Retail Sales Forecasting for Corporación La Favorita

Evaluating Predictive Frameworks for the Guayas Region

Target Period: August 16, 2016 – August 15, 2017

Models: XGBoost, Facebook Prophet, SARIMAX

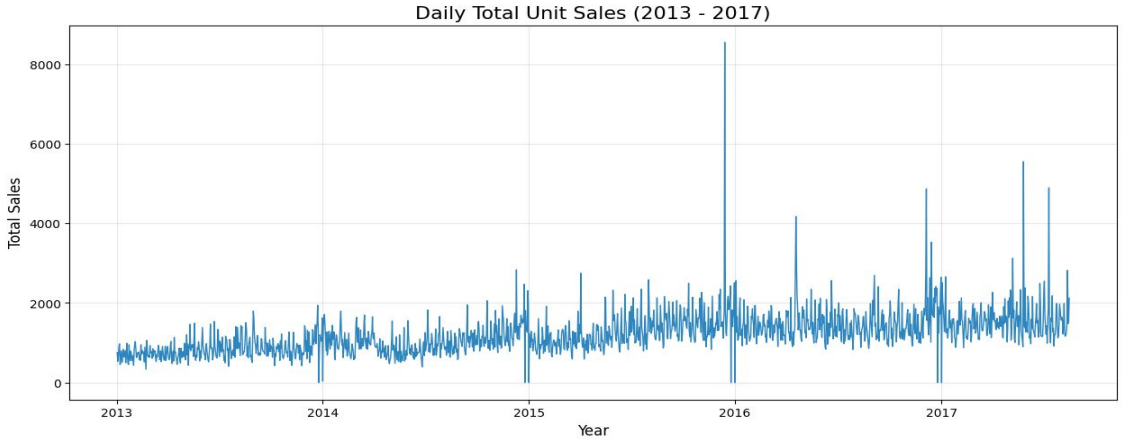
Project Objectives

Scalability: Process 125M+ rows of transaction data efficiently.

Strategic Focus: Isolate demand patterns in the Guayas region for Top 3 product families

Benchmarking: Compare Modern ML (XGBoost) vs. Additive Models (Prophet) vs. Statistical Models (SARIMAX).

Goal: Identify the most accurate engine for automated replenishment and inventory planning.



Data Strategy & Filter Logic

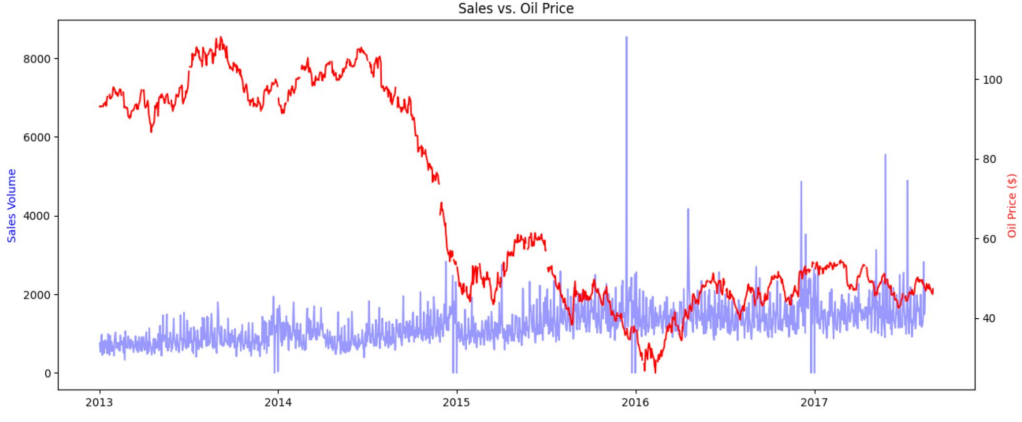
Chunk Processing: Handled 5GB dataset in 1M-row increments for system stability.

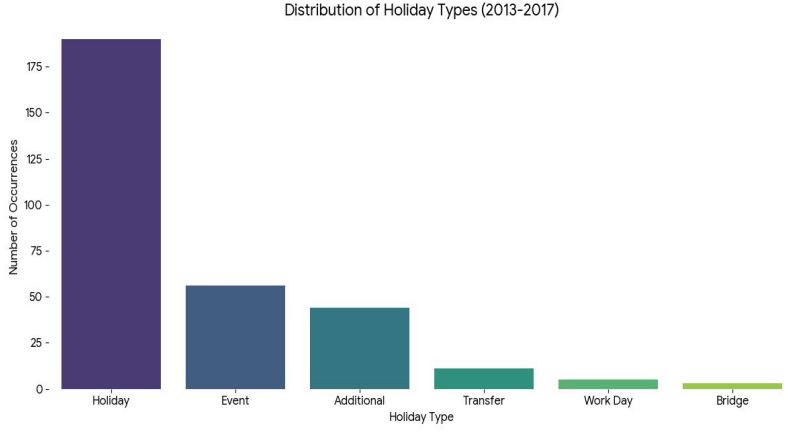
Segmentation: Focused on Grocery I, Beverages, and Cleaning categories.

