



TODO LO QUE TU BANCO SABE DE VOS



PRÁCTICO 1:
ANÁLISIS Y
VISUALIZACIÓN

Grupo 1

Integrantes:

Gonzalez, Patricia
Peralta, Agustín
Porcel, Carolina
Sosa, Manuel

CONTENIDO

**Análisis estadístico
de variables**

01.

02.

**Distribución de
Frecuencias**

**Análisis de clientes
con más de un
préstamo**

03.

04.

**Probabilidades
condicionales e
independencia entre
dos variables**

Análisis Estadístico



1. Distribución de variables

Monto de Capital: Distribución asimétrica hacia la derecha. La mayoría de los préstamos se concentran en montos menores a \$ 100.000

Edad: Distribución asimétrica hacia la izquierda. Esto podría deberse al hecho de que la mayoría de los clientes son jubilados, que tienen en el banco su cuenta para cobrar sus haberes.

Años de antigüedad: La distribución se asemeja a una distribución normal. Existe una medida de moda en el valor de 13 años. Para mayor inferencia deberíamos ahondar más en el análisis.

Cantidad de préstamos personales: Distribución asimétrica hacia la derecha.



Ingreso: Distribución asimétrica hacia la derecha. La mayoría de los clientes tienen ingresos menores a \$ 100.000. Tiene concordancia con lo visto en el monto de capital. A primera vista y por lógica, estas variables podrían estar relacionadas.

Total Adeudado: La distribución es asimétrica hacia la derecha. La mayoría de los clientes debe menos de \$ 200.000.

Clasificación de cliente: La mayoría de los clientes son jubilados con su cuenta de haberes en el banco (esto está relacionado con la variable edad) y aquellos que cobran sus salarios a través de cuentas sueldo.

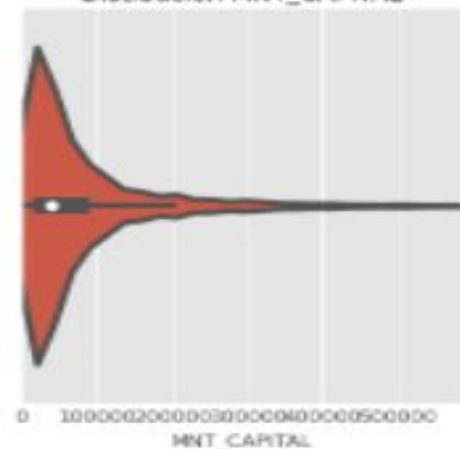
Sexo: Hay mayor cantidad de clientes de sexo femenino. En principio habría algún tipo de error en el registro H, que con posterioridad habría que analizar.

Clasificación valor económico del cliente: La mayor parte de los clientes tienen un valor económico muy alto para el banco

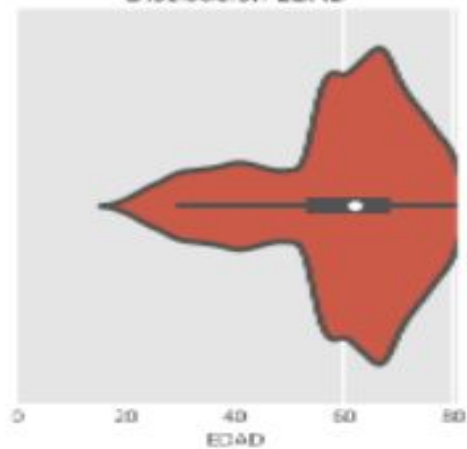
2. Estadísticos descriptivos para variables numéricas

	MNT_CAPITAL	EDAD	ANIOS_ANTIGUEDAD	CANTIDAD_PRESTAMOS_PERSONALES	INGRESO	TOTAL_ADEUDADO
count	62,473.00	142,473.00	142,473.00	142,473.00	142,473.00	142,473.00
mean	80,283.73	59.02	14.74	1.48	60,888.90	156.88
std	117,268.64	14.22	6.74	2.73	50,996.18	366.41
min	0.00	18.00	0.00	0.00	0.00	0.00
25%	17,201.00	53.00	13.00	0.00	21,409.00	12.00
50%	40,626.00	62.00	13.00	0.00	50,080.00	75.00
75%	92,263.00	69.00	18.00	2.00	81,467.00	193.00
max	2,232,452.00	82.00	67.00	37.00	1,362,310.00	39,539.00

