Smart Inventory and Promotion Optimization for Walmart using Predictive Sales Analytics

Group 9

Team Members:

Divyasri Yelubolu

Manasvini Edukulla

Rishitha Thatipalli

Srinidh Vudityala

Charan Sai Nalam

Introduction

- This project is about using data and machine learning to help Walmart make smarter inventory and promotion decisions.

- By analyzing three years of weekly sales data from 45 stores and 81 departments, along with economic, holiday, and weather factors, to build a high-accuracy forecasting model.

These forecasts were then used to:

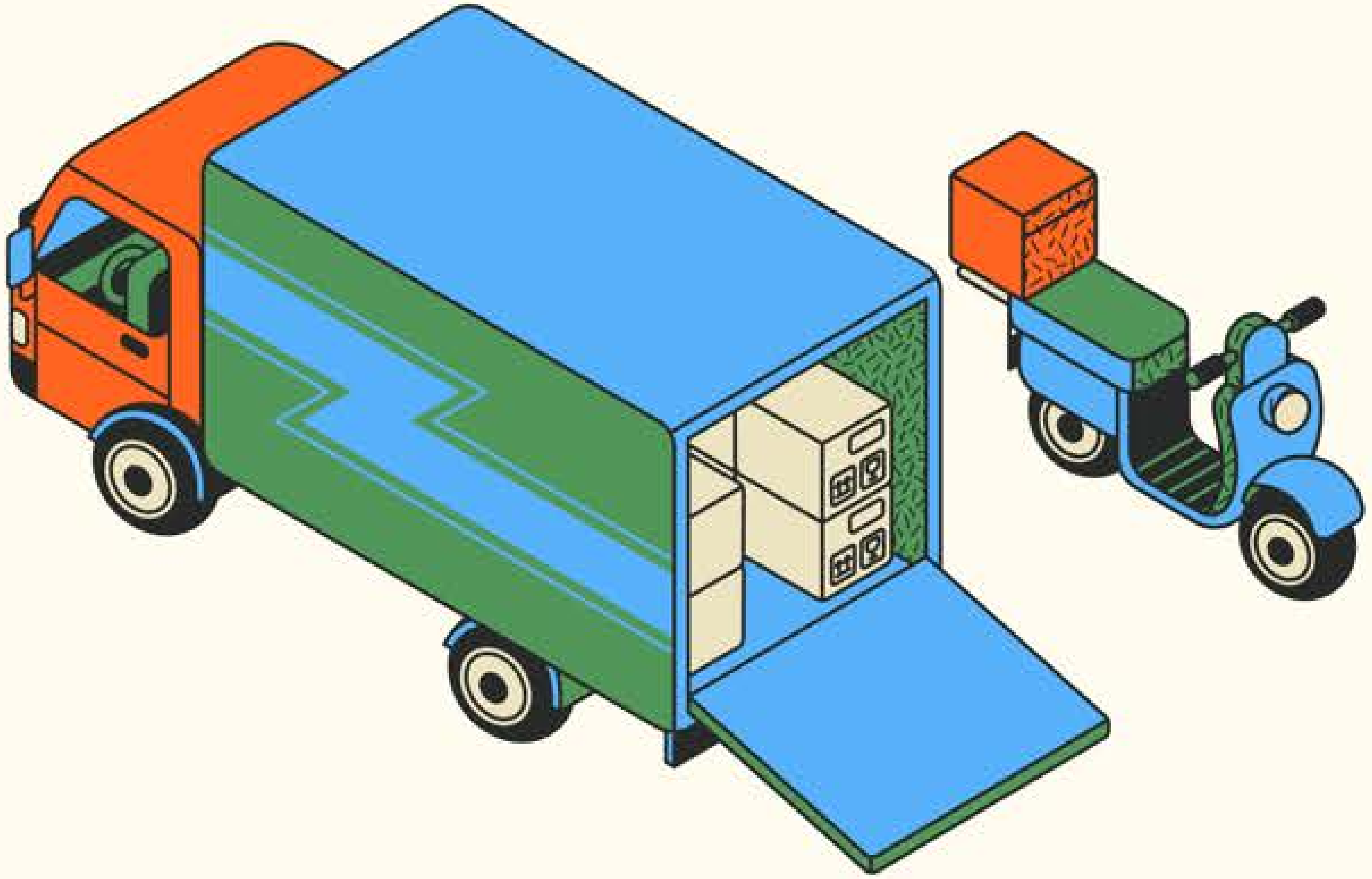
- Set inventory levels with built-in safety buffers
- Identify optimal weeks for promotions
- Reduce overstocking and stock-out risks.

In short, we turned raw sales data into actionable insights that can improve Walmart's efficiency, customer satisfaction, and profitability.

WHAT DID WE SOLVE?

Statistics

Inventory distortion costs retailers over \$1.75 trillion globally every year, driven by overstocking, stock-outs, and missed demand signals.



Problem

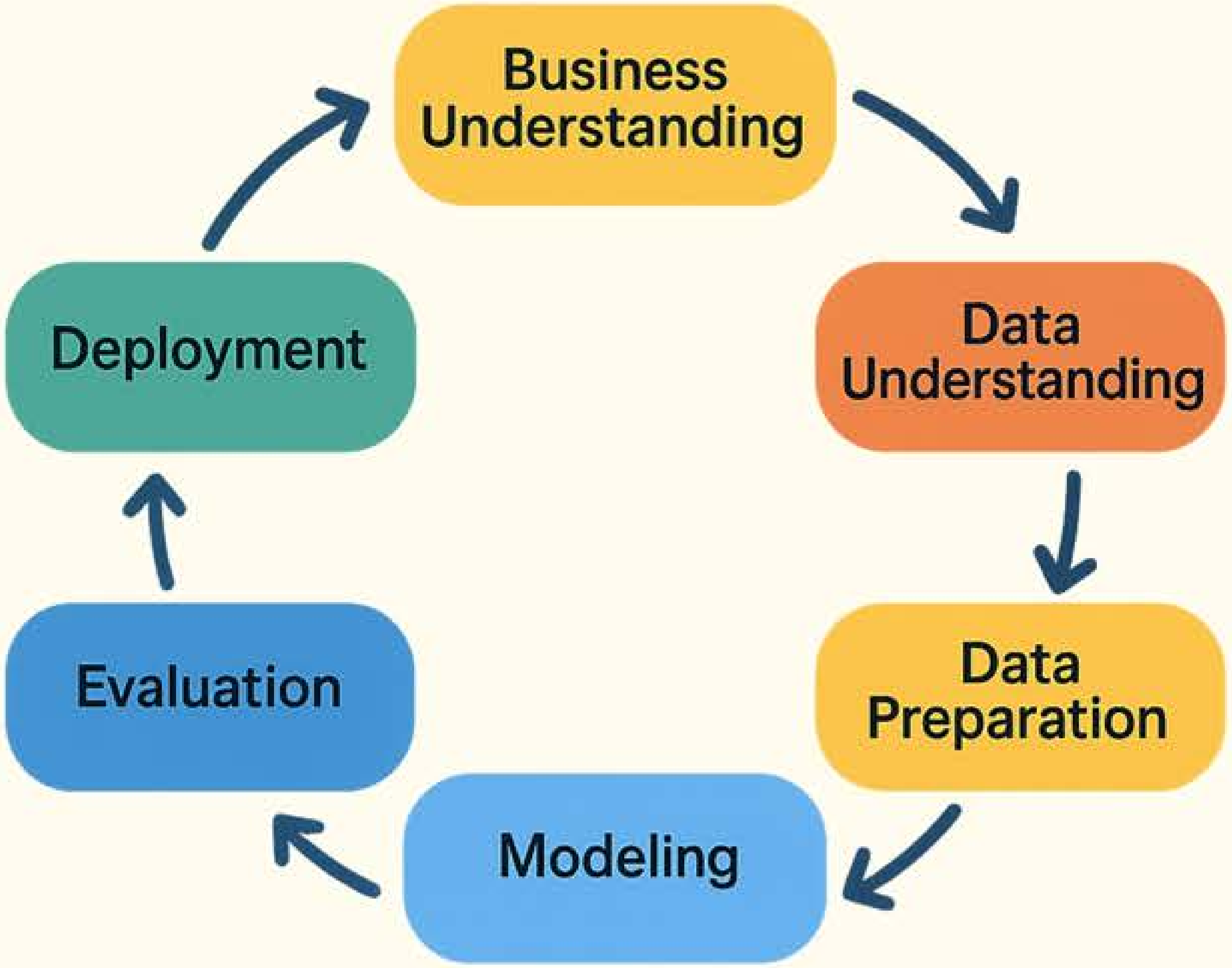
Walmart must forecast sales across 45 stores and 81 departments weekly, but traditional planning fails to capture promotion spikes, local variation, and holiday trends.

