Ejercicios basicos de modelamiento

Santiago Franco

12/5/2022

```
library(ggplot2)
```

```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to
## register S3 method.
```

```
library(grid)
library(gridExtra)
library(dplyr)
```

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

```
## The following object is masked from 'package:gridExtra':
##
##     combine
```

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

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

```
library(magrittr)
library(stringr)
```

```
mapa_de_calor <- function(categoria1, categoria2, relleno){
  p <- ggplot(aes(x=categoria1, y=categoria2, fill= relleno)) + geom_tile()
  return(p)
}
```

```
datos <- read.table("C:\\Users\\SantiagoFranco\\OneDrive - DATAKNOW S.A.S\\Documentos\\Ejercicios-Gesti
tabla_mostrar <- head(datos) %>%
  mutate_if(is.numeric, round, digits=2)
names(tabla_mostrar) <- str_replace_all(names(tabla_mostrar), "\\.", ' ')
(names(tabla_mostrar))
```

```
##  [1] "tamaño"              "sector"              "numero cifin"
##  [4] "Ebitda ActivoTotal"  "UtilNeta Patrimonio" "Razón Corriente"
##  [7] "Ciclo Efectivo"      "Pasivo Ventas"       "Pasivo Activos"
## [10] "Pasivo total"        "Default"             "Saldos"
```

```
grid.table(tabla_mostrar, theme=ttheme_default(base_size = 4))
```

	tamaño	sector	numero cifin	Ebitda ActivoTotal	UtilNeta Patrimonio	Razón Corriente	Ciclo Efectivo	Pasivo Ventas	Pasivo Activos	Pasivo total	Default	Saldos
1	G	2	10	0.12	0.12	1.16	151.32	0.49	0.60	−0.04	0	2187164000
2	P	1	1	−0.05	−0.05	0.41	206.33	2.67	1.20	0.23	0	74561800
3	M	4	10	0.05	0.05	2.23	109.48	0.41	0.73	−0.38	0	118593800
4	P	2	10	0.16	0.16	1.07	43.78	0.47	0.69	0.99	0	426824300
5	P	4	10	0.06	0.06	2.36	75.92	0.13	0.38	0.00	0	735468
6	P	4	10	−0.04	−0.04	1.32	43.89	0.15	0.72	0.40	0	251092600

Preparación datos para el modelo:

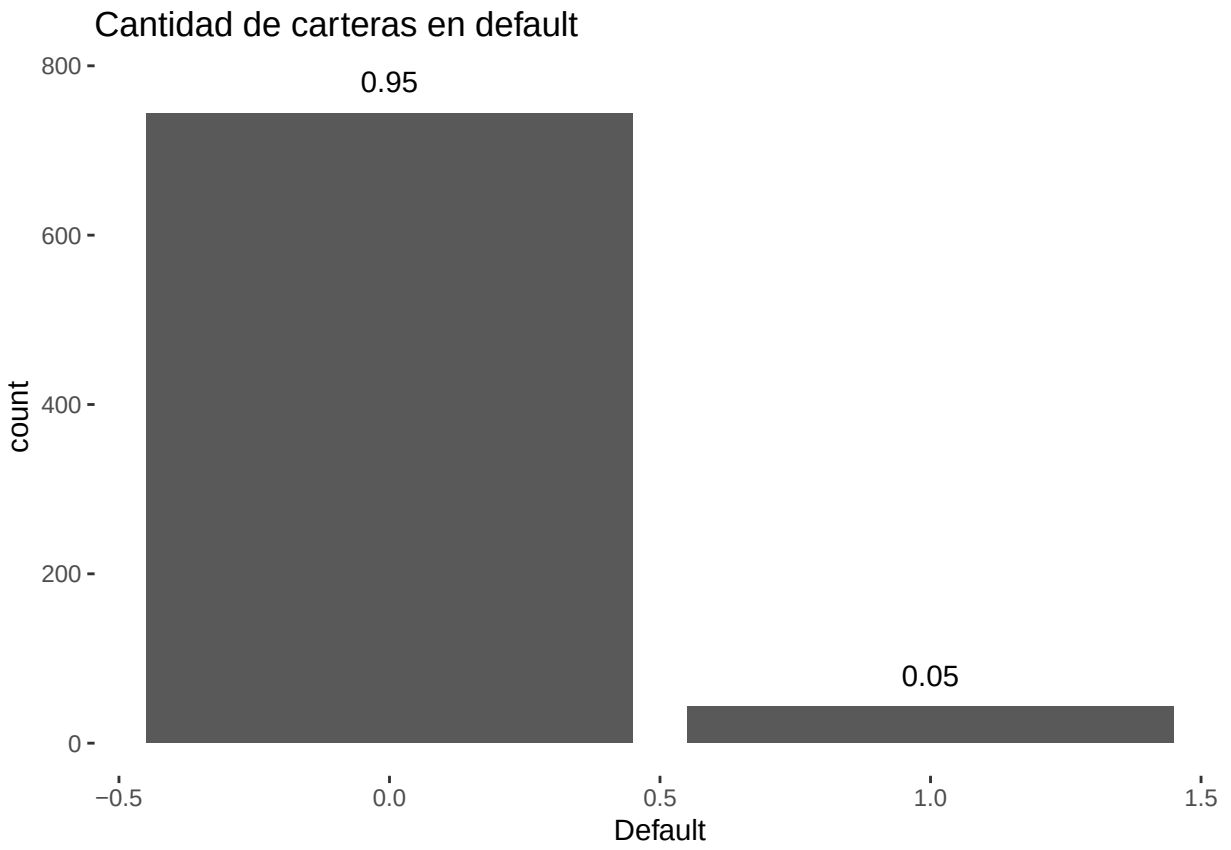
```
## Variables categóricas:
datos$tamaño <- as.factor(datos$tamaño)
datos$sector <- as.factor(datos$sector)

## Datos para el modelo
datos_modelo <- datos[-c(3)]
datos_modelo$Default <- as.factor(datos_modelo$Default)
```

