0. Scenario

I am an entry-level data analyst at an online retail company, helping interpret real-world data to help make key business decisions. Through visualizations and statistical analysis, I will uncover key trends, such as the busiest sales months, best-selling products, and the store's most valuable customers. Ultimately, this project aims to provide actionable insights that can drive strategic business decisions and enhance the store's overall performance in the competitive online retail market.

1. Ask

The first step of the analysis is to define the problem we are trying to solve:

By examining the store's top selling countries, best-selling products, and busiest sales months and days, what strategic business decisions can be made based on these findings?

2. Prepare

The store operates its own webstore, so all sales data is collected directly from its internal database. The first step is to perform a ROCCC assessment of the dataset:

- **R – Reliable:** The sales data comes directly from the webstore's database, making it a reliable data source.

- **O – Original:** Because the data is generated and stored by the company's own system, it is considered original.

- **C – Comprehensive:** The dataset contains all key information required for analysis, including stock codes, descriptions, quantities, unit prices, and more.

- **C – Current:** The data is extracted immediately before analysis and can be refreshed at any time thanks to a fully automated Python pipeline.

- **C – Cited:** The dataset originates entirely from the company's internal data collection processes.
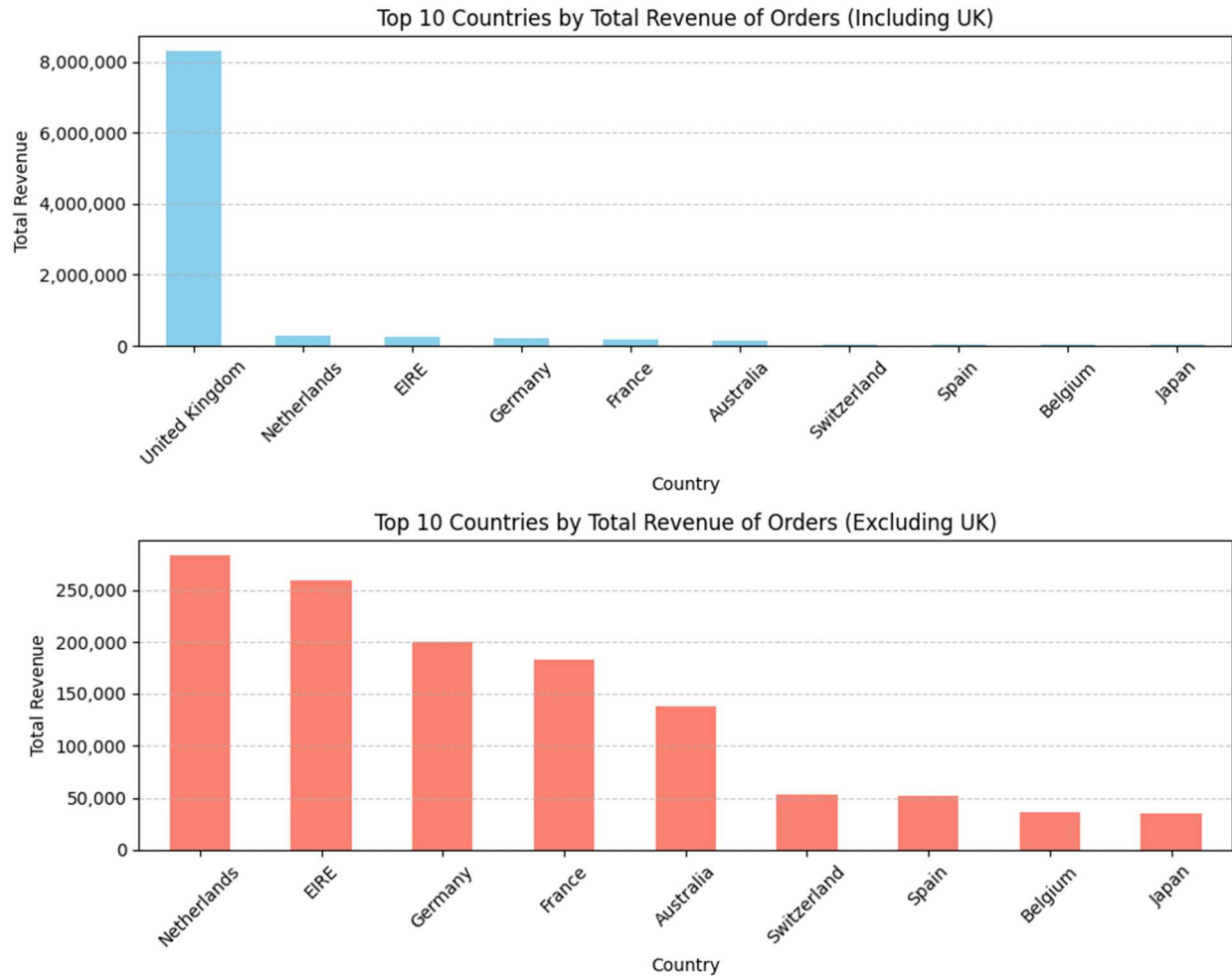
3. Process

