



Resumen de Hallazgos

Universidad Cenfotec
Certificación XTOL Data Analytics & Big
Data

Grettel Rojas Conejo

Contenido

Introducción	2
Análisis de Valor Comercial	2
Lecciones Aprendidas	3
Recomendaciones.....	3

Introducción

El requerimiento de este análisis se basa en que se ha identificado que han aumentado las tasas de incumplimiento de los clientes, lo cual es algo negativo para Credit One, ya que su labor es aprobar los préstamos de acuerdo con información crediticia y sus respectivos análisis.

En el presente informe se muestran los principales hallazgos encontrados en el análisis previo de los datos, realizado mediante programación en Python. Estos hallazgos describen el comportamiento de los datos, en cuanto a los atributos que aportan valor, y generan una idea premiliar de hacia dónde generar esfuerzos para solventar la problemática.

Por otra parte, se establecen los hallazgos sobre lecciones aprendidas y recomendaciones que se han identificado en el desarrollo de este análisis, para efectos de que sean puntos de valor en futuros análisis.

Análisis de Valor Comercial

Se analizaron los datos en Python mediante gráficos, tablas de distribución de frecuencias y matriz de correlación de variables, como puntos de relevantes para efectos de este proyecto se identificó lo siguiente:

1. El 21% de los clientes presentan incumplimiento en el próximo pago.
2. La cantidad de personas divorciadas es despreciable, con respecto a la población de casados y solteros.
3. Para efectos del cumplimiento de pago según el estado civil, los clientes casados, en su mayoría los hombres, presentan un mayor incumplimiento, lo que podría tener sentido, por los gastos en los que puede incurrir una persona al estar establecido en un matrimonio.
4. Para efectos del cumplimiento de pago según el género, las mujeres presentan una mayor tendencia al incumplimiento en comparación con los hombres.
5. Entre más alto el nivel académico, mayor cantidad de personas presentan incumplimiento de pago.
6. El rango de edad con más deudores es de los 20 a los 26 años, seguido de los 26 a los 30 años.
7. Mediante la matriz de correlación se puede observar que existe relación significativa entre los saldos de los meses consecutivos. De esto se puede entender que, si una persona tiene una deuda en un mes, es probable que el siguiente mes presente una deuda que tenga relación, lo que indica que no se puede recuperar de un mes a otro, sino que las deudas son proporcionalmente crecientes.

Lecciones Aprendidas

1. Como en todo análisis de datos, lo primero es iniciar realizando una inspección general del data set, esto es, verificar qué datos se están dando, qué significa cada una de las variables, ya que si esto no está claro desde un inicio es probable que el análisis no agregue valor.
2. La estadística descriptiva nos ayuda a entender el comportamiento de los datos, y nos brinda una idea inicial sobre los pasos siguientes a realizar en el análisis.
3. Es importante entender el tipo de variables que se tienen, y si es requerido convertir éstas para poder ajustar los datos en los gráficos y análisis que se requieran realizar. Por ejemplo, que existan variables numéricas cuando realmente deberían ser categóricas, esto podría inducir un error en los resultados.
4. Se debe analizar si existen variables que no agregan valor y que más bien generan ruido, en caso de realizar por ejemplo un análisis de correlación. Los ID por lo general son variables que no se utilizan, ya que son registros por lo general numéricos que no indican nada respecto al data set. En caso de encontrar variables que no agregan valor, estas deben ser eliminadas.

Recomendaciones

1. Antes de realizar cualquier análisis del que se nos pide realizar inferencias, es importante conocer el negocio, el contexto y la problemática, esto para poder ser empáticos con los datos. Así mismo, se debe tener claro cuál es el objetivo del análisis, ya que si no sabemos con exactitud qué es lo que buscamos, es probable que no encontremos nada relevante. De igual forma, el conocer el cliente nos permite setear la expectativa, comprender cuál es el grado de detalle y de tecnicismos que podemos utilizar para efectos de un entregable o una presentación.
2. En los proyectos de análisis de datos, se debe considerar tiempo y esfuerzo en actividades de preprocesamiento, ya que de esto depende en gran parte el éxito del análisis. Si bien es cierto, estas actividades no nos van a dar un resultado final que entregar al cliente, pero son necesarias para tener en el futuro un entregable de calidad.
3. Es de vital importancia, conocer que significa cada una de las variables con las que contamos en un set de datos, por lo que de no estar seguros que significa un dato, lo recomendable es investigar qué significan los valores, de donde proviene el dato, por qué se llevan registros de estos datos, su contexto y si el cliente lo considera de valor.