

# Project\_01

Group\_1

March 21, 2017

```
#include = FALSE, echo = FALSE
loadData = function(path){
  data = read.csv(path)
  data = data[-1]

  data$MSSubClass = as.factor(data$MSSubClass);#Factor encoded as a numeric value

  #Median or 0
  data$LotFrontage[is.na(data$LotFrontage) == TRUE] = median(data$LotFrontage[!is.na(data$LotFrontage)])
  #data$LotFrontage[is.na(data$LotFrontage) == TRUE] = median(data$LotFrontage[!is.na(data$LotFrontage)])

  data$Alley = factor(data$Alley, levels=c(NA, "None", "Grvl", "Pave"))
  data$Alley[is.na(data$Alley)] = "None"

  data$MasVnrType[is.na(data$MasVnrType) == TRUE] = "None"

  data$MasVnrArea[is.na(data$MasVnrArea) == TRUE] = "None"

  data$BsmtQual = factor(data$BsmtQual, levels=c("None", levels(data$BsmtQual)))
  data$BsmtQual[is.na(data$BsmtQual)] = "None"

  data$BsmtCond = factor(data$BsmtCond, levels=c("None", levels(data$BsmtCond)))
  data$BsmtCond[is.na(data$BsmtCond)] = "None"

  data$BsmtExposure = factor(data$BsmtExposure, levels=c("None", levels(data$BsmtExposure)))
  data$BsmtExposure[is.na(data$BsmtExposure)] = "None"

  data$BsmtFinType1 = factor(data$BsmtFinType1, levels=c("None", levels(data$BsmtFinType1)))
  data$BsmtFinType1[is.na(data$BsmtFinType1)] = "None"

  data$BsmtFinType2 = factor(data$BsmtFinType2, levels=c("None", levels(data$BsmtFinType2)))
  data$BsmtFinType2[is.na(data$BsmtFinType2)] = "None"

  data$Electrical[is.na(data$Electrical) == TRUE] = "SBrkr"

  data$FireplaceQu = factor(data$FireplaceQu, levels=c("None", levels(data$FireplaceQu)))
  data$FireplaceQu[is.na(data$FireplaceQu)] = "None"

  data$GarageType = factor(data$GarageType, levels=c("None", levels(data$GarageType)))
  data$GarageType[is.na(data$GarageType)] = "None"

  data$GarageYrBlt[is.na(data$GarageYrBlt) == TRUE] = 0

  data$GarageFinish = factor(data$GarageFinish, levels=c("None", levels(data$GarageFinish)))
  data$GarageFinish[is.na(data$GarageFinish)] = "None"

  data$GarageQual = factor(data$GarageQual, levels=c("None", levels(data$GarageQual)))
```

```

data$GarageQual[is.na(data$GarageQual)] = "None"

data$GarageCond = factor(data$GarageCond, levels=c("None", levels(data$GarageCond)))
data$GarageCond[is.na(data$GarageCond)] = "None"

data$PoolQC = factor(data$PoolQC, levels=c("None", levels(data$PoolQC)))
data$PoolQC[is.na(data$PoolQC)] = "None"

data$Fence = factor(data$Fence, levels=c("None", levels(data$Fence)))
data$Fence[is.na(data$Fence)] = "None"

data$MiscFeature = factor(data$MiscFeature, levels=c("None", levels(data$MiscFeature)))
data$MiscFeature[is.na(data$MiscFeature)] = "None"

return (data)
}

data = loadData("./train.csv")
summary(data)

