

Assignment 3 Part 3 Report

Subin Shrestha

On analysis of the house prices data, we can see that most of the numerical value is positively skewed. From Figure 1, we can see that the value of value of miscellaneous feature (MiscVal) has the highest skewness followed by Pool Area, Lot Area, three season porch area (3ssnPorch) and Low quality finished square feet (LowQualFinSF). On the other hand values like, YearBuilt, YearRemodAdd, GarageYrBlt and GarageCars has negative correlation with sale price.

Similarly, from Figure 2, the overall rate of material and finish of the house given by OverallQual has the highest correlation with sale price. GarageCars and GarageArea is the second and third most feature correlated with sale price. Both features have similar values as GarageCars and GarageArea have high correlation between the two. TotalBsmtSF is the forth most correlated feature with similar value to that of GarageCars and GarageArea.

From Figure 3, we can see that, only OverallQual(Rates the overall material and finish of the house) and OverallCond (Rates the overall condition of the house) have positive relationship with sale price. GarageCars (Size of garage in car capacity) has a positive relationship until 3 cars and then it falls off. All of the other features has fluctuations in sale prices even for same value.

Similarly, Northridge, Northridge Heights, Stone Brook are the most expensive neighborhood (Figure 4) with the highest mean and median house prices. House prices did not change between years 2006 and 2010 (Figure 5). There does not seem to any change in mean and median house prices. Foundations used to build the house also had some effect in the price of the house (Figure 6). Poured Concrete and Wood were the most expensive with both having the highest mean and median house prices. In the similar fashion, the type of housing (MSSubClass) also affected the sale price (Figure 7). 2-STORY 1946 & NEWER (60) and 1-STORY PUD (Planned Unit Development) - 1946 & NEWER (120) were the most expensive ones due to its high mean and median sale price.

