Above\_Median Project

Tabitha Hagen

2022-10-15

# Project for BAN 502 Predictive Analytics

# Read in the clean\_ames\_table which was created in “THagenBAN502Project\_Phase1.rmd”

clean\_ames\_table <- readRDS("ames\_cleaned\_table.rds")

str (clean\_ames\_table)

## 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(clean\_ames\_table)

## 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   
##

new\_ames\_table = clean\_ames\_table %>% dplyr::select("Above\_Median", "Lot\_Shape", "Overall\_Qual", "Mas\_Vnr\_Type", "Exter\_Qual", "Foundation", "Bsmt\_Qual", "Heating\_QC", "Kitchen\_Qual", "Fireplace\_Qu", "Garage\_Type", "Garage\_Finish", "Mas\_Vnr\_Area", "Second\_Flr\_SF", "Low\_Qual\_Fin\_SF", "Half\_Bath", "Fireplaces", "Wood\_Deck\_SF", "Open\_Porch\_SF", "Enclosed\_Porch", "Screen\_Porch", "Neighborhood")

Build regression models with all of the variables.

ames\_glm\_model =   
 logistic\_reg() %>% #note the use of logistic\_reg  
 set\_engine("glm") #standard logistic regression engine is glm  
  
ames\_recipe = recipe(Above\_Median ~ Lot\_Shape + Overall\_Qual + Mas\_Vnr\_Type + Exter\_Qual + Foundation + Bsmt\_Qual + Heating\_QC + Kitchen\_Qual + Fireplace\_Qu + Garage\_Type + Garage\_Finish + Mas\_Vnr\_Area + Second\_Flr\_SF + Low\_Qual\_Fin\_SF + Half\_Bath + Fireplaces + Wood\_Deck\_SF + Open\_Porch\_SF + Enclosed\_Porch + Screen\_Porch + Neighborhood, new\_ames\_table) %>% #survived by p class  
 step\_dummy(all\_nominal(), -all\_outcomes()) #exclude the response variable from being dummy converted  
 # survived was changed to a factor which is required do not dummify the binary response variable  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_recipe) %>%   
 add\_model(ames\_glm\_model)  
  
