House Price Compition

Ankita Gairola

2023-08-06

1. Introduction

Goal: Estimate the sale price of property

Dataset: Data from Kaggle. In this dataset, 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa.

2. Data Assessment

2.1 Set the Directory

getwd()

setwd("E:/Kaggle/Competition_1)

2.2 Install packages

install.packages("readr")

install.packages("ggplot2")

install.packages("gplots")

install.packages("repr")

install.packages("tidyverse")

install.packages("skimr")

install.packages("dplyr")

install.packages("plyr")

```
library(readr)

library(ggplot2)

library(repr)

library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ stringr   1.5.0
## ✓ forcats   1.0.0     ✓ tibble    3.2.1
```

```
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ─────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(gplots)
```

```
##
## Attaching package: 'gplots'
##
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(skimr)
```

```
library(dplyr)
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------
----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first,
then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------
----
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
## The following object is masked from 'package:purrr':
##
##     compact
```

2.2 Import Data

```
Test <- read.csv("E:/Kaggle/Competition_1/house-prices-advanced-regression-
techniques/test.csv")

Train <- read.csv("E:/Kaggle/Competition_1/house-prices-advanced-regression-
techniques/train.csv")
```

2.3 Exploring the Data

```
colnames(Test)
```

```
##  [1] "Id"            "MSSubClass"    "MSZoning"      "LotFrontage"
##  [5] "LotArea"       "Street"        "Alley"         "LotShape"
##  [9] "LandContour"   "Utilities"     "LotConfig"     "LandSlope"
## [13] "Neighborhood"  "Condition1"    "Condition2"    "BldgType"
## [17] "HouseStyle"    "OverallQual"   "OverallCond"   "YearBuilt"
## [21] "YearRemodAdd"  "RoofStyle"     "RoofMatl"      "Exterior1st"
## [25] "Exterior2nd"   "MasVnrType"    "MasVnrArea"    "ExterQual"
## [29] "ExterCond"     "Foundation"    "BsmtQual"      "BsmtCond"
## [33] "BsmtExposure"  "BsmtFinType1"  "BsmtFinSF1"    "BsmtFinType2"
## [37] "BsmtFinSF2"    "BsmtUnfSF"     "TotalBsmtSF"   "Heating"
## [41] "HeatingQC"     "CentralAir"    "Electrical"    "X1stFlrSF"
## [45] "X2ndFlrSF"     "LowQualFinSF"  "GrLivArea"     "BsmtFullBath"
## [49] "BsmtHalfBath"  "FullBath"      "HalfBath"      "BedroomAbvGr"
## [53] "KitchenAbvGr"  "KitchenQual"   "TotRmsAbvGrd"  "Functional"
## [57] "Fireplaces"    "FireplaceQu"   "GarageType"    "GarageYrBlt"
## [61] "GarageFinish"  "GarageCars"    "GarageArea"    "GarageQual"
## [65] "GarageCond"    "PavedDrive"    "WoodDeckSF"    "OpenPorchSF"
## [69] "EnclosedPorch" "X3SsnPorch"    "ScreenPorch"   "PoolArea"
## [73] "PoolQC"        "Fence"         "MiscFeature"   "MiscVal"
## [77] "MoSold"        "YrSold"        "SaleType"      "SaleCondition"
```

```
colnames(Train)
```

```
##  [1] "Id"            "MSSubClass"    "MSZoning"      "LotFrontage"
##  [5] "LotArea"       "Street"        "Alley"         "LotShape"
##  [9] "LandContour"   "Utilities"     "LotConfig"     "LandSlope"
## [13] "Neighborhood"  "Condition1"    "Condition2"    "BldgType"
## [17] "HouseStyle"    "OverallQual"   "OverallCond"   "YearBuilt"
## [21] "YearRemodAdd"  "RoofStyle"     "RoofMatl"      "Exterior1st"
## [25] "Exterior2nd"   "MasVnrType"    "MasVnrArea"    "ExterQual"
## [29] "ExterCond"     "Foundation"    "BsmtQual"      "BsmtCond"
## [33] "BsmtExposure"  "BsmtFinType1"  "BsmtFinSF1"    "BsmtFinType2"
## [37] "BsmtFinSF2"    "BsmtUnfSF"     "TotalBsmtSF"   "Heating"
## [41] "HeatingQC"     "CentralAir"    "Electrical"    "X1stFlrSF"
## [45] "X2ndFlrSF"     "LowQualFinSF"  "GrLivArea"     "BsmtFullBath"
## [49] "BsmtHalfBath"  "FullBath"      "HalfBath"      "BedroomAbvGr"
## [53] "KitchenAbvGr"  "KitchenQual"   "TotRmsAbvGrd"  "Functional"
## [57] "Fireplaces"    "FireplaceQu"   "GarageType"    "GarageYrBlt"
## [61] "GarageFinish"  "GarageCars"    "GarageArea"    "GarageQual"
## [65] "GarageCond"    "PavedDrive"    "WoodDeckSF"    "OpenPorchSF"
## [69] "EnclosedPorch" "X3SsnPorch"    "ScreenPorch"   "PoolArea"
## [73] "PoolQC"        "Fence"         "MiscFeature"   "MiscVal"
## [77] "MoSold"        "YrSold"        "SaleType"      "SaleCondition"
## [81] "SalePrice"
```

```
dim(Test)
```

```
## [1] 1459   80
```

```
dim(Train)
```

```
## [1] 1460   81
```

```
skim_without_charts(Test)
```

Data summary

Name	Test
Number of rows	1459
Number of columns	80

_____

Column type frequency:	
character	43
numeric	37

_____

Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
MSZoning	4	1.00	2	7	0	5	0
Street	0	1.00	4	4	0	2	0
Alley	1352	0.07	4	4	0	2	0
LotShape	0	1.00	3	3	0	4	0
LandContour	0	1.00	3	3	0	4	0
Utilities	2	1.00	6	6	0	1	0
LotConfig	0	1.00	3	7	0	5	0
LandSlope	0	1.00	3	3	0	3	0
Neighborhood	0	1.00	5	7	0	25	0
Condition1	0	1.00	4	6	0	9	0
Condition2	0	1.00	4	6	0	5	0
BldgType	0	1.00	4	6	0	5	0
HouseStyle	0	1.00	4	6	0	7	0
RoofStyle	0	1.00	3	7	0	6	0
RoofMatl	0	1.00	7	7	0	4	0
Exterior1st	1	1.00	6	7	0	13	0
Exterior2nd	1	1.00	5	7	0	15	0
MasVnrType	16	0.99	4	7	0	4	0
ExterQual	0	1.00	2	2	0	4	0
ExterCond	0	1.00	2	2	0	5	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Foundation	0	1.00	4	6	0	6	0
BsmtQual	44	0.97	2	2	0	4	0
BsmtCond	45	0.97	2	2	0	4	0
BsmtExposure	44	0.97	2	2	0	4	0
BsmtFinType1	42	0.97	3	3	0	6	0
BsmtFinType2	42	0.97	3	3	0	6	0
Heating	0	1.00	4	4	0	4	0
HeatingQC	0	1.00	2	2	0	5	0
CentralAir	0	1.00	1	1	0	2	0
Electrical	0	1.00	5	5	0	4	0
KitchenQual	1	1.00	2	2	0	4	0
Functional	2	1.00	3	4	0	7	0
FireplaceQu	730	0.50	2	2	0	5	0
GarageType	76	0.95	6	7	0	6	0
GarageFinish	78	0.95	3	3	0	3	0
GarageQual	78	0.95	2	2	0	4	0
GarageCond	78	0.95	2	2	0	5	0
PavedDrive	0	1.00	1	1	0	3	0
PoolQC	1456	0.00	2	2	0	2	0
Fence	1169	0.20	4	5	0	4	0
MiscFeature	1408	0.03	4	4	0	3	0
SaleType	1	1.00	2	5	0	9	0
SaleCondition	0	1.00	6	7	0	6	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1.00	2190.00	421.32	1461	1825.50	2190.0	2554.50	2919
MSSubClass	0	1.00	57.38	42.75	20	20.00	50.0	70.00	190
LotFrontage	227	0.84	68.58	22.38	21	58.00	67.0	80.00	200
LotArea	0	1.00	9819.16	4955.52	1470	7391.00	9399.0	11517.50	56600
OverallQual	0	1.00	6.08	1.44	1	5.00	6.0	7.00	10
OverallCond	0	1.00	5.55	1.11	1	5.00	5.0	6.00	9
YearBuilt	0	1.00	1971.	30.39	187	1953.	1973	2001.0	201

