

Proyecto Final



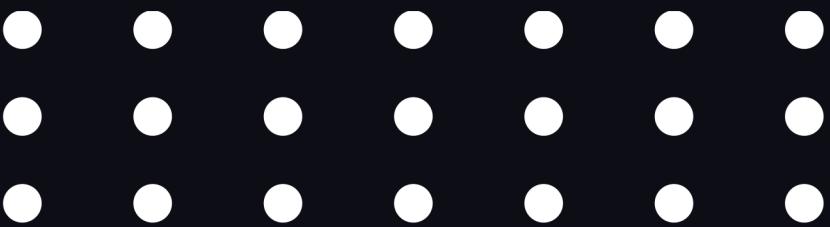
DATA SCIENCE | Norberto Stelmach

PROFESOR | Arturo Tapia Sanchez



índice

- 
 - 1 Introducción
 - 2 Presentación del Caso
 - 3 Preguntas y Objetivos
 - 4 EDA [Exploratory Data Analysis]
 - 5 Algoritmos de Machine Learning
 - 6 Conclusión



Introducción ①

¿ Quién soy ?

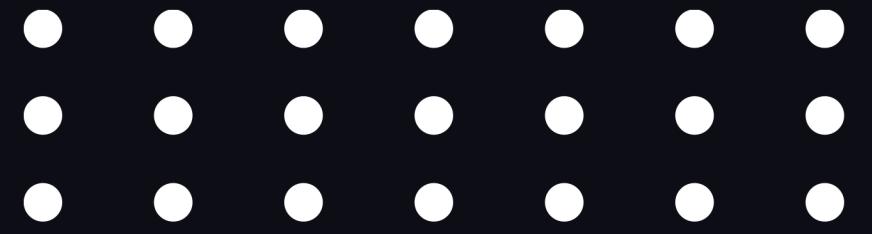
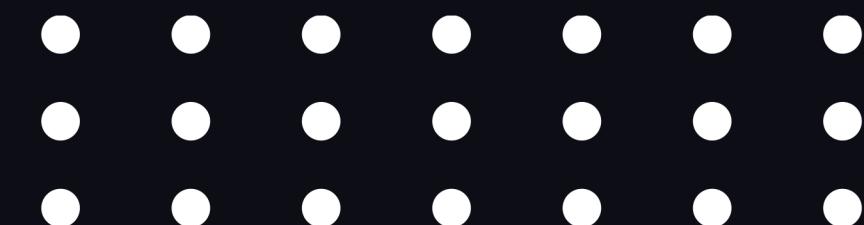
Norberto Stelmach

- Founder/CEO de Clasifreelance
- Diseñador gráfico
- Editor de video / animación
- Analista de datos.



2

Presentación del Caso



Presentación del Caso 2

Analizamos datos de clientes bancarios para entender sus necesidades y comportamientos. El objetivo es mejorar la retención y reducir el abandono, impactando en la rentabilidad del banco. El análisis también informará decisiones estratégicas y mejoras en productos y servicios. En resumen, buscamos mejorar la satisfacción del cliente y el éxito financiero del banco mediante el análisis de datos en la industria bancaria.

RowNumber	CustomerId	Surname	CreditScore	Country	Card_Category	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	
0	1	15634602	Hargrave	619	France	Blue	Female	42	2	0.00	1	1	1	101348.88	1
1	2	15647311	Hill	608	Spain	Blue	Female	41	1	83807.86	1	0	1	112542.58	0
2	3	15619304	Onio	502	France	Blue	Female	42	8	159660.80	3	1	0	113931.57	1
3	4	15701354	Boni	699	France	Blue	Female	39	1	0.00	2	0	0	93826.63	0
4	5	15737888	Mitchell	850	Spain	Blue	Female	43	2	125510.82	1	1	1	79084.10	0

Hipótesis

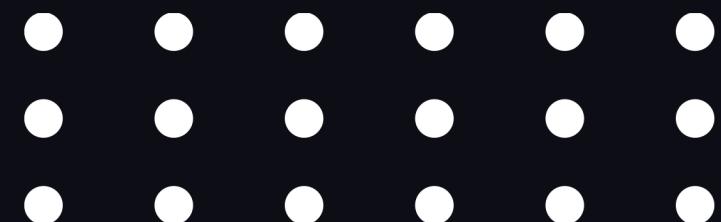
2

Con el análisis de este conjunto de datos podemos observar el comportamiento de los clientes en términos de género, edad, historial crediticio, tarjetas de crédito, estatus actual, salario y productos contratados en esta entidad financiera.

