MTH404 R Project

Mustafif Khan

Loading Required Packages

Before the dataset can be analyzed, the following packages must be imported:

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.1     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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
```

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
library(ggplot2)
library(lubridate)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library(caTools)
library(GGally)
```

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

```r
library(dplyr)
```

Reading the Training and Test Data

From the kaggle competition, we are provided the CSV data for training and testing, and the dataset is imported like the following:

```
train_data <- read.csv("train.csv")
test_data <- read.csv("test.csv")
```

With the data, we can check if we are missing values, let's first check our training data:

```
# get the missing training values
missing_training_values <- sapply(train_data, function(x) sum(is.na(x)))
# create a data frame with column names and missing values
mtv_df <- data.frame(column = names(missing_training_values),
                     missing_values = missing_training_values)
mtv_df
```

```
##                      column missing_values
## Id                       Id              0
## MSSubClass       MSSubClass              0
## MSZoning           MSZoning              0
## LotFrontage     LotFrontage            259
## LotArea             LotArea              0
## Street               Street              0
## Alley                 Alley           1369
## LotShape           LotShape              0
## LandContour     LandContour              0
## Utilities         Utilities              0
## LotConfig         LotConfig              0
## LandSlope         LandSlope              0
## Neighborhood   Neighborhood              0
## Condition1       Condition1              0
## Condition2       Condition2              0
## BldgType           BldgType              0
## HouseStyle       HouseStyle              0
## OverallQual     OverallQual              0
## OverallCond     OverallCond              0
## YearBuilt         YearBuilt              0
## YearRemodAdd   YearRemodAdd              0
## RoofStyle         RoofStyle              0
## RoofMatl           RoofMatl              0
## Exterior1st     Exterior1st              0
## Exterior2nd     Exterior2nd              0
## MasVnrType       MasVnrType              8
## MasVnrArea       MasVnrArea              8
## ExterQual         ExterQual              0
## ExterCond         ExterCond              0
## Foundation       Foundation              0
## BsmtQual           BsmtQual             37
## BsmtCond           BsmtCond             37
## BsmtExposure   BsmtExposure             38
## BsmtFinType1   BsmtFinType1             37
## BsmtFinSF1       BsmtFinSF1              0
## BsmtFinType2   BsmtFinType2             38
## BsmtFinSF2       BsmtFinSF2              0
## BsmtUnfSF         BsmtUnfSF              0
## TotalBsmtSF     TotalBsmtSF              0
## Heating             Heating              0
## HeatingQC         HeatingQC              0
## CentralAir       CentralAir              0
```

```
## Electrical       Electrical              1
## X1stFlrSF         X1stFlrSF              0
## X2ndFlrSF         X2ndFlrSF              0
## LowQualFinSF   LowQualFinSF              0
## GrLivArea         GrLivArea              0
## BsmtFullBath   BsmtFullBath              0
## BsmtHalfBath   BsmtHalfBath              0
## FullBath           FullBath              0
## HalfBath           HalfBath              0
## BedroomAbvGr   BedroomAbvGr              0
## KitchenAbvGr   KitchenAbvGr              0
## KitchenQual     KitchenQual              0
## TotRmsAbvGrd   TotRmsAbvGrd              0
## Functional       Functional              0
## Fireplaces       Fireplaces              0
## FireplaceQu     FireplaceQu            690
## GarageType       GarageType             81
## GarageYrBlt     GarageYrBlt             81
## GarageFinish   GarageFinish             81
## GarageCars       GarageCars              0
## GarageArea       GarageArea              0
## GarageQual       GarageQual             81
## GarageCond       GarageCond             81
## PavedDrive       PavedDrive              0
## WoodDeckSF       WoodDeckSF              0
## OpenPorchSF     OpenPorchSF              0
## EnclosedPorch EnclosedPorch              0
## X3SsnPorch       X3SsnPorch              0
## ScreenPorch     ScreenPorch              0
## PoolArea           PoolArea              0
## PoolQC               PoolQC           1453
## Fence                 Fence           1179
## MiscFeature     MiscFeature           1406
## MiscVal             MiscVal              0
## MoSold               MoSold              0
## YrSold               YrSold              0
## SaleType           SaleType              0
## SaleCondition SaleCondition              0
## SalePrice         SalePrice              0
```

As you can notice there are data in this dataset that contains large amounts of missing values, such as columns like `PoolQC`, `Fence`, etc. We will be removing the columns with missing values like the following:

```
train_data <- select(train_data, -c(LotFrontage, Alley, MasVnrType, MasVnrArea, BsmtQual ,
                                    BsmtCond,BsmtExposure, BsmtFinType1, BsmtFinType2,
                                    FireplaceQu, GarageType, GarageYrBlt, GarageFinish,
                                    GarageQual, GarageCond, PoolQC, Fence , MiscFeature ))
```