Figures

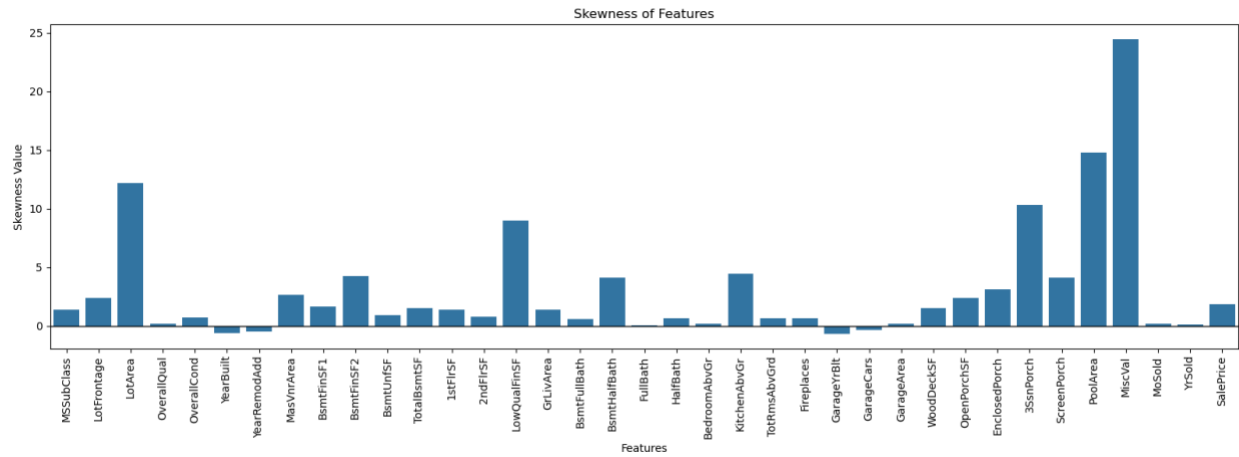


Figure 1: Skewness of Features

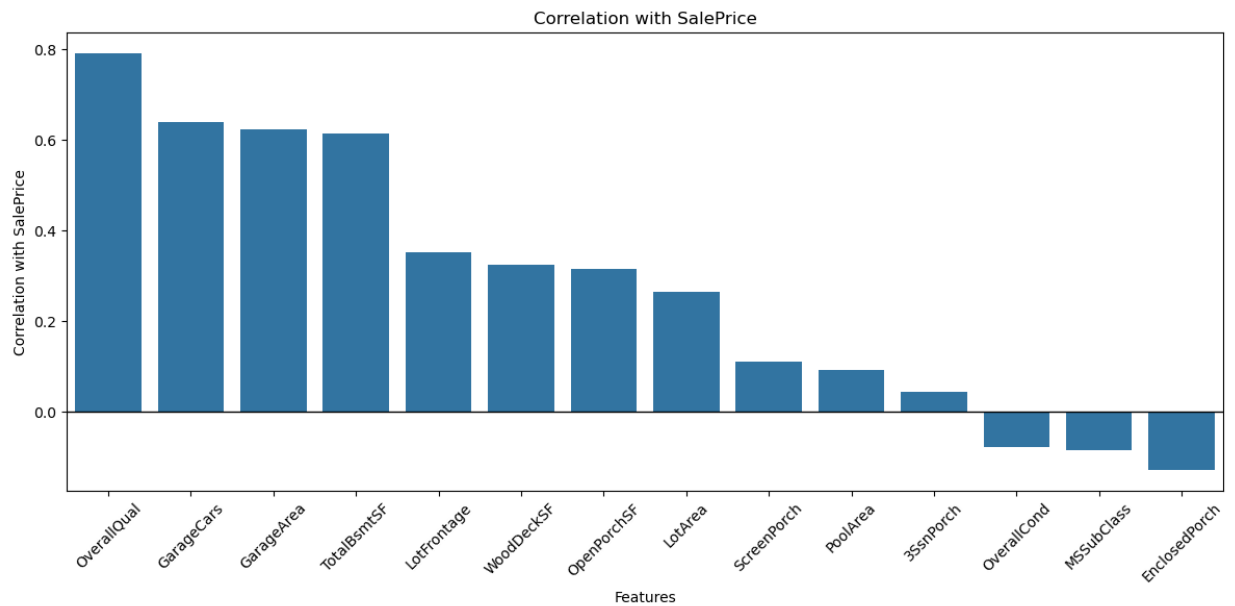


Figure 2: Correlation with Sale Price

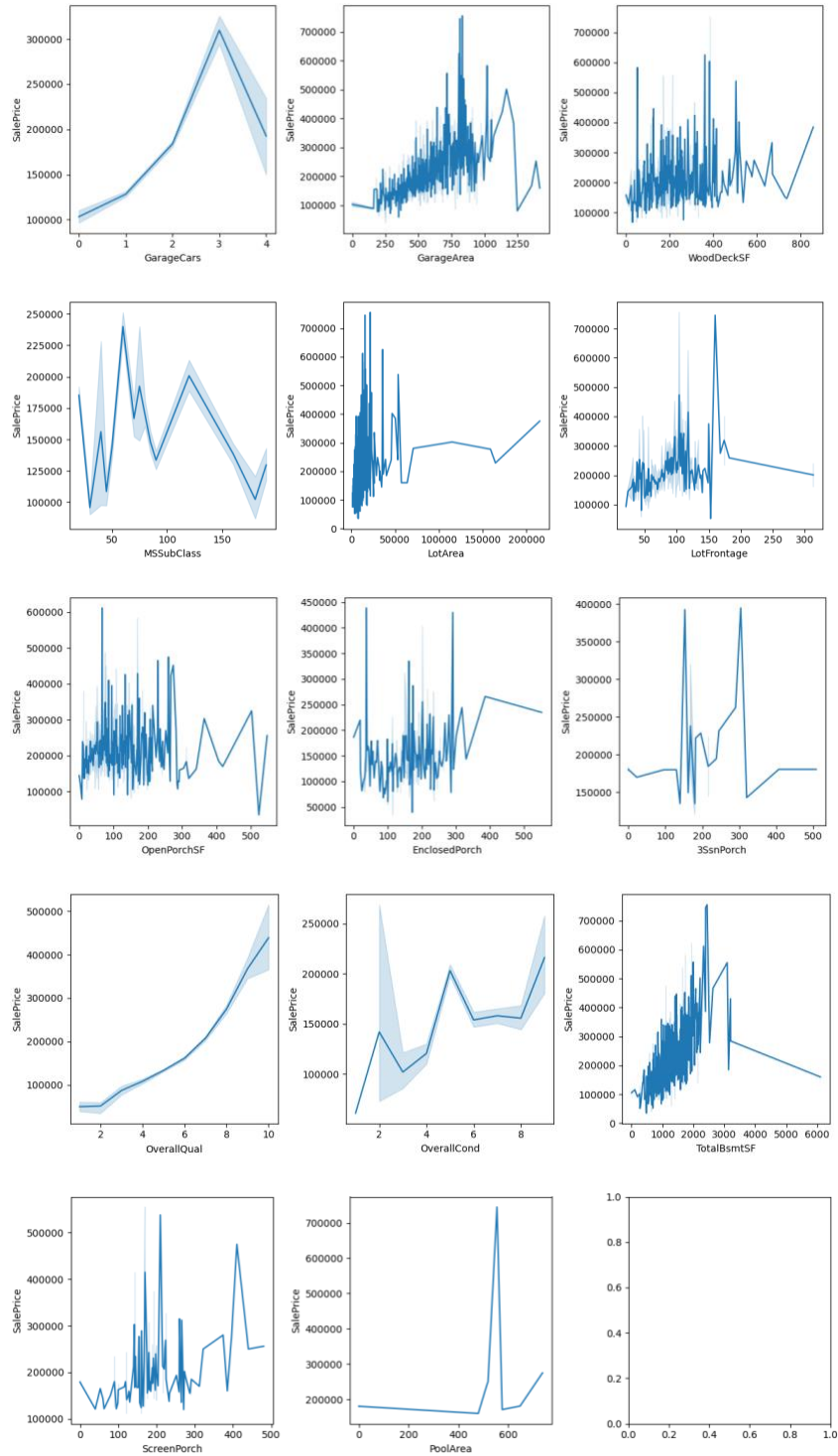


