
Análisis de Perfil de Nivel Crediticio

Caso de Estudio:
Deudores Banco "Buenos Amigos"

Contenido

01

Conformación
de Equipo de
Trabajo

02

Presentación del
Caso

03

Preguntas y
Objetivos de la
Investigación

04

Exploratory
Data Analysis

05

Aplicación de
algoritmos de
ML

06

Conclusiones

Conformación de Equipo de Trabajo



Banco
Buenos Amigos

Somos los estudiantes **Renzo Tenaglia** y **Ernesto Castro**. Conformamos el **equipo de Data Science del banco "Buenos Amigos"**, cuya misión es guardar el dinero de los depositantes pagando una tasa de interés.

Además de prestar a quienes necesiten financiamiento cobrando una tasa.

Presentación del Caso



Nos han proporcionado **una base con 8838 deudores**, suponiendo que en promedio estos pidieron un préstamo de 100 U\$D, el **banco incurre en una pérdida de 883.800 mil dolares.**

Nuestro objetivo es poder **reconocer a los posibles deudores antes de desembolsar el préstamo.**



¿Cómo saber la calidad crediticia de un cliente en base a sus características?

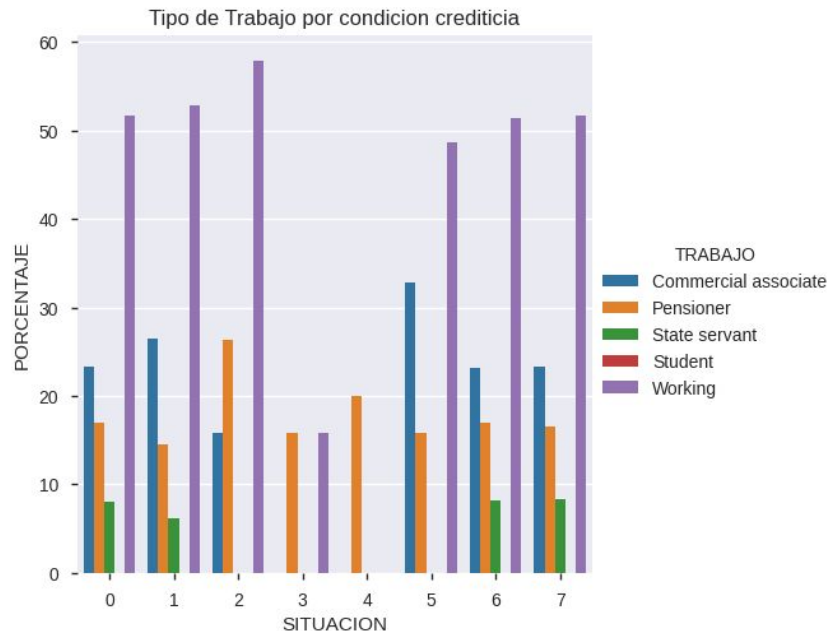
Pregunta de Investigación

¿Cual es el estado
de la base de datos?

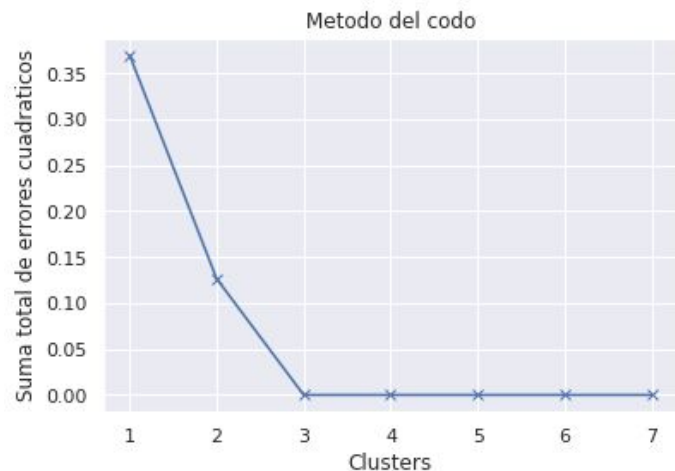
Base de Datos Desbalanceada

La base de datos está separada en tipo según condición crediticia (del 1 al 7).

Se observa una **gran escasez de casos para los grupos 3 y 4** por lo que la información de los clientes es dispersa puede **arrojar interpretaciones erróneas**

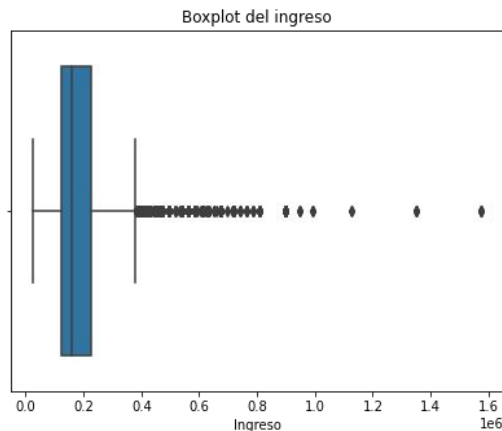
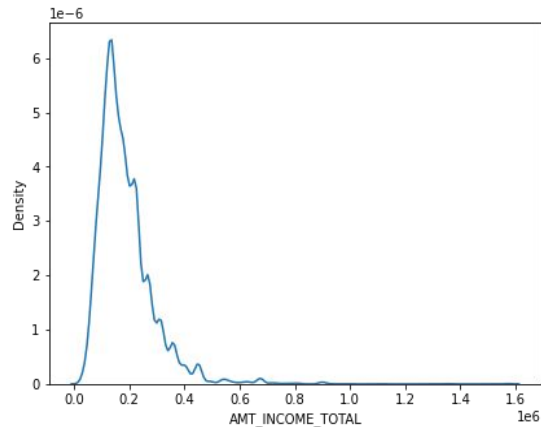


Agrupamiento de mediante Método del Codo



Mediante este método se plantea una clusterización eficiente para poder agrupar los tipos de cliente.

