ANEXO

I) Frecuencia de aparición de cada uno de los valores de las variables categóricas

```
# Comprobación de la frecuencia de aparición de los valores de las variables categóri
cas
lapply(Filter(is.factor, datos), FUN = freq)

## $Genero
##           n    % val%
## Female 2963 46.6 49.4
## Male   3038 47.8 50.6
## NA      352  5.5   NA
##
## $Mayor65
##       n    % val%
## 0  5045 79.4 84.1
## 1   956 15.0 15.9
## NA  352  5.5   NA
##
## $Conyuge
##      n    % val%
## 0 3287 51.7 51.7
## 1 3066 48.3 48.3
##
## $PersCargo
##      n    % val%
## 0 4439 69.9 69.9
## 1 1914 30.1 30.1
##
## $Telf_serv
##       n    % val%
## 0   605  9.5  9.7
## 1  5656 89.0 90.3
## NA   92  1.4   NA
##
## $VariasLineas
##      n  % val%
## 0 3687 58   58
## 1 2666 42   42
##
## $Int_serv
##                n    % val%
## DSL         2122 33.4 34.4
## Fiber optic 2702 42.5 43.8
## No          1339 21.1 21.7
## NA           190  3.0   NA
##
## $Seguridad
##      n    % val%
## 0 4532 71.3 71.3
## 1 1821 28.7 28.7
##
## $CopiaSeguridad
##      n    % val%
## 0 4180 65.8 65.8
## 1 2173 34.2 34.2
##
## $Antivirus_disp
##      n    % val%
## 0 4180 65.8 65.8
## 1 2173 34.2 34.2
##
```

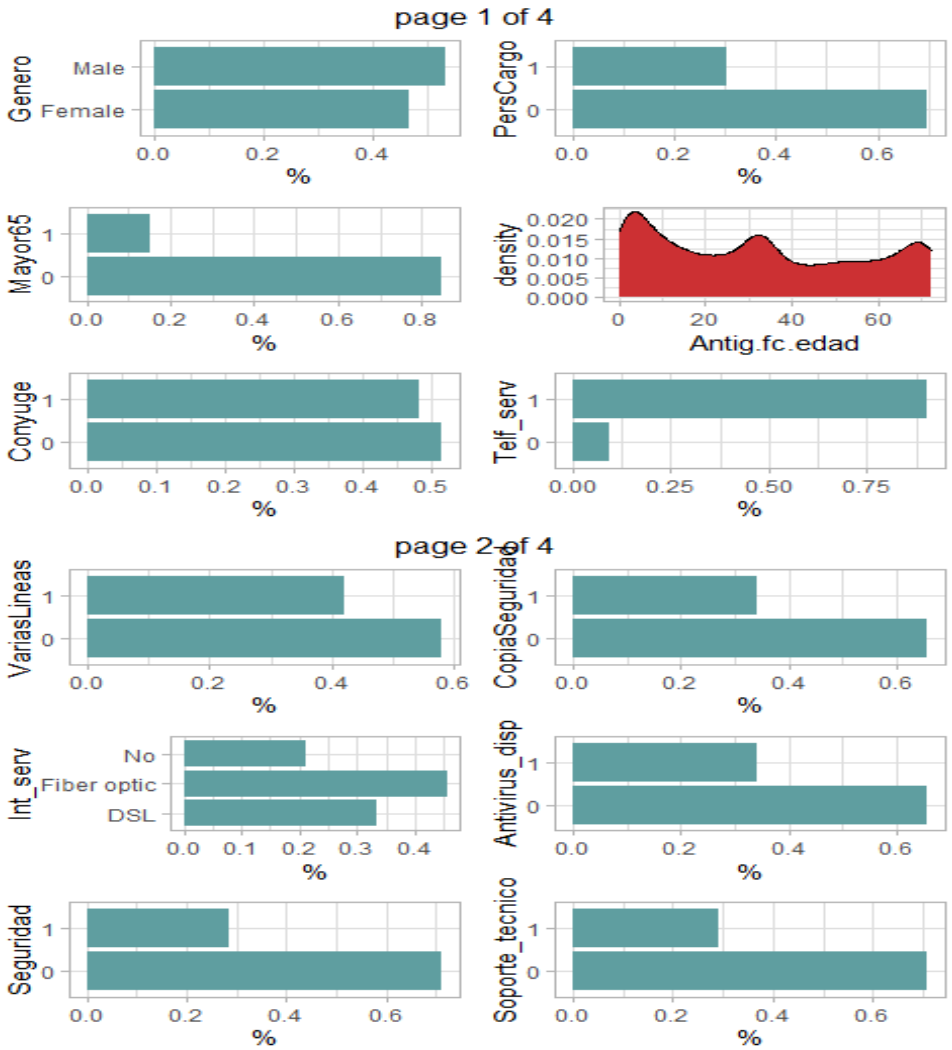
```
## $Soporte_tecnico
##      n  % val%
## 0 4509 71   71
## 1 1844 29   29
##
## $TV_streaming
##      n  % val%
## 0 3936 62   62
## 1 2417 38   38
##
## $Peliculas
##      n    % val%
## 0 3902 61.4 61.4
## 1 2451 38.6 38.6
##
## $Contrato
##                    n    % val%
## Month-to-month 3285 51.7 55.1
## One year       1261 19.8 21.2
## Two year       