Data Analytics & Predictive Modeling Project Worksheet

# Sales Performance and Customer Behavior Analysis for Britania Retail

Dataset: Britania retail sales.csv

Domain: Retail Sales, Marketing & Customer Behavior

**1. Problem Statement**

Britania Retail wants to understand:

* What factors affect **monthly sales revenue**?
* Is there any **correlation** between marketing spend, customer visits, discounts, and sales?
* Can we **predict next month’s sales** using regression?
* How can management **visualize insights** in an interactive **Power BI dashboard**?

# Section A — Understanding the Dataset

* Load the dataset retail\_sales.csv into your preferred tool (SQL / Python / Excel / Power BI).
* Identify the following:  
  - Number of records and variables  
  - Data types of each variable  
  - Range (min, max, mean) of numerical variables (Store\_Visits, Marketing\_Spend, Discount\_Percentage, Competitor\_Price\_Index, Sales\_Revenue).
* Are there any missing or unusual values in the dataset? If yes, how would you handle them?

# Section B — SQL Exploration

* Write SQL queries to answer:
* a. What is the average monthly Sales Revenue?
* b. In which month was the highest Sales Revenue recorded?
* c. What is the total Marketing Spend over the dataset period?
* d. List months where Discount Percentage > 20%.
* e. Which months had Competitor Price Index < 100 (competitors cheaper)?

# Section C — Exploratory Data Analysis (Python)

* Import the dataset into Python (pandas).
* Generate summary statistics using df.describe().
* Plot:  
  - Sales Revenue trend over time (line chart)  
  - Scatter plot of Marketing Spend vs Sales Revenue  
  - Histogram of Store Visits
* Create a correlation heatmap for all numerical variables.
* Answer: Which variables are most positively correlated with Sales Revenue? Which variable is negatively correlated?

# Section D — Feature Engineering

* Create the following new features in Python:  
  - Sales\_Lag\_1: Previous month’s Sales Revenue  
  - Sales\_Roll3: 3-month rolling average of Sales Revenue

# Section E — Regression Modeling

* Standardize the input features using StandardScaler.
* Train a Multiple Linear Regression model to predict Sales\_Revenue using:  
  Store Visits, Marketing Spend, Discount Percentage, Competitor Price Index, Sales\_Lag\_1, Sales\_Roll3
* Interpret the coefficients: Which variables have the strongest positive or negative influence on Sales Revenue?
* Plot Predicted vs Actual Sales Revenue for the set.
* Create a Residual Plot. What do you observe?

# Section F — Forecasting

* Using the trained model, forecast Sales Revenue for the next month (e.g., Jan 2025).
* Assume Marketing Spend increases by 10% compared to the last month. Keep other features the same as last month.
* What is the predicted Sales Revenue?

# Section G — Power BI Dashboard

* Import the dataset into Power BI.
* Create the following visuals:  
  - Line Chart: Sales Revenue trend over time  
  - Bar Chart: Discount Percentage vs Sales Revenue  
  - Scatter Plot: Marketing Spend vs Sales Revenue  
  - Card Visuals: Total Sales, Avg Discount, Avg Marketing Spend  
  - Add a slicer for Year and Month
* Add a Forecast to the Sales Revenue line chart using Power BI’s Analytics pane.

Bonus: Import the model’s predicted values into Power BI and plot Actual vs Predicted.

# Section H — Insights

* Based on your analysis, answer:  
  - How does Marketing Spend impact Sales Revenue?  
  - Does offering higher discounts always increase Sales?  
  - How sensitive are Sales to competitor prices?
* Suggest 3 actionable recommendations to increase sales for Britania Retail Company.

# ✨ Deliverables

- SQL query file (.sql)

- Python notebook/script (.ipynb or .py)

- Power BI dashboard (.pbix)

- Final report with insights (2–3 pages)