Caso III

Abigail Monterroso

2023-07-02

1. Preparar variable a clasificar:

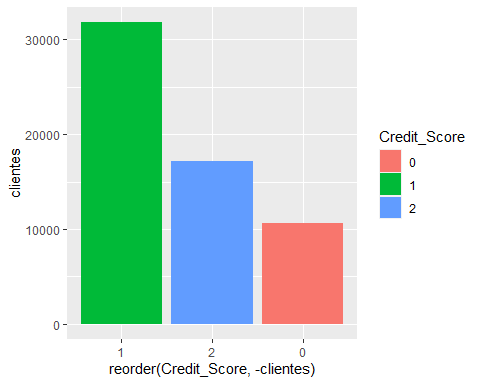
bank$Credit\_Score = as.factor(bank$Credit\_Score)  
bank$Credit\_Score = relevel(bank$Credit\_Score,ref = "0")

1. Exploración variables cualitativas:

credit1 <- bank %>%   
 group\_by(Credit\_Score) %>%   
 summarise(clientes = n())  
credit1

## # A tibble: 3 × 2  
## Credit\_Score clientes  
## <fct> <int>  
## 1 0 10618  
## 2 1 31846  
## 3 2 17203

ggplot(credit1,aes(x=reorder(Credit\_Score,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Credit\_Score))

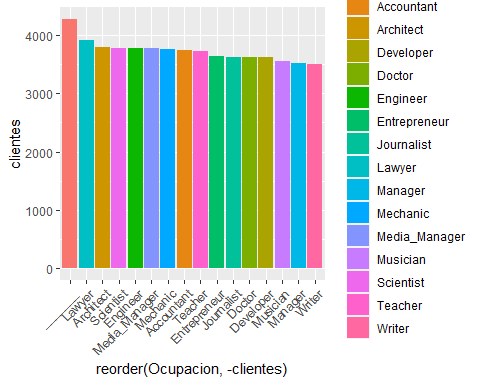


Existe una mayor cantidad de clientes con un nivel de credito regular, seguidos por los malos y una menro cantidade de personas con una buena calificación de credito. Por lo tanto se espera que los resultados reflejen esta tendencia.

ocupaciont1 <- bank %>%   
 group\_by(Ocupacion) %>%   
 summarise(clientes = n())  
ocupaciont1

## # A tibble: 16 × 2  
## Ocupacion clientes  
## <chr> <int>  
## 1 Accountant 3740  
## 2 Architect 3794  
## 3 Developer 3621  
## 4 Doctor 3624  
## 5 Engineer 3775  
## 6 Entrepreneur 3649  
## 7 Journalist 3629  
## 8 Lawyer 3925  
## 9 Manager 3519  
## 10 Mechanic 3765  
## 11 Media\_Manager 3775  
## 12 Musician 3561  
## 13 Scientist 3778  
## 14 Teacher 3733  
## 15 Writer 3504  
## 16 \_\_\_\_\_\_\_ 4275

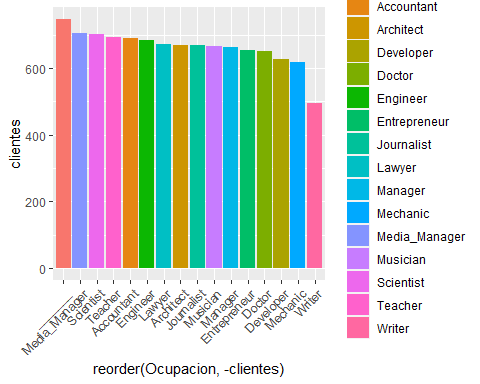
ggplot(ocupaciont1,aes(x=reorder(Ocupacion,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Ocupacion))+ scale\_x\_discrete(guide=guide\_axis(angle=45))



ocupacion\_CS0 <- bank[bank$Credit\_Score == 0,] %>%   
 group\_by(Ocupacion) %>%   
 summarise(clientes = n())  
ocupacion\_CS0

## # A tibble: 16 × 2  
## Ocupacion clientes  
## <chr> <int>  
## 1 Accountant 692  
## 2 Architect 671  
## 3 Developer 628  
## 4 Doctor 652  
## 5 Engineer 685  
## 6 Entrepreneur 655  
## 7 Journalist 670  
## 8 Lawyer 672  
## 9 Manager 664  
## 10 Mechanic 619  
## 11 Media\_Manager 706  
## 12 Musician 668  
## 13 Scientist 702  
## 14 Teacher 693  
## 15 Writer 494  
## 16 \_\_\_\_\_\_\_ 747

ggplot(ocupacion\_CS0,aes(x=reorder(Ocupacion,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Ocupacion))+ scale\_x\_discrete(guide=guide\_axis(angle=45))

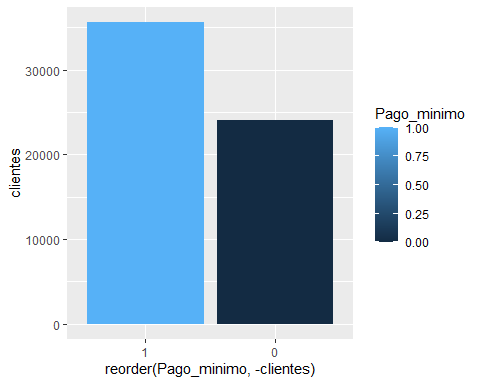


La cantidad de clientes por cada profesión parece uniforme, indicando que no existen una cantidad de clientes similares por cada profesión. La grafica de las profesiones de clientes con un buen record crediticio se mantiene similar a la anterior, pero en un distinto orden y con una marcada reducción de clientes para un par de profesiones como writter. De entre las variables que se vieron más afectadas por el filtro se encuentra: Lawyer y Mechanic

pago\_min1 <- bank %>%   
 group\_by(Pago\_minimo) %>%   
 summarise(clientes = n())  
pago\_min1