1413 22.2 23.7
## NA              394  6.2   NA
##
## $Fact_sinPapel
##       n    % val%
## 0  2483 39.1 40.8
## 1  3609 56.8 59.2
## NA  261  4.1   NA
##
## $MetodoPago
##                               n    % val%
## Bank transfer (automatic) 1283 20.2 21.8
## Credit card (automatic)   1290 20.3 21.9
## Electronic check          1972 31.0 33.5
## Mailed check              1336 21.0 22.7
## NA                         472  7.4   NA
##
## $Fuga
##      n    % val%
## 0 4667 73.5 73.5
## 1 1686 26.5 26.5
```

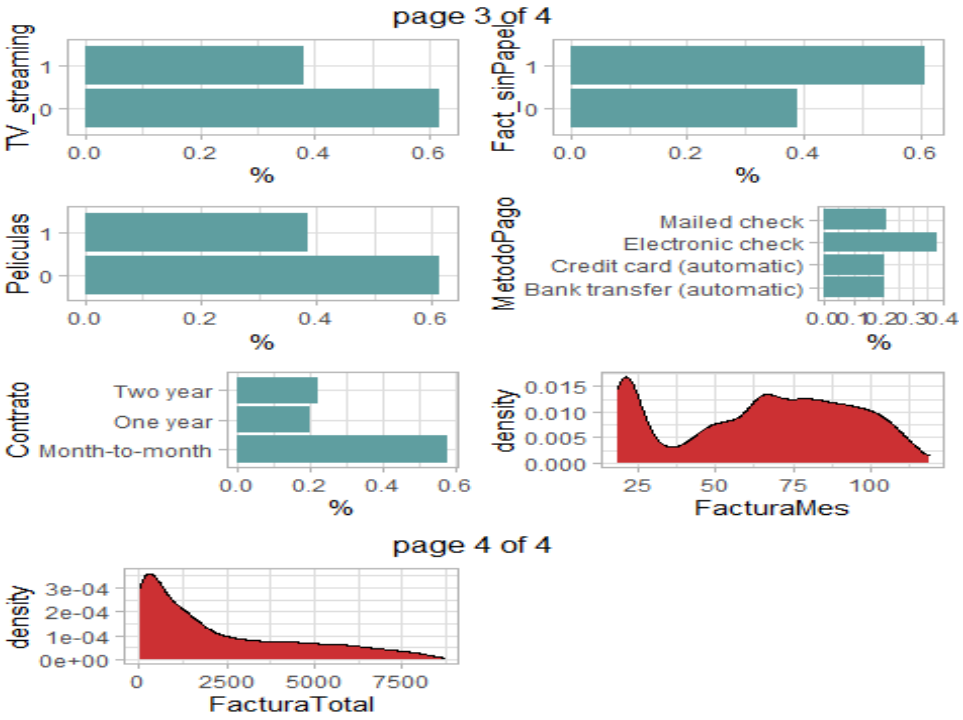
II) Observación de los estadísticos básicos de las variables predictoras e inspección gráfica de los datos depurados

```
# Observamos a detalle los valores de las variables

summary(input)
```

```
##     Genero     Mayor65  Conyuge  PersCargo Antig.fc.edad   Telf_serv
##  Female:2963   0:5397   0:3287   0:4439    Min.   : 0.00   0: 605
##  Male  :3390   1: 956   1:3066   1:1914    1st Qu.:10.00   1:5748
##                                            Median :32.00
##                                            Mean   :32.38
##                                            3rd Qu.:54.00
##                                            Max.   :72.00
##  VariasLineas        Int_serv    Seguridad CopiaSeguridad Antivirus_disp
##  0:3687       DSL        :2122   0:4532    0:4180         0:4180
##  1:2666       Fiber optic:2892   1:1821    1:2173         1:2173
##               No         :1339
##
##
##
##  Soporte_tecnico TV_streaming Peliculas           Contrato    Fact_sinPapel
##  0:4509          0:3936       0:3902    Month-to-month:3679   0:2483
##  1:1844          1:2417       1:2451    One year      :1261   1:3870
```

