CPSC_4800_Assignment3_Part3

Explanatory Data Analysis

1)

By the analysis for the dataset:

-There are 1460 rows and 81 columns.

-There are columns with missing values.

-The sale price is the target variable which indicates the price of the house.

-The mean sale price is approximately $180,921 with a standard deviation of $79,442.

#Hypothesis Analysis.

As sale price is the target variable , will compare it with three other variables which can show with which reason we can sell the houses more.

-(1) Sale price versus Lotarea

The variables have a postive relation. And mostly houses are sold upto the lot size 50000.

-(2) Sale price versus Neighborhood

There is no correlation between both the variables and the selling of houses does not depend on the neighborhood.

-(3)Sale price versus Overall Quality

It clearly shows the relation is positive and the sale price depends on the overall quality of houses.