

# Proyecto de Machine Learning: Predicción de Cancelación de Clientes

## 1. Introducción

---

Al operador de telecomunicaciones **Interconnect** le gustaría poder pronosticar su tasa de cancelación de clientes. Si se descubre que un usuario o usuaria planea irse, se le ofrecerán códigos promocionales y opciones de planes especiales. El equipo de marketing de Interconnect ha recopilado algunos de los datos personales de sus clientes, incluyendo información sobre sus planes y contratos.

En este proyecto se realizara un anlysis exploratorio de los datos proporcionados, se manipularan y procesaran de manera que se pueda crear al final de una etapa de preprocesamiento, un dataset para poder entrenar diferentes modelos, la métrica clave será 'AUC-ROC' y se buscara un modelo que cumpla con el siguiente objetivo:

### Criterios de evaluación:

- **AUC-ROC < 0.75** — 0 SP
- **$0.75 \leq \text{AUC-ROC} < 0.81$**  — 4 SP
- **$0.81 \leq \text{AUC-ROC} < 0.85$**  — 4.5 SP
- **$0.85 \leq \text{AUC-ROC} < 0.87$**  — 5 SP
- **$0.87 \leq \text{AUC-ROC} < 0.88$**  — 5.5 SP
- **AUC-ROC  $\geq 0.88$**  — 6 SP

Usare:

**Clasificación:** Random Forest, XGBoost, Logistic Regression, Redes Neuronales.

```
# Importar librerías para el preprocesamiento de datos
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline

# Importar librerías para la modelización
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, roc_auc_score, accuracy_score
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
import numpy as np
import pandas as pd
import re

# Importar librerías para visualización
import matplotlib.pyplot as plt
import seaborn as sns
```

```
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots as sp

# Importar librerías adicionales (si es necesario)
import datetime
```

## 2. Análisis Exploratorio de Datos (EDA)

Comenzaremos con el analisis de los datasets en este orden:

1. contract.csv
2. internet.csv
3. personal.csv
4. phone.csv

```
# Contract.csv
```

```
df_contract = pd.read_csv('datasets/contract.csv')
```

```
df_contract.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customerID            7043 non-null  object
1   BeginDate             7043 non-null  object
2   EndDate               7043 non-null  object
3   Type                  7043 non-null  object
4   PaperlessBilling      7043 non-null  object
5   PaymentMethod         7043 non-null  object
6   MonthlyCharges        7043 non-null  float64
7   TotalCharges          7043 non-null  object
dtypes: float64(1), object(7)
memory usage: 440.3+ KB
```

```
df_contract.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customerID            7043 non-null  object
```

1	BeginDate	7043	non-null	object
2	EndDate	7043	non-null	object
3	Type	7043	non-null	object
4	PaperlessBilling	7043	non-null	object
5	PaymentMethod	7043	non-null	object
6	MonthlyCharges	7043	non-null	float64
7	TotalCharges	7043	non-null	object

dtypes: float64(1), object(7)

memory usage: 440.3+ KB

La primera impresión es positiva, no hay columnas con entradas nulas, solo hay que cambiar algunos tipos de dato como por ejemplo, **BeginDate** a tipo **datetime**, **TotalCharges** a tipo **float64** y modificar los nombres

```
def ColumName_change(df):
    new_cols = []

    for col_name in df.columns:
        # Paso 1: Insertar guion bajo antes de mayúsculas, pero manejar adecuadamente secuencias de escape
        new_col_name = re.sub(r'(?<=[a-z0-9])([A-Z])', r'_\1', col_name) # Insertar _ antes de mayúsculas

        # Paso 2: Convertir todo a minúsculas
        new_col_name = new_col_name.lower()

        # Paso 3: Eliminar cualquier guion bajo inicial o duplicado
        new_col_name = re.sub(r'_+', '_', new_col_name).strip('_')

        new_cols.append(new_col_name)

    df.columns = new_cols
    return df
```

```
# Aplicación de la función a un DataFrame
df_contract = ColumName_change(df_contract)
df_contract
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
0	7590-VHVEG	2020-01-01	No	Month-to-month	Yes	Electronic check	29.85	29.85
1	5575-GNVDE	2017-04-01	No	One year	No	Mailed check	56.95	1889.5
2	3668-QPYBK	2019-10-01	2019-12-01 00:00:00	Month-to-month	Yes	Mailed check	53.85	108.15
3	7795-	2016-05-01	No	One	No	Bank transfer	42.30	1840.75

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
	CFOCW			year		(automatic)		
4	9237-HQITU	2019-09-01	2019-11-01 00:00:00	Month-to-month	Yes	Electronic check	70.70	151.65
...	...	...	...	...	...	...	...	...
7038	6840-RESVB	2018-02-01	No	One year	Yes	Mailed check	84.80	1990.5
7039	2234-XADUH	2014-02-01	No	One year	Yes	Credit card (automatic)	103.20	7362.9
7040	4801-JAZL	2019-03-01	No	Month-to-month	Yes	Electronic check	29.60	346.45
7041	8361-LTMKD	2019-07-01	2019-11-01 00:00:00	Month-to-month	Yes	Mailed check	74.40	306.6
7042	3186-AJIEK	2014-08-01	No	Two year	Yes	Bank transfer (automatic)	105.65	6844.5

7043 rows × 8 columns

```
# Cambios de tipo de dato en la columnas

df_contract['begin_date'] = pd.to_datetime(df_contract['begin_date'])
df_contract['total_charges'] = pd.to_numeric(df_contract['total_charges'], errors='coerce')

df_contract.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
0   customer_id           7043 non-null  object
1   begin_date            7043 non-null  datetime64[ns]
2   end_date              7043 non-null  object
3   type                  7043 non-null  object
4   paperless_billing     7043 non-null  object
5   payment_method        7043 non-null  object
6   monthly_charges       7043 non-null  float64
7   total_charges         7032 non-null  float64
dtypes: datetime64[ns](1), float64(2), object(5)
memory usage: 440.3+ KB
```

En la linea modificada de **total\_charges** lo converti a numerico y los datos que no pudiesen convertirse a float se hicieron 'Nan', por lo tanto hay que rellenar esos datos, como son pocos lo haré a continuación

pues se puede buscar la posibilidad de calcularlo mediante el tiempo que llevan y su cargo mensual

```
# Observando los valores nulos de de la columna total_charges

df_contract[df_contract['total_charges'].isnull()]
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
488	4472-LVYGI	2020-02-01	No	Two year	Yes	Bank transfer (automatic)	52.55	NaN
753	3115-CZMZD	2020-02-01	No	Two year	No	Mailed check	20.25	NaN
936	5709-LVOEQ	2020-02-01	No	Two year	No	Mailed check	80.85	NaN
1082	4367-NUYAO	2020-02-01	No	Two year	No	Mailed check	25.75	NaN
1340	1371-DWPAZ	2020-02-01	No	Two year	No	Credit card (automatic)	56.05	NaN
3331	7644-OMVMY	2020-02-01	No	Two year	No	Mailed check	19.85	NaN
3826	3213-VVOLG	2020-02-01	No	Two year	No	Mailed check	25.35	NaN
4380	2520-SGTTA	2020-02-01	No	Two year	No	Mailed check	20.00	NaN
5218	2923-ARZLG	2020-02-01	No	One year	Yes	Mailed check	19.70	NaN
6670	4075-WKNIU	2020-02-01	No	Two year	No	Mailed check	73.35	NaN
6754	2775-SEFEE	2020-02-01	No	Two year	Yes	Bank transfer (automatic)	61.90	NaN

```
df_contract[df_contract['type'] == 'Two year']
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
11	7469-LKBCI	2018-10-01	No	Two year	No	Credit card (automatic)	18.95	326.80
15	3655-SNQYZ	2014-05-01	No	Two year	No	Credit card (automatic)	113.25	7895.15
17	9959-WOFKT	2014-03-01	No	Two year	No	Bank transfer (automatic)	106.70	7382.25

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
23	3638-WEABW	2015-04-01	No	Two year	Yes	Credit card (automatic)	59.90	3505.10
28	5248-YGIJN	2014-02-01	No	Two year	Yes	Credit card (automatic)	90.25	6369.45
...	...	...	...	...	...	...	...	...
7017	4807-IZYOZ	2015-11-01	No	Two year	No	Bank transfer (automatic)	20.65	1020.75
7019	9710-NJERN	2016-11-01	No	Two year	No	Mailed check	20.15	826.00
7028	9281-CEDRU	2014-06-01	No	Two year	No	Bank transfer (automatic)	64.10	4326.25
7037	2569-WGERO	2014-02-01	No	Two year	Yes	Bank transfer (automatic)	21.15	1419.40
7042	3186-AJIEK	2014-08-01	No	Two year	Yes	Bank transfer (automatic)	105.65	6844.50

1695 rows × 8 columns

```
df_contract[df_contract['begin_date'] == '2020-02-01']
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
488	4472-LVYGI	2020-02-01	No	Two year	Yes	Bank transfer (automatic)	52.55	NaN
753	3115-CZMZD	2020-02-01	No	Two year	No	Mailed check	20.25	NaN
936	5709-LVOEQ	2020-02-01	No	Two year	No	Mailed check	80.85	NaN
1082	4367-NUYAO	2020-02-01	No	Two year	No	Mailed check	25.75	NaN
1340	1371-DWPAZ	2020-02-01	No	Two year	No	Credit card (automatic)	56.05	NaN
3331	7644-OMVMY	2020-02-01	No	Two year	No	Mailed check	19.85	NaN
3826	3213-VVOLG	2020-02-01	No	Two year	No	Mailed check	25.35	NaN
4380	2520-SGTTA	2020-02-01	No	Two year	No	Mailed check	20.00	NaN
5218	2923-ARZLG	2020-02-01	No	One year	Yes	Mailed check	19.70	NaN

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
6670	4075-WKNIU	2020-02-01	No	Two year	No	Mailed check	73.35	NaN
6754	2775-SEFEE	2020-02-01	No	Two year	Yes	Bank transfer (automatic)	61.90	NaN

```
# Encontrar la fecha más actual
fecha_mas_actual = df_contract['begin_date'].max()

print("La fecha más actual es:", fecha_mas_actual)
```

La fecha más actual es: 2020-02-01 00:00:00

Explorando los datos veo que las peronas que tienen el registro más actual son de la fecha : 2020-02-01

También observo que los clientes que iniciaron en esa fecha no tienen **total\_charge**, por lo tanto para calcular el pago hay que multiplicar la carga mensual que si viene registrada por los 24 meses del contrato

```
def Total_chargeCalculate(df):
    # Convertir 'type' a minúsculas para consistencia
    df['type'] = df['type'].str.lower()

    # Crear máscara para filas con 'total_charges' nulo y tipo de contrato
    mask_two_year = (df['total_charges'].isnull()) & (df['type'] == 'two year')
    mask_one_year = (df['total_charges'].isnull()) & (df['type'] == 'one year')

    # Actualizar 'total_charges' para 'two year'
    df.loc[mask_two_year, 'total_charges'] = df.loc[mask_two_year, 'monthly_charges'] * 24

    # Actualizar 'total_charges' para 'one year'
    df.loc[mask_one_year, 'total_charges'] = df.loc[mask_one_year, 'monthly_charges'] * 12

    return df
```

```
df_contract = Total_chargeCalculate(df_contract)

df_contract[df_contract['total_charges'].isnull()]
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
--	-------------	------------	----------	------	-------------------	----------------	-----------------	---------------

Con esto confirmamos que se calculo correctamente todos lo valores, los cargos que no habían sido registrados ya se calcularon

```
# Entradas unicas en cada fila

df_contract.nunique()
```

```
customer_id      7043
begin_date       77
end_date         5
type             3
paperless_billing 2
payment_method   4
monthly_charges  1585
total_charges    6541
dtype: int64
```

```
# Imprimiendo entradas unicas de columnas
cat_cols = ['end_date', 'type', 'paperless_billing', 'payment_method']

for col in cat_cols:
    print(df_contract[col].unique())
```

```
['No' '2019-12-01 00:00:00' '2019-11-01 00:00:00' '2019-10-01 00:00:00'
 '2020-01-01 00:00:00']
['month-to-month' 'one year' 'two year']
['Yes' 'No']
['Electronic check' 'Mailed check' 'Bank transfer (automatic)'
 'Credit card (automatic)']
```

```
for i in df_contract['end_date'].unique():
    print("Hay {} filas que tienen a '{}' como entrada en 'end_date'.format(len(df_contract[df_con
```

```
Hay 5174 filas que tienen a 'No' como entrada en 'end_date'
Hay 466 filas que tienen a '2019-12-01 00:00:00' como entrada en 'end_date'
Hay 485 filas que tienen a '2019-11-01 00:00:00' como entrada en 'end_date'
Hay 458 filas que tienen a '2019-10-01 00:00:00' como entrada en 'end_date'
Hay 460 filas que tienen a '2020-01-01 00:00:00' como entrada en 'end_date'
```

```
df_contract = df_contract.sort_values(by='begin_date')
df_contract.head(10)
```

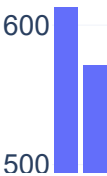
	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
4513	8580-QVLOC	2013-10-01	2019-10-01 00:00:00	two year	No	Credit card (automatic)	92.45	6440.25
4610	2889-FPWRM	2013-10-01	2019-10-01 00:00:00	one year	Yes	Bank transfer (automatic)	117.80	8684.80
3439	0917-EZOLA	2013-10-01	2019-10-01 00:00:00	two year	Yes	Bank transfer (automatic)	104.15	7689.95

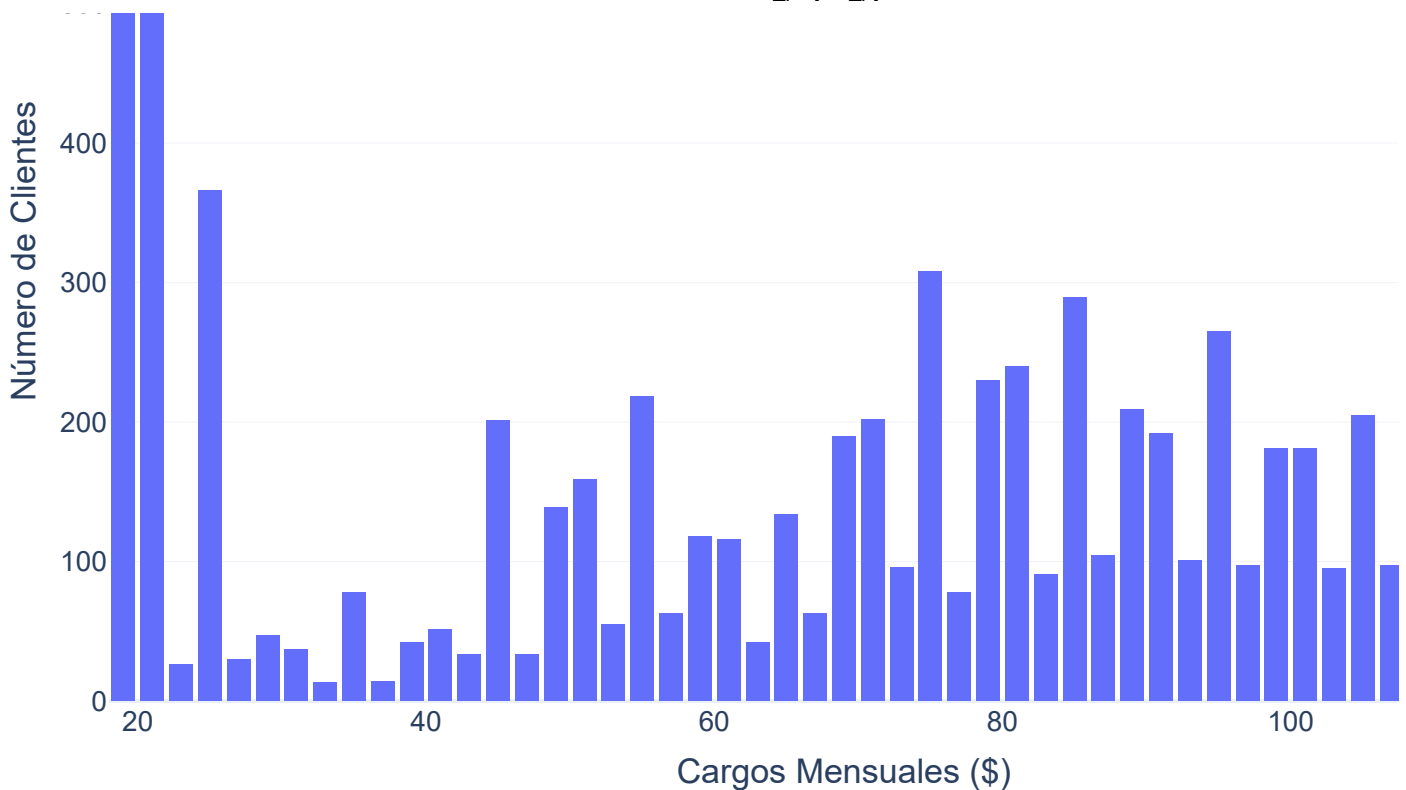


	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
975	2834-JRTUA	2013-11-01	2019-10-01 00:00:00	two year	Yes	Electronic check	108.05	7532.15
3040	7317-GGVPB	2013-11-01	2019-10-01 00:00:00	two year	Yes	Credit card (automatic)	108.60	7690.90
6038	1555-DJEQW	2013-12-01	2019-10-01 00:00:00	two year	Yes	Bank transfer (automatic)	114.20	7723.90
6290	2530-ENDWQ	2013-12-01	2019-11-01 00:00:00	two year	Yes	Bank transfer (automatic)	93.70	6585.35
5441	3512-IZIKN	2013-12-01	2019-10-01 00:00:00	two year	No	Credit card (automatic)	65.30	4759.75
6424	6034-ZRYCV	2014-01-01	2020-01-01 00:00:00	two year	Yes	Electronic check	54.20	3937.45
4684	6305-YLBMM	2014-01-01	2019-10-01 00:00:00	one year	Yes	Bank transfer (automatic)	104.05	7262.00

```
fig = px.histogram(df_contract,
x='monthly_charges',
title='Distribución de Cargos Mensuales',
color_discrete_sequence=['#636EFA'], # Color moderno
template='plotly_white' # Diseño limpio
)
fig.update_layout(
title_font_size=24,
xaxis_title='Cargos Mensuales ($)',
yaxis_title='Número de Clientes',
font=dict(family="Arial", size=14),
bargap=0.2
)
fig.show()
```

## Distribución de Cargos Mensuales





Se observa que la distribución entre el cargo mensual y el número de clientes es predominante en los 18-22 dolares, existe a partir de los 48 dolares un cargo más estable

```
fig = px.scatter(
    df_contract,
    x='monthly_charges',
    y='total_charges',
    title='Cargos Mensuales vs. Cargos Totales',
    labels={'monthly_charges': 'Cargos Mensuales ($)', 'total_charges': 'Cargos Totales ($)'},
    color='type',
    color_discrete_sequence=px.colors.qualitative.Plotly,
    template='plotly_white',
    opacity=0.6 # Transparencia para reducir sobreposición
)

# Ajuste del tamaño de los puntos y línea de tendencia
fig.update_traces(marker=dict(size=8, line=dict(width=1, color='DarkSlateGrey'))