Solution

We built a machine learning model (R^2 = 0.9745) that predicts weekly sales and powers smarter inventory decisions and promotion targeting, reducing waste and boosting ROI.

How did we approach the problem?

Followed the CRISP-DM framework:



Iterative approach helped refine forecasts and policies

Data Overview

Four datasets were sourced from Kaggle, covering sales, store attributes, external factors, and holidays to support accurate forecasting.

Contains weekly sales data for each store–department combination, used to train the forecasting model.

Provides store-level details such as type (A/B/C) and physical size in square feet.

Matches the structure of train.csv but without sales figures—used to generate predictions.

Adds external variables like holidays, markdowns, temperature, fuel price, CPI, and unemployment.

EDA Insights: What we discovered before modeling

Seasonal Demand Peaks:

- Significant sales spikes observed in Week 47 (Black Friday) and Week 51 (Christmas) across all three years.

Store Type Differences:

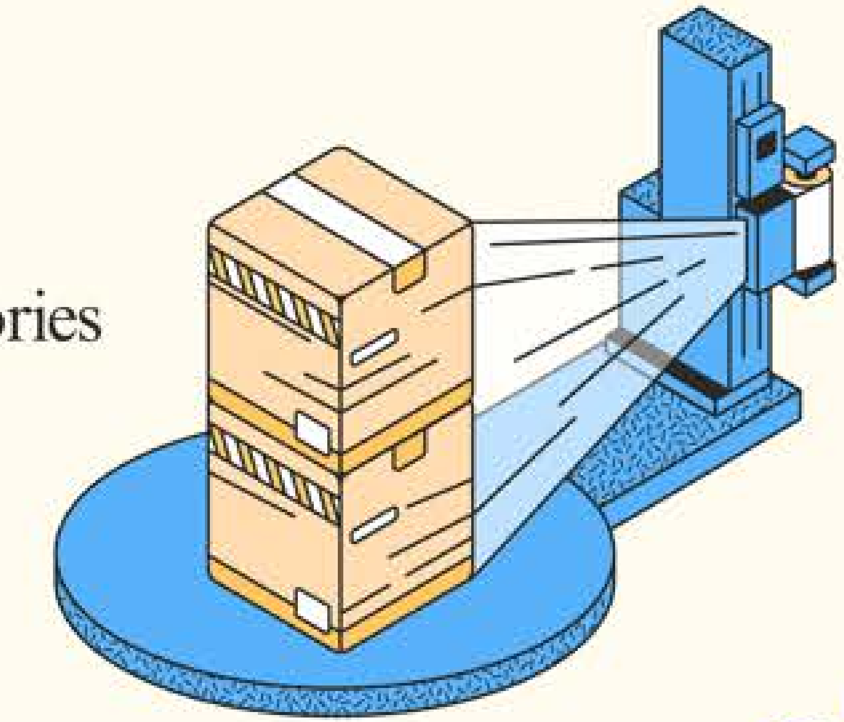
- Type A stores generally reported higher sales volumes, though some Type B and C stores outperformed expectations, indicating local factors at play.

Holiday Uplift:

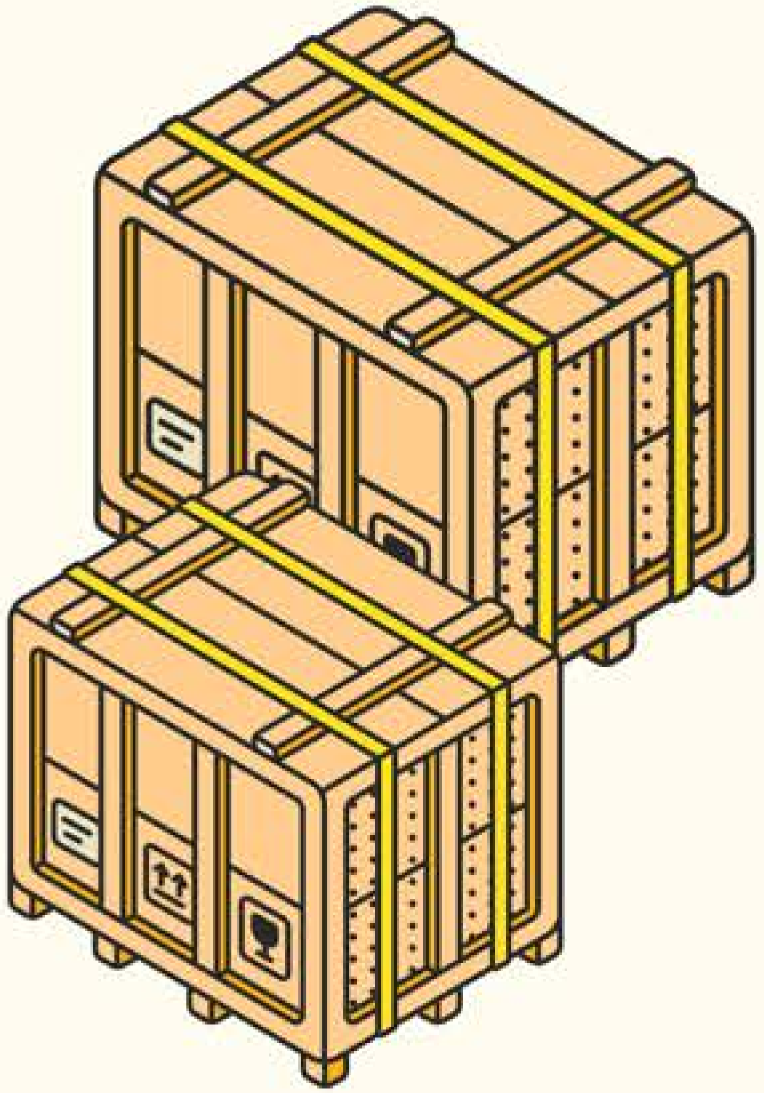
- Sales during holiday weeks were on average 7.1% higher than regular weeks, confirming their importance for planning.

Markdown Effectiveness:

- Promotional markdowns had varying impacts across departments—not all categories responded equally to discounts.

