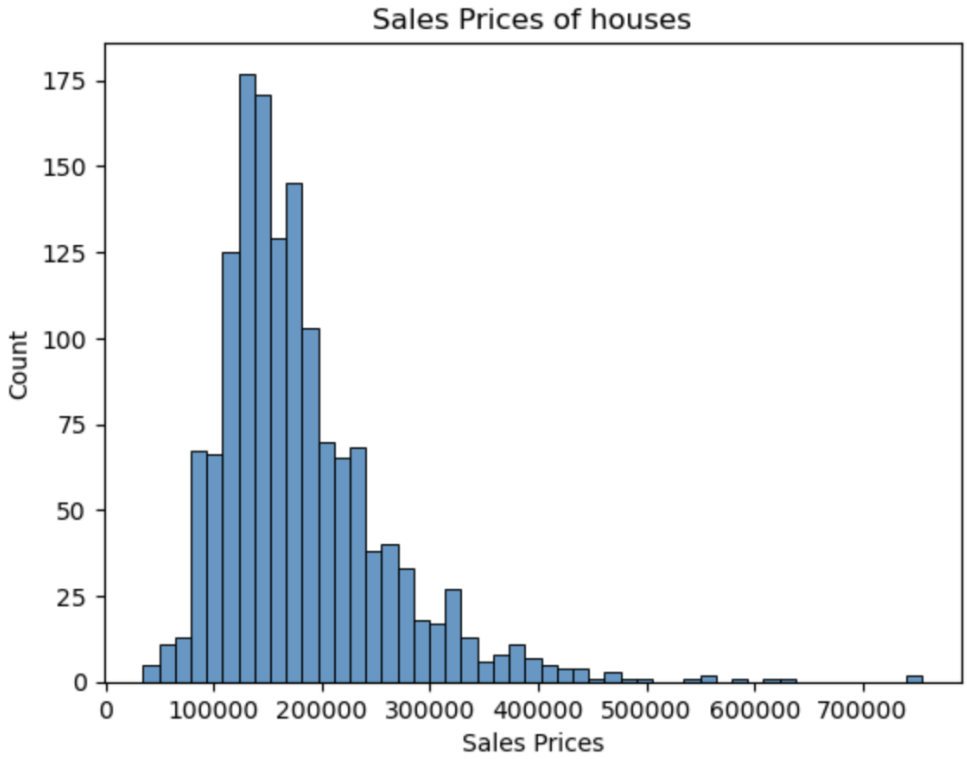
Gautham Jayadeep
PDD Data Analytics
Langara College

Assignment 3-part 3

Source_dataset: Housing_Price

1) Analysis of the distribution of SalePrice

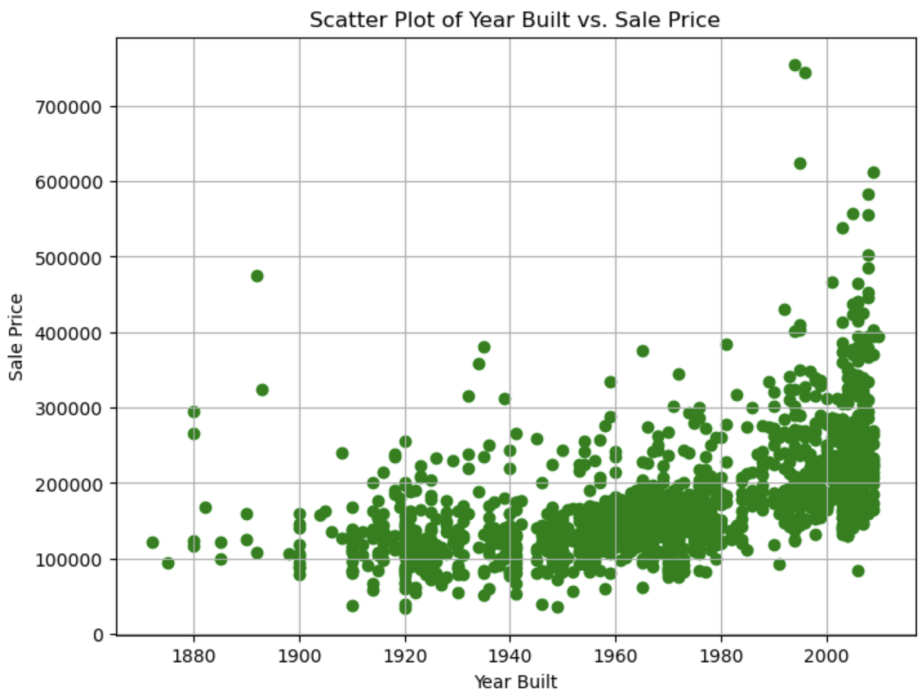


From the histogram, we can see the graph is **Right skewed** i.e. most of the data falls to the left part of the graph and the graph has a single peak.