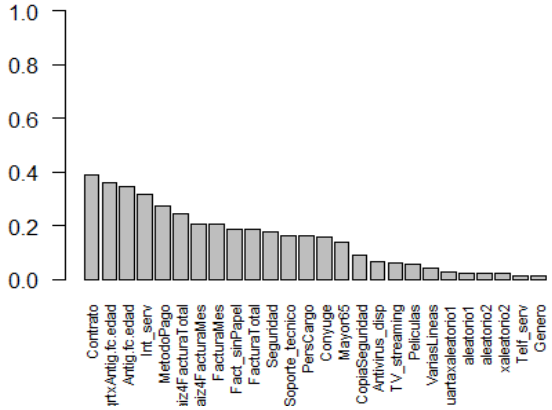
```
##                                         Two year      :1413
##
##
##
##                        MetodoPago     FacturaMes      FacturaTotal
##  Bank transfer (automatic):1283   Min.   : 18.25   Min.   :  18.8
##  Credit card (automatic)  :1290   1st Qu.: 41.10   1st Qu.: 396.3
##  Electronic check         :2444   Median : 68.75   Median :1387.3
##  Mailed check             :1336   Mean   : 64.68   Mean   :2268.7
##                                   3rd Qu.: 89.00   3rd Qu.:3778.0
##                                   Max.   :118.75   Max.   :8672.5

# Graficar el Dataset
par(mfrow = c(3,3))
lista_his <- dfplot_his(input)
gridExtra::marrangeGrob(lista_his, nrow=3, ncol=2)
```





III) Influencia de las variables originales y transformadas sobre la variable objetivo

```
# Evaluación sobre la importancia de las variables predictoras sobre la variable objetivo
graficoVcramer(input_obj, varObj)
```



IV) Modelo de referencia en el modelado manual de variables originales

```
# Primer modelo sin las variables transformadas
modeloInicial <- glm(varObj~., data = data_train[ ,c(1:21, 27)], family = binomial)
summary(modeloInicial)
```

```
##
## Call:
## glm(formula = varObj ~ ., family = binomial, data = data_train[,
##     c(1:21, 27)])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9365  -0.6693  -0.2951   0.7121   3.2456
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -3.556e-01  2.537e-01  -1.402 0.161037
