

## One Credit

**Problema:** Un aumento en las tasas de incumplimiento de los clientes es malo para Credit One ya que su negocio está aprobando préstamos para clientes en primer lugar. Es probable que este resulte en la pérdida de los clientes comerciales de Credit One.

### Preguntas:

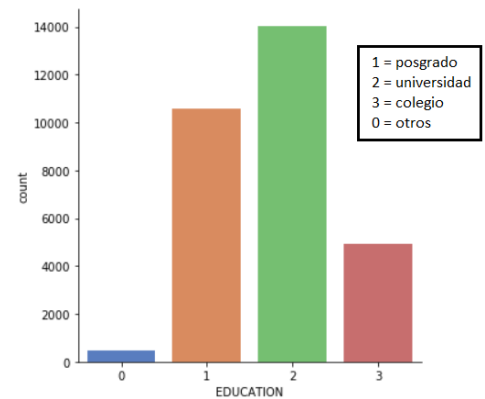
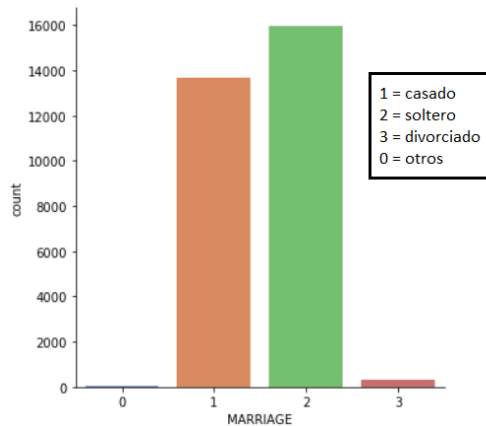
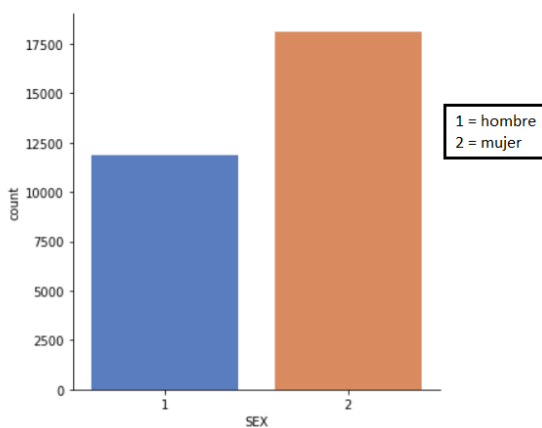
- Cómo asegurarse que el cliente pueda/vaya a pagar el préstamo?
- Se puede aprobar clientes con alta certeza?

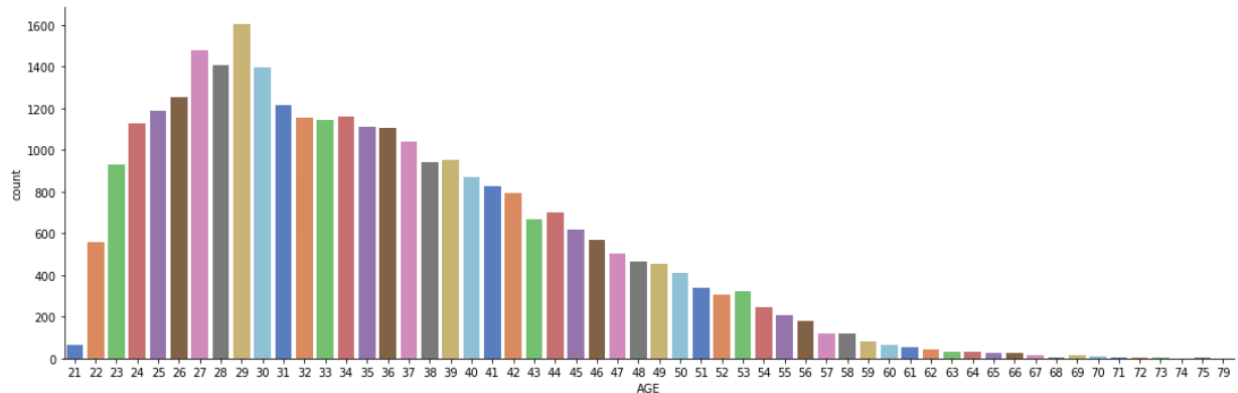
El cliente provee un set de datos de clientes actuales que serán utilizados durante todo el proceso de Data Science. A partir de visualización y modelado se intentarán responder las preguntas en cuestión y dar una respuesta a Credit One.

En un reporte anterior se presentó a Credit One el proceso de primer análisis de los datos y limpieza de los mismos, este reporte no presentará los mismos resultados pero se mostrarán las conclusiones principales. (Estos archivos pueden ser consultados dentro de este proyecto y sus nombres son: CreditOne.ipynb y CreditOne\_EDA\_Report.pdf).

