

RETAIL SALES FORECASTING USING MULTIPLE REGRESSION WITH INTERACTION TERMS IN EXCEL

1. Background and Problem Statement:

A national retail brand with seasonal fluctuations in demand wanted to forecast **monthly sales** based on advertising spend and seasonal factors. The marketing team suspected that the **impact of advertising varied by month** and needed a method to model this interaction. The goal was to build a regression model in Excel that includes both continuous predictors and **interaction terms** with month dummies to capture seasonality-adjusted advertising effects.

2. Objectives:

- Forecast monthly sales using advertising spend, month (as dummy), and interaction terms
- Build the entire model in Excel using native formulas and Data Analysis Toolpak
- Visualize seasonal effects and differences in advertising impact by month
- Create a dynamic forecasting sheet to support monthly campaign planning

3. Methodology:

3.1 Dataset Overview

- **Observations:** 12 months of sales data from Jan–Dec 2023
- **Variables Collected:**
 - Sales (in INR Lakhs)
 - Ad_Spend (in INR Lakhs)
 - Month (converted into 11 dummy variables; January as baseline)

3.2 Variable Engineering

- Created 11 month dummies: Feb, Mar, ..., Dec
- Created **11 interaction terms**: $Ad_Feb = Ad_Spend \times Feb$, $Ad_Mar = Ad_Spend \times Mar$, ..., $Ad_Dec = Ad_Spend \times Dec$
- Each interaction term captures the **effect of ad spend in a specific month** beyond the baseline (January)

3.3 Model Estimation

- Used **Excel Data Analysis Toolpak** → **Regression**
- Independent variables: Ad_Spend, 11 month dummies, 11 interaction terms
- Dependent variable: Sales
- Conducted multicollinearity check using manual R^2 regressions

4. Results and Interpretation:

4.1 Regression Summary (Highlights)

Variable	Coefficient	p-Value	Interpretation
Ad_Spend	1.95	0.011	In January, each ₹1L ad spend generates ₹1.95L sales
Feb (Dummy)	4.2	0.030	Feb baseline sales higher than Jan by ₹4.2L
Ad_Feb (Interaction)	-0.65	0.045	Ad effectiveness lower in Feb: net effect = 1.30 (=1.95-0.65)
Ad_Dec (Interaction)	+0.90	0.022	December ads much more effective than January: net = 2.85

- **Adjusted $R^2 = 0.92$**
- Interaction terms showed **advertising worked best in December and poorly in April-May**
- Model passed all residual and linearity checks; no major outliers

5. Deliverables in Excel:

- Full dataset with dummy and interaction columns auto-calculated
- Regression summary table and coefficients interpreted in a separate sheet
- **Forecasting Tool:**
 - Input: Month + planned Ad Spend
 - Output: Predicted Sales (using correct base + interaction term)
- Visualizations:
 - Sales vs. Ad Spend by Month

- Line chart showing predicted vs. actual sales
- Bar chart of ad effectiveness per month (slope values)

6. Recommendations:

- Increase ad budget for **Q4 months**, especially **December**, to maximize ROI
- Reduce ad spend in months where interaction coefficients are significantly negative
- Update the model quarterly with new sales and advertising data
- Use simplified monthly models for tactical short-term campaigns

7. Stakeholder Relevance:

Academic:

- Ideal example to teach **interaction effects in multiple regression using Excel**
- Shows manual construction of dummy and interaction variables in spreadsheets

Corporate:

- Helps retail and eCommerce teams identify **month-wise ROI on advertising**
- Gives marketers a clear, Excel-native tool for campaign planning and budgeting