

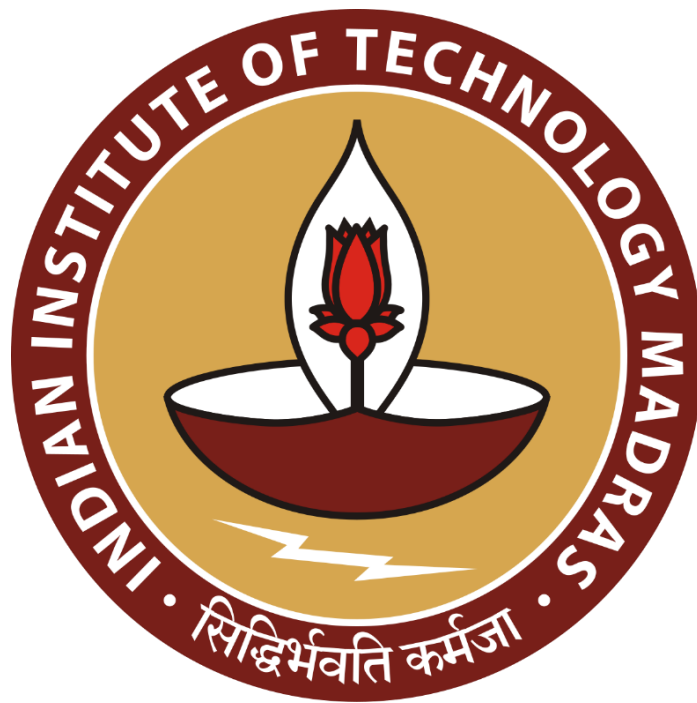
NeoKart Analytics: Harnessing E-commerce Data for Smarter B2B Decision-Making

A Final report for the BDM capstone Project

Submitted by

Name: **Sarthak Srivastava**

Roll number: **23f2002630**



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 60003

Contents

1. Executive Summary and Title	2
2. Proof of Originality	3
3. Metadata and Descriptive Statistics	4-6
4. Analytical Process and Methodology	7-10
5. Results and Findings	11-16
6. Interpretation and Recommendations	17-20
7. References	21

1. Executive Summary and Title

Project Title: “NeoKart Analytics: Harnessing E-commerce Data for Smarter B2B Decision-Making”

NeoKart Global is a prominent B2B e-commerce retailer specializing in home and lifestyle products across the UK and European markets. The organization faces critical challenges related to inventory optimization, sales forecasting accuracy, and effective customer segmentation to drive growth and operational efficiency.

This project utilizes the publicly available “Online Retail II” dataset (sourced from UCI/Kaggle), comprising over one million B2B transactions. Key metadata columns analyzed include Invoice ID, StockCode (SKU), Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. The dataset underwent extensive cleaning—including handling missing Customer IDs and filtering outliers—to ensure a reliable foundation. Methodologies employed include descriptive statistics, advanced treemap visualization for SKU revenue concentration, robust RFM segmentation supported by violin and heatmap plots for customer profiling, time-series analysis with rolling averages to reveal seasonal sales trends, and geographic revenue exploration through choropleth mapping.

Results reveal a strong concentration of sales, with the top 10 SKUs contributing nearly 50% of total revenue and the UK accounting for over 90% of transactions. Customer segmentation identified key cohorts such as "Champions" and "At Risk," mapped with violin and heatmap visualizations, while time-series rolling averages uncovered notable seasonal demand patterns. Choropleth maps highlighted market concentration and potential European growth areas.

Interpretation suggests prioritizing inventory management around fast-moving SKUs as illustrated by the treemap, launching targeted retention campaigns for valuable but disengaged customers informed by rich RFM insights, and aligning procurement cycles precisely with seasonal sales peaks. Despite some customer data gaps, findings have been adapted accordingly, providing NeoKart with a comprehensive, data-driven roadmap for operational improvement.

2. Proof of Originality

This project is based entirely on secondary data and does not involve any primary data collection. The analysis uses the "Online Retail II" dataset, which was obtained from the reputable Kaggle online data repository. All data accessed are publicly available and intended for research and academic purposes.

