E-commerce Sales Analysis – Documentation Guideline

Project Overview:

Project Goals:
The goal of this project is to build a Python-based data analysis pipeline that:

- Loads and cleans an e-commerce sales dataset.
- Validates and converts date fields for time-series analysis.
- Calculates key metrics such as total revenue, best-selling product, and regional sales distribution.
- Creates multiple visualizations (bar, pie, line charts) to highlight insights.
- Generates a formatted Markdown report with findings and recommendations.
- Provides a user-friendly and professional documentation structure for recruiters and reviewers.

Objectives:

- **Data Cleaning:** Handle missing values and duplicates.
- **Date Conversion:** Convert string dates into datetime for daily/monthly analysis.
- **Metric Calculation:** Compute total revenue, product sales, and regional distribution.
- **Visualization:** Create bar, pie, and line charts using matplotlib.
- **Reporting:** Export results into a Markdown report with charts and insights.
- **Documentation & Testing:** Provide README, test cases, and screenshots to demonstrate functionality.

Setup Instructions:

- **Install Python**
  - Download and install Python 3.8 or higher from [python.org](https://www.python.org/downloads/).
  - Verify installation:
    Bash python –version
- **Install VS Code**
  - Download and install [Visual Studio Code](https://code.visualstudio.com/).
  - Open VS Code after installation.
- **Create a Project Folder**
  Make a new folder for your project.
- **Open Folder in VS Code**
  In VS Code, go to File → Open Folder and select ecommerce_sales_analysis

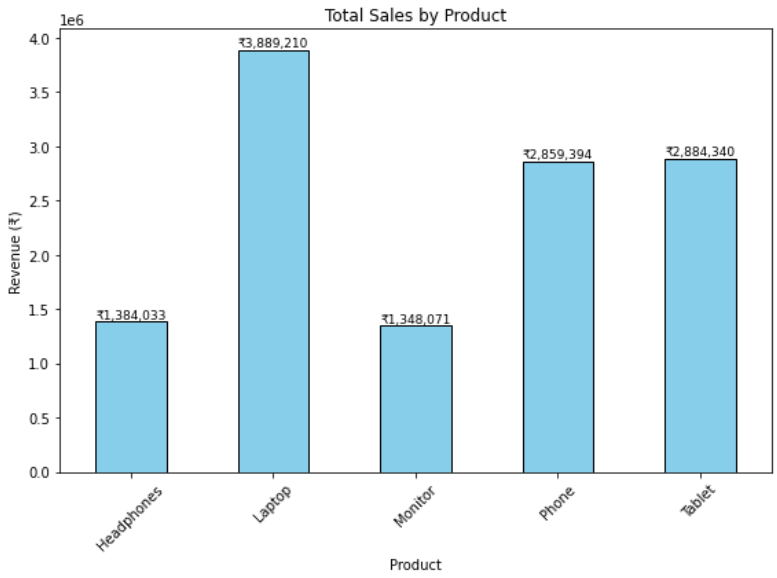
- **Add Files**
  Inside the folder, create the following files:
    - analysis.ipynb→ main Python program
    - README.md → documentation
    - requirements.txt → dependencies
    - report/sales_analysis.md → analysis report
    - visualizations/ → charts (bar, pie, line)
    - tests/test_sales_analysis.py → validation tests
    - screenshot.png → screenshot of program output

Code Structure:

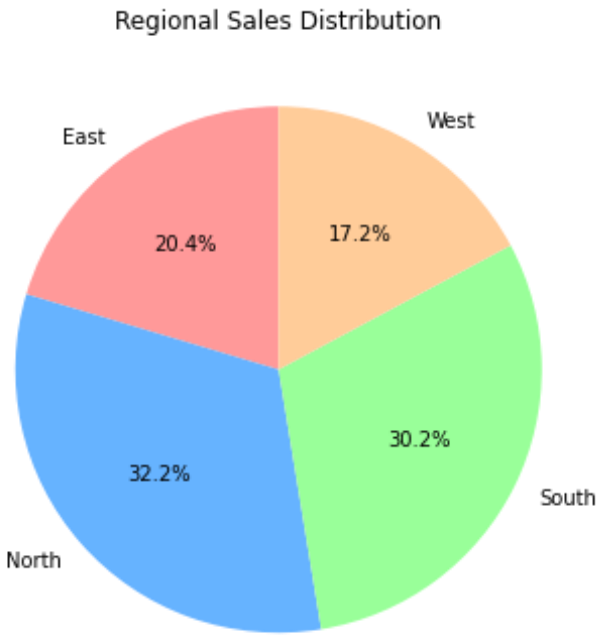
- analysis.ipynb
    - load_data() → loads CSV file.
    - clean_data() → handles missing values and duplicates.
    - analyze_sales() → calculates revenue, product, and regional metrics.
    - create_visualizations() → generates bar, pie, and line charts.
    - generate_report() → saves Markdown report with insights.
- visualizations/ → images showing charts.
- report/ → Markdown report with findings

