FORECASTING SUPPLEMENT SALES AT WOMART

SAIRAM JAMMU

EXECUTIVE SUMMARY

- Developed predictive models to forecast daily sales across 365 WOMart stores
- Used XGBoost, ARIMA, Prophet, LightGBM
- Built Shiny and Power BI dashboards
- Enabled data-driven inventory and promotion planning

PROJECT OBJECTIVES

- Predict 61 days of future daily supplement sales
- Identify sales trends by store type, region, promotions
- Optimize stock levels and resource allocation
- Deliver actionable insights via interactive dashboards

DATASET OVERVIEW

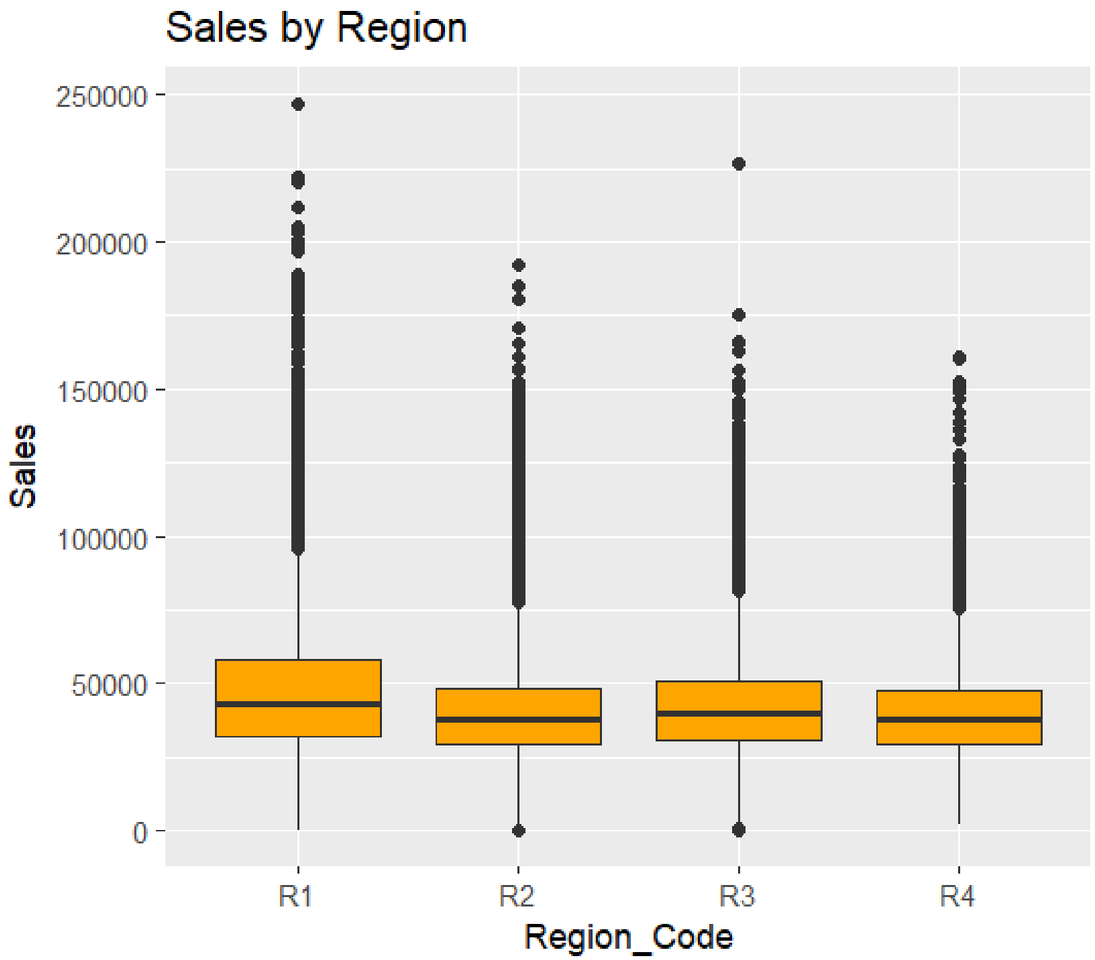
- TRAIN.csv: Daily sales records for 18 months
- TEST_FINAL.csv: Same structure without sales data
- SAMPLE.csv: Submission template
- Key features: Store_ID, Store_Type, Region_Code, Holiday, Discount, Date

