

```
library(clickR)
summary(datos_ana)
```

```
clickR:::numeros(datos_ana$antiguedad_poliza)
```

```
datos_ana$antiguedad_poliza<-clickR:::numeros(datos_ana$antiguedad_poliza)
datos_ana<-fix_factors(datos_ana, k=55)
datos_ana<-fix_NA(datos_ana)
track_changes(datos_ana)
mine.plot(datos_ana)
descriptive(datos_ana)
```

```
#Comparación de las variables con la variable respuesta Fraude.
plot(datos_ana$edad_conductor1, datos_ana$antiguedad_poliza)
plot(datos_ana$edad_conductor1, datos_ana$potencia)
plot(datos_ana$edad_conductor1, datos_ana$cilindrada)
plot(datos_ana$aceptoculpasinantecedentes, datos_ana$antiguedad_poliza)
log(importesiniestro)
#variables num?ricas
boxplot(datos_ana$fechaintervencion~datos_ana$respuesta_dicot1)
boxplot(datos_ana$importesiniestro~datos_ana$respuesta_dicot1)
boxplot(datos_ana$importeinvertido~datos_ana$respuesta_dicot1)
plot(datos_ana$importeahorrado~datos_ana$respuesta_dicot1)
boxplot(datos_ana$edad_conductor1~datos_ana$respuesta_dicot1)
boxplot(datos_ana$antiguedad_poliza~datos_ana$respuesta_dicot1)
boxplot(datos_ana$dias_notificacion~datos_ana$respuesta_dicot1)
plot(datos_ana$provinciaid~datos_ana$respuesta_dicot1)
boxplot(datos_ana$valorvehiculomercado~datos_ana$respuesta_dicot1)
boxplot(datos_ana$valorvehiculofabrica~datos_ana$respuesta_dicot1)
boxplot(datos_ana$potencia~datos_ana$respuesta_dicot1)
boxplot(datos_ana$cilindrada~datos_ana$respuesta_dicot1)
boxplot(datos_ana$pesovehiculo~datos_ana$respuesta_dicot1)
boxplot(datos_ana$longitud~datos_ana$respuesta_dicot1)
plot(datos_ana$aceptoculpasinantecedentes~datos_ana$respuesta_dicot1)

plot(datos_ana$longitud~datos_ana$respuesta_dicot1)
```

```
#variables categ?ricas
plot(datos_ana$sexo~datos_ana$respuesta_dicot1)
plot(datos_ana$tipodocumento~datos_ana$respuesta_dicot1)
plot(datos_ana$garantia_agrupada~datos_ana$respuesta_dicot1)
plot(datos_ana$formapago_agrupado~datos_ana$respuesta_dicot1)
plot(datos_ana$scoring~datos_ana$respuesta_dicot1)
```

```
# Mostrar la tabla de contingencia
tabla_contingencia <- table(datos_ana$sexo, datos_ana$respuesta_dicot1)
print(tabla_contingencia)
```

```
tabla_contingencia <- table(datos_ana$tipodocumento, datos_ana$respuesta_dicot1)
print(tabla_contingencia)
```

```

rm()
tabla_contingencia <- table(datos_ana$garantia_agrupada, datos_ana$respuesta_dicot1)
print(tabla_contingencia)

tabla_contingencia <- table(datos_ana$formapago_agrupado, datos_ana$respuesta_dicot1)
print(tabla_contingencia)

tabla_contingencia <- table(datos_ana$scoring, datos_ana$respuesta_dicot1)
print(tabla_contingencia)

tabla_contingencia <- table(datos_ana$provinciaid, datos_ana$respuesta_dicot1)
print(tabla_contingencia)

tabla_contingencia <- table(datos_ana$acceptoculpasinantecedentes, datos_ana$respuesta_dicot1)
print(tabla_contingencia)

#eliminar observaciones
Base1<-datos_ana[datos_ana$edad_conductor1 >= 18, ]
Base2<-Base1[Base1$cilindrada >=60, ]
Base3<-Base2[Base2$pesovehiculo >= 500 & Base2$pesovehiculo <= 3000, ]
Base4<-Base3[Base3$valorvehiculofabrica <=85000, ]
Base5<-Base4[Base4$importesiniestro >0, ]

bananew<-Base5[,c(1,6:12,14:22)]

#imputación de datos
library(missForest)
bbddimp<-missForest(bananew)
summary(bbddimp)
bbddimp_final<-(bbddimp$ximp)
summary(bbddimp_final)
descriptive(bbddimp_final)

##Comparación de las variables con la variable respuesta Fraude despues de imputación
plot(bbddimp_final$provinciaid~bbddimp_final$respuesta_dicot1)
boxplot(bbddimp_final$valorvehiculomercado~bbddimp_final$respuesta_dicot1)
boxplot(bbddimp_final$valorvehiculofabrica~bbddimp_final$respuesta_dicot1)
boxplot(bbddimp_final$potencia~bbddimp_final$respuesta_dicot1)
boxplot(bbddimp_final$cilindrada~bbddimp_final$respuesta_dicot1)
boxplot(bbddimp_final$pesovehiculo~bbddimp_final$respuesta_dicot1)
boxplot(bbddimp_final$longitud~bbddimp_final$respuesta_dicot1)

plot(bbddimp_final$longitud~bbddimp_final$respuesta_dicot1)

#variables categóricas

