Above\_Median Project

Tabitha Hagen

2022-10-10

# Project for BAN 502 Predictive Analytics

# Read in the “ames\_student.csv” dataset. This dataset comes from a larger dataset which includes all residential home sales in Ames, Iowa between 2006 and 2010. The data set contains many explanatory variables on the quality and quantity of physical attributes of residential homes in Iowa sold between 2006 and 2010. Most of the variables describe information a typical home buyer would like to know about a property (square footage, number of bedrooms and bathrooms, size of lot, etc.). : <https://www.tmwr.org/ames.html> A good description of the dataset can be found here: <http://jse.amstat.org/v19n3/decock.pdf>

ames\_table <- read\_csv("ames\_student.csv")

## Rows: 2053 Columns: 81  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (47): MS\_SubClass, MS\_Zoning, Street, Alley, Lot\_Shape, Land\_Contour, Ut...  
## dbl (34): Lot\_Frontage, Lot\_Area, Year\_Built, Year\_Remod\_Add, Mas\_Vnr\_Area, ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

str(ames\_table)

## spec\_tbl\_df [2,053 × 81] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : chr [1:2053] "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" ...  
## $ MS\_Zoning : chr [1:2053] "Residential\_Low\_Density" "Residential\_High\_Density" "Residential\_Low\_Density" "Residential\_Low\_Density" ...  
## $ Lot\_Frontage : num [1:2053] 141 80 81 93 74 78 43 39 0 85 ...  
## $ Lot\_Area : num [1:2053] 31770 11622 14267 11160 13830 ...  
## $ Street : chr [1:2053] "Pave" "Pave" "Pave" "Pave" ...  
## $ Alley : chr [1:2053] "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" ...  
## $ Lot\_Shape : chr [1:2053] "Slightly\_Irregular" "Regular" "Slightly\_Irregular" "Regular" ...  
## $ Land\_Contour : chr [1:2053] "Lvl" "Lvl" "Lvl" "Lvl" ...  
## $ Utilities : chr [1:2053] "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ Lot\_Config : chr [1:2053] "Corner" "Inside" "Corner" "Corner" ...  
## $ Land\_Slope : chr [1:2053] "Gtl" "Gtl" "Gtl" "Gtl" ...  
## $ Neighborhood : chr [1:2053] "North\_Ames" "North\_Ames" "North\_Ames" "North\_Ames" ...  
## $ Condition\_1 : chr [1:2053] "Norm" "Feedr" "Norm" "Norm" ...  
## $ Condition\_2 : chr [1:2053] "Norm" "Norm" "Norm" "Norm" ...  
## $ Bldg\_Type : chr [1:2053] "OneFam" "OneFam" "OneFam" "OneFam" ...  
## $ House\_Style : chr [1:2053] "One\_Story" "One\_Story" "One\_Story" "One\_Story" ...  
## $ Overall\_Qual : chr [1:2053] "Above\_Average" "Average" "Above\_Average" "Good" ...  
## $ Overall\_Cond : chr [1:2053] "Average" "Above\_Average" "Above\_Average" "Average" ...  
## $ Year\_Built : num [1:2053] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add : num [1:2053] 1960 1961 1958 1968 1998 ...  
## $ Roof\_Style : chr [1:2053] "Hip" "Gable" "Hip" "Hip" ...  
## $ Roof\_Matl : chr [1:2053] "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior\_1st : chr [1:2053] "BrkFace" "VinylSd" "Wd Sdng" "BrkFace" ...  
## $ Exterior\_2nd : chr [1:2053] "Plywood" "VinylSd" "Wd Sdng" "BrkFace" ...  
## $ Mas\_Vnr\_Type : chr [1:2053] "Stone" "None" "BrkFace" "None" ...  
## $ Mas\_Vnr\_Area : num [1:2053] 112 0 108 0 0 20 0 0 0 0 ...  
## $ Exter\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Good" ...  
## $ Exter\_Cond : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Foundation : chr [1:2053] "CBlock" "CBlock" "CBlock" "CBlock" ...  
## $ Bsmt\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Cond : chr [1:2053] "Good" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Exposure : chr [1:2053] "Gd" "No" "No" "No" ...  
## $ BsmtFin\_Type\_1 : chr [1:2053] "BLQ" "Rec" "ALQ" "ALQ" ...  
## $ BsmtFin\_SF\_1 : num [1:2053] 2 6 1 1 3 3 1 3 1 3 ...  
## $ BsmtFin\_Type\_2 : chr [1:2053] "Unf" "LwQ" "Unf" "Unf" ...  
## $ BsmtFin\_SF\_2 : num [1:2053] 0 144 0 0 0 0 0 0 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:2053] 441 270 406 1045 137 ...  
## $ Total\_Bsmt\_SF : num [1:2053] 1080 882 1329 2110 928 ...  
## $ Heating : chr [1:2053] "GasA" "GasA" "GasA" "GasA" ...  
## $ Heating\_QC : chr [1:2053] "Fair" "Typical" "Typical" "Excellent" ...  
## $ Central\_Air : chr [1:2053] "Y" "Y" "Y" "Y" ...  
## $ Electrical : chr [1:2053] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...  
## $ First\_Flr\_SF : num [1:2053] 1656 896 1329 2110 928 ...  
## $ Second\_Flr\_SF : num [1:2053] 0 0 0 0 701 678 0 0 0 0 ...  
## $ Low\_Qual\_Fin\_SF : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:2053] 1656 896 1329 2110 1629 ...  
## $ Bsmt\_Full\_Bath : num [1:2053] 1 0 0 1 0 0 0 1 1 1 ...  
## $ Bsmt\_Half\_Bath : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:2053] 1 1 1 2 2 2 2 2 2 1 ...  
## $ Half\_Bath : num [1:2053] 0 0 1 1 1 1 0 0 0 1 ...  
## $ Bedroom\_AbvGr : num [1:2053] 3 2 3 3 3 3 2 2 3 2 ...  
## $ Kitchen\_AbvGr : num [1:2053] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : chr [1:2053] "Typical" "Typical" "Good" "Excellent" ...  
## $ TotRms\_AbvGrd : num [1:2053] 7 5 6 8 6 7 5 5 6 5 ...  
## $ Functional : chr [1:2053] "Typ" "Typ" "Typ" "Typ" ...  
## $ Fireplaces : num [1:2053] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Fireplace\_Qu : chr [1:2053] "Good" "No\_Fireplace" "No\_Fireplace" "Typical" ...  
## $ Garage\_Type : chr [1:2053] "Attchd" "Attchd" "Attchd" "Attchd" ...  
## $ Garage\_Finish : chr [1:2053] "Fin" "Unf" "Unf" "Fin" ...  
## $ Garage\_Cars : num [1:2053] 2 1 1 2 2 2 2 2 2 2 ...  
## $ Garage\_Area : num [1:2053] 528 730 312 522 482 470 506 608 420 506 ...  
## $ Garage\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Garage\_Cond : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Paved\_Drive : chr [1:2053] "Partial\_Pavement" "Paved" "Paved" "Paved" ...  
## $ Wood\_Deck\_SF : num [1:2053] 210 140 393 0 212 360 0 237 483 192 ...  
## $ Open\_Porch\_SF : num [1:2053] 62 0 36 0 34 36 82 152 21 0 ...  
## $ Enclosed\_Porch : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:2053] 0 120 0 0 0 0 144 0 0 0 ...  
## $ Pool\_Area : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : chr [1:2053] "No\_Pool" "No\_Pool" "No\_Pool" "No\_Pool" ...  
## $ Fence : chr [1:2053] "No\_Fence" "Minimum\_Privacy" "No\_Fence" "No\_Fence" ...  
## $ Misc\_Feature : chr [1:2053] "None" "None" "Gar2" "None" ...  
## $ Misc\_Val : num [1:2053] 0 0 12500 0 0 0 0 0 500 0 ...  
## $ Mo\_Sold : num [1:2053] 5 6 6 4 3 6 1 3 3 2 ...  
## $ Year\_Sold : num [1:2053] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : chr [1:2053] "WD" "WD" "WD" "WD" ...  
## $ Sale\_Condition : chr [1:2053] "Normal" "Normal" "Normal" "Normal" ...  
## $ Longitude : num [1:2053] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:2053] 42.1 42.1 42.1 42.1 42.1 ...  
## $ Above\_Median : chr [1:2053] "Yes" "No" "Yes" "Yes" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. MS\_SubClass = col\_character(),  
## .. MS\_Zoning = col\_character(),  
## .. Lot\_Frontage = col\_double(),  
## .. Lot\_Area = col\_double(),  
## .. Street = col\_character(),  
## .. Alley = col\_character(),  
## .. Lot\_Shape = col\_character(),  
## .. Land\_Contour = col\_character(),  
## .. Utilities = col\_character(),  
## .. Lot\_Config = col\_character(),  
## .. Land\_Slope = col\_character(),  
## .. Neighborhood = col\_character(),  
## .. Condition\_1 = col\_character(),  
## .. Condition\_2 = col\_character(),  
## .. Bldg\_Type = col\_character(),  
## .. House\_Style = col\_character(),  
## .. Overall\_Qual = col\_character(),  
## .. Overall\_Cond = col\_character(),  
## .. Year\_Built = col\_double(),  
## .. Year\_Remod\_Add = col\_double(),  
## .. Roof\_Style = col\_character(),  
## .. Roof\_Matl = col\_character(),  
## .. Exterior\_1st = col\_character(),  
## .. Exterior\_2nd = col\_character(),  
## .. Mas\_Vnr\_Type = col\_character(),  
## .. Mas\_Vnr\_Area = col\_double(),  
## .. Exter\_Qual = col\_character(),  
## .. Exter\_Cond = col\_character(),  
## .. Foundation = col\_character(),  
## .. Bsmt\_Qual = col\_character(),  
## .. Bsmt\_Cond = col\_character(),  
## .. Bsmt\_Exposure = col\_character(),  
## .. BsmtFin\_Type\_1 = col\_character(),  
## .. BsmtFin\_SF\_1 = col\_double(),  
## .. BsmtFin\_Type\_2 = col\_character(),  
## .. BsmtFin\_SF\_2 = col\_double(),  
## .. Bsmt\_Unf\_SF = col\_double(),  
## .. Total\_Bsmt\_SF = col\_double(),  
## .. Heating = col\_character(),  
## .. Heating\_QC = col\_character(),  
## .. Central\_Air = col\_character(),  
## .. Electrical = col\_character(),  
## .. First\_Flr\_SF = col\_double(),  
## .. Second\_Flr\_SF = col\_double(),  
## .. Low\_Qual\_Fin\_SF = col\_double(),  
## .. Gr\_Liv\_Area = col\_double(),  
## .. Bsmt\_Full\_Bath = col\_double(),  
## .. Bsmt\_Half\_Bath = col\_double(),  
## .. Full\_Bath = col\_double(),  
## .. Half\_Bath = col\_double(),  
## .. Bedroom\_AbvGr = col\_double(),  
## .. Kitchen\_AbvGr = col\_double(),  
## .. Kitchen\_Qual = col\_character(),  
## .. TotRms\_AbvGrd = col\_double(),  
## .. Functional = col\_character(),  
## .. Fireplaces = col\_double(),  
## .. Fireplace\_Qu = col\_character(),  
## .. Garage\_Type = col\_character(),  
## .. Garage\_Finish = col\_character(),  
## .. Garage\_Cars = col\_double(),  
## .. Garage\_Area = col\_double(),  
## .. Garage\_Qual = col\_character(),  
## .. Garage\_Cond = col\_character(),  
## .. Paved\_Drive = col\_character(),  
## .. Wood\_Deck\_SF = col\_double(),  
## .. Open\_Porch\_SF = col\_double(),  
## .. Enclosed\_Porch = col\_double(),  
## .. Three\_season\_porch = col\_double(),  
## .. Screen\_Porch = col\_double(),  
## .. Pool\_Area = col\_double(),  
## .. Pool\_QC = col\_character(),  
## .. Fence = col\_character(),  
## .. Misc\_Feature = col\_character(),  
## .. Misc\_Val = col\_double(),  
## .. Mo\_Sold = col\_double(),  
## .. Year\_Sold = col\_double(),  
## .. Sale\_Type = col\_character(),  
## .. Sale\_Condition = col\_character(),  
## .. Longitude = col\_double(),  
## .. Latitude = col\_double(),  
## .. Above\_Median = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