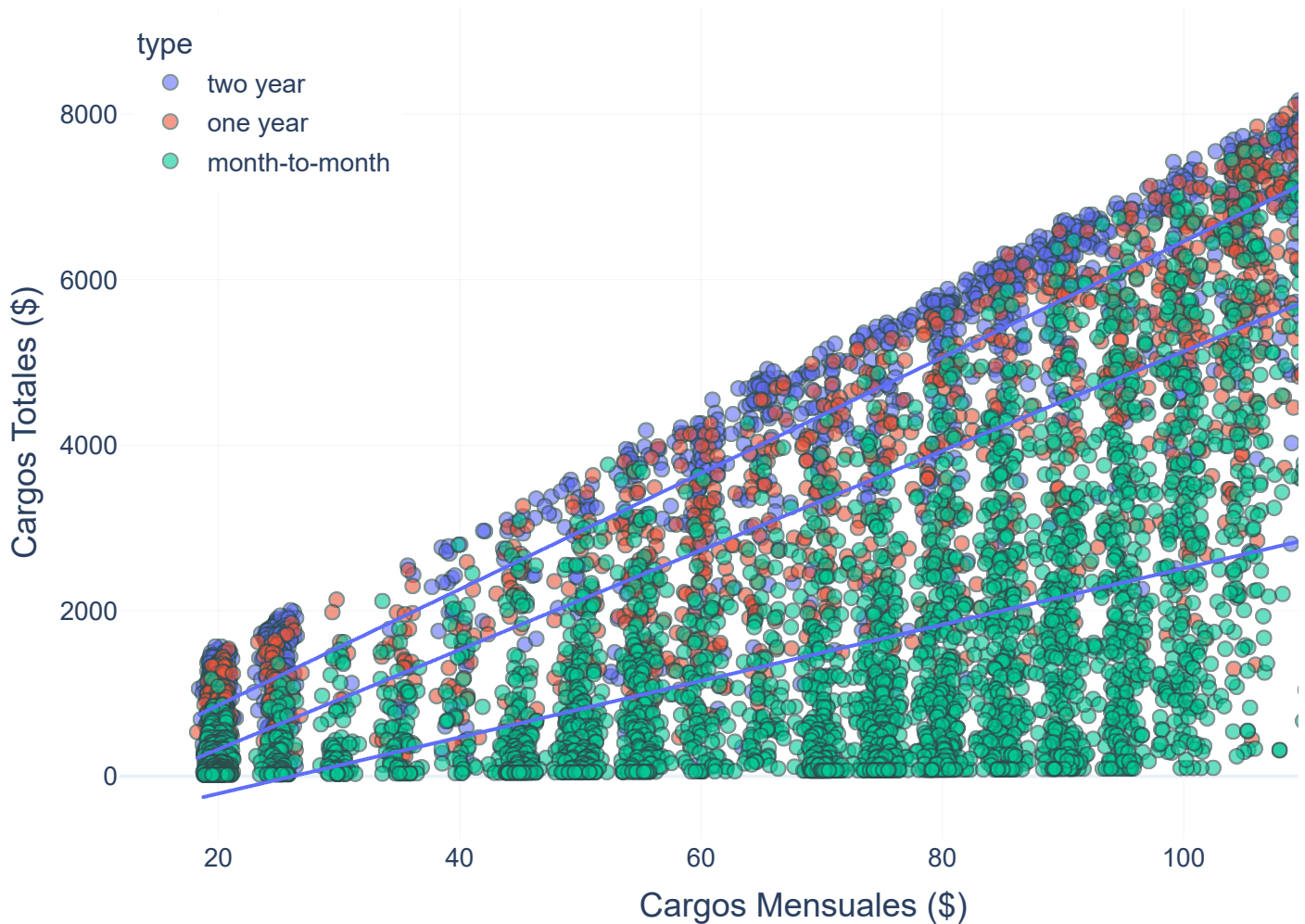
# Mejora del layout
fig.update_layout(
    title_font_size=24,
    xaxis_title_font=dict(size=18),
    yaxis_title_font=dict(size=18),
    legend_title_font=dict(size=16),
    font=dict(family="Arial", size=14),
    legend=dict(yanchor="top", y=0.99, xanchor="left", x=0.01), # Mover la leyenda
)

# Agregar línea de tendencia para cada tipo de contrato
```

```
for contract_type in df_contract['type'].unique():
    df_subset = df_contract[df_contract['type'] == contract_type]
    fig.add_traces(px.scatter(df_subset, x='monthly_charges', y='total_charges',
                              trendline="ols").data[1])

fig.show()
```

## Cargos Mensuales vs. Cargos Totales



Existe una clara relación entre cargos mensuales y totales

```
df_contract_copy_1 = df_contract.copy()
df_contract_copy_1['begin_date'] = pd.to_datetime(df_contract_copy_1['begin_date'])
df_contract_copy_1['end_date'] = pd.to_datetime(df_contract_copy_1['end_date'], errors='coerce')
df_contract_copy_1['contract_duration'] = (df_contract_copy_1['end_date'] - df_contract['begin_da

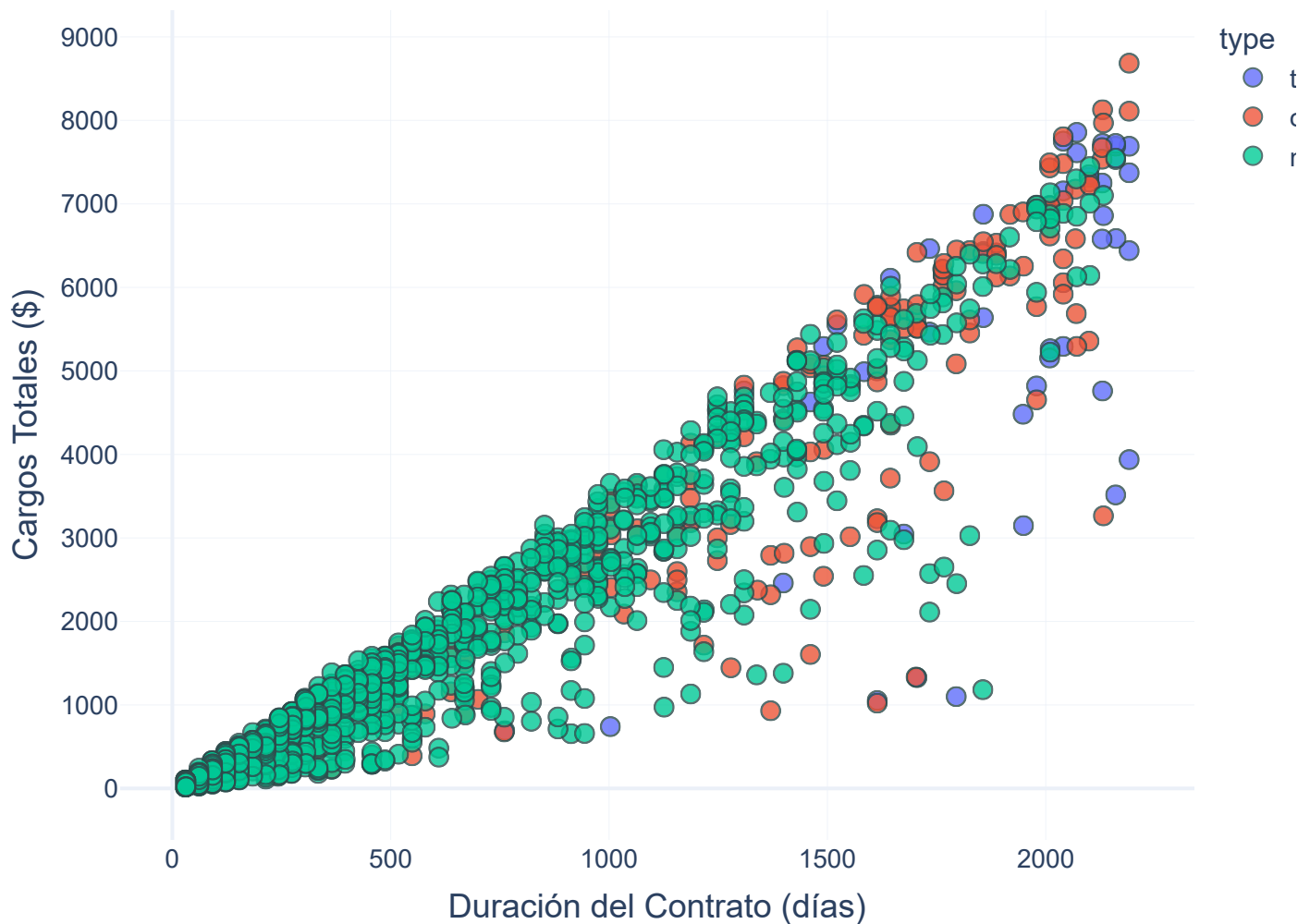
fig = px.scatter(
    df_contract_copy_1,
    x='contract_duration',
    y='total_charges',
    title='Duración del Contrato vs. Cargos Totales',
    labels={'contract_duration': 'Duración del Contrato (días)', 'total_charges': 'Cargos Totales',
    color='type', # Diferenciar por tipo de contrato
    color_continuous_scale=px.colors.sequential.Viridis,
    template='plotly_white'
```

```

)
fig.update_traces(marker=dict(size=10, opacity=0.8, line=dict(width=1, color='DarkSlateGrey')))
fig.update_layout(
    title_font_size=24,
    xaxis_title_font=dict(size=18),
    yaxis_title_font=dict(size=18),
    legend_title_font=dict(size=16),
    font=dict(family="Arial", size=14)
)
fig.show()

```

## Duración del Contrato vs. Cargos Totales



Vemos que hay una reacción entre la duración del contrato y los cargos totales, parece un grosor de datos como línea recta tendencial

```

fig = px.box(
    df_contract,
    x='payment_method',
    y='monthly_charges',
    title='Cargos Mensuales por Método de Pago',
    labels={'payment_method': 'Método de Pago', 'monthly_charges': 'Cargos Mensuales ($)'},
    color='payment_method',

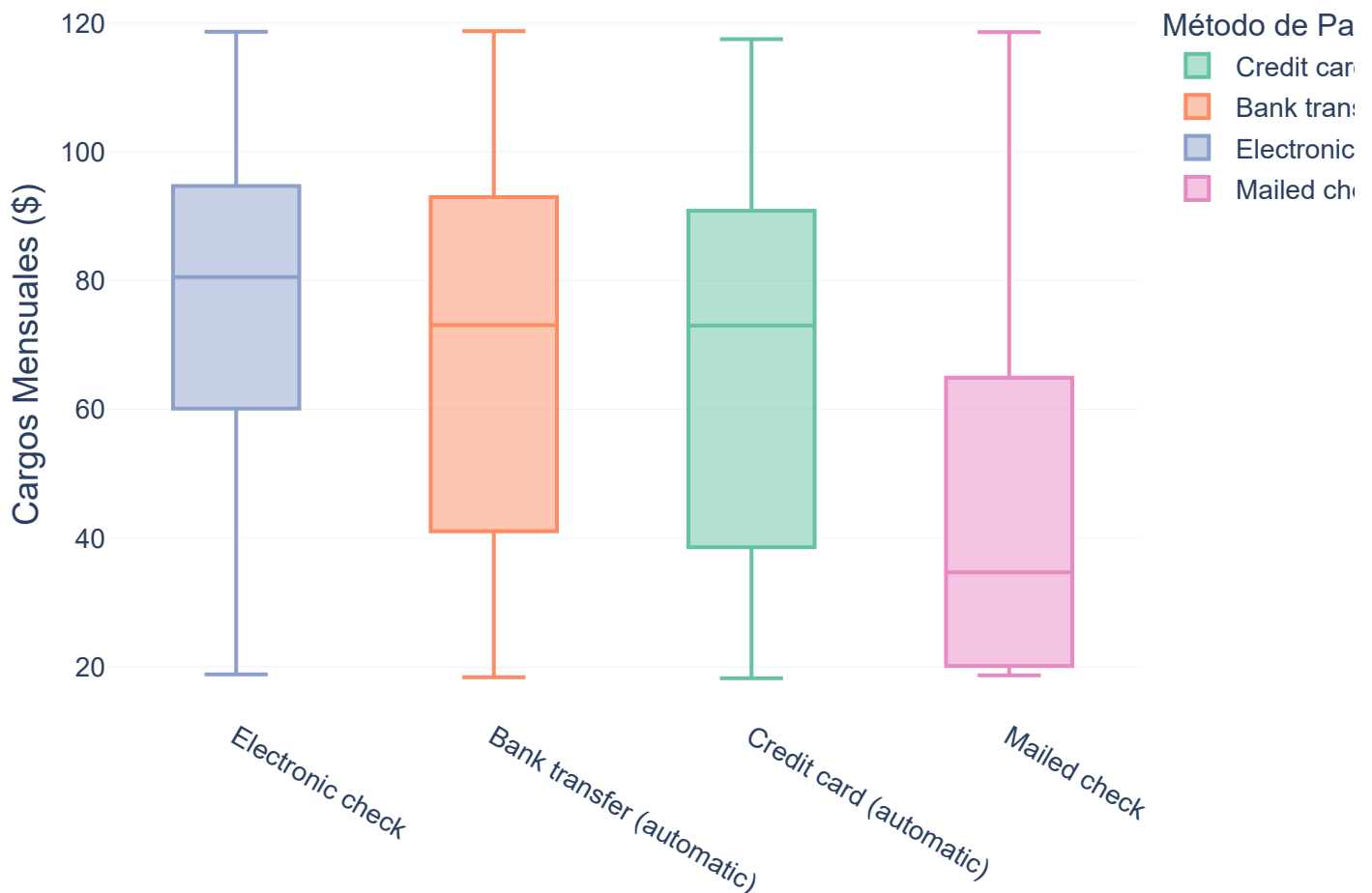
```

```

color_discrete_sequence=px.colors.qualitative.Set2, # Colores suaves y modernos
template='plotly_white'
)
fig.update_layout(
    title_font_size=24,
    xaxis_title_font=dict(size=18),
    yaxis_title_font=dict(size=18),
    legend_title_font=dict(size=16),
    font=dict(family="Arial", size=14),
    xaxis={'categoryorder':'total descending'} # Ordenar categorías
)
fig.show()

```

## Cargos Mensuales por Método de Pago



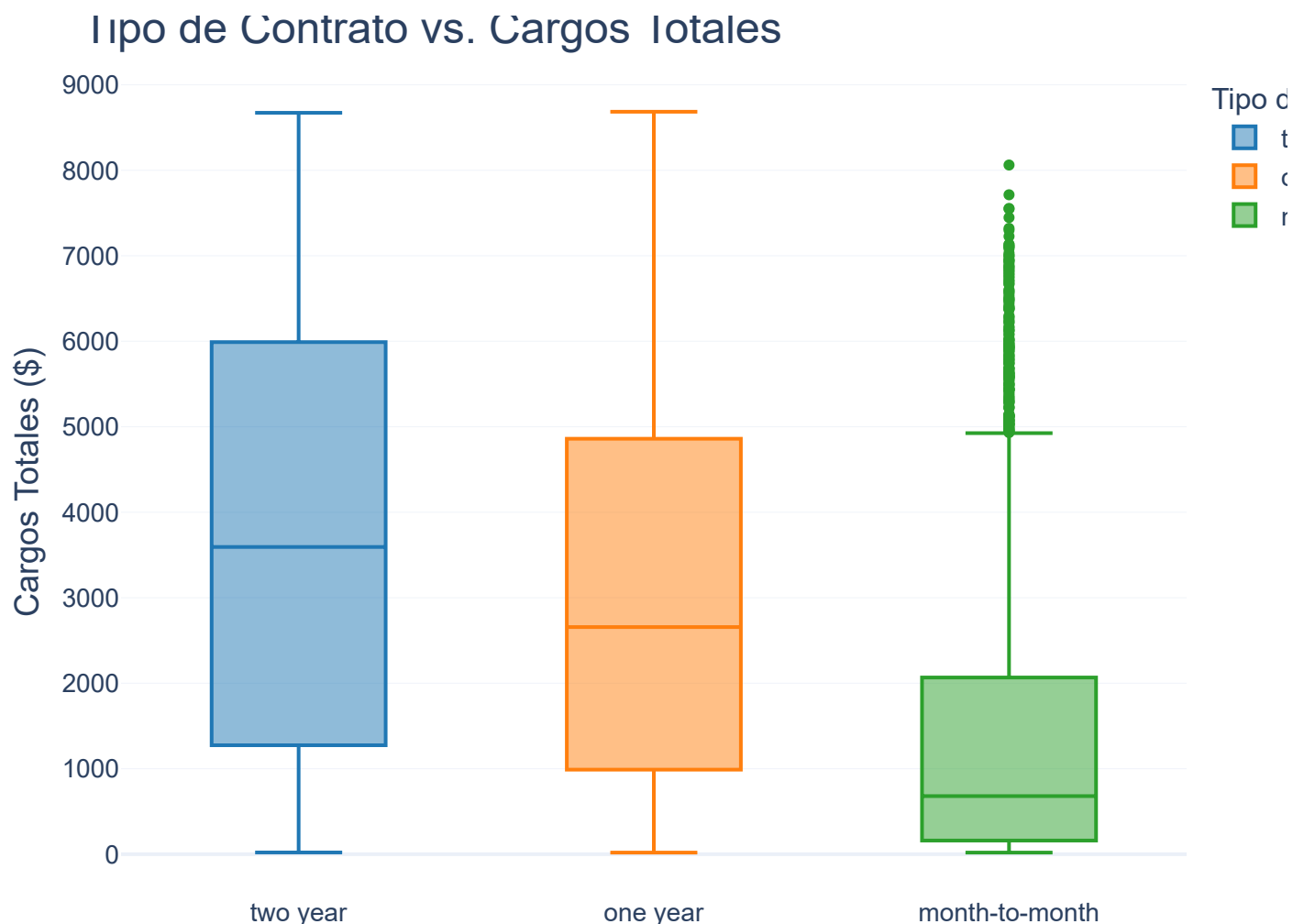
### Método de Pago

Se observa en los boxplots entre método de pago y cargos mensuales que 'Bank transfer' y 'Credit card' tienen cajas muy similares, puede ser debido a que el método de pago es algo similar, las otras 2 difieren bastante.

Hablando de el método de pago por chequeo electrónico vemos que los cargos mensuales entre el cuartil 1 y 3 están más comprimidos, la mediana es la más alta de todas (\$80.55), el 75% de los cargos mensuales por ese método son inferiores a \$94.7.

En cuanto a los cheques por correo es todo lo contrario, el 25% de los datos es menor a \$18.7 que de hecho coincide con el dato mínimo por este metodo, el 75% tienen cargos mensuales menores a los \$65 por los que puede ser que haya una tendencia a pagar por este metodo cuando las caragas mensuales son inferiores.