Feature Engineering:

- **Excluded:** Oil prices (correlation analysis proved negligible impact).
- **Included:** National/Local holidays and 2016 Earthquake recovery data.

Outcome: Created a high-signal dataset focused on actual consumer drivers.





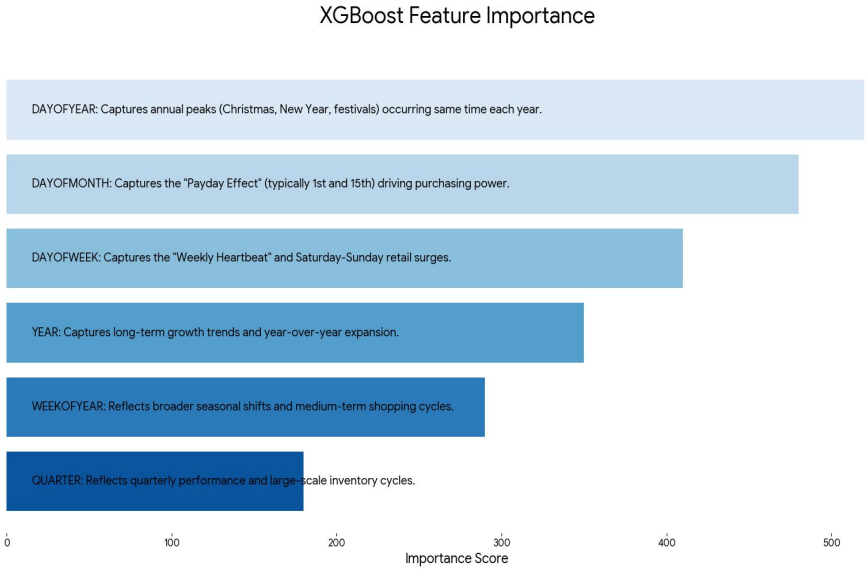
XGBoost – The Champion

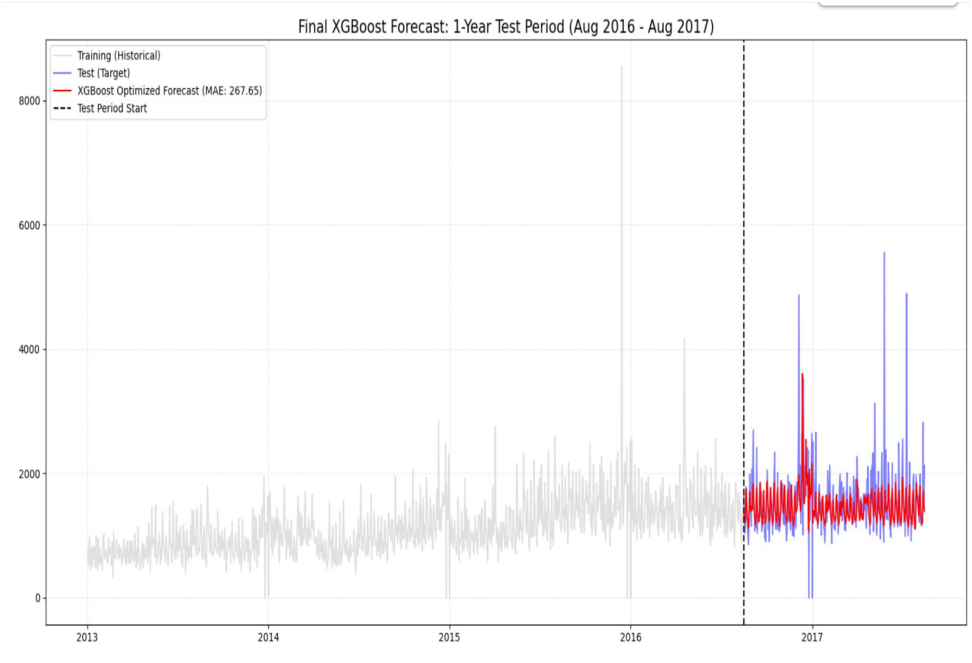
Technical Setup: Up to 5,000 Estimators | Max Depth 3 | 100-round Early Stopping.

