Datathon 2025

Adelin Ma [*], Kevin Seng [†], Oscar Song [‡]

April 2025

# Analysis Report in Retail: Accelerating the Sales of the Modecraft Ecommerce Store

[*] mayidan@uw.edu
[†] senkevkk@uw.edu
[‡] oscarsjs@uw.edu

# 1 Introduction

The following analysis concerns Modecraft, an e-commerce store offering a wide range of household items for a diverse global clientele. As consultants tasked with analyzing the data of the online retail store, we took a deep dive into their annual dataset to offer them a clear view of their business performance and to identify the best strategies for sustainable growth. Given the data of all transactions registered by this multinational firm from 1 December 2010 to 31 December 2011, we conducted a thorough process of data cleaning, data grouping, graphical plotting, and interpolation. This rigorous process forms the basis of our thesis about the company's overall sight. Therefore, in this document, you will find the statistics about the transactions of the Modecraft E-Commerce Store for the fiscal year 2011, how we reached those numbers with our own metrics, and what they mean for the business. Our analysis also points towards promising opportunities for future growth.

# 2 Methodology

Data processing and visualization tasks were executed in parallel using Python and Microsoft Excel. Transaction data were ingested into a pandas DataFrame, where boolean indexing removed non-positive quantities and text fields were normalized. Concurrently, the same CSV extracts were loaded into Excel workbooks, where filters and conditional formatting verified data quality and consistency before further analysis.

In both environments we computed aggregate metrics (total quantity and total revenue by product, country and period) using pandas' groupby, multi-indexing and pivot operations alongside Excel PivotTables and functions such as SUMIFS and AVERAGEIFS. Percentage-change calculations and interpolation for gap-filling leveraged NumPy arrays and pandas' built-in methods, with Excel formulas mirroring these computations for cross-validation.

Visual outputs were produced in Python via Matplotlib for figure and axis customization and Seaborn's catplot and relplot for statistical charts, while Excel's charting engine generated interactive dashboards for stakeholder review. This dual-platform approach ensured reproducibility, rigorous cross-checking and accessibility, allowing each analyst to work with the toolset most familiar to them.

# 3 Data Cleaning

Upon initial inspection of Modecraft's transaction log, we observed a number of entries whose 'Description' field indicated inventory adjustments rather than bona fide sales. Examples included values such as "thrown away," "lost" and "wet," all of which correspond to stock discrepancies or spoilage. Because our analysis is explicitly concerned with sales volume and revenue generation, these adjustment records were extraneous and would have distorted aggregate metrics.

All of the adjustment records lacked a valid 'CustomerID', suggesting they were never associated with a paying customer. In order to maintain consistency and avoid counting non-revenue transactions, we removed every row with a missing 'CustomerID'. This step also eliminated orphaned returns or system-generated entries that were outside the scope of our sales report.

In addition to the description-based and customer-ID filters, we verified that all remaining records had strictly positive 'Quantity' values. Any zero or negative quantities—representing returns or data errors were excluded, ensuring that our subsequent aggregations of unit counts and revenue would reflect only completed purchases.

Finally, in preparation for detailed analysis, we enriched the cleaned dataset with several derived columns. Specifically, we computed Revenue as Quantity times UnitPrice, extracted the Day of Week and Quarter from the invoice timestamp, and categorized each transaction by Time of Day (morning, afternoon, evening) and Weekday or Weekend. For clarity in reporting, we also propagated the Product Name into its own field.