summary(ames\_table)

## MS\_SubClass MS\_Zoning Lot\_Frontage Lot\_Area   
## Length:2053 Length:2053 Min. : 0.00 Min. : 1300   
## Class :character Class :character 1st Qu.: 43.00 1st Qu.: 7500   
## Mode :character Mode :character Median : 62.00 Median : 9548   
## Mean : 57.38 Mean : 10258   
## 3rd Qu.: 78.00 3rd Qu.: 11600   
## Max. :313.00 Max. :215245   
## Street Alley Lot\_Shape Land\_Contour   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Utilities Lot\_Config Land\_Slope Neighborhood   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Condition\_1 Condition\_2 Bldg\_Type House\_Style   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Overall\_Qual Overall\_Cond Year\_Built Year\_Remod\_Add  
## Length:2053 Length:2053 Min. :1875 Min. :1950   
## Class :character Class :character 1st Qu.:1953 1st Qu.:1965   
## Mode :character Mode :character Median :1972 Median :1993   
## Mean :1971 Mean :1984   
## 3rd Qu.:2000 3rd Qu.:2004   
## Max. :2010 Max. :2010   
## Roof\_Style Roof\_Matl Exterior\_1st Exterior\_2nd   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual Exter\_Cond   
## Length:2053 Min. : 0.0 Length:2053 Length:2053   
## Class :character 1st Qu.: 0.0 Class :character Class :character   
## Mode :character Median : 0.0 Mode :character Mode :character   
## Mean : 103.8   
## 3rd Qu.: 164.0   
## Max. :1600.0   
## Foundation Bsmt\_Qual Bsmt\_Cond Bsmt\_Exposure   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2 BsmtFin\_SF\_2   
## Length:2053 Min. :1.00 Length:2053 Min. : 0.00   
## Class :character 1st Qu.:3.00 Class :character 1st Qu.: 0.00   
## Mode :character Median :3.00 Mode :character Median : 0.00   
## Mean :4.21 Mean : 52.57   
## 3rd Qu.:7.00 3rd Qu.: 0.00   
## Max. :7.00 Max. :1526.00   
## Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating Heating\_QC   
## Min. : 0.0 Min. : 0 Length:2053 Length:2053   
## 1st Qu.: 226.0 1st Qu.: 793 Class :character Class :character   
## Median : 460.0 Median : 988 Mode :character Mode :character   
## Mean : 561.2 Mean :1055   
## 3rd Qu.: 801.0 3rd Qu.:1304   
## Max. :2336.0 Max. :5095   
## Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## Length:2053 Length:2053 Min. : 432 Min. : 0.0   
## Class :character Class :character 1st Qu.: 882 1st Qu.: 0.0   
## Mode :character Mode :character Median :1088 Median : 0.0   
## Mean :1168 Mean : 326.1   
## 3rd Qu.:1402 3rd Qu.: 701.0   
## Max. :5095 Max. :1862.0   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## Min. : 0.000 Min. : 480 Min. :0.0000 Min. :0.00000   
## 1st Qu.: 0.000 1st Qu.:1137 1st Qu.:0.0000 1st Qu.:0.00000   
## Median : 0.000 Median :1447 Median :0.0000 Median :0.00000   
## Mean : 4.973 Mean :1499 Mean :0.4301 Mean :0.05796   
## 3rd Qu.: 0.000 3rd Qu.:1737 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1064.000 Max. :5095 Max. :3.0000 Max. :2.00000   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## Min. :0.000 Min. :0.0000 Min. :0.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000   
## Median :2.000 Median :0.0000 Median :3.000 Median :1.000   
## Mean :1.564 Mean :0.3751 Mean :2.855 Mean :1.047   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :4.000 Max. :2.0000 Max. :6.000 Max. :3.000   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## Length:2053 Min. : 3.000 Length:2053 Min. :0.000   
## Class :character 1st Qu.: 5.000 Class :character 1st Qu.:0.000   
## Mode :character Median : 6.000 Mode :character Median :1.000   
## Mean : 6.442 Mean :0.603   
## 3rd Qu.: 7.000 3rd Qu.:1.000   
## Max. :15.000 Max. :4.000   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## Length:2053 Length:2053 Length:2053 Min. :0.000   
## Class :character Class :character Class :character 1st Qu.:1.000   
## Mode :character Mode :character Mode :character Median :2.000   
## Mean :1.774   
## 3rd Qu.:2.000   
## Max. :5.000   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## Min. : 0 Length:2053 Length:2053 Length:2053   
## 1st Qu.: 320 Class :character Class :character Class :character   
## Median : 478 Mode :character Mode :character Mode :character   
## Mean : 472   
## 3rd Qu.: 576   
## Max. :1488   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 27.00 Median : 0.00 Median : 0.000   
## Mean : 93.52 Mean : 48.17 Mean : 23.02 Mean : 2.799   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1424.00 Max. :742.00 Max. :584.00 Max. :407.000   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## Min. : 0.00 Min. : 0.000 Length:2053 Length:2053   
## 1st Qu.: 0.00 1st Qu.: 0.000 Class :character Class :character   
## Median : 0.00 Median : 0.000 Mode :character Mode :character   
## Mean : 16.68 Mean : 1.339   
## 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :576.00 Max. :800.000   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold   
## Length:2053 Min. : 0.00 Min. : 1.000 Min. :2006   
## Class :character 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007   
## Mode :character Median : 0.00 Median : 6.000 Median :2008   
## Mean : 60.12 Mean : 6.189 Mean :2008   
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009   
## Max. :17000.00 Max. :12.000 Max. :2010   
## Sale\_Type Sale\_Condition Longitude Latitude   
## Length:2053 Length:2053 Min. :-93.69 Min. :41.99   
## Class :character Class :character 1st Qu.:-93.66 1st Qu.:42.02   
## Mode :character Mode :character Median :-93.64 Median :42.03   
## Mean :-93.64 Mean :42.03   
## 3rd Qu.:-93.62 3rd Qu.:42.05   
## Max. :-93.58 Max. :42.06   
## Above\_Median   
## Length:2053   
## Class :character   
## Mode :character   
##   
##   
##