```

```

##      MSSubClass      MSZoning      LotFrontage      LotArea
## 20      :536      C (all): 10      Min.      : 21.00      Min.      : 1300
## 60      :299      FV      : 65      1st Qu.: 60.00      1st Qu.: 7554
## 50      :144      RH      : 16      Median : 69.00      Median : 9478
## 120     : 87      RL      :1151      Mean   : 69.86      Mean   : 10517
## 30      : 69      RM      : 218      3rd Qu.: 79.00      3rd Qu.: 11602
## 160     : 63                      Max.   :313.00      Max.   :215245
## (Other):262
##      Street      Alley      LotShape      LandContour      Utilities
## Grv1: 6      None:1369      IR1:484      Bnk: 63      AllPub:1459
## Pave:1454      Grv1: 50      IR2: 41      HLS: 50      NoSeWa: 1
##                      Pave: 41      IR3: 10      Low: 36
##                      Reg:925      Lvl:1311
##
##
##
##      LotConfig      LandSlope      Neighborhood      Condition1      Condition2
## Corner : 263      Gtl:1382      NNames :225      Norm :1260      Norm :1445
## CulDSac: 94      Mod: 65      CollgCr:150      Feedr : 81      Feedr : 6
## FR2 : 47      Sev: 13      OldTown:113      Artery : 48      Artery : 2
## FR3 : 4                      Edwards:100      RRAn : 26      PosN : 2
## Inside :1052      Somerst: 86      PosN : 19      RRNn : 2
##                      Gilbert: 79      RRAe : 11      PosA : 1
##                      (Other):707      (Other): 15      (Other): 2
##
##      BldgType      HouseStyle      OverallQual      OverallCond
## 1Fam :1220      1Story :726      Min. : 1.000      Min. :1.000
## 2fmCon: 31      2Story :445      1st Qu.: 5.000      1st Qu.:5.000
## Duplex: 52      1.5Fin :154      Median : 6.000      Median :5.000
## Twnhs : 43      SLvl : 65      Mean : 6.099      Mean :5.575
## TwnhsE: 114      SFoyer : 37      3rd Qu.: 7.000      3rd Qu.:6.000
##                      1.5Unf : 14      Max. :10.000      Max. :9.000
##                      (Other): 19
##
##      YearBuilt      YearRemodAdd      RoofStyle      RoofMatl      Exterior1st

```

```

## Min. :1872 Min. :1950 Flat : 13 CompShg:1434 VinylSd:515
## 1st Qu.:1954 1st Qu.:1967 Gable :1141 Tar&Grv: 11 HdBoard:222
## Median :1973 Median :1994 Gambrel: 11 WdShngl: 6 MetalSd:220
## Mean :1971 Mean :1985 Hip : 286 WdShake: 5 Wd Sdng:206
## 3rd Qu.:2000 3rd Qu.:2004 Mansard: 7 ClyTile: 1 Plywood:108
## Max. :2010 Max. :2010 Shed : 2 Membran: 1 CemntBd: 61
## (Other): 2 (Other):128
## Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond
## VinylSd:504 BrkCmn : 15 Length:1460 Ex: 52 Ex: 3
## MetalSd:214 BrkFace:445 Class :character Fa: 14 Fa: 28
## HdBoard:207 None :872 Mode :character Gd:488 Gd: 146
## Wd Sdng:197 Stone :128 TA:906 Po: 1
## Plywood:142 TA:1282
## CmentBd: 60
## (Other):136
## Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1
## BrkTil:146 None: 37 None: 37 None: 38 None: 37
## CBlock:634 Ex :121 Fa : 45 Av :221 ALQ :220
## PConc :647 Fa : 35 Gd : 65 Gd :134 BLQ :148
## Slab : 24 Gd :618 Po : 2 Mn :114 GLQ :418
## Stone : 6 TA :649 TA :1311 No :953 LwQ : 74
## Wood : 3 Rec :133
## Unf :430
## BsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF
## Min. : 0.0 None: 38 Min. : 0.00 Min. : 0.0
## 1st Qu.: 0.0 ALQ : 19 1st Qu.: 0.00 1st Qu.: 223.0
## Median : 383.5 BLQ : 33 Median : 0.00 Median : 477.5
## Mean : 443.6 GLQ : 14 Mean : 46.55 Mean : 567.2
## 3rd Qu.: 712.2 LwQ : 46 3rd Qu.: 0.00 3rd Qu.: 808.0
## Max. :5644.0 Rec : 54 Max. :1474.00 Max. :2336.0
## Unf :1256
## TotalBsmtSF Heating HeatingQC CentralAir Electrical
## Min. : 0.0 Floor: 1 Ex:741 N: 95 FuseA: 94
## 1st Qu.: 795.8 GasA :1428 Fa: 49 Y:1365 FuseF: 27
## Median : 991.5 GasW : 18 Gd:241 FuseP: 3
## Mean :1057.4 Grav : 7 Po: 1 Mix : 1
## 3rd Qu.:1298.2 OthW : 2 TA:428 SBrkr:1335
## Max. :6110.0 Wall : 4
## X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea
## Min. : 334 Min. : 0 Min. : 0.000 Min. : 334
## 1st Qu.: 882 1st Qu.: 0 1st Qu.: 0.000 1st Qu.:1130
## Median :1087 Median : 0 Median : 0.000 Median :1464
## Mean :1163 Mean : 347 Mean : 5.845 Mean :1515
## 3rd Qu.:1391 3rd Qu.: 728 3rd Qu.: 0.000 3rd Qu.:1777
## Max. :4692 Max. :2065 Max. :572.000 Max. :5642
## BsmtFullBath BsmtHalfBath FullBath HalfBath
## Min. :0.0000 Min. :0.00000 Min. :0.000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:1.000 1st Qu.:0.0000
## Median :0.0000 Median :0.00000 Median :2.000 Median :0.0000
## Mean :0.4253 Mean :0.05753 Mean :1.565 Mean :0.3829
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:2.000 3rd Qu.:1.0000
## Max. :3.0000 Max. :2.00000 Max. :3.000 Max. :2.0000