Como resultado separamos a la base de datos en las situaciones: **“Pagador, No pagador y No pidió préstamo”**

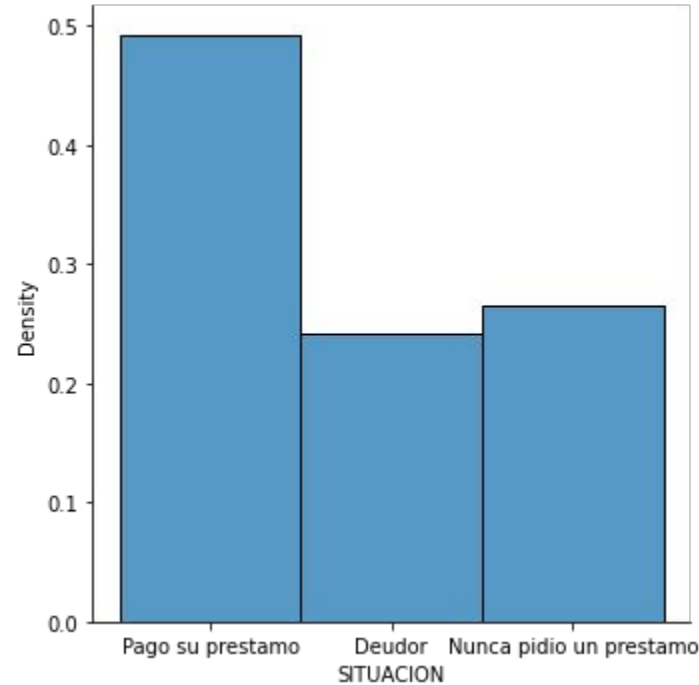


Descarte de Outliers

Se filtra la base de datos tomando los valores **presentes entre el percentil 25 y 75 de la muestra.**

Con esto evitamos resultados que alteren las etapas de Machine LeARNING.

Distribución de casos tras Reagrupación de Muestra



Exploratory Data Analysis



Empezamos el **Análisis Exploratorio de Datos** visualizando la información de **todas las características** presentes en la base de datos del banco “Buenos Amigos”

