



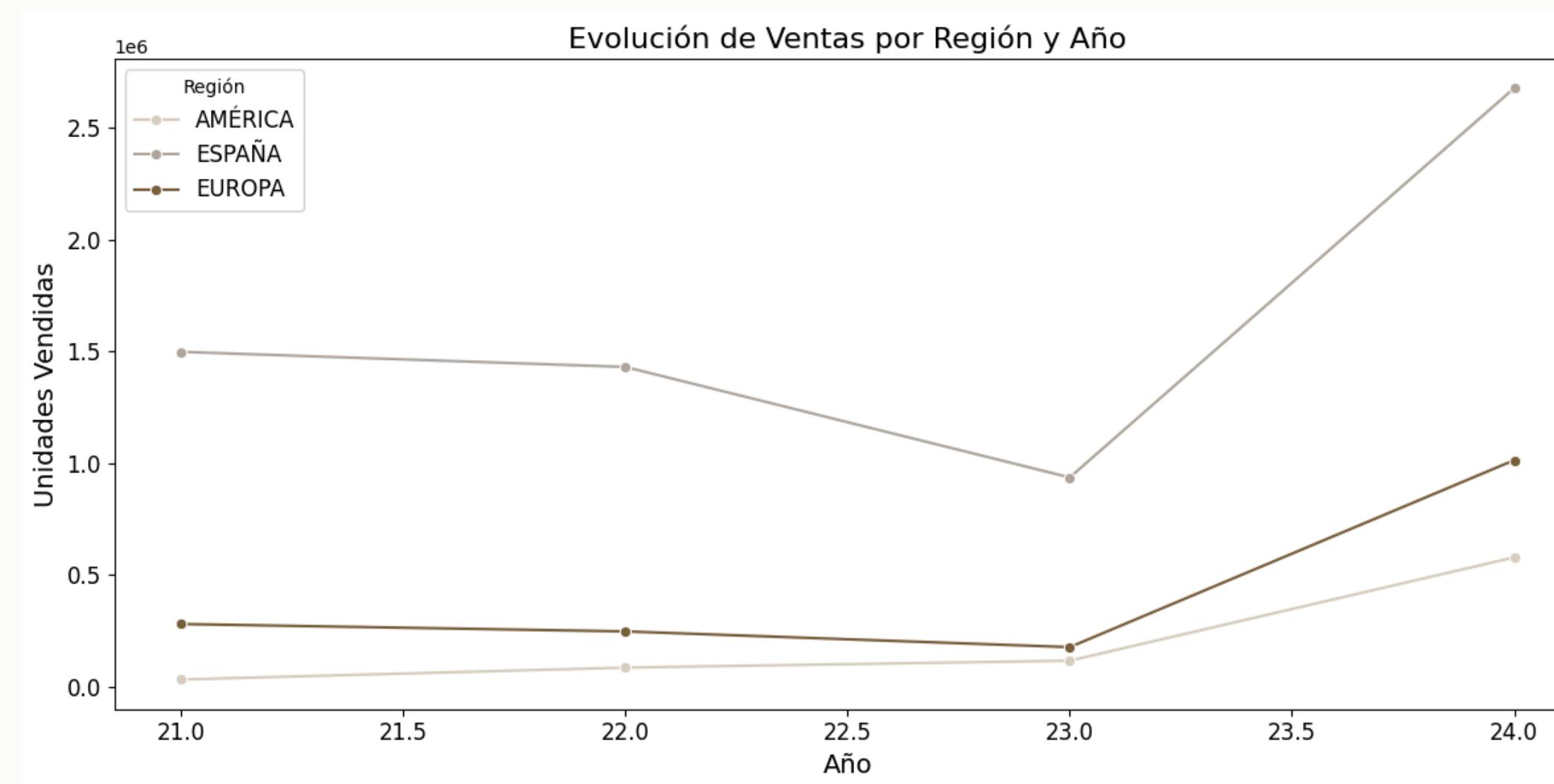
# *PREDICCIÓN DE VENTAS POR REGIÓN Y PRODUCTO USANDO MACHINE LEARNING*



# Introducción

**Objetivo:** el objetivo inicial era desarrollar un modelo que permitiera predecir las ventas totales del 2025 teniendo en cuenta las ventas del año 2021 hasta 2024