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
			36		9	00	.0	0	0
YearRemodAdd	0	1.00	1983.66	21.13	1950	1963.00	1992.0	2004.00	2010
MasVnrArea	15	0.99	100.71	177.63	0	0.00	0.0	164.00	1290
BsmtFinSF1	1	1.00	439.20	455.27	0	0.00	350.5	753.50	4010
BsmtFinSF2	1	1.00	52.62	176.75	0	0.00	0.0	0.00	1526
BsmtUnfSF	1	1.00	554.29	437.26	0	219.25	460.0	797.75	2140
TotalBsmtSF	1	1.00	1046.12	442.90	0	784.00	988.0	1305.00	5095
X1stFlrSF	0	1.00	1156.53	398.17	407	873.50	1079.0	1382.50	5095
X2ndFlrSF	0	1.00	325.97	420.61	0	0.00	0.0	676.00	1862
LowQualFinSF	0	1.00	3.54	44.04	0	0.00	0.0	0.00	1064
GrLivArea	0	1.00	1486.05	485.57	407	1117.50	1432.0	1721.00	5095
BsmtFullBath	2	1.00	0.43	0.53	0	0.00	0.0	1.00	3
BsmtHalfBath	2	1.00	0.07	0.25	0	0.00	0.0	0.00	2
FullBath	0	1.00	1.57	0.56	0	1.00	2.0	2.00	4
HalfBath	0	1.00	0.38	0.50	0	0.00	0.0	1.00	2
BedroomAbvGr	0	1.00	2.85	0.83	0	2.00	3.0	3.00	6
KitchenAbvGr	0	1.00	1.04	0.21	0	1.00	1.0	1.00	2
TotRmsAbvGrd	0	1.00	6.39	1.51	3	5.00	6.0	7.00	15
Fireplaces	0	1.00	0.58	0.65	0	0.00	0.0	1.00	4
GarageYrBlt	78	0.95	1977.72	26.43	1895	1959.00	1979.0	2002.00	2207
GarageCars	1	1.00	1.77	0.78	0	1.00	2.0	2.00	5
GarageArea	1	1.00	472.77	217.05	0	318.00	480.0	576.00	1488

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
WoodDeckSF	0	1.00	93.17	127.74	0	0.00	0.0	168.00	1424
OpenPorchSF	0	1.00	48.31	68.88	0	0.00	28.0	72.00	742
EnclosedPorch	0	1.00	24.24	67.23	0	0.00	0.0	0.00	1012
X3SsnPorch	0	1.00	1.79	20.21	0	0.00	0.0	0.00	360
ScreenPorch	0	1.00	17.06	56.61	0	0.00	0.0	0.00	576
PoolArea	0	1.00	1.74	30.49	0	0.00	0.0	0.00	800
MiscVal	0	1.00	58.17	630.81	0	0.00	0.0	0.00	17000
MoSold	0	1.00	6.10	2.72	1	4.00	6.0	8.00	12
YrSold	0	1.00	2007.77	1.30	2006	2007.00	2008.0	2009.00	2010

```
skim_without_charts(Train)
```

Data summary

Name	Train
Number of rows	1460
Number of columns	81
_____	
Column type frequency:	
character	43
numeric	38
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
MSZoning	0	1.00	2	7	0	5	0
Street	0	1.00	4	4	0	2	0
Alley	1369	0.06	4	4	0	2	0
LotShape	0	1.00	3	3	0	4	0
LandContour	0	1.00	3	3	0	4	0
Utilities	0	1.00	6	6	0	2	0
LotConfig	0	1.00	3	7	0	5	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
LandSlope	0	1.00	3	3	0	3	0
Neighborhood	0	1.00	5	7	0	25	0
Condition1	0	1.00	4	6	0	9	0
Condition2	0	1.00	4	6	0	8	0
BldgType	0	1.00	4	6	0	5	0
HouseStyle	0	1.00	4	6	0	8	0
RoofStyle	0	1.00	3	7	0	6	0
RoofMatl	0	1.00	4	7	0	8	0
Exterior1st	0	1.00	5	7	0	15	0
Exterior2nd	0	1.00	5	7	0	16	0
MasVnrType	8	0.99	4	7	0	4	0
ExterQual	0	1.00	2	2	0	4	0
ExterCond	0	1.00	2	2	0	5	0
Foundation	0	1.00	4	6	0	6	0
BsmtQual	37	0.97	2	2	0	4	0
BsmtCond	37	0.97	2	2	0	4	0
BsmtExposure	38	0.97	2	2	0	4	0
BsmtFinType1	37	0.97	3	3	0	6	0
BsmtFinType2	38	0.97	3	3	0	6	0
Heating	0	1.00	4	5	0	6	0
HeatingQC	0	1.00	2	2	0	5	0
CentralAir	0	1.00	1	1	0	2	0
Electrical	1	1.00	3	5	0	5	0
KitchenQual	0	1.00	2	2	0	4	0
Functional	0	1.00	3	4	0	7	0
FireplaceQu	690	0.53	2	2	0	5	0
GarageType	81	0.94	6	7	0	6	0
GarageFinish	81	0.94	3	3	0	3	0
GarageQual	81	0.94	2	2	0	5	0
GarageCond	81	0.94	2	2	0	5	0
PavedDrive	0	1.00	1	1	0	3	0
PoolQC	1453	0.00	2	2	0	3	0
Fence	1179	0.19	4	5	0	4	0
MiscFeature	1406	0.04	4	4	0	4	0
SaleType	0	1.00	2	5	0	9	0
SaleCondition	0	1.00	6	7	0	6	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1.00	730.50	421.61	1	365.75	730.5	1095.25	1460
MSSubClass	0	1.00	56.90	42.30	20	20.00	50.0	70.00	190
LotFrontage	259	0.82	70.05	24.28	21	59.00	69.0	80.00	313
LotArea	0	1.00	10516.83	9981.26	1300	7553.50	9478.5	11601.50	215245
OverallQual	0	1.00	6.10	1.38	1	5.00	6.0	7.00	10
OverallCond	0	1.00	5.58	1.11	1	5.00	5.0	6.00	9
YearBuilt	0	1.00	1971.27	30.20	1872	1954.00	1973.0	2000.00	2010
YearRemodAdd	0	1.00	1984.87	20.65	1950	1967.00	1994.0	2004.00	2010
MasVnrArea	8	0.99	103.69	181.07	0	0.00	0.0	166.00	1600
BsmtFinSF1	0	1.00	443.64	456.10	0	0.00	383.5	712.25	5644
BsmtFinSF2	0	1.00	46.55	161.32	0	0.00	0.0	0.00	1474
BsmtUnfSF	0	1.00	567.24	441.87	0	223.00	477.5	808.00	2336
TotalBsmtSF	0	1.00	1057.43	438.71	0	795.75	991.5	1298.25	6110
X1stFlrSF	0	1.00	1162.63	386.59	334	882.00	1087.0	1391.25	4692
X2ndFlrSF	0	1.00	346.99	436.53	0	0.00	0.0	728.00	2065
LowQualFinSF	0	1.00	5.84	48.62	0	0.00	0.0	0.00	572
GrLivArea	0	1.00	1515.46	525.48	334	1129.50	1464.0	1776.75	5642
BsmtFullBath	0	1.00	0.43	0.52	0	0.00	0.0	1.00	3
BsmtHalfBath	0	1.00	0.06	0.24	0	0.00	0.0	0.00	2