# This dataset requires quite a bit of cleaning and preparation before analysis.

# check to see if there are missing values  
na\_count <-sapply(ames\_table, function(ames\_table) sum(length(which(is.na(ames\_table)))))  
na\_count

## MS\_SubClass MS\_Zoning Lot\_Frontage Lot\_Area   
## 0 0 0 0   
## Street Alley Lot\_Shape Land\_Contour   
## 0 0 0 0   
## Utilities Lot\_Config Land\_Slope Neighborhood   
## 0 0 0 0   
## Condition\_1 Condition\_2 Bldg\_Type House\_Style   
## 0 0 0 0   
## Overall\_Qual Overall\_Cond Year\_Built Year\_Remod\_Add   
## 0 0 0 0   
## Roof\_Style Roof\_Matl Exterior\_1st Exterior\_2nd   
## 0 0 0 0   
## Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual Exter\_Cond   
## 0 0 0 0   
## Foundation Bsmt\_Qual Bsmt\_Cond Bsmt\_Exposure   
## 0 0 0 0   
## BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2 BsmtFin\_SF\_2   
## 0 0 0 0   
## Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating Heating\_QC   
## 0 0 0 0   
## Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## 0 0 0 0   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## 0 0 0 0   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## 0 0 0 0   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## 0 0 0 0   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## 0 0 0 0   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## 0 0 0 0   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch   
## 0 0 0 0   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## 0 0 0 0   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold   
## 0 0 0 0   
## Sale\_Type Sale\_Condition Longitude Latitude   
## 0 0 0 0   
## Above\_Median   
## 0

# It appears that the missing Values have been removed from the student version of the data.

# Looking at how many unique values there are for each variable  
sapply(ames\_table, function(x) n\_distinct(x))

## MS\_SubClass MS\_Zoning Lot\_Frontage Lot\_Area   
## 16 7 120 1458   
## Street Alley Lot\_Shape Land\_Contour   
## 2 3 4 4   
## Utilities Lot\_Config Land\_Slope Neighborhood   
## 2 5 3 28   
## Condition\_1 Condition\_2 Bldg\_Type House\_Style   
## 9 8 5 8   
## Overall\_Qual Overall\_Cond Year\_Built Year\_Remod\_Add   
## 10 9 114 61   
## Roof\_Style Roof\_Matl Exterior\_1st Exterior\_2nd   
## 6 6 16 17   
## Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual Exter\_Cond   
## 5 382 4 5   
## Foundation Bsmt\_Qual Bsmt\_Cond Bsmt\_Exposure   
## 6 6 6 5   
## BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2 BsmtFin\_SF\_2   
## 7 7 7 210   
## Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating Heating\_QC   
## 947 892 6 5   
## Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## 2 5 942 508   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## 28 1057 4 3   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## 5 3 7 3   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## 5 11 8 5   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## 6 7 4 6   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## 523 6 6 3   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch   
## 321 226 150 24   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## 97 7 5 5   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold   
## 5 33 12 5   
## Sale\_Type Sale\_Condition Longitude Latitude   
## 10 6 1970 1965   
## Above\_Median   
## 2

# Considering which variables to change to factors below:

## char Variables w/ 2 values: Street Utilities Central\_Air Above\_Median

## char Variables w/ 3 values: Alley Land\_Slope Paved\_Drive

## char Variables w/ 4 values: Lot\_Shape Land\_Contour Exter\_Qual Garage\_Finish

## char Variables w/ 5 values: Lot\_Config Bldg\_Type Mas\_Vnr\_Type Exter\_Cond Bsmt\_Exposure Heating\_QC Electrical Kitchen\_Qual Pool\_QC Fence Misc\_Feature

## char Variables w/ 6 values: Roof\_Style Roof\_Matl Foundation Bsmt\_Qual Bsmt\_Cond Heating Fireplace\_Qu Garage\_Qual Garage\_Cond Sale\_Condition

## char Variables w/ 7 values: MS\_Zoning BsmtFin\_Type\_1 BsmtFin\_Type\_2 Garage\_Type

## char Variables w/ 8 Values: Condition\_2 House\_Style Functional

## char Variables w/ 9 values: Condition\_1 Overall\_Cond

## char Variables w/ 10 values: Overall\_Qual

## char Variables w/ 11+ values: MS\_SubClass Neighborhood Exterior\_1st Exterior\_2nd

# View Unique Values in columns  
Column\_2values <- unique(ames\_table["Street"])  
Column\_2values # Pave Grvl

## # A tibble: 2 × 1  
## Street  
## <chr>   
## 1 Pave   
## 2 Grvl

Column\_2values <- unique(ames\_table["Utilities"])  
Column\_2values # AllPub NoSewr

## # A tibble: 2 × 1  
## Utilities  
## <chr>   
## 1 AllPub   
## 2 NoSewr

Column\_2values <- unique(ames\_table["Central\_Air"])  
Column\_2values # Y N

## # A tibble: 2 × 1  
## Central\_Air  
## <chr>   
## 1 Y   
## 2 N

Column\_2values <- unique(ames\_table["Above\_Median"])  
Column\_2values #Yes No

## # A tibble: 2 × 1  
## Above\_Median  
## <chr>   
## 1 Yes   
## 2 No

Column\_3values <- unique(ames\_table["Alley"])  
Column\_3values # No\_Alley\_Access Paved Gravel

## # A tibble: 3 × 1  
## Alley   
## <chr>   
## 1 No\_Alley\_Access  
## 2 Paved   
## 3 Gravel

Column\_3values <- unique(ames\_table["Land\_Slope"])  
Column\_3values # Gtl Mod Sev

## # A tibble: 3 × 1  
## Land\_Slope  
## <chr>   
## 1 Gtl   
## 2 Mod   
## 3 Sev

Column\_3values <- unique(ames\_table["Paved\_Drive"])  
Column\_3values # Partial\_Pavement Paved Dirt\_Gravel

## # A tibble: 3 × 1  
## Paved\_Drive   
## <chr>   
## 1 Partial\_Pavement  
## 2 Paved   
## 3 Dirt\_Gravel

Column\_4values <- unique(ames\_table["Lot\_Shape"])  
Column\_4values # Slightly\_Irregular Regular Moderately\_Irregular Irregular

## # A tibble: 4 × 1  
## Lot\_Shape   
## <chr>   
## 1 Slightly\_Irregular   
## 2 Regular   
## 3 Moderately\_Irregular  
## 4 Irregular

Column\_4values <- unique(ames\_table["Land\_Contour"])  
Column\_4values # Lvl HLS Bnk Low

## # A tibble: 4 × 1  
## Land\_Contour  
## <chr>   
## 1 Lvl   
## 2 HLS   
## 3 Bnk   
## 4 Low

Column\_4values <- unique(ames\_table["Exter\_Qual"]) #Scale-4  
Column\_4values # Typical Good Excellent Fair

## # A tibble: 4 × 1  
## Exter\_Qual  
## <chr>   
## 1 Typical   
## 2 Good   
## 3 Excellent   
## 4 Fair

Column\_4values <- unique(ames\_table["Garage\_Finish"])  
Column\_4values # Fin Unf RFn No\_Garage

## # A tibble: 4 × 1  
## Garage\_Finish  
## <chr>   
## 1 Fin   
## 2 Unf   
## 3 RFn   
## 4 No\_Garage

# View Unique Values in columns  
  
Column\_5values <- unique(ames\_table["Lot\_Config"])  
Column\_5values # Corner Inside CulDSac FR2 FR3

## # A tibble: 5 × 1  
## Lot\_Config  
## <chr>   
## 1 Corner   
## 2 Inside   
## 3 CulDSac   
## 4 FR2   
## 5 FR3

Column\_5values <- unique(ames\_table["Bldg\_Type"]) #DUPLICATES  
Column\_5values # OneFam TwnhsE Twnhs Duplex TwoFmCon

## # A tibble: 5 × 1  
## Bldg\_Type  
## <chr>   
## 1 OneFam   
## 2 TwnhsE   
## 3 Twnhs   
## 4 Duplex   
## 5 TwoFmCon

Column\_5values <- unique(ames\_table["Mas\_Vnr\_Type"])  
Column\_5values # Stone None BrkFace BrkCmn CBlock

## # A tibble: 5 × 1  
## Mas\_Vnr\_Type  
## <chr>   
## 1 Stone   
## 2 None   
## 3 BrkFace   
## 4 BrkCmn   
## 5 CBlock

Column\_5values <- unique(ames\_table["Exter\_Cond"]) #Scale-5  
Column\_5values # Typical Good Fair Excellent Poor

