Name: Pratama Alethea Agung Yodha

Student ID: 24072262

Dashboard

Link to dashboard: https://data-analytics-assignment-yodha.streamlit.app/
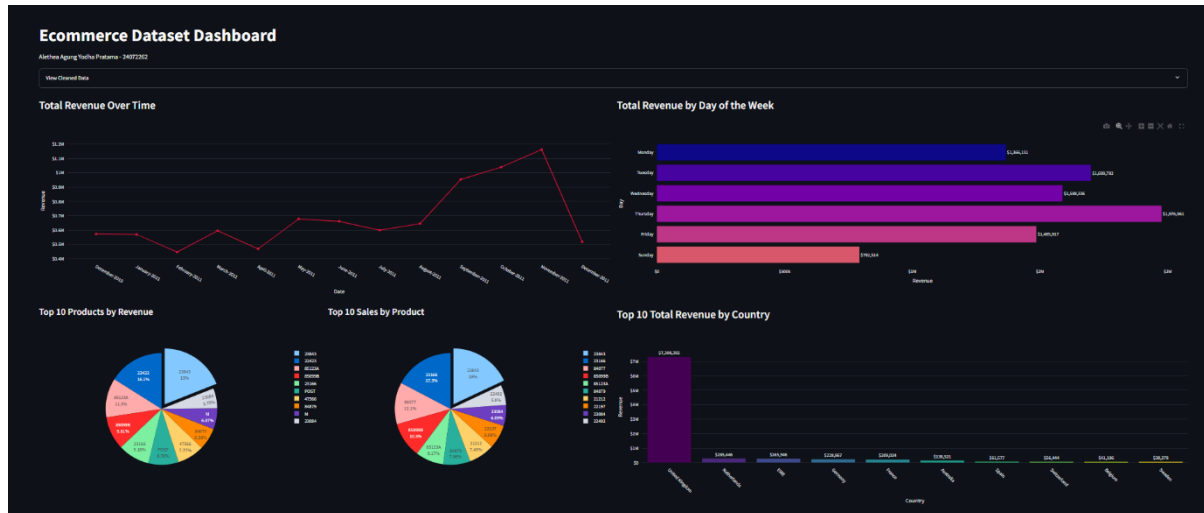
Link to github repo for the code: https://github.com/yodhasu/Data-Analytics/tree/main/Assignment%201

Report

1. **Introduction and Objectives**

   In today's hyper-competitive digital economy, e-commerce retailers must leverage data analytics to navigate rapidly shifting consumer behaviors, optimize operational efficiency, and identify new avenues for growth. This report examines an anonymized **Ecommerce Dataset**—sourced from Kaggle (lissetteg/ecommerce-dataset)—which contains approximately half a million transactions recorded from December 2010 through December 2011. Each transaction includes key attributes such as InvoiceDate, InvoiceNo, StockCode (our SKU identifier), Country, Quantity, and UnitPrice.

   The overarching aim of this analysis is twofold. First, we seek to **diagnose critical business challenges** by uncovering patterns in temporal performance (monthly and weekday sales), product-level concentration (both revenue- and volume-based), and geographic distribution. These insights will spotlight where the business is most vulnerable—be it seasonal troughs, weekend underperformance, over-reliance on a handful of SKUs, or dependence on a single market.

Second, we intend to **translate these diagnostic findings into actionable recommendations**. Grounded in data-driven reasoning, our objectives include:

1. **Characterizing seasonality and volatility** in monthly revenues, pinpointing periods of pronounced growth as well as problematic dips.
2. **Quantifying the weekday versus weekend revenue gap** to identify under-leveraged time windows for targeted interventions.
3. **Assessing product portfolio risk** by comparing the revenue share of top SKUs against their unit-sales profiles, thus revealing opportunities for diversification and bundling.
4. **Mapping international sales** to highlight geographic dependencies and uncover under-penetrated markets ripe for expansion.

By systematically addressing these four analytical pillars, we aim to equip the business with a clear problem statement and a set of data-backed strategies—ranging from promotional calendar adjustments and product-mix optimization to dynamic pricing and market localizations—that collectively can stabilize revenue, mitigate concentration risks, and unlock sustainable, long-term growth.