ames\_fit = fit(logreg\_wf, new\_ames\_table)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(ames\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.0671 -0.1986 0.0000 0.2502 3.5275   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 2.026e+00 8.661e-01  
## Mas\_Vnr\_Area -2.056e-03 9.896e-04  
## Second\_Flr\_SF -1.655e-03 3.397e-04  
## Low\_Qual\_Fin\_SF 7.734e-04 2.502e-03  
## Half\_Bath 1.705e-01 2.221e-01  
## Fireplaces -9.419e-01 3.082e-01  
## Wood\_Deck\_SF -3.593e-03 7.762e-04  
## Open\_Porch\_SF -3.485e-03 1.313e-03  
## Enclosed\_Porch -8.456e-04 1.519e-03  
## Screen\_Porch -3.664e-03 1.417e-03  
## Lot\_Shape\_Regular 6.280e-01 2.064e-01  
## Lot\_Shape\_Moderately\_Irregular -1.979e-01 6.893e-01  
## Lot\_Shape\_Irregular -2.644e+00 1.409e+00  
## Overall\_Qual\_Average 8.103e-01 2.295e-01  
## Overall\_Qual\_Good -1.014e+00 2.718e-01  
## Overall\_Qual\_Very\_Good -3.334e+00 9.768e-01  
## Overall\_Qual\_Excellent 2.239e+00 1.300e+00  
## Overall\_Qual\_Below\_Average 1.341e+00 5.508e-01  
## Overall\_Qual\_Fair 1.581e+01 1.708e+03  
## Overall\_Qual\_Poor 1.423e+01 4.153e+03  
## Overall\_Qual\_Very\_Excellent -1.146e+01 1.399e+03  
## Overall\_Qual\_Very\_Poor 1.348e+01 6.788e+03  
## Mas\_Vnr\_Type\_None -7.225e-01 4.624e-01  
## Mas\_Vnr\_Type\_BrkFace -5.934e-01 4.308e-01  
## Mas\_Vnr\_Type\_BrkCmn -3.653e-01 8.666e-01  
## Mas\_Vnr\_Type\_CBlock 1.608e+01 1.075e+04  
## Exter\_Qual\_Good -6.310e-01 3.253e-01  
## Exter\_Qual\_Excellent -1.639e+01 9.603e+02  
## Exter\_Qual\_Fair -9.164e-02 1.198e+00  
## Foundation\_PConc -6.720e-01 2.959e-01  
## Foundation\_Wood 1.627e+00 2.708e+00  
## Foundation\_BrkTil 6.943e-01 3.997e-01  
## Foundation\_Slab -1.716e+01 1.802e+03  
## Foundation\_Stone 4.661e-01 1.613e+00  
## Bsmt\_Qual\_Good -1.672e-01 2.726e-01  
## Bsmt\_Qual\_Excellent -4.534e-01 7.809e-01  
## Bsmt\_Qual\_No\_Basement 1.729e+01 1.802e+03  
## Bsmt\_Qual\_Fair -8.840e-02 7.297e-01  
## Bsmt\_Qual\_Poor 1.477e+01 1.075e+04  
## Heating\_QC\_Typical 5.841e-01 5.377e-01  
## Heating\_QC\_Excellent 2.031e-01 5.395e-01  
## Heating\_QC\_Good -1.869e-01 5.601e-01  
## Heating\_QC\_Poor 1.280e+00 1.272e+04  
## Kitchen\_Qual\_Good -3.790e-01 2.442e-01  
## Kitchen\_Qual\_Excellent -1.544e+00 7.625e-01  
## Kitchen\_Qual\_Fair 1.396e+00 1.017e+00  
## Kitchen\_Qual\_Poor 1.709e+01 1.075e+04  
## Fireplace\_Qu\_No\_Fireplace 5.007e-01 4.400e-01  
## Fireplace\_Qu\_Typical -2.878e-01 2.816e-01  
## Fireplace\_Qu\_Poor 8.627e-01 6.159e-01  
## Fireplace\_Qu\_Excellent -1.287e+01 1.688e+03  
## Fireplace\_Qu\_Fair 6.387e-01 4.916e-01  
## Garage\_Type\_BuiltIn -7.525e-01 6.976e-01  
## Garage\_Type\_Basment 7.157e-01 5.936e-01  
## Garage\_Type\_Detchd 8.016e-01 2.886e-01  
## Garage\_Type\_No\_Garage 2.304e+01 1.075e+04  
## Garage\_Type\_CarPort 6.260e-01 1.440e+00  
## Garage\_Type\_More\_Than\_Two\_Types -1.021e+00 7.002e-01  
## Garage\_Finish\_Unf 7.678e-01 3.036e-01  
## Garage\_Finish\_RFn 1.659e-01 2.934e-01  
## Garage\_Finish\_No\_Garage -2.138e+01 1.075e+04  
## Neighborhood\_Gilbert -1.926e+00 7.099e-01  
## Neighborhood\_Stone\_Brook -1.599e+01 1.256e+03  
## Neighborhood\_Northwest\_Ames -1.391e+00 4.019e-01  
## Neighborhood\_Somerset -1.188e+00 5.729e-01  
## Neighborhood\_Briardale 1.704e+01 2.151e+03  
## Neighborhood\_Northpark\_Villa 1.887e+01 2.617e+03  
## Neighborhood\_Northridge\_Heights -1.301e+00 1.163e+00  
## Neighborhood\_Bloomington\_Heights -3.234e-01 9.246e-01  
## Neighborhood\_Northridge -1.366e+01 1.266e+03  
## Neighborhood\_Sawyer\_West -8.860e-01 5.349e-01  
## Neighborhood\_Sawyer 1.124e-02 4.031e-01  
## Neighborhood\_Greens 2.702e+00 1.583e+00  
## Neighborhood\_Old\_Town 1.050e+00 5.530e-01  
## Neighborhood\_Brookside -5.829e-01 5.331e-01  
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road 1.711e-01 8.339e-01  
## Neighborhood\_Clear\_Creek -1.185e+00 6.392e-01  
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -4.406e-01 6.310e-01  
## Neighborhood\_Edwards -1.879e-01 4.284e-01  
## Neighborhood\_College\_Creek -9.808e-01 4.446e-01  
## Neighborhood\_Crawford -1.704e+00 4.819e-01  
## Neighborhood\_Mitchell -1.280e+00 4.408e-01  
## Neighborhood\_Timberland -3.126e+00 1.114e+00  
## Neighborhood\_Meadow\_Village 1.744e+01 1.838e+03  
## Neighborhood\_Veenker -9.352e-01 1.213e+00  
## Neighborhood\_Blueste 1.897e+00 1.299e+00  
## Neighborhood\_Landmark 1.797e+01 1.075e+04  
## Neighborhood\_Green\_Hills -3.662e+01 1.090e+04  
## z value Pr(>|z|)   
## (Intercept) 2.340 0.019305 \*   
## Mas\_Vnr\_Area -2.078 0.037723 \*   
## Second\_Flr\_SF -4.871 1.11e-06 \*\*\*  
## Low\_Qual\_Fin\_SF 0.309 0.757267   
## Half\_Bath 0.768 0.442679   
## Fireplaces -3.056 0.002242 \*\*   
## Wood\_Deck\_SF -4.629 3.67e-06 \*\*\*  
## Open\_Porch\_SF -2.653 0.007970 \*\*   
## Enclosed\_Porch -0.557 0.577817   
## Screen\_Porch -2.586 0.009706 \*\*   
## Lot\_Shape\_Regular 3.043 0.002343 \*\*   
## Lot\_Shape\_Moderately\_Irregular -0.287 0.774009   
## Lot\_Shape\_Irregular -1.876 0.060691 .   
## Overall\_Qual\_Average 3.531 0.000414 \*\*\*  
## Overall\_Qual\_Good -3.732 0.000190 \*\*\*  
## Overall\_Qual\_Very\_Good -3.413 0.000643 \*\*\*  
## Overall\_Qual\_Excellent 1.723 0.084949 .   
## Overall\_Qual\_Below\_Average 2.435 0.014902 \*   
## Overall\_Qual\_Fair 0.009 0.992615   
## Overall\_Qual\_Poor 0.003 0.997267   
## Overall\_Qual\_Very\_Excellent -0.008 0.993469   
## Overall\_Qual\_Very\_Poor 0.002 0.998415   
## Mas\_Vnr\_Type\_None -1.562 0.118218   
## Mas\_Vnr\_Type\_BrkFace -1.377 0.168449   
## Mas\_Vnr\_Type\_BrkCmn -0.422 0.673327   
## Mas\_Vnr\_Type\_CBlock 0.001 0.998807   
## Exter\_Qual\_Good -1.940 0.052421 .   
## Exter\_Qual\_Excellent -0.017 0.986382   
## Exter\_Qual\_Fair -0.077 0.939020   
## Foundation\_PConc -2.271 0.023173 \*   
## Foundation\_Wood 0.601 0.547863   
## Foundation\_BrkTil 1.737 0.082373 .   
## Foundation\_Slab -0.010 0.992402   
## Foundation\_Stone 0.289 0.772671   
## Bsmt\_Qual\_Good -0.613 0.539775   
## Bsmt\_Qual\_Excellent -0.581 0.561472   
## Bsmt\_Qual\_No\_Basement 0.010 0.992345   
## Bsmt\_Qual\_Fair -0.121 0.903570   
## Bsmt\_Qual\_Poor 0.001 0.998905   
## Heating\_QC\_Typical 1.086 0.277317   
## Heating\_QC\_Excellent 0.376 0.706569   
## Heating\_QC\_Good -0.334 0.738589   
## Heating\_QC\_Poor 0.000 0.999920   
## Kitchen\_Qual\_Good -1.552 0.120669   
## Kitchen\_Qual\_Excellent -2.026 0.042803 \*   
## Kitchen\_Qual\_Fair 1.372 0.169939   
## Kitchen\_Qual\_Poor 0.002 0.998732   
## Fireplace\_Qu\_No\_Fireplace 1.138 0.255192   
## Fireplace\_Qu\_Typical -1.022 0.306706   
## Fireplace\_Qu\_Poor 1.401 0.161327   
## Fireplace\_Qu\_Excellent -0.008 0.993918   
## Fireplace\_Qu\_Fair 1.299 0.193818   
## Garage\_Type\_BuiltIn -1.079 0.280698   
## Garage\_Type\_Basment 1.206 0.227908   
## Garage\_Type\_Detchd 2.777 0.005482 \*\*   
## Garage\_Type\_No\_Garage 0.002 0.998291   
## Garage\_Type\_CarPort 0.435 0.663703   
## Garage\_Type\_More\_Than\_Two\_Types -1.459 0.144669   
## Garage\_Finish\_Unf 2.529 0.011436 \*   
## Garage\_Finish\_RFn 0.565 0.571871   
## Garage\_Finish\_No\_Garage -0.002 0.998414   
## Neighborhood\_Gilbert -2.714 0.006654 \*\*   
## Neighborhood\_Stone\_Brook -0.013 0.989837   
## Neighborhood\_Northwest\_Ames -3.460 0.000539 \*\*\*  
## Neighborhood\_Somerset -2.074 0.038114 \*   
## Neighborhood\_Briardale 0.008 0.993678   
## Neighborhood\_Northpark\_Villa 0.007 0.994246   
## Neighborhood\_Northridge\_Heights -1.119 0.263151   
## Neighborhood\_Bloomington\_Heights -0.350 0.726525   
## Neighborhood\_Northridge -0.011 0.991390   
## Neighborhood\_Sawyer\_West -1.656 0.097624 .   
## Neighborhood\_Sawyer 0.028 0.977763   
## Neighborhood\_Greens 1.707 0.087915 .   
## Neighborhood\_Old\_Town 1.899 0.057625 .   
## Neighborhood\_Brookside -1.093 0.274236   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road 0.205 0.837405   
## Neighborhood\_Clear\_Creek -1.854 0.063746 .   
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -0.698 0.484973   
## Neighborhood\_Edwards -0.439 0.660982   
## Neighborhood\_College\_Creek -2.206 0.027393 \*   
## Neighborhood\_Crawford -3.535 0.000407 \*\*\*  
## Neighborhood\_Mitchell -2.905 0.003673 \*\*   
## Neighborhood\_Timberland -2.806 0.005012 \*\*   
## Neighborhood\_Meadow\_Village 0.009 0.992430   
## Neighborhood\_Veenker -0.771 0.440631   
## Neighborhood\_Blueste 1.461 0.144099   
## Neighborhood\_Landmark 0.002 0.998667   
## Neighborhood\_Green\_Hills -0.003 0.997321   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2845.53 on 2052 degrees of freedom  
## Residual deviance: 864.55 on 1965 degrees of freedom  
## AIC: 1040.6  
##   
## Number of Fisher Scoring iterations: 18