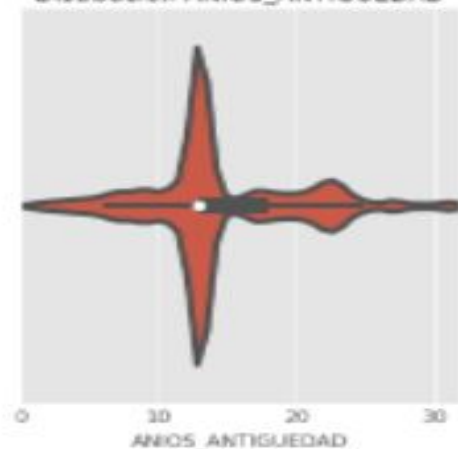
Distribución MNT_CAPITAL



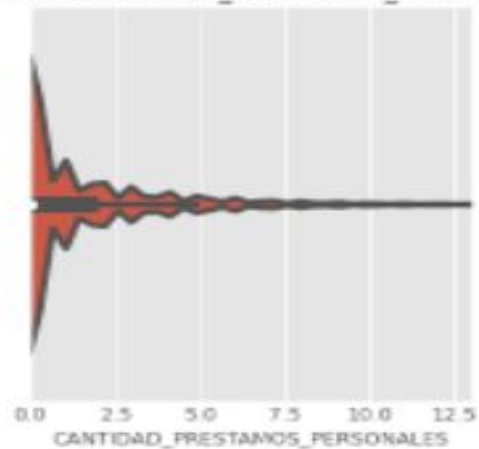
Distribución EDAD



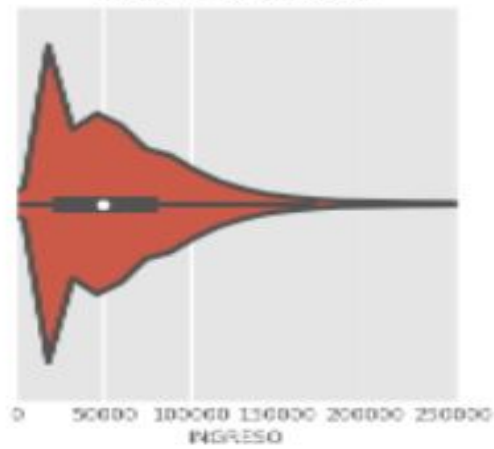
Distribución AÑOS_ANTIQUEDAD



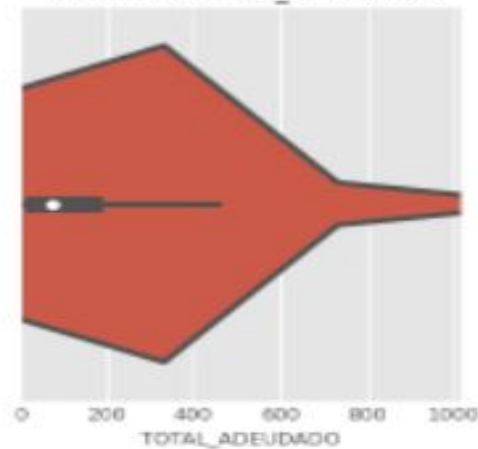
Distribución CANTIDAD_PRESTAMOS_PERSONALES