```

```

plot(bbddimp_final$scoring~bbddimp_final$respuesta_dicot1)

# Mostrar la tabla de contingencia
tabla_contingencia <- table(bbddimp_final$scoring, bbddimp_final$respuesta_dicot1)
print(tabla_contingencia)

tabla_contingencia <- table(bbddimp_final$provinciaid, bbddimp_final$respuesta_dicot1)
print(tabla_contingencia)

#Modelo logistico en R

mod1 <- glm( respuesta_dicot1~ sexo + tipodocumento + edad_conductor1 + garantia_agrupada +
antiguedad_poliza + dias_notificacion + formapago_agrupado + provinciaid +
valorvehiculomercado + valorvehiculofabrica + potencia + cilindrada +pesovehiculo + longitud +
scoring + aceptoculpasinantecedentes, data=bbddimp_final, family=binomial)
#resumen del modelo
summary(mod1)
#resumen del modelo, s?lo coeficientes
summary(mod1)$coefficients
library(repmo)
report(mod1)
library(car)
vif_mod1 <- vif(mod1)
print(vif_mod1)

#Modelo 2
mod2 <- glm( respuesta_dicot1~ sexo + tipodocumento + edad_conductor1 + garantia_agrupada +
antiguedad_poliza + dias_notificacion + formapago_agrupado + valorvehiculomercado +
valorvehiculofabrica + potencia + cilindrada +pesovehiculo + longitud + scoring +
aceptoculpasinantecedentes, data=bbddimp_final, family=binomial)
#resumen del modelo
summary(mod2)
report(mod2)
vif_mod2 <- vif(mod2)
print(vif_mod2)

#Modelo 3
mod3 <- glm( respuesta_dicot1~ sexo + tipodocumento + edad_conductor1 + garantia_agrupada +
antiguedad_poliza + dias_notificacion + formapago_agrupado + valorvehiculofabrica + potencia +
cilindrada +pesovehiculo + longitud + scoring + aceptoculpasinantecedentes, data=bbddimp_final,
family=binomial)
#resumen del modelo
summary(mod3)
report(mod3)
vif_mod3 <- vif(mod3)
print(vif_mod3)

#Modelo 4
mod4 <- glm( respuesta_dicot1~ sexo + tipodocumento + edad_conductor1 + garantia_agrupada +

```

```

antiguedad_poliza + dias_notificacion + formapago_agrupado + valorvehiculofabrica + potencia +
cilindrada + longitud + scoring + aceptoculpasinantecedentes, data=bbddimp_final,
family=binomial)
#resumen del modelo
summary(mod4)
report(mod4)
vif_mod4 <- vif(mod4)
print(vif_mod4)

#Modelo 5
mod5 <- glm( respuesta_dicot1~ edad_conductor1 + garantia_agrupada + dias_notificacion +
cilindrada + potencia + scoring + aceptoculpasinantecedentes, data=bbddimp_final,
family=binomial)
#resumen del modelo
summary(mod5)
report(mod5)
vif_mod5 <- vif(mod5)
print(vif_mod5)

#Modelo 7
mod6 <- glm( respuesta_dicot1~ edad_conductor1 + garantia_agrupada + potencia + cilindrada +
scoring + aceptoculpasinantecedentes, data=bbddimp_final, family=binomial)
#resumen del modelo
summary(mod7)
report(mod7)
vif_mod7 <- vif(mod7)
print(vif_mod7)

#Modelo 6
mod6 <- glm( respuesta_dicot1~ edad_conductor1 + garantia_agrupada + dias_notificacion +
potencia + cilindrada + scoring + aceptoculpasinantecedentes, data=bbddimp_final,
family=binomial)
#resumen del modelo
summary(mod6)
report(mod6)
vif_mod6 <- vif(mod6)
print(vif_mod6)

#Modelo 7
mod7 <- glm( respuesta_dicot1~ garantia_agrupada, data=bbddimp_final, family=binomial)
#resumen del modelo
summary(mod7)
report(mod7)
vif_mod7 <- vif(mod7)
print(vif_mod7)

library(rms)
mod9 <- logistf( respuesta_dicot1~ edad_conductor1 + garantia_agrupada + dias_notificacion +
potencia + cilindrada + aceptoculpasinantecedentes, data=bbddimp_final)
report(mod9)

```

```

exp(mod9$coefficients)
exp(mod9$ci.lower)
exp(mod9$ci.upper)
summary(mod9)

#eliminación de separación perfecta
xtabs(~bbddimp_final$respuesta_dicot1+bbddimp_final$garantia_agrupada)
xtabs(~bbddimp_final$respuesta_dicot1+bbddimp_final$scoring)
library(logistf)
mod8 <- logistf( respuesta_dicot1~ edad_conductor1 + garantia_agrupada + dias_notificacion +
potencia + cilindrada + scoring + aceptoculpasinantecedentes, data=bbddimp_final)
report(mod8)
exp(mod8$coefficients)
exp(mod8$ci.lower)
exp(mod8$ci.upper)
summary(mod8)
extractAIC(mod8)
drop1(mod8)
library(MuMIn)
AICc(mod8)
AICc(mod6)

predict(mod8)
predicciones<- (predict(mod8))
library(pROC)
roc(bbddimp_final$respuesta_dicot1~predicciones)
curvaroc<-(roc(bbddimp_final$respuesta_dicot1~predicciones))
plot(curvaroc)

```