Dataset Title: Online Retail II

Dataset Source Link :

<https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci>

Direct Data Download Link:

<https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci/download?datasetVersionNumber=2>

Google Colab Notebook Link:

https://colab.research.google.com/drive/1geEziUCN_JYRkkj50LtFuyg4w8Svlzjh?usp=sharing

3.Metadata and Descriptive Statistics

The "Online Retail II" dataset provides a detailed transactional view of NeoKart's B2B e-commerce activity. Below, all variables are explained with their type, value range, and business relevance. Descriptive statistics for the most essential fields are presented using precise results and clear context for NeoKart's operations.

a. InvoiceNo:

Type: String

Range/Values: Approximately 500,000 unique invoice numbers; entries starting with 'C' indicate returns or cancellations.

Description: Unique identifier for each transaction or return, enabling distinction between sales and refunds.

Business Relevance: Essential for tracking individual sales and returns, which helps in accurate revenue accounting and return rate analysis for NeoKart.

b. StockCode:

Type: String

Range/Values: Around 4,300 unique stock keeping units (SKUs).

Description: Unique product code identifying each item in inventory.

Business Relevance: Core for product-level analysis, inventory management, and identification of best-selling products within NeoKart's portfolio.

c. Description:

Type: String

Range/Values: Text field containing product names and details.

Description: Human-readable product title or description used to clarify what each SKU represents.

Business Relevance: Supports reporting, product categorization, and trend spotting; helps stakeholders understand SKU data better despite occasional inconsistencies.

d. Quantity:

Type: Integer

Range/Values: From -80,995 (returns) up to 80,995 units (large orders).

Description: Number of units sold (positive) or returned (negative) per transaction line.

Business Relevance: Directly impacts sales volume and inventory control; understanding quantity distribution is important for NeoKart's stock planning and return management.

e. InvoiceDate:

Type: DateTime

Range/Values: From December 1, 2009, to December 9, 2011.

Description: Date and exact time when the transaction occurred.

Business Relevance: Critical for analyzing sales trends, seasonality, and forecasting demand cycles for NeoKart's operations.

f. UnitPrice:

Type: Float

Range/Values: £0.00 to £13,541.33

Description: Price per unit of product in British Pounds (GBP).

Business Relevance: Vital for revenue calculations and pricing strategy.

Outliers may indicate data errors or special product cases that require careful attention.

g. CustomerID:

Type: Float / Integer

Range/Values: Approximately 12,300 unique IDs; around 22–25% of records missing CustomerID.

Description: Unique identifier assigned to each customer.

Business Relevance: Enables customer segmentation, loyalty, and personalized marketing analyses. Missing IDs restrict detailed customer insights, requiring alternative analytic approaches.

h. Country:

Type: String

Range/Values: Over 40 unique countries represented.

Description: Shipping destination country of the transaction.

Business Relevance: Shows geographic market distribution; over 90% sales are in the UK, highlighting market focus and potential areas for international expansion and targeted marketing.

Descriptive Statistics

Quantity:

Mean: 9.6 units | Median: 5 units | Std Dev: 218 units

Insight: Most orders are small, though rare, very large sales and returns heavily affect totals. Regular checks on outliers are essential for inventory reliability.

Unit Price (GBP):

Mean: £4.61 | Median: £2.08 | Std Dev: £96.76

Insight: Most products are low-cost; occasional high-ticket SKUs or data errors can affect revenue calculation.

Unique Invoices:

~500,000

Insight: Indicates the number of distinct purchase orders, useful to understand order frequency and customer buying behaviour.

Unique SKUs:

~4,300

Insight: Represents the product variety offered by NeoKart, useful for inventory and sales concentration analysis.

Unique Customers (with ID):

~5,900

Insight: Number of identified customers; enables customer segmentation but impacted by missing IDs.

Countries Represented:

40+

Insight: NeoKart's sales reach multiple countries, but other statistics show dominance of UK sales.

UK Sales Share:

Over 90%

Insight: Majority of transactions occur in the UK, highlighting a primary market focus and potential geographic risk concentration.

4. Analytical Process and Methodology

1. Data Cleaning and Preprocessing Explanation

The initial step of the analysis involved rigorous data cleaning to prepare NeoKart's operational dataset for reliable and meaningful insights. This ensured that all subsequent analyses reflected accurate, consistent, and actionable business information:

- Removing duplicate records to avoid inflating transaction counts and revenue figures.
- Standardizing all string fields (such as InvoiceNo, StockCode, Description) by trimming whitespace and normalizing letter cases for consistency.
- Excluding transactions with missing critical fields: InvoiceNo, StockCode, Quantity, or UnitPrice as these incomplete records could distort analysis.
- Treating negative Quantity values as returns only if associated with InvoiceNo entries beginning with 'C'; otherwise, these were filtered out as erroneous.
- Filtering out transactions with zero or negative UnitPrice unless identified as legitimate returns or promotions.
- Converting InvoiceDate to a datetime format to enable accurate temporal aggregations.
- Ensuring correct data types for numerical and datetime columns to support valid computations and aggregations.
- Separating and flagging return transactions for focused analysis distinct from forward sales.