Now it would be good to also check our training data for missing values it may contain:

```
# get the missing values for the testing dataset
missing_testing_values <- sapply(test_data, function (x) sum(is.na(x)))
# create a data frame of the missing testing values
mtestv_df <- data.frame(column = names(missing_testing_values),
                        missing_values = missing_testing_values)
mtestv_df
```

```
##                          column missing_values
## Id                           Id              0
## MSSubClass           MSSubClass              0
## MSZoning               MSZoning              4
## LotFrontage         LotFrontage            227
## LotArea                 LotArea              0
## Street                   Street              0
## Alley                     Alley           1352
## LotShape               LotShape              0
## LandContour         LandContour              0
## Utilities             Utilities              2
## LotConfig             LotConfig              0
## LandSlope             LandSlope              0
## Neighborhood       Neighborhood              0
## Condition1           Condition1              0
## Condition2           Condition2              0
## BldgType               BldgType              0
## HouseStyle           HouseStyle              0
## OverallQual         OverallQual              0
## OverallCond         OverallCond              0
## YearBuilt             YearBuilt              0
## YearRemodAdd       YearRemodAdd              0
## RoofStyle             RoofStyle              0
## RoofMatl               RoofMatl              0
## Exterior1st         Exterior1st              1
## Exterior2nd         Exterior2nd              1
## MasVnrType           MasVnrType             16
## MasVnrArea           MasVnrArea             15
## ExterQual             ExterQual              0
## ExterCond             ExterCond              0
## Foundation           Foundation              0
## BsmtQual               BsmtQual             44
## BsmtCond               BsmtCond             45
## BsmtExposure       BsmtExposure             44
## BsmtFinType1       BsmtFinType1             42
## BsmtFinSF1           BsmtFinSF1              1
## BsmtFinType2       BsmtFinType2             42
## BsmtFinSF2           BsmtFinSF2              1
## BsmtUnfSF             BsmtUnfSF              1
## TotalBsmtSF         TotalBsmtSF              1
## Heating                 Heating              0
## HeatingQC             HeatingQC              0
## CentralAir           CentralAir              0
## Electrical           Electrical              0
## X1stFlrSF             X1stFlrSF              0
## X2ndFlrSF             X2ndFlrSF              0
## LowQualFinSF       LowQualFinSF              0
## GrLivArea             GrLivArea              0
## BsmtFullBath       BsmtFullBath              2
## BsmtHalfBath       BsmtHalfBath              2
## FullBath               FullBath              0
## HalfBath               HalfBath              0
```

```
## BedroomAbvGr   BedroomAbvGr             0
## KitchenAbvGr   KitchenAbvGr             0
## KitchenQual     KitchenQual             1
## TotRmsAbvGrd   TotRmsAbvGrd             0
## Functional       Functional             2
## Fireplaces       Fireplaces             0
## FireplaceQu     FireplaceQu           730
## GarageType       GarageType            76
## GarageYrBlt     GarageYrBlt            78
## GarageFinish   GarageFinish            78
## GarageCars       GarageCars             1
## GarageArea       GarageArea             1
## GarageQual       GarageQual            78
## GarageCond       GarageCond            78
## PavedDrive       PavedDrive             0
## WoodDeckSF       WoodDeckSF             0
## OpenPorchSF     OpenPorchSF             0
## EnclosedPorch EnclosedPorch             0
## X3SsnPorch       X3SsnPorch             0
## ScreenPorch     ScreenPorch             0
## PoolArea           PoolArea             0
## PoolQC               PoolQC          1456
## Fence                 Fence          1169
## MiscFeature     MiscFeature          1408
## MiscVal             MiscVal             0
## MoSold               MoSold             0
## YrSold               YrSold             0
## SaleType           SaleType             1
## SaleCondition SaleCondition             0
```

As you can see in our testing dataset we can notice columns that have high amount of missing values such as columns like `LotFrontage`, `Alley`, etc. As before we will be removing the columns with missing values like the following:

```
test_data <- select(test_data, -c(LotFrontage, Alley , MasVnrType, MasVnrArea, BsmtQual,
                                  BsmtCond, BsmtExposure, BsmtFinType1, BsmtFinType2,
                                  FireplaceQu, GarageType, GarageYrBlt, GarageFinish,
                                  GarageQual, GarageCond, PoolQC, Fence, MiscFeature))
```

