

**FEMSA**



# **Datathon 2025**

## **Reto Femsa**

NaN CREW

# Metodología

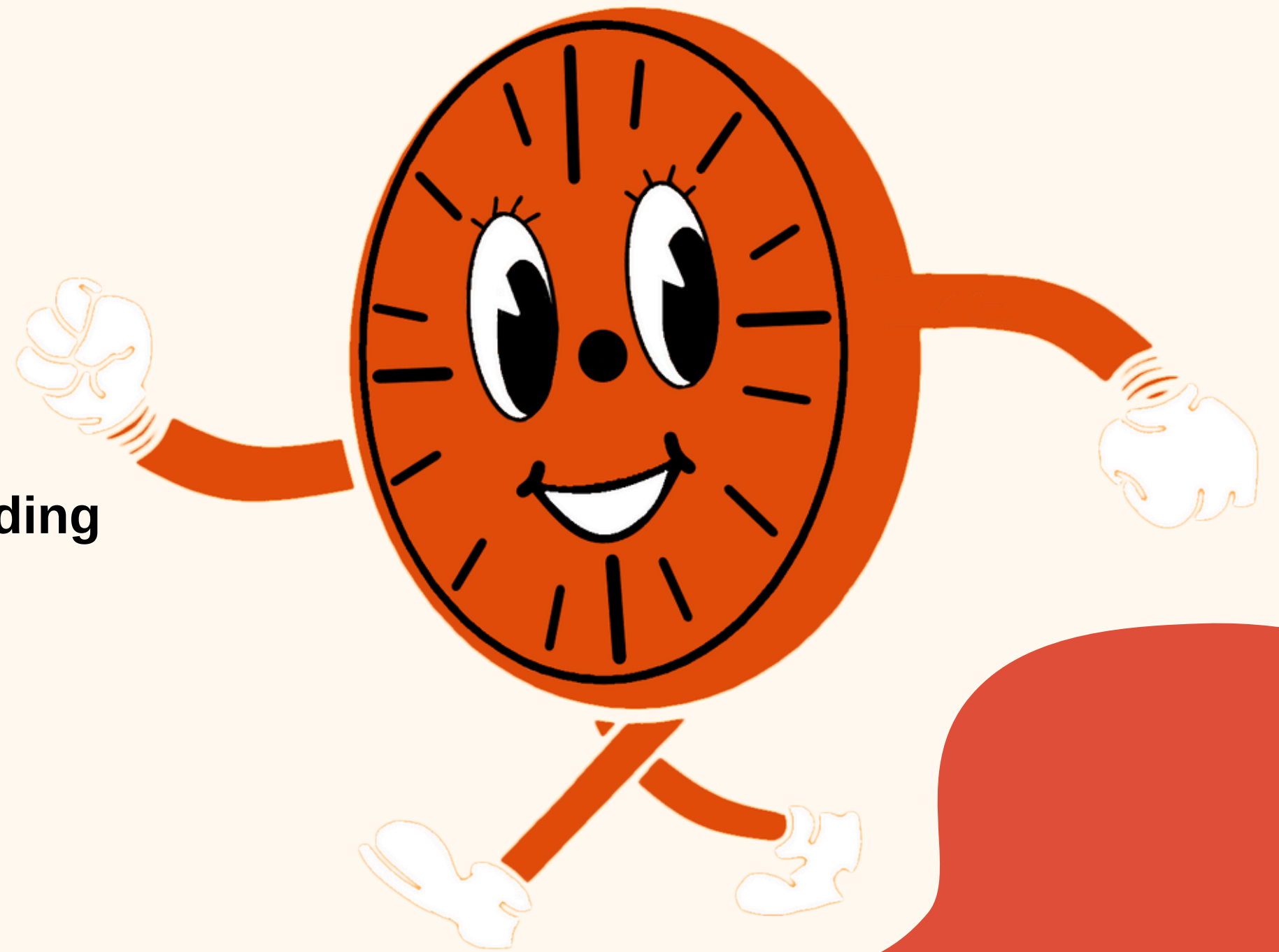
**1** Problem Definition & Data Understanding

**2** Data Augmentation

**3** Data Preprocessing & Cleaning

**4** Feature Engineering

**5** Model Selection



# Problem Definition

Desarrollar un modelo de predicción que a partir de una ubicación (latitud/longitud) determine si una tienda Oxxo tiene potencial de éxito, cumpla con su meta de venta.