**Buscábamos optimizar la planificación de inventarios y reducir los costes operativos. Inicialmente, asumimos que un único modelo podría capturar las tendencias generales de ventas**



# PRIMER MODELO

1) *20697 rows × 14 columns ---> 747 rows × 5 columns*

year	month	region	product_id	sold_units

2) *REGION (unidades)*

*47 países separados en 5 regiones*

*Zona Geográfica*

*region\_AMÉRICA* 175

*region\_ESPAÑA* 287

*region\_EUROPA* 264

*region\_RESTO DEL MUNDO* 21

**REGRESIÓN LINEAL, RANDOM FOREST Y  
XGBOOST -> R2 MUY BAJO  
LAS VARIABLES NO EXPLICAN EL OBJETIVO**

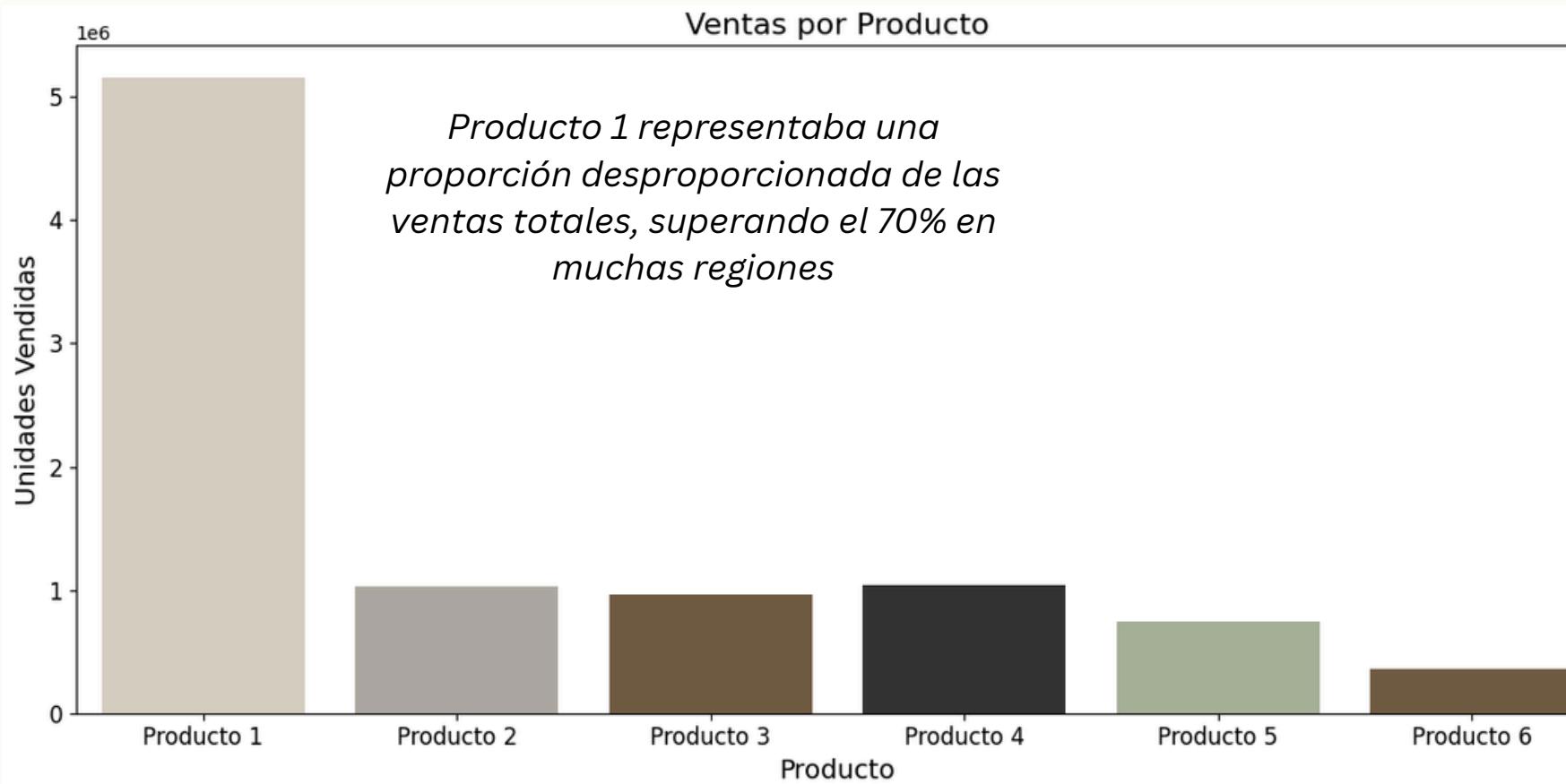
3) *PRODUCT\_ID*

*5 GRUPOS PARA EL PRECIO*

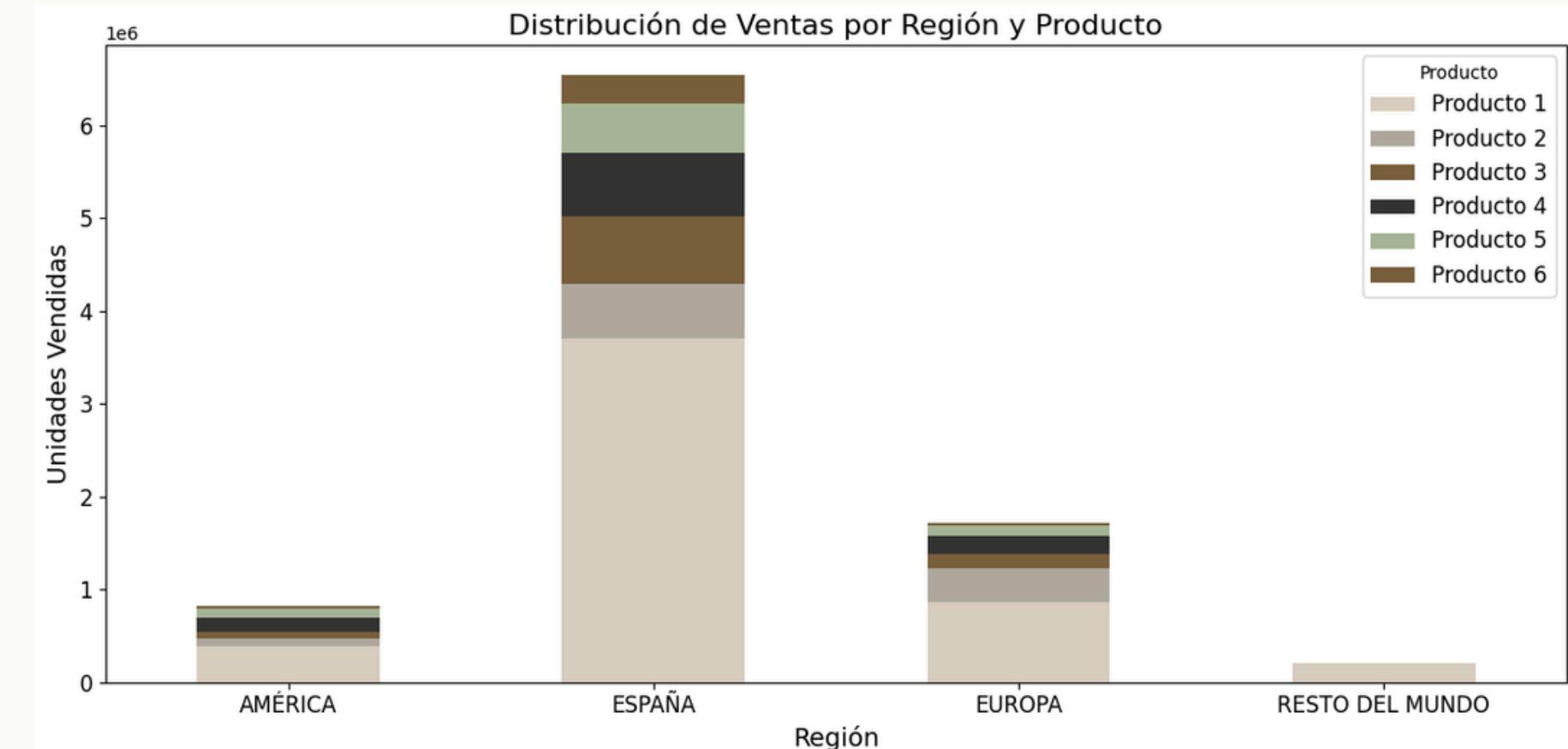
1. Grupo 1: 0-6.3 → Precios muy bajos
2. Grupo 2: 6.3-10.377 → Precios bajos
3. Grupo 3: 10.377-13.57 → Precios medios
4. Grupo 4: 13.57-100 → Precios altos
5. Grupo 5: > 100 → Precios muy altos (outliers)

