House Prediction Report

2024-04-16

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'tibble' was built under R version 4.3.3
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```
## Warning: package 'readr' was built under R version 4.3.3
```

```
## Warning: package 'purrr' was built under R version 4.3.2
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
## Warning: package 'forcats' was built under R version 4.3.3
```

```
## Warning: package 'lubridate' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.92 loaded
```

```r
library(lubridate)
library(readr)
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 4.3.3
```

```r
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.3.3
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.3.3
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.3.3
```

```r
library(zoo)
```

```
## Warning: package 'zoo' was built under R version 4.3.3
```

```
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
train_data <- read.csv("C:/Users/matt1/OneDrive/Desktop/MTH404/House Prices Data/train.csv")
test_data <- read.csv("C:/Users/matt1/OneDrive/Desktop/MTH404/House Prices Data/test.csv")
head(train_data)
```

```
##   Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape LandContour
## 1  1         60       RL          65    8450   Pave  <NA>      Reg         Lvl
## 2  2         20       RL          80    9600   Pave  <NA>      Reg         Lvl
```

```
## 3  3         60       RL          68   11250   Pave  <NA>      IR1         Lvl
## 4  4         70       RL          60    9550   Pave  <NA>      IR1         Lvl
## 5  5         60       RL          84   14260   Pave  <NA>      IR1         Lvl
## 6  6         50       RL          85   14115   Pave  <NA>      IR1         Lvl
##   Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType
## 1    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 2    AllPub       FR2       Gtl      Veenker      Feedr       Norm     1Fam
## 3    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 4    AllPub    Corner       Gtl      Crawfor       Norm       Norm     1Fam
## 5    AllPub       FR2       Gtl      NoRidge       Norm       Norm     1Fam
## 6    AllPub    Inside       Gtl      Mitchel       Norm       Norm     1Fam
##   HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl
## 1     2Story           7           5      2003         2003     Gable  CompShg
## 2     1Story           6           8      1976         1976     Gable  CompShg
## 3     2Story           7           5      2001         2002     Gable  CompShg
## 4     2Story           7           5      1915         1970     Gable  CompShg
## 5     2Story           8           5      2000         2000     Gable  CompShg
## 6     1.5Fin           5           5      1993         1995     Gable  CompShg
##   Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond Foundation
## 1     VinylSd     VinylSd    BrkFace        196        Gd        TA      PConc
## 2     MetalSd     MetalSd       None          0        TA        TA     CBlock
## 3     VinylSd     VinylSd    BrkFace        162        Gd        TA      PConc
## 4     Wd Sdng     Wd Shng       None          0        TA        TA     BrkTil
## 5     VinylSd     VinylSd    BrkFace        350        Gd        TA      PConc
## 6     VinylSd     VinylSd       None          0        TA        TA       Wood
##   BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
## 1       Gd       TA           No          GLQ        706          Unf
## 2       Gd       TA           Gd          ALQ        978          Unf
## 3       Gd       TA           Mn          GLQ        486          Unf
## 4       TA       Gd           No          ALQ        216          Unf
## 5       Gd       TA           Av          GLQ        655          Unf
## 6       Gd       TA           No          GLQ        732          Unf
##   BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir Electrical
## 1          0       150         856    GasA        Ex          Y      SBrkr
## 2          0       284        1262    GasA        Ex          Y      SBrkr
## 3          0       434         920    GasA        Ex          Y      SBrkr
## 4          0       540         756    GasA        Gd          Y      SBrkr
## 5          0       490        1145    GasA        Ex          Y      SBrkr
## 6          0        64         796    GasA        Ex          Y      SBrkr
##   X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath
## 1       856       854            0      1710            1            0        2
## 2      1262         0            0      1262            0            1        2
## 3       920       866            0      1786            1            0        2
## 4       961       756            0      1717            1            0        1
## 5      1145      1053            0      2198            1            0        2
## 6       796       566            0      1362            1            0        1
##   HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## 1        1            3            1          Gd            8        Typ
## 2        0            3            1          TA            6        Typ
## 3        1            3            1          Gd            6        Typ
## 4        0            3            1          Gd            7        Typ
## 5        1            4            1          Gd            9        Typ
## 6        1            1            1          TA            5        Typ
##   Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars
```

```
## 1          0        <NA>     Attchd        2003          RFn          2
## 2          1          TA     Attchd        1976          RFn          2
## 3          1          TA     Attchd        2001          RFn          2
## 4          1          Gd     Detchd        1998          Unf          3
## 5          1          TA     Attchd        2000          RFn          3
## 6          0        <NA>     Attchd        1993          Unf          2
##   GarageArea GarageQual GarageCond PavedDrive WoodDeckSF OpenPorchSF
## 1        548         TA         TA          Y          0          61
## 2        460         TA         TA          Y        298           0
## 3        608         TA         TA          Y          0          42
## 4        642         TA         TA          Y          0          35
## 5        836         TA         TA          Y        192          84
## 6        480         TA         TA          Y         40          30
##   EnclosedPorch X3SsnPorch ScreenPorch PoolArea PoolQC Fence MiscFeature
## 1             0          0           0        0   <NA>  <NA>        <NA>
## 2             0          0           0        0   <NA>  <NA>        <NA>
## 3             0          0           0        0   <NA>  <NA>        <NA>
## 4           272          0           0        0   <NA>  <NA>        <NA>
## 5             0          0           0        0   <NA>  <NA>        <NA>
## 6             0        320           0        0   <NA> MnPrv        Shed
##   MiscVal MoSold YrSold SaleType SaleCondition SalePrice
## 1       0      2   2008       WD        Normal    208500
## 2       0      5   2007       WD        Normal    181500
## 3       0      9   2008       WD        Normal    223500
## 4       0      2   2006       WD       Abnorml    140000
## 5       0     12   2008       WD        Normal    250000
## 6     700     10   2009       WD        Normal    143000
```

```
head(test_data)
```

```
##     Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape
## 1 1461         20       RH          80   11622   Pave  <NA>      Reg
## 2 1462         20       RL          81   14267   Pave  <NA>      IR1
## 3 1463         60       RL          74   13830   Pave  <NA>      IR1
## 4 1464         60       RL          78    9978   Pave  <NA>      IR1
## 5 1465        120       RL          43    5005   Pave  <NA>      IR1
## 6 1466         60       RL          75   10000   Pave  <NA>      IR1
##   LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2
## 1         Lvl    AllPub    Inside       Gtl        NAmes      Feedr       Norm
## 2         Lvl    AllPub    Corner       Gtl        NAmes       Norm       Norm
## 3         Lvl    AllPub    Inside       Gtl      Gilbert       Norm       Norm
## 4         Lvl    AllPub    Inside       Gtl      Gilbert       Norm       Norm
## 5         HLS    AllPub    Inside       Gtl      StoneBr       Norm       Norm
## 6         Lvl    AllPub    Corner       Gtl      Gilbert       Norm       Norm
##   BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle
## 1     1Fam     1Story           5           6      1961         1961     Gable
## 2     1Fam     1Story           6           6      1958         1958       Hip
## 3     1Fam     2Story           5           5      1997         1998     Gable
## 4     1Fam     2Story           6           6      1998         1998     Gable
## 5   TwnhsE     1Story           8           5      1992         1992     Gable
## 6     1Fam     2Story           6           5      1993         1994     Gable
##   RoofMatl Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond
## 1  CompShg     VinylSd     VinylSd       None          0        TA        TA
## 2  CompShg     Wd Sdng     Wd Sdng    BrkFace        108        TA        TA
```

```
## 3  CompShg     VinylSd     VinylSd       None          0        TA        TA
## 4  CompShg     VinylSd     VinylSd    BrkFace         20        TA        TA
## 5  CompShg     HdBoard     HdBoard       None          0        Gd        TA
## 6  CompShg     HdBoard     HdBoard       None          0        TA        TA
##   Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1
## 1      CBlock       TA       TA           No          Rec        468
## 2      CBlock       TA       TA           No          ALQ        923
## 3       PConc       Gd       TA           No          GLQ        791
## 4       PConc       TA       TA           No          GLQ        602
## 5       PConc       Gd       TA           No          ALQ        263
## 6       PConc       Gd       TA           No          Unf          0
##   BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir
## 1          LwQ        144       270         882    GasA        TA          Y
## 2          Unf          0       406        1329    GasA        TA          Y
## 3          Unf          0       137         928    GasA        Gd          Y
## 4          Unf          0       324         926    GasA        Ex          Y
## 5          Unf          0      1017        1280    GasA        Ex          Y
## 6          Unf          0       763         763    GasA        Gd          Y
##   Electrical X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath
## 1      SBrkr       896         0            0       896            0
## 2      SBrkr      1329         0            0      1329            0
## 3      SBrkr       928       701            0      1629            0
## 4      SBrkr       926       678            0      1604            0
## 5      SBrkr      1280         0            0      1280            0
## 6      SBrkr       763       892            0      1655            0
##   BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual
## 1            0        1        0            2            1          TA
## 2            0        1        1            3            1          Gd
## 3            0        2        1            3            1          TA
## 4            0        2        1            3            1          Gd
## 5            0        2        0            2            1          Gd
## 6            0        2        1            3            1          TA
##   TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt
## 1            5        Typ          0       <NA>     Attchd        1961
## 2            6        Typ          0       <NA>     Attchd        1958
## 3            6        Typ          1         TA     Attchd        1997
## 4            7        Typ          1         Gd     Attchd        1998
## 5            5        Typ          0       <NA>     Attchd        1992
## 6            7        Typ          1         TA     Attchd        1993
##   GarageFinish GarageCars GarageArea GarageQual GarageCond PavedDrive
## 1          Unf          1        730         TA         TA          Y
## 2          Unf          1        312         TA         TA          Y
## 3          Fin          2        482         TA         TA          Y
## 4          Fin          2        470         TA         TA          Y
## 5          RFn          2        506         TA         TA          Y
## 6          Fin          2        440         TA         TA          Y
##   WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch PoolArea PoolQC
## 1        140           0             0          0         120        0   <NA>
## 2        393          36             0          0           0        0   <NA>
## 3        212          34             0          0           0        0   <NA>
## 4        360          36             0          0           0        0   <NA>
## 5          0          82             0          0         144        0   <NA>
## 6        157          84             0          0           0        0   <NA>
##   Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition
```