```
fig = px.box(
    df_contract,
    x='type',
    y='total_charges',
    title='Tipo de Contrato vs. Cargos Totales',
    labels={'type': 'Tipo de Contrato', 'total_charges': 'Cargos Totales ($)'},
    color='type',
    color_discrete_sequence=px.colors.qualitative.D3, # Colores vibrantes
    template='plotly_white'
)
fig.update_layout(
    title_font_size=24,
    xaxis_title_font=dict(size=18),
    yaxis_title_font=dict(size=18),
    legend_title_font=dict(size=16),
    font=dict(family="Arial", size=14)
)
fig.show()
```



## Tipo de Contrato

A primera vista podemos ver que la caja de los que tienen su contrato mes a mes esta comprimida a cargos totales inferiores, también tiene una cantidad considerable de datos anomalos como es de esperarse,

```
# Copia del dataset original
df_contract_copy = df_contract.copy()

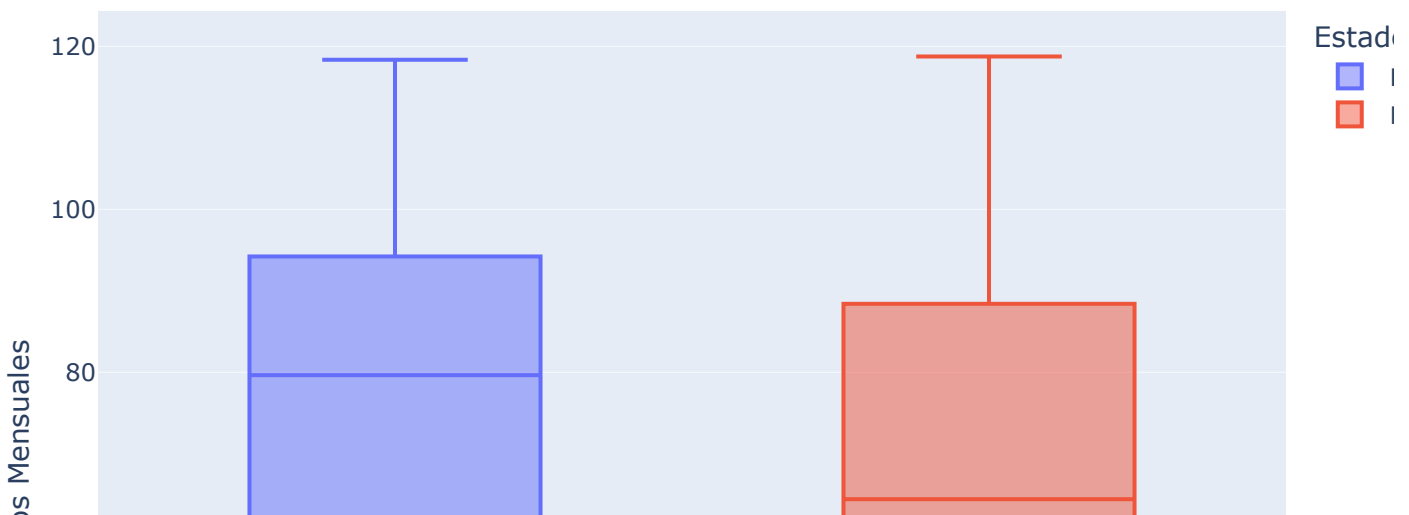
# Crear la columna 'end_status'
df_contract_copy['end_status'] = df_contract_copy['end_date'].apply(lambda x: 'No finalizado' if x < '2023-01-01' else 'Finalizado')

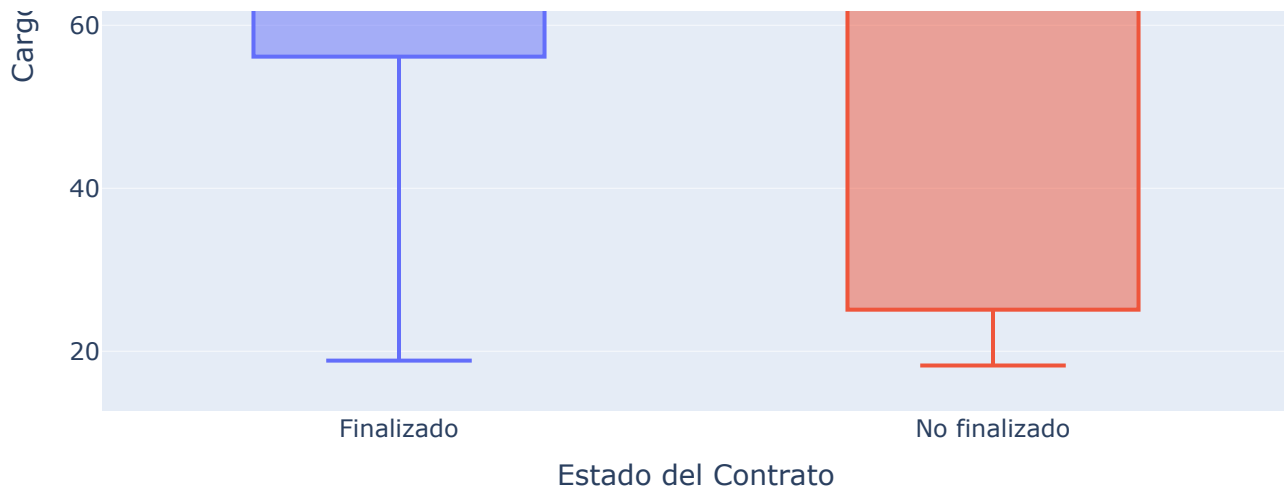
# Gráfico de distribución de cargos mensuales en función del estado del contrato
fig_monthly_charges = px.box(df_contract_copy,
                             x='end_status',
                             y='monthly_charges',
                             title='Distribución de Cargos Mensuales según Estado del Contrato',
                             color='end_status',
                             labels={'end_status': 'Estado del Contrato', 'monthly_charges': 'Cargos Mensuales'},
                             color_discrete_sequence=['#636EFA', '#EF553B'])

# Gráfico de distribución de cargos totales en función del estado del contrato
fig_total_charges = px.box(df_contract_copy,
                            x='end_status',
                            y='total_charges',
                            title='Distribución de Cargos Totales según Estado del Contrato',
                            color='end_status',
                            labels={'end_status': 'Estado del Contrato', 'total_charges': 'Cargos Totales'},
                            color_discrete_sequence=['#636EFA', '#EF553B'])

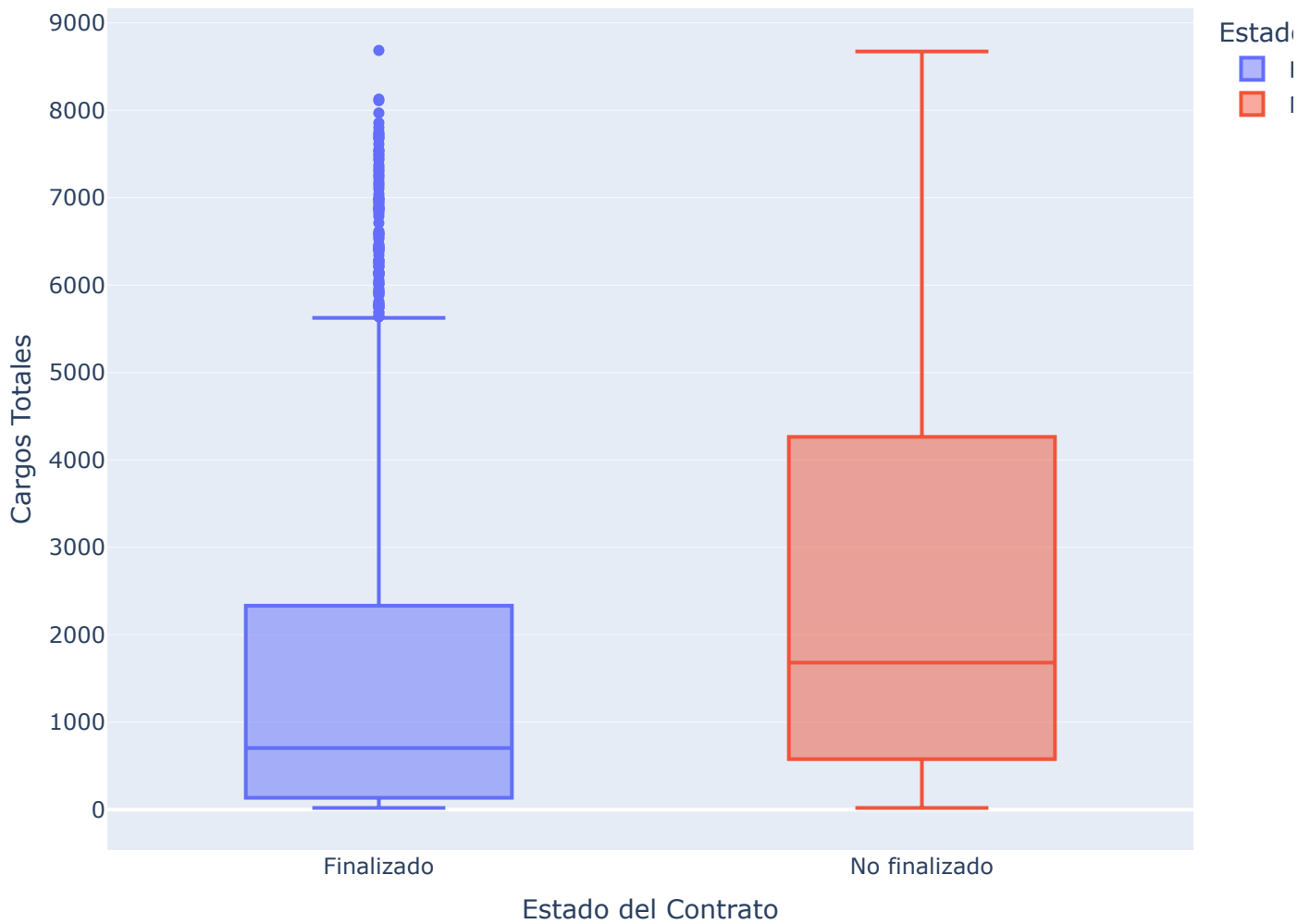
# Mostrar los gráficos
fig_monthly_charges.show()
fig_total_charges.show()
```

Distribución de Cargos Mensuales según Estado del Contrato





Distribución de Cargos Totales según Estado del Contrato



Vemos 2 graficos que comparan a las personas que no han finalizado su contrato y las que ya lo hicieron

En el primer gradico observamos estos 2 tipos clientes por Cargos mensuales observamos que los que no han finalizado su contrato tienen cargos mensuales más variados y los que ya lo finalizaron suelen tener cargos más altos, esto se puede observar en las diferencias del primer quartil de ambos tipos de clientes, los que ya lo finalizaron es de \$56 y los que no \$25.1, el 3er quartil también existe una diferencia que puede hacer pensar que los que finalizas su contrato suelen tener caragas mensuales más altas.



En el segundo grafico se observa que los clientes que finalizan suelen tener cargas totales más bajas, las cargas altas suelen ser anomalías, se podría decir que son clientes de corto plazo, mientras que los que no han finalizado tienen cargas más altas, quizá clientes que están contentos con el servicio a largo plazo no suelen finalizar

```
# internet.csv

df_internet = pd.read_csv('datasets/internet.csv')
```

```
df_internet.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5517 entries, 0 to 5516
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customerID            5517 non-null   object
1   InternetService        5517 non-null   object
2   OnlineSecurity         5517 non-null   object
3   OnlineBackup           5517 non-null   object
4   DeviceProtection       5517 non-null   object
5   TechSupport            5517 non-null   object
6   StreamingTV            5517 non-null   object
7   StreamingMovies        5517 non-null   object
dtypes: object(8)
memory usage: 344.9+ KB
```

Este dataset da información sobre los servicios de internet, los tipos de datos parecen estar bien únicamente sería modificar los nombres de las columnas

```
df_internet = ColumName_change(df_internet)

df_internet
```

	customer_id	internet_service	online_security	online_backup	device_protection	tech_support	streaming_tv
0	7590-VHVEG	DSL	No	Yes	No	No	No
1	5575-GNVDE	DSL	Yes	No	Yes	No	No
2	3668-QPYBK	DSL	Yes	Yes	No	No	No
3	7795-CFOCW	DSL	Yes	No	Yes	Yes	No
4	9237-HQITU	Fiber optic	No	No	No	No	No
...	...	...	...	...	...	...	...
5512	6840-RESVB	DSL	Yes	No	Yes	Yes	Yes

	customer_id	internet_service	online_security	online_backup	device_protection	tech_support	streaming_tv
5513	2234-XADUH	Fiber optic	No	Yes	Yes	No	Yes
5514	4801-JAZL	DSL	Yes	No	No	No	No
5515	8361-LTMKD	Fiber optic	No	No	No	No	No
5516	3186-AJIEK	Fiber optic	Yes	No	Yes	Yes	Yes

5517 rows × 8 columns

```
df_internet.nunique()
```

```
customer_id      5517
internet_service      2
online_security      2
online_backup        2
device_protection     2
tech_support          2
streaming_tv          2
streaming_movies      2
dtype: int64
```

Se observa que principalmente se contrata el servicio de fibra optica

```
import pandas as pd
import plotly.graph_objects as go
import plotly.express as px
from plotly.subplots import make_subplots

# Crear el DataFrame para contar las características por tipo de servicio de internet
df_counts = pd.DataFrame()
services = ['online_security', 'online_backup', 'device_protection', 'tech_support', 'streaming_tv']

for service in services:
    service_counts = df_internet.groupby(['internet_service', service]).size().reset_index(name='count')
    df_counts = pd.concat([df_counts, service_counts], axis=0)

# Crear los subplots
fig = make_subplots(rows=3, cols=1,
                    subplot_titles=('Distribución de Servicios de Internet',
                                    'Distribución de Servicios Adicionales',
                                    'Distribución de Servicios Adicionales por Tipo de Servicio de Internet'),
                    row_heights=[0.3, 0.3, 0.4])

# Gráfico 1: Distribución de Servicios de Internet
internet_service_counts = df_internet['internet_service'].value_counts()
fig.add_trace(go.Bar(
    x=internet_service_counts.index,
```

```

y=internet_service_counts.values,
name='Servicios de Internet',
marker_color=px.colors.qualitative.Set1[0]
), row=1, col=1)

# Gráfico 2: Distribución de Características Adicionales agrupadas por servicio
for response in ['Yes', 'No']:
    for service in services:
        service_counts = df_internet[df_internet[service] == response][service].value_counts()
        fig.add_trace(go.Bar(
            x=[service.replace("_", " ").title()],
            y=service_counts.values,
            name=f'{service.replace("_", " ").title()} ({response})',
            marker_color=px.colors.qualitative.Set1[services.index(service)],
            offsetgroup=response # Agrupa "Yes" y "No" juntas
        ), row=2, col=1)

# Cambiar el modo de barras a 'group' para no apilar en el segundo gráfico
fig.update_xaxes(type='category', row=2, col=1)
fig.update_yaxes(title_text="Conteo", row=2, col=1)
fig.update_layout(barmode='group') # Configura el modo de barras a 'group' para este gráfico

# Gráfico 3: Distribución de Servicios Adicionales por Tipo de Servicio de Internet
for service in services:
    service_data = df_counts[df_counts[service] == 'Yes']
    fig.add_trace(go.Bar(
        x=service_data['internet_service'],
        y=service_data['count'],
        name=service.replace("_", " ").title(),
        marker_color=px.colors.qualitative.Set1[services.index(service)]
    ), row=3, col=1)

# Actualizar la disposición de los gráficos y aumentar la altura
fig.update_layout(
    title_text='Análisis de Servicios de Internet y Características Adicionales',
    showlegend=True,
    height=1000 # Ajusta esta altura según tus necesidades
)

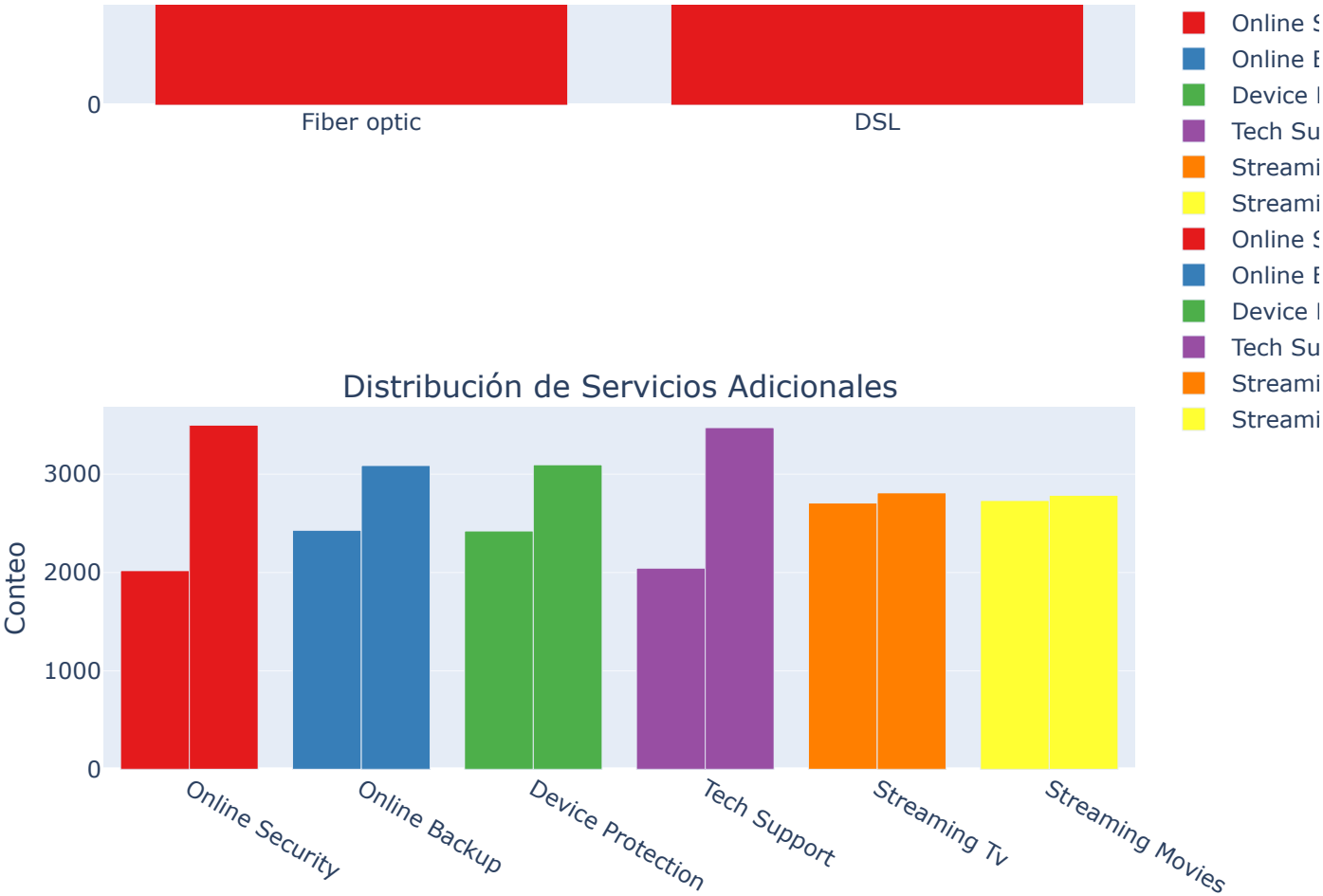
fig.show()

```

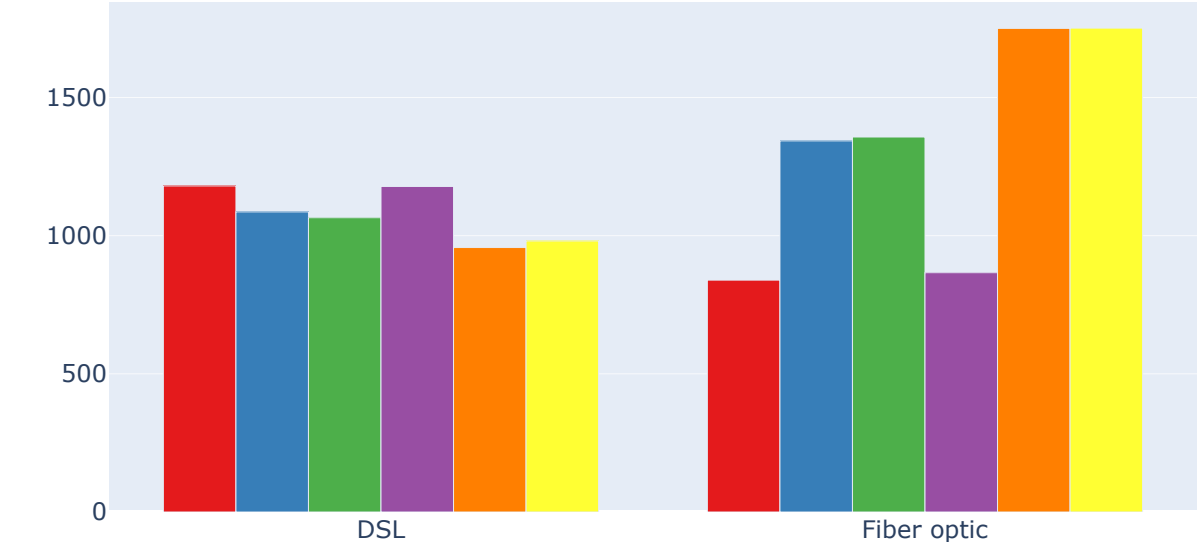
## Análisis de Servicios de Internet y Características Adicionales

### Distribución de Servicios de Internet





Distribución de Servicios Adicionales por Tipo de Servicio de Internet



En estos 3 graficos podemos ver infomación valiosa, en el primero se puede observar que hay una preferencia de contratos de internet para la fibra optica

En el segundo grafico se observa que para seguridad online, apoyo online y protección del hardware es más usual que no se contraten, sin embargo para servicios de streaming de tv y peliculas hay un igualdad entre los que si piden y los que no

En el ultimo grafico se suele pedir bastante la seguridad online y bastante parecido para los demás servicios, en la fibra ooptoca los servicios de streaming de tv y peliculas son los mas solicitados y en segundo lugarla protección del hardware y el apoyo online

