



Santander Customer Satisfaction



Jose Domingo Mateo Vázquez
Supervisor - J.M. Caro Bravo

Noviembre de 2017
Máster MECOFIN

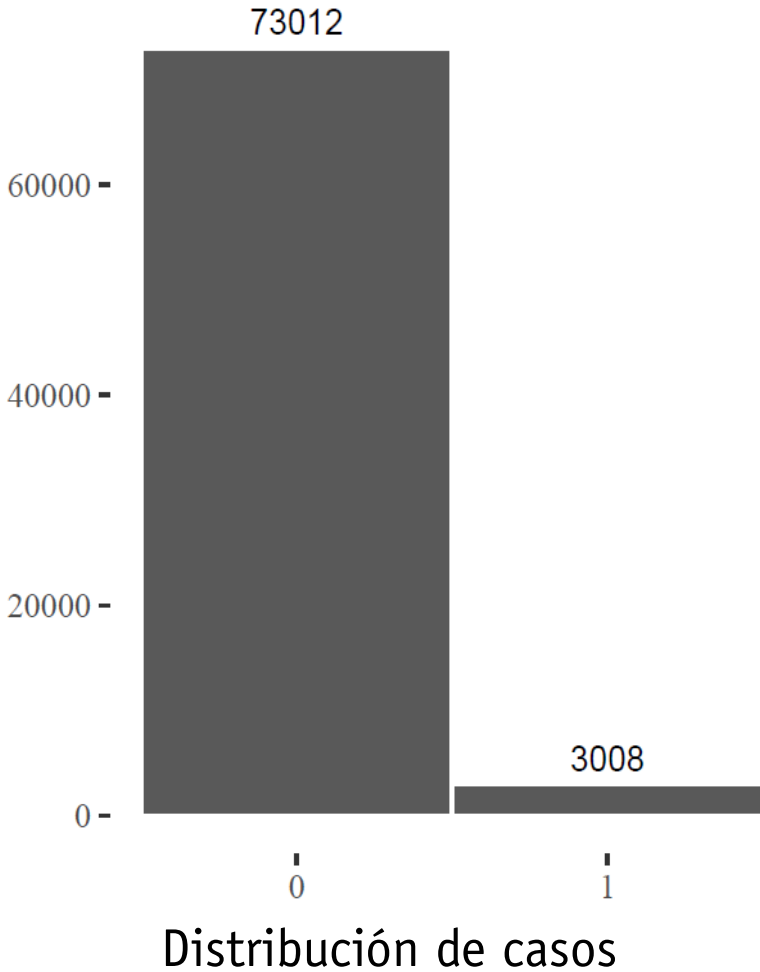
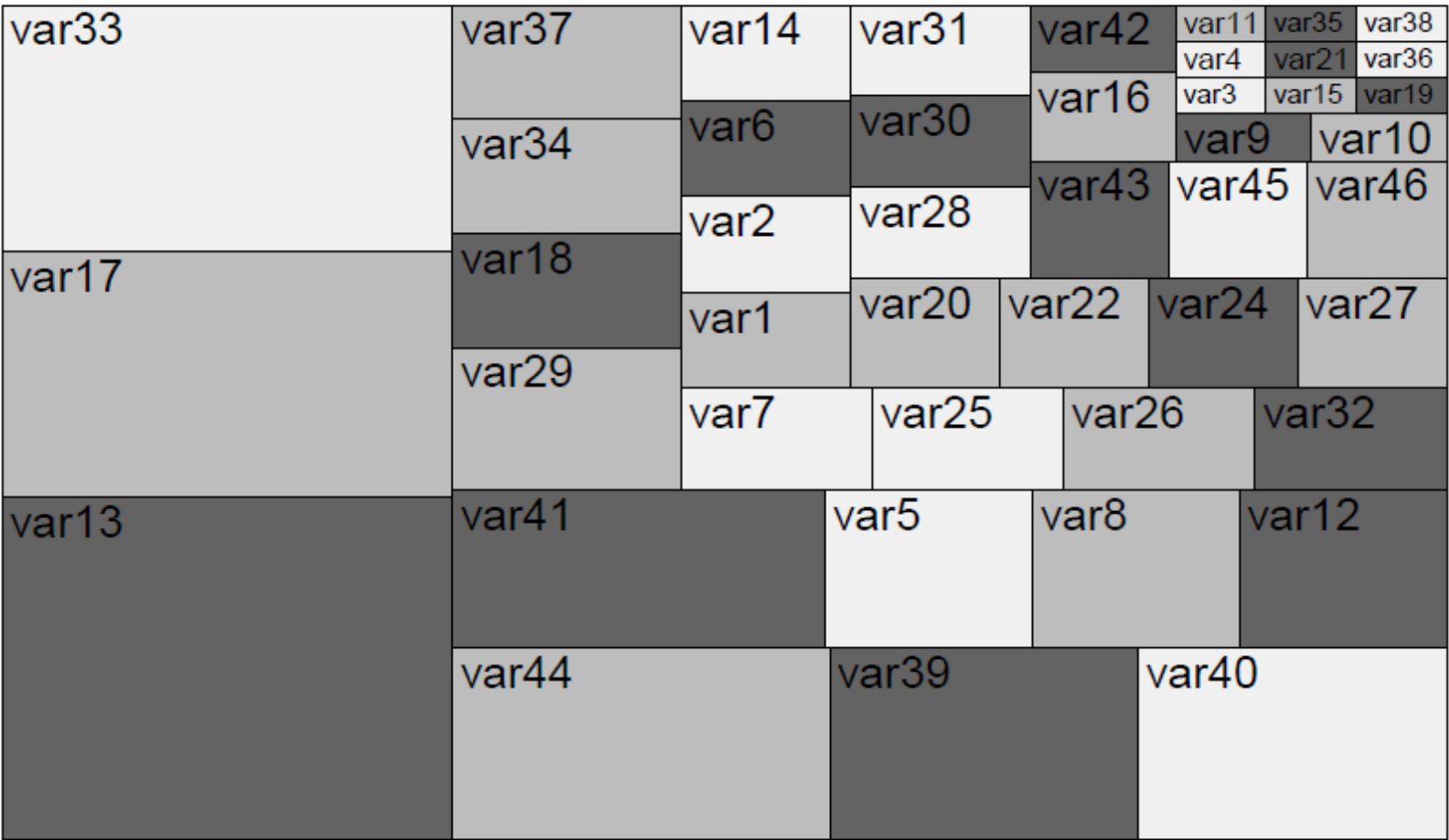
Introducción



