



CHURN



-
- Ana María Bolaños
 - Wendy Calero
 - Soledad Suarez
 - Evelyn Nina

El problema



15% y 30%

**Las empresas de telecomunicaciones globales
enfrentaban tasas de CHURN**



**Harvard
Business
School**

5 a 25 veces

**Puede costar adquirir nuevos clientes que
retener a los ya existentes**



Modelo de predicción de churn

Identificar a los clientes con mayor riesgo de deserción

Permitiendo a la empresa tomar acciones proactivas para retenerlos. Esto se traduce en un aumento de la fidelización de clientes



Soluciones existentes



Programa de Fidelización de Clientes

Utiliza big data y ML para ofrecer recompensas y beneficios personalizados a los clientes



Modelos Predictivos de Churn

Analizan datos en tiempo real y predicen quienes están en riesgo de abandonar la empresa



Patrones de uso y quejas

Empleo modelos de análisis predictivo para reducir el *churn* entre sus clientes

Contexto de nuestros datos



CLIENTES

7043



UNITED STATES

CALIFORNIA



TRIMESTRE

3



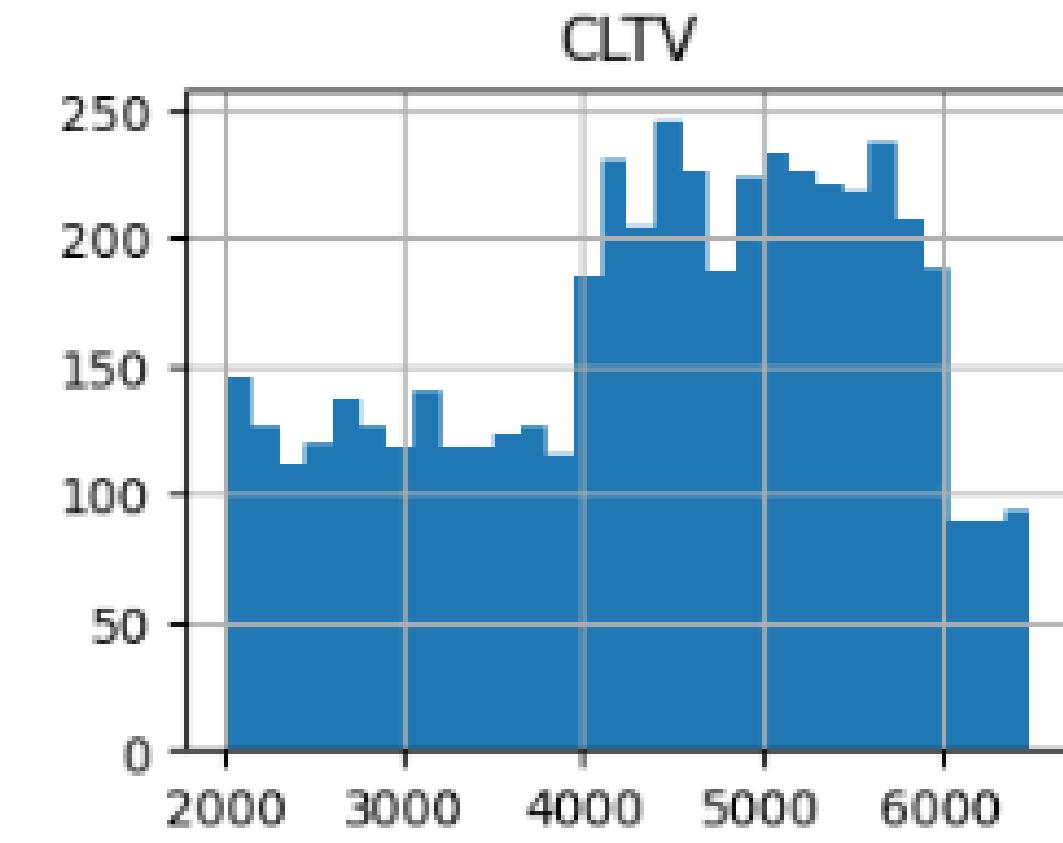
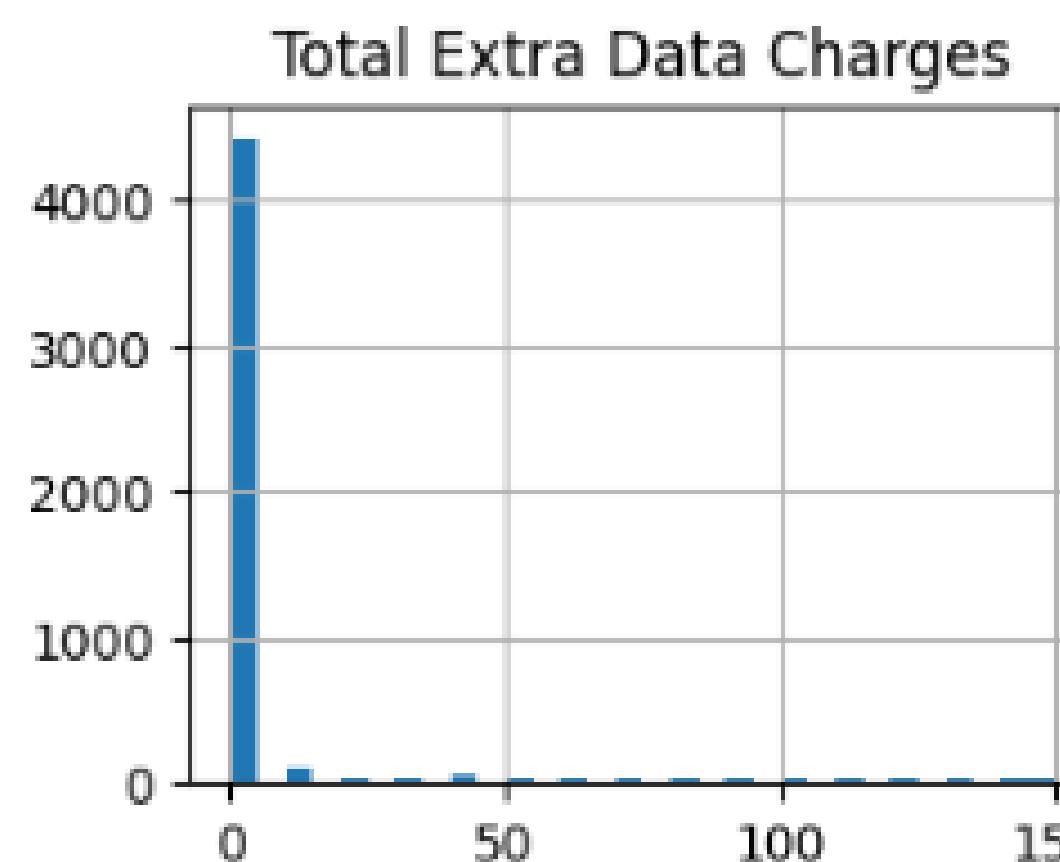
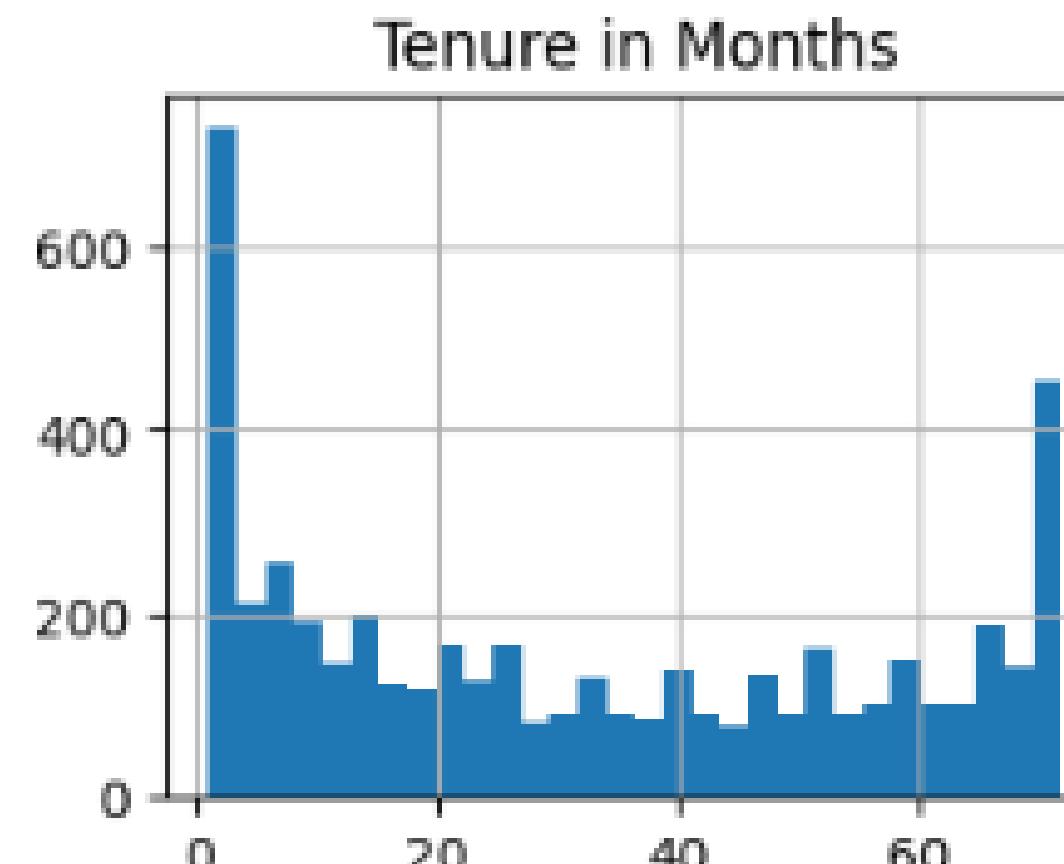
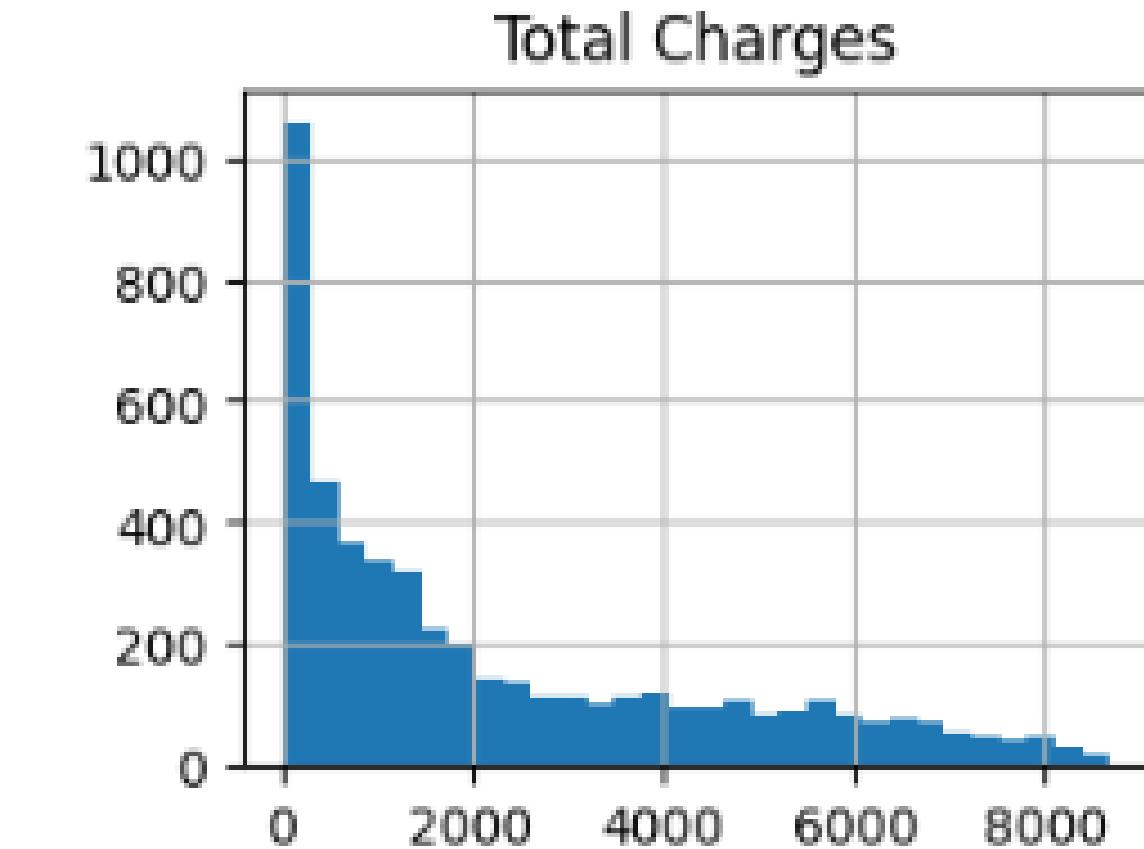
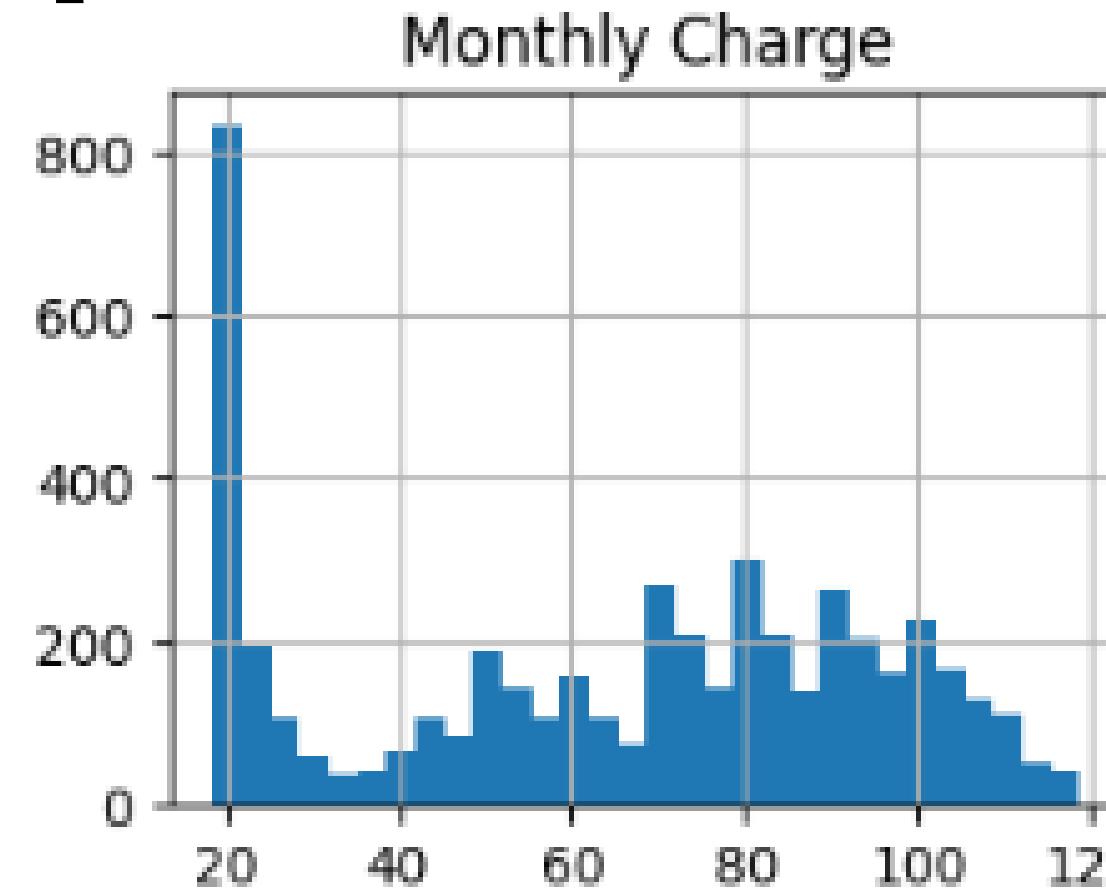
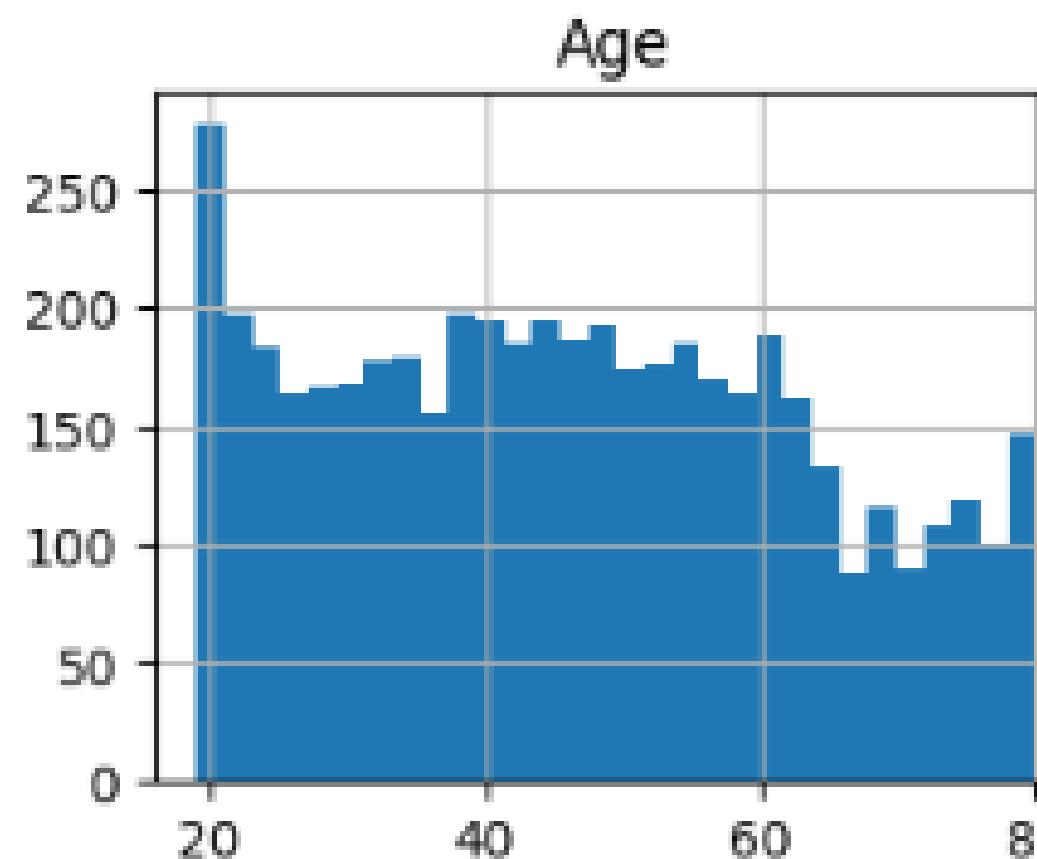
VALORES FALTANTES

