Task 1 - Exploratory Data Analysis (EDA)

Internship at Nexus AI Digital — Data Science & Machine Learning
Author: Syed Huzaifa Bin Khamis

Objective:
Before building predictive models, it is important to understand the data.
This task explores the **Online Retail dataset** through cleaning, descriptive statistics, and visualization.
The goal is to answer: *What are customers buying? When are they buying it?

Steps Taken:
1. Data Loading— Loaded the CSV file using pandas.
2. Data Cleaning — Removed missing `CustomerID`, dropped duplicates, and converted date fields.
3. Descriptive Stats — Calculated mean, median, mode, and spread for `Quantity`, `UnitPrice`, and `TotalSales`.
4. Visualizations
   - Histogram of customer sales
   - Bar chart of sales by product
   - Line chart of monthly sales
   - Sales by country
   - Customer spend distribution

Key Insights:
1. A small number of products generate the majority of sales.
2. The UK contributes the highest revenue, while only a few other countries show significant activity.
3. Seasonal patterns exist: noticeable peaks in November and December.
4. Customer spending is highly skewed — a few high-value customers dominate revenue.
5. Many purchases are small-quantity, low-value transactions, indicating impulse buys.

How to Run:
1. Clone this repository or download the files.
2. Install dependencies:
   ```bash
   pip install -r requirements.txt
```