```

```

##
## BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## Min. :0.000 Min. :0.000 Ex:100 Min. : 2.000 Maj1: 14
## 1st Qu.:2.000 1st Qu.:1.000 Fa: 39 1st Qu.: 5.000 Maj2: 5
## Median :3.000 Median :1.000 Gd:586 Median : 6.000 Min1: 31
## Mean :2.866 Mean :1.047 TA:735 Mean : 6.518 Min2: 34
## 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.: 7.000 Mod : 15
## Max. :8.000 Max. :3.000 Max. :14.000 Sev : 1
## Typ :1360
## Fireplaces FireplaceQu GarageType GarageYrBltd GarageFinish
## Min. :0.000 None:690 None : 81 Min. : 0 None: 81
## 1st Qu.:0.000 Ex : 24 2Types : 6 1st Qu.:1958 Fin :352
## Median :1.000 Fa : 33 Attchd :870 Median :1977 RFn :422
## Mean :0.613 Gd :380 Basment: 19 Mean :1869 Unf :605
## 3rd Qu.:1.000 Po : 20 BuiltIn: 88 3rd Qu.:2001
## Max. :3.000 TA :313 CarPort: 9 Max. :2010
## Detchd :387
## GarageCars GarageArea GarageQual GarageCond PavedDrive
## Min. :0.000 Min. : 0.0 None: 81 None: 81 N: 90
## 1st Qu.:1.000 1st Qu.: 334.5 Ex : 3 Ex : 2 P: 30
## Median :2.000 Median : 480.0 Fa : 48 Fa : 35 Y:1340
## Mean :1.767 Mean : 473.0 Gd : 14 Gd : 9
## 3rd Qu.:2.000 3rd Qu.: 576.0 Po : 3 Po : 7
## Max. :4.000 Max. :1418.0 TA :1311 TA :1326
##
## WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.00
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00
## Median : 0.00 Median : 25.00 Median : 0.00 Median : 0.00
## Mean : 94.24 Mean : 46.66 Mean : 21.95 Mean : 3.41
## 3rd Qu.:168.00 3rd Qu.: 68.00 3rd Qu.: 0.00 3rd Qu.: 0.00
## Max. :857.00 Max. :547.00 Max. :552.00 Max. :508.00
##
## ScreenPorch PoolArea PoolQC Fence MiscFeature
## Min. : 0.00 Min. : 0.000 None:1453 None :1179 None:1406
## 1st Qu.: 0.00 1st Qu.: 0.000 Ex : 2 GdPrv: 59 Gar2: 2
## Median : 0.00 Median : 0.000 Fa : 2 GdWo : 54 Othr: 2
## Mean : 15.06 Mean : 2.759 Gd : 3 MnPrv: 157 Shed: 49
## 3rd Qu.: 0.00 3rd Qu.: 0.000 MnWw : 11 TenC: 1
## Max. :480.00 Max. :738.000
##
## MiscVal MoSold YrSold SaleType
## Min. : 0.00 Min. : 1.000 Min. :2006 WD :1267
## 1st Qu.: 0.00 1st Qu.: 5.000 1st Qu.:2007 New : 122
## Median : 0.00 Median : 6.000 Median :2008 COD : 43
## Mean : 43.49 Mean : 6.322 Mean :2008 ConLD : 9
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009 ConLI : 5
## Max. :15500.00 Max. :12.000 Max. :2010 ConLw : 5
## (Other): 9
## SaleCondition SalePrice
## Abnorml: 101 Min. : 34900
## AdjLand: 4 1st Qu.:129975
## Alloca : 12 Median :163000
## Family : 20 Mean :180921

```

```
## Normal :1198    3rd Qu.:214000
## Partial: 125    Max.      :755000
##
```