Global Super Store

Project Objective

- The Global Super Store dataset contains 24 features like sales, profits, customer, and product details.
- The goal of this project is to **predict profits** for future sales transactions based on various features like product category, shipping mode, region, and discount.

The Technologies I used

- **For Data Preprocessing**
  - **Handling Missing Values:** Checked for null values and imputed/removed them
  - **Feature Engineering:** Derived new features like Total profit (Profit – Shipping cost).
  - **Encoding Categorical Variables**: Used **label encoding** for categorical data to convert numerical values so that ml models can process them effectively
  - **Outlier Handling**: Applied **IQR method** to remove outlier
  - **Data Scaling**: Standardized numerical columns to improve model performance.

- **Model Selection & Training**
  Tried multiple regression models
  - **Linear Regression:** Baseline model.
  - **Decision Tree Regressor:** Captured non-linearity but prone to overfitting.
  - **Random Forest Regressor:** Improved generalization.
  - **XGBoost:** Provided the best results with feature importance analysis.

- **Model Evaluation**
  - Used Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 Score for evaluation.
  - XGBoost achieved lowest MAE & MSE, indicating best performance.

Challenges I faced

- **Handling Outliers :** used IQR Method
- **Overfitting :** used Cross-Validation