ModelosDeRegresionLineal

Cristopher Barrios, Carlos Daniel Estrada

2023-03-06

```
library(haven)
library(rpart)
library(stats)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(cluster)
library(rpart.plot)
library(fpc)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --

## v tibble  3.1.8     v purrr   1.0.1
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#train <- read.csv("train.csv")
#test <- read.csv("test.csv")
```

1. Descargue los conjuntos de datos de la plataforma kaggle.

librerías

```r
datos = read.csv("./train.csv")
test<- read.csv("./test.csv", stringsAsFactors = FALSE)

#Columnas
house <-select(datos, LotFrontage, LotArea, YearBuilt, YearRemodAdd, MasVnrArea, BsmtFinSF1,BsmtFinSF2,

#Data
house <- na.omit(house)

# Resumen
summary(house)
```

```
##   LotFrontage        LotArea         YearBuilt     YearRemodAdd
##  Min.   : 21.00   Min.   :  1300   Min.   :1880   Min.   :1950
##  1st Qu.: 60.00   1st Qu.:  7590   1st Qu.:1953   1st Qu.:1966
##  Median : 70.00   Median :  9416   Median :1974   Median :1995
##  Mean   : 70.67   Mean   : 10123   Mean   :1972   Mean   :1986
##  3rd Qu.: 80.00   3rd Qu.: 11361   3rd Qu.:2003   3rd Qu.:2005
##  Max.   :313.00   Max.   :215245   Max.   :2010   Max.   :2010
##    MasVnrArea       BsmtFinSF1       BsmtFinSF2        BsmtUnfSF
##  Min.   :   0.0   Min.   :   0.0   Min.   :   0.00   Min.   :   0.0
##  1st Qu.:   0.0   1st Qu.:   0.0   1st Qu.:   0.00   1st Qu.: 250.0
##  Median :   0.0   Median : 374.0   Median :   0.00   Median : 506.0
##  Mean   : 108.5   Mean   : 438.4   Mean   :  44.59   Mean   : 594.1
##  3rd Qu.: 170.0   3rd Qu.: 702.0   3rd Qu.:   0.00   3rd Qu.: 840.0
##  Max.   :1600.0   Max.   :5644.0   Max.   :1474.00   Max.   :2336.0
##   TotalBsmtSF     X1stFlrSF      X2ndFlrSF       LowQualFinSF
##  Min.   :   0   Min.   : 438   Min.   :   0.0   Min.   :  0.000
##  1st Qu.: 803   1st Qu.: 894   1st Qu.:   0.0   1st Qu.:  0.000
##  Median :1008   Median :1097   Median :   0.0   Median :  0.000
##  Mean   :1077   Mean   :1174   Mean   : 353.3   Mean   :  4.568
##  3rd Qu.:1324   3rd Qu.:1411   3rd Qu.: 728.0   3rd Qu.:  0.000
##  Max.   :6110   Max.   :4692   Max.   :2065.0   Max.   :572.000
##    GrLivArea     TotRmsAbvGrd      Fireplaces      GarageYrBlt
##  Min.   : 438   Min.   : 3.000   Min.   :0.0000   Min.   :1900
##  1st Qu.:1155   1st Qu.: 5.000   1st Qu.:0.0000   1st Qu.:1959
##  Median :1479   Median : 6.000   Median :1.0000   Median :1981
##  Mean   :1531   Mean   : 6.576   Mean   :0.6039   Mean   :1978
##  3rd Qu.:1776   3rd Qu.: 7.000   3rd Qu.:1.0000   3rd Qu.:2003
##  Max.   :5642   Max.   :12.000   Max.   :3.0000   Max.   :2010
##    GarageCars      GarageArea      WoodDeckSF      OpenPorchSF
##  Min.   :1.000   Min.   : 160   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:1.000   1st Qu.: 360   1st Qu.:  0.00   1st Qu.:  0.00
##  Median :2.000   Median : 484   Median :  0.00   Median : 27.00
##  Mean   :1.879   Mean   : 503   Mean   : 92.61   Mean   : 46.13
##  3rd Qu.:2.000   3rd Qu.: 600   3rd Qu.:168.00   3rd Qu.: 68.00
##  Max.   :4.000   Max.   :1418   Max.   :857.00   Max.   :547.00
##  EnclosedPorch     ScreenPorch       PoolArea           MoSold
##  Min.   :  0.00   Min.   :  0.0   Min.   :  0.000   Min.   : 1.00
##  1st Qu.:  0.00   1st Qu.:  0.0   1st Qu.:  0.000   1st Qu.: 5.00
##  Median :  0.00   Median :  0.0   Median :  0.000   Median : 6.00
##  Mean   : 21.84   Mean   : 16.1   Mean   :  2.935   Mean   : 6.34
##  3rd Qu.:  0.00   3rd Qu.:  0.0   3rd Qu.:  0.000   3rd Qu.: 8.00
```

```
##  Max.   :552.00   Max.   :480.0   Max.   :648.000   Max.   :12.00
##      YrSold       SalePrice
##  Min.   :2006   Min.   : 35311
##  1st Qu.:2007   1st Qu.:131000
##  Median :2008   Median :164900
##  Mean   :2008   Mean   :185506
##  3rd Qu.:2009   3rd Qu.:219500
##  Max.   :2010   Max.   :755000
```