2. <u>Data Overview & Methodology</u>

- **Data Cleaning:**
  - I removed cancelled orders (identified by "C" in the InvoiceNo).
  - Filtered out zero or negative quantities and unit prices as likely returns or data errors.
  - Aggregated daily revenues (Quantity × UnitPrice) and then rolled up to monthly, weekday, SKU, and country levels.
- **Visualization Tools:**
  - Used Python's pandas for aggregation.
  - Created plots with Plotly in a Streamlit dashboard on Kaggle for interactive slicing and dicing.
  - Exported static charts for this report.
- **Analytical Approach:**
  - **Time Series**: Line chart of monthly totals to detect seasonality.
  - **Weekday Analysis**: Bar chart of summed revenue by weekday.
  - **Product Concentration**: Two pie charts—one ranked by revenue, one by units sold.
  - **Geo-Analysis**: Bar chart of total revenue by country.

Throughout, I reflect on the limitations of the dataset (e.g., no customer demographics) but also leverage its strengths—large transaction volume and diverse SKUs.
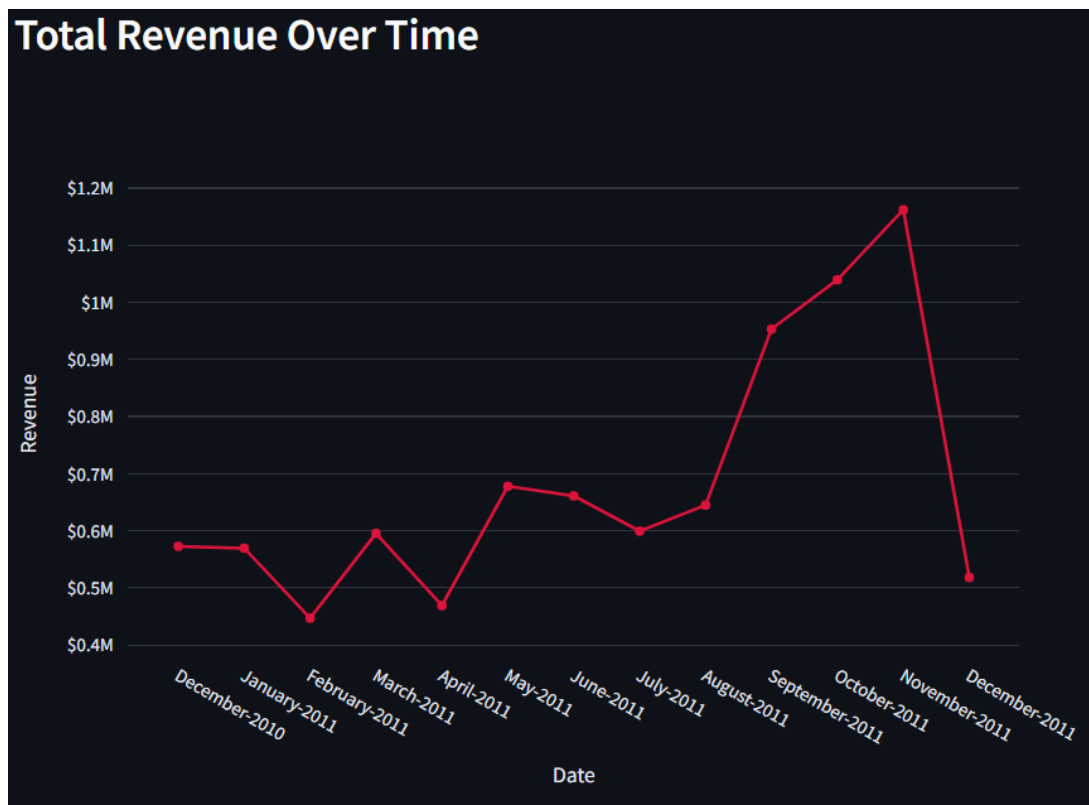
3. **Analysis of "Total Revenue Over Time"**

Total Revenue Over Time

Chart Description:
A line chart traces monthly revenue from December 2010 to December 2011.

Observations:

- Low Start, Mid-Year Plateau, Late Surge:
    - December 2010 begins near $575 K.
    - A trough in February 2011 (~$480 K) likely reflects post-holiday lulls.
    - May–July show a flatter band around $620 K–$650 K.
    - A sharp climb in August–November peaks above $1.1 M.
- Unusual December Collapse:
  December 2011 unexpectedly dips back near $500 K, despite typical holiday boost.

I suspect the August–November uptick corresponds to a targeted marketing push—perhaps back-to-school promotions or new product launches mid-summer. The dramatic November peak may align with Black Friday/Cyber Monday events. As a student, I find it fascinating that the business—like many real-world ventures—rides the waves of seasonality. Yet the December collapse puzzles me: I would expect holiday sales to carry momentum through year-end. Potential explanations include:

1. Stock-Outs or Supply Issues: If inventory ran low on hot sellers, shoppers may have looked elsewhere.
2. Technical Downtime: A site outage or payment gateway failure over the crucial holiday shopping weekend.

3. Data Gaps: Maybe some invoices weren't captured in the dataset for December 2011.

If I had access to qualitative notes or server-log data, I'd merge them with transaction logs to confirm.

Key Business Implications:

- **Revenue Volatility:** Difficult to forecast cash flow and staffing needs when month-to-month swings exceed 70 percent.
- **Campaign Timing:** Over-reliance on late-year promotions could mask growth opportunities earlier in the calendar.
- **Operational Gaps:** The December anomaly is a red flag for possible logistical or technical failures.

As a recommendation, I'd propose investigating root causes of the December drop and running pilot spring campaigns to smooth the early-year trough. Ensuring robust infrastructure during peak season is also critical.

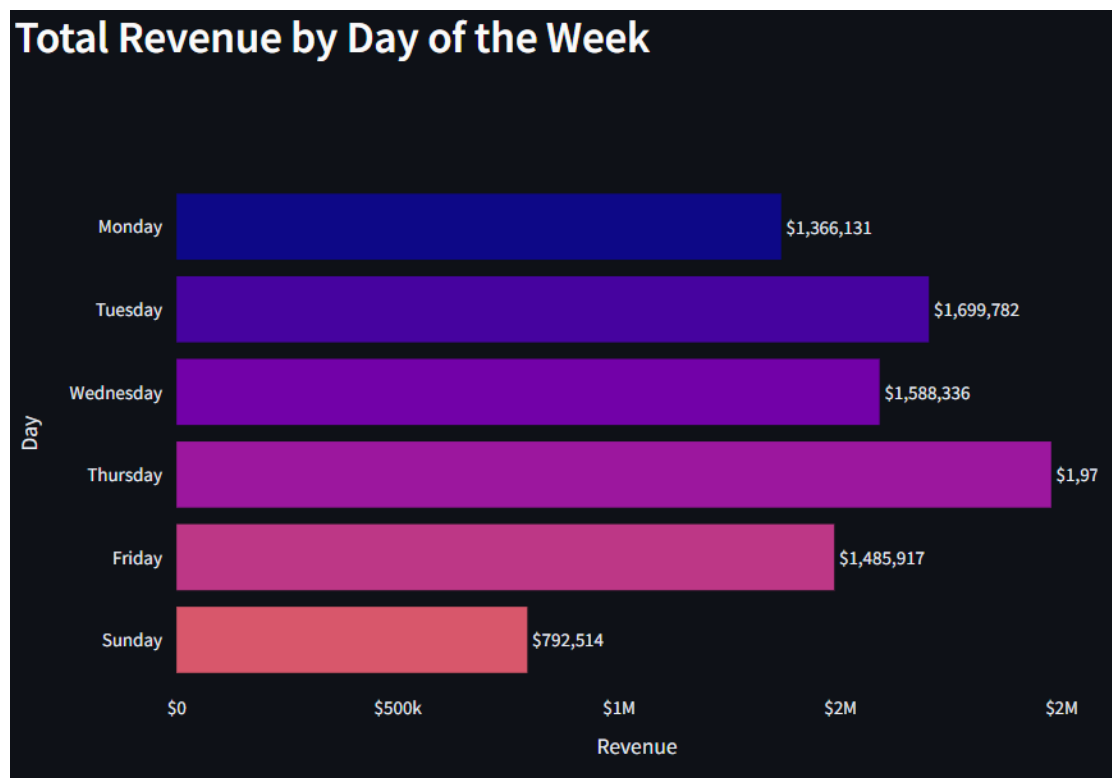4. **Analysis of "Total Revenue by Day of the Week"**