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
FullBath	0	1.00	1.57	0.55	0	1.00	2.0	2.00	3
HalfBath	0	1.00	0.38	0.50	0	0.00	0.0	1.00	2
BedroomAbvGr	0	1.00	2.87	0.82	0	2.00	3.0	3.00	8
KitchenAbvGr	0	1.00	1.05	0.22	0	1.00	1.0	1.00	3
TotRmsAbvGrd	0	1.00	6.52	1.63	2	5.00	6.0	7.00	14
Fireplaces	0	1.00	0.61	0.64	0	0.00	1.0	1.00	3
GarageYrBlt	81	0.94	1978.51	24.69	1900	1961.00	1980.0	2002.00	2010
GarageCars	0	1.00	1.77	0.75	0	1.00	2.0	2.00	4
GarageArea	0	1.00	472.98	213.80	0	334.50	480.0	576.00	1418
WoodDeckSF	0	1.00	94.24	125.34	0	0.00	0.0	168.00	857
OpenPorchSF	0	1.00	46.66	66.26	0	0.00	25.0	68.00	547
EnclosedPorch	0	1.00	21.95	61.12	0	0.00	0.0	0.00	552
X3SsnPorch	0	1.00	3.41	29.32	0	0.00	0.0	0.00	508
ScreenPorch	0	1.00	15.06	55.76	0	0.00	0.0	0.00	480
PoolArea	0	1.00	2.76	40.18	0	0.00	0.0	0.00	738
MiscVal	0	1.00	43.49	496.12	0	0.00	0.0	0.00	15500
MoSold	0	1.00	6.32	2.70	1	5.00	6.0	8.00	12
YrSold	0	1.00	2007.82	1.33	2006	2007.00	2008.0	2009.00	2010
SalePrice	0	1.00	180921.20	79442.50	34900	129975.00	163000.0	214000.00	755000

```
str(Test)
```

```
## 'data.frame':    1459 obs. of  80 variables:
##  $ Id           : int  1461 1462 1463 1464 1465 1466 1467 1468 1469 1470
...
##  $ MSSubClass   : int  20 20 60 60 120 60 20 60 20 20 ...
##  $ MSZoning     : chr  "RH" "RL" "RL" "RL" ...
```

```
##  $ LotFrontage  : int  80 81 74 78 43 75 NA 63 85 70 ...
##  $ LotArea      : int  11622 14267 13830 9978 5005 10000 7980 8402 10176
8400 ...
##  $ Street       : chr  "Pave" "Pave" "Pave" "Pave" ...
##  $ Alley        : chr  NA NA NA NA ...
##  $ LotShape     : chr  "Reg" "IR1" "IR1" "IR1" ...
##  $ LandContour  : chr  "Lvl" "Lvl" "Lvl" "Lvl" ...
##  $ Utilities    : chr  "AllPub" "AllPub" "AllPub" "AllPub" ...
##  $ LotConfig    : chr  "Inside" "Corner" "Inside" "Inside" ...
##  $ LandSlope    : chr  "Gtl" "Gtl" "Gtl" "Gtl" ...
##  $ Neighborhood : chr  "NAmes" "NAmes" "Gilbert" "Gilbert" ...
##  $ Condition1   : chr  "Feedr" "Norm" "Norm" "Norm" ...
##  $ Condition2   : chr  "Norm" "Norm" "Norm" "Norm" ...
##  $ BldgType     : chr  "1Fam" "1Fam" "1Fam" "1Fam" ...
##  $ HouseStyle   : chr  "1Story" "1Story" "2Story" "2Story" ...
##  $ OverallQual  : int  5 6 5 6 8 6 6 6 7 4 ...
##  $ OverallCond  : int  6 6 5 6 5 5 7 5 5 5 ...
##  $ YearBuilt    : int  1961 1958 1997 1998 1992 1993 1992 1998 1990 1970
...
##  $ YearRemodAdd : int  1961 1958 1998 1998 1992 1994 2007 1998 1990 1970
...
##  $ RoofStyle    : chr  "Gable" "Hip" "Gable" "Gable" ...
##  $ RoofMatl     : chr  "CompShg" "CompShg" "CompShg" "CompShg" ...
##  $ Exterior1st  : chr  "VinylSd" "Wd Sdng" "VinylSd" "VinylSd" ...
##  $ Exterior2nd  : chr  "VinylSd" "Wd Sdng" "VinylSd" "VinylSd" ...
##  $ MasVnrType   : chr  "None" "BrkFace" "None" "BrkFace" ...
##  $ MasVnrArea   : int  0 108 0 20 0 0 0 0 0 0 ...
##  $ ExterQual    : chr  "TA" "TA" "TA" "TA" ...
##  $ ExterCond    : chr  "TA" "TA" "TA" "TA" ...
##  $ Foundation   : chr  "CBlock" "CBlock" "PConc" "PConc" ...
##  $ BsmtQual     : chr  "TA" "TA" "Gd" "TA" ...
##  $ BsmtCond     : chr  "TA" "TA" "TA" "TA" ...
##  $ BsmtExposure : chr  "No" "No" "No" "No" ...
##  $ BsmtFinType1 : chr  "Rec" "ALQ" "GLQ" "GLQ" ...
##  $ BsmtFinSF1   : int  468 923 791 602 263 0 935 0 637 804 ...
##  $ BsmtFinType2 : chr  "LwQ" "Unf" "Unf" "Unf" ...
##  $ BsmtFinSF2   : int  144 0 0 0 0 0 0 0 0 78 ...
##  $ BsmtUnfSF    : int  270 406 137 324 1017 763 233 789 663 0 ...
##  $ TotalBsmtSF  : int  882 1329 928 926 1280 763 1168 789 1300 882 ...
##  $ Heating      : chr  "GasA" "GasA" "GasA" "GasA" ...
##  $ HeatingQC    : chr  "TA" "TA" "Gd" "Ex" ...
##  $ CentralAir   : chr  "Y" "Y" "Y" "Y" ...
##  $ Electrical   : chr  "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
##  $ X1stFlrSF    : int  896 1329 928 926 1280 763 1187 789 1341 882 ...
##  $ X2ndFlrSF    : int  0 0 701 678 0 892 0 676 0 0 ...
##  $ LowQualFinSF : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GrLivArea    : int  896 1329 1629 1604 1280 1655 1187 1465 1341 882 ...
##  $ BsmtFullBath : int  0 0 0 0 0 0 1 0 1 1 ...
##  $ BsmtHalfBath : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FullBath     : int  1 1 2 2 2 2 2 2 1 1 ...
##  $ HalfBath     : int  0 1 1 1 0 1 0 1 1 0 ...
```

```
##  $ BedroomAbvGr : int  2 3 3 3 2 3 3 3 2 2 ...
##  $ KitchenAbvGr : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ KitchenQual  : chr  "TA" "Gd" "TA" "Gd" ...
##  $ TotRmsAbvGrd : int  5 6 6 7 5 7 6 7 5 4 ...
##  $ Functional   : chr  "Typ" "Typ" "Typ" "Typ" ...
##  $ Fireplaces   : int  0 0 1 1 0 1 0 1 1 0 ...
##  $ FireplaceQu  : chr  NA NA "TA" "Gd" ...
##  $ GarageType   : chr  "Attchd" "Attchd" "Attchd" "Attchd" ...
##  $ GarageYrBlt  : int  1961 1958 1997 1998 1992 1993 1992 1998 1990 1970
...
##  $ GarageFinish : chr  "Unf" "Unf" "Fin" "Fin" ...
##  $ GarageCars   : int  1 1 2 2 2 2 2 2 2 2 ...
##  $ GarageArea   : int  730 312 482 470 506 440 420 393 506 525 ...
##  $ GarageQual   : chr  "TA" "TA" "TA" "TA" ...
##  $ GarageCond   : chr  "TA" "TA" "TA" "TA" ...
##  $ PavedDrive   : chr  "Y" "Y" "Y" "Y" ...
##  $ WoodDeckSF   : int  140 393 212 360 0 157 483 0 192 240 ...
##  $ OpenPorchSF  : int  0 36 34 36 82 84 21 75 0 0 ...
##  $ EnclosedPorch: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ X3SsnPorch   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ScreenPorch  : int  120 0 0 0 144 0 0 0 0 0 ...
##  $ PoolArea     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolQC       : chr  NA NA NA NA ...
##  $ Fence        : chr  "MnPrv" NA "MnPrv" NA ...
##  $ MiscFeature  : chr  NA "Gar2" NA NA ...
##  $ MiscVal      : int  0 12500 0 0 0 0 500 0 0 0 ...
##  $ MoSold       : int  6 6 3 6 1 4 3 5 2 4 ...
##  $ YrSold       : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010
...
##  $ SaleType     : chr  "WD" "WD" "WD" "WD" ...
##  $ SaleCondition: chr  "Normal" "Normal" "Normal" "Normal" ...
```

```
str(Train)
```

```
## 'data.frame':    1460 obs. of  81 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" ...
##  $ LotFrontage  : int  65 80 68 60 84 85 75 NA 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120
7420 ...
##  $ Street       : chr  "Pave" "Pave" "Pave" "Pave" ...
##  $ Alley        : chr  NA NA NA NA ...
##  $ 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" ...
##  $ MasVnrType   : chr  "BrkFace" "None" "BrkFace" "None" ...
##  $ MasVnrArea   : int  196 0 162 0 350 0 186 240 0 0 ...
##  $ ExterQual    : chr  "Gd" "TA" "Gd" "TA" ...
##  $ ExterCond    : chr  "TA" "TA" "TA" "TA" ...
##  $ Foundation   : chr  "PConc" "CBlock" "PConc" "BrkTil" ...
##  $ BsmtQual     : chr  "Gd" "Gd" "Gd" "TA" ...
##  $ BsmtCond     : chr  "TA" "TA" "TA" "Gd" ...
##  $ BsmtExposure : chr  "No" "Gd" "Mn" "No" ...
##  $ BsmtFinType1 : chr  "GLQ" "ALQ" "GLQ" "ALQ" ...
##  $ BsmtFinSF1   : int  706 978 486 216 655 732 1369 859 0 851 ...
##  $ BsmtFinType2 : chr  "Unf" "Unf" "Unf" "Unf" ...
##  $ 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 ...
##  $ FireplaceQu  : chr  NA "TA" "TA" "Gd" ...
##  $ GarageType   : chr  "Attchd" "Attchd" "Attchd" "Detchd" ...
##  $ GarageYrBlt  : int  2003 1976 2001 1998 2000 1993 2004 1973 1931 1939
...
##  $ GarageFinish : chr  "RFn" "RFn" "RFn" "Unf" ...
##  $ GarageCars   : int  2 2 2 3 3 2 2 2 2 1 ...
##  $ GarageArea   : int  548 460 608 642 836 480 636 484 468 205 ...
```

```
##  $ GarageQual   : chr  "TA" "TA" "TA" "TA" ...
##  $ GarageCond   : chr  "TA" "TA" "TA" "TA" ...
##  $ 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 ...
##  $ PoolQC       : chr  NA NA NA NA ...
##  $ Fence        : chr  NA NA NA NA ...
##  $ MiscFeature  : chr  NA NA NA NA ...
##  $ 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 ...
```