Offer e Internet Type



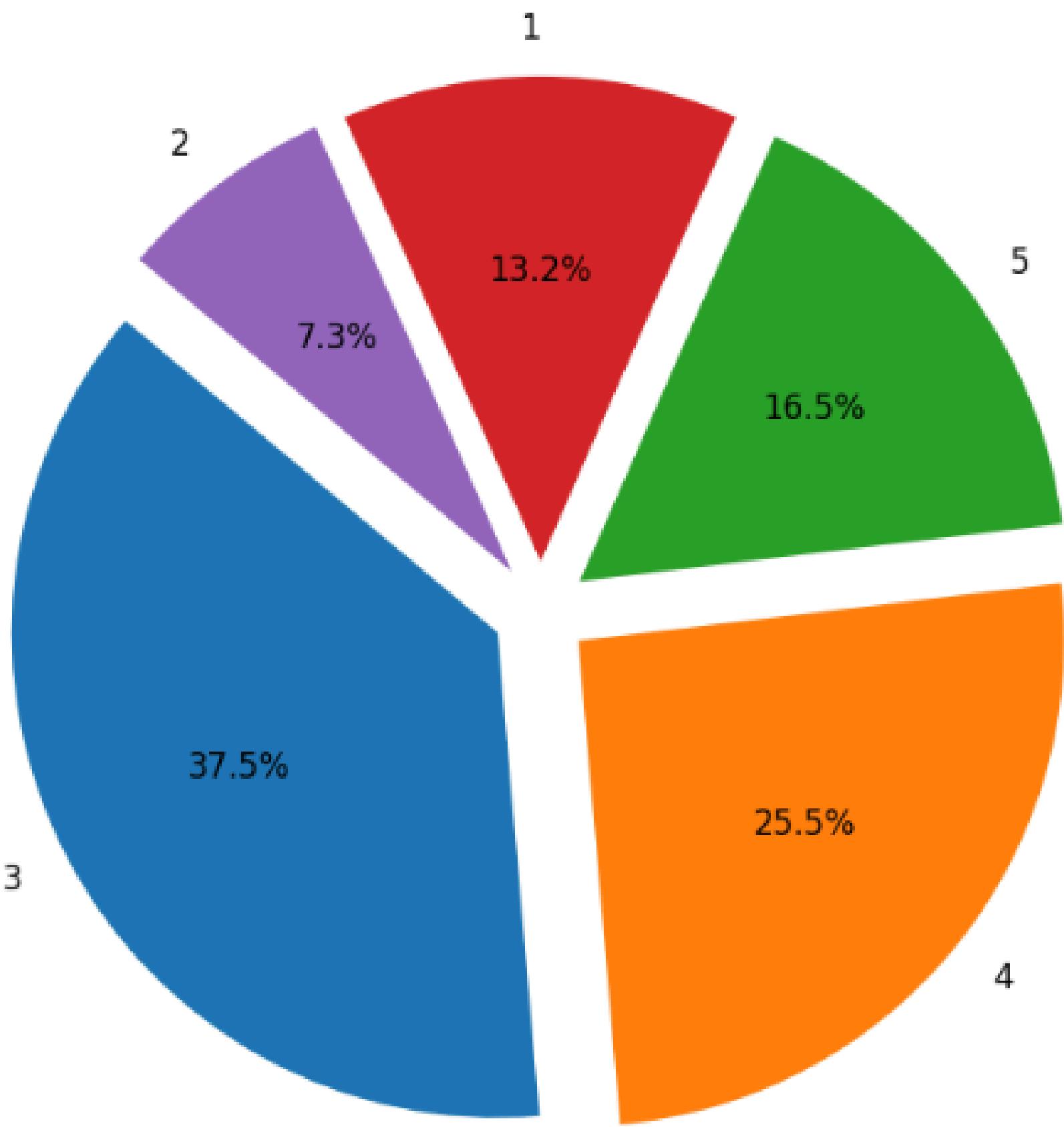
Metodología

Análisis exploratorio

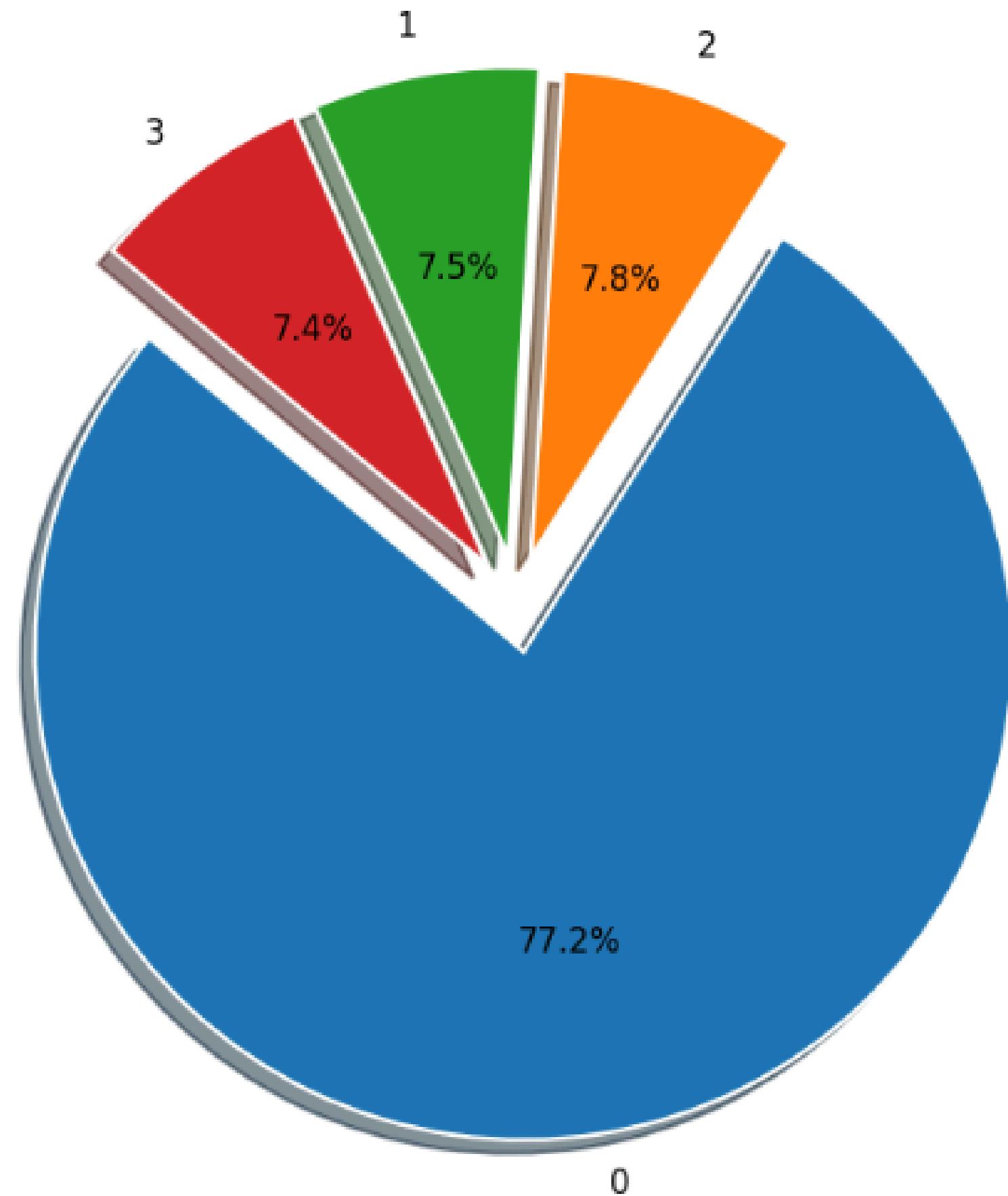


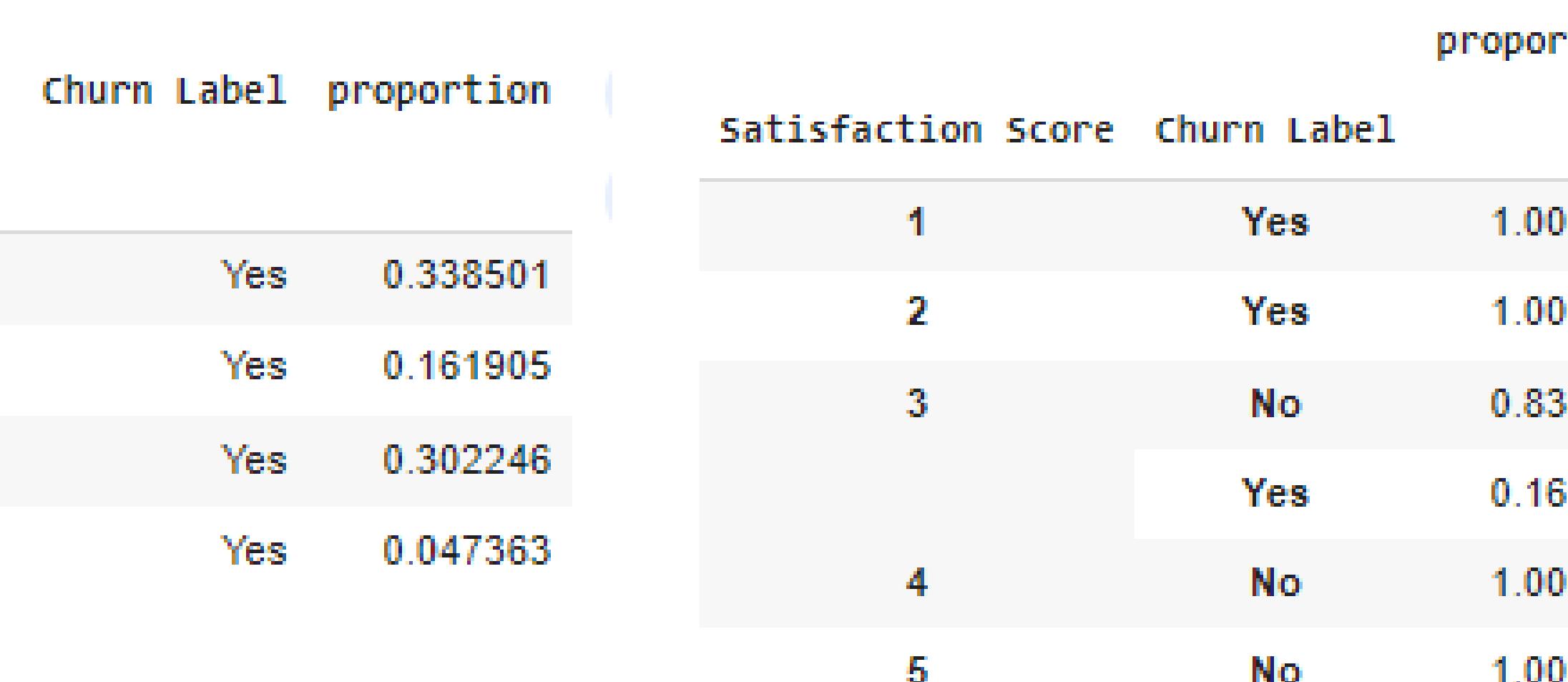
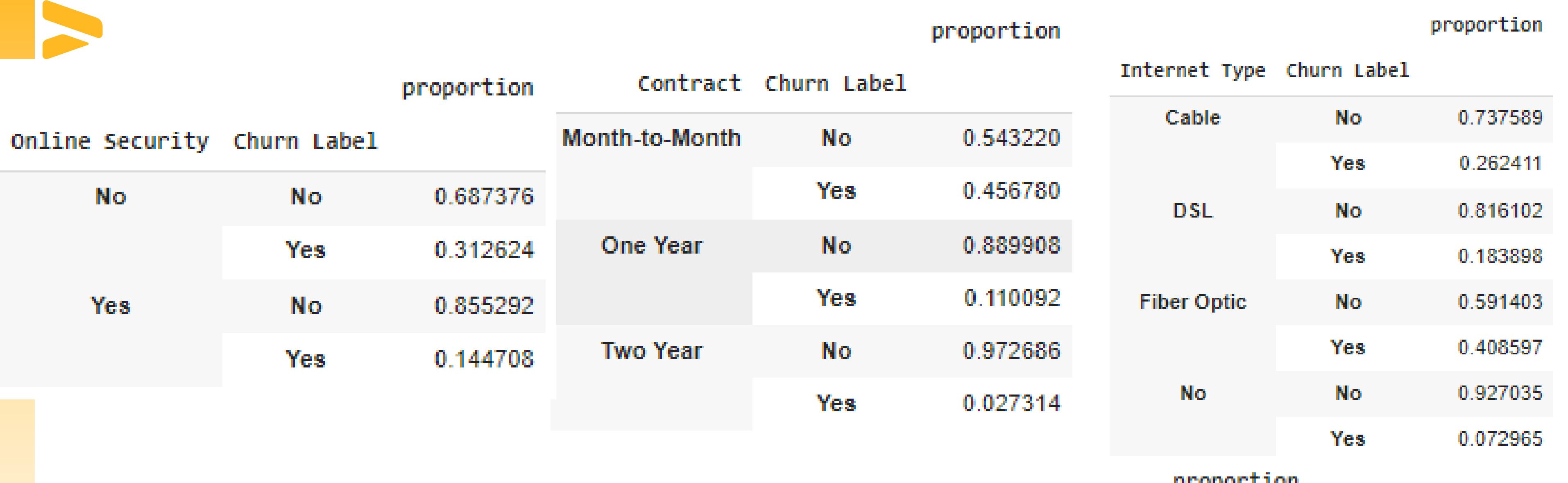


Distribución de Satisfaction Scores

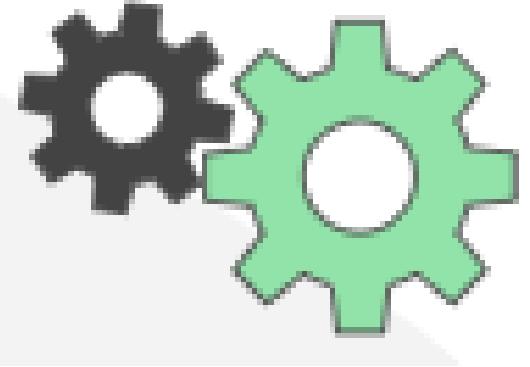


Distribución de Número de Dependientes





Procesamiento de datos



TRAIN DATA

Los datos que se usará para entrenar el modelo y es probable que el modelo tenga suficiente información para aprender de manera efectiva.

70%

TEST DATA

Independientes para evaluar el rendimiento final del modelo

VALIDATION DATA

Usados para optimizar el modelo y ajustar hiperparámetros

15%

Feature Engineering

■ Extra Charges Ratio

(Total Extra Data Charges)/(Total Charges)

■ Has Streaming Service

'Streaming TV', 'Streaming Movies',
'Streaming Music'

■ Región California

Norte de California: Si la latitud es mayor a 38.

Centro de California: Si la latitud está entre 36 y 38 (inclusive).

Sur de California: Si la latitud es menor a 36.

