## 1️⃣ Tools Used

📌 **Programming & Data Tools:**

- **Python (VS Code)** - Data wrangling, cleaning, and visualization
- **Excel** - Additional data manipulation & formatting

📌 **Python Libraries Used:**

- **pandas** - Data manipulation
- **numpy** - Numerical operations
- **seaborn & matplotlib** - Data visualization
- **os & io** - File management
- **colorama** - Terminal color formatting (optional, but makes output more readable)

---

## 2️⃣ Data Wrangling Process

I developed a **Python wrangling script** to automate key preprocessing steps:

✔ **Automated CSV File Detection**

- Instead of manually specifying file paths, the script **walks the OS** to locate the dataset.
- Helps streamline my workflow since I store datasets separately from Python scripts.

✔ **Initial Data Overview**

- Loads data as a **pandas DataFrame**.
- Prints a **colorized summary** of the dataset (for readability).

✔ **Handling Duplicates**

- Detects duplicate rows and **prompts the user** to remove them (`y/n` input).

✔ **Data Exploration & Column Filtering**

- **Generates a histogram** of the dataset for quick insights.
- Prompts the user to drop unnecessary columns.
- Allows multiple column selections via **comma-separated input**.

✔ **Handling Missing Values**

- Offers three options:
    1. **Normalize** missing values
    2. **Drop** missing data
    3. **Fill with zero**

✔ **Final Cleanup & Export**

- Saves the cleaned dataset as `wrangled_{fname}.csv` to maintain version control and avoid overwriting the original file.

---

## 3️⃣ Key Data Cleaning & Transformations

🟢 **Dropped the 'Employee' Column** - It didn't seem relevant to product sales.
🟢 **Fixed Data Types** - Converted **'Amount'** from an object to an integer for analysis.
🟢 **Corrected Typos** - Fixed a recurring typo in the **'Product'** column (`Sliky -> Silky`).
🟢 **Grouped Data for Analysis** - Created different **aggregations** (`Country`, `Product`, `Amount`).
🟢 **Adjusted Date Format** - Changed **'Date'** column to `DD.MMM.YYYY` for better readability.

---

## 4️⃣ Visualizations & Insights

📊 **Experimented with Multiple Groupings:**

- Some visualization ideas worked better than others.
- Lesson learned: **Not all visualizations are useful**—some seemed great in theory but didn't provide meaningful insights.

❌ **Mistake:** I didn't screenshot my code while working on visualizations.
🔹 **Lesson Learned:** In future projects, I'll document my analysis better to track what worked and what didn't.

---

## 5️⃣ Final Analysis & Takeaways

- **Did not restore the 'Employee' column** - It didn't add value to my analysis.
- **Would appreciate feedback** - Was dropping this column a mistake for the data story?

---

### 🔹 How This Can Be Improved Further

💡 **Next Steps:**
✅ Capture screenshots of analysis steps for better documentation.
✅ Consider interactive visualizations (e.g., **Plotly** or **Streamlit**).
✅ Improve README structure in GitHub for better clarity.