## GeneroMale                         2.993e-02  7.657e-02   0.391 0.695855
## Mayor651                           1.425e-01  1.021e-01   1.395 0.162938
## Conyuge1                          -7.541e-02  9.119e-02  -0.827 0.408290
## PersCargo1                        -1.943e-01  1.052e-01  -1.846 0.064887 .
## Antig.fc.edad                     -3.055e-02  4.703e-03  -6.495 8.29e-11 ***
## Telf_serv1                        -5.350e-01  1.750e-01  -3.057 0.002234 **
## VariasLineas1                      1.565e-01  9.651e-02   1.621 0.104999
## Int_servFiber optic                9.115e-01  1.465e-01   6.223 4.87e-10 ***
## Int_servNo                        -8.179e-01  1.906e-01  -4.291 1.78e-05 ***
## Seguridad1                        -4.153e-01  1.034e-01  -4.015 5.95e-05 ***
## CopiaSeguridad1                   -1.737e-01  9.457e-02  -1.837 0.066238 .
## Antivirus_disp1                   -6.495e-02  9.633e-02  -0.674 0.500196
## Soporte_tecnico1                  -3.417e-01  1.027e-01  -3.328 0.000875 ***
## TV_streaming1                      2.551e-01  1.056e-01   2.415 0.015731 *
## Peliculas1                         2.294e-01  1.043e-01   2.200 0.027804 *
## ContratoOne year                  -7.581e-01  1.247e-01  -6.079 1.21e-09 ***
## ContratoTwo year                  -1.619e+00  2.162e-01  -7.490 6.91e-14 ***
## Fact_sinPapel1                     3.923e-01  8.711e-02   4.503 6.69e-06 ***
## MetodoPagoCredit card (automatic) -6.882e-02  1.401e-01  -0.491 0.623264
## MetodoPagoElectronic check         3.815e-01  1.124e-01   3.393 0.000690 ***
## MetodoPagoMailed check             1.213e-01  1.364e-01   0.889 0.373942
## FacturaMes                         2.083e-03  4.697e-03   0.444 0.657361
## FacturaTotal                      -6.154e-06  5.616e-05  -0.110 0.912741
## aleatorio1                         1.437e-01  1.316e-01   1.091 0.275139
## aleatorio2                        -7.623e-02  1.316e-01  -0.579 0.562316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5882.3  on 5082  degrees of freedom
## Residual deviance: 4233.1  on 5057  degrees of freedom
## AIC: 4285.1
##
## Number of Fisher Scoring iterations: 6
```

```r
# Volver a ajustar el modelo inicial
modeloInicial <- glm(varObj~., data = data_train[ ,c(1:21, 27)], family=binomial)
summary(modeloInicial)
```

```
##
## Call:
## glm(formula = varObj ~ ., family = binomial, data = data_train[,
##     c(1:21, 27)])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9314  -0.6691  -0.2952   0.7208   3.2513
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -4.488e-01  2.408e-01  -1.864 0.062351 .
## GeneroMale                       2.750e-02  7.647e-02   0.360 0.719122
## Mayor651                         1.552e-01  1.019e-01   1.523 0.127772
## Conyuge1                        -6.675e-02  9.102e-02  -0.733 0.463352
## PersCargo1                      -2.051e-01  1.050e-01  -1.952 0.050920 .
## Antig.fc.edad                   -3.008e-02  4.701e-03  -6.397 1.58e-10 ***
## Telf_serv1                      -5.533e-01  1.743e-01  -3.173 0.001507 **
## VariasLineas1                    1.615e-01  9.641e-02   1.676 0.093819 .
## Int_servFiber optic              9.316e-01  1.459e-01   6.386 1.70e-10 ***
## Int_servNo                      -8.580e-01  1.896e-01  -4.526 6.00e-06 ***
## Seguridad1                      -4.273e-01  1.032e-01  -4.142 3.45e-05 ***
## CopiaSeguridad1                 -1.735e-01  9.448e-02  -1.837 0.066260 .
## Antivirus_disp1                 -7.233e-02  9.613e-02  -0.752 0.451781
## Soporte_tecnico1                -3.579e-01  1.024e-01  -3.495 0.000474 ***
## TV_streaming1                    2.599e-01  1.054e-01   2.465 0.013685 *
## Peliculas1                       2.404e-01  1.040e-01   2.310 0.020876 *
## ContratoOne year                -7.623e-01  1.247e-01  -6.115 9.67e-10 ***
## ContratoTwo year                -1.627e+00  2.160e-01  -7.530 5.09e-14 ***
## Fact_sinPapel1                   4.128e-01  8.664e-02   4.765 1.89e-06 ***
## MetodoPagoNot Automatic Payments  3.449e-01  8.637e-02   3.994 6.50e-05 ***
## FacturaMes                       2.758e-03  4.674e-03   0.590 0.555214
## FacturaTotal                    -1.052e-05  5.612e-05  -0.187 0.851283
## aleatorio1                       1.386e-01  1.313e-01   1.055 0.291324
## aleatorio2                      -7.133e-02  1.314e-01  -0.543 0.587106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5882.3  on 5082  degrees of freedom
## Residual deviance: 4238.8  on 5059  degrees of freedom
## AIC: 4286.8
##
## Number of Fisher Scoring iterations: 6
```

```r
# Segundo modelo
modelo2 <- update(modeloInicial,.~.-FacturaTotal-aleatorio2-Genero-FacturaMes-
Conyuge-Antivirus_disp-aleatorio1)

# Tercer modelo
modelo3 <- update(modelo2,.~.-Mayor65)

# Cuarto modelo
modelo4 <- update(modelo3,.~.-CopiaSeguridad)

# Quinto modelo
modelo5 <- update(modelo4,.~.-VariasLineas)

# Sexto modelo
modelo6 <- update(modelo5,.~.-Peliculas)
```

```r
# Séptimo modelo
modelo7 <- update(modelo6,.~.-PersCargo)

# Octavo modelo
modelo8 <- update(modelo7,.~.-Telf_serv)