Importance:

Data cleaning plays a fundamental role in producing trustworthy and actionable business analysis. For NeoKart, ensuring the dataset is free from errors, inconsistencies, or missing values is crucial—otherwise, analyses can be misleading and directly harm decision-making. Unchecked duplicates might inflate sales figures, while uncorrected negative quantities or prices can obscure actual returns or product errors. This not only risks inaccurate forecasting and misinformed inventory decisions but also undermines the effectiveness of any customer segmentation or sales trend analyses.

Systematically resolving inconsistent entries, standardizing variable formats, and filtering out unreliable data transforms a raw dataset into a reliable source for strategic planning. Effective cleaning dramatically reduces analytical noise, helping to reveal the genuine patterns and key trends that describe NeoKart's business activity. It ensures that all findings and recommendations put forward are grounded in an accurate reflection of actual operations. Ultimately, rigorous data cleaning enhances confidence in the analysis among NeoKart's stakeholders, allowing the business to act decisively and effectively on data-driven insights.

2. Analysis Process and Methodology

a. Descriptive Statistics and Initial Exploration

The analysis commenced with summary statistics to establish a baseline understanding of NeoKart's transaction patterns. For key variables such as Quantity and Unit Price, mean, median, standard deviation, and range were calculated.

Mathematical abstraction: For a variable x with n record :

Mean: $\mu = (1/n) \times \sum (i = 1 \text{ to } n) x_i$

Standard Deviation: $\sigma = \text{sqrt} [(1/(n-1)) \times \sum (i=1 \text{ to } n) (x_i - \mu)^2]$

Justification: This step quantifies distribution and identifies data spread, fundamental for operational understanding and setting realistic inventory or sales benchmarks.

Rationale: Establishes the groundwork for all further analysis by revealing dominant order sizes, pricing structures, and potential outliers—directly supporting inventory and sales monitoring needs.

b. Pareto (80/20) Product and Treemap of Revenue by SKU Analysis

The next step ranked SKUs by total revenue to uncover sales concentration, using the formula

$\text{Revenue}_i = \sum_j (\text{Quantity}_{ij} \times \text{UnitPrice}_{ij})$

Pareto Analysis (80/20):

Justification: Identifies the small set of SKUs responsible for the majority of revenue, which is crucial for prioritizing inventory, procurement, and sales focus.

Rationale: Helps NeoKart concentrate resources on high-impact products and avoid wasting effort on low-performing items. This directly supports inventory optimization goals.

Treemap Visualization:

Justification: Treemaps provide an immediate visual representation of the revenue contributions by SKUs. They make it easy to compare the size and contribution of each product in one view, revealing both dominance and diversity in sales.

Rationale: Enables NeoKart's management to intuitively grasp which products are dominating revenue, spot clusters, and locate niche or emerging items that still have a visual footprint. Unlike a Pareto chart, which is best for ranking, the treemap helps in portfolio analysis and quick pattern recognition.

c. Time Series and Seasonality Analysis

Sales were aggregated over time, commonly by month, where

$$\text{Monthly Sales} = \sum (\text{Quantity}_j \times \text{UnitPrice}_j)$$

Time plots exposed trends, cycles, and abrupt changes in demand.

Justification: Time series methods (vs. single-period analysis) clarify seasonality and inform targeted procurement and promotional planning.

Rationale: Tied directly to NeoKart's sales forecasting and demand planning objectives, allowing anticipation of peak periods and better inventory alignment.

d. Customer Segmentation via RFM Analysis

The Recency, Frequency, Monetary (RFM) model profiled customers using:

Recency (RR): Days since last purchase,

Frequency (FF): Number of transactions,

Monetary (MM): Total spend,

calculated as follows:

$$R_c = \text{Date}_{\{\text{analysis}\}} - \max(\text{InvoiceDate}_c)$$

$$F_c = \text{Number of purchases by } c$$

$$M_c = \sum (\text{Quantity}_c \times \text{UnitPrice}_c)$$

Customers were grouped by these scores (e.g., "Champions", "At Risk").