Podremos generar informes segmentados y obtener conclusiones específicas sobre la retención y abandono de clientes.

Estas conclusiones servirán como base para tomar decisiones estratégicas sobre la expansión de la cartera de clientes y la fidelización, ofreciendo servicios y beneficios personalizados.

Al analizar las columnas (Age, Country, Gender, Exited), obtendremos información relevante para perfilar a los clientes de la entidad.



Posibles problemas



Tasa de abandono: Identificar y reducir la tasa de abandono de clientes para preservar la rentabilidad.

Satisfacción del cliente: Mejorar productos y servicios para aumentar la satisfacción de los clientes.

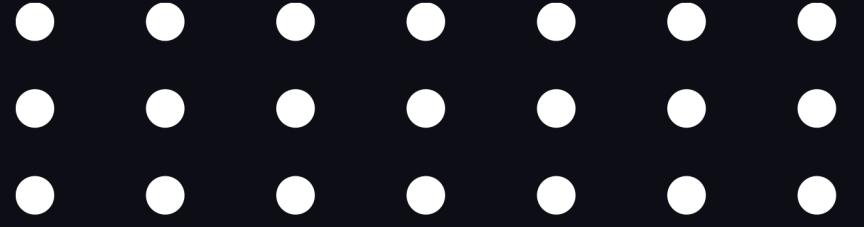
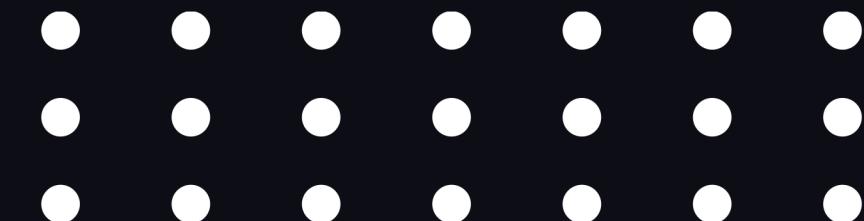
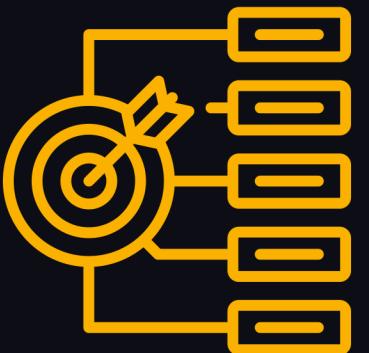
Expansión a nuevos mercados: Identificar mercados potenciales y tomar decisiones informadas sobre la expansión.

Retención de clientes: Identificar factores clave para la retención y aplicar medidas para mejorarla.

Mejora de productos y servicios: Identificar áreas de mejora y tomar decisiones informadas para actualizar servicios y productos.

3

Preguntas y objetivos



Preguntas

¿Cómo se distribuye la cantidad de clientes de cada país?

Explorar la diversidad cultural de los clientes bancarios revela patrones únicos en la distribución por país. Este análisis ofrece insights sobre la realidad social y económica, identifica oportunidades para mejorar la inclusión financiera en regiones subrepresentadas, y permite adaptar productos y servicios bancarios a las necesidades específicas de cada país.

¿Cómo se distribuyen los diferentes segmentos etarios?

Explorar las necesidades financieras a lo largo de la vida revela patrones distintivos en diferentes segmentos etarios. Desde jóvenes que ahorran para educación hasta mayores que protegen su patrimonio, cada grupo tiene características y necesidades únicas. El análisis de la distribución por segmentos etarios permite adaptar productos y servicios bancarios para ayudar a los clientes a alcanzar sus objetivos financieros a lo largo de su vida.

¿Cuál es el género más frecuente de clientes?

Explorar la distribución de poder financiero entre hombres y mujeres revela patrones que reflejan desigualdades de género en el acceso a servicios financieros e ingresos. El análisis de género permite identificar oportunidades para promover la inclusión financiera y empoderar a las mujeres en la toma de decisiones financieras.