Breakthrough: Feature Elimination Study showed that removing redundant "weekend flags" improved precision.

Key Finding: XGBoost is very good at identifying non-linear spikes.

Final Performance: **MAE: 267.65 | RMSE: 483.87.**





Facebook Prophet – Explaining the "Why"

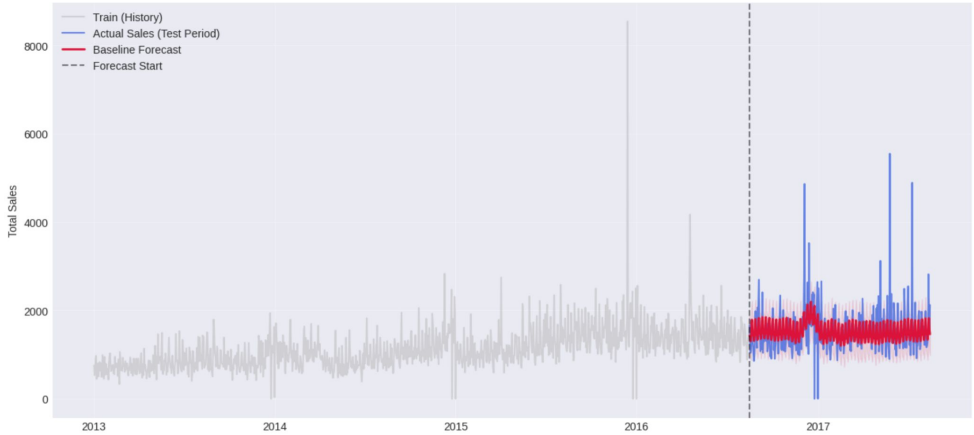
Core Logic: Additive decomposition of Trend, Weekly, and Yearly cycles.

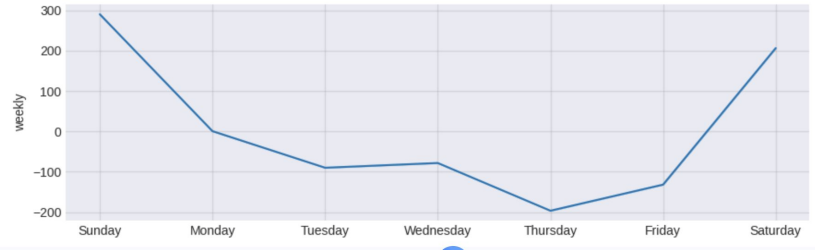
The "Weekly Heartbeat": Successfully isolated Saturday and Sunday as the primary volume drivers.

Comparison: Baseline (Calendar-focus) vs. Enhanced (External Regressors).

Verdict: The **Baseline model** generalized better, proving that regular shopping rhythms are more reliable than specific holiday weights.

