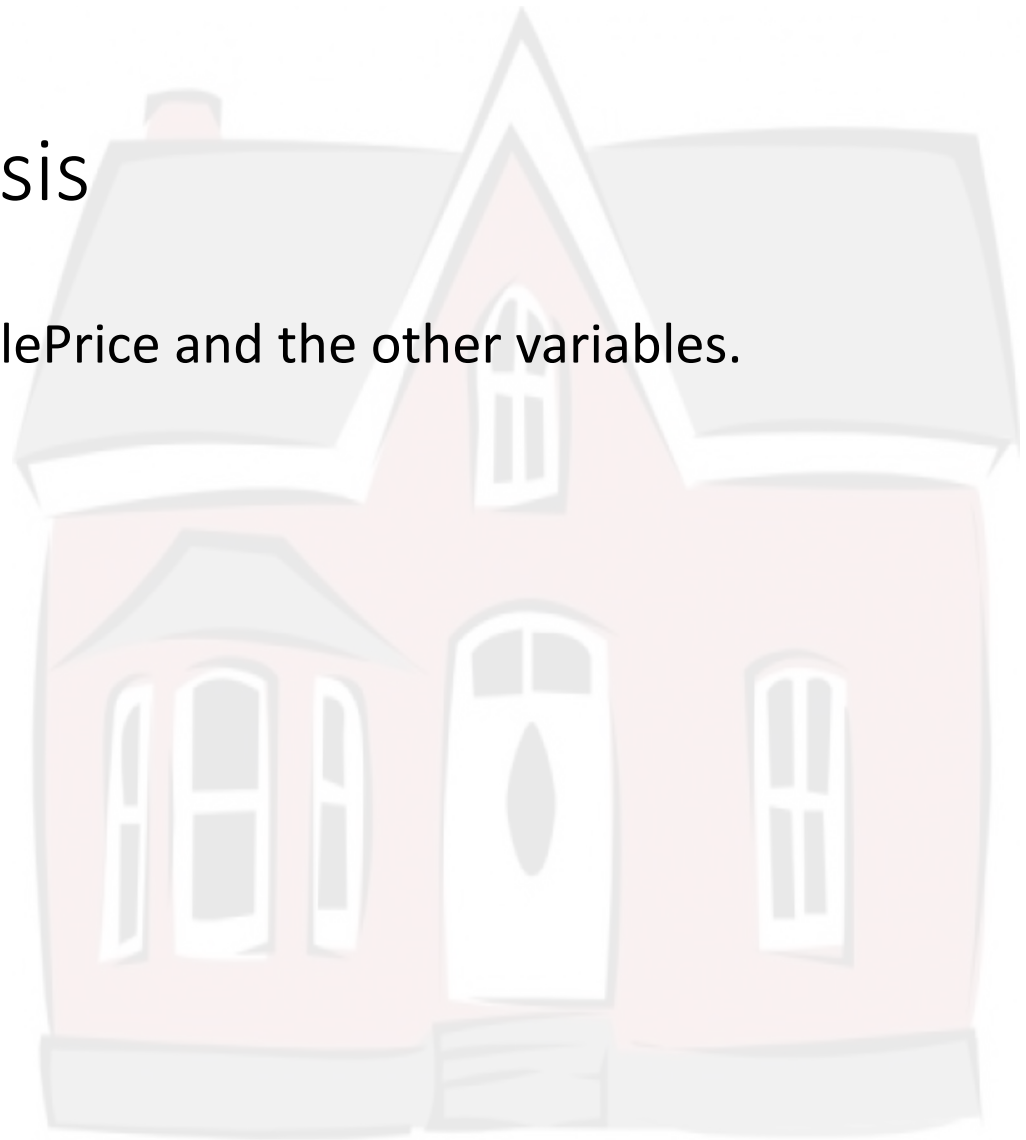
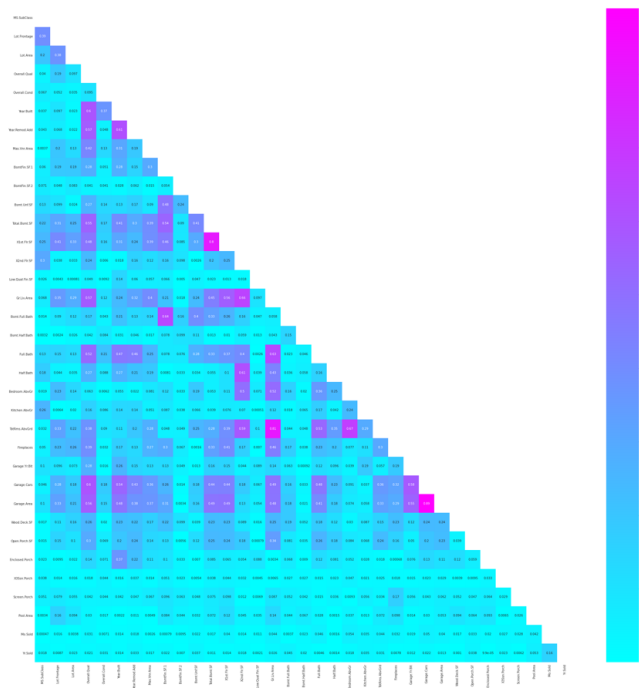
Exploratory Data Analysis

- There are 80 different Columns (variables)
- 2930 different data points.
- There are categorical and numerical variables in the data set.
- Categorical data can range from 2 options to about 6 options all describing the quality or features of the house.
- The average sales price is \$180,796.00 The most expensive sales price in this data set is \$755,000



Exploratory Data Analysis

Looking at Correlations between SalePrice and the other variables.