```
## 1 MnPrv        <NA>       0      6   2010       WD        Normal
## 2  <NA>        Gar2   12500      6   2010       WD        Normal
## 3 MnPrv        <NA>       0      3   2010       WD        Normal
## 4  <NA>        <NA>       0      6   2010       WD        Normal
## 5  <NA>        <NA>       0      1   2010       WD        Normal
## 6  <NA>        <NA>       0      4   2010       WD        Normal
```

```
summary(train_data)
```

```
##        Id           MSSubClass      MSZoning          LotFrontage
##  Min.   :   1.0   Min.   : 20.0   Length:1460        Min.   : 21.00
##  1st Qu.: 365.8   1st Qu.: 20.0   Class :character   1st Qu.: 59.00
##  Median : 730.5   Median : 50.0   Mode  :character   Median : 69.00
##  Mean   : 730.5   Mean   : 56.9                      Mean   : 70.05
##  3rd Qu.:1095.2   3rd Qu.: 70.0                      3rd Qu.: 80.00
##  Max.   :1460.0   Max.   :190.0                      Max.   :313.00
##                                                      NA's   :259
##     LotArea          Street             Alley             LotShape
##  Min.   :  1300   Length:1460        Length:1460        Length:1460
##  1st Qu.:  7554   Class :character   Class :character   Class :character
##  Median :  9478   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 10517
##  3rd Qu.: 11602
##  Max.   :215245
##
##  LandContour         Utilities          LotConfig          LandSlope
##  Length:1460        Length:1460        Length:1460        Length:1460
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Neighborhood        Condition1         Condition2          BldgType
##  Length:1460        Length:1460        Length:1460        Length:1460
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   HouseStyle         OverallQual      OverallCond      YearBuilt
##  Length:1460        Min.   : 1.000   Min.   :1.000   Min.   :1872
##  Class :character   1st Qu.: 5.000   1st Qu.:5.000   1st Qu.:1954
##  Mode  :character   Median : 6.000   Median :5.000   Median :1973
##                     Mean   : 6.099   Mean   :5.575   Mean   :1971
##                     3rd Qu.: 7.000   3rd Qu.:6.000   3rd Qu.:2000
##                     Max.   :10.000   Max.   :9.000   Max.   :2010
##
##   YearRemodAdd   RoofStyle           RoofMatl         Exterior1st
##  Min.   :1950   Length:1460        Length:1460        Length:1460
##  1st Qu.:1967   Class :character   Class :character   Class :character
##  Median :1994   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1985
```

```
##  3rd Qu.:2004
##  Max.   :2010
##
##  Exterior2nd         MasVnrType          MasVnrArea      ExterQual
##  Length:1460        Length:1460        Min.   :   0.0   Length:1460
##  Class :character   Class :character   1st Qu.:   0.0   Class :character
##  Mode  :character   Mode  :character   Median :   0.0   Mode  :character
##                                        Mean   : 103.7
##                                        3rd Qu.: 166.0
##                                        Max.   :1600.0
##                                        NA's   :8
##   ExterCond          Foundation          BsmtQual           BsmtCond
##  Length:1460        Length:1460        Length:1460        Length:1460
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  BsmtExposure       BsmtFinType1         BsmtFinSF1     BsmtFinType2
##  Length:1460        Length:1460        Min.   :   0.0   Length:1460
##  Class :character   Class :character   1st Qu.:   0.0   Class :character
##  Mode  :character   Mode  :character   Median : 383.5   Mode  :character
##                                        Mean   : 443.6
##                                        3rd Qu.: 712.2
##                                        Max.   :5644.0
##
##    BsmtFinSF2        BsmtUnfSF       TotalBsmtSF       Heating
##  Min.   :   0.00   Min.   :   0.0   Min.   :   0.0   Length:1460
##  1st Qu.:   0.00   1st Qu.: 223.0   1st Qu.: 795.8   Class :character
##  Median :   0.00   Median : 477.5   Median : 991.5   Mode  :character
##  Mean   :  46.55   Mean   : 567.2   Mean   :1057.4
##  3rd Qu.:   0.00   3rd Qu.: 808.0   3rd Qu.:1298.2
##  Max.   :1474.00   Max.   :2336.0   Max.   :6110.0
##
##   HeatingQC          CentralAir         Electrical          X1stFlrSF
##  Length:1460        Length:1460        Length:1460        Min.   : 334
##  Class :character   Class :character   Class :character   1st Qu.: 882
##  Mode  :character   Mode  :character   Mode  :character   Median :1087
##                                                           Mean   :1163
##                                                           3rd Qu.:1391
##                                                           Max.   :4692
##
##    X2ndFlrSF     LowQualFinSF       GrLivArea     BsmtFullBath
##  Min.   :   0   Min.   :  0.000   Min.   : 334   Min.   :0.0000
##  1st Qu.:   0   1st Qu.:  0.000   1st Qu.:1130   1st Qu.:0.0000
##  Median :   0   Median :  0.000   Median :1464   Median :0.0000
##  Mean   : 347   Mean   :  5.845   Mean   :1515   Mean   :0.4253
##  3rd Qu.: 728   3rd Qu.:  0.000   3rd Qu.:1777   3rd Qu.:1.0000
##  Max.   :2065   Max.   :572.000   Max.   :5642   Max.   :3.0000
##
##   BsmtHalfBath        FullBath        HalfBath       BedroomAbvGr
##  Min.   :0.00000   Min.   :0.000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:2.000
```

```
##  Median :0.00000   Median :2.000   Median :0.0000   Median :3.000
##  Mean   :0.05753   Mean   :1.565   Mean   :0.3829   Mean   :2.866
##  3rd Qu.:0.00000   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:3.000
##  Max.   :2.00000   Max.   :3.000   Max.   :2.0000   Max.   :8.000
##
##   KitchenAbvGr   KitchenQual         TotRmsAbvGrd     Functional
##  Min.   :0.000   Length:1460        Min.   : 2.000   Length:1460
##  1st Qu.:1.000   Class :character   1st Qu.: 5.000   Class :character
##  Median :1.000   Mode  :character   Median : 6.000   Mode  :character
##  Mean   :1.047                      Mean   : 6.518
##  3rd Qu.:1.000                      3rd Qu.: 7.000
##  Max.   :3.000                      Max.   :14.000
##
##    Fireplaces    FireplaceQu         GarageType         GarageYrBlt
##  Min.   :0.000   Length:1460        Length:1460        Min.   :1900
##  1st Qu.:0.000   Class :character   Class :character   1st Qu.:1961
##  Median :1.000   Mode  :character   Mode  :character   Median :1980
##  Mean   :0.613                                         Mean   :1979
##  3rd Qu.:1.000                                         3rd Qu.:2002
##  Max.   :3.000                                         Max.   :2010
##                                                        NA's   :81
##  GarageFinish         GarageCars      GarageArea      GarageQual
##  Length:1460        Min.   :0.000   Min.   :   0.0   Length:1460
##  Class :character   1st Qu.:1.000   1st Qu.: 334.5   Class :character
##  Mode  :character   Median :2.000   Median : 480.0   Mode  :character
##                     Mean   :1.767   Mean   : 473.0
##                     3rd Qu.:2.000   3rd Qu.: 576.0
##                     Max.   :4.000   Max.   :1418.0
##
##   GarageCond         PavedDrive          WoodDeckSF      OpenPorchSF
##  Length:1460        Length:1460        Min.   :  0.00   Min.   :  0.00
##  Class :character   Class :character   1st Qu.:  0.00   1st Qu.:  0.00
##  Mode  :character   Mode  :character   Median :  0.00   Median : 25.00
##                                        Mean   : 94.24   Mean   : 46.66
##                                        3rd Qu.:168.00   3rd Qu.: 68.00
##                                        Max.   :857.00   Max.   :547.00
##
##  EnclosedPorch      X3SsnPorch      ScreenPorch        PoolArea
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :  0.000
##  1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.000
##  Median :  0.00   Median :  0.00   Median :  0.00   Median :  0.000
##  Mean   : 21.95   Mean   :  3.41   Mean   : 15.06   Mean   :  2.759
##  3rd Qu.:  0.00   3rd Qu.:  0.00   3rd Qu.:  0.00   3rd Qu.:  0.000
##  Max.   :552.00   Max.   :508.00   Max.   :480.00   Max.   :738.000
##
##     PoolQC             Fence           MiscFeature          MiscVal
##  Length:1460        Length:1460        Length:1460        Min.   :    0.00
##  Class :character   Class :character   Class :character   1st Qu.:    0.00
##  Mode  :character   Mode  :character   Mode  :character   Median :    0.00
##                                                           Mean   :   43.49
##                                                           3rd Qu.:    0.00
##                                                           Max.   :15500.00
##
##      MoSold           YrSold       SaleType         SaleCondition
```

```
##  Min.   : 1.000   Min.   :2006   Length:1460        Length:1460
##  1st Qu.: 5.000   1st Qu.:2007   Class :character   Class :character
##  Median : 6.000   Median :2008   Mode  :character   Mode  :character
##  Mean   : 6.322   Mean   :2008
##  3rd Qu.: 8.000   3rd Qu.:2009
##  Max.   :12.000   Max.   :2010
##
##    SalePrice
##  Min.   : 34900
##  1st Qu.:129975
##  Median :163000
##  Mean   :180921
##  3rd Qu.:214000
##  Max.   :755000
##
```

DATA DESCRIPTION

The data contains about 80 independent variables, with the dependent variable being the sales price. In total there are 1460 observations.

