

## Tarea 01

# Predicción de Demanda en Retail con Machine Learning

BLANCA AZUCENA ORDUÑA LÓPEZ  
DANIEL MIRANDA BADILLO

## Resumen del problema

Se presenta un caso de estudio para el análisis de datos de la compañía rusa **\*\*1C Company\*\*** que opera una cadena de tiendas en múltiples ciudades. La compañía tiene el reto de modernizar su proceso de inventarios para eficientar los costos operativos, de almacén y maximizar sus operaciones.

Los problemas están asociados a:

- **Sobrecostos por sobrestock:** genera costos de almacenamiento y dificulta la rotación de mercancía.
- **Escasez de producto (quiebre de stock):** cuando no se tiene stock son ventas perdidas, generando un impacto en las ganancias y en satisfacción del cliente.

Actualmente se utilizan métodos tradicionales como promedios móviles, que no cuentan con la suficiente capacidad para analizar correctamente el comportamiento del stock en las tiendas y su comportamiento a lo largo de los meses del año. Se busca mejorar la métrica del error de predicción RMSE DE 11 unidades a menos de 5 unidades para alcanzar el primer objetivo de margen operativo que consiste en mejorar la rotación del inventario de 6.2x a 9x anual.

## Solución propuesta

Se evaluaron dos modelos de machine learning: **regresión lineal** y **random forest**. La variable objetivo es el **número de artículos por item por mes** que se deben tener en inventario para cubrir la demanda sin caer sobre stock con base en la información histórica. El detalle se analiza a **nivel producto-tienda**.

Los datos:

**2,935,849**  
Transacciones

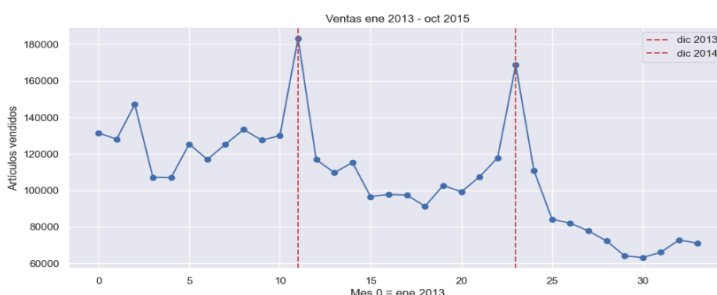
**22,170**  
Productos diferentes

**84**  
Categorías

**34**  
meses

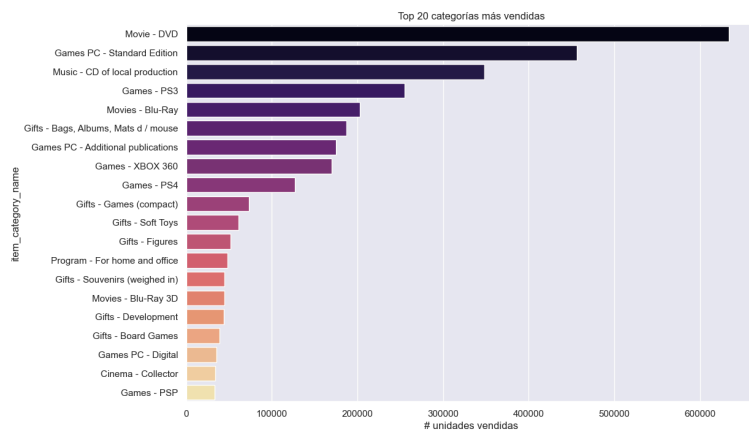
**1,034**  
días

## Análisis exploratorio: hallazgos relevantes

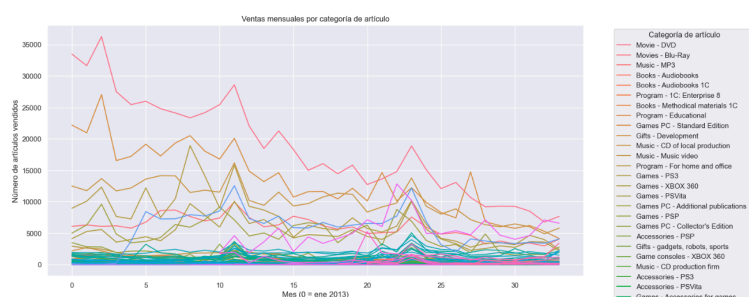


Vemos una tendencia estacional, con decremento de ventas los últimos 3 años.