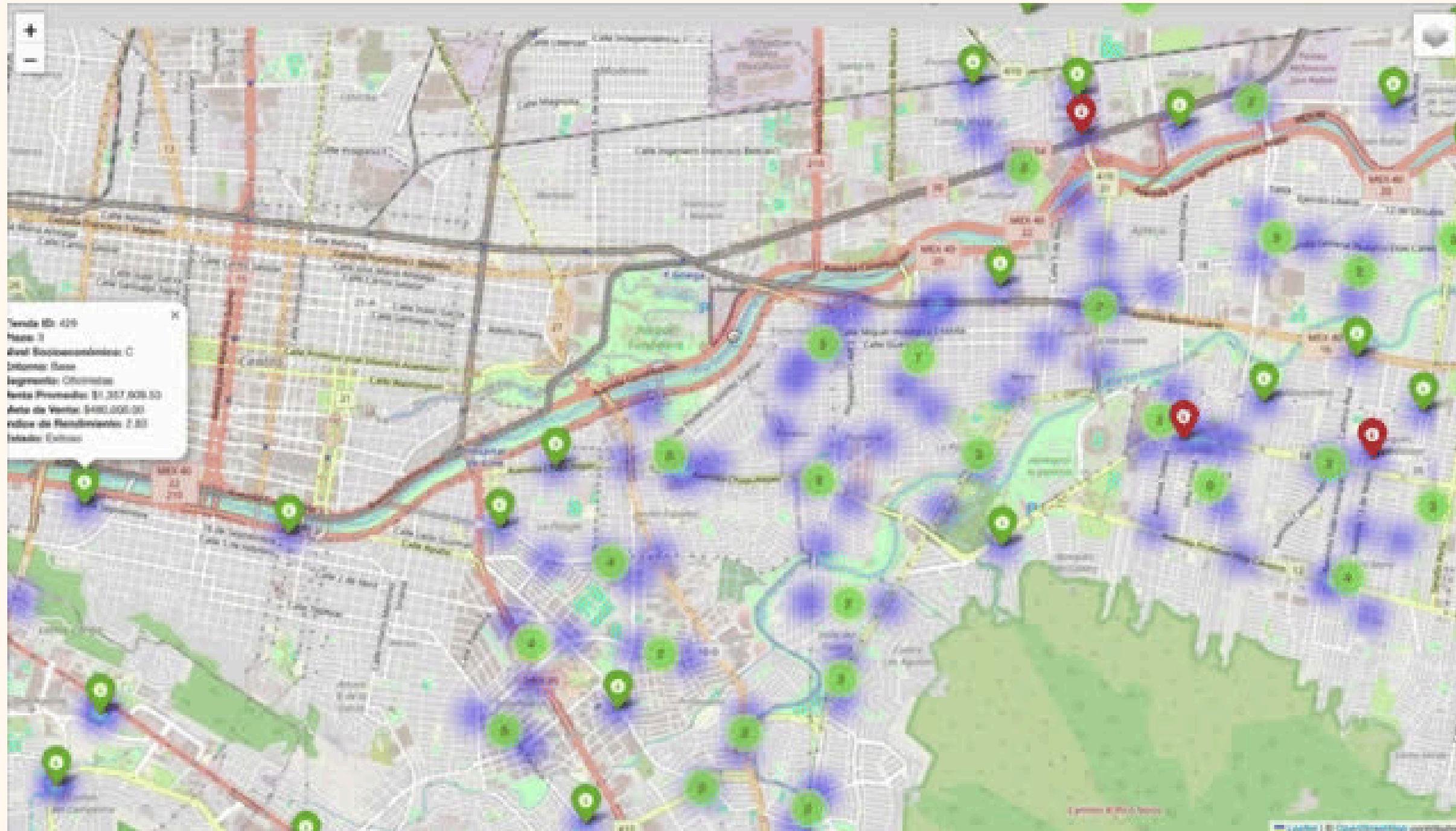




La base de datos proporcionada

table	title	tipo	description
DIM_TIENDA	TIENDA_ID	string	<p>Hace referencia al código que se utiliza para identificar una tienda</p>
DIM_TIENDA	PLAZA_CVE	string	<p>Hace referencia al nombre de la plaza donde esta asignada la tienda</p>
DIM_TIENDA	NIVELSOCIOECONOMICO_DES	string	<p>Hace referencia al promedio del nivel socioeconómico del entorno en que se ubica la tienda</p>
DIM_TIENDA	ENTORNO_DES	string	<p>Hace referencia al entorno en que se ubica la tienda</p>
DIM_TIENDA	MTS2VENTAS_NUM	float	<p>Hace referencia a la medida del área del piso de ventas en la tienda y cuarto frio de exhibición al cliente, el cual es expresado en metros cuadrados (m2)</p>
DIM_TIENDA	PUERTASREFRIG_NUM	float	<p>Hace referencia al número de puertas del cuarto frio</p>
DIM_TIENDA	CAJONESESTACIONAMIENTO_NUM	float	<p>Hace referencia a la cantidad de cajones de estacionamiento de la tienda</p>
DIM_TIENDA	LATITUD_NUM	float	<p>Hace referencia a la latitud de la tienda</p>
DIM_TIENDA	LONGITUD_NUM	float	<p>Hace referencia a la longitud de la tienda</p>
DIM_TIENDA	SEGMENTO_MAESTRO_DESC	string	<p>Hace refrencia al tipo de segmento a la cual pertenece la tienda </p>
DIM_TIENDA	LID_UBICACION_TIENDA	string	<p>Hace referencia al tipo de ubicación de la tienda</p>
DIM_TIENDA	Dataset	string	<p>Hace referencia al conjunto al que pertenece la tienda TRAIN o TEST</p>
Venta	TIENDA_ID	string	<p>Hace referencia al código que se utiliza para identificar una tienda</p>
Venta	MES_ID	string	<p>Hace referencia al año y mes </p>