```
NA_values <- data.frame(no_of_na_values=colSums(is.na(train_data)))
head(NA_values,80)
```

```
##               no_of_na_values
## Id                          0
## MSSubClass                  0
## MSZoning                    0
## LotFrontage               259
## LotArea                     0
## Street                      0
## Alley                    1369
## LotShape                    0
## LandContour                 0
## Utilities                   0
## LotConfig                   0
## LandSlope                   0
## Neighborhood                0
## Condition1                  0
## Condition2                  0
## BldgType                    0
## HouseStyle                  0
## OverallQual                 0
## OverallCond                 0
## YearBuilt                   0
## YearRemodAdd                0
## RoofStyle                   0
## RoofMatl                    0
## Exterior1st                 0
## Exterior2nd                 0
## MasVnrType                  8
## MasVnrArea                  8
## ExterQual                   0
```

```
## ExterCond                0
## Foundation               0
## BsmtQual                37
## BsmtCond                37
## BsmtExposure            38
## BsmtFinType1            37
## BsmtFinSF1               0
## BsmtFinType2            38
## BsmtFinSF2               0
## BsmtUnfSF                0
## TotalBsmtSF              0
## Heating                  0
## HeatingQC                0
## CentralAir               0
## Electrical               1
## X1stFlrSF                0
## X2ndFlrSF                0
## LowQualFinSF             0
## GrLivArea                0
## BsmtFullBath             0
## BsmtHalfBath             0
## FullBath                 0
## HalfBath                 0
## BedroomAbvGr             0
## KitchenAbvGr             0
## KitchenQual              0
## TotRmsAbvGrd             0
## Functional               0
## Fireplaces               0
## FireplaceQu            690
## GarageType              81
## GarageYrBlt             81
## GarageFinish            81
## GarageCars               0
## GarageArea               0
## GarageQual              81
## GarageCond              81
## PavedDrive               0
## WoodDeckSF               0
## OpenPorchSF              0
## EnclosedPorch            0
## X3SsnPorch               0
## ScreenPorch              0
## PoolArea                 0
## PoolQC                1453
## Fence                 1179
## MiscFeature           1406
## MiscVal                  0
## MoSold                   0
## YrSold                   0
## SaleType                 0
## SaleCondition            0
```

There are several missing values in the data, most notably Alley, PoolQC, Fence, and MiscFeature. We will

drop these features with the most missing values.

```
train_data <- train_data[, -c(7, 60, 73, 74, 75)]
head(train_data)
```

```
##   Id MSSubClass MSZoning LotFrontage LotArea Street LotShape LandContour
## 1  1         60       RL          65    8450   Pave      Reg         Lvl
## 2  2         20       RL          80    9600   Pave      Reg         Lvl
## 3  3         60       RL          68   11250   Pave      IR1         Lvl
## 4  4         70       RL          60    9550   Pave      IR1         Lvl
## 5  5         60       RL          84   14260   Pave      IR1         Lvl
## 6  6         50       RL          85   14115   Pave      IR1         Lvl
##   Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType
## 1    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 2    AllPub       FR2       Gtl      Veenker      Feedr       Norm     1Fam
## 3    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 4    AllPub    Corner       Gtl      Crawfor       Norm       Norm     1Fam
## 5    AllPub       FR2       Gtl      NoRidge       Norm       Norm     1Fam
## 6    AllPub    Inside       Gtl      Mitchel       Norm       Norm     1Fam
##   HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl
## 1     2Story           7           5      2003         2003     Gable  CompShg
## 2     1Story           6           8      1976         1976     Gable  CompShg
## 3     2Story           7           5      2001         2002     Gable  CompShg
## 4     2Story           7           5      1915         1970     Gable  CompShg
## 5     2Story           8           5      2000         2000     Gable  CompShg
## 6     1.5Fin           5           5      1993         1995     Gable  CompShg
##   Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond Foundation
## 1     VinylSd     VinylSd    BrkFace        196        Gd        TA      PConc
## 2     MetalSd     MetalSd       None          0        TA        TA     CBlock
## 3     VinylSd     VinylSd    BrkFace        162        Gd        TA      PConc
## 4     Wd Sdng     Wd Shng       None          0        TA        TA     BrkTil
## 5     VinylSd     VinylSd    BrkFace        350        Gd        TA      PConc
## 6     VinylSd     VinylSd       None          0        TA        TA       Wood
##   BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
## 1       Gd       TA           No          GLQ        706          Unf
## 2       Gd       TA           Gd          ALQ        978          Unf
## 3       Gd       TA           Mn          GLQ        486          Unf
## 4       TA       Gd           No          ALQ        216          Unf
## 5       Gd       TA           Av          GLQ        655          Unf
## 6       Gd       TA           No          GLQ        732          Unf
##   BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir Electrical
## 1          0       150         856    GasA        Ex          Y      SBrkr
## 2          0       284        1262    GasA        Ex          Y      SBrkr
## 3          0       434         920    GasA        Ex          Y      SBrkr
## 4          0       540         756    GasA        Gd          Y      SBrkr
## 5          0       490        1145    GasA        Ex          Y      SBrkr
## 6          0        64         796    GasA        Ex          Y      SBrkr
##   X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath
## 1       856       854            0      1710            1            0        2
## 2      1262         0            0      1262            0            1        2
## 3       920       866            0      1786            1            0        2
## 4       961       756            0      1717            1            0        1
## 5      1145      1053            0      2198            1            0        2
## 6       796       566            0      1362            1            0        1
```

```
##   HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## 1        1            3            1          Gd            8        Typ
## 2        0            3            1          TA            6        Typ
## 3        1            3            1          Gd            6        Typ
## 4        0            3            1          Gd            7        Typ
## 5        1            4            1          Gd            9        Typ
## 6        1            1            1          TA            5        Typ
##   Fireplaces FireplaceQu GarageType GarageFinish GarageCars GarageArea
## 1          0       <NA>     Attchd          RFn          2        548
## 2          1         TA     Attchd          RFn          2        460
## 3          1         TA     Attchd          RFn          2        608
## 4          1         Gd     Detchd          Unf          3        642
## 5          1         TA     Attchd          RFn          3        836
## 6          0       <NA>     Attchd          Unf          2        480
##   GarageQual GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch
## 1         TA         TA          Y          0          61             0
## 2         TA         TA          Y        298           0             0
## 3         TA         TA          Y          0          42             0
## 4         TA         TA          Y          0          35           272
## 5         TA         TA          Y        192          84             0
## 6         TA         TA          Y         40          30             0
##   X3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold SaleType SaleCondition
## 1          0           0        0       0      2   2008       WD        Normal
## 2          0           0        0       0      5   2007       WD        Normal
## 3          0           0        0       0      9   2008       WD        Normal
## 4          0           0        0       0      2   2006       WD       Abnorml
## 5          0           0        0       0     12   2008       WD        Normal
## 6        320           0        0     700     10   2009       WD        Normal
##   SalePrice
## 1    208500
## 2    181500
## 3    223500
## 4    140000
## 5    250000
## 6    143000
```

For the rest of the missing values, we will use the mode to estimate them. For LotFrontage, we will use the mean for a better estimate.

```
mode_value <- names(sort(table(train_data$BsmtCond), decreasing = TRUE))[1]
train_data$BsmtCond <- ifelse(is.na(train_data$BsmtCond), mode_value, train_data$BsmtCond)
mode_value2 <- names(sort(table(train_data$BsmtQual), decreasing = TRUE))[1]
train_data$BsmtQual <- ifelse(is.na(train_data$BsmtQual), mode_value2, train_data$BsmtQual)
mode_value3 <- names(sort(table(train_data$GarageType), decreasing = TRUE))[1]
train_data$GarageType <- ifelse(is.na(train_data$GarageType), mode_value3, train_data$GarageType)
mode_value4 <- names(sort(table(train_data$GarageQual), decreasing = TRUE))[1]
train_data$GarageQual <- ifelse(is.na(train_data$GarageQual), mode_value4, train_data$GarageQual)
mode_value5 <- names(sort(table(train_data$GarageCond), decreasing = TRUE))[1]
train_data$GarageCond <- ifelse(is.na(train_data$GarageCond), mode_value5, train_data$GarageCond)
mode_value6 <- names(sort(table(train_data$GarageFinish), decreasing = TRUE))[1]
train_data$GarageFinish <- ifelse(is.na(train_data$GarageFinish), mode_value6, train_data$GarageFinish)
mode_value7 <- names(sort(table(train_data$FireplaceQu), decreasing = TRUE))[1]
train_data$FireplaceQu <- ifelse(is.na(train_data$FireplaceQu), mode_value7, train_data$FireplaceQu)
train_data$LotFrontage <- na.aggregate(train_data$LotFrontage, FUN = mean, na.rm = TRUE)
```

