CREDIT ONE

Clasificacion de clientes

Oscar Quesada Ávalos

UNIVERSIDAD CENFOTEC

Contents

[Abstracto 2](#_Toc37010870)

[1. Limpieza de datos 2](#_Toc37010871)

[2. EDA 2](#_Toc37010872)

[3. RFE 5](#_Toc37010873)

[4. Matriz de Correlación 6](#_Toc37010874)

[5. Modelos de Clasificación 6](#_Toc37010875)

[6. Preguntas 7](#_Toc37010876)

# Abstracto

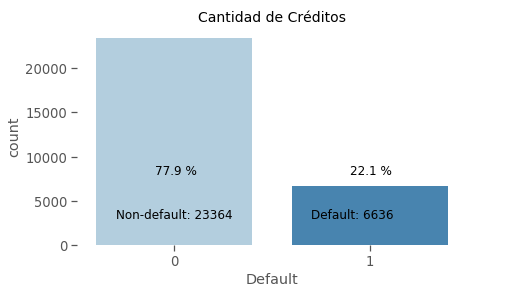
Análisis de clasificación de los clientes que incumplen con los pagos de la tarjeta de crédito.

## Limpieza de datos

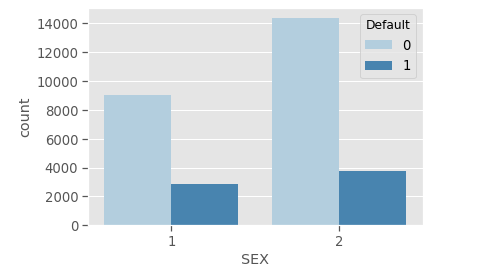
Se ajusta nombre “default payment next month ” a “Default”, atributo “EDUCATION” se remplazan valores 0;5 y 6 con el valor 4 para dejar este valor a la clasificacion “Others”.

## EDA

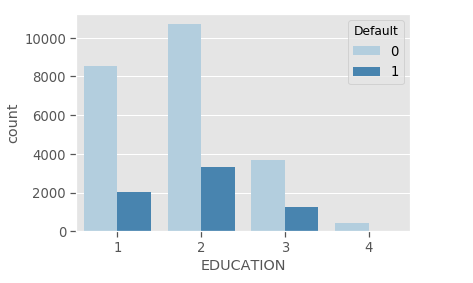
Data set conformado por 30 mil registros con un 22,1% (6636) correspondiendo a clientes con incumplimiento con el pago de la tarjeta de crédito.



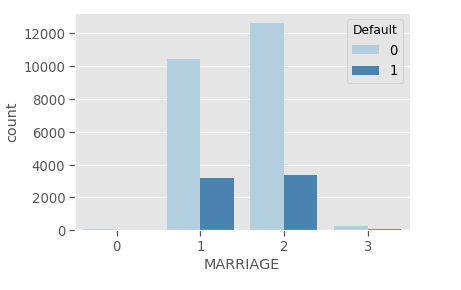
Vamos a verificar si el incumplimiento tiene relación con el género y esto nos brinda una muy poca diferencia.



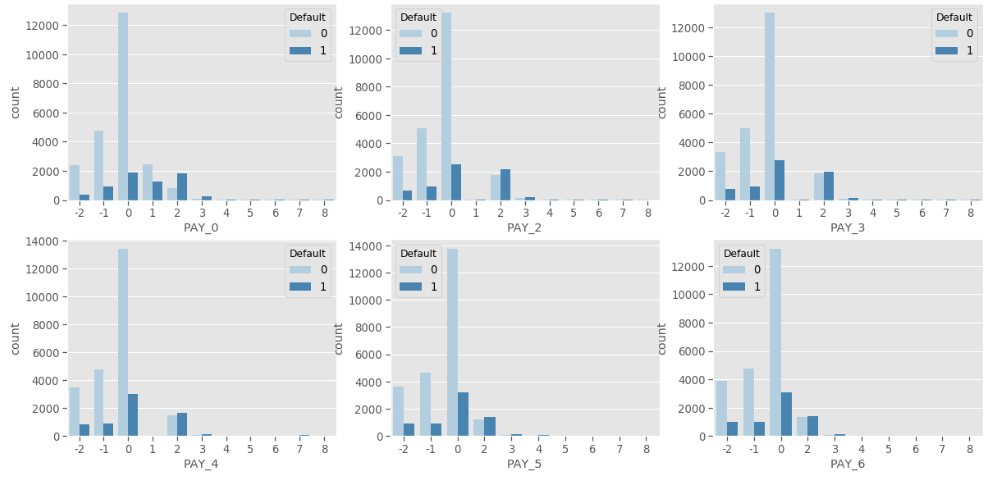
También se realiza análisis del grado de educación. El grafico muestra que los deudores que incumplen son los que cuentan con grado de escolaridad universitario (2) seguidos por los que finalizaron la escuela (1) y de ultimo los que finalizaron el colegio (3).



