



**Facultad de  
Ciencias**  
UNAM

ESTADÍSTICA 3. MODELOS DE SUPERVIVENCIA Y SERIES DE TIEMPO

---

## Proyecto final

ANÁLISIS DE DATOS DE CLIENTES BANCARIOS

---

Enríquez Hernández Leobardo  
Tlahuiz Tenorio Saúl Giovanni

6 de mayo de 2024

# Índice

<b>Introducción.</b>	<b>1</b>
<b>1. Estadística descriptiva.</b>	<b>1</b>
Selección de variables y verificación de datos faltantes. . . . .	1
Resumen de los datos numéricos. . . . .	2
Variables categóricas. . . . .	3
<b>2. Procesamiento de datos.</b>	<b>4</b>
<b>3. El problema de supervivencia.</b>	<b>4</b>
<b>4. Conclusiones</b>	<b>4</b>
<b>Referencias</b>	<b>5</b>

## Introducción.

Este documento tiene dos principales objetivos generales, mostrar algunos elementos estadísticos del análisis de supervivencia y al mismo tiempo con base en unos datos de clientes bancarios hacer algunas predicciones sobre la pérdida de clientes.

En la primera sección se hace la estadística descriptiva de los datos con los que se trabajará, para dar un contexto y un panorama general de la naturaleza y características de las variables. En la sección dos, haremos procesamiento de los datos en caso de que sea necesario por ejemplo tratar con valores perdidos, valores atípicos, etc. En la tercera sección se plantea y desarrolla el problema de supervivencia. Y finalmente, en la cuarta sección se presentan las principales conclusiones del trabajo.

## 1. Estadística descriptiva.

La base de datos es de clientes de un banco con las siguientes variables:

- id: número de fila de la observación, comenzando por el 0.
- CustomerId: número de cuenta del cliente.
- Surname: apellido.
- CreditScore: puntaje de crédito.
- Geography: país de residencia.
- Gender: género del cliente.
- Age: edad del cliente.
- Tenure: cuántos años ha tenido cuenta bancaria en el Banco.
- Balance: saldo de la cuenta.
- NumOfProducts: número de productos bancarios en el Banco.
- HasCrCard: si tiene o no tarjeta de crédito (sí=1).
- IsActiveMember: si es miembro activo del banco (sí=1).
- EstimatedSalary: salario estimado.
- Exited: si el cliente ha dejado el banco por algún periodo (sí=1).

### Selección de variables y verificación de datos faltantes.

Primero tomaremos un subconjunto del conjunto total de variables, omitiremos variables que no utilizaremos en el análisis tales como id, CustomerId, y Surname. Luego mostraremos en el siguiente cuadro que no hay datos faltantes (NA's) para las variables elegidas.

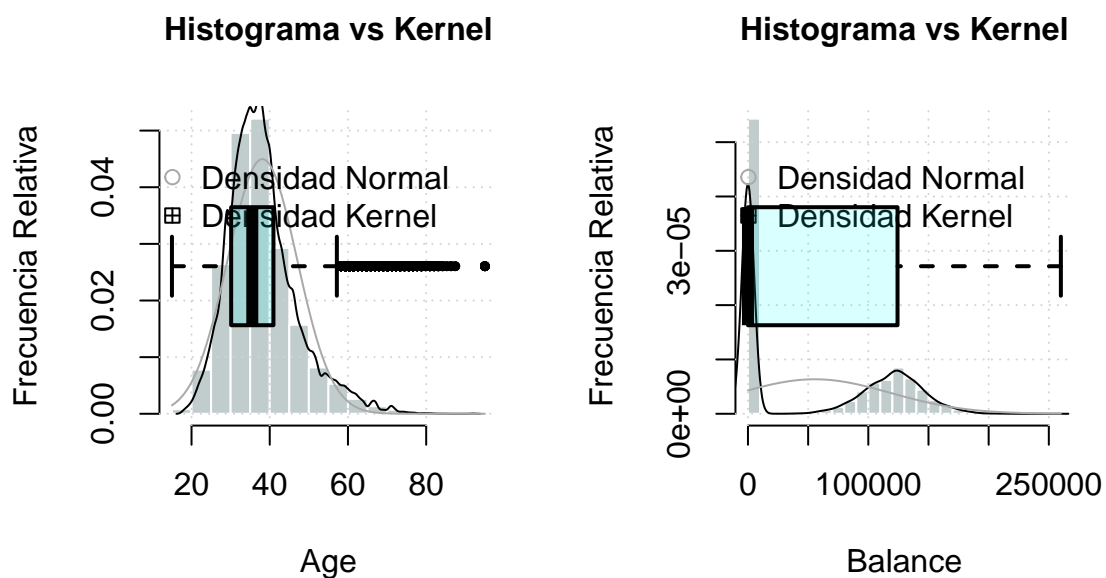
CreditScore : 0	Geography: 0	Gender: 0	Age: 0
Tenure : 0	Balance: 0	NumOfProducts: 0	HasCrCard: 0
IsActiveMember : 0	EstimatedSalary: 0	Exited: 0	

