

Datathon 2025

Reto OXXO

Gustavo Hernández, Cesar Aguirre, Uziel Lujan, Diego Paniagua





Problemas/Objetivos

- Desarrollar un modelo predictivo robusto que, a partir de coordenadas geográficas (latitud y longitud) y características del entorno, determine si una nueva tienda OXXO tiene un alto potencial de éxito (definido como alcanzar su meta de ventas).
- Abordar el fuerte desbalance de clases con técnicas de preprocesamiento y la elección de una métrica adecuada para etiquetar correctamente las tiendas exitosas y mejorar el rendimiento del modelo.





Metodologia

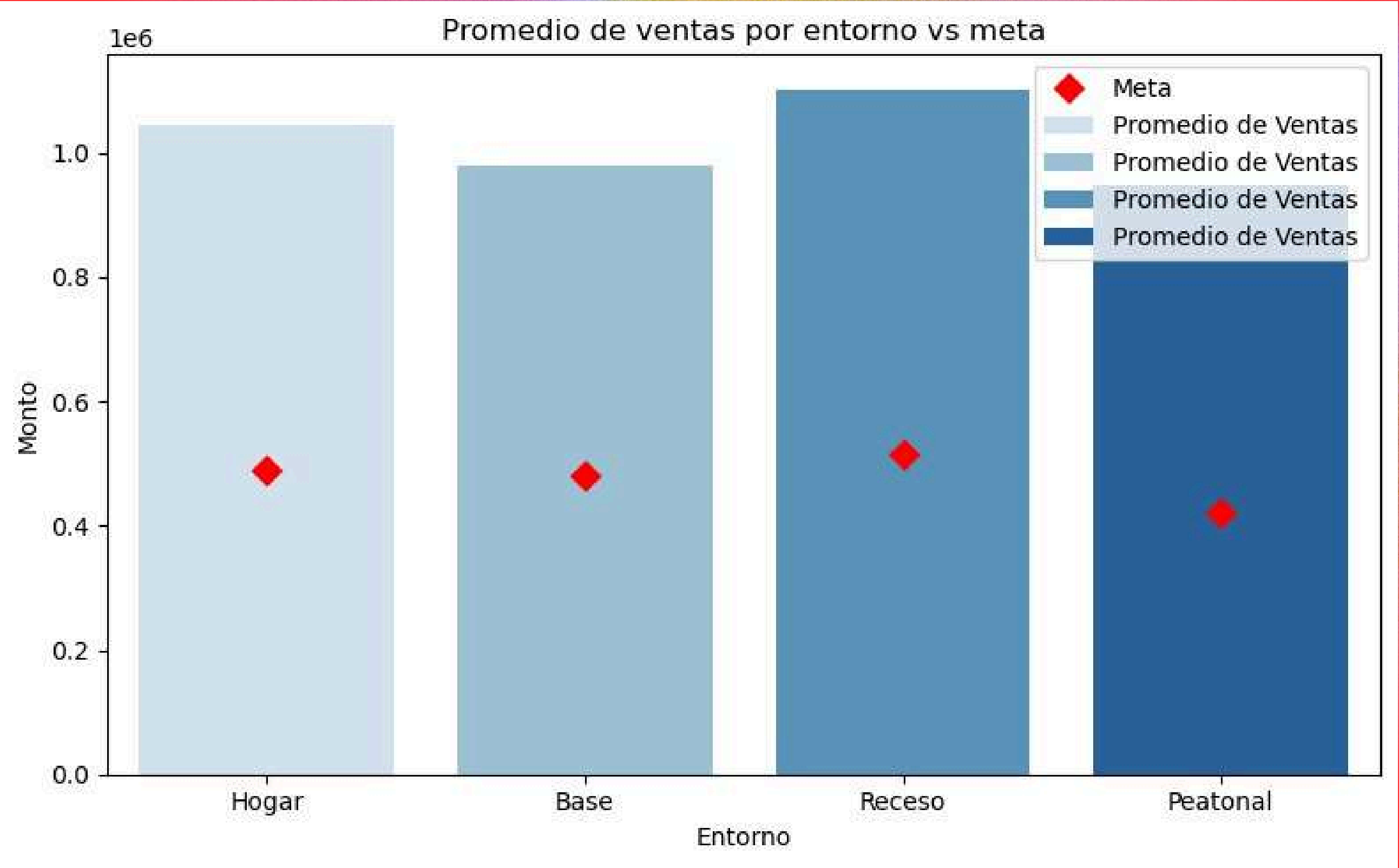
- **Etapa 1. Limpieza y transformación:**
Imputación de datos faltantes, codificación de variables categóricas, normalización.
- **Etapa 2. Elección de una métrica de éxito:**
La importancia de la definición de una tienda exitosa para evitar el desbalance en los datos.
- **Etapa 3. Representación de los datos:**
Propuestas para mejorar la codificación de la información de las variables
- **Etapa 4. Modelado predictivo:**
Comparativa entre Random Forest, XGBoost, y redes neuronales.



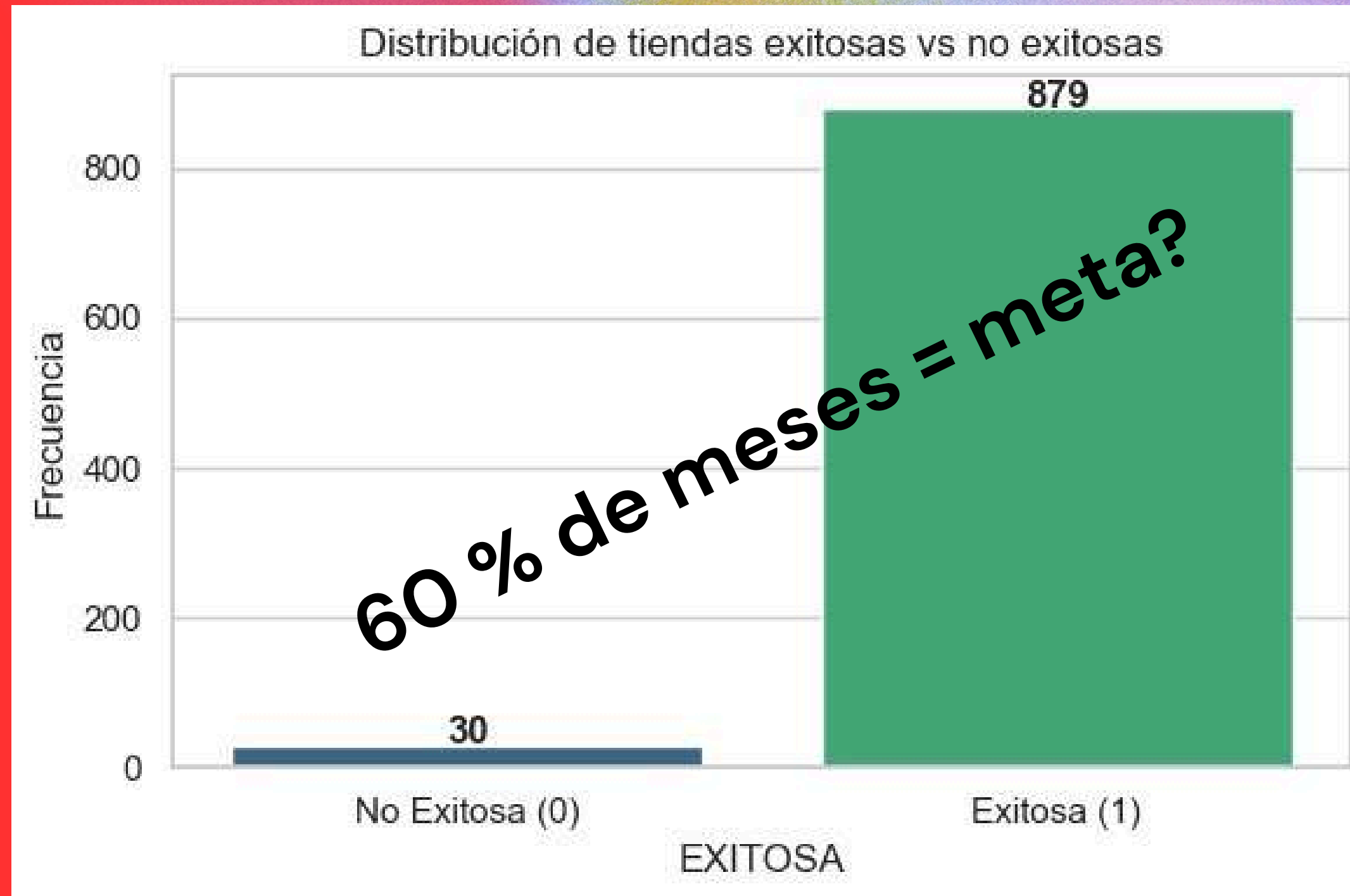
Problemática: Etapa 1 y 2

Historiales de Venta

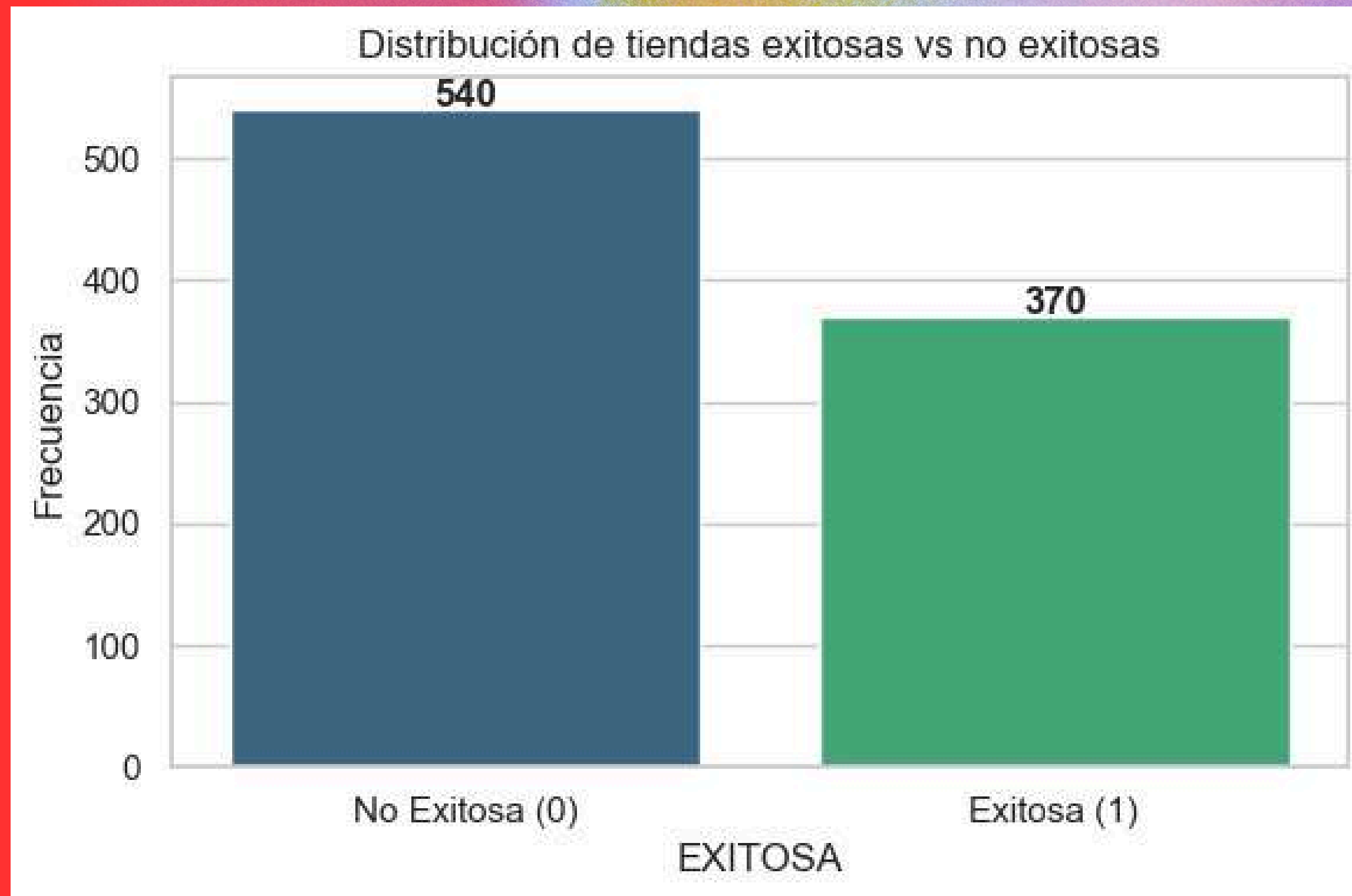




Definición de Tienda Exitosa