The Second\_Flr\_SF variable is significant at 1.11e-06 because the p-value is less than 0.05. However, it has a negative coefficient of -1.655e-03.

The Wood\_Deck\_SF variable is significant at 3.67e-06 because the p-value is less than 0.05. However, it has a negative coefficient of -3.593e-03.

\*\*\* The Overall\_Qual\_Average variable is significant at 0.000414 because the p-value is less than 0.05. It has a positive coefficient of 8.103e-01, which is good.

The Overall\_Qual\_Good variable is significant at 0.000190 because the p-value is less than 0.05. However, it has a negative coefficient of -1.014e+00. This negative value is concerning.

The Overall\_Qual\_Very\_Good variable is significant at 0.000643 because the p-value is less than 0.05. However, it has a negative coefficient of -3.334e+00. This negative value is concerning.

Neighborhood\_Northwest\_Ames variable is significant at 0.000539 because the p-value is less than 0.05. However, it has a negative coefficient of -1.391e+00. This negative value is concerning.

Neighborhood\_Crawford variable is significant at 0.000407 because the p-value is less than 0.05. However, it has a negative coefficient of -1.704e+00. This negative value is concerning.

All these negative coefficients suggest multicollinearity.

Note the AIC of this model (a measure of model quality) is 1040.6 which actually improved as variables were added. We can use this value to compare this model to others. Smaller AIC is better.

car::vif(ames\_fit$fit$fit$fit)

## Mas\_Vnr\_Area   
## 2.555793e+00   
## Second\_Flr\_SF   
## 2.152845e+00   
## Low\_Qual\_Fin\_SF   
## 1.082709e+00   
## Half\_Bath   
## 1.538331e+00   
## Fireplaces   
## 5.679663e+00   
## Wood\_Deck\_SF   
## 1.261175e+00   
## Open\_Porch\_SF   
## 1.207949e+00   
## Enclosed\_Porch   
## 1.384365e+00   
## Screen\_Porch   
## 1.104261e+00   
## Lot\_Shape\_Regular   
## 1.241435e+00   
## Lot\_Shape\_Moderately\_Irregular   
## 1.122154e+00   
## Lot\_Shape\_Irregular   
## 1.153632e+00   
## Overall\_Qual\_Average   
## 1.509028e+00   
## Overall\_Qual\_Good   
## 1.552242e+00   
## Overall\_Qual\_Very\_Good   
## 1.940314e+00   
## Overall\_Qual\_Excellent   
## 1.442851e+00   
## Overall\_Qual\_Below\_Average   
## 1.204819e+00   
## Overall\_Qual\_Fair   
## 1.000000e+00   
## Overall\_Qual\_Poor   
## 1.000000e+00   
## Overall\_Qual\_Very\_Excellent   
## 1.000002e+00   
## Overall\_Qual\_Very\_Poor   
## 1.398453e+00   
## Mas\_Vnr\_Type\_None   
## 6.761182e+00   
## Mas\_Vnr\_Type\_BrkFace   
## 5.405881e+00   
## Mas\_Vnr\_Type\_BrkCmn   
## 1.632367e+00   
## Mas\_Vnr\_Type\_CBlock   
## 1.000000e+00   
## Exter\_Qual\_Good   
## 2.440889e+00   
## Exter\_Qual\_Excellent   
## 1.000001e+00   
## Exter\_Qual\_Fair   
## 1.133554e+00   
## Foundation\_PConc   
## 2.555691e+00   
## Foundation\_Wood   
## 1.059724e+00   
## Foundation\_BrkTil   
## 1.906422e+00   
## Foundation\_Slab   
## 3.564583e+06   
## Foundation\_Stone   
## 1.126284e+00   
## Bsmt\_Qual\_Good   
## 2.339657e+00   
## Bsmt\_Qual\_Excellent   
## 1.479288e+00   
## Bsmt\_Qual\_No\_Basement   
## 3.564583e+06   
## Bsmt\_Qual\_Fair   
## 1.071110e+00   
## Bsmt\_Qual\_Poor   
## 1.000000e+00   
## Heating\_QC\_Typical   
## 8.239643e+00   
## Heating\_QC\_Excellent   
## 9.433452e+00   
## Heating\_QC\_Good   
## 6.215594e+00   
## Heating\_QC\_Poor   
## 1.398453e+00   
## Kitchen\_Qual\_Good   
## 1.787242e+00   
## Kitchen\_Qual\_Excellent   
## 1.297980e+00   
## Kitchen\_Qual\_Fair   
## 1.075822e+00   
## Kitchen\_Qual\_Poor   
## 1.000000e+00   
## Fireplace\_Qu\_No\_Fireplace   
## 6.262248e+00   
## Fireplace\_Qu\_Typical   
## 1.725582e+00   
## Fireplace\_Qu\_Poor   
## 1.248429e+00   
## Fireplace\_Qu\_Excellent   
## 1.000000e+00   
## Fireplace\_Qu\_Fair   
## 1.299527e+00   
## Garage\_Type\_BuiltIn   
## 1.184588e+00   
## Garage\_Type\_Basment   
## 1.150363e+00   
## Garage\_Type\_Detchd   
## 1.986410e+00   
## Garage\_Type\_No\_Garage   
## 3.253545e+08   
## Garage\_Type\_CarPort   
## 1.266919e+00   
## Garage\_Type\_More\_Than\_Two\_Types   
## 1.189044e+00   
## Garage\_Finish\_Unf   
## 3.007889e+00   
## Garage\_Finish\_RFn   
## 2.397169e+00   
## Garage\_Finish\_No\_Garage   
## 3.253545e+08   
## Neighborhood\_Gilbert   
## 1.306263e+00   
## Neighborhood\_Stone\_Brook   
## 1.000001e+00   
## Neighborhood\_Northwest\_Ames   
## 1.425282e+00   
## Neighborhood\_Somerset   
## 2.072236e+00   
## Neighborhood\_Briardale   
## 1.000000e+00   
## Neighborhood\_Northpark\_Villa   
## 1.000000e+00   
## Neighborhood\_Northridge\_Heights   
## 1.252132e+00   
## Neighborhood\_Bloomington\_Heights   
## 1.411072e+00   
## Neighborhood\_Northridge   
## 1.000000e+00   
## Neighborhood\_Sawyer\_West   
## 1.705813e+00   
## Neighborhood\_Sawyer   
## 1.405445e+00   
## Neighborhood\_Greens   
## 1.945244e+00   
## Neighborhood\_Old\_Town   
## 2.136953e+00   
## Neighborhood\_Brookside   
## 1.598501e+00   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road   
## 1.175328e+00   
## Neighborhood\_Clear\_Creek   
## 1.214497e+00   
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University   
## 1.572431e+00   
## Neighborhood\_Edwards   
## 1.528827e+00   
## Neighborhood\_College\_Creek   
## 2.196606e+00   
## Neighborhood\_Crawford   
## 1.866130e+00   
## Neighborhood\_Mitchell   
## 1.667974e+00   
## Neighborhood\_Timberland   
## 1.103904e+00   
## Neighborhood\_Meadow\_Village   
## 1.000000e+00   
## Neighborhood\_Veenker   
## 1.134342e+00   
## Neighborhood\_Blueste   
## 1.193581e+00   
## Neighborhood\_Landmark   
## 1.000000e+00   
## Neighborhood\_Green\_Hills   
## 1.028086e+00

