PROFITABILITY MODELING FOR A MULTI-OUTLET RETAIL CHAIN USING PYTHON REGRESSION

1. Introduction

A regional retail chain operating 50+ outlets across urban and semi-urban areas faced difficulty in explaining profit variation across stores. Leadership sought a data-driven solution to identify what combination of store characteristics and product categories drive profitability. Our task was to build a Python-based multiple linear regression model that would uncover key drivers and guide strategic decisions on store layout, product focus, and staffing.

2. Objective

- To identify how store size, location, product mix, and staffing contribute to monthly profit
- To build an interpretable regression model using Python that includes dummy variables and interaction effects
- To deliver a reusable Python script and visual report for use by the business intelligence team

3. Data Provided by Client

The client shared three months of performance data across 58 stores. Each row represented one store per month and contained:

- Monthly_Profit (in INR)
- Store_Size (in sq. ft.)
- Total_Staff_Count
- %Electronics_Sales
- %Groceries_Sales
- %Clothing_Sales
- Location_Type (Urban / Semi-Urban / Rural)

Dummy variables were created for Location_Type, and an interaction term was created between Store_Size and %Electronics_Sales to test format-product synergy.

4. Methodology

4.1 Data Preparation

- Cleaned the data, handled minor missing values
- Verified that percentage columns summed to ~100% to ensure product mix consistency
- Encoded Location_Type using one-hot encoding
- Created interaction variables such as Store_Size × %Electronics_Sales

4.2 Model Building

- Built the model using Statsmodels.OLS
- Conducted stepwise variable selection to reduce overfitting
- Validated assumptions via residual plots, Q-Q plots, and VIF diagnostics

4.3 Deliverables

- Python script with code comments and reusability for future datasets
- Visual report summarizing drivers of profitability
- Excel dashboard to test hypothetical scenarios (e.g., what if we increase staff count by 2?)

5. Key Results

- Model R-squared: **0.71**, Adjusted R-squared: **0.69**
- **Store_Size** and **%Electronics_Sales** had the highest positive effect on profit
- **Total_Staff_Count** was positive but only up to a threshold (diminishing returns observed)
- **Urban stores** outperformed rural and semi-urban stores even after controlling for store size
- The interaction between store size and electronics sales was significant, indicating synergy in larger-format stores selling electronics

6. Business Impact

- Helped regional heads redesign staffing and product mix based on model predictions
- Used model output to justify downsizing 3 underperforming rural stores and upgrading 4 urban ones

- Provided evidence-based input into real estate planning and merchandising for new locations
- Estimated annual profit improvement potential of **₹12–15 million** by applying model-based recommendations

7. Future Scope Suggested

- Integrate monthly marketing spend as an additional variable
- Build a forecasting module to estimate next quarter profit by store
- Apply logistic regression to model store-level success probability (binary outcome)
- Develop a Power BI dashboard using the same model logic for executive access