

Predicción de Demanda en Retail con Machine Learning

Andrés Padrón Quintana. Manuel de la Tejera González.

1) Decisión ejecutiva

Este proyecto implementa un modelo predictivo de demanda mensual a nivel tienda–producto para reducir el error de pronóstico y habilitar decisiones operativas más rápidas y consistentes en inventario. Con un pipeline end-to-end (datos, features, entrenamiento, validación, generación de predicciones), se mejoró el desempeño frente a un baseline, reduciendo el error en Kaggle (proxy de generalización) de 1.21907 a 1.18463 en RMSE público (mejora relativa $\approx 2.8\%$).

En términos operativos, esta mejora implica menor desviación entre lo planificado y lo vendido, reduciendo costos por sobrestock y disminuyendo la probabilidad de quiebres de stock en productos clave.

2) Qué hicimos

- Consolidamos las ventas históricas diarias en un dataset mensual por tienda y producto, alineado con el objetivo del reto (pronóstico del mes 34).
- Construimos variables predictoras (features) para capturar patrones de demanda: historial reciente (lags), promedios agregados por grupos y señales de estacionalidad por mes.
- Entrenamos y evaluamos usando un split temporal (entrenar con meses anteriores y validar con el mes 33), evitando “ver el futuro”, lo que si hicimos fue que generamos predicciones finales para el set de prueba y las exportamos c

3) Datos y hallazgos del EDA (antes de modelar)

Escala del problema y granularidad operativa.

El histórico contiene 2,935,849 registros diarios; al agregarse a nivel mensual se obtienen 1,609,124 observaciones (pares tienda–producto por mes), con meses 0 a 33, 60 tiendas y 21,807 productos únicos. Esto confirma un escenario de alta dimensionalidad y heterogeneidad por tienda.

Comportamiento del target (*item_cnt_month*).

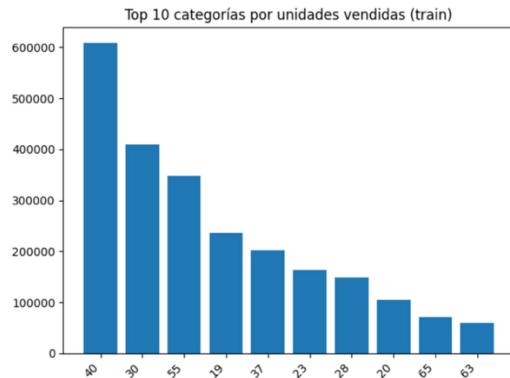
La demanda mensual por par tienda–producto está concentrada en valores bajos, lo cual es típico: mediana 1, percentil 75.2, media ≈ 2.03 y desviación estándar ≈ 2.58 . Para estabilidad del entrenamiento y alineación con la métrica del reto, se utiliza el *clipping* del target/predicciones en el rango [0, 20], evitando que outliers dominen el RMSE.

Métrica de demanda mensual (<i>item_cnt_month</i>)	Valor
Media	2.03
Desv. estándar	2.58
Mediana	1

Percentil 75	2
Máximo (clip)	20

Insight visual por categorías.

La figura de “Top 10 categorías por unidades vendidas (train)” muestra una concentración importante del volumen en pocas categorías (por ejemplo, IDs **40, 30, 55, 19, 37, 23, 28, 20, 65, 63**). Esto sugiere que un modelo global debe capturar bien patrones dominantes, pero también generalizar a categorías con menos volumen (donde hay más riesgo de error y *cold-start* relativo).



Implicación para negocio: dado que la mayoría de pares tienda–producto tienen ventas bajas en un mes, el valor viene de capturar bien la señal temporal (persistencia/estacionalidad) con features tipo **lags** y agregaciones, reduciendo errores que se traducen en sobrestock o quiebres.

4) Resultados (métrica y comparación)

Métrica: RMSE (Root Mean Squared Error). Menor es mejor.

Modelo	Validación (mes 33) RMSE	Kaggle Public RMSE	Notas
Baseline LightGBM	≈ 0.4982	1.21907	Base rápida y estable
XGBoost (final)	≈ 0.4773	1.18463	Mejor generalización en leaderboard

Ganancia vs baseline (Kaggle Public):

- Delta absoluto: $1.21907 - 1.18463 = 0.03444$
- Mejora relativa: $0.03444 / 1.21907 \approx 2.8\%$

Análisis del error y comportamiento del modelo.

