# CAPSTONE PROJECT REPORT

## Walmart Sales Forecasting & Online Retail Analysis

1. **Problem Statement**

Retail businesses face challenges in managing inventory due to fluctuating customer demand.
In Problem Statement 1, a retail store with multiple outlets (Walmart) wants to analyze historical sales data and forecast future sales to improve inventory planning.

In Problem Statement 2, an online retail store aims to understand customer purchasing behavior to identify buying patterns, popular products, country-wise sales, and customer value.

**2. Project Objective**

The main objectives of this project are:

- To analyze historical sales data and identify trends
- To forecast future sales for better inventory management
- To understand customer buying behavior in online retail
- To identify high-value customers and popular products
- To provide data-driven insights for business decision-making

**3. Data Description**

Dataset 1: Walmart Sales Data

- File Name: Walmart (1).csv
- Total Rows: 6435
- Total Columns: 8

**Features:**

- Store: Store number
- Date: Week of sales
- Weekly_Sales: Sales amount
- Holiday_Flag: Holiday indicator
- Temperature: Temperature on sale day

- Fuel_Price: Fuel cost
- CPI: Consumer Price Index
- Unemployment: Unemployment rate

### Dataset 2: Online Retail Data

- File Name: Online Retail.csv
- Total Rows: 387961
- Total Columns: 8

Features:

- InvoiceNo, StockCode, Description
- Quantity, InvoiceDate
- UnitPrice, CustomerID, Country

## 4. Data Pre-processing Steps and Inspiration

The following data preprocessing steps were performed:

- Removal of missing and null values
- Conversion of date columns into datetime format
- Removal of cancelled transactions
- Feature engineering such as Year, Month, and TotalAmount
- Outlier handling for extreme values

These steps were inspired to improve data quality and model performance.

## 5. Choosing the Algorithm for the Project

- For Walmart Sales Forecasting:
  **Linear Regression / Random Forest Regression** was used to predict weekly sales.
- For Online Retail Analysis:
  **Exploratory Data Analysis (EDA)** and aggregation techniques were used to analyze customer behavior.

## 6. Motivation and Reasons for Choosing the Algorithm

- Linear Regression is simple, interpretable, and effective for continuous target variables.
- Random Forest handles non-linearity and improves accuracy.
- EDA techniques help in understanding customer patterns without complex modeling.
- These methods provide clear and actionable business insights.

## 7. Assumptions

- Historical sales patterns will continue in the future
- External factors such as economy remain stable
- Data provided is accurate and reliable
- Customer behavior patterns do not change drastically in short periods

## 8. Model Evaluation and Techniques

The model performance was evaluated using:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Comparison of actual vs predicted sales

  Visualization techniques such as line charts and bar plots were also used to validate results.

## 9. Inferences from the Same

- Sales show seasonal trends and holiday impact
- A small group of customers contributes to high revenue
- Certain products are consistently popular
- UK contributes the highest sales in online retail
- Sales forecasting helps in better inventory planning

## 10. Future Possibilities of the Project

- Incorporating advanced time-series models like ARIMA or LSTM
- Adding promotional and marketing data
- Real-time sales forecasting dashboards
- Customer segmentation using clustering techniques
- Integration with business intelligence tools