

# PREDICTIVE MODELING FOR RETAIL SALES FORECASTING USING INFERENTIAL AND EXPLORATORY STATISTICS

## 1. Client Context

A regional chain of retail stores in the Midwest U.S. approached us with the challenge of improving their monthly sales forecasts. Their current manual estimation process lacked statistical grounding, leading to inaccurate inventory planning and cash flow mismatches. They needed a statistical model that could predict sales and identify key seasonal, promotional, and location-based drivers.

## 2. Problem Statement

The client wanted to:

- Forecast total sales per store for the next three months
- Understand how variables like store size, customer footfall, product mix, promotions, and seasonality influenced monthly revenue
- Integrate these insights into their existing Excel-based reporting structure

## 3. Tools and Techniques Used

- **Software:** Excel (for final integration and automation), SPSS (for statistical modeling)
- **Statistical Techniques:**
  - Multiple Linear Regression with interaction terms (store size  $\times$  promotion intensity)
  - One-way ANOVA to test differences in average sales between store locations
  - Time series decomposition (trend, seasonality) using moving averages
  - Correlation analysis and scatter plots for exploratory insight
  - Residual analysis and variance inflation factor (VIF) to test assumptions
- **Reporting:** Executive-friendly PDF report plus automated Excel dashboard with embedded formulas

## 4. Data Overview

- **Dataset Size:** 36 stores  $\times$  12 months = 432 records
- **Variables Included:**
  - Monthly\_Sales (continuous, target variable)
  - Store\_Size\_sqft (continuous)
  - Footfall (continuous)
  - Product\_Diversity\_Index (0–1 scaled)
  - Promo\_Intensity (index from 0 to 10)
  - Month (categorical – encoded for seasonality)
  - Region (categorical – Midwest North, Midwest South)

Data was cleaned to remove outliers (e.g., one store under renovation) and imputed where missing (e.g., footfall data missing in 2 months was replaced using regional mean).

## 5. Key Findings

- **Regression Model Performance:** The final model had an  $R^2$  of **0.76**, indicating that 76% of the variance in sales could be explained by the selected variables.
- **Significant Predictors:**
  - **Footfall** was the strongest predictor ( $p < 0.001$ ), with every additional 100 visitors contributing an average of \$1,020 in revenue
  - **Promotion intensity** significantly interacted with store size; smaller stores saw higher proportional gains from promotions
  - **Product diversity** positively influenced sales in higher-income ZIP codes
  - **Seasonality:** November and December contributed 28% of the annual revenue, confirming holiday-driven sales spikes
- **Visualization Outputs:**
  - Regression plots, bar charts by month, and correlation heatmaps were built into Excel for weekly executive briefings

## 6. Deliverables

- 18-page analysis report covering:
  - Model design, assumptions, and testing
  - Interpretation of coefficients and interaction effects
  - Recommendations for variable optimization (e.g., increase product variety in Q4)
  - A model evaluation checklist for internal use
- An **Excel dashboard** with:
  - Dynamic inputs for variable adjustment
  - Sales forecast output in dollar value and percentage change
  - Conditional formatting to flag stores at inventory risk

## 7. Client Outcome

The forecasting model led to a **17% improvement** in inventory allocation accuracy and reduced excess inventory costs by **\$92,000 over 3 months**. The client began using the dashboard in weekly planning sessions, and requested a Phase 2 extension to include product-level forecasting. Their internal team also began using SPSS for independent quarterly reports based on the training provided during the project handover.