## # A tibble: 2 × 2  
## Pago\_minimo clientes  
## <int> <int>  
## 1 0 24044  
## 2 1 35623

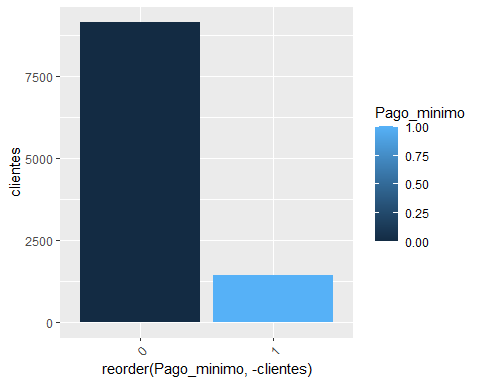
ggplot(pago\_min1,aes(x=reorder(Pago\_minimo,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Pago\_minimo))



pago\_min.CS0 <- bank[bank$Credit\_Score == 0,] %>%   
 group\_by(Pago\_minimo) %>%   
 summarise(clientes = n())  
pago\_min.CS0

## # A tibble: 2 × 2  
## Pago\_minimo clientes  
## <int> <int>  
## 1 0 9164  
## 2 1 1454

ggplot(pago\_min.CS0,aes(x=reorder(Pago\_minimo,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Pago\_minimo))+ scale\_x\_discrete(guide=guide\_axis(angle=45))

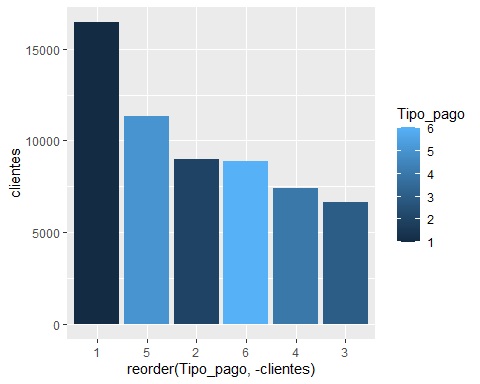


Aunque existe una mayor cantidad de clientes de los cuales su ultimo pago fue el minimo, se nota que aquellos que no dieron el minimo tienen un mejor record crediticio.

pagost1 <- bank %>%   
 group\_by(Tipo\_pago) %>%   
 summarise(clientes = n())  
pagost1

## # A tibble: 6 × 2  
## Tipo\_pago clientes  
## <int> <int>  
## 1 1 16458  
## 2 2 8985  
## 3 3 6650  
## 4 4 7387  
## 5 5 11313  
## 6 6 8874

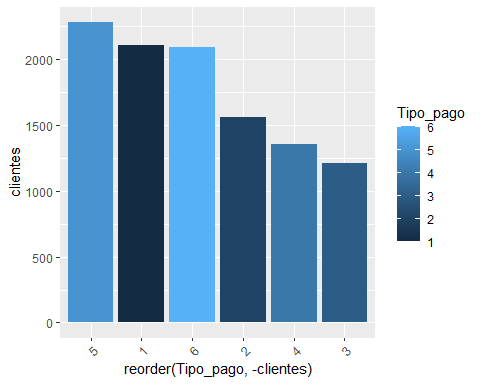
ggplot(pagost1,aes(x=reorder(Tipo\_pago,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Tipo\_pago))



pagos.CS0 <- bank[bank$Credit\_Score == 0,] %>%   
 group\_by(Tipo\_pago) %>%   
 summarise(clientes = n())  
pagos.CS0

## # A tibble: 6 × 2  
## Tipo\_pago clientes  
## <int> <int>  
## 1 1 2110  
## 2 2 1564  
## 3 3 1215  
## 4 4 1354  
## 5 5 2284  
## 6 6 2091

ggplot(pagos.CS0,aes(x=reorder(Tipo\_pago,-clientes),y=clientes))+  
 geom\_bar(stat = "identity",aes(fill=Tipo\_pago))+ scale\_x\_discrete(guide=guide\_axis(angle=45))



Aqui se puede ver variación entre los datos generales por tipo de pagos, y la tendencia de aquellos con un buen record crediticio. Se nota que los que hacer pocos pagos de valor pequeño y los de muchos pagos a valor mediano tienen una mayor probabilidad de tener un buen credito.

1. Evaluación de variables cualitativas:

bank2 <- dummy\_cols(bank,select\_columns = c("Tipo\_pago", "Pago\_minimo", "Ocupacion"))  
  
pagost2 <- table(Pago1 = bank2$Tipo\_pago\_1, Credito = bank2$Credit\_Score)  
pagost2

## Credito  
## Pago1 0 1 2  
## 0 8508 23359 11342  
## 1 2110 8487 5861

prop.table(pagost2,1)

## Credito  
## Pago1 0 1 2  
## 0 0.1969034 0.5406050 0.2624916  
## 1 0.1282051 0.5156763 0.3561186

summary(pagost2)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 692.3, df = 2, p-value = 4.604e-151

pagost3 <- table(Pago5 = bank2$Tipo\_pago\_5, Credito = bank2$Credit\_Score)  
pagost3

## Credito  
## Pago5 0 1 2  
## 0 8334 25674 14346  
## 1 2284 6172 2857

prop.table(pagost3,1)

## Credito  
## Pago5 0 1 2  
## 0 0.1723539 0.5309592 0.2966869  
## 1 0.2018916 0.5455670 0.2525413

summary(pagost3)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 110.58, df = 2, p-value = 9.702e-25

Para los clientes con muchos pagos de valor mediano se rechaza la Ho con una confianza de 95%, por lo que se concluye que si existe una relación entre esta variable y credit score. Como se esperaba, la mayoría de daots se encuentran con un credito regular, aunque si muestra una mayor proporción de clientes con un buen record crediticio mayor al de los de tipo 1 con 20%

pago\_min2 <- table(Pago\_min\_si = bank2$Pago\_minimo\_1, Credito = bank2$Credit\_Score)  
pago\_min2

