**VISA - Clasificación de clientes (good/bad) (5-8)**

En el mundo financiero, especialmente en el sector de tarjetas de crédito, la gestión del riesgo crediticio es fundamental para mantener la salud económica de las instituciones. Uno de los mayores desafíos que enfrentan las entidades financieras es la identificación y el manejo de clientes que son malos pagadores, es decir, aquellos que no cumplen con sus obligaciones de pago en los plazos establecidos.

VISA una de las marcas más reconocidas y utilizadas en el mundo de los pagos electrónicos le ha encargado este proyecto. Las tarjetas VISA permiten a los consumidores realizar compras, retirar efectivo en cajeros automáticos y efectuar pagos en línea. Su aceptación universal y la seguridad en las transacciones son dos de sus principales fortalezas. El banco emisor de los clientes de este dataset es Banco Santander.

El banco quiere potenciar la emisión de tarjetas VISA para una campaña. Para ello se le pide analizar el comportamiento de pago de un grupo de clientes, e identificar patrones y tendencias que permitan segmentar y gestionar de manera proactiva el riesgo asociado a estos clientes.

Usted deberá cumplir con al menos los siguientes puntos:

Definición de mal pagador:

1. Crear una definición de ‘mal\_pagador’. Esta definición deberá contar con un componente temporal y un componente en dias, por ejemplo:
   1. Durante 12 meses el cliente superó mas de 30 días al menos una vez.
   2. Durante 24 meses el cliente superó mas de 30 días al menos 2 veces.
2. Cree un dataset a nivel de cliente, en donde cada fila sea un cliente único. Este dataset contará con los clientes que entrarán al análisis, es probable que usted excluya ciertos clientes porque no cumplen con criterios que usted considere relevantes, por ejemplo: excluir del análisis a clientes que nunca recibieron un crédito o a clientes que recién llevan x cantidad de meses con el crédito.
3. El dataset deberá constar de dos columnas CLIENTE\_ID y mal\_pagador.

Join con información adicional del cliente:

1. Con el dataset que creó anteriormente, unifíquelo con el archivo de información adicional del cliente. Asegurese de que solo se encuentren los clientes que quiere analizar

Data quality:

1. Analice si existen valores duplicados, y si es así aborde el problema.
2. Analice si existen valores nulos, y si es así aborde el problema. Puede usar cualquiera de las alternativas que vimos en clase

Feature transformation:

1. En este apartado usted deberá crear nuevas columnas apartir de las que tiene. Pueden ser:
   1. Variables indicadores (1 si cumple x condición, 0 caso contrario)
   2. Variables categóricas apartir de una variable nominal

Analisis correlacional:

1. Identifique si tiene variables correlacionadas
2. En caso de ser así, encuentre una explicación lógica para esa correlación. Si no encuentra un argumento, puede mencionarlo.

Análisis descriptivo:

1. Usando estatísticos descriptivos univariados, analice las variables que considere relevantes y aportan información.
2. Usando estatísticos descriptivos bivariados, analice las variables que considere relevantes y aportan información.
3. Con variables categóricas, realice análisis con tablas cruzadas.

**Es importante que cree preguntas de negocio y las responda cualquiera de los puntos 1,2,3. Por ejemplo:**

1. Los clientes más jóvenes son los más riesgosos?
2. Los clientes más jóvenes que no tienen hijos, son más riesgosos?

**Es importante que siempre use su variable de interés(mal\_pagador).**

* 1. Por ejemplo: Los clientes que tienen mas de x años de vida son mas/menos propensos a ser malos pagadores …

**Es importante que use las variables que creó en Feature transformation**

**Es importante que use la mayor cantidad de visualizaciones posibles para los puntos anteriores. Por ejemplo:**

* Diagrama de cajas y bigotes
* Histogramas
* Gráficos de dispersión
* Gráficos pairplots

Conclusiones y Recomendaciones

1. En este apartado usted deberá sugerirle a la empresa a qué clientes se debería enfocar para la campaña, argumentando que esos clientes son los menos riesgosos.

**Entregables**

1. Archivo notebook en donde se evidencie el desarrollo de cada uno de los puntos que deberá subirse hasta el sábado 29 de junio en aula virtual.
2. Ppt o pdf que deberá subirse hasta el sábado 29 de junio en aula virtual.
3. BONUS: Los animamos a que creen un repositorio en github para su proyecto. No suban datos, solo el notebook y la presentación (ppt). Esto les permitirá ir construyendo su portafolio de proyectos.

**Evaluación y Calificación:**

Un jurado de 3 personas evaluará al grupo considerando el uso efectivo del tiempo; aplicación de lo aprendido en el módulo y presentación efectiva de la idea.

**Consideraciones finales:**

1. Cada equipo decidirá el número de personas que quieren presentar el caso, desde una sola hasta todos. Esto no aumentará ni disminuirá el tiempo disponible.
2. Si alguna de las personas del equipo no está presente en la sesión de presentación del proyecto, no podrá tener la nota correspondiente.
3. Los miembros del equipo pueden reportar hasta el Sábado 29 de Junio si otro miembro del equipo no ha trabajado. Si no se notifica al facilitador hasta ese día, se entenderá que todas las personas del grupo han participado en igualdad de condiciones y aportaciones.

DATASET

<https://drive.google.com/drive/folders/147uPfLaRzFssOSyNrt0-_4TwP_XPi8RX?usp=drive_link>