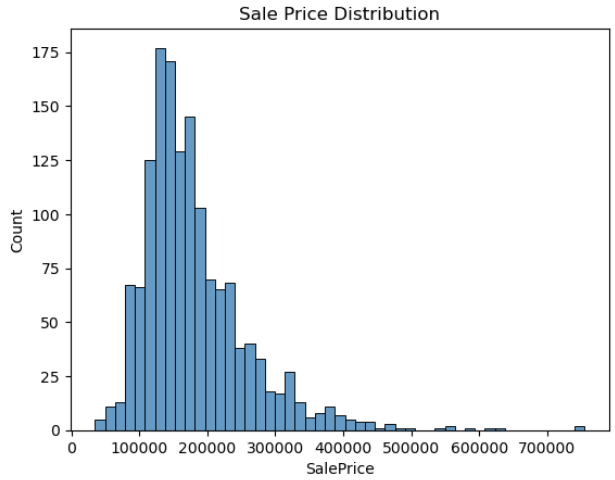
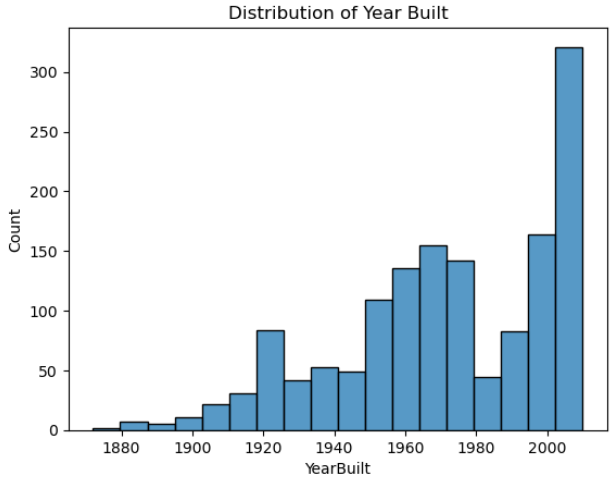
Determine the distribution of *SalePrice* and compute the appropriate measure of centre.



The histogram of the column *SalePrice* appears to be heavily right skewed. Due to the presence of some outliers, the appropriate measure of centre would be the median value of `163000`.

Determine the distribution of *YearBuilt*.

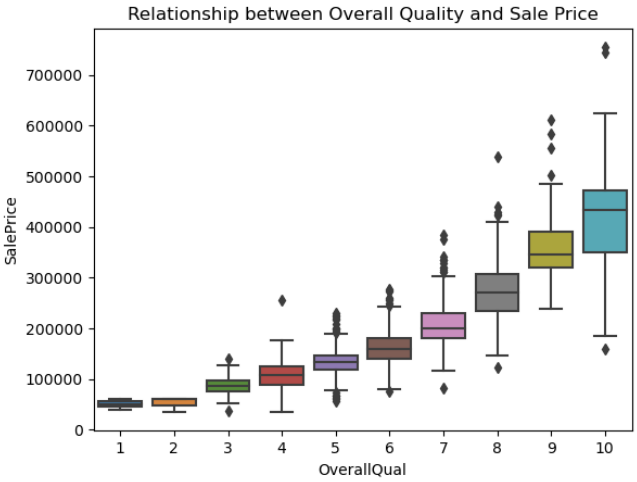


The distribution of *YearBuilt* is left-skewed and is multimodal. The data contain higher number of houses built after the year 2000 compared to other years, but most houses were built in the mid 1900's.

Plot *GrLivArea* and *SalePrice*. Are the variables correlated with one another?

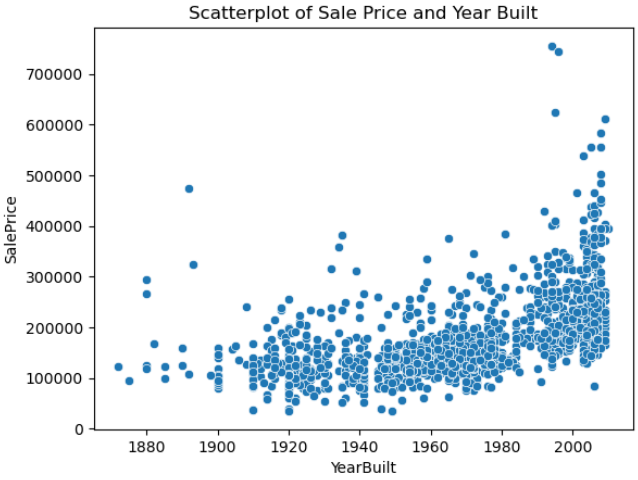
The scatterplot of *GrLivArea* and *SalePrice* appears to be positively correlated with outliers present. The correlation coefficient of the variables is approximately `0.71`, which indicates that there is a significant correlation between *GrLivArea* and *SalePrice*.

Is the variable *SalePrice* dependent on *OverallQual* of houses?



The boxplots of *OverallQual* show that houses with higher quality appear to sell at higher price than those with lower quality on average. However, houses with higher quality appear to have higher *SalePrice* range compared to those with lower quality.

Is there relationship between *SalePrice* and *YearBuilt*?



The scatterplot of *SalePrice* and *YearBuilt* appears to have a positive slope. There are outliers present in addition. The variables have a correlation coefficient of `0.53`, which indicates a moderate positive correlation.