univariados

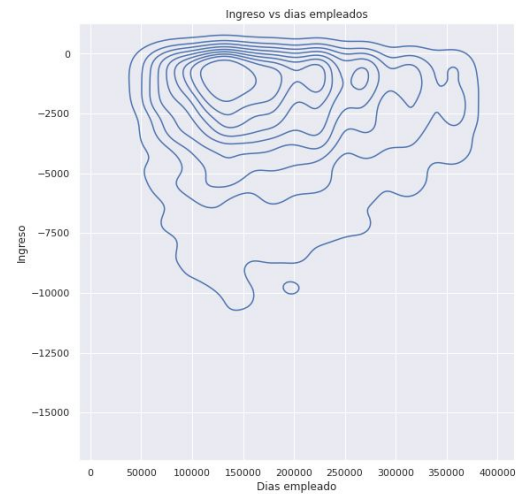
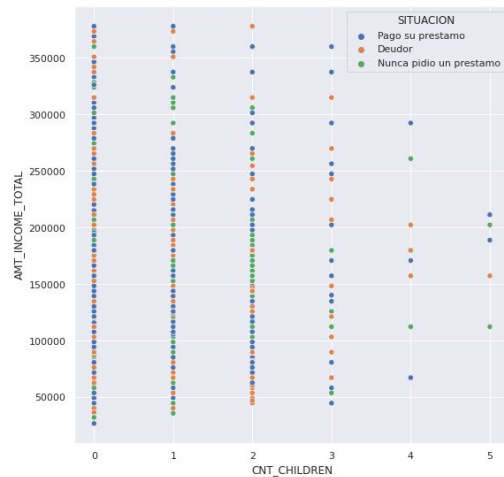
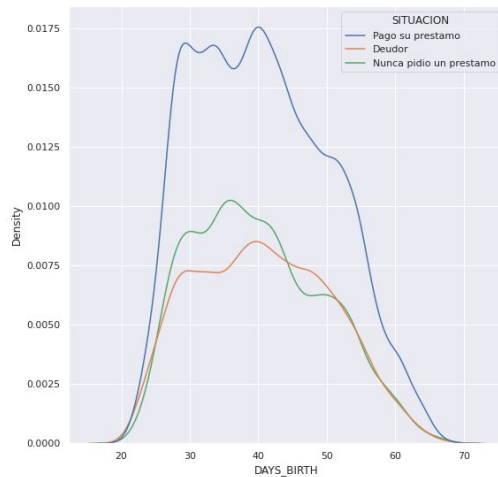
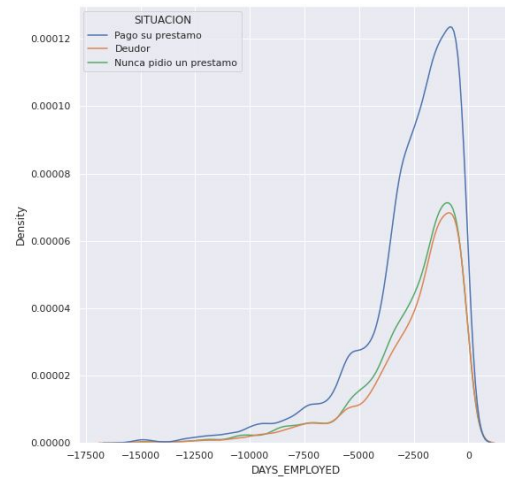
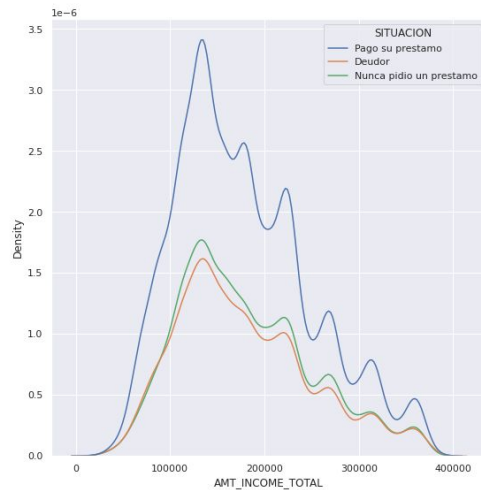
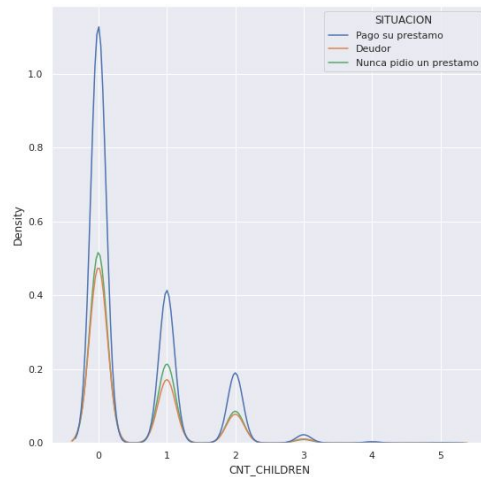