In general, seeing variables with VIF values greater than 4 indicates the presence of multicollinearity. There were almost all the variables with large VIF values.

Next, build smaller regression models with subsets of the variables to look for more clues of multicollinearity

ames\_glm\_model =   
 logistic\_reg() %>% #note the use of logistic\_reg  
 set\_engine("glm") #standard logistic regression engine is glm  
  
ames\_recipe = recipe(Above\_Median ~ Overall\_Qual + Heating\_QC + Kitchen\_Qual + Fireplace\_Qu + Second\_Flr\_SF + Low\_Qual\_Fin\_SF + Half\_Bath + Fireplaces, new\_ames\_table) %>% #survived by p class  
 step\_dummy(all\_nominal(), -all\_outcomes()) #exclude the response variable from being dummy converted  
 # survived was changed to a factor which is required do not dummify the binary response variable  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_recipe) %>%   
 add\_model(ames\_glm\_model)  
  
ames\_fit = fit(logreg\_wf, new\_ames\_table)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(ames\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.74328 -0.40324 -0.00018 0.42241 2.97553   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.161e+00 5.386e-01 2.156 0.031080 \*   
## Second\_Flr\_SF -5.071e-04 2.262e-04 -2.242 0.024982 \*   
## Low\_Qual\_Fin\_SF 1.670e-03 1.819e-03 0.918 0.358442   
## Half\_Bath -5.001e-01 1.699e-01 -2.944 0.003237 \*\*   
## Fireplaces -9.760e-01 2.629e-01 -3.713 0.000205 \*\*\*  
## Overall\_Qual\_Average 1.320e+00 1.719e-01 7.681 1.58e-14 \*\*\*  
## Overall\_Qual\_Good -1.484e+00 2.045e-01 -7.253 4.08e-13 \*\*\*  
## Overall\_Qual\_Very\_Good -3.430e+00 6.054e-01 -5.665 1.47e-08 \*\*\*  
## Overall\_Qual\_Excellent -2.582e+00 1.139e+00 -2.266 0.023430 \*   
## Overall\_Qual\_Below\_Average 2.228e+00 4.269e-01 5.219 1.80e-07 \*\*\*  
## Overall\_Qual\_Fair 1.598e+01 7.060e+02 0.023 0.981936   
## Overall\_Qual\_Poor 1.566e+01 1.564e+03 0.010 0.992013   
## Overall\_Qual\_Very\_Excellent -1.527e+01 7.581e+02 -0.020 0.983932   
## Overall\_Qual\_Very\_Poor 1.534e+01 2.738e+03 0.006 0.995529   
## Heating\_QC\_Typical 8.766e-01 4.430e-01 1.979 0.047829 \*   
## Heating\_QC\_Excellent -1.042e-01 4.415e-01 -0.236 0.813399   
## Heating\_QC\_Good 8.933e-02 4.533e-01 0.197 0.843766   
## Heating\_QC\_Poor -9.956e-01 4.811e+03 0.000 0.999835   
## Kitchen\_Qual\_Good -1.066e+00 1.703e-01 -6.261 3.84e-10 \*\*\*  
## Kitchen\_Qual\_Excellent -1.322e+00 5.758e-01 -2.296 0.021698 \*   
## Kitchen\_Qual\_Fair 1.798e+00 8.658e-01 2.076 0.037866 \*   
## Kitchen\_Qual\_Poor 1.549e+01 3.956e+03 0.004 0.996876   
## Fireplace\_Qu\_No\_Fireplace 2.608e-01 3.644e-01 0.716 0.474146   
## Fireplace\_Qu\_Typical -8.731e-01 2.165e-01 -4.033 5.52e-05 \*\*\*  
## Fireplace\_Qu\_Poor 4.900e-01 4.789e-01 1.023 0.306219   
## Fireplace\_Qu\_Excellent -1.322e+01 7.970e+02 -0.017 0.986762   
## Fireplace\_Qu\_Fair 1.314e-01 3.947e-01 0.333 0.739237   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2845.5 on 2052 degrees of freedom  
## Residual deviance: 1272.1 on 2026 degrees of freedom  
## AIC: 1326.1  
##   
## Number of Fisher Scoring iterations: 16

\*\*\* The Overall\_Qual\_Average variable is significant at 1.58e-14 because the p-value is less than 0.05. It has a positive coefficient of 1.320e+00, which is still good.

The Overall\_Qual\_Good variable is significant at 4.08e-13 because the p-value is less than 0.05. However, it has a negative coefficient of -1.484e+00. This negative value is concerning.

The Overall\_Qual\_Very\_Good variable is significant at 1.47e-08 because the p-value is less than 0.05. However, it has a negative coefficient of -3.430e+00. This negative value is concerning.