Table 1: First 25 Rows of Cleaned and Modified Data

| InvoiceNo | StockCode | Description | Quantity | InvoiceDate | UnitPrice | CustomerID | Country | Revenue | DayOfWeek | TimeOfDay | WeekdayOrWeekend | Month | Quarter | ProductName | YearMonth | Hour |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 536365 | 84406B | CREAM CUPID HEARTS COAT HANGER | 8 | 2010-12-01 08:26:00 | 2.75 | 17850.0 | United Kingdom | 22.0 | Wednesday | Morning | Weekday | December | 4 | CREAM CUPID HEARTS COAT HANGER | 2010-12 | 8 |
| 536365 | 21730 | GLASS STAR FROSTED T-LIGHT HOLDER | 6 | 2010-12-01 08:26:00 | 4.25 | 17850.0 | United Kingdom | 25.5 | Wednesday | Morning | Weekday | December | 4 | GLASS STAR FROSTED T-LIGHT HOLDER | 2010-12 | 8 |
| 536365 | 84029G | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | 2010-12-01 08:26:00 | 3.39 | 17850.0 | United Kingdom | 20.34 | Wednesday | Morning | Weekday | December | 4 | KNITTED UNION FLAG HOT WATER BOTTLE | 2010-12 | 8 |
| 536365 | 84029E | RED WOOLLY HOTTIE WHITE HEART. | 6 | 2010-12-01 08:26:00 | 3.39 | 17850.0 | United Kingdom | 20.34 | Wednesday | Morning | Weekday | December | 4 | RED WOOLLY HOTTIE WHITE HEART. | 2010-12 | 8 |
| 536365 | 22752 | SET 7 BABUSHKA NESTING BOXES | 2 | 2010-12-01 08:26:00 | 7.65 | 17850.0 | United Kingdom | 15.3 | Wednesday | Morning | Weekday | December | 4 | SET 7 BABUSHKA NESTING BOXES | 2010-12 | 8 |
| 536365 | 85123A | WHITE HANGING HEART T-LIGHT HOLDER | 6 | 2010-12-01 08:26:00 | 2.55 | 17850.0 | United Kingdom | 15.30 | Wednesday | Morning | Weekday | December | 4 | WHITE HANGING HEART T-LIGHT HOLDER | 2010-12 | 8 |
| 536365 | 71053 | WHITE METAL LANTERN | 6 | 2010-12-01 08:26:00 | 3.39 | 17850.0 | United Kingdom | 20.34 | Wednesday | Morning | Weekday | December | 4 | WHITE METAL LANTERN | 2010-12 | 8 |
| 536366 | 22632 | HAND WARMER RED POLKA DOT | 6 | 2010-12-01 08:28:00 | 1.85 | 17850.0 | United Kingdom | 11.10 | Wednesday | Morning | Weekday | December | 4 | HAND WARMER RED POLKA DOT | 2010-12 | 8 |
| 536366 | 22633 | HAND WARMER UNION JACK | 6 | 2010-12-01 08:28:00 | 1.85 | 17850.0 | United Kingdom | 11.10 | Wednesday | Morning | Weekday | December | 4 | HAND WARMER UNION JACK | 2010-12 | 8 |
| 536367 | 84879 | ASSORTED COLOUR BIRD ORNAMENT | 32 | 2010-12-01 08:34:00 | 1.69 | 13047.0 | United Kingdom | 54.08 | Wednesday | Morning | Weekday | December | 4 | ASSORTED COLOUR BIRD ORNAMENT | 2010-12 | 8 |
| 536367 | 84969 | BOX OF 6 ASSORTED COLOUR TEASPOONS | 6 | 2010-12-01 08:34:00 | 4.25 | 13047.0 | United Kingdom | 25.5 | Wednesday | Morning | Weekday | December | 4 | BOX OF 6 ASSORTED COLOUR TEASPOONS | 2010-12 | 8 |
| 536367 | 22622 | BOX OF VINTAGE ALPHABET BLOCKS | 2 | 2010-12-01 08:34:00 | 9.95 | 13047.0 | United Kingdom | 19.9 | Wednesday | Morning | Weekday | December | 4 | BOX OF VINTAGE ALPHABET BLOCKS | 2010-12 | 8 |
| 536367 | 22623 | BOX OF VINTAGE JIGSAW BLOCKS | 3 | 2010-12-01 08:34:00 | 4.95 | 13047.0 | United Kingdom | 14.85 | Wednesday | Morning | Weekday | December | 4 | BOX OF VINTAGE JIGSAW BLOCKS | 2010-12 | 8 |
| 536367 | 48187 | DOORMAT NEW ENGLAND | 4 | 2010-12-01 08:34:00 | 7.95 | 13047.0 | United Kingdom | 31.8 | Wednesday | Morning | Weekday | December | 4 | DOORMAT NEW ENGLAND | 2010-12 | 8 |
| 536367 | 22749 | FELTCRAFT PRINCESS CHARLOTTE DOLL | 8 | 2010-12-01 08:34:00 | 3.75 | 13047.0 | United Kingdom | 30.0 | Wednesday | Morning | Weekday | December | 4 | FELTCRAFT PRINCESS CHARLOTTE DOLL | 2010-12 | 8 |
| 536367 | 21754 | HOME BUILDING BLOCK WORD | 3 | 2010-12-01 08:34:00 | 5.95 | 13047.0 | United Kingdom | 17.85 | Wednesday | Morning | Weekday | December | 4 | HOME BUILDING BLOCK WORD | 2010-12 | 8 |
| 536367 | 22310 | IVORY KNITTED MUG COSY | 6 | 2010-12-01 08:34:00 | 1.65 | 13047.0 | United Kingdom | 9.90 | Wednesday | Morning | Weekday | December | 4 | IVORY KNITTED MUG COSY | 2010-12 | 8 |
| 536367 | 21755 | LOVE BUILDING BLOCK WORD | 3 | 2010-12-01 08:34:00 | 5.95 | 13047.0 | United Kingdom | 17.85 | Wednesday | Morning | Weekday | December | 4 | LOVE BUILDING BLOCK WORD | 2010-12 | 8 |
| 536367 | 22745 | POPPY'S PLAYHOUSE BEDROOM | 6 | 2010-12-01 08:34:00 | 2.1 | 13047.0 | United Kingdom | 12.60 | Wednesday | Morning | Weekday | December | 4 | POPPY'S PLAYHOUSE BEDROOM | 2010-12 | 8 |
| 536367 | 22748 | POPPY'S PLAYHOUSE KITCHEN | 6 | 2010-12-01 08:34:00 | 2.1 | 13047.0 | United Kingdom | 12.60 | Wednesday | Morning | Weekday | December | 4 | POPPY'S PLAYHOUSE KITCHEN | 2010-12 | 8 |
| 536367 | 21777 | RECIPE BOX WITH METAL HEART | 4 | 2010-12-01 08:34:00 | 7.95 | 13047.0 | United Kingdom | 31.8 | Wednesday | Morning | Weekday | December | 4 | RECIPE BOX WITH METAL HEART | 2010-12 | 8 |
| 536368 | 22914 | BLUE COAT RACK PARIS FASHION | 3 | 2010-12-01 08:34:00 | 4.95 | 13047.0 | United Kingdom | 14.85 | Wednesday | Morning | Weekday | December | 4 | BLUE COAT RACK PARIS FASHION | 2010-12 | 8 |
| 536368 | 22960 | JAM MAKING SET WITH JARS | 6 | 2010-12-01 08:34:00 | 4.25 | 13047.0 | United Kingdom | 25.5 | Wednesday | Morning | Weekday | December | 4 | JAM MAKING SET WITH JARS | 2010-12 | 8 |
| 536368 | 22913 | RED COAT RACK PARIS FASHION | 3 | 2010-12-01 08:34:00 | 4.95 | 13047.0 | United Kingdom | 14.85 | Wednesday | Morning | Weekday | December | 4 | RED COAT RACK PARIS FASHION | 2010-12 | 8 |

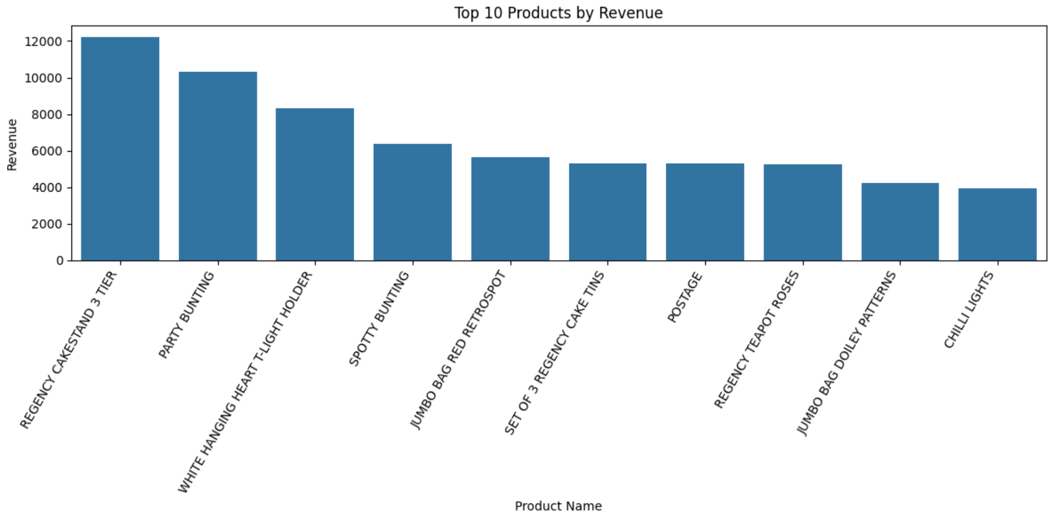
The outcome of these cleaning operations was a pared-down dataset comprising exclusively of valid sales transactions, each tied to a customer and contributing positively to the company's top-line metrics.