DATA CLEANING & FEATURE ENGINEERING

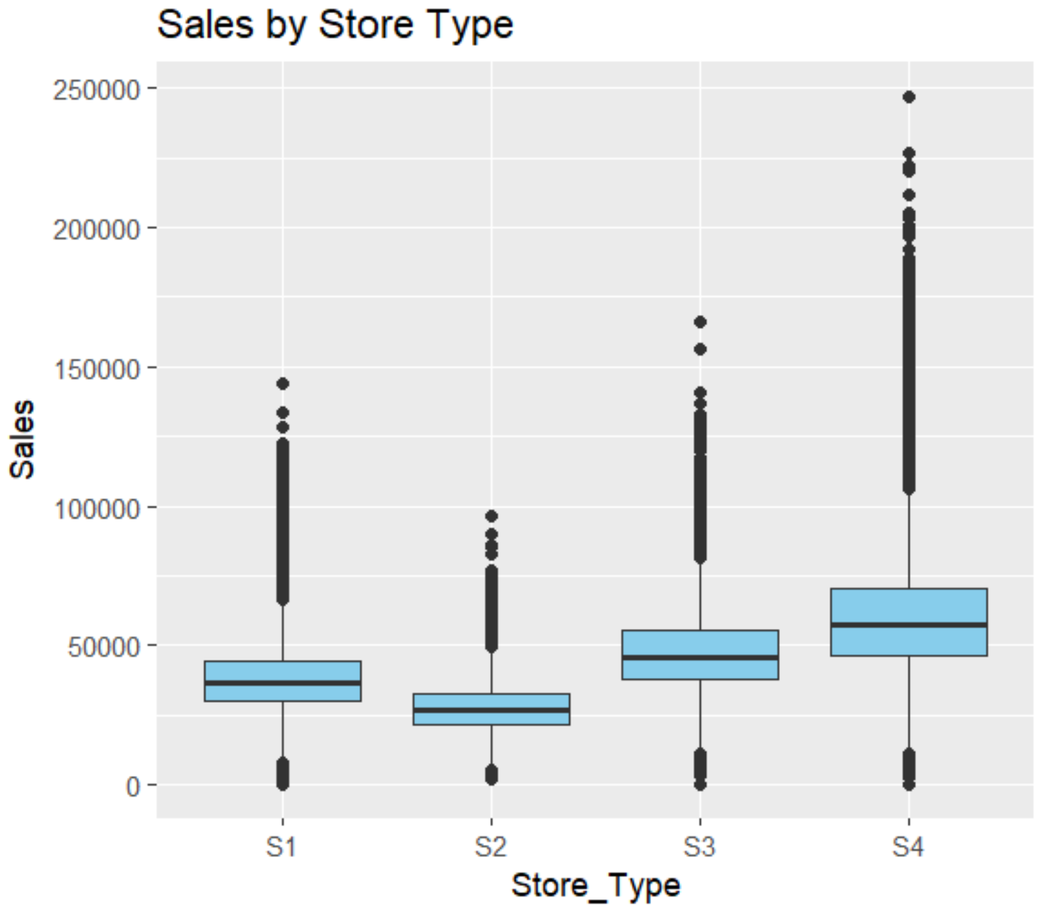
- Converted date columns using lubridate
- Engineered features: Day, Week, Month, DayOfWeek
- Created interaction variables: Holiday × Discount
- Encoded categorical variables