2.3.1 Any missing value

Missing value in Rows

```
missing_row_train <- Train[!complete.cases(Train),]

head(missing_row_train)
```

```
##   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

missing_row_test <- Test[!complete.cases(Test),]

head(missing_row_test)
```

```
##     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
```

```
nrow(Train)
```

```
## [1] 1460
```

```
nrow(Test)
```

```
## [1] 1459
```

```
nrow(missing_row_test)
```

```
## [1] 1459
```

```
nrow(missing_row_train)
```

```
## [1] 1460
```

Combined Dataset

```
Complete_Data <- merge(Test, Train, all = TRUE)

view(Complete_Data)

head(Complete_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
```

```
dim(Complete_Data)
```

```
## [1] 2919   81
```

Selecting important variable for working

```
variable_name <- names(Complete_Data)

variable_name
```

```
##  [1] "Id"            "MSSubClass"    "MSZoning"      "LotFrontage"
##  [5] "LotArea"       "Street"        "Alley"         "LotShape"
##  [9] "LandContour"   "Utilities"     "LotConfig"     "LandSlope"
## [13] "Neighborhood"  "Condition1"    "Condition2"    "BldgType"
## [17] "HouseStyle"    "OverallQual"   "OverallCond"   "YearBuilt"
## [21] "YearRemodAdd"  "RoofStyle"     "RoofMatl"      "Exterior1st"
## [25] "Exterior2nd"   "MasVnrType"    "MasVnrArea"    "ExterQual"
## [29] "ExterCond"     "Foundation"    "BsmtQual"      "BsmtCond"
## [33] "BsmtExposure"  "BsmtFinType1"  "BsmtFinSF1"    "BsmtFinType2"
## [37] "BsmtFinSF2"    "BsmtUnfSF"     "TotalBsmtSF"   "Heating"
## [41] "HeatingQC"     "CentralAir"    "Electrical"    "X1stFlrSF"
## [45] "X2ndFlrSF"     "LowQualFinSF"  "GrLivArea"     "BsmtFullBath"
## [49] "BsmtHalfBath"  "FullBath"      "HalfBath"      "BedroomAbvGr"
## [53] "KitchenAbvGr"  "KitchenQual"   "TotRmsAbvGrd"  "Functional"
## [57] "Fireplaces"    "FireplaceQu"   "GarageType"    "GarageYrBlt"
## [61] "GarageFinish"  "GarageCars"    "GarageArea"    "GarageQual"
## [65] "GarageCond"    "PavedDrive"    "WoodDeckSF"    "OpenPorchSF"
## [69] "EnclosedPorch" "X3SsnPorch"    "ScreenPorch"   "PoolArea"
## [73] "PoolQC"        "Fence"         "MiscFeature"   "MiscVal"
## [77] "MoSold"        "YrSold"        "SaleType"      "SaleCondition"
## [81] "SalePrice"
```

Variables for new dataset

```
 select_var <- c("Id", "MSZoning", "LotArea", "Utilities", "BldgType",
"HouseStyle", "OverallQual", "OverallCond", "YearBuilt", "ExterQual",
"ExterCond", "BsmtQual", "BsmtCond", "Heating", "HeatingQC", "CentralAir",
"Electrical", "GrLivArea", "BedroomAbvGr", "KitchenAbvGr","KitchenQual",
"TotRmsAbvGrd", "Functional", "Fireplaces", "GarageType",
"PoolArea","Fence","MoSold",   "YrSold", "SaleType", "SaleCondition",
"SalePrice")
```

2.4 New dataset for prediction

```
new_dataset <- Complete_Data[, select_var]

head(new_dataset)
```

```
##   Id MSZoning LotArea Utilities BldgType HouseStyle OverallQual
OverallCond
## 1  1       RL    8450    AllPub     1Fam     2Story           7
5
## 2  2       RL    9600    AllPub     1Fam     1Story           6
8
## 3  3       RL   11250    AllPub     1Fam     2Story           7
5
## 4  4       RL    9550    AllPub     1Fam     2Story           7
5
## 5  5       RL   14260    AllPub     1Fam     2Story           8
5
## 6  6       RL   14115    AllPub     1Fam     1.5Fin           5
5
##   YearBuilt ExterQual ExterCond BsmtQual BsmtCond Heating HeatingQC
CentralAir
## 1      2003        Gd        TA       Gd       TA    GasA        Ex
Y
## 2      1976        TA        TA       Gd       TA    GasA        Ex
Y
## 3      2001        Gd        TA       Gd       TA    GasA        Ex
Y
## 4      1915        TA        TA       TA       Gd    GasA        Gd
Y
## 5      2000        Gd        TA       Gd       TA    GasA        Ex
Y
## 6      1993        TA        TA       Gd       TA    GasA        Ex
Y
##   Electrical GrLivArea BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd
## 1      SBrkr      1710            3            1          Gd            8
## 2      SBrkr      1262            3            1          TA            6
## 3      SBrkr      1786            3            1          Gd            6
## 4      SBrkr      1717            3            1          Gd            7
## 5      SBrkr      2198            4            1          Gd            9
## 6      SBrkr      1362            1            1          TA            5
##   Functional Fireplaces GarageType PoolArea Fence MoSold YrSold SaleType
## 1        Typ          0     Attchd        0  <NA>      2   2008       WD
## 2        Typ          1     Attchd        0  <NA>      5   2007       WD
## 3        Typ          1     Attchd        0  <NA>      9   2008       WD
## 4        Typ          1     Detchd        0  <NA>      2   2006       WD
## 5        Typ          1     Attchd        0  <NA>     12   2008       WD
## 6        Typ          0     Attchd        0 MnPrv     10   2009       WD
##   SaleCondition SalePrice
## 1        Normal    208500
## 2        Normal    181500
## 3        Normal    223500
## 4       Abnorml    140000
## 5        Normal    250000
## 6        Normal    143000
```

```
summary(new_dataset)
```

```
##        Id           MSZoning            LotArea        Utilities
##  Min.   :   1.0   Length:2919        Min.   :  1300   Length:2919
##  1st Qu.: 730.5   Class :character   1st Qu.:  7478   Class :character
##  Median :1460.0   Mode  :character   Median :  9453   Mode  :character
##  Mean   :1460.0                      Mean   : 10168
##  3rd Qu.:2189.5                      3rd Qu.: 11570
##  Max.   :2919.0                      Max.   :215245
##
##    BldgType          HouseStyle         OverallQual      OverallCond
##  Length:2919        Length:2919        Min.   : 1.000   Min.   :1.000
##  Class :character   Class :character   1st Qu.: 5.000   1st Qu.:5.000
##  Mode  :character   Mode  :character   Median : 6.000   Median :5.000
##                                        Mean   : 6.089   Mean   :5.565
##                                        3rd Qu.: 7.000   3rd Qu.:6.000
##                                        Max.   :10.000   Max.   :9.000
##
##    YearBuilt     ExterQual          ExterCond           BsmtQual
##  Min.   :1872   Length:2919        Length:2919        Length:2919
##  1st Qu.:1954   Class :character   Class :character   Class :character
##  Median :1973   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1971
##  3rd Qu.:2001
##  Max.   :2010
##
##    BsmtCond           Heating           HeatingQC          CentralAir
##  Length:2919        Length:2919        Length:2919        Length:2919
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Electrical          GrLivArea     BedroomAbvGr   KitchenAbvGr
##  Length:2919        Min.   : 334   Min.   :0.00   Min.   :0.000
##  Class :character   1st Qu.:1126   1st Qu.:2.00   1st Qu.:1.000
##  Mode  :character   Median :1444   Median :3.00   Median :1.000
##                     Mean   :1501   Mean   :2.86   Mean   :1.045
##                     3rd Qu.:1744   3rd Qu.:3.00   3rd Qu.:1.000
##                     Max.   :5642   Max.   :8.00   Max.   :3.000
##
##  KitchenQual         TotRmsAbvGrd     Functional          Fireplaces
##  Length:2919        Min.   : 2.000   Length:2919        Min.   :0.0000
##  Class :character   1st Qu.: 5.000   Class :character   1st Qu.:0.0000
##  Mode  :character   Median : 6.000   Mode  :character   Median :1.0000
##                     Mean   : 6.452                      Mean   :0.5971
##                     3rd Qu.: 7.000                      3rd Qu.:1.0000
##                     Max.   :15.000                      Max.   :4.0000
##
##   GarageType           PoolArea          Fence               MoSold
##  Length:2919        Min.   :  0.000   Length:2919        Min.   : 1.000
##  Class :character   1st Qu.:  0.000   Class :character   1st Qu.: 4.000
```

```
##  Mode  :character   Median :  0.000   Mode  :character   Median : 6.000
##                     Mean   :  2.252                     Mean   : 6.213
##                     3rd Qu.:  0.000                     3rd Qu.: 8.000
##                     Max.   :800.000                     Max.   :12.000
##
##      YrSold       SaleType         SaleCondition        SalePrice
##  Min.   :2006   Length:2919        Length:2919        Min.   : 34900
##  1st Qu.:2007   Class :character   Class :character   1st Qu.:129975
##  Median :2008   Mode  :character   Mode  :character   Median :163000
##  Mean   :2008                                         Mean   :180921
##  3rd Qu.:2009                                         3rd Qu.:214000
##  Max.   :2010                                         Max.   :755000
##                                                       NA's   :1459
```

```
cleaned_data <- new_dataset[complete.cases(new_dataset), ]
```