The mean for the distribution =180921.195890
The median = 163000.000000

Mean > Median, indicating the data is right skewed.

2) Determine if the year of built affects the sale price of the house



From the graph we can see that during the year 1880-1980, the process of houses was steady. But after 1980, the sale price grew exponentially.

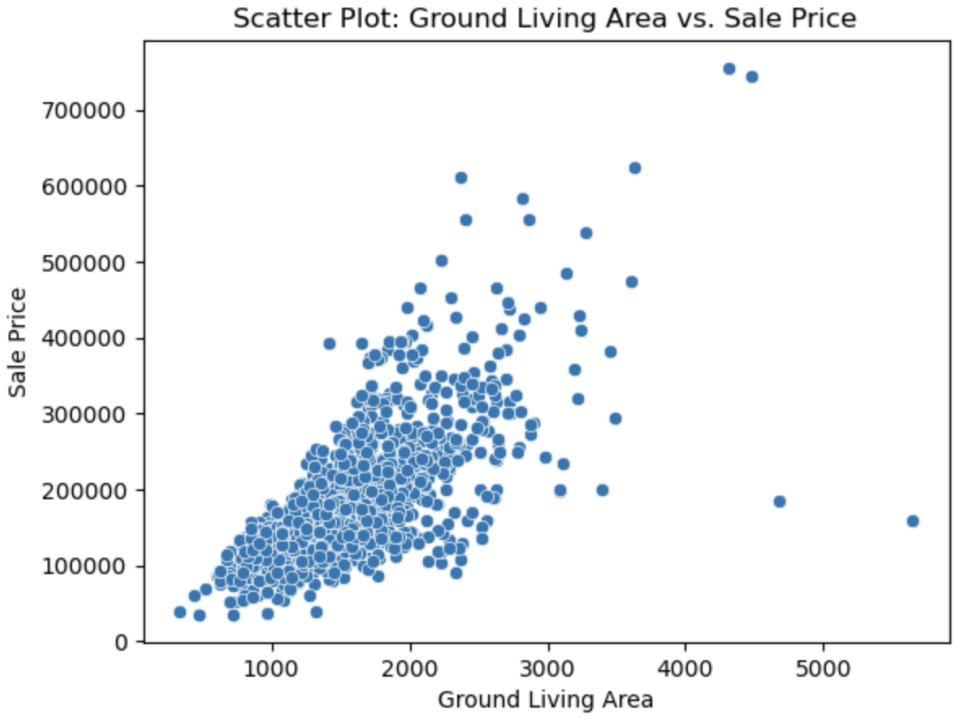
Chi_square test:

Chi-squared value: 411.88818868825433
Degree of freedom: 8
P-value: 5.358640559461605e-84

From the above test, we can compare the chi-square value and the decision point to understand the association. The degree of freedom corresponds to the decision point(DP) of 15.51. Since the chi-square value is greater than DP, thus we have significant data to conclude that the year of built and the sale price are not significantly independent of each other.

Hence we can conclude that **the year of built is associated with the sale price of the house**.