# Noveno modelo
modelo9 <- update(modelo8,.~.-Soporte_tecnico)
# Decimo modelo
modelo10 <- update(modelo9,.~.-TV_streaming)

# Undecimo modelo
modelo11 <- update(modelo10,.~.-MetodoPago)

# Duodecimo modelo
modelo12 <- update(modelo11,.~.-Seguridad)

# Decimotercer modelo
modelo13 <- update(modelo12,.~.-Fact_sinPapel)
```

V) Parámetros del modelo para su interpretación final

```r
modFinal <- glm(formula(modeloManual),data = todo_obj, family = binomial)

coef(modFinal)
```

```
##                       (Intercept)                      Antig.fc.edad
##                         -0.7029580                         -0.0311027
##               Int_servFiber optic                         Int_servNo
##                          0.9340172                         -1.0611609
##                         Seguridad1                  Soporte_tecnico1
##                         -0.4592007                         -0.4041569
##                       TV_streaming1                  ContratoOne year
##                          0.4213377                         -0.7793340
##                    ContratoTwo year                    Fact_sinPapel1
##                         -1.5827798                          0.4387844
## MetodoPagoNot Automatic Payments
##                          0.3312746
```

```r
logistic.display(modFinal)
```

```
##
## Logistic regression predicting varObj : 1 vs 0
##
##                                                                    crude OR(95%CI)
## Antig.fc.edad (cont. var.)                                         0.96 (0.96,0.96)
##
## Int_serv: ref.=DSL
##    Fiber optic                                                     3.07 (2.69,3.5)
##    No                                                              0.36 (0.28,0.45)
##
## Seguridad: 1 vs 0                                                  0.36 (0.32,0.42)
##
## Soporte_tecnico: 1 vs 0                                            0.4 (0.34,0.46)
##
## TV_streaming: 1 vs 0                                               1.31 (1.17,1.47)
##
## Contrato: ref.=Month-to-month
##    One year                                                        0.18 (0.15,0.21)
##    Two year                                                        0.04 (0.03,0.05)
##
## Fact_sinPapel: 1 vs 0                                              2.53 (2.23,2.87)
##
## MetodoPago: Not Automatic Payments vs Automatic Payments           2.73 (2.41,3.1)
##
##                                                                    adj. OR(95%CI)
```

```
## Antig.fc.edad (cont. var.)                                  0.97 (0.97,0.97)
##
## Int_serv: ref.=DSL
##    Fiber optic                                              2.54 (2.18,2.98)
##    No                                                       0.35 (0.27,0.45)
##
## Seguridad: 1 vs 0                                           0.63 (0.53,0.75)
##
## Soporte_tecnico: 1 vs 0                                     0.67 (0.56,0.79)
##
## TV_streaming: 1 vs 0                                        1.52 (1.31,1.78)
##
## Contrato: ref.=Month-to-month
##    One year                                                 0.46 (0.37,0.57)
##    Two year                                                 0.21 (0.14,0.3)
##
## Fact_sinPapel: 1 vs 0                                       1.55 (1.33,1.8)
##
## MetodoPago: Not Automatic Payments vs Automatic Payments 1.39 (1.2,1.62)
##
##                                                             P(Wald's test)
## Antig.fc.edad (cont. var.)                                  < 0.001
##
## Int_serv: ref.=DSL
##    Fiber optic                                              < 0.001
##    No                                                       < 0.001
##
## Seguridad: 1 vs 0                                           < 0.001
##
## Soporte_tecnico: 1 vs 0                                     < 0.001
##
## TV_streaming: 1 vs 0                                        < 0.001
##
## Contrato: ref.=Month-to-month
##    One year                                                 < 0.001
##    Two year                                                 < 0.001
##
## Fact_sinPapel: 1 vs 0                                       < 0.001
##
## MetodoPago: Not Automatic Payments vs Automatic Payments < 0.001
##
##                                                             P(LR-test)
## Antig.fc.edad (cont. var.)                                  < 0.001
##
## Int_serv: ref.=DSL                                          < 0.001
##    Fiber optic
##    No
##
## Seguridad: 1 vs 0                                           < 0.001
##
## Soporte_tecnico: 1 vs 0                                     < 0.001
##
## TV_streaming: 1 vs 0                                        < 0.001
##
## Contrato: ref.=Month-to-month                               < 0.001
##    One year
##    Two year
##
## Fact_sinPapel: 1 vs 0                                       < 0.001
##
## MetodoPago: Not Automatic Payments vs Automatic Payments < 0.001
##
## Log-likelihood = -2674.2628
## No. of observations = 6353
## AIC value = 5370.5256
```