2.4.1 Analyze form sales price

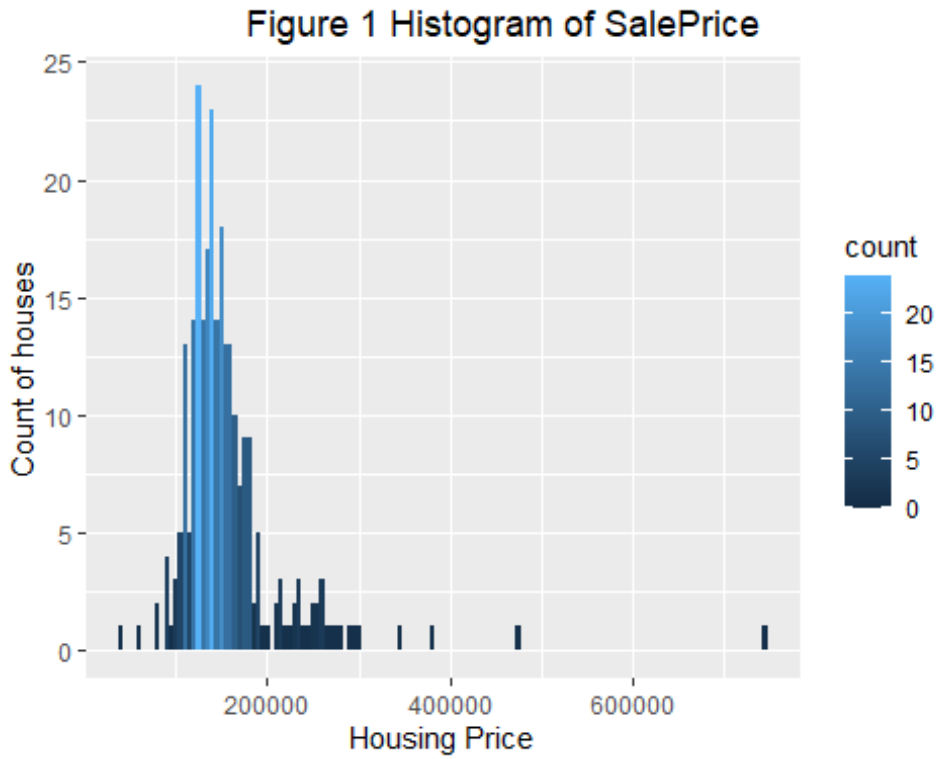
```
summary(cleaned_data$SalePrice)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   40000  125500  144000  157209  168500  745000
```

Distribustion of sales price

```
options(scipen=10000)
ggplot(cleaned_data, aes(x = SalePrice, fill = ..count..)) +
  geom_histogram(binwidth = 5000) +
  ggtitle("Figure 1 Histogram of SalePrice") +
  ylab("Count of houses") +
  xlab("Housing Price") +
  theme(plot.title = element_text(hjust = 0.8))
```

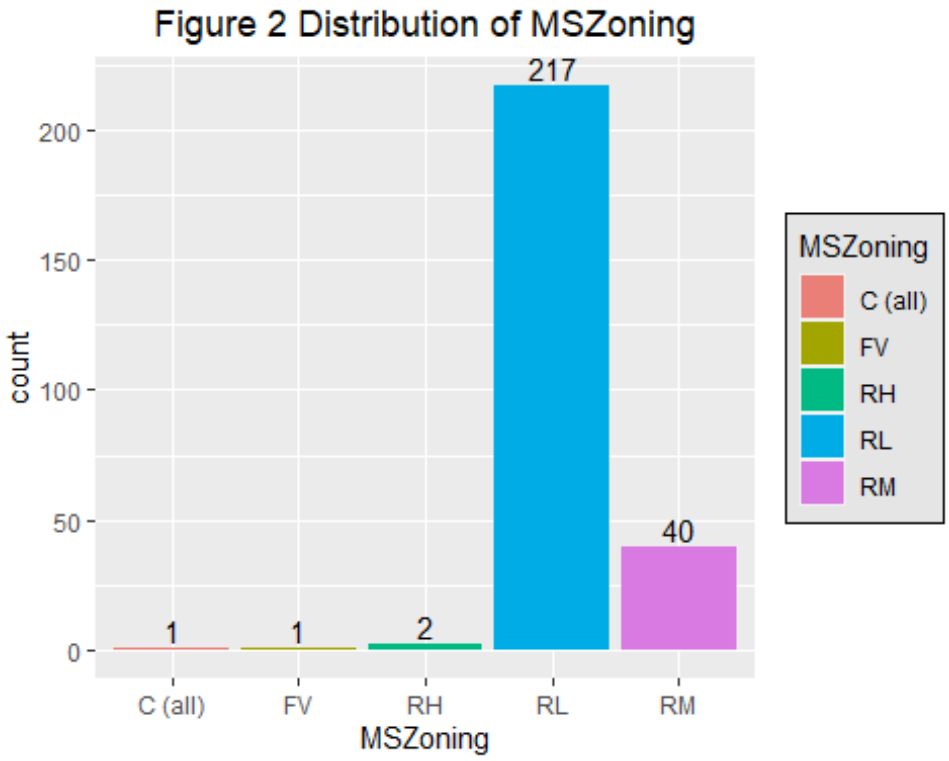
```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2
3.4.0.
## ℹ Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Distribution of Sales price by MSZoing

```
options(repr.plot.width=5, repr.plot.height=4)
ggplot(cleaned_data, aes(x = MSZoning, fill = MSZoning )) +
geom_bar()+
scale_fill_hue(c = 80)+
ggtitle("Figure 2 Distribution of MSZoning")+
theme(plot.title = element_text(hjust = 0.5),legend.position="right",
legend.background = element_rect(fill="grey90",

size=0.5, linetype="solid",

colour ="black"))+
geom_text(stat='count',aes(label=..count..),vjust=-0.25)
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of
ggplot2 3.4.0.
## ℹ Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

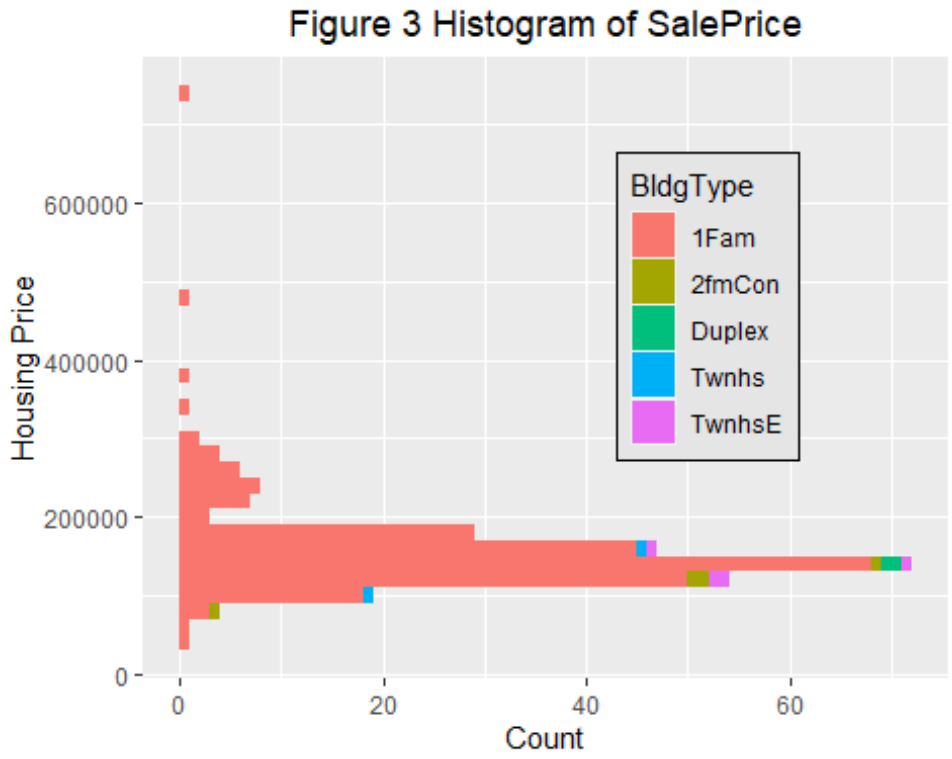


Distribution of sale price by BldfType

```
ddply(cleaned_data, .(BldgType), summarize,Total =
length(BldgType),Max_price=max(SalePrice),Min_price=min(SalePrice))
```