Chart Description:
A horizontal bar chart sums all transactions by weekday (Monday–Sunday) across the entire period.

Observations:

- **Thursday Highest:** ~$1.98 M.

- **Tuesday & Wednesday Strong:** ~$1.70 M and $1.59 M respectively.
- **Monday & Friday Moderate:** ~$1.36 M and $1.49 M.
- **Saturday Solid:** ~$1.48 M.
- **Sunday Slumps:** Only ~$792 K—less than half of Thursday's revenue.

Interpretation & Student-Style Reflection:
This pattern surprised me at first: I'd assumed weekends would outperform weekdays, reflecting casual browsing. Instead, Monday–Friday dominate. Possible reasons:

- **Business Customers:** If many transactions come from corporate or B2B buyers, they naturally operate in a weekday rhythm.
- **Weekend Friction:** Customer service, live chat, and shipping operations may be reduced on Sundays, deterring last-minute purchases.
- **Promotional Cadence:** Marketing emails or ad campaigns might drop early in the week, fading by Sunday.

I might further test whether average order value (AOV) differs by weekday or whether email-open rates drive the weekday spike. If Sunday conversion rates lag because fewer campaigns land in inboxes, adjusting send schedules could help.

Key Business Implications:

1. **Untapped Weekend Opportunity:** Sunday is nearly a blank canvas—there's room for **≥50 percent** revenue gain.
2. **Operational Alignment:** If we can staff chatbots or customer support on weekends, shopper confidence may rise.
3. **Campaign Extension:** Rolling out "weekend-only" deals or loyalty point bonuses could stimulate off-peak buying.

I'd propose A/B tests on weekend-specific incentives—e.g., free shipping for Sunday orders—and measure lift in orders and revenue. Additionally, surveying customers about weekend pain points could uncover actionable insights.

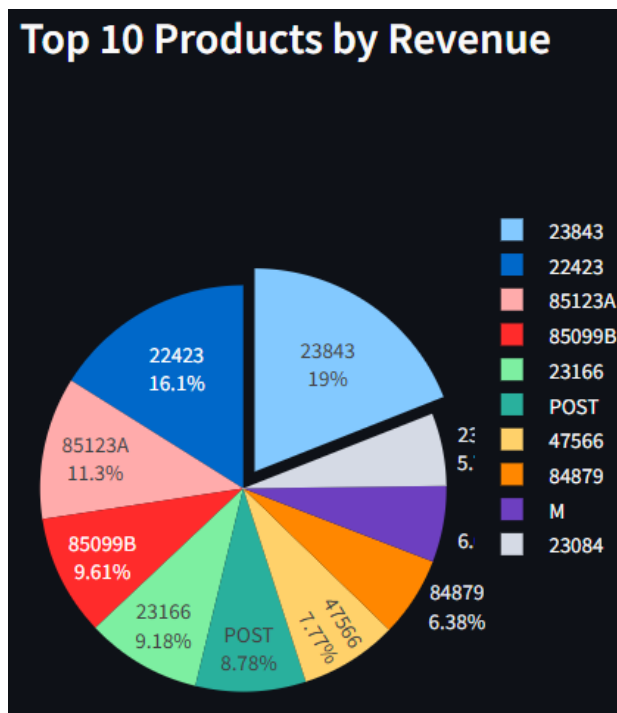## 5. Analysis of "Top 10 Products by Revenue"

Chart Description:
A pie chart displays each of the ten highest-revenue SKUs (e.g., 23843, 22423), with percentage shares of combined revenue.

