**Caso de Uso: Aplican crédito a plazos**

**Grupo**: II

**Integrantes**:

* Céspedes Chaupis Roberto
* Guzmán Quintan Deby
* Tacure Purizaca Leonela

**Objetivo**:

Predecir el incumplimiento dentro de los 12 meses al momento de la solicitud del crédito a plazos.

**Análisis de Variables**

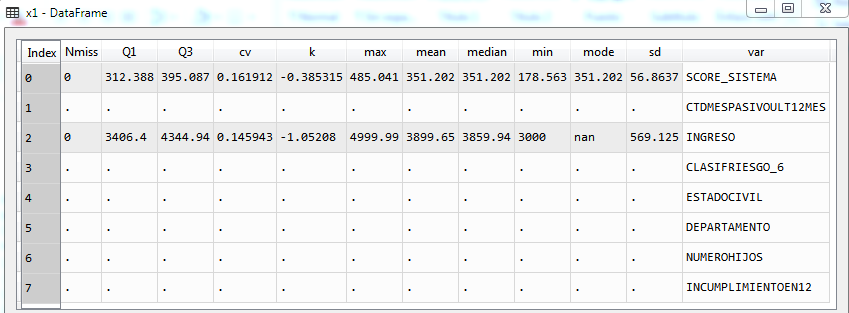
Contamos con la Data de 2052 clientes para entrenar un modelo para predecir el incumplimiento y ayudar en la toma de decisión de aprobación o negación de las solicitudes del crédito.

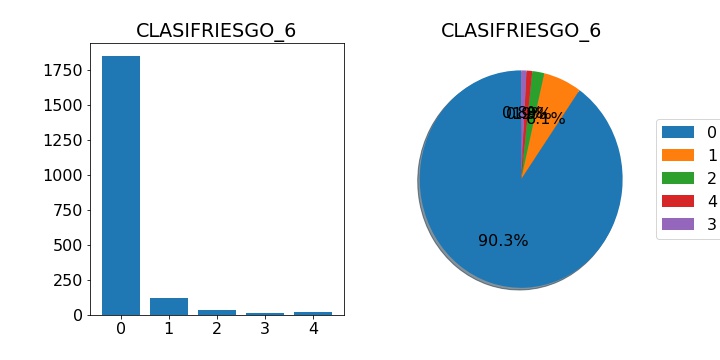
Para el desarrollo del caso solo se utilizaron las siguientes variables:

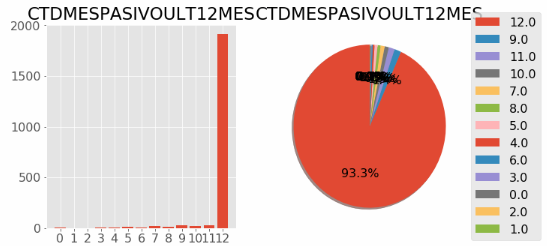
* SCORE\_SISTEMA
* CTDMESPASIVOULT12MES
* INGRESO
* CLASIFRIESGO\_6
* ESTADOCIVIL
* DEPARTAMENTO
* NUMEROHIJOS
* INCUMPLIMIENTOEN12 (**Variable Target**)

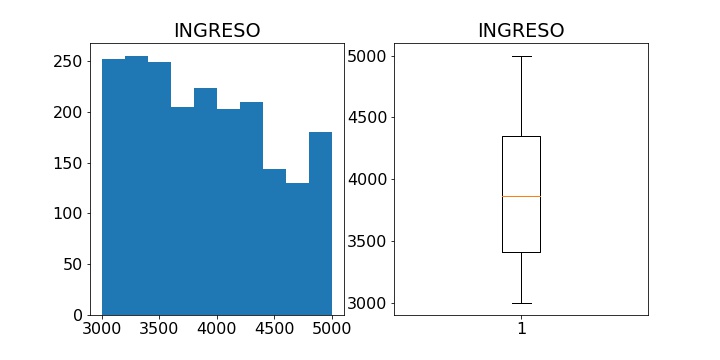
Las demás variables como ID, MES, DELTA\_BANCOS y TIPO se omitieron por tener un valor constante y no influir en el tratamiento de la data.

