FinalProjectPhase1

2023-06-18

## Subbu Subramanian

# Analysis of Homes sold Above median value in Ames, Iowa:

## Libraries and Data

#Loading Libraries and reading the csv file  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tidymodels)

## ── Attaching packages ────────────────────────────────────── tidymodels 1.1.0 ──  
## ✔ broom 1.0.4 ✔ rsample 1.1.1  
## ✔ dials 1.2.0 ✔ tune 1.1.1  
## ✔ infer 1.0.4 ✔ workflows 1.1.3  
## ✔ modeldata 1.1.0 ✔ workflowsets 1.0.1  
## ✔ parsnip 1.1.0 ✔ yardstick 1.2.0  
## ✔ recipes 1.0.6   
## ── Conflicts ───────────────────────────────────────── tidymodels\_conflicts() ──  
## ✖ scales::discard() masks purrr::discard()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ recipes::fixed() masks stringr::fixed()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ yardstick::spec() masks readr::spec()  
## ✖ recipes::step() masks stats::step()  
## • Learn how to get started at https://www.tidymodels.org/start/

library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(ggcorrplot)  
HomeSales = read\_csv("ames\_student-1-2.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.

## Initial Analysis

#initial Analysis of the given data  
summary(HomeSales)

## 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   
##   
##   
##

## Data Wrangling

# Data Wrangling to remove the insignificant columns and factoring them as needed  
HomeSales2 = HomeSales %>%  
 select(MS\_Zoning, Lot\_Area, Neighborhood, Bldg\_Type, Year\_Built, Year\_Remod\_Add, Gr\_Liv\_Area, TotRms\_AbvGrd, Garage\_Cars,Above\_Median) %>% mutate(Neighborhood = as\_factor(Neighborhood)) %>% mutate(MS\_Zoning = as\_factor(MS\_Zoning)) %>% mutate(Bldg\_Type = as\_factor(Bldg\_Type)) %>% mutate(Above\_Median = as\_factor(Above\_Median)) %>% filter(Above\_Median == "Yes")

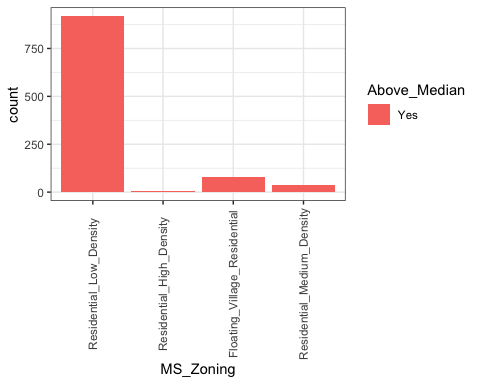
## Second Level Analysis

#Summarizing for Second Level Analysis  
summary(HomeSales2)

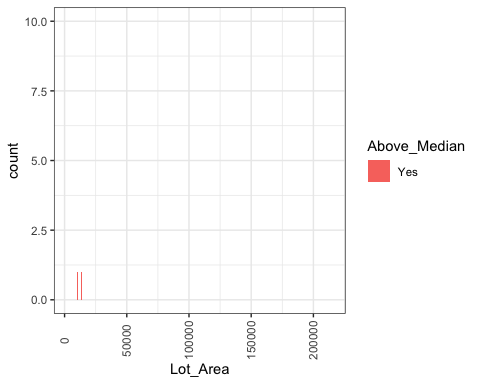
## MS\_Zoning Lot\_Area Neighborhood  
## Residential\_Low\_Density :919 Min. : 2117 College\_Creek :134   
## Residential\_High\_Density : 6 1st Qu.: 8738 Somerset :110   
## Floating\_Village\_Residential: 79 Median : 10437 Gilbert :106   
## Residential\_Medium\_Density : 39 Mean : 11829 Northridge\_Heights:104   
## C\_all : 0 3rd Qu.: 12837 Northwest\_Ames : 76   
## A\_agr : 0 Max. :215245 North\_Ames : 72   
## I\_all : 0 (Other) :441   
## Bldg\_Type Year\_Built Year\_Remod\_Add Gr\_Liv\_Area TotRms\_AbvGrd   
## OneFam :905 Min. :1879 Min. :1950 Min. : 964 Min. : 4.00   
## TwnhsE :100 1st Qu.:1977 1st Qu.:1991 1st Qu.:1456 1st Qu.: 6.00   
## Twnhs : 19 Median :1999 Median :2000 Median :1672 Median : 7.00   
## Duplex : 13 Mean :1988 Mean :1995 Mean :1770 Mean : 7.01   
## TwoFmCon: 6 3rd Qu.:2005 3rd Qu.:2006 3rd Qu.:1984 3rd Qu.: 8.00   
## Max. :2010 Max. :2010 Max. :5095 Max. :15.00   
##   
## Garage\_Cars Above\_Median  
## Min. :0.000 Yes:1043   
## 1st Qu.:2.000 No : 0   
## Median :2.000   
## Mean :2.193   
## 3rd Qu.:2.000   
## Max. :4.000   
##