Final Performance: **MAE 285.75 | RMSE 485.06**





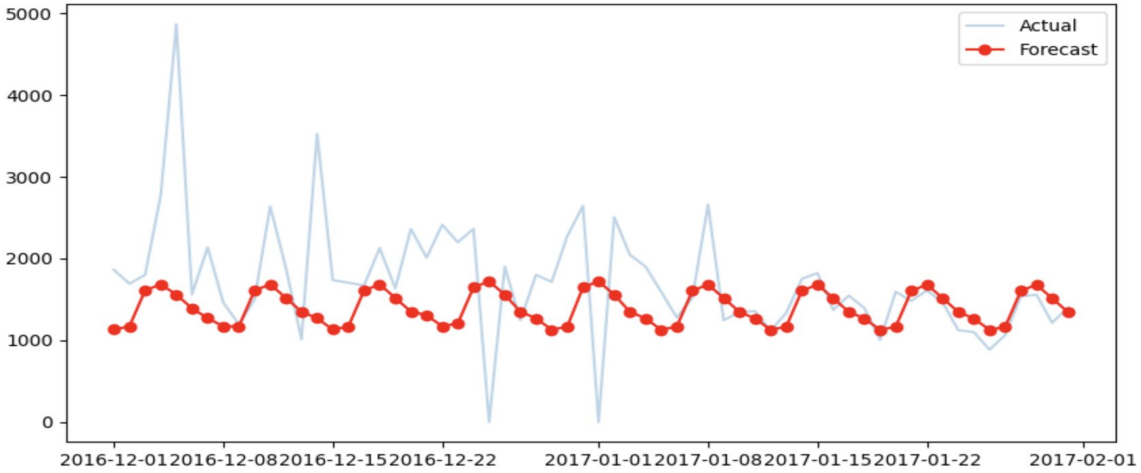
SARIMAX – The Statistical Safety Net

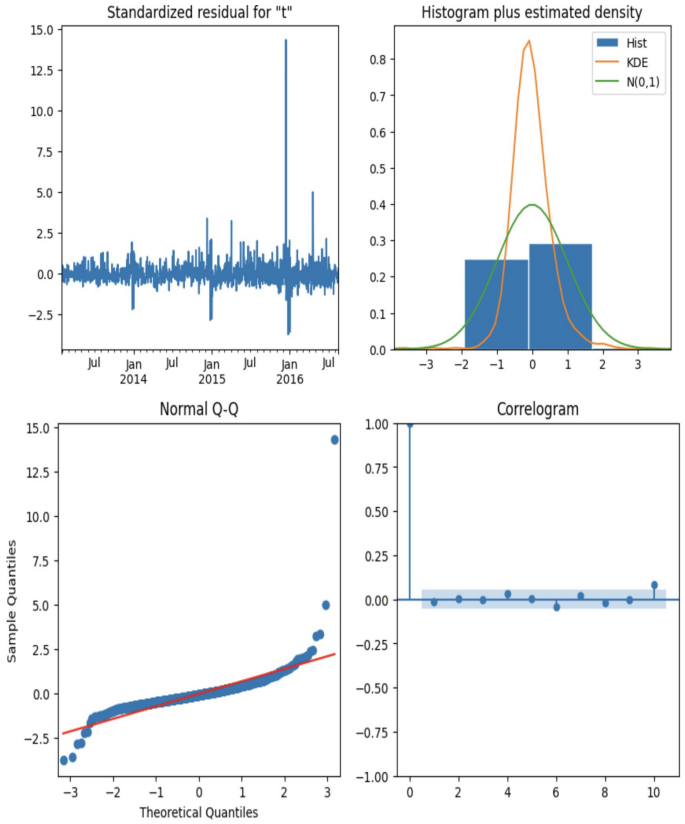
Stability: Applied first-order differencing (d=1) to stabilize the mean.

Technical Refinement: Manually introduced **Seasonal Autoregression (p=1)** to maintain the 7-day sales cycle.

Impact: Manual tuning resulted in a **12% accuracy improvement** over automated settings.

Final Performance: **MAE: 301.25 | RMSE: 520.22.**





Quick Performance Comparison

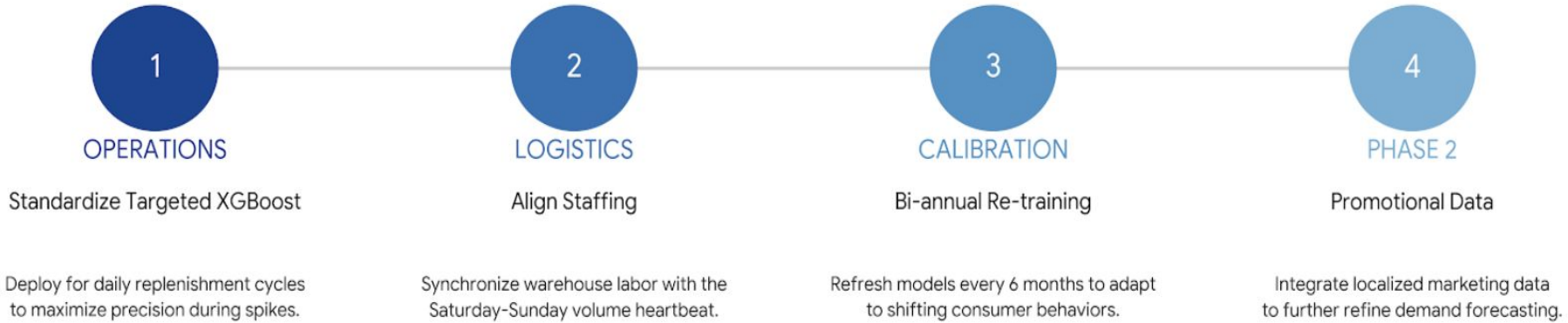
MAE & RMSE Comparison — XGBoost Leads

Model Strategy	MAE (Error)	RMSE (Sensitivity)	Strategic Verdict
XGBoost (Targeted)	267.65	483.87	**Champion**: Most Precise
Prophet (Baseline)	285.75	485.06	**Explainer**: Best Trend Logic
SARIMAX (Manual)	301.25	520.22	**Benchmark**: Statistical Floor