2. Haga un análisis exploratorio extenso de los datos. Explique bien todos los hallazgos. No ponga solo gráficas y código. Debe llegar a conclusiones interesantes para poder predecir. Explique el preprocesamiento que necesitó hacer.

Analisis exploratorio

—Exploración rápida de datos— train

```
#summary(train)
```

test

```
#summary(test)
```

```
NombreDeLaVariables <- c("SalePrice", "MSSubClass", "MSZoning", "LotFrontage", "LotArea", "Street", "Al
Cualitativa <- c("", "", "x", "", "", "x", "x", "x", "x", "x", "x", "x", "x", "x", "x", "x", "x", "", "
Cuantitativa <- c("x", "x", "", "x", "x", "", "", "", "", "", "", "", "", "", "", "", "", "x", "x", "x"
DataFrame.NombreDeLaVariables <- data.frame(NombreDeLaVariables, Cualitativa,Cuantitativa)
print(DataFrame.NombreDeLaVariables)
```

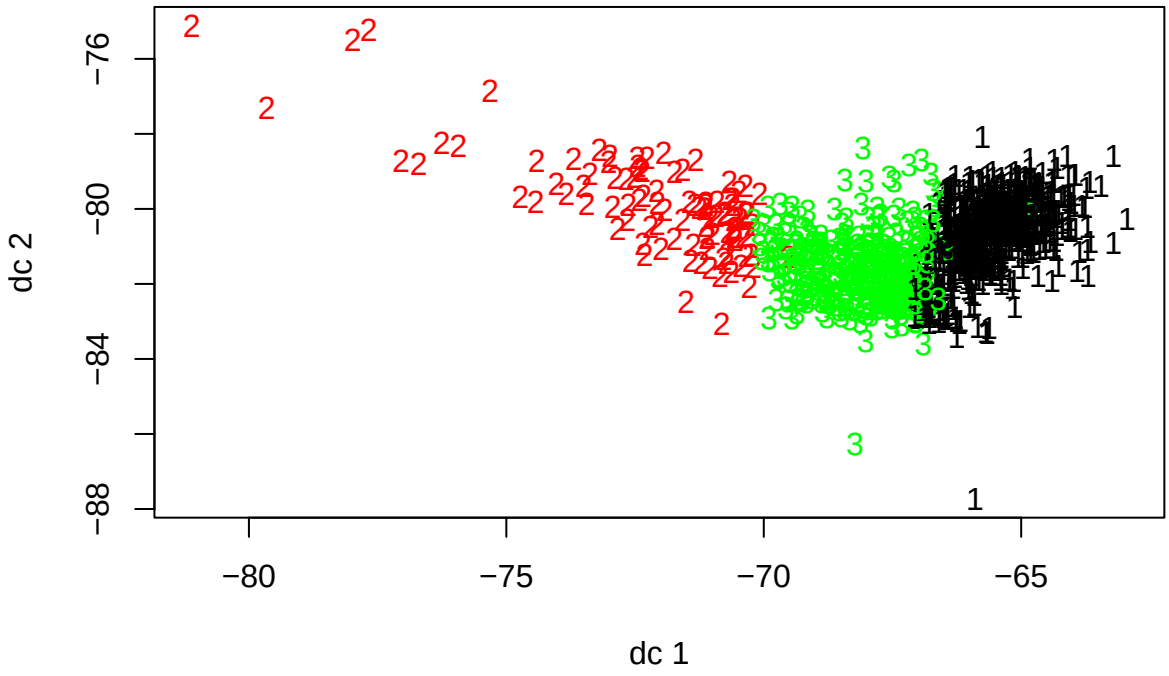
```
##    NombreDeLaVariables Cualitativa Cuantitativa
## 1            SalePrice                        x
## 2           MSSubClass                        x
## 3             MSZoning           x
## 4          LotFrontage                        x
## 5              LotArea                        x
## 6               Street           x
## 7                Alley           x
## 8             LotShape           x
## 9          LandContour           x
## 10           Utilities           x
## 11           LotConfig           x
## 12           LandSlope           x
## 13        Neighborhood           x
## 14          Condition1           x
## 15          Condition2           x
## 16            BldgType           x
## 17          HouseStyle           x
## 18         OverallQual                        x
## 19         OverallCond                        x
## 20           YearBuilt                        x
## 21           YearBuilt                        x
## 22           RoofStyle           x
```

```