Grafico de barras por tipo de ingreso

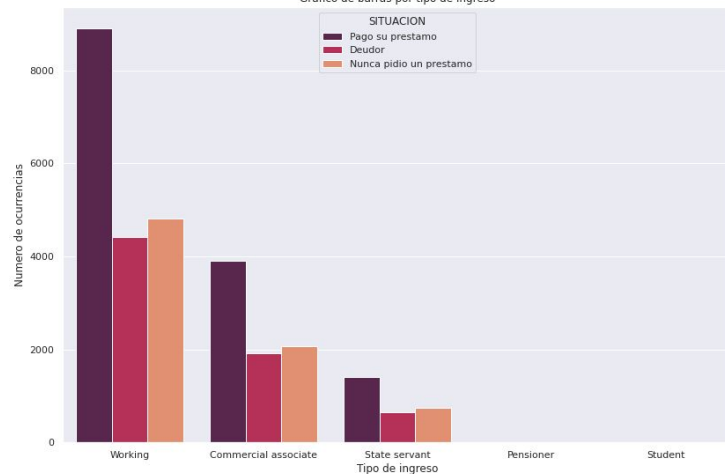


Grafico de barras por nivel de educativo

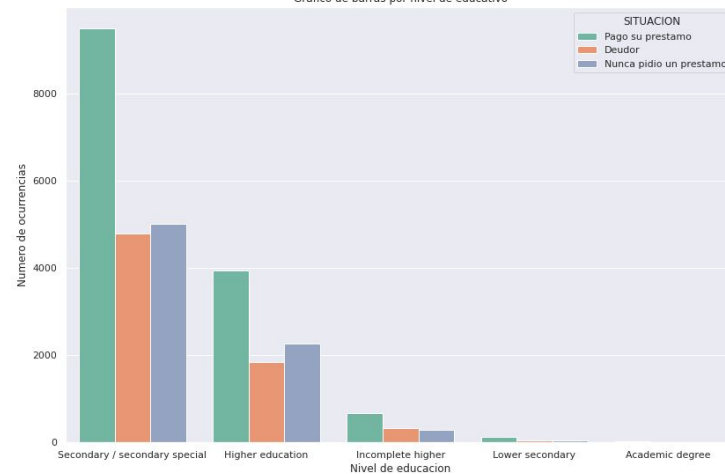


Grafico de barras por situacion civil

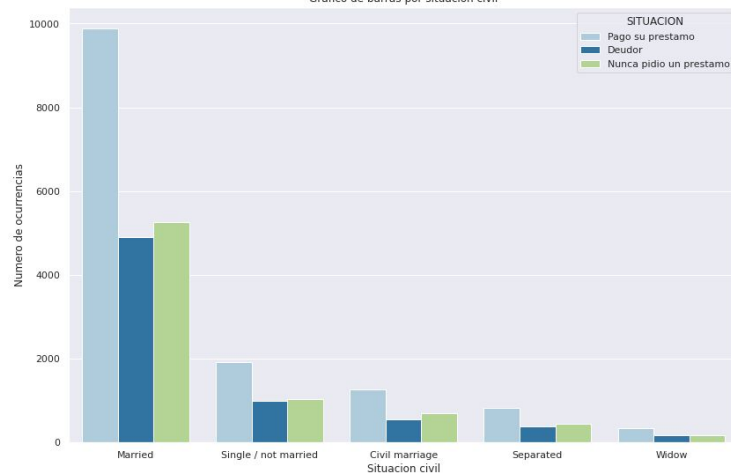
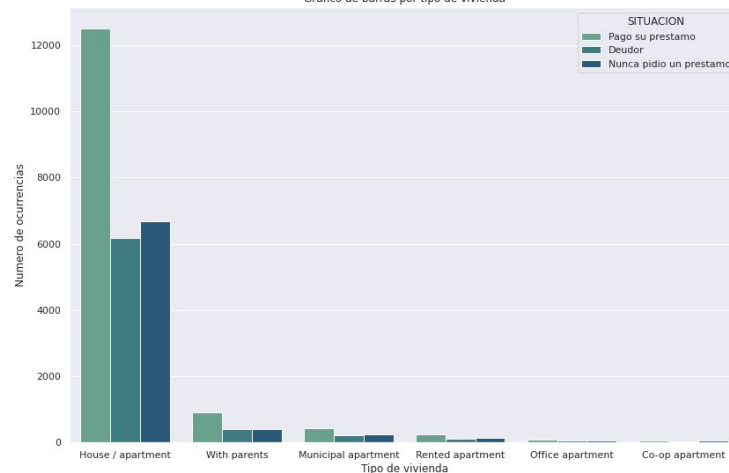
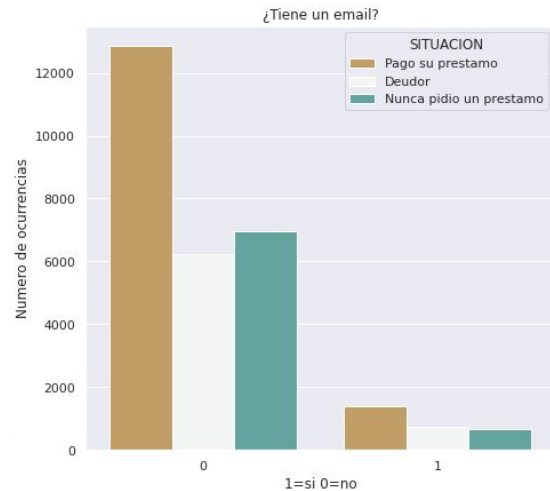
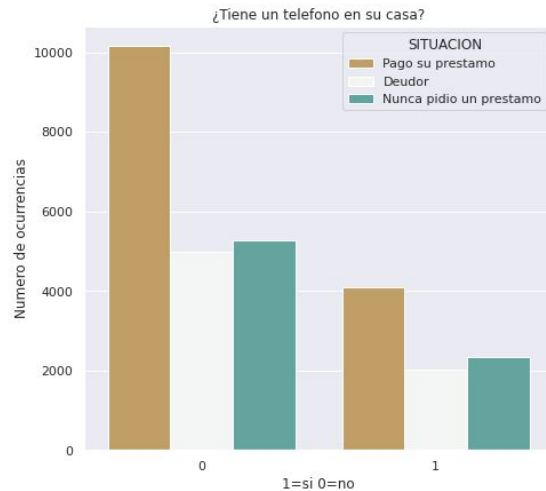
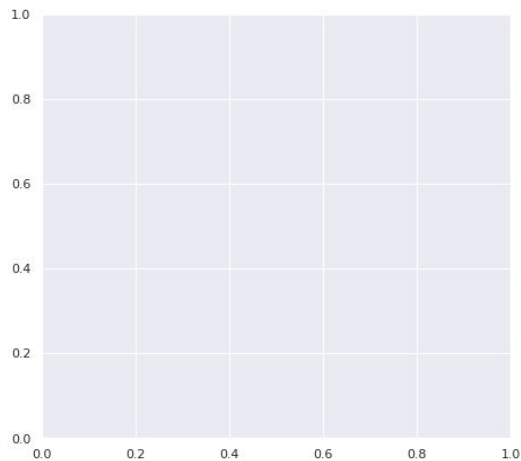
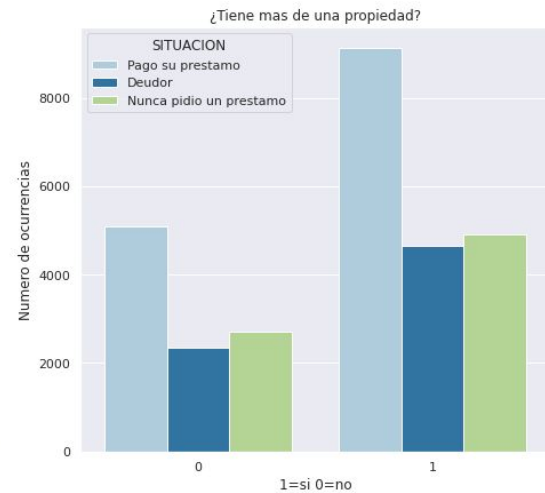
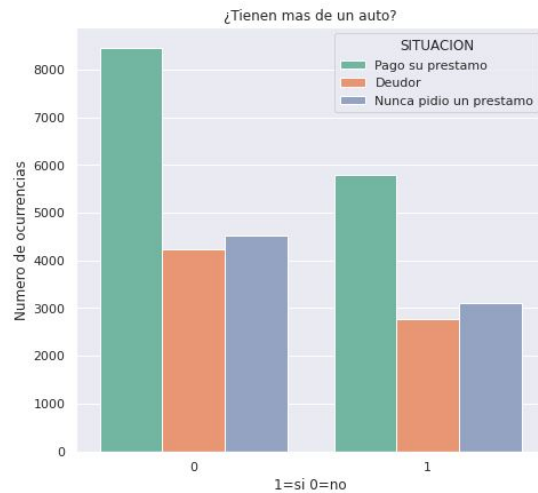
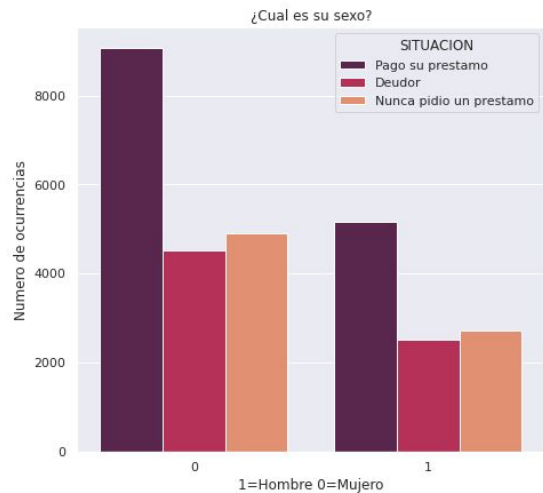


