

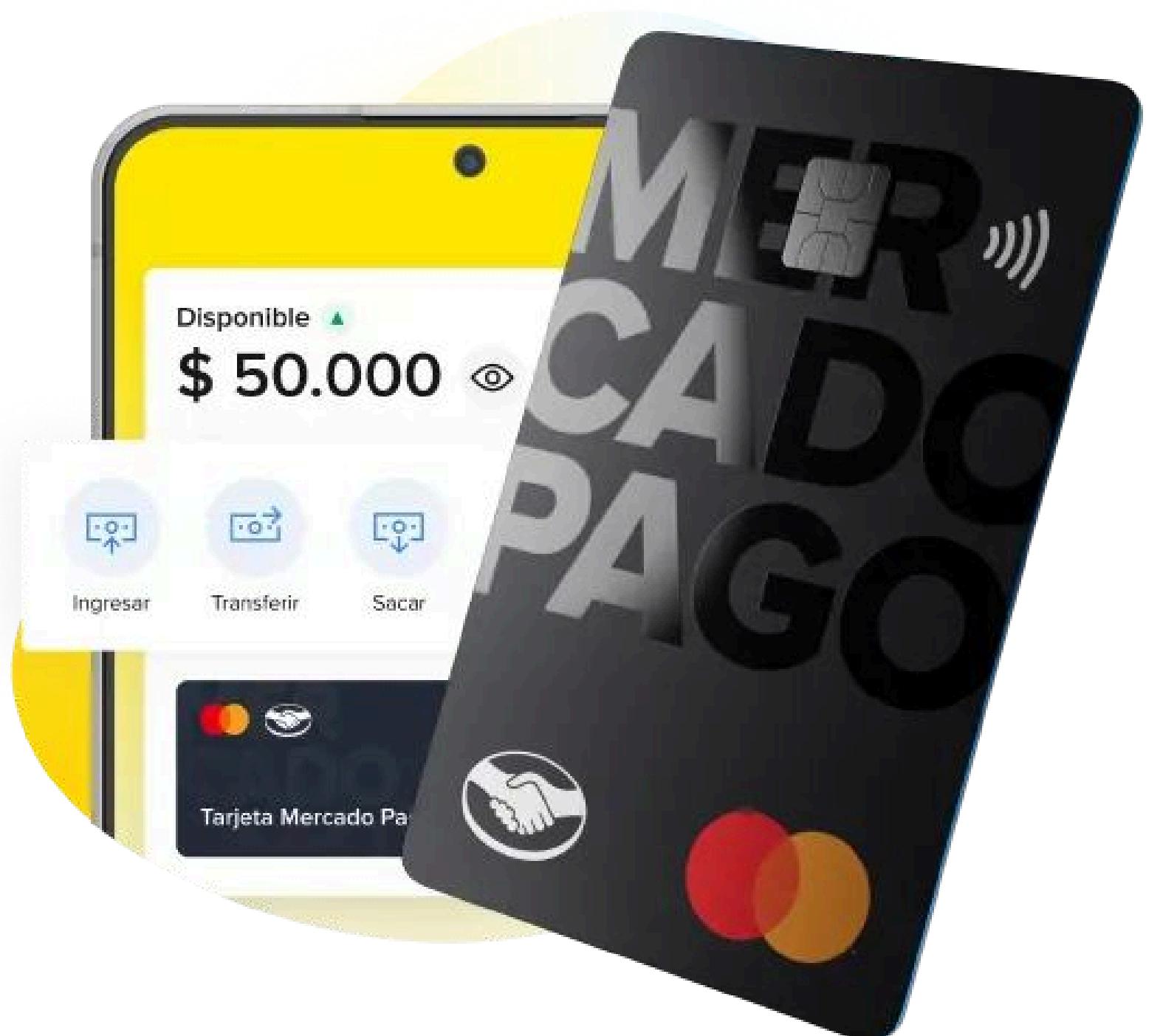
IGNACIO MUÑOZ GOMEÑUKA

# Churn Analysis and Prediction

Proyecto exploratorio y modelo predictivo para detectar y prevenir la pérdida de clientes, este análisis fue realizado con mock data basado en un dataset fintech con información acerca de clientes ficticios.