```
# personal.csv
```

```
df_personal = pd.read_csv('datasets/personal.csv')
```

```
df_personal.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   customerID      7043 non-null   object
1   gender          7043 non-null   object
2   SeniorCitizen   7043 non-null   int64
3   Partner         7043 non-null   object
4   Dependents      7043 non-null   object
dtypes: int64(1), object(4)
memory usage: 275.2+ KB
```

```
df_personal.nunique()
```

```
customerID      7043
gender           2
SeniorCitizen    2
Partner          2
Dependents       2
dtype: int64
```

Viendo el dataset, son datos personales se hara un cambio de nombre de columnas y algunas visualizaciones de los datos

```
df_personal = ColumName_change(df_personal)
df_personal
```

	<b>customer_id</b>	<b>gender</b>	<b>senior_citizen</b>	<b>partner</b>	<b>dependents</b>
0	7590-VHVEG	Female	0	Yes	No
1	5575-GNVDE	Male	0	No	No
2	3668-QPYBK	Male	0	No	No
3	7795-CFOCW	Male	0	No	No
4	9237-HQITU	Female	0	No	No
...	...	...	...	...	...

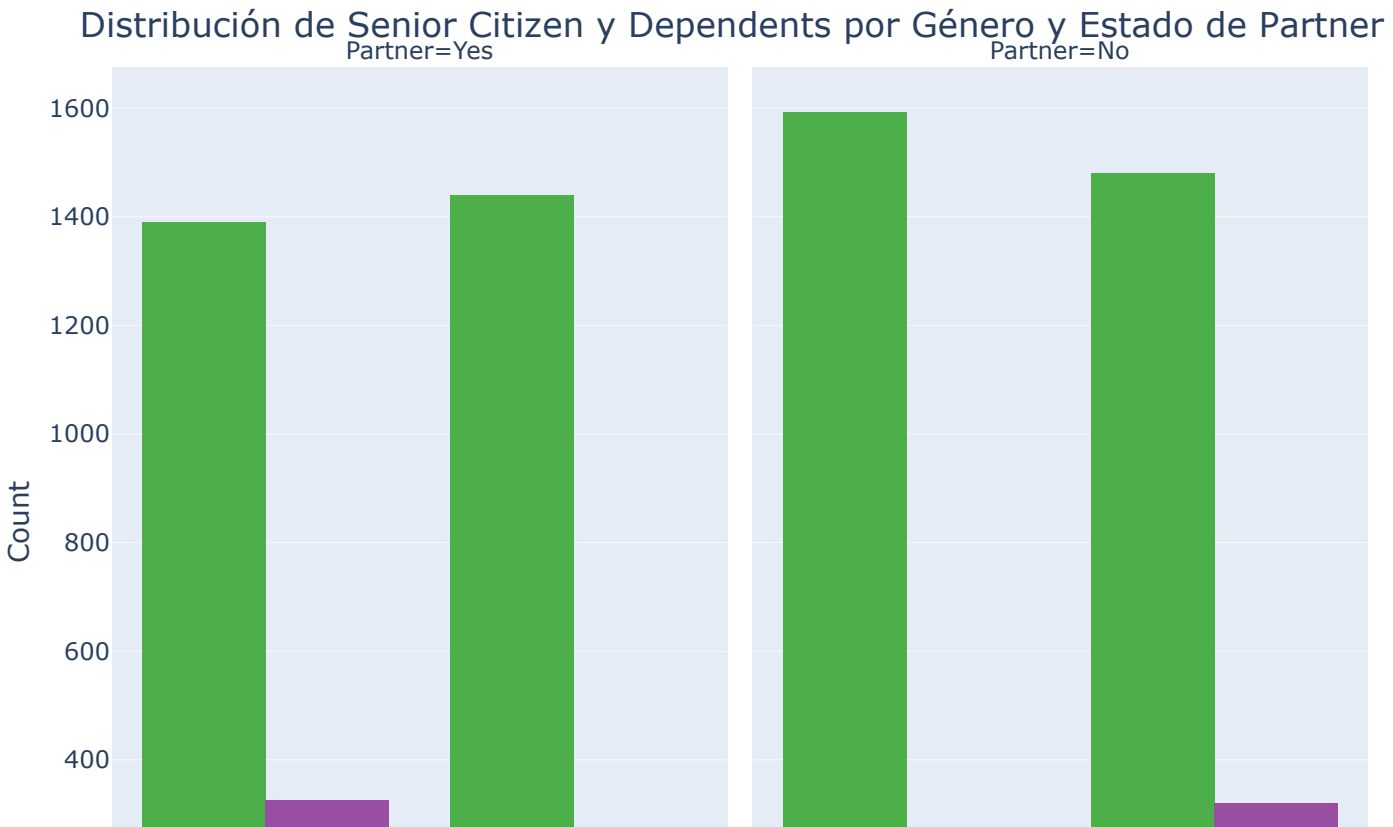
	customer_id	gender	senior_citizen	partner	dependents
7038	6840-RESVB	Male	0	Yes	Yes
7039	2234-XADUH	Female	0	Yes	Yes
7040	4801-JZAZL	Female	0	Yes	Yes
7041	8361-LTMKD	Male	1	Yes	No
7042	3186-AJIEK	Male	0	No	No

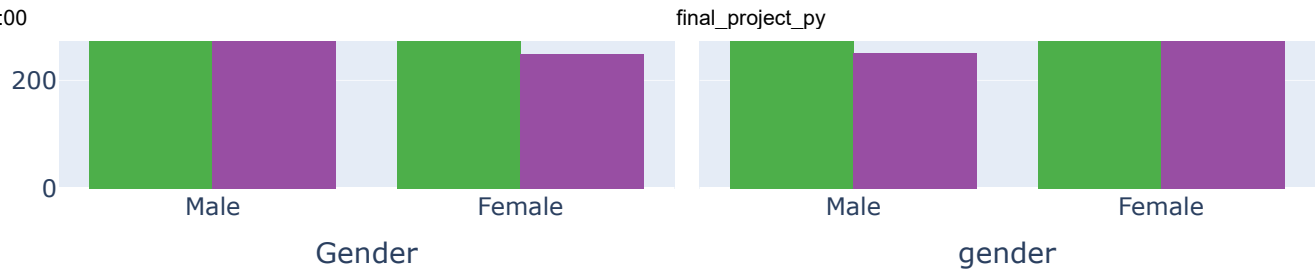
7043 rows × 5 columns

```
# Crear un gráfico de barras apiladas
fig = px.histogram(df_personal,
                   x='gender',
                   color='senior_citizen',
                   barmode='group',
                   category_orders={"gender": ["Male", "Female"], "senior_citizen": ["0", "1"]},
                   facet_col='partner',
                   title='Distribución de Senior Citizen y Dependents por Género y Estado de Partner',
                   labels={'senior_citizen': 'Senior Citizen', 'dependents': 'Dependents', 'partner': 'Partner'},
                   color_discrete_sequence=px.colors.qualitative.Set1)

fig.update_layout(barmode='group',
                  xaxis_title='Gender',
                  yaxis_title='Count')

fig.show()
```





**La mayoría de los clientes no son senior citizens** (`Senior Citizen = 0`), sin importar género o estado de pareja.

**El estado de pareja no influye significativamente** en ser o no un senior citizen.

**No hay diferencias notables entre géneros** en la proporción de senior citizens.

**Posible menor cantidad de dependents entre senior citizens**, pero requiere análisis adicional.

```
# phone.csv
```

```
df_phone = pd.read_csv('datasets/phone.csv')
```

```
df_phone.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6361 entries, 0 to 6360
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   customerID      6361 non-null   object
1   MultipleLines    6361 non-null   object
dtypes: object(2)
memory usage: 99.5+ KB
```

Es un simple df que contiene si los clientes cuentan con líneas múltiples y su id

```
df_phone = ColumName_change(df_phone)
```

```
df_phone
```

	customer_id	multiple_lines
0	5575-GNVDE	No
1	3668-QPYBK	No
2	9237-HQITU	No
3	9305-CDSKC	Yes
4	1452-KIOVK	Yes
...	...	...

	customer_id	multiple_lines
6356	2569-WGERO	No
6357	6840-RESVB	Yes
6358	2234-XADUH	Yes
6359	8361-LTMKD	Yes
6360	3186-AJIEK	No

6361 rows × 2 columns

```
# Contar la distribución de los servicios de "multiple_lines"
multiple_lines_service_counts = df_phone['multiple_lines'].value_counts()

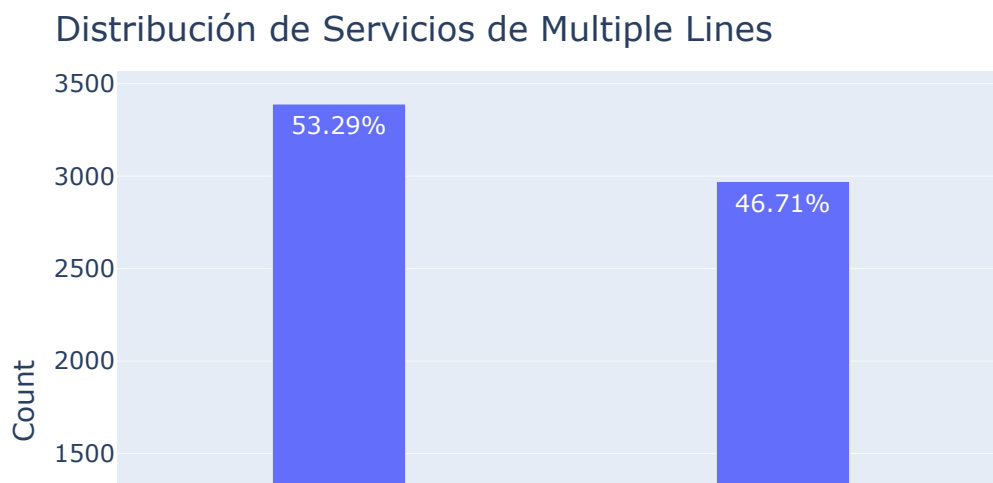
# Calcular los porcentajes
percentages = (multiple_lines_service_counts / multiple_lines_service_counts.sum()) * 100

# Crear el gráfico con barras más delgadas y un tamaño ajustado
fig = px.bar(
    x=multiple_lines_service_counts.index,
    y=multiple_lines_service_counts.values,
    labels={'x': 'Multiple Lines', 'y': 'Count'},
    title='Distribución de Servicios de Multiple Lines'
)

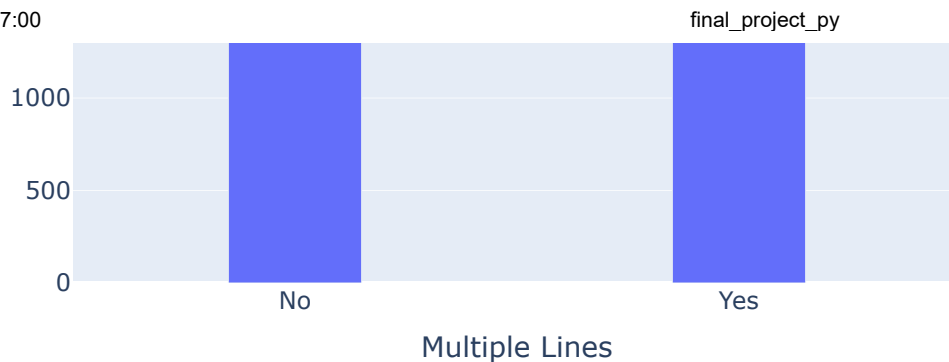
# Agregar los porcentajes como texto en las barras
fig.update_traces(
    text=percentages.round(2).astype(str) + '%',
    textposition='auto',
    width=0.3 # Hacer las barras más delgadas
)

# Ajustar el tamaño del gráfico para hacerlo menos horizontal
fig.update_layout(width=500, height=400)

# Mostrar el gráfico
fig.show()
```







Se observa en el grafico que la mayoría no cuenta con multiples lineas aunque no por mucho, el 53% no tiene y el 46.7% si tiene

### 3. Preprocesamiento de Datos

Para esta sección prepare el dataset final para poder entrenar el modelo, tomare las variables mas importantes de cada dataset y usare ingeniería de características para datos categoricos, datos numericos, para obtener un dataset limpio que se dividira en entrenamiento y validación para comenzar a entrenar los modelos.

```
# Crear un diccionario para almacenar los nombres de las columnas
data = {
    'df_contract': df_contract.columns,
    'df_internet': df_internet.columns,
    'df_phone': df_phone.columns,
    'df_personal': df_personal.columns
}

# Convertir el diccionario en un DataFrame para visualizarlo como una tabla
columns_df = pd.DataFrame.from_dict(data, orient='index').transpose()

# Mostrar la tabla
columns_df
```

	df_contract	df_internet	df_phone	df_personal
0	customer_id	customer_id	customer_id	customer_id
1	begin_date	internet_service	multiple_lines	gender
2	end_date	online_security	None	senior_citizen
3	type	online_backup	None	partner
4	paperless_billing	device_protection	None	dependents
5	payment_method	tech_support	None	None
6	monthly_charges	streaming_tv	None	None
7	total_charges	streaming_movies	None	None

- En el dataset **contract**, las columnas importantes son:

- 27/8/24, 17:00
- final\_project\_py
- begin\_date, end\_date(target), paperless\_billing, type, paymenth\_method, monthly\_charge, total\_charge

- En el dataset **internet**, las columnas importantes son:

internet\_service, online\_security, online\_backup device\_protection, tech\_support, streaming\_tv, streaming\_movies
  - En el dataset **phone**, las columnas importantes son:

multiple\_lines
  - En el dataset **personal**, las columnas importantes son:

senior\_citizen, partner, dependents
- El Genero no lo inclui por un tema de que en el analisis del dataset personal no fue muy relevante costumer\_id servira para cuando se necsiten unir los datasets pero no para un fin predictivo.

Construcción del dataset para los modelos

```
from functools import reduce

df_personal_model = df_personal.drop('gender', axis=1)
# Lista de DataFrames
dfs = [df_contract, df_internet, df_phone, df_personal_model]

# Realizar merge en cadena usando 'outer' para no perder filas
df_merged = reduce(lambda left, right: pd.merge(left, right, on='customer_id', how='outer'), dfs)
df_merged
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charg
0	0002-ORFBO	2019-05-01	No	one year	Yes	Mailed check	65.60	593.30
1	0003-MKNFE	2019-05-01	No	month-to-month	No	Mailed check	59.90	542.40
2	0004-TLHLJ	2019-09-01	2020-01-01 00:00:00	month-to-month	Yes	Electronic check	73.90	280.85
3	0011-IGKFF	2018-12-01	2020-01-01 00:00:00	month-to-month	Yes	Electronic check	98.00	1237.85
4	0013-EXCHZ	2019-09-01	2019-12-01 00:00:00	month-to-month	Yes	Mailed check	83.90	267.40
...	...	...	...	...	...	...	...	...

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
7038	9987-LUTYD	2019-01-01	No	one year	No	Mailed check	55.15	742.90
7039	9992- RRAMN	2018-02-01	2019-12- 01 00:00:00	month- to- month	Yes	Electronic check	85.10	1873.70
7040	9992-UJOEL	2019-12-01	No	month- to- month	Yes	Mailed check	50.30	92.75
7041	9993-LHIEB	2014-07-01	No	two year	No	Mailed check	67.85	4627.65
7042	9995- HOTOH	2014-11-01	No	two year	No	Electronic check	59.00	3707.60

7043 rows × 19 columns

```
df_merged.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customer_id           7043 non-null   object
1   begin_date            7043 non-null   datetime64[ns]
2   end_date              7043 non-null   object
3   type                  7043 non-null   object
4   paperless_billing     7043 non-null   object
5   payment_method        7043 non-null   object
6   monthly_charges       7043 non-null   float64
7   total_charges         7043 non-null   float64
8   internet_service      5517 non-null   object
9   online_security       5517 non-null   object
10  online_backup          5517 non-null   object
11  device_protection     5517 non-null   object
12  tech_support           5517 non-null   object
13  streaming_tv           5517 non-null   object
14  streaming_movies       5517 non-null   object
15  multiple_lines         6361 non-null   object
16  senior_citizen        7043 non-null   int64
17  partner               7043 non-null   object
18  dependents            7043 non-null   object
dtypes: datetime64[ns](1), float64(2), int64(1), object(15)
memory usage: 1.0+ MB
```

Vamos a rellenar las columnas primeramente multiple\_lines y posteriormente con las demás

```
columns_full = ['internet_service', 'online_security', 'online_backup', 'device_protection', 'tech
```

```
for df in dfs:
    print(df.shape)
    print(df['customer_id'].nunique())
```

```
(7043, 8)
```

```
7043
```

```
(5517, 8)
```

```
5517
```

```
(6361, 2)
```

```
6361
```

```
(7043, 4)
```

```
7043
```

Vemos que solo df\_personal y df\_contract son los unicos datasets con todos los ids rellenare los datos faltantes con 'unknown' para pasarlos por labelencoder posteriormente a las columnas categoricas y finalmente aplicar un StandarScaler

```
df_merged[columns_full] = df_merged[columns_full].fillna('unknown')
df_merged
```

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
0	0002-ORFBO	2019-05-01	No	one year	Yes	Mailed check	65.60	593.30
1	0003-MKNFE	2019-05-01	No	month-to-month	No	Mailed check	59.90	542.40
2	0004-TLHLJ	2019-09-01	2020-01-01 00:00:00	month-to-month	Yes	Electronic check	73.90	280.85
3	0011-IGKFF	2018-12-01	2020-01-01 00:00:00	month-to-month	Yes	Electronic check	98.00	1237.85
4	0013-EXCHZ	2019-09-01	2019-12-01 00:00:00	month-to-month	Yes	Mailed check	83.90	267.40
...	...	...	...	...	...	...	...	...
7038	9987-LUTYD	2019-01-01	No	one year	No	Mailed check	55.15	742.90
7039	9992-RRAMN	2018-02-01	2019-12-01 00:00:00	month-to-month	Yes	Electronic check	85.10	1873.70

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
7040	9992-UJOEL	2019-12-01	No	month-to-month	Yes	Mailed check	50.30	92.75
7041	9993-LHIEB	2014-07-01	No	two year	No	Mailed check	67.85	4627.65
7042	9995-HOTOH	2014-11-01	No	two year	No	Electronic check	59.00	3707.60

7043 rows × 19 columns

```
df_merged.nunique()
```

```
customer_id      7043
begin_date        77
end_date          5
type              3
paperless_billing  2
payment_method    4
monthly_charges  1585
total_charges     6541
internet_service   3
online_security    3
online_backup      3
device_protection  3
tech_support       3
streaming_tv       3
streaming_movies   3
multiple_lines     3
senior_citizen     2
partner            2
dependents         2
dtype: int64
```

```
fecha_actual = df_merged['begin_date'].max()
print(f"La fecha más reciente en el dataset es: {fecha_actual}")
```

La fecha más reciente en el dataset es: 2020-02-01 00:00:00

```
# Agrego el tiempo en días que ha estado el cliente desde el inicio de su contrato
df_merged['time_in_company'] = (fecha_actual - df_merged['begin_date']).dt.days
df_merged.describe()
```

	begin_date	monthly_charges	total_charges	senior_citizen	time_in_company
count	7043	7043.000000	7043.000000	7043.000000	7043.000000

