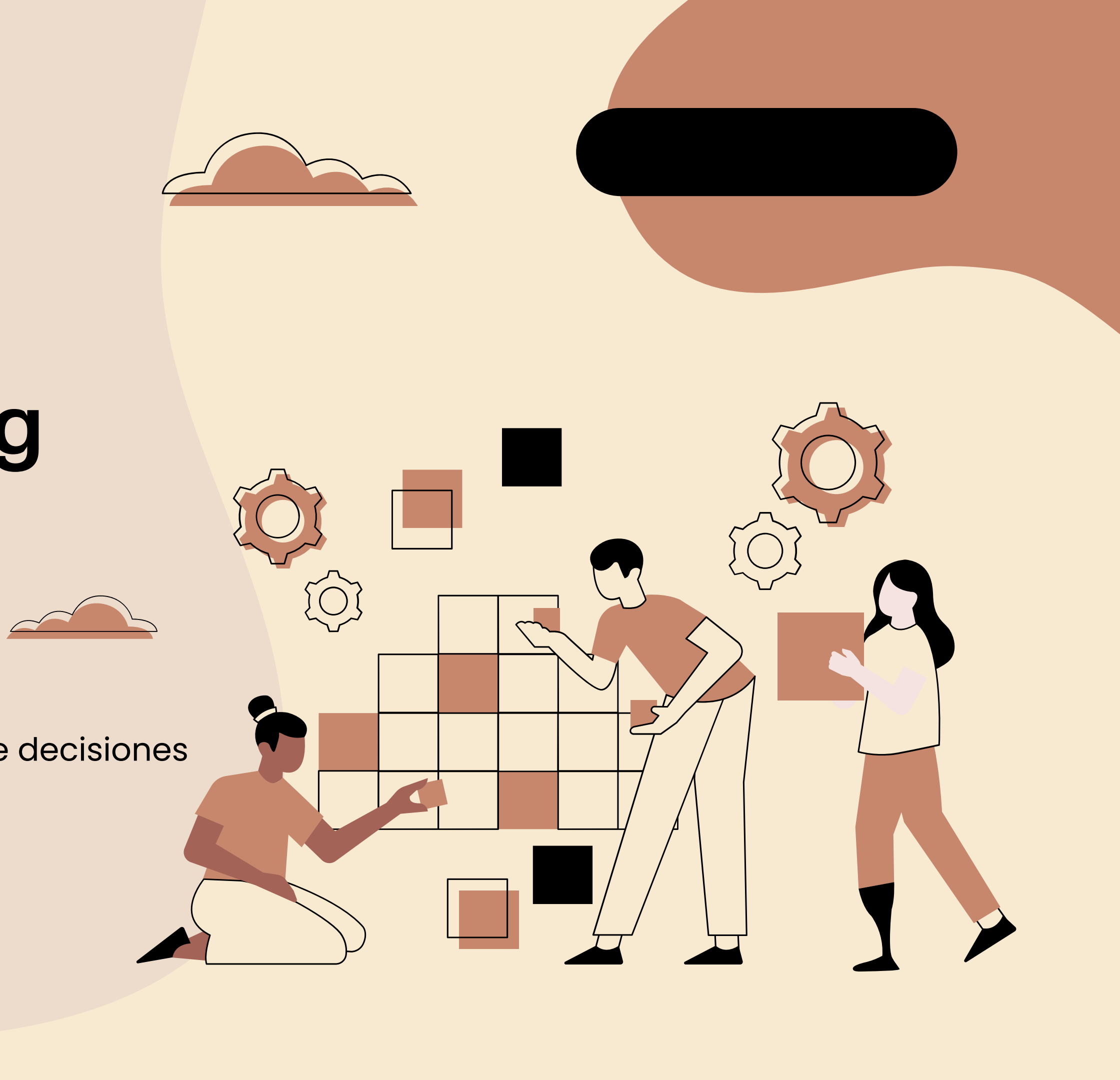


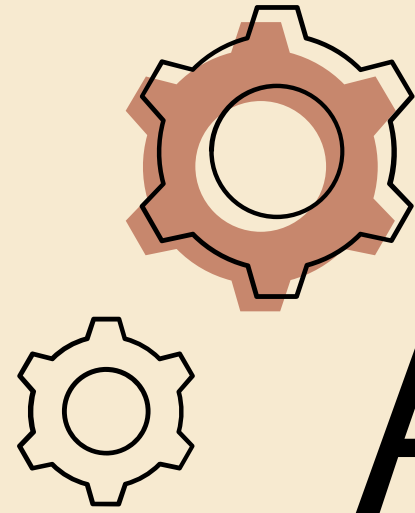
# Predicción de Ventas con Machine Learning

Mirlenis Tovar

Junio 2025

“Una herramienta de apoyo a la toma de decisiones basada en datos históricos de ventas.”