```
head(train_data)
```

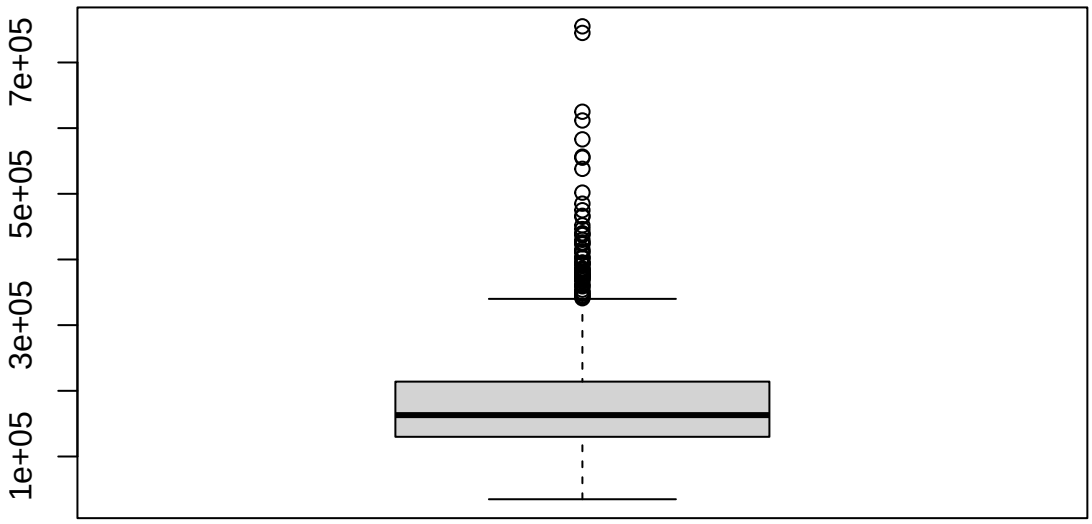
```
##   Id MSSubClass MSZoning LotFrontage LotArea Street LotShape LandContour
## 1  1         60       RL          65    8450   Pave      Reg         Lvl
## 2  2         20       RL          80    9600   Pave      Reg         Lvl
## 3  3         60       RL          68   11250   Pave      IR1         Lvl
## 4  4         70       RL          60    9550   Pave      IR1         Lvl
## 5  5         60       RL          84   14260   Pave      IR1         Lvl
## 6  6         50       RL          85   14115   Pave      IR1         Lvl
##   Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType
## 1    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 2    AllPub       FR2       Gtl      Veenker      Feedr       Norm     1Fam
## 3    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 4    AllPub    Corner       Gtl      Crawfor       Norm       Norm     1Fam
## 5    AllPub       FR2       Gtl      NoRidge       Norm       Norm     1Fam
## 6    AllPub    Inside       Gtl      Mitchel       Norm       Norm     1Fam
##   HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl
## 1     2Story           7           5      2003         2003     Gable  CompShg
## 2     1Story           6           8      1976         1976     Gable  CompShg
## 3     2Story           7           5      2001         2002     Gable  CompShg
## 4     2Story           7           5      1915         1970     Gable  CompShg
## 5     2Story           8           5      2000         2000     Gable  CompShg
## 6     1.5Fin           5           5      1993         1995     Gable  CompShg
##   Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond Foundation
## 1     VinylSd     VinylSd    BrkFace        196        Gd        TA      PConc
## 2     MetalSd     MetalSd       None          0        TA        TA     CBlock
## 3     VinylSd     VinylSd    BrkFace        162        Gd        TA      PConc
## 4     Wd Sdng     Wd Shng       None          0        TA        TA     BrkTil
## 5     VinylSd     VinylSd    BrkFace        350        Gd        TA      PConc
## 6     VinylSd     VinylSd       None          0        TA        TA       Wood
##   BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
## 1       Gd       TA           No          GLQ        706          Unf
## 2       Gd       TA           Gd          ALQ        978          Unf
## 3       Gd       TA           Mn          GLQ        486          Unf
## 4       TA       Gd           No          ALQ        216          Unf
## 5       Gd       TA           Av          GLQ        655          Unf
## 6       Gd       TA           No          GLQ        732          Unf
##   BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir Electrical
## 1          0       150         856    GasA        Ex          Y      SBrkr
## 2          0       284        1262    GasA        Ex          Y      SBrkr
## 3          0       434         920    GasA        Ex          Y      SBrkr
## 4          0       540         756    GasA        Gd          Y      SBrkr
## 5          0       490        1145    GasA        Ex          Y      SBrkr
## 6          0        64         796    GasA        Ex          Y      SBrkr
##   X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath
## 1       856       854            0      1710            1            0        2
## 2      1262         0            0      1262            0            1        2
## 3       920       866            0      1786            1            0        2
## 4       961       756            0      1717            1            0        1
## 5      1145      1053            0      2198            1            0        2
## 6       796       566            0      1362            1            0        1
##   HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## 1        1            3            1          Gd            8        Typ
```

```
## 2        0            3            1          TA            6        Typ
## 3        1            3            1          Gd            6        Typ
## 4        0            3            1          Gd            7        Typ
## 5        1            4            1          Gd            9        Typ
## 6        1            1            1          TA            5        Typ
##   Fireplaces FireplaceQu GarageType GarageFinish GarageCars GarageArea
## 1          0          Gd     Attchd          RFn          2        548
## 2          1          TA     Attchd          RFn          2        460
## 3          1          TA     Attchd          RFn          2        608
## 4          1          Gd     Detchd          Unf          3        642
## 5          1          TA     Attchd          RFn          3        836
## 6          0          Gd     Attchd          Unf          2        480
##   GarageQual GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch
## 1         TA         TA          Y          0          61             0
## 2         TA         TA          Y        298           0             0
## 3         TA         TA          Y          0          42             0
## 4         TA         TA          Y          0          35           272
## 5         TA         TA          Y        192          84             0
## 6         TA         TA          Y         40          30             0
##   X3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold SaleType SaleCondition
## 1          0           0        0       0      2   2008       WD        Normal
## 2          0           0        0       0      5   2007       WD        Normal
## 3          0           0        0       0      9   2008       WD        Normal
## 4          0           0        0       0      2   2006       WD       Abnorml
## 5          0           0        0       0     12   2008       WD        Normal
## 6        320           0        0     700     10   2009       WD        Normal
##   SalePrice
## 1    208500
## 2    181500
## 3    223500
## 4    140000
## 5    250000
## 6    143000
```

EXPLORATORY DATA ANALYSIS

We start off by creating a boxplot and histogram of the data.