# DIFICULTADES

## Desequilibrio en los datos



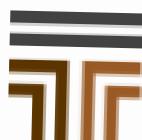
## Patrones regionales diferentes



Generó un fuerte sesgo en el modelo, que tendía a enfocarse únicamente en este producto y descuidaba otros productos menos representados.

Producto 1 dominaba claramente las ventas, mientras que en América y Europa, las ventas estaban más distribuidas entre varios productos. Esto hacía que un modelo único no pudiera capturar estos patrones regionales contradictorios.

MODELO UNICO NO CAPTURA BIEN LAS DIFERENCIAS REGIONALES NI LA DISTRIBUCIÓN DESIGUAL DE LOS PRODUCTOS





# AJUSTES

*Decidimos entrenar modelos separados para cada región. Esto permitió que cada modelo se enfocara en patrones locales, eliminando el ruido causado por patrones contradictorios entre regiones."*

## 1) REGIÓN RESTO DEL MUNDO NO LA VOY A ESTUDIAR

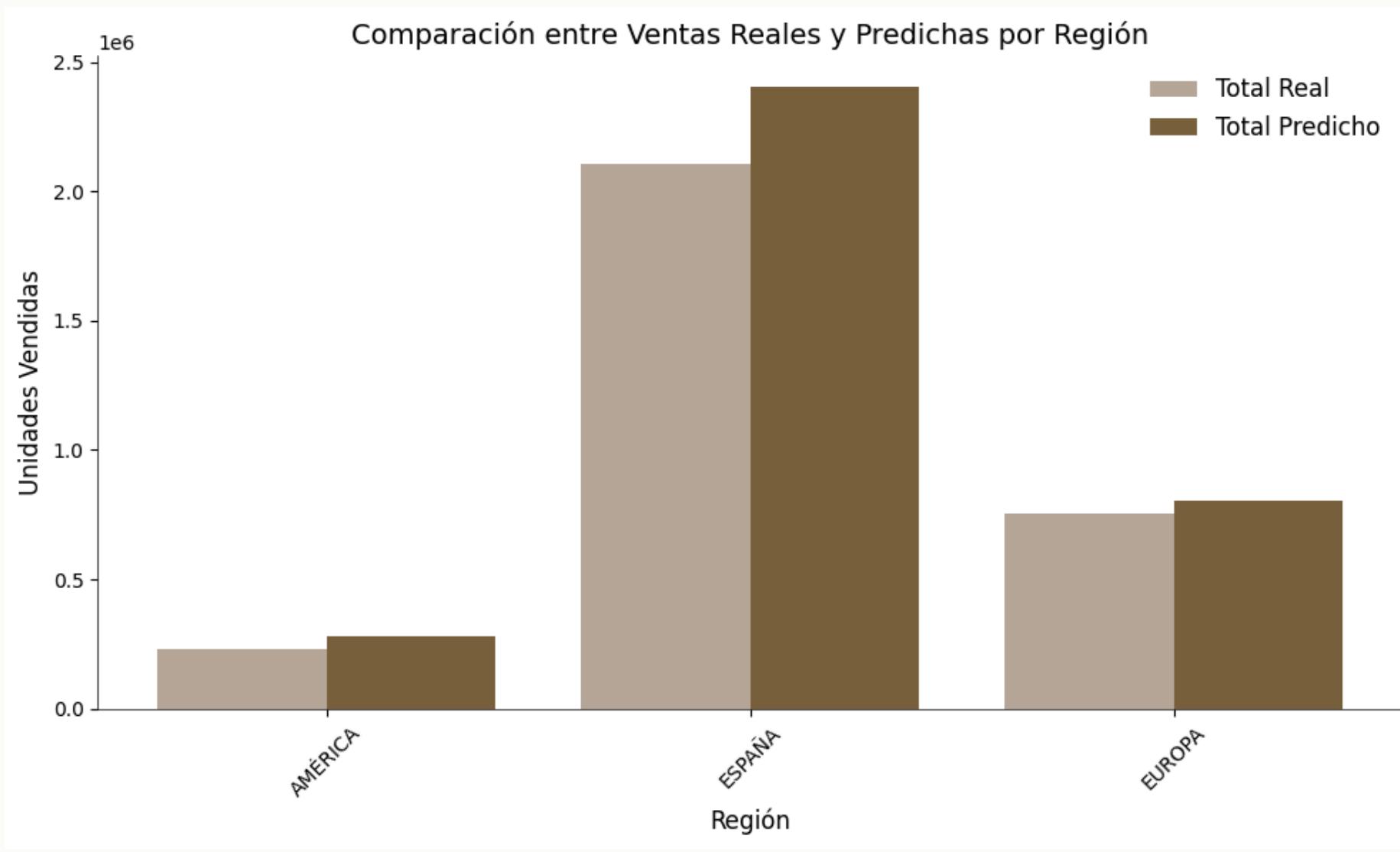
- dispongo de 21 observaciones
- no comparte patrones de venta con ninguna región