## 23           RoofMatl          x
## 24        Exterior1st          x
## 25        Exterior2nd          x
## 26         MasVnrType          x
## 27         MasVnrArea                    x
## 28          ExterQual          x
## 29          ExterCond          x
## 30         Foundation          x
## 31           BsmtQual          x
## 32           BsmtCond          x
## 33       BsmtExposure          x
## 34       BsmtFinType1          x
## 35         BsmtFinSF1                    x
## 36       BsmtFinType2          x
## 37         BsmtFinSF2                    x
## 38          BsmtUnfSF                    x
## 39        TotalBsmtSF                    x
## 40            Heating          x
## 41          HeatingQC          x
## 42         CentralAir          x
## 43         Electrical          x
## 44           1stFlrSF                    x
## 45           2ndFlrSF                    x
## 46       LowQualFinSF                    x
## 47          GrLivArea                    x
## 48       BsmtFullBath                    x
## 49       BsmtHalfBath                    x
## 50           FullBath          f         xf
## 51           HalfBath                    x
## 52            Bedroom                    x
## 53            Kitchen                    x
## 54        KitchenQual          x
## 55       TotRmsAbvGrd                    x
## 56         Functional          x
## 57         Fireplaces                    x
## 58        FireplaceQu          x
## 59         GarageType          x
## 60        GarageYrBlt                    x
## 61       GarageFinish          x
## 62         GarageCars                    x
## 63         GarageArea                    x
## 64         GarageQual          x
## 65         GarageCond          x
## 66         PavedDrive          x
## 67         WoodDeckSF                    x
## 68        OpenPorchSF                    x
## 69      EnclosedPorch                    x
## 70         3SsnPorch                    x
## 71        ScreenPorch                    x
## 72           PoolArea                    x
## 73             PoolQC          x
## 74              Fence          x
## 75        MiscFeature          x
## 76            MiscVal                    x
```

```
## 77              MoSold                      x
## 78              YrSold                      x
## 79            SaleType           x
## 80       SaleCondition           x
```