## # A tibble: 5 × 1  
## Exter\_Cond  
## <chr>   
## 1 Typical   
## 2 Good   
## 3 Fair   
## 4 Excellent   
## 5 Poor

Column\_5values <- unique(ames\_table["Bsmt\_Exposure"]) #DUPLICATES  
Column\_5values # Gd No Av Mn No\_Basement

## # A tibble: 5 × 1  
## Bsmt\_Exposure  
## <chr>   
## 1 Gd   
## 2 No   
## 3 Av   
## 4 Mn   
## 5 No\_Basement

Column\_5values <- unique(ames\_table["Heating\_QC"]) #Scale-5  
Column\_5values # Fair Typical Excellent Good Poor

## # A tibble: 5 × 1  
## Heating\_QC  
## <chr>   
## 1 Fair   
## 2 Typical   
## 3 Excellent   
## 4 Good   
## 5 Poor

Column\_5values <- unique(ames\_table["Electrical"])  
Column\_5values # SBrkr FuseA FuseF FuseP Unknown

## # A tibble: 5 × 1  
## Electrical  
## <chr>   
## 1 SBrkr   
## 2 FuseA   
## 3 FuseF   
## 4 FuseP   
## 5 Unknown

Column\_5values <- unique(ames\_table["Kitchen\_Qual"]) #Scale-5  
Column\_5values # Fair Typical Excellent Good Poor

## # A tibble: 5 × 1  
## Kitchen\_Qual  
## <chr>   
## 1 Typical   
## 2 Good   
## 3 Excellent   
## 4 Fair   
## 5 Poor

Column\_5values <- unique(ames\_table["Pool\_QC"]) #Scale-5  
Column\_5values # No\_Pool Excellent Typical Fair Good

## # A tibble: 5 × 1  
## Pool\_QC   
## <chr>   
## 1 No\_Pool   
## 2 Excellent  
## 3 Typical   
## 4 Fair   
## 5 Good

Column\_5values <- unique(ames\_table["Fence"])  
Column\_5values # No\_Fence Minimum\_Privacy Good\_Privacy Good\_Wood Minimum\_Wood\_Wire

## # A tibble: 5 × 1  
## Fence   
## <chr>   
## 1 No\_Fence   
## 2 Minimum\_Privacy   
## 3 Good\_Privacy   
## 4 Good\_Wood   
## 5 Minimum\_Wood\_Wire

Column\_5values <- unique(ames\_table["Misc\_Feature"])  
Column\_5values # None Gar2 Shed Othr Elev

## # A tibble: 5 × 1  
## Misc\_Feature  
## <chr>   
## 1 None   
## 2 Gar2   
## 3 Shed   
## 4 Othr   
## 5 Elev

# View Unique Values in columns  
Column\_6values <- unique(ames\_table["Roof\_Style"])  
Column\_6values # Hip Gable Mansard Gambrel Shed Flat

## # A tibble: 6 × 1  
## Roof\_Style  
## <chr>   
## 1 Hip   
## 2 Gable   
## 3 Mansard   
## 4 Gambrel   
## 5 Shed   
## 6 Flat

Column\_6values <- unique(ames\_table["Roof\_Matl"])  
Column\_6values # CompShg WdShake Tar&Grv WdShngl Roll Metal

## # A tibble: 6 × 1  
## Roof\_Matl  
## <chr>   
## 1 CompShg   
## 2 WdShake   
## 3 Tar&Grv   
## 4 WdShngl   
## 5 Roll   
## 6 Metal

Column\_6values <- unique(ames\_table["Foundation"])  
Column\_6values # CBlock PConc Wood BrkTil Slab Stone

## # A tibble: 6 × 1  
## Foundation  
## <chr>   
## 1 CBlock   
## 2 PConc   
## 3 Wood   
## 4 BrkTil   
## 5 Slab   
## 6 Stone

Column\_6values <- unique(ames\_table["Bsmt\_Qual"]) #Scale-6  
Column\_6values # Typical Good Excellent No\_Basement Fair Poor

## # A tibble: 6 × 1  
## Bsmt\_Qual   
## <chr>   
## 1 Typical   
## 2 Good   
## 3 Excellent   
## 4 No\_Basement  
## 5 Fair   
## 6 Poor

Column\_6values <- unique(ames\_table["Bsmt\_Cond"]) #Scale-6  
Column\_6values # Good Typical Poor No\_Basement Fair Excellent

## # A tibble: 6 × 1  
## Bsmt\_Cond   
## <chr>   
## 1 Good   
## 2 Typical   
## 3 Poor   
## 4 No\_Basement  
## 5 Fair   
## 6 Excellent

Column\_6values <- unique(ames\_table["Heating"])  
Column\_6values # GasA GasW Grav Wall Floor OthW

## # A tibble: 6 × 1  
## Heating  
## <chr>   
## 1 GasA   
## 2 GasW   
## 3 Grav   
## 4 Wall   
## 5 Floor   
## 6 OthW

Column\_6values <- unique(ames\_table["Fireplace\_Qu"]) #Scale-6  
Column\_6values # Good No\_Fireplace Typical Poor Excellent Fair

## # A tibble: 6 × 1  
## Fireplace\_Qu  
## <chr>   
## 1 Good   
## 2 No\_Fireplace  
## 3 Typical   
## 4 Poor   
## 5 Excellent   
## 6 Fair

Column\_6values <- unique(ames\_table["Garage\_Qual"]) #Scale-6  
Column\_6values # Typical No\_Garage Fair Good Excellent Poor

## # A tibble: 6 × 1  
## Garage\_Qual  
## <chr>   
## 1 Typical   
## 2 No\_Garage   
## 3 Fair   
## 4 Good   
## 5 Excellent   
## 6 Poor

Column\_6values <- unique(ames\_table["Garage\_Cond"]) #Scale-6  
Column\_6values # Typical No\_Garage Fair Excellent Poor Good

## # A tibble: 6 × 1  
## Garage\_Cond  
## <chr>   
## 1 Typical   
## 2 No\_Garage   
## 3 Fair   
## 4 Excellent   
## 5 Poor   
## 6 Good

Column\_6values <- unique(ames\_table["Sale\_Condition"])  
Column\_6values # Normal Partial Family Abnorml Alloca AdjLand

## # A tibble: 6 × 1  
## Sale\_Condition  
## <chr>   
## 1 Normal   
## 2 Partial   
## 3 Family   
## 4 Abnorml   
## 5 Alloca   
## 6 AdjLand

# View Unique Values in columns  
Column\_7values <- unique(ames\_table["MS\_Zoning"])  
Column\_7values # Residential\_Low\_Density Residential\_High\_Density Floating\_Village\_Residential Residential\_Medium\_Density C\_all A\_agr I\_all

## # A tibble: 7 × 1  
## MS\_Zoning   
## <chr>   
## 1 Residential\_Low\_Density   
## 2 Residential\_High\_Density   
## 3 Floating\_Village\_Residential  
## 4 Residential\_Medium\_Density   
## 5 C\_all   
## 6 A\_agr   
## 7 I\_all

Column\_7values <- unique(ames\_table["BsmtFin\_Type\_1"])  
Column\_7values # BLQ Rec ALQ GLQ Unf LwQ No\_Basement

## # A tibble: 7 × 1  
## BsmtFin\_Type\_1  
## <chr>   
## 1 BLQ   
## 2 Rec   
## 3 ALQ   
## 4 GLQ   
## 5 Unf   
## 6 LwQ   
## 7 No\_Basement

Column\_7values <- unique(ames\_table["BsmtFin\_Type\_2"])  
Column\_7values # Unf LwQ BLQ Rec GLQ No\_Basement ALQ

## # A tibble: 7 × 1  
## BsmtFin\_Type\_2  
## <chr>   
## 1 Unf   
## 2 LwQ   
## 3 BLQ   
## 4 Rec   
## 5 GLQ   
## 6 No\_Basement   
## 7 ALQ

Column\_7values <- unique(ames\_table["Garage\_Type"])  
Column\_7values # Attchd BuiltIn Basment Detchd No\_Garage CarPort More\_Than\_Two\_Types

## # A tibble: 7 × 1  
## Garage\_Type   
## <chr>   
## 1 Attchd   
## 2 BuiltIn   
## 3 Basment   
## 4 Detchd   
## 5 No\_Garage   
## 6 CarPort   
## 7 More\_Than\_Two\_Types

Column\_8values <- unique(ames\_table["Condition\_2"])  
Column\_8values # Norm Feedr PosA PosN Artery RRNn RRAe RRAn

## # A tibble: 8 × 1  
## Condition\_2  
## <chr>   
## 1 Norm   
## 2 Feedr   
## 3 PosA   
## 4 PosN   
## 5 Artery   
## 6 RRNn   
## 7 RRAe   
## 8 RRAn

Column\_8values <- unique(ames\_table["House\_Style"])  
Column\_8values #One\_Story Two\_Story One\_and\_Half\_Fin SLvl SFoyer One\_and\_Half\_Unf Two\_and\_Half\_Unf Two\_and\_Half\_Fin