\*\*\*The Overall\_Qual\_Below\_Average variable is significant at 1.80e-07 because the p-value is less than 0.05. It has a positive coefficient of 2.228e+00, which is now good.

The Kitchen\_Qual\_Good variable is significant at 3.84e-10 because the p-value is less than 0.05. However, it has a negative coefficient of -1.066e+00. This negative value is concerning.

The Fireplace\_Qu\_Typical variable is significant at 5.52e-05 because the p-value is less than 0.05. However, it has a negative coefficient of -8.731e-01. This negative value is concerning.

Note the AIC of this model (a measure of model quality) is now 1326.1 compared with the original 1040.6 which actually worsened in this subset. We can use this value to compare this model to others. Smaller AIC is better.

car::vif(ames\_fit$fit$fit$fit)

## Second\_Flr\_SF Low\_Qual\_Fin\_SF   
## 1.462120 1.012381   
## Half\_Bath Fireplaces   
## 1.430206 5.668882   
## Overall\_Qual\_Average Overall\_Qual\_Good   
## 1.198521 1.253940   
## Overall\_Qual\_Very\_Good Overall\_Qual\_Excellent   
## 1.045688 1.259540   
## Overall\_Qual\_Below\_Average Overall\_Qual\_Fair   
## 1.042787 1.000000   
## Overall\_Qual\_Poor Overall\_Qual\_Very\_Excellent   
## 1.000000 1.000000   
## Overall\_Qual\_Very\_Poor Heating\_QC\_Typical   
## 1.478980 7.972006   
## Heating\_QC\_Excellent Heating\_QC\_Good   
## 9.582701 6.542116   
## Heating\_QC\_Poor Kitchen\_Qual\_Good   
## 1.478980 1.333457   
## Kitchen\_Qual\_Excellent Kitchen\_Qual\_Fair   
## 1.289424 1.021831   
## Kitchen\_Qual\_Poor Fireplace\_Qu\_No\_Fireplace   
## 1.000000 6.517486   
## Fireplace\_Qu\_Typical Fireplace\_Qu\_Poor   
## 1.575892 1.117121   
## Fireplace\_Qu\_Excellent Fireplace\_Qu\_Fair   
## 1.000000 1.172778

In general, seeing variables with VIF values greater than 4 indicates the presence of multicollinearity. Initially, almost all the variables had large VIF values.

Now, Second\_Flr\_SF has 1.462120, Low\_Qual\_Fin\_SF has 1.012381, and Half\_Bath has 1.430206. These show no multicollinearity but are not significant above.

\*\*\* Overall\_Qual wit its 9 Levels has a VIF range from 1.000000 to 1.478980. These show no multicollinearity, and are significant.

Heating\_Poor has 1.478980 but the other 3 levels are above a 4. This suggests that it isn’t a good variable to predict Above\_Median.

\*\*\* Kitchen\_Qual has 4 levels ranging from 1.000000 to 1.333457. These show no multicollinearity and are significant.

\*\* Fireplace\_Qu has 5 levels with 4 levels ranging from 1.000000 to 1.575892 (Fireplace\_Qu\_No\_Fireplace had 6.517486). Fireplace\_Qu might be a good variable to predict Above\_Median.

Fireplaces was high at 5.668882. This suggests that it isn’t a good variable to predict Above\_Median.

ames\_glm\_model =   
 logistic\_reg() %>% #note the use of logistic\_reg  
 set\_engine("glm") #standard logistic regression engine is glm  
  
ames\_recipe = recipe(Above\_Median ~ Mas\_Vnr\_Area + Mas\_Vnr\_Type + Lot\_Shape + Neighborhood, new\_ames\_table) %>% #survived by p class  
 step\_dummy(all\_nominal(), -all\_outcomes()) #exclude the response variable from being dummy converted  
 # survived was changed to a factor which is required do not dummify the binary response variable  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_recipe) %>%   
 add\_model(ames\_glm\_model)  
  
ames\_fit = fit(logreg\_wf, new\_ames\_table)

summary(ames\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3777 -0.5099 -0.0001 0.5160 3.3639   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 8.409e-01 3.644e-01  
## Mas\_Vnr\_Area -2.696e-03 7.844e-04  
## Mas\_Vnr\_Type\_None 5.302e-01 3.411e-01  
## Mas\_Vnr\_Type\_BrkFace 2.006e-01 3.128e-01  
## Mas\_Vnr\_Type\_BrkCmn 5.562e-01 7.069e-01  
## Mas\_Vnr\_Type\_CBlock 1.796e+01 3.956e+03  
## Lot\_Shape\_Regular 5.782e-01 1.448e-01  
## Lot\_Shape\_Moderately\_Irregular -1.101e+00 5.151e-01  
## Lot\_Shape\_Irregular -1.639e+00 9.911e-01  
## Neighborhood\_Gilbert -4.890e+00 6.081e-01  
## Neighborhood\_Stone\_Brook -1.839e+01 5.931e+02  
## Neighborhood\_Northwest\_Ames -2.458e+00 3.004e-01  
## Neighborhood\_Somerset -3.882e+00 3.829e-01  
## Neighborhood\_Briardale 1.691e+01 8.363e+02  
## Neighborhood\_Northpark\_Villa 1.571e+01 1.017e+03  
## Neighborhood\_Northridge\_Heights -5.116e+00 1.023e+00  
## Neighborhood\_Bloomington\_Heights -3.778e+00 7.774e-01  
## Neighborhood\_Northridge -1.790e+01 5.306e+02  
## Neighborhood\_Sawyer\_West -2.257e+00 2.879e-01  
## Neighborhood\_Sawyer 3.366e-01 3.069e-01  
## Neighborhood\_Greens -3.163e+00 1.096e+00  
## Neighborhood\_Old\_Town 5.265e-01 3.033e-01  
## Neighborhood\_Brookside 1.176e-01 3.768e-01  
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road 8.163e-01 5.495e-01  
## Neighborhood\_Clear\_Creek -2.468e+00 4.686e-01  
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -6.912e-01 4.205e-01  
## Neighborhood\_Edwards 1.753e-01 2.940e-01  
## Neighborhood\_College\_Creek -2.245e+00 2.252e-01  
## Neighborhood\_Crawford -2.378e+00 3.061e-01  
## Neighborhood\_Mitchell -1.280e+00 2.779e-01  
## Neighborhood\_Timberland -5.103e+00 1.023e+00  
## Neighborhood\_Meadow\_Village 1.569e+01 8.036e+02  
## Neighborhood\_Veenker -3.865e+00 1.058e+00  
## Neighborhood\_Blueste -4.314e-01 1.135e+00  
## Neighborhood\_Landmark 1.619e+01 3.956e+03  
## Neighborhood\_Green\_Hills -1.894e+01 3.956e+03  
## z value Pr(>|z|)   
## (Intercept) 2.308 0.021024 \*   
## Mas\_Vnr\_Area -3.438 0.000587 \*\*\*  
## Mas\_Vnr\_Type\_None 1.555 0.120042   
## Mas\_Vnr\_Type\_BrkFace 0.641 0.521349   
## Mas\_Vnr\_Type\_BrkCmn 0.787 0.431390   
## Mas\_Vnr\_Type\_CBlock 0.005 0.996378   
## Lot\_Shape\_Regular 3.993 6.52e-05 \*\*\*  
## Lot\_Shape\_Moderately\_Irregular -2.137 0.032621 \*   
## Lot\_Shape\_Irregular -1.653 0.098230 .   
## Neighborhood\_Gilbert -8.041 8.90e-16 \*\*\*  
## Neighborhood\_Stone\_Brook -0.031 0.975260   
## Neighborhood\_Northwest\_Ames -8.183 2.78e-16 \*\*\*  
## Neighborhood\_Somerset -10.138 < 2e-16 \*\*\*  
## Neighborhood\_Briardale 0.020 0.983865   
## Neighborhood\_Northpark\_Villa 0.015 0.987681   
## Neighborhood\_Northridge\_Heights -5.000 5.74e-07 \*\*\*  
## Neighborhood\_Bloomington\_Heights -4.860 1.17e-06 \*\*\*  
## Neighborhood\_Northridge -0.034 0.973095   
## Neighborhood\_Sawyer\_West -7.841 4.47e-15 \*\*\*  
## Neighborhood\_Sawyer 1.097 0.272769   
## Neighborhood\_Greens -2.885 0.003917 \*\*   
## Neighborhood\_Old\_Town 1.736 0.082611 .   
## Neighborhood\_Brookside 0.312 0.755023   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road 1.486 0.137376   
## Neighborhood\_Clear\_Creek -5.268 1.38e-07 \*\*\*  
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -1.644 0.100243   
## Neighborhood\_Edwards 0.596 0.551140   
## Neighborhood\_College\_Creek -9.969 < 2e-16 \*\*\*  
## Neighborhood\_Crawford -7.768 7.96e-15 \*\*\*  
## Neighborhood\_Mitchell -4.608 4.07e-06 \*\*\*  
## Neighborhood\_Timberland -4.989 6.06e-07 \*\*\*  
## Neighborhood\_Meadow\_Village 0.020 0.984424   
## Neighborhood\_Veenker -3.654 0.000259 \*\*\*  
## Neighborhood\_Blueste -0.380 0.704019   
## Neighborhood\_Landmark 0.004 0.996734   
## Neighborhood\_Green\_Hills -0.005 0.996181   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2845.5 on 2052 degrees of freedom  
## Residual deviance: 1488.9 on 2017 degrees of freedom  
## AIC: 1560.9  
##   
## Number of Fisher Scoring iterations: 16

