# Reporte de lecciones aprendidas – Caso Credit One.

* La experiencia inicial en el uso de Python ha sido muy agradable y provechosa. Es una herramienta muy fácil de usar y cuenta con mucha ayuda disponible para cualquier problema presentado.
* Al igual que en cualquiera de las otras herramientas utilizadas anteriormente como R y rapid miner, en python también es imprescindible respetar los tipos de datos de cada variable a utilizar en el set de datos.
* El set de datos con el que se dispone para el desarrollo de este proyecto contiene información muy valiosa y que puede ser de gran utilidad para concretar el objetivo principal de dicho proyecto. Por ejemplo, contamos con información referente al comportamiento que han tenido los clientes en sus pagos en los últimos seis meses. Lo que nos permite crear diferentes visualizaciones utilizando parámetros como el género, edad, estado civil, nivel de educación para conocer cómo se comportan los clientes en sus pagos dependiendo de una categoría u otra.
* Esta claro que no podemos controlar los hábitos de los clientes al momento de gastar, pero si podemos evaluar el nivel de riesgo que representa cada uno de ellos al momento de otorgar un crédito, esto mediante el estudio y procesamiento de información crediticia histórica.

**Did you learn anything of potential business value from this analysis?**

Analizar un caso como este nos permite darnos cuenta de que no es lo mas recomendable basarnos en una única variable para concluir si un cliente va a pagar o no. La unión y estudio de las diferentes variables nos van a portar un dato más exacto y real en dicha predicción.

Por ejemplo, basarnos únicamente en la información referente al estado del último pago del cliente no sería muy recomendable porque pudo suceder que el cliente ese mes en particular tuvo una situación esporádica que le impidió realizar el pago a tiempo. Pero no significa que sea un cliente que debemos clasificar como de alto riesgo.

**What are the main lessons you've learned from this experience?**

Cada caso de estudio debe ser abarcado de forma independiente, tratando siempre de sacar el máximo potencial de la información con la que contamos, entendiendo el problema y creando visualizaciones que nos permitan evidenciar los resultados de nuestra investigación para entendimiento y tranquilizad del interesado final.

**What recommendations would you give to the Data Science team regarding your findings?**

Inicialmente estudiar y entender el set de datos, el cual contiene información potencial, pero a primera vista puede resultar un poco difícil de entender y podríamos cometer el error de excluir variables que nos pueden resultar muy útiles simplemente porque no entendemos el contenido de dicha variable.

Sacar provecho de las diferentes bibliotecas que nos ofrece Python para realizar visualizaciones y comparar diferentes variables con el objetivo de comprender la relación entre las mismas.

La limpieza y categorización de los datos es imprescindible y facilita el entendimiento durante el desarrollo de las diferentes etapas del proyecto.

Las columnas **BILL\_AMT4, BILL\_AMT5 y BILL\_AMT6** fueron eliminadas del set de datos ya que presentaban una correlación superior a 90% con otras variables.

En el caso de la columna **AGE**, definí cuatro categorías para esta: ['20-35', '35-50', '50-65', '65+']. Con ello, se reduce el numero de posibles combinaciones al momento de aplicar los modelos.

Para las columnas **EDUCATION, SEX, MARRIAGE**, **PAY\_0 - PAY\_6** con el fin de evitar usar valores numéricos que resultan poco legibles, los convertí en columnas categóricas que permiten un mejor entendimiento al momento de graficar los datos. Ejemplo: Para la columna **SEX**, en lugar de mostrar 1,2 los valores a utilizar serán Male y Female respectivamente.

Experimentar con diferentes combinaciones de variables para poder identificar con qué grupo de estas podremos lograr una mayor exactitud en el resultado.