3. Incluya un análisis de grupos en el análisis exploratorio. Explique las características de los grupos.

```
# con k-medias
cluster <- house
km<-kmeans(house,3)
house$grupo<-km$cluster
plotcluster(cluster,km$cluster) # los cluster
```




```
#método de la silueta
silkm<-silhouette(km$cluster,dist(house))
mean(silkm[,3]) #Silueta
```

```
## [1] 0.561677
```

```
g1<- house[house$grupo==1,]
prop.table(table(g1$Species))*100
```

```
## numeric(0)
```

```
g2<- house[house$grupo==2,]
prop.table(table(g2$Species))*100
```

```
## numeric(0)
```

```
g3<- house[house$grupo==3,]
prop.table(table(g3$Species))*100
```

```
## numeric(0)
```

```
summary(g1)
```

```
##   LotFrontage        LotArea        YearBuilt     YearRemodAdd
##  Min.   : 21.00   Min.   : 1300   Min.   :1880   Min.   :1950
##  1st Qu.: 55.00   1st Qu.: 7000   1st Qu.:1940   1st Qu.:1956
##  Median : 64.00   Median : 8593   Median :1959   Median :1972
##  Mean   : 65.45   Mean   : 8680   Mean   :1957   Mean   :1976
##  3rd Qu.: 75.00   3rd Qu.:10172   3rd Qu.:1973   3rd Qu.:1998
##  Max.   :313.00   Max.   :63887   Max.   :2009   Max.   :2009
##    MasVnrArea       BsmtFinSF1       BsmtFinSF2        BsmtUnfSF
##  Min.   :   0.00   Min.   :   0.0   Min.   :   0.00   Min.   :   0.0
##  1st Qu.:   0.00   1st Qu.:   0.0   1st Qu.:   0.00   1st Qu.: 192.0
##  Median :   0.00   Median : 319.0   Median :   0.00   Median : 450.0
##  Mean   :  60.21   Mean   : 351.2   Mean   :  51.82   Mean   : 491.1
##  3rd Qu.:  66.50   3rd Qu.: 600.0   3rd Qu.:   0.00   3rd Qu.: 719.0
##  Max.   :1129.00   Max.   :5644.0   Max.   :1085.00   Max.   :1907.0
##   TotalBsmtSF       X1stFlrSF      X2ndFlrSF       LowQualFinSF
##  Min.   :   0.0   Min.   : 438   Min.   :   0.0   Min.   :  0.000
##  1st Qu.: 728.0   1st Qu.: 834   1st Qu.:   0.0   1st Qu.:  0.000
##  Median : 894.0   Median : 981   Median :   0.0   Median :  0.000
##  Mean   : 894.1   Mean   :1020   Mean   : 256.5   Mean   :  4.513
##  3rd Qu.:1060.5   3rd Qu.:1150   3rd Qu.: 588.0   3rd Qu.:  0.000
##  Max.   :6110.0   Max.   :4692   Max.   :1230.0   Max.   :481.000
##    GrLivArea     TotRmsAbvGrd   Fireplaces      GarageYrBlt     GarageCars
##  Min.   : 438   Min.   : 3    Min.   :0.0000   Min.   :1900   Min.   :1.000
##  1st Qu.:1032   1st Qu.: 5    1st Qu.:0.0000   1st Qu.:1953   1st Qu.:1.000
##  Median :1224   Median : 6    Median :0.0000   Median :1966   Median :2.000
##  Mean   :1281   Mean   : 6    Mean   :0.4137   Mean   :1967   Mean   :1.568
##  3rd Qu.:1484   3rd Qu.: 7    3rd Qu.:1.0000   3rd Qu.:1983   3rd Qu.:2.000
##  Max.   :5642   Max.   :12    Max.   :3.0000   Max.   :2009   Max.   :4.000
##    GarageArea       WoodDeckSF      OpenPorchSF     EnclosedPorch
##  Min.   : 160.0   Min.   :  0.00   Min.   :  0.00   Min.   :  0.00
##  1st Qu.: 288.0   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00
##  Median : 416.0   Median :  0.00   Median :  0.00   Median :  0.00
##  Mean   : 417.4   Mean   : 63.95   Mean   : 26.44   Mean   : 30.07
##  3rd Qu.: 506.0   3rd Qu.:106.00   3rd Qu.: 35.00   3rd Qu.:  0.00
##  Max.   :1418.0   Max.   :736.00   Max.   :418.00   Max.   :330.00
##   ScreenPorch        PoolArea           MoSold           YrSold
##  Min.   :  0.00   Min.   :  0.000   Min.   : 1.000   Min.   :2006
##  1st Qu.:  0.00   1st Qu.:  0.000   1st Qu.: 4.000   1st Qu.:2007
```

```