EXPLORATORY DATA ANALYSIS

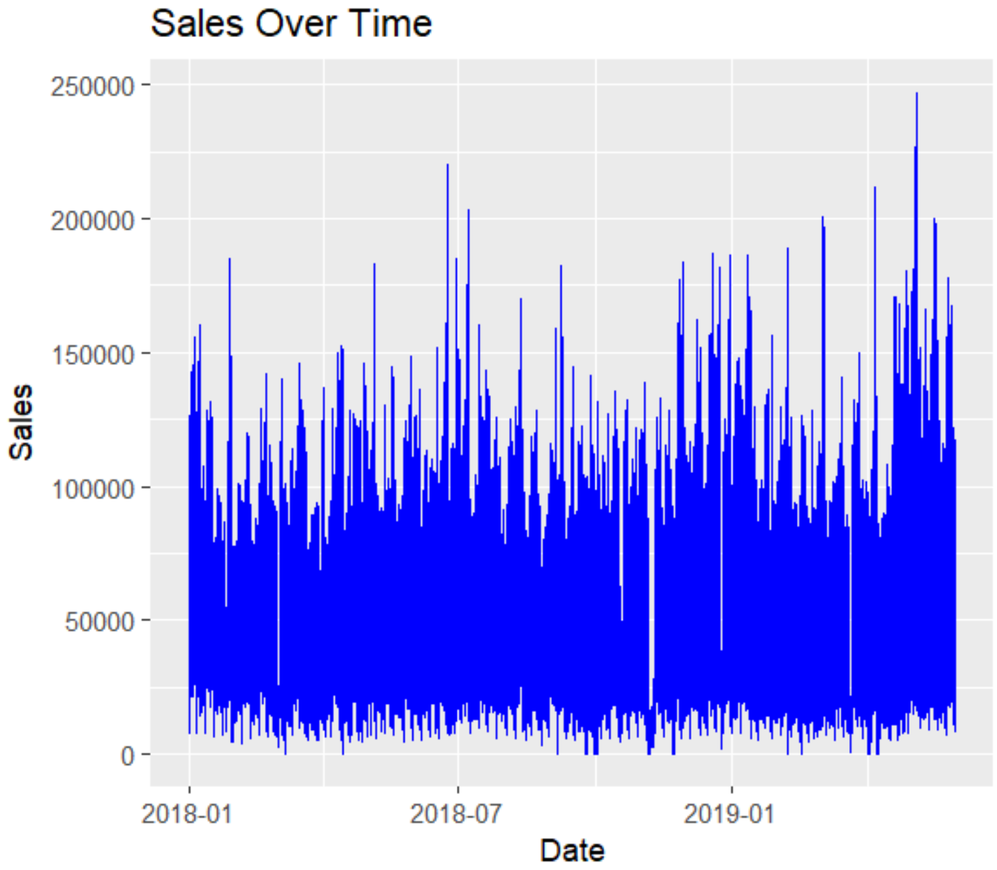
- S1 and S4 store types had higher sales
- Discounts and holidays strongly influenced revenue
- R1 and R3 regions showed higher sales variance.



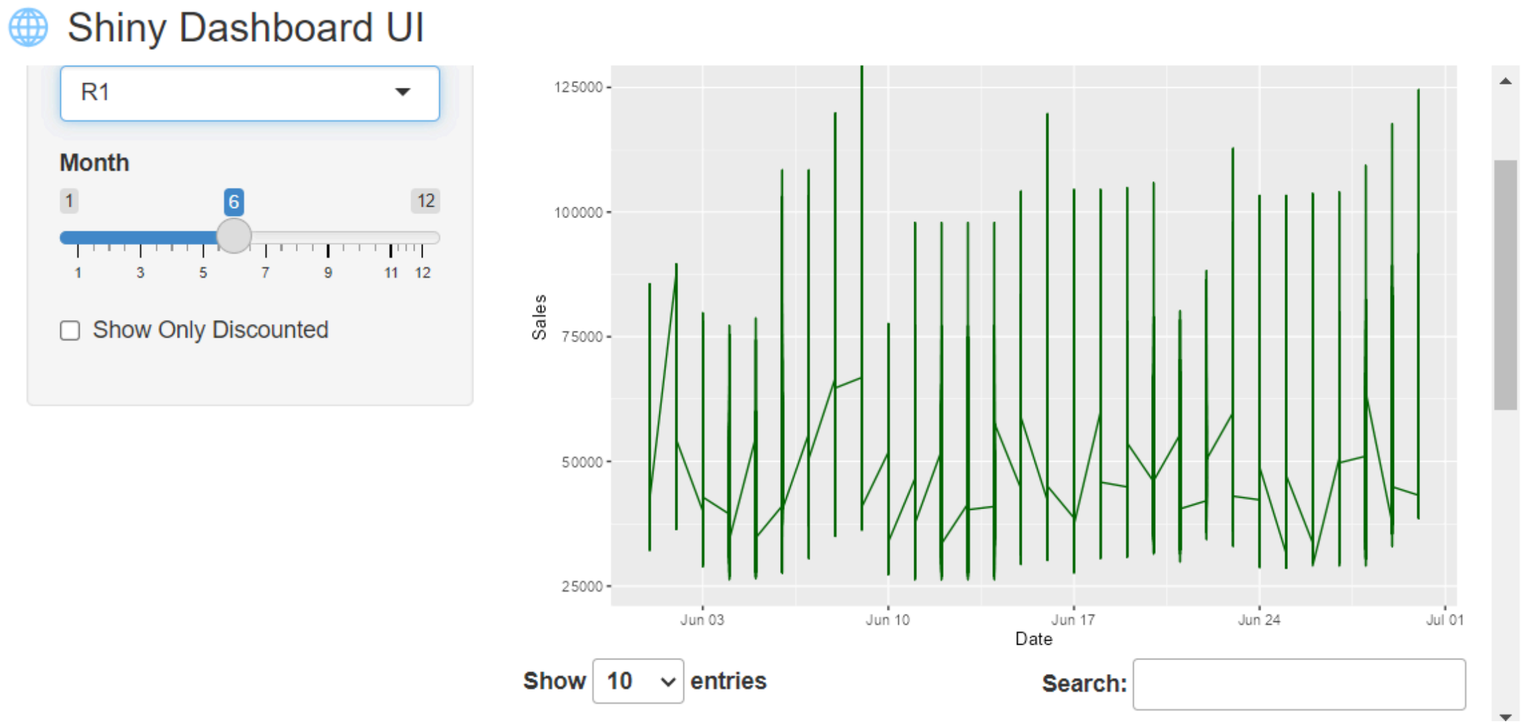
- REGION R1 DEMONSTRATES THE HIGHEST MEDIAN AND THE WIDEST SALES RANGE
- R2 AND R4 SHOW LOWER AND MORE STABLE SALES
- FREQUENT OUTLIERS ACROSS ALL REGIONS



- **STORE TYPE S4 HAS THE HIGHEST MEDIAN AND WIDEST SPREAD IN SALE**
- **STORE TYPE S2 CONSISTENTLY UNDERPERFORMS**
- **S1 AND S3 EXHIBIT MODERATE SALES DISTRIBUTIONS**



- SALES SHOW STRONG SEASONALITY AND RECURRING SPIKES
- AN UPWARD TREND IS NOTICEABLE POST MID-2018, INDICATING GRADUAL GROWTH IN CUSTOMER DEMAND OR IMPROVED MARKETING EFFECTIVENESS.
- SHARP DIPS ARE OBSERVED OCCASIONALLY, POSSIBLY DUE TO STOCKOUTS, SYSTEM DOWNTIME, OR EXTERNAL FACTORS LIKE HOLIDAYS WITHOUT PROMOTIONS—THESE NEED FURTHER INVESTIGATION.



- SPIKES AROUND MID-JUNE
- THE INTERACTIVE SHINY DASHBOARD ENABLES REAL-TIME FILTERING, ALLOWING USERS TO VISUALIZE SALES PERFORMANCE BY REGION, MONTH, AND DISCOUNT STATUS
- DAILY SALES IN REGION R1 FOR JUNE SHOW A CONSISTENT PATTERN WITH PEAKS AND TROUGHS, SUGGESTING WEEKLY CYCLES OR PROMOTIONAL TIMING.

MODELING APPROACH

- Used XGBoost for fast, accurate predictions
- Created DMatrix for training/testing
- Tuned hyperparameters using MSLE
- Handled overfitting using regularization

MODEL EVALUATION

- Models evaluated using MSLE
- XGBoost outperformed other models
- Forecasts aligned with historical demand patterns

CHALLENGES AND SOLUTIONS

- Sparse sales data → zero-filling, aggregation
- Region-based variations → interaction features
- Overfitting → L1/L2 regularization

RECOMMENDATIONS

- FOCUS ON PROMOTIONS DURING HOLIDAYS IN LOW-SALES REGIONS
- EXPAND SUCCESSFUL STORE FORMATS (S1/S4)
- USE DEMAND FORECASTS FOR WORKFORCE AND STOCK PLANNING

CONCLUSION

- ROBUST FORECASTING IMPROVES OPERATIONAL EFFICIENCY"
- DATA-DRIVEN STRATEGY SUPPORTS GROWTH AND INNOVATION"
- SCALABLE MODELS CAN BE DEPLOYED CHAIN-WIDE

THANK YOU