	<b>begin_date</b>	<b>monthly_charges</b>	<b>total_charges</b>	<b>senior_citizen</b>	<b>time_in_company</b>
mean	2017-04-30 13:01:50.918642688	64.761692	2281.253259	0.162147	1006.457050
min	2013-10-01 00:00:00	18.250000	18.800000	0.000000	0.000000
25%	2015-06-01 00:00:00	35.500000	401.900000	0.000000	306.000000
50%	2017-09-01 00:00:00	70.350000	1396.250000	0.000000	883.000000
75%	2019-04-01 00:00:00	89.850000	3786.600000	0.000000	1706.000000
max	2020-02-01 00:00:00	118.750000	8684.800000	1.000000	2314.000000
std	NaN	30.090047	2265.703526	0.368612	736.596428

De media vemos que los clientes tienen 1000 días en la compañía

De media los cargo totales son de \$2281.25

De media los cargos mensuales son de \$64.76

```
df_merged['begin_year'] = df_merged['begin_date'].dt.year
df_merged['begin_month'] = df_merged['begin_date'].dt.month
df_merged['begin_day'] = df_merged['begin_date'].dt.day
```

```
df_merged['end_date'] = np.where(df_merged['end_date'] == 'No', 0, 1)
```

```
print(list(df_merged.columns))
```

```
['customer_id', 'begin_date', 'end_date', 'type', 'paperless_billing', 'payment_method',
'monthly_charges', 'total_charges', 'internet_service', 'online_security', 'online_backup',
'device_protection', 'tech_support', 'streaming_tv', 'streaming_movies', 'multiple_lines',
'senior_citizen', 'partner', 'dependents', 'time_in_company', 'begin_year', 'begin_month',
'begin_day']
```

Eliminamos las columnas que ya no son necesarias

```
df_merged.head()
```

	<b>customer_id</b>	<b>begin_date</b>	<b>end_date</b>	<b>type</b>	<b>paperless_billing</b>	<b>payment_method</b>	<b>monthly_charges</b>	<b>total_charges</b>
0	0002-ORFBO	2019-05-01	0	one year	Yes	Mailed check	65.6	593.30
1	0003-MKNFE	2019-05-01	0	month-to-month	No	Mailed check	59.9	542.40
2	0004-TLHLJ	2019-09-01	1	month-to-month	Yes	Electronic check	73.9	280.85
3	0011-IGKFF	2018-12-01	1	month-to-	Yes	Electronic check	98.0	1237.85

	customer_id	begin_date	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges
				month				
4	0013-EXCHZ	2019-09-01	1	month- to- month	Yes	Mailed check	83.9	267.40

5 rows × 23 columns

```
# Elimino begin_day puesgto que todos son 1, entonces no tiene mucha importancia
df_merged = df_merged.drop(['customer_id', 'begin_date', 'begin_day'], axis=1)
df_merged
```

	end_date	type	paperless_billing	payment_method	monthly_charges	total_charges	internet_service	onlin
0	0	one year	Yes	Mailed check	65.60	593.30	DSL	No
1	0	month- to- month	No	Mailed check	59.90	542.40	DSL	No
2	1	month- to- month	Yes	Electronic check	73.90	280.85	Fiber optic	No
3	1	month- to- month	Yes	Electronic check	98.00	1237.85	Fiber optic	No
4	1	month- to- month	Yes	Mailed check	83.90	267.40	Fiber optic	No
...	...	...	...	...	...	...	...	...
7038	0	one year	No	Mailed check	55.15	742.90	DSL	Yes
7039	1	month- to- month	Yes	Electronic check	85.10	1873.70	Fiber optic	No
7040	0	month- to- month	Yes	Mailed check	50.30	92.75	DSL	No
7041	0	two year	No	Mailed check	67.85	4627.65	DSL	Yes
7042	0	two year	No	Electronic check	59.00	3707.60	DSL	Yes

7043 rows × 20 columns

```
features = df_merged.drop('end_date', axis=1)
targets = df_merged['end_date']

features_train, features_valid, targets_train, targets_valid = train_test_split(features, targets,
```

```
cat_cols = ['type', 'paperless_billing', 'payment_method', 'internet_service', 'online_security',
            'online_backup', 'device_protection', 'tech_support', 'streaming_tv', 'streaming_movies',
            'multiple_lines', 'partner', 'dependents', 'begin_year', 'begin_month']
```

```
num_cols = ['monthly_charges', 'total_charges', 'time_in_company', 'senior_citizen']
```

```
# Haciendo un OneHotEncoder a las columnas categoricas

# Preprocesamiento para características categóricas
categorical_transformer = Pipeline(steps=[
    ('onehot', OneHotEncoder(drop='first', handle_unknown='ignore', sparse_output=False))])

# Preprocesamiento para características numéricas
numeric_transformer = Pipeline(steps=[
    ('scaler', StandardScaler())])

# Combinación de transformadores
preprocessor = ColumnTransformer(
    transformers=[
        ('num', numeric_transformer, num_cols),
        ('cat', categorical_transformer, cat_cols)])
```

```
df_models = df_merged.drop('end_date', axis=1)

df_models= preprocessor.fit_transform(df_models)
```

```
print(df_models.shape)
```

(7043, 46)

Creado el pipeline nos queda un array de 7043 entradas y 45 columnas tomando begin\_month y begin\_year como categoricas haciendoles un Ohe

## 4. Modelado



Para esta tarea usaremos los siguientes modelos

**Clasificación:** Random Forest, XGBoost, Logistic Regression, Redes Neuronales.

```
# Transformacion de características
```

```
features_train = preprocessor.fit_transform(features_train)
features_valid = preprocessor.transform(features_valid)
```

```
features_train.shape
```

(5282, 46)

```
# Regresión Logística
```

```
log_reg = LogisticRegression()
```

```
param_grid = {
    'C': [0.01, 0.1, 1],
    'solver': ['liblinear', 'saga', 'lbfgs'],
    'max_iter': [100, 200, 300]
}
```

```
grid_search = GridSearchCV(
    estimator=log_reg,
    param_grid=param_grid,
    cv=5,
    scoring='roc_auc'
)
```

```
grid_search.fit(features_train, targets_train)
```

```
best_model = grid_search.best_estimator_
```

```
probabilities = best_model.predict_proba(features_valid)[:, 1]
```

```
predictions = best_model.predict(features_valid)
```

```
print(f"Mejores parámetros encontrados: {grid_search.best_params_}")
```

```
# Mejor puntuación en la búsqueda en rejilla
print(f"Mejor puntuación: {grid_search.best_score_:.4f}")
print()
```

```
# Calcular AUC-ROC usando las probabilidades
auc_roc = roc_auc_score(targets_valid, probabilities)
print(f'AUC-ROC: {auc_roc:.2f}')

# Calcular Accuracy usando las predicciones binarias
accuracy = accuracy_score(targets_valid, predictions)
print(f'Accuracy: {accuracy:.2f}')
```

C:\Users\evolu\AppData\Roaming\Python\Python312\site-packages\sklearn\linear\_model\\_sag.py:349:  
ConvergenceWarning:

The max\_iter was reached which means the coef\_ did not converge

C:\Users\evolu\AppData\Roaming\Python\Python312\site-packages\sklearn\linear\_model\\_sag.py:349:  
ConvergenceWarning:

The max\_iter was reached which means the coef\_ did not converge

Mejores parámetros encontrados: {'C': 1, 'max\_iter': 100, 'solver': 'lbfgs'}  
Mejor puntuación: 0.8529

AUC-ROC: 0.87

Accuracy: 0.81

```
# Random forest

Rand_forest_class = RandomForestClassifier(random_state=12345)

param_grid = {
    'n_estimators': [300, 500],
    'max_depth': [None, 10, 20],
    'min_samples_split': [2, 10, 20],
    'min_samples_leaf': [1, 4],
}

grid_search = GridSearchCV(
    estimator=Rand_forest_class,
    param_grid=param_grid,
    cv=5,
    n_jobs=-2,
    scoring='roc_auc'
)

grid_search.fit(features_train, targets_train)

best_model = grid_search.best_estimator_
```

```

probabilities = best_model.predict_proba(features_valid)[: , 1]

predictions = best_model.predict(features_valid)

print(f"Mejores parámetros encontrados: {grid_search.best_params_}")

# Mejor puntuación en la búsqueda en rejilla
print(f"Mejor puntuación: {grid_search.best_score_:.4f}")
print()

# Calcular AUC-ROC usando las probabilidades
auc_roc = roc_auc_score(targets_valid, probabilities)
print(f'AUC-ROC: {auc_roc:.2f}')

# Calcular Accuracy usando las predicciones binarias
accuracy = accuracy_score(targets_valid, predictions)
print(f'Accuracy: {accuracy:.2f}')

```

Mejores parámetros encontrados: {'max\_depth': 20, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 500}

Mejor puntuación: 0.8886

AUC-ROC: 0.89

Accuracy: 0.86

```

# XGBoost

Gran_Boost_Class = GradientBoostingClassifier(random_state=12345)

param_grid = {
    'n_estimators': [300, 500],
    'max_depth': [6, 10],
    'learning_rate': [0.01, 0.1],
    'subsample': [0.7, 0.8, 1.0]
}

grid_search = GridSearchCV(
    estimator=Gran_Boost_Class,
    param_grid=param_grid,
    cv=5,
    n_jobs=-2,
    scoring='roc_auc'
)

grid_search.fit(features_train, targets_train)

```

```

best_model = grid_search.best_estimator_

probabilities = best_model.predict_proba(features_valid)[:, 1]

predictions = best_model.predict(features_valid)

print(f"Mejores parámetros encontrados: {grid_search.best_params_}")

# Mejor puntuación en la búsqueda en rejilla
print(f"Mejor puntuación: {grid_search.best_score_:.4f}")
print()

# Calcular AUC-ROC usando las probabilidades
auc_roc = roc_auc_score(targets_valid, probabilities)
print(f'AUC-ROC: {auc_roc:.2f}')

# Calcular Accuracy usando las predicciones binarias
accuracy = accuracy_score(targets_valid, predictions)
print(f'Accuracy: {accuracy:.2f}')

```

Mejores parámetros encontrados: {'learning\_rate': 0.1, 'max\_depth': 6, 'n\_estimators': 500, 'subsample': 0.8}

Mejor puntuación: 0.9272

AUC-ROC: 0.94

Accuracy: 0.90

```

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, BatchNormalization
from tensorflow.keras.optimizers import AdamW
from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint
from sklearn.metrics import roc_auc_score, accuracy_score

# Crear el modelo
model = Sequential()

# Mantener la complejidad pero con ajustes en Dropout y regularización
model.add(Dense(units=512, activation='relu', input_dim=features_train.shape[1], kernel_regularizer=keras.regularizers.l2))
model.add(BatchNormalization())
model.add(Dropout(0.5)) # Aumentar Dropout

model.add(Dense(units=256, activation='relu', kernel_regularizer='l2'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(units=128, activation='relu', kernel_regularizer='l2'))

```

```
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(units=64, activation='relu', kernel_regularizer='l2'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

# Capa de salida para clasificación binaria
model.add(Dense(units=1, activation='sigmoid'))

# Compilar el modelo con una tasa de aprendizaje ajustada
optimizer = AdamW(learning_rate=0.0001)
model.compile(
    loss='binary_crossentropy',
    optimizer=optimizer,
    metrics=['AUC']
)

# Mostrar el resumen del modelo
model.summary()

# Restaurar EarlyStopping y agregar ModelCheckpoint para guardar los mejores pesos
early_stopping = EarlyStopping(
    monitor='val_loss',
    patience=10,
    restore_best_weights=True
)

# Guardar los mejores pesos del modelo
checkpoint = ModelCheckpoint(
    filepath='best_model.weights.h5', # Cambiar el nombre del archivo a .weights.h5
    monitor='val_loss',
    save_best_only=True,
    save_weights_only=True
)

# Entrenar el modelo con EarlyStopping y ModelCheckpoint
history = model.fit(
    features_train,
    targets_train,
    epochs=1000, # Aumentar significativamente el número de épocas
    batch_size=64, # Mantener el tamaño del batch
    validation_data=(features_valid, targets_valid),
    verbose=1,
    callbacks=[early_stopping, checkpoint] # Aplicar callbacks
)

# Cargar los mejores pesos del modelo
model.load_weights('best_model.weights.h5')

# Obtener predicciones
```

```
probabilities = model.predict(features_valid).flatten()

# Convertir probabilidades a predicciones binarias
predictions = (probabilities > 0.5).astype(int)

# Calcular AUC-ROC usando las probabilidades
auc_roc = roc_auc_score(targets_valid, probabilities)
print(f'AUC-ROC: {auc_roc:.2f}')

# Calcular Accuracy usando las predicciones binarias
accuracy = accuracy_score(targets_valid, predictions)
print(f'Accuracy: {accuracy:.2f}')
```

C:\Users\evolu\AppData\Roaming\Python\Python312\site-packages\keras\src\layers\core\dense.py:87: UserWarning:

Do not pass an `input\_shape`/`input\_dim` argument to a layer. When using Sequential models, prefer using an `Input(shape)` object as the first layer in the model instead.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 512)	24,064
batch_normalization (BatchNormalization)	(None, 512)	2,048
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131,328
batch_normalization_1 (BatchNormalization)	(None, 256)	1,024
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32,896
batch_normalization_2 (BatchNormalization)	(None, 128)	512
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8,256
batch_normalization_3 (BatchNormalization)	(None, 64)	256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 1)	65

Total params: 200,449 (783.00 KB)

Trainable params: 198,529 (775.50 KB)

**Non-trainable params:** 1,920 (7.50 KB)

Epoch 1/1000

[illegible]

Epoch 2/1000

```

1/83 _____ 2s 29ms/step - AUC: 0.6752 - loss: _____
7.4005_____14/83
_____ 0s 4ms/step - AUC: 0.6020 - loss: 7.4841
_____27/83 _____ 0s
4ms/step - AUC: 0.6032 - loss:
7.4881_____40/83
_____ 0s 4ms/step - AUC: 0.6040 - loss:
7.4859_____53/83
_____ 0s 4ms/step - AUC: 0.6056 - loss:
7.4795_____67/83
_____ 0s 4ms/step - AUC: 0.6090 - loss:
7.4694_____80/83
_____ 0s 4ms/step - AUC: 0.6129 - loss:
7.4592_____83/83
_____ 0s 5ms/step - AUC: 0.6139 - loss: 7.4563 - val_AUC: 0.7954 - val_loss:
7.0402