- A 3x6 grid of white circles on a black background. The circles are arranged in three rows and six columns, forming a pattern of dots.

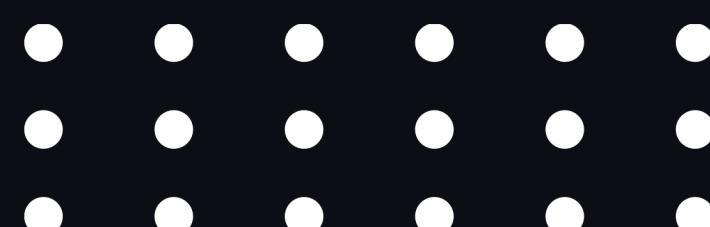
Objetivos



3

Los datos que analizaremos se refieren a clientes de una entidad bancaria. El contexto comercial de estos datos es, por lo tanto, la industria bancaria. El objetivo de analizar estos datos es comprender mejor el comportamiento y las necesidades de los clientes para poder ofrecer productos y servicios que satisfagan sus necesidades y mejoren la experiencia del cliente. También se busca mejorar la retención de clientes y reducir la tasa de abandono del banco, ya que esto puede tener un impacto significativo en la rentabilidad y la reputación del banco.

Además, el análisis de estos datos puede ayudar a la entidad bancaria a tomar decisiones estratégicas, como la expansión a nuevos mercados o la mejora de los productos y servicios existentes. En resumen, el contexto comercial de estos datos es la industria bancaria y el objetivo es mejorar la satisfacción del cliente y la rentabilidad del banco a través del análisis de datos.