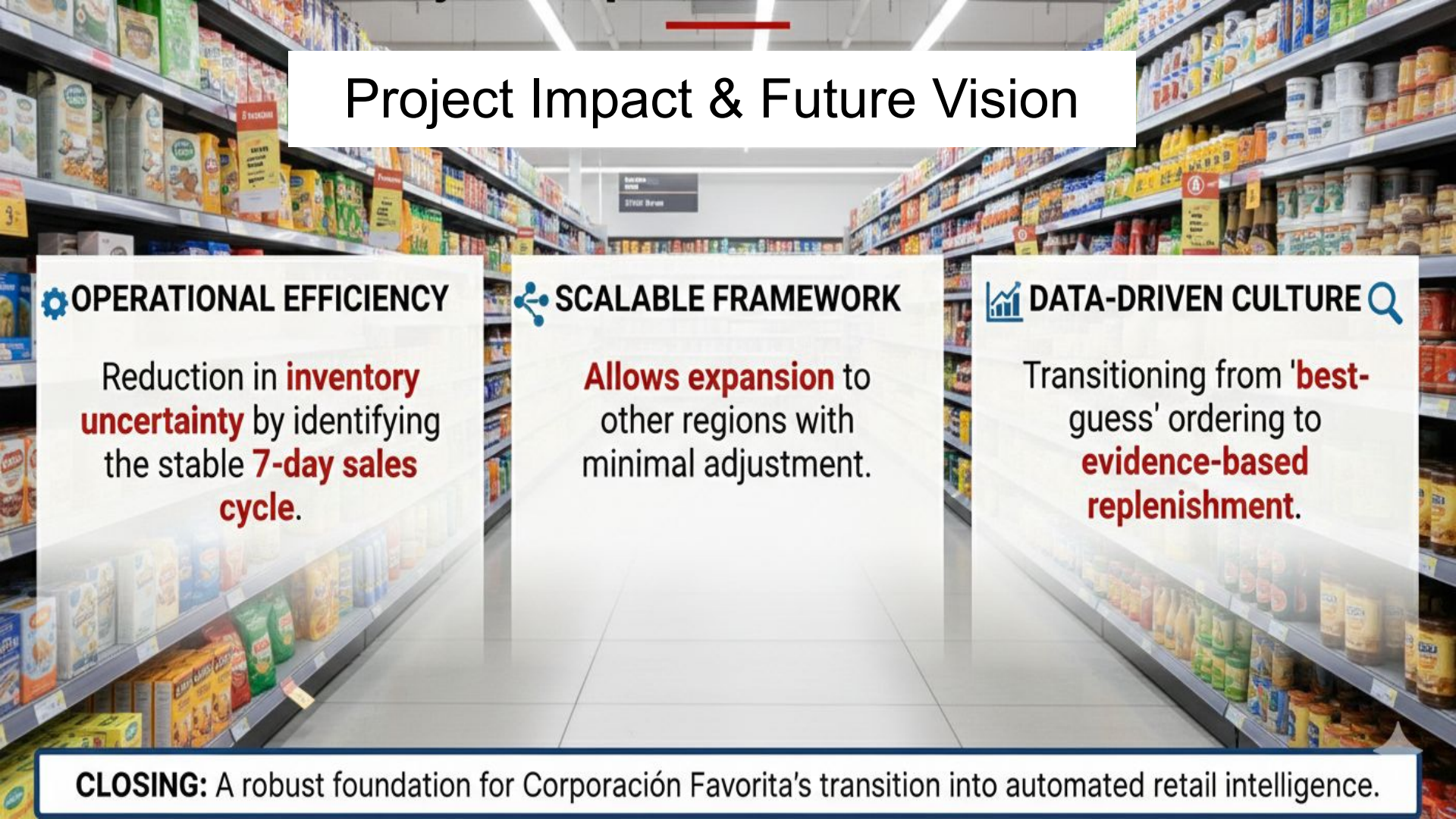
Key Strategic Insights

- **The Weekly Heartbeat:** Supply chain and staffing should be prioritized for Friday–Sunday peaks.
- **Simplicity Over Complexity:** In both XGBoost and Prophet, leaner feature sets generalized better during the 1-year test period.
- **Shock Resilience:** The models successfully distinguished between standard seasonal dips and unique events like the 2016 Earthquake.

Recommendations & Next Steps



Project Impact & Future Vision



OPERATIONAL EFFICIENCY

Reduction in **inventory uncertainty** by identifying the stable **7-day sales cycle**.

SCALABLE FRAMEWORK

Allows expansion to other regions with minimal adjustment.

DATA-DRIVEN CULTURE

Transitioning from '**best-guess**' ordering to **evidence-based replenishment**.

CLOSING: A robust foundation for Corporación Favorita's transition into automated retail intelligence.