Distribución INGRESO

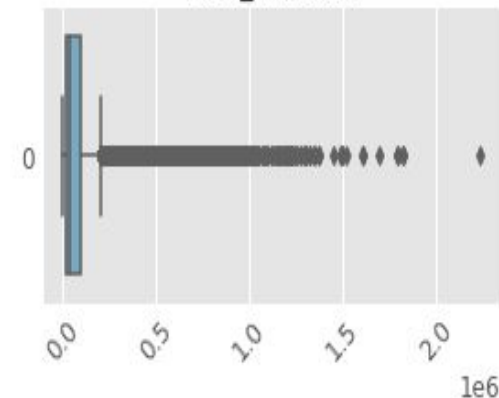


Distribución TOTAL_ADEUDADO

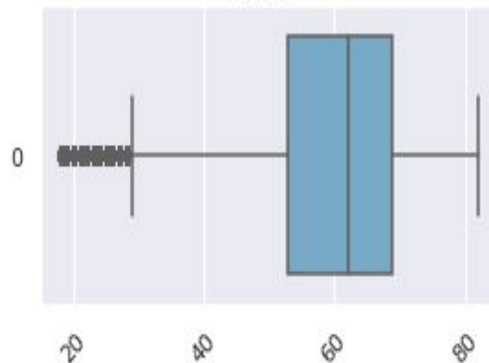


3. Gráficos Boxplot

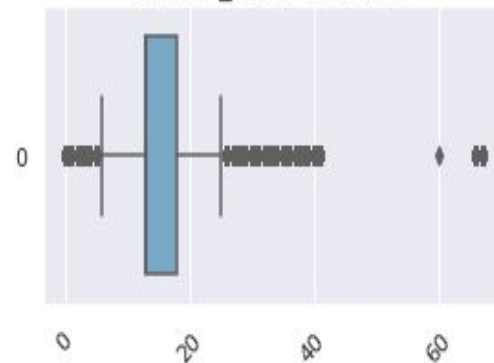
MNT_CAPITAL



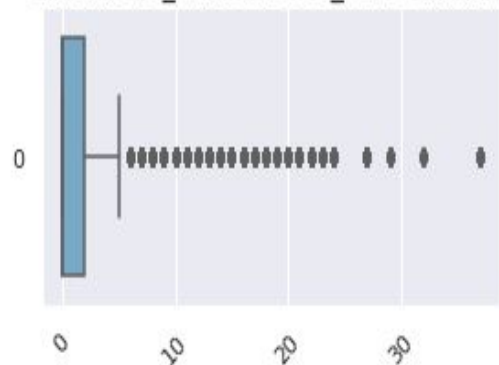
EDAD



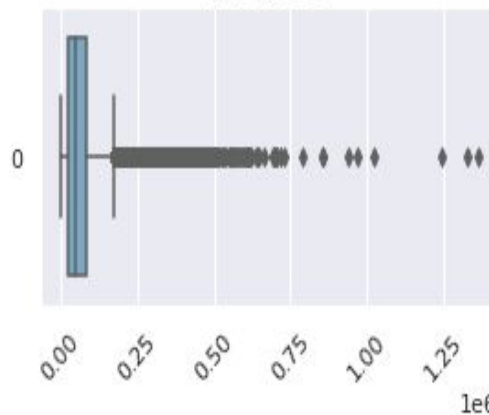
ANIOS_ANTIGUEDAD



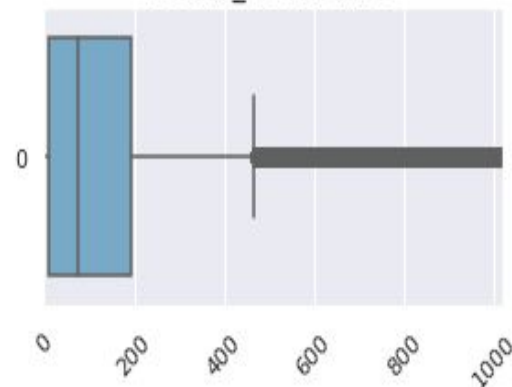
CANTIDAD_PRESTAMOS_PERSONALES



INGRESO



TOTAL_ADEUDADO



CALCULO DE VALORES ATIPICOS

MONTO CAPITAL

Maximo no atípico 204856.00
Cantidad mayores a Maximo 5583.00
Porcentaje de mayores a Maximo 3.92 %

AÑOS DE ANTIGÜEDAD

Maximo no atípico 25.50
Minimo no atípico 5.50
Cantidad mayores a Maximo 6522.00
Cantidad menores a Minimo 6670.00
Porcentaje de mayores a Maximo 4.58 %
Porcentaje de menores a Mínimo 4.68 %

INGRESOS

Maximo no atípico 171554.00
Cantidad mayores a Maximo 4120.00
Porcentaje de mayores a Maximo 2.89 %

CANTIDAD DE PRESTAMOS PERSONALES

Maximo no atípico 5.00
Cantidad mayores a Maximo 10740.00
Porcentaje de mayores a Maximo 7.54 %

EDAD

Maximo no atípico 93.00
Minimo no atípico 29.00
Cantidad mayores a Maximo 0.00
Cantidad menores a Minimo 5548.00
Porcentaje de mayores a Maximo 0.00 %
Porcentaje de menores a Mínimo 3.89 %

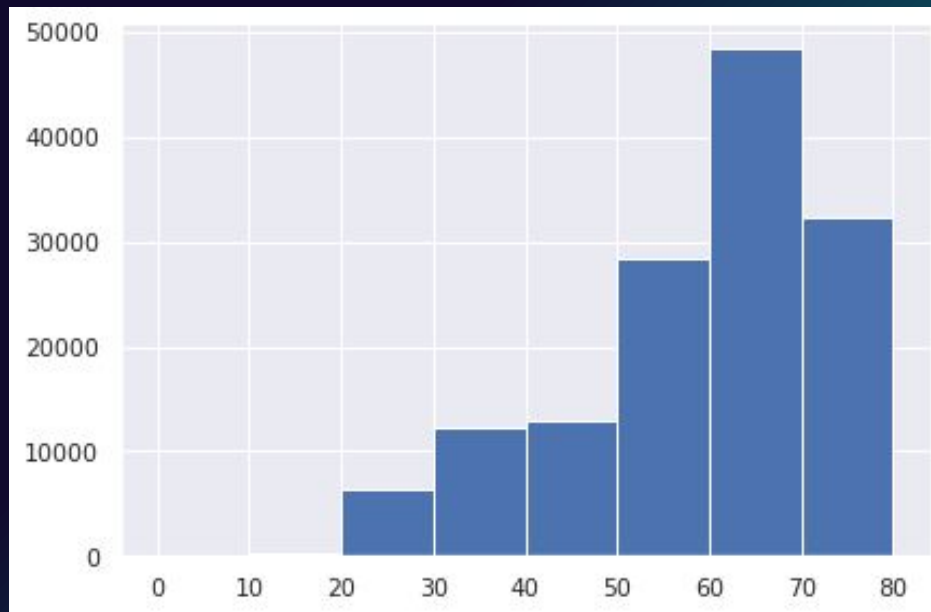
TOTAL ADEUDADO

Maximo no atípico 464.50
Cantidad mayores a Maximo 9950.00
Porcentaje de mayores a Maximo 6.98 %