3) Determine whether Living area is associated with Sale Price



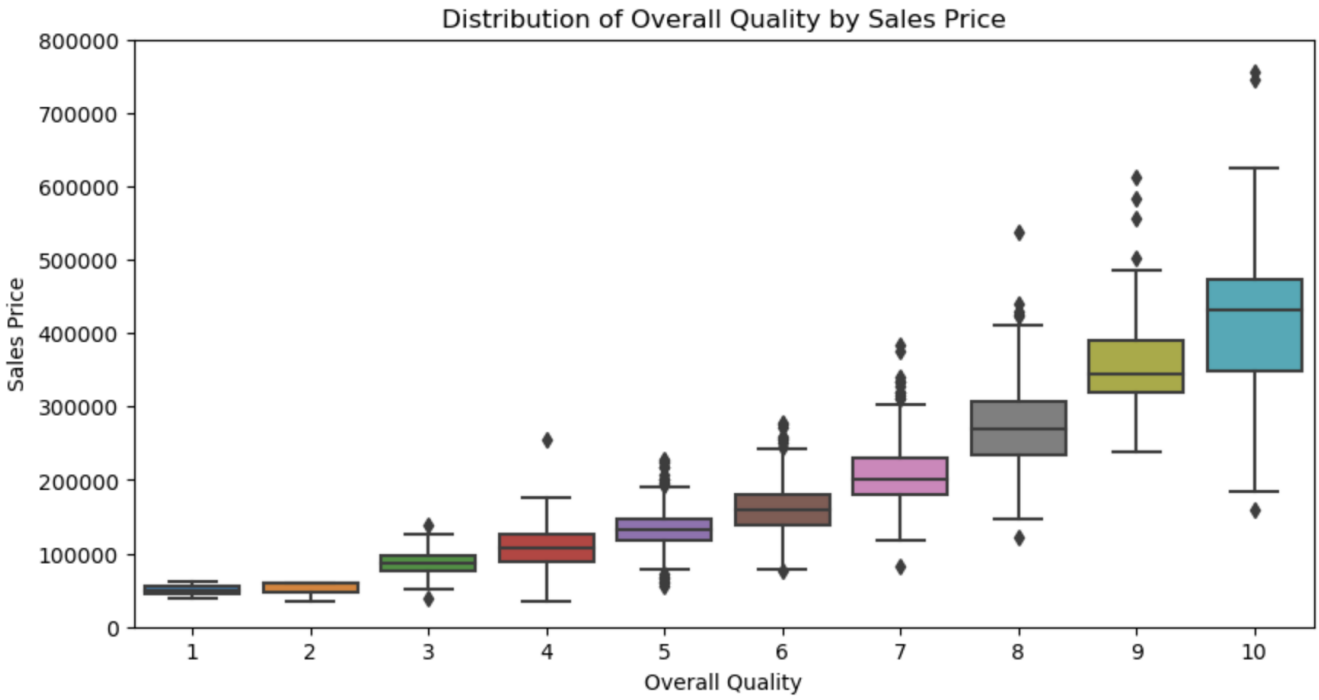
The graph shows a strong, positive linear relationship. **As the living area increases, the price of the house increases.**

Correlation Coefficient

Correlation between Ground Living Area and Sale Price: 0.71.
A correlation of 0.71 is relatively high, suggesting that changes in Ground Living Area are associated with significant changes in Sale Price.

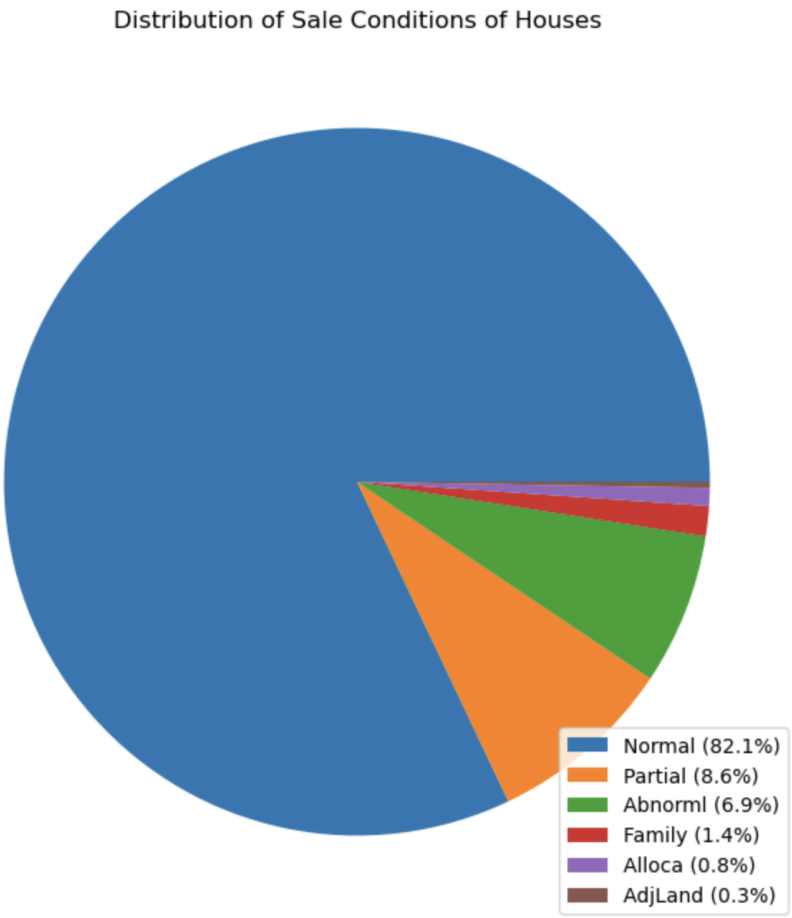
Hence we can conclude that **there is a strong positive association between the size of the ground living area and the sale price of houses**.

4) Determine if the overall quality of the house is associated with the Sale Price



From the graph above, we can infer that as the overall quality of the house increases, the median sale price increases. Also, an increase in the variance of price is seen from the different interquartile ranges.

5) Analyze the distribution of sale condition of the houses.



From the above pie chart, we can understand that Normal houses are sold the most while Adjacent Land House.