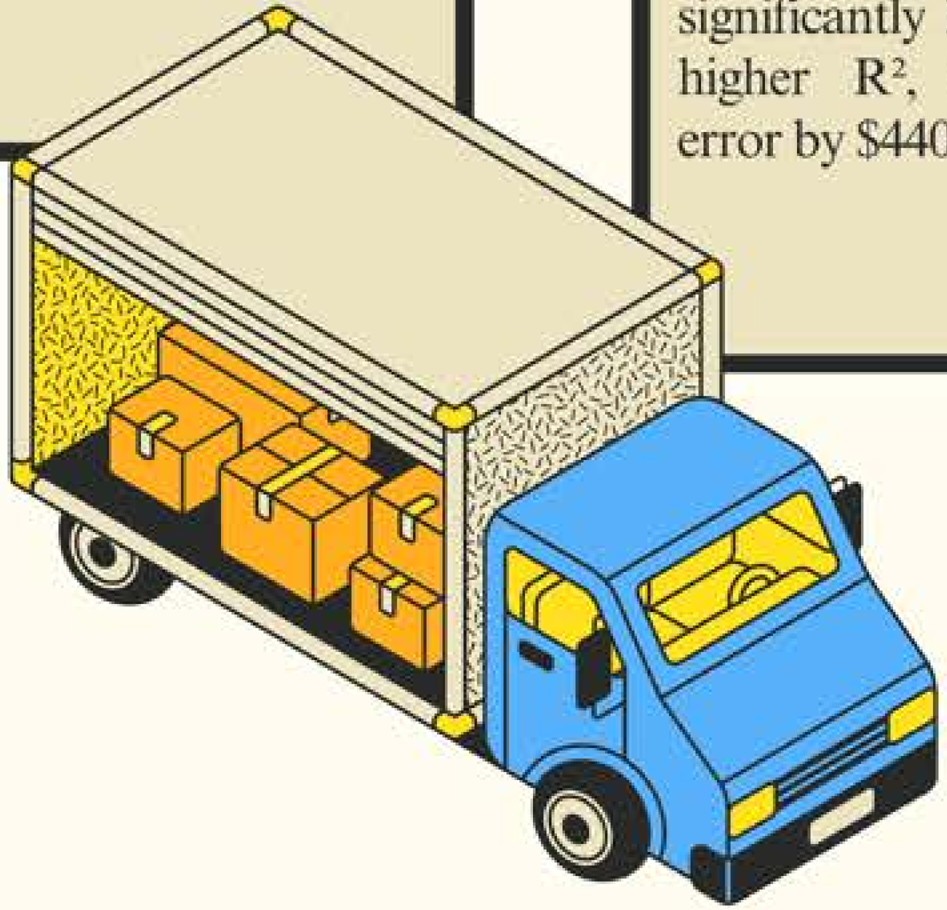


Modeling Approach



Baseline Model – Decision Tree Regressor

Served as a simple, interpretable starting point to validate feature relevance and structure.



Final Model – Random Forest (100 Trees)

Outperformed baseline with significantly lower WMAE and higher R^2, reducing forecast error by $440+.

Holiday Weighting Strategy

Forecasts were evaluated using Weighted Mean Absolute Error (WMAE), where holiday weeks received 5× more weight to prioritize accuracy during high-impact periods like Black Friday and Christmas

Modeling Performance (Decision Tree)

Model Strengths:

- Offers easy interpretability by showing how features like department, store size, and holidays split into decision paths
- Helped validate that store characteristics and calendar variables are meaningful sales drivers

Performance Metrics:

- $R^2 = 0.9575$ Explained 95.75% of weekly sales variance
- RMSE = $4,709.55 Average forecast error per week was relatively high
- WMAE = $2,075.53 Underperformed especially during holidays, due to lack of depth in learning complex interactions

Takeaway:

- While helpful for understanding feature influence, the Decision Tree lacked accuracy and consistency, especially in capturing promotional or seasonal spikes—leading to the adoption of Random Forest for final deployment.

Model Performance

Key Metrics (Random Forest Regressor)

R^2 = 0.9745

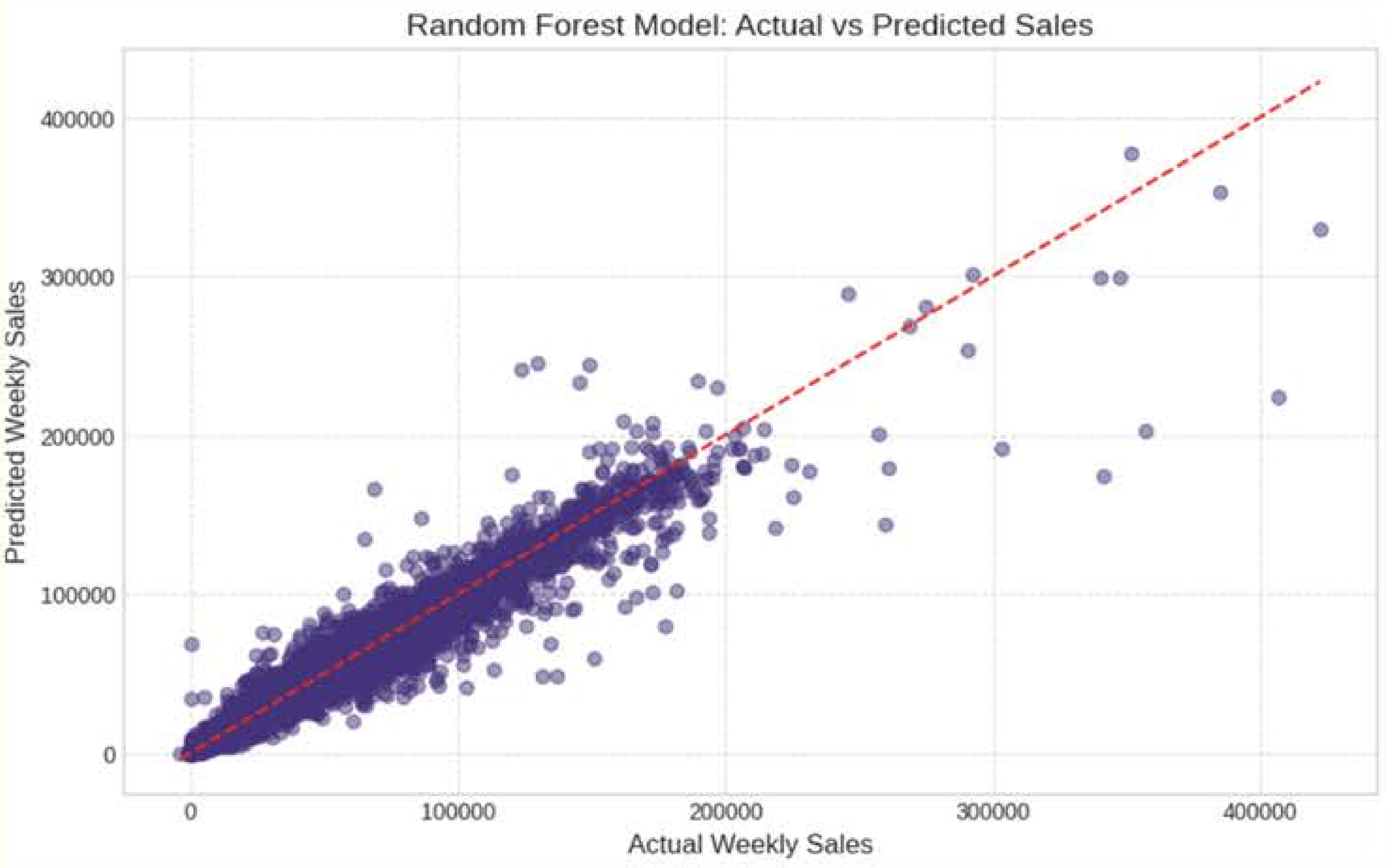
- Captures 97.45% of sales variation — indicates strong model accuracy and reliability.