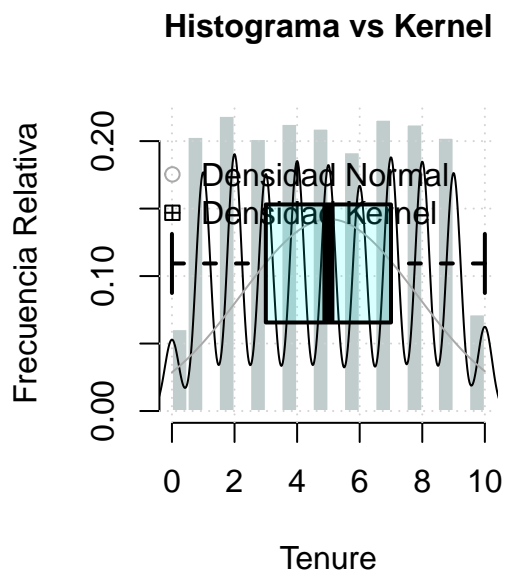
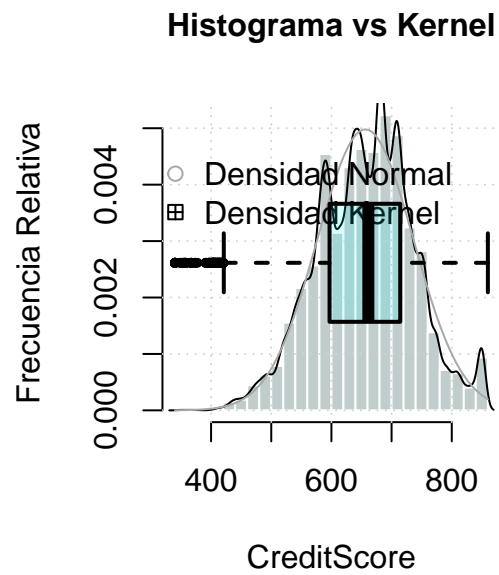
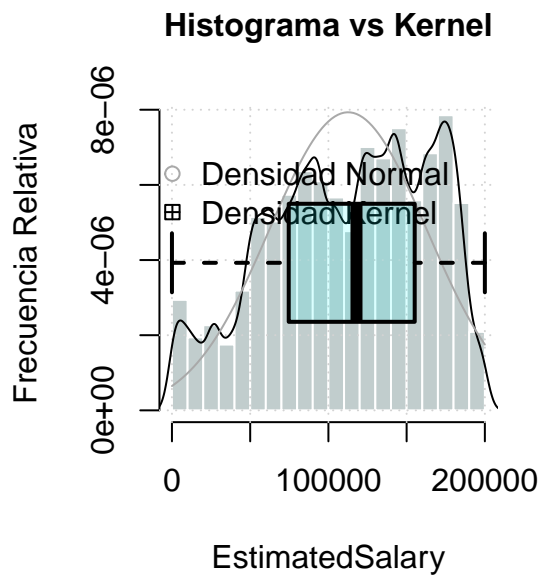
## Resumen de los datos numéricos.

A continuación se muestra la estadística descriptiva de los valores numéricos relevantes.

