

Figure 1: Training data histograms

From training data histogram, for example, 3 bedroom houses are most commonly sold followed by 4 bedroom; 2 size of garage in car capacity are most commonly sold; The overall condition of the house is average (5 out of 10), but the overall quality of the house is below average( 4 out of 10).

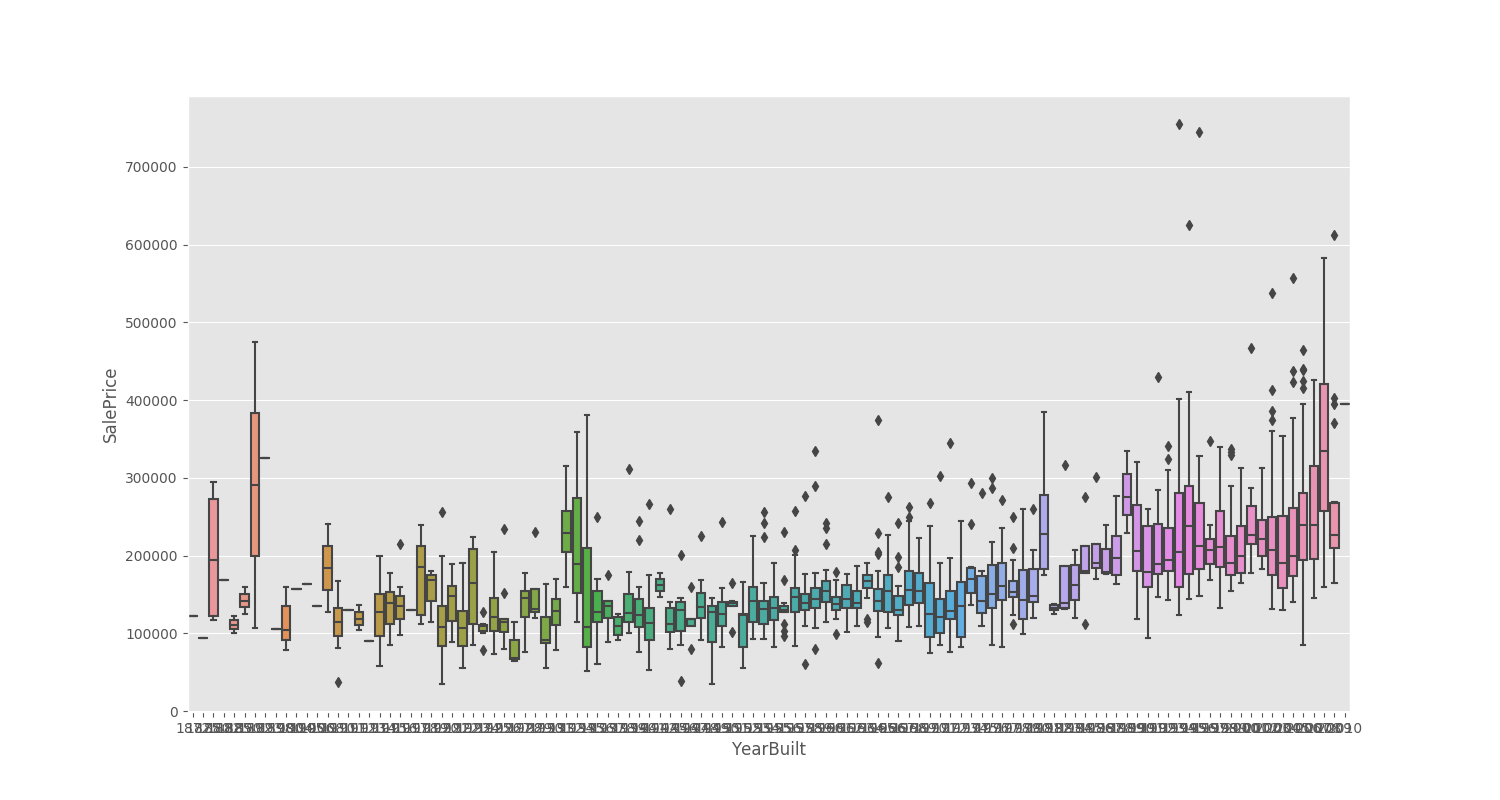


Figure 2: Sale Price against Year Built

Form figure 2, it shows that the price of recent-built houses is higher.