Análisis descriptivo de los datos

Variable respuesta

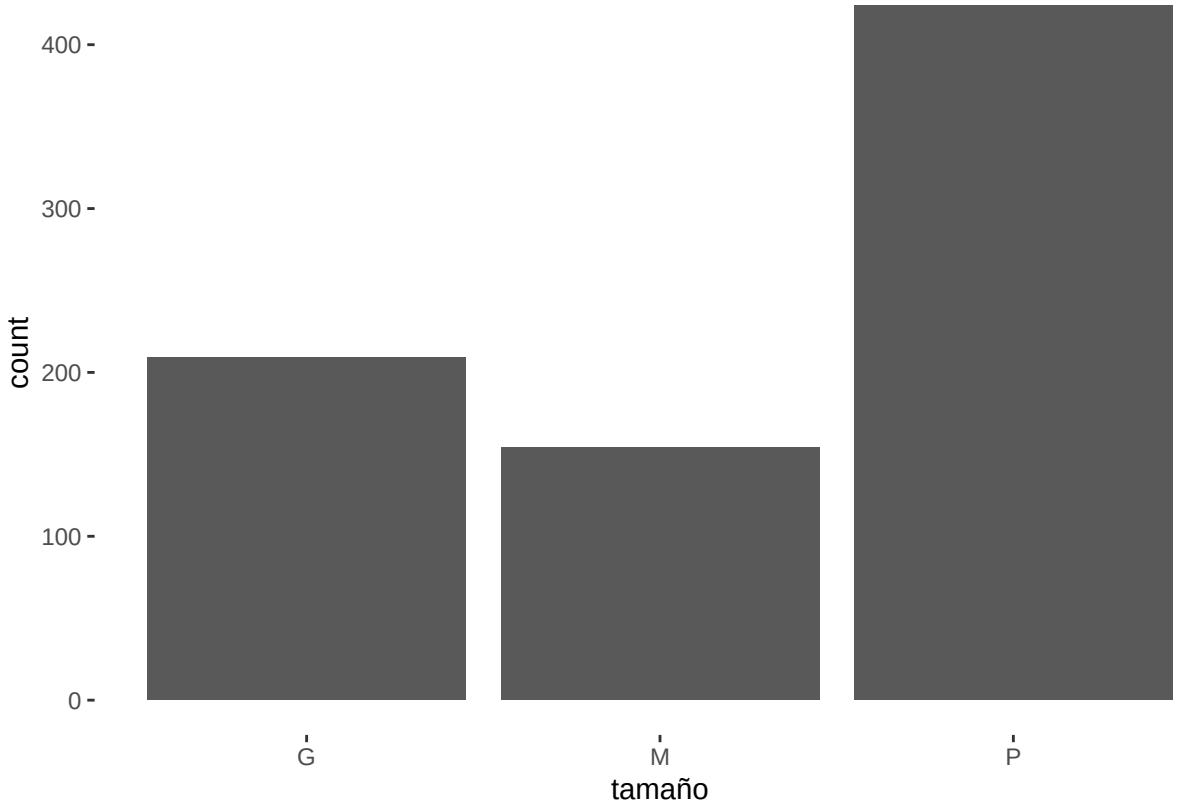
```
ggplot(data=datos, aes(x=Default)) +
  geom_bar() +
  geom_text(stat='count', aes(label=round(..count../sum(..count..),2)), vjust=-1) +
  ylim(0,770) +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank()) + ggtitle('Cantidad de carteras en default')
```



A simple vista se observa que se tienen dos categorías desbalanceadas en la base de datos para establecer el modelo.