Kaggle.com
Santander Customer Satisfaction
Objetivos



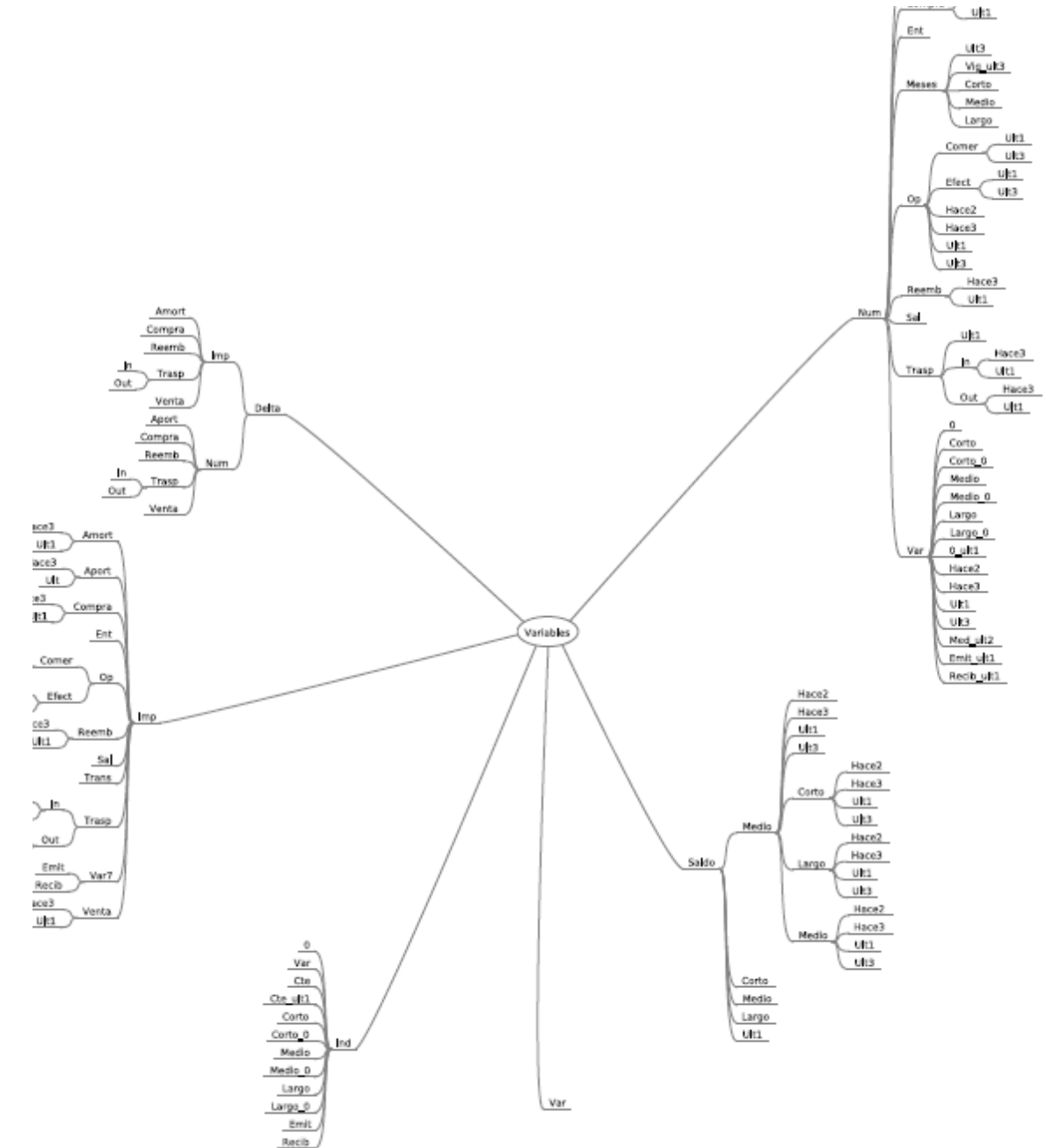