Venta	VENTA_AMT	float	<p>Hace referencia a la venta mensual de la tienda</p>
Meta_Venta	ENTORNO_DES	string	<p>Hace referencia al entorno en que se ubica la tienda</p>
Meta_Venta	Meta_venta	float	<p>Hace referencia a la venta mensual requerida para alcanzar su meta de venta</p>

# Data Understanding



**1**

## Open Street Maps

Conteo de POIs (escuelas, bancos, parques) y vialidades cercanas.

**2**

## DENUE

Directorio de unidades económicas para identificar competencia o generadores de tráfico.

**3**

## Censo INEGI

El Censo del INEGI aporta una radiografía detallada del entorno socioeconómico de cada ubicación, permitiendo al modelo comprender el perfil demográfico, educativo y habitacional de los potenciales clientes.

# Data Preprocessing & Cleaning



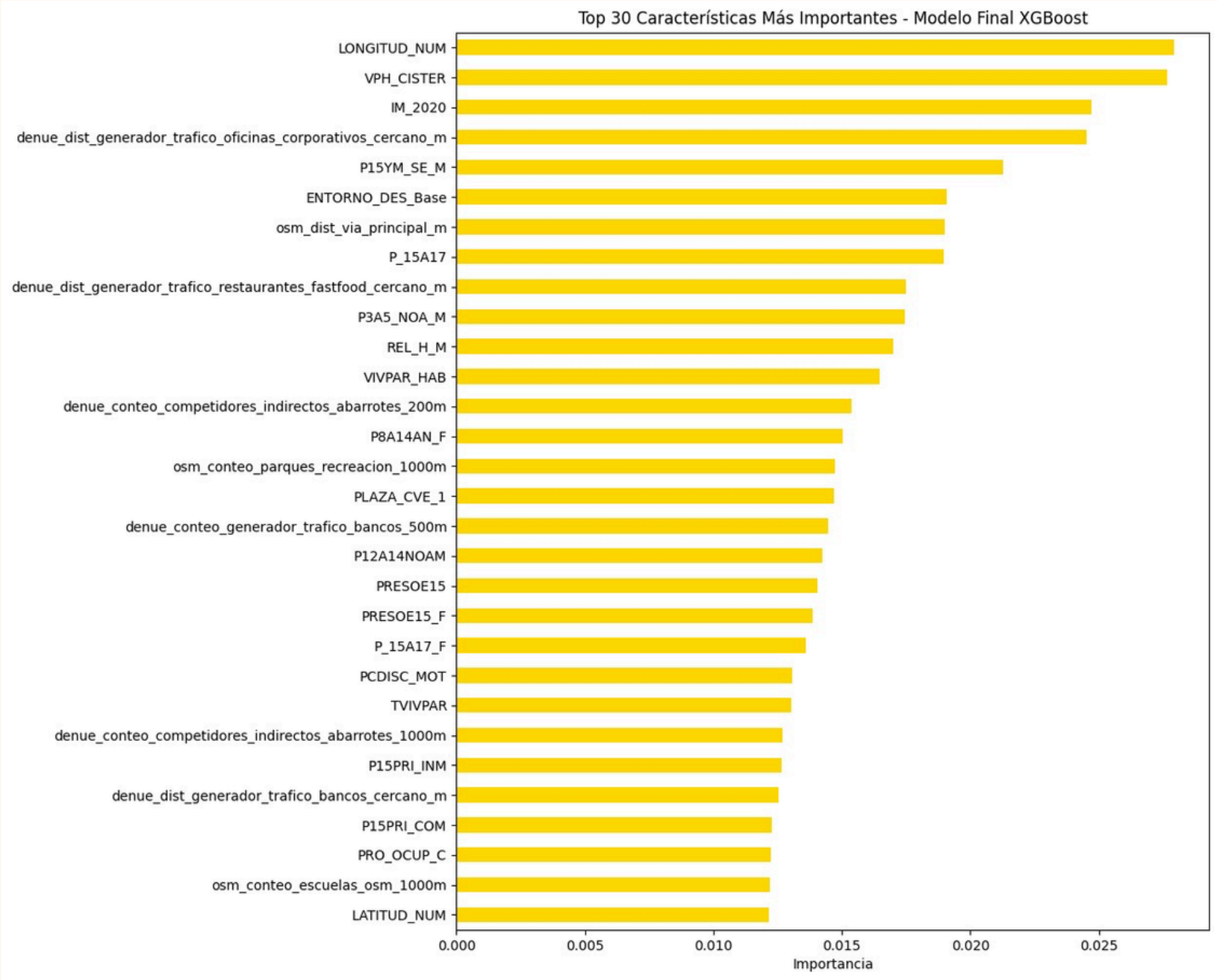
```
# --- 2. Selección Inicial de Características y Separación Target/Features ---
cols_a_eliminar = []
'TIENDA_ID',
'geometry',
'VENTA_PROMEDIO_MENSUAL', # Solo en train_gdf
'Meta_venta',            # Solo en train_gdf
'N_MESES_CON_VENTA_EN_PERIODO', # Solo en train_gdf
'CVEGEO',
'NOM_ENT', 'NOM_MUN', 'NOM_LOC', 'NOM_AGEB', # Nombres geográficos
'CVE_AGEB_ENT', # Si se coló del IMU y es redundante

# NUEVAS COLUMNAS A AÑADIR PARA ELIMINAR (basado en tu warning)
'AGEB',          # Del warning
'DATASET',       # Del warning, definitivamente no es una feature
'ENTIDAD',       # Del warning
'LOC',           # Del warning
'MUN',           # Del warning
'PROMEDIO TOTAL', # Del warning
'TOTAL'          # Del warning
```

