

Proyecto 2 Análisis Exploratorio

Incumplimiento de Pago de Tarjeta de Crédito

Problema Científico

En los años de la pandemia, en los Estados Unidos de América, el comportamiento del uso de tarjetas de crédito ha variado bastante. El promedio de la deuda de tarjeta de crédito creció en 52% entre el 2018 y el 2019, sin embargo, este porcentaje ha caído significativamente en el 2020 (ValuePenguin)(Household Debt and Credit Report, NewYorkFed).

Estados Unidos tiene una deuda de \$807 mil millones distribuidas en 506 millones de tarjetas de crédito mientras que la deuda promedio de una familia Estadounidense es de \$6,270. Estos números incrementan cada año hasta el 2020 donde se vio un decremento debido al COVID-19. En el 2020 la deuda promedio por familia disminuyó a \$5,315, casi \$1,000. La pandemia tuvo un gran impacto en la vida de todos lo cual causó cambios que atribuyen estos cambios de deudas. Entre estos cambios está el hecho que la gente utilizaba menos su dinero en compras y gastos al igual que muchos recibieron dinero suplementario por desempleo o por otras razones que pueden utilizar para estas deudas (Resendiz, 2021).

Además, para poder conocer más sobre las posibles causas del decremento de deudas se investigó sobre diferentes métodos para resolver dichas situaciones económicas. La primera es a través de las compañías de liquidación de deudas. Estas compañías ofrecen negociar con los emisores de tarjetas de crédito para que el endeudado tenga la opción de disminuir la cantidad que debe pero debe tomar en cuenta que esa negociación puede tomar tiempo y debe pagarle a dicha compañía de liquidación. Además, debe tener cuidado con compañías fraudulentas que simplemente lo pondrán en una peor situación (Comision Federal de Comercio).

Otra opción sería que el individuo con impago negocie directamente con los acreedores. Es posible que le den una tasa de interés más baja la cual facilita el pago de deuda. Una tasa más baja indica que una menor cantidad de cada pago mensual que realice se consuma a causa de los cargos por intereses acumulados y por ende se podrá pagar la deuda de manera más rápida.

Cabe mencionar que maximizar su flujo de efectivo es crucial. Ya sea al conseguir nuevos/mejores empleos o minimizar sus gastos cada cantidad ayuda. Además, debe organizar y priorizar sus deudas. Al analizar y ordenar todas las deudas que tiene lo ayudará desarrollar un plan que indique cómo y en qué orden irá pagando (Debt).

Universidad del Valle de Guatemala
Depto. Ingeniería
Data Science
Sección 20
Catedrático: Luis Furlan

Nombre	Apellido	Carné
Juan Diego	Solorzano	18151
Ricardo	Valenzuela	18762
Sara	Zavala	18893
Jose Amado	Garcia	181469

Se puede observar que estos datos y soluciones dependen mucho de los individuos que tienen la deuda. Por lo mismo es importante saber que características tiene un cliente que cae en incumplimiento de pago de tarjeta de crédito.

El objetivo de este proyecto es crear modelos de aprendizaje automático que permita predecir qué tipo de clientes podrán caer en impago, o no. Para esto se cuenta con un conjunto de datos que contiene información sobre pagos predeterminados, factores demográficos, datos crediticios, historial de pagos y extractos de cuentas de clientes de tarjetas de crédito en Taiwán desde abril de 2005 hasta septiembre de 2005.

Objetivos

Generales:

- Identificar a los tipos de clientes que pueden caer en impago

Específicos:

- Identificar qué variable(s) tienen el mayor impacto en las deudas de tarjeta de crédito
- Desarrollar modelos vistos en clase y predicciones sobre los clientes con deuda de tarjeta de crédito
- Identificar a qué se debe el comportamiento que demuestran los porcentajes de deuda en los últimos años

Análisis Exploratorio

Descripción del dataset previo a la limpieza

Hay 25 variables (22 cuantitativas, 3 categóricas):

- ID: ID de cada cliente
- LIMIT_BAL: cantidad del crédito otorgado en dólares NT (incluye crédito individual y familiar / suplementario)
- SEX: Género (1 = masculino, 2 = femenino)
- EDUCATION: (1 = estudios de posgrado, 2 = universidad, 3 = bachillerato, 4 = otros, 5 = desconocidos, 6 = desconocidos)



Universidad del Valle de Guatemala
Depto. Ingeniería
Data Science
Sección 20
Catedrático: Luis Furlan

Nombre	Apellido	Carné
Juan Diego	Solorzano	18151
Ricardo	Valenzuela	18762
Sara	Zavala	18893
Jose Amado	Garcia	181469