Realizamos análisis al estado civil, sin tener mayor diferencia casados (1) y solteros (2).

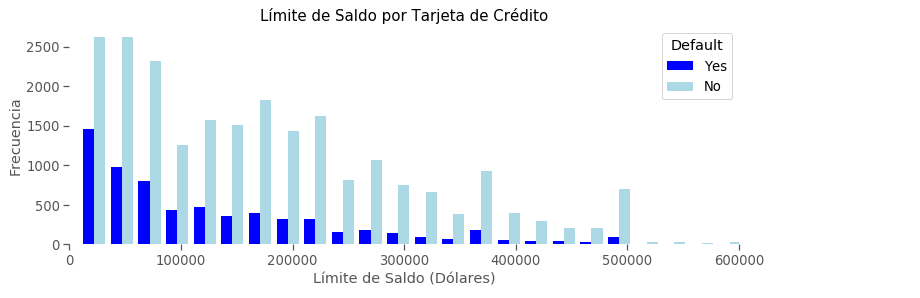


Para finalizar realizamos análisis del comportamiento de pago de deudores durante 6 meses consecutivos. Los atrasos mensuales en los pagos, se puede observar que había datos que se repetían de los atributos PAY\_2, PAY\_3 y PAY\_4 tambien PAY\_5 y PAY\_6.



Como pudimos observar se puede concluir que posiblemente los datos de abril a agosto fueron obtenidos de manera equivocada. Con lo cual nos lleva a consultar con nuestro cliente si la información es correcta.

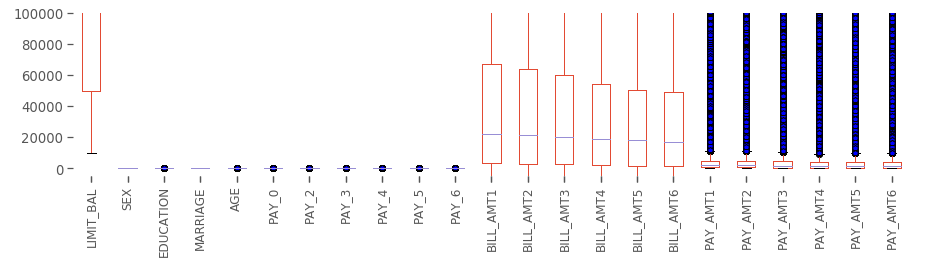
En el siguiente análisis observamos un comportamiento de incumplimiento según su límite de crédito aprobado.