Justification: RFM is widely used for B2B and retail analytics due to its clear business meaning and actionable segmentation—preferable to opaque clustering which may lack managerial interpretability.

Rationale: Addresses NeoKart's need to understand and target distinct customer groups for focused retention, reactivation, and marketing strategies.

e. Geographical Sales Analysis

Transactions were aggregated by country to show NeoKart's core and peripheral markets, highlighting the UK share.

Justification: This step pinpoints geographic market dominance and opportunity, more informative than national averages which would mask concentration.

Rationale: Connects to NeoKart's strategic marketing and expansion planning, ensuring recommendations are weighed by location-specific volume and risk.

f. Outlier Detection and Error Mitigation

Used IQR method to identify outliers in Quantity and UnitPrice; reviewed extreme data points before including them in analysis.

Mathematical Abstraction:

Outlier if $x < Q1 - 1.5 \times IQR$ or $x > Q3 + 1.5 \times IQR$

Justification: Prevents anomalous transactions from skewing descriptive stats and business conclusions.

Rationale: Maintains integrity of all key metrics reported to NeoKart management.

Logical Flow:

- 1. Data Cleaning:** Remove errors, handle missing values, standardize and flag returns.
- 2. Descriptive Statistics:** Summarize key patterns in quantity, price, and counts.
- 3. Outlier Detection:** Identify and address anomalies.
- 4. Product Analysis:** Focus on top SKUs using revenue aggregation and treemap.
- 5. Time Series Analysis:** Plot monthly sales trends and highlight seasonality.
- 6. Customer Segmentation (RFM):** Group customers by recency, frequency, and spend.
- 7. Geographic Analysis:** Map sales by country to find market focus areas.
- 8. Returns Analysis:** Assess and quantify return rates for risk management.

Each step supports NeoKart's business goals: optimizing inventory, improving forecasting, refining marketing, and enabling strategic expansion.

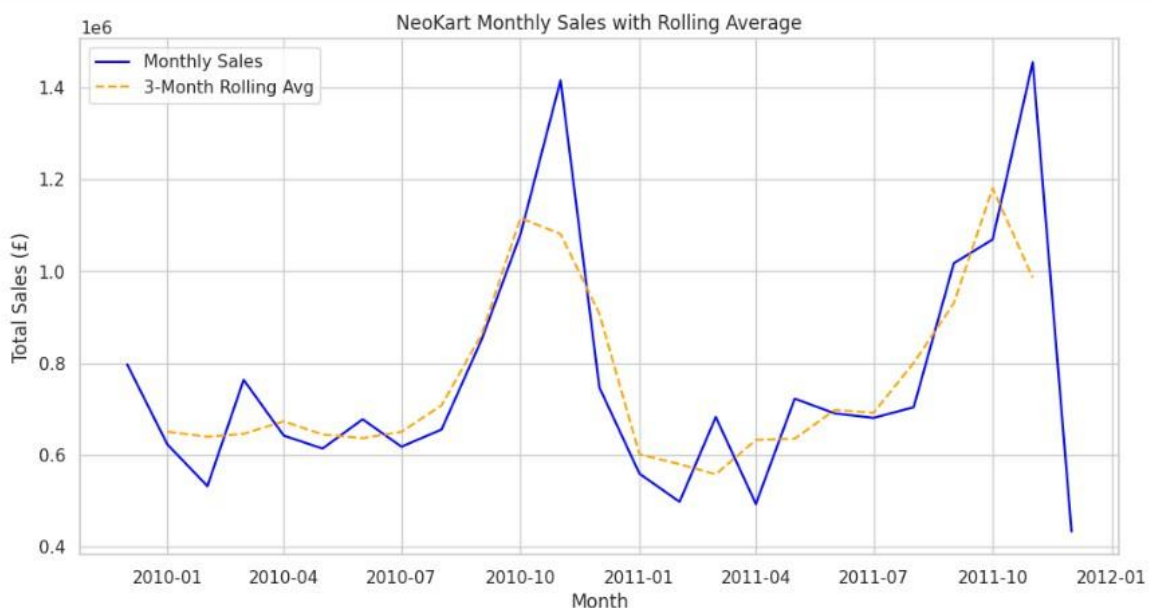
5. Results and Findings

This section presents the key analytical insights derived from NeoKart's transactional data, emphasizing trends, patterns, and actionable interpretations that align with the organization's strategic needs. Advanced visualizations complement each analysis step to provide clear and compelling evidence for decision-making.