Distribución de frecuencias



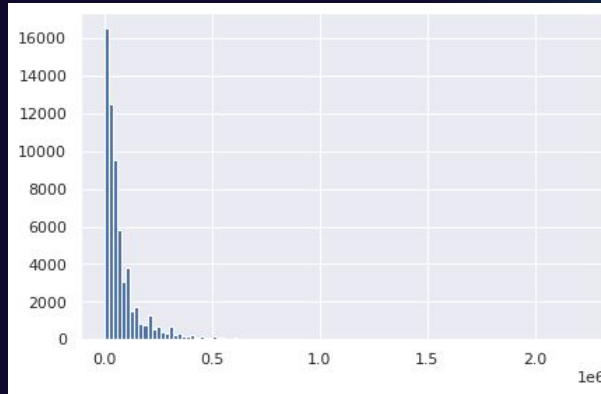
DISTRIBUCIÓN ETARIA DE LOS CLIENTES POR CORTES DE EDAD CADA 10 AÑOS



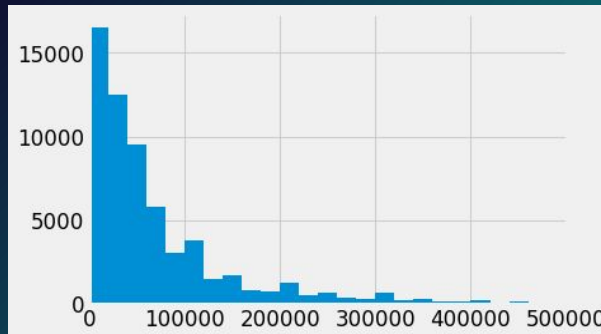
Los clientes entre 60 y 70 años es el grupo más frecuente.

DISTRIBUCIÓN DE MONTOS DE VENTA POR CORTES CADA 20.000 PESOS

(1)