```

Epoch 3/1000

```
1/83 ----- 2s 29ms/step - AUC: 0.5250 - loss:
7.4691-----12/83
----- 0s 5ms/step - AUC: 0.5956 - loss: 7.3738
-----21/83 ----- 0s
6ms/step - AUC: 0.6096 - loss:
7.3542-----35/83
----- 0s 5ms/step - AUC: 0.6254 - loss:
7.3266-----47/83
----- 0s 5ms/step - AUC: 0.6327 - loss:
7.3090-----59/83
----- 0s 5ms/step - AUC: 0.6367 - loss:
7.2963-----72/83
```

```
----- 0s 4ms/step - AUC: 0.6395 - loss:
7.2849-----83/83
----- 0s 6ms/step - AUC: 0.6418 - loss: 7.2748 - val_AUC: 0.8234 - val_loss:
6.8367
Epoch 4/1000
 1/83 ----- 2s 30ms/step - AUC: 0.7348 - loss:
6.9340-----14/83
----- 0s 4ms/step - AUC: 0.6878 - loss: 7.0260
-----27/83 ----- 0s
4ms/step - AUC: 0.6726 - loss:
7.0477-----41/83
----- 0s 4ms/step - AUC: 0.6733 - loss:
7.0440-----54/83
----- 0s 4ms/step - AUC: 0.6742 - loss:
7.0390-----65/83
----- 0s 4ms/step - AUC: 0.6751 - loss:
7.0340-----76/83
----- 0s 4ms/step - AUC: 0.6762 - loss:
7.0287-----83/83
----- 0s 5ms/step - AUC: 0.6766 - loss: 7.0251 - val_AUC: 0.8346 - val_loss:
6.6339
Epoch 5/1000
 1/83 ----- 2s 28ms/step - AUC: 0.6471 - loss:
6.8994-----14/83
----- 0s 4ms/step - AUC: 0.6879 - loss: 6.9068
-----25/83 ----- 0s
4ms/step - AUC: 0.6917 - loss:
6.8921-----37/83
----- 0s 4ms/step - AUC: 0.6923 - loss:
6.8827-----50/83
----- 0s 4ms/step - AUC: 0.6930 - loss:
6.8753-----61/83
----- 0s 4ms/step - AUC: 0.6931 - loss:
6.8685-----74/83
----- 0s 4ms/step - AUC: 0.6932 - loss:
6.8611-----83/83
----- 0s 6ms/step - AUC: 0.6932 - loss: 6.8557 - val_AUC: 0.8402 - val_loss:
6.4453
Epoch 6/1000
 1/83 ----- 2s 31ms/step - AUC: 0.7143 - loss:
6.6743-----13/83
----- 0s 4ms/step - AUC: 0.6833 - loss: 6.7267
-----27/83 ----- 0s
4ms/step - AUC: 0.6877 - loss:
6.7167-----40/83
----- 0s 4ms/step - AUC: 0.6904 - loss:
6.7073-----54/83
----- 0s 4ms/step - AUC: 0.6937 - loss:
6.6957-----68/83
----- 0s 4ms/step - AUC: 0.6949 - loss:
6.6860-----82/83
```



```
----- 0s 4ms/step - AUC: 0.6956 - loss:
6.6769-----83/83
----- 0s 5ms/step - AUC: 0.6957 - loss: 6.6756 - val_AUC: 0.8426 - val_loss:
6.2739
Epoch 7/1000
 1/83 ----- 2s 29ms/step - AUC: 0.7472 - loss:
6.5282-----13/83
----- 0s 5ms/step - AUC: 0.7382 - loss: 6.5106
-----25/83 ----- 0s
4ms/step - AUC: 0.7332 - loss:
6.4997-----36/83
----- 0s 5ms/step - AUC: 0.7312 - loss:
6.4924-----49/83
----- 0s 4ms/step - AUC: 0.7292 - loss:
6.4848-----61/83
----- 0s 4ms/step - AUC: 0.7284 - loss:
6.4780-----75/83
----- 0s 4ms/step - AUC: 0.7281 - loss:
6.4698-----83/83
----- 0s 5ms/step - AUC: 0.7277 - loss: 6.4649 - val_AUC: 0.8447 - val_loss:
6.1149
Epoch 8/1000
 1/83 ----- 2s 31ms/step - AUC: 0.6729 - loss:
6.4470-----13/83
----- 0s 4ms/step - AUC: 0.7137 - loss: 6.3605
-----26/83 ----- 0s
4ms/step - AUC: 0.7130 - loss:
6.3475-----39/83
----- 0s 4ms/step - AUC: 0.7134 - loss:
6.3358-----53/83
----- 0s 4ms/step - AUC: 0.7145 - loss:
6.3249-----67/83
----- 0s 4ms/step - AUC: 0.7156 - loss:
6.3155-----77/83
----- 0s 4ms/step - AUC: 0.7167 - loss:
6.3089-----83/83
----- 0s 6ms/step - AUC: 0.7172 - loss: 6.3045 - val_AUC: 0.8463 - val_loss:
5.9592
Epoch 9/1000
 1/83 ----- 2s 31ms/step - AUC: 0.6930 - loss:
6.2758-----9/83
----- 0s 6ms/step - AUC: 0.7185 - loss: 6.1965
-----18/83 ----- 0s
6ms/step - AUC: 0.7154 - loss:
6.1865-----28/83
----- 0s 6ms/step - AUC: 0.7102 - loss:
6.1826-----39/83
----- 0s 6ms/step - AUC: 0.7092 - loss:
6.1751-----48/83
----- 0s 6ms/step - AUC: 0.7099 - loss:
6.1685-----58/83
```

```
----- 0s 5ms/step - AUC: 0.7102 - loss:
6.1617-----68/83
----- 0s 5ms/step - AUC: 0.7107 - loss:
6.1547-----78/83
----- 0s 5ms/step - AUC: 0.7110 - loss:
6.1480-----83/83
----- 1s 7ms/step - AUC: 0.7112 - loss: 6.1442 - val_AUC: 0.8478 - val_loss:
5.8061
Epoch 10/1000
  1/83 ----- 2s 27ms/step - AUC: 0.7403 - loss:
6.0279-----14/83
----- 0s 4ms/step - AUC: 0.7325 - loss: 6.0066
----- 0s
4ms/step - AUC: 0.7340 - loss:
5.9944-----40/83
----- 0s 4ms/step - AUC: 0.7342 - loss:
5.9828-----53/83
----- 0s 4ms/step - AUC: 0.7343 - loss:
5.9739-----64/83
----- 0s 4ms/step - AUC: 0.7336 - loss:
5.9671-----76/83
----- 0s 4ms/step - AUC: 0.7331 - loss:
5.9601-----83/83
----- 0s 5ms/step - AUC: 0.7329 - loss: 5.9554 - val_AUC: 0.8493 - val_loss:
5.6487
Epoch 11/1000
  1/83 ----- 2s 30ms/step - AUC: 0.7564 - loss:
5.8300-----13/83
----- 0s 4ms/step - AUC: 0.7256 - loss: 5.8363
----- 0s
6ms/step - AUC: 0.7256 - loss:
5.8331-----32/83
----- 0s 5ms/step - AUC: 0.7287 - loss:
5.8244-----46/83
----- 0s 5ms/step - AUC: 0.7298 - loss:
5.8175-----60/83
----- 0s 5ms/step - AUC: 0.7309 - loss:
5.8092-----71/83
----- 0s 5ms/step - AUC: 0.7314 - loss:
5.8027-----83/83
----- 0s 6ms/step - AUC: 0.7319 - loss: 5.7951 - val_AUC: 0.8498 - val_loss:
5.4943
Epoch 12/1000
  1/83 ----- 2s 28ms/step - AUC: 0.6595 - loss:
5.8223-----12/83
----- 0s 5ms/step - AUC: 0.7343 - loss: 5.6610
----- 0s
5ms/step - AUC: 0.7463 - loss:
5.6421-----35/83
----- 0s 5ms/step - AUC: 0.7509 - loss:
5.6301-----49/83
```

```
----- 0s 4ms/step - AUC: 0.7515 - loss:
5.6212????????????????????????????????????????????????????????????????63/83
----- 0s 4ms/step - AUC: 0.7502 - loss:
5.6151????????????????????????????????????????????????????????????????77/83
----- 0s 4ms/step - AUC: 0.7495 - loss:
5.6087????????????????????????????????????????????????????????????????83/83
----- 0s 5ms/step - AUC: 0.7492 - loss: 5.6055 - val_AUC: 0.8517 - val_loss:
5.3465
Epoch 13/1000
 1/83 ----- 2s 28ms/step - AUC: 0.8704 - loss:
5.3574????????????????????????????????????????????????????????????????14/83
----- 0s 4ms/step - AUC: 0.7612 - loss: 5.4725
????????????????????????????????????????????????????????????????28/83 ----- 0s
4ms/step - AUC: 0.7554 - loss:
5.4740????????????????????????????????????????????????????????????????40/83
----- 0s 4ms/step - AUC: 0.7506 - loss:
5.4757????????????????????????????????????????????????????????????????50/83
----- 0s 4ms/step - AUC: 0.7472 - loss:
5.4753????????????????????????????????????????????????????????????????61/83
----- 0s 4ms/step - AUC: 0.7457 - loss:
5.4719????????????????????????????????????????????????????????????????72/83
----- 0s 4ms/step - AUC: 0.7446 - loss:
5.4683????????????????????????????????????????????????????????????????83/83
----- 0s 6ms/step - AUC: 0.7437 - loss: 5.4643 - val_AUC: 0.8523 - val_loss:
5.1961
Epoch 14/1000
 1/83 ----- 2s 32ms/step - AUC: 0.8035 - loss:
5.2101????????????????????????????????????????????????????????????????14/83
----- 0s 4ms/step - AUC: 0.7663 - loss: 5.2900
????????????????????????????????????????????????????????????????28/83 ----- 0s
4ms/step - AUC: 0.7605 - loss:
5.3024????????????????????????????????????????????????????????????????41/83
----- 0s 4ms/step - AUC: 0.7554 - loss:
5.3061????????????????????????????????????????????????????????????????55/83
----- 0s 4ms/step - AUC: 0.7500 - loss:
5.3088????????????????????????????????????????????????????????????????69/83
----- 0s 4ms/step - AUC: 0.7469 - loss:
5.3075????????????????????????????????????????????????????????????????81/83
----- 0s 4ms/step - AUC: 0.7455 - loss:
5.3043????????????????????????????????????????????????????????????????83/83
----- 0s 5ms/step - AUC: 0.7453 - loss: 5.3034 - val_AUC: 0.8542 - val_loss:
5.0523
Epoch 15/1000
 1/83 ----- 2s 28ms/step - AUC: 0.8456 - loss:
5.0626????????????????????????????????????????????????????????????????14/83
----- 0s 4ms/step - AUC: 0.7387 - loss: 5.2008
????????????????????????????????????????????????????????????????28/83 ----- 0s
4ms/step - AUC: 0.7347 - loss:
5.2016????????????????????????????????????????????????????????????????41/83
----- 0s 4ms/step - AUC: 0.7368 - loss:
5.1939????????????????????????????????????????????????????????????????54/83
```

```
----- 0s 4ms/step - AUC: 0.7377 - loss:
5.1875-----63/83
----- 0s 4ms/step - AUC: 0.7384 - loss:
5.1828-----77/83
----- 0s 4ms/step - AUC: 0.7396 - loss:
5.1748-----83/83
----- 0s 5ms/step - AUC: 0.7401 - loss: 5.1709 - val_AUC: 0.8549 - val_loss:
4.9110
Epoch 16/1000
 1/83 ----- 2s 31ms/step - AUC: 0.5932 - loss:
5.2732-----13/83
----- 0s 4ms/step - AUC: 0.7308 - loss: 5.0930
----- 0s
----- 0s 5ms/step - AUC: 0.7386 - loss:
5.0673-----37/83
----- 0s 4ms/step - AUC: 0.7443 - loss:
5.0518-----50/83
----- 0s 4ms/step - AUC: 0.7464 - loss:
5.0429-----65/83
----- 0s 4ms/step - AUC: 0.7472 - loss:
5.0347-----79/83
----- 0s 4ms/step - AUC: 0.7469 - loss:
5.0281-----83/83
----- 0s 5ms/step - AUC: 0.7469 - loss: 5.0257 - val_AUC: 0.8547 - val_loss:
4.7712
Epoch 17/1000
 1/83 ----- 2s 26ms/step - AUC: 0.6931 - loss:
4.9747-----14/83
----- 0s 4ms/step - AUC: 0.7175 - loss: 4.9455
----- 0s
----- 0s 4ms/step - AUC: 0.7250 - loss:
4.9287-----39/83
----- 0s 4ms/step - AUC: 0.7278 - loss:
4.9181-----53/83
----- 0s 4ms/step - AUC: 0.7292 - loss:
4.9098-----64/83
----- 0s 4ms/step - AUC: 0.7300 - loss:
4.9044-----76/83
----- 0s 4ms/step - AUC: 0.7315 - loss:
4.8977-----83/83
----- 0s 5ms/step - AUC: 0.7327 - loss: 4.8927 - val_AUC: 0.8555 - val_loss:
4.6367
Epoch 18/1000
 1/83 ----- 2s 31ms/step - AUC: 0.8420 - loss:
4.6739-----12/83
----- 0s 5ms/step - AUC: 0.7886 - loss: 4.7222
----- 0s
----- 0s 4ms/step - AUC: 0.7762 - loss:
4.7280-----38/83
----- 0s 4ms/step - AUC: 0.7700 - loss:
4.7288-----50/83
```

```
----- 0s 4ms/step - AUC: 0.7671 - loss:
4.7266-----63/83
----- 0s 4ms/step - AUC: 0.7651 - loss:
4.7234-----77/83
----- 0s 4ms/step - AUC: 0.7637 - loss:
4.7194-----83/83
----- 0s 5ms/step - AUC: 0.7631 - loss: 4.7173 - val_AUC: 0.8568 - val_loss:
4.5005
Epoch 19/1000
  1/83 ----- 2s 29ms/step - AUC: 0.6528 - loss:
4.7538-----14/83
----- 0s 4ms/step - AUC: 0.7304 - loss: 4.6567
----- 0s
----- 0s 4ms/step - AUC: 0.7374 - loss:
4.6399-----42/83
----- 0s 4ms/step - AUC: 0.7415 - loss:
4.6284-----56/83
----- 0s 4ms/step - AUC: 0.7437 - loss:
4.6195-----69/83
----- 0s 4ms/step - AUC: 0.7450 - loss:
4.6124-----80/83
----- 0s 4ms/step - AUC: 0.7460 - loss:
4.6068-----83/83
----- 0s 5ms/step - AUC: 0.7464 - loss: 4.6049 - val_AUC: 0.8574 - val_loss:
4.3713
Epoch 20/1000
  1/83 ----- 2s 27ms/step - AUC: 0.8275 - loss:
4.4511-----14/83
----- 0s 4ms/step - AUC: 0.7873 - loss: 4.4552
----- 0s
----- 0s 4ms/step - AUC: 0.7821 - loss:
4.4611-----41/83
----- 0s 4ms/step - AUC: 0.7780 - loss:
4.4622-----55/83
----- 0s 4ms/step - AUC: 0.7759 - loss:
4.4605-----68/83
----- 0s 4ms/step - AUC: 0.7748 - loss:
4.4572-----81/83
----- 0s 4ms/step - AUC: 0.7734 - loss:
4.4542-----83/83
----- 0s 5ms/step - AUC: 0.7731 - loss: 4.4535 - val_AUC: 0.8572 - val_loss:
4.2457
Epoch 21/1000
  1/83 ----- 2s 27ms/step - AUC: 0.7453 - loss:
4.3298-----15/83
----- 0s 4ms/step - AUC: 0.7766 - loss: 4.3336
----- 0s
----- 0s 4ms/step - AUC: 0.7755 - loss:
4.3328-----42/83
----- 0s 4ms/step - AUC: 0.7728 - loss:
4.3323-----56/83
```

```
----- 0s 4ms/step - AUC: 0.7702 - loss:
4.3316----- 0s 4ms/step - AUC: 0.7683 - loss:
4.3289----- 0s 5ms/step - AUC: 0.7672 - loss: 4.3254 - val_AUC: 0.8566 - val_loss:
4.1243
Epoch 22/1000
 1/83 ----- 2s 31ms/step - AUC: 0.8047 - loss:
4.1922----- 0s 5ms/step - AUC: 0.7942 - loss: 4.2046
----- 0s 5ms/step - AUC: 0.7918 - loss:
4.2029----- 0s 5ms/step - AUC: 0.7932 - loss:
4.1965----- 0s 5ms/step - AUC: 0.7930 - loss:
4.1914----- 0s 4ms/step - AUC: 0.7919 - loss:
4.1881----- 0s 4ms/step - AUC: 0.7904 - loss:
4.1860----- 0s 5ms/step - AUC: 0.7887 - loss:
4.1841----- 1s 6ms/step - AUC: 0.7883 - loss: 4.1834 - val_AUC: 0.8573 - val_loss:
4.0025
Epoch 23/1000
 1/83 ----- 2s 28ms/step - AUC: 0.8920 - loss:
3.9787----- 0s 4ms/step - AUC: 0.7827 - loss: 4.0964
----- 0s 4ms/step - AUC: 0.7827 - loss:
4.0883----- 0s 4ms/step - AUC: 0.7832 - loss:
4.0817----- 0s 4ms/step - AUC: 0.7830 - loss:
4.0769----- 0s 4ms/step - AUC: 0.7834 - loss:
4.0719----- 0s 4ms/step - AUC: 0.7833 - loss:
4.0675----- 0s 5ms/step - AUC: 0.7833 - loss: 4.0672 - val_AUC: 0.8578 - val_loss:
3.8838
Epoch 24/1000
 1/83 ----- 2s 27ms/step - AUC: 0.9186 - loss:
3.9015----- 0s 5ms/step - AUC: 0.8133 - loss: 3.9643
----- 0s 5ms/step - AUC: 0.7949 - loss:
3.9748----- 0s 4ms/step - AUC: 0.7865 - loss:
3.9778-----
```

```
----- 0s 5ms/step - AUC: 0.7844 - loss:
3.9760-----57/83
----- 0s 5ms/step - AUC: 0.7824 - loss:
3.9727-----68/83
----- 0s 5ms/step - AUC: 0.7816 - loss:
3.9682-----79/83
----- 0s 5ms/step - AUC: 0.7804 - loss:
3.9645-----83/83
----- 1s 6ms/step - AUC: 0.7800 - loss: 3.9628 - val_AUC: 0.8590 - val_loss:
3.7687
Epoch 25/1000
 1/83 ----- 2s 27ms/step - AUC: 0.8374 - loss:
3.7920-----12/83
----- 0s 5ms/step - AUC: 0.7932 - loss: 3.8543
----- 0s
----- 5ms/step - AUC: 0.7913 - loss:
3.8541-----35/83
----- 0s 5ms/step - AUC: 0.7851 - loss:
3.8588-----48/83
----- 0s 5ms/step - AUC: 0.7816 - loss:
3.8591-----62/83
----- 0s 4ms/step - AUC: 0.7792 - loss:
3.8568-----76/83
----- 0s 4ms/step - AUC: 0.7780 - loss:
3.8535-----83/83
----- 0s 5ms/step - AUC: 0.7774 - loss: 3.8515 - val_AUC: 0.8590 - val_loss:
3.6570
Epoch 26/1000
 1/83 ----- 2s 30ms/step - AUC: 0.8442 - loss:
3.7295-----12/83
----- 0s 5ms/step - AUC: 0.7962 - loss: 3.7473
----- 0s
----- 5ms/step - AUC: 0.7945 - loss:
3.7421-----36/83
----- 0s 4ms/step - AUC: 0.7936 - loss:
3.7366-----49/83
----- 0s 4ms/step - AUC: 0.7929 - loss:
3.7324-----63/83
----- 0s 4ms/step - AUC: 0.7919 - loss:
3.7278-----77/83
----- 0s 4ms/step - AUC: 0.7913 - loss:
3.7233-----83/83
----- 0s 5ms/step - AUC: 0.7911 - loss: 3.7211 - val_AUC: 0.8600 - val_loss:
3.5471
Epoch 27/1000
 1/83 ----- 2s 29ms/step - AUC: 0.8576 - loss:
3.5274-----12/83
----- 0s 5ms/step - AUC: 0.7764 - loss: 3.6410
----- 0s
----- 5ms/step - AUC: 0.7811 - loss:
3.6412-----28/83
```

```
----- 0s 6ms/step - AUC: 0.7825 - loss:
3.6388-----37/83
----- 0s 6ms/step - AUC: 0.7827 - loss:
3.6364-----50/83
----- 0s 5ms/step - AUC: 0.7836 - loss:
3.6314-----64/83
----- 0s 5ms/step - AUC: 0.7850 - loss:
3.6255-----78/83
----- 0s 5ms/step - AUC: 0.7852 - loss:
3.6209-----83/83
----- 1s 6ms/step - AUC: 0.7852 - loss: 3.6190 - val_AUC: 0.8605 - val_loss:
3.4412
Epoch 28/1000
 1/83 ----- 2s 27ms/step - AUC: 0.7414 - loss:
3.5288-----15/83
----- 0s 4ms/step - AUC: 0.7795 - loss: 3.5409
----- 0s
----- 0s 4ms/step - AUC: 0.7824 - loss:
3.5360-----41/83
----- 0s 4ms/step - AUC: 0.7843 - loss:
3.5297-----54/83
----- 0s 4ms/step - AUC: 0.7851 - loss:
3.5241-----64/83
----- 0s 4ms/step - AUC: 0.7861 - loss:
3.5198-----72/83
----- 0s 4ms/step - AUC: 0.7870 - loss:
3.5165-----83/83
----- 0s 4ms/step - AUC: 0.7879 - loss:
3.5123-----83/83
----- 0s 6ms/step - AUC: 0.7879 - loss: 3.5119 - val_AUC: 0.8600 - val_loss:
3.3393
Epoch 29/1000
 1/83 ----- 2s 26ms/step - AUC: 0.7757 - loss:
3.4479-----14/83
----- 0s 4ms/step - AUC: 0.7795 - loss: 3.4460
----- 0s
----- 0s 4ms/step - AUC: 0.7783 - loss:
3.4424-----41/83
----- 0s 4ms/step - AUC: 0.7776 - loss:
3.4384-----54/83
----- 0s 4ms/step - AUC: 0.7782 - loss:
3.4332-----66/83
----- 0s 4ms/step - AUC: 0.7797 - loss:
3.4277-----80/83
----- 0s 4ms/step - AUC: 0.7811 - loss:
3.4213-----83/83
----- 0s 5ms/step - AUC: 0.7815 - loss: 3.4194 - val_AUC: 0.8581 - val_loss:
3.2425
Epoch 30/1000
 1/83 ----- 2s 28ms/step - AUC: 0.8861 - loss:
3.2117-----14/83
```



```

0s 4ms/step - AUC: 0.7995 - loss: 3.3194
28/83 ----- 0s
4ms/step - AUC: 0.7872 - loss:
3.3262----- 0s 4ms/step - AUC: 0.7871 - loss:
3.3240----- 0s 4ms/step - AUC: 0.7872 - loss:
3.3210----- 0s 4ms/step - AUC: 0.7879 - loss:
3.3168----- 0s 4ms/step - AUC: 0.7888 - loss:
3.3125----- 0s 5ms/step - AUC: 0.7898 - loss: 3.3083 - val_AUC: 0.8602 - val_loss:
3.1448
Epoch 31/1000
 1/83 ----- 2s 32ms/step - AUC: 0.8421 - loss:
3.1751----- 0s 6ms/step - AUC: 0.8258 - loss: 3.1751
21/83 ----- 0s
5ms/step - AUC: 0.8158 - loss:
3.1893----- 0s 5ms/step - AUC: 0.8132 - loss:
3.1915----- 0s 4ms/step - AUC: 0.8105 - loss:
3.1938----- 0s 4ms/step - AUC: 0.8090 - loss:
3.1928----- 0s 4ms/step - AUC: 0.8083 - loss:
3.1908----- 0s 5ms/step - AUC: 0.8080 - loss:
3.1888----- 1s 6ms/step - AUC: 0.8079 - loss: 3.1884 - val_AUC: 0.8595 - val_loss:
3.0499
Epoch 32/1000
 1/83 ----- 2s 34ms/step - AUC: 0.8860 - loss:
3.0425----- 0s 5ms/step - AUC: 0.8401 - loss: 3.0999
23/83 ----- 0s
5ms/step - AUC: 0.8253 - loss:
3.1122----- 0s 4ms/step - AUC: 0.8161 - loss:
3.1167----- 0s 4ms/step - AUC: 0.8113 - loss:
3.1154----- 0s 4ms/step - AUC: 0.8089 - loss:
3.1126----- 0s 4ms/step - AUC: 0.8073 - loss:
3.1095----- 0s 5ms/step - AUC: 0.8067 - loss: 3.1082 - val_AUC: 0.8611 - val_loss:
2.9576
Epoch 33/1000