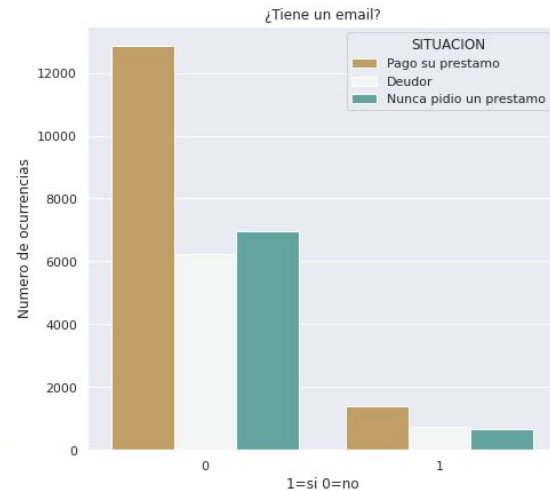
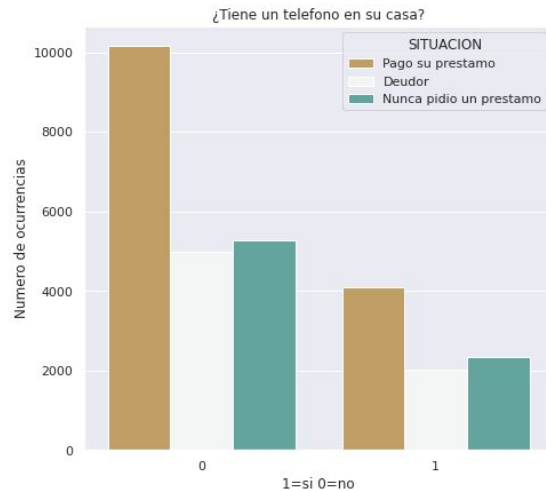
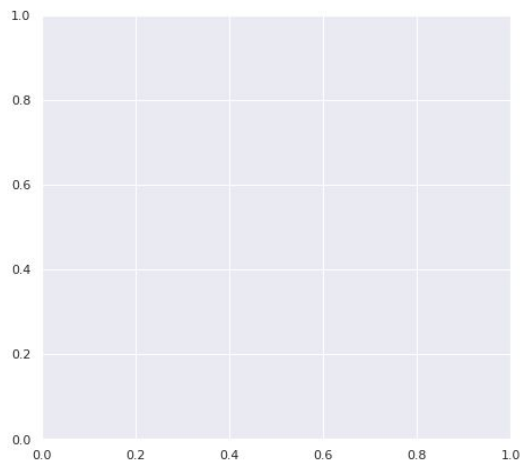
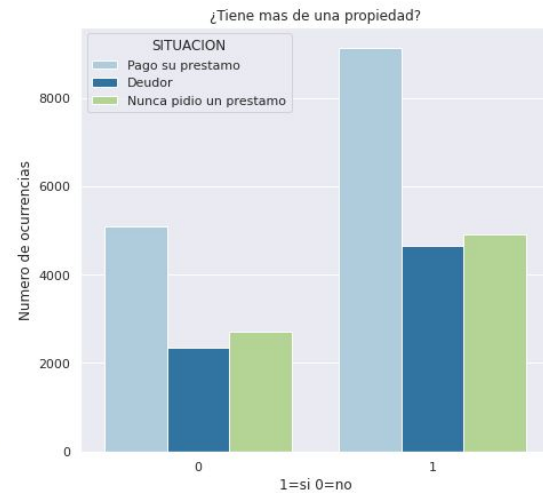
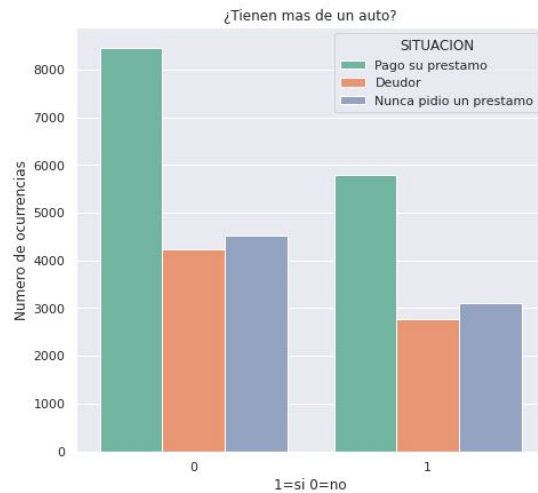
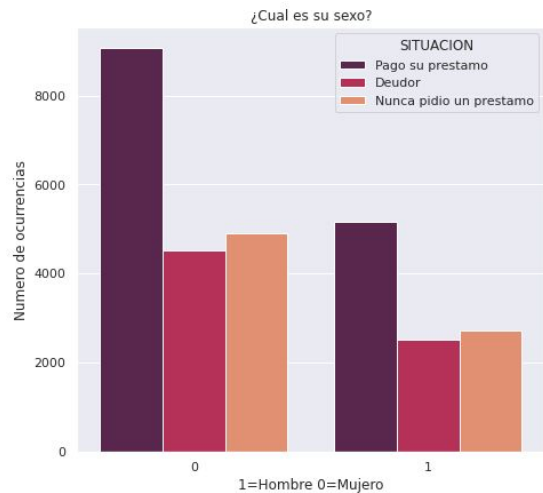
Grafico de barras por tipo de vivienda