Preprocessing

- Cleaning Data
- Normalizing data



Data Cleaning – Data that was removed

- Variables with over 90% null values
- Variables that have over 95% correlation with Sale Price
- Variables that are under 1% correlated with Sale Price



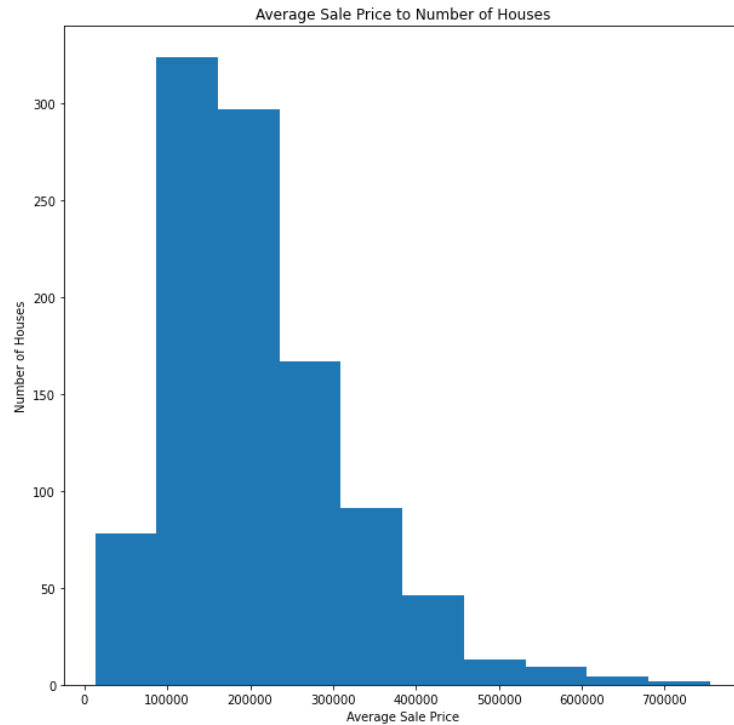
Filling in Null Values

- Variables with categorical values null values were substituted with 0 or their categorical counterpart of 0
- Variables with null values that contain a numerical values would be filled with the mean.



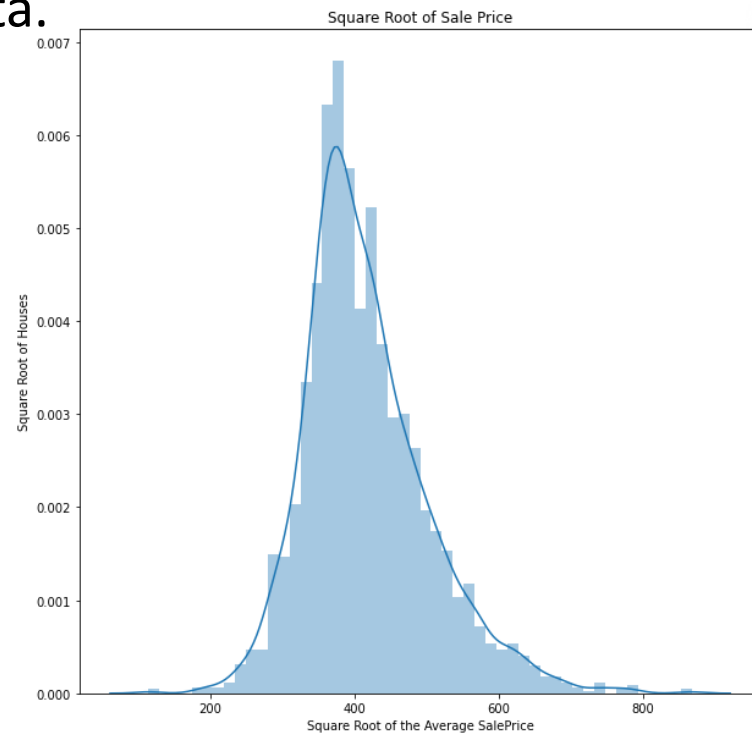
Normalizing the data set

- Looking Sale Price the data was skewed towards the right.



Sale Price

- Taking the Square root of Sale Price improved the distribution of the data.



Changing data types

- Changed all object items in to categorical data so that it can be converted into dummie variables.



Models and Model Processing

- Models used to predict Sales Price:

1. Ordinary Least-Squares Regression
2. Linear Regression
3. Lasso
4. Ridge
5. Elastic Net
6. Gradient Boosting Regressor
7. Random Forest Regressor



Model Metrics

| | Explained Variance (r ² score) | mean ² error | root mean ² error | mean absolute error |
|--------------------------------|--|-------------------------|------------------------------|---------------------|
| Linear Regression | -1.14E+21 | 8.36E+24 | 2.8921E+12 | 1.1149E+11 |
| Ridge | 0.917 | 605.997 | 24.617 | 16.968 |
| Gradient Boosting Regressor | 0.909 | 664.556 | 25.779 | 17.445 |
| Random Forest Regressor | 0.901 | 727.977 | 26.981 | 18.304 |
| Lasso(alpha=95) | 0.919 | 595.885 | 24.411 | 16.791 |
| ElasticNet(ratio=.95) | 0.917 | 605.997 | 24.617 | 16.968 |

Future Recommendations and Next Best steps

- Minimize the variance of Sale Price by creating a new variable with a range of prices.
- Find the top 10 variables that contribute the Sale Price
- Collect more data