RMSE = $3,644.76 ($1,064.79)

- Lower average prediction error — helps reduce inventory misalignment and waste.

WMAE = $1,634.84 ($440.69)

- Prioritizes accuracy during holidays — ensures better stock planning in high-demand weeks.



Performance Highlights:

- Improved R^2 by 1.7 percentage points over Decision Tree
- Reduced prediction error across both regular and holiday weeks
- Passed all sanity checks, including holiday uplift, markdown impact, and department volatility

Feature Importance Insights:

- Store type, size, and department ID explained over 88% of model variance
- Calendar features (month, week) captured seasonality trends
- Economic indicators (fuel price, CPI, unemployment) added complementary signal

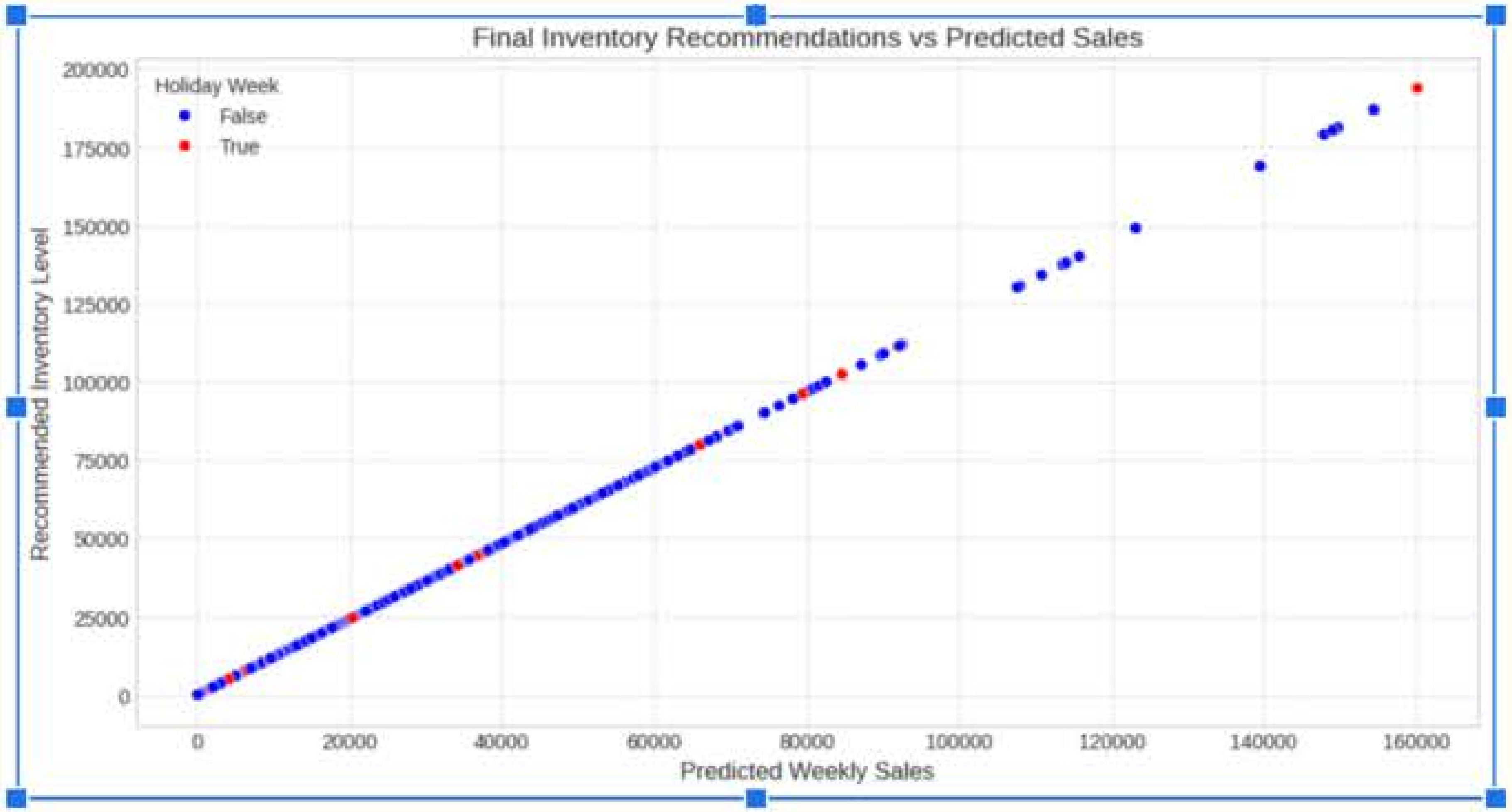
Inventory Optimization – Forecast-Driven Stocking

To convert weekly sales forecasts into stocking targets, a simple three-tier rule was applied:

- **Base Inventory = Predicted Sales × 1.10**
  Adds a 10% operating buffer to cover normal demand noise

- **Safety Stock = Base Inventory × 0.10**
  Adds a second 10% layer to hedge against forecast error

- **Final Inventory = Base + Safety**
  Leads to a total 21% inventory cushion over the forecasted sales



The scatterplot compares predicted weekly sales to final recommended inventory levels.

Promotion Optimization – Targeting High-Demand Weeks

Peak Demand Identification:

- Weeks with sales forecasts above the 95th percentile ($144.7M) were tagged as promotion opportunities.
- Weeks 6, 22, and 51 were consistently flagged, aligning with Super Bowl, Memorial Day, and pre-Christmas periods.