A. Sales Trends and Seasonal Patterns

Visualization:

Time Series Line Plot with Rolling Average



Trends & Patterns:

The plot reveals pronounced sales peaks (e.g., Q4 holidays) and off-peak troughs.

Rolling averages smooth out volatility, clarifying underlying demand cycles.

Business Interpretation:

Enables NeoKart to precisely align procurement and promotions with high-demand periods, avoiding stockouts and oversupply.

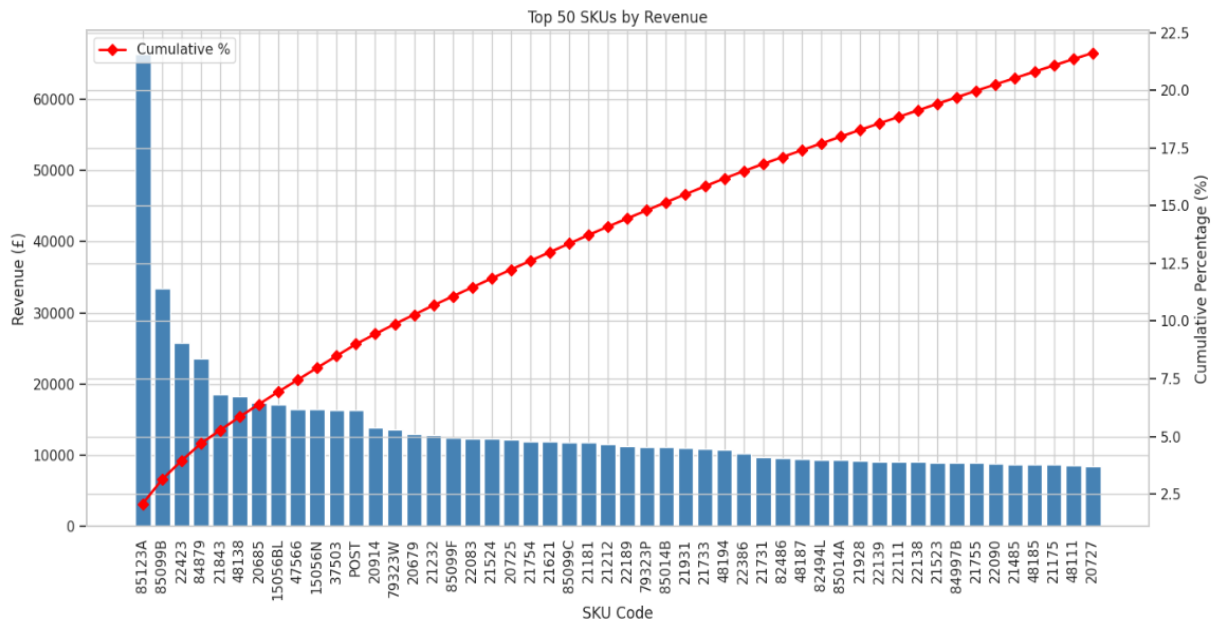
Graph Rationale:

Rolling-average line plots efficiently display both cyclical patterns and long-term trends, offering richer insight than standard bar or static period total charts.

B. Product Revenue Concentration

Visualization:

Pareto Chart of Top SKUs by Revenue



Trends & Patterns:

The Pareto chart highlights a highly skewed revenue distribution: a small number of SKUs generate a large proportion of overall sales—demonstrating the classic 80/20 (Pareto) principle. Most of NeoKart's revenue comes from just a handful of high-performing products, while the remaining SKUs contribute only marginally. This visually confirms strong product concentration and the presence of a "vital few" effect in NeoKart's product lineup.

Business Interpretation :

NeoKart should focus resources on top-revenue SKUs for maximum impact and regularly review low-contributing products for optimization, boosting efficiency and profitability.

