Project Overview

The **Sales Data Analysis Project** is designed to demonstrate a complete workflow for **cleaning, analyzing, and visualizing sales data** using Python. The main objectives of the project are:

1. **Data Cleaning** – Handle missing values, remove duplicates, and ensure the dataset is ready for analysis.
2. **Revenue Calculation** – Add a Revenue column (Quantity × Price) to assess financial performance.
3. **Sales Insights** – Identify top-performing products, calculate total revenue, and aggregate revenue per product.
4. **Visualization (Optional)** – Generate bar charts, pie charts, and heatmaps for clear and insightful data representation.
5. **Professional Output** – Save cleaned and analyzed datasets as CSV files for future reporting or dashboard integration.

This project is ideal for demonstrating **data manipulation, exploratory data analysis, and basic business analytics** skills in Python.

Setup Instructions

Follow these steps to set up and run the project locally:

1️⃣ Clone the Repository

git clone https://github.com/your-username/sales-data-analysis.git

cd sales-data-analysis

2️⃣ Install Required Packages

Ensure you have Python 3.x installed, then run:

pip install -r requirements.txt

3️⃣ Prepare the Data

- Place your raw sales CSV file in the data/ folder (e.g., data/sales_data.csv).
- Ensure the CSV contains at least the following columns:
    - Product

- Quantity
- Price
- (Optional) Other columns like Category, Region, etc.

4️⃣ Run Scripts

Data Cleaning

python scripts/clean_sales_data.py

- Cleans missing values and duplicates.
- Saves cleaned_sales_data.csv in outputs/.

Revenue Analysis

python scripts/analyze_sales_data.py

- Adds a Revenue column.
- Aggregates revenue per product and identifies the best product.
- Saves sales_with_revenue.csv in outputs/.

Optional Visualization

python scripts/visualize_sales_data.py

- Generates charts (bar charts, pie charts, heatmaps) and saves them in outputs/.