```

50/99

```
1/83 ----- 2s 28ms/step - AUC: 0.8123 - loss:
2.8005----- 0s 5ms/step - AUC: 0.8238 - loss: 2.7374
----- 0s
5ms/step - AUC: 0.8203 - loss:
2.7442----- 0s 5ms/step - AUC: 0.8185 - loss:
----- 0s 5ms/step - AUC: 0.8159 - loss:
2.7506----- 0s 5ms/step - AUC: 0.8144 - loss:
----- 0s 5ms/step - AUC: 0.8135 - loss:
2.7502----- 0s 5ms/step - AUC: 0.8130 - loss:
----- 1s 6ms/step - AUC: 0.8128 - loss: 2.7468 - val_AUC: 0.8616 - val_loss:
2.6168
Epoch 37/1000
1/83 ----- 2s 29ms/step - AUC: 0.7474 - loss:
2.7720----- 0s 4ms/step - AUC: 0.8021 - loss: 2.7081
----- 0s
5ms/step - AUC: 0.8022 - loss:
2.7037----- 0s 6ms/step - AUC: 0.8030 - loss:
----- 0s 6ms/step - AUC: 0.8037 - loss:
2.7001----- 0s 5ms/step - AUC: 0.8052 - loss:
----- 0s 5ms/step - AUC: 0.8058 - loss:
2.6896----- 0s 5ms/step - AUC: 0.8058 - loss:
----- 1s 6ms/step - AUC: 0.8058 - loss: 2.6767 - val_AUC: 0.8610 - val_loss:
2.5387
Epoch 38/1000
1/83 ----- 2s 28ms/step - AUC: 0.8977 - loss:
2.5399----- 0s 4ms/step - AUC: 0.8055 - loss: 2.6365
----- 0s
4ms/step - AUC: 0.8047 - loss:
2.6252----- 0s 4ms/step - AUC: 0.8087 - loss:
----- 0s 4ms/step - AUC: 0.8120 - loss:
2.6124----- 0s 4ms/step - AUC: 0.8136 - loss:
----- 0s 4ms/step - AUC: 0.8137 - loss:
```

52/99

53/99

[illegible]

55/99

[illegible]



```
1.6410██████████ ██████████ 0s 4ms/step - AUC: 0.8372 - loss:
1.6405██████████ ██████████ 0s 4ms/step - AUC: 0.8376 - loss:
1.6391██████████ ██████████ 0s 5ms/step - AUC: 0.8379 - loss: 1.6377 - val_AUC: 0.8674 - val_loss:
1.5559
Epoch 54/1000
1/83 ██████████ 2s 30ms/step - AUC: 0.8145 - loss:
1.6529██████████ ██████████ 0s 6ms/step - AUC: 0.8261 - loss: 1.6212
██████████ ██████████ 21/83 ██████████ 0s
5ms/step - AUC: 0.8320 - loss:
1.6084██████████ ██████████ 0s 5ms/step - AUC: 0.8338 - loss:
1.6034██████████ ██████████ 0s 5ms/step - AUC: 0.8360 - loss:
1.5987██████████ ██████████ 0s 5ms/step - AUC: 0.8374 - loss:
1.5952██████████ ██████████ 0s 5ms/step - AUC: 0.8384 - loss:
1.5922██████████ ██████████ 0s 5ms/step - AUC: 0.8389 - loss:
1.5898██████████ ██████████ 1s 6ms/step - AUC: 0.8390 - loss: 1.5891 - val_AUC: 0.8683 - val_loss:
1.5084
Epoch 55/1000
1/83 ██████████ 2s 27ms/step - AUC: 0.8385 - loss:
1.5886██████████ ██████████ 0s 4ms/step - AUC: 0.8448 - loss: 1.5423
██████████ ██████████ 27/83 ██████████ 0s
4ms/step - AUC: 0.8404 - loss:
1.5462██████████ ██████████ 0s 4ms/step - AUC: 0.8371 - loss:
1.5488██████████ ██████████ 0s 4ms/step - AUC: 0.8353 - loss:
1.5496██████████ ██████████ 0s 4ms/step - AUC: 0.8342 - loss:
1.5496██████████ ██████████ 0s 4ms/step - AUC: 0.8343 - loss:
1.5480██████████ ██████████ 0s 5ms/step - AUC: 0.8344 - loss: 1.5475 - val_AUC: 0.8695 - val_loss:
1.4646
Epoch 56/1000
1/83 ██████████ 2s 30ms/step - AUC: 0.6947 - loss:
1.7732██████████ ██████████ 0s 5ms/step - AUC: 0.7938 - loss: 1.5763
██████████ ██████████ 21/83 ██████████ 0s
5ms/step - AUC: 0.8146 - loss:
1.5456██████████ ██████████ 0s 5ms/step - AUC: 0.8219 - loss:
```

58/99

127.0.0.1:5500/Experiments/Final project py.html

60/99

```

26/83 ----- 0s
4ms/step - AUC: 0.8438 - loss:
1.1776----- 0s 4ms/step - AUC: 0.8448 - loss:
1.1750----- 0s 4ms/step - AUC: 0.8461 - loss:
1.1720----- 0s 4ms/step - AUC: 0.8472 - loss:
1.1699----- 0s 4ms/step - AUC: 0.8480 - loss:
1.1681----- 0s 5ms/step - AUC: 0.8486 - loss: 1.1669 - val_AUC: 0.8773 - val_loss:
1.1009
Epoch 66/1000
 1/83 ----- 2s 28ms/step - AUC: 0.8054 - loss:
1.1740----- 0s 4ms/step - AUC: 0.8313 - loss: 1.1548
----- 0s
27/83 ----- 0s
4ms/step - AUC: 0.8383 - loss:
1.1512----- 0s 4ms/step - AUC: 0.8433 - loss:
1.1462----- 0s 4ms/step - AUC: 0.8453 - loss:
1.1427----- 0s 4ms/step - AUC: 0.8465 - loss:
1.1402----- 0s 4ms/step - AUC: 0.8475 - loss:
1.1380----- 0s 5ms/step - AUC: 0.8483 - loss: 1.1365 - val_AUC: 0.8772 - val_loss:
1.0736
Epoch 67/1000
 1/83 ----- 2s 28ms/step - AUC: 0.9094 - loss:
1.0375----- 0s 4ms/step - AUC: 0.8572 - loss: 1.1062
----- 0s
29/83 ----- 0s
4ms/step - AUC: 0.8609 - loss:
1.1043----- 0s 4ms/step - AUC: 0.8594 - loss:
1.1050----- 0s 4ms/step - AUC: 0.8583 - loss:
1.1046----- 0s 4ms/step - AUC: 0.8581 - loss:
1.1030----- 0s 4ms/step - AUC: 0.8586 - loss:
1.1009----- 0s 5ms/step - AUC: 0.8591 - loss:
1.0992----- 1s 6ms/step - AUC: 0.8593 - loss: 1.0988 - val_AUC: 0.8786 - val_loss:
1.0431
Epoch 68/1000
 1/83 ----- 2s 29ms/step - AUC: 0.8842 - loss:
```

```
1.04400s 4ms/step - AUC: 0.8649 - loss: 1.0774
24/83 0s
5ms/step - AUC: 0.8636 - loss:
1.072934/83
0s 5ms/step - AUC: 0.8622 - loss:
1.072545/83
0s 5ms/step - AUC: 0.8619 - loss:
1.071356/83
0s 5ms/step - AUC: 0.8611 - loss:
1.071169/83
0s 5ms/step - AUC: 0.8604 - loss:
1.070783/83
0s 4ms/step - AUC: 0.8602 - loss:
1.069583/83
0s 6ms/step - AUC: 0.8602 - loss: 1.0694 - val_AUC: 0.8797 - val_loss:
1.0175
Epoch 69/1000
1/83 2s 28ms/step - AUC: 0.8822 - loss:
1.021912/83
0s 5ms/step - AUC: 0.8724 - loss: 1.0364
23/83 0s
5ms/step - AUC: 0.8675 - loss:
1.043134/83
0s 5ms/step - AUC: 0.8663 - loss:
1.043847/83
0s 4ms/step - AUC: 0.8641 - loss:
1.045661/83
0s 4ms/step - AUC: 0.8630 - loss:
1.046175/83
0s 4ms/step - AUC: 0.8622 - loss:
1.045683/83
0s 5ms/step - AUC: 0.8618 - loss: 1.0452 - val_AUC: 0.8805 - val_loss:
0.9909
Epoch 70/1000
1/83 2s 27ms/step - AUC: 0.9595 - loss:
0.918314/83
0s 4ms/step - AUC: 0.9005 - loss: 0.9730
26/83 0s
4ms/step - AUC: 0.8882 - loss:
0.984637/83
0s 4ms/step - AUC: 0.8825 - loss:
0.991348/83
0s 4ms/step - AUC: 0.8791 - loss:
0.995359/83
0s 4ms/step - AUC: 0.8766 - loss:
0.998472/83
0s 4ms/step - AUC: 0.8744 - loss:
1.000783/83
0s 5ms/step - AUC: 0.8732 - loss: 1.0015 - val_AUC: 0.8810 - val_loss:
0.9653
```

[illegible]

```
----- 0s 5ms/step - AUC: 0.8575 - loss: 0.9490
21/83 ----- 0s
5ms/step - AUC: 0.8651 - loss:
0.9373----- 30/83
----- 0s 5ms/step - AUC: 0.8689 - loss:
0.9323----- 39/83
----- 0s 5ms/step - AUC: 0.8702 - loss:
0.9308----- 48/83
----- 0s 5ms/step - AUC: 0.8715 - loss:
0.9290----- 57/83
----- 0s 5ms/step - AUC: 0.8723 - loss:
0.9279----- 68/83
----- 0s 5ms/step - AUC: 0.8725 - loss:
0.9274----- 78/83
----- 0s 5ms/step - AUC: 0.8723 - loss:
0.9275----- 83/83
----- 1s 7ms/step - AUC: 0.8721 - loss: 0.9277 - val_AUC: 0.8815 - val_loss:
0.8966
Epoch 74/1000
 1/83 ----- 2s 34ms/step - AUC: 0.9275 - loss:
0.8584----- 10/83
----- 0s 6ms/step - AUC: 0.9057 - loss: 0.8840
21/83 ----- 0s
5ms/step - AUC: 0.8953 - loss:
0.8914----- 33/83
----- 0s 5ms/step - AUC: 0.8913 - loss:
0.8930----- 44/83
----- 0s 5ms/step - AUC: 0.8887 - loss:
0.8940----- 54/83
----- 0s 5ms/step - AUC: 0.8864 - loss:
0.8957----- 63/83
----- 0s 5ms/step - AUC: 0.8849 - loss:
0.8968----- 69/83
----- 0s 5ms/step - AUC: 0.8842 - loss:
0.8974----- 79/83
----- 0s 5ms/step - AUC: 0.8830 - loss:
0.8986----- 83/83
----- 1s 7ms/step - AUC: 0.8825 - loss: 0.8989 - val_AUC: 0.8839 - val_loss:
0.8757
Epoch 75/1000
 1/83 ----- 2s 30ms/step - AUC: 0.9545 - loss:
0.7897----- 11/83
----- 0s 5ms/step - AUC: 0.9007 - loss: 0.8640
19/83 ----- 0s
6ms/step - AUC: 0.8913 - loss:
0.8778----- 28/83
----- 0s 6ms/step - AUC: 0.8853 - loss:
0.8866----- 37/83
----- 0s 6ms/step - AUC: 0.8828 - loss:
0.8890----- 46/83
----- 0s 6ms/step - AUC: 0.8803 - loss:
```



[illegible]

```
----- 0s 5ms/step - AUC: 0.8716 - loss:
0.8550????????????????????????????????????????????????????????????83/83
----- 1s 7ms/step - AUC: 0.8717 - loss: 0.8549 - val_AUC: 0.8885 - val_loss:
0.8156
Epoch 78/1000
 1/83 ----- 2s 33ms/step - AUC: 0.8579 - loss:
0.9393????????????????????????????????????????????????????????????11/83
----- 0s 5ms/step - AUC: 0.8761 - loss: 0.8490
????????????????????????????????????????????????????????????22/83 ----- 0s
5ms/step - AUC: 0.8844 - loss:
0.8328????????????????????????????????????????????????????????????33/83
----- 0s 5ms/step - AUC: 0.8849 - loss:
0.8295????????????????????????????????????????????????????????????44/83
----- 0s 5ms/step - AUC: 0.8843 - loss:
0.8282????????????????????????????????????????????????????????????56/83
----- 0s 5ms/step - AUC: 0.8831 - loss:
0.8281????????????????????????????????????????????????????????????67/83
----- 0s 5ms/step - AUC: 0.8828 - loss:
0.8277????????????????????????????????????????????????????????????74/83
----- 0s 5ms/step - AUC: 0.8826 - loss:
0.8275????????????????????????????????????????????????????????????83/83
----- 0s 5ms/step - AUC: 0.8827 - loss:
0.8273????????????????????????????????????????????????????????????83/83
----- 1s 7ms/step - AUC: 0.8827 - loss: 0.8272 - val_AUC: 0.8886 - val_loss:
0.8013
Epoch 79/1000
 1/83 ----- 3s 40ms/step - AUC: 0.8135 - loss:
0.9484????????????????????????????????????????????????????????????11/83
----- 0s 5ms/step - AUC: 0.8949 - loss: 0.8038
????????????????????????????????????????????????????????????21/83 ----- 0s
5ms/step - AUC: 0.8917 - loss:
0.8103????????????????????????????????????????????????????????????30/83
----- 0s 5ms/step - AUC: 0.8872 - loss:
0.8158????????????????????????????????????????????????????????????42/83
----- 0s 5ms/step - AUC: 0.8833 - loss:
0.8192????????????????????????????????????????????????????????????53/83
----- 0s 5ms/step - AUC: 0.8815 - loss:
0.8200????????????????????????????????????????????????????????????65/83
----- 0s 5ms/step - AUC: 0.8803 - loss:
0.8200????????????????????????????????????????????????????????????76/83
----- 0s 5ms/step - AUC: 0.8797 - loss:
0.8197????????????????????????????????????????????????????????????83/83
----- 1s 6ms/step - AUC: 0.8796 - loss: 0.8192 - val_AUC: 0.8902 - val_loss:
0.7801
Epoch 80/1000
 1/83 ----- 2s 28ms/step - AUC: 0.8854 - loss:
0.8162????????????????????????????????????????????????????????????12/83
----- 0s 5ms/step - AUC: 0.8735 - loss: 0.8151
????????????????????????????????????????????????????????????24/83 ----- 0s
5ms/step - AUC: 0.8774 - loss:
0.8065????????????????????????????????????????????????????????????35/83
```

```
----- 0s 5ms/step - AUC: 0.8776 - loss:
0.8071-----44/83
----- 0s 5ms/step - AUC: 0.8778 - loss:
0.8073-----53/83
----- 0s 5ms/step - AUC: 0.8773 - loss:
0.8077-----63/83
----- 0s 5ms/step - AUC: 0.8773 - loss:
0.8072-----75/83
----- 0s 5ms/step - AUC: 0.8774 - loss:
0.8063-----83/83
----- 1s 6ms/step - AUC: 0.8777 - loss: 0.8053 - val_AUC: 0.8900 - val_loss:
0.7634
Epoch 81/1000
  1/83 ----- 2s 29ms/step - AUC: 0.9300 - loss:
0.6981-----12/83
----- 0s 5ms/step - AUC: 0.9163 - loss: 0.7218
----- 0s
----- 0s 5ms/step - AUC: 0.9106 - loss:
0.7339-----33/83
----- 0s 5ms/step - AUC: 0.9061 - loss:
0.7419-----45/83
----- 0s 5ms/step - AUC: 0.9030 - loss:
0.7475-----53/83
----- 0s 5ms/step - AUC: 0.9014 - loss:
0.7502-----64/83
----- 0s 5ms/step - AUC: 0.8999 - loss:
0.7526-----76/83
----- 0s 5ms/step - AUC: 0.8986 - loss:
0.7543-----83/83
----- 1s 6ms/step - AUC: 0.8978 - loss: 0.7554 - val_AUC: 0.8901 - val_loss:
0.7480
Epoch 82/1000
  1/83 ----- 2s 29ms/step - AUC: 0.8833 - loss:
0.8284-----10/83
----- 0s 6ms/step - AUC: 0.8919 - loss: 0.7654
----- 0s
----- 0s 5ms/step - AUC: 0.8923 - loss:
0.7572-----31/83
----- 0s 5ms/step - AUC: 0.8887 - loss:
0.7601-----43/83
----- 0s 5ms/step - AUC: 0.8874 - loss:
0.7602-----56/83
----- 0s 5ms/step - AUC: 0.8874 - loss:
0.7592-----66/83
----- 0s 5ms/step - AUC: 0.8873 - loss:
0.7589-----76/83
----- 0s 5ms/step - AUC: 0.8874 - loss:
0.7583-----83/83
----- 1s 6ms/step - AUC: 0.8874 - loss: 0.7581 - val_AUC: 0.8953 - val_loss:
0.7292
Epoch 83/1000
```

```
1/83 ----- 2s 29ms/step - AUC: 0.8670 - loss:
0.7873----- 0s 4ms/step - AUC: 0.8723 - loss: 0.7720
----- 0s 4ms/step - AUC: 0.8777 - loss:
0.7626----- 0s 4ms/step - AUC: 0.8808 - loss:
0.7574----- 0s 4ms/step - AUC: 0.8817 - loss:
0.7564----- 0s 4ms/step - AUC: 0.8823 - loss:
0.7554----- 0s 4ms/step - AUC: 0.8831 - loss:
0.7541----- 0s 5ms/step - AUC: 0.8833 - loss: 0.7534 - val_AUC: 0.8946 - val_loss:
0.7188
Epoch 84/1000
1/83 ----- 2s 27ms/step - AUC: 0.9004 - loss:
0.7081----- 0s 4ms/step - AUC: 0.8867 - loss: 0.7275
----- 0s 4ms/step - AUC: 0.8854 - loss:
0.7343----- 0s 4ms/step - AUC: 0.8857 - loss:
0.7348----- 0s 5ms/step - AUC: 0.8860 - loss:
0.7339----- 0s 5ms/step - AUC: 0.8862 - loss:
0.7337----- 0s 5ms/step - AUC: 0.8864 - loss:
0.7334----- 0s 5ms/step - AUC: 0.8864 - loss:
0.7332----- 1s 6ms/step - AUC: 0.8865 - loss: 0.7332 - val_AUC: 0.8921 - val_loss:
0.7050
Epoch 85/1000
1/83 ----- 2s 29ms/step - AUC: 0.8859 - loss:
0.7389----- 0s 5ms/step - AUC: 0.8654 - loss: 0.7541
----- 0s 4ms/step - AUC: 0.8753 - loss:
0.7421----- 0s 4ms/step - AUC: 0.8795 - loss:
0.7355----- 0s 4ms/step - AUC: 0.8814 - loss:
0.7324----- 0s 4ms/step - AUC: 0.8825 - loss:
0.7308----- 0s 4ms/step - AUC: 0.8828 - loss:
0.7301----- 0s 6ms/step - AUC: 0.8832 - loss: 0.7289 - val_AUC: 0.8972 - val_loss:
```

———— 0s 4ms/step - AUC: 0.9011 - loss:

70/99

71/99

[illegible]



73/99

74/99

[illegible]

76/99

[illegible]

```
0.48730s 4ms/step - AUC: 0.9318 - loss: 0.4657
0.46880s 4ms/step - AUC: 0.9299 - loss:
0.46830s 4ms/step - AUC: 0.9304 - loss:
0.46770s 4ms/step - AUC: 0.9310 - loss:
0.46750s 4ms/step - AUC: 0.9311 - loss:
0.46700s 5ms/step - AUC: 0.9312 - loss: 0.4667 - val_AUC: 0.9131 - val_loss:
0.4851
Epoch 113/1000
1/83 2s 29ms/step - AUC: 0.8822 - loss:
0.54280s 4ms/step - AUC: 0.9190 - loss: 0.4881
0.47490s 6ms/step - AUC: 0.9280 - loss:
0.46970s 6ms/step - AUC: 0.9300 - loss:
0.46530s 6ms/step - AUC: 0.9319 - loss:
0.46140s 6ms/step - AUC: 0.9332 - loss:
0.45900s 5ms/step - AUC: 0.9339 - loss:
0.45740s 6ms/step - AUC: 0.9341 - loss: 0.4567 - val_AUC: 0.9122 - val_loss:
0.4905
Epoch 114/1000
1/83 2s 26ms/step - AUC: 0.9468 - loss:
0.42800s 4ms/step - AUC: 0.9454 - loss: 0.4391
0.44060s 4ms/step - AUC: 0.9405 - loss:
0.44860s 4ms/step - AUC: 0.9384 - loss:
0.45210s 4ms/step - AUC: 0.9372 - loss:
0.45360s 4ms/step - AUC: 0.9364 - loss:
0.45440s 5ms/step - AUC: 0.9363 - loss: 0.4545 - val_AUC: 0.9109 - val_loss:
0.4857
```