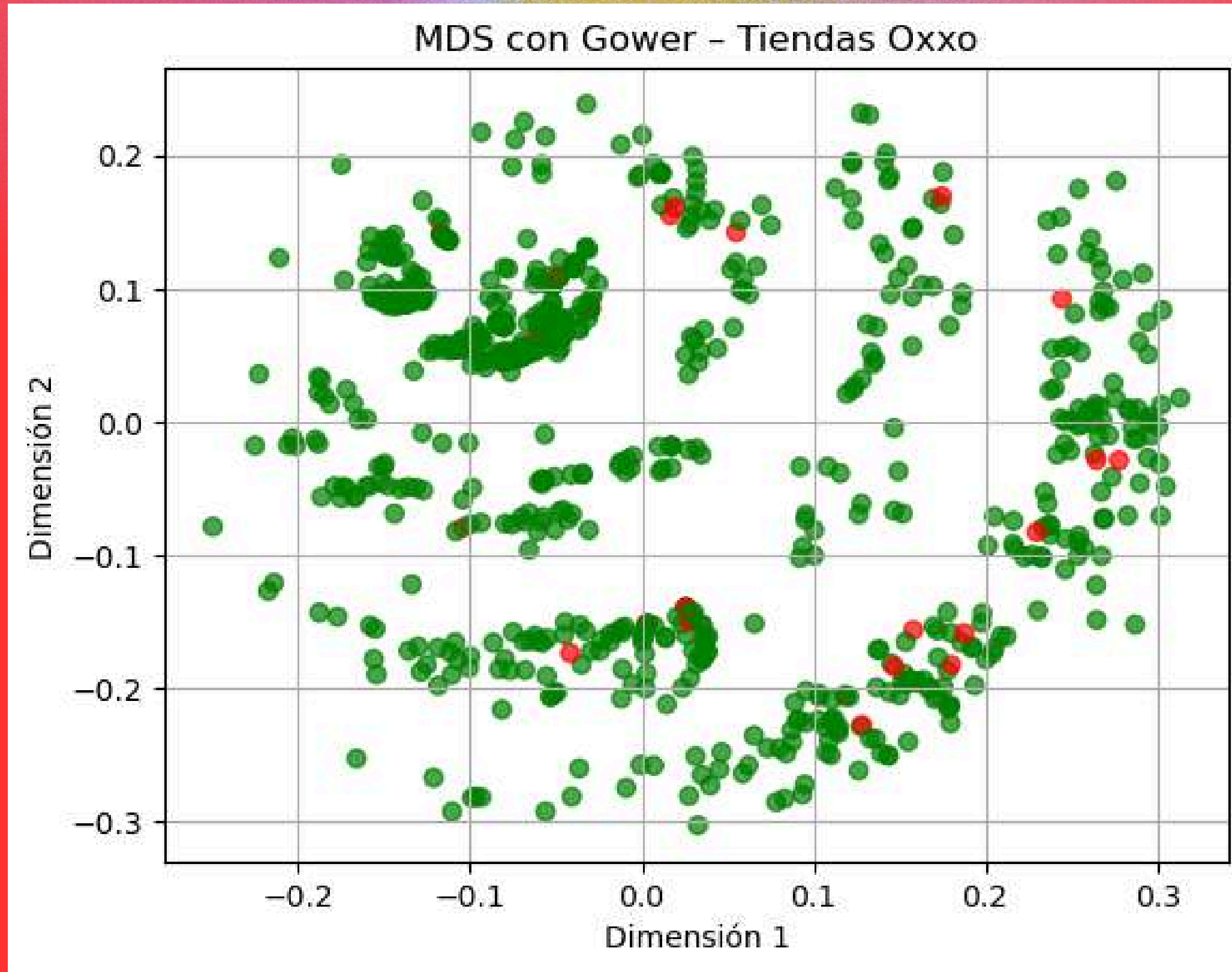


¿Por qué no algo más?...

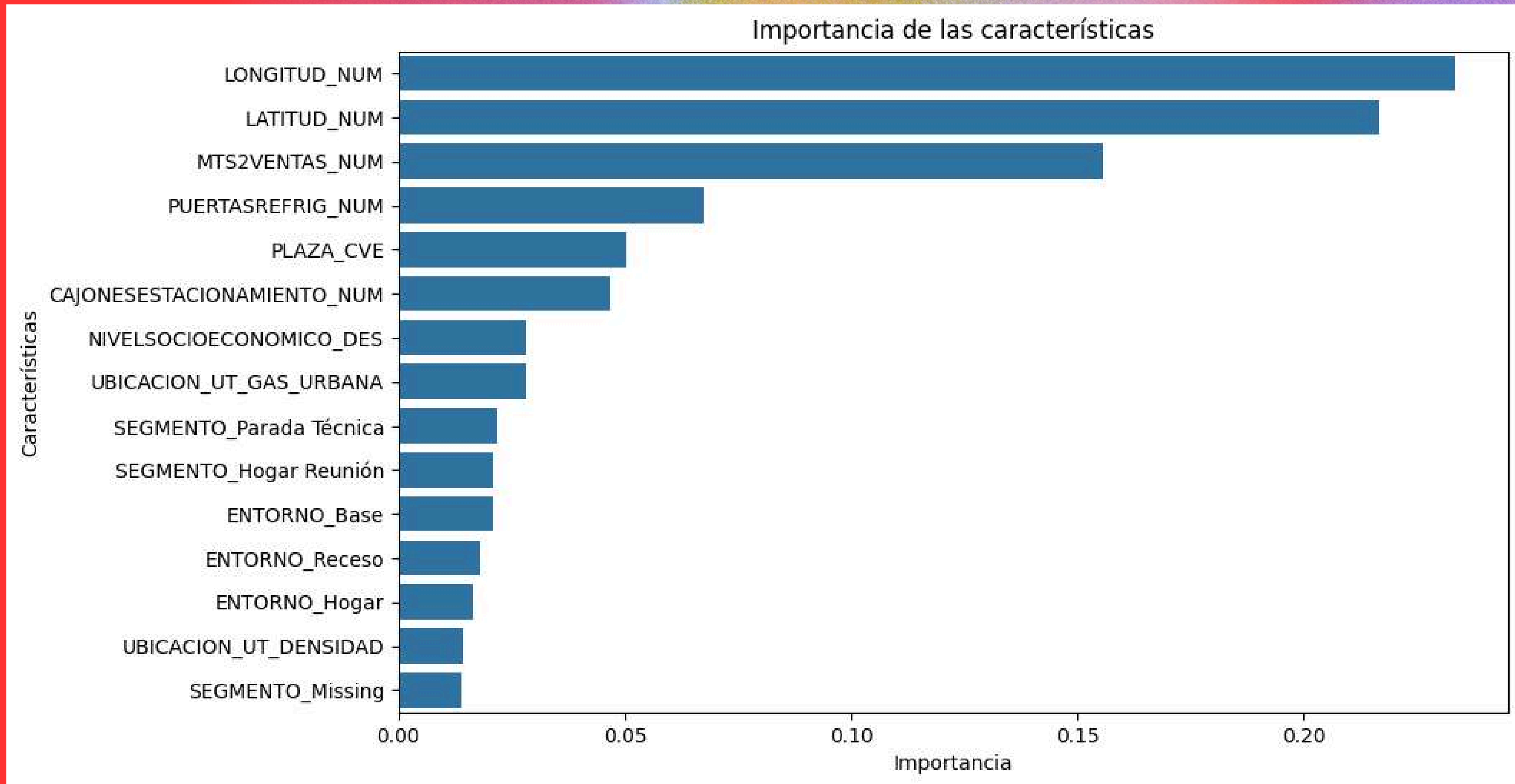


Etapas 3. Las variables

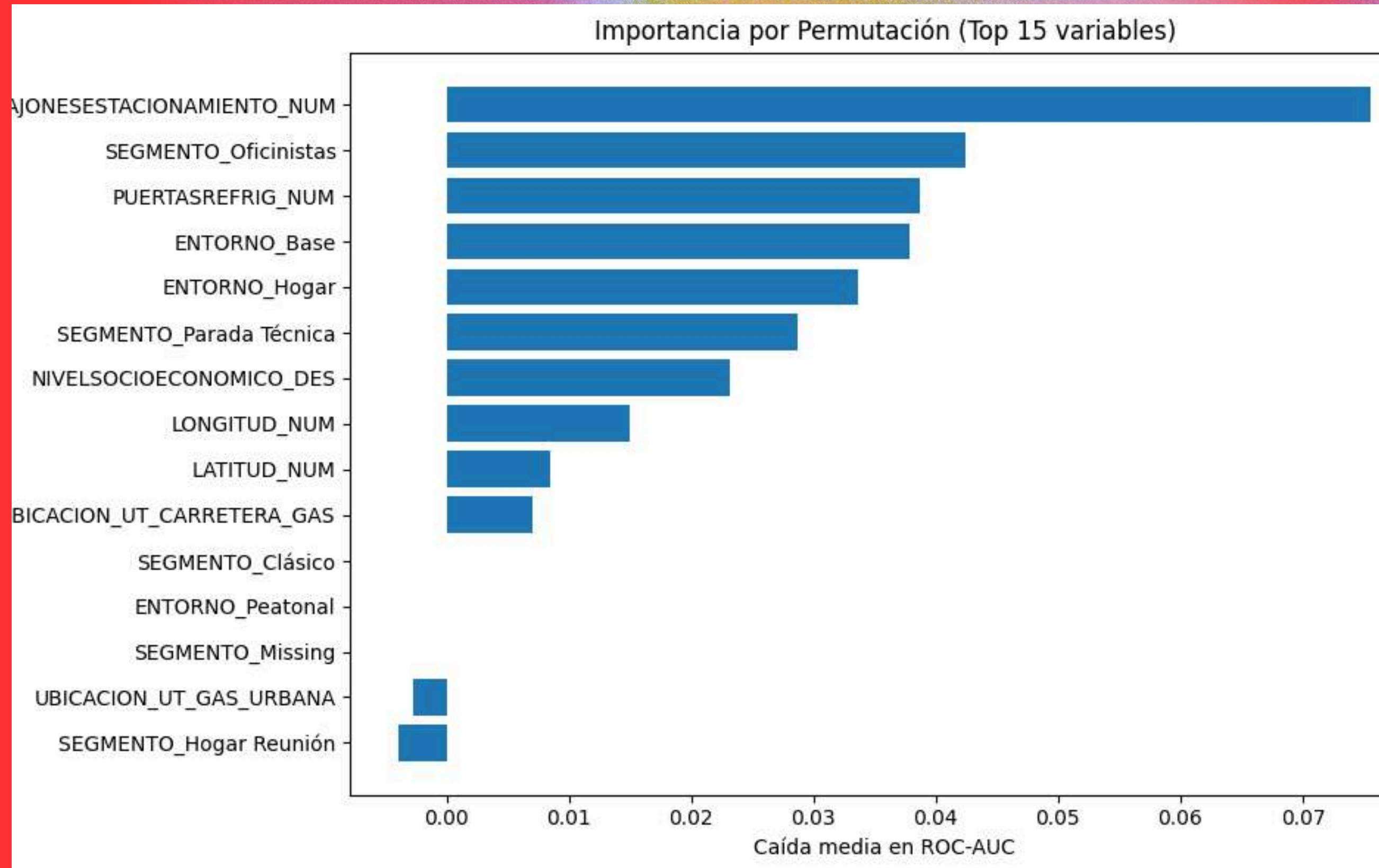
¿Qué nos dicen los datos?



Random Forest: Feature extraction



Random Forest: Feature extraction



Etapas 4. El modelo

¿Que modelo?

- Random Forest
- XGBoost
- Redes Neuronales
- Regresion Logistica

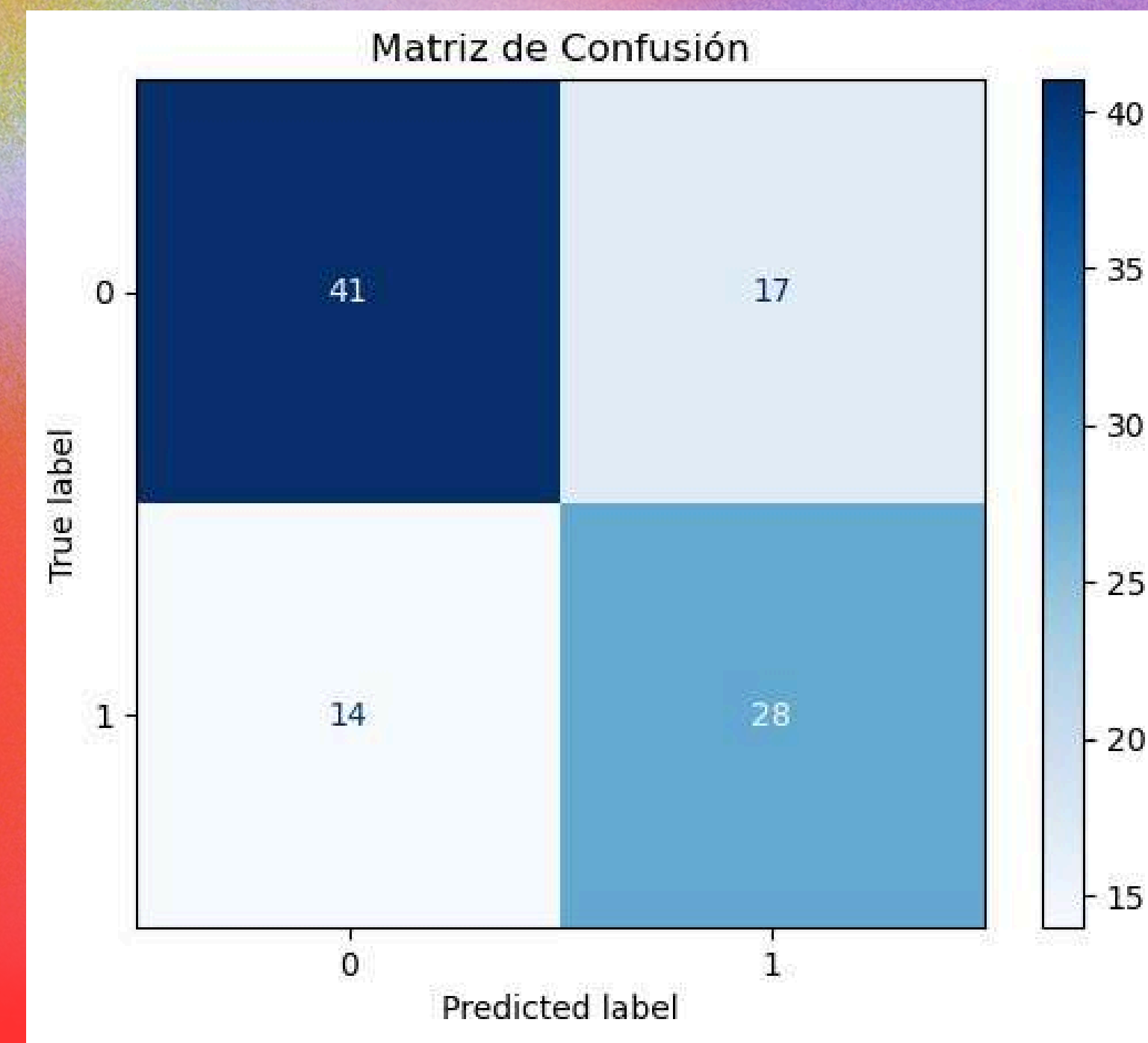
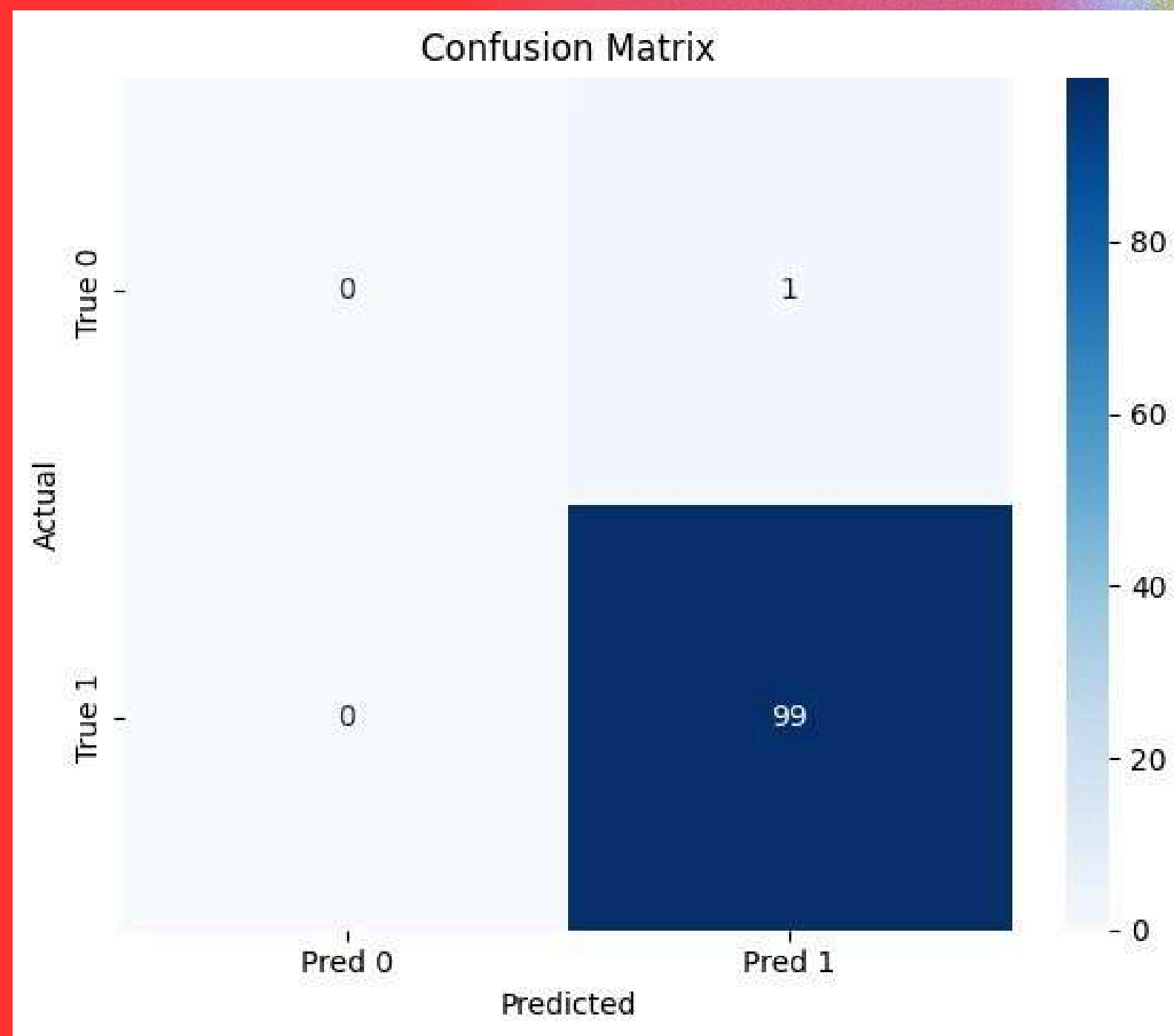
...

Resultados

```
== Classification Report ==
              precision    recall  f1-score   support

     0.0         0.7455         0.7069         0.7257         58
     1.0         0.6222         0.6667         0.6437         42

 accuracy              0.6900         100
  macro avg           0.6838         0.6868         0.6847         100
 weighted avg           0.6937         0.6900         0.6912         100
```

¡Gracias!