##  Median :  0.00   Median :  0.000   Median : 6.000   Median :2008
##  Mean   : 13.26   Mean   :  1.612   Mean   : 6.157   Mean   :2008
##  3rd Qu.:  0.00   3rd Qu.:  0.000   3rd Qu.: 7.000   3rd Qu.:2009
##  Max.   :385.00   Max.   :576.000   Max.   :12.000   Max.   :2010
##    SalePrice          grupo
##  Min.   : 35311   Min.   :1
##  1st Qu.:118450   1st Qu.:1
##  Median :135000   Median :1
##  Mean   :134136   Mean   :1
##  3rd Qu.:154400   3rd Qu.:1
##  Max.   :178000   Max.   :1
```

```
summary(g2)
```

```
##   LotFrontage        LotArea          YearBuilt     YearRemodAdd
##  Min.   : 32.00   Min.   :  5119   Min.   :1892   Min.   :1965
##  1st Qu.: 76.00   1st Qu.: 11003   1st Qu.:1998   1st Qu.:2000
##  Median : 86.00   Median : 12444   Median :2005   Median :2006
##  Mean   : 87.82   Mean   : 16048   Mean   :1999   Mean   :2003
##  3rd Qu.: 99.00   3rd Qu.: 14601   3rd Qu.:2007   3rd Qu.:2007
##  Max.   :174.00   Max.   :215245   Max.   :2010   Max.   :2010
##    MasVnrArea       BsmtFinSF1       BsmtFinSF2       BsmtUnfSF
##  Min.   :   0.0   Min.   :   0.0   Min.   :   0.0   Min.   :   0.0
##  1st Qu.: 160.0   1st Qu.: 240.0   1st Qu.:   0.0   1st Qu.: 294.0
##  Median : 302.0   Median :1163.0   Median :   0.0   Median : 528.0
##  Mean   : 345.6   Mean   : 939.3   Mean   :  28.7   Mean   : 705.9
##  3rd Qu.: 466.0   3rd Qu.:1406.0   3rd Qu.:   0.0   3rd Qu.:1117.0
##  Max.   :1378.0   Max.   :2188.0   Max.   :1474.0   Max.   :2336.0
##   TotalBsmtSF     X1stFlrSF      X2ndFlrSF       LowQualFinSF
##  Min.   : 853   Min.   :1026   Min.   :   0.0   Min.   :  0.000
##  1st Qu.:1410   1st Qu.:1470   1st Qu.:   0.0   1st Qu.:  0.000
##  Median :1702   Median :1718   Median : 568.0   Median :  0.000
##  Mean   :1674   Mean   :1699   Mean   : 593.6   Mean   :  5.448
##  3rd Qu.:1926   3rd Qu.:1940   3rd Qu.:1177.0   3rd Qu.:  0.000
##  Max.   :3094   Max.   :2444   Max.   :2065.0   Max.   :572.000
##    GrLivArea     TotRmsAbvGrd      Fireplaces   GarageYrBlt     GarageCars
##  Min.   :1419   Min.   : 5.000   Min.   :0.0   Min.   :1932   Min.   :2.000
##  1st Qu.:1869   1st Qu.: 7.000   1st Qu.:1.0   1st Qu.:1998   1st Qu.:3.000
##  Median :2224   Median : 8.000   Median :1.0   Median :2005   Median :3.000
##  Mean   :2298   Mean   : 8.457   Mean   :1.2   Mean   :2001   Mean   :2.781
##  3rd Qu.:2610   3rd Qu.:10.000   3rd Qu.:1.0   3rd Qu.:2007   3rd Qu.:3.000
##  Max.   :4476   Max.   :12.000   Max.   :3.0   Max.   :2010   Max.   :3.000
##    GarageArea       WoodDeckSF   OpenPorchSF     EnclosedPorch
##  Min.   : 380.0   Min.   :  0   Min.   :  0.00   Min.   :  0.000
##  1st Qu.: 672.0   1st Qu.:113   1st Qu.: 45.00   1st Qu.:  0.000
##  Median : 774.0   Median :186   Median : 72.00   Median :  0.000
##  Mean   : 762.2   Mean   :192   Mean   : 88.22   Mean   :  5.876
##  3rd Qu.: 842.0   3rd Qu.:250   3rd Qu.:111.00   3rd Qu.:  0.000
##  Max.   :1220.0   Max.   :857   Max.   :502.00   Max.   :216.000
##   ScreenPorch       PoolArea           MoSold           YrSold
##  Min.   :  0.0   Min.   :  0.000   Min.   : 1.000   Min.   :2006
##  1st Qu.:  0.0   1st Qu.:  0.000   1st Qu.: 5.000   1st Qu.:2007
##  Median :  0.0   Median :  0.000   Median : 7.000   Median :2008
##  Mean   : 22.3   Mean   :  5.286   Mean   : 6.743   Mean   :2008
```

```