# 4 Monthly Performance

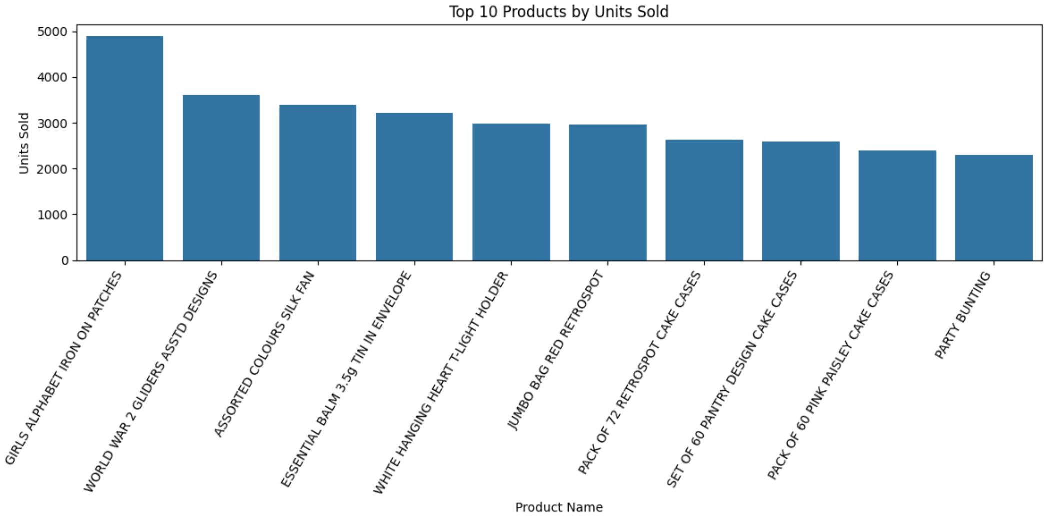
In order to capture short-term dynamics and identify actionable opportunities, we developed a Monthly Analysis Report for any selected month and year on Microsoft Excel. This report surfaces both volume and value-based metrics, such as the top and bottom selling products by units and by revenue, and customer and temporal patterns that drive performance. By examining a single month in isolation, stakeholders can assess the immediate impact of promotions, price changes and operational adjustments before they become obscured in year-to-date aggregates.

Prior to analysis, we applied the same data-cleaning rules described earlier. In order to exemplify the effectiveness of our model, we selected July 2011 as our exemplar period because it sits squarely in the middle of the fiscal year and typically delivers strong, varied transaction volumes, capturing both summer driven peaks and baseline ordering behavior. This makes it an ideal month to demonstrate our SQL reporting templates on a dataset that is large enough to reflect meaningful patterns, yet not skewed by year-end or launch anomalies.

The Monthly Product Sales report groups by product, year and month, then computes total units sold and total revenue. This dual aggregation quickly surfaces both high-volume and high-value SKUs.




The first chart focuses on revenue generation, presenting a different set of leading products. The "REGENCY CAKESTAND 3 TIER" stands out as the highest revenue generator, bringing in approximately \$12,000, despite not appearing on the top units sold list. "PARTY BUNTING" is second in revenue ( \$10,000+), showing strong performance in both volume and value. Decorative items like the "WHITE HANGING HEART T-LIGHT HOLDER" and other bunting variations also rank highly in revenue. This chart highlights products that, while potentially selling in lower quantities than the unit leaders, command higher prices, thus contributing more significantly to the overall revenue.

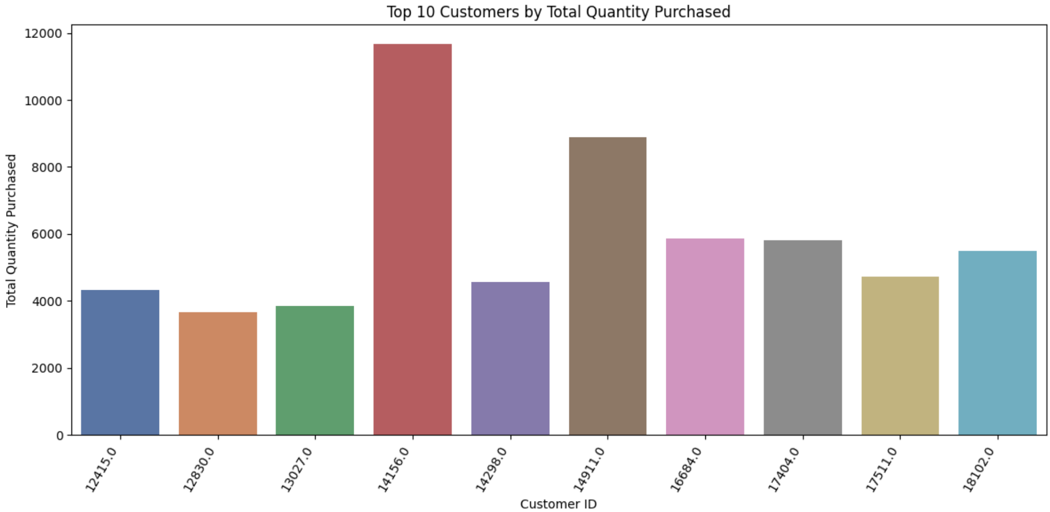



Conversely, the second chart reveals the top 10 products based on the quantity sold. Leading this list significantly are "GIRLS ALPHABET IRON ON PATCHES," with nearly 5,000 units sold. Following this are items like "WORLD WAR 2 GLIDERS ASSTD DESIGNS," "ASSORTED COLOURS SILK FAN," and "ESSENTIAL BALM 3.5G TIN IN ENVELOPE," each selling over 3,000 units. The list primarily features smaller, potentially lower-priced items like patches, gliders, fans, and small accessories. Even the 10th product, "PARTY BUNTING," achieved over 2,000 units sold, indicating substantial volume movement for these specific products.

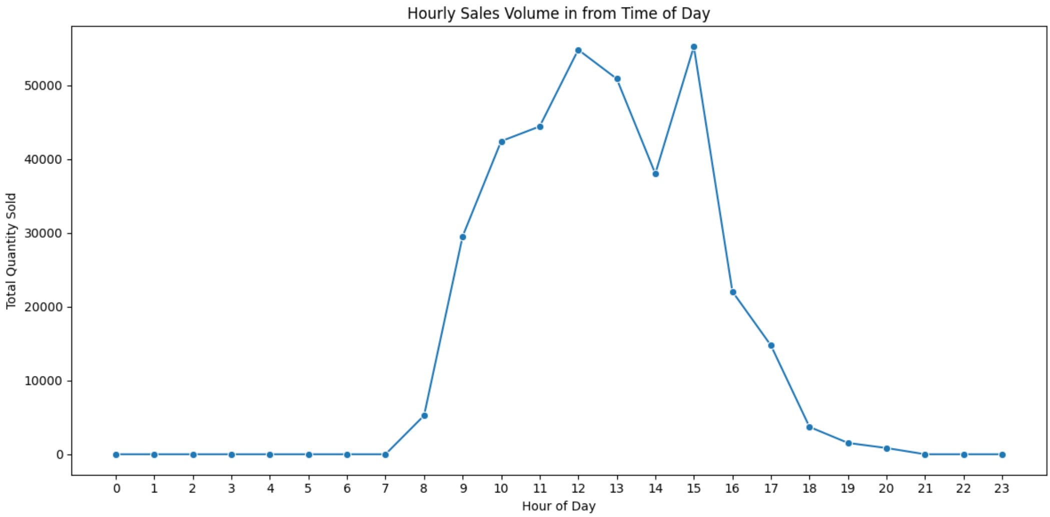
Comparing these two perspectives offers valuable insights. It's clear the business relies on a mix of high-volume, likely lower-margin items (like patches and gliders) and lower-volume, higher-margin products (like the cake stand). Products appearing on both lists, such as "PARTY BUNTING," "WHITE HANGING HEART T-LIGHT HOLDER," and "JUMBO BAG RED RETROSPOT," represent key items that balance popularity with significant revenue contribution. This dual approach seems successful, but understanding the profit margins on the high-volume items is crucial. Future expectations suggest these trends may continue, emphasizing the need to manage inventory effectively for both high-unit movers and top-revenue generators, while potentially exploring marketing strategies to upsell customers from lower-priced items or bundle products effectively.