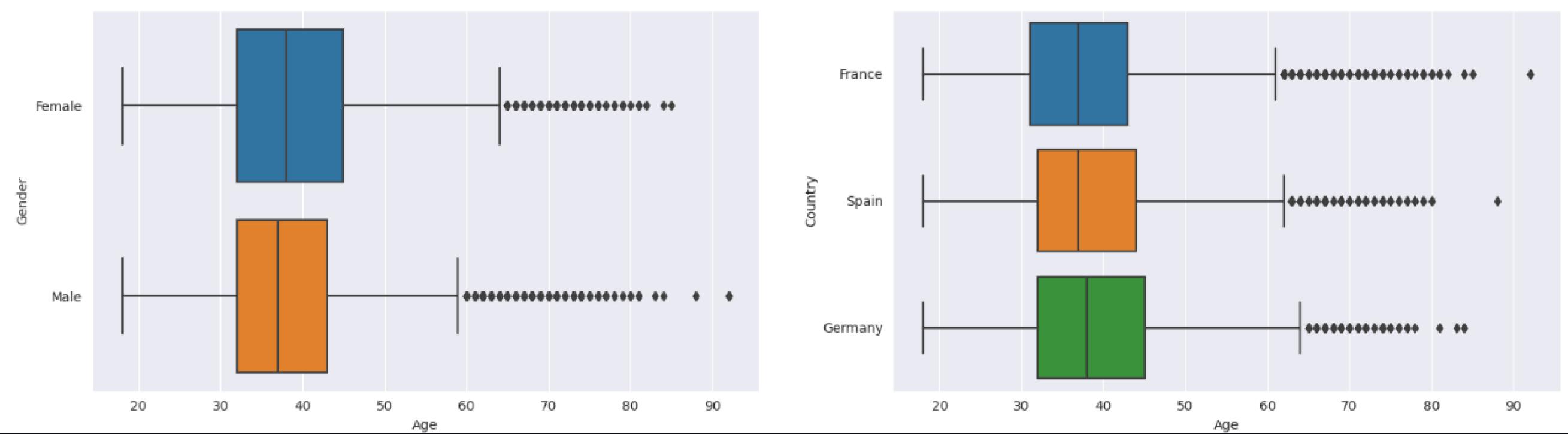
4

EDA

Exploratory Data Analysis

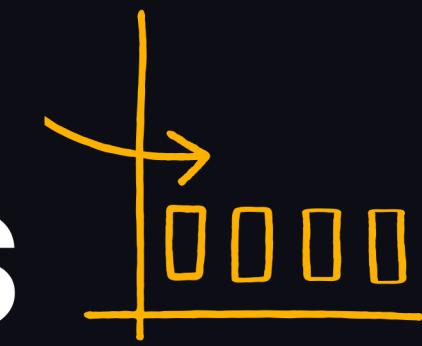


Outliers 4

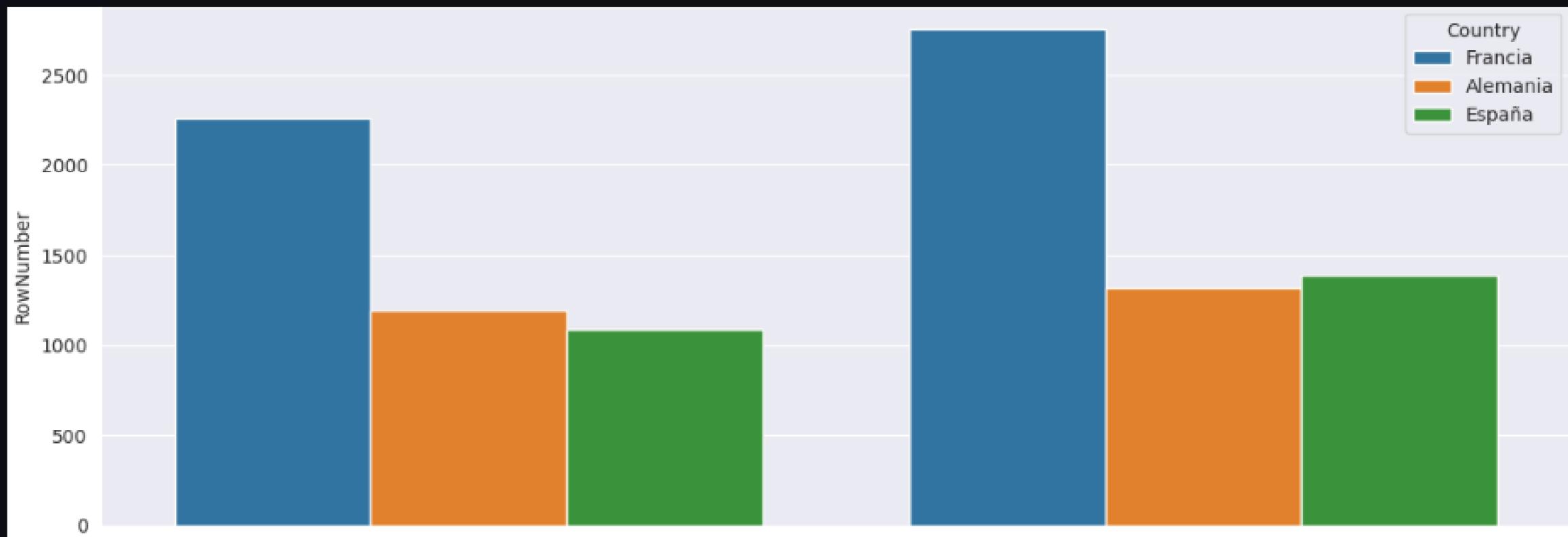


El análisis del último gráfico revela que la concentración principal de clientes en cartera está entre los 32 y 45 años. Se observa un rango menor en mujeres, con algunos valores atípicos en hombres mayores de 90 años. Francia tiene clientes más longevos, seguido de España y Alemania. Alemania muestra un rango etario mayor, mientras que en España y Francia disminuye. Este panorama proporciona información valiosa para la toma de decisiones independientes por región y género.