## # A tibble: 8 × 1  
## House\_Style   
## <chr>   
## 1 One\_Story   
## 2 Two\_Story   
## 3 One\_and\_Half\_Fin  
## 4 SLvl   
## 5 SFoyer   
## 6 One\_and\_Half\_Unf  
## 7 Two\_and\_Half\_Unf  
## 8 Two\_and\_Half\_Fin

Column\_8values <- unique(ames\_table["Functional"])  
Column\_8values # Typ Mod Min2 Min1 Maj1 Maj2 Sev Sal

## # A tibble: 8 × 1  
## Functional  
## <chr>   
## 1 Typ   
## 2 Mod   
## 3 Min2   
## 4 Min1   
## 5 Maj1   
## 6 Maj2   
## 7 Sev   
## 8 Sal

Column\_9values <- unique(ames\_table["Condition\_1"])  
Column\_9values # Norm Feedr RRAe PosA Artery PosN RRAn RRNn RRNe

## # A tibble: 9 × 1  
## Condition\_1  
## <chr>   
## 1 Norm   
## 2 Feedr   
## 3 RRAe   
## 4 PosA   
## 5 Artery   
## 6 PosN   
## 7 RRAn   
## 8 RRNn   
## 9 RRNe

Column\_9values <- unique(ames\_table["Overall\_Cond"]) #Scale-9  
Column\_9values # Average Above\_Average Good Poor Very\_Good Below\_Average Excellent Fair Very\_Poor

## # A tibble: 9 × 1  
## Overall\_Cond   
## <chr>   
## 1 Average   
## 2 Above\_Average  
## 3 Good   
## 4 Poor   
## 5 Very\_Good   
## 6 Below\_Average  
## 7 Excellent   
## 8 Fair   
## 9 Very\_Poor

Column\_10values <- unique(ames\_table["Overall\_Qual"]) #Scale-10  
Column\_10values # Above\_Average Average Good Very\_Good Excellent Below\_Average Fair Poor Very\_Excellent Very\_Poor

## # A tibble: 10 × 1  
## Overall\_Qual   
## <chr>   
## 1 Above\_Average   
## 2 Average   
## 3 Good   
## 4 Very\_Good   
## 5 Excellent   
## 6 Below\_Average   
## 7 Fair   
## 8 Poor   
## 9 Very\_Excellent  
## 10 Very\_Poor

# Note: There are 1 scale w/ four criteria, 4 scales w/ five criteria, 5 scales with 6 criteria, 1 scale w/ 9 criteria and 1 scale w/ ten criteria. A better way of receiving input would be to have a single measurement scale with a Not Applicable choice.

# Always do this prior to splitting # Create factors from character variable  
  
ames\_table\_clean = ames\_table %>%   
 # Create and order factors w/ 2 levels  
 mutate(Street = as\_factor(Street)) %>%   
 mutate(Street = fct\_recode(Street,"Pave"="0","Grvl"="1")) %>%   
 mutate(Utilities = as\_factor(Utilities)) %>%   
 mutate(Utilities = fct\_recode(Utilities,"AllPub"="0","NoSewr"="1")) %>%   
 mutate(Central\_Air = as\_factor(Central\_Air)) %>%   
 mutate(Central\_Air = fct\_recode(Central\_Air,"N"="0","Y"="1")) %>%   
 mutate(Above\_Median = as\_factor(Above\_Median)) %>%   
 mutate(Above\_Median = fct\_recode(Above\_Median,"No"="0","Yes"="1")) %>%   
   
 # Create and order factors w/ 3 levels   
 mutate(Alley = as\_factor(Alley)) %>%   
 mutate(Alley = fct\_recode(Alley,"No\_Alley\_Access"="0","Paved"="1", "Gravel" = "2")) %>%   
 mutate(Land\_Slope = as\_factor(Land\_Slope)) %>%   
 mutate(Land\_Slope = fct\_recode(Land\_Slope,"Gtl"="0","Mod"="1", "Sev" = "2")) %>%   
 mutate(Paved\_Drive = as\_factor(Paved\_Drive)) %>%   
 mutate(Paved\_Drive = fct\_recode(Paved\_Drive,"Partial\_Pavement"="0","Paved"="1", "Dirt\_Gravel" = "2")) %>%  
  
 # Create and order factors w/ 4 levels  
 mutate(Lot\_Shape = as\_factor(Lot\_Shape)) %>%   
 mutate(Lot\_Shape = fct\_recode(Lot\_Shape,"Slightly\_Irregular"="0","Regular"="1", "Moderately\_Irregular" = "2", "Irregular" = "3")) %>%   
 mutate(Land\_Contour = as\_factor(Land\_Contour)) %>%   
 mutate(Land\_Contour = fct\_recode(Land\_Contour,"Lvl"="0","HLS"="1", "Bnk" = "2", "Low" = "3")) %>%   
 mutate(Exter\_Qual = as\_factor(Exter\_Qual)) %>%   
 mutate(Exter\_Qual = fct\_recode(Exter\_Qual,"Typical"="0","Good"="1", "Excellent" = "2", "Fair" = "3")) %>%   
 mutate(Garage\_Finish = as\_factor(Garage\_Finish)) %>%   
 mutate(Garage\_Finish = fct\_recode(Garage\_Finish,"No\_Garage"="0","Unf"="1", "Fin" = "2", "RFn" = "3")) %>%   
  
 # Fix Duplicate by combining factors TwnhsE and Twnhs   
 mutate(Bldg\_Type = as\_factor(Bldg\_Type)) %>%   
 mutate(Bldg\_Type = fct\_recode(Bldg\_Type,"OneFam"="0","TwnhsE"="1", "Twnhs" = "1", "Duplex" = "2", "TwoFmCon" = "3")) %>%   
   
 # Fix Duplicate by combining factors No and No\_Basement   
 mutate(Bsmt\_Exposure = as\_factor(Bsmt\_Exposure)) %>%   
 mutate(Bsmt\_Exposure = fct\_recode(Bsmt\_Exposure,"No"="0","No\_Basement"="0", "Mn" = "1", "Av" = "2", "Gd" = "3")) %>%   
   
 # Create and order factors w/ 5 levels  
 mutate(Lot\_Config = as\_factor(Lot\_Config)) %>%   
 mutate(Lot\_Config = fct\_recode(Lot\_Config,"Corner"="0","Inside"="1", "CulDSac" = "2", "FR2" = "3", "FR3" = "4")) %>%  
 mutate(Mas\_Vnr\_Type = as\_factor(Mas\_Vnr\_Type)) %>%   
 mutate(Mas\_Vnr\_Type = fct\_recode(Mas\_Vnr\_Type,"None"="0","Stone"="1", "BrkFace" = "2", "BrkCmn" = "3", "Cblock" = "4")) %>%  
 mutate(Exter\_Cond = as\_factor(Exter\_Cond)) %>%  
 mutate(Exter\_Cond = fct\_recode(Exter\_Cond,"Poor"="0","Fair"="1", "Typical" = "2", "Good" = "3", "Excellent" = "4")) %>%  
 mutate(Heating\_QC = as\_factor(Heating\_QC)) %>%   
 mutate(Heating\_QC = fct\_recode(Heating\_QC,"Poor"="0","Fair"="1", "Typical" = "2", "Good" = "3", "Excellent" = "4")) %>%  
 mutate(Electrical = as\_factor(Electrical)) %>%   
 mutate(Electrical = fct\_recode(Electrical,"SBrkr"="0","FuseA"="1", "FuseF" = "2", "FuseP" = "3", "Unknown" = "4")) %>%  
 mutate(Kitchen\_Qual = as\_factor(Kitchen\_Qual)) %>%   
 mutate(Kitchen\_Qual = fct\_recode(Kitchen\_Qual,"Poor"="0","Fair"="1", "Typical" = "2", "Good" = "3", "Excellent" = "4")) %>%  
 mutate(Pool\_QC = as\_factor(Pool\_QC)) %>%   
 mutate(Pool\_QC = fct\_recode(Pool\_QC,"No\_Pool"="0","Fair"="1", "Typical" = "2", "Good" = "3", "Excellent" = "4")) %>%  
 mutate(Fence = as\_factor(Fence)) %>%   
 mutate(Fence = fct\_recode(Fence,"No\_Fence"="0","Minimum\_Privacy"="1", "Minimum\_Wood\_Wire" = "2", "Good\_Privacy" = "3", "Good\_Wood" = "4")) %>%   
 mutate(Misc\_Feature = as\_factor(Misc\_Feature)) %>%   
 mutate(Misc\_Feature = fct\_recode(Misc\_Feature,"None"="0","Shed"="1", "Gar2" = "2", "Othr" = "3", "Elev" = "4")) %>%   
  