## Objetivo

- Analizar los factores asociados al abandono de clientes
- Identificar patrones de comportamiento relevantes para retención
- Construir un modelo predictivo de churn
- Aplicar técnicas de análisis de datos alineadas a un entorno real de fintech (Mercado Pago)



# Índice

1. Objetivo del Proyecto
2. Descripción del Dataset
3. Tasa General de Churn
4. Análisis Exploratorio
  - Género, Edad, Balance
  - Número de Productos, Satisfacción, Quejas
5. Construcción del Modelo Predictivo
  - Logistic Regression
  - Random Forest
6. Evaluación y Métricas
7. Importancia de Variables
8. Conclusiones

# Descripción general del dataset

10.000 clientes simulados de una entidad financiera

Variables disponibles:

- Edad, saldo, productos contratados, geografía, género
- Score de satisfacción, tarjeta de crédito, quejas
- Variable target: Exited (1 = churn)

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

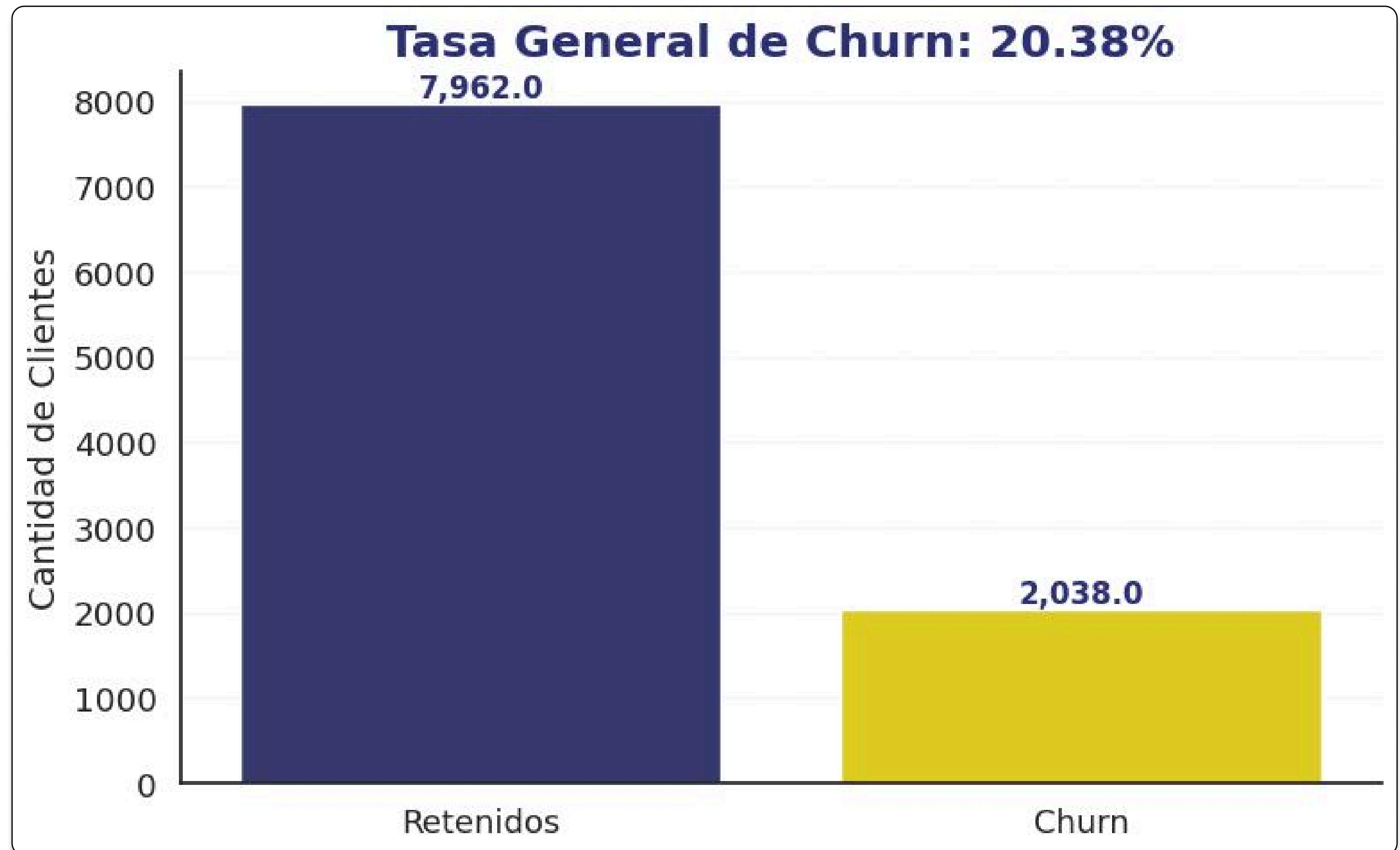
# Tasa de Churn

¿Cuál es la magnitud del problema de abandono de clientes?