- MARRIAGE: Estado civil (1 = casado, 2 = soltero, 3 = otros)
- AGE: Edad en años
- PAY_0: Estado de reembolso en septiembre de 2005 (-1 = pago debidamente, 1 = retraso en el pago durante un mes, 2 = retraso en el pago durante dos meses,... 8 = retraso en el pago durante ocho meses, 9 = retraso en el pago durante nueve meses o más)
- PAY_2: Estado de reembolso en agosto de 2005 (escala igual a la anterior)
- PAY_3: Estado de reembolso en julio de 2005 (escala igual a la anterior)
- PAY_4: Estado de reembolso en junio de 2005 (escala igual a la anterior)
- PAY_5: Estado de reembolso en mayo de 2005 (escala igual a la anterior)
- PAY_6: Estado de reembolso en abril de 2005 (escala igual a la anterior)
- BILL_AMT1: Monto de la factura en septiembre de 2005 (dólar NT)
- BILL_AMT2: Monto de la factura en agosto de 2005 (dólar NT)
- BILL_AMT3: Monto de la factura en julio de 2005 (dólar NT)
- BILL_AMT4: Monto de la factura en junio de 2005 (dólar NT)
- BILL_AMT5: Monto de la factura en mayo de 2005 (dólar NT)
- BILL_AMT6: Monto de la factura en abril de 2005 (dólar NT)
- PAY_AMT1: Monto del pago anterior en septiembre de 2005 (dólar NT)
- PAY_AMT2: Monto del pago anterior en agosto de 2005 (dólar NT)
- PAY_AMT3: Monto del pago anterior en julio de 2005 (dólar NT)
- PAY_AMT4: Monto del pago anterior en junio de 2005 (dólar NT)
- PAY_AMT5: Monto del pago anterior en mayo de 2005 (dólar NT)
- PAY_AMT6: Monto del pago anterior en abril de 2005 (dólar NT)
- default.payment.next.month: incumplimiento de pago al mes siguiente (1 = sí, 0 = no)

Utilizando diferentes librerías de Python, especialmente `pandas_profiling`, generamos varias gráficas y tablas que nos permiten ver el estado del dataset y sus variables. Inmediatamente observamos que algunas variables tienen alta correlación y otras características que nos mostraban la necesidad de limpieza. Los procesos que realizamos fueron:

- Estandarizar nombres de las columnas
- Cambiar el tipo a categórica de las variables `sex`, `education`, `marriage` y todas las `pay`
- Se eliminó la columna de `Id` ya que no la necesitábamos
- En la variable “`education`”, los valores 0, 5 y 6 representan el valor desconocido por lo que se juntaron para que tomen solamente el valor 5
- En la variable “`marriage`” se cambio el valor 0 a 3 ya que ambos representan el valor otros
- Se creó una columna `prom_bill_amt` que toma los valores promedios de todas las columnas `bill_amt` ya que tenían alta correlación
- Se eliminaron algunos datos atípicos y nos quedamos con 77% de los datos originales lo cual aún es una cantidad adecuada para nuestro análisis

Durante la limpieza se investigó sobre los valores -2 y 0 de las columnas `pay` ya que a primera vista se veían como datos nulos que debíamos modificar. Sin embargo, observamos que estos valores si tienen un significado en el dataset con respecto a la realización de los pagos por lo que no se pueden modificar a nulo.

Descripción y Conclusión Final

El conjunto de datos tras las operaciones de limpieza cuenta con 23241 observaciones y 19 variables. De estas variables:

- 10 son categóricas: `sex`, `education`, `marriage`, `pay_1` ... `pay_6`, `default`, `payment`, `next`, `month`
- 9 son cuantitativas: `limit_bal`, `age`, `prom_bill_amt`, `pay_amt1` ... `pay_amt6`

Al analizar las variables cualitativas podemos notar ciertas características importantes como:

- El nivel de educación que predomina en los datos es universidad seguido de postgrado.
- El estado civil predominante es soltero



Universidad del Valle de Guatemala
Depto. Ingeniería
Data Science
Sección 20
Catedrático: Luis Furlan

Nombre	Apellido	Carné
Juan Diego	Solorzano	18151
Ricardo	Valenzuela	18762
Sara	Zavala	18893
Jose Amado	Garcia	181469

- El sexo mas presente es el femenino

De las variables cuantitativos podemos observar lo siguiente:

- La media de edad de las personas es de 35 años.
- La media de cantidad de crédito es alrededor de 143000.
- Los datos de pay_amt están muy sesgados, cuentan con altos valores atípicos

Al cruzar variables podemos observar lo siguiente:

- Los que más probablemente paguen su deuda son personas alrededor de 23 a 36 años.
- Hay un buen porcentaje de impago en todas las edades.
- El estado civil casado presenta una cantidad de impagos muy similar a la de soltero a pesar de tener menos observaciones.
- El sexo femenino presenta tener una cantidad de impagos muy similar al masculino a pesar de tener mas observaciones.
- Gran cantidad del crédito está en un intervalo de 0 a 500000 valores arriba de esto pueden considerarse atípicos
- Los niveles de educación estudios de posgrado, universidad, bachillerato suelen pedir un valor más alto de crédito que otros y desconocidos.
- Gran cantidad del monto de la factura general de los meses está entre 0 a 300000, sin embargo, para el caso de impago hay valores atípicos muy arriba de esta cifra y hay datos negativos

Referencias

Debt.com (2018). Como reducir la deuda de tarjeta de crédito en 5 simples pasos.
<https://www.debt.com/es/como-hacer-para-reducir-tu-deuda-de-tarjetas-en-5-simples-pasos/>



Universidad del Valle de Guatemala
Depto. Ingeniería
Data Science
Sección 20
Catedrático: Luis Furlan

Nombre	Apellido	Carné
Juan Diego	Solorzano	18151
Ricardo	Valenzuela	18762
Sara	Zavala	18893
Jose Amado	Garcia	181469

La Comisión Federal de Comercio (2010). Liquidando su deuda de tarjeta de crédito.
<https://www.consumidor.ftc.gov/articulos/s0145-liquidando-su-deuda-de-tarjeta-de-credito>

Resendiz, J. (2021). Average Credit Card Debt in America: 2021.
<https://www.valuepenguin.com/average-credit-card-debt>