 # Create and order factors w/ 6 levels  
 mutate(Roof\_Style = as\_factor(Roof\_Style)) %>%   
 mutate(Roof\_Style = fct\_recode(Roof\_Style,"Tar&Grv"="0","WdShake"="1", "WdShngl" = "2", "Roll" = "3", "CompShg" = "4", "Metal" = "5")) %>%  
 mutate(Roof\_Matl = as\_factor(Roof\_Matl)) %>%   
 mutate(Roof\_Matl = fct\_recode(Roof\_Matl,"No\_Pool"="0","Poor"="1", "Fair" = "2", "Typical" = "3", "Good" = "4", "Excellent" = "5")) %>%  
 mutate(Foundation = as\_factor(Foundation)) %>%   
 mutate(Foundation = fct\_recode(Foundation,"CBlock"="0","PConc"="1", "Wood" = "2", "BrkTil" = "3", "Slab" = "4", "Stone" = "5")) %>%  
 mutate(Bsmt\_Qual = as\_factor(Bsmt\_Qual)) %>%   
 mutate(Bsmt\_Qual = fct\_recode(Bsmt\_Qual,"No\_Basement"="0","Poor"="1", "Fair" = "2", "Typical" = "3", "Good" = "4", "Excellent" = "5")) %>%  
 mutate(Bsmt\_Cond = as\_factor(Bsmt\_Cond)) %>%   
 mutate(Bsmt\_Cond = fct\_recode(Bsmt\_Cond,"No\_Basement"="0","Poor"="1", "Fair" = "2", "Typical" = "3", "Good" = "4", "Excellent" = "5")) %>%  
 mutate(Heating = as\_factor(Heating)) %>%   
 mutate(Heating = fct\_recode(Heating,"GasA"="0","GasW"="1", "Grav" = "2", "Wall" = "3", "Floor" = "4", "OtherW" = "5")) %>%  
 mutate(Fireplace\_Qu = as\_factor(Fireplace\_Qu)) %>%   
 mutate(Fireplace\_Qu = fct\_recode(Fireplace\_Qu,"No\_Fireplace"="0","Poor"="1", "Fair" = "2", "Typical" = "3", "Good" = "4", "Excellent" = "5")) %>%  
 mutate(Garage\_Qual = as\_factor(Garage\_Qual)) %>%   
 mutate(Garage\_Qual = fct\_recode(Garage\_Qual,"No\_Garage"="0","Poor"="1", "Fair" = "2", "Typical" = "3", "Good" = "4", "Excellent" = "5")) %>%  
 mutate(Garage\_Cond = as\_factor(Garage\_Cond)) %>%   
 mutate(Garage\_Cond = fct\_recode(Garage\_Cond,"No\_Garage"="0","Poor"="1", "Fair" = "2", "Typical" = "3", "Good" = "4", "Excellent" = "5")) %>%  
 mutate(Sale\_Condition = as\_factor(Sale\_Condition)) %>%   
 mutate(Sale\_Condition = fct\_recode(Sale\_Condition,"Abnorml"="0","Partial"="1", "Normal" = "2", "Alloca" = "3", "AdjLand" = "4", "Excellent" = "5")) %>%   
  
 # Create and order factors w/ 7 levels   
 mutate(MS\_Zoning = as\_factor(MS\_Zoning)) %>%   
 mutate(MS\_Zoning = fct\_recode(MS\_Zoning,"I\_all"="0","A\_agr"="1", "C\_all" = "2", "Floating\_Village" = "3", "Residential\_High\_Density" = "4", "Residential\_Medium\_Density" = "5","Residential\_Low\_Density"="6")) %>%  
 mutate(BsmtFin\_Type\_1 = as\_factor(BsmtFin\_Type\_1)) %>%   
 mutate(BsmtFin\_Type\_1 = fct\_recode(BsmtFin\_Type\_1,"No\_Basement"="0","Unf"="1", "LwO" = "2", "GLQ" = "3", "ALQ" = "4", "BLQ" = "5","Rec"="6")) %>%  
 mutate(BsmtFin\_Type\_2 = as\_factor(BsmtFin\_Type\_2)) %>%   
 mutate(BsmtFin\_Type\_2 = fct\_recode(BsmtFin\_Type\_2,"No\_Basement"="0","Unf"="1", "LwO" = "2", "GLQ" = "3", "ALQ" = "4", "BLQ" = "5","Rec"="6")) %>%  
 mutate(Garage\_Type = as\_factor(Garage\_Type)) %>%   
 mutate(Garage\_Type = fct\_recode(Garage\_Type,"No\_Garage"="0","CarPort"="1", "Detchd" = "2", "Attchd" = "3", "BuiltIn" = "4", "Basment" = "5","More\_Than\_Two\_Types"="6")) %>%  
   
 # Create and order factors w/ 8 levels   
 mutate(Condition\_2 = as\_factor(Condition\_2)) %>%   
 mutate(Condition\_2 = fct\_recode(Condition\_2,"RRAn"="0","RRAe"="1", "RRNn" = "2", "Artery" = "3", "Feedr" = "4", "PosA" = "5","PosN"="6","Norm"="7")) %>%  
 mutate(House\_Style = as\_factor(House\_Style)) %>%   
 mutate(House\_Style = fct\_recode(House\_Style,"Slvl"="0","One\_Story"="1","SFoyer"="7", "One\_and\_Half\_Unf" = "3", "One\_and\_Half\_Fin" = "4", "Two\_Story" = "5", "Two\_and\_Half\_Unf" = "6","Two\_and\_Half\_Fin"="7")) %>%  
 mutate(Functional = as\_factor(Functional)) %>%   
 mutate(Functional = fct\_recode(Functional,"Sal"="0","Sev"="1", "Maj2" = "2", "Maj1" = "3", "Mod" = "4", "Typical" = "5","Min2"="6","Min1"="7")) %>%  
   
 # Create and order factors w/ 9 levels   
 mutate(Condition\_1 = as\_factor(Condition\_1)) %>%   
 mutate(Condition\_1 = fct\_recode(Condition\_1,"RRAn"="0","RRAe"="1", "RRNn" = "2", "RRNe" = "3", "Artery" = "4", "Feeder" = "5","PosA"="6","PosN"="7", "Normal"="0")) %>%   
 mutate(Overall\_Cond = as\_factor(Overall\_Cond)) %>%   
 mutate(Overall\_Cond = fct\_recode(Overall\_Cond,"Very\_Poor"="0","Poor"="1", "Below\_Average" = "2", "Fair" = "3", "Average" = "4", "Above\_Average" = "5","Good"="6","Very\_Good"="7", "Excellent"="8")) %>%   
   
 # Create and order factors w/ 10 levels   
 mutate(Overall\_Qual = as\_factor(Overall\_Qual)) %>%   
 mutate(Overall\_Qual = fct\_recode(Overall\_Qual,"Very\_Poor"="0","Poor"="1", "Below\_Average" = "2", "Fair" = "3", "Average" = "4", "Above\_Average" = "5","Good"="6","Very\_Good"="7", "Excellent"="8", "Very\_Excellent"="9")) %>%   
   
 # Change anything left that is character to a factor  
 mutate\_if(is.character, as\_factor)

## Warning: Unknown levels in `f`: 0, 1  
  
## Warning: Unknown levels in `f`: 0, 1  
  
## Warning: Unknown levels in `f`: 0, 1  
  
## Warning: Unknown levels in `f`: 0, 1

## Warning: Unknown levels in `f`: 0, 1, 2  
  
## Warning: Unknown levels in `f`: 0, 1, 2  
  
## Warning: Unknown levels in `f`: 0, 1, 2

## Warning: Unknown levels in `f`: 0, 1, 2, 3  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3

## Warning: Unknown levels in `f`: 0, 1, 1, 2, 3

## Warning: Unknown levels in `f`: 0, 0, 1, 2, 3

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6  
  
## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6, 7

## Warning: Unknown levels in `f`: 0, 1, 7, 3, 4, 5, 6, 7

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6, 7

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6, 7, 0

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6, 7, 8

## Warning: Unknown levels in `f`: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

str (ames\_table\_clean)