## Credito  
## Pago\_min\_si 0 1 2  
## 0 9164 11665 3215  
## 1 1454 20181 13988

prop.table(pago\_min2,1)

## Credito  
## Pago\_min\_si 0 1 2  
## 0 0.38113459 0.48515222 0.13371319  
## 1 0.04081633 0.56651601 0.39266766

summary(pago\_min2)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 12859, df = 2, p-value = 0

pago\_min3 <- table(Pago\_min\_No = bank2$Pago\_minimo\_0, Credito = bank2$Credit\_Score)  
pago\_min3

## Credito  
## Pago\_min\_No 0 1 2  
## 0 1454 20181 13988  
## 1 9164 11665 3215

prop.table(pago\_min3,1)

## Credito  
## Pago\_min\_No 0 1 2  
## 0 0.04081633 0.56651601 0.39266766  
## 1 0.38113459 0.48515222 0.13371319

summary(pago\_min3)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 12859, df = 2, p-value = 0

Con un nivel de confianza de 95%, se concluye que ambas variables parecen tener una relación con credit score. Asímismo, se puede notar que aquellos que su ultimo pago no fue el minimo tienen una mayor proporción de clientes con un buen record crediticio.

Abogados <- table(Abogado = bank2$Ocupacion\_Lawyer, Credito = bank2$Credit\_Score)  
Abogados

## Credito  
## Abogado 0 1 2  
## 0 9946 29666 16130  
## 1 672 2180 1073

prop.table(Abogados,1)

## Credito  
## Abogado 0 1 2  
## 0 0.1784292 0.5322019 0.2893689  
## 1 0.1712102 0.5554140 0.2733758

summary(Abogados)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 8.028, df = 2, p-value = 0.01806

Mecanico <- table(Mecanico = bank2$Ocupacion\_Mechanic, Credito = bank2$Credit\_Score)  
Abogados

## Credito  
## Abogado 0 1 2  
## 0 9946 29666 16130  
## 1 672 2180 1073

prop.table(Mecanico,1)

## Credito  
## Mecanico 0 1 2  
## 0 0.1788666 0.5338807 0.2872527  
## 1 0.1644090 0.5314741 0.3041169

summary(Mecanico)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 7.661, df = 2, p-value = 0.0217

Cientifico <- table(Cientifico = bank2$Ocupacion\_Scientist, Credito = bank2$Credit\_Score)  
Abogados

## Credito  
## Abogado 0 1 2  
## 0 9946 29666 16130  
## 1 672 2180 1073

prop.table(Cientifico,1)

## Credito  
## Cientifico 0 1 2  
## 0 0.1774231 0.5350248 0.2875521  
## 1 0.1858126 0.5145580 0.2996294

summary(Cientifico)

## Number of cases in table: 59667   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 5.967, df = 2, p-value = 0.05061

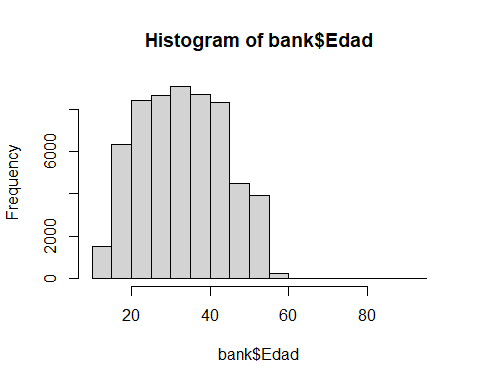
Con un nivel de confianza de 95% se concluye que todas las variables mantiene una relación con el credit score, por lo que pueden ser utilizadas para formar un modelo.

1. Exploración de variables numéricas:

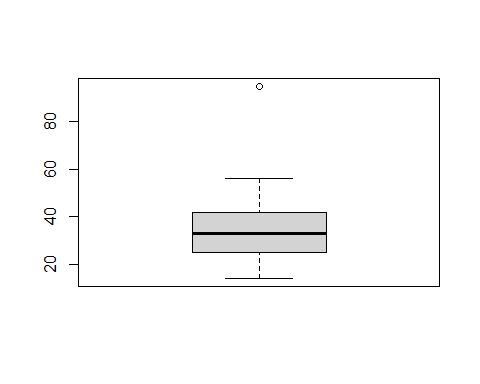
summary(bank$Edad)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 14.00 25.00 33.00 33.46 42.00 95.00

hist(bank$Edad)



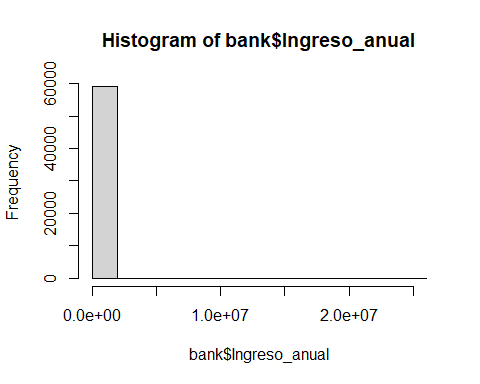
boxplot(bank$Edad)



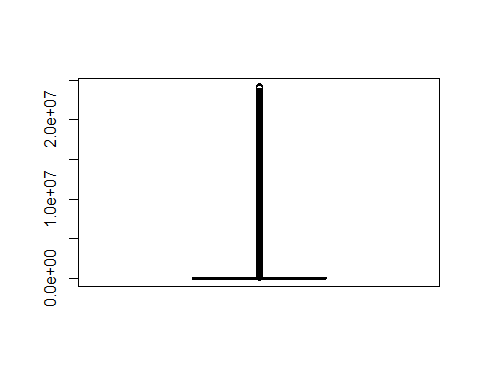
#No existen muchos datos atipicos, la mayoría parece estar entre 20-40 años de edad. Parece seguir una forma normal.   
  
summary(bank$Ingreso\_anual)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 7006 20058 40760 183663 83158 24198062

hist(bank$Ingreso\_anual)