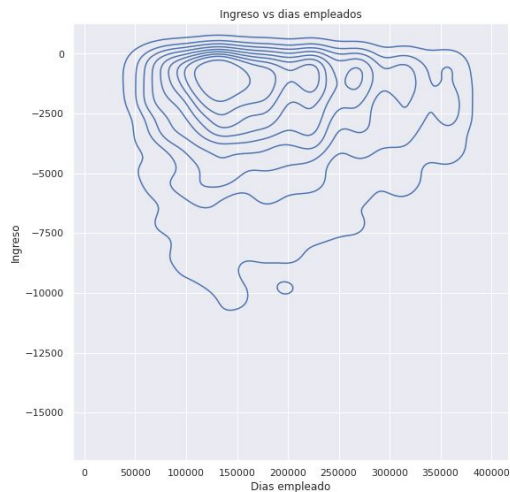
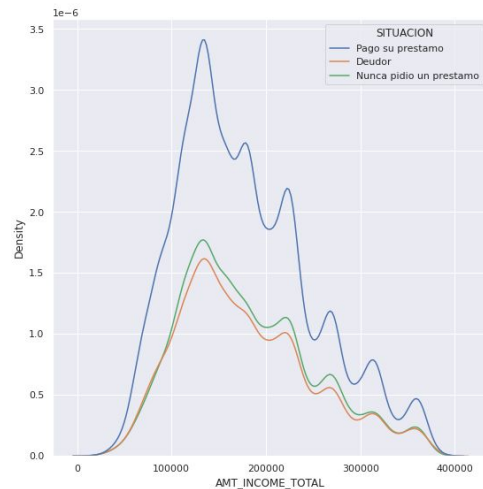


univariados

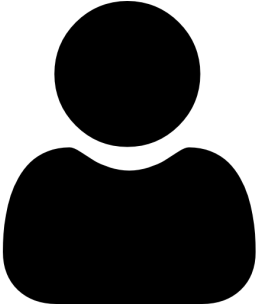


Bivariados





- Vemos como los **clientes que pagaron su préstamo tienen un ingreso más alto** en comparación a los otros grupos. Los picos dentro del gráfico de densidad de la variable ingreso nos hacen pensar que existen varios subgrupos (quintiles).
- La variable **Days employed** (Dias empleados) **presenta una asimetría negativa**, es entendible ya que eliminamos las observaciones con dias de empleo negativos.