zoom

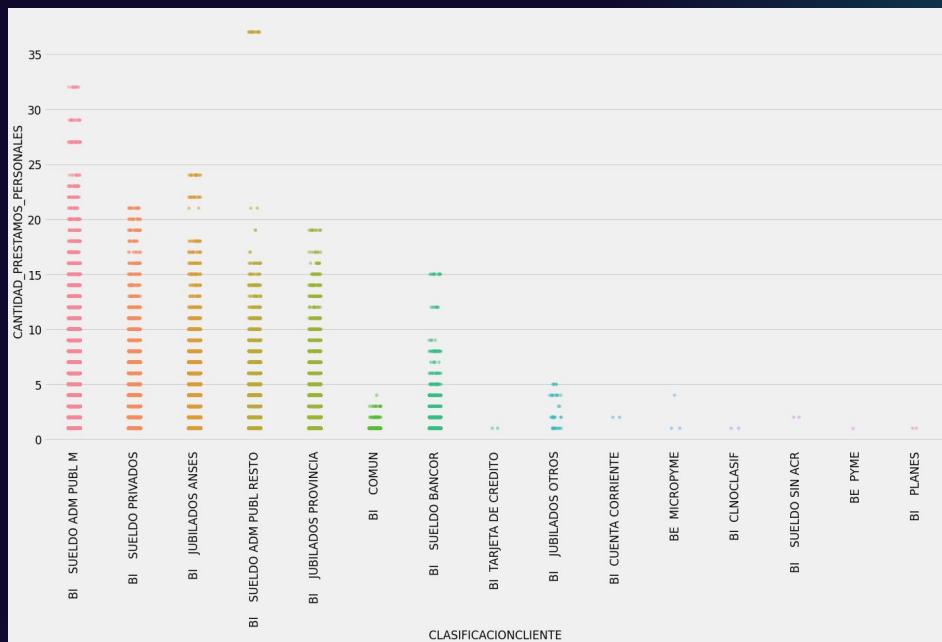


La mayor parte de los préstamos están entre \$20000 y \$60000

La distribución está sesgada hacia la derecha

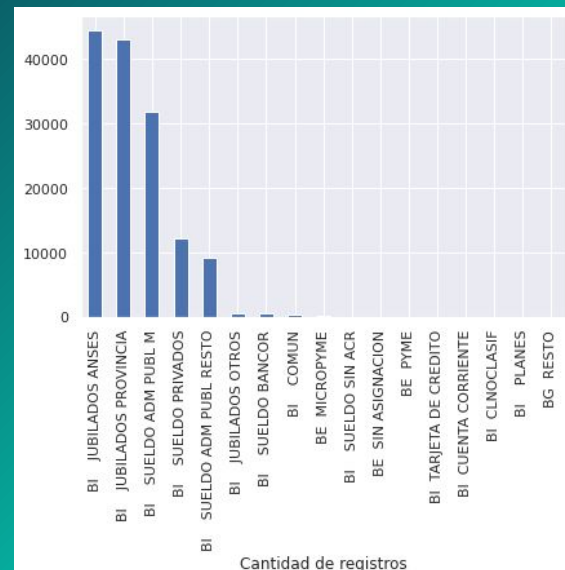
Esta variable tiene valores extremos (1).

DISTRIBUCIÓN DE PRESTAMOS DADOS DE ALTA POR TIPO DE CONVENIO(jubilado nacional, jubilado provincial, etc)

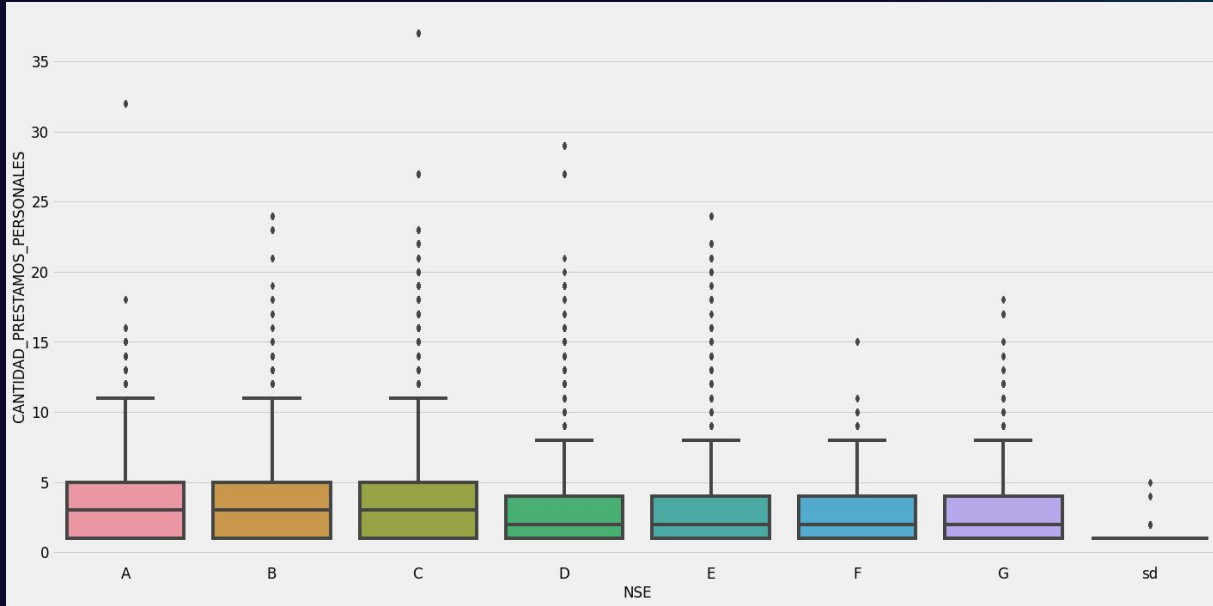


La categoría “BI Sueldo adm publ M” tiene mayor cantidad de préstamos dados de alta en relación a las otras.

Es importante tener en cuenta que hay 5 categorías que concentran la mayor parte de los registros.



DISTRIBUCION DE PRÉSTAMOS DADOS DE ALTA POR NIVEL SOCIO ECONÓMICO.



Los niveles socioeconómicos A,B y C tienen una distribución con mayor cantidad de préstamos, pero la diferencia no es significativa.

No se puede asegurar que los niveles socioeconómicos afectan a la distribución de la cantidad de préstamos.