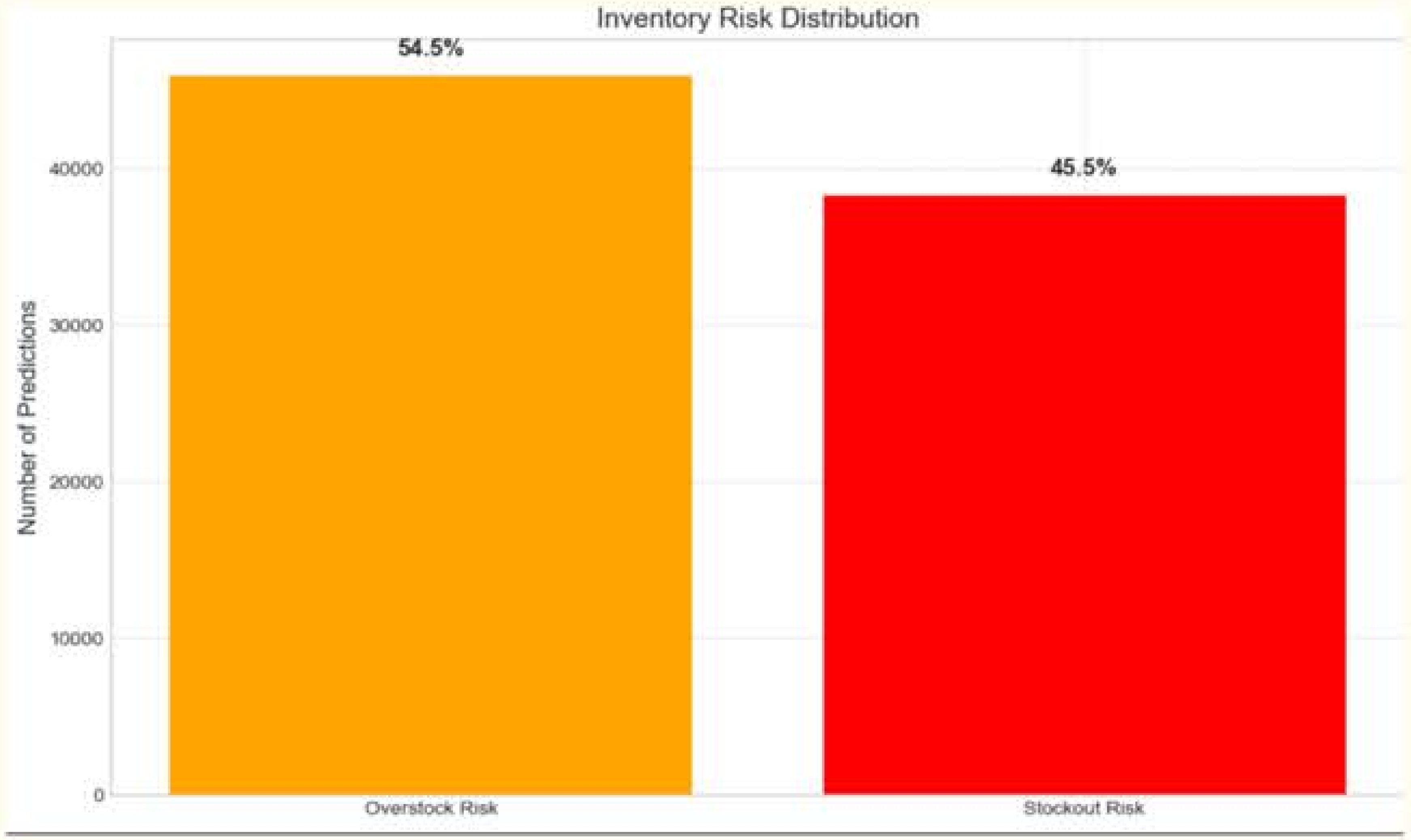
Model Performance (Random Forest Classifier):

- Accuracy, Precision, Recall, Specificity = 1.000
- Classified 100% of promotion and regular weeks correctly with no false positives or negatives.

Business Impact:

- Enables precise timing of marketing campaigns without risk of false alarms.
- Aligns with the 21% inventory buffer to ensure timely replenishment and maximize sell-through and margin during high-sales weeks.

Risk Analysis – Inventory Forecasting Gaps

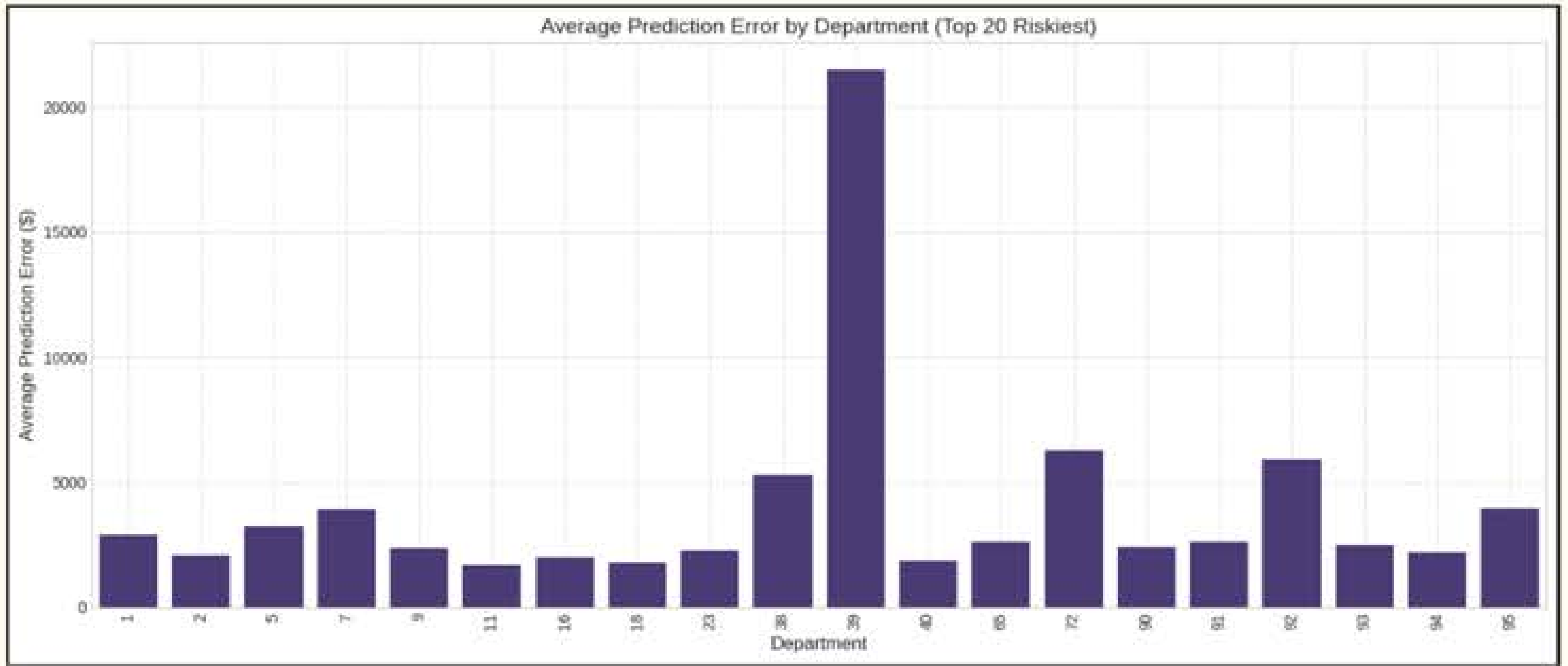


Inventory Risk Distribution (Bar Chart)

- Forecast error was split into two buckets:
  - Overstock (54.5%): Model overestimated demand, average surplus = $1,310
  - Stock-out (45.5%): Model underestimated demand, average deficit = $1,587
- Slight bias toward overstock is intentional to align with Walmart's "never out of stock" policy

Prediction Error by Department (Top 20)