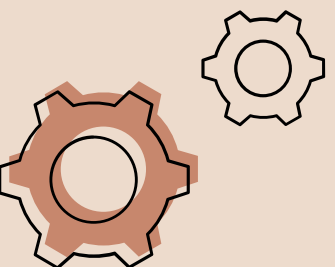
# Abstracto / Motivación

¿Por qué este proyecto?

En el contexto actual de la industria farmacéutica, predecir las ventas es esencial para:

- Optimizar la planificación de stock
- Mejorar la eficiencia comercial
- Detectar tendencias y tomar decisiones proactivas

Este modelo predice el importe vendido neto por producto, usando datos históricos entre 2019 y 2025.



# PROBLEMA Y AUDIENCIA



## Problema de negocio:

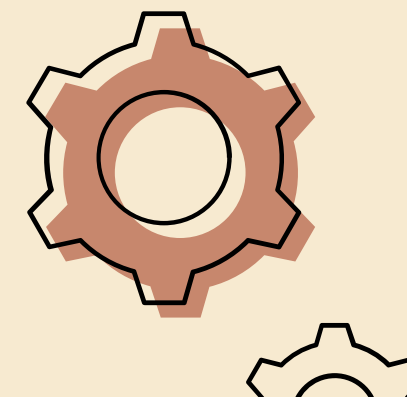
¿Cómo podemos predecir las ventas de un producto?

## Objetivo:

Construir un modelo predictivo que permita anticipar montos de venta según múltiples variables: artículo, vendedor, cliente, ubicación y tipo de cliente.

## Audiencia objetivo:

Gerentes comerciales y de ventas



# DATOS UTILIZADOS



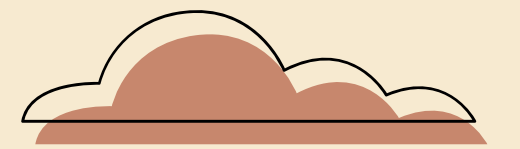
- **Origen:** Base de datos propia (XLSX)

- **Período:** 2019 – mayo 2025

- **Registros:** 531.893 filas

- **Variables clave:**
  - Artículo y familia
  - Tipo de cliente
  - Provincia / Localidad
  - Vendedor
  - Fecha
  - Importe vendido neto

# PREGUNTAS / HIPÓTESIS DEL ANÁLISIS



Se plantearon preguntas clave que guiaron el desarrollo del modelo y las visualizaciones:

- ¿Cuáles son los productos más vendidos por importe neto?

Para identificar qué artículos impulsan las ventas totales.

- ¿Qué vendedores generan mayor volumen de ventas?

Para reconocer patrones de rendimiento comercial.

- ¿Qué provincias o localidades concentran la mayor facturación?

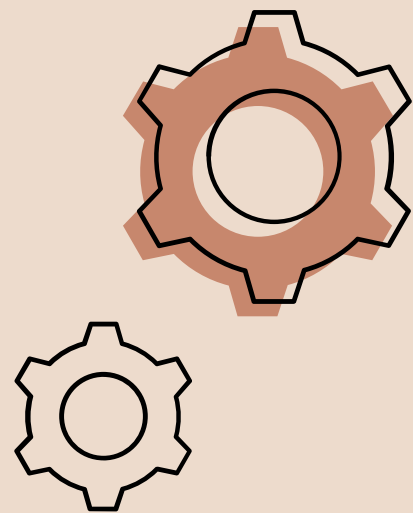
Para detectar oportunidades de crecimiento geográfico.

- ¿El tipo de cliente influye en el volumen de ventas?

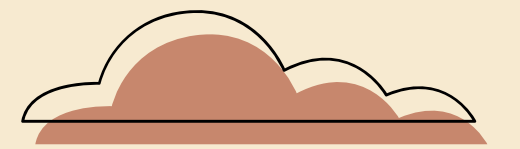
Para segmentar estrategias comerciales.

- ¿Podemos predecir el importe vendido neto con variables como producto, cliente, vendedor, zona y fecha?

Para automatizar decisiones futuras.