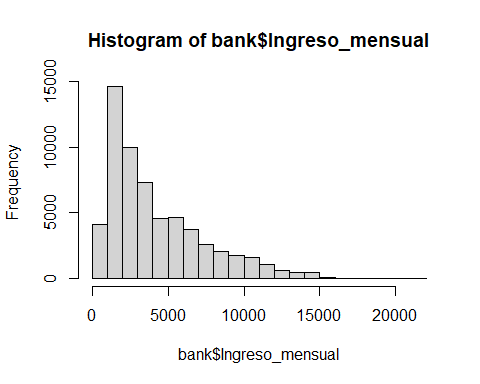
boxplot(bank$Ingreso\_anual)



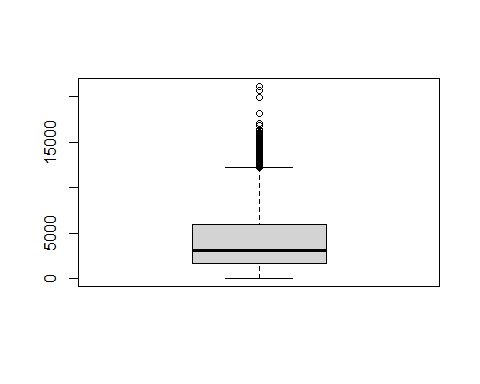
#Demasiados datos atipicos, salario muy por arriba de la media.  
  
summary(bank$Ingreso\_mensual)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 1688 3098 4179 5888 21167

hist(bank$Ingreso\_mensual)



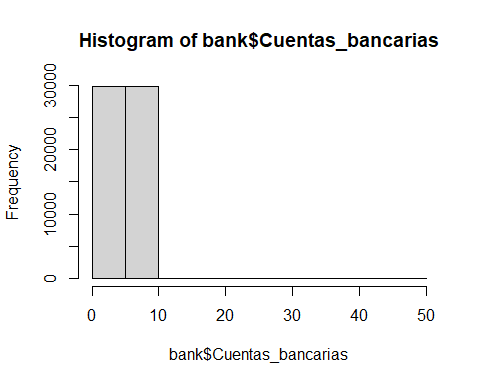
boxplot(bank$Ingreso\_mensual)



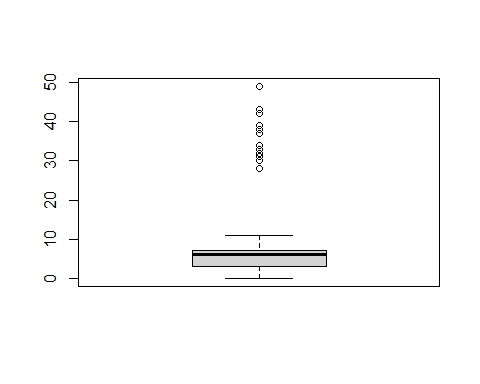
#No tantos datos atipicos.  
  
summary(bank$Cuentas\_bancarias)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 3.000 6.000 5.385 7.000 49.000

hist(bank$Cuentas\_bancarias)



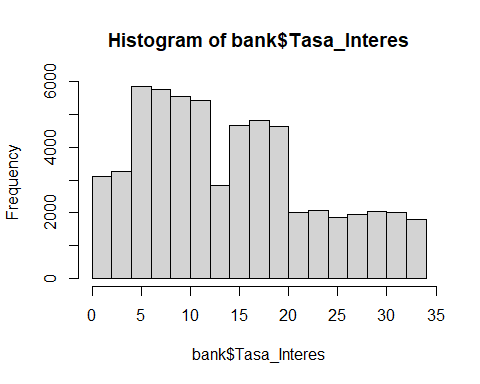
boxplot(bank$Cuentas\_bancarias)



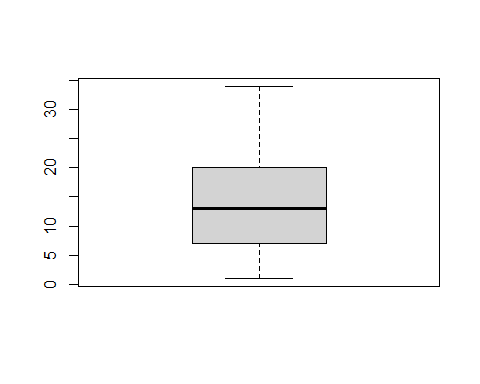
#No tantos datos atipicos, pero los que existen se alejan demasiado impactando fuertemente a la media. La mayoría parece centrarse entre 1-10 cuenstas.  
  
summary(bank$Tasa\_Interes)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 7.00 13.00 14.55 20.00 34.00

hist(bank$Tasa\_Interes)



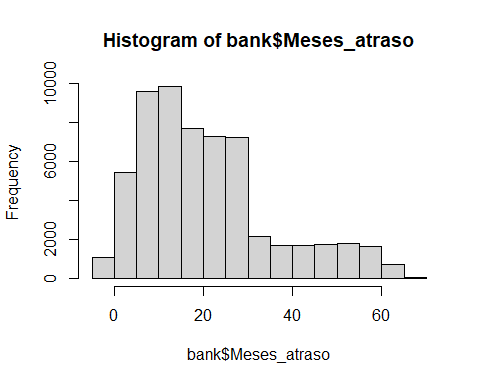
boxplot(bank$Tasa\_Interes)



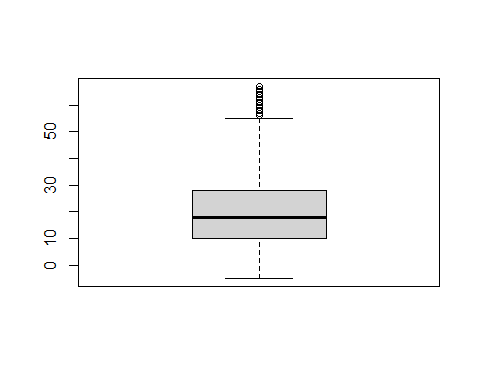
#0 datos atipicos, parece más uniforme que otra cosa.  
  