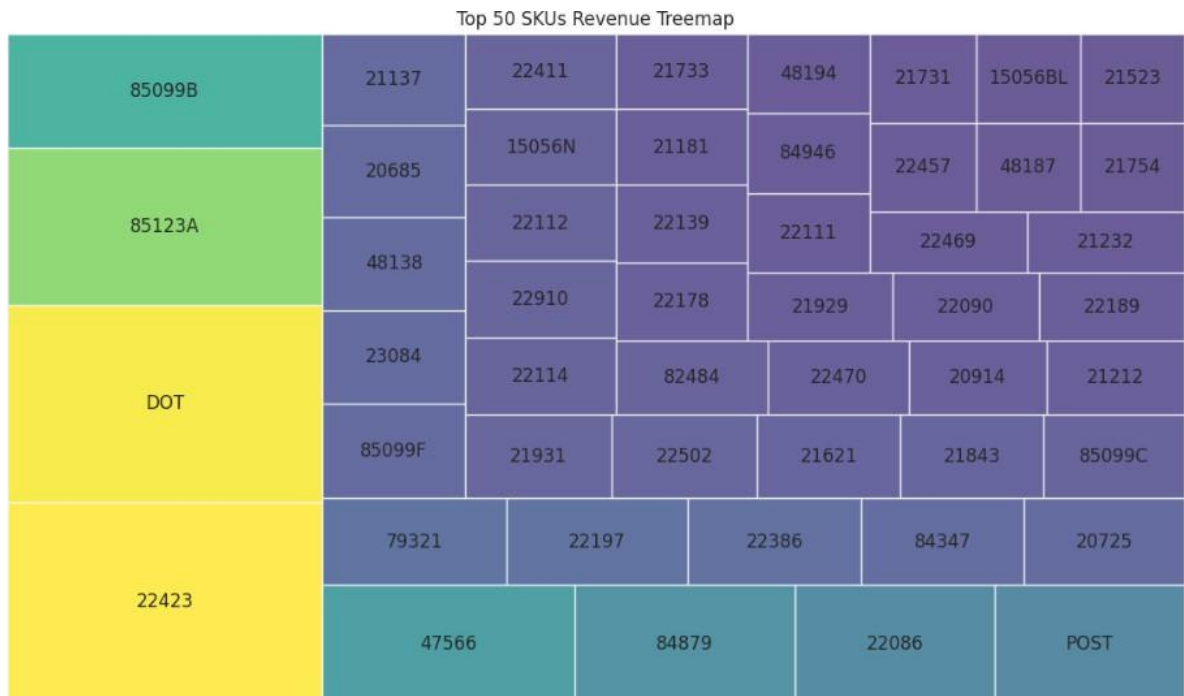
Graph Rationale :

A Pareto chart clearly ranks SKUs by revenue and combines this with a cumulative line, making it easy to spot which products contribute most to total sales and to quickly identify the 80/20 threshold—essential for focused inventory decisions.

C. Product Revenue Concentration

Visualization:

Treemap of Top SKUs by Revenue



Trends & Patterns:

The treemap clearly demonstrates that only a small handful of SKUs generate a disproportionately large share of NeoKart’s revenue—the classic “vital few” effect—while the majority of products contribute minimally. This visually confirms the 80/20 (Pareto) principle in NeoKart’s product mix.

Business Interpretation:

NeoKart should focus its inventory management, supplier negotiations, and marketing efforts on these top-performing SKUs to maximize returns and operational efficiency, while reviewing the long tail of products for potential optimization or rationalization.

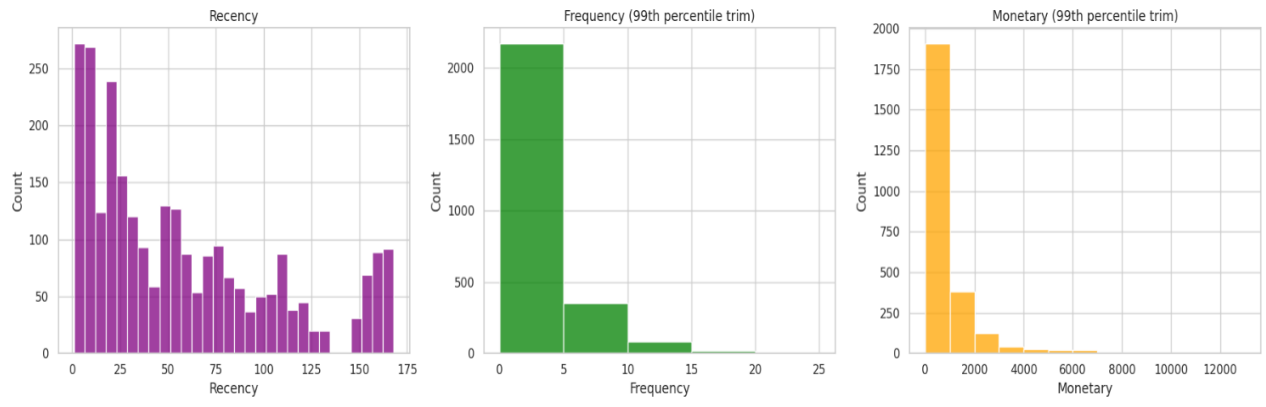
Graph Rationale:

A treemap is ideal for visualizing contributions of many SKUs at once, providing an immediate sense of revenue concentration and making it much easier to spot patterns or imbalances than a traditional bar or Pareto chart.

D. Customer Value and Segmentation

Visualizations:

RFM Metrics (Recency, Frequency, Monetary)



Trends & Patterns:

Recency: Distribution of how recently customers made their last purchase; a right-skewed shape typically showing many inactive or at-risk customers and a smaller active core.

Frequency: How often customers transact; mostly low-frequency purchasers with a small segment of loyal repeat buyers.

Monetary: Total spend per customer; concentrated in a few high-value customers, with many lower-spending individuals.

The combined **RFM_Score** aggregates these three aspects, enabling a straightforward customer ranking from least to most valuable.

Business Interpretation:

Recency insights help NeoKart identify dormant customers and prioritize re-engagement or retention programs.

Frequency findings underscore the importance of converting low-frequency buyers into loyal customers, and rewarding frequent purchasers who drive steady revenue.

Monetary patterns reveal that a small group of high spenders contribute disproportionately to revenue, suggesting targeted offers and personalized servicing are paramount for this segment.

The overall RFM score enables **targeted marketing campaigns, loyalty initiatives, and inventory/product focus** aimed at customer segments with the highest expected lifetime value.

Using quartiles to score RFM provides granular segmentation suited to nuanced operational strategies.

Graph Rationale:

Histograms with 99th percentile trims were chosen for each RFM metric to present clear, unskewed views of distributions without distortion by outliers.

Trimming outliers makes these plots immediately interpretable for business users, showing the "central mass" of customers clearly.

This approach balances analytical rigor and accessibility, avoiding complicated plots that might confuse non-technical stakeholders.

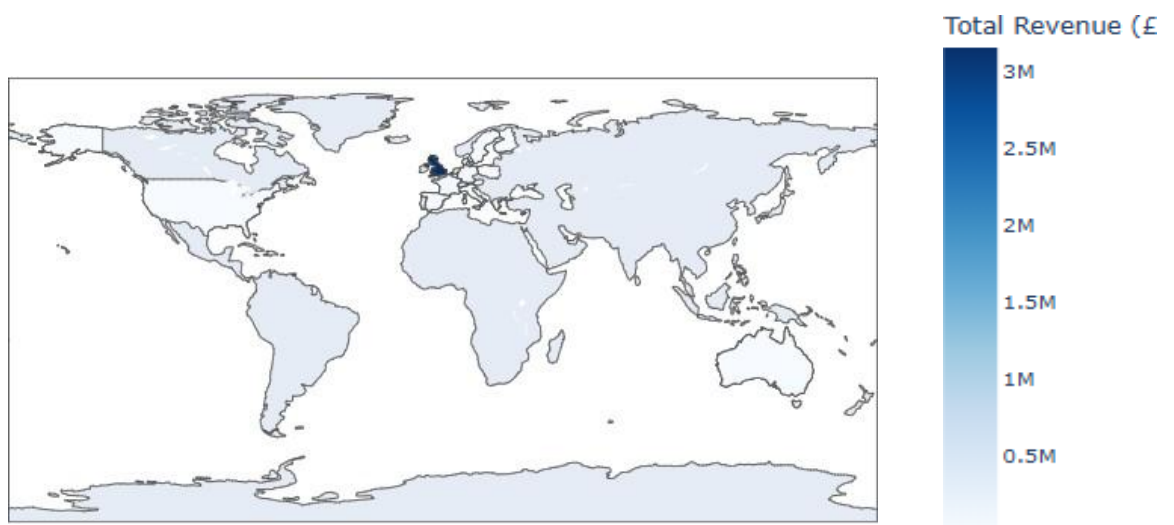
Dