Python Project Documentation

1. Project Overview

This project is designed to demonstrate Python data analysis skills using **Pandas** and **Matplotlib**.
The dataset contains raw sales records, and the goal is to **clean, explore, and extract useful business insights**.
This simulates the role of a Data Analyst — transforming messy data into clear, actionable findings.

2. Objectives

1. Import the dataset into Pandas and understand its structure.
2. Perform **data cleaning**: handle missing values, remove duplicates, fix formatting issues, and set correct data types.
3. Use **descriptive statistics and aggregation** to explore trends and summaries.
4. Answer business questions related to sales performance.
5. Create **visualizations** (bar charts, line graphs, pie charts) for better understanding.
6. Document findings in a clear report.

3. Tools & Libraries

- **Python** (3.x)
- **Pandas** – data manipulation
- **Matplotlib** – data visualization
- **Jupyter Notebook** – analysis & reporting

4. Dataset Description

The dataset contains the following columns:

- **Date** – Date of transaction
- **Order ID** – The ordering number
- **Product** – Name of the product sold
- **Quantity** – Number of units sold
- **Price** – Price per unit
- **Revenue** – Total sales revenue (calculated as *Quantity* × *Price*)
- **Payment Method** – Mode of payment (Cash, Card, Transfer, etc.)
- **City** – Location of sale
- **Manager** – Sales manager responsible for transaction

5. Data Cleaning

Steps taken to clean the dataset:

- Imported the data and remove the first row to make the second row the header with (header =1)
- Removed empty/unnamed columns.
- Dropped duplicates.
- Handled missing values (dropped or filled depending on column).
- Converted Date column to datetime format.
- Trimmed extra spaces in string columns (Product, City, Payment Method).
- Created a new **Revenue column** from Price × Quantity.

6. Exploratory Data Analysis (EDA)

- Checked summary statistics (.describe()).
- Verified data types (.info()).
- Counted unique products, cities, managers, and payment methods.
- Visualized distributions of products, revenue, and payment methods.

7. Business Questions & Answers

1. **Most Preferred Payment Method**
    - Found using value_counts() on Payment Method.
    - Pie chart created to show distribution.
2. **Most Selling Product (Quantity & Revenue)**
    - Grouped by Product, aggregated sum() of Quantity and Revenue.
    - Bar chart (descending order) for both.
3. **City with Maximum Revenue & Manager with Maximum Revenue**
    - Grouped by City and Manager.
    - Extracted maximum values.
4. **Average Revenue**
    - Computed using .mean().
5. **Average Revenue of November & December**
    - Filtered dates with .dt.month in [11, 12], then .mean().
6. **Standard Deviation of Revenue and Quantity**
    - Used .std().
7. **Variance of Revenue and Quantity**
    - Used .var().
8. **Revenue Trend Over Time**

- Grouped by month (.dt.to_period("M")) and plotted line chart.

9. **Average Quantity Sold & Average Revenue per Product**
    - Used groupby('Product').mean().
10. **Total Number of Orders/Sales Made**
    - Counted rows with df.shape[0].

8. Visualizations

- **Bar Charts** – for top products, cities, managers.
- **Pie Charts** – for payment method distribution.
- **Line Chart** – for monthly revenue trend.
- **Histograms** – for distribution of quantity & revenue.

All charts were sorted (ascending/descending) for clarity and included **percentages/values** where applicable.

9. Key Insights

- The most preferred payment method was **Credit card**.
- The top-selling product by quantity was **Beverages**, while by revenue it was **Burgers**.
- **Lisbon** generated the highest revenue, managed by **Joao Silva**.
- Average revenue was **3029.589996850394**, while Nov–Dec average was **2939.0677734513274**
- Revenue showed an **increasing trend** over time.

10. Conclusion

This analysis provided actionable insights into sales performance across products, payment methods, cities, and managers. Businesses can use this information to:

- Focus on top-performing products.
- Promote the most popular payment method.
- Strengthen sales in top revenue-generating cities.
- Track revenue trends to plan marketing and promotions.