\*\*\* The Mas\_Vnr\_Area variable is significant at 0.000587 because the p-value is less than 0.05. However, it has a negative coefficient of -2.696e-03. This negative value is concerning.

Mas\_Vnr\_Type was not significant.

\*\*\* The Lot\_Shape\_Regular variable is significant at 6.52e-05 because the p-value is less than 0.05. However, and it has a positive coefficient of 5.782e-01, which is good. However, the other 2 levels are not significant. This negative value is concerning.

The Neighborhood\_Gilbert, Neighborhood\_Northwest\_Ames, Neighborhood\_Somerset, Neighborhood\_Northridge\_Heights, Neighborhood\_Bloomington\_Heights , Neighborhood\_Sawyer\_West, Neighborhood\_Clear\_Creek, Neighborhood\_College\_Creek, Neighborhood\_Crawford, Neighborhood\_Mitchell, Neighborhood\_Timberland , and Neighborhood\_Veenker are variables that are significant because the p-value is less than 0.05. However, those 12/28 levels have a negative coefficient. The negative coefficients suggest multicollinearity.

Note the AIC of this model (a measure of model quality) is now 1560.9 compared with the original 1040.6 which actually worsened in this subset. We can use this value to compare this model to others. Smaller AIC is better.

car::vif(ames\_fit$fit$fit$fit)

## Mas\_Vnr\_Area   
## 2.407828   
## Mas\_Vnr\_Type\_None   
## 6.329171   
## Mas\_Vnr\_Type\_BrkFace   
## 4.837955   
## Mas\_Vnr\_Type\_BrkCmn   
## 1.202629   
## Mas\_Vnr\_Type\_CBlock   
## 1.000000   
## Lot\_Shape\_Regular   
## 1.122660   
## Lot\_Shape\_Moderately\_Irregular   
## 1.053955   
## Lot\_Shape\_Irregular   
## 1.045948   
## Neighborhood\_Gilbert   
## 1.058874   
## Neighborhood\_Stone\_Brook   
## 1.000000   
## Neighborhood\_Northwest\_Ames   
## 1.215290   
## Neighborhood\_Somerset   
## 1.148655   
## Neighborhood\_Briardale   
## 1.000000   
## Neighborhood\_Northpark\_Villa   
## 1.000000   
## Neighborhood\_Northridge\_Heights   
## 1.028219   
## Neighborhood\_Bloomington\_Heights   
## 1.084332   
## Neighborhood\_Northridge   
## 1.000000   
## Neighborhood\_Sawyer\_West   
## 1.257505   
## Neighborhood\_Sawyer   
## 1.212723   
## Neighborhood\_Greens   
## 1.026631   
## Neighborhood\_Old\_Town   
## 1.308198   
## Neighborhood\_Brookside   
## 1.161033   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road   
## 1.075715   
## Neighborhood\_Clear\_Creek   
## 1.086422   
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University   
## 1.124706   
## Neighborhood\_Edwards   
## 1.265560   
## Neighborhood\_College\_Creek   
## 1.456718   
## Neighborhood\_Crawford   
## 1.244205   
## Neighborhood\_Mitchell   
## 1.278913   
## Neighborhood\_Timberland   
## 1.016768   
## Neighborhood\_Meadow\_Village   
## 1.000000   
## Neighborhood\_Veenker   
## 1.017331   
## Neighborhood\_Blueste   
## 1.017431   
## Neighborhood\_Landmark   
## 1.000000   
## Neighborhood\_Green\_Hills   
## 1.000000

Only Mas\_Vnr\_Type\_None and Mas\_Vnr\_Type\_BrkFace have VIF values over 4. They were not significant and therefore Mas\_Vnr\_Type is not a good variable to predict Above\_Median.

\*\*Mas\_Vnr\_Area had a VIR score of 2.407828 and was significant with a negative coefficient. Mas\_Vnr\_Area might be a good variable to predict Above\_Median.

\*\*Lot\_Shape had a VIR score between 1.053955 to 1.122660 and had 2 out of 3 variables significant with negative coefficients.Lot\_Shape might be a good variable to predict Above\_Median.

The neighborhood variable is a mixed bag with less than 50% significance and low VIF values. It’s probably not a goo variable to predict Above\_Median.

ames\_glm\_model =   
 logistic\_reg() %>% #note the use of logistic\_reg  
 set\_engine("glm") #standard logistic regression engine is glm  
  
ames\_recipe = recipe(Above\_Median ~ Wood\_Deck\_SF + Exter\_Qual + Foundation + Bsmt\_Qual + Garage\_Type + Garage\_Finish + Open\_Porch\_SF + Enclosed\_Porch + Screen\_Porch, new\_ames\_table) %>% #survived by p class  
 step\_dummy(all\_nominal(), -all\_outcomes()) #exclude the response variable from being dummy converted  
 # survived was changed to a factor which is required do not dummify the binary response variable  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_recipe) %>%   
 add\_model(ames\_glm\_model)  
  
ames\_fit = fit(logreg\_wf, new\_ames\_table)

summary(ames\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1654 -0.3739 -0.0001 0.5307 2.7356   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 9.932e-01 2.125e-01 4.673 2.97e-06 \*\*\*  
## Wood\_Deck\_SF -3.251e-03 5.606e-04 -5.800 6.63e-09 \*\*\*  
## Open\_Porch\_SF -4.944e-03 9.816e-04 -5.037 4.73e-07 \*\*\*  
## Enclosed\_Porch -2.336e-03 1.078e-03 -2.167 0.0303 \*   
## Screen\_Porch -6.161e-03 1.142e-03 -5.396 6.82e-08 \*\*\*  
## Exter\_Qual\_Good -1.595e+00 1.864e-01 -8.560 < 2e-16 \*\*\*  
## Exter\_Qual\_Excellent -1.560e+01 4.172e+02 -0.037 0.9702   
## Exter\_Qual\_Fair 8.519e-01 1.126e+00 0.757 0.4492   
## Foundation\_PConc -9.967e-01 1.805e-01 -5.522 3.34e-08 \*\*\*  
## Foundation\_Wood -1.211e-01 1.117e+00 -0.108 0.9136   
## Foundation\_BrkTil -4.014e-01 2.580e-01 -1.556 0.1197   
## Foundation\_Slab -1.896e+00 1.270e+00 -1.493 0.1355   
## Foundation\_Stone 7.101e-01 1.274e+00 0.557 0.5773   
## Bsmt\_Qual\_Good -8.188e-01 1.685e-01 -4.858 1.18e-06 \*\*\*  
## Bsmt\_Qual\_Excellent -1.136e+00 5.195e-01 -2.186 0.0288 \*   
## Bsmt\_Qual\_No\_Basement 2.231e+00 1.113e+00 2.004 0.0451 \*   
## Bsmt\_Qual\_Fair 6.174e-01 5.595e-01 1.104 0.2698   
## Bsmt\_Qual\_Poor 1.534e+01 3.956e+03 0.004 0.9969   
## Garage\_Type\_BuiltIn -7.292e-01 3.880e-01 -1.880 0.0602 .   
## Garage\_Type\_Basment 7.214e-01 4.522e-01 1.595 0.1106   
## Garage\_Type\_Detchd 1.329e+00 1.987e-01 6.690 2.23e-11 \*\*\*  
## Garage\_Type\_No\_Garage 2.194e+01 3.956e+03 0.006 0.9956   
## Garage\_Type\_CarPort 2.025e+00 1.191e+00 1.701 0.0890 .   
## Garage\_Type\_More\_Than\_Two\_Types 3.875e-02 5.527e-01 0.070 0.9441   
## Garage\_Finish\_Unf 8.957e-01 2.078e-01 4.310 1.63e-05 \*\*\*  
## Garage\_Finish\_RFn 3.595e-01 2.045e-01 1.758 0.0788 .   
## Garage\_Finish\_No\_Garage -1.826e+01 3.956e+03 -0.005 0.9963   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2845.5 on 2052 degrees of freedom  
## Residual deviance: 1405.6 on 2026 degrees of freedom  
## AIC: 1459.6  
##   
## Number of Fisher Scoring iterations: 16

Wood\_Deck\_SF variable is significant at 6.63e-09because the p-value is less than 0.05. However, it has a negative coefficient of -3.251e-03. This negative value is concerning.

Open\_Porch\_SF variable is significant at 4.73e-07 because the p-value is less than 0.05. However, it has a negative coefficient of -4.944e-03. This negative value is concerning.

Screen\_Porch variable is significant at 6.82e-08 because the p-value is less than 0.05. However, it has a negative coefficient of -6.161e-03-6.161e-03. This negative value is concerning.

Exter\_Qual\_Good variable is significant at 1.63e-05 because the p-value is less than 0.05. However, it has a negative coefficient of -1.595e+00. This negative value is concerning. It was only 1 out of 4 levels.

Foundation\_PConc variable is significant at < 2e-16 because the p-value is less than 0.05. However, it has a negative coefficient of -9.967e-01. This negative value is concerning. It was 1 out of 6 levels.

Bsmt\_Qual\_Good variable is significant at 1.18e-06 because the p-value is less than 0.05. However, it has a negative coefficient of -8.188e-01. This negative value is concerning. It was 1 out of 6 levels.

\*\*\* The Garage\_Type\_Detchd variable is significant at 2.23e-11 because the p-value is less than 0.05. It has a positive coefficient of 1.329e+00, which is good. However, only 1 out of 7 levels were significant. This is probably not a good variable to predict Above\_Median.

Garage\_Finish\_Unf variable is significant at 1.63e-05 because the p-value is less than 0.05. However, it has a negative coefficient of 8.957e-01. This negative value is concerning. It was 1 out of 4 levels.

All these negative coefficients suggest multicollinearity.

Note the AIC of this model (a measure of model quality) is now 1459.6 compared with the original 1040.6 which actually worsened in this subset. We can use this value to compare this model to others. Smaller AIC is better.

car::vif(ames\_fit$fit$fit$fit)

## Wood\_Deck\_SF Open\_Porch\_SF   
## 1.071776e+00 1.047369e+00   
## Enclosed\_Porch Screen\_Porch   
## 1.173383e+00 1.059621e+00   
## Exter\_Qual\_Good Exter\_Qual\_Excellent   
## 1.276398e+00 1.000000e+00   
## Exter\_Qual\_Fair Foundation\_PConc   
## 1.015388e+00 1.565108e+00   
## Foundation\_Wood Foundation\_BrkTil   
## 1.045992e+00 1.353110e+00   
## Foundation\_Slab Foundation\_Stone   
## 4.299376e+00 1.045171e+00   
## Bsmt\_Qual\_Good Bsmt\_Qual\_Excellent   
## 1.489286e+00 1.135606e+00   
## Bsmt\_Qual\_No\_Basement Bsmt\_Qual\_Fair   
## 4.302472e+00 1.036922e+00   
## Bsmt\_Qual\_Poor Garage\_Type\_BuiltIn   
## 1.000000e+00 1.052301e+00   
## Garage\_Type\_Basment Garage\_Type\_Detchd   
## 1.057028e+00 1.571095e+00   
## Garage\_Type\_No\_Garage Garage\_Type\_CarPort   
## 5.142446e+07 1.017579e+00   
## Garage\_Type\_More\_Than\_Two\_Types Garage\_Finish\_Unf   
## 1.037057e+00 2.349633e+00   
## Garage\_Finish\_RFn Garage\_Finish\_No\_Garage   
## 1.953484e+00 5.142446e+07

In general, seeing variables with VIF values greater than 4 indicates the presence of multicollinearity. There were almost all the variables with large VIF values. There was only 1 variable level with a positive coefficient so this model did not produce any viable predictors for Above\_Median.