Figure 3: Feature Line Plots

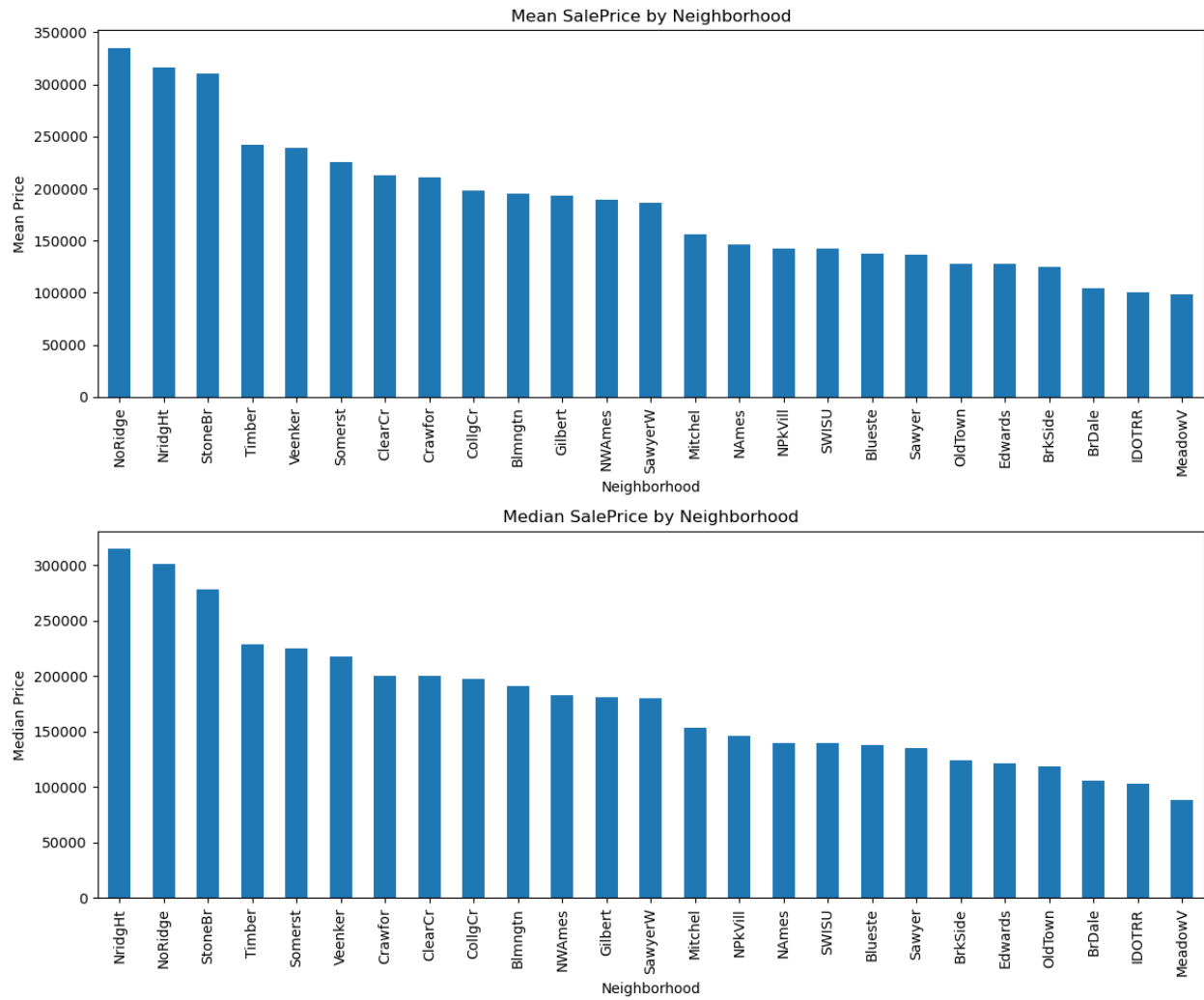


Figure 4: Mean and Median neighborhood house prices

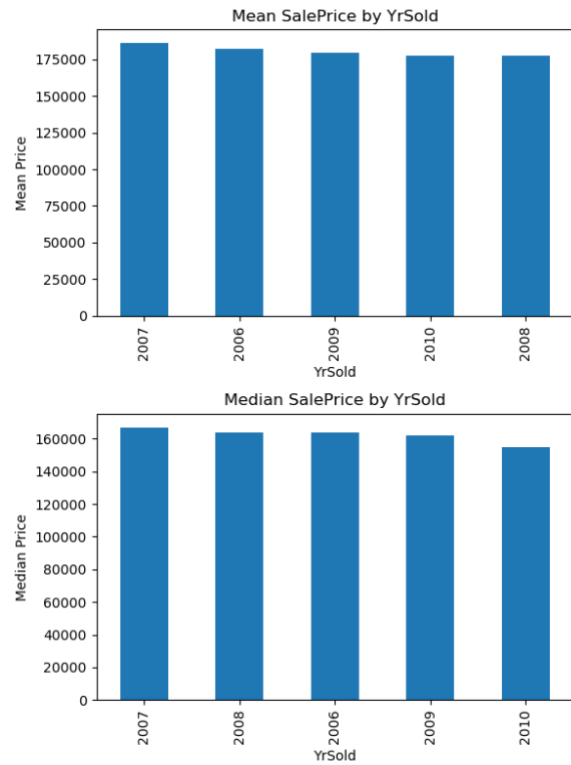