Analysis of the Train Data

Before we create a model to analyze our training data, we will need to look at the structure of our different columns using the `str()` function:

```
str(train_data)
```

```
## 'data.frame':    1460 obs. of  63 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : int  60 20 60 70 60 50 20 60 50 190 ...
##  $ MSZoning     : chr  "RL" "RL" "RL" "RL" ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7420 ...
##  $ Street       : chr  "Pave" "Pave" "Pave" "Pave" ...
##  $ LotShape     : chr  "Reg" "Reg" "IR1" "IR1" ...
##  $ LandContour  : chr  "Lvl" "Lvl" "Lvl" "Lvl" ...
##  $ Utilities    : chr  "AllPub" "AllPub" "AllPub" "AllPub" ...
```

```
##  $ LotConfig    : chr  "Inside" "FR2" "Inside" "Corner" ...
##  $ LandSlope    : chr  "Gtl" "Gtl" "Gtl" "Gtl" ...
##  $ Neighborhood : chr  "CollgCr" "Veenker" "CollgCr" "Crawfor" ...
##  $ Condition1   : chr  "Norm" "Feedr" "Norm" "Norm" ...
##  $ Condition2   : chr  "Norm" "Norm" "Norm" "Norm" ...
##  $ BldgType     : chr  "1Fam" "1Fam" "1Fam" "1Fam" ...
##  $ HouseStyle   : chr  "2Story" "1Story" "2Story" "2Story" ...
##  $ OverallQual  : int  7 6 7 7 8 5 8 7 7 5 ...
##  $ OverallCond  : int  5 8 5 5 5 5 5 6 5 6 ...
##  $ YearBuilt    : int  2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 ...
##  $ YearRemodAdd : int  2003 1976 2002 1970 2000 1995 2005 1973 1950 1950 ...
##  $ RoofStyle    : chr  "Gable" "Gable" "Gable" "Gable" ...
##  $ RoofMatl     : chr  "CompShg" "CompShg" "CompShg" "CompShg" ...
##  $ Exterior1st  : chr  "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...
##  $ Exterior2nd  : chr  "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...
##  $ ExterQual    : chr  "Gd" "TA" "Gd" "TA" ...
##  $ ExterCond    : chr  "TA" "TA" "TA" "TA" ...
##  $ Foundation   : chr  "PConc" "CBlock" "PConc" "BrkTil" ...
##  $ BsmtFinSF1   : int  706 978 486 216 655 732 1369 859 0 851 ...
##  $ BsmtFinSF2   : int  0 0 0 0 0 0 0 32 0 0 ...
##  $ BsmtUnfSF    : int  150 284 434 540 490 64 317 216 952 140 ...
##  $ TotalBsmtSF  : int  856 1262 920 756 1145 796 1686 1107 952 991 ...
##  $ Heating      : chr  "GasA" "GasA" "GasA" "GasA" ...
##  $ HeatingQC    : chr  "Ex" "Ex" "Ex" "Gd" ...
##  $ CentralAir   : chr  "Y" "Y" "Y" "Y" ...
##  $ Electrical   : chr  "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
##  $ X1stFlrSF    : int  856 1262 920 961 1145 796 1694 1107 1022 1077 ...
##  $ X2ndFlrSF    : int  854 0 866 756 1053 566 0 983 752 0 ...
##  $ LowQualFinSF : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GrLivArea    : int  1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 ...
##  $ BsmtFullBath : int  1 0 1 1 1 1 1 1 0 1 ...
##  $ BsmtHalfBath : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FullBath     : int  2 2 2 1 2 1 2 2 2 1 ...
##  $ HalfBath     : int  1 0 1 0 1 1 0 1 0 0 ...
##  $ BedroomAbvGr : int  3 3 3 3 4 1 3 3 2 2 ...
##  $ KitchenAbvGr : int  1 1 1 1 1 1 1 1 2 2 ...
##  $ KitchenQual  : chr  "Gd" "TA" "Gd" "Gd" ...
##  $ TotRmsAbvGrd : int  8 6 6 7 9 5 7 7 8 5 ...
##  $ Functional   : chr  "Typ" "Typ" "Typ" "Typ" ...
##  $ Fireplaces   : int  0 1 1 1 1 0 1 2 2 2 ...
##  $ GarageCars   : int  2 2 2 3 3 2 2 2 2 1 ...
##  $ GarageArea   : int  548 460 608 642 836 480 636 484 468 205 ...
##  $ PavedDrive   : chr  "Y" "Y" "Y" "Y" ...
##  $ WoodDeckSF   : int  0 298 0 0 192 40 255 235 90 0 ...
##  $ OpenPorchSF  : int  61 0 42 35 84 30 57 204 0 4 ...
##  $ EnclosedPorch: int  0 0 0 272 0 0 0 228 205 0 ...
##  $ X3SsnPorch   : int  0 0 0 0 0 320 0 0 0 0 ...
##  $ ScreenPorch  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolArea     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ MiscVal      : int  0 0 0 0 0 700 0 350 0 0 ...
##  $ MoSold       : int  2 5 9 2 12 10 8 11 4 1 ...
##  $ YrSold       : int  2008 2007 2008 2006 2008 2009 2007 2009 2008 2008 ...
##  $ SaleType     : chr  "WD" "WD" "WD" "WD" ...
##  $ SaleCondition: chr  "Normal" "Normal" "Normal" "Abnorml" ...
```

```
##  $ SalePrice    : int  208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
```