Epoch 116/1000

Epoch 117/1000

[illegible]

```
----- 0s 4ms/step - AUC: 0.9422 - loss:
0.4275-----83/83
----- 0s 5ms/step - AUC: 0.9422 - loss: 0.4275 - val_AUC: 0.9154 - val_loss:
0.4689
Epoch 118/1000
 1/83 ----- 2s 27ms/step - AUC: 0.9479 - loss:
0.3880-----11/83
----- 0s 5ms/step - AUC: 0.9512 - loss: 0.4035
-----19/83 ----- 0s
6ms/step - AUC: 0.9487 - loss:
0.4093-----27/83
----- 0s 6ms/step - AUC: 0.9459 - loss:
0.4148-----36/83
----- 0s 6ms/step - AUC: 0.9437 - loss:
0.4192-----44/83
----- 0s 6ms/step - AUC: 0.9427 - loss:
0.4212-----51/83
----- 0s 6ms/step - AUC: 0.9421 - loss:
0.4225-----57/83
----- 0s 7ms/step - AUC: 0.9419 - loss:
0.4230-----67/83
----- 0s 6ms/step - AUC: 0.9416 - loss:
0.4236-----77/83
----- 0s 6ms/step - AUC: 0.9411 - loss:
0.4247-----83/83
----- 1s 8ms/step - AUC: 0.9408 - loss: 0.4254 - val_AUC: 0.9130 - val_loss:
0.4685
Epoch 119/1000
 1/83 ----- 4s 54ms/step - AUC: 0.9382 - loss:
0.4277----- 9/83
----- 0s 7ms/step - AUC: 0.9255 - loss: 0.4542
-----17/83 ----- 0s
7ms/step - AUC: 0.9269 - loss:
0.4526-----28/83
----- 0s 6ms/step - AUC: 0.9277 - loss:
0.4534-----42/83
----- 0s 5ms/step - AUC: 0.9285 - loss:
0.4526-----56/83
----- 0s 5ms/step - AUC: 0.9296 - loss:
0.4504-----70/83
----- 0s 5ms/step - AUC: 0.9305 - loss:
0.4486-----83/83
----- 0s 4ms/step - AUC: 0.9313 - loss:
0.4469-----83/83
----- 1s 6ms/step - AUC: 0.9313 - loss: 0.4468 - val_AUC: 0.9154 - val_loss:
0.4598
Epoch 120/1000
 1/83 ----- 2s 30ms/step - AUC: 0.9766 - loss:
0.3470-----15/83
----- 0s 4ms/step - AUC: 0.9481 - loss: 0.4011
-----29/83 ----- 0s
```



[illegible]

82/99

83/99

84/99

85/99

86/99

```

16/83 ----- 0s
7ms/step - AUC: 0.9668 - loss:
0.3387----- 24/83
----- 0s 7ms/step - AUC: 0.9659 - loss:
0.3396----- 35/83
----- 0s 6ms/step - AUC: 0.9639 - loss:
0.3432----- 48/83
----- 0s 5ms/step - AUC: 0.9623 - loss:
0.3468----- 62/83
----- 0s 5ms/step - AUC: 0.9610 - loss:
0.3499----- 73/83
----- 0s 5ms/step - AUC: 0.9603 - loss:
0.3515----- 83/83
----- 1s 6ms/step - AUC: 0.9598 - loss: 0.3527 - val_AUC: 0.9225 - val_loss:
0.4242
Epoch 138/1000
 1/83 ----- 3s 38ms/step - AUC: 0.9604 - loss:
0.4439----- 14/83
----- 0s 4ms/step - AUC: 0.9661 - loss: 0.3448
----- 28/83 ----- 0s
4ms/step - AUC: 0.9666 - loss:
0.3388----- 40/83
----- 0s 4ms/step - AUC: 0.9648 - loss:
0.3416----- 52/83
----- 0s 4ms/step - AUC: 0.9631 - loss:
0.3448----- 64/83
----- 0s 4ms/step - AUC: 0.9620 - loss:
0.3468----- 78/83
----- 0s 4ms/step - AUC: 0.9612 - loss:
0.3484----- 83/83
----- 0s 5ms/step - AUC: 0.9609 - loss: 0.3490 - val_AUC: 0.9207 - val_loss:
0.4242
Epoch 139/1000
 1/83 ----- 2s 25ms/step - AUC: 0.9875 - loss:
0.2700----- 15/83
----- 0s 4ms/step - AUC: 0.9629 - loss: 0.3427
----- 29/83 ----- 0s
4ms/step - AUC: 0.9589 - loss:
0.3516----- 43/83
----- 0s 4ms/step - AUC: 0.9578 - loss:
0.3543----- 54/83
----- 0s 4ms/step - AUC: 0.9575 - loss:
0.3552----- 64/83
----- 0s 4ms/step - AUC: 0.9573 - loss:
0.3558----- 76/83
----- 0s 4ms/step - AUC: 0.9572 - loss:
0.3559----- 83/83
----- 0s 5ms/step - AUC: 0.9570 - loss: 0.3563 - val_AUC: 0.9209 - val_loss:
0.4266
Epoch 140/1000
 1/83 ----- 2s 30ms/step - AUC: 0.9898 - loss:
```

[illegible]



```

26/83 ----- 0s
4ms/step - AUC: 0.9603 - loss:
0.341126/83
----- 0s 4ms/step - AUC: 0.9588 - loss:
0.344926/83
----- 0s 4ms/step - AUC: 0.9581 - loss:
0.347626/83
----- 0s 4ms/step - AUC: 0.9575 - loss:
0.349426/83
----- 0s 4ms/step - AUC: 0.9575 - loss:
0.349726/83
----- 0s 5ms/step - AUC: 0.9576 - loss: 0.3495 - val_AUC: 0.9212 - val_loss:
0.4213
Epoch 144/1000
 1/83 ----- 2s 31ms/step - AUC: 0.9537 - loss:
0.388226/83
----- 0s 5ms/step - AUC: 0.9561 - loss: 0.3784
22/83 ----- 0s
5ms/step - AUC: 0.9555 - loss:
0.371326/83
----- 0s 6ms/step - AUC: 0.9555 - loss:
0.368026/83
----- 0s 6ms/step - AUC: 0.9558 - loss:
0.364626/83
----- 0s 6ms/step - AUC: 0.9559 - loss:
0.362226/83
----- 0s 6ms/step - AUC: 0.9560 - loss:
0.360626/83
----- 0s 6ms/step - AUC: 0.9563 - loss:
0.358526/83
----- 0s 5ms/step - AUC: 0.9567 - loss:
0.356426/83
----- 1s 6ms/step - AUC: 0.9567 - loss: 0.3561 - val_AUC: 0.9154 - val_loss:
0.4348
Epoch 145/1000
 1/83 ----- 2s 29ms/step - AUC: 0.9614 - loss:
0.331926/83
----- 0s 4ms/step - AUC: 0.9705 - loss: 0.3138
22/83 ----- 0s
5ms/step - AUC: 0.9696 - loss:
0.316626/83
----- 0s 5ms/step - AUC: 0.9683 - loss:
0.319326/83
----- 0s 5ms/step - AUC: 0.9665 - loss:
0.323826/83
----- 0s 5ms/step - AUC: 0.9649 - loss:
0.327626/83
----- 0s 5ms/step - AUC: 0.9636 - loss:
0.330526/83
----- 0s 5ms/step - AUC: 0.9627 - loss:
0.332426/83
```

```
----- 1s 6ms/step - AUC: 0.9626 - loss: 0.3328 - val_AUC: 0.9209 - val_loss:
0.4198
Epoch 146/1000
 1/83 ----- 2s 33ms/step - AUC: 0.9772 - loss:
0.2978????????????????????????????????????????????????????????????????????????????????????13/83
----- 0s 4ms/step - AUC: 0.9665 - loss: 0.3319
????????????????????????????????????????????????????????????????????????????????????26/83 ----- 0s
4ms/step - AUC: 0.9623 - loss:
0.3378????????????????????????????????????????????????????????????????????????????????????39/83
----- 0s 4ms/step - AUC: 0.9616 - loss:
0.3370????????????????????????????????????????????????????????????????????????????????????52/83
----- 0s 4ms/step - AUC: 0.9606 - loss:
0.3382????????????????????????????????????????????????????????????????????????????????????65/83
----- 0s 4ms/step - AUC: 0.9596 - loss:
0.3399????????????????????????????????????????????????????????????????????????????????????75/83
----- 0s 4ms/step - AUC: 0.9591 - loss:
0.3409????????????????????????????????????????????????????????????????????????????????????83/83
----- 0s 5ms/step - AUC: 0.9589 - loss: 0.3415 - val_AUC: 0.9178 - val_loss:
0.4218
Epoch 147/1000
 1/83 ----- 2s 31ms/step - AUC: 0.9736 - loss:
0.3142????????????????????????????????????????????????????????????????????????????????????11/83
----- 0s 5ms/step - AUC: 0.9652 - loss: 0.3263
????????????????????????????????????????????????????????????????????????????????????21/83 ----- 0s
5ms/step - AUC: 0.9679 - loss:
0.3183????????????????????????????????????????????????????????????????????????????????????30/83
----- 0s 5ms/step - AUC: 0.9672 - loss:
0.3197????????????????????????????????????????????????????????????????????????????????????40/83
----- 0s 5ms/step - AUC: 0.9664 - loss:
0.3222????????????????????????????????????????????????????????????????????????????????????47/83
----- 0s 6ms/step - AUC: 0.9657 - loss:
0.3243????????????????????????????????????????????????????????????????????????????????????54/83
----- 0s 6ms/step - AUC: 0.9652 - loss:
0.3258????????????????????????????????????????????????????????????????????????????????????61/83
----- 0s 6ms/step - AUC: 0.9648 - loss:
0.3272????????????????????????????????????????????????????????????????????????????????????68/83
----- 0s 6ms/step - AUC: 0.9644 - loss:
0.3283????????????????????????????????????????????????????????????????????????????????????75/83
----- 0s 7ms/step - AUC: 0.9641 - loss:
0.3292????????????????????????????????????????????????????????????????????????????????????78/83
----- 0s 7ms/step - AUC: 0.9640 - loss:
0.3295????????????????????????????????????????????????????????????????????????????????????83/83
----- 0s 7ms/step - AUC: 0.9639 - loss:
0.3298????????????????????????????????????????????????????????????????????????????????????83/83
----- 1s 10ms/step - AUC: 0.9639 - loss: 0.3299 - val_AUC: 0.9189 - val_loss:
0.4198
Epoch 148/1000
 1/83 ----- 3s 42ms/step - AUC: 0.9591 - loss:
0.3449???????????????????????????????????????????????????????????????????????????????????? 8/83
----- 0s 8ms/step - AUC: 0.9567 - loss: 0.3494
????????????????????????????????????????????????????????????????????????????????????15/83 ----- 0s
```

\_\_\_\_\_ 0s 5ms/step - AUC: 0.9643 - loss:

92/99

```
0.30500s 4ms/step - AUC: 0.9714 - loss: 0.30490s 4ms/step - AUC: 0.9709 - loss: 0.30600s 5ms/step - AUC: 0.9704 - loss: 0.3072 - val_AUC: 0.9160 - val_loss: 0.4299
Epoch 154/1000
1/83 2s 28ms/step - AUC: 0.9417 - loss: 0.46510s 5ms/step - AUC: 0.9496 - loss: 0.3631
0s 6ms/step - AUC: 0.9546 - loss: 0.34770s 6ms/step - AUC: 0.9572 - loss: 0.34070s 6ms/step - AUC: 0.9587 - loss: 0.33720s 6ms/step - AUC: 0.9603 - loss: 0.33300s 5ms/step - AUC: 0.9614 - loss: 0.33010s 5ms/step - AUC: 0.9616 - loss: 0.32991s 6ms/step - AUC: 0.9617 - loss: 0.3296 - val_AUC: 0.9216 - val_loss: 0.4163
Epoch 155/1000
1/83 2s 28ms/step - AUC: 0.9650 - loss: 0.32560s 4ms/step - AUC: 0.9609 - loss: 0.3328
0s 4ms/step - AUC: 0.9604 - loss: 0.33060s 4ms/step - AUC: 0.9612 - loss: 0.32800s 4ms/step - AUC: 0.9617 - loss: 0.32700s 4ms/step - AUC: 0.9619 - loss: 0.32630s 5ms/step - AUC: 0.9622 - loss: 0.3255 - val_AUC: 0.9200 - val_loss: 0.4140
Epoch 156/1000
1/83 2s 29ms/step - AUC: 0.9955 - loss: 0.22020s 4ms/step - AUC: 0.9728 - loss: 0.2908
0s 4ms/step - AUC: 0.9719 - loss: 0.29520s 4ms/step - AUC: 0.9701 - loss: 0.30080s 4ms/step - AUC: 0.9693 - loss:
```

[illegible]

```

0.30890s 4ms/step - AUC: 0.9713 - loss:
0.31100s 4ms/step - AUC: 0.9698 - loss:
0.31170s 4ms/step - AUC: 0.9691 - loss:
0.31240s 5ms/step - AUC: 0.9685 - loss:
0.3128 - val_AUC: 0.9195 - val_loss:
0.4219
Epoch 160/1000
1/83 2s 27ms/step - AUC: 0.9723 - loss:
0.30750s 4ms/step - AUC: 0.9743 - loss: 0.2965
0.29810s 4ms/step - AUC: 0.9723 - loss:
0.29740s 4ms/step - AUC: 0.9719 - loss:
0.29840s 4ms/step - AUC: 0.9713 - loss:
0.29960s 5ms/step - AUC: 0.9707 - loss: 0.3014 - val_AUC: 0.9196 - val_loss:
0.4224
Epoch 161/1000
1/83 2s 26ms/step - AUC: 1.0000 - loss:
0.20360s 4ms/step - AUC: 0.9857 - loss: 0.2527
0.27210s 4ms/step - AUC: 0.9757 - loss:
0.28370s 4ms/step - AUC: 0.9732 - loss:
0.29110s 4ms/step - AUC: 0.9718 - loss:
0.29550s 4ms/step - AUC: 0.9712 - loss:
0.29730s 5ms/step - AUC: 0.9710 - loss: 0.2980 - val_AUC: 0.9145 - val_loss:
0.4323
Epoch 162/1000
1/83 2s 30ms/step - AUC: 0.9919 - loss:
0.23160s 4ms/step - AUC: 0.9819 - loss: 0.2633
0.27310s 4ms/step - AUC: 0.9755 - loss:

```

96/99



## 5. Resultados

El resultado fue exitoso, se sobrepaso el objetivo más alto que era conseguir un 'AUC-ROC' de más de 0.89 y en 2 modelos se pudo sobrepasar que fueron: **El Claificador por descenso de Gradiente** y **La red neuronal**

- **Métrica principal:** AUC-ROC.
- **Métrica adicional:** Exactitud.

### Criterios de evaluación:

- **AUC-ROC < 0.75** — 0 SP
- **$0.75 \leq \text{AUC-ROC} < 0.81$**  — 4 SP
- **$0.81 \leq \text{AUC-ROC} < 0.85$**  — 4.5 SP
- **$0.85 \leq \text{AUC-ROC} < 0.87$**  — 5 SP
- **$0.87 \leq \text{AUC-ROC} < 0.88$**  — 5.5 SP
- **$\text{AUC-ROC} \geq 0.88$**  — 6 SP

## Regresión Logística:

- **Mejores parámetros encontrados:** {'C': 10, 'l1\_ratio': 1.0, 'max\_iter': 300, 'penalty': 'l1', 'solver': 'liblinear'}
- **Mejor puntuación:** 0.8537
- **AUC-ROC:** 0.87
- **Accuracy:** 0.82

## Clasificador por Bosque Aleatorio:

- **Mejores parámetros encontrados:** {'bootstrap': True, 'max\_depth': 20, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 1000}
  - **Mejor puntuación:** 0.8903
  - **AUC-ROC:** 0.89
  - **Accuracy:** 0.85
- 

## Clasificador por Descenso de Gradiente:

- **Mejores parámetros encontrados:** {'learning\_rate': 0.2, 'max\_depth': 3, 'n\_estimators': 500, 'subsample': 0.8}
  - **Mejor puntuación:** 0.9276
  - **AUC-ROC:** 0.94
  - **Accuracy:** 0.90
- 

## Red Neuronal:

- **Arquitectura:** | Layer (type) | Output Shape | Param # | |-----|-----|  
-----| | **Dense (dense\_75)** | (None, 512) | 23,552 | | **BatchNormalization (batch\_51)** | (None, 512) | 2,048 | | **Dropout (dropout\_51)** | (None, 512) | 0 | | **Dense (dense\_76)** | (None, 256) | 131,328 | | **BatchNormalization (batch\_52)** | (None, 256) | 1,024 | | **Dropout (dropout\_52)** | (None, 256) | 0 | | **Dense (dense\_77)** | (None, 128) | 32,896 | | **BatchNormalization (batch\_53)** | (None, 128) | 512 | | **Dropout (dropout\_53)** | (None, 128) | 0 | | **Dense (dense\_78)** | (None, 64) | 8,256 | | **BatchNormalization (batch\_54)** | (None, 64) | 256 | | **Dropout (dropout\_54)** | (None, 64) | 0 | | **Dense (dense\_79)** | (None, 1) | 65 |
  - **Entrenamiento:**
    - **Epoch 1/1000:** AUC: 0.5244, loss: 7.8016, val\_AUC: 0.7564, val\_loss: 7.4223
    - **Epoch 2/1000:** AUC: 0.6461, loss: 7.4874, val\_AUC: 0.7904, val\_loss: 7.2717
    - **Epoch 3/1000:** AUC: 0.6852, loss: 7.2402, val\_AUC: 0.8100, val\_loss: 7.0595
    - **Epoch 4/1000:** AUC: 0.7044, loss: 7.0418, val\_AUC: 0.8184, val\_loss: 6.8342
    - **Epoch 5/1000:** AUC: 0.7000, loss: 6.8556, val\_AUC: 0.8299, val\_loss: 6.6247 ...
    - **Última Epoch:** AUC: 0.9708, loss: 0.2998, val\_AUC: 0.9294, val\_loss: 0.3969
  - **Resultados:**
    - **AUC-ROC:** 0.93
    - **Accuracy:** 0.89
- 

## Conclusión

---

En este proyecto se realizó un modelo capaz de predecir si un cliente ha terminado su contrato o no además de un análisis exploratorio de datos, se llegaron a las siguientes conclusiones:

- La mayor cantidad de contratos son de pagos mensuales de entre \$18 - \$22 dólares
- Existe una relación entre cantidad de cargos mensuales y totales de los clientes
- Los métodos de pago por transferencia bancaria y tarjeta de crédito tienen cargos mensuales similares, el pago por cheque electrónico tiene cargos concentrados con una mediana alta (\$80.55), y los cheques por correo muestran una gran dispersión, sugiriendo que se usan más para cargos bajos.
- Los clientes con contrato finalizado muestran cargos mensuales más altos y menos variados, con mayores diferencias en los cuartiles comparados con los clientes activos.
- Los clientes que han finalizado su contrato tienden a tener cargos totales más bajos y presentan más anomalías, sugiriendo que podrían ser clientes de corto plazo, mientras que los activos muestran cargos más altos, indicando posible satisfacción a largo plazo.
- La seguridad online es bastante solicitada, y la fibra óptica junto con los servicios de streaming son los más requeridos, seguidos por protección del hardware y apoyo online.

## Conclusiones de los modelos:

En conclusión, el claro ganador es la red neuronal, no solo por su métrica AUC-ROC, sino porque supera significativamente al único competidor, que es el modelo de descenso de gradiente. La red neuronal se entrena mucho más rápido. Sin embargo, es importante mencionar que el modelo de descenso de gradiente también es muy bueno y ambos logran cumplir el umbral objetivo del proyecto.