El análisis de los datos muestra que la mayoría de los productos vende pocas unidades al mes, lo que implica que en muchos casos existe información histórica limitada para anticipar

cambios en la demanda. En estos productos, el modelo tiene mayor dificultad para ajustar con precisión, mientras que en aquellos con ventas más constantes el pronóstico es más estable.

Adicionalmente, el modelo se apoya principalmente en el comportamiento pasado de las ventas y no incorpora información sobre promociones, cambios de precio o eventos extraordinarios. Esto hace que el pronóstico funcione bien para la demanda habitual, pero que tienda a quedarse corto cuando ocurren aumentos inesperados. En conjunto, los errores más relevantes se concentran en situaciones poco frecuentes y difíciles de anticipar, lo cual es consistente con el enfoque utilizado y refuerza la necesidad de incorporar nuevas señales en futuras iteraciones para mejorar la precisión en estos casos.

5) Interpretación para negocio

Una reducción de RMSE significa que, en promedio, el pronóstico está más cerca de la demanda real. En retail, mejoras de este tipo se traducen en:

- **Menos sobrestock:** menor necesidad de inventario “colchón”, reduciendo costos de almacenamiento y liquidaciones.
- **Menos quiebres de stock:** mejor anticipación de picos de demanda, reduciendo ventas perdidas y mejorando experiencia del cliente.
- **Planeación operativa más consistente:** el pronóstico mensual por tienda–producto habilita decisiones de reposición, asignación y logística con reglas más estandarizadas.
- **Reproducibilidad del proceso:** se reemplaza un enfoque reactivo/manual por un pipeline automático que puede correrse de manera periódica.

6) Riesgos y limitaciones

1. **Cold-start (poca historia):** pares tienda–producto con pocas observaciones tienen menos señal para aprender patrones robustos.
2. **Validación interna limitada:** validar solo en el mes 33 puede no representar toda la variabilidad; Kaggle funciona como prueba adicional de generalización.
3. **Clipping [0,20]:** restringir predicciones al rango del reto mejora estabilidad en la métrica, pero en producción debe alinearse a políticas de negocio (p. ej. mínimos de abastecimiento, límites por capacidad).

7) Próximos pasos (acciones concretas)

1. **Validación temporal robusta (rolling/backtesting):** evaluar múltiples ventanas (p. ej. validar en meses 30–33) para medir estabilidad y evitar sobreajuste a un solo mes.
2. **Features de mayor señal:** incorporar proxies de **precio/promoción**, disponibilidad, calendario (festivos/eventos) y enriquecer agregaciones por categoría.
3. **Segmentación operativa:** entrenar modelos por clusters (tiendas similares / familias de producto) y comparar contra el modelo global; suele mejorar desempeño y facilita decisiones específicas.

8) Anexo: evidencia de Kaggle

- Screenshot leaderboard + timestamp:

The screenshot shows the 'Submissions' page of a Kaggle competition. The left sidebar includes links for Home, Competitions (which is selected), Datasets, Models, Benchmarks, Game Arena, Code, Discussions, Learn, More, Your Work, Viewed, and Predict Future Sales. The main content area displays a table of submissions. At the top right, there is a circular badge with '0/2'. Below it, a note states: 'You selected 0 of 2 submissions to be evaluated for your final leaderboard score. Since you selected less than 2 submissions, Kaggle auto-selected up to 2 submissions from among your public best-scoring unselected submissions for evaluation. The evaluated submission with the best Private Score is used for your final score.' A legend indicates that checked boxes represent 'Submissions evaluated for final score'. The table has columns for Submission and Description, Private Score, Public Score, and Selected. Three submissions are listed:

Submission and Description	Private Score	Public Score	Selected
submission_xgb.csv Complete (after deadline) · 14h ago · xgboost, submission TAREA 1 METODOS DE GRAN E...	1.18434	1.18463	<input type="checkbox"/>
submission_best.csv Complete (after deadline) · 1d ago · TAREA 1 METODOS DE GRAN ESCALA ITAM MCD	1.20930	1.22045	<input type="checkbox"/>
submission_lgb.csv Complete (after deadline) · 3d ago · PRUEBA TAREA 1_ METODOS GRAN ESCALA ITAM	1.20813	1.21907	<input type="checkbox"/>