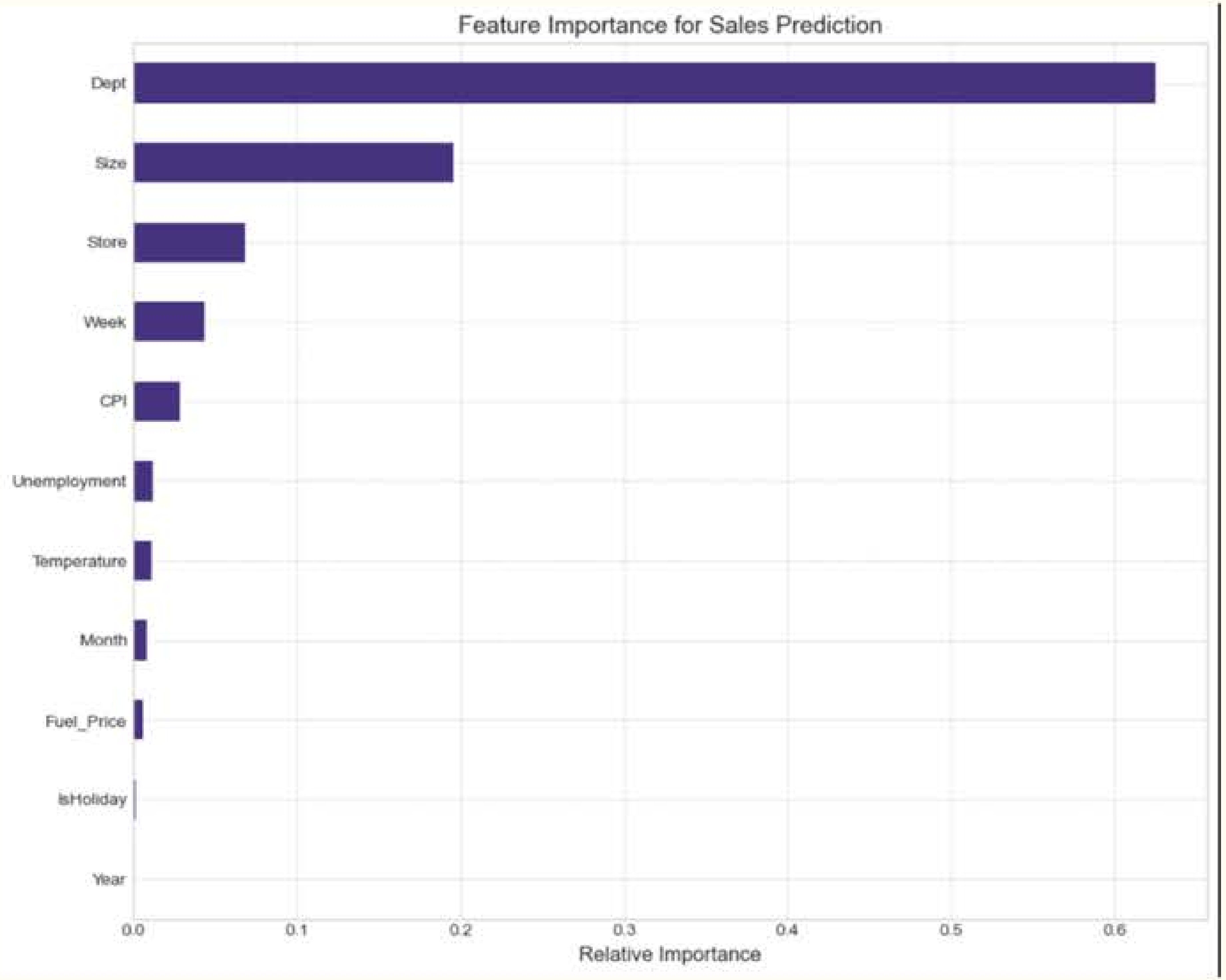
- Departments like 39 , 72 , 92 show highest forecast error
- These high-variance departments justify increased safety buffers (25–30%) based on observed forecast error, compared to 15–20% for more stable ones.



Feature Importance

Feature Importance (Random Forest):

- Department code was the strongest predictor (62.5% relative weight)
- Store size and store ID also had meaningful influence
- Calendar variables (month, holiday, year) and macroeconomic indicators had minimal impact.

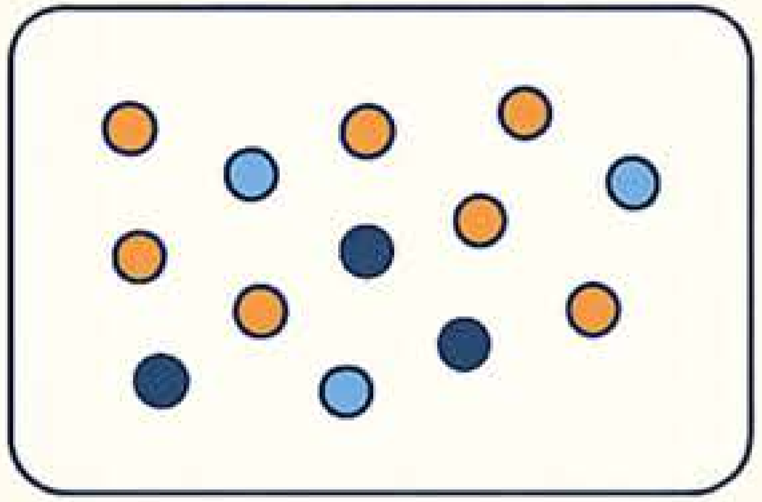


Key Takeaway:

Sales are primarily driven by what is sold (department) and where it's sold (format & size), not by macroeconomic or time-based variables.

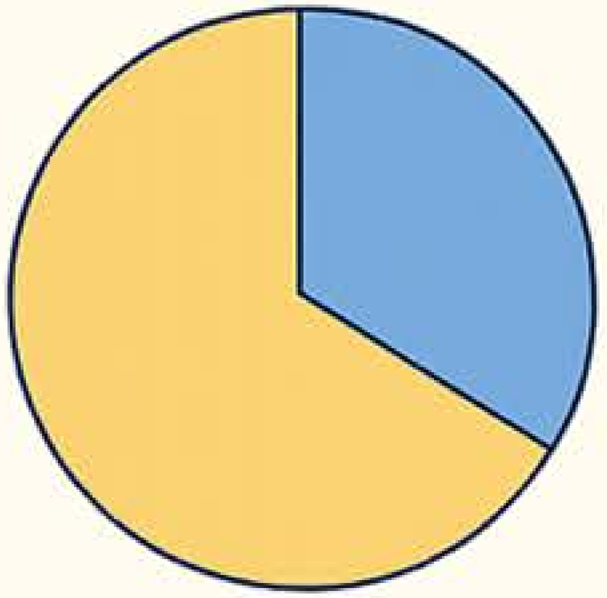
External Drivers

EXTERNA FACTOR INSIGHTS



Most macro and weather variables had weak overall correlation (jpl < 0.20)

Fuel price elasticity = −0.084, suggesting demand is slightly inelastic overall



Higher sensitivity inferred in likely discretionary categories (e.g. out-door, seasonal) based on observed monthly sales patterns

Do External Factors Impact Sales?

What we found:

- Temperature has a small positive effect on weekly sales overall
- Likely stronger in summer-sensitive categories (e.g., garden/outdoor), inferred from seasonal trends
- Fuel prices show a weak negative impact
- This suggests that higher fuel costs may slightly reduce overall consumer spending.
- CPI (inflation) shows a weak positive correlation with sales
- Likely reflects price effects—nominal sales rise with inflation even if purchase volume doesn't
- CPI moves with sales, while unemployment moves in the opposite direction.

Key Insight:

These macro factors do influence sales, but only slightly. The model focuses more on what is sold and where, rather than on external conditions.

Fuel Price Elasticity – Measuring Sales Sensitivity

Key Insight:

- A log-log regression reveals a fuel price elasticity of –0.084
- A 1% rise in fuel prices leads to a 0.08% drop in weekly sales.

