

Machine Learning Project

Michal Heydel

29 November 2017

```
setwd("C:/Users/Michal/Documents/01- Master Degree/GitHub/ST443-Project-group9/Housing price data")
getwd()
```

```
## [1] "C:/Users/Michal/Documents/01- Master Degree/GitHub/ST443-Project-group9/Housing price data"
```

```
train = read.csv("train.csv", row.names = "Id", stringsAsFactors=FALSE)
testing_kaggle = read.csv("test.csv", row.names = "Id", stringsAsFactors=FALSE)
```

```
#combining train and test data for quicker data prep
```

```
testing_kaggle$SalePrice <- NA
```

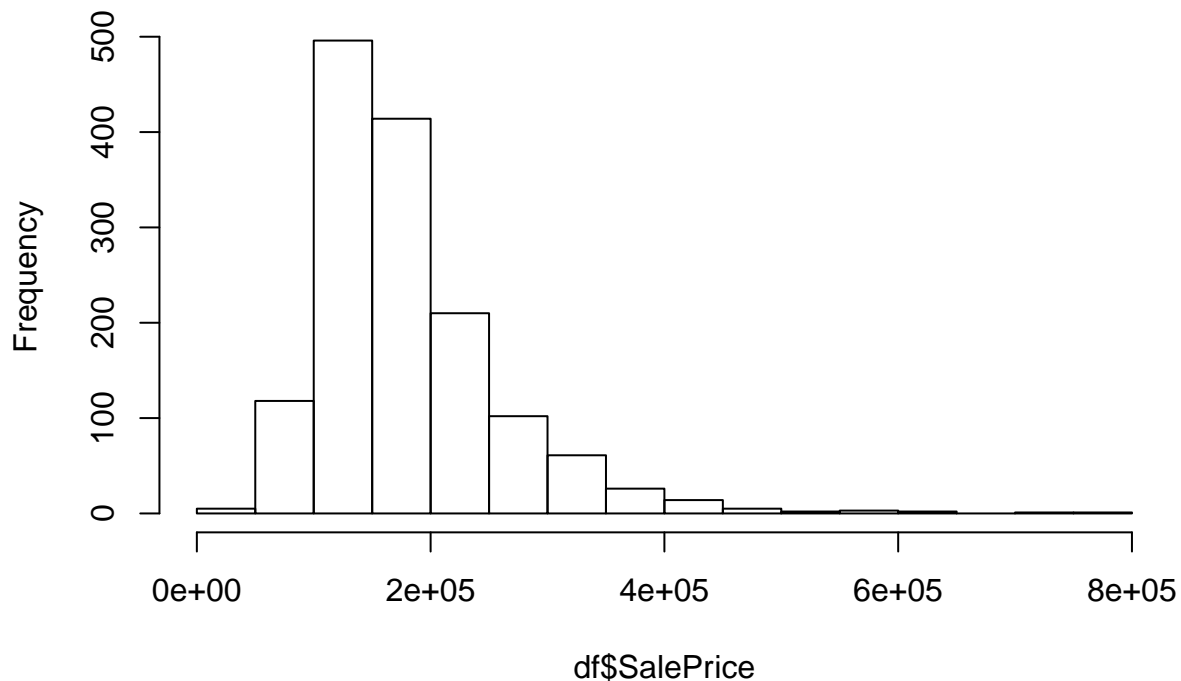
```
train$isTrain <- 1
```

```
testing_kaggle$isTrain <- 0
```

```
df <- rbind(train,testing_kaggle)
```

```
hist(df$SalePrice)
```