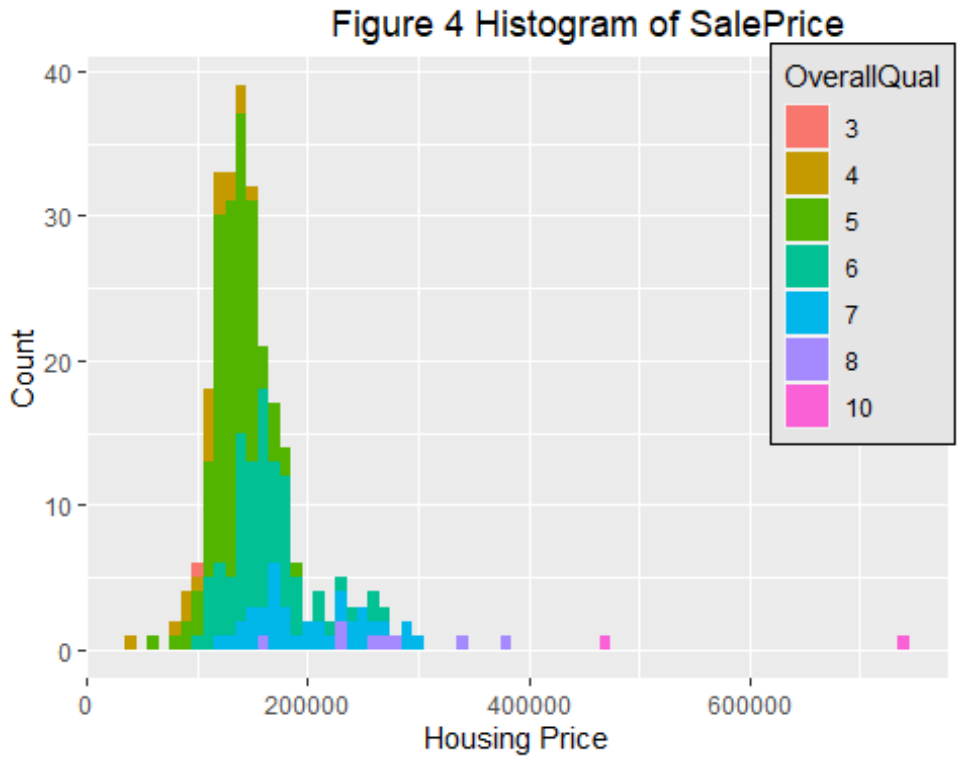
```
##   BldgType Total Max_price Min_price
## 1     1Fam   249    745000     40000
## 2   2fmCon     4    140000     80000
## 3   Duplex     2    144000    132500
## 4    Twnhs     2    151000     91000
## 5   TwnhsE     4    151400    124000
```

```
ggplot(cleaned_data, aes(SalePrice)) +
 geom_histogram(aes(fill = BldgType), position = position_stack(reverse =
TRUE), binwidth = 20000) +
 coord_flip() + ggtitle("Figure 3 Histogram of SalePrice") +
 ylab("Count") +
 xlab("Housing Price") +
 theme(plot.title = element_text(hjust = 0.5),legend.position=c(0.7,0.6),
legend.background = element_rect(fill="grey90",

size=0.6, linetype="solid",

colour ="black"))
```



Distribution of price by OverallQual

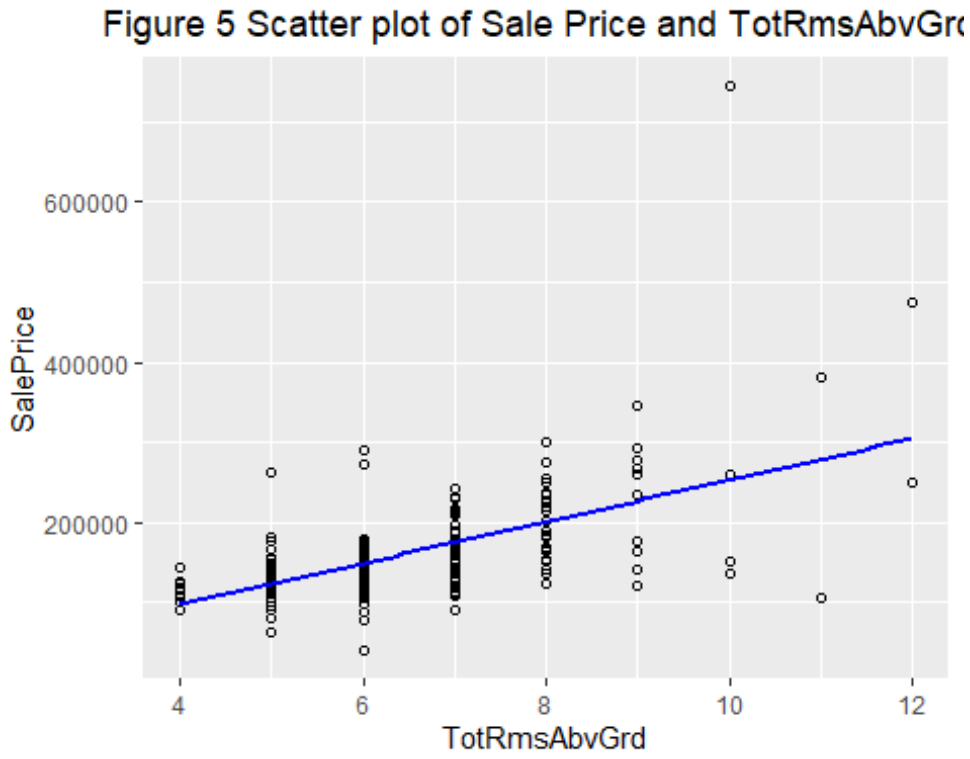
```
ggplot(cleaned_data, aes(x = SalePrice,fill = as.factor(OverallQual))) +
  geom_histogram(position = "stack", binwidth = 10000) +
  ggtitle("Figure 4 Histogram of SalePrice") +
  ylab("Count") +
  xlab("Housing Price") +
  scale_fill_discrete(name="OverallQual")+
  theme(plot.title = element_text(hjust = 0.7), legend.position=c(0.9,0.7),
legend.background = element_rect(fill="grey90",

size=0.7, linetype="solid",

colour ="black"))
```



2.5. Corelation Exploreation

Correlation between Sales price and TotRmsAbvGrd

```
ggplot(cleaned_data, aes(x=TotRmsAbvGrd, y=SalePrice)) +
  geom_point(shape=1) +
  geom_smooth(method=lm , color="blue", se=FALSE)+
  ggtitle("Figure 5 Scatter plot of Sale Price and TotRmsAbvGrd") +
  theme(plot.title = element_text(hjust = 0.6))

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



Correlation between Sales price and heating

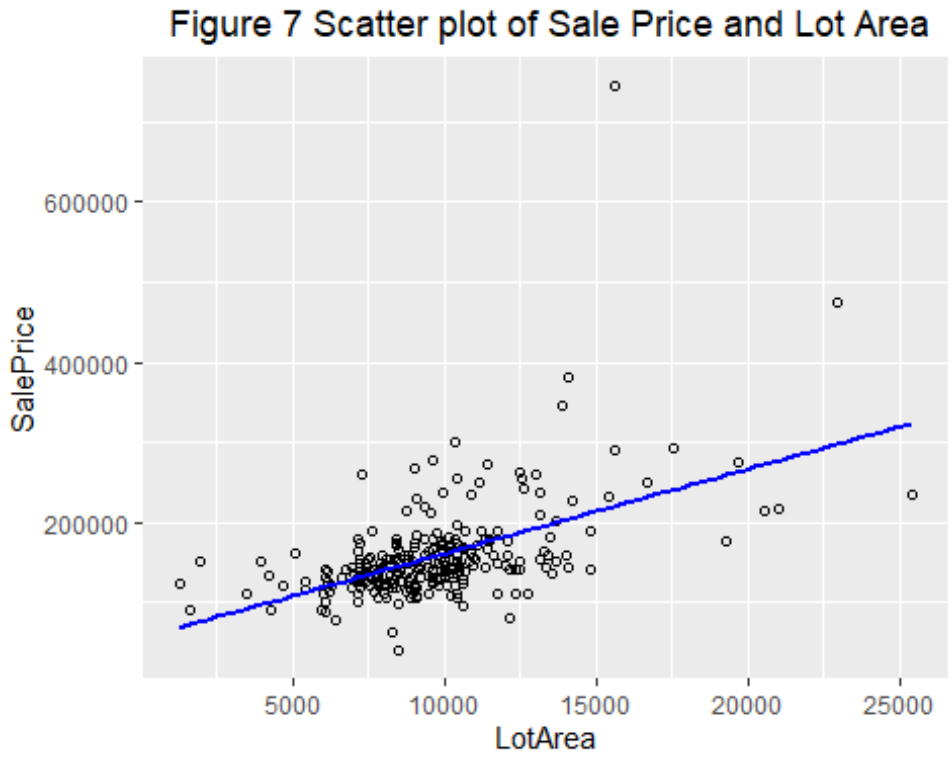
```
ggplot(cleaned_data, aes(x=Heating, y=SalePrice)) +
  geom_point(shape=1) +
  geom_smooth(method=lm , color="blue", se=FALSE)+
  ggtitle("Figure 6 Scatter plot of Sale Price and Heating") +
  theme(plot.title = element_text(hjust = 0.6))
```

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

Correlation between Sales price and Lot Area