Gráfico de barras

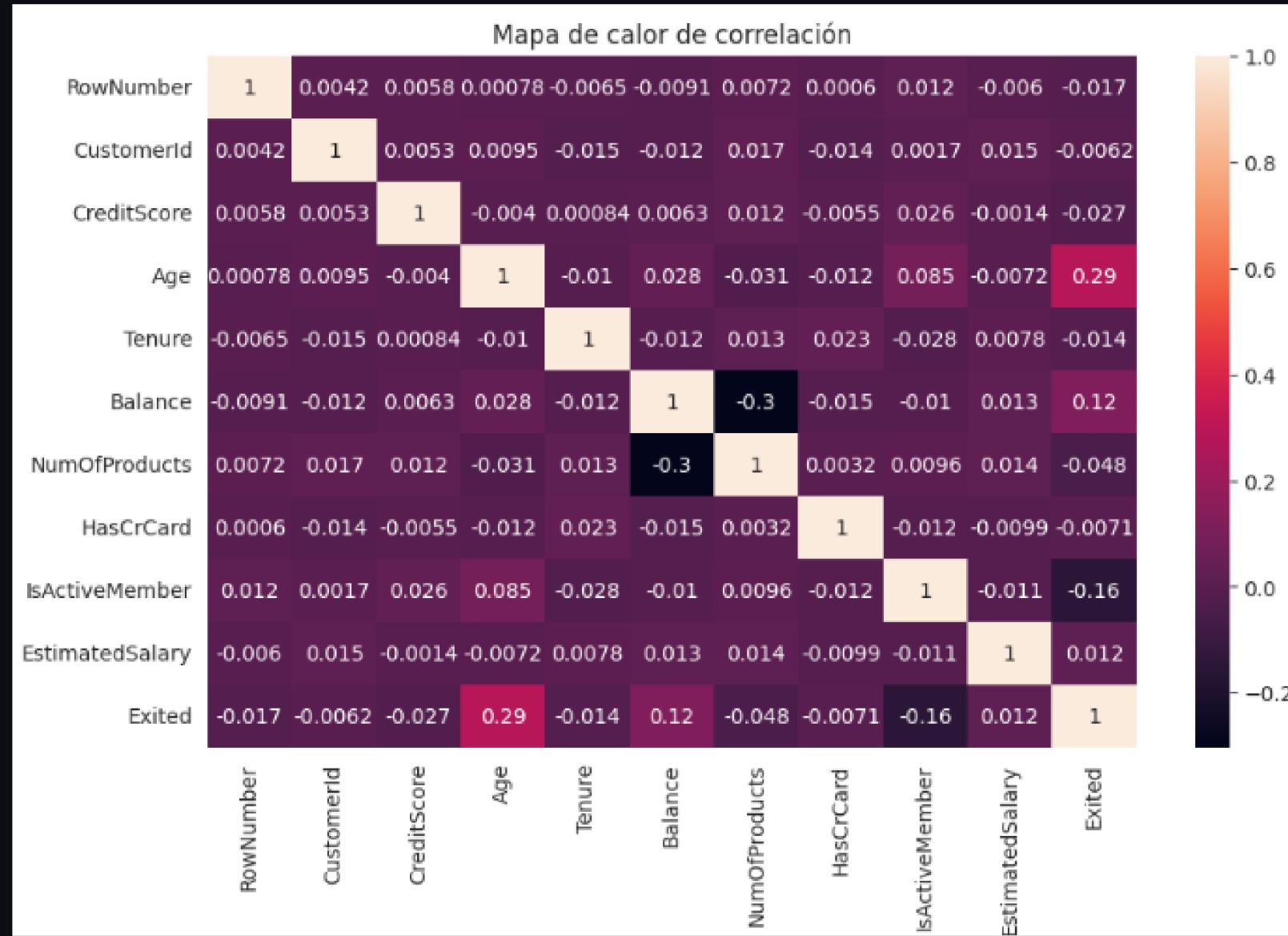


4



Gender	Country	Age
Female	France	18
		19
		20
		21
		22
		...
Male	Spain	76
		77
		78
		80
		88
		...

Gráfico de correlación 4



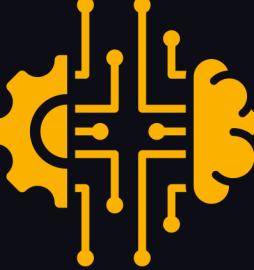
Podemos observar que existe correlación entre las variables **Row Number** y **Exited**. Lo que nos ayudará a detectar motivos por los que los clientes abandonan la entidad.