Al hacer un estudio superficial sobre los clientes de Credit One destaca la siguiente información concreta:

- Hay más mujeres en la muestra que hombres.
- La mayoría de los clientes de Credit One están solteros, seguido por los casados.
- La mayoría de los clientes se han graduado de un nivel académico. Los más comunes son aquellos con un bachillerato seguidos por aquellos que alcanzaron estudios avanzados.
- Al analizar los pagos de los últimos meses lo más común es que los clientes paguen el crédito revolving o que hayan pagado el costo total.
- El rango de edades de los clientes va de 21 a 79 con una media de ~35. Es posible observar que la mayoría de los clientes tiene una edad entre los 25 y 40 años. Esto se puede tomar como un mercado meta y en proyectos futuros considerar otros elementos específicos para este rango de edades.





Al hacer un análisis de correlación entre las variables independientes y dependientes no es posible identificar atributos que tengan alta correlación con la variable dependiente, por lo tanto no es posible identificar cuales atributos son estadísticamente relevantes con respecto al problema. Esto quiere decir que no hay una sola variable que nos permita identificar si un cliente es susceptible a incurrir en una falta de pago o no. A partir de este momento es evidente que para poder identificar si una persona incurrirá en una falta, o no, es indispensable crear un modelo de clasificación que permita ver el set de datos como un conjunto.

Antes de construir los modelos se aplica un modelo RFE para valorar la eliminación recursiva de datos de manera que se pueda disminuir la cantidad de atributos del data set. Al aplicar el mismo se muestra que la cantidad de atributos con respecto al data set en análisis es el indicado y se continúa el estudio con la totalidad de los datos.

Se decide entrenar 3 modelos distintos entre los cuales se tomará el de mayor rendimiento y se intentará ajustar sus parámetros para obtener una mejor clasificación. Lo que queremos es saber si se puede construir un modelo que permita, a partir de los datos proporcionados por Credit One, clasificar si un cliente incurrirá en una falta de pago. Por lo tanto, se podrá determinar si este es un buen candidato para un préstamo o no.

Los tres modelos construidos fueron Random Forest, Support Vector Machine y Stochastic Gradient Descent. El modelo de Random Forest presentó un accuracy de 0.999, como sabemos un modelo que es prácticamente perfecto es una señal de que el modelo puede estar aprendiendo los datos de memoria y que hay overfitting. El segundo modelo en tener el mayor nivel de accuracy es el SVM y es elegido para continuar.

Se toma el modelo de clasificación de Support Vector Machine y se ajustan sus parámetros para alcanzar el modelo de mayor rendimiento. En este caso se ajusta el tipo de Kernel y el C a varios valores hasta encontrar el modelo con mayor accuracy que es de 0.88. Este modelo permitirá clasificar con bastante certeza si un cliente incurrirá en una falta de pago o no.

A partir del proceso completo de Data Science se puede concluir lo siguiente:

- Para la respuesta de las dos preguntas: Cómo asegurarse que el cliente pueda/vaya a pagar el préstamo? Se puede aprobar clientes con alta certeza?
  - Los datos nunca nos permitirán acertar al 100% un comportamiento que es representado por muchos atributos o cualidades del cliente. En este caso es evidente dado que ningún atributo muestra una alta correlación con respecto a la variable dependiente y es la combinación de todos estos factores los cuales nos puede ayudar a predecir si un cliente pagará o no un préstamo.
  - Es importante destacar que los datos que tiene Credit One solo contemplan una serie de datos que pueden ayudar a describir el problema pero no considera otros temas que pueden ser de importancia. Por ejemplo, si las parejas tienen hijos u otras deudas que deben de pagar.
  - Aunque podamos predecir un comportamiento a partir de ciertos atributos no sucederá siempre de la misma manera. El modelo creado provee un alto rendimiento y certeza, sin embargo, esto es a partir de los datos que provee Credit One pero no se sabe si habrá algún otro atributo que podría cambiar la dinámica del problema. No se toma en cuenta por ejemplo la posibilidad de que el deudor fallezca o de que el deudor pierda su trabajo. Tampoco contempla si el cliente es asalariado o no, solamente por mencionar algunos ejemplos.

## **Recomendaciones**

- Se recomienda a Credit One tomar el modelo creado para predecir si un cliente incurrirá en una falta o no. Sin embargo, debe tener en cuenta que siempre cabe una posibilidad de que la clasificación no sea 100% certera o que por un factor externo (que no se prevé en la data contemplada para generar el modelo) haya un cambio.
- Dado que el comportamiento humano es sumamente incierto y cambiante se recomienda que al hacer préstamos de muy alto riesgo se haga un mejor análisis del cliente antes de concederle el dinero.
- Es importante que Credit One tome en consideración que hay tendencias en el mercado durante los años y que es imprescindible que después de unos años se haga un ajuste de los modelos con respecto a datos más actualizados, esto con el fin de que los modelos hagan una clasificación basada en las tendencias del mercado más recientes.