Codificación de features categóricos

Valores excluyentes

Population

Zip Code

Customer ID

Se utilizo:

➤ OrdinalEncoder

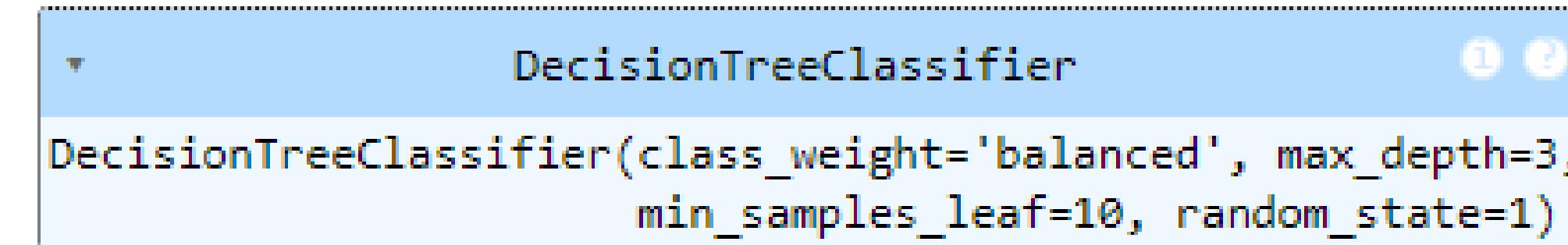
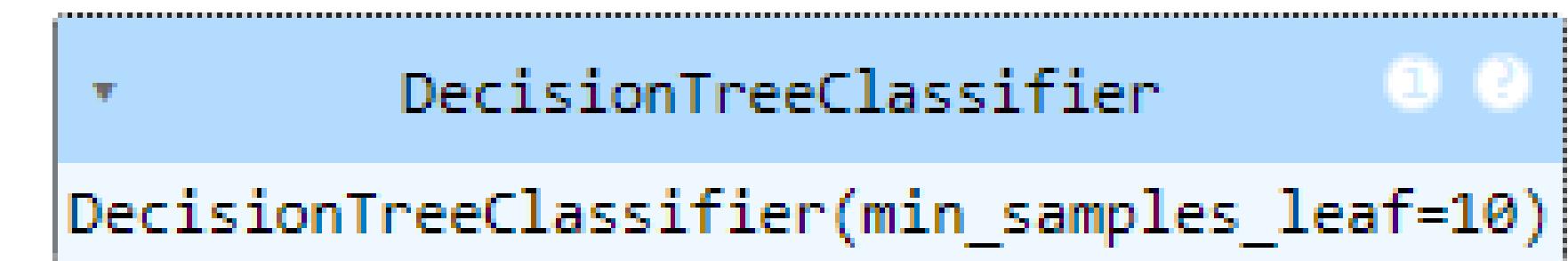
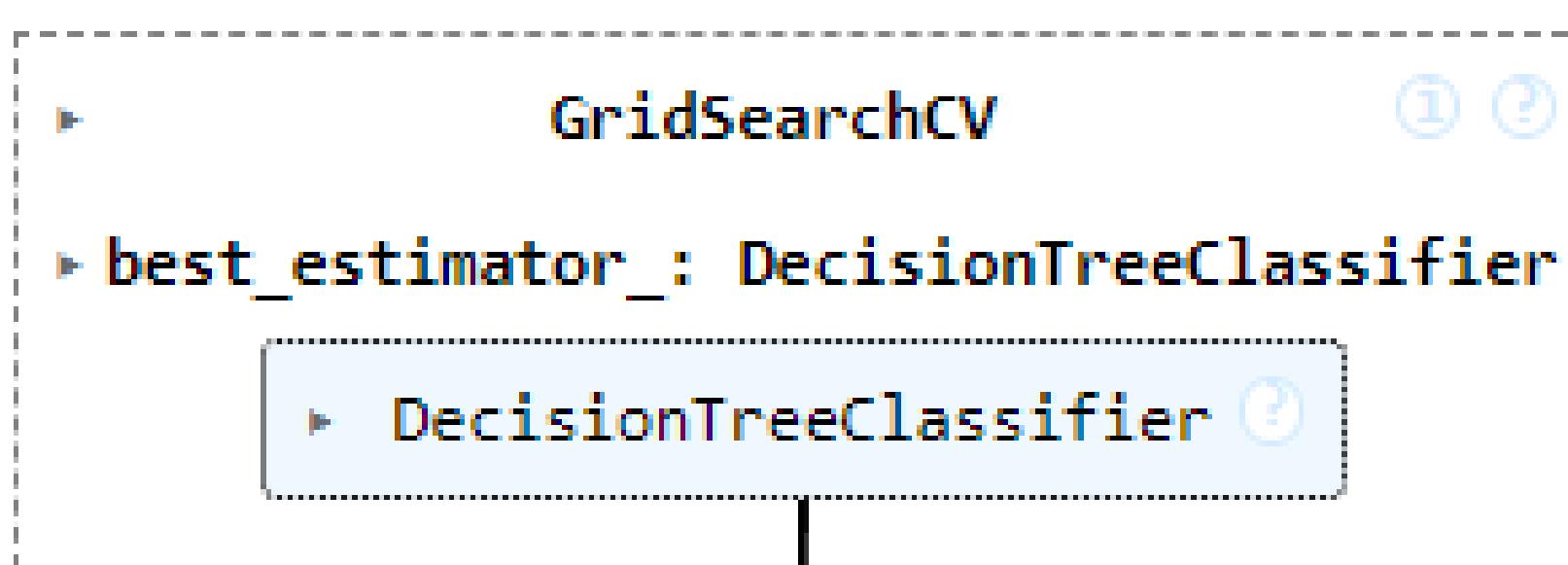
➤ Funciones para iterar y acelerar el proceso de reemplazo en valores booleanos

#	Column	Non-Null Count	Dtype
0	Gender	4930 non-null	int64
1	Age	4930 non-null	int64
2	Under 30	4930 non-null	int64
3	Senior Citizen	4930 non-null	int64
4	Married	4930 non-null	int64
5	Dependents	4930 non-null	int64
6	Number of Dependents	4930 non-null	int64
7	Latitude	4930 non-null	float64
8	Referred a Friend	4930 non-null	int64
9	Number of Referrals	4930 non-null	int64
10	Tenure in Months	4930 non-null	int64
11	Offer	4930 non-null	int64
12	Phone Service	4930 non-null	int64
13	Avg Monthly Long Distance Charges	4930 non-null	float64
14	Multiple Lines	4930 non-null	int64
15	Internet Service	4930 non-null	int64
16	Internet Type	4930 non-null	int64
17	Avg Monthly GB Download	4930 non-null	int64
18	Online Security	4930 non-null	int64
19	Online Backup	4930 non-null	int64
20	Device Protection Plan	4930 non-null	int64
21	Premium Tech Support	4930 non-null	int64
22	Streaming TV	4930 non-null	int64
23	Streaming Movies	4930 non-null	int64
24	Streaming Music	4930 non-null	int64
25	Unlimited Data	4930 non-null	int64
26	Contract	4930 non-null	int64
27	Paperless Billing	4930 non-null	int64
28	Payment Method	4930 non-null	int64
29	Monthly Charge	4930 non-null	float64
30	Total Charges	4930 non-null	float64
31	Total Refunds	4930 non-null	float64
32	Total Extra Data Charges	4930 non-null	int64
33	Total Long Distance Charges	4930 non-null	float64
34	Total Revenue	4930 non-null	float64
35	Satisfaction Score	4930 non-null	int64
36	Churn Label	4930 non-null	int64
37	CLTV	4930 non-null	int64
38	Extra Charges Ratio	4930 non-null	float64
39	Has Streaming Service	4930 non-null	int64
40	Region_CA	4930 non-null	int64

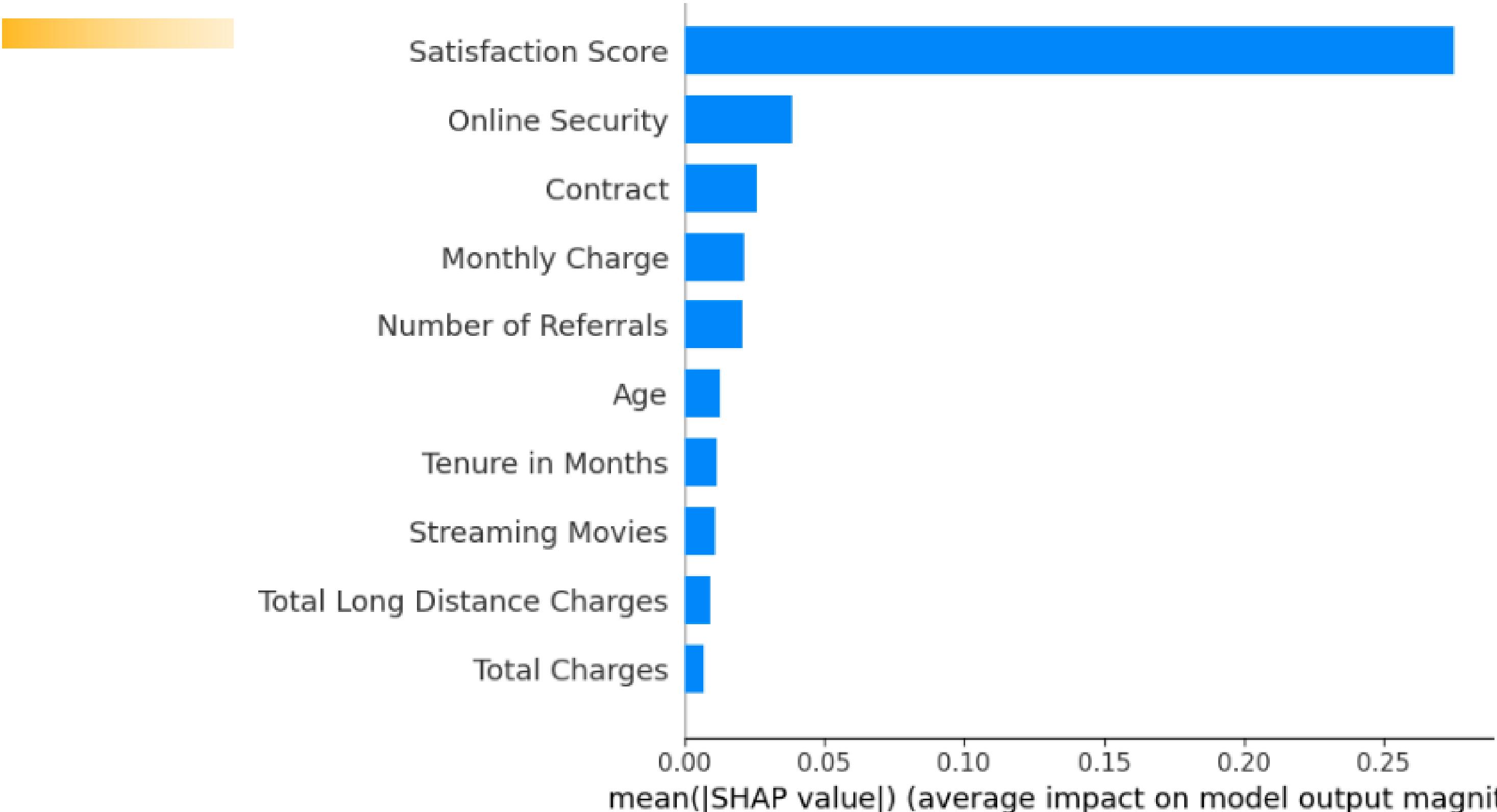
Entrenamiento y tuneo de hiperparámetros