## tibble [2,053 × 81] (S3: tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : Factor w/ 16 levels "One\_Story\_1946\_and\_Newer\_All\_Styles",..: 1 1 1 1 2 2 3 3 1 1 ...  
## $ MS\_Zoning : Factor w/ 7 levels "Residential\_Low\_Density",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Frontage : num [1:2053] 141 80 81 93 74 78 43 39 0 85 ...  
## $ Lot\_Area : num [1:2053] 31770 11622 14267 11160 13830 ...  
## $ Street : Factor w/ 2 levels "Pave","Grvl": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Alley : Factor w/ 3 levels "No\_Alley\_Access",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Shape : Factor w/ 4 levels "Slightly\_Irregular",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ Land\_Contour : Factor w/ 4 levels "Lvl","HLS","Bnk",..: 1 1 1 1 1 1 2 1 1 1 ...  
## $ Utilities : Factor w/ 2 levels "AllPub","NoSewr": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Config : Factor w/ 5 levels "Corner","Inside",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Land\_Slope : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Neighborhood : Factor w/ 28 levels "North\_Ames","Gilbert",..: 1 1 1 1 2 2 3 3 2 2 ...  
## $ Condition\_1 : Factor w/ 9 levels "Norm","Feedr",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ Condition\_2 : Factor w/ 8 levels "Norm","Feedr",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Bldg\_Type : Factor w/ 5 levels "OneFam","TwnhsE",..: 1 1 1 1 1 1 2 2 1 1 ...  
## $ House\_Style : Factor w/ 8 levels "One\_Story","Two\_Story",..: 1 1 1 1 2 2 1 1 1 1 ...  
## $ Overall\_Qual : Factor w/ 10 levels "Above\_Average",..: 1 2 1 3 2 1 4 4 1 3 ...  
## $ Overall\_Cond : Factor w/ 9 levels "Average","Above\_Average",..: 1 2 2 1 1 2 1 1 3 1 ...  
## $ Year\_Built : num [1:2053] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add : num [1:2053] 1960 1961 1958 1968 1998 ...  
## $ Roof\_Style : Factor w/ 6 levels "Hip","Gable",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Roof\_Matl : Factor w/ 6 levels "CompShg","WdShake",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Exterior\_1st : Factor w/ 16 levels "BrkFace","VinylSd",..: 1 2 3 1 2 2 4 5 4 4 ...  
## $ Exterior\_2nd : Factor w/ 17 levels "Plywood","VinylSd",..: 1 2 3 4 2 2 5 6 5 5 ...  
## $ Mas\_Vnr\_Type : Factor w/ 5 levels "Stone","None",..: 1 2 3 2 2 3 2 2 2 2 ...  
## $ Mas\_Vnr\_Area : num [1:2053] 112 0 108 0 0 20 0 0 0 0 ...  
## $ Exter\_Qual : Factor w/ 4 levels "Typical","Good",..: 1 1 1 2 1 1 2 2 1 1 ...  
## $ Exter\_Cond : Factor w/ 5 levels "Typical","Good",..: 1 1 1 1 1 1 1 1 2 1 ...  
## $ Foundation : Factor w/ 6 levels "CBlock","PConc",..: 1 1 1 1 2 2 2 2 2 2 ...  
## $ Bsmt\_Qual : Factor w/ 6 levels "Typical","Good",..: 1 1 1 1 2 1 2 2 2 2 ...  
## $ Bsmt\_Cond : Factor w/ 6 levels "Good","Typical",..: 1 2 2 2 2 2 2 2 2 2 ...  
## $ Bsmt\_Exposure : Factor w/ 5 levels "Gd","No","Av",..: 1 2 2 2 2 2 2 2 2 1 ...  
## $ BsmtFin\_Type\_1 : Factor w/ 7 levels "BLQ","Rec","ALQ",..: 1 2 3 3 4 4 3 4 3 4 ...  
## $ BsmtFin\_SF\_1 : num [1:2053] 2 6 1 1 3 3 1 3 1 3 ...  
## $ BsmtFin\_Type\_2 : Factor w/ 7 levels "Unf","LwQ","BLQ",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ BsmtFin\_SF\_2 : num [1:2053] 0 144 0 0 0 0 0 0 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:2053] 441 270 406 1045 137 ...  
## $ Total\_Bsmt\_SF : num [1:2053] 1080 882 1329 2110 928 ...  
## $ Heating : Factor w/ 6 levels "GasA","GasW",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Heating\_QC : Factor w/ 5 levels "Fair","Typical",..: 1 2 2 3 4 3 3 3 3 4 ...  
## $ Central\_Air : Factor w/ 2 levels "Y","N": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Electrical : Factor w/ 5 levels "SBrkr","FuseA",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ First\_Flr\_SF : num [1:2053] 1656 896 1329 2110 928 ...  
## $ Second\_Flr\_SF : num [1:2053] 0 0 0 0 701 678 0 0 0 0 ...  
## $ Low\_Qual\_Fin\_SF : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:2053] 1656 896 1329 2110 1629 ...  
## $ Bsmt\_Full\_Bath : num [1:2053] 1 0 0 1 0 0 0 1 1 1 ...  
## $ Bsmt\_Half\_Bath : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:2053] 1 1 1 2 2 2 2 2 2 1 ...  
## $ Half\_Bath : num [1:2053] 0 0 1 1 1 1 0 0 0 1 ...  
## $ Bedroom\_AbvGr : num [1:2053] 3 2 3 3 3 3 2 2 3 2 ...  
## $ Kitchen\_AbvGr : num [1:2053] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : Factor w/ 5 levels "Typical","Good",..: 1 1 2 3 1 2 2 2 1 2 ...  
## $ TotRms\_AbvGrd : num [1:2053] 7 5 6 8 6 7 5 5 6 5 ...  
## $ Functional : Factor w/ 8 levels "Typ","Mod","Min2",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Fireplaces : num [1:2053] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Fireplace\_Qu : Factor w/ 6 levels "Good","No\_Fireplace",..: 1 2 2 3 3 1 2 3 2 4 ...  
## $ Garage\_Type : Factor w/ 7 levels "Attchd","BuiltIn",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Garage\_Finish : Factor w/ 4 levels "Fin","Unf","RFn",..: 1 2 2 1 1 1 3 3 1 2 ...  
## $ Garage\_Cars : num [1:2053] 2 1 1 2 2 2 2 2 2 2 ...  
## $ Garage\_Area : num [1:2053] 528 730 312 522 482 470 506 608 420 506 ...  
## $ Garage\_Qual : Factor w/ 6 levels "Typical","No\_Garage",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Garage\_Cond : Factor w/ 6 levels "Typical","No\_Garage",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Paved\_Drive : Factor w/ 3 levels "Partial\_Pavement",..: 1 2 2 2 2 2 2 2 2 2 ...  
## $ Wood\_Deck\_SF : num [1:2053] 210 140 393 0 212 360 0 237 483 192 ...  
## $ Open\_Porch\_SF : num [1:2053] 62 0 36 0 34 36 82 152 21 0 ...  
## $ Enclosed\_Porch : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:2053] 0 120 0 0 0 0 144 0 0 0 ...  
## $ Pool\_Area : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : Factor w/ 5 levels "No\_Pool","Excellent",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Fence : Factor w/ 5 levels "No\_Fence","Minimum\_Privacy",..: 1 2 1 1 2 1 1 1 3 1 ...  
## $ Misc\_Feature : Factor w/ 5 levels "None","Gar2",..: 1 1 2 1 1 1 1 1 3 1 ...  
## $ Misc\_Val : num [1:2053] 0 0 12500 0 0 0 0 0 500 0 ...  
## $ Mo\_Sold : num [1:2053] 5 6 6 4 3 6 1 3 3 2 ...  
## $ Year\_Sold : num [1:2053] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : Factor w/ 10 levels "WD","New","COD",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Sale\_Condition : Factor w/ 6 levels "Normal","Partial",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Longitude : num [1:2053] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:2053] 42.1 42.1 42.1 42.1 42.1 ...  
## $ Above\_Median : Factor w/ 2 levels "Yes","No": 1 2 1 1 1 1 1 1 1 1 ...

summary(ames\_table\_clean)