Hay un pico de ventas muy marcado en diciembre



El comportamiento de las categorías top ha cambiado a lo largo de los años.



Por ejemplo, los DVD y blue-ray han disminuido, probablemente por la transición a plataformas digitales.



Identificamos 14 tiendas sin ventas en los últimos 2 meses. Debemos verificar si son tiendas que han cerrado.

Además en el **85% de las combinaciones tienda-producto no presentan ventas en un mes dado**, es decir, existe una demanda intermitente que pueden estar causando sobrestock.

### Limpieza y outliers:

- Para evitar predicciones erróneas o extremas se excluyeron datos con cantidades de venta mayores a 1,000 unidades por día y precios por ítem mayores a 100,000

### Datos interesantes:

- **Transacciones por gastos de envío (EMS).** (Es la transacción más alta y no influye en el stock, solo genera ruido)
- **Transacciones de categorías** [Pagos en línea, pago de tarjetas, suscripciones, costos de envío, tickets de conciertos]. (No influyen en la cantidad de stock, generan ruido)
- **Compra de artículos digitales desde Tienda en Línea.** (Los videojuegos digitales, licencias de software digitales, o libros digitales. No forman parte de un stock físico )

## Evaluación detallada del modelo

RMSE objetivo : menos de **5 unidades**

Modelo	RMSE	Observaciones
Actual	11	Deficiente para grandes volúmenes de información y variaciones.
Regresión lineal	0.9824	Deficiente para capturar estacionalidad.
Random Forest	<b>0.9743</b>	Captura patrones complejos y estacionalidad.

Con la implementación de un modelo de **Random Forest** el error se reduce en promedio a **1 unidad por producto al mes**.

### Fortalezas del modelo

- Precisión, se logra un error promedio menor a 1 en ítem por mes.
- Efectivo para capturar comportamientos no lineales como la estacionalidad mensual y el comportamiento por tienda.
- Muy bueno para productos de rotación estable y tiendas con ventas consistentes.

### Debilidades y limitaciones

- Depende de los datos históricos, si hay datos faltantes o erróneos, la calidad de las predicciones baja.
- Al introducir nuevos productos, las predicciones no serán las mejores pues solo considera el comportamiento de la categoría.
- El modelo no tiene visibilidad para predecir picos de demanda atípicos, como los que generan las promociones.
- No es recomendable utilizar el modelo para la apertura de nuevas tiendas, porque depende de las ventas históricas.

### Recomendaciones y accionables

- Verificar las condiciones de las 14 tiendas sin ventas, para evitar el envío de stock innecesario.
- Implementación de un proceso híbrido para el cálculo de inventarios, con enfoque de la fuerza **manual** de trabajo en:
  - Cálculo de stock para nuevos productos.
  - Cálculo de inventario para la apertura de nuevas tiendas usando tiendas espejo.
  - Mapeo de causa-raíz de productos con demanda intermitente por tienda.

Predict Future Sales

Late Submission



- Overview
- Data
- Code
- Models
- Discussion
- Leaderboard
- Rules
- Team
- Submissions

Leaderboard

Raw Data Refresh

YOUR RECENT SUBMISSION



submission 23 ene 2026, 19:20:32 GMT-6 · 2s Execution Duration

Submitted by Azucena Orduña · Submitted 11 minutes ago

Score: 1.02184  
Private score: 1.02444

Jump to your leaderboard position

Search leaderboard

Public Private

This leaderboard is calculated with approximately 35% of the test data. The final results will be based on the other 65%, so the final standings may be different.