```
boxplot(train_data$SalePrice)
```

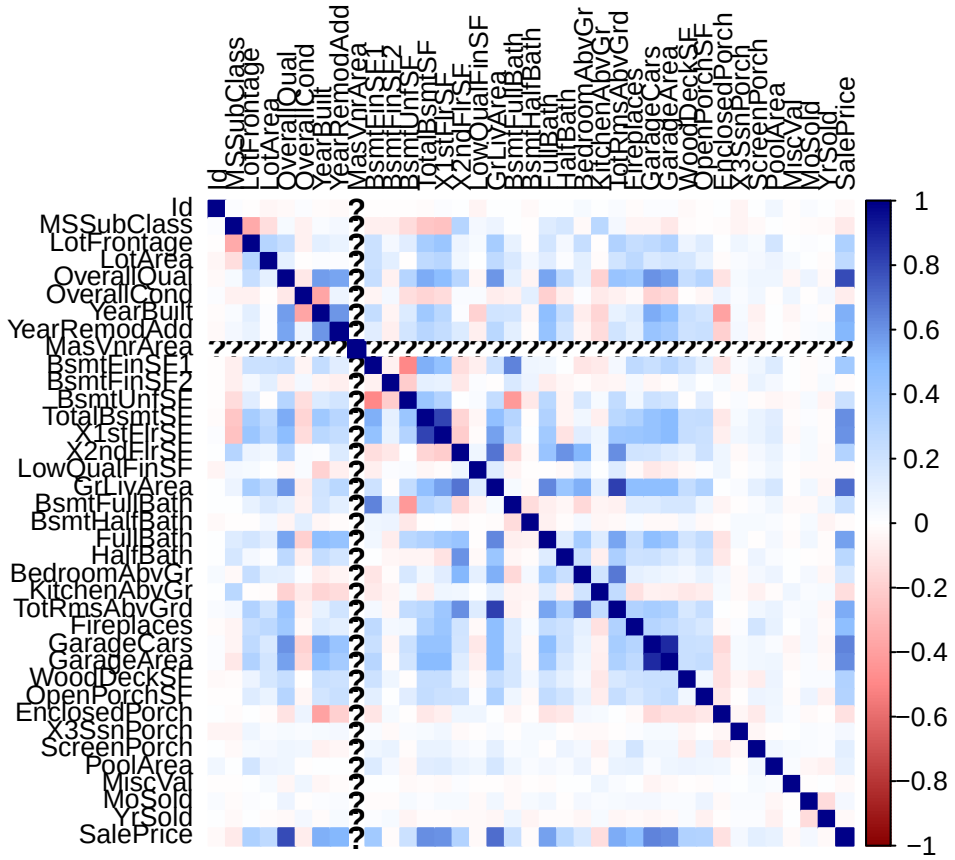



```
hist(train_data$SalePrice)
```

The price is skewed to the right with several very high prices. Since the response is not normally distributed, we may consider to use neural network which does not need to use the normality assumption of the response, or use log transformation of the response.

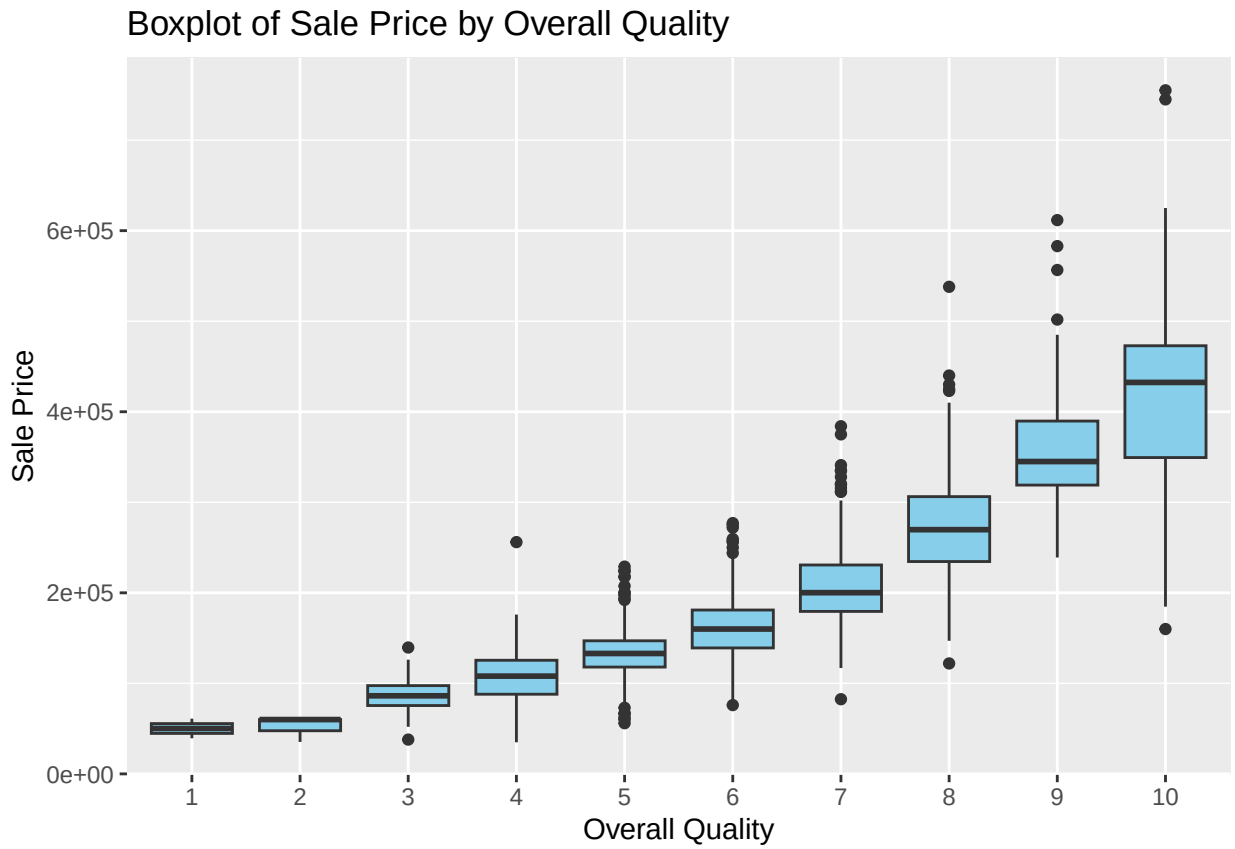
Next we will create a correlation plot to determine the association between the dependent variable SalesPrice and the independent variables.

```
data_num <- train_data %>% dplyr::select(where(is.numeric))
correlation <- cor(data_num)
col <- colorRampPalette(c("#880000", "#FF8888", "#FFFFFF", "#88BBFF", "#000088")) # Darker blue and red
corrplot(correlation, method = "color", col = col(200), tl.cex = 0.8, tl.col = "black")
```



From the correlation plot, we see the Overall Quality of a house (OverallQual) is most strongly correlated with the Sales Price. We may further analyze this relationship through a boxplot.

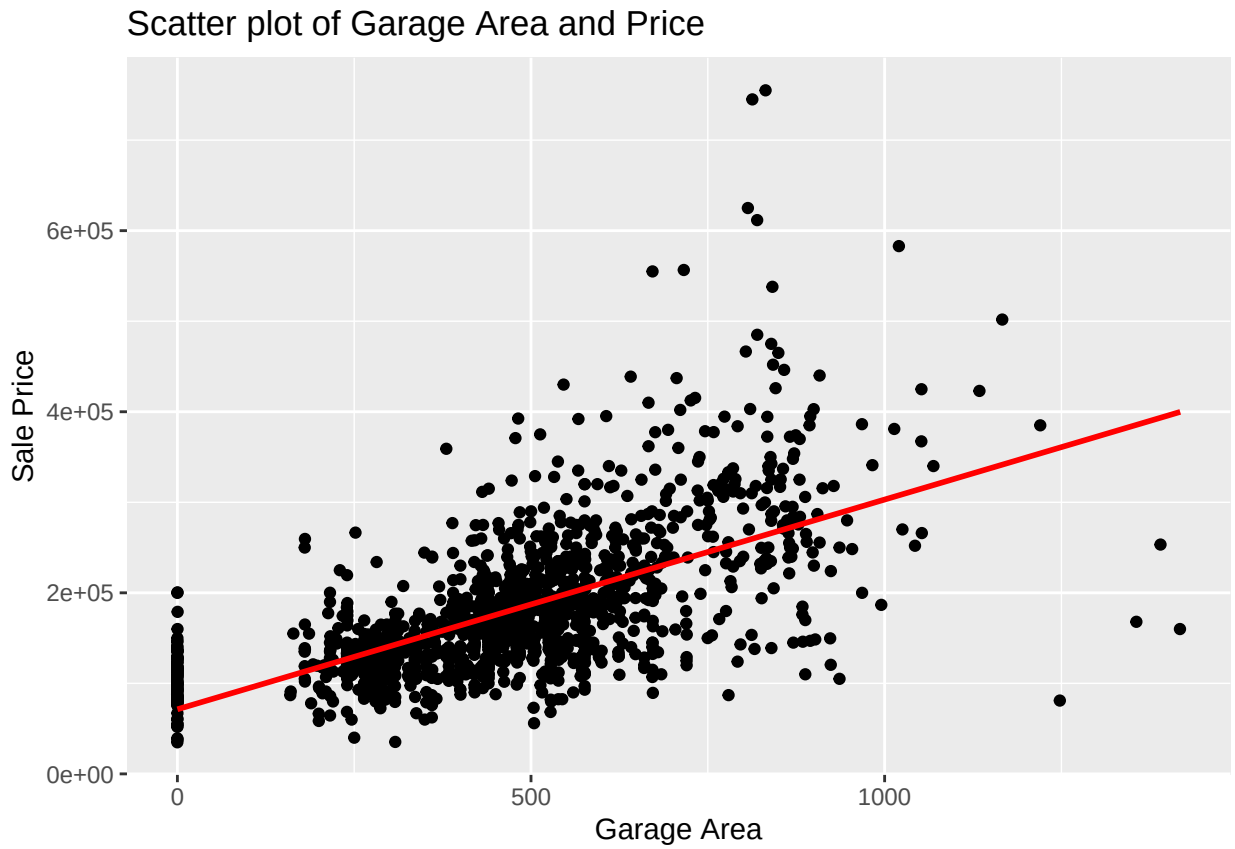
```
ggplot(train_data, aes(x = as.factor(OverallQual), y = SalePrice)) +
 geom_boxplot(fill = "skyblue") +
 labs(x = "Overall Quality", y = "Sale Price") +
 ggtitle("Boxplot of Sale Price by Overall Quality")
```



Let us now see the Scatter Plots and Regression Lines of a few other variables with respect to the Sale Price.

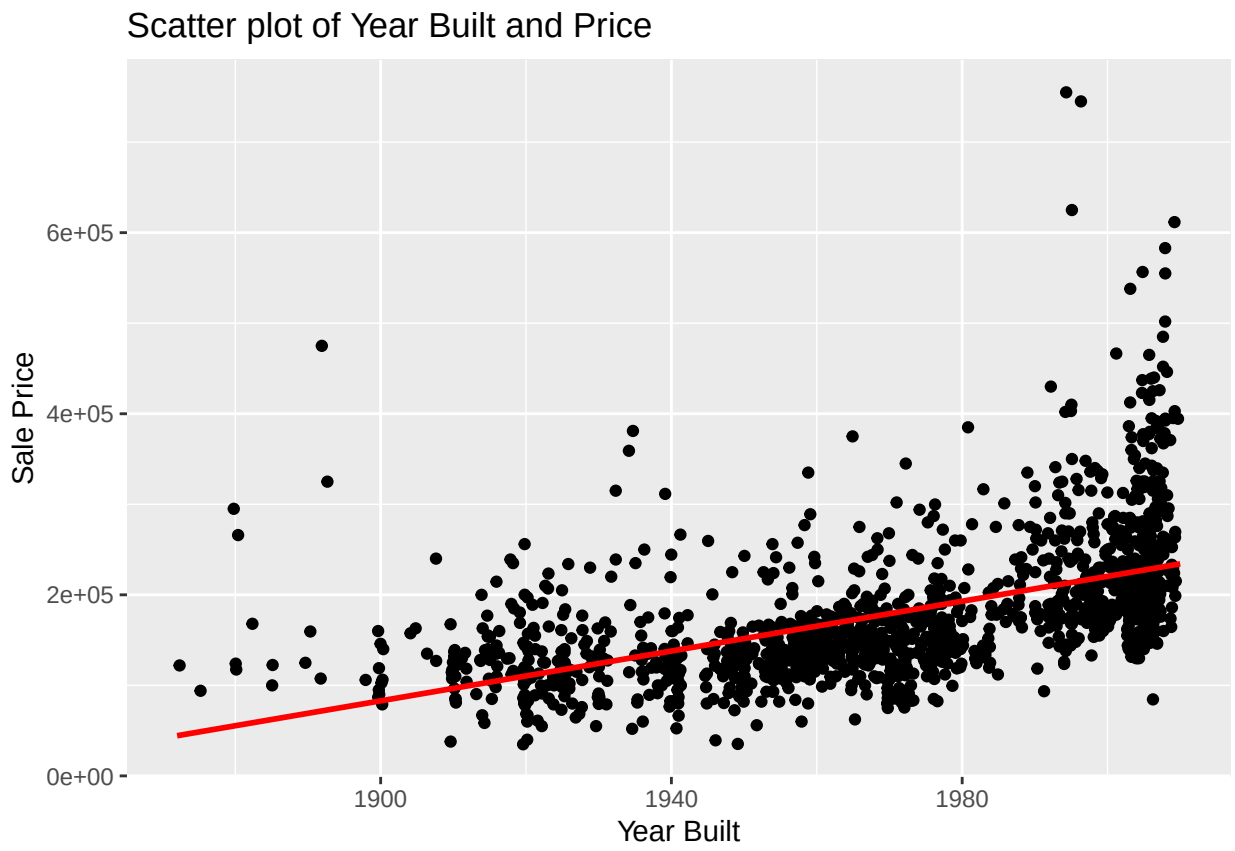
```
ggplot(data = train_data, aes(x = GarageArea, y = SalePrice)) +
  geom_jitter() +  geom_smooth(method = "lm", se = FALSE, color = 'red')+labs(title="Scatter plot of Ga
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



