setwd("C:/Viraj/Bridgeport/FALL-18-Project/Kaggle-competitiom/Predicting-house-prices")

dataset\_train <- read.csv("train.csv")

dataset\_test <- read.csv("test.csv")

dataset\_train$Alley <- NULL

dataset\_train$MiscFeature <- NULL

dataset\_train$PoolQC <- NULL

dataset\_train$FireplaceQu<- NULL

dataset\_train$Fence <- NULL

dataset\_train$LotFrontage[is.na(dataset\_train$LotFrontage)]<- 0

dataset\_train$MasVnrType[is.na(dataset\_train$MasVnrType)]<- "None"

# Levels setting - BsmtQual

levels <- levels(dataset\_train$BsmtQual)

levels[length(levels) + 1] <- "UN"

dataset\_train$BsmtQual <- factor(dataset\_train$BsmtQual, levels = levels)

# Levels setting - BsmtCond

levels <- levels(dataset\_train$BsmtCond)

levels[length(levels) + 1] <- "UN"

dataset\_train$BsmtCond <- factor(dataset\_train$BsmtCond, levels = levels)

# Levels setting - BsmtExposure

levels <- levels(dataset\_train$BsmtExposure)

levels[length(levels) + 1] <- "UN"

dataset\_train$BsmtExposure <- factor(dataset\_train$BsmtExposure, levels = levels)

# Levels setting - BsmtFinType1

levels <- levels(dataset\_train$BsmtFinType1)

levels[length(levels) + 1] <- "UN"

dataset\_train$BsmtFinType1 <- factor(dataset\_train$BsmtFinType1, levels = levels)

# Levels setting - BsmtFinType1

levels <- levels(dataset\_train$BsmtFinType2)

levels[length(levels) + 1] <- "UN"

dataset\_train$BsmtFinType2 <- factor(dataset\_train$BsmtFinType2, levels = levels)

dataset\_train$BsmtQual[is.na(dataset\_train$BsmtQual)] <- "UN"

dataset\_train$BsmtCond[is.na(dataset\_train$BsmtCond)] <- "UN"

dataset\_train$BsmtExposure[is.na(dataset\_train$BsmtExposure)] <- "UN"

dataset\_train$BsmtFinType1[is.na(dataset\_train$BsmtFinType1)] <- "UN"

dataset\_train$BsmtFinType2[is.na(dataset\_train$BsmtFinType2)] <- "UN"

# Levels setting - GarageType

levels <- levels(dataset\_train$GarageType)

levels[length(levels) + 1] <- "UN"

dataset\_train$GarageType <- factor(dataset\_train$GarageType, levels = levels)

# Levels setting - GarageFinish

levels <- levels(dataset\_train$GarageFinish)

levels[length(levels) + 1] <- "UN"

dataset\_train$GarageFinish <- factor(dataset\_train$GarageFinish, levels = levels)

# Levels setting - GarageQuality

levels <- levels(dataset\_train$GarageQual)

levels[length(levels) + 1] <- "UN"

dataset\_train$GarageQual <- factor(dataset\_train$GarageQual, levels = levels)

# Levels setting - Garagecondition

levels <- levels(dataset\_train$GarageCond)

levels[length(levels) + 1] <- "UN"

dataset\_train$GarageCond <- factor(dataset\_train$GarageCond, levels = levels)

# Levels setting - Electrical

levels <- levels(dataset\_train$Electrical)

levels[length(levels) + 1] <- "UN"

dataset\_train$Electrical <- factor(dataset\_train$Electrical, levels = levels)

dataset\_train$GarageType[is.na(dataset\_train$GarageType)] <- "UN"

dataset\_train$GarageYrBlt[is.na(dataset\_train$GarageYrBlt)] <- 0000

dataset\_train$GarageFinish[is.na(dataset\_train$GarageFinish)] <- "UN"

dataset\_train$GarageQual[is.na(dataset\_train$GarageQual)] <- "UN"

dataset\_train$GarageCond[is.na(dataset\_train$GarageCond)] <- "UN"

dataset\_train$MasVnrArea[is.na(dataset\_train$MasVnrArea)] <- 0

dataset\_train$Electrical[is.na(dataset\_train$Electrical)] <- "UN"

dataset\_train$SalePrice <- log10(dataset\_train$SalePrice)

colSums(is.na(dataset\_train))

View(dataset\_train)

############################# Test #########################################

dataset\_test$Alley <- NULL

dataset\_test$MiscFeature <- NULL

dataset\_test$PoolQC <- NULL

dataset\_test$FireplaceQu<- NULL

dataset\_test$Fence <- NULL

dataset\_test$LotFrontage[is.na(dataset\_test$LotFrontage)]<- 0

dataset\_test$MasVnrType[is.na(dataset\_test$MasVnrType)]<- "None"

dataset\_test$MasVnrArea[is.na(dataset\_test$MasVnrArea)]<- 0

# Levels setting - BsmtQual

levels <- levels(dataset\_test$BsmtQual)

levels[length(levels) + 1] <- "UN"

dataset\_test$BsmtQual <- factor(dataset\_test$BsmtQual, levels = levels)

# Levels setting - BsmtCond

levels <- levels(dataset\_test$BsmtCond)

levels[length(levels) + 1] <- "UN"

dataset\_test$BsmtCond <- factor(dataset\_test$BsmtCond, levels = levels)

# Levels setting - BsmtExposure

levels <- levels(dataset\_test$BsmtExposure)

levels[length(levels) + 1] <- "UN"

dataset\_test$BsmtExposure <- factor(dataset\_test$BsmtExposure, levels = levels)