Cientes con más de un préstamo



Cantidad de clientes con más de un préstamo: 41.690

Cantidad promedio de préstamos por cliente: 1.66

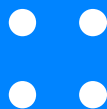
Análisis descriptivo de clientes con más de un préstamo:

El monto de capital promedio asociado es de \$ 53.896

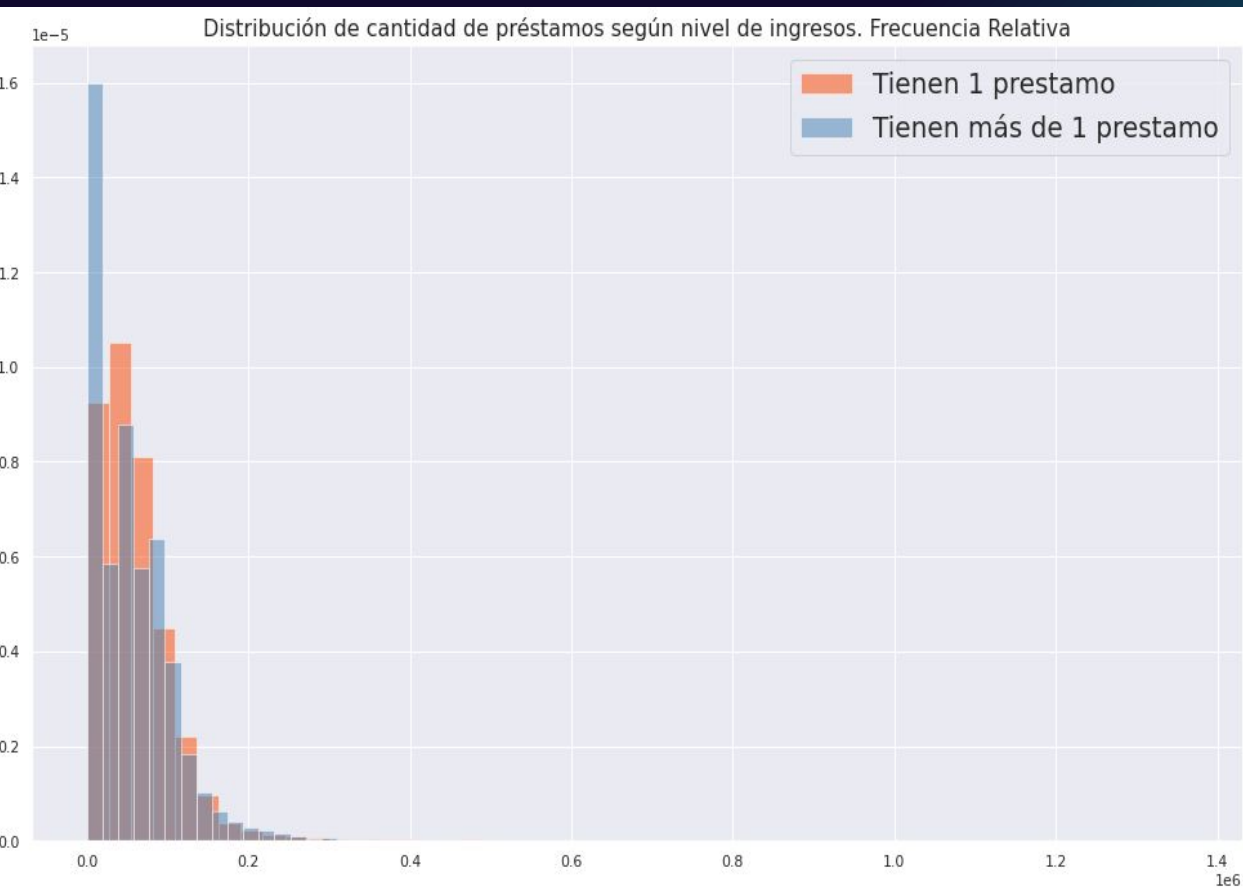
El ingreso promedio asociado es de \$ 59.166

La edad promedio de los clientes es cercana a los 50.

La antigüedad promedio de los clientes es entre 10 y 11

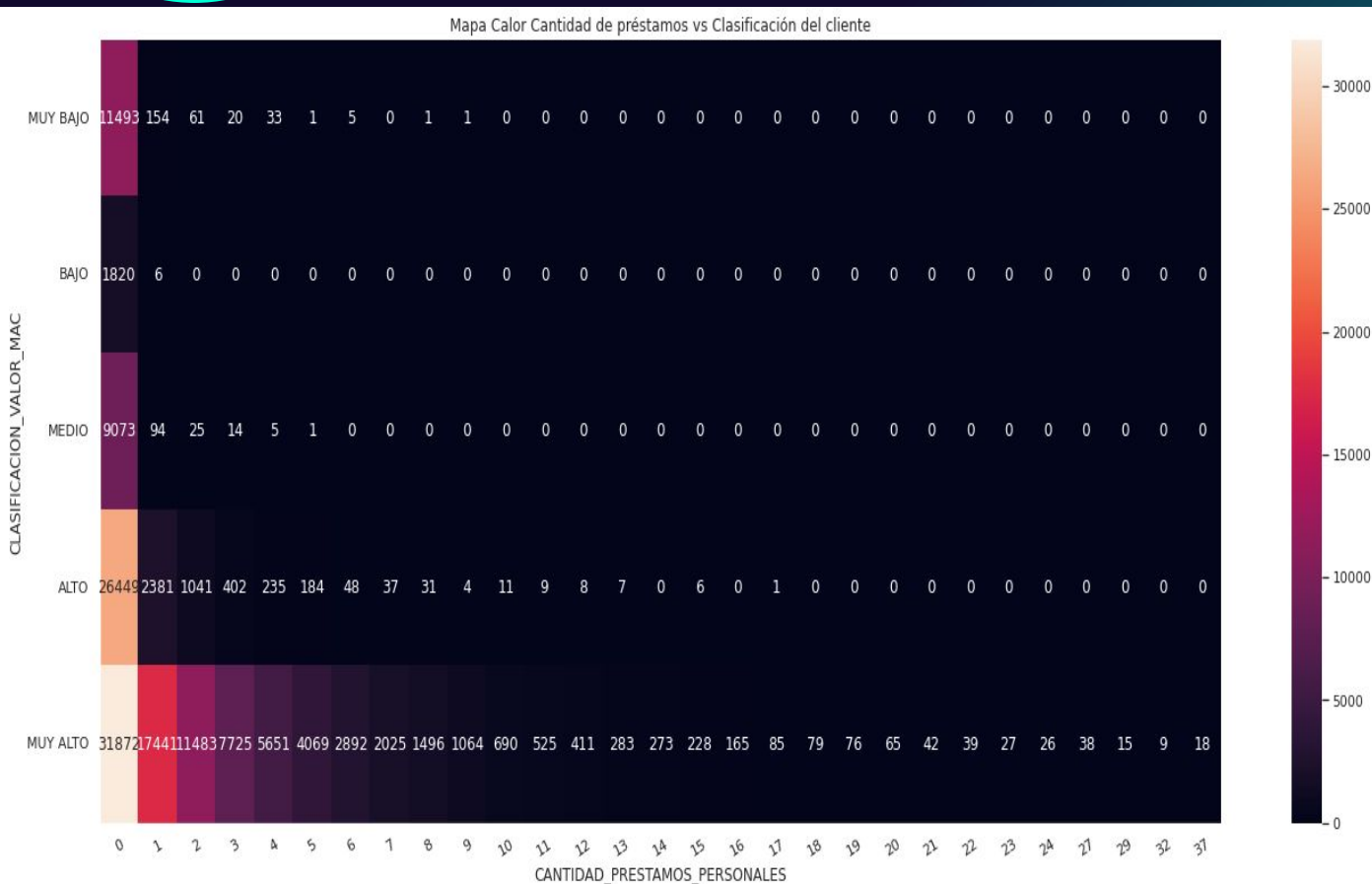


Cantidad de préstamos e ingresos



El hecho de tener un préstamo o más de 1, no parece en principio estar siendo afectado por el nivel de ingresos del cliente

Cantidad de préstamos y clasificación económica del cliente



Aquellos clientes que poseen más de 1 préstamo, son en una gran mayoría clientes con un valor económico muy alto para el banco.

Análisis de probabilidades condicionales e independencia entre dos variables



¿Existe algún tipo de relación entre el alta de préstamos y el nivel de endeudamiento?

$$P(A)=0.43$$

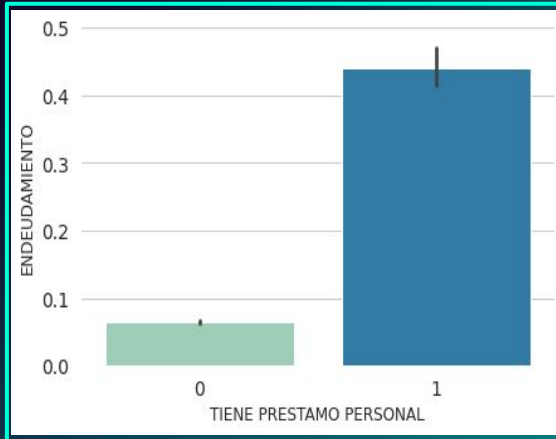
La probabilidad de tener un préstamo es del 43%