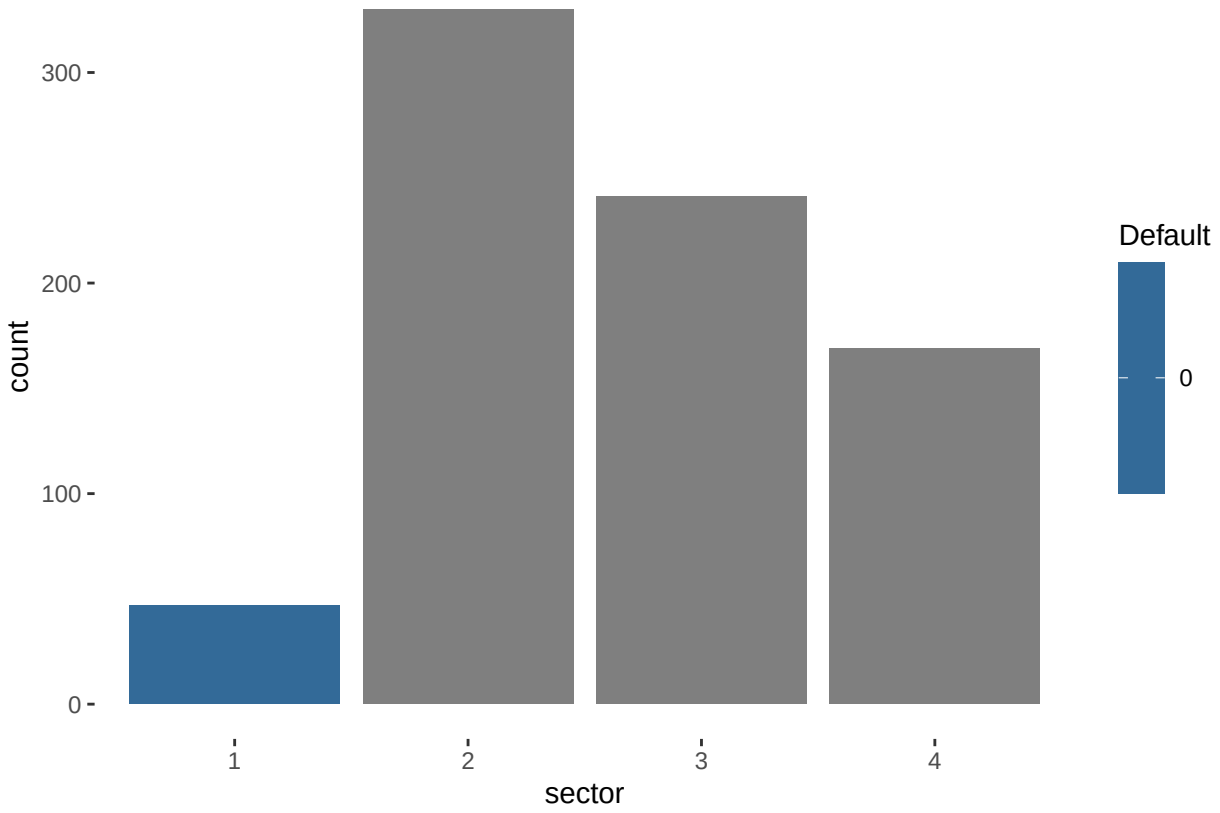
Tamaño de la empresa

```
### Heatmap
p <- ggplot(data=datos, aes(x=tamaño, fill=Default)) +
  geom_bar(position = 'dodge') +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank())
p
```



