Precio de la vivienda - Ejercicio SURA

EBadillo

2023-10

Variables:

names(data)

## [1] "SalePrice" "LotArea" "BedroomAbvGr" "FullBath" "GarageCars"

# 1. ¿Determine el precio promedio de venta de casas a partir del número de habitaciones?

df1 <- data |> group\_by(BedroomAbvGr) |>   
 summarise(media\_precio = mean(SalePrice))  
df1

## # A tibble: 5 × 2  
## BedroomAbvGr media\_precio  
## <dbl> <dbl>  
## 1 1 178371.  
## 2 2 158110.  
## 3 3 181057.  
## 4 4 220421.  
## 5 5 180819.

# 2. Determine la probabilidad de comprar una casa que tenga la capacidad de almacenar dos autos en su garaje si su presupuesto son 150,000 USD

df2 <- data |> filter(SalePrice==150000)   
freq(df2$GarageCars)

## Frequencies   
## df2$GarageCars   
## Type: Numeric   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 0 1 25.00 25.00 25.00 25.00  
## 1 1 25.00 50.00 25.00 50.00  
## 2 2 50.00 100.00 50.00 100.00  
## <NA> 0 0.00 100.00  
## Total 4 100.00 100.00 100.00 100.00

# La probabilidad es del 50%

# 3. ¿Cuánta plata ofrecería para comprar una casa que cuenta con garaje, tres habitaciones y dos baños?

freq(data$BedroomAbvGr)

## Frequencies   
## data$BedroomAbvGr   
## Type: Numeric   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 48 3.33 3.33 3.33 3.33  
## 2 356 24.69 28.02 24.69 28.02  
## 3 804 55.76 83.77 55.76 83.77  
## 4 213 14.77 98.54 14.77 98.54  
## 5 21 1.46 100.00 1.46 100.00  
## <NA> 0 0.00 100.00  
## Total 1442 100.00 100.00 100.00 100.00

freq(data$FullBath)

## Frequencies   
## data$FullBath   
## Type: Numeric   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 649 45.01 45.01 45.01 45.01  
## 2 760 52.70 97.71 52.70 97.71  
## 3 33 2.29 100.00 2.29 100.00  
## <NA> 0 0.00 100.00  
## Total 1442 100.00 100.00 100.00 100.00

df3 <- data |> filter(GarageCars>0 & BedroomAbvGr==3 & FullBath==2)  
summary(df3$SalePrice)

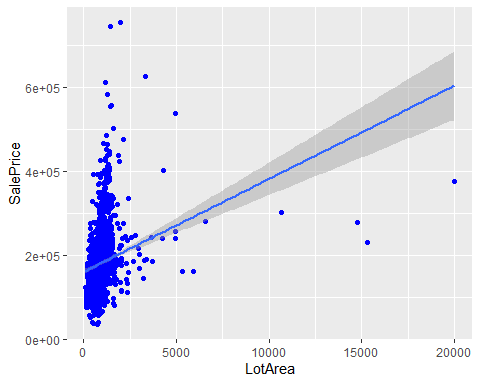
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 82500 175925 200750 214921 243375 451950

# 4. ¿Considera que existe una relación entre el precio de la vivienda y el tamaño del terreno?

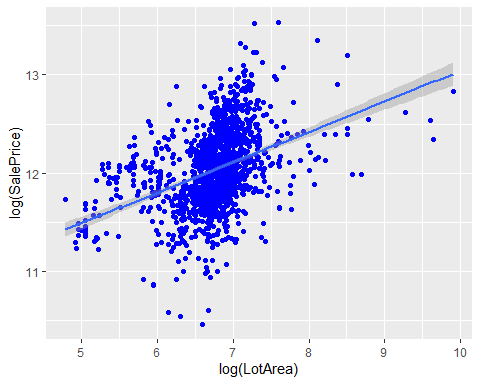
cor.test(data$SalePrice,data$LotArea)

##   
## Pearson's product-moment correlation  
##   
## data: data$SalePrice and data$LotArea  
## t = 10.184, df = 1440, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2103868 0.3067112  
## sample estimates:  
## cor   
## 0.2591934

ggplot(data, aes(x=LotArea,y=SalePrice)) +  
 geom\_point(color="blue") +  
 geom\_smooth(formula = y~x, method = lm)



ggplot(data, aes(x=log(LotArea),y=log(SalePrice))) +  
 geom\_point(color="blue") +  
 geom\_smooth(formula = y~x, method = lm)



# 5. Implemente un modelo a través del cual se estime el precio de venta de una casa en función de las variables que considere relevantes

Modelo con área en metros cuadrados como X

modelo1 <- lm(log(SalePrice) ~ LotArea +   
 BedroomAbvGr +   
 FullBath +   
 GarageCars,  
 data=data)  
summary(modelo1)

##   
## Call:  
## lm(formula = log(SalePrice) ~ LotArea + BedroomAbvGr + FullBath +   
## GarageCars, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.10946 -0.12812 0.00803 0.14711 0.87538   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.104e+01 3.026e-02 364.977 < 2e-16 \*\*\*  
## LotArea 5.402e-05 7.414e-06 7.287 5.21e-13 \*\*\*  
## BedroomAbvGr 2.562e-02 9.545e-03 2.684 0.00737 \*\*   
## FullBath 2.451e-01 1.514e-02 16.185 < 2e-16 \*\*\*  
## GarageCars 2.666e-01 1.051e-02 25.366 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2555 on 1437 degrees of freedom  
## Multiple R-squared: 0.5878, Adjusted R-squared: 0.5867   
## F-statistic: 512.3 on 4 and 1437 DF, p-value: < 2.2e-16

Tiene más sentido tener la Y como el precio por metros cuadrados, que sería el precio en forma relativa al tamaño del área

modelo2 <- lm(log(SalePrice/LotArea) ~ BedroomAbvGr + FullBath + GarageCars,  
 data=data)  
summary(modelo2)

##   
## Call:  
## lm(formula = log(SalePrice/LotArea) ~ BedroomAbvGr + FullBath +   
## GarageCars, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.70480 -0.21876 -0.00019 0.21206 1.63929   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.17015 0.05547 93.203 < 2e-16 \*\*\*  
## BedroomAbvGr -0.17054 0.01741 -9.796 < 2e-16 \*\*\*  
## FullBath 0.28062 0.02778 10.102 < 2e-16 \*\*\*  
## GarageCars 0.09541 0.01918 4.974 7.34e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4689 on 1438 degrees of freedom  
## Multiple R-squared: 0.1543, Adjusted R-squared: 0.1525   
## F-statistic: 87.45 on 3 and 1438 DF, p-value: < 2.2e-16