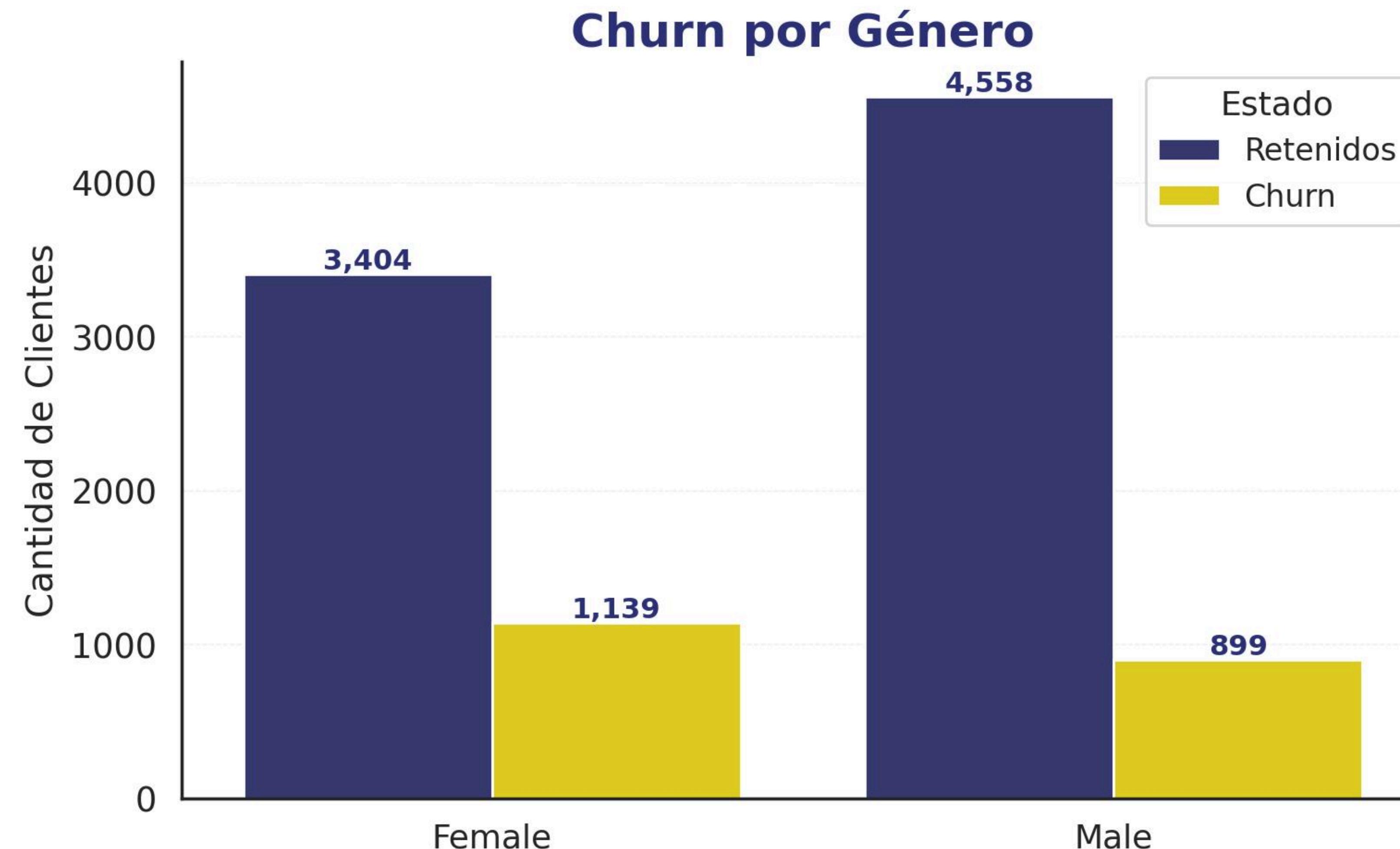
El análisis parte de una base de 10.000 clientes, donde identificamos una tasa de churn del 20,38%.