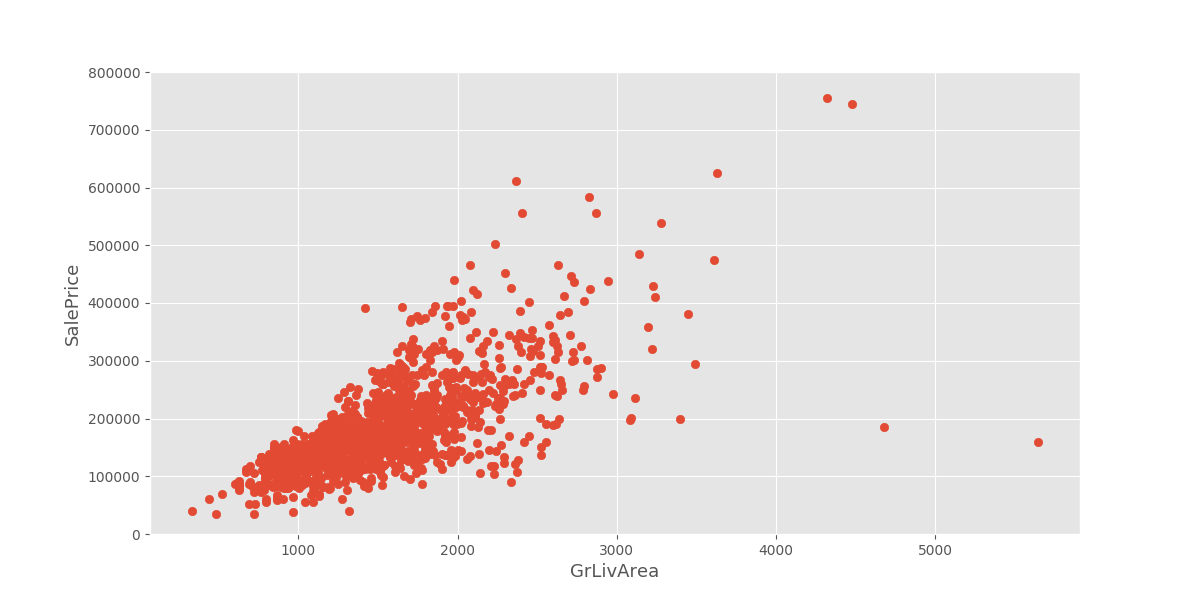


Figure 3: Sale Price against GrLivArea

When the Sale Price against GrLivArea (Above grade (ground) living area square feet) is plotted, we found that there are some outliers. For example, those houses whose area is over 4,000 but the sale price is under 300,000.

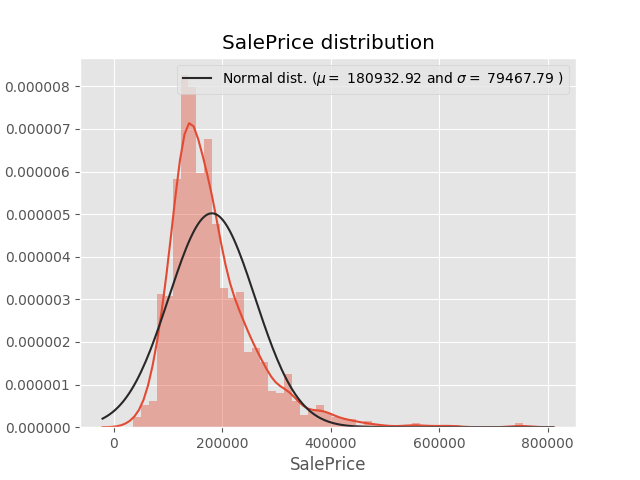


Figure 4: Sale Price distribution



Figure 5: Log Normal Sale Price distribution

The Sale Price distribution is plotted, and it is found that the distribution is right skewed. Therefore, the Sale Price would be log-normalized before training.