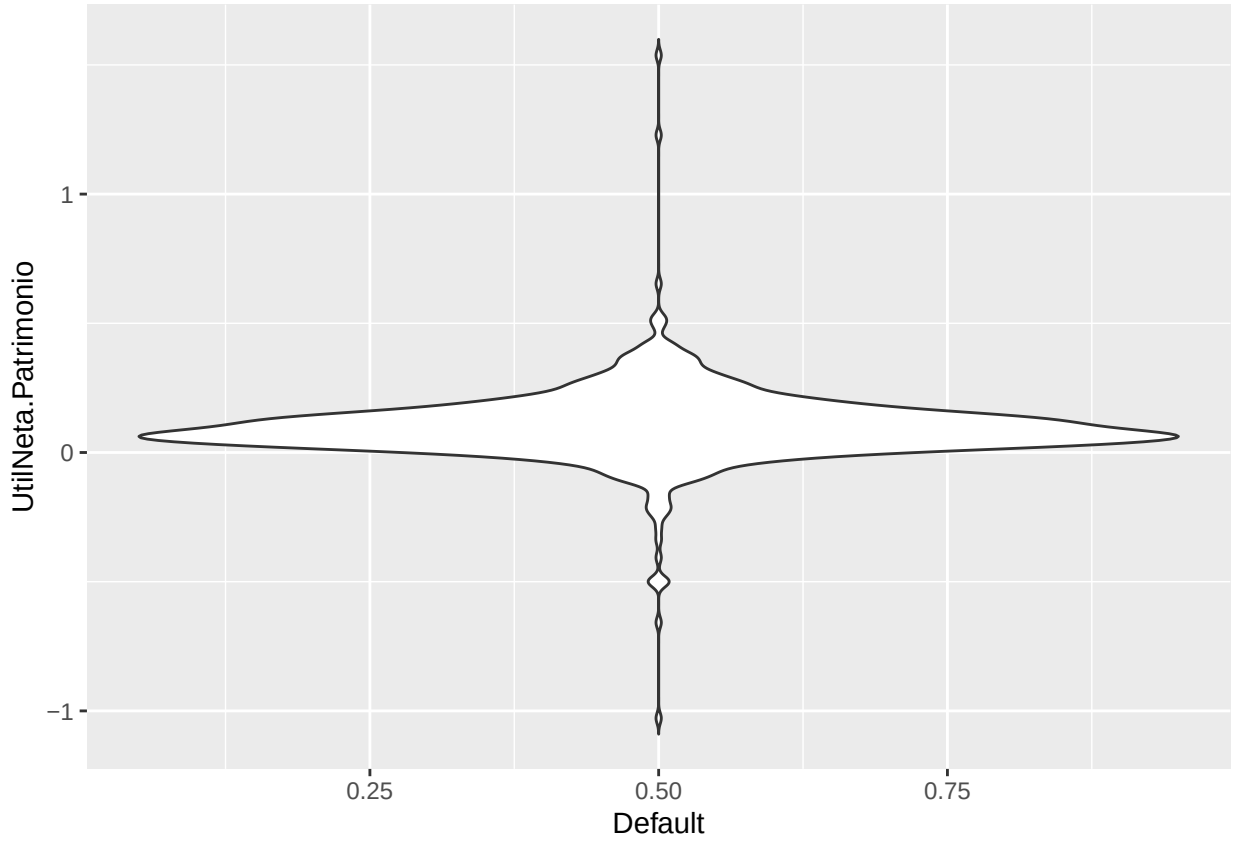
Sector

```
### Heatmap
p <- ggplot(data=datos, aes(x=sector, fill=Default)) +
  geom_bar(position = 'dodge') +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank())
p
```