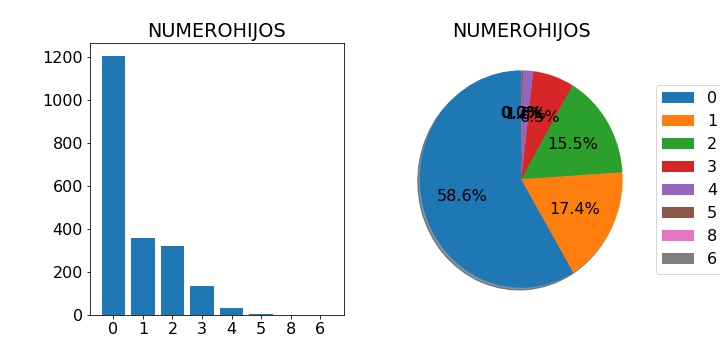
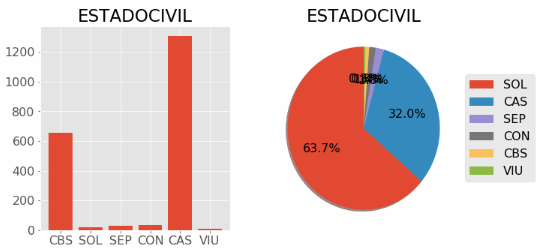
1. **Análisis Univariado**

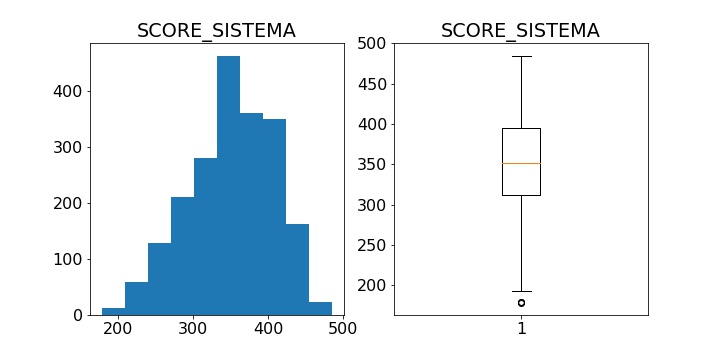
* Variables no categóricas:
* Variables categóricas.

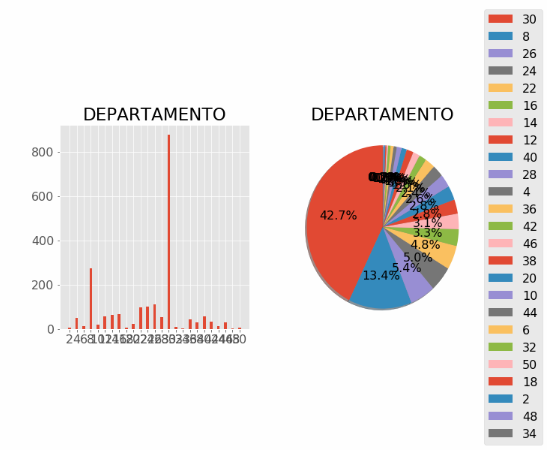




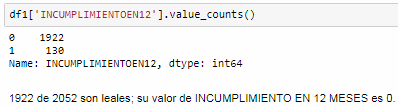


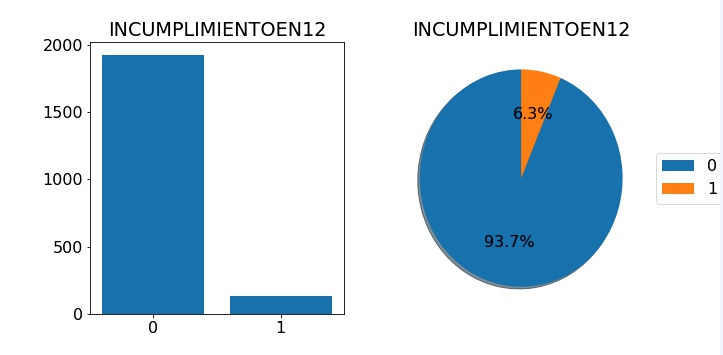
 





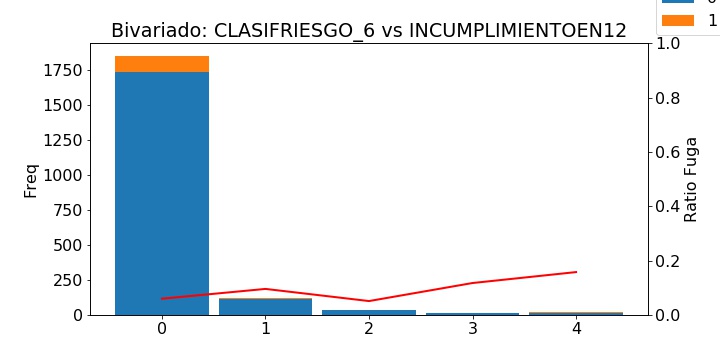
* Variable Categórica Target: *Observamos que 1922 de 2052 clientes son leales; su valor de INCUMPLIMIENTO EN 12 MESES es 0. Contando con un 6.3% de Tasa de Deserción.*