Figure 5: Year Sold Mean Median Prices

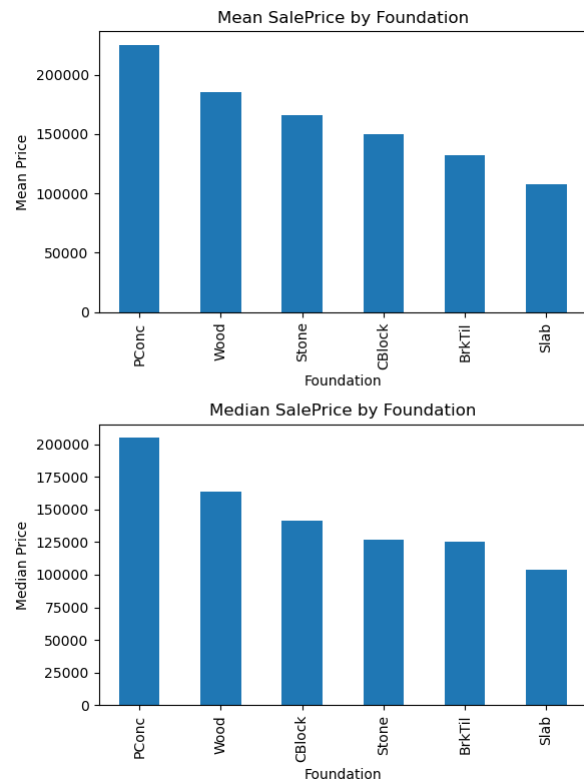


Figure 6 Foundations Mean and Median Sale Price

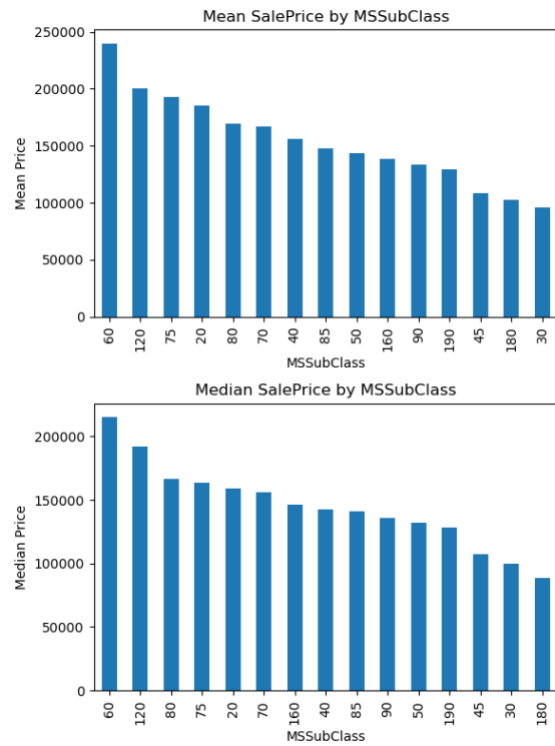


Figure 7 MSSubClass Mean Median Sale Price