Interpretation:

- Core Walmart traffic is largely resilient to fuel price changes
- However, variation across departments suggests that some areas of business may respond more strongly and should be monitored for sensitivity

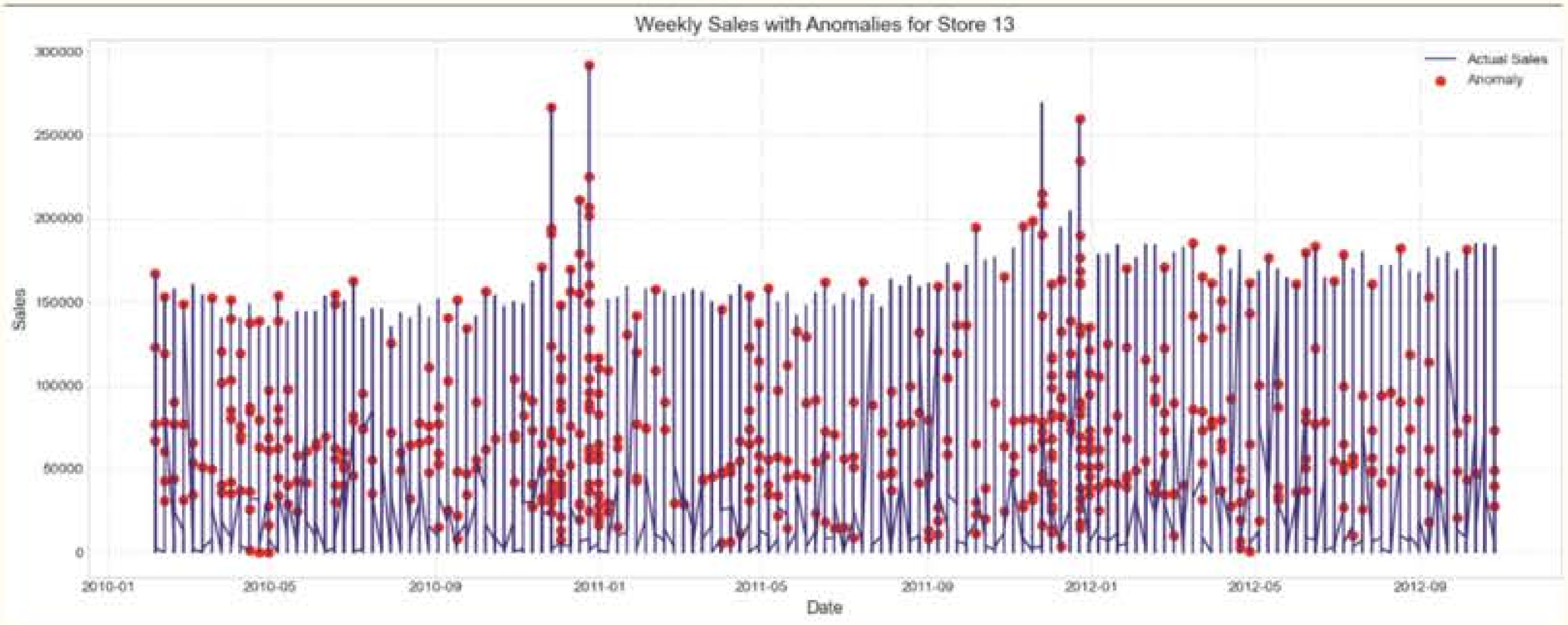
Why This Matters:

- Adds depth to forecasting by highlighting economic sensitivity
- Helps tailor promotion planning, assortment strategy, and risk controls in fuel-sensitive categories

Anomaly Surveillance – Detecting Irregular Sales Patterns

Overview:

- A 2-sigma threshold flagged 12,077 anomalies, or 2.87% of all data points.



Key Observations:

- Holiday spikes (e.g., mid-November, Week 51) were expected and correctly identified
- Unexpected anomalies (e.g., early 2011) hint at local disruptions, clearance events, or data errors.

Why It Matters:

- These outliers help loss-prevention and replenishment teams proactively investigate markdowns, POS errors, or regional events
- Early detection ensures future forecasts remain accurate and reliable

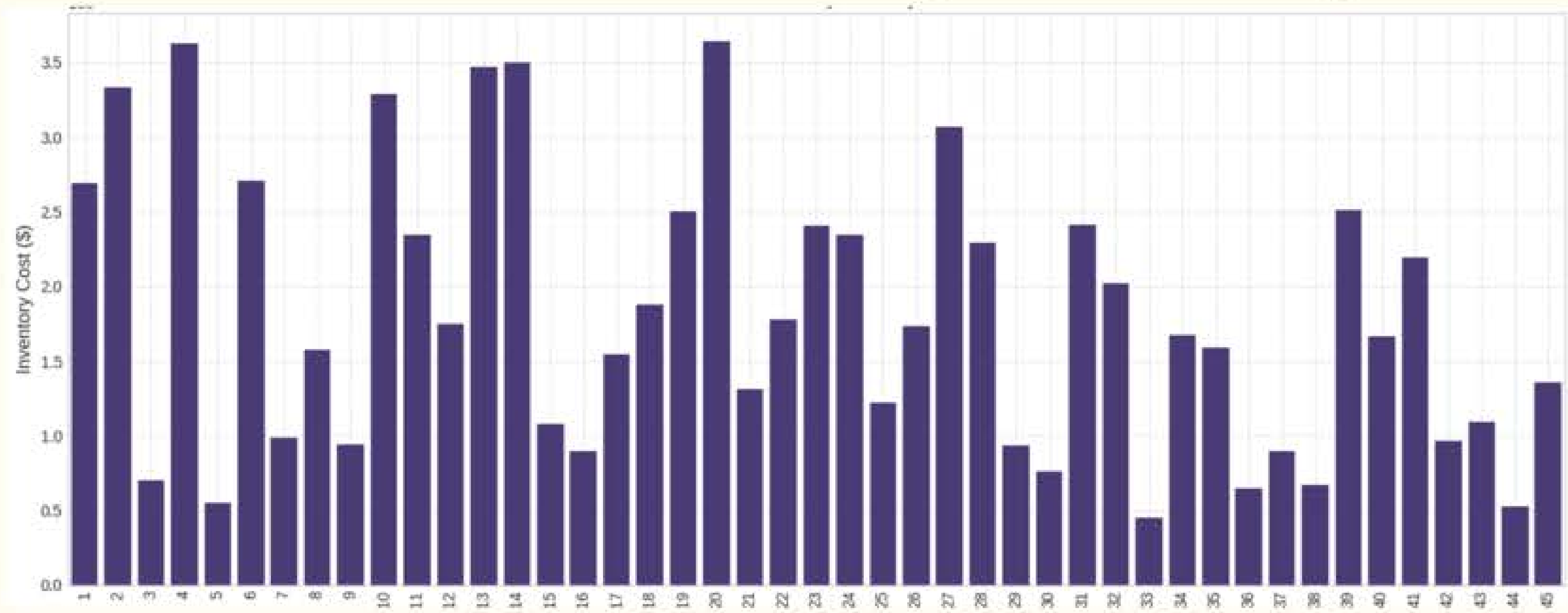
Inventory & Promotion Efficiency Analysis

Key Findings:

- Total Inventory Plan:
- Forecast-driven plan projected 8.15 billion units totaling $81.5B,
- with each store managing ~$1.81B in value

Store-Level Variation:

- Inventory demand varied widely by store, emphasizing the need for localized planning
- (See bar chart: Forecast Inventory Cost by Store)