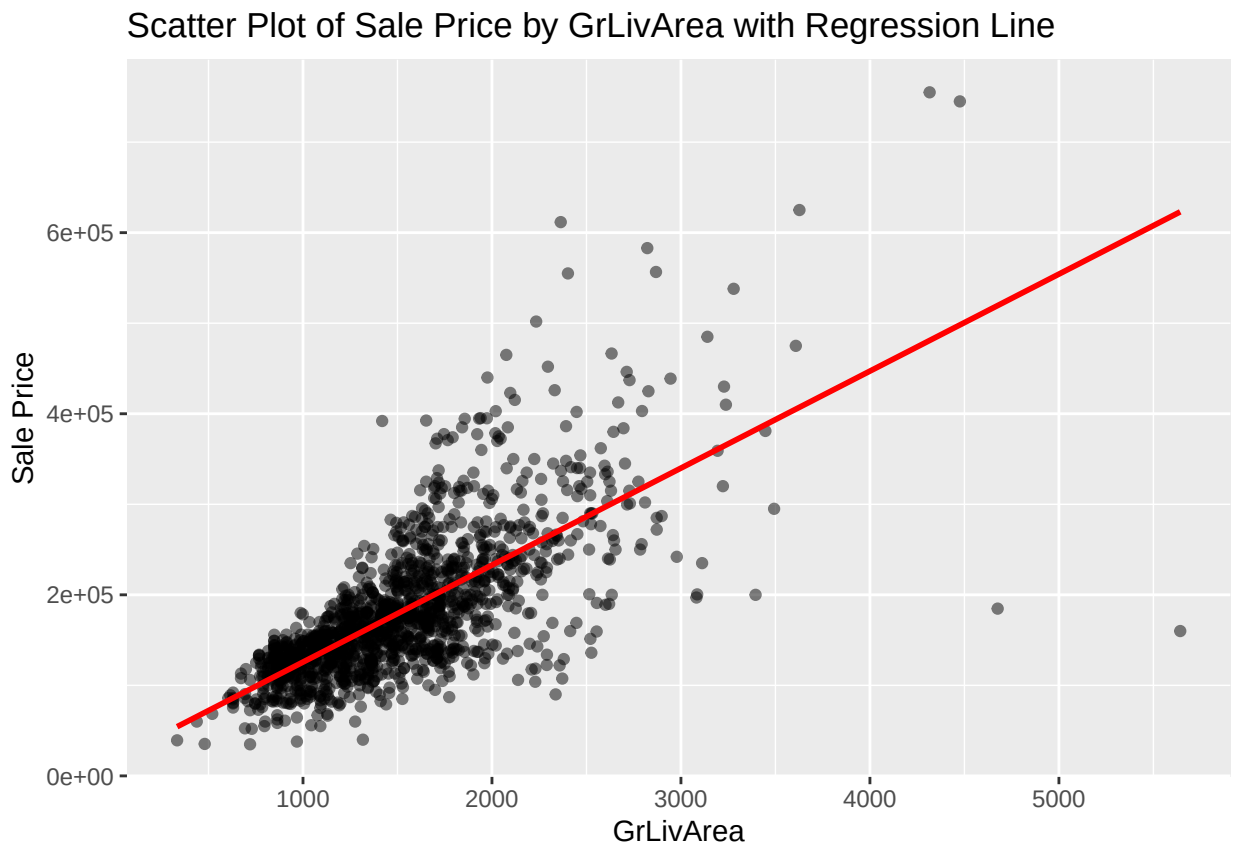

```
ggplot(data = train_data, aes(x = YearBuilt, y = SalePrice)) +
  geom_jitter() +  geom_smooth(method = "lm", se = FALSE, color = 'red')+labs(title="Scatter plot of Yea
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(train_data, aes(x = GrLivArea, y = SalePrice)) +
 geom_point(color = "black", alpha = 0.5) +
 geom_smooth(method = "lm", se = FALSE, color = "red") +
 labs(x = "GrLivArea", y = "Sale Price") +
 ggtitle("Scatter Plot of Sale Price by GrLivArea with Regression Line")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



We see that we have a significantly large number of outliers on each plot.

Treating or altering the outlier/extreme values in genuine observations is not a standard operating procedure. However, it is essential to understand their impact on our predictive models.

To better understand the implications of outliers better, we should compare the fit of a simple linear regression model on the dataset with and without outliers. For this we first extract outliers from the data and then obtain the data without the outliers.

```
outliers=boxplot(train_data$SalePrice,plot=FALSE)$out
outliers_data=train_data[which(train_data$SalePrice %in% outliers),]
train_data1= train_data[-which(train_data$SalePrice %in% outliers),]
```

Using the GrLivArea variable, which is strongly correlated with Sales Price, we can demonstrate the impact of outliers in the data

```
q1 <- quantile(train_data$GrLivArea, 0.25)
q3 <- quantile(train_data$GrLivArea, 0.75)
iqr <- q3 - q1
lower_bound <- q1 - 1.5 * iqr
upper_bound <- q3 + 1.5 * iqr
filtered_data <- train_data %>%
filter(GrLivArea >= lower_bound & GrLivArea <= upper_bound)
```

```
ggplot(filtered_data, aes(x = GrLivArea, y = SalePrice)) +
 geom_point(color = "black", alpha = 0.5) +
```

```
  geom_smooth(method = "lm", se = FALSE, color = "red") + # Add regression line
  labs(x = "GrLivArea", y = "Sale Price") +
  ggtitle("Scatter Plot of Sale Price by GrLivArea with Regression Line (Outliers Removed)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

