

---

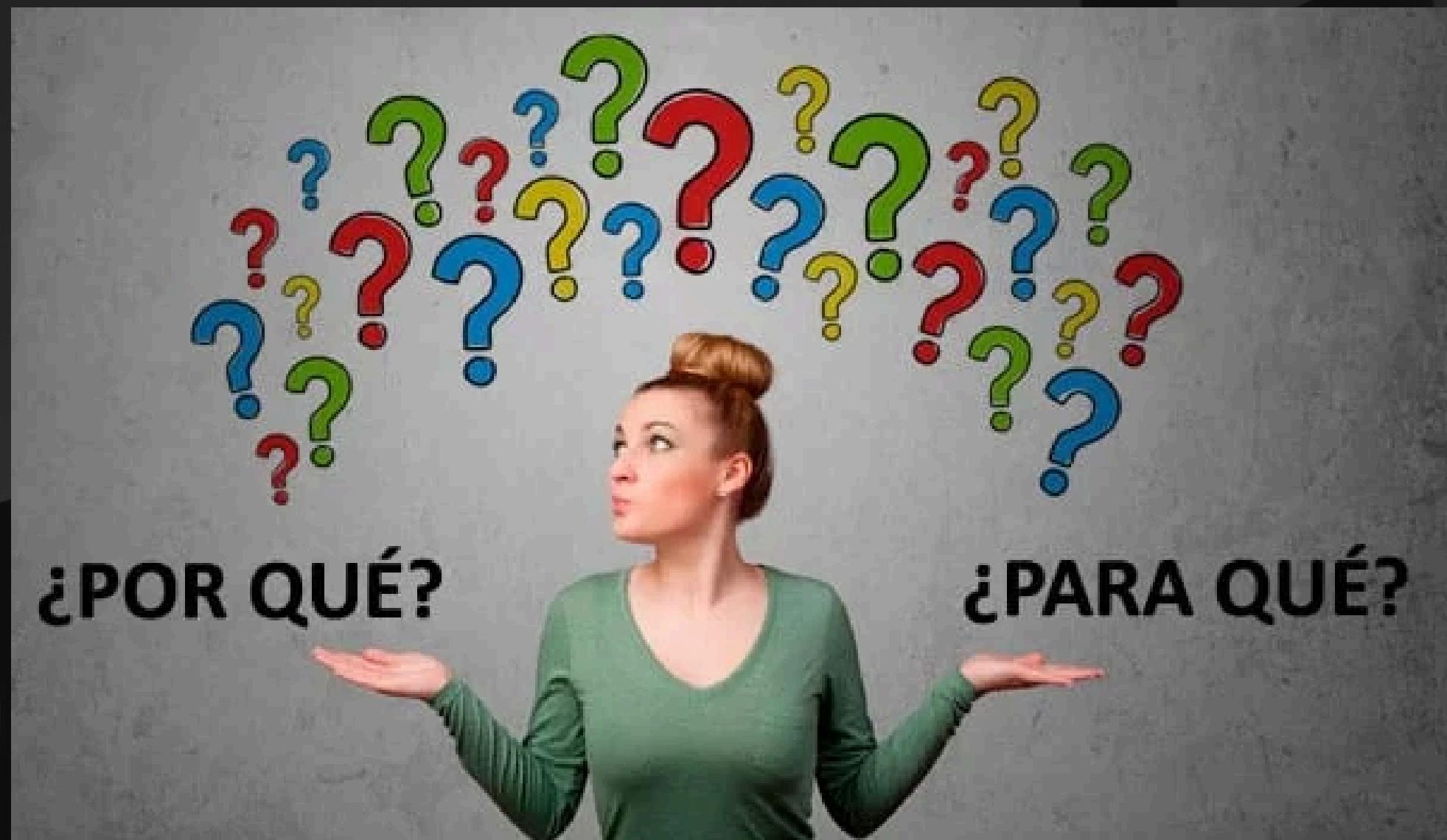
# ChurnShield

---

Ludwig Alvarado Becerra  
Laura Daniela Pineda

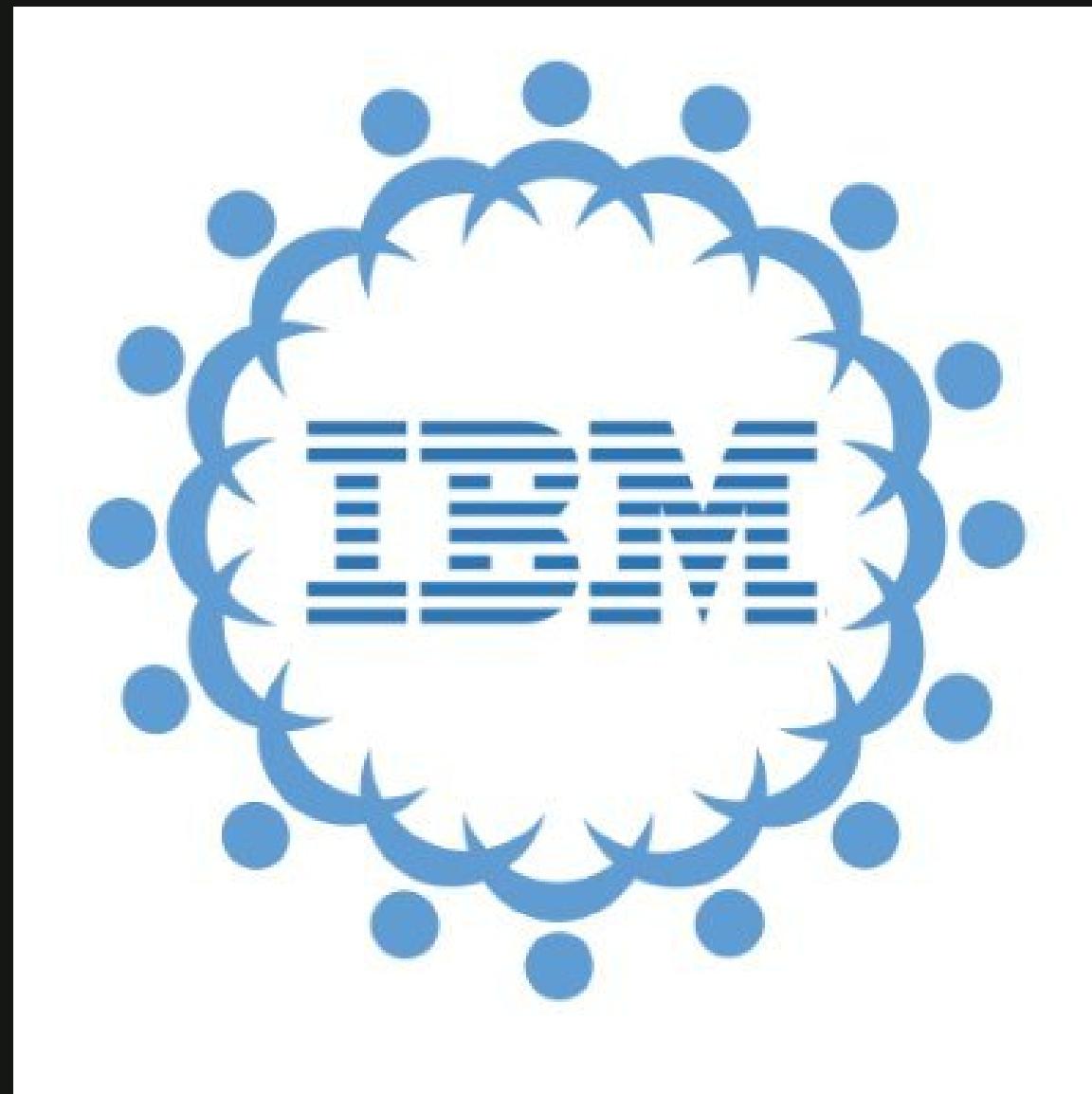
“

# 1. Entendimiento del Negocio



”

# Origen del dataset



Dataset para Business Analytics de IBM  
Community

## **Telco customer churn (11.1.3+)**

Limited-Time Offer: Get on the waiting list now for the 2025 Conference,

⌚ community.ibm.com / Jul 11, 2019

# Origen del dataset



- Compañía de telecomunicaciones.
- Se requiere saber las razones de por qué los clientes se van de la empresa.

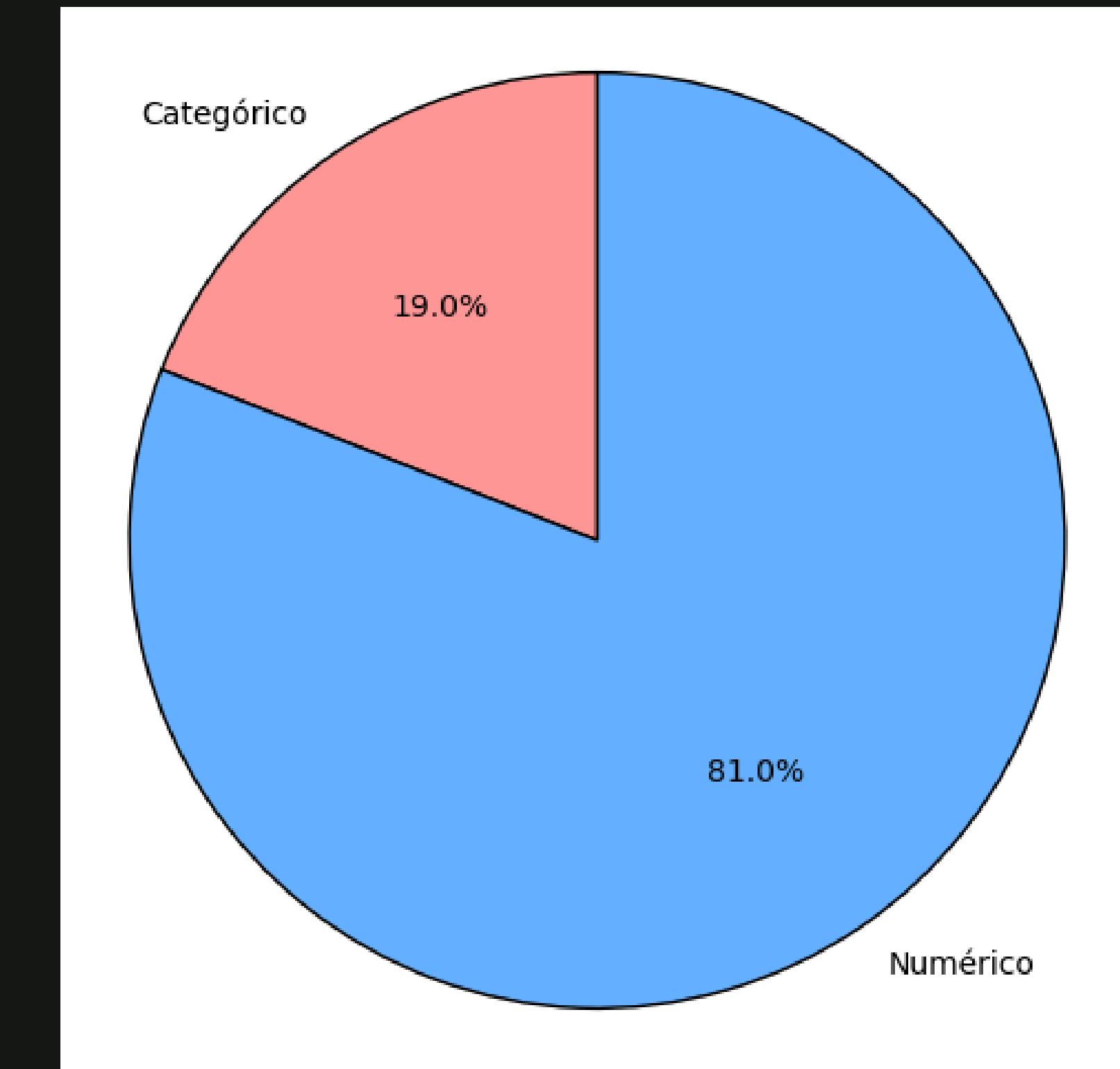
“

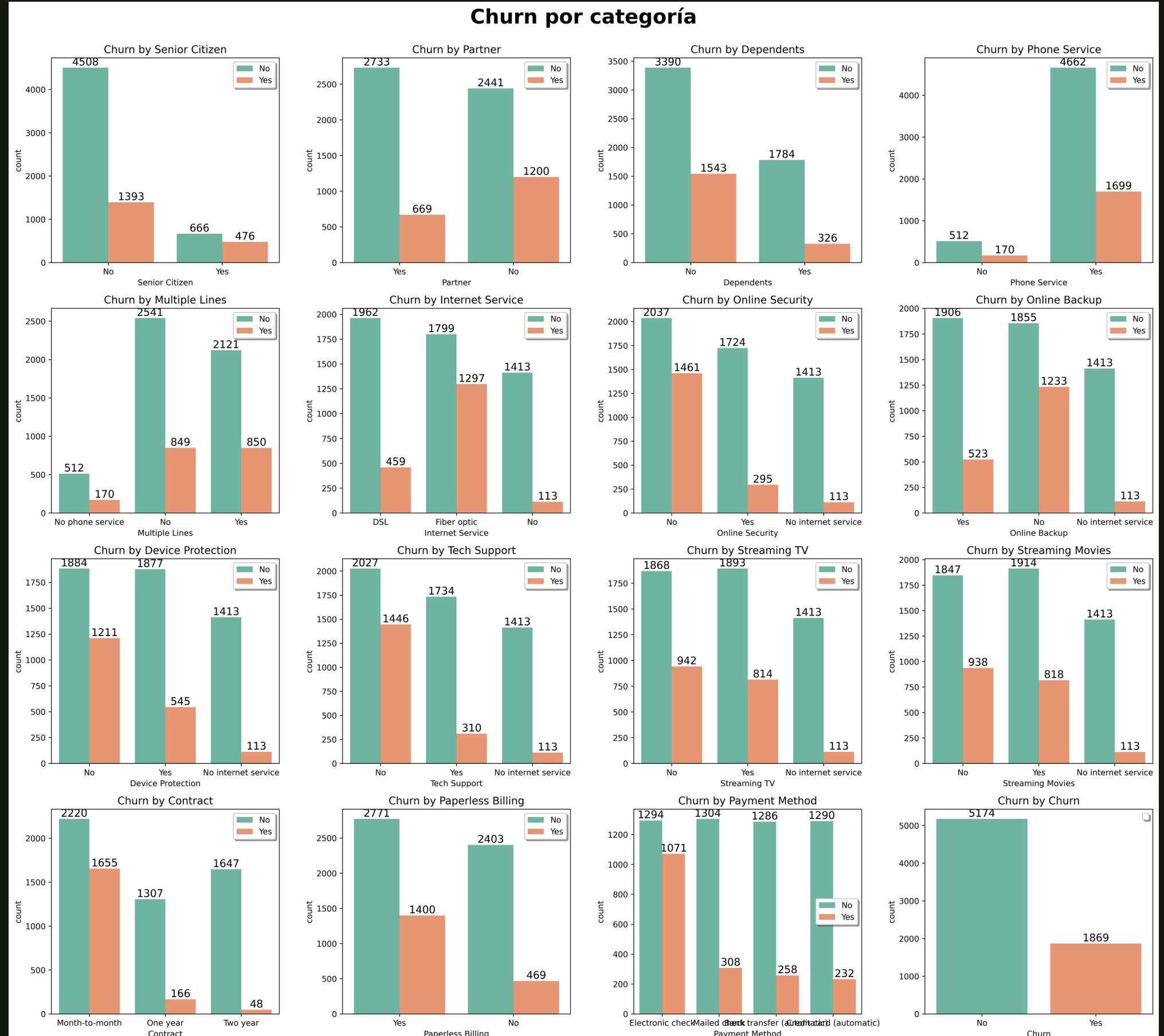
## 2. Entender los Datos:



”

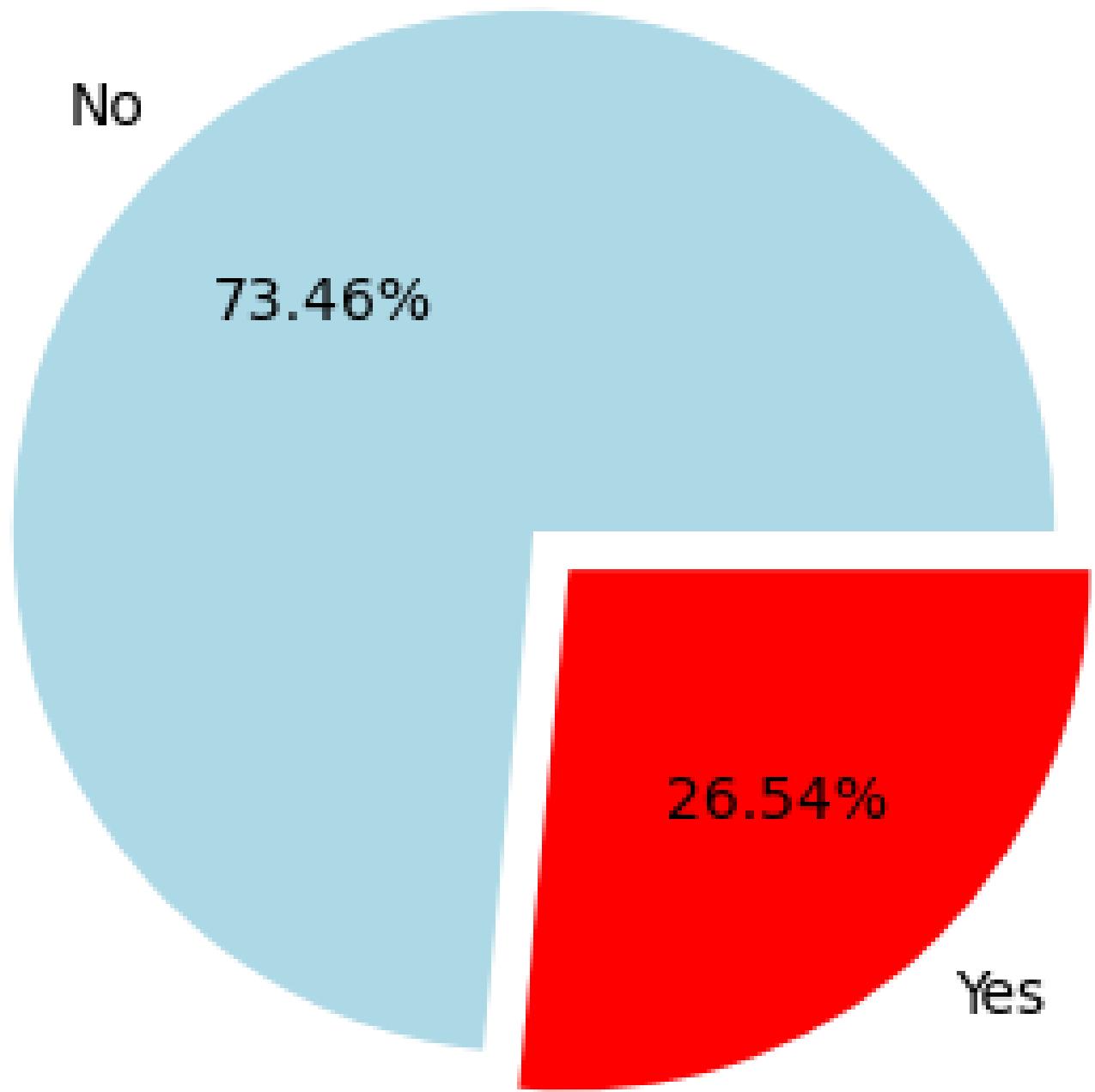
# Tamaño y estructura





- La mayoría de clientes no cuentan con estado senior, es decir, mayores de edad.
- Los clientes que más cancelan son aquellos que no cuentan con servicio de soporte y de seguridad en línea.

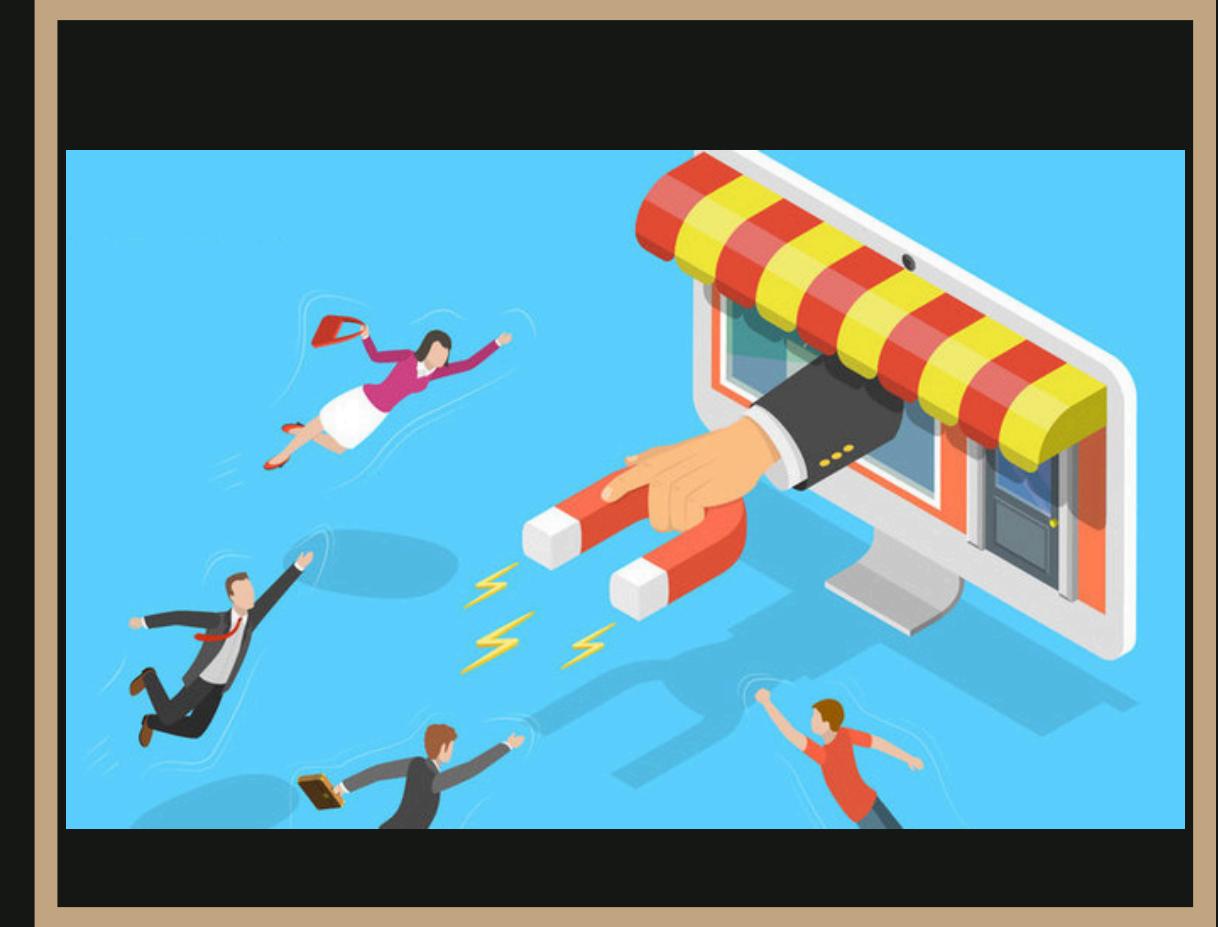
Gráfico de torta para el *Churn*



El 73.46% de los registros aún son clientes de la empresa, mientras que el 26.54% ya no lo son.



### 3. Cómo se Relaciona el Negocio y los Datos



# Aplicación del Modelo



# Procesamiento de datos



# Limpiar valores nulos

```
1 df.isnull().sum()
✓ 0.0s

LoyaltyID          0
Customer ID        0
Senior Citizen     0
Partner            0
Dependents         0
Tenure              0
Phone Service       0
Multiple Lines      0
Internet Service    0
Online Security     0
Online Backup        0
Device Protection   0
Tech Support         0
Streaming TV         0
Streaming Movies     0
Contract             0
Paperless Billing    0
Payment Method       0
Monthly Charges      0
Total Charges        11
Churn                0
dtype: int64
```

Los 11 valores nulos encontrados en la columna Total Charges fueron llenados con el promedio de los valores de toda la columna

# Aplicación pd.get\_dummies

```
3 df[categorical]
2 data_encoded = pd.get_dummies(df[categorical], categorical, drop_first=True)
1
4 data_encoded = pd.concat([data_encoded, numeric_columns, df['Churn']], axis=1)
1 data_encoded
✓ 0.0s
```

	Senior Citizen_Yes	Partner_Yes	Dependents_Yes	Phone Service_Yes	Multiple Lines_No phone service	Multiple Lines_Yes	Internet Service_Fiber optic	Internet Service_No
0	False	True	False	False	True	False	False	False
1	False	False	False	True	False	False	False	False
2	False	False	False	True	False	False	False	False
3	False	False	False	False	True	False	False	False
4	False	False	False	True	False	False	True	False
...	...	...	...	...	...	...	...	...
7038	False	True	True	True	False	True	False	False
7039	False	True	True	True	False	True	True	False
7040	False	True	True	False	True	False	False	False
7041	True	True	False	True	False	True	True	False
7042	False	False	False	True	False	False	True	False

7043 rows × 31 columns

Resultando en un dataframe de 31 columnas en total combinadas con las numéricas y el *Churn*.

# Dividir los datos



# Entrenamiento del modelo

```
3 rf_model = RandomForestClassifier(n_estimators=600, random_state=42)
2 rf_model.fit(X_train, y_train)
1
4 y_pred = rf_model.predict(X_test)
✓ 2.9s
```

RandomForest de la librería scikit-learn  
con 600 árboles de decisión.

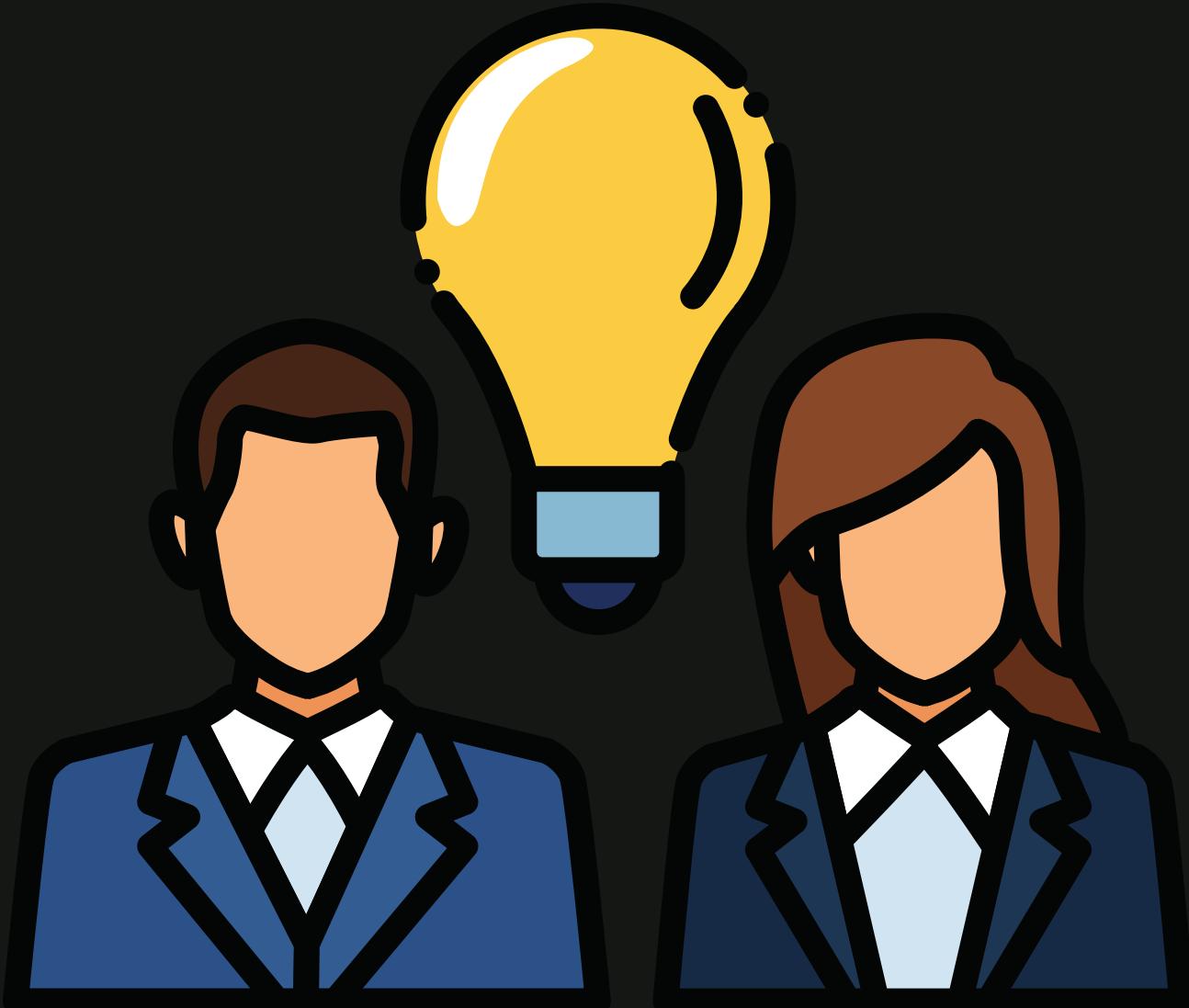
# Evaluación del modelo

```
1 print(classification_report(y_test, y_pred))  
✓ 0.0s
```

	precision	recall	f1-score	support
No	0.84	0.90	0.87	1559
Yes	0.64	0.52	0.58	554
accuracy			0.80	2113
macro avg	0.74	0.71	0.72	2113
weighted avg	0.79	0.80	0.79	2113