Observations:

- **SKU 23843 leads** at ~19 percent.
- **SKU 22423 next** at ~16 percent.
- Together, the top three (23843, 22423, 85123 A) contribute ≈**46 percent** of top-10 revenue.
- The bottom five each hover between 5 percent–9 percent.
- Collectively, these ten SKUs account for ~**80 percent** of total product revenue.

Interpretation & Student-Style Reflection:
This high concentration reminds me of the **Pareto Principle** (80/20 rule). A few heavy hitters drive the majority of sales. But in a real-world e-commerce context, overdependence on a small set of SKUs is risky:

- **Stock-Out Vulnerability:** If one flagship product goes out of stock, revenue could drop by nearly 20 percent in an instant.
- **Competitive Exposure:** Rivals grilling on those high-revenue SKUs could poach our customers.
- **Innovation Stagnation:** The company may underinvest in novel or "long-tail" products that collectively could drive incremental revenue.

I'd dive deeper into SKU-level margins—are these top revenue generators also our top profit drivers? If not, a high-revenue, low-margin item could be less attractive than a stable mid-revenue, high-margin SKU.

Key Business Implications:

- **Concentration Risk:** The "all-eggs-in-few-baskets" phenomenon demands diversification.
- **Bundling Strategies:** We could package bestsellers with underperformers at a slight discount to boost long-tail visibility.
- **SKU Lifecycle Management:** Are any of these bestsellers nearing end-of-life? Product roadmap alignment is crucial.

I would recommend a cross-sell experiment: at checkout for SKU 23843, suggest a complementary lower-volume SKU at a 10 percent bundle discount. Tracking incremental uplift would quantify long-tail potential.

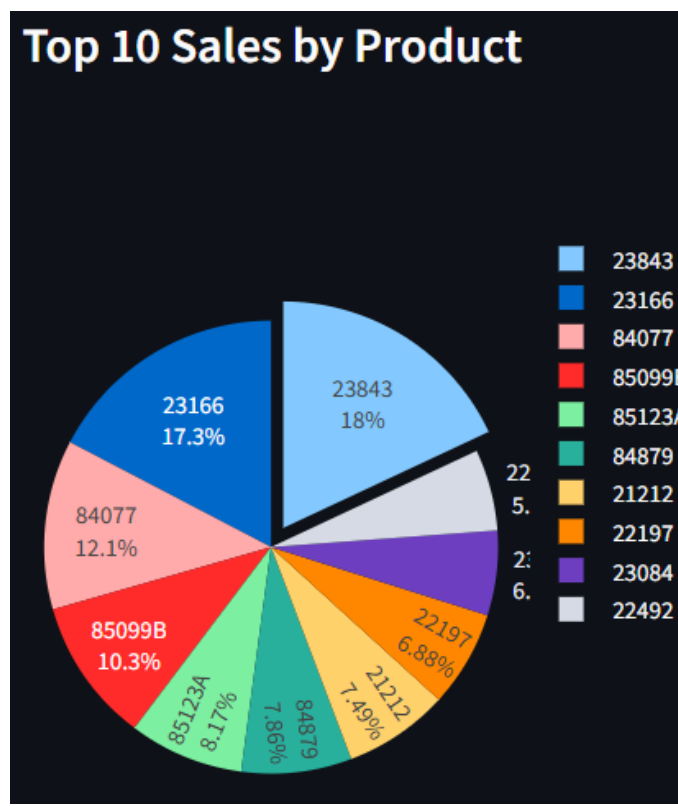6. **Analysis of "Top 10 Sales by Product"**

Chart Description:
A second pie chart ranks SKUs by **unit count sold**, not revenue.

Observations:

- **SKU 23843** again leads in volume (18 percent).
- **SKU 23166** takes second in units (17.3 percent), yet held a smaller revenue share in the previous chart—indicating a lower average selling price (ASP).
- **SKUs like 84077 and 850998** appear higher here relative to their revenue slice.
- **SKU 22423**, a top revenue earner, sells fewer units—suggesting it's a premium-priced item with higher ASP.

