Task: Sales Performance Analysis and Forecasting

Objective:

The objective is to perform in-depth sales analysis to uncover patterns, identify top and low-performing products or regions, and forecast future trends. The analysis will include data cleaning, transformation, exploratory data analysis (EDA), and visualization to support strategic decision-making.

Steps:

1. **Data Collection and Exploration**
    - Dataset Loading:
      Load the sales dataset using Pandas.
    - Initial Exploration:
      Review the structure (columns, data types, time granularity).
      Generate summary statistics to understand sales distribution and seasonal effects.
      Identify missing or inconsistent records.
2. **Data Cleaning and Transformation**
    - Handle Missing Values:
      Use interpolation or removal based on time continuity or data sparsity.
    - Outlier Detection and Treatment:
      Identify anomalies in sales or units sold using box plots or z-scores.
    - Data Transformation:
        - Encode categorical variables like product categories or regions.
        - Convert date columns to datetime format and create derived features (month, quarter, year).
        - Aggregate data for time series analysis if needed.
3. **Exploratory Data Analysis (EDA)**
    - Descriptive Statistics:
      Understand revenue distribution, product sales variance, and customer region behavior.
    - Visualization:
        - Line charts for trends.
        - Bar charts for top-performing products.

        - Heatmaps to correlate regions/products/time.
        - Seasonality and cyclic pattern analysis.
    - Key Insights:
      Identify peak sales periods, seasonal dependencies, and high-performing segments.
      Detect declining trends or underperforming markets.

4. **Documentation**
    - Describe dataset structure, cleaning steps, and key findings.
    - Comment code for each transformation and EDA process.
5. **Version Control**
    - Use Git to manage versions.
    - Create a GitHub repository with datasets, notebooks, and a README file.

Deliverables

- Jupyter Notebook or Python Script:
  Including data wrangling, visualizations, and insights.
- GitHub Repository:
  Organized with data, scripts, and project documentation.