## Plot Analysis across different Variables

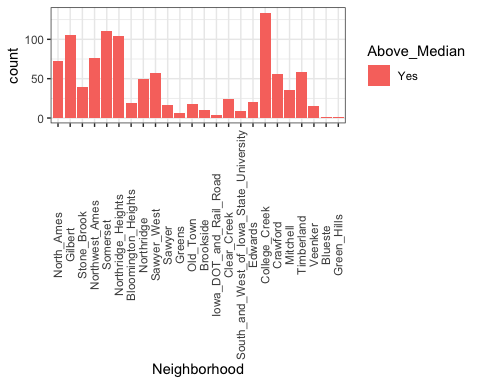
#Analyzing Above median sold homes based on Zones  
ggplot(HomeSales2, aes(x=MS\_Zoning, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



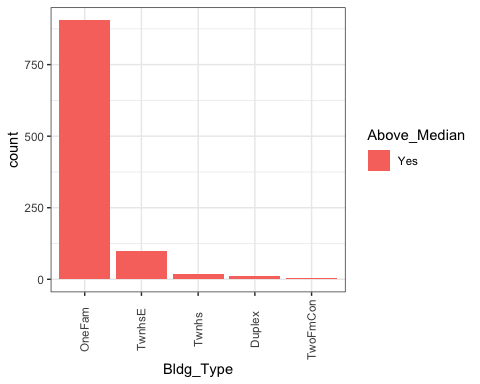
#Analyzing Above median sold homes based on Lot Area  
ggplot(HomeSales2, aes(x=Lot\_Area, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



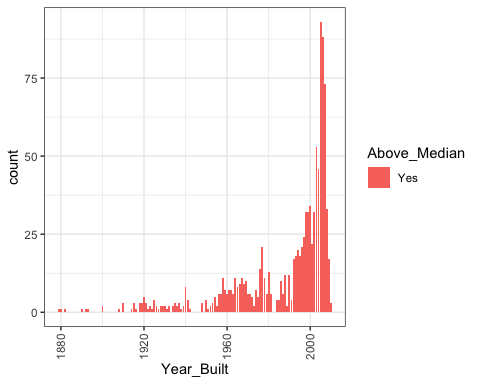
#Analyzing Above median sold homes based on Neighborhood  
ggplot(HomeSales2, aes(x=Neighborhood, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



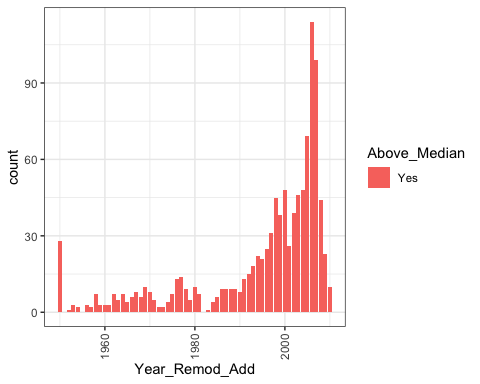
#Analyzing Above median sold homes based on the Type of Buildings  
ggplot(HomeSales2, aes(x=Bldg\_Type, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



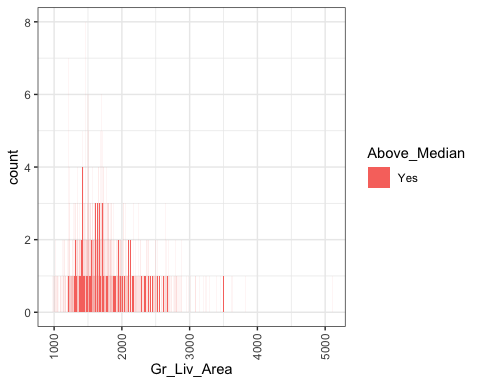
#Analyzing Above median sold homes based on Built year  
ggplot(HomeSales2, aes(x=Year\_Built, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