To analyze the data, we will need to create a model using the `lm()` function, where will be using `SalesPrice` to compare our other fields. Using the model, we can use `summary()` to find better information about our model, such as R^2.

```
model <- lm(SalePrice ~., data=train_data)
summary(model)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = train_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -174309  -10535       0    9742  174309
##
## Coefficients: (3 not defined because of singularities)
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1.199e+06  1.066e+06  -1.125 0.260855
## Id                    6.358e-01  1.603e+00   0.397 0.691694
## MSSubClass           -7.830e+00  8.553e+01  -0.092 0.927077
## MSZoningFV            3.101e+04  1.225e+04   2.532 0.011470 *
## MSZoningRH            2.371e+04  1.232e+04   1.925 0.054412 .
## MSZoningRL            2.588e+04  1.051e+04   2.463 0.013909 *
## MSZoningRM            2.503e+04  9.852e+03   2.541 0.011174 *
## LotArea               7.006e-01  1.083e-01   6.470 1.40e-10 ***
## StreetPave            3.867e+04  1.229e+04   3.148 0.001685 **
## LotShapeIR2           4.618e+03  4.323e+03   1.068 0.285624
## LotShapeIR3           4.584e+03  9.063e+03   0.506 0.613065
## LotShapeReg           5.589e+02  1.667e+03   0.335 0.737481
## LandContourHLS        1.350e+04  5.308e+03   2.543 0.011098 *
## LandContourLow       -4.241e+03  6.530e+03  -0.650 0.516119
## LandContourLvl        7.073e+03  3.822e+03   1.851 0.064427 .
## UtilitiesNoSeWa      -3.037e+04  2.663e+04  -1.141 0.254268
## LotConfigCulDSac      7.629e+03  3.328e+03   2.293 0.022033 *
## LotConfigFR2         -5.847e+03  4.160e+03  -1.406 0.160105
## LotConfigFR3         -1.349e+04  1.309e+04  -1.031 0.302943
## LotConfigInside      -1.241e+03  1.811e+03  -0.685 0.493217
## LandSlopeMod          1.046e+04  4.044e+03   2.586 0.009824 **
## LandSlopeSev         -2.560e+04  1.111e+04  -2.305 0.021326 *
## NeighborhoodBlueste  -2.735e+03  1.935e+04  -0.141 0.887666
## NeighborhoodBrDale    8.372e+03  1.113e+04   0.752 0.452231
## NeighborhoodBrkSide  -2.092e+03  9.509e+03  -0.220 0.825905
## NeighborhoodClearCr  -1.277e+04  9.432e+03  -1.354 0.176093
## NeighborhoodCollgCr  -9.717e+03  7.337e+03  -1.324 0.185610
## NeighborhoodCrawfor   9.596e+03  8.674e+03   1.106 0.268795
## NeighborhoodEdwards  -1.675e+04  8.085e+03  -2.072 0.038477 *
## NeighborhoodGilbert  -1.388e+04  7.854e+03  -1.767 0.077484 .
## NeighborhoodIDOTRR   -7.806e+03  1.087e+04  -0.718 0.472893
## NeighborhoodMeadowV  -1.365e+03  1.141e+04  -0.120 0.904805
## NeighborhoodMitchel  -2.038e+04  8.281e+03  -2.461 0.013996 *
## NeighborhoodNAmes    -1.447e+04  7.905e+03  -1.831 0.067377 .
## NeighborhoodNoRidge   2.877e+04  8.406e+03   3.423 0.000640 ***
## NeighborhoodNPkVill   8.163e+03  1.434e+04   0.569 0.569256
```

```
## NeighborhoodNridgHt   2.459e+04  7.383e+03   3.330 0.000893 ***
## NeighborhoodNWAmes   -2.053e+04  8.154e+03  -2.517 0.011950 *
## NeighborhoodOldTown  -1.301e+04  9.681e+03  -1.344 0.179263
## NeighborhoodSawyer   -1.005e+04  8.236e+03  -1.221 0.222449
## NeighborhoodSawyerW  -6.145e+03  7.857e+03  -0.782 0.434278
## NeighborhoodSomerst  -1.607e+01  8.984e+03  -0.002 0.998573
## NeighborhoodStoneBr   3.896e+04  8.390e+03   4.643 3.78e-06 ***
## NeighborhoodSWISU    -9.589e+03  9.837e+03  -0.975 0.329853
## NeighborhoodTimber   -6.108e+03  8.419e+03  -0.726 0.468254
## NeighborhoodVeenker   3.134e+03  1.074e+04   0.292 0.770436
## Condition1Feedr       2.827e+03  5.118e+03   0.552 0.580839
## Condition1Norm        1.210e+04  4.226e+03   2.862 0.004278 **
## Condition1PosA        7.386e+03  1.031e+04   0.716 0.473947
## Condition1PosN        7.856e+03  7.634e+03   1.029 0.303635
## Condition1RRAe       -1.708e+04  9.381e+03  -1.820 0.068945 .
## Condition1RRAn        6.184e+03  7.041e+03   0.878 0.379890
## Condition1RRNe       -7.318e+03  1.839e+04  -0.398 0.690706
## Condition1RRNn        3.829e+03  1.312e+04   0.292 0.770531
## Condition2Feedr      -9.499e+03  2.307e+04  -0.412 0.680617
## Condition2Norm       -7.535e+03  1.967e+04  -0.383 0.701655
## Condition2PosA        2.022e+04  3.803e+04   0.532 0.594962
## Condition2PosN       -2.303e+05  2.764e+04  -8.331  < 2e-16 ***
## Condition2RRAe       -1.288e+05  4.688e+04  -2.747 0.006091 **
## Condition2RRAn       -1.208e+04  3.197e+04  -0.378 0.705522
## Condition2RRNn       -8.718e+03  2.713e+04  -0.321 0.747988
## BldgType2fmCon       -6.293e+03  1.288e+04  -0.489 0.625255
## BldgTypeDuplex       -1.020e+03  7.466e+03  -0.137 0.891399
## BldgTypeTwnhs        -2.560e+04  1.017e+04  -2.517 0.011949 *
## BldgTypeTwnhsE       -2.342e+04  9.215e+03  -2.542 0.011144 *
## HouseStyle1.5Unf      1.116e+04  7.941e+03   1.405 0.160212
## HouseStyle1Story      8.915e+03  4.365e+03   2.042 0.041324 *
## HouseStyle2.5Fin     -1.711e+04  1.232e+04  -1.388 0.165250
## HouseStyle2.5Unf     -1.188e+04  9.396e+03  -1.265 0.206181
## HouseStyle2Story     -6.353e+03  3.558e+03  -1.785 0.074426 .