$$P(A/B)=0.63$$

La probabilidad de tener un préstamo dado que tengo una deuda es del 63%

$$P(A/B) \neq P(A)$$

Ya que son diferentes, se presume que las variables NO son independientes



PODRÍA EXISTIR UNA RELACIÓN ENTRE VARIABLES

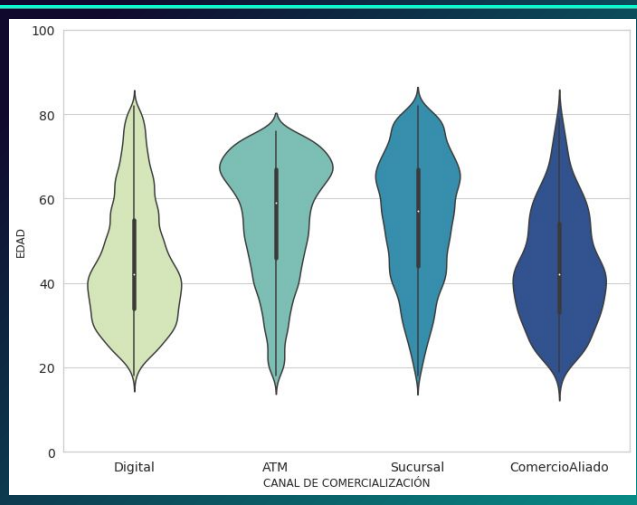
La probabilidad de tener un préstamo dado que se tiene alguna deuda es mayor que la probabilidad de tener un préstamo.

CONCLUSIONES

- Se esperaría que la relación fuera revés, es decir, a mayor deuda menor probabilidad de obtener un préstamo.
- Sin embargo, se requiere más información sobre la variable ENDEUDAMIENTO para interpretar el resultado.
- Una de las causas podría ser que el pago de las deudas le otorga al banco más seguridad respecto al cliente.
- Es necesario realizar alguna prueba de hipótesis para corroborar estadísticamente esta relación.

¿Existe algún tipo de relación entre la edad del cliente y el canal por el cual solicita el préstamo?

CANAL DE COMERCIALIZACIÓN	MEDIANA EDAD
DIGITAL	42
ATM	59
SUCURSAL	57
COMERCIO A.	43



ESTAS VARIABLES PODRÍAN ESTAR RELACIONADAS.
El canal de comercialización elegido depende de la edad del cliente.

CONCLUSIONES

- Los clientes de alrededor de 60 años utilizan más los canales ATM y SUCURSAL para solicitar el préstamo.
- Los clientes cercanos a la edad de 40 años prefieren los medios DIGITAL y COMERCIO ALIADO.
- En el gráfico se representan medidas de tendencia central, por lo cual solo se pueden hacer observaciones, se debería usar otra herramienta estadística para determinar si está relación es significativa.

¿ Existe alguna relación entre el alta de préstamos y el nivel de ingresos?

$$P(A)=0.43$$

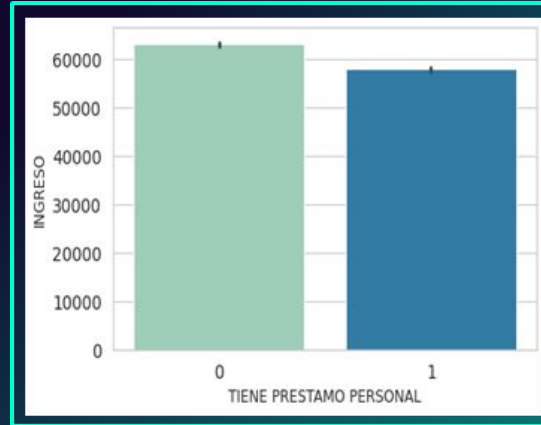
La probabilidad de tener un préstamo es del 43%

$$P(A/B)=0.41$$

La prob. de tener un préstamo dado que se tiene un ingreso superior al promedio es de 41%

$$P(A/B) \sim P(A)$$

Son similares, se interpreta que las variables son independientes

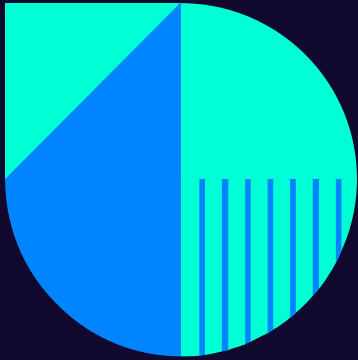


ES PROBABLE QUE SEAN VARIABLES INDEPENDIENTES

La probabilidad de tener un préstamo dado que se tiene un ingreso superior al promedio es similar (levemente menor) que la probabilidad de tener un préstamo.

CONCLUSIONES

- El ingreso promedio para la población es de 60888.9 pesos (std 50,996.18 , min=0 y max=1,362,310.00).
- Deberían limpiarse los datos antes de realizar el análisis.
- Quizás otras asociaciones, como por ejemplo el monto otorgado en función del ingreso, permitan explicar mejor la relación entre estas variables.
- Es necesario realizar alguna prueba estadística para demostrar que estas variables efectivamente son independientes.



GRACIAS!

Para mayor detalle de lo trabajado:

<https://colab.research.google.com/drive/15On9fLYUhgmLQF6IN2YnGcYg2chirfB#scrollTo=zIH6mUTpWZeS>

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon and infographics & images by Freepik