- Data Loading & Initial Setup: Load Excel dataset, configure pandas display options, suppress warnings

- Sampling for External Analysis: Random sampling of raw data, export raw sample to Excel
- Missing Value Analysis & Cleaning: Identify missing values, analyze missing Description rows, remove rows with missing Description
- Product Description Validation (Data Cleaning): Define product validation rule (uppercase letters), identify non-product descriptions, inspect suspicious descriptions, filter dataset to valid products only, standardize StockCode datatype, remove known non-product stock codes
- CustomerID Handling: Analyze missing CustomerID values, retain rows with missing CustomerID
- Export Cleaned Dataset Sample: Sample cleaned data, export cleaned sample to Excel
- Feature Engineering: Create total transaction value (TotalPrice), extract Year–Month feature, extract day-of-week feature
- Aggregation-Ready Data Transformations (Manipulation): Prepare country-level revenue aggregations, prepare product-level revenue aggregations
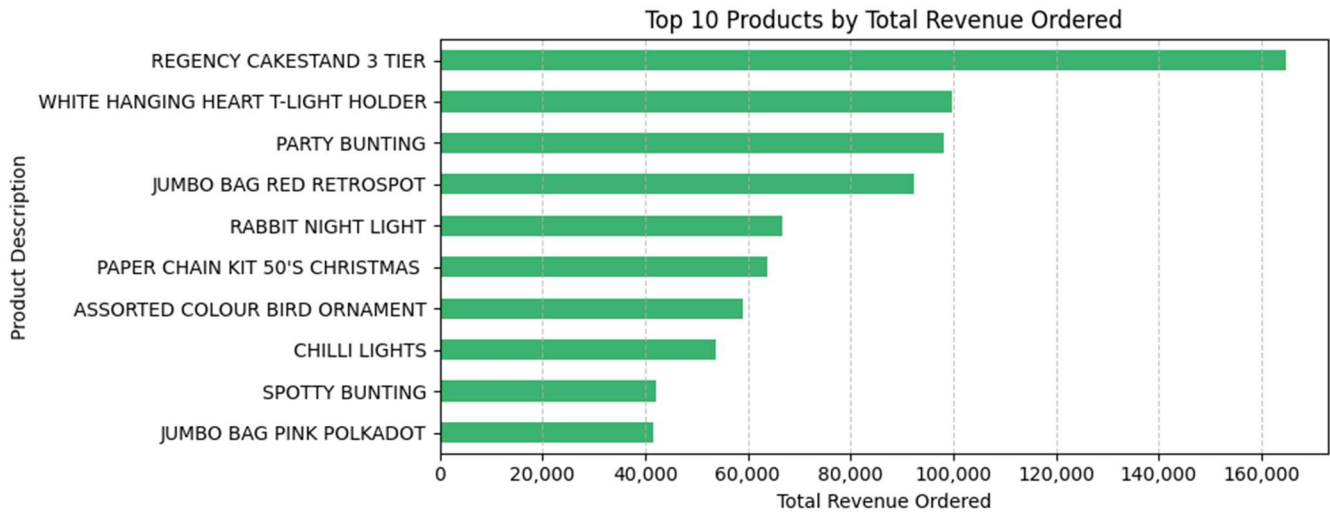
4. Analysis and Share

To identify the top 10 countries by total revenue, we first ranked all countries based on their sales performance. The United Kingdom overwhelmingly dominates the revenue, which makes it challenging to compare the remaining countries on the same scale. To provide clearer insight, a second chart is created that excludes the UK, allowing for a more meaningful comparison among the other top-performing countries.
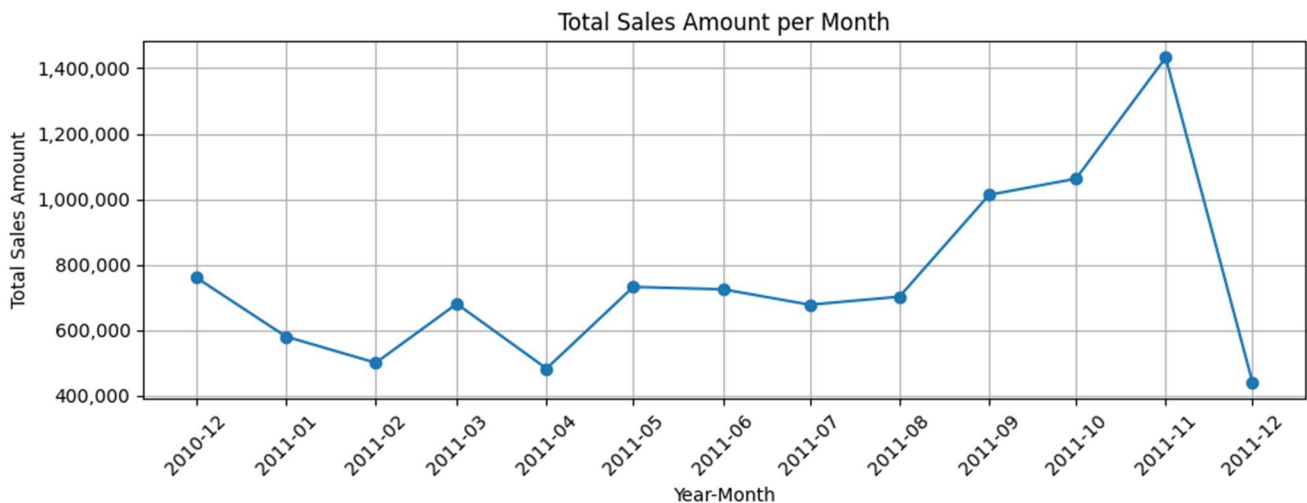
The analysis shows that the United Kingdom overwhelmingly dominates total revenue, generating more than 8 million in sales - far exceeding every other country. When the UK is removed from the comparison, the Netherlands and EIRE emerge as the next strongest performers, each contributing around 250,000, which highlights the sharp drop-off in revenue relative to the UK. This contrast underscores the UK's central role in overall sales performance and the comparatively modest contributions of the remaining top countries.
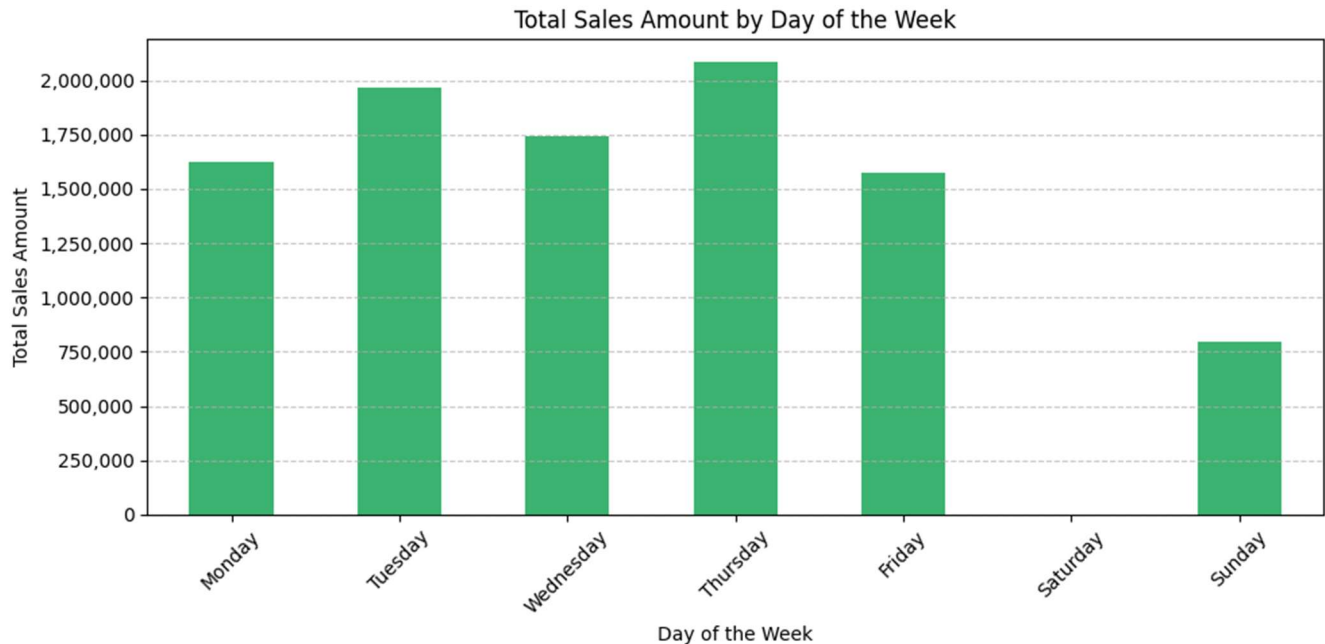
Top 10 Countries by Total Revenue of Orders (Including UK)