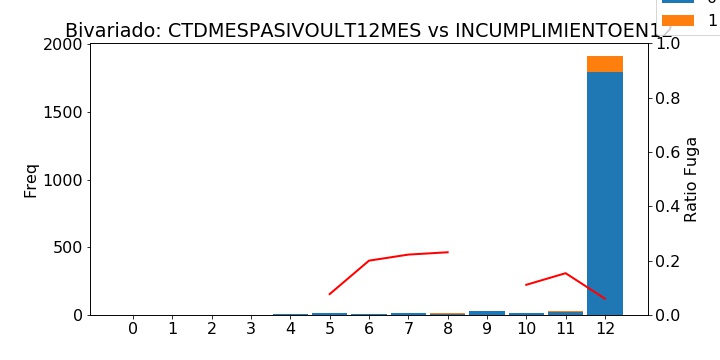


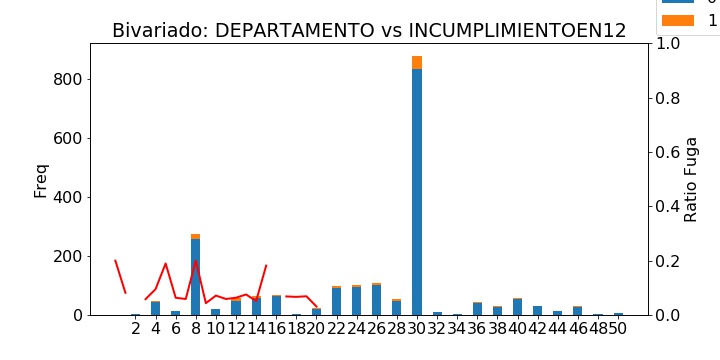


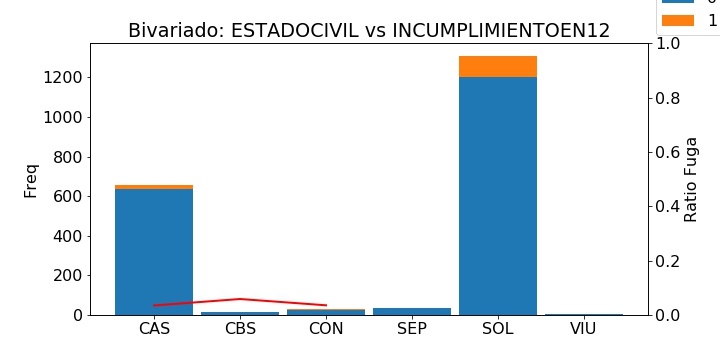
1. **Análisis Bivariado**

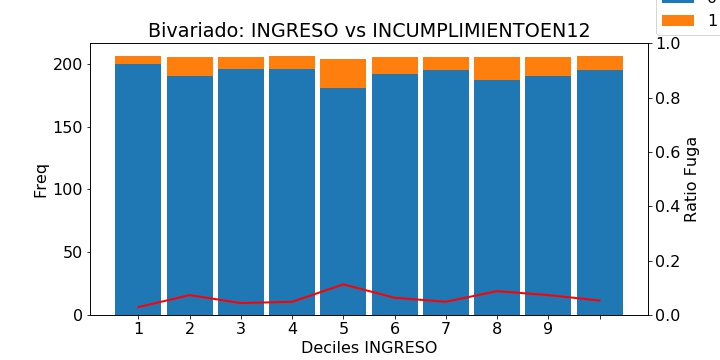
En este análisis realizamos un comparativo de las variables independientes con la variable predicha (Target).