Visual Documentation:

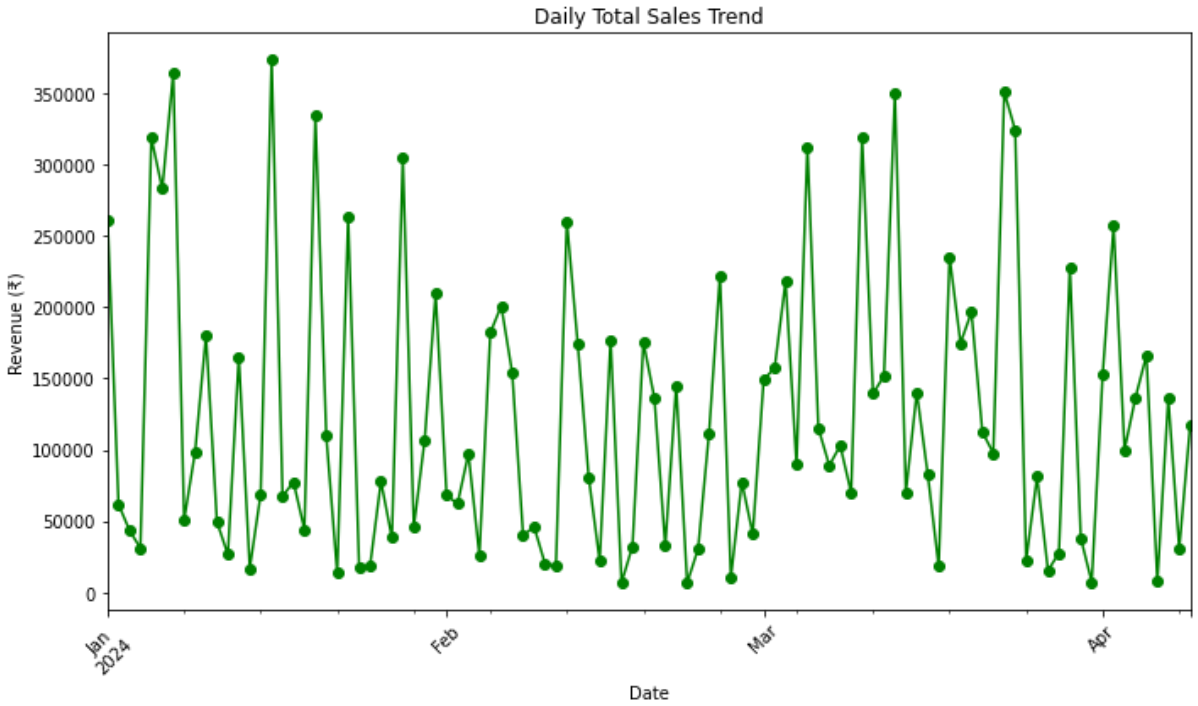
Product Sales Bar Chart:



Regional Sales Pie Chart:



Sales Trend Over Time:



Technical Details:

- **Algorithm:** Load → Clean → Aggregate → Visualize → Report.
- **Data Structures:** pandas DataFrame for data.
- **Architecture:** Single-file script with modular functions; outputs charts and a Markdown report.
- **Libraries:** pandas, matplotlib.

Testing Evidence:

- **Valid Data:** Input: Complete dataset → Output: Correct total revenue and product rankings.

```
Total Revenue: ₹12,365,048.00
Best-Selling Product: Laptop (₹3,889,210.00)

 Regional Sales Distribution:
Region
East     2519639
North    3983635
South    3737852
West     2123922
Name: Total_Sales, dtype: int64
```

- **Missing Values:** Ensure no NULL values after cleaning.

```
Null Values Check:
Date           0
Product        0
Quantity       0
Price          0
Customer_ID    0
Region         0
Total_Sales    0
dtype: int64
```