■ Sales Forecasting Internship Tasks

Project Overview

This repository contains multiple tasks completed during the internship. The main goal is to analyze sales data, preprocess it, apply machine learning models, and forecast future sales using statistical and ML approaches.

Tasks Breakdown

- **Task 1 - Data Collection**: Loaded the dataset (Superstore Sales Data). Performed basic exploration and understood dataset structure.
- **Task 2 - Data Cleaning & Preprocessing**: Handled missing values and duplicates. Converted date columns into datetime format. Extracted useful features and encoded categorical variables.
- **Task 3 - Exploratory Data Analysis (EDA)**: Performed statistical summaries and created visualizations: sales distribution, trends, and top-selling products.
- **Task 4 - Feature Engineering**: Generated new features such as daily sales, monthly aggregated sales, and customer-level purchase history.
- **Task 5 - Model Training & Evaluation**: Applied regression models (Linear Regression, Random Forest). Evaluated with MAE, RMSE, MAPE, R^2 Score.
- **Task 6 - Forecasting with Time-Series Models**: Applied ARIMA/SARIMA models. Used pmdarima for auto selection. Compared forecasts against actual sales.

How to Run Each Task

1. Navigate to the respective task folder.
2. Install dependencies using **pip install -r requirements.txt**.
3. Run the task using **python task_X.py**.

Dependencies

Each task has its own requirements.txt. Key libraries used across tasks:

- pandas
- numpy
- matplotlib
- seaborn
- scikit-learn
- statsmodels
- pmdarima

Results

✔■ Cleaned & structured dataset ready for ML.
✔■ Detailed EDA insights for decision making.
✔■ Trained regression & time-series models.
✔■ Generated sales forecasts for business planning.

Author

■■■ Muhammad Umar Bilal
Internship Project – Sales Forecasting