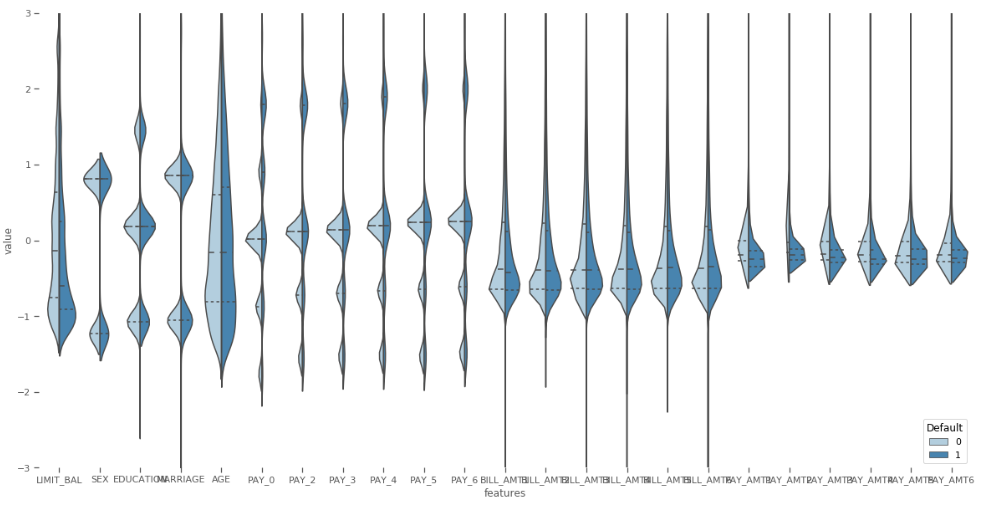
Podemos identificar que las tarjetas de crédito con limite más bajo son las que poseen mayor incidencia con el incumplimiento de los pagos.

## RFE

Podemos observar anomalías en los datos desde 3 hasta 100.000, por lo que debemos normalizar la información.

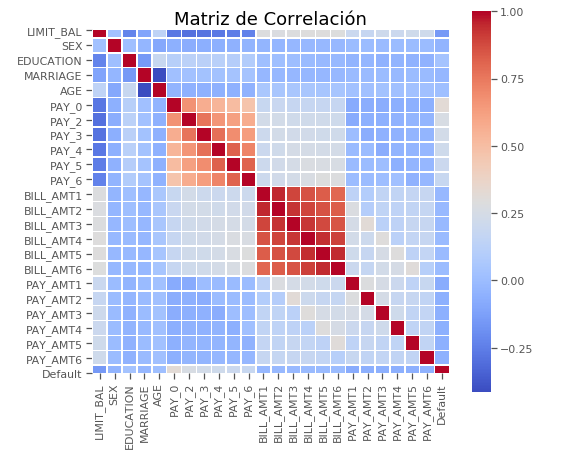


Luego de tratar la información al ver la anomalía podemos visualizar datos entre rango -3 y 3:



## Matriz de Correlación

En dicha matriz podemos visualizar una correlación que se encuentra en los atributos de pagos PAY\_0 PAY\_6 y los BILL\_AMT1 BILL\_AMT6



## Modelos de Clasificación

Modelos Seleccionados:

**Logistic Regression** = 0.8078 (**Accuracy**)

**Decision Tree** = 0.808 (**Accuracy**)

**Random Forest** = 0.8078 (**Accuracy**)

Notas: se realizan modificaciones para los modelos.

* Funcion RandomizedSearchCV para buscar parámetros óptimos.
* Cross Validation, se configura de 5 a 10 sin exito

Modelo Seleccionado 🡪 **Random Forest con mayor precisión.**

## Preguntas

¿Cómo se asegura de que los clientes puedan pagar sus préstamos?

Según los análisis realizados, podemos brindar sugerencias como:

* Clientes con menor límite de crédito son los que más deben
* Clientes con mayor grado universitario son los que incurren más en incumplir

Para brindar una respuesta con mayor precisión y claridad sería bueno estudiar más características de los clientes para poder sacar un común denominador. Asi podríamos sugerir basados en los clientes.

¿Podemos aprobar clientes con alta certeza?

Según el modelo seleccionado tenemos una precisión de 81.7% para sugerir clientes.

¿Qué atributos en los datos podemos considerar estadísticamente significativos para el problema en cuestión?

* Se necesita más información – N/A

¿Qué información concreta podemos derivar de los datos que tenemos?

* Podemos derivar los posibles clientes con alto riesgo de incumplimiento como los que poseen bajo limite de crédito y también con alto grado de escolaridad.

¿Qué métodos probados podemos usar para descubrir más información y por qué?

* N/A