As for the products that sold the least and generated the least revenue, the quantity and revenue will simply be 0. After counting the total number of unique products (3885) and the total number of unique products that were sold in July 2011 (2370), we see that there were 1515 products that were not sold at all in that month.

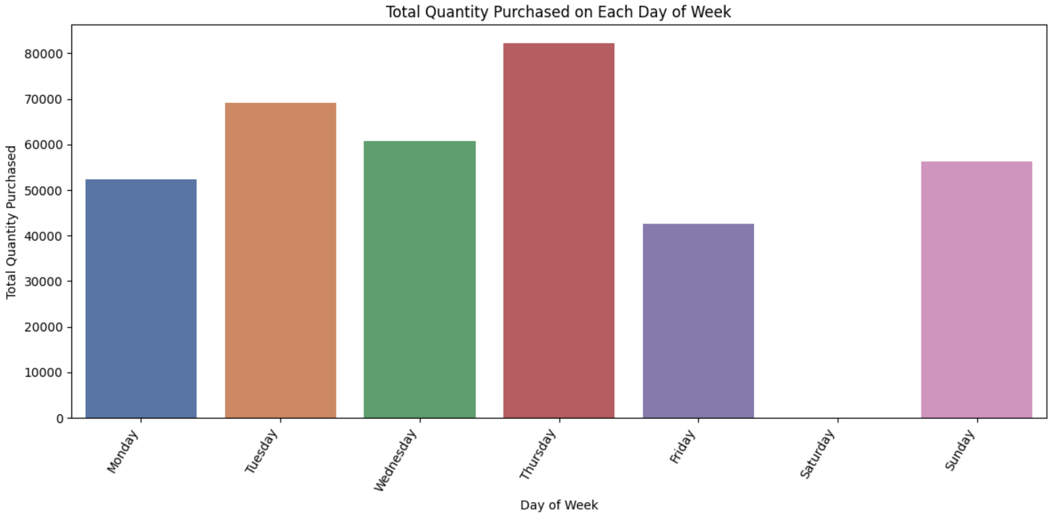
For customer analysis, we group by CustomerID (after discarding nulls) and sum quantity to rank buyers by purchase volume. The resulting list identifies the accounts that generate the largest order quantities. We have summarised the top 10 customers in the diagram below.




Temporal insights are derived by bucketing invoice timestamps into time-of-day segments and day-of-week labels. Summing revenue across these dimensions highlights the most lucrative time slots and markets.




The first graph illustrates hourly sales volume throughout a typical day. It reveals a distinct pattern where purchasing activity is minimal during the late night and early morning hours (approximately 00:00 to 07:00). Sales begin a steep climb around 8:00 AM, reaching a peak period between 10:00 AM and 3:00 PM (15:00), with the absolute highest point occurring around 3 PM. Following this peak, sales decline sharply into the evening, becoming negligible again after 9:00 PM (21:00). This indicates that the core business hours for customer purchasing are concentrated in the late morning and afternoon.

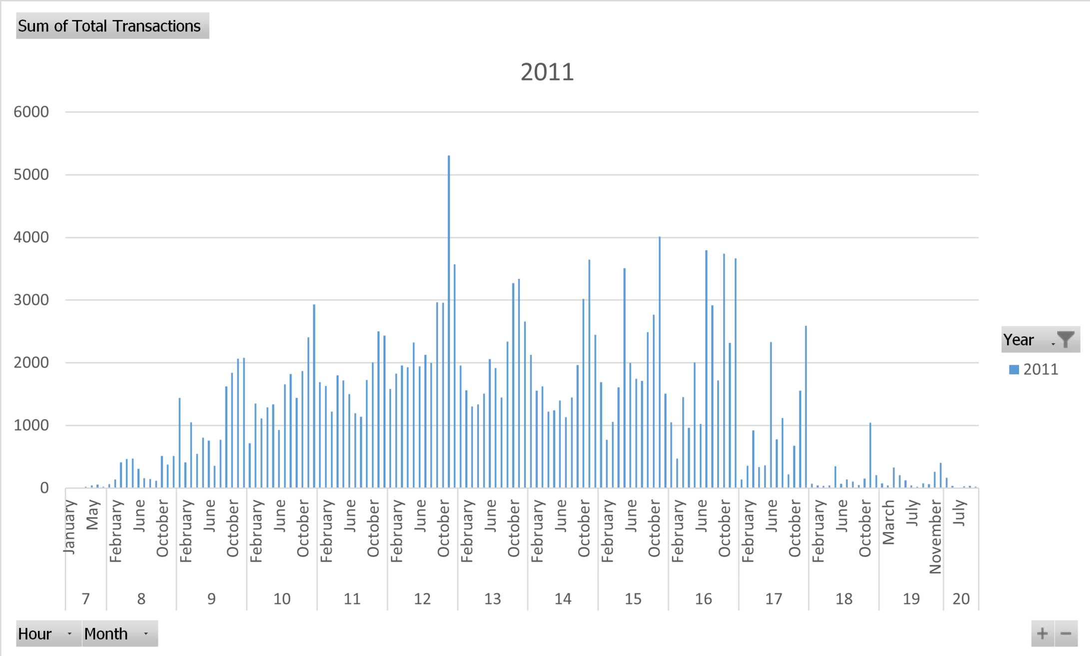



The second graph shows the total quantity purchased across different days of the week. Thursday stands out with the highest volume, exceeding 80,000 units. Tuesday and Wednesday also demonstrate strong sales activity. Monday and Sunday show moderate volumes, while Friday has notably lower sales, and Saturday's volume appears extremely low, potentially indicating that the store is not open on Saturdays, and no transactions occurred.

Combining these insights suggests that the company's peak transaction period occurs during midday on weekdays, particularly mid-week, with Thursday being the most active day. The concentration of sales between 10:00 AM and 3:00 PM implies that customers are most active during standard daytime hours. The unusually low volume recorded for Saturday requires investigation to confirm data accuracy and understand the underlying cause, as weekends are often strong periods for e-commerce. Based on the general trend (excluding the Saturday anomaly), staffing for order processing and customer support should be prioritized during the late morning to mid-afternoon hours on weekdays, especially Tuesday through Thursday, to handle the peak load effectively. Marketing initiatives might yield better results if timed to align with these established high-activity periods.

Each SQL template accepts year and month (or any date range) as parameters, providing a flexible, repeatable framework that can generate up-to-date product, customer, temporal and regional reports without altering the core query logic.

# 5 Consumer Insights




The graph illustrates a clear interplay between the time of day and the volume of transactions across different months. We can observe that the overall pattern of transactions increasing towards midday and then declining in the late afternoon holds true for most months displayed. However, the magnitude of these hourly fluctuations varies significantly depending on the specific month. For instance, months like January and May show a more pronounced surge in transactions during peak hours (around 12:00 to 16:00) compared to months like February or October, which exhibit a more moderate level of activity throughout the day.