5

Algorítmos de Machine Learning



Machine Learning

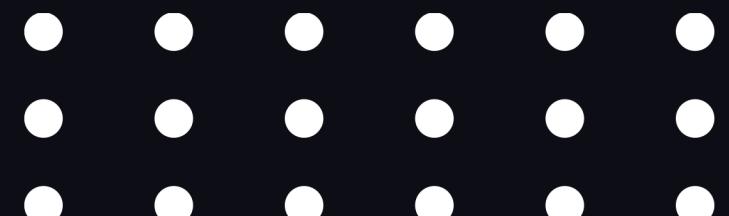
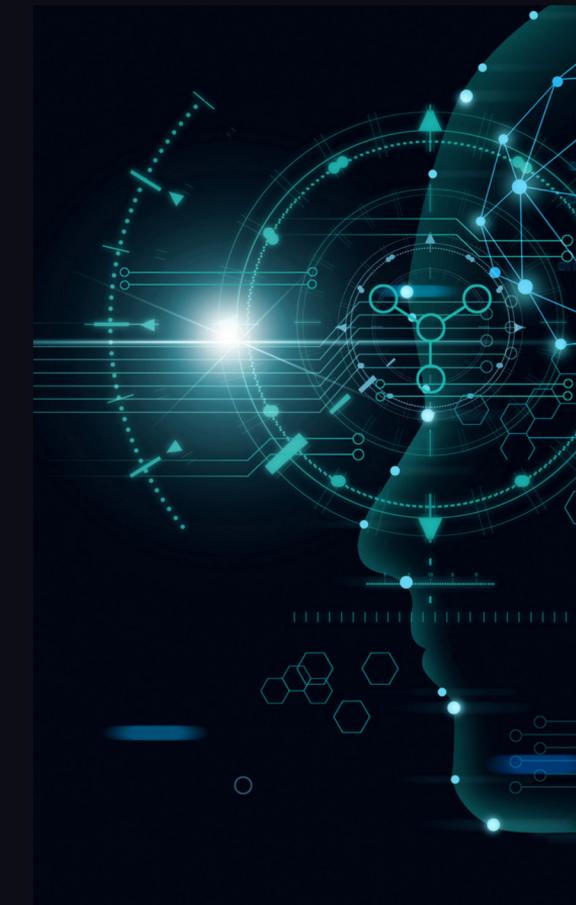


5

Aplicamos distintos modelos, métodos y métricas

C clientes salientes / clientes estables (Variable 'Exited')

- Random Forest 0.865
- Validación Cruzada 0.8725
- F1 Score 0.8005 de ajuste
- Recall Score 0.5996860282574569
- Presición Score 0.4860050890585242
- Validación Cruzada 0.7827868852459017