```
ggplot(cleaned_data, aes(x=LotArea, y=SalePrice)) +
  geom_point(shape=1) +
  geom_smooth(method=lm , color="blue", se=FALSE)+
  ggtitle("Figure 7 Scatter plot of Sale Price and Lot Area") +
  theme(plot.title = element_text(hjust = 0.6))
```

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



Correlation between Sales price and House Style

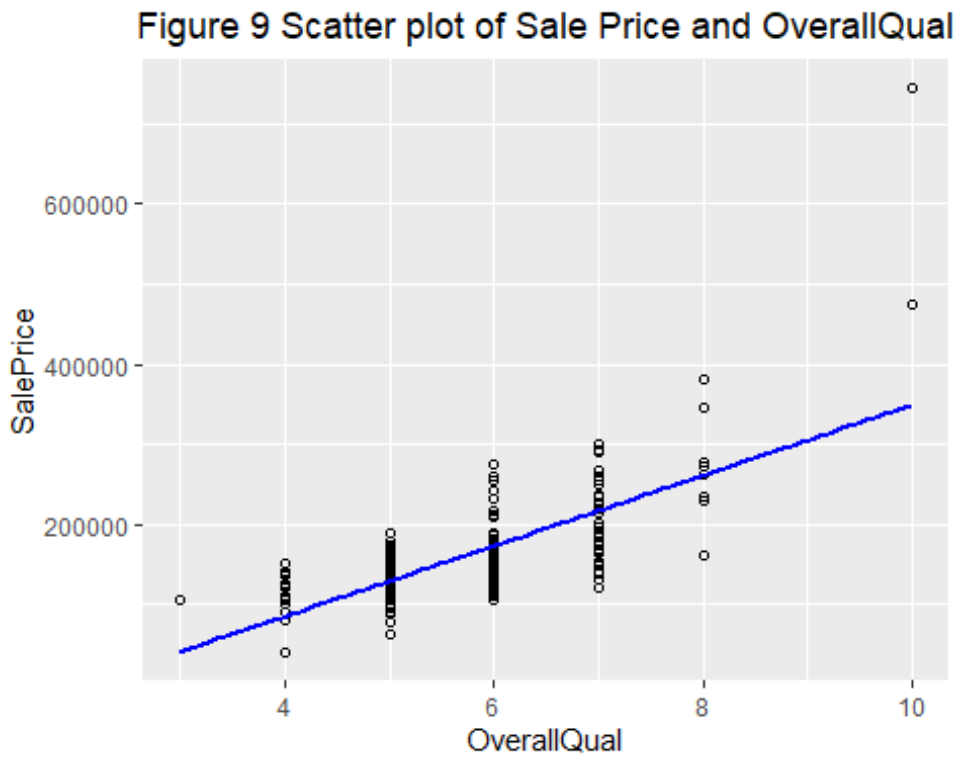
```
ggplot(cleaned_data, aes(x=HouseStyle, y=SalePrice)) +
  geom_point(shape=1) +
  geom_smooth(method=lm , color="blue", se=FALSE)+
  ggtitle("Figure 8 Scatter plot of Sale Price and House Style") +
  theme(plot.title = element_text(hjust = 0.6))
```

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

Correlation between Sales price and OverallQual

```
ggplot(cleaned_data, aes(x=OverallQual, y=SalePrice)) +
  geom_point(shape=1) +
  geom_smooth(method=lm , color="blue", se=FALSE)+
  ggtitle("Figure 9 Scatter plot of Sale Price and OverallQual") +
  theme(plot.title = element_text(hjust = 0.6))
```

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

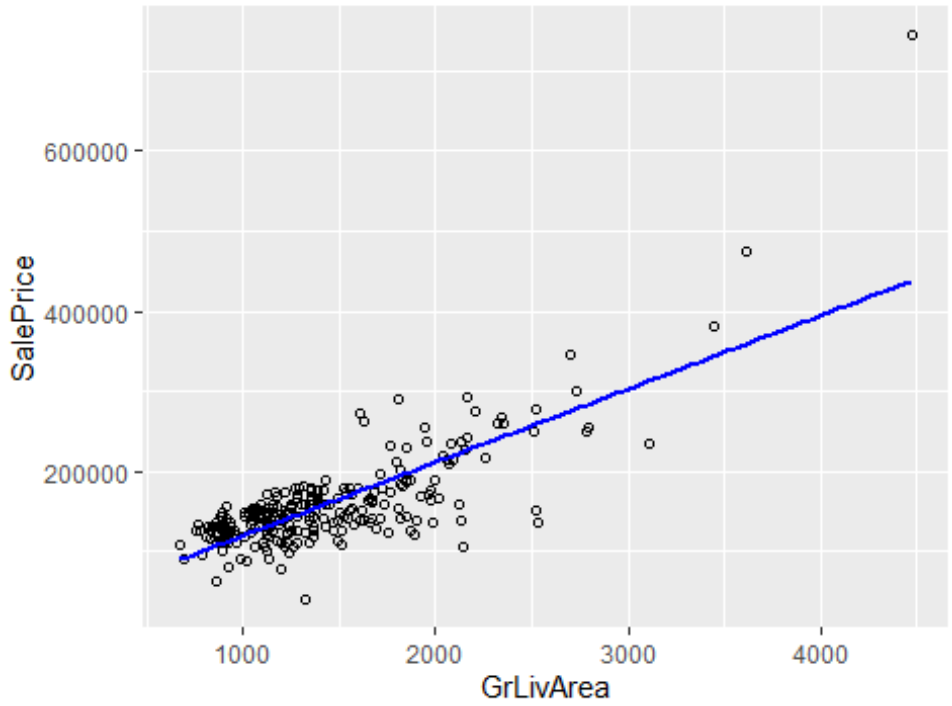


Correlation between Sales price and GrLivArea

```
ggplot(cleaned_data, aes(x=GrLivArea, y=SalePrice)) +
  geom_point(shape=1) +
  geom_smooth(method=lm , color="blue", se=FALSE)+
  ggtitle("Figure 10 Scatter plot of Sale Price and GrLivArea") +
  theme(plot.title = element_text(hjust = 0.6))
```

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

Figure 10 Scatter plot of Sale Price and GrLivArea



3. Linear Regression Model

3.1 Varibles for regression model

```
model_variables <- c("Id", "OverallQual", "OverallCond", "YearBuilt",
"ExterQual", "ExterCond", "BsmtQual", "BsmtCond", "HeatingQC", "CentralAir",
"Electrical", "GrLivArea", "BedroomAbvGr", "KitchenAbvGr","KitchenQual",
"TotRmsAbvGrd", "Functional", "Fireplaces", "GarageType",
"PoolArea","Fence","SalePrice")

Newmodel_Dataset <- new_dataset[, model_variables]

Newmodel_Dataset$SalePrice1 <- log(Newmodel_Dataset$SalePrice)
```

Dividing dataset in two parts. one is model_data_1 and other is model_data_2

```
set.seed(10000)
model.index <- sample(c(1:dim(Newmodel_Dataset)[1]),
dim(Newmodel_Dataset)[1]*0.6)
model_data_1 = Newmodel_Dataset[model.index,]
model_data_2 <- Newmodel_Dataset[-model.index,]
```

Run Regression

```
Regdata <- lm(SalePrice1~.-SalePrice, data = model_data_1)
summary(Regdata)