Top 10 Countries by Total Revenue of Orders (Excluding UK)

The analysis of the top 10 products by total revenue shows that **"REGENCY CAKESTAND 3 TIER"** is the clear standout, generating nearly 160,000 and significantly outperforming every other product. Several other strong performers follow, including **"WHITE HANGING HEART T-LIGHT HOLDER"** and **"PARTY BUNTING,"** each exceeding 100,000 in revenue. The sharp decline in revenue after the leading product highlights its exceptional popularity or pricing advantage compared to the rest of the top 10.

Top 10 Products by Total Revenue Ordered

The monthly revenue analysis shows stable sales from December 2010 to August 2011, followed by a sharp spike in November 2011, reaching about 1,400,000. This surge is immediately followed by a steep drop in December, suggesting the peak was driven by early holiday purchasing - likely influenced by promotions or stock concerns—before demand fell back to normal levels.

Total Sales Amount per Month

The analysis of sales by day of the week reveals a clear weekday-driven pattern. Thursday is the strongest performer, with total sales reaching nearly 2,000,000, while Sunday shows the lowest activity at just over 750,000. Sales remain consistently high from Monday through Friday but drop sharply over the weekend. Notably, the dataset contains no transactions at all on Saturdays, resulting in a zero value on the chart. This absence may indicate that the business does not operate on Saturdays or that sales for that day are not being recorded due to a data or logging issue.

Total Sales Amount by Day of the Week

5. Act

Here are **five actionable steps** the store could take based on the sales-pattern findings:

1. Strengthen Thursday Promotions and Product Launches

Since Thursday is the highest-revenue day, the store can schedule major promotions, new product releases, and marketing campaigns on Thursdays to maximize impact.

2. Improve Weekend Engagement—Especially Sunday

With sales dropping sharply on weekends, the store could introduce weekend-only discounts, bundles, or free-shipping offers to boost activity on low-performing days like Sunday.

3. Investigate and Address the Saturday Data Gap

Because no Saturday transactions appear in the dataset, the store should verify whether:

- the business is unintentionally missing sales opportunities, or

- there is a technical issue preventing Saturday sales from being recorded. Fixing this could unlock an additional revenue day.

4. Plan Inventory and Staffing Around Weekday Peaks

Since Monday–Friday consistently outperform weekends, the store can optimize staffing, stock replenishment, and customer support availability during these high-traffic days.

5. Leverage Seasonal and Early-Purchase Behavior

The massive November spike suggests customers buy early for the holiday season. The store can capitalize on this by:

- launching holiday promotions earlier,

- offering pre-Christmas bundles in November, and

- ensuring high-demand items are well-stocked before the surge.