This suggests that the timing of transactions is not uniform throughout the year, and that certain months experience more concentrated periods of activity. This could be attributed to various factors, such as seasonal events, pay cycles, or specific industry trends that influence transaction patterns. For example, if the graph represents retail data, January might show high activity due to post-holiday sales, while May could be influenced by spring-related shopping. Understanding these monthly variations in hourly transaction patterns is crucial for resource allocation and operational planning.

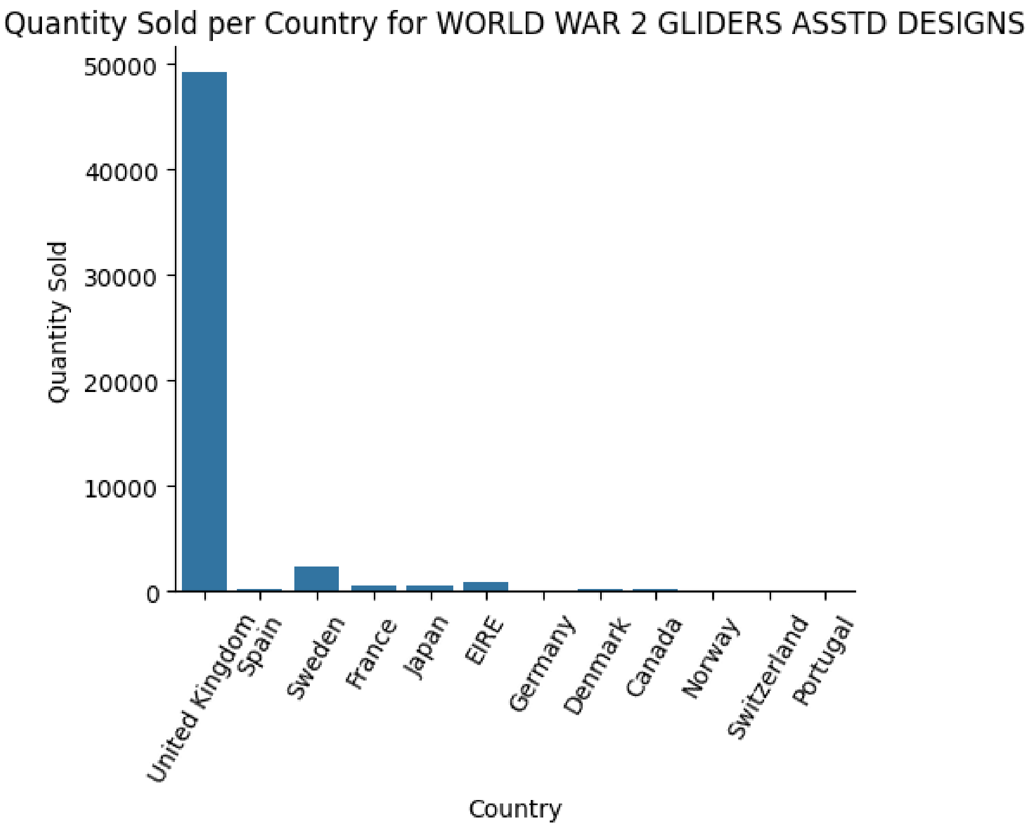
In essence, the graph reveals a dynamic relationship between hourly transaction volume and the time of year. While the general trend of increased activity around midday persists across most months, the intensity of this trend differs. This highlights the importance of considering both the time of day and the specific month when analyzing transaction data, as it provides a more nuanced understanding of the underlying patterns and drivers of activity.

# 6 Item Analysis

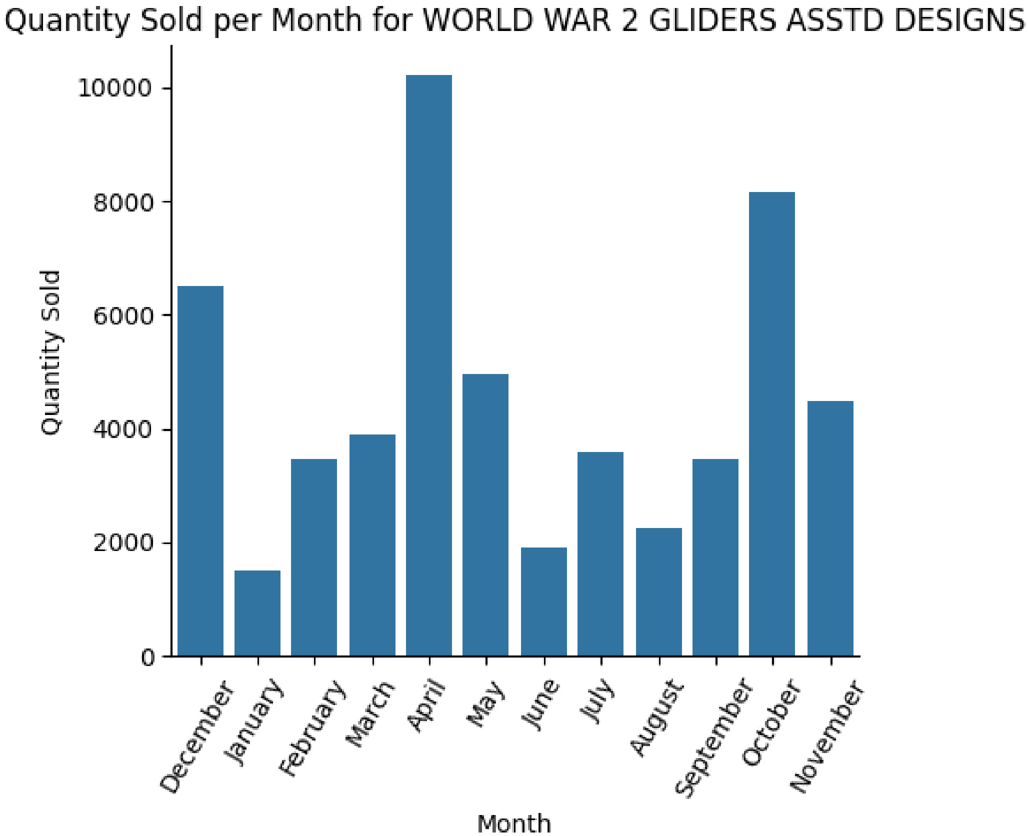
Item level analysis constitutes a cornerstone of our sales intelligence framework because it enables precise identification of high-impact products and supports data-driven decision making across inventory planning, marketing investment and promotional design. By disaggregating aggregate sales figures into product-specific performance indicators, we obtain clarity on which items drive revenue and volume, and we uncover patterns that might be obscured when viewing data only at the category or total revenue level. This granular perspective reduces risk by highlighting underperforming products early and by validating assumptions about product appeal in different markets and time periods.

To support actionable insights we developed a standardized reporting framework for any given product using Python and its libraries. This framework comprises six core metrics Quantity Sold by Country, Quantity Sold by Month, Quantity Sold by Time of Day, Quantity Sold by Day of Week, Total Quantity Sold and Total Revenue Generated. By encapsulating these dimensions in a single report template we ensure consistency, comparability and transparency across the product portfolio.