##
## Call:
## lm(formula = SalePrice1 ~ . - SalePrice, data = model_data_1)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58093 -0.05092  0.00868  0.06903  0.28433
##
## Coefficients:
##                       Estimate  Std. Error t value     Pr(>|t|)
## (Intercept)        4.73487646  1.54030498   3.074     0.002578 **
## Id                 0.00001410  0.00002587   0.545     0.586596
## OverallQual        0.04313412  0.01880166   2.294     0.023397 *
## OverallCond        0.04407245  0.01233601   3.573     0.000498 ***
## YearBuilt          0.00332614  0.00074869   4.443 0.0000189215 ***
## ExterQualFa        0.11415256  0.25201331   0.453     0.651337
## ExterQualGd       -0.03536793  0.14881714  -0.238     0.812522
## ExterQualTA       -0.06991311  0.14804031  -0.472     0.637541
## ExterCondGd       -0.15777158  0.10242499  -1.540     0.125922
## ExterCondTA       -0.14712725  0.09845090  -1.494     0.137507
## BsmtQualFa        -0.06322107  0.21183575  -0.298     0.765845
## BsmtQualGd         0.00567761  0.17907941   0.032     0.974757
## BsmtQualTA        -0.01275464  0.18375763  -0.069     0.944771
## BsmtCondGd         0.09426064  0.11068649   0.852     0.396014
## BsmtCondTA         0.09363921  0.10126322   0.925     0.356843
## HeatingQCFa       -0.00244646  0.06280144  -0.039     0.968986
## HeatingQCGd       -0.02124396  0.03342498  -0.636     0.526182
## HeatingQCTA       -0.01775013  0.02921331  -0.608     0.544518
## CentralAirY        0.17016282  0.06201102   2.744     0.006933 **
## ElectricalFuseF   -0.07096834  0.14805267  -0.479     0.632505
## ElectricalSBrkr   -0.02465211  0.06279745  -0.393     0.695288
## GrLivArea          0.00033579  0.00005545   6.056 0.0000000142 ***
## BedroomAbvGr      -0.03890662  0.02475438  -1.572     0.118468
## KitchenAbvGr      -0.08013346  0.07386629  -1.085     0.280014
## KitchenQualFa     -0.04351024  0.14945526  -0.291     0.771423
## KitchenQualGd     -0.12820636  0.08228433  -1.558     0.121662
## KitchenQualTA     -0.16019266  0.08528122  -1.878     0.062583 .
## TotRmsAbvGrd       0.00061777  0.01797881   0.034     0.972643
## FunctionalMin1     0.01030984  0.14251305   0.072     0.942441
## FunctionalMin2     0.02243113  0.15051169   0.149     0.881761
## FunctionalMod      0.18996124  0.20470958   0.928     0.355164
## FunctionalTyp      0.02854819  0.13709851   0.208     0.835377
## Fireplaces         0.02020372  0.01950781   1.036     0.302292
## GarageTypeAttchd   0.01753320  0.16232019   0.108     0.914151
## GarageTypeBasment  0.02667108  0.18709295   0.143     0.886864
## GarageTypeBuiltIn -0.16329184  0.18829562  -0.867     0.387437
## GarageTypeDetchd  -0.03447723  0.16380968  -0.210     0.833632
## PoolArea           0.00012923  0.00010904   1.185     0.238142
## FenceGdWo         -0.02987566  0.03658566  -0.817     0.415666
## FenceMnPrv        -0.00936012  0.02973914  -0.315     0.753467
## FenceMnWw         -0.06031814  0.05702184  -1.058     0.292120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

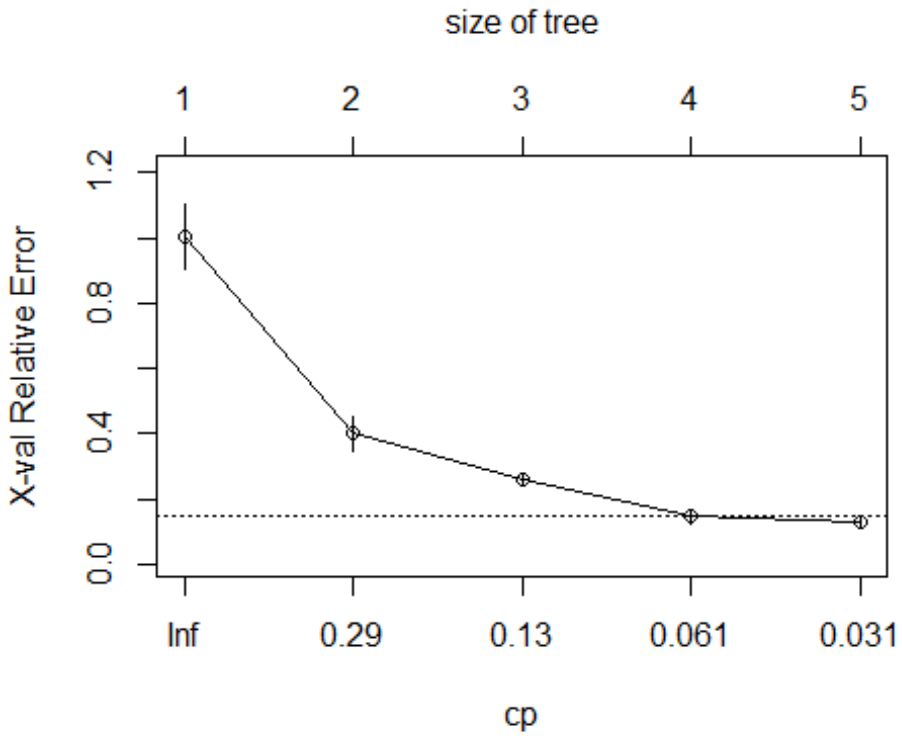
```
## Residual standard error: 0.1254 on 129 degrees of freedom
##   (1581 observations deleted due to missingness)
## Multiple R-squared:  0.8686, Adjusted R-squared:  0.8279
## F-statistic: 21.32 on 40 and 129 DF,  p-value: < 0.00000000000000022
```

3.2 Classification of Tree

install.package("rpart")

install.packages("rpart.plot")

```
library(rpart)

library(rpart.plot)

Newmodel.tree <- rpart(SalePrice~.-SalePrice,
                    data = model_data_1,control = rpart.control(cp = 0.03))

plotcp(Newmodel.tree)
```




```
printcp(Newmodel.tree)

##
## Regression tree:
## rpart(formula = SalePrice ~ . - SalePrice, data = model_data_1,
##     control = rpart.control(cp = 0.03))
##
## Variables actually used in tree construction:
## [1] SalePrice1
```

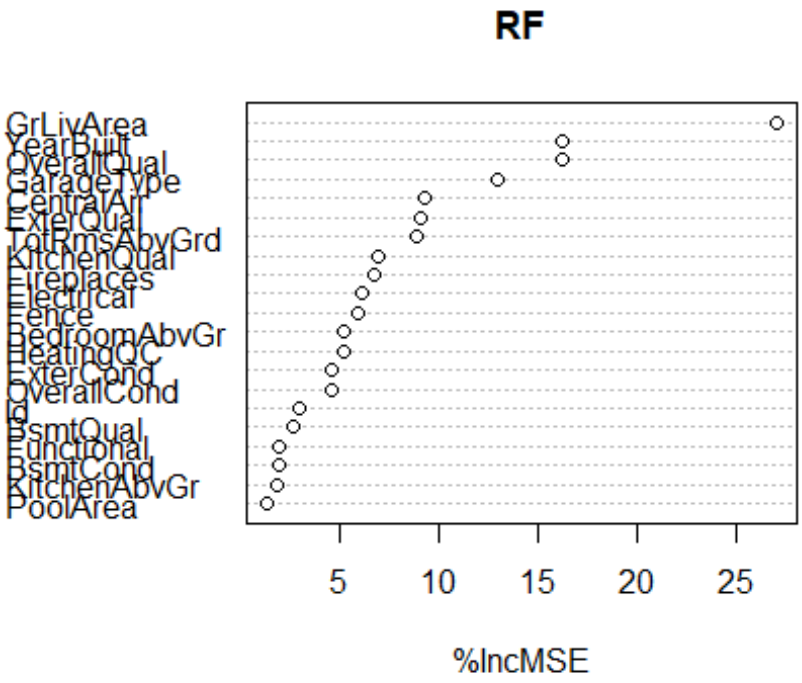
```
##
## Root node error: 5503053366469/908 = 6060631461
##
## n=908 (843 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.60424      0   1.00000 1.00329 0.099290
## 2 0.14180      1   0.39576 0.40023 0.053885
## 3 0.11439      2   0.25396 0.25925 0.020086
## 4 0.03232      3   0.13957 0.14511 0.018960
## 5 0.03000      4   0.10725 0.13094 0.018802
```

3.3 Random Forest

install.packages("randomForest")

install.packages("yardstick")

```
library(randomForest)

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

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

## The following object is masked from 'package:ggplot2':
##
##     margin

library(yardstick)

##
## Attaching package: 'yardstick'

## The following object is masked from 'package:readr':
##
##     spec

model_data_3 <-na.omit(model_data_1)

RF <- randomForest(SalePrice1 ~ . - SalePrice, data = model_data_3,
                   importance = TRUE, ntree = 500, nodesize = 7)

options(repr.plot.width = 9, repr.plot.height = 6)
varImpPlot(RF, type = 1)
```




```
rf.pred <- predict(RF, newdata=model_data_2)

plot(rf.pred, model_data_2$lSalePrice1, main = "Figure 10 Predicted vs.
Actual log SalePrice")
abline(0,1)
```