## 2) DISPONGO DE 726 OBSERVACIONES SEPARADAS EN LAS 3 REGIONES

- *region\_AMÉRICA: Mejor modelo: Regresión Lineal R^2: 0.17*
- *region\_ESPAÑA: Mejor modelo: Regresión Lineal R^2: 0.39*
- *region\_EUROPA: Mejor modelo: Random Forest R^2: 0.37*

## 3) AJUSTO RANGOS DE PRECIOS

- *Se crearon nuevas categorías basadas en las ventas totales, aumentando de 6 a 10 categorías.*
- *Objetivo: Mejorar la granularidad en los rangos de precios para reflejar mejor las diferencias en los patrones de venta.*



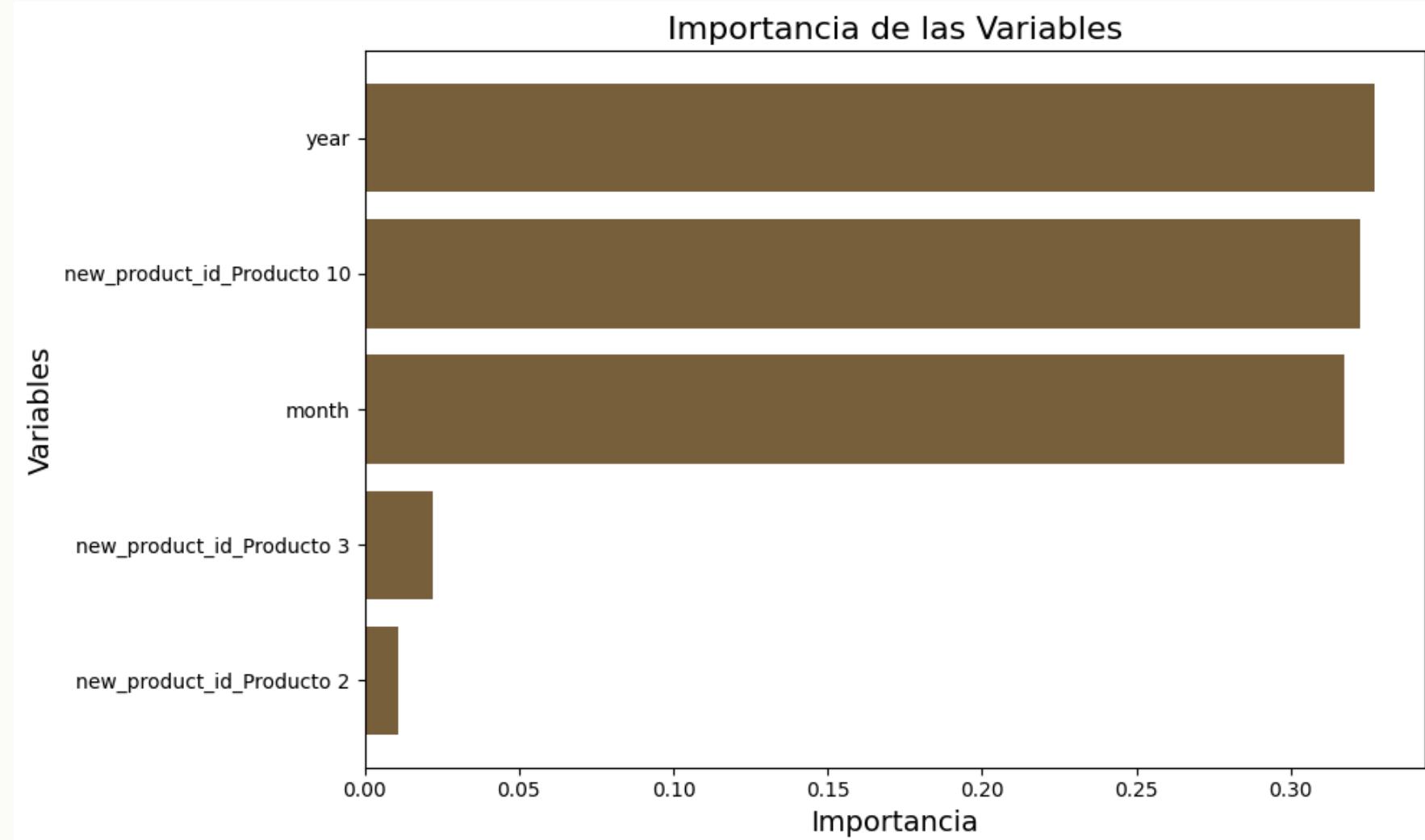
**Región: AMÉRICA RMSE: 6219.33, R<sup>2</sup>: 0.65**  
**Región: ESPAÑA RMSE: 25941.27, R<sup>2</sup>: 0.40**  
**Región: EUROPA RMSE: 15070.50, R<sup>2</sup>: 0.63**



# MODELO FINAL

Regresión Lineal, Random Forest, XGBoost

El modelo de Random Forest fue el mejor para capturar interacciones no lineales



# IMPACTO EN EL NEGOCIO



## GESTIÓN DE INVENTARIOS

América: el modelo predice 276,994 unidades vendidas, un 20% más que las ventas reales actuales (230,042 unidades). proyectada.

En Europa: 803,956 unidades predichas frente a las 752,463 reales -> ajustar el inventario



## MARKETING DIRIGIDO

Diseñar campañas promocionales para regiones con menores ventas predichas o para productos menos representados como para el resto del mundo