Statistic	N	Mean	St. Dev.	Min	Max
Age	165,034	38.126	8.867	18.000	92.000
Balance	165,034	55,478.090	62,817.660	0.000	250,898.100
EstimatedSalary	165,034	112,574.800	50,292.870	11.580	199,992.500
CreditScore	165,034	656.454	80.103	350	850
Tenure	165,034	5.020	2.806	0	10

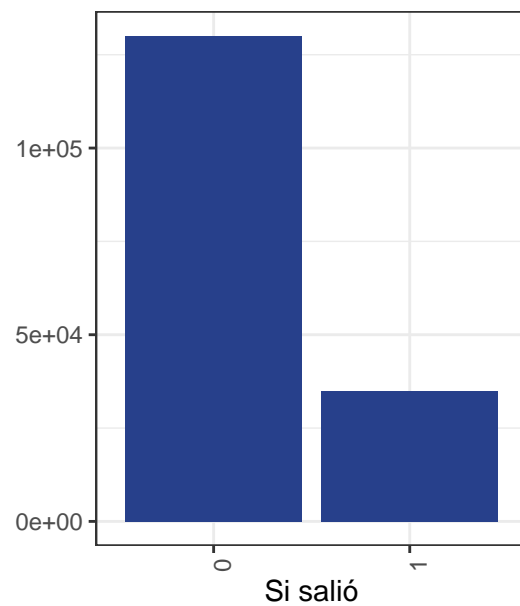
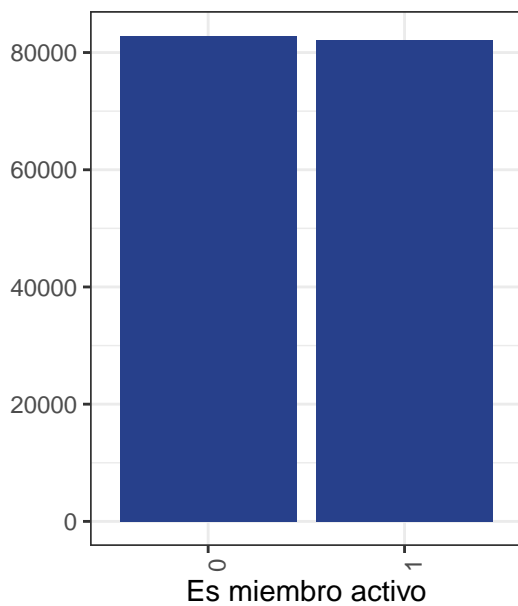
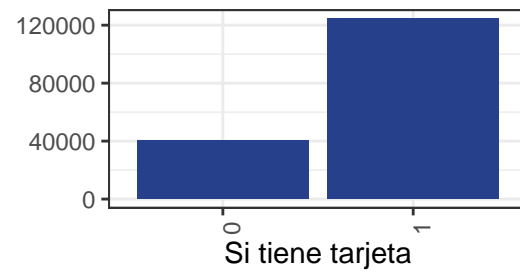
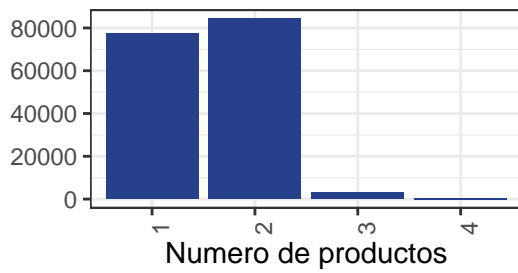
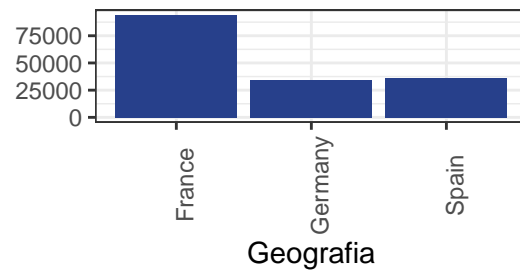
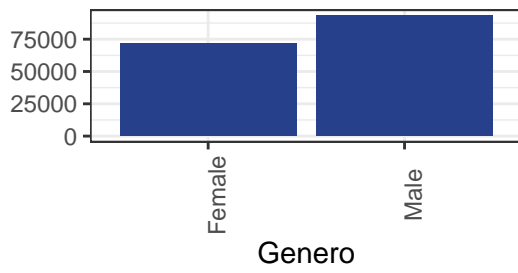
A continuación se muestran los histogramas, distribuciones y boxplot para las variables numéricas.





### Variables categóricas.

A continuación presentaremos los datos categóricos, por su frecuencia absoluta en la base de datos.



2. Procesamiento de datos.

3. El problema de supervivencia.

4. Conclusiones

## Referencias