



El DataSet elegido es un base de datos de aproximadamente diez mil usuarios de una institución bancaria la cual he reunido una cantidad considerable de datos acerca de la situación de cada uno de ellos, tanto personal como en relación los paquetes que mantienen con el banco y su situación frente a ellos.

Se puede observar que cada cliente cuenta con un número único de identificación, además tenemos información personal como ser la edad, el sexo, nivel de educación, estado marital, nivel de ingresos. Con respecto al paquete que tiene cada cliente contractualmente con el banco conocemos la cantidad de tarjetas adeherentes y las categorías de sus tarjetas.

En lo que concierne directamente con la utilización de sus tarjetas, el banco otorga información de cada cliente que detalla:

* La antigüedad (en meses) que posee el cliente en la entidad
* El total de productos que posee con la entidad
* Los meses que estuve inactivo en el último año
* El límite de crédito otorgado por la entidad al cliente
* El monto total adeudado a la fecha
* El límite disponible para utilizar
* El total de operaciones realizadas
* El total en dinero de las operación realizadas
* El “Average credit utilization ratio” que es un índice basados es scores que indica la situación del cliente.

Además debemos destacar como principal información la categorización con la contamos en “Existing Customer” y “Attrited Customer”. Esta atribución nos dice en qué situación se encuentra el cliente actualmente con respecto al banco al día de la fecha si sigue siendo un cliente activo o es un cliente en baja.

A partir de esta categorización y con la información pertinente en conocimiento, es que se planteará una análisis del DataSet para evaluar si es posible establecer algún modelo que nos indique en base a alguna de las otras variables si es posible predecir la futura baja de un cliente y en base a eso donde se debe apuntar para mejor las condiciones de ese cliente para poder evitarla y mantenerlo como cliente de la entidad.

**Preguntas e hipótesis**

* ¿Es posible predecir la futura baja de un cliente si posee una cantidad de transacciones en su tarjeta por debajo de cierto umbral?
* ¿Tiene inferencia el nivel de ingresos del cliente en el mantenimiento de la cuenta más allá de su nivel de utilización?
* ¿Es posible que el sexo del titular y/o la cantidad de tarjetas adherentes influyan a la hora de decidir darse de baja como cliente de la entidad?
* ¿Es posible predecir una futura baja mediante el “Average credit utilization ratio”?
* El límite de crédito otorgado al cliente y su porcentaje de utilización de ese límite es un indicador del posible estado de este?