Histogram of df\$SalePrice



```
colSums(sapply(df, is.na))
```

```
##    MSSubClass    MSZoning  LotFrontage    LotArea    Street
##         0         4         486         0         0
##    Alley    LotShape  LandContour  Utilities    LotConfig
```

```
##      2721      0      0      2      0
##      LandSlope Neighborhood Condition1 Condition2 BldgType
##      0      0      0      0      0
##      HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd
##      0      0      0      0      0
##      RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType
##      0      0      1      1      24
##      MasVnrArea ExterQual ExterCond Foundation BsmtQual
##      23      0      0      0      81
##      BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
##      82      82      79      1      80
##      BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC
##      1      1      1      0      0
##      CentralAir Electrical X1stFlrSF X2ndFlrSF LowQualFinSF
##      0      1      0      0      0
##      GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath
##      0      2      2      0      0
##      BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
##      0      0      1      0      2
##      Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish
##      0      1420      157      159      159
##      GarageCars GarageArea GarageQual GarageCond PavedDrive
##      1      1      159      159      0
##      WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch
##      0      0      0      0      0
##      PoolArea PoolQC Fence MiscFeature MiscVal
##      0      2909      2348      2814      0
##      MoSold YrSold SaleType SaleCondition SalePrice
##      0      0      1      0      1459
##      isTrain
##      0
```

```
for(i in colnames(df[,sapply(df, is.character)])){
  df[,i][which(is.na(df[,i]))] <- "None"
}
```

```
colSums(sapply(df, is.na))
```

```
##      MSSubClass MSZoning LotFrontage LotArea Street
##      0      0      486      0      0
##      Alley LotShape LandContour Utilities LotConfig
##      0      0      0      0      0
##      LandSlope Neighborhood Condition1 Condition2 BldgType
##      0      0      0      0      0
##      HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd
##      0      0      0      0      0
##      RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType
##      0      0      0      0      0
##      MasVnrArea ExterQual ExterCond Foundation BsmtQual
##      23      0      0      0      0
##      BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
##      0      0      0      1      0
##      BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC
##      1      1      1      0      0
```

```
##      CentralAir      Electrical      X1stFlrSF      X2ndFlrSF      LowQualFinSF
##           0           0           0           0           0
##      GrLivArea      BsmtFullBath      BsmtHalfBath      FullBath      HalfBath
##           0           2           2           0           0
##      BedroomAbvGr      KitchenAbvGr      KitchenQual      TotRmsAbvGrd      Functional
##           0           0           0           0           0
##      Fireplaces      FireplaceQu      GarageType      GarageYrBlt      GarageFinish
##           0           0           0           159           0
##      GarageCars      GarageArea      GarageQual      GarageCond      PavedDrive
##           1           1           0           0           0
##      WoodDeckSF      OpenPorchSF      EnclosedPorch      X3SsnPorch      ScreenPorch
##           0           0           0           0           0
##      PoolArea      PoolQC      Fence      MiscFeature      MiscVal
##           0           0           0           0           0
##      MoSold      YrSold      SaleType      SaleCondition      SalePrice
##           0           0           0           0           1459
##      isTrain
##           0
```

```
df$LotFrontage[which(is.na(df$LotFrontage))] <- mean(df$LotFrontage,na.rm = T)
```

```
df$MasVnrArea[which(is.na(df$MasVnrArea))] <- mean(df$LotFrontage,na.rm = T)
```

```
table(df$GarageYrBlt)
```

```
##
## 1895 1896 1900 1906 1908 1910 1914 1915 1916 1917 1918 1919 1920 1921 1922
##    1    1    6    1    1   10    2    7    6    2    3    1   33    5    8
## 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937
##    6    8   15   15    5    7    2   27    4    4    1    4    8    7    6
## 1938 1939 1940 1941 1942 1943 1945 1946 1947 1948 1949 1950 1951 1952 1953
##   11   21   25   14    6    1   10    9    5   19   14   51   17   16   23
## 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968
##   37   24   41   34   42   36   37   31   35   34   35   34   39   36   48
## 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983
##   32   32   24   27   29   35   28   50   66   41   35   32   15    9   11
## 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998
##   19   18   12   18   20   19   26   17   27   49   39   35   40   44   58
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2207
##   54   55   41   53   92   99  142  115  115   61   29    5    1
```

```
df$GarageYrBlt[which(is.na(df$GarageYrBlt))] <- 0
```

```
table(df$GarageYrBlt)
```

```
##
##    0 1895 1896 1900 1906 1908 1910 1914 1915 1916 1917 1918 1919 1920 1921
## 159    1    1    6    1    1   10    2    7    6    2    3    1   33    5
## 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936
##    8    6    8   15   15    5    7    2   27    4    4    1    4    8    7
## 1937 1938 1939 1940 1941 1942 1943 1945 1946 1947 1948 1949 1950 1951 1952
##    6   11   21   25   14    6    1   10    9    5   19   14   51   17   16
## 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967
##   23   37   24   41   34   42   36   37   31   35   34   35   34   39   36
## 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982
##   48   32   32   24   27   29   35   28   50   66   41   35   32   15    9
## 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
```

```
##      11      19      18      12      18      20      19      26      17      27      49      39      35      40      44
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2207
##      58      54      55      41      53      92      99     142     115     115      61      29       5       1
```

```
x = c("BsmtFinSF1", "BsmtFinSF2", "BsmtUnfSF", "TotalBsmtSF", "BsmtFullBath", "BsmtHalfBath", "GarageYrBlt")
```

```
for(i in x){
```

```
  df[,i][which(is.na(df[,i]))] <- 0
}
```

```
colSums(sapply(df, is.na))
```

```
##      MSSubClass      MSZoning  LotFrontage      LotArea      Street
##           0           0           0           0           0
##      Alley      LotShape  LandContour  Utilities      LotConfig
##           0           0           0           0           0
##      LandSlope  Neighborhood  Condition1  Condition2      BldgType
##           0           0           0           0           0
##      HouseStyle  OverallQual  OverallCond  YearBuilt  YearRemodAdd
##           0           0           0           0           0
##      RoofStyle      RoofMatl  Exterior1st  Exterior2nd  MasVnrType
##           0           0           0           0           0
##      MasVnrArea  ExterQual  ExterCond  Foundation  BsmtQual
##           0           0           0           0           0
##      BsmtCond  BsmtExposure  BsmtFinType1  BsmtFinSF1  BsmtFinType2
##           0           0           0           0           0
##      BsmtFinSF2  BsmtUnfSF  TotalBsmtSF  Heating  HeatingQC
##           0           0           0           0           0
##      CentralAir  Electrical  X1stFlrSF  X2ndFlrSF  LowQualFinSF
##           0           0           0           0           0
##      GrLivArea  BsmtFullBath  BsmtHalfBath  FullBath  HalfBath
##           0           0           0           0           0
##      BedroomAbvGr  KitchenAbvGr  KitchenQual  TotRmsAbvGrd  Functional
##           0           0           0           0           0
##      Fireplaces  FireplaceQu  GarageType  GarageYrBlt  GarageFinish
##           0           0           0           0           0
##      GarageCars  GarageArea  GarageQual  GarageCond  PavedDrive
##           0           0           0           0           0
##      WoodDeckSF  OpenPorchSF  EnclosedPorch  X3SsnPorch  ScreenPorch
##           0           0           0           0           0
##      PoolArea      PoolQC      Fence  MiscFeature  MiscVal
##           0           0           0           0           0
##      MoSold      YrSold      SaleType  SaleCondition  SalePrice
##           0           0           0           0           1459
##      isTrain
##           0
```