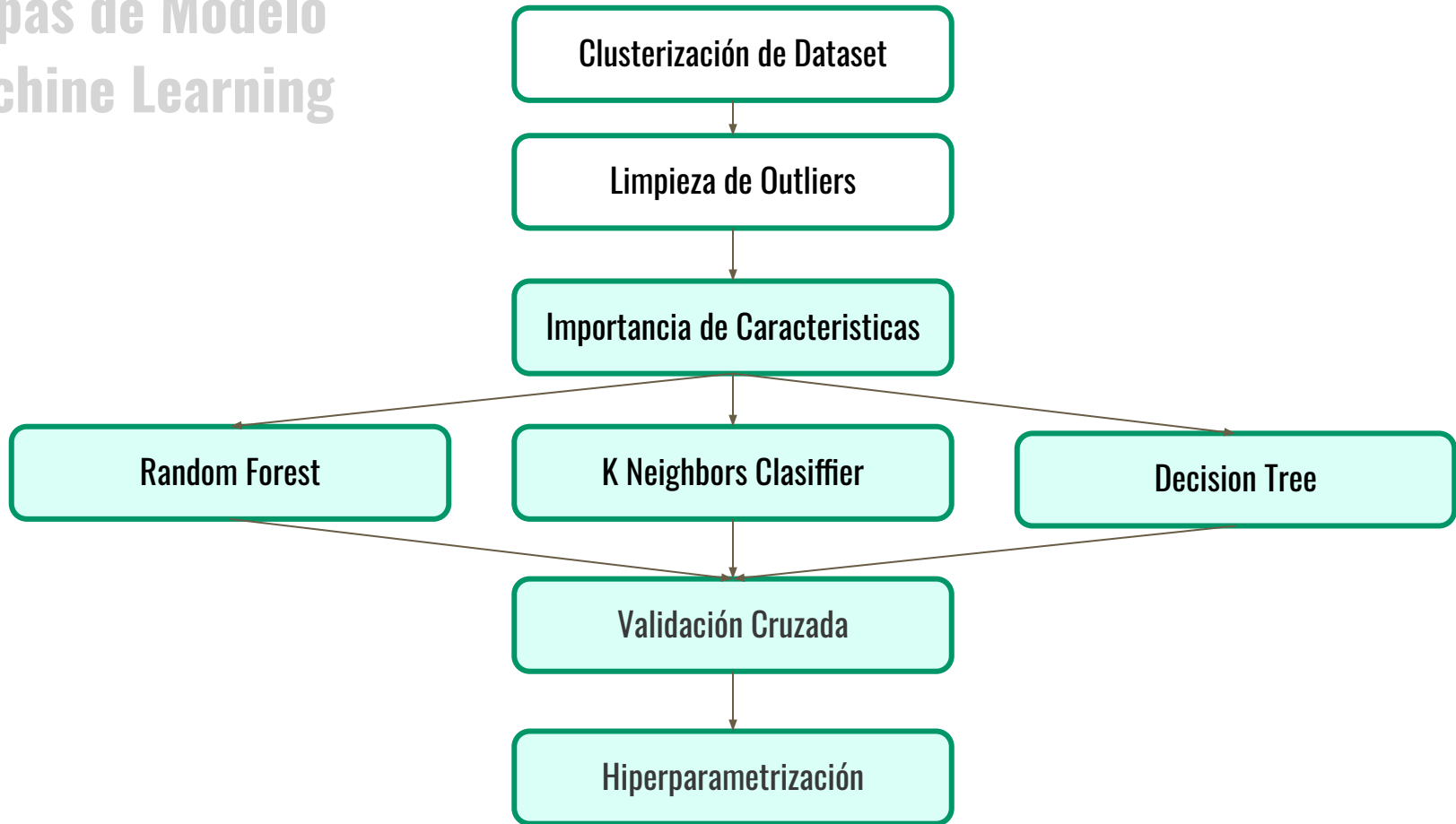


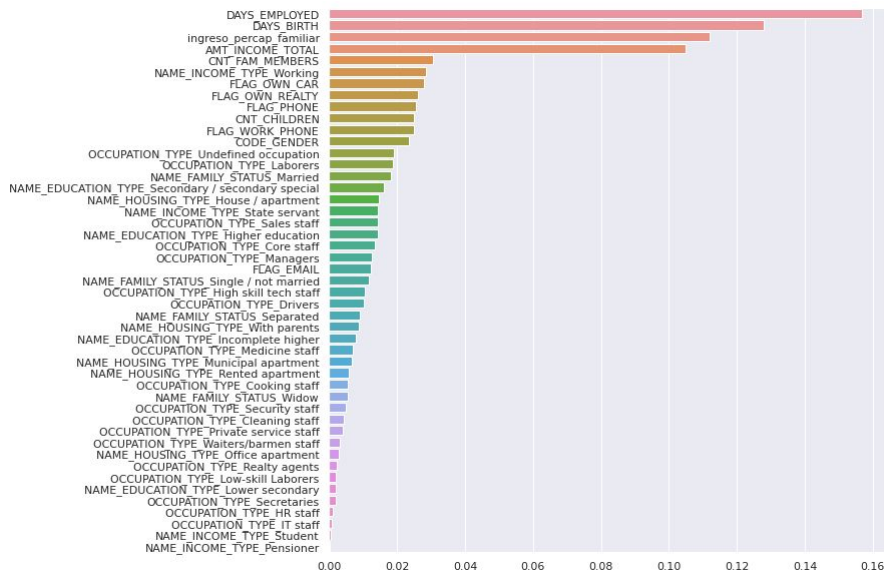
_ Casa / Apartamento
_ Casada
_ Secundaria Completa
_ Mujer
_ 2+ Propiedades

- La mayor parte de nuestras observaciones vive en una casa **o apartamento**, bajo un estado civil **casado**, son trabajadores con un nivel educativo máximo de **secundaria completa**. No vemos a simple vista una categoría dentro de las variables categóricas que por sí sola tenga una mayor probabilidad de default.
 - A priori en la muestra tenemos **más mujeres que pagan sus préstamos** que hombres que lo hacen. También vemos que contar con **más de una propiedad** lleva a pagar más los préstamos.
-

Aplicación de algoritmos de ML

Etapas de Modelo Machine Learning





- + Tiempo trabajado
- + Ingreso per cápita
- + Edad
- + Ingreso Total

- Nivel Universitario
- Nivel Secundario
- Nivel Primario
- Tipo de Ocupación

Importancia de Características

El modelo se basa mayormente en los días trabajados para definir a una **persona como deudora potencial o no**.

Seguido por la edad, el ingreso per cápita, el ingreso total, el tipo de trabajo, el estado civil, en que tipo de vivienda se establecen, y la condición de empleo.

Random Forest

	precision	recall	f1-score	support
0	0.45	0.24	0.32	2097
1	0.57	0.42	0.48	2264
2	0.60	0.82	0.70	4301
accuracy			0.58	8662
macro avg	0.54	0.50	0.50	8662
weighted avg	0.56	0.58	0.55	8662

K-Neighbours

	precision	recall	f1-score	support
0	0.36	0.36	0.36	2097
1	0.45	0.46	0.46	2264
2	0.63	0.62	0.62	4301
accuracy			0.52	8662
macro avg	0.48	0.48	0.48	8662
weighted avg	0.52	0.52	0.52	8662

Decision Tree

	precision	recall	f1-score	support
0	0.35	0.22	0.27	2097
1	0.47	0.34	0.39	2264
2	0.56	0.75	0.64	4301
accuracy			0.51	8662
macro avg	0.46	0.44	0.44	8662
weighted avg	0.49	0.51	0.49	8662

Selección del Modelo

Se trabajan con tres modelos de MachineLearning: **Random Forest**, **K-neighbours** y **Decision Tree**. En base a ellos, se toma el Accuracy para determinar con cual de ellos continuar la validación y optimización

Se termina **seleccionando el método Random Forest** dando un resultado parcil de 0.58 en Accuracy.

Evaluación del modelo

Tras un proceso de validación (K-folds) y optimización (Hiperparámetros), **nuestro modelo predice con un accuracy del 54%**, no está mal, pero como la muestra está desbalanceada no es la mejor métrica para evaluarlo.

Por lo que **nos centramos en el recall (de los positivos , cuántos efectivamente acertó el modelo)**, vemos que para los pagadores puntuales y en menor medida para los que nunca tomaron préstamos el recall es alto. Pudiendo discernir por descarte quien será un deudor.

Categorías	Precision	Recall	f1-Score	Obs
<i>Pagadores puntuales</i>	0.60	0.74	0.66	2610
<i>Deudores</i>	0.39	0.28	0.32	2929
<i>Nunca tomaron préstamo</i>	0.51	0.42	0.46	5399

<i>Accuracy</i>			0.58	10938
<i>macro avg</i>	0.50	0.48	0.48	10938
<i>weighted avg</i>	0.52	0.54	0.53	10938

Caso de Prueba: Jessica Perez

Para ejemplificar el modelo se procede a crear el caso del potencial cliente Jessica Perez, quien cuenta con las siguientes características:

- No tiene más de un auto
- Tiene más de una casa
- No tiene hijos
- Ingresa 216,000 dólares al año
- Tiene trabajo
- Tiene secundaria completa
- Está casada
- Vive en una casa o apartamento

```
CODE_GENDER      0.0
FLAG_OWN_CAR      0.0
FLAG_OWN_REALTY   1.0
CNT_CHILDREN      0.0
AMT_INCOME_TOTAL  216000.0
FLAG_WORK_PHONE   0.0
FLAG_PHONE        0.0
FLAG_EMAIL        0.0
CNT_FAM_MEMBERS   2.0
EMPLEADO          1.0
SECUNDARIA        1.0
PRIMARIA          0.0
SUPERIOR          0.0
SUPERIOR_INCOMPLETO 0.0
ACADEMICO         0.0
TRABAJO           1.0
SITUACION_CIVIL   1.0
DONDE_VIVEN       1.0
Name: 5135273, dtype: float64
2
```

Jessica pertenece al
grupo de **Pagadores
Puntuales**



Conclusiones



Banco
Buenos Amigos

1. El **ingreso del cliente es fundamental** a la hora de decidir si darle un crédito.
 2. Los **clientes que llevan más tiempo en situación de default tienen menores ingresos** que el resto de observaciones.
 3. El **ingreso es más alto en las familias poco numerosas.**
 4. En promedio, **contar con un automóvil es un buen indicativo** de que tienen un ingreso alto.
 5. Las personas **casadas tienen menos ingreso** con respecto a las solteras o divorciadas en promedio
-



Banco
Buenos Amigos

6. En promedio, **Los pensionados** es el grupo **con más ingresos** en comparación a los demás.

7. Mayores **niveles de educación completos** **llevan a obtener un mayor ingreso.**

8. Bajo qué tipo de **condición habitacional** subsiste el cliente (su propia casa/ alquiler/ casa de sus padres etc) **no es significativo para explicar el ingreso.**
