Project Title

Supermart Profit Prediction & Business Analytics System

Business Problem

Retail businesses often struggle to predict profitability due to fluctuating sales, varying discount strategies, seasonal effects, and product-level differences. Incorrect pricing or discount decisions can significantly impact profit margins.

Business Questions Addressed:

- How do sales, discounts, categories, and regions affect profit?
- Can profit be predicted in advance for a given order?
- How can business users simulate scenarios before making pricing decisions?

Objective

The main objective of this project was to:

- Analyze historical supermarket sales data
- Identify key factors influencing profit
- Build a machine learning model to predict profit
- Deploy the solution as an interactive dashboard for real-time decision support

---

Dataset Overview

The dataset contains historical grocery sales data with the following key attributes:

- Order information (Order ID, Customer Name)
- Product details (Category, Sub-category)
- Location details (City, Region, State)
- Transaction details (Sales, Discount, Profit)
- Time-related information (Order Date)

This dataset represents real-world retail transactions, making it suitable for applied business analytics.

Step 1: Data Cleaning & Preprocessing

The following preprocessing steps were performed:

- Verified column names and data types
- Converted numerical columns such as Sales, Discount, and Profit to appropriate formats
- Removed irrelevant columns (Order ID, Customer Name, State) for modeling
- Handled date-related inconsistencies and extracted useful time features

---

☐ Step 2: Exploratory Data Analysis (EDA)

Exploratory analysis was performed to understand:

- Distribution of sales and profit
- Relationship between discount and profit
- Category-wise and region-wise profit contribution
- Seasonal patterns based on year and month

This step helped uncover important business insights such as:

- Higher discounts often reduce profitability
- Certain categories consistently generate higher profit
- Sales volume plays a major role in profit variability

☐ Step 3: Feature Engineering

To improve model learning, additional features were created:

- Year, Month, Quarter from order date
- Weekend indicator (weekday vs weekend)
- Encoded categorical variables using one-hot encoding
- Ensured all features were numeric for ML models

Feature engineering helped capture temporal and categorical effects more effectively.

☐ Step 4: Model Development

Baseline Model: Linear Regression

- Used as a baseline to understand linear relationships
- Provided interpretability but limited ability to model complex patterns

Advanced Model: Random Forest Regressor

- Implemented to capture non-linear relationships
- Better handled interactions between sales, discount, category, and region
- Provided feature importance insights

Both models were evaluated using:

- Mean Absolute Error (MAE)
- R² Score

☐ Step 5: Model Evaluation & Interpretation

The evaluation showed:

- Moderate predictive performance due to data noise and missing cost-related variables
- Strong influence of discount, sales volume, and product category on profit
- Realistic performance aligned with real-world retail forecasting challenges

The results emphasized that **data quality and business context** are as important as model complexity.

☐ Step 6: Model Deployment (Streamlit)

The final Random Forest model was deployed using **Streamlit Cloud**:

- Users can input order details such as sales, discount, category, region, and time
- The app provides real-time profit predictions
- Business warnings and insights are displayed for high discounts or risky scenarios
- A clean, user-friendly dashboard enables non-technical users to interact with the model

This transformed the model from a technical artifact into a **decision-support tool**.

☐ Business Impact & Insights

Key insights derived from the project:

- High discounts significantly reduce profitability
- Certain categories and regions consistently perform better
- Profit prediction can support pricing, promotion, and sales planning
- Scenario-based analysis helps reduce business risk