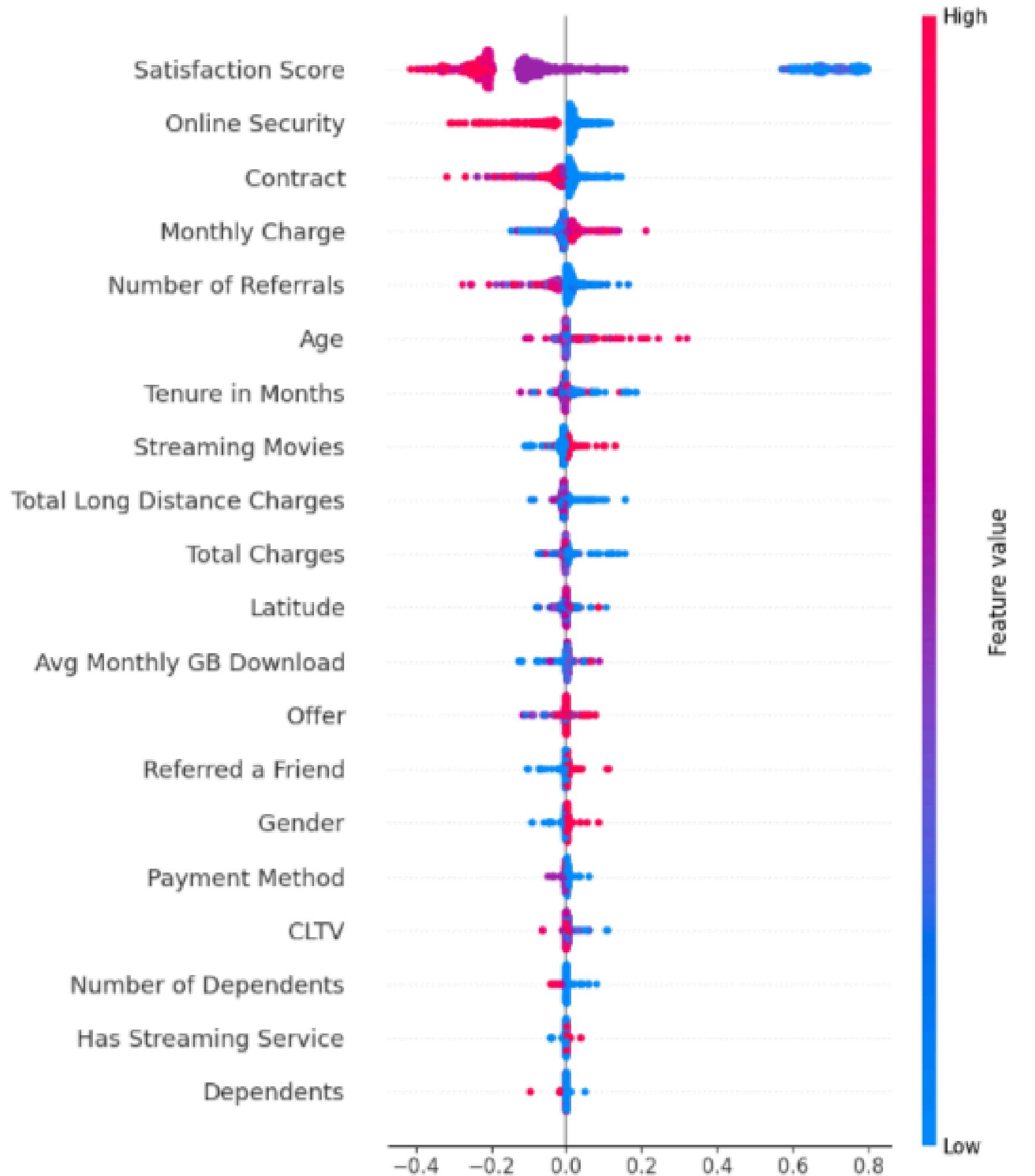
La métrica que se busco optimizar fue el f1 para nuestro primer modelo por **dos razones**:

- Balance entre precisión y sensibilidad
- Efectividad en situaciones de clases desbalanceadas



Explicabilidad del modelo







Resultados

Rendimiento

Medidas

El objetivo es minimizar los falsos positivos y asegurarse de que las predicciones negativas sean correctas

	precision	recall	f1-score	support
0	0.98	0.93	0.95	776
1	0.83	0.95	0.89	281
accuracy			0.94	1057
macro avg	0.91	0.94	0.92	1057
weighted avg	0.94	0.94	0.94	1057