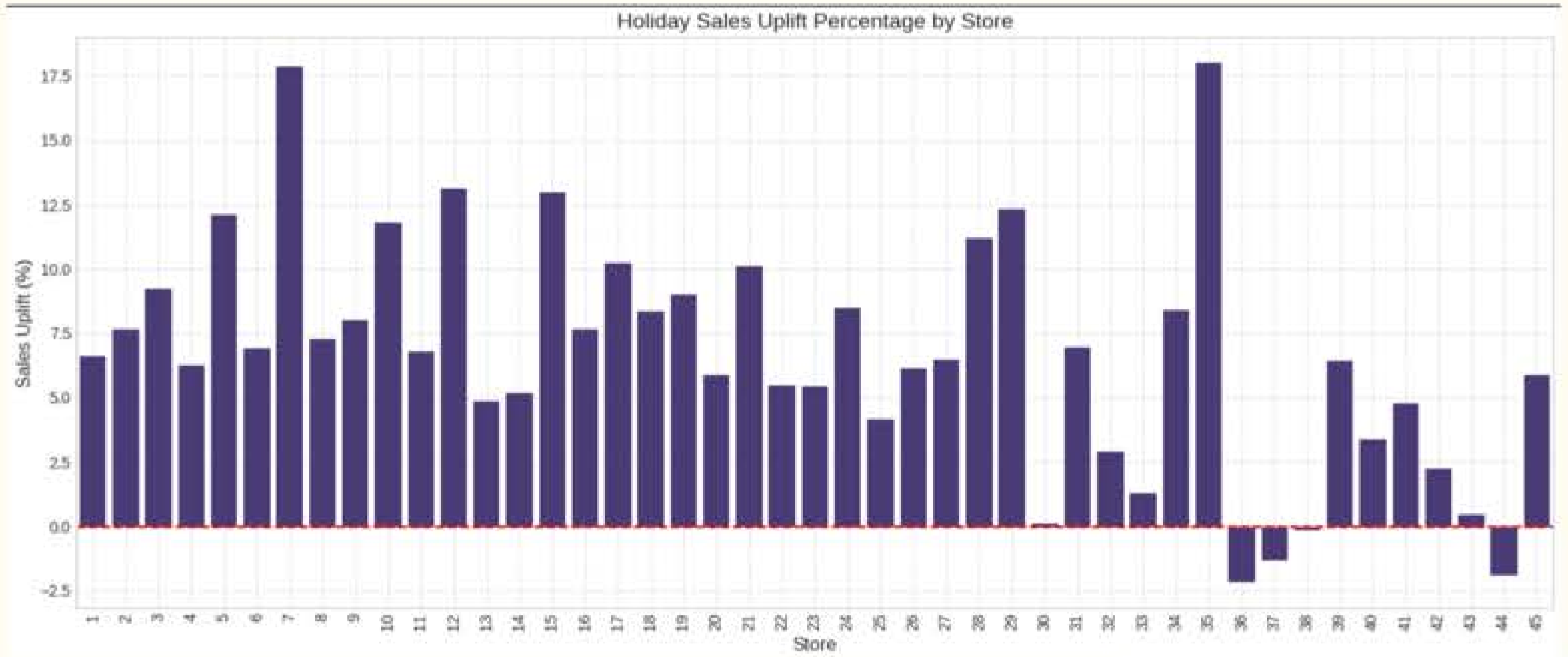
Promotion Impact:

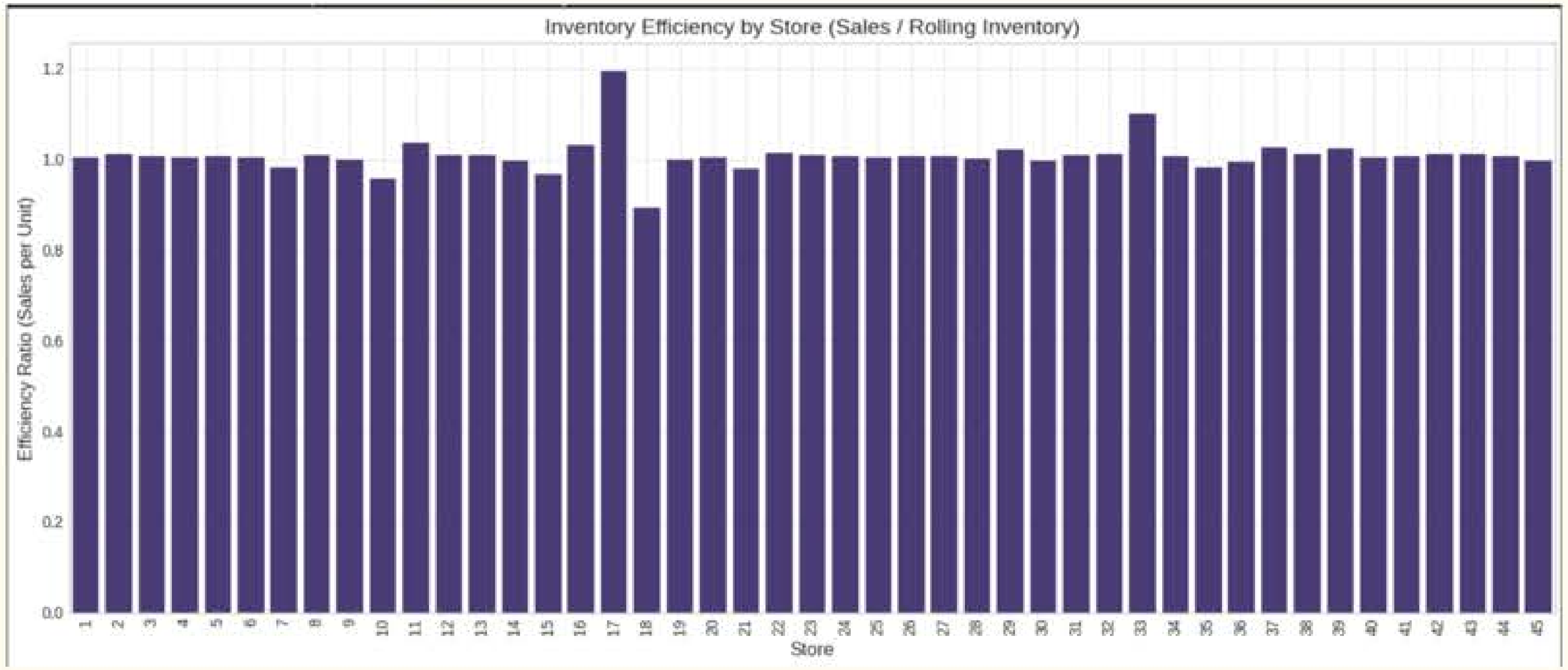
- Store 35 and 7 saw 17–18% uplift in holiday weeks
- Store 36 and 44 saw minimal benefit
- Highlights the need for targeted campaign planning

Efficiency Gaps:

- Store 17: High performer (1.2 sales per unit of inventory)
- Store 18: Underperformer (below 0.9)
- Calls for stock optimization review

Inventory & Promotion Efficiency Analysis





CONCLUSION

Forecast Accuracy

Robust model developed using CRiSP DM and Random Forest

Inventory Optimization

Implemented a 21% buffer in inventory levels

Promotion Targeting

100% accuracy in identifying high-opportunity weeks

Smarter Dectsions

Improves inventory management ana promotions

RECOMMENDATIONS

Tailor Inventory to Store Needs

Adjust stock levels based on store type, size, and local demand. Avoid one-size-fits-all rules.

January Reset Strategy

Clear excess stock post-holidays to free up space and cash for spring inventory.

Buffer by Department Volatility

Apply 25–30% buffers for seasonal items (e.g., décor, school supplies); 15–20% for stable goods.

Holiday Execution Planning

Prioritize staffing and promotions around proven high-demand weeks like Black Friday (Week 47) and Christmas (Week 51).

Respond to Macro Conditions

During fuel price hikes or economic stress, shift focus to essentials and reduce discretionary inventory.

Future Scope

Daily Forecasting

Move from weekly to daily forecasts for more precise inventory and staffing decisions.

Real-Time Integration

Connect sales forecasts directly with POS systems and real-time dashboards to support quicker, more accurate decisions on the ground.

Deep Learning Models

Explore advanced techniques like LSTM to better capture patterns in seasonal or highly fluctuating sales data.

Sustainability Focus:

Incorporate sustainability metrics to track excess inventory, reduce waste, and support eco-friendly retail operations.

THANK YOU

WE ARE NOW OPEN FOR QUESTIONS!