Análisis exploratorio



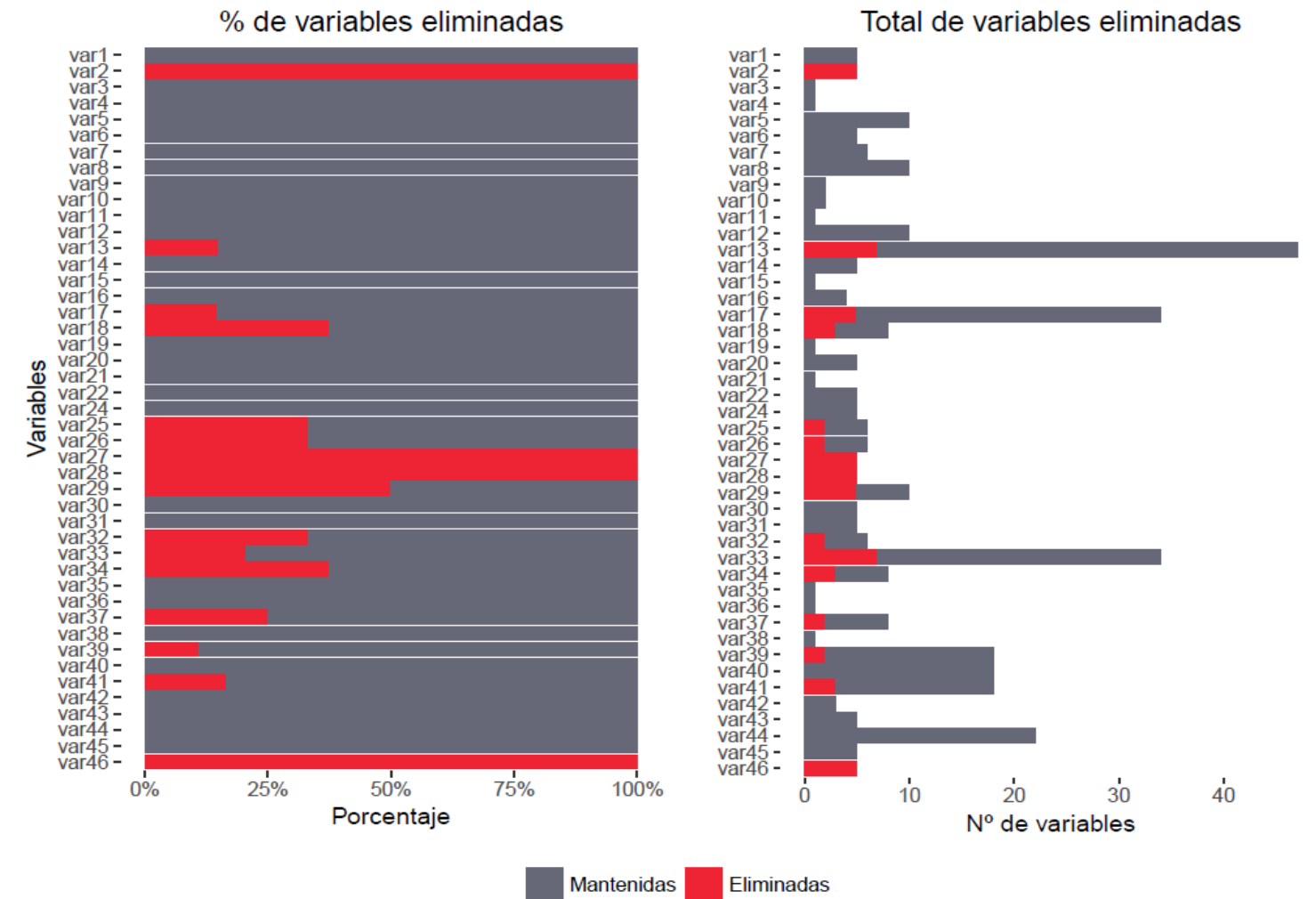