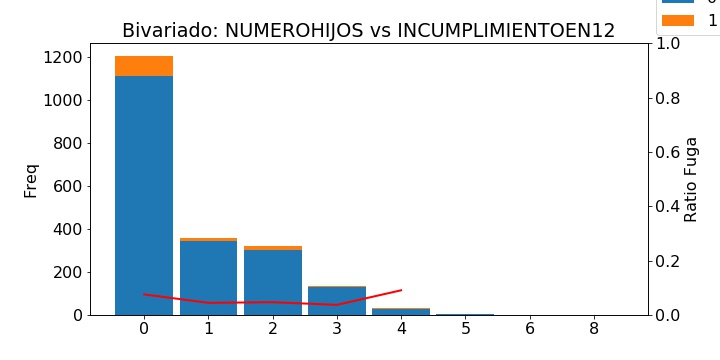


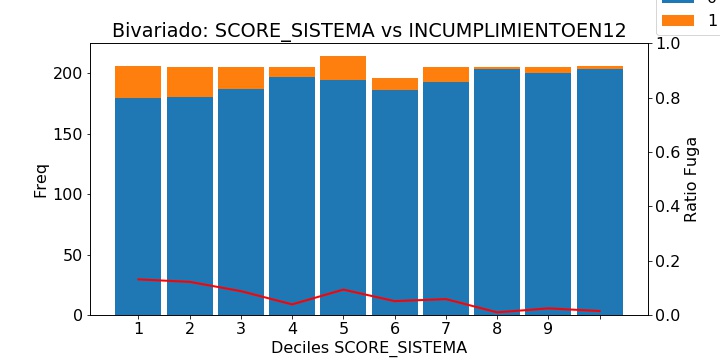








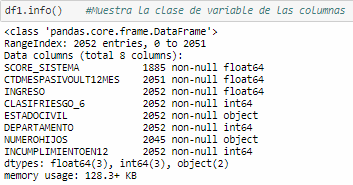


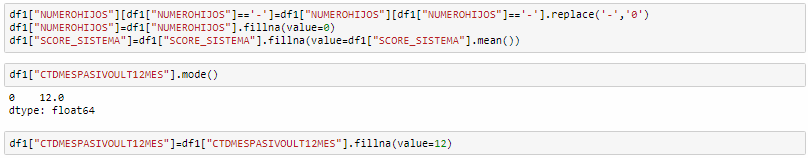


1. **Tratamiento de los datos**

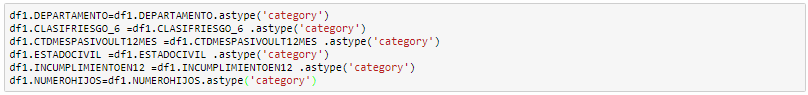
Como observamos en los análisis univariado y bivariado tenemos casos donde existe outliers, inclusive hay etiquetas en algunas variables donde no existe incumplimiento, para todos estos casos realizamos los siguientes tratamientos:

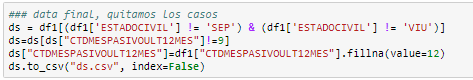
1° Tratamiento de datos nulos e inexistentes dentro de nuestras variables





2° Transformación de variables



3° Limpiamos la data

1. **Elección del Modelo para la trata de la data**

MODELO ÁRBOLES DE DECISIÓN

Trabajaremos con el modelo de árboles porque nos permite predecir y clasificar una variable objetivo independiente a través de la combinación de variables independientes.

Es un algoritmo de aprendizaje automático, que nos ayuda a explorar los datos, el reconocimiento de patrones de importancia y tendencia del riesgo de nuestro objetivo.

PRIMER ENTRENAMIENTO DEL MODELO

1° Visualizamos la importancia de las variables

INGRESO : 39.1262777721

SCORE\_SISTEMA : 37.5961566311

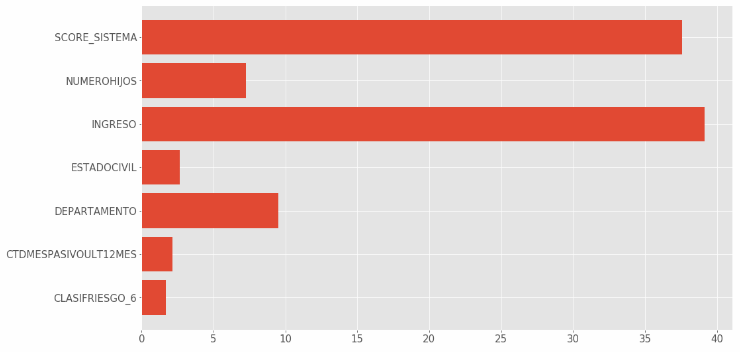
DEPARTAMENTO : 9.49138166587

NUMEROHIJOS : 7.26050887348

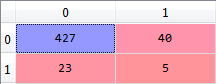
ESTADOCIVIL : 2.6645882913

CTDMESPASIVOULT12MES : 2.15204140102

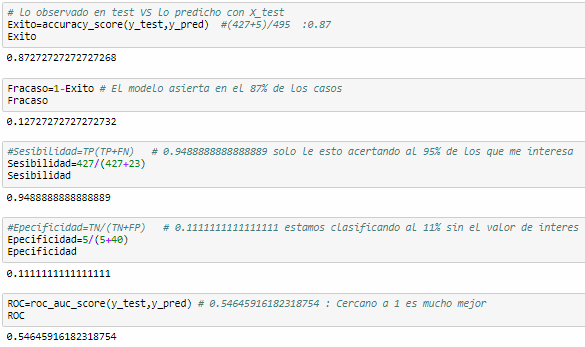
CLASIFRIESGO\_6 : 1.70904536513

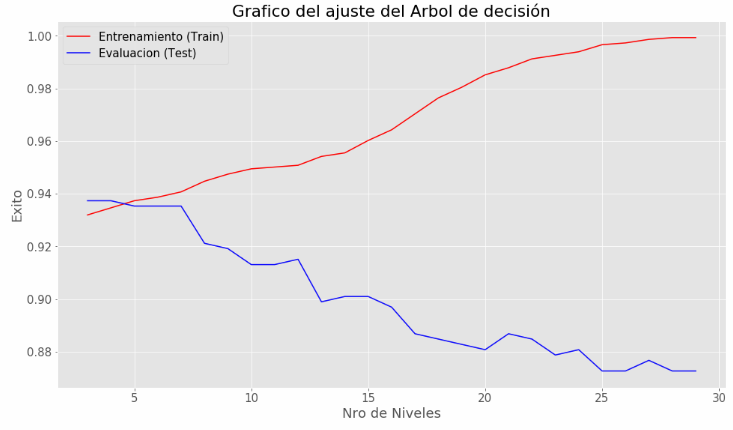


2° Obteniendo en el primer entrenamiento la matriz de confusión



Con un éxito de 87%, acertando en un 94% en las variables que me interesan, clasificando un 11% sin valor de interés y un ROC de 55%



Al realizar los múltiples entrenamientos obtenemos la siguiente gráfica 

Sé observa un sobre entrenamiento, el modelo esta 100% entrenado para el train, mientras más va creciendo la profundidad de mi árbol la predicción de mi test va bajando. Tomando como punto óptimo la intersección (X=5).

MODELO FINAL ENTRENADO

Con la profundidad igual a 5 realizamos nuestro modelo final, agregando la condición **class\_weight = “balanced”** dado que nuestro modelo no esta balanceado, obteniendo un nuevo orden de importancia de las variables:

SCORE\_SISTEMA : 50.1984036002

INGRESO : 27.0328160565

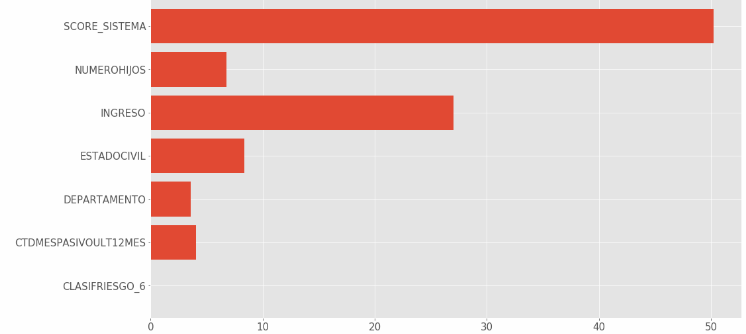
ESTADOCIVIL : 8.35502610419

NUMEROHIJOS : 6.76156777133

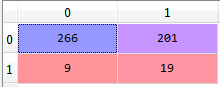
CTDMESPASIVOULT12MES : 4.03842689041

DEPARTAMENTO : 3.61375957741

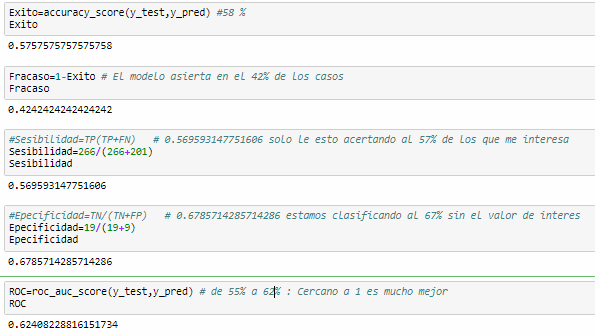
CLASIFRIESGO\_6 : 0.0



Obteniendo una nueva matriz de confusión



Con un éxito de 58%, acertando en un 57% en las variables que me interesan, clasificando un 62% sin valor de interés y un ROC de 62% aumentando de 55% y siendo más cercano a 1.



Creándose el árbol de decisión que considera como variable principal en base a mi Traget SCORE\_SISTEMA de donde parte para la ramificación.