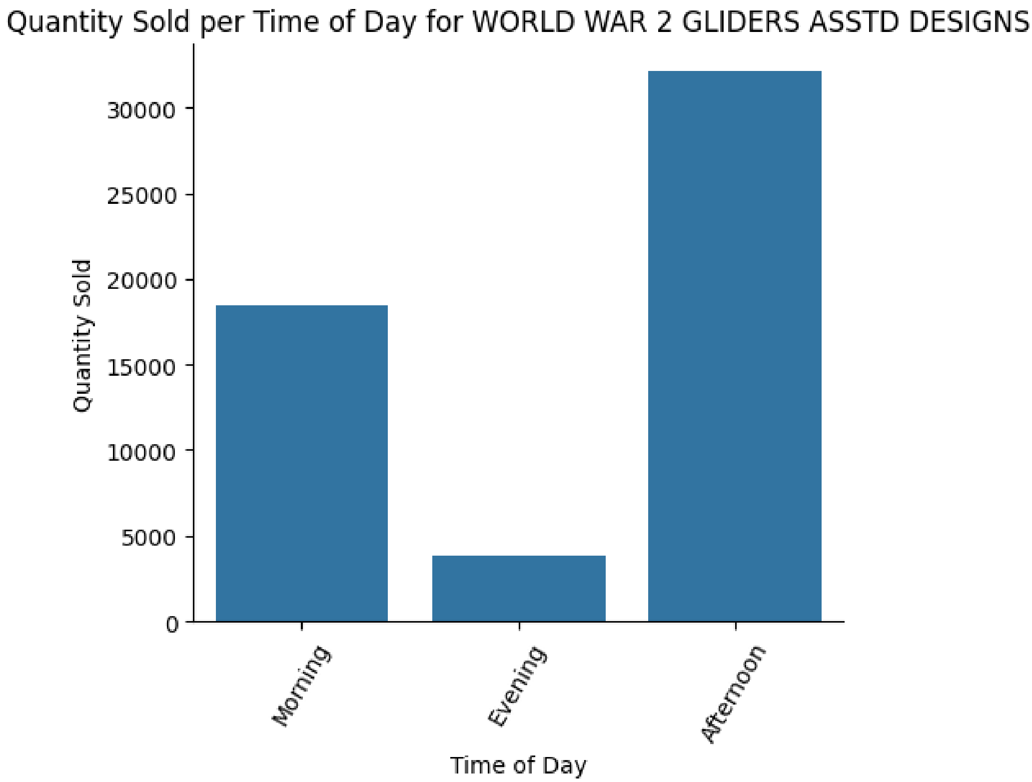
WORLD WAR 2 GLIDERS ASSTD DESIGNS was selected for the initial demonstration of this reporting framework because it recorded the highest cumulative units sold (53215) among all items in our dataset. The Total Quantity of WORLD WAR 2 GLIDERS ASSTD DESIGNS sold is 54415. The Total Revenue from selling WORLD WAR 2 GLIDERS ASSTD DESIGNS is 13586.25. Its prominence in sales volume makes it a natural candidate to validate the utility of the report template. The sales performance of WORLD WAR 2 GLIDERS ASSTD DESIGNS was examined across multiple dimensions to uncover the product's core markets and temporal demand patterns.



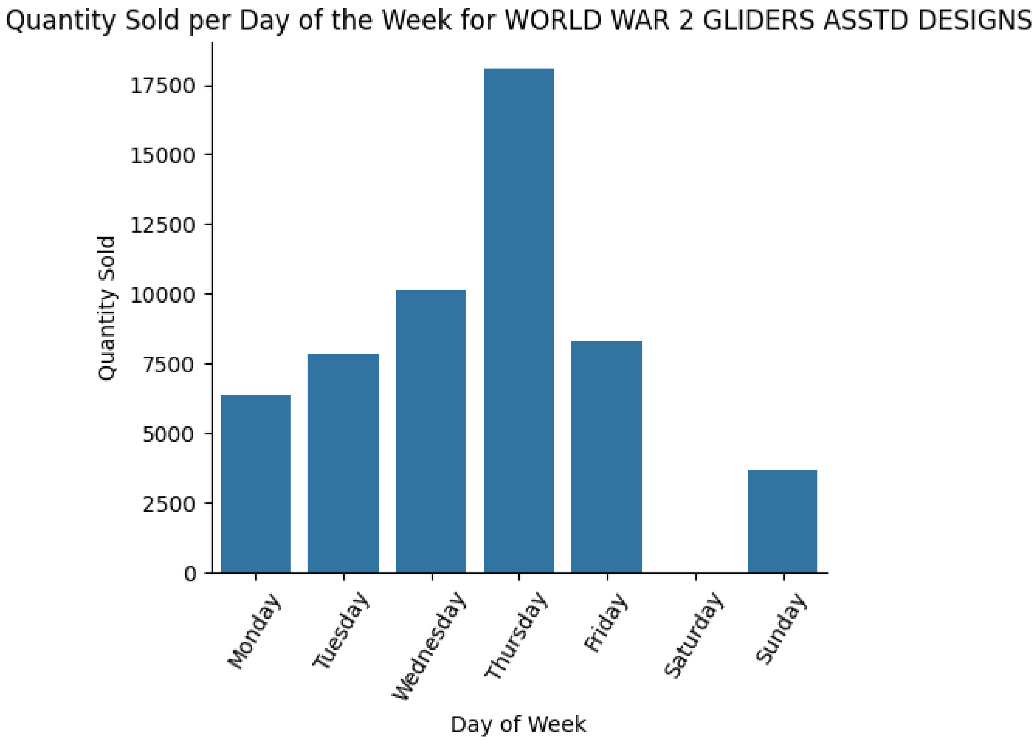

Geographically, the United Kingdom accounts for 47982 sales of the 53215 units sold. Secondary markets combined contribute less than 10 percent of global sales, underscoring the UK as the principal focus for inventory planning, promotional investment, and distribution efforts for this product line.




Seasonal analysis reveals that April represents the highest volume month, with unit sales significantly exceeding those in any other period. October follows as the second-strongest month, driven perhaps by pre-holiday interest or specific marketing campaigns. December ranks third, suggesting that year-end gifting or inventory clearance activities also bolster demand. Off-peak months fall well below these levels, indicating strong seasonality that should inform production scheduling and cash-flow forecasting.




Within each calendar day, purchasing behavior for this item is skewed toward the afternoon window. These intraday patterns suggest that customer engagement initiatives, such as time-targeted email or paid search campaigns, may achieve higher ROI when scheduled for the mid-day period.




Analysis by day of week shows a steady ascent from Monday through Thursday. Monday volume begins at 6360 thousand units, rising each day to reach a weekly maximum of 17963 units on Thursday. Sales then taper off toward the weekend. There are no sales made on Saturdays.

This structured reporting approach creates a repeatable playbook for product evaluation and performance monitoring. By standardizing the six key metrics in a single template and validating the approach on our highest volume item we establish a clear methodology that can be extended to every product. Such rigor supports more accurate forecasting, more efficient inventory planning and more targeted marketing investments. Ultimately this granular insight fosters a disciplined, evidence based decision making culture throughout the organization.