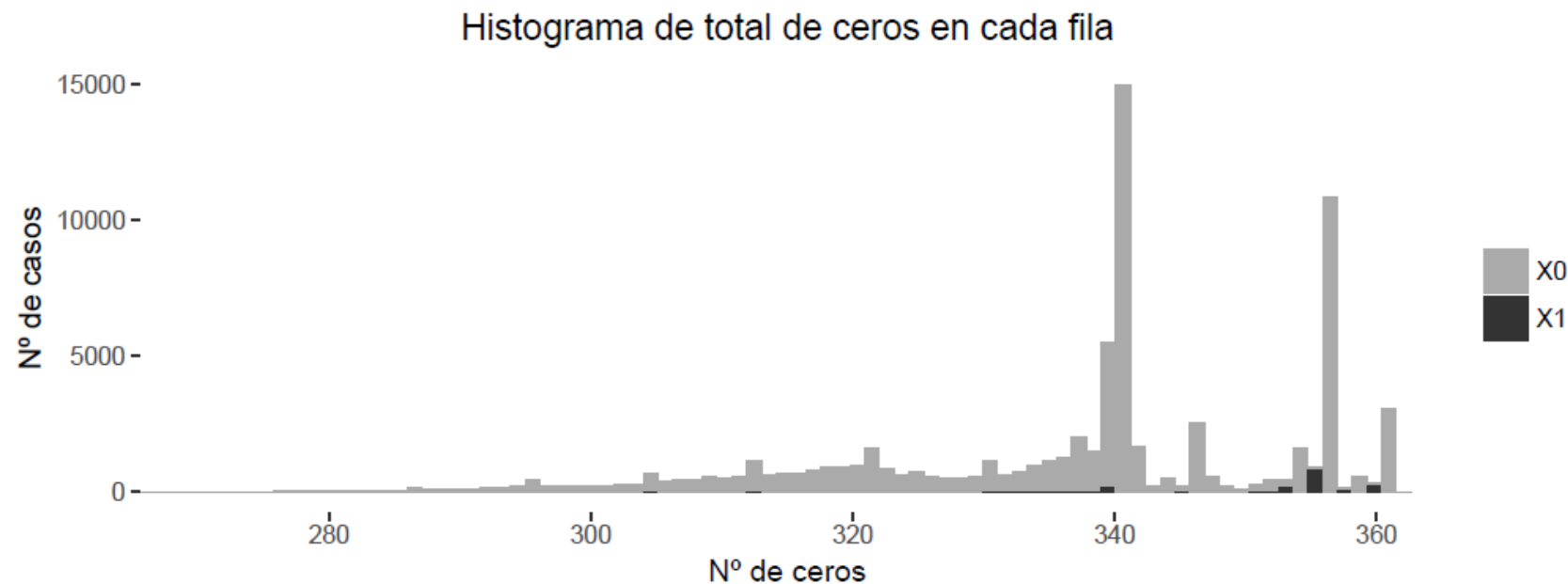
150.000 observaciones
370 Variables
Forman 46 Grupos de variables