##  3rd Qu.:  0.0   3rd Qu.:  0.000   3rd Qu.: 9.000   3rd Qu.:2009
##  Max.   :410.0   Max.   :555.000   Max.   :12.000   Max.   :2010
##    SalePrice          grupo
##  Min.   :301000   Min.   :2
##  1st Qu.:320000   1st Qu.:2
##  Median :348000   Median :2
##  Mean   :377265   Mean   :2
##  3rd Qu.:395000   3rd Qu.:2
##  Max.   :755000   Max.   :2
```

```
summary(g3)
```

```
##   LotFrontage        LotArea        YearBuilt     YearRemodAdd
##  Min.   : 24.00   Min.   : 2280   Min.   :1880   Min.   :1950
##  1st Qu.: 63.00   1st Qu.: 8737   1st Qu.:1990   1st Qu.:1996
##  Median : 74.00   Median :10206   Median :2002   Median :2003
##  Mean   : 75.13   Mean   :11018   Mean   :1991   Mean   :1998
##  3rd Qu.: 85.00   3rd Qu.:12011   3rd Qu.:2005   3rd Qu.:2006
##  Max.   :313.00   Max.   :70761   Max.   :2009   Max.   :2010
##    MasVnrArea       BsmtFinSF1     BsmtFinSF2       BsmtUnfSF
##  Min.   :   0.0   Min.   :   0   Min.   :   0.0   Min.   :   0.0
##  1st Qu.:   0.0   1st Qu.:   0   1st Qu.:   0.0   1st Qu.: 314.0
##  Median :  40.0   Median : 410   Median :   0.0   Median : 725.0
##  Mean   : 127.1   Mean   : 451   Mean   :  36.1   Mean   : 748.5
##  3rd Qu.: 210.0   3rd Qu.: 767   3rd Qu.:   0.0   3rd Qu.:1141.0
##  Max.   :1600.0   Max.   :2260   Max.   :1127.0   Max.   :2042.0
##   TotalBsmtSF     X1stFlrSF      X2ndFlrSF       LowQualFinSF
##  Min.   :   0   Min.   : 495   Min.   :   0.0   Min.   :  0.000
##  1st Qu.: 915   1st Qu.:1020   1st Qu.:   0.0   1st Qu.:  0.000
##  Median :1240   Median :1307   Median :   0.0   Median :  0.000
##  Mean   :1236   Mean   :1299   Mean   : 458.9   Mean   :  4.413
##  3rd Qu.:1496   3rd Qu.:1552   3rd Qu.: 872.0   3rd Qu.:  0.000
##  Max.   :3206   Max.   :3138   Max.   :1818.0   Max.   :420.000
##    GrLivArea     TotRmsAbvGrd      Fireplaces      GarageYrBlt     GarageCars
##  Min.   :1146   Min.   : 4.000   Min.   :0.0000   Min.   :1908   Min.   :1.00
##  1st Qu.:1502   1st Qu.: 6.000   1st Qu.:0.0000   1st Qu.:1991   1st Qu.:2.00
##  Median :1684   Median : 7.000   Median :1.0000   Median :2002   Median :2.00
##  Mean   :1763   Mean   : 7.075   Mean   :0.7756   Mean   :1993   Mean   :2.18
##  3rd Qu.:1947   3rd Qu.: 8.000   3rd Qu.:1.0000   3rd Qu.:2005   3rd Qu.:2.00
##  Max.   :4676   Max.   :12.000   Max.   :3.0000   Max.   :2010   Max.   :4.00
##    GarageArea       WoodDeckSF     OpenPorchSF     EnclosedPorch
##  Min.   : 180.0   Min.   :  0.0   Min.   :  0.00   Min.   :  0.00
##  1st Qu.: 484.0   1st Qu.:  0.0   1st Qu.: 30.00   1st Qu.:  0.00
##  Median : 552.0   Median :120.0   Median : 54.00   Median :  0.00
##  Mean   : 582.9   Mean   :115.7   Mean   : 69.62   Mean   : 11.56
##  3rd Qu.: 662.0   3rd Qu.:192.0   3rd Qu.: 96.00   3rd Qu.:  0.00
##  Max.   :1390.0   Max.   :635.0   Max.   :547.00   Max.   :552.00
##   ScreenPorch        PoolArea           MoSold           YrSold
##  Min.   :  0.00   Min.   :  0.000   Min.   : 1.000   Min.   :2006
##  1st Qu.:  0.00   1st Qu.:  0.000   1st Qu.: 5.000   1st Qu.:2007
##  Median :  0.00   Median :  0.000   Median : 6.000   Median :2008
##  Mean   : 19.45   Mean   :  4.651   Mean   : 6.554   Mean   :2008
##  3rd Qu.:  0.00   3rd Qu.:  0.000   3rd Qu.: 8.000   3rd Qu.:2009
##  Max.   :480.00   Max.   :648.000   Max.   :12.000   Max.   :2010
```

```
##    SalePrice          grupo
##  Min.   :178900   Min.   :3
##  1st Qu.:193000   1st Qu.:3
##  Median :216837   Median :3
##  Mean   :222938   Mean   :3
##  3rd Qu.:248900   3rd Qu.:3
##  Max.   :297000   Max.   :3
```