## MS\_SubClass MS\_Zoning   
## One\_Story\_1946\_and\_Newer\_All\_Styles :772 Residential\_Low\_Density :1600   
## Two\_Story\_1946\_and\_Newer :383 Residential\_High\_Density : 20   
## One\_and\_Half\_Story\_Finished\_All\_Ages:204 Floating\_Village\_Residential: 87   
## One\_Story\_PUD\_1946\_and\_Newer :129 Residential\_Medium\_Density : 326   
## One\_Story\_1945\_and\_Older : 98 C\_all : 17   
## Two\_Story\_1945\_and\_Older : 95 A\_agr : 2   
## (Other) :372 I\_all : 1   
## Lot\_Frontage Lot\_Area Street Alley   
## Min. : 0.00 Min. : 1300 Pave:2046 No\_Alley\_Access:1914   
## 1st Qu.: 43.00 1st Qu.: 7500 Grvl: 7 Paved : 45   
## Median : 62.00 Median : 9548 Gravel : 94   
## Mean : 57.38 Mean : 10258   
## 3rd Qu.: 78.00 3rd Qu.: 11600   
## Max. :313.00 Max. :215245   
##   
## Lot\_Shape Land\_Contour Utilities Lot\_Config   
## Slightly\_Irregular : 714 Lvl:1833 AllPub:2052 Corner : 359   
## Regular :1275 HLS: 94 NoSewr: 1 Inside :1495   
## Moderately\_Irregular: 53 Bnk: 81 CulDSac: 135   
## Irregular : 11 Low: 45 FR2 : 56   
## FR3 : 8   
##   
##   
## Land\_Slope Neighborhood Condition\_1 Condition\_2 Bldg\_Type   
## Gtl:1951 North\_Ames : 327 Norm :1771 Norm :2027 OneFam :1706   
## Mod: 89 College\_Creek: 183 Feedr : 113 Feedr : 12 TwnhsE : 157   
## Sev: 13 Old\_Town : 181 Artery : 67 PosA : 4 Twnhs : 67   
## Edwards : 129 RRAn : 35 Artery : 4 Duplex : 76   
## Somerset : 119 PosN : 24 PosN : 3 TwoFmCon: 47   
## Gilbert : 109 RRAe : 19 RRNn : 1   
## (Other) :1005 (Other): 24 (Other): 2   
## House\_Style Overall\_Qual Overall\_Cond   
## One\_Story :1052 Average :587 Average :1143   
## Two\_Story : 590 Above\_Average:518 Above\_Average: 376   
## One\_and\_Half\_Fin: 225 Good :411 Good : 286   
## SLvl : 90 Very\_Good :237 Very\_Good : 98   
## SFoyer : 56 Below\_Average:169 Below\_Average: 73   
## Two\_and\_Half\_Unf: 19 Excellent : 70 Fair : 35   
## (Other) : 21 (Other) : 61 (Other) : 42   
## Year\_Built Year\_Remod\_Add Roof\_Style Roof\_Matl Exterior\_1st  
## Min. :1875 Min. :1950 Hip : 404 CompShg:2023 VinylSd:705   
## 1st Qu.:1953 1st Qu.:1965 Gable :1607 WdShake: 8 MetalSd:319   
## Median :1972 Median :1993 Mansard: 9 Tar&Grv: 17 Wd Sdng:313   
## Mean :1971 Mean :1984 Gambrel: 14 WdShngl: 3 HdBoard:303   
## 3rd Qu.:2000 3rd Qu.:2004 Shed : 5 Roll : 1 Plywood:151   
## Max. :2010 Max. :2010 Flat : 14 Metal : 1 CemntBd: 90   
## (Other):172   
## Exterior\_2nd Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual   
## VinylSd:699 Stone : 166 Min. : 0.0 Typical :1272   
## MetalSd:317 None :1231 1st Qu.: 0.0 Good : 682   
## Wd Sdng:302 BrkFace: 638 Median : 0.0 Excellent: 78   
## HdBoard:277 BrkCmn : 17 Mean : 103.8 Fair : 21   
## Plywood:190 CBlock : 1 3rd Qu.: 164.0   
## CmentBd: 90 Max. :1600.0   
## (Other):178   
## Exter\_Cond Foundation Bsmt\_Qual Bsmt\_Cond   
## Typical :1787 CBlock:880 Typical :911 Good : 80   
## Good : 213 PConc :911 Good :849 Typical :1833   
## Fair : 43 Wood : 4 Excellent :178 Poor : 4   
## Excellent: 9 BrkTil:216 No\_Basement: 57 No\_Basement: 57   
## Poor : 1 Slab : 36 Fair : 57 Fair : 76   
## Stone : 6 Poor : 1 Excellent : 3   
##   
## Bsmt\_Exposure BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2  
## Gd : 199 BLQ :196 Min. :1.00 Unf :1740   
## No :1331 Rec :216 1st Qu.:3.00 LwQ : 64   
## Av : 284 ALQ :298 Median :3.00 BLQ : 47   
## Mn : 179 GLQ :578 Mean :4.21 Rec : 79   
## No\_Basement: 60 Unf :602 3rd Qu.:7.00 GLQ : 23   
## LwQ :106 Max. :7.00 No\_Basement: 58   
## No\_Basement: 57 ALQ : 42   
## BsmtFin\_SF\_2 Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating   
## Min. : 0.00 Min. : 0.0 Min. : 0 GasA :2019   
## 1st Qu.: 0.00 1st Qu.: 226.0 1st Qu.: 793 GasW : 21   
## Median : 0.00 Median : 460.0 Median : 988 Grav : 6   
## Mean : 52.57 Mean : 561.2 Mean :1055 Wall : 5   
## 3rd Qu.: 0.00 3rd Qu.: 801.0 3rd Qu.:1304 Floor: 1   
## Max. :1526.00 Max. :2336.0 Max. :5095 OthW : 1   
##   
## Heating\_QC Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## Fair : 61 Y:1916 SBrkr :1887 Min. : 432 Min. : 0.0   
## Typical : 618 N: 137 FuseA : 126 1st Qu.: 882 1st Qu.: 0.0   
## Excellent:1040 FuseF : 33 Median :1088 Median : 0.0   
## Good : 333 FuseP : 6 Mean :1168 Mean : 326.1   
## Poor : 1 Unknown: 1 3rd Qu.:1402 3rd Qu.: 701.0   
## Max. :5095 Max. :1862.0   
##   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## Min. : 0.000 Min. : 480 Min. :0.0000 Min. :0.00000   
## 1st Qu.: 0.000 1st Qu.:1137 1st Qu.:0.0000 1st Qu.:0.00000   
## Median : 0.000 Median :1447 Median :0.0000 Median :0.00000   
## Mean : 4.973 Mean :1499 Mean :0.4301 Mean :0.05796   
## 3rd Qu.: 0.000 3rd Qu.:1737 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1064.000 Max. :5095 Max. :3.0000 Max. :2.00000   
##   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## Min. :0.000 Min. :0.0000 Min. :0.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000   
## Median :2.000 Median :0.0000 Median :3.000 Median :1.000   
## Mean :1.564 Mean :0.3751 Mean :2.855 Mean :1.047   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :4.000 Max. :2.0000 Max. :6.000 Max. :3.000   
##   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## Typical :1070 Min. : 3.000 Typ :1896 Min. :0.000   
## Good : 790 1st Qu.: 5.000 Min2 : 54 1st Qu.:0.000   
## Excellent: 142 Median : 6.000 Min1 : 51 Median :1.000   
## Fair : 50 Mean : 6.442 Mod : 27 Mean :0.603   
## Poor : 1 3rd Qu.: 7.000 Maj1 : 15 3rd Qu.:1.000   
## Max. :15.000 Maj2 : 6 Max. :4.000   
## (Other): 4   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## Good :538 Attchd :1204 Fin :509 Min. :0.000   
## No\_Fireplace:993 BuiltIn : 127 Unf :872 1st Qu.:1.000   
## Typical :409 Basment : 29 RFn :563 Median :2.000   
## Poor : 36 Detchd : 549 No\_Garage:109 Mean :1.774   
## Excellent : 21 No\_Garage : 108 3rd Qu.:2.000   
## Fair : 56 CarPort : 15 Max. :5.000   
## More\_Than\_Two\_Types: 21   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## Min. : 0 Typical :1839 Typical :1872 Partial\_Pavement: 42   
## 1st Qu.: 320 No\_Garage: 109 No\_Garage: 109 Paved :1848   
## Median : 478 Fair : 85 Fair : 53 Dirt\_Gravel : 163   
## Mean : 472 Good : 16 Excellent: 1   
## 3rd Qu.: 576 Excellent: 2 Poor : 8   
## Max. :1488 Poor : 2 Good : 10   
##   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 27.00 Median : 0.00 Median : 0.000   
## Mean : 93.52 Mean : 48.17 Mean : 23.02 Mean : 2.799   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1424.00 Max. :742.00 Max. :584.00 Max. :407.000   
##   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## Min. : 0.00 Min. : 0.000 No\_Pool :2047 No\_Fence :1661   
## 1st Qu.: 0.00 1st Qu.: 0.000 Excellent: 2 Minimum\_Privacy : 225   
## Median : 0.00 Median : 0.000 Typical : 2 Good\_Privacy : 81   
## Mean : 16.68 Mean : 1.339 Fair : 1 Good\_Wood : 77   
## 3rd Qu.: 0.00 3rd Qu.: 0.000 Good : 1 Minimum\_Wood\_Wire: 9   
## Max. :576.00 Max. :800.000   
##   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold Sale\_Type   
## None:1978 Min. : 0.00 Min. : 1.000 Min. :2006 WD :1789   
## Gar2: 5 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007 New : 163   
## Shed: 66 Median : 0.00 Median : 6.000 Median :2008 COD : 54   
## Othr: 3 Mean : 60.12 Mean : 6.189 Mean :2008 ConLD : 16   
## Elev: 1 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009 ConLI : 8   
## Max. :17000.00 Max. :12.000 Max. :2010 CWD : 8   
## (Other): 15   
## Sale\_Condition Longitude Latitude Above\_Median  
## Normal :1712 Min. :-93.69 Min. :41.99 Yes:1043   
## Partial: 169 1st Qu.:-93.66 1st Qu.:42.02 No :1010   
## Family : 30 Median :-93.64 Median :42.03   
## Abnorml: 121 Mean :-93.64 Mean :42.03   
## Alloca : 16 3rd Qu.:-93.62 3rd Qu.:42.05   
## AdjLand: 5 Max. :-93.58 Max. :42.06   
##

# Data Exploration

I chose to use the Esquisser app because it was less memory intensive. This required me to save the graphs in another document and then analyze. The PDF document of these graphs can be found here: <https://drive.google.com/file/d/1rhv-ButSuqcScUmLIQRzTw5Bm9ZuT07v/view?usp=sharing>

# Create and save in another document the pictures of the bar graphs comparing Above\_Median to all the other variables. Will use this to choose which variables have a marked difference between Yes-AboveMedian and No-AboveMedian.  
#esquisser() # commented out so it doesn't get knitted