At the end of analyzing the three subsets, Overall\_Qu and Kitchen\_Qu are the best predictors with Fireplace\_Qu, Lot\_Shape, and Mas\_Vnr\_Area strong Maybe’s.

Testing our conclusion with a final GLM model:

ames\_glm\_model =   
 logistic\_reg() %>% #note the use of logistic\_reg  
 set\_engine("glm") #standard logistic regression engine is glm  
  
ames\_recipe = recipe(Above\_Median ~ Overall\_Qual + Kitchen\_Qual + Lot\_Shape + Mas\_Vnr\_Area, new\_ames\_table) %>% #survived by p class  
 step\_dummy(all\_nominal(), -all\_outcomes()) #exclude the response variable from being dummy converted  
 # survived was changed to a factor which is required do not dummify the binary response variable  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_recipe) %>%   
 add\_model(ames\_glm\_model)  
  
ames\_fit = fit(logreg\_wf, new\_ames\_table)

summary(ames\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.7156 -0.4736 -0.0003 0.4786 3.4782   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.172e-01 1.497e-01 -0.783 0.4337   
## Mas\_Vnr\_Area -2.300e-03 4.914e-04 -4.680 2.87e-06 \*\*\*  
## Overall\_Qual\_Average 1.516e+00 1.568e-01 9.670 < 2e-16 \*\*\*  
## Overall\_Qual\_Good -1.636e+00 1.930e-01 -8.479 < 2e-16 \*\*\*  
## Overall\_Qual\_Very\_Good -3.356e+00 5.994e-01 -5.599 2.16e-08 \*\*\*  
## Overall\_Qual\_Excellent -2.387e+00 1.118e+00 -2.136 0.0327 \*   
## Overall\_Qual\_Below\_Average 2.562e+00 4.055e-01 6.318 2.64e-10 \*\*\*  
## Overall\_Qual\_Fair 1.677e+01 7.241e+02 0.023 0.9815   
## Overall\_Qual\_Poor 1.642e+01 1.601e+03 0.010 0.9918   
## Overall\_Qual\_Very\_Excellent -1.563e+01 7.360e+02 -0.021 0.9831   
## Overall\_Qual\_Very\_Poor 1.625e+01 2.235e+03 0.007 0.9942   
## Kitchen\_Qual\_Good -1.127e+00 1.489e-01 -7.569 3.76e-14 \*\*\*  
## Kitchen\_Qual\_Excellent -1.465e+00 5.420e-01 -2.702 0.0069 \*\*   
## Kitchen\_Qual\_Fair 1.139e+00 7.735e-01 1.473 0.1409   
## Kitchen\_Qual\_Poor 1.617e+01 3.956e+03 0.004 0.9967   
## Lot\_Shape\_Regular 1.124e+00 1.435e-01 7.835 4.70e-15 \*\*\*  
## Lot\_Shape\_Moderately\_Irregular -6.511e-01 5.195e-01 -1.253 0.2101   
## Lot\_Shape\_Irregular -5.293e-01 9.482e-01 -0.558 0.5767   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2845.5 on 2052 degrees of freedom  
## Residual deviance: 1440.9 on 2035 degrees of freedom  
## AIC: 1476.9  
##   
## Number of Fisher Scoring iterations: 16

The Mas\_Vnr\_Area variable is significant at 2.87e-06 because the p-value is less than 0.05. However, it has a negative coefficient of -2.300e-03 which doesn’t make sense.

\*\*\* The Overall\_Qual\_Average variable is significant at < 2e-16 because the p-value is less than 0.05. It has a positive coefficient of 1.516e+00, which is good.

The Overall\_Qual\_Good variable is significant at < 2e-16 because the p-value is less than 0.05. However, it has a negative coefficient of -1.636e+00 which doesn’t make sense.

Overall\_Qual\_Very\_Good variable is significant at 2.16e-08 because the p-value is less than 0.05. However, it has a negative coefficient of -3.356e+00 which doesn’t make sense.

The Overall\_Qual\_Below\_Average variable is significant at 2.64e-10 because the p-value is less than 0.05. It has a positive coefficient of 2.562e+00, which is doesn’t make sense.

The Kitchen\_Qual\_Good variable is significant at 3.76e-14 because the p-value is less than 0.05. However, it has a negative coefficient of -1.127e+00 which doesn’t make sense.

\*\*\* The Lot\_Shape\_Regular variable is significant at 4.70e-15 because the p-value is less than 0.05. It has a positive coefficient of 1.124e+00 , which is good.

All these logically opposite coefficients suggest multicollinearity.

Note the AIC of this model (a measure of model quality) is now 1476.9 compared with the original 1040.6 which actually worsened in this subset. We can use this value to compare this model to others. Smaller AIC is better.

car::vif(ames\_fit$fit$fit$fit)

## Mas\_Vnr\_Area Overall\_Qual\_Average   
## 1.050362 1.162503   
## Overall\_Qual\_Good Overall\_Qual\_Very\_Good   
## 1.201851 1.041456   
## Overall\_Qual\_Excellent Overall\_Qual\_Below\_Average   
## 1.216158 1.041567   
## Overall\_Qual\_Fair Overall\_Qual\_Poor   
## 1.000000 1.000000   
## Overall\_Qual\_Very\_Excellent Overall\_Qual\_Very\_Poor   
## 1.000000 1.000000   
## Kitchen\_Qual\_Good Kitchen\_Qual\_Excellent   
## 1.133535 1.223053   
## Kitchen\_Qual\_Fair Kitchen\_Qual\_Poor   
## 1.009608 1.000000   
## Lot\_Shape\_Regular Lot\_Shape\_Moderately\_Irregular   
## 1.068087 1.037861   
## Lot\_Shape\_Irregular   
## 1.014007

The VIF values show no multicollinearity.

This leaves Overall\_Qual and Lot\_Shape the best predictors, but they are not without problems.

saveRDS(new\_ames\_table, file="new\_ames\_table.rds")