## PRODUCCIÓN EFICIENTE

Ajustar la producción para enfocarse en los productos y regiones más demandados  
Producto 10 y 3: productos deben ser priorizados en la producción para satisfacer la demanda



## OPTIMIZACIÓN DE PRECIOS POR PRODUCTO

Ajustar los precios de los productos en función de las regiones donde tienen mayor impacto en las ventas o donde las predicciones muestran menor demanda

# CONCLUSIÓN

01

## ÉXITO DEL MODELO

El modelo Random Forest logró predicciones más precisas en regiones como América ( $R^2=0.65$ ) y Europa ( $R^2=0.63$ ), ayudando a identificar tendencias claras en estas áreas.

02

## ESPAÑA

España: mayor volumen de ventas reales (2,106,046 unidades), mostró una menor precisión en las predicciones ( $R^2=0.40$ ) lo que indica que sus patrones de venta son más complejos

Nature  
Tecno

04

## AJUSTES EN EL MODELO

Integrar aprendizaje continuo para actualizar las predicciones en tiempo real con nuevos datos

03

## AÑADIR FACTORES EXTERNOS

Añadir factores externos como datos macroeconómicos (PIB, inflación) o eventos estacionales (festividades, cambios climáticos) para mejorar la precisión





GRACIAS, HASTA LA PRÓXIMA!

