**Project Synopsis: Walmart Data Analysis**

**Batch Code:** ANP-C8220

**Project Partner:** Harshada Pawar and Pooja Ingale

**Trainer:** Rajshri Thete ma’am

**Tools and Technologies:**

**Programming Language:** Python

**Libraries:** Numpy, Pandas, Matplotlib, Scipy

**IDE:** VS Code

**Source of Data:** <https://www.kaggle.com/datasets/asahu40/walmart-data-analysis-and-forcasting>

1. **Title:**

Walmart DataAnalysis

1. **Introduction:**

Walmart, being one of the largest retail chains in the world, has a vast range of products and sales data. Analysing this data provides insights into consumer behaviour, sales patterns, and business performance. This project aims to analyse a dataset of Walmart sales to identify the key factors influencing sales performance. By leveraging data analysis techniques, we can extract valuable insights to help Walmart enhance its business strategies and decision-making processes.

1. **Objectives:**

The primary objectives of this project are:

* + To explore and understand the features of the Walmart sales dataset.
  + To perform data preprocessing, including handling missing values and outliers.
  + To identify key factors that affect Walmart's sales using statistical analysis.
  + To build predictive models that can forecast Walmart's sales performance accurately.
  + To visualize the results and provide actionable insights for business improvement.

1. **Scope of Work:**

The project will involve the following tasks:

* + **Data Exploration:** Understanding the Walmart dataset, including the features and target variables such as sales and dates.
  + **Data Preprocessing:** Cleaning the dataset by handling missing values, removing outliers, and normalizing/standardizing the data.
  + **Feature Selection:** Identifying the most significant features influencing Walmart's sales (e.g., store location, holiday effects, promotions, etc.).
  + **Data Visualization:** Using plots and graphs to visualize sales trends, seasonality, and relationships between variables.
  + **Model Building:** Building and evaluating machine learning models to predict sales trends and performance.
  + **Interpretation of Results:** Analyzing model outputs and drawing conclusions on factors impacting Walmart sales.
  + **Reporting:** Documenting the findings and preparing a final report for strategic insights.

1. **Methodology:**

The project will follow a structured approach:

* 1. **Data Collection:** The dataset will be sourced from a public repository like Kaggle (Walmart Sales Dataset).
  2. **Data Preprocessing:**
     + Handle missing data using imputation techniques.
     + Detect and remove outliers.
     + Normalize or standardize the data if necessary.
  3. **Exploratory Data Analysis (EDA):**
     + Use descriptive statistics to summarize the dataset.
     + Create visualizations like line plots, box plots, and heatmaps to understand sales patterns and relationships.
  4. **Feature Selection:**
     + Use correlation analysis to identify relevant features (e.g., holidays, promotions).
     + Apply dimensionality reduction techniques like PCA if necessary.
  5. **Modelling:**
     + Split the data into training and testing sets.
     + Train multiple models (e.g., Linear Regression, Decision Trees, Random Forest, etc.) and evaluate their performance using metrics like R-squared, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).
     + Tune hyperparameters to optimize model performance.

1. **Evaluation and Interpretation:**
   * + Compare model performance.
     + Interpret results to understand the impact of different features on sales performance.
2. **Visualization:** o Generate charts and graphs to visualize the findings.
3. **Reporting:**

o Compile the analysis, results, and insights into a comprehensive report.

1. **Tools and Technologies:**

The project will utilize the following tools and technologies:

* + **Programming Language:** Python
  + **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
  + **IDE:** Jupyter Notebook or any Python-compatible Integrated Development Environment (IDE)
  + **Data Source:** Kaggle (Walmart Sales Dataset)

1. **Expected Outcomes:**
   * Identification of the most significant factors influencing Walmart sales.
   * Development of a predictive model with high accuracy for sales forecasting.
   * Visualization of sales data and model results to provide actionable insights for Walmart's business strategies.
   * A comprehensive report documenting the analysis process, findings, and recommendations for improving sales performance.
2. **Timeline:**

The project is expected to be completed within a [specific timeframe, e.g., 4 weeks], with the following milestones:

* + **Week 1:** Data Collection and Preprocessing
  + **Week 2:** Exploratory Data Analysis and Feature Selection
  + **Week 3:** Model Building and Evaluation
  + **Week 4:** Visualization, Reporting, and Final Submission

1. **Conclusion:**

This project will provide valuable insights into the factors that determine Walmart's sales performance, leveraging data analysis techniques. The results of this analysis can help Walmart optimize its sales strategies, improve customer satisfaction, and enhance overall business performance.