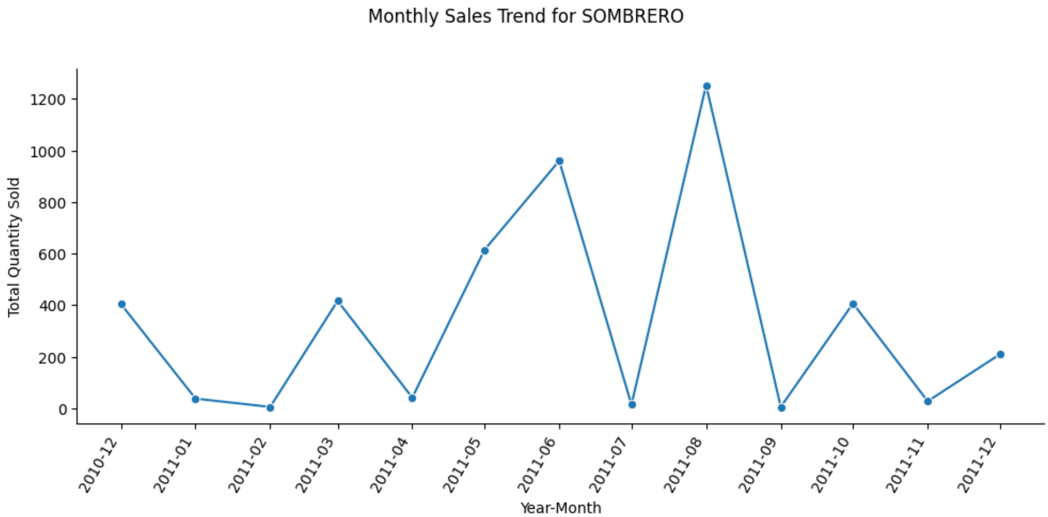
# 7 Significant Fluctuations

Our analysis of monthly sales trends naturally led to an investigation of which products exhibit unusually large swings in demand. Significant month-to-month variations can indicate emerging market opportunities, seasonal promotions, supply chain disruptions or data irregularities. By identifying the items that fluctuate most dramatically, we gain the ability to flag potential issues early and to allocate marketing or inventory resources where they will have the greatest impact.

## 7.1 Frequent Fluctuations

To quantify these fluctuations we first aggregated unit sales for each product by distinct year-month periods. We then calculated the percentage change from one month to the next for every product. A tolerance threshold of 1000% was applied to isolate only those month-over-month increases or decreases that exceeded this magnitude. This threshold remains fully adjustable, allowing the analysis to be tailored to more conservative or more aggressive definitions of "significant" change depending on business priorities.

When this methodology was applied across the entire product catalogue, the item with the greatest number of drastic monthly shifts was SOMBRERO, which registered five separate periods of change exceeding our 1000% threshold. This finding highlights SOMBRERO not only as a high-volatility SKU but also as a potential bellwether for broader market or operational dynamics affecting the range.




Inspecting SOMBRERO's month-to-month trajectory reveals a pronounced rise in sales from April through June, followed by a sharp decline in July. August then saw another substantial surge, before volume retreated again in September. The precise drivers of these swings remain unclear. Possible explanations include time-limited promotions, shipment timing, restocking delays or shifts in customer interest. Further investigation into pricing, campaign calendars and supply schedules will be required to determine the underlying causes.

## 7.2 Largest Increment

An equally important dimension of our sales volatility analysis was to identify which single product experienced the largest absolute shift in units sold from one month to the next. While our percentage-change methodology highlighted multiple dramatic swings across the portfolio, pinpointing the single largest month-to-month movement provides a clear signal for where operational or marketing intervention may be most urgently required.


The product that emerged at the top of this ranking was PLEASE ONE PERSON METAL SIGN, which exhibited the single largest month-to-month swing in our dataset. A closer look at its monthly trajectory reveals a persistent decline in sales beginning in March and continuing through August, suggesting waning demand or possible stock constraints. In September, however, the item's sales rebounded sharply, registering an increase that exceeded any other month-to-month jump across the entire catalogue.

Understanding the drivers behind this dramatic reversal is critical. Potential explanations include the conclusion of a marketing promotion in late summer, the resolution of inventory shortages, or the introduction of complementary products that renewed customer attention. Investigating promotional calendars, supply-chain records, and competitive activity around the August–September boundary will be necessary to determine the root cause. By isolating this single most volatile SKU, we can prioritize our follow-up analysis and rapidly deploy targeted strategies to stabilize and optimize its performance.

# 8 Metric System

## 8.1 Rationale for Adjusted Product Health Metric

While traditional metrics like total units sold or total revenue provide useful snapshots of historical performance, they fail to capture the dynamic trajectory of a product over time. Static aggregates risk favoring items that were once popular but are now in decline, or overlooking emerging products that have recently gained momentum. Our Adjusted Product Health metric addresses these shortcomings by integrating unit sales momentum, revenue growth percentage, and time-weighted recency into a single comprehensive score.

Specifically, we prioritize products that exhibit strong increases in unit sales (Momentum) while maintaining stable or moderate revenue growth rates. Products undergoing extreme revenue surges or collapses are treated with caution, as such volatility may indicate unsustainable trends. The Health Ratio ($H$) for a given month is therefore computed as:

$$H = \frac{Q_{\text{current}} - Q_{\text{previous}}}{1 + \left| \frac{R_{\text{current}} - R_{\text{previous}}}{R_{\text{previous}}} \right|}$$

where: - $Q$ denotes quantity sold, and - $R$ denotes revenue.

Dividing Momentum by $1 + |\text{Growth\%}|$ ensures that steady, sustainable sales expansion is preferred over erratic spikes. This normalization penalizes products with excessive volatility, stabilizing the ranking system toward reliable performers.

## 8.2 Rolling Health Computation

To capture both short-term dynamics and long-term momentum, we developed a recursive model for Rolling Health. Let $H_{\text{current}}$ represent the current month's Health Ratio, $RH_{\text{previous}}$ represent the most recent previous Rolling Health Score, and $m$ represent the number of months elapsed since the previous transaction. The Rolling Health ($RH_{\text{current}}$) is calculated as:

$$RH_{\text{current}} = \begin{cases} \sqrt{|H_{\text{current}} + (0.9^m \times RH_{\text{previous}})|} & \text{if } H_{\text{current}} + (0.9^m \times RH_{\text{previous}}) \geq 0 \\ -\sqrt{|H_{\text{current}} + (0.9^m \times RH_{\text{previous}})|} & \text{otherwise} \end{cases}$$

This formulation combines the current performance with a decayed memory of past performance, applying a 10% monthly depreciation factor to prior health. The square root transformation moderates extreme swings, creating a more stable and interpretable health score trajectory over time.

