**Excel Project - Retail Sales Analysis**

**Scenario**

You are a Data Analyst at an e-commerce company aiming to enhance sales strategies and customer satisfaction.  
Your objective is to analyze online retail transaction data to uncover insights into sales performance, customer behaviour, and product trends.

**Dataset Description**

- Invoice No: Unique identifier for each transaction  
- Stock Code: Product/item code  
- Description: Product description  
- Quantity: Number of items purchased  
- Invoice Date: Date and time of transaction  
- Unit Price: Price per unit in GBP  
- Customer: Unique identifier for each customer  
- Country: Country of the customer

**Objectives**

Clean and prepare retail sales data in Excel by fixing errors, creating new features, and organizing it properly. Analyze sales trends, customer behavior, product performance, and geographical insights through simple calculations and visualizations.  
Create a clean, professional dashboard with filters to easily explore and update the sales data.

**Data Cleaning Steps:**

* **Customer ID Column:** Handled blank values in the Customer ID column. If the analysis focuses on customers, blank values must be removed. If the focus is on products or inventory, keeping or removing them is optional unless they cause issues during analysis.
* **Invoice No column:** 2nd thing is that in invoice no column some values were numeric, and some were like alphabet at the start (C) so i flag them as deliver order & Cancel order.
* **Stock code column**: Removed text values (e.g., Post, D, M) from Stock Code column because their descriptions were unclear and did not correspond to valid products. Since working independently, unable to verify with team; decision made to improve data quality for product sales analysis.
* **Date column:** In date column i extract hours, day of the week, Month & Year for better analysis.
* **Quantity Column: Identified** and removed 9 extreme outliers (both high positive and high negative values) in the Quantity column. For the remaining values, applied a threshold of 1,500 to flag unusually large transactions and reduce confusion. Created two custom columns: **Quantity Status** (Sales, Return) to classify transaction type, and **Quantity Outlier Flag** (Flag, Normal) to indicate whether a value exceeds the threshold.
* **Country Column:** I found some "Unspecified" values in the Country column, so I did not impute them with the mode because that would artificially increase the share of a particular country and affect the analysis. My dataset has 400K rows, and the "Unspecified" values total just 244, which is about 0.06%. Therefore, I removed these rows, thinking that this small number would not impact future analysis.
* I created a total of **10 custom columns** from the dataset using Excel formulas for better and deeper analysis:
* **Order Status**
* **Stock/Product Code Status**
* **Quantity Status**
* **Quantity Outlier Flag**
* **Date Breakdown** (Hour, Day of Week, Month, Year)
* **Total Revenue** and **Revenue Status**

**Insights**

**General Questions**

1. Which month had the highest sales, and what was the revenue?

November recorded the highest sales revenue at £1,142,145.77, followed by October with £1,004,634.90.  
The lowest sales revenue was in February, at £443,346.02.

2. Which top 10 countries sold the most products?

The United Kingdom sold the most products, with 354,001 units, followed by Germany (8,659 units) and France (8,034 units). The lowest sales among the top 10 countries were in Australia (1,184 units) and Portugal (1,425 units). The large gap between the UK and the other countries suggests a strong home-market bias, while the relatively low sales in Australia and Portugal may require further investigation into factors such as product demand, marketing strategies, or distribution efficiency.

3. At what time of day do we have the highest sales?

Sales peaked at **12 PM** with revenue of £1,364,948.91, followed by **10 AM** with £1,166,648.98. The lowest sales occurred at **6 PM**, with only £4.25 in revenue.  
*Possible reason:* Midday and late morning could align with higher customer activity (e.g., lunch breaks or coffee runs), while evenings see fewer transactions.

“12 PM is likely the peak time for lunch and tea, which may explain the highest sales. Similarly, 10 AM could be the peak time for coffee lovers, contributing to elevated sales during that hour.

**Note**: it’s **a hypothesis**, not a confirmed fact. And let me tell me this insight is based on observed data patterns and represents a possible explanation further validation is recommended.

4. Are there specific days where the number of items sold peaks or dips significantly?

Thursday has the highest sales with approximately **1,145,169** items sold, followed by Wednesday with **960,199**. The lowest sales occur on Sunday, likely because it is a rest day when people tend to stay home. The higher sales on Thursday may be due to customers preparing for the upcoming weekend.  
*Note: This is a hypothesis based on observed data patterns and may require further validation.* *I am working on independently on this project I might be wrong.*

5. Which top 10 products sold the most?

Stock code **22423** generated the highest sales revenue at **£142,567.45**, while stock code **22386** had the lowest s**ale**s among the top 10, with **£37,289.59**.

**Sales and Revenue Insights**

1. What is the total revenue over time? (daily and monthly trends)

Thursday has the highest sales at approximately £1,941,266, followed by Tuesday (£1,597,358) and Wednesday (£1,563,522). The lowest sales occur on Sunday (£786,763), possibly because people prefer to stay home. Higher sales on Thursday may be due to early weekend shopping preparations.

2 What are the top-selling products by quantity sold?

Product codes **84077** and **22197** have very similar quantities sold, with **49,615** and **49,183** units respectively. The product with the lowest quantity sold among the top listed is **23084**, with **27,202** units sold.

**Customer Behavior Analysis**

1 How frequently do customers make purchases on average?

Since your data covers 2 years, here’s how to interpret 92.9 average purchases per customer:

* On average, each customer made about 93 purchases over 2 years.
* That’s roughly 46.5 purchases per year per customer (92.9 ÷ 2).

This means customers buy frequently, almost once every week!

**Thanks! 😊.**