summary(bank$Meses\_atraso)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -5.00 10.00 18.00 21.11 28.00 67.00

hist(bank$Meses\_atraso)



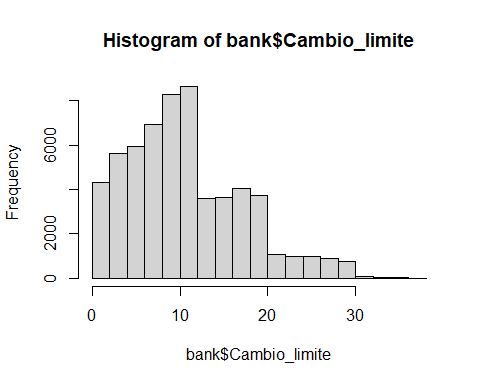
boxplot(bank$Meses\_atraso)



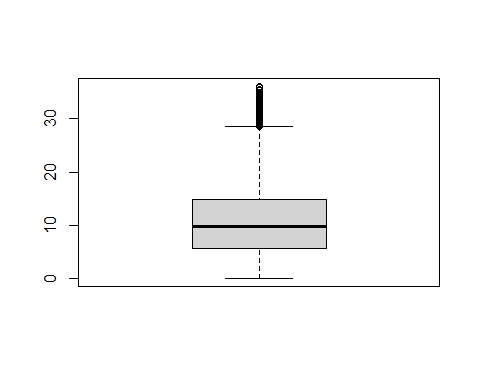
summary(bank$Cambio\_limite)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.01 5.68 9.70 10.61 14.84 36.09

hist(bank$Cambio\_limite)



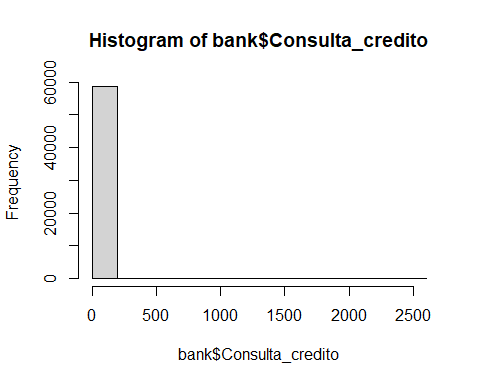
boxplot(bank$Cambio\_limite)



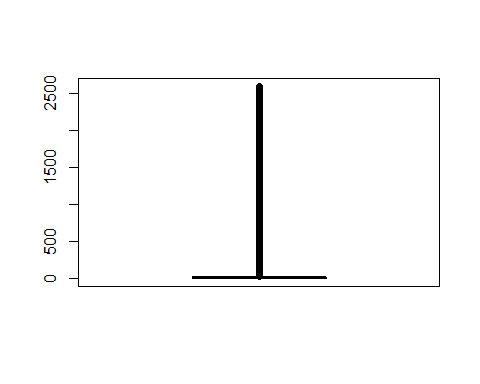
summary(bank$Consulta\_credito)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 3.00 6.00 28.37 9.00 2597.00

hist(bank$Consulta\_credito)



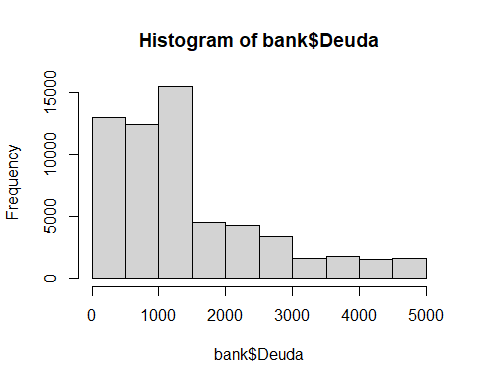
boxplot(bank$Consulta\_credito)



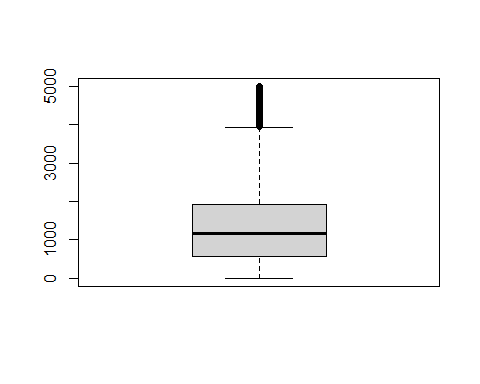
summary(bank$Deuda)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.23 569.38 1173.99 1425.41 1919.27 4998.07

hist(bank$Deuda)



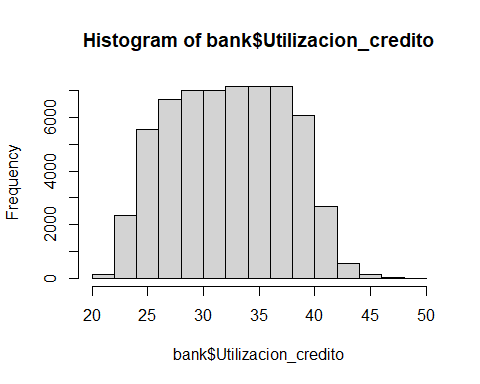
boxplot(bank$Deuda)



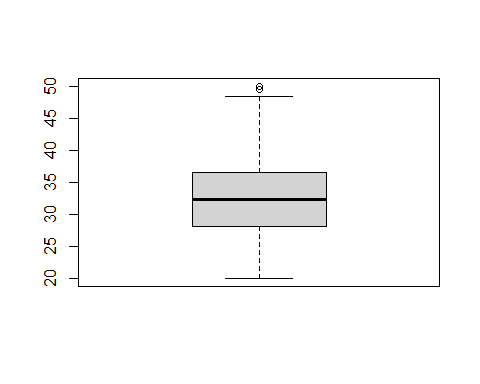
#No tantos datos atipicos.  
  
summary(bank$Utilizacion\_credito)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 20.00 28.06 32.32 32.28 36.49 50.00

hist(bank$Utilizacion\_credito)



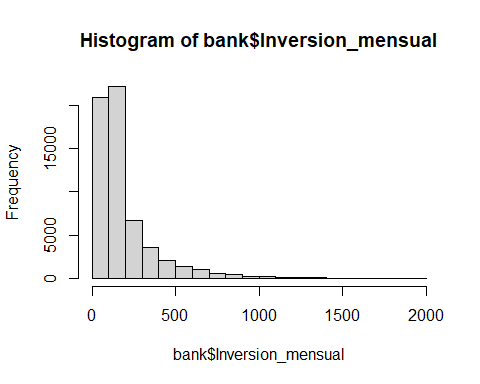
boxplot(bank$Utilizacion\_credito)



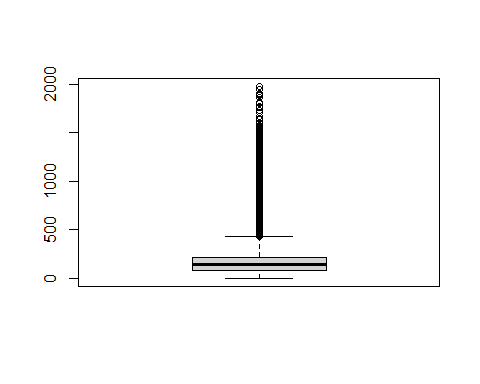
summary(bank$Inversion\_mensual)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 76.81 142.45 194.37 218.67 1977.33

hist(bank$Inversion\_mensual)



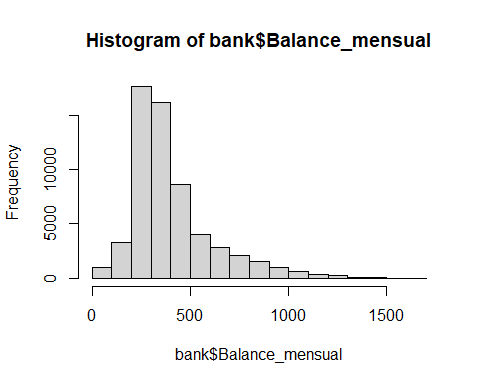
boxplot(bank$Inversion\_mensual)



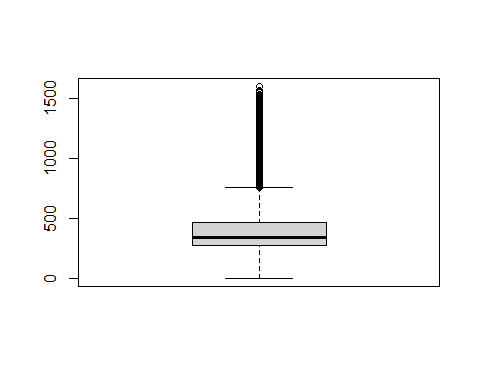
#Demasiados outliers  
  
summary(bank$Balance\_mensual)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0886 270.9549 338.8716 402.5058 466.9801 1602.0405

hist(bank$Balance\_mensual)



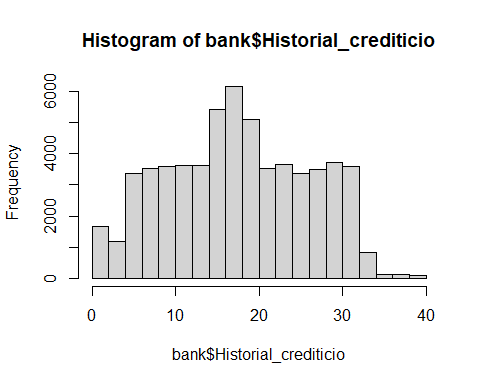
boxplot(bank$Balance\_mensual)



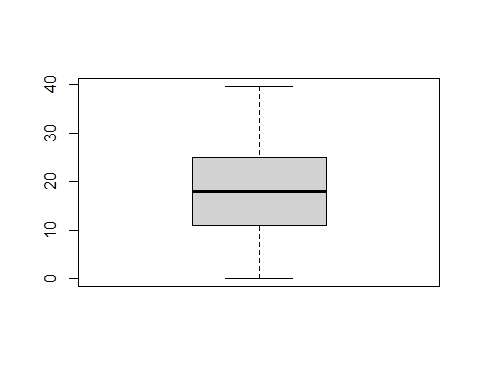
#Demasiados outliers  
  
summary(bank$Historial\_crediticio)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 11.00 18.00 17.95 25.00 39.69

hist(bank$Historial\_crediticio)



boxplot(bank$Historial\_crediticio)



#Sin outliers parece inclinarse a ser más uniforme

De esta eploración se filtran aquellas que muestran una menor cantidad de outliers, así como una forma no tan uniforme, ya que se espera que tengan una mayor relación con la variable de credit score.

1. Evaluación de variables númericas:

## Setting direction: controls > cases  
## Setting direction: controls > cases  
## Setting direction: controls > cases

##   
## Call:  
## multiclass.roc.formula(formula = bank2$Credit\_Score ~ bank2$Edad)  
##   
## Data: bank2$Edad with 3 levels of bank2$Credit\_Score: 0, 1, 2.  
## Multi-class area under the curve: 0.5921

## Setting direction: controls > cases  
## Setting direction: controls > cases  
## Setting direction: controls > cases

##   
## Call:  
## multiclass.roc.formula(formula = bank2$Credit\_Score ~ bank2$Ingreso\_mensual)  
##   
## Data: bank2$Ingreso\_mensual with 3 levels of bank2$Credit\_Score: 0, 1, 2.  
## Multi-class area under the curve: 0.6194

