RETAIL STORE SALES FORECASTING USING MULTIPLE LINEAR REGRESSION IN R FOR A U.S. APPAREL CHAIN

1. Background

A regional apparel retailer with over 80 physical stores across the U.S. was preparing for an expansion and wanted to optimize its resource planning. The corporate team lacked clarity on which operational levers—store size, employee count, customer traffic, or marketing spend—were actually driving weekly sales.

They hired us to use R for building a multiple linear regression model that could help forecast weekly sales and serve as a reference for improving ROI on promotional budgets and in-store operations.

2. Objective

- To model weekly store sales using operational and marketing data
- To identify and quantify the influence of key drivers on sales performance
- To generate actionable insights for promotional budgeting and staff deployment

3. Data Used

Source: Internal ERP and marketing systems (Q1 to Q3, 2023)

Dataset Details:

- 3,240 observations (weekly data × 80 stores × 9 months)
- Fields:
    - Store_ID, Week, Store_Size_SqFt, Staff_Count, Foot_Traffic, Promo_Spend, Weekly_Sales
    - Region (categorical, 5 zones), Holiday_Week_Flag (dummy)

4. Methodology

4.1 Data Preparation

- Cleaned and merged data using dplyr

- Scaled continuous variables with scale()
- Created dummy variables for Region and Holiday_Week_Flag

4.2 Regression Modeling

- Built a multiple linear regression model using lm() in R
- Checked for interaction effects: Foot_Traffic × Promo_Spend
- Assessed multicollinearity via VIF (car package)
- Conducted diagnostic tests for residual normality, homoscedasticity, and leverage points

5. Model Output and Diagnostics

Predictor	**Coefficient (β)**	**p-value**	**Interpretation**
Staff_Count	+146.7	< 0.001	More staff positively impacts sales
Foot_Traffic	+5.42	< 0.001	Each additional visitor adds \$5.42 on average
Promo_Spend	+0.89	< 0.001	Promotional ROI: ~\$0.89 added per \$1 spent
Store_Size_SqFt	+0.017	0.033	Larger stores generate higher baseline sales
Holiday_Week_Flag	+1125.5	< 0.001	Sales spike during holidays
Region_Northeast (ref: SE)	+743.1	0.006	Stores in NE outperform SE baseline
Traffic × Promo_Spend	+0.0032	0.010	Combined effect: promotions more effective with high foot traffic

- **Adjusted R^2** = 0.72
- Residual diagnostics: Q-Q normality confirmed, no significant heteroskedasticity
- Cook's Distance: < 0.5 for all, indicating no extreme influence points

6. Interpretation and Recommendations

- **Staffing levels** are critical, particularly in high-traffic stores
- **Foot traffic** is a strong predictor but only effective when paired with targeted **promo spend**
- **Holiday weeks** significantly boost sales and should trigger **temporary staffing and inventory increases**
- **Regional disparities** suggest adjusting expectations and budgets by geography
- Recommended creating a **dynamic budgeting tool** based on regression coefficients

7. Reporting Output

- **R Markdown Report (PDF, 18 pages)**:
  - Model results
  - Coefficient plots and diagnostics
  - Actionable summaries and implementation checklist
- **Excel Dashboard**:
  - Predictive sales tool per store-week
  - Inputs for traffic, spend, and staffing
  - ROI calculator for marketing team
- **Slides (Optional)**:
  - Visual report for business unit heads and retail managers

8. Business Impact

- Used model outputs to reallocate **20% of promo budget** toward high-traffic stores
- Resulted in a **9.8% sales lift** during Q4 promotional periods
- Staff deployment guidelines revised using modeled headcount thresholds
- Regression tool adopted as part of **quarterly planning dashboard** for operations team