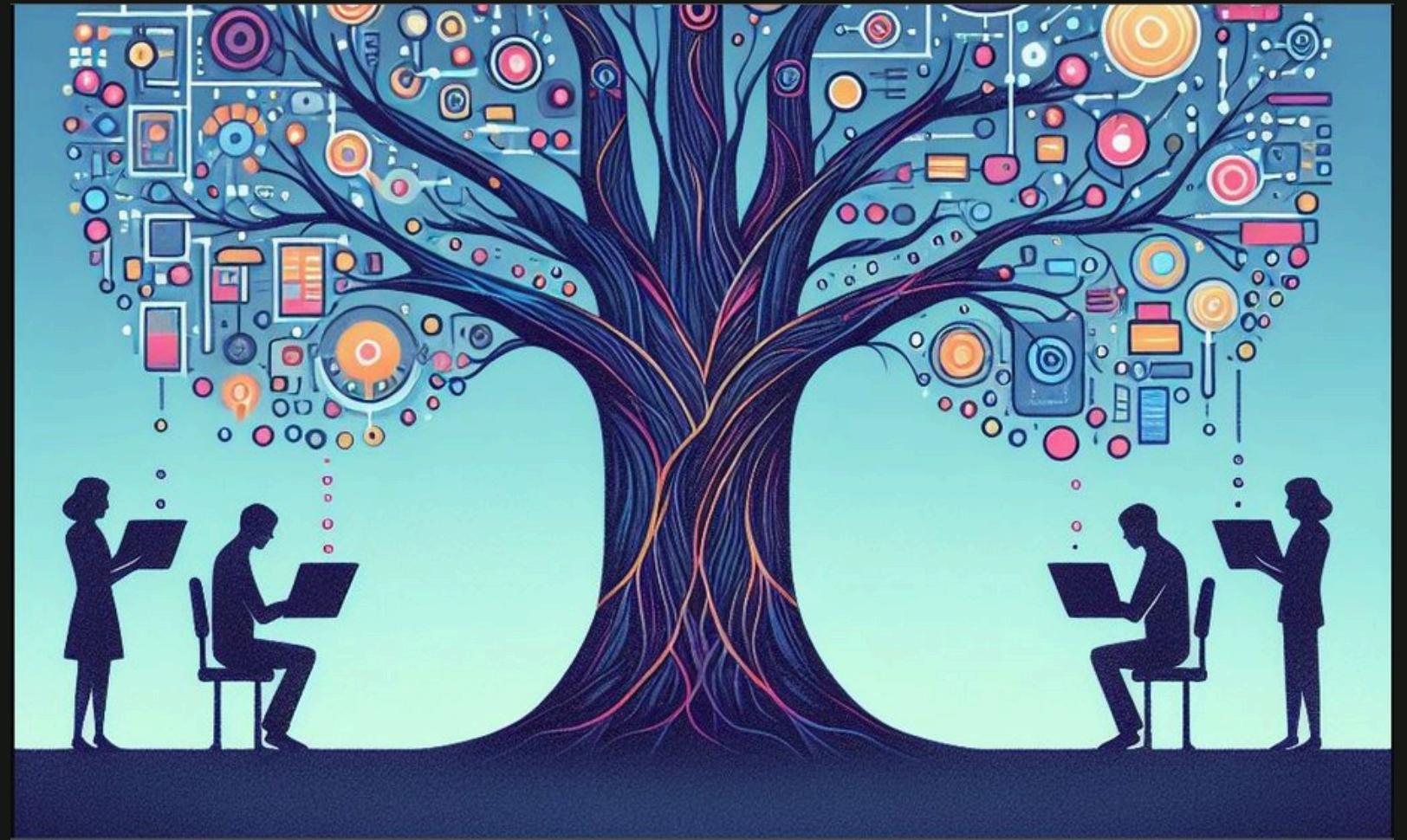
- Se desempeña mejor con clientes que ya no están en la empresa.
- El F1-Score está equilibrado.

# Estrategias que se pueden implementar



- Enfocarse en planes a largo plazo.
- Mejorar el soporte técnico.
- Incluir planes con seguridad informática.

# Conclusiones



- Se realizó un proyecto aplicando metodologías vistas en clase.
- Se siguió el proceso de: análisis -> limpieza -> Modelo -> Interpretación.
- Propuestas para la empresa

# Este proyecto se encuentra en:



*Free as in Freedom*

TheLudway/  
**Modelos-Churn**



1 Contributor 0 Issues 0 Stars 0 Forks

---

**TheLudway/Modelos-Churn**  
Contribute to TheLudway/Modelos-Churn development by creating an account on GitHub.  
[GitHub](#)