Interpretation & Student-Style Reflection:
By comparing the two pie charts side by side, I learn that:

- **High-Volume/Low-ASP Items** (e.g. 23166): Great for driving site traffic and volume-based metrics. Potentially lower margins, but excellent for customer acquisition and collecting reviews.
- **Low-Volume/High-ASP Items** (e.g. 22423): Premium segment—fewer customers but higher revenue per sale. Ideal for margin maximization but riskier if over-discounted.

I'd run a simple profitability estimate: multiply unit sales × (UnitPrice – Cost) for each SKU (assuming cost data were available). That would reveal true margin drivers, which might differ from pure revenue or volume leaders.

Key Business Implications:

1. **Segment-Specific Strategies:**
   - Use high-volume SKUs in loss-leader promotions to bring customers in.
   - Offer premium bundles or financing options for high-ASP SKUs to ease purchase friction.
2. **Inventory Optimization:**
   - Forecast high-volume items carefully to avoid stockouts.
   - Consider MOQ (minimum order quantity) adjustments for slower-moving big-ticket products.
3. **Personalized Recommendations:**
   - On product pages for 23166, suggest upsells of 22423.
   - For buyers of 22423, recommend smaller accessories (e.g. 23166) to increase cart value.

A next step in my analysis would be to build a simple machine-learning recommendation model (e.g. association rules) to automate these cross-sell suggestions.

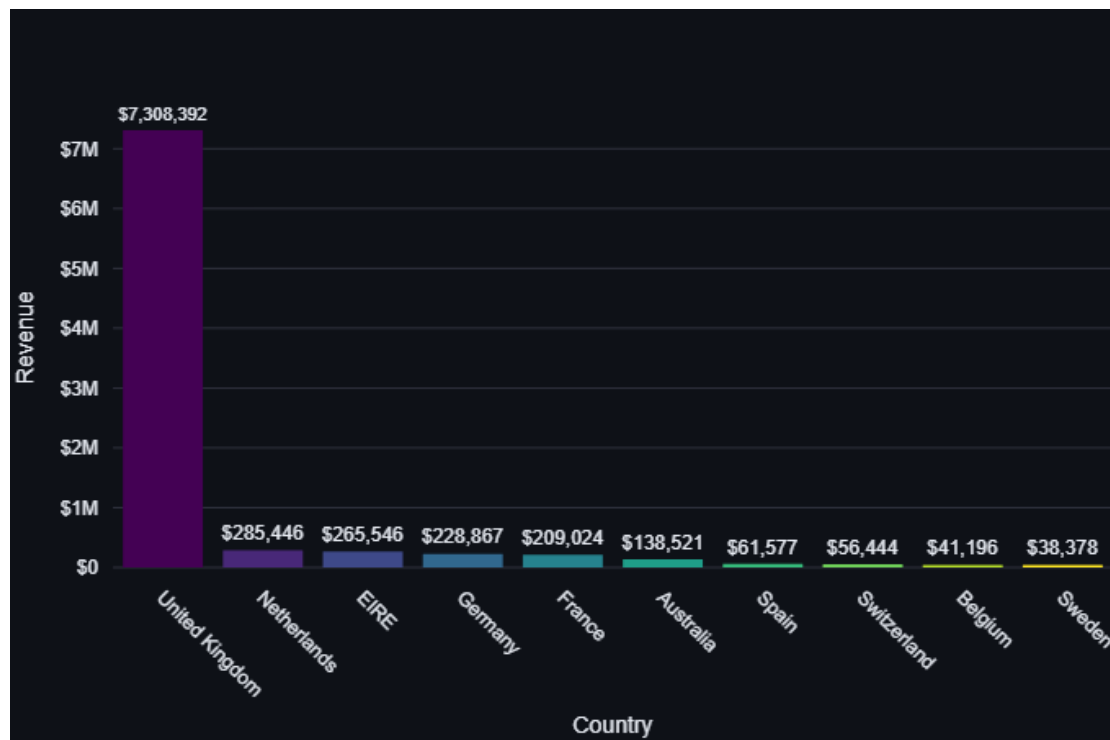7. **Analysis of "Top 10 Total Revenue by Country"**

Chart Description:
A vertical bar chart shows total revenue per country; the U.K., Netherlands, USA, Germany, France, Australia, Spain, Switzerland, Belgium, and Sweden comprise the top 10.