— prueba 1 —

```
##Similitud en las variables independientes y los precios de venta:
cor(house$YearBuilt, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.5253936
```

```
cor(house$YearRemodAdd, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.5212533
```

```
cor(house$TotalBsmtSF, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6156122
```

```
cor(house$X1stFlrSF, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6079691
```

```
cor(house$GrLivArea, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.7051536
```

```
cor(house$TotRmsAbvGrd, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.5470674
```

```
cor(house$GarageCars, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6470336
```

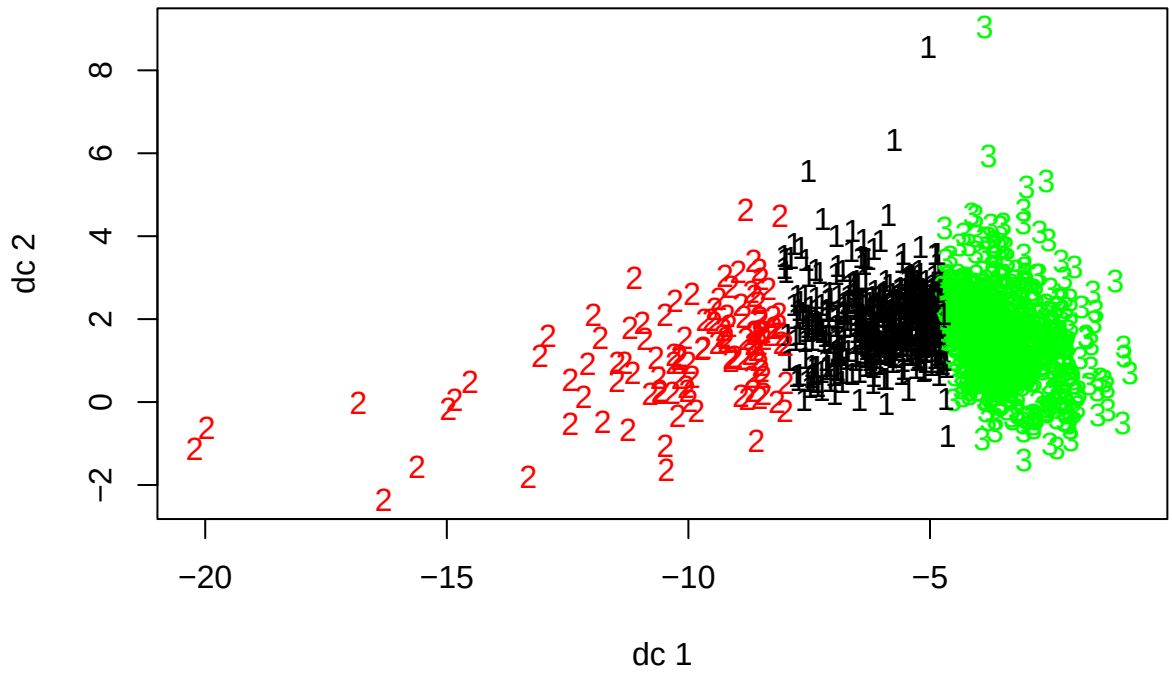
```
cor(house$GarageArea, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6193296
```

```
#Columnas
house <-select(datos,TotalBsmtSF,X1stFlrSF,GrLivArea,GarageCars,GarageArea,SalePrice)

#limpiamos
house <- na.omit(house)

#k-medias
cluster <- house
km<-kmeans(house,3)
house$grupo<-km$cluster
plotcluster(cluster,km$cluster) #graficamos ubicacion de cluster
```




```
#Silueta
silkm<-silhouette(km$cluster,dist(house))
mean(silkm[,3])
```

```
## [1] 0.5612983
```

