

# Exploratory Analysis

*Raul Ossada and Lucas Hamaguchi*

*11 de julho de 2017*

## Análise Exploratória dos Dados

```
setwd("~/Documentos/00_IME_BE/KaggleStat/HousePrices/Data");  
training_data <- read.csv(file="train.csv");
```

```
summary(training_data);
```

```
##           Id           MSSubClass           MSZoning           LotFrontage  
## Min.      : 1.0      Min.      : 20.0      C (all): 10      Min.      : 21.00  
## 1st Qu.: 365.8      1st Qu.: 20.0      FV       : 65      1st Qu.: 59.00  
## Median : 730.5      Median : 50.0      RH       : 16      Median : 69.00  
## Mean    : 730.5      Mean    : 56.9      RL      :1151      Mean    : 70.05  
## 3rd Qu.:1095.2      3rd Qu.: 70.0      RM       : 218      3rd Qu.: 80.00  
## Max.    :1460.0      Max.    :190.0                      Max.    :313.00  
##                                           NA's     :259  
##           LotArea           Street           Alley           LotShape           LandContour  
## Min.      : 1300      Grvl: 6      Grvl: 50      IR1:484      Bnk: 63  
## 1st Qu.: 7554      Pave:1454      Pave: 41      IR2: 41      HLS: 50  
## Median : 9478                      NA's:1369      IR3: 10      Low: 36  
## Mean    : 10517                      Reg:925      Lvl:1311  
## 3rd Qu.: 11602  
## Max.    :215245  
##  
##           Utilities           LotConfig           LandSlope           Neighborhood           Condition1  
## AllPub:1459      Corner : 263      Gtl:1382      NNames :225      Norm :1260  
## NoSeWa: 1      CulDSac: 94      Mod: 65      CollgCr:150      Feedr : 81  
## FR2 : 47      Sev: 13      OldTown:113      Artery : 48  
## FR3 : 4      Edwards:100      RRAn : 26  
## Inside :1052      Somerst: 86      PosN : 19  
## Gilbert: 79      RRAe : 11  
## (Other):707      (Other): 15  
##  
##           Condition2           BldgType           HouseStyle           OverallQual  
## Norm :1445      1Fam :1220      1Story :726      Min. : 1.000  
## Feedr : 6      2fmCon: 31      2Story :445      1st Qu.: 5.000  
## Artery : 2      Duplex: 52      1.5Fin :154      Median : 6.000  
## PosN : 2      Twnhs : 43      SLvl : 65      Mean : 6.099  
## RRNn : 2      TwnhsE: 114      SFoyer : 37      3rd Qu.: 7.000  
## PosA : 1      1.5Unf : 14      Max. :10.000  
## (Other): 2      (Other): 19  
##  
##           OverallCond           YearBuilt           YearRemodAdd           RoofStyle  
## Min. :1.000      Min. :1872      Min. :1950      Flat : 13  
## 1st Qu.:5.000      1st Qu.:1954      1st Qu.:1967      Gable :1141  
## Median :5.000      Median :1973      Median :1994      Gambrel: 11  
## Mean :5.575      Mean :1971      Mean :1985      Hip : 286  
## 3rd Qu.:6.000      3rd Qu.:2000      3rd Qu.:2004      Mansard: 7  
## Max. :9.000      Max. :2010      Max. :2010      Shed : 2
```

```

##
##      RoofMatl      Exterior1st  Exterior2nd      MasVnrType      MasVnrArea
## CompShg:1434 VinylSd:515 VinylSd:504 BrkCmn : 15 Min. : 0.0
## Tar&Grv: 11 HdBoard:222 MetalSd:214 BrkFace:445 1st Qu.: 0.0
## WdShngl: 6 MetalSd:220 HdBoard:207 None :864 Median : 0.0
## WdShake: 5 Wd Sdng:206 Wd Sdng:197 Stone :128 Mean : 103.7
## ClyTile: 1 Plywood:108 Plywood:142 NA's : 8 3rd Qu.: 166.0
## Membran: 1 CemntBd: 61 CmentBd: 60 Max. :1600.0
## (Other): 2 (Other):128 (Other):136 NA's :8
## ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure
## Ex: 52 Ex: 3 BrkTil:146 Ex :121 Fa : 45 Av :221
## Fa: 14 Fa: 28 CBlock:634 Fa : 35 Gd : 65 Gd :134
## Gd:488 Gd: 146 PConc :647 Gd :618 Po : 2 Mn :114
## TA:906 Po: 1 Slab : 24 TA :649 TA :1311 No :953
## TA:1282 Stone : 6 NA's: 37 NA's: 37 NA's: 38
## Wood : 3
##
## BsmtFinType1 BsmtFinSF1 BsmtFinType2 BsmtFinSF2
## ALQ :220 Min. : 0.0 ALQ : 19 Min. : 0.00
## BLQ :148 1st Qu.: 0.0 BLQ : 33 1st Qu.: 0.00
## GLQ :418 Median : 383.5 GLQ : 14 Median : 0.00
## LwQ : 74 Mean : 443.6 LwQ : 46 Mean : 46.55
## Rec :133 3rd Qu.: 712.2 Rec : 54 3rd Qu.: 0.00
## Unf :430 Max. :5644.0 Unf :1256 Max. :1474.00
## NA's: 37 NA's: 38
## BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir
## Min. : 0.0 Min. : 0.0 Floor: 1 Ex:741 N: 95
## 1st Qu.: 223.0 1st Qu.: 795.8 GasA :1428 Fa: 49 Y:1365
## Median : 477.5 Median : 991.5 GasW : 18 Gd:241
## Mean : 567.2 Mean :1057.4 Grav : 7 Po: 1
## 3rd Qu.: 808.0 3rd Qu.:1298.2 OthW : 2 TA:428
## Max. :2336.0 Max. :6110.0 Wall : 4
##
## Electrical X1stFlrSF X2ndFlrSF LowQualFinSF
## FuseA: 94 Min. : 334 Min. : 0 Min. : 0.000
## FuseF: 27 1st Qu.: 882 1st Qu.: 0 1st Qu.: 0.000
## FuseP: 3 Median :1087 Median : 0 Median : 0.000
## Mix : 1 Mean :1163 Mean : 347 Mean : 5.845
## SBrkr:1334 3rd Qu.:1391 3rd Qu.: 728 3rd Qu.: 0.000
## NA's : 1 Max. :4692 Max. :2065 Max. :572.000
##
## GrLivArea BsmtFullBath BsmtHalfBath FullBath
## Min. : 334 Min. :0.0000 Min. :0.00000 Min. :0.000
## 1st Qu.:1130 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:1.000
## Median :1464 Median :0.0000 Median :0.00000 Median :2.000
## Mean :1515 Mean :0.4253 Mean :0.05753 Mean :1.565
## 3rd Qu.:1777 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:2.000
## Max. :5642 Max. :3.0000 Max. :2.00000 Max. :3.000
##
## HalfBath BedroomAbvGr KitchenAbvGr KitchenQual
## Min. :0.0000 Min. :0.000 Min. :0.000 Ex:100
## 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000 Fa: 39
## Median :0.0000 Median :3.000 Median :1.000 Gd:586
## Mean :0.3829 Mean :2.866 Mean :1.047 TA:735