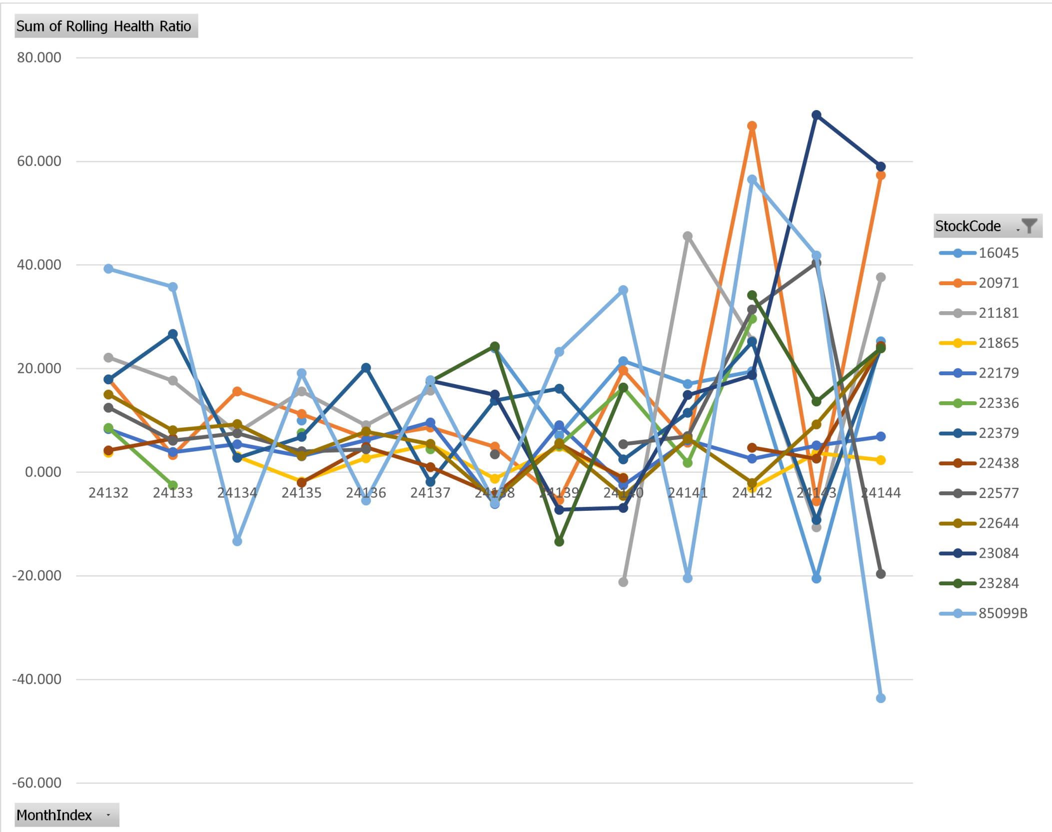
## 8.3 Adjusted Product Health Computation

After computing the Rolling Health up to the last observed sale for a product, we apply an additional adjustment to account for the passage of time since the product's last transaction. Specifically, if $M_{\text{latest}}$ denotes the maximum MonthIndex observed across the dataset, and $M_{\text{current}}$ denotes the last month the product recorded a sale, then the Adjusted Product Health $(APH)$ is calculated as:

$$APH = RH_{\text{current}} \times 0.9^{(M_{\text{latest}} - M_{\text{current}})}$$

This compounding decay ensures that products which have not sold in several months naturally decline in ranking, reflecting the diminishing business relevance of stale inventory. Recent activity is thus more heavily rewarded, while older momentum is appropriately penalized, providing a forward-looking framework for prioritizing products based on both current and emerging performance trends.

## 8.4 Sample Chart and Table



Rolling health ratios with respect to months. X-axis represent Month Index, starting at 24132 for December 2010, Y-axis for rolling health ratios, Different colored lines for different products per its stock code.

Table 2: First 25 Rows of Metric Time Tracking

| StockCode | Description | Adjusted Product Health | Date of Last Purchase | Last Purchase Quantity | Price |
|---|---|---|---|---|---|
| 22197 | SMALL POPCORN HOLDER | 67.183 | 24144 | 67.183 | December 2011 |
| 23084 | RABBIT NIGHT LIGHT | 59.020 | 24144 | 59.020 | December 2011 |
| 20971 | PINK BLUE FELT CRAFT TRINKET BOX | 57.318 | 24144 | 57.318 | December 2011 |
| 22273 | FELTCRAFT DOLL MOLLY | 44.649 | 24144 | 44.649 | December 2011 |
| 23582 | VINTAGE DOILY JUMBO BAG RED | 44.611 | 24144 | 44.611 | December 2011 |
| 22147 | FELTCRAFT BUTTERFLY HEARTS | 43.697 | 24144 | 43.697 | December 2011 |
| 21181 | PLEASE ONE PERSON METAL SIGN | 37.624 | 24144 | 37.624 | December 2011 |
| 22413 | METAL SIGN TAKE IT OR LEAVE IT | 37.609 | 24144 | 37.609 | December 2011 |
| 22178 | VICTORIAN GLASS HANGING T-LIGHT | 36.783 | 24144 | 36.783 | December 2011 |
| 23552 | BICYCLE PUNCTURE REPAIR KIT | 35.199 | 24144 | 35.199 | December 2011 |
| 21787 | RAIN PONCHO RETROSPOT | 34.300 | 24144 | 34.300 | December 2011 |
| 21137 | BLACK RECORD COVER FRAME | 33.965 | 24144 | 33.965 | December 2011 |
| 23498 | CLASSIC BICYCLE CLIPS | 33.541 | 24144 | 33.541 | December 2011 |
| 23307 | SET OF 60 PANTRY DESIGN CAKE CASES | 32.888 | 24144 | 32.888 | December 2011 |
| 85152 | HAND OVER THE CHOCOLATE SIGN | 32.367 | 24144 | 32.367 | December 2011 |
| 23497 | CLASSIC CHROME BICYCLE BELL | 31.561 | 24144 | 31.561 | December 2011 |
| 20668 | DISCO BALL CHRISTMAS DECORATION | 31.241 | 24144 | 31.241 | December 2011 |
| 22961 | JAM MAKING SET PRINTED | 29.324 | 24144 | 29.324 | December 2011 |
| 22386 | JUMBO BAG PINK POLKADOT | 28.822 | 24144 | 28.822 | December 2011 |
| 22579 | WOODEN TREE CHRISTMAS SCANDINAVIAN | 31.807 | 24143 | 28.627 | November 2011 |
| 21705 | BAG 500g SWIRLY MARBLES | 28.432 | 24144 | 28.432 | December 2011 |
| 20975 | 12 PENCILS SMALL TUBE RED RETROSPOT | 27.403 | 24144 | 27.403 | December 2011 |
| 23543 | KEEP CALM WALL ART | 27.401 | 24144 | 27.401 | December 2011 |

# 9 Conclusion

This report has presented a rigorous analysis of Modecraft's 2011 transaction dataset, leveraging both Python and Excel to ensure data integrity, reproducibility and accessibility.

Looking ahead, this dual approach, detailed profiling plus change detection, can be extended to customer segments, channel analyses and price elasticity studies. We recommend adopting a rolling-window cadence for these reports, integrating external data such as marketing campaigns and supply metrics, and automating key SQL templates to keep insights current. Together, these enhancements will equip Modecraft with a scalable, data-driven decision-support system that drives operational efficiency and sustainable growth.

# 10 Appendix

Link to our Github Repository: `https://github.com/Oscariano/Datathon2025.git`