## HouseStyleSFoyer      7.609e+03  6.207e+03   1.226 0.220443
## HouseStyleSLvl        7.253e+03  5.500e+03   1.319 0.187483
## OverallQual           8.042e+03  1.022e+03   7.867 7.70e-15 ***
## OverallCond           5.439e+03  8.759e+02   6.210 7.18e-10 ***
## YearBuilt             3.307e+02  7.400e+01   4.469 8.58e-06 ***
## YearRemodAdd          1.065e+02  5.571e+01   1.911 0.056168 .
## RoofStyleGable        1.518e+03  1.877e+04   0.081 0.935529
## RoofStyleGambrel      4.328e+03  2.052e+04   0.211 0.832985
## RoofStyleHip          3.112e+03  1.882e+04   0.165 0.868667
## RoofStyleMansard      1.724e+04  2.186e+04   0.789 0.430398
## RoofStyleShed         8.755e+04  3.554e+04   2.463 0.013905 *
## RoofMatlCompShg       6.504e+05  3.306e+04  19.672  < 2e-16 ***
## RoofMatlMembran       7.377e+05  4.782e+04  15.427  < 2e-16 ***
## RoofMatlMetal         6.978e+05  4.724e+04  14.770  < 2e-16 ***
## RoofMatlRoll          6.496e+05  4.168e+04  15.583  < 2e-16 ***
## RoofMatlTar&Grv       6.558e+05  3.801e+04  17.257  < 2e-16 ***
## RoofMatlWdShake       6.309e+05  3.683e+04  17.132  < 2e-16 ***
## RoofMatlWdShngl       7.282e+05  3.430e+04  21.229  < 2e-16 ***
## Exterior1stAsphShn   -1.265e+04  3.422e+04  -0.370 0.711798
## Exterior1stBrkComm   -1.327e+04  2.869e+04  -0.463 0.643757
```

```
## Exterior1stBrkFace    5.438e+03  1.287e+04   0.422 0.672763
## Exterior1stCBlock    -2.812e+04  2.761e+04  -1.018 0.308705
## Exterior1stCemntBd   -1.490e+04  1.947e+04  -0.765 0.444353
## Exterior1stHdBoard   -1.379e+04  1.299e+04  -1.062 0.288621
## Exterior1stImStucc   -6.922e+04  2.863e+04  -2.418 0.015758 *
## Exterior1stMetalSd   -3.215e+03  1.483e+04  -0.217 0.828473
## Exterior1stPlywood   -1.796e+04  1.288e+04  -1.395 0.163280
## Exterior1stStone     -1.511e+04  2.438e+04  -0.620 0.535573
## Exterior1stStucco    -4.992e+03  1.418e+04  -0.352 0.724787
## Exterior1stVinylSd   -1.764e+04  1.347e+04  -1.310 0.190513
## Exterior1stWd Sdng   -1.358e+04  1.243e+04  -1.092 0.274945
## Exterior1stWdShing   -6.428e+03  1.344e+04  -0.478 0.632616
## Exterior2ndAsphShn    8.090e+03  2.282e+04   0.354 0.723059
## Exterior2ndBrk Cmn    1.496e+04  2.075e+04   0.721 0.470967
## Exterior2ndBrkFace   -8.279e+02  1.331e+04  -0.062 0.950396
## Exterior2ndCBlock            NA         NA      NA       NA
## Exterior2ndCmentBd    1.303e+04  1.920e+04   0.678 0.497621
## Exterior2ndHdBoard    8.081e+03  1.251e+04   0.646 0.518448
## Exterior2ndImStucc    3.368e+04  1.447e+04   2.327 0.020112 *
## Exterior2ndMetalSd    2.897e+03  1.448e+04   0.200 0.841498
## Exterior2ndOther     -6.218e+03  2.822e+04  -0.220 0.825635
## Exterior2ndPlywood    9.189e+03  1.215e+04   0.756 0.449761
## Exterior2ndStone     -1.015e+04  1.738e+04  -0.584 0.559203
## Exterior2ndStucco     2.426e+03  1.366e+04   0.178 0.859056
## Exterior2ndVinylSd    1.646e+04  1.301e+04   1.266 0.205865
## Exterior2ndWd Sdng    1.052e+04  1.200e+04   0.877 0.380774
## Exterior2ndWd Shng    3.419e+03  1.251e+04   0.273 0.784665
## ExterQualFa          -8.622e+03  1.089e+04  -0.792 0.428642
## ExterQualGd          -3.080e+04  4.794e+03  -6.425 1.86e-10 ***
## ExterQualTA          -3.069e+04  5.365e+03  -5.720 1.33e-08 ***
## ExterCondFa          -2.614e+03  1.888e+04  -0.138 0.889928
## ExterCondGd          -7.990e+03  1.802e+04  -0.443 0.657627
## ExterCondPo           1.218e+04  3.286e+04   0.371 0.711022
## ExterCondTA          -5.344e+03  1.799e+04  -0.297 0.766449
## FoundationCBlock      1.760e+03  3.200e+03   0.550 0.582379
## FoundationPConc       4.829e+03  3.509e+03   1.376 0.169008
## FoundationSlab        8.498e+03  7.864e+03   1.081 0.280108
## FoundationStone       2.503e+03  1.118e+04   0.224 0.822878
## FoundationWood       -3.336e+04  1.513e+04  -2.205 0.027642 *
## BsmtFinSF1            3.710e+01  4.424e+00   8.386  < 2e-16 ***
## BsmtFinSF2            2.458e+01  5.800e+00   4.238 2.42e-05 ***
## BsmtUnfSF             1.497e+01  4.072e+00   3.676 0.000247 ***
## TotalBsmtSF                  NA         NA      NA       NA
## HeatingGasA          -7.036e+03  2.547e+04  -0.276 0.782451
## HeatingGasW          -1.559e+04  2.627e+04  -0.594 0.552905
## HeatingGrav          -1.540e+04  2.765e+04  -0.557 0.577754
## HeatingOthW          -4.575e+04  3.174e+04  -1.442 0.149665
## HeatingWall           8.292e+03  2.951e+04   0.281 0.778794
## HeatingQCFa          -1.594e+03  4.832e+03  -0.330 0.741498
## HeatingQCGd          -3.627e+03  2.153e+03  -1.685 0.092279 .
## HeatingQCPo           8.341e+03  2.775e+04   0.301 0.763758
## HeatingQCTA          -4.402e+03  2.123e+03  -2.074 0.038305 *
## CentralAirY          -3.675e+03  3.999e+03  -0.919 0.358319
## ElectricalFuseF      -1.221e+03  5.993e+03  -0.204 0.838579
```

```
## ElectricalFuseP      -1.003e+04  1.745e+04  -0.574 0.565821    
## ElectricalMix         3.604e+03  2.893e+04   0.125 0.900860    
## ElectricalSBrkr      -1.335e+03  3.028e+03  -0.441 0.659257    
## X1stFlrSF             5.503e+01  5.337e+00  10.311  < 2e-16 ***
## X2ndFlrSF             6.998e+01  5.279e+00  13.257  < 2e-16 ***
## LowQualFinSF          2.514e+01  1.873e+01   1.342 0.179809    
## GrLivArea                    NA         NA      NA       NA    
## BsmtFullBath          1.551e+03  1.969e+03   0.788 0.431109    
## BsmtHalfBath          3.517e+02  3.118e+03   0.113 0.910199    
## FullBath              2.601e+03  2.247e+03   1.158 0.247210    
## HalfBath             -1.508e+02  2.141e+03  -0.070 0.943878    
## BedroomAbvGr         -5.506e+03  1.385e+03  -3.975 7.45e-05 ***
## KitchenAbvGr         -1.576e+04  5.776e+03  -2.729 0.006432 ** 
## KitchenQualFa        -2.065e+04  6.417e+03  -3.218 0.001324 ** 
## KitchenQualGd        -2.775e+04  3.490e+03  -7.953 4.01e-15 ***
## KitchenQualTA        -2.522e+04  3.999e+03  -6.307 3.93e-10 ***
## TotRmsAbvGrd          1.343e+03  9.765e+02   1.375 0.169414    
## FunctionalMaj2       -5.309e+02  1.480e+04  -0.036 0.971397    
## FunctionalMin1        4.452e+03  8.670e+03   0.514 0.607659    
## FunctionalMin2        8.559e+03  8.584e+03   0.997 0.318926    
## FunctionalMod        -7.249e+03  1.057e+04  -0.686 0.492877    
## FunctionalSev        -5.986e+04  2.759e+04  -2.169 0.030236 *  
## FunctionalTyp         1.970e+04  7.422e+03   2.654 0.008048 ** 
## Fireplaces            2.821e+03  1.374e+03   2.052 0.040326 *  
## GarageCars            4.257e+03  2.222e+03   1.916 0.055637 .  
## GarageArea            1.337e+01  7.653e+00   1.748 0.080774 .  
## PavedDriveP          -3.332e+03  5.577e+03  -0.598 0.550237    
## PavedDriveY          -2.109e+03  3.459e+03  -0.610 0.542108    
## WoodDeckSF            1.370e+01  5.958e+00   2.300 0.021622 *  
## OpenPorchSF           1.212e+01  1.185e+01   1.023 0.306696    
## EnclosedPorch         5.632e+00  1.285e+01   0.438 0.661311    
## X3SsnPorch            2.423e+01  2.316e+01   1.046 0.295536    
## ScreenPorch           3.717e+01  1.260e+01   2.950 0.003238 ** 
## PoolArea              7.127e+01  1.836e+01   3.883 0.000109 ***
## MiscVal              -3.146e-01  1.470e+00  -0.214 0.830518    
## MoSold               -6.394e+02  2.542e+02  -2.516 0.011996 *  
## YrSold               -1.758e+02  5.250e+02  -0.335 0.737816    
## SaleTypeCon           3.537e+04  1.839e+04   1.923 0.054645 .  
## SaleTypeConLD         1.673e+04  1.002e+04   1.669 0.095317 .  
## SaleTypeConLI         9.984e+03  1.192e+04   0.837 0.402591    
## SaleTypeConLw        -2.443e+03  1.243e+04  -0.197 0.844222    
## SaleTypeCWD           2.316e+04  1.337e+04   1.732 0.083534 .  
## SaleTypeNew           3.440e+04  1.606e+04   2.142 0.032353 *  
## SaleTypeOth           1.856e+04  1.504e+04   1.234 0.217488    
## SaleTypeWD            4.603e+02  4.344e+03   0.106 0.915627    
## SaleConditionAdjLand  1.047e+04  1.506e+04   0.696 0.486792    
## SaleConditionAlloca   5.064e+03  8.787e+03   0.576 0.564546    
## SaleConditionFamily  -1.338e+03  6.330e+03  -0.211 0.832625    
## SaleConditionNormal   6.618e+03  2.994e+03   2.210 0.027273 *  
## SaleConditionPartial -9.139e+03  1.547e+04  -0.591 0.554858    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 24010 on 1268 degrees of freedom
```

```
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9206, Adjusted R-squared:  0.9087
## F-statistic: 77.37 on 190 and 1268 DF,  p-value: < 2.2e-16
```

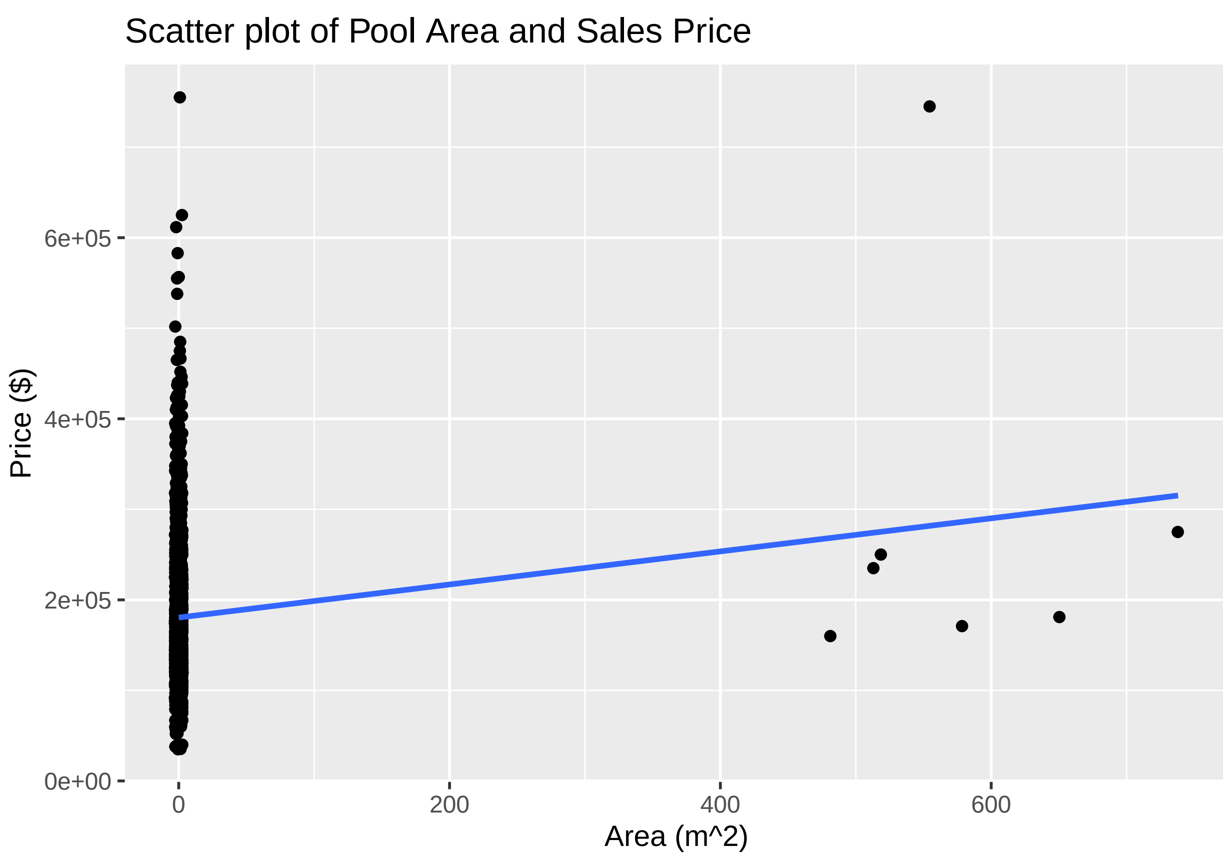
With the summary we can see that we have $R^2 = 0.9206$, and we can make some of the following interpretations when comparing `SalesPrice` with some of the most significant fields marked with `***`: Whenever the sales price increases by one dollar…

- `PoolArea`: the pool's area will increase by $71.27m^2$.
- `OverallQual`: the overall quality will increase by \$8042.
- `OverallCond`: the overall condition will increase by \$5439.
- `LotArea`: the lot's area will increase by $0.701m^2$.

For a better visual we will be plotting graphs between the sale price and significant fields.

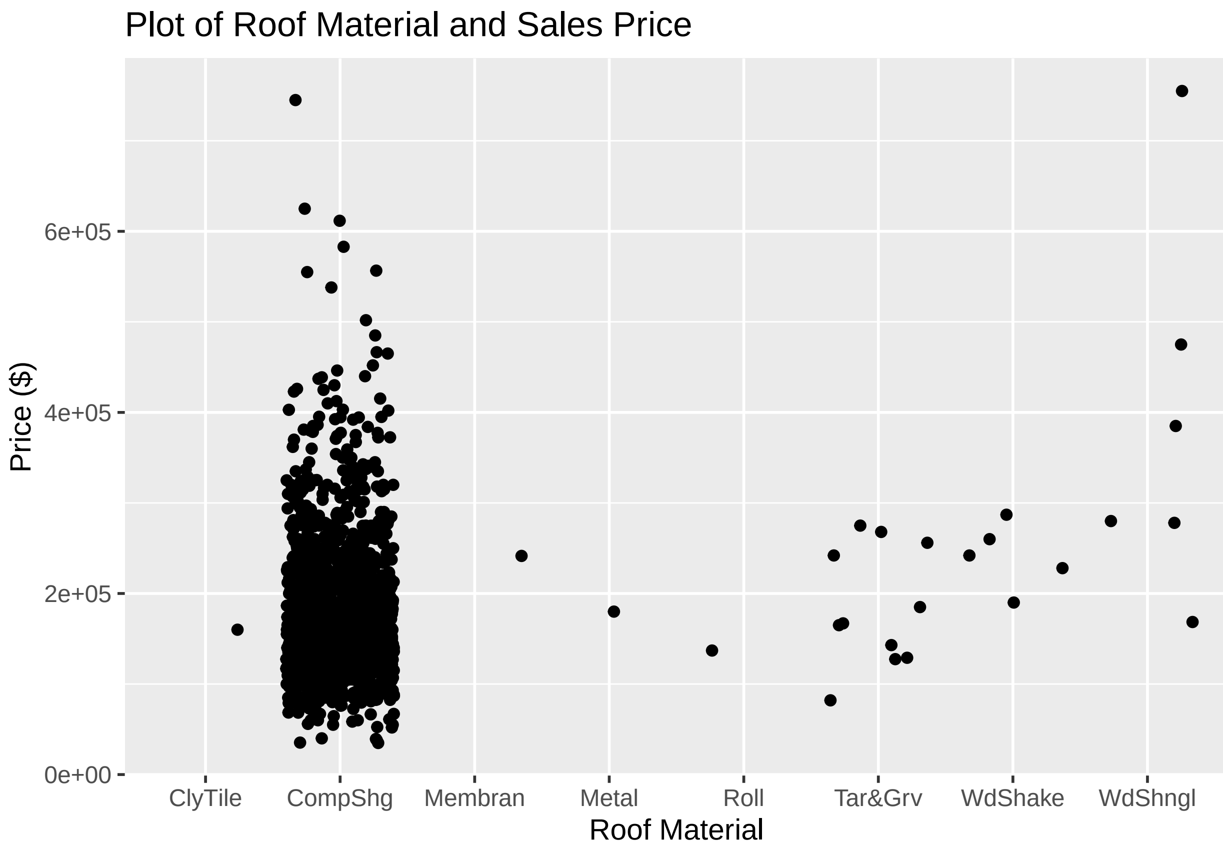
```r
# Plot between PoolArea and SalePrice
ggplot(data = train_data, aes(x = PoolArea, y = SalePrice)) +
  geom_jitter() +  geom_smooth(method = "lm", se = FALSE) +
  labs(title="Scatter plot of Pool Area and Sales Price", x="Area (m^2)",y="Price ($)")
```

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




```r
# Plot between RoofMatl and SalePrice
ggplot(data = train_data, aes(x = RoofMatl, y = SalePrice)) +
  geom_jitter() +  geom_smooth(method = "lm", se = FALSE) +
  labs(title="Plot of Roof Material and Sales Price", x="Roof Material",y="Price ($)")
```

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


Plot of Roof Material and Sales Price
Price ($)
6e+05
4e+05
2e+05
0e+00
ClyTile
CompShg
Membran
Metal
Roll
Tar&Grv
WdShake
WdShngl
Roof Material