```

```

## 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000
## Max. :2.0000 Max. :8.000 Max. :3.000
##
## TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType
## Min. : 2.000 Maj1: 14 Min. :0.000 Ex : 24 2Types : 6
## 1st Qu.: 5.000 Maj2: 5 1st Qu.:0.000 Fa : 33 Attchd :870
## Median : 6.000 Min1: 31 Median :1.000 Gd :380 Basment: 19
## Mean : 6.518 Min2: 34 Mean :0.613 Po : 20 BuiltIn: 88
## 3rd Qu.: 7.000 Mod : 15 3rd Qu.:1.000 TA :313 CarPort: 9
## Max. :14.000 Sev : 1 Max. :3.000 NA's:690 Detchd :387
## Typ :1360 NA's : 81
## GarageYrBlt GarageFinish GarageCars GarageArea GarageQual
## Min. :1900 Fin :352 Min. :0.000 Min. : 0.0 Ex : 3
## 1st Qu.:1961 RFn :422 1st Qu.:1.000 1st Qu.: 334.5 Fa : 48
## Median :1980 Unf :605 Median :2.000 Median : 480.0 Gd : 14
## Mean :1979 NA's: 81 Mean :1.767 Mean : 473.0 Po : 3
## 3rd Qu.:2002 3rd Qu.:2.000 3rd Qu.: 576.0 TA :1311
## Max. :2010 Max. :4.000 Max. :1418.0 NA's: 81
## NA's :81
## GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch
## Ex : 2 N: 90 Min. : 0.00 Min. : 0.00 Min. : 0.00
## Fa : 35 P: 30 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00
## Gd : 9 Y:1340 Median : 0.00 Median : 25.00 Median : 0.00
## Po : 7 Mean : 94.24 Mean : 46.66 Mean : 21.95
## TA :1326 3rd Qu.:168.00 3rd Qu.: 68.00 3rd Qu.: 0.00
## NA's: 81 Max. :857.00 Max. :547.00 Max. :552.00
##
## X3SsnPorch ScreenPorch PoolArea PoolQC
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Ex : 2
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000 Fa : 2
## Median : 0.00 Median : 0.00 Median : 0.000 Gd : 3
## Mean : 3.41 Mean : 15.06 Mean : 2.759 NA's:1453
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000
## Max. :508.00 Max. :480.00 Max. :738.000
##
## Fence MiscFeature MiscVal MoSold
## GdPrv: 59 Gar2: 2 Min. : 0.00 Min. : 1.000
## GdWo : 54 Othr: 2 1st Qu.: 0.00 1st Qu.: 5.000
## MnPrv: 157 Shed: 49 Median : 0.00 Median : 6.000
## MnWw : 11 TenC: 1 Mean : 43.49 Mean : 6.322
## NA's :1179 NA's:1406 3rd Qu.: 0.00 3rd Qu.: 8.000
## Max. :15500.00 Max. :12.000
##
## YrSold SaleType SaleCondition SalePrice
## Min. :2006 WD :1267 Abnorml: 101 Min. : 34900
## 1st Qu.:2007 New : 122 AdjLand: 4 1st Qu.:129975
## Median :2008 COD : 43 Alloca : 12 Median :163000
## Mean :2008 ConLD : 9 Family : 20 Mean :180921
## 3rd Qu.:2009 ConLI : 5 Normal :1198 3rd Qu.:214000
## Max. :2010 ConLw : 5 Partial: 125 Max. :755000
## (Other): 9

```