Tipos de variables y etiquetas

Tipo de variable	Categóricas	Continuas	Discretas
Total de variables	3	123	244
Prefijos asociados	Sin prefijo	imp saldo delta	imp ind num delta

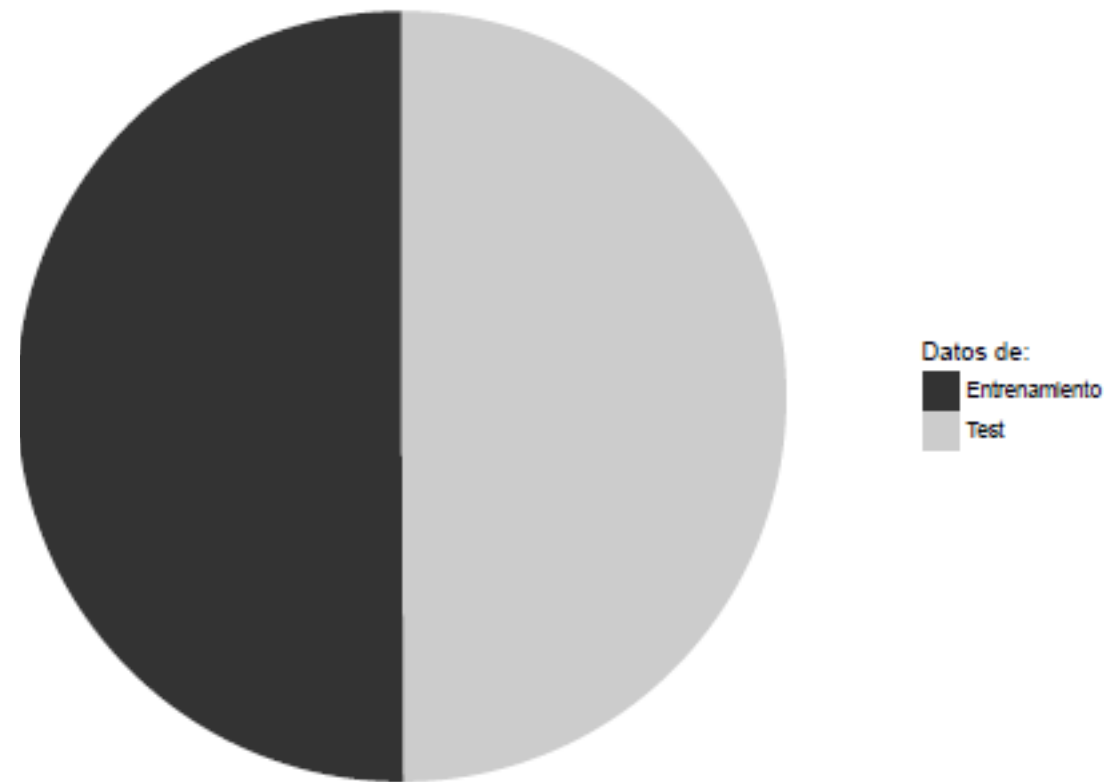


Transformaciones

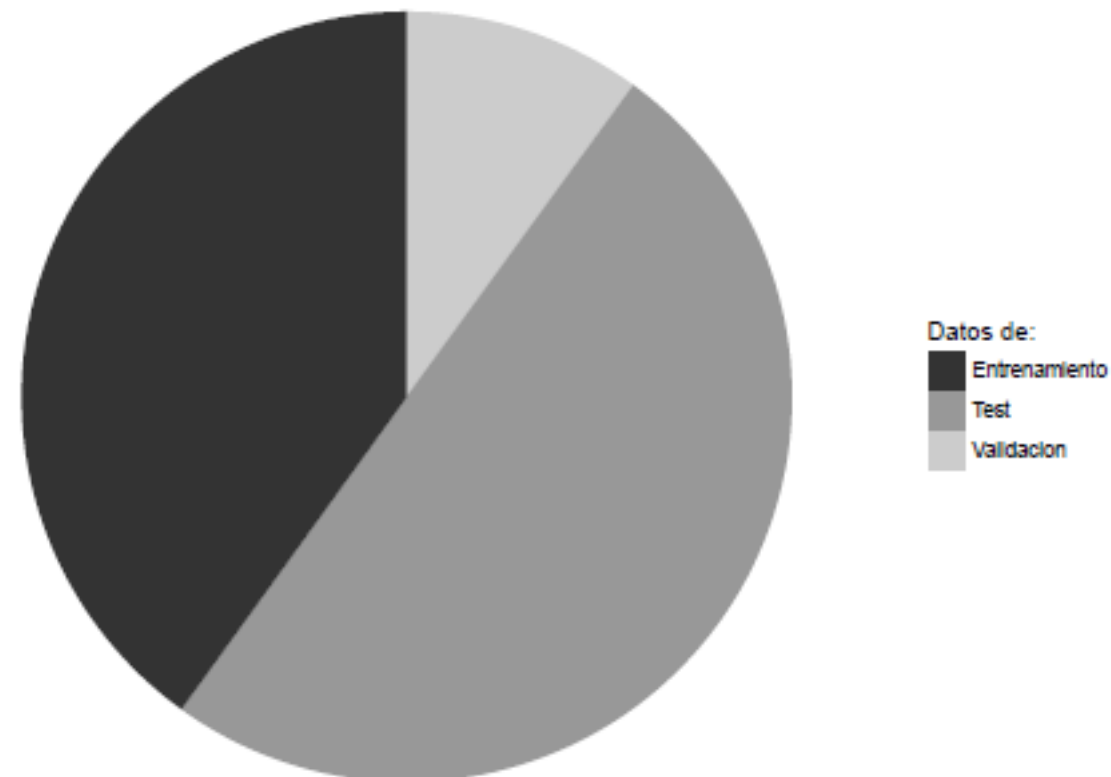


Variables iguales a 0
Creación de nuevas variables
Eliminación de variables
Tratamiento de los casos perdidos
Tratamiento de la descompensación

División Test-Entrenamiento



División Validación-Test-Entrenamiento



Métodos de compensación

- Downsampling
- Upsampling
- Smote
- ROSE

Métodos de validación

- K-Fold Cross Validation
- Random Cross Validation

	Regresión logística	KNN	Naïve Bayes	SVM	Random Forest	XGBoost
TRANSFORMACIONES						
Estandarizado	X	X	X	X		
Logaritmo	X					
Eliminación constantes	X					
Datos perdidos	X					X
Centrado/Normalizado	X					
DESCOMPENSACIÓN						
Downsampling				X		
Upsampling		X				
Smote			X			
ROSE	X					
SELECCIÓN VARIABLES						
Selección sucesiva	X	X	X	X		
Variables importantes					X	
Todas las variables						X
VALIDACIÓN						
K-fold CV				X	X	X
Random CV	X	X	X			
PUNTUACIÓN OBTENIDA						
Pública	0.797883	0.504230	0.660662	0.677769	0.619606	0.840522
Privada	0.781125	0.500277	0.688245	0.653400	0.598594	0.826565

Resultados

Mejor modelo - XGBoost

Posición 601/5123

AUC 0.826565

-0.002507 que el ganador