# Levels setting - BsmtFinType1

levels <- levels(dataset\_test$BsmtFinType1)

levels[length(levels) + 1] <- "UN"

dataset\_test$BsmtFinType1 <- factor(dataset\_test$BsmtFinType1, levels = levels)

# Levels setting - BsmtFinType1

levels <- levels(dataset\_test$BsmtFinType2)

levels[length(levels) + 1] <- "UN"

dataset\_test$BsmtFinType2 <- factor(dataset\_test$BsmtFinType2, levels = levels)

dataset\_test$BsmtQual[is.na(dataset\_test$BsmtQual)] <- "UN"

dataset\_test$BsmtCond[is.na(dataset\_test$BsmtCond)] <- "UN"

dataset\_test$BsmtExposure[is.na(dataset\_test$BsmtExposure)] <- "UN"

dataset\_test$BsmtFinType1[is.na(dataset\_test$BsmtFinType1)] <- "UN"

dataset\_test$BsmtFinType2[is.na(dataset\_test$BsmtFinType2)] <- "UN"

# Levels setting - GarageType

levels <- levels(dataset\_test$GarageType)

levels[length(levels) + 1] <- "UN"

dataset\_test$GarageType <- factor(dataset\_test$GarageType, levels = levels)

# Levels setting - GarageFinish

levels <- levels(dataset\_test$GarageFinish)

levels[length(levels) + 1] <- "UN"

dataset\_test$GarageFinish <- factor(dataset\_test$GarageFinish, levels = levels)

# Levels setting - GarageQuality

levels <- levels(dataset\_test$GarageQual)

levels[length(levels) + 1] <- "UN"

dataset\_test$GarageQual <- factor(dataset\_test$GarageQual, levels = levels)

# Levels setting - Garagecondition

levels <- levels(dataset\_test$GarageCond)

levels[length(levels) + 1] <- "UN"

dataset\_test$GarageCond <- factor(dataset\_test$GarageCond, levels = levels)

# Levels setting - MsZoning

levels <- levels(dataset\_test$MSZoning)

levels[length(levels) + 1] <- "UN"

dataset\_test$MSZoning <- factor(dataset\_test$MSZoning, levels = levels)

# Levels setting - Utilites

levels <- levels(dataset\_test$Utilities)

levels[length(levels) + 1] <- "UN"

dataset\_test$Utilities <- factor(dataset\_test$Utilities, levels = levels)

# Levels setting - Exterior1st

levels <- levels(dataset\_test$Exterior1st)

levels[length(levels) + 1] <- "UN"

dataset\_test$Exterior1st <- factor(dataset\_test$Exterior1st, levels = levels)

# Levels setting - Exterior1st

levels <- levels(dataset\_test$Exterior2nd)

levels[length(levels) + 1] <- "UN"

dataset\_test$Exterior2nd <- factor(dataset\_test$Exterior2nd, levels = levels)

# Levels setting - KitchenQual

levels <- levels(dataset\_test$KitchenQual)

levels[length(levels) + 1] <- "UN"

dataset\_test$KitchenQual <- factor(dataset\_test$KitchenQual, levels = levels)

# Levels setting - Functional

levels <- levels(dataset\_test$Functional)

levels[length(levels) + 1] <- "UN"

dataset\_test$Functional <- factor(dataset\_test$Functional, levels = levels)

# Levels setting - SaleType

levels <- levels(dataset\_test$SaleType)

levels[length(levels) + 1] <- "UN"

dataset\_test$SaleType <- factor(dataset\_test$SaleType, levels = levels)

dataset\_test$GarageType[is.na(dataset\_test$GarageType)] <- "UN"

dataset\_test$GarageYrBlt[is.na(dataset\_test$GarageYrBlt)] <- 0000

dataset\_test$GarageFinish[is.na(dataset\_test$GarageFinish)] <- "UN"

dataset\_test$GarageQual[is.na(dataset\_test$GarageQual)] <- "UN"

dataset\_test$GarageCond[is.na(dataset\_test$GarageCond)] <- "UN"

dataset\_test$MSZoning[is.na(dataset\_test$MSZoning)] <- "UN"

dataset\_test$Utilities[is.na(dataset\_test$Utilities)] <- "UN"

dataset\_test$Exterior1st[is.na(dataset\_test$Exterior1st)] <- "UN"

dataset\_test$Exterior2nd[is.na(dataset\_test$Exterior2nd)] <- "UN"

dataset\_test$BsmtFinSF1[is.na(dataset\_test$BsmtFinSF1)] <- 0

dataset\_test$BsmtFinSF2[is.na(dataset\_test$BsmtFinSF2)] <- 0

dataset\_test$BsmtUnfSF[is.na(dataset\_test$BsmtUnfSF)] <- 0

dataset\_test$TotalBsmtSF[is.na(dataset\_test$TotalBsmtSF)] <- 0

dataset\_test$BsmtFullBath[is.na(dataset\_test$BsmtFullBath)] <- 0

dataset\_test$BsmtHalfBath[is.na(dataset\_test$BsmtHalfBath)] <- 0

dataset\_test$KitchenQual[is.na(dataset\_test$KitchenQual)] <- "UN"

dataset\_test$Functional[is.na(dataset\_test$Functional)] <- "UN"

dataset\_test$GarageCars[is.na(dataset\_test$GarageCars)] <- 0

dataset\_test$GarageArea[is.na(dataset\_test$GarageArea)] <- 0

dataset\_test$SaleType[is.na(dataset\_test$SaleType)] <- "UN"

colSums(is.na(dataset\_test))

View(dataset\_test)