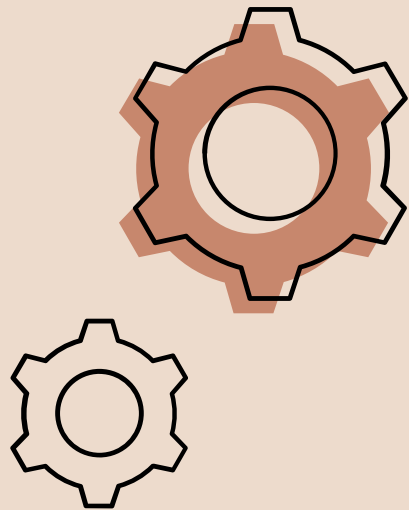


# METODOLOGÍA



## Proceso aplicado:

- Limpieza y transformación de datos
- Codificación de variables categóricas
- División en entrenamiento y prueba (80/20)
- Entrenamiento de modelos:
- Regresión Lineal (modelo base)
- Random Forest Regressor (modelo final)



## Métricas utilizadas:

- MSE (Error Cuadrático Medio)
- $R^2$  (Coeficiente de Determinación)

# RESULTADOS DE LOS MODELOS

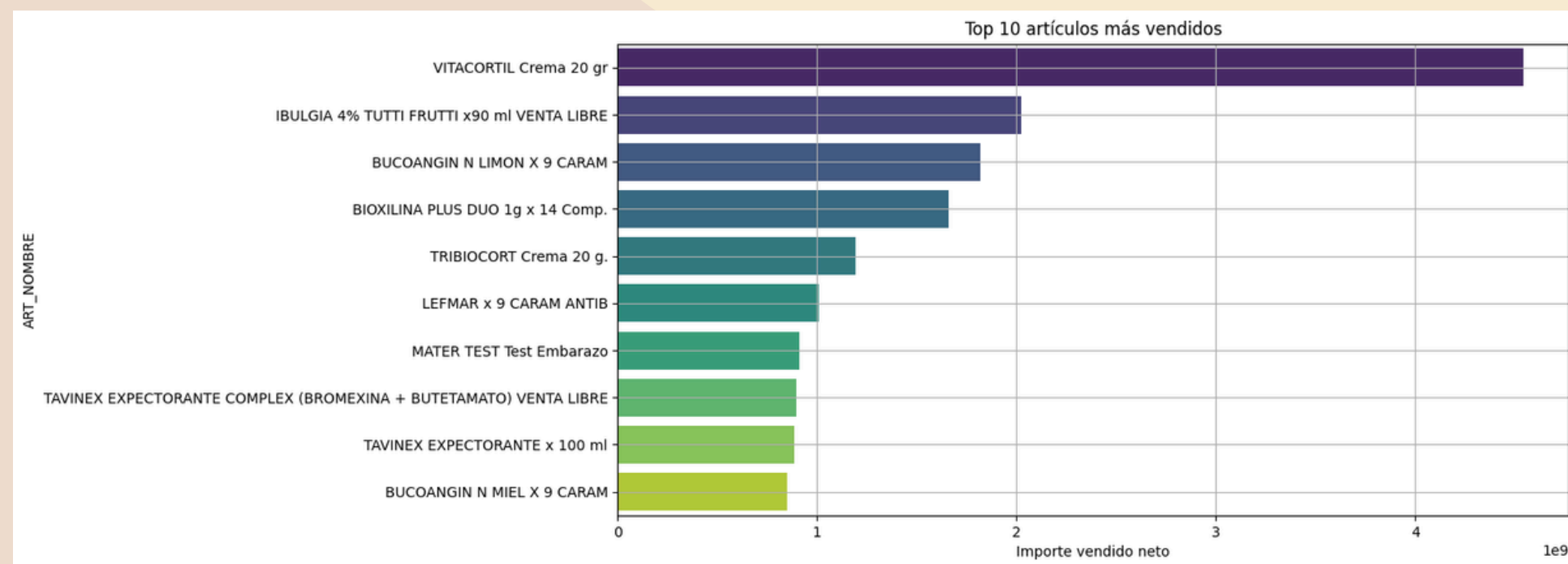
Modelo	MSE	R <sup>2</sup>
Regresión Lineal	119.840.733.939	0.096
Random Forest	98.833.459.198	0.255 <input checked="" type="checkbox"/>