## Setting direction: controls < cases

## Setting direction: controls < cases  
## Setting direction: controls < cases

##   
## Call:  
## multiclass.roc.formula(formula = bank2$Credit\_Score ~ bank2$Meses\_atraso)  
##   
## Data: bank2$Meses\_atraso with 3 levels of bank2$Credit\_Score: 0, 1, 2.  
## Multi-class area under the curve: 0.7519

## Setting direction: controls < cases  
## Setting direction: controls < cases  
## Setting direction: controls < cases

##   
## Call:  
## multiclass.roc.formula(formula = bank2$Credit\_Score ~ bank2$Deuda)  
##   
## Data: bank2$Deuda with 3 levels of bank2$Credit\_Score: 0, 1, 2.  
## Multi-class area under the curve: 0.7305

## Setting direction: controls > cases

## Setting direction: controls > cases  
## Setting direction: controls > cases

##   
## Call:  
## multiclass.roc.formula(formula = bank2$Credit\_Score ~ bank2$Balance\_mensual)  
##   
## Data: bank2$Balance\_mensual with 3 levels of bank2$Credit\_Score: 0, 1, 2.  
## Multi-class area under the curve: 0.6264

Se puede notar que las mejores variables para utilizar en el modelo son meses de atraso y Deuda, con un area bajo la curva de más de 70, lo cual las hace buenas candidatas para predecir lo valores de credito.

1. Evaluar el modelo con 2 datasets, train y test:

index=sort(sample(nrow(bank2),nrow(bank2)\*0.70))  
train= bank2[index,]  
test= bank2[-index,]  
  
  
modelo1 = multinom(Credit\_Score~ Ocupacion\_Lawyer+Tipo\_pago\_5+Ingreso\_mensual+Meses\_atraso+Deuda+Balance\_mensual,data = train, na.action = na.exclude)

## # weights: 24 (14 variable)  
## initial value 45884.640849   
## iter 10 value 39603.977678  
## iter 20 value 36261.289865  
## final value 36260.074928   
## converged

summary(modelo1)

## Call:  
## multinom(formula = Credit\_Score ~ Ocupacion\_Lawyer + Tipo\_pago\_5 +   
## Ingreso\_mensual + Meses\_atraso + Deuda + Balance\_mensual,   
## data = train, na.action = na.exclude)  
##   
## Coefficients:  
## (Intercept) Ocupacion\_Lawyer Tipo\_pago\_5 Ingreso\_mensual Meses\_atraso  
## 1 -0.1455612 0.1362794 -0.1314084 -5.432351e-05 0.07571551  
## 2 -1.8325459 0.1029557 -0.2380695 -8.693858e-05 0.10396938  
## Deuda Balance\_mensual  
## 1 0.0003566506 1.494792e-04  
## 2 0.0006951841 7.507254e-05  
##   
## Std. Errors:  
## (Intercept) Ocupacion\_Lawyer Tipo\_pago\_5 Ingreso\_mensual Meses\_atraso  
## 1 3.743466e-05 2.429001e-06 5.097013e-06 5.713358e-06 0.001562224  
## 2 2.240117e-05 1.631612e-06 1.735607e-06 7.358006e-06 0.001678174  
## Deuda Balance\_mensual  
## 1 2.026255e-05 7.386317e-05  
## 2 2.148883e-05 9.359824e-05  
##   
## Residual Deviance: 72520.15   
## AIC: 72548.15

modelo1$AIC

## [1] 72548.15

tidy(modelo1)

## # A tibble: 14 × 6  
## y.level term estimate std.error statistic p.value  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 1 (Intercept) -0.146 0.0000374 -3888. 0   
## 2 1 Ocupacion\_Lawyer 0.136 0.00000243 56105. 0   
## 3 1 Tipo\_pago\_5 -0.131 0.00000510 -25781. 0   
## 4 1 Ingreso\_mensual -0.0000543 0.00000571 -9.51 1.94e- 21  
## 5 1 Meses\_atraso 0.0757 0.00156 48.5 0   
## 6 1 Deuda 0.000357 0.0000203 17.6 2.40e- 69  
## 7 1 Balance\_mensual 0.000149 0.0000739 2.02 4.30e- 2  
## 8 2 (Intercept) -1.83 0.0000224 -81806. 0   
## 9 2 Ocupacion\_Lawyer 0.103 0.00000163 63101. 0   
## 10 2 Tipo\_pago\_5 -0.238 0.00000174 -137168. 0   
## 11 2 Ingreso\_mensual -0.0000869 0.00000736 -11.8 3.25e- 32  
## 12 2 Meses\_atraso 0.104 0.00168 62.0 0   
## 13 2 Deuda 0.000695 0.0000215 32.4 1.35e-229  
## 14 2 Balance\_mensual 0.0000751 0.0000936 0.802 4.23e- 1

Para este modelo se toma la varible de lawyer, ya que muestra la mayor cantidad de datos y un buen porcentaje de clientes con un buen credito. Asímismo, meses atraso, deuda y balance mensual mostraron el area bajo la curva mayor de los datos analizados, y por lo tanto se incluyen en el modelo. El tipo de pago 5 es donde se encuentran mayor número de clientes con un buen credito.

Tanto para creditos regulares como malos se ve una relación directamente proporcional con cada una de las variables escogidas, donde la probabilidad de que un cliente perteneca a una de estas aumenta por cada unidad que suba o cualidad a la que pertenezca. Deuda es la que parece tener más impacto en los resultados, con el coeficiente más alto. No se consideraría un modelo efectivo con un AIC bastante alto.

modelo2 = multinom(Credit\_Score~Ingreso\_mensual+Meses\_atraso+Ocupacion\_Mechanic+Tipo\_pago\_1+Pago\_minimo\_0+Ocupacion\_Scientist+Deuda,data = train, na.action = na.exclude)

## # weights: 27 (16 variable)  
## initial value 45884.640849   
## iter 10 value 38263.688543  
## iter 20 value 34897.377756  
## final value 34816.894758   
## converged

summary(modelo2)

## Call:  
## multinom(formula = Credit\_Score ~ Ingreso\_mensual + Meses\_atraso +   
## Ocupacion\_Mechanic + Tipo\_pago\_1 + Pago\_minimo\_0 + Ocupacion\_Scientist +   
## Deuda, data = train, na.action = na.exclude)  
##   
## Coefficients:  
## (Intercept) Ingreso\_mensual Meses\_atraso Ocupacion\_Mechanic Tipo\_pago\_1  
## 1 1.6676424 -1.039539e-05 0.04692361 0.1427465 0.1820407  
## 2 -0.1522495 -3.649873e-05 0.07523513 0.1966725 0.3956439  
## Pago\_minimo\_0 Ocupacion\_Scientist Deuda  
## 1 -1.933296 -0.03842566 -9.246705e-06  
## 2 -1.969242 0.03319393 3.253746e-04  
##   
## Std. Errors:  
## (Intercept) Ingreso\_mensual Meses\_atraso Ocupacion\_Mechanic Tipo\_pago\_1  
## 1 4.207537e-05 3.194379e-06 0.001568892 1.914586e-06 1.250088e-05  
## 2 2.850222e-05 4.194740e-06 0.001673270 1.478012e-06 1.163074e-05  
## Pago\_minimo\_0 Ocupacion\_Scientist Deuda  
## 1 2.795237e-05 1.463054e-06 2.124009e-05  
## 2 1.077063e-05 1.658109e-06 2.234264e-05  
##   
## Residual Deviance: 69633.79   
## AIC: 69665.79

modelo2$AIC

## [1] 69665.79

tidy(modelo2)

## # A tibble: 16 × 6  
## y.level term estimate std.error statistic p.value  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 1 (Intercept) 1.67 0.0000421 39635. 0   
## 2 1 Ingreso\_mensual -0.0000104 0.00000319 -3.25 1.14e- 3  
## 3 1 Meses\_atraso 0.0469 0.00157 29.9 1.51e-196  
## 4 1 Ocupacion\_Mechanic 0.143 0.00000191 74557. 0   
## 5 1 Tipo\_pago\_1 0.182 0.0000125 14562. 0   
## 6 1 Pago\_minimo\_0 -1.93 0.0000280 -69164. 0   
## 7 1 Ocupacion\_Scientist -0.0384 0.00000146 -26264. 0   
## 8 1 Deuda -0.00000925 0.0000212 -0.435 6.63e- 1  
## 9 2 (Intercept) -0.152 0.0000285 -5342. 0   
## 10 2 Ingreso\_mensual -0.0000365 0.00000419 -8.70 3.29e- 18  
## 11 2 Meses\_atraso 0.0752 0.00167 45.0 0   
## 12 2 Ocupacion\_Mechanic 0.197 0.00000148 133066. 0   
## 13 2 Tipo\_pago\_1 0.396 0.0000116 34017. 0   
## 14 2 Pago\_minimo\_0 -1.97 0.0000108 -182834. 0   
## 15 2 Ocupacion\_Scientist 0.0332 0.00000166 20019. 0   
## 16 2 Deuda 0.000325 0.0000223 14.6 4.83e- 48

Para el modelo 2 se agregaron la ocupación mecanico y cientifico, que fueron analizadas conjuntamente con abogados anteriormente. Por otra parte tambien se incluyó el tip de pago 1 donde se vio una mayor cantidad de clientes y el pago minimo, que se consideró relevante para el analisis ya que muestra un gran cambio con respecto a la calidad creditica de un cliente, por lo que se asume que es una acción que hace una diferencia.

1. Predicción con modelo para train:

train$predictedp = predict(modelo2,newdata = train,"probs")  
train$predictedc = predict(modelo2,newdata = train,"class")  
  
  
train$original= train$Credit\_Score  
train$nuevo=train$predictedc  
  
comparacion1 = table(original=train$original,nuevo=train$nuevo)  
comparacion1

## nuevo  
## original 0 1 2  
## 0 2319 5001 127  
## 1 1843 17400 3059  
## 2 391 7269 4357

prop.table(comparacion1)

## nuevo  
## original 0 1 2  
## 0 0.055523632 0.119738543 0.003040751  
## 1 0.044126802 0.416606809 0.073241393  
## 2 0.009361682 0.174041086 0.104319303

accuracy1 <- sum(diag(comparacion1))/sum(comparacion1)\*100  
accuracy1

## [1] 57.64497

Tiene un accuracy bastante bajo, por lo que se sigue considerando un mal modelo de predicción, aunque muestre un mejor akaike que el modelo 1.

1. Predicción con modelo para test:

test$predictedp = predict(modelo2,newdata = test,"probs")  
test$predictedc = predict(modelo2,newdata = test,"class")  
  
  
test$original= test$Credit\_Score  
test$nuevo=test$predictedc  
  
comparacion2 = table(original=test$original,nuevo=test$nuevo)  
comparacion2

## nuevo  
## original 0 1 2  
## 0 947 2171 53  
## 1 774 7387 1383  
## 2 163 3102 1921

prop.table(comparacion2)

## nuevo  
## original 0 1 2  
## 0 0.052902073 0.121278141 0.002960728  
## 1 0.043237808 0.412658511 0.077258254  
## 2 0.009105637 0.173286409 0.107312441

accuracy2 <- sum(diag(comparacion2))/sum(comparacion2)\*100  
accuracy2

## [1] 57.2873

Se llega al nivel de accuracy esperado con la data de entrenamiento.

1. Uso de modelo para predicción de datos en segundo dataset de bancos:

setwd("C:/Users/abiga/Documents/DATAR")  
bank.test <- read.csv("bank2.csv")  
bank.test2 <- dummy\_cols(bank.test,select\_columns = c("Tipo\_pago", "Pago\_minimo", "Ocupacion"))  
  
  
bank.test2$predictedp = predict(modelo2,newdata = bank.test2,"probs")  
bank.test2$predictedc = predict(modelo2,newdata = bank.test2,"class")