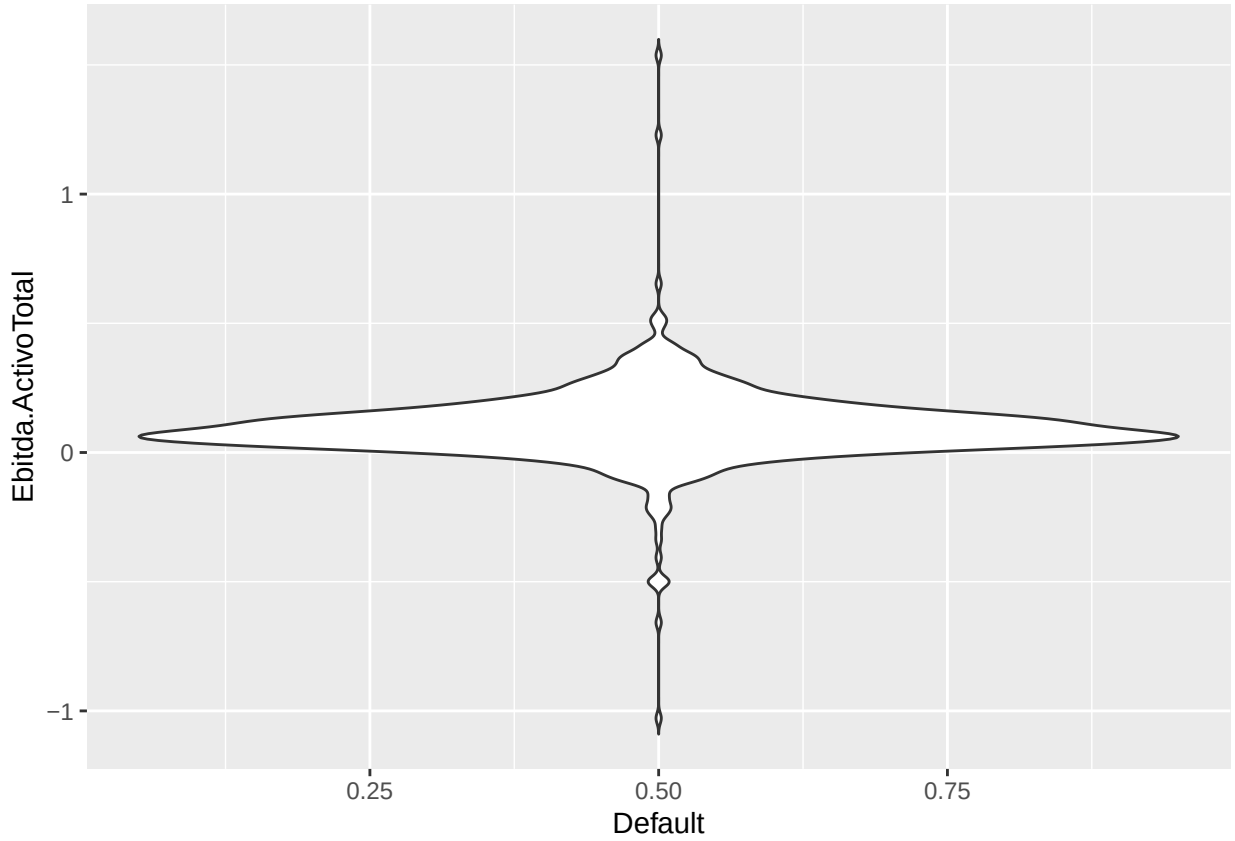
Margen Operativo

```
### Grafica violín
ggplot(data=datos, aes(x=Default,
                       y=UtilNeta.Patrimonio,
                       fill=Default)) +
  geom_violin(trim = F)
```



Margen Ebitda

```
### Grafica violín
ggplot(data=datos, aes(x=Default,
                       y=Ebitda.ActivoTotal,
                       fill=Default)) +
  geom_violin(trim = F)
```



Ebitda/ Gasto Financiero

Pasivo Fcro. Total / (Ebitda - Gasto Fcro)

Margen Neto

Ebitda / Activo Total

SALDO DEUDA

##Implementacion modelo logit

```
modelo_logit <- glm(Default ~ ., family = binomial, data=datos_modelo)
```

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

```
summary(modelo_logit)
```

```
##
## Call:
## glm(formula = Default ~ ., family = binomial, data = datos_modelo)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.0121  -0.3249  -0.2438  -0.1550   2.8272
##
## Coefficients: (1 not defined because of singularities)
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.875e+01  9.074e+02  -0.021   0.9835
## tamañoM              5.121e-01  5.519e-01   0.928   0.3535
## tamañoP              5.315e-01  4.646e-01   1.144   0.2526
## sector2              1.596e+01  9.074e+02   0.018   0.9860
## sector3              1.544e+01  9.074e+02   0.017   0.9864
## sector4              1.618e+01  9.074e+02   0.018   0.9858
## Ebitda.ActivoTotal  -5.530e+00  1.355e+00  -4.082 4.46e-05 ***
## UtilNeta.Patrimonio         NA         NA      NA       NA
## Razón.Corriente     -2.735e-01  2.923e-01  -0.936   0.3494
## Ciclo.Efectivo      -4.192e-03  2.484e-03  -1.688   0.0915 .
## Pasivo.Ventas        1.148e-01  8.208e-02   1.399   0.1619
## Pasivo.Activos       1.219e+00  8.016e-01   1.521   0.1283
## Pasivo.total        -1.743e-01  1.788e-01  -0.975   0.3294
## Saldos              -8.461e-12  1.472e-10  -0.057   0.9542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 333.61  on 786  degrees of freedom
## Residual deviance: 262.21  on 774  degrees of freedom
## AIC: 288.21
##
## Number of Fisher Scoring iterations: 17
```

###Tabla de coeficientes

summary(modelo_logit) ###Valores P

###Pruebas de ajuste

###Curva ROC

##Implementacion de random forest

###Curva ROC

##Implementacion de Knn

###Curva ROC

##Implementacion de SVM

###Curva ROC

Curva ROC para todos los modelos

AUC de todos los modelos

##Simulacion con la peor estimacion

##Simulacion con la mejor estimacion

Comparacion de resultados