— prueba 2 —

```
cor(house$TotalBsmtSF, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6135806
```

```
cor(house$X1stFlrSF, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6058522
```

```
cor(house$GrLivArea, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.7086245
```

```
cor(house$GarageCars, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6404092
```

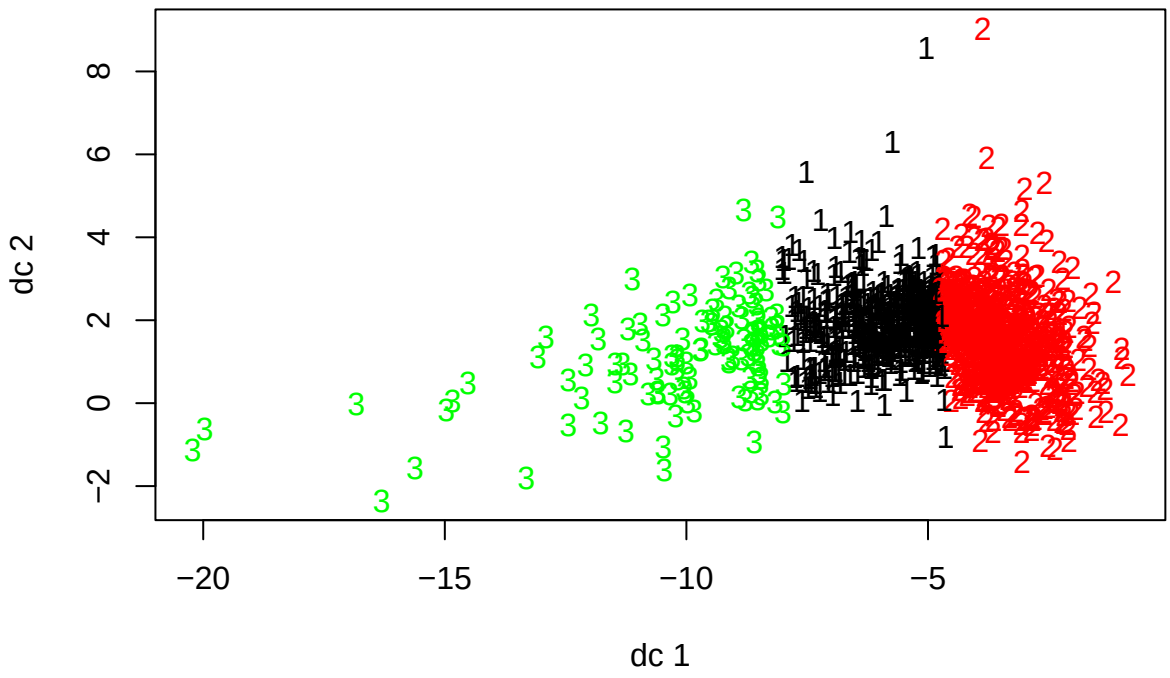
```
cor(house$GarageArea, house$SalePrice, method = c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6234314
```

```
#Columnas
house <-select(datos,TotalBsmtSF,X1stFlrSF,GrLivArea,GarageCars,GarageArea,SalePrice)

#limpiamos
house <- na.omit(house)

#k-medias
cluster <- house
km<-kmeans(house,3)
house$grupo<-km$cluster
plotcluster(cluster,km$cluster) #Graficamos clusters
```




```
#silueta
silkm<-silhouette(km$cluster,dist(house))
mean(silkm[,3])
```

```
## [1] 0.5612983
```

4. Divida el set de datos preprocesados en dos conjuntos: Entrenamiento y prueba. Describa el criterio que usó para crear los conjuntos: número de filas de cada uno, estratificado o no, balanceado o no, etc. Si le proveen un conjunto de datos de prueba y tiene suficientes datos, tómelo como de validación, pero haga sus propios conjuntos de prueba.

5. Haga ingeniería de características, ¿qué variables cree que puedan ser mejores predictores para el precio de las casas? Explique en que basó la selección o no de las variables.

6. Todos los resultados deben ser reproducible por lo que debe fijar que los conjuntos de entrenamiento y prueba sean los mismos siempre que se ejecute el código.

7. Seleccione una de las variables y haga un modelo univariado de regresión lineal para predecir el precio de las casas. Analice el m