##Modelling

From here, we can clean our data and create our regression model

```
cleaned_data <- na.omit(train_data)
cleaned_data <- as.data.frame(model.matrix(~ . - 1, data = cleaned_data))
lm_model <- lm(SalePrice ~ ., data = cleaned_data)
summary(lm_model)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = cleaned_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -182595   -9332     608    9615  182595
##
## Coefficients: (4 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          -8.618e+05  1.077e+06  -0.801 0.423575
## Id                    8.013e-01  1.592e+00   0.503 0.614824
## MSSubClass           -4.577e+01  8.558e+01  -0.535 0.592875
## `MSZoningC (all)`    -2.517e+04  9.727e+03  -2.588 0.009773 **
## MSZoningFV            1.240e+04  7.577e+03   1.636 0.102139
## MSZoningRH            2.155e+03  7.940e+03   0.271 0.786116
## MSZoningRL            4.002e+03  3.880e+03   1.031 0.302553
## MSZoningRM                   NA         NA      NA       NA
## LotFrontage           5.662e+01  4.413e+01   1.283 0.199791
## LotArea               7.725e-01  1.111e-01   6.956 5.78e-12 ***
## StreetPave            3.098e+04  1.231e+04   2.517 0.011982 *
## LotShapeIR2           3.648e+03  4.351e+03   0.838 0.401945
## LotShapeIR3           2.433e+03  9.128e+03   0.267 0.789880
## LotShapeReg           1.600e+03  1.644e+03   0.973 0.330709
## LandContourHLS        8.846e+03  5.265e+03   1.680 0.093181 .
## LandContourLow       -8.640e+03  6.748e+03  -1.280 0.200646
## LandContourLvl        6.432e+03  3.841e+03   1.674 0.094299 .
## UtilitiesNoSeWa      -3.052e+04  2.678e+04  -1.140 0.254675
## LotConfigCulDSac      9.036e+03  3.404e+03   2.655 0.008046 **
## LotConfigFR2         -8.102e+03  4.123e+03  -1.965 0.049613 *
## LotConfigFR3         -1.519e+04  1.275e+04  -1.192 0.233562
## LotConfigInside      -1.588e+03  1.830e+03  -0.868 0.385696
## LandSlopeMod          7.134e+03  4.097e+03   1.741 0.081912 .
## LandSlopeSev         -4.413e+04  1.163e+04  -3.794 0.000156 ***
## NeighborhoodBlueste  -2.730e+01  1.928e+04  -0.001 0.998871
## NeighborhoodBrDale    1.891e+02  1.117e+04   0.017 0.986489
## NeighborhoodBrkSide  -2.900e+03  9.641e+03  -0.301 0.763657
## NeighborhoodClearCr  -1.546e+04  9.440e+03  -1.638 0.101669
## NeighborhoodCollgCr  -1.055e+04  7.383e+03  -1.428 0.153475
## NeighborhoodCrawfor   1.236e+04  8.726e+03   1.416 0.156915
## NeighborhoodEdwards  -1.913e+04  8.195e+03  -2.334 0.019771 *
## NeighborhoodGilbert  -1.197e+04  7.821e+03  -1.530 0.126303
## NeighborhoodIDOTRR   -8.092e+03  1.096e+04  -0.738 0.460603
## NeighborhoodMeadowV  -7.235e+03  1.140e+04  -0.635 0.525852
## NeighborhoodMitchel  -2.242e+04  8.345e+03  -2.687 0.007318 **
## NeighborhoodNAmes    -1.697e+04  7.956e+03  -2.133 0.033160 *
## NeighborhoodNoRidge   2.399e+04  8.594e+03   2.792 0.005330 **
## NeighborhoodNPkVill   1.409e+04  1.430e+04   0.985 0.324726
## NeighborhoodNridgHt   1.698e+04  7.662e+03   2.216 0.026902 *
## NeighborhoodNWAmes   -1.907e+04  8.162e+03  -2.336 0.019650 *
## NeighborhoodOldTown  -1.217e+04  9.814e+03  -1.240 0.215194
## NeighborhoodSawyer   -1.107e+04  8.271e+03  -1.339 0.180869
## NeighborhoodSawyerW  -4.886e+03  7.965e+03  -0.614 0.539656
## NeighborhoodSomerst  -2.670e+03  9.180e+03  -0.291 0.771173
## NeighborhoodStoneBr   3.688e+04  8.448e+03   4.365 1.38e-05 ***
## NeighborhoodSWISU    -6.633e+03  9.845e+03  -0.674 0.500568
## NeighborhoodTimber   -1.086e+04  8.289e+03  -1.310 0.190539
## NeighborhoodVeenker   1.002e+03  1.069e+04   0.094 0.925335
## Condition1Feedr       5.409e+03  5.170e+03   1.046 0.295679
## Condition1Norm        1.462e+04  4.290e+03   3.409 0.000675 ***
## Condition1PosA        6.480e+03  1.017e+04   0.637 0.524185
## Condition1PosN        1.221e+04  7.588e+03   1.609 0.107877
## Condition1RRAe       -1.597e+04  9.649e+03  -1.656 0.098076 .
## Condition1RRAn        8.367e+03  6.968e+03   1.201 0.230085
```

```
## Condition1RRNe       -1.795e+03  1.781e+04  -0.101 0.919764
## Condition1RRNn        6.784e+03  1.310e+04   0.518 0.604654
## Condition2Feedr      -2.434e+03  2.414e+04  -0.101 0.919710
## Condition2Norm       -3.788e+03  2.097e+04  -0.181 0.856666
## Condition2PosA        4.017e+04  3.783e+04   1.062 0.288507
## Condition2PosN       -2.385e+05  2.836e+04  -8.409  < 2e-16 ***
## Condition2RRAe       -1.224e+05  4.720e+04  -2.592 0.009646 **
## Condition2RRAn       -1.391e+04  3.233e+04  -0.430 0.667100
## Condition2RRNn        6.076e+03  2.783e+04   0.218 0.827242
## BldgType2fmCon       -4.582e+03  1.292e+04  -0.355 0.722911
## BldgTypeDuplex       -1.068e+04  7.650e+03  -1.396 0.163016
## BldgTypeTwnhs        -1.856e+04  1.030e+04  -1.802 0.071746 .
## BldgTypeTwnhsE       -1.530e+04  9.322e+03  -1.642 0.100923
## HouseStyle1.5Unf      1.645e+04  8.069e+03   2.039 0.041710 *
## HouseStyle1Story      9.149e+03  4.526e+03   2.021 0.043479 *
## HouseStyle2.5Fin     -2.354e+04  1.238e+04  -1.902 0.057474 .
## HouseStyle2.5Unf     -6.893e+03  9.371e+03  -0.736 0.462171
## HouseStyle2Story     -4.779e+03  3.625e+03  -1.319 0.187588
## HouseStyleSFoyer      5.813e+03  6.596e+03   0.881 0.378317
## HouseStyleSLvl        4.925e+03  5.723e+03   0.861 0.389603
## OverallQual           6.402e+03  1.043e+03   6.137 1.15e-09 ***
## OverallCond           5.745e+03  8.929e+02   6.434 1.81e-10 ***
## YearBuilt             3.256e+02  7.695e+01   4.231 2.50e-05 ***
## YearRemodAdd          1.046e+02  5.669e+01   1.845 0.065341 .
## RoofStyleGable        8.127e+03  1.877e+04   0.433 0.665158
## RoofStyleGambrel      9.995e+03  2.046e+04   0.489 0.625265
## RoofStyleHip          8.242e+03  1.884e+04   0.438 0.661781
## RoofStyleMansard      2.199e+04  2.185e+04   1.006 0.314601
## RoofStyleShed         9.952e+04  3.523e+04   2.825 0.004805 **
## RoofMatlCompShg       6.778e+05  3.418e+04  19.833  < 2e-16 ***
## RoofMatlMembran       7.734e+05  4.894e+04  15.804  < 2e-16 ***
## RoofMatlMetal         7.465e+05  4.773e+04  15.642  < 2e-16 ***
## RoofMatlRoll          6.670e+05  4.276e+04  15.599  < 2e-16 ***
## `RoofMatlTar&Grv`     6.803e+05  3.899e+04  17.450  < 2e-16 ***
## RoofMatlWdShake       6.692e+05  3.786e+04  17.678  < 2e-16 ***
## RoofMatlWdShngl       7.306e+05  3.540e+04  20.636  < 2e-16 ***
## Exterior1stBrkComm   -4.126e+04  3.376e+04  -1.222 0.221928
## Exterior1stBrkFace    3.722e+03  1.361e+04   0.273 0.784613
## Exterior1stCBlock    -2.150e+04  2.784e+04  -0.772 0.440076
## Exterior1stCemntBd   -1.855e+04  1.978e+04  -0.938 0.348641
## Exterior1stHdBoard   -1.889e+04  1.373e+04  -1.377 0.168924
## Exterior1stImStucc   -4.588e+04  2.850e+04  -1.610 0.107660
## Exterior1stMetalSd   -1.079e+04  1.533e+04  -0.703 0.481962
## Exterior1stPlywood   -1.975e+04  1.360e+04  -1.452 0.146707
## Exterior1stStone     -1.132e+04  2.687e+04  -0.421 0.673795
## Exterior1stStucco    -1.707e+04  1.501e+04  -1.137 0.255574
## Exterior1stVinylSd   -1.966e+04  1.390e+04  -1.415 0.157444
## `Exterior1stWd Sdng` -1.832e+04  1.320e+04  -1.387 0.165571
## Exterior1stWdShing   -1.604e+04  1.421e+04  -1.129 0.259209
## Exterior2ndAsphShn    1.173e+04  2.294e+04   0.511 0.609219
## `Exterior2ndBrk Cmn`  9.604e+03  2.084e+04   0.461 0.645062
## Exterior2ndBrkFace    6.861e+03  1.416e+04   0.485 0.628007
## Exterior2ndCBlock            NA         NA      NA       NA
## Exterior2ndCmentBd    1.881e+04  1.951e+04   0.964 0.335177
```

```