- Limpieza de Datos Externos
- Unión de Datos Censales a AGEBS
- Unión Espacial de Tiendas con AGEBS+Censo
- Manejo de Inconsistencias Post-Unión



# Feature Engineering



Etiqueta	Description
LONGITUD_NUM	Longitud Geográfica de la Tienda
VPH_CISTER	Viviendas Particulares Habitadas con Cisterna/Aljibe
IM_2020	Índice de Marginación 2020
denue_dist_generador_trafico_oficinas_corporativos_cercano_m	Distancia a Oficinas/Corporativos Cercanos (metros)
P15YM_SE_M	Población de 15+ Años Sin Escolaridad (Hombres)
ENTORNO_DES_Base	Indicador: Entorno de Tienda es "Base"
osm_dist_via_principal_m	Distancia a Vía Principal Cercana (metros)
P_15A17	Población de 15 a 17 Años
denue_dist_generador_trafico_restaurantes_fastfood_cercano_m	Distancia a Restaurantes/Comida Rápida Cercanos (metros)
P3A5_NOA_M	Población de 3-5 Años que NO Asiste a Escuela (Hombres)



# #1era Opción

## XGBoost

Se seleccionó un modelo predictivo XGBoost por las siguientes razones:

- Regularización
- Boosting secuencial automatizado
- Control granular
- Útil cuando se manejan datasets grandes con grandes cantidades de features



97.9%  
PRECISIÓN

93.1%  
SENSIBILIDAD

95.4%  
F1-SCORE

91.4%  
EXACTITUD

**Nota:** La matriz muestra el rendimiento del modelo XGBoost optimizado. La alta precisión (98.9%) indica pocas predicciones incorrectas de la clase positiva, mientras que la sensibilidad del 87.1% sugiere que el modelo identifica correctamente la mayoría de los casos positivos reales.

## # 2da Opción

### XGBoost Refinado - Dual Feature Set

Como modelo alternativo se presenta un XGBoost utilizando parametros mas conservadores, este modelo sacrifica un poco de sensibilidad para ser más preciso en lo que predice como positivo.



98.9%

PRECISIÓN

87.1%

SENSIBILIDAD

92.6%

F1-SCORE

86.7%

EXACTITUD

**Nota:** La matriz muestra el rendimiento del modelo XGBoost optimizado. La alta precisión (98.9%) indica pocas predicciones incorrectas de la clase positiva, mientras que la sensibilidad del 87.1% sugiere que el modelo identifica correctamente la mayoría de los casos positivos reales.

# Páginas de recursos

<https://www.inegi.org.mx/app/scitel/Default?ev=9>

<https://www.openstreetmap.org/#map=5/23.94/-102.58>

<https://nl.semaforo.com.mx/>

<https://www.gob.mx/conapo/documentos/indices-de-marginacion-2020-284372>

<https://codemaker2016.medium.com/introducing-uv-next-gen-python-package-manager-b78ad39c95d7>

<https://platform.openai.com/docs/overview>