STORE SALES DIVE ANALYSIS AND ACTION PLAN

DN4: Rachakonda Srikanth, Sameer Yelamarthi & Valerie Robert

INTRODUCTION & BUSINESS PROBLEM

This report details the development of a machine learning pipeline on Google Cloud Platform to predict daily sales for each product family across all stores. Using the DIVE (Discover, Investigate, Validate, Extend) framework, we translate the technical results from our final predictive model into a strategic analysis and an actionable plan designed to improve operational efficiency and drive business value

DIVE FRAMEWORK INSIGHTS

DISCOVER: INITIAL FINDINGS

The primary story these results tell is that **store sales operate on a highly predictable, structured rhythm.** Our BOOSTED_TREE_REGRESSOR model proved this by successfully explaining **79.3% of the variance** in daily sales with a Mean Absolute Error of **~$186**. Our forecast for the next 14 days predicts a sales volume of **~$8.6M**, a decrease from the prior period's holiday-inflated sales of ~$11.4M. This predictability is driven by a few key patterns: the consistent weekly shopping cycle, fundamental differences between store types, and the unique demand profile of each product family.

INVESTIGATE: DEEP 'WHY' EXPLORATION OF SALES PATTERNS

Our model predicts a sales decrease primarily because the previous 14-day period contained several major holidays which are absent in the forecast period. The deeper patterns that emerged show why specific factors are so influential:

- **The Weekly Rhythm:** Customers adhere to a strict weekly shopping routine. This is the core business mechanism explaining why the day of the week is a powerful predictor. Evidence shows average Sunday sales (~$825k) are **64% higher** than on the slowest day, Thursday (~$505k).
- **Store Strategy:** Store characteristics are critical because they represent different business strategies for store size, format, and target market. The data proves this, with high-volume **Type A** stores generating **3.5 times more sales** than low-volume **Type C** stores.
- **Varying Predictability:** The model's accuracy is not uniform. It is most accurate for low-volume stores (Type C, MAE ~$109) and least accurate for high-volume stores (Type A, MAE ~$324). This indicates that our highest-volume stores have more complex sales patterns that are harder to predict, likely due to a wider range of influencing factors.

VALIDATE: CRITICAL EVALUATION OF THE MODEL

The model's main assumption is that past patterns will predict the future. Its key blind spots are real-time events and missing data.

- **What could make these predictions wrong?** The model is blind to external events like a competitor's surprise marketing campaign, severe weather, or a hyper-local event not in our holidays dataset.
- **What are the data limitations?** Our data ends in 2017, so the model is unaware of recent shifts in consumer behavior. It also lacks key operational data like real-time inventory levels, staffing, and marketing spend.
- **When would this model fail?** The model would fail in any scenario that breaks from historical precedent. This includes major supply chain disruptions, a store renovation that impacts operations, or an unprecedented event like the COVID-19 pandemic.

EXTEND: ACTIONABLE RECOMMENDATION

Given our model shows that sales are driven by predictable, micro-level patterns, and considering its main risk is blindness to real-time events, our strategy must shift from broad initiatives to targeted, data-driven operational tactics. This leads to the following action plan.

ACTION PLAN

Given our model's insights and limitations, the following specific actions are recommended for store managers:

1. Next Week: Implement Forecast-Driven Operations

- **Action:** Use the model's daily and family-level forecasts to align staff schedules and key inventory with the proven weekend sales peak.
- **Success Metric:** Reduce weekend stock-outs in the top 5 product families by 10%; maintain or reduce overtime labor costs.
- **Risk Mitigation:** Mitigate the risk of unexpected demand by maintaining an on-call staff list for weekends and a defined safety stock for the top 20 best-selling items.

2. Next Month: Develop Differentiated Store-Level Strategies

- **Action:** For high-volume, high-error **Type A** stores, use the forecast as a quantitative baseline but empower managers to make ordering adjustments based on local knowledge. For low-volume, high-accuracy **Type C** stores, leverage the forecast to implement more automated inventory replenishment for non-perishable goods.
- **Success Metric:** Decrease inventory waste in Type C stores by 5%; improve in-stock availability for the top 20% of items in Type A stores.
- **Risk Mitigation:** To refine managerial adjustments, a bi-weekly review process will be established to compare manual overrides against the model's forecast, identifying and correcting for potential biases.

3. Long-Term Planning: Establish an Anomaly Detection System

- **Action:** Use the model's predictions as a baseline for an early warning system. An automated alert will flag any store whose actual sales consistently deviate from the forecast by more than a set threshold (e.g., 25% for three consecutive days). This turns the model's weakness (blindness to the unknown) into a strategic strength.
- **Success Metric:** The number of actionable alerts generated per quarter that lead to the discovery of a significant, previously unknown business event (e.g., a new local competitor).
- **Risk Mitigation:** To prevent "alert fatigue," the deviation threshold will initially be set high to identify major strategic shifts, not minor operational noise.

APPENDIX:

COST OPTIMIZATION ANALYSIS

The cost optimization analysis identifies stores with the highest prediction errors to uncover potential inefficiencies:

- **Store 52** shows the highest total forecast error (~$279k), with an average daily error of ~$565.
- **Store 3** has the highest percentage error (522%), indicating volatility or missing data factors.
- **Top 10 stores** contribute over **$1.8M** in total error across the test period.

Recommendations:

- Focus retraining and additional feature engineering on these top-error stores.
- Use prediction accuracy as a proxy to optimize inventory levels, staffing, and promotional planning.
- Implement ongoing monitoring to reduce forecast-driven inefficiencies and improve model ROI.

ADDITIONAL ANALYSIS

Validate - Additional Evaluation:

- A cost-focused error analysis revealed that the top 10 stores account for over **$1.8M in total forecast error** over 495 days, with Store 52 alone having a cumulative error of ~$279k. This highlights the need for store-specific improvements and operational reviews to reduce forecast-driven inefficiencies.

SOURCES

Store Sales - Time Series Forecasting. (2025). @Kaggle. https://www.kaggle.com/competitions/store-sales-time-series-forecasting/data