Este valor representa un nivel de abandono significativo que justifica el desarrollo de estrategias predictivas y preventivas.

Visualmente, observamos una distribución equilibrada entre clientes retenidos y perdidos, lo cual permite entrenar modelos de clasificación sin sesgo severo.



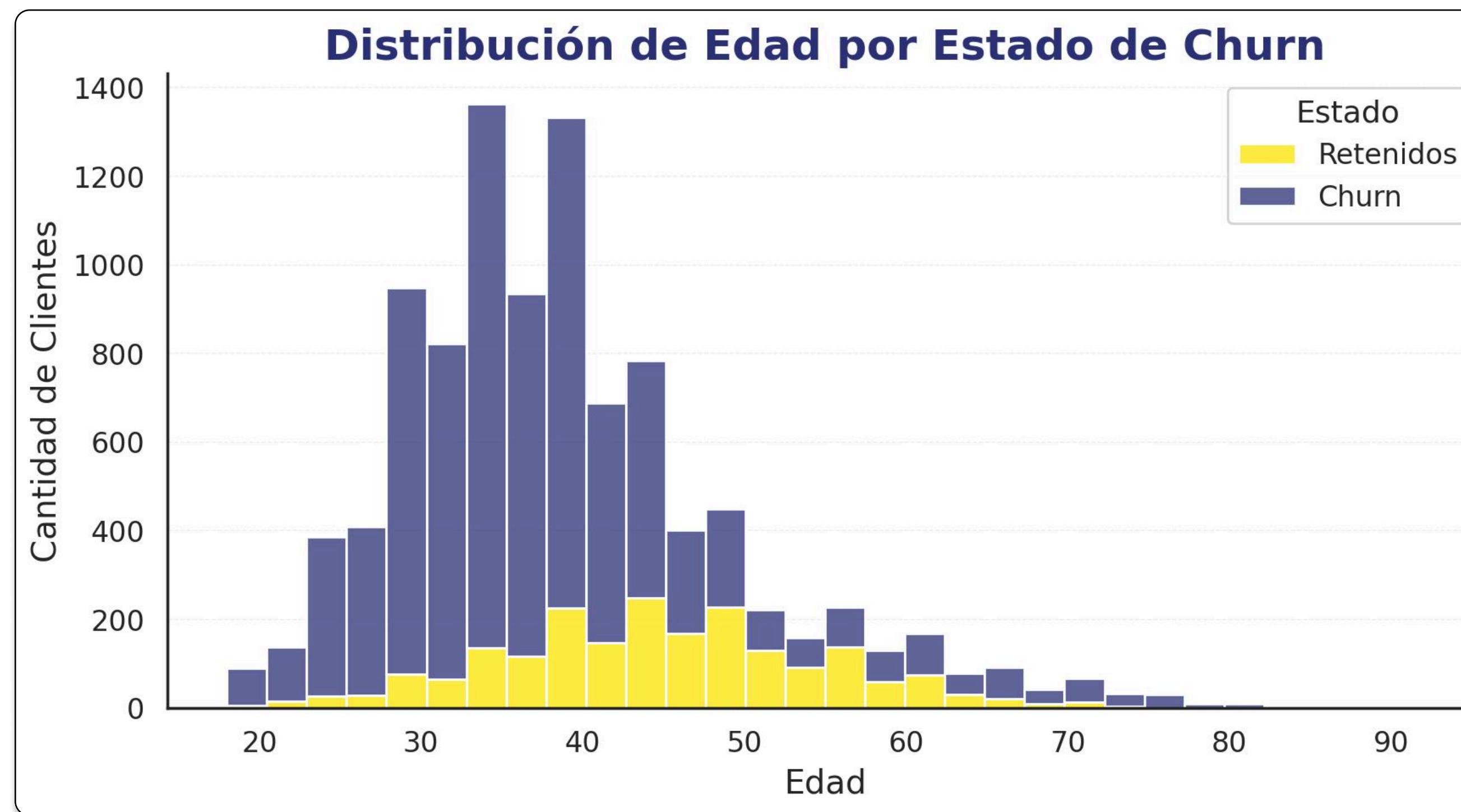
# Churn por género



# Distribución de edad por estado de churn

¿Que rango etario presenta la mayor tasa de churn?

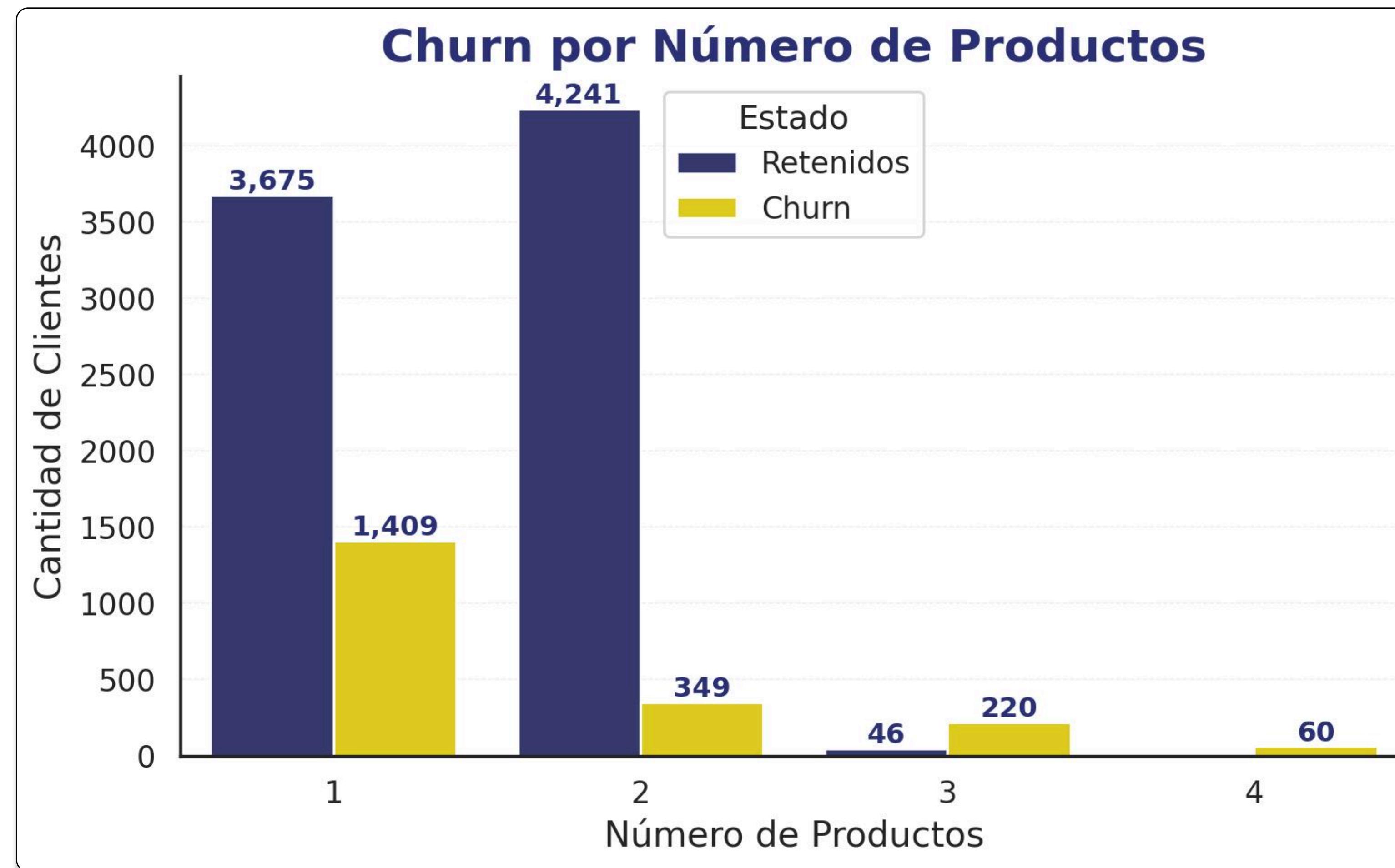
Edad: los clientes mayores de 50 años son los que más abandonan, con tasas superiores al 50%.



Grupo Etario	Tasa de Churn
0	0.075203
1	0.121096
2	0.339655
3	0.562108
4	0.247845

# Churn por número de productos

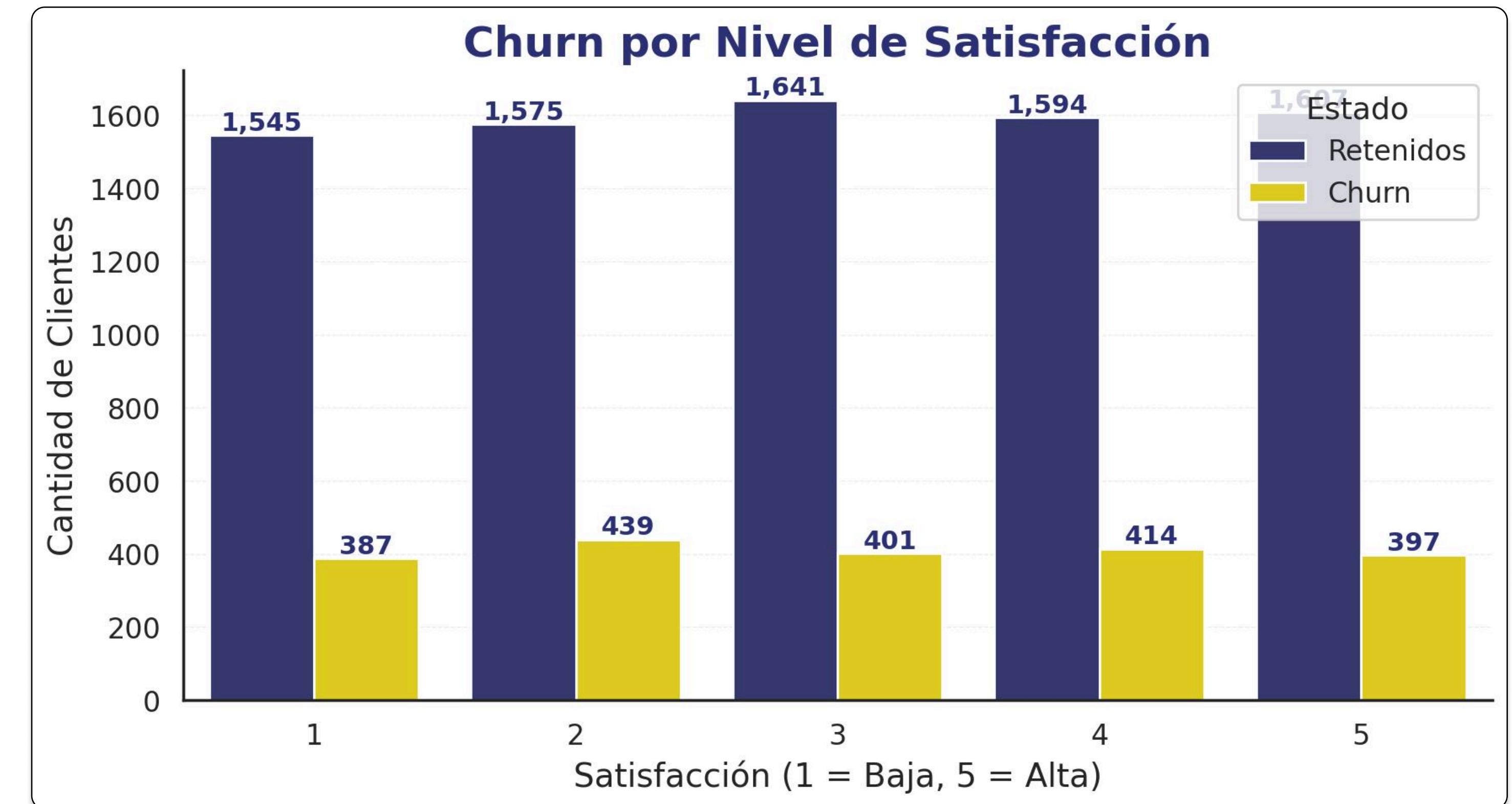
Productos contratados: los clientes con 2 productos tienen alta retención, mientras que aquellos con 3 o más productos tienen tasas de churn mayores al 80%.



# Churn basado en satisfacción del cliente

Quejas: el 99,5% de los clientes que presentaron una queja abandonaron, marcando una correlación casi determinante.

Nivel de Satisfacción	Tasa de Churn
0	1 0.200311
1	2 0.217974
2	3 0.196376
3	4 0.206175
4	5 0.198104



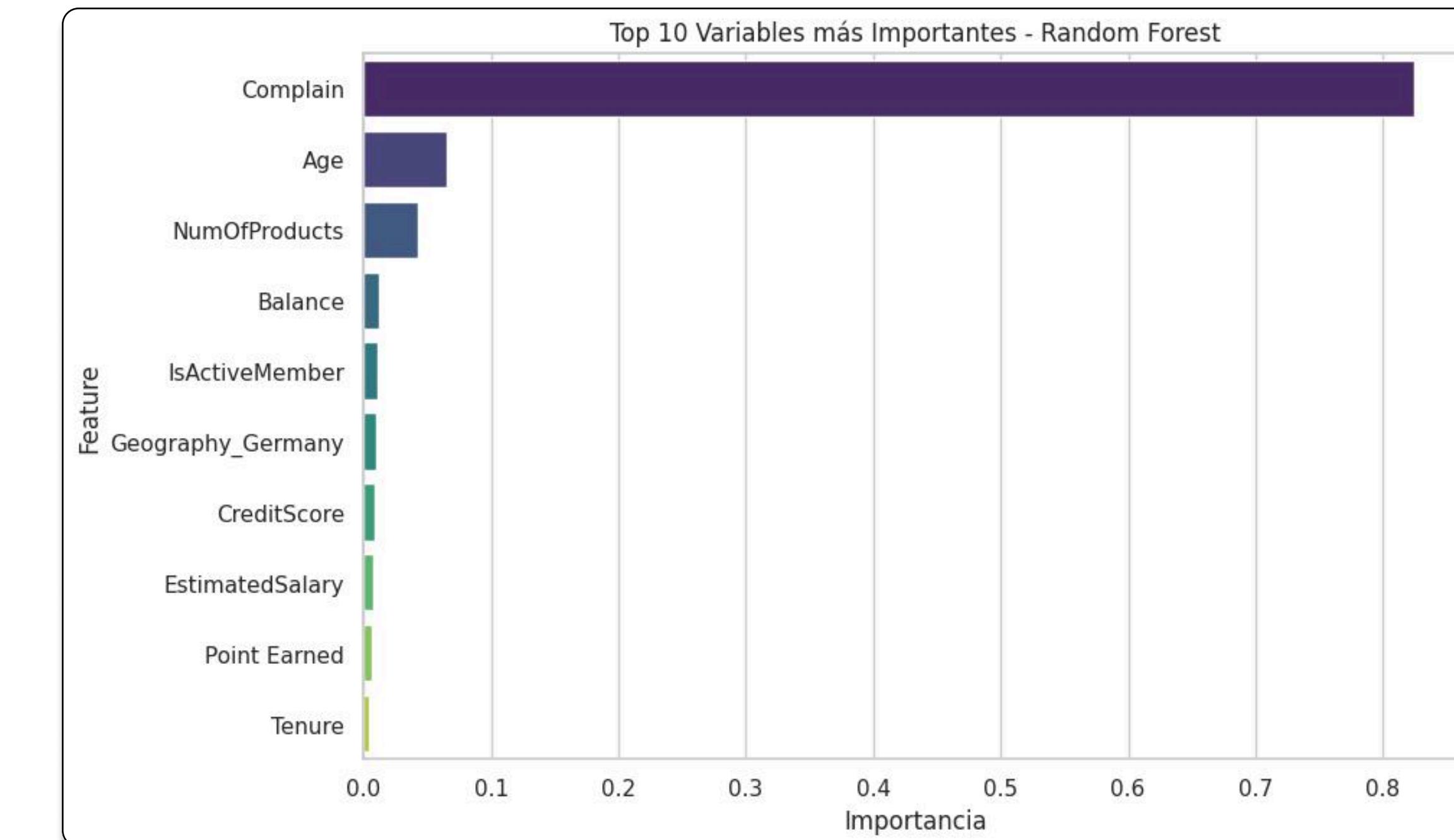
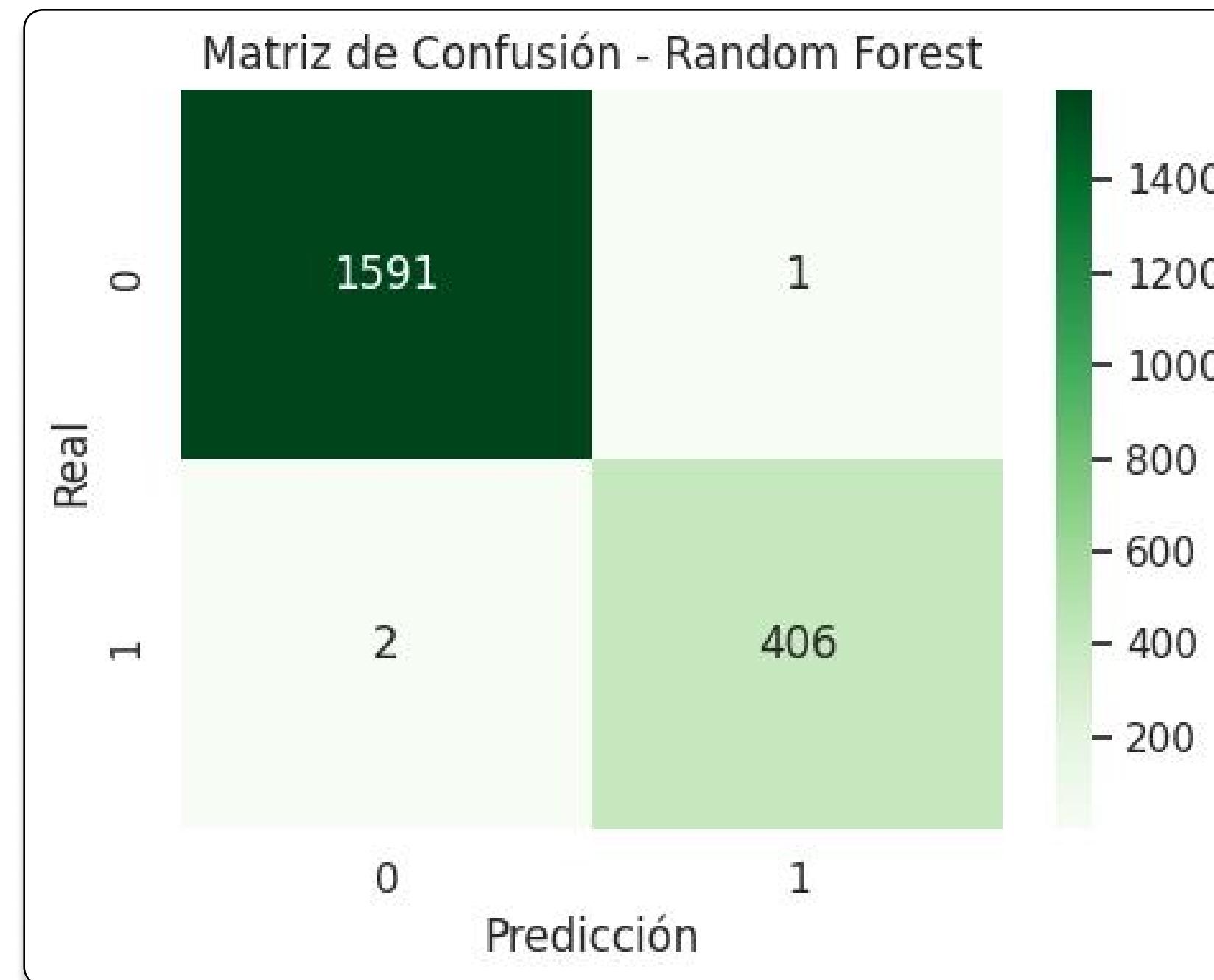
# Construcción del Modelo Predictivo

¿Qué variables influyen más en la predicción del churn?

El gráfico de importancia muestra claramente que el modelo basa sus decisiones principalmente en la variable Complain, que representa si el cliente presentó una queja.

Top variables más importantes:

1. Complain: representa más del 80% del peso del modelo
2. Age: segunda variable en importancia
3. NumOfProducts: vinculado con sobreventa y frustración



# ¿Porque es importante para las fintech?

## ¿Qué aprendizajes accionables surgen del análisis?

- El análisis y modelado permiten identificar con antelación los clientes en riesgo de abandonar.
- Existen perfiles específicos altamente sensibles al churn, como usuarios con quejas o sobreexposición a productos.
- Implementar campañas de retención basadas en estos insights puede tener un impacto directo en el LTV y en la rentabilidad de la vertical.
- La metodología puede escalarse y ajustarse dinámicamente a otras unidades de negocio (cuentas, QR, tarjetas).

Thank you!

mercado  
pago