Observations:

- **U.K. dominates** at $7.3 M—**~75 percent** of total.
- **All other countries combined** barely top $1.6 M.
- The next-highest markets—Netherlands ($285 K) and USA ($265 K)—are each under 4 percent.

Interpretation & Student-Style Reflection:
This geographic skew is striking. It tells me:

- **Market Dependence Risk:** A single-market focus subjects the business to local economic cycles, currency fluctuations (GBP volatility post-Brexit), and potentially regulatory changes (VAT or consumer protection laws).
- **Localization Shortfalls:** If the site were truly global, I'd expect broader spread— e.g. the U.S. as a major e-commerce market.
- **Marketing and Fulfillment Gaps:** High shipping costs or long delivery times abroad may deter non-U.K. shoppers.

I'd augment transaction data with shipping-cost estimates to see if that correlates with low uptake in countries further afield. I might also look at time-on-site or cart-abandonment rates by region if available.

Key Business Implications:

1. **Diversification Imperative:** Relying on one market stunts international growth and increases risk.
2. **Localization Investments:** Translate key pages, show local currencies, and integrate regional payment gateways (e.g. iDEAL in the Netherlands, Klarna in Germany).
3. **Regional Promotions:** Align with local holidays—e.g. Fourth of July sales in the U.S., Singles' Day in China (if data extended), and Boxing Day in Australia.

I recommend piloting a "Global Expansion Sprint": choose two adjacent European markets (e.g. Germany and France), localize the storefront, run targeted Facebook/Instagram ads, and measure lift in new-customer acquisition versus a control group.

8. <u>Strategic Solutions for Mitigating E-commerce Concentration Risks</u>

Our e-commerce operation faces a critical concentration dilemma: we rely heavily on a narrow seasonal window for revenue spikes, a small set of flagship SKUs for most sales, and a single geographic market for the bulk of our income. This manifests as pronounced early-year troughs and an anomalous December collapse in monthly revenue; a stark Sunday underperformance versus weekdays; roughly 80 percent of product revenue coming from just ten SKUs; and about 75 percent of total sales generated in the U.K. alone. These interlinked vulnerabilities not only hinder reliable forecasting and cash-flow planning but also expose us to supply-chain, competitive, and regional economic risks that could rapidly erode our topline if left unaddressed.

To mitigate these risks and unlock growth, we propose a multi-pronged strategy. First, we'll level out seasonal troughs and weekend lulls by launching spring-and-summer "Early-Bird Specials" and dedicated Sunday promotions with enhanced loyalty incentives. Second, we'll diversify our product portfolio by curating "Under the Radar" bundles that pair long-tail SKUs with bestsellers, while also investing in adjacent product variants informed by customer feedback. Third, we'll implement a basic dynamic-pricing engine to optimize average selling price and roll out a simple association-rule recommender at checkout to drive cross-sell uplifts. Finally, we'll expand internationally with localized microsites for Germany and France—complete with translated content, local payment options, and region-targeted ad campaigns— while monitoring performance via real-time dashboards and conducting quarterly holiday-season stress tests to avert technical or logistical failures.

Reflecting on this analysis underscores the power of data to surface actionable insights: visualizing month-to-month volatility revealed calendar imbalances; comparing revenue versus volume charted our product and pricing spectrum; and mapping sales by country highlighted untapped markets. As a student of data analytics, I've learned that coupling robust quantitative findings with targeted experiments (A/B tests, pilot campaigns) transforms raw numbers into business-critical decisions. Moving forward, integrating richer customer-level and session-level data would allow deeper personalization and a full-blown recommendation engine—further enhancing both customer experience and sustainable growth.