# VISUALIZACIONES CLAVE

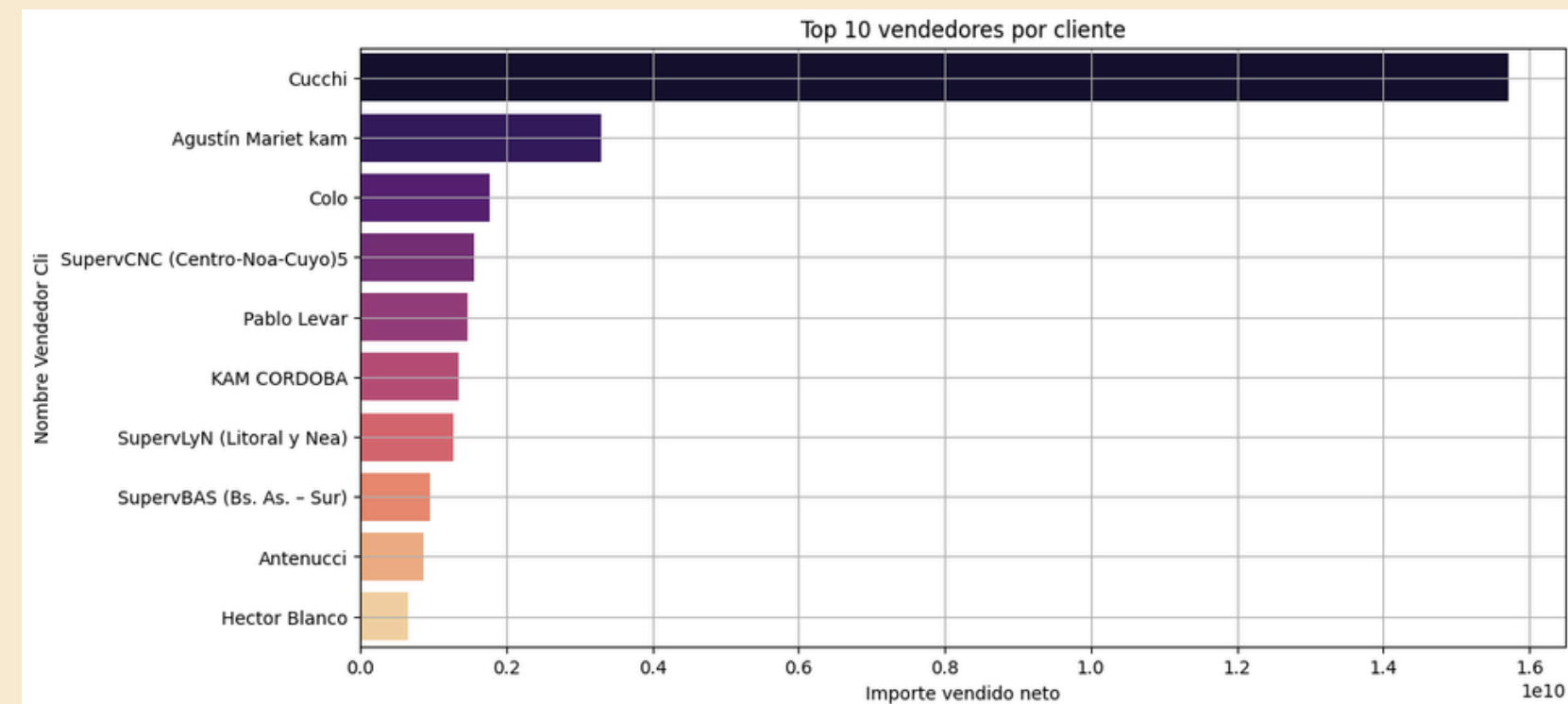
1.

Top 10 artículos más vendidos:



2.

Top 10 vendedores por cliente:

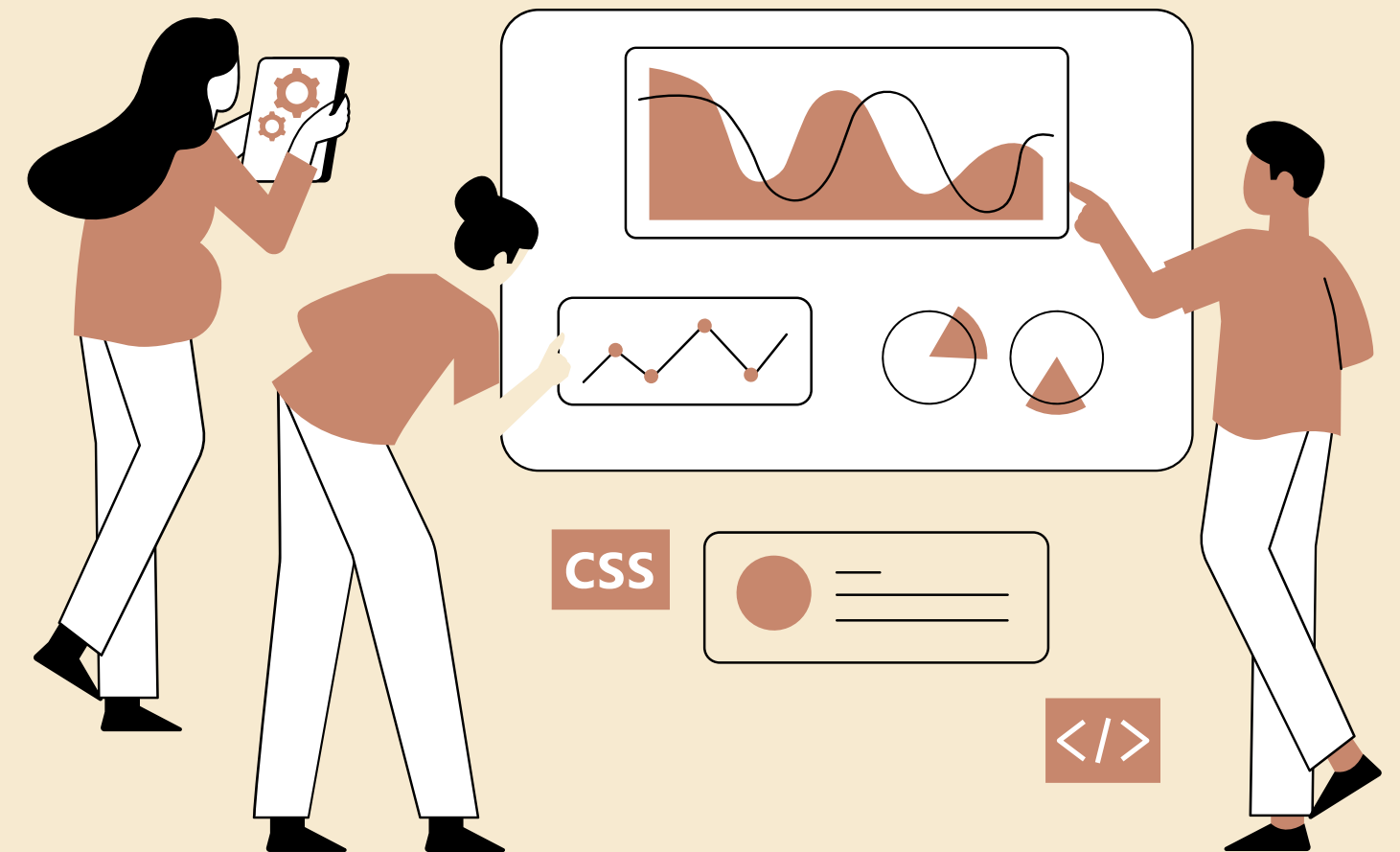


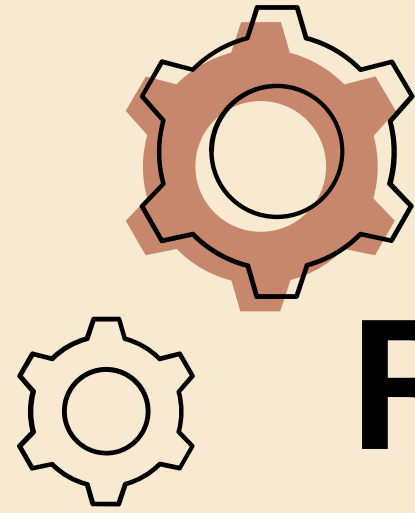


# INSIGHTS EJECUTIVOS

## Hallazgos principales:

- Los productos con mayor venta representan el 60% del volumen total.
- Ciertos vendedores concentran más del 40% del importe vendido.
- El modelo predice mejor en artículos con comportamiento estable y clientes frecuentes.
- Hay oportunidades en provincias con baja presencia de vendedores top.





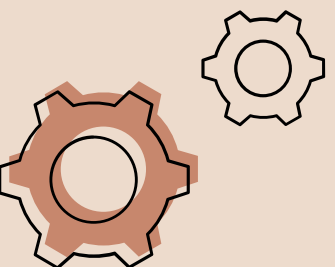
# RECOMENDACIONES

¿Qué podemos hacer con esto?

- Enfocar las campañas en los artículos más predictivos.
- Asignar vendedores top a zonas menos desarrolladas.
- Optimizar el inventario por provincia según estacionalidad.
- Continuar mejorando el modelo con más variables: clima, promociones, días hábiles.

Este proyecto demuestra cómo el uso de Machine Learning permite:

- Prever ingresos
- Detectar patrones de venta
- Tomar decisiones comerciales más informadas



# GRACIAS

Presentado por: Mirlenis Tovar  
Industria farmacéutica – Data Science II  
Junio 2025