## Exterior2ndHdBoard    1.266e+04  1.328e+04   0.953 0.340699
## Exterior2ndImStucc    3.074e+04  1.499e+04   2.051 0.040508 *
## Exterior2ndMetalSd    8.900e+03  1.498e+04   0.594 0.552517
## Exterior2ndOther     -1.071e+04  2.776e+04  -0.386 0.699672
## Exterior2ndPlywood    9.958e+03  1.296e+04   0.769 0.442328
## Exterior2ndStone     -8.773e+03  2.443e+04  -0.359 0.719541
## Exterior2ndStucco     1.489e+04  1.439e+04   1.035 0.301082
## Exterior2ndVinylSd    1.709e+04  1.342e+04   1.273 0.203139
## `Exterior2ndWd Sdng`  1.437e+04  1.282e+04   1.121 0.262530
## `Exterior2ndWd Shng`  1.004e+04  1.330e+04   0.755 0.450420
## MasVnrTypeBrkFace     8.418e+03  6.937e+03   1.213 0.225192
## MasVnrTypeNone        1.162e+04  7.001e+03   1.660 0.097257 .
## MasVnrTypeStone       1.346e+04  7.335e+03   1.835 0.066685 .
## MasVnrArea            2.082e+01  5.864e+00   3.550 0.000401 ***
## ExterQualFa          -6.565e+03  1.178e+04  -0.557 0.577512
## ExterQualGd          -1.902e+04  4.886e+03  -3.892 0.000105 ***
## ExterQualTA          -1.987e+04  5.399e+03  -3.681 0.000242 ***
## ExterCondFa          -2.436e+03  1.845e+04  -0.132 0.894956
## ExterCondGd          -9.243e+03  1.741e+04  -0.531 0.595668
## ExterCondPo           6.720e+03  3.200e+04   0.210 0.833722
## ExterCondTA          -6.667e+03  1.738e+04  -0.384 0.701357
## FoundationCBlock      2.914e+03  3.259e+03   0.894 0.371407
## FoundationPConc       4.511e+03  3.500e+03   1.289 0.197640
## FoundationStone       6.914e+03  1.123e+04   0.615 0.538401
## FoundationWood       -2.662e+04  1.497e+04  -1.778 0.075666 .
## BsmtQualFa           -1.231e+04  6.465e+03  -1.904 0.057102 .
## BsmtQualGd           -1.852e+04  3.394e+03  -5.456 5.92e-08 ***
## BsmtQualTA           -1.519e+04  4.203e+03  -3.614 0.000314 ***
## BsmtCondGd           -1.633e+03  5.392e+03  -0.303 0.762079
## BsmtCondPo            6.935e+04  3.093e+04   2.242 0.025140 *
## BsmtCondTA            2.582e+03  4.354e+03   0.593 0.553380
## BsmtExposureGd        1.315e+04  3.049e+03   4.312 1.75e-05 ***
## BsmtExposureMn       -4.227e+03  3.071e+03  -1.376 0.168975
## BsmtExposureNo       -5.837e+03  2.222e+03  -2.628 0.008710 **
## BsmtFinType1BLQ       1.650e+03  2.827e+03   0.584 0.559658
## BsmtFinType1GLQ       5.600e+03  2.559e+03   2.189 0.028819 *
## BsmtFinType1LwQ      -3.611e+03  3.798e+03  -0.951 0.341833
## BsmtFinType1Rec      -1.936e+02  3.036e+03  -0.064 0.949162
## BsmtFinType1Unf       2.976e+03  2.964e+03   1.004 0.315437
## BsmtFinSF1            3.965e+01  5.611e+00   7.067 2.69e-12 ***
## BsmtFinType2BLQ      -1.474e+04  7.667e+03  -1.923 0.054702 .
## BsmtFinType2GLQ      -3.148e+03  9.458e+03  -0.333 0.739334
## BsmtFinType2LwQ      -1.653e+04  7.481e+03  -2.210 0.027324 *
## BsmtFinType2Rec      -1.177e+04  7.212e+03  -1.632 0.102871
## BsmtFinType2Unf      -1.140e+04  7.656e+03  -1.489 0.136797
## BsmtFinSF2            2.894e+01  9.320e+00   3.105 0.001947 **
## BsmtUnfSF             2.119e+01  5.178e+00   4.093 4.55e-05 ***
## TotalBsmtSF                  NA         NA      NA       NA
## HeatingGasW          -5.351e+03  7.186e+03  -0.745 0.456611
## HeatingGrav          -2.812e+03  1.199e+04  -0.235 0.814527
## HeatingOthW          -2.109e+04  1.880e+04  -1.122 0.262176
## HeatingQCFa           3.314e+02  4.908e+03   0.068 0.946176
## HeatingQCGd          -3.391e+03  2.113e+03  -1.605 0.108743
## HeatingQCPo           4.644e+03  2.719e+04   0.171 0.864422
```

```
## HeatingQCTA          -3.205e+03  2.122e+03  -1.510 0.131230
## CentralAirY           6.552e+02  4.053e+03   0.162 0.871605
## ElectricalFuseF      -3.238e+03  6.425e+03  -0.504 0.614349
## ElectricalFuseP      -3.742e+02  2.259e+04  -0.017 0.986786
## ElectricalMix        -4.317e+04  4.590e+04  -0.940 0.347220
## ElectricalSBrkr      -1.739e+03  3.060e+03  -0.568 0.570074
## X1stFlrSF             4.724e+01  6.003e+00   7.871 7.93e-15 ***
## X2ndFlrSF             6.821e+01  5.741e+00  11.880  < 2e-16 ***
## LowQualFinSF          1.468e+01  1.860e+01   0.789 0.430078
## GrLivArea                    NA         NA      NA       NA
## BsmtFullBath          1.094e+03  2.009e+03   0.544 0.586206
## BsmtHalfBath         -9.672e+01  3.067e+03  -0.032 0.974848
## FullBath              3.729e+03  2.279e+03   1.636 0.102125
## HalfBath              1.037e+03  2.151e+03   0.482 0.629930
## BedroomAbvGr         -3.152e+03  1.414e+03  -2.229 0.025987 *
## KitchenAbvGr         -1.166e+04  5.920e+03  -1.970 0.049015 *
## KitchenQualFa        -1.886e+04  6.347e+03  -2.972 0.003022 **
## KitchenQualGd        -2.444e+04  3.533e+03  -6.916 7.57e-12 ***
## KitchenQualTA        -2.180e+04  4.001e+03  -5.449 6.17e-08 ***
## TotRmsAbvGrd          9.697e+02  9.768e+02   0.993 0.321000
## FunctionalMaj2       -4.203e+03  1.477e+04  -0.284 0.776083
## FunctionalMin1        6.766e+03  9.020e+03   0.750 0.453320
## FunctionalMin2        7.858e+03  9.065e+03   0.867 0.386204
## FunctionalMod        -3.783e+03  1.162e+04  -0.326 0.744825
## FunctionalSev        -4.050e+04  3.069e+04  -1.319 0.187311
## FunctionalTyp         1.795e+04  7.803e+03   2.300 0.021619 *
## Fireplaces            2.524e+03  1.558e+03   1.620 0.105504
## FireplaceQuFa        -4.513e+03  7.023e+03  -0.643 0.520650
## FireplaceQuGd         9.232e+02  5.494e+03   0.168 0.866571
## FireplaceQuPo         5.229e+03  8.070e+03   0.648 0.517199
## FireplaceQuTA         1.463e+03  5.699e+03   0.257 0.797416
## GarageTypeAttchd      1.816e+04  1.115e+04   1.629 0.103555
## GarageTypeBasment     1.910e+04  1.293e+04   1.477 0.140002
## GarageTypeBuiltIn     1.718e+04  1.159e+04   1.482 0.138656
## GarageTypeCarPort     2.400e+04  1.616e+04   1.485 0.137703
## GarageTypeDetchd      2.017e+04  1.113e+04   1.813 0.070077 .
## GarageFinishRFn      -2.473e+03  2.004e+03  -1.234 0.217411
## GarageFinishUnf       9.667e+02  2.460e+03   0.393 0.694388
## GarageCars            2.640e+03  2.243e+03   1.177 0.239451
## GarageArea            1.511e+01  7.789e+00   1.939 0.052690 .
## GarageQualFa         -1.193e+05  3.040e+04  -3.924 9.20e-05 ***
## GarageQualGd         -1.122e+05  3.122e+04  -3.594 0.000339 ***
## GarageQualPo         -1.377e+05  3.931e+04  -3.504 0.000476 ***
## GarageQualTA         -1.133e+05  3.014e+04  -3.759 0.000179 ***
## GarageCondFa          1.056e+05  3.512e+04   3.006 0.002704 **
## GarageCondGd          1.065e+05  3.624e+04   2.938 0.003371 **
## GarageCondPo          1.078e+05  3.774e+04   2.857 0.004345 **
## GarageCondTA          1.092e+05  3.478e+04   3.141 0.001725 **
## PavedDriveP          -4.012e+03  5.694e+03  -0.705 0.481152
## PavedDriveY          -8.614e+02  3.624e+03  -0.238 0.812176
## WoodDeckSF            1.291e+01  5.951e+00   2.169 0.030299 *
## OpenPorchSF           1.075e+00  1.178e+01   0.091 0.927308
## EnclosedPorch         4.833e+00  1.280e+01   0.378 0.705752
## X3SsnPorch            2.865e+01  2.301e+01   1.245 0.213289
```

```
## ScreenPorch           2.808e+01  1.247e+01   2.251 0.024558 *
## PoolArea              8.115e+01  1.882e+01   4.311 1.76e-05 ***
## MiscVal              -2.332e-02  1.455e+00  -0.016 0.987216
## MoSold               -3.873e+02  2.516e+02  -1.539 0.124064
## YrSold               -3.519e+02  5.289e+02  -0.665 0.505904
## SaleTypeCon           2.654e+04  1.785e+04   1.487 0.137330
## SaleTypeConLD         1.892e+04  1.054e+04   1.796 0.072803 .
## SaleTypeConLI         5.139e+03  1.176e+04   0.437 0.662137
## SaleTypeConLw        -7.053e+02  1.239e+04  -0.057 0.954617
## SaleTypeCWD           1.468e+04  1.310e+04   1.120 0.262876
## SaleTypeNew           2.210e+04  1.576e+04   1.402 0.161109
## SaleTypeOth           8.557e+03  1.464e+04   0.585 0.558864
## SaleTypeWD           -3.006e+02  4.333e+03  -0.069 0.944708
## SaleConditionAdjLand  1.463e+04  1.609e+04   0.909 0.363604
## SaleConditionAlloca   1.065e+04  1.018e+04   1.046 0.295760
## SaleConditionFamily  -1.716e+03  6.165e+03  -0.278 0.780819
## SaleConditionNormal   4.626e+03  2.952e+03   1.567 0.117351
## SaleConditionPartial -2.745e+03  1.514e+04  -0.181 0.856135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23000 on 1182 degrees of freedom
## Multiple R-squared:  0.9295, Adjusted R-squared:  0.9159
## F-statistic: 68.06 on 229 and 1182 DF,  p-value: < 2.2e-16
```

The adjusted R squared value is 0.914, indicating that 91.4% of the variability can be explained by the model. Finally, we can calculate the MSE (mean square error)

```
residuals <- residuals(lm_model)
RMSE <- sqrt(mean(residuals^2))
RMSE
```

```
## [1] 21039.04
```

```
x <- model.matrix(SalePrice ~ . - 1, data = train_data) %>% scale ()
y <- train_data$SalePrice
```

In conclusion, with an adjusted R squared value of 0.914, the model has a strong fit to the data, meaning the variables chosen are good indicators of the price of a house.