Online Retail Transaction Analysis

Prabudh Raghuram

1. Project Objective

The objective of this project was to conduct a comprehensive analysis of a transactional dataset from a UK-based online retailer. Using advanced features within Microsoft Excel, the goal was to transform raw transactional data into actionable commercial insights. The primary focus was on identifying key revenue drivers and flagging underperforming product lines to support strategic decision-making in inventory management, procurement, and marketing.

2. Methodology & Technical Skills Demonstrated

This project was executed in three distinct phases, each utilising specific advanced Excel functionalities to ensure data integrity and produce meaningful analysis.

Phase 1: Data ETL & Preparation with Power Query

The initial dataset contained over 540,000 rows and required significant cleaning before analysis. **Power Query** was chosen as the primary tool for this Extract, Transform, and Load (ETL) process due to its robustness and ability to create a repeatable, automated data-cleaning workflow.

Key Data Transformation Steps:

- **Data Ingestion:** The raw .csv file was loaded into Power Query directly, maintaining the integrity of the source data.
- **Data Type Correction:** Ensured all columns were assigned the correct data type (e.g., InvoiceDate to DateTime, CustomerID to Text to prevent numerical aggregation).
- **Filtering Invalid Data:**
    - **Cancelled Orders:** Transactions with an InvoiceNo starting with "C" were systematically filtered out to exclude returns from the revenue analysis. This was achieved using a **Text Filter**.
    - **Null Customer IDs:** Records without a CustomerID were removed to ensure that all analysed transactions could be attributed to a specific customer, which would be vital for future customer-level analysis.
- **Handling Erroneous Entries:** Transactions with a negative Quantity or a zero UnitPrice were removed to prevent data skewing. This was handled using **Number Filters**.
- **Feature Engineering:** A new column, **Revenue**, was created by multiplying Quantity and UnitPrice using a **Custom Column** formula. This is a critical calculated field for the subsequent analysis.

- **Duplicate Removal:** All fully duplicate rows were removed to ensure each transaction line was unique.

The cleaned dataset was then loaded into an Excel Table, creating a direct and refreshable link to the Power Query workflow.

Phase 2: In-Depth Analysis with Pivot Tables

With a clean and structured dataset, **Pivot Tables** were employed to perform a multi-dimensional analysis.

- **Pareto Analysis (80/20 Rule):**
    1. A Pivot Table was created with Description (Product) in the Rows and Sum of Revenue in the Values.
    2. The data was sorted in descending order of revenue to identify top-performing products.
    3. A **Calculated Field** showing the **% Running Total In** was added. This allowed for the precise identification of the top 20% of products that contributed to 80% of the total revenue, providing clear, data-driven evidence for inventory prioritisation.
- **Slow-Moving Goods Identification:**
    1. A second Pivot Table was built analysing the Sum of Quantity for each Description.
    2. This table was sorted in ascending order, immediately highlighting products with the lowest sales volume over the period. This forms the basis of a report for inventory clearance or marketing focus.

Phase 3: Interactive Dashboard & Visualisation

The findings were consolidated into a user-friendly, interactive dashboard on a dedicated worksheet. This dashboard translates the raw analysis into easily digestible visual insights.

- **Pivot Charts:**
    - A **Line Chart** visualises Revenue over Time (monthly), showing sales trends and seasonality.
    - A **Bar Chart** displays the Top 10 Products by Revenue, making it easy to see the most valuable items at a glance.
- **Slicers for Dynamic Filtering:**
    - A **Slicer** for the Country field was implemented and connected to all Pivot Tables and Charts using **Report Connections**.

- This empowers a non-technical end-user to dynamically filter the entire dashboard by country, allowing for targeted geographic analysis without needing to manipulate the underlying data.

3. Key Commercial Insights & Recommendations

- **Insight 1: Top 20% of Products Drive 80% of Revenue.**
  - **Recommendation:** The business should prioritise stock availability for these high-value products. Marketing efforts and budget should be disproportionately allocated to these items to maximise ROI.
- **Insight 2: A Long Tail of Slow-Moving Stock Exists.**
  - **Recommendation:** The slow-moving goods report should be reviewed quarterly. Consider strategies such as promotional bundling, targeted discounts, or discontinuing the lowest-performing items to free up warehouse space and capital.
- **Insight 3: The UK is the Dominant Market.**
  - **Recommendation:** While the UK is the core market, the dashboard reveals significant revenue from other European countries like Germany and France. This suggests that targeted marketing campaigns in these regions could yield substantial growth.

4. Potential Enhancements & Future Analysis

To further develop this project, the following enhancements could be made:

- **Customer Segmentation (RFM Analysis):** Analyse **R**ecency, **F**requency, and **M**onetary value for each CustomerID to segment customers into loyalty tiers (e.g., high-value, at-risk, new).
- **Market Basket Analysis:** Identify which products are frequently purchased together to inform product bundling strategies and cross-selling opportunities.
- **Time-Series Forecasting:** Use Excel's forecasting tools to predict future sales based on the historical monthly revenue data.