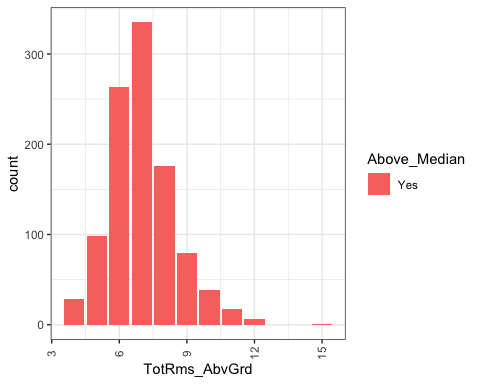
#Analyzing Above median sold homes based on Remodeled year   
ggplot(HomeSales2, aes(x=Year\_Remod\_Add, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



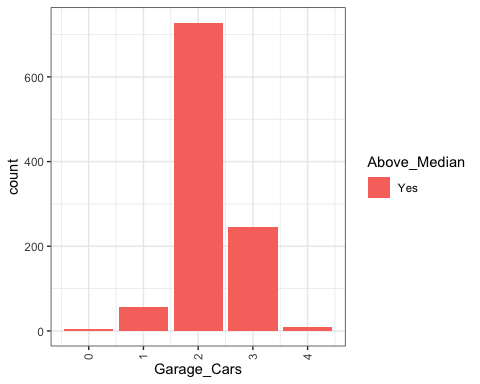
#Analyzing Above median sold homes based on Living Area  
ggplot(HomeSales2, aes(x=Gr\_Liv\_Area, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



#Analyzing Above median sold homes based on Total number of Rooms above Ground  
ggplot(HomeSales2, aes(x=TotRms\_AbvGrd, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



#Analyzing Above median sold homes based on the Garage Car capacity  
ggplot(HomeSales2, aes(x=Garage\_Cars, fill = Above\_Median)) + geom\_bar() + theme\_bw() +  
theme(axis.text.x = element\_text(angle = 90,  
 vjust = 0.5,  
 hjust = 0.5))



## Finalizing

#As evident from the analysis so far, College Creek neighborhood has the highest number of homes sold above median price. So filtering the other factors which aid the above median value of the homes that are sold in College Creek  
  
HomeSales3 = HomeSales2 %>% filter(Neighborhood == "College\_Creek")  
   
summary(HomeSales3)

## MS\_Zoning Lot\_Area Neighborhood  
## Residential\_Low\_Density :134 Min. : 6762 College\_Creek :134   
## Residential\_High\_Density : 0 1st Qu.: 8817 North\_Ames : 0   
## Floating\_Village\_Residential: 0 Median :10142 Gilbert : 0   
## Residential\_Medium\_Density : 0 Mean :10408 Stone\_Brook : 0   
## C\_all : 0 3rd Qu.:11418 Northwest\_Ames: 0   
## A\_agr : 0 Max. :21533 Somerset : 0   
## I\_all : 0 (Other) : 0   
## Bldg\_Type Year\_Built Year\_Remod\_Add Gr\_Liv\_Area TotRms\_AbvGrd   
## OneFam :134 Min. :1975 Min. :1977 Min. :1009 Min. : 5.000   
## TwnhsE : 0 1st Qu.:1999 1st Qu.:2000 1st Qu.:1479 1st Qu.: 6.000   
## Twnhs : 0 Median :2002 Median :2003 Median :1661 Median : 7.000   
## Duplex : 0 Mean :2002 Mean :2002 Mean :1660 Mean : 6.791   
## TwoFmCon: 0 3rd Qu.:2005 3rd Qu.:2006 3rd Qu.:1797 3rd Qu.: 7.000   
## Max. :2010 Max. :2010 Max. :2574 Max. :10.000   
##   
## Garage\_Cars Above\_Median  
## Min. :2.000 Yes:134   
## 1st Qu.:2.000 No : 0   
## Median :2.000   
## Mean :2.201   
## 3rd Qu.:2.000   
## Max. :3.000   
##

## Summary of the Analysis

Highest number Homes in College Creek neighborhood are the ones which are sold above median value. These above median value homes are all of One Family homes with a Low Density Residential Zoning. They all have been built on a median lot area of 10142 sq.ft with a median living area of 1661 sq.ft. All of these homes are either mostly built after 2002 or remodeled after 2003. Two car garages are a minimum for these homes and they have a median of 7 rooms abouve ground level.