🧠 Capstone Project Title:

“Optimizing Walmart Store Operations Using Machine Learning”

🎯 Objective:

Leverage classical ML techniques to **segment stores**, **predict key performance indicators**, and **detect operational anomalies** using Walmart store-level data.

📦 Dataset Description:

You can provide a dummy or publicly available dataset with the following (or similar) structure:

Store_ID	Weekly_Sales	Store_Size	Region	Avg_Income	Holiday_Flag	Footfall	Inventory_Level	Promo_Spend	Date

(You may simulate this with dummy data in CSV or use the Walmart Sales Forecasting dataset on Kaggle for realism.)

📚 Project Phases:

Phase 1: Exploratory Data Analysis & Problem Framing

- Perform EDA to understand trends in sales, promotions, and footfall.
- Identify missing values, outliers, and basic statistical summaries.
- Define the business problems:
    - **Sales forecasting** (regression)

- **Store segmentation** (clustering)
- **Anomaly detection** (monitoring unusual store behavior)

Phase 2: Supervised Learning – Regression & Classification

1. **Regression Task**:
   - **Objective**: Predict **weekly sales** based on store attributes and promotions.
   - Apply: Linear Regression, Random Forest Regressor
   - Evaluate: RMSE, MAE, R^2 Score
2. **Classification Task**:
   - **Objective**: Predict whether a store will have a **high sales week** (binary classification).
   - Convert weekly sales into binary target using a threshold.
   - Apply: Logistic Regression, Decision Trees
   - Evaluate: Accuracy, Precision, Recall, F1-Score, Confusion Matrix

---

Phase 3: Unsupervised Learning – Clustering

- **Objective**: Segment stores based on `Store_Size`, `Region`, `Avg_Income`, `Footfall`, etc.
- Techniques:
  - K-Means, DBSCAN
  - PCA for dimensionality reduction before clustering
- Visualize clusters and describe common characteristics of each segment.

- **Business Goal**: Recommend regional marketing or stocking strategies per segment.

Phase 4: Anomaly Detection

- **Objective**: Detect anomalous store behavior such as:
    - Sudden drop in sales
    - Unusual footfall
    - Promo budget overspending
- Techniques:
    - Z-Score/Box Plot for statistical anomalies
    - Isolation Forest or One-Class SVM for multivariate anomalies
- Output: List of flagged stores and explanation of detected anomalies.

📊 Deliverables

Participants must submit the following:

1. **Jupyter Notebook or Python scripts** with:
    - Data loading and cleaning
    - EDA and visualizations
    - Model training, evaluation, and interpretation
    - Cluster visualizations and anomaly explanations
2. **Capstone Report/Presentation** (optional):
    - Business insights from model results

- Recommendations for Walmart's operations team
- Challenges faced and how they were addressed

🧰 Tools/Frameworks to Use

- **Python**, **Pandas**, **NumPy**
- **scikit-learn**, **matplotlib**, **seaborn**
- **Jupyter Notebook** or **VS Code**

🏁 Optional Bonus Extension

"Deploy a simple REST API to serve the weekly sales prediction model using Flask or FastAPI."