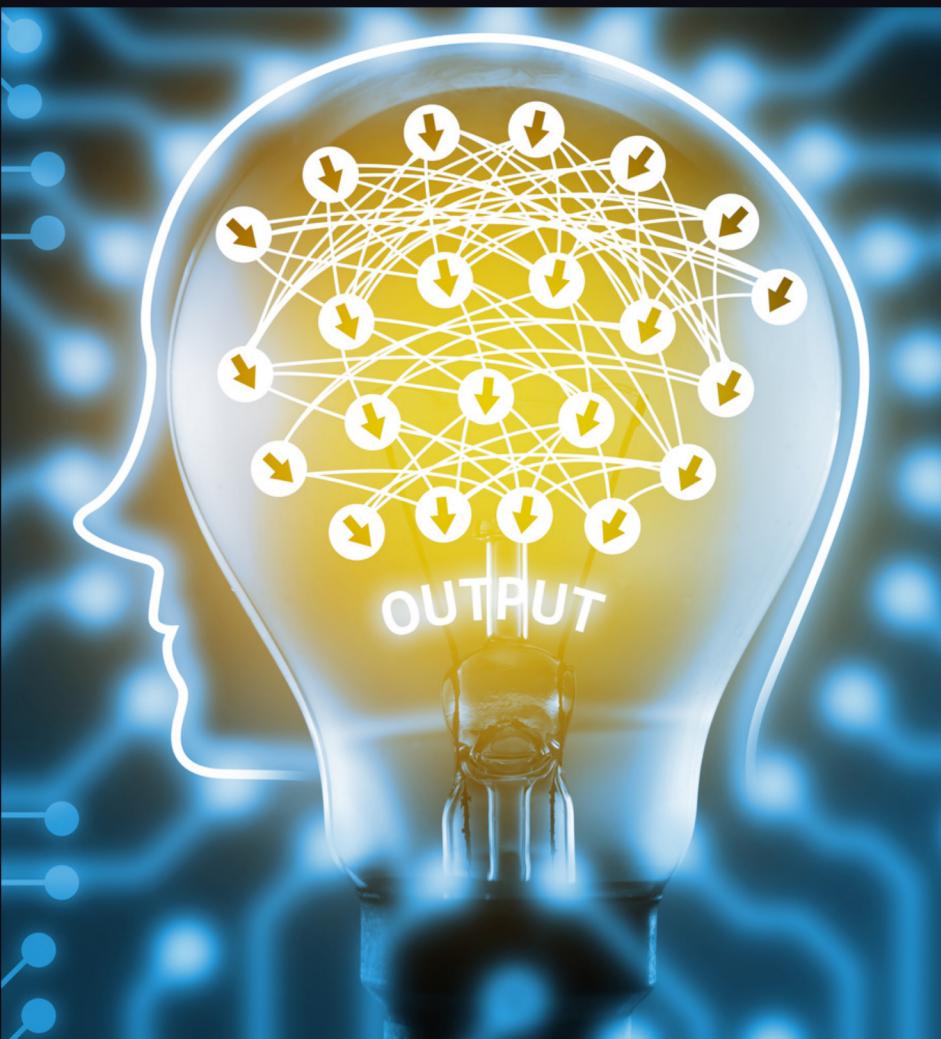
Resumen ML



5

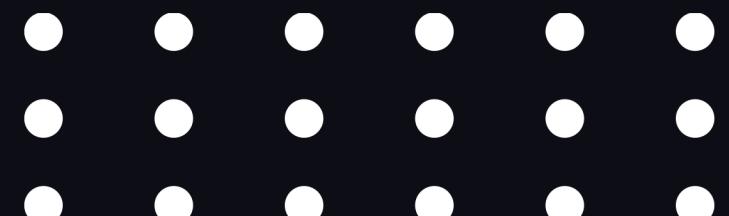
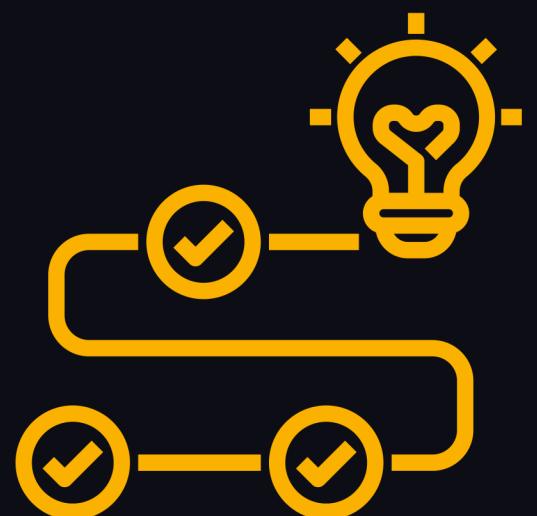
La mayor precisión obtenida en el conjunto de prueba fue de 87.25%, lo que significa que el 87.25% de las predicciones fueron correctas en comparación con las etiquetas reales en el conjunto de prueba.

En resumen, la validación cruzada nos permitió obtener una evaluación más precisa y confiable del rendimiento de nuestro modelo, lo que facilita la toma de decisiones informadas sobre su aplicabilidad y generalización a nuevos datos.



6

Conclusiones



Conclusiones generales 6

El análisis de datos bancarios revela historias fascinantes sobre la diversidad y las necesidades financieras de la sociedad. La exploración por país, segmento etario y género proporciona valiosos patrones e identifica oportunidades para mejorar la inclusión financiera. Además, el enfoque comercial aborda problemas clave como la tasa de abandono, la satisfacción del cliente, la retención y la mejora de productos y servicios. El contexto bancario y las hipótesis guían el análisis, destacando la importancia de la toma de decisiones estratégicas. El modelo de Random Forest, validado con precisión del 87.25%, demuestra ser efectivo para generalizar y respaldar decisiones informadas en la industria bancaria.

¡Muchas
Gracias!

Norberto Stelmach