specificity : 92.91%

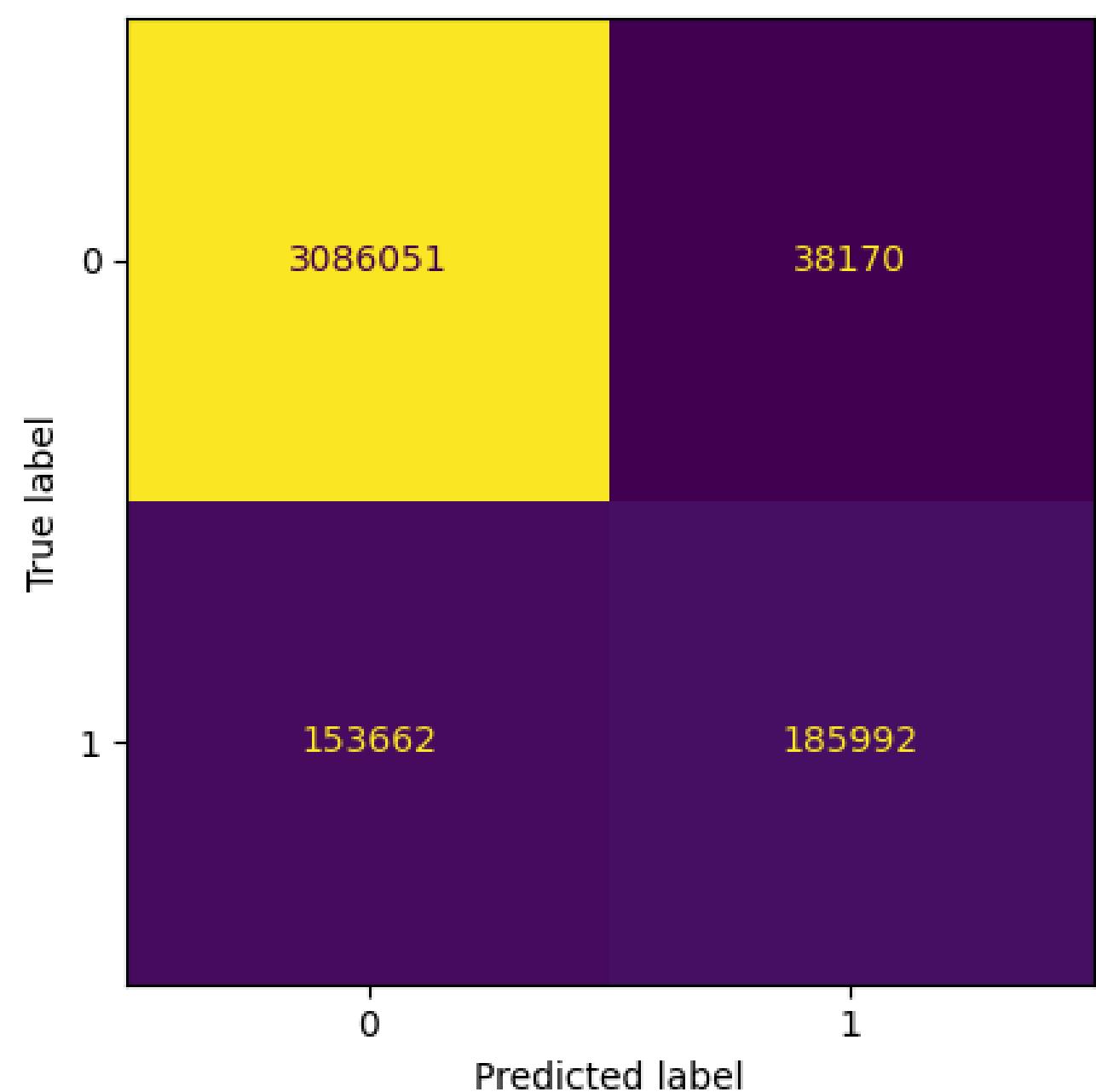
NPV (Valor Predictivo Negativo): 0.98

Evan Bailyn de First Page Sage
Una consultora especializada en marketing y
análisis de adquisición de clientes.

**En el sector de telecomunicaciones,
el costo promedio de adquisición de clientes
(CAC) es de aproximadamente \$694 USD**



Matriz de Confusión



\$224.162

True Negative (TN):

La empresa ha ahorrado o evitado costos por un valor de \$3,086,051 USD, ya que estos clientes no abandonaron y no fue necesario gastar dinero para retenerlos o adquirir nuevos clientes.

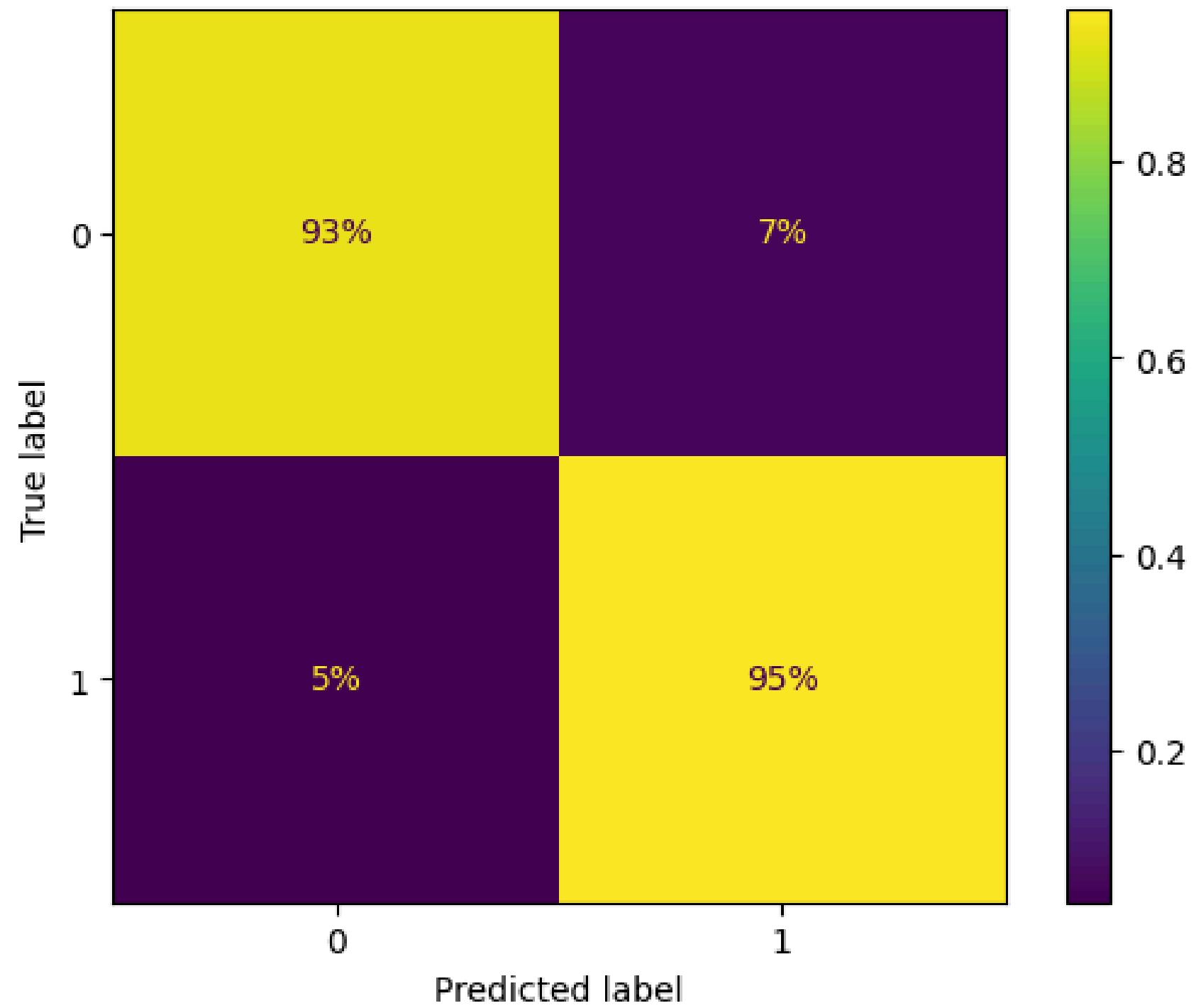
False Negative (FN):

Este valor es crítico, ya que la empresa no identificó a estos clientes como en riesgo de churn, lo que resultó en \$153,662 USD en costos de adquisición para reemplazar a los clientes que abandonaron. Reducir los falsos negativos debería ser una prioridad para disminuir estos costos.

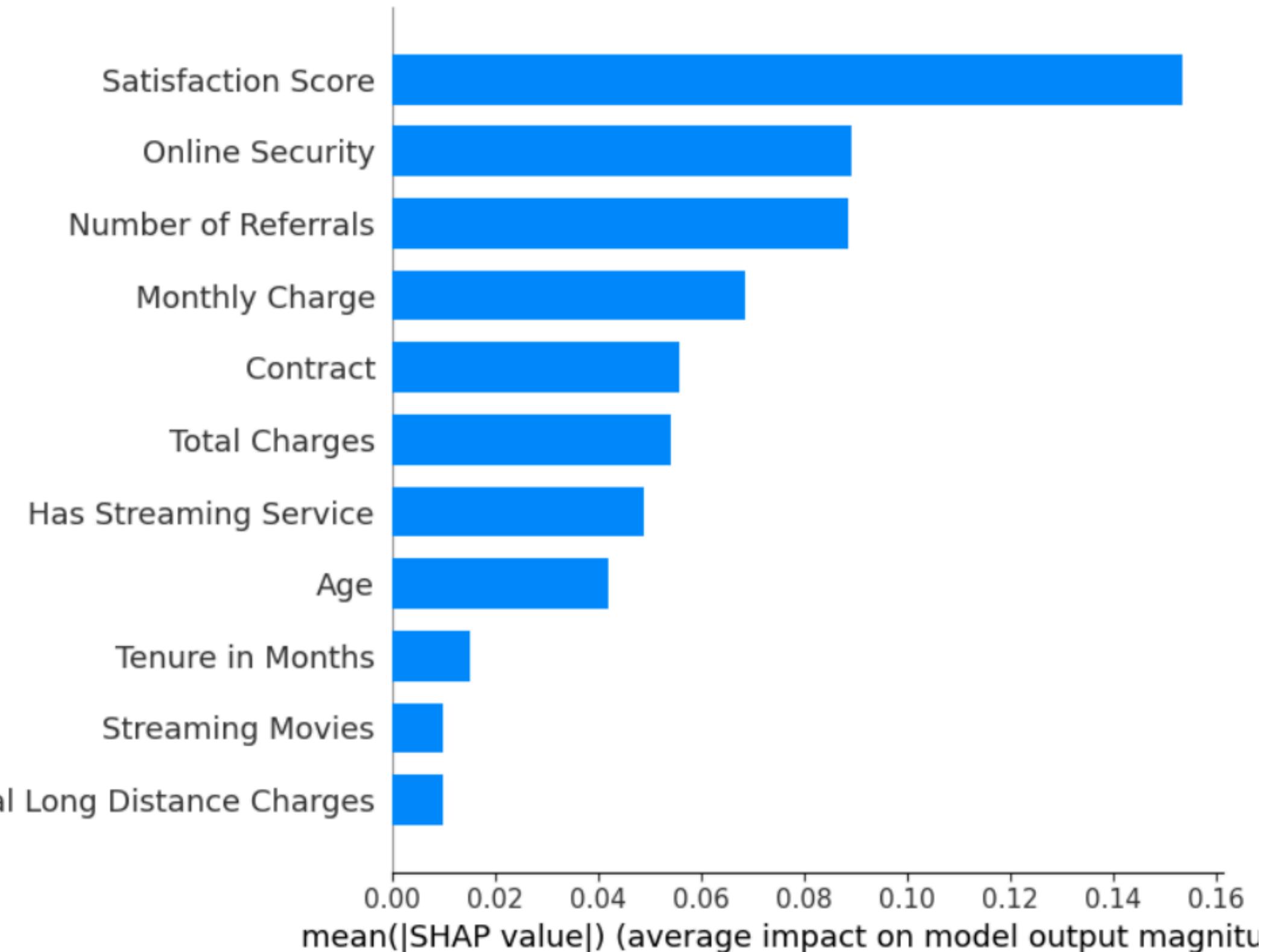
True Positive (TP):

La empresa pudo haber evitado un costo de \$185,992 USD, ya que identificó correctamente a los clientes en riesgo de churn y probablemente aplicó estrategias para retenerlos.

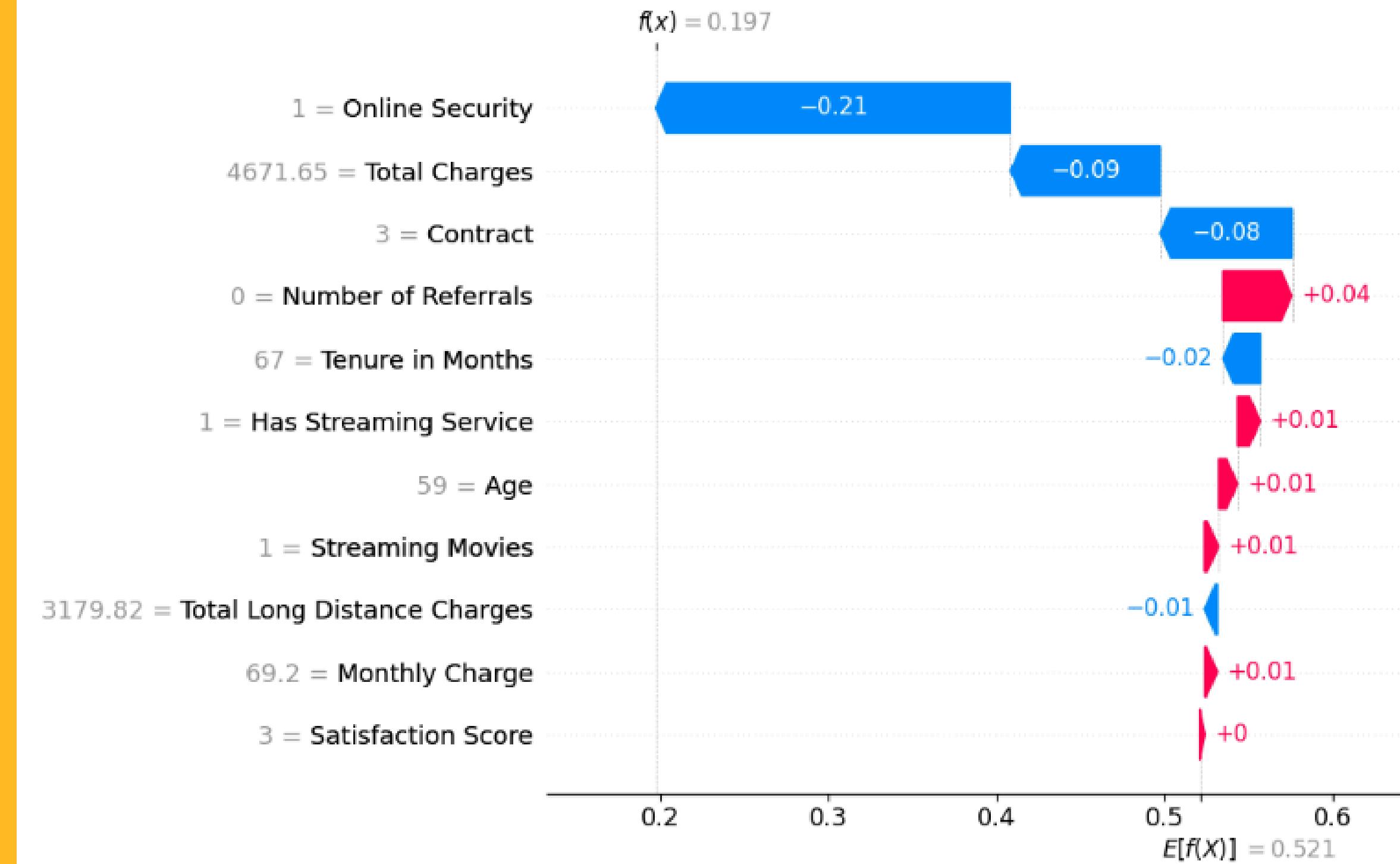
Matriz de Confusión



Test Shap



Feature importance





Implementación en el negocio

Proceso de implementación



Integración al Sistema CRM

- Llamadas de ventas
- Servicio de atención al cliente
- Correos electrónicos (contenido marketing)



Segmentación de clientes

- Creación de grupos alto, medio y bajo riesgo en función de su probabilidad de churn.
- Ofertas personalizadas, descuentos, contacto servicio al cliente con nivel alto.
- Comunicaciones de lealtad, programas de fidelización y recordatorio de beneficios a nivel medio.

•Feedback y Monitoreo

- KPI
- Satisfacción servicio al cliente
- Net Promoter Score (NPS)



Rol del modelo en la solución



■ Decisiones supervisadas por humanos

Informes revisados por el equipo AC y marketing

Acciones a tomar estarán basados en la predicción del modelo



Comunicación y supervisión

■ Automatización de acciones

Respuestas automatizadas como envío de correos electrónicos a clientes de alto riesgo.

Ofertas personalizadas, descuentos, contacto servicio al cliente



Alineación con el Rendimiento del modelo y negocio



Evaluación continua

Garantizar que las predicciones sean precisas y relevantes

- Reentrenamiento del modelo con datos nuevos
- Ajustes en las estrategias de retención

Retorno de la inversión

Cálculo del ROI de las acciones de retención basados en predicciones del modelo.

Justificación de futuras inversiones en mejoras tecnológicas y en equipo SAC

Alineación estratégica

Retención de clientes

Mejora de satisfacción al cliente

Permite un crecimiento sostenible a largo plazo



Límitaciones

Autocrítica sobre la Solución Implementada



- Falta de variables predictivas relacionadas con factores externos o temporales que podrían influir en la variable objetivo.
- Falta de datos que expliquen una temporalidad estacional o cíclicos.
- El tamaño del conjunto de datos limita la capacidad del modelo para aprender patrones generalizables.
- La interacción entre ciertas variables no fue explorada a fondo.



Conclusiones y recomendaciones

- Los contratos de mes a mes tienen una gran cantidad de clientes que abandonan el servicio a diferencia de aquellos que tienen dos años; por lo que se recomienda promocionar contratos que sean de al menos 1 año para disminuir la deserción de los clientes.
- Mantener una puntuación de satisfacción de nivel 3 es muy importante porque si esta baja hay una probabilidad de más del 90% que el cliente decida dejar la empresa.
- Servicios como la fibra óptica de la compañía tiene que ser revisada porque sus niveles de deserción son altos en comparación con las otras opciones.
- A partir de las variables más importantes que determinó el modelo, se recomienda incorporar en los planes promocionales aquellos que aporten una retención de clientes.



Future work

Mejoras del proyecto



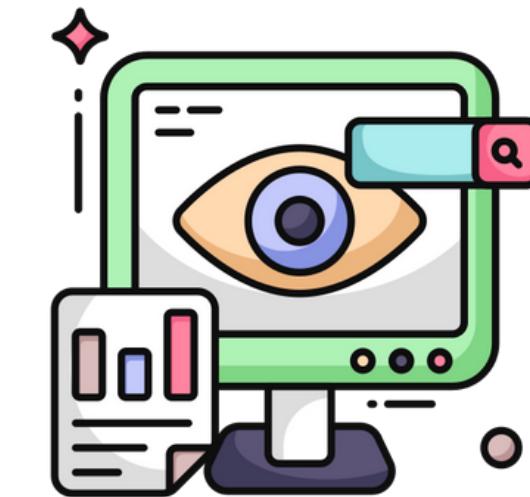
Incluir datos

Incluir datos de llamadas a servicio al cliente, promociones aplicadas y patrones de uso podría aumentar la capacidad predictiva del modelo.



Tiempo real

La predicción en tiempo real podría detectar comportamientos de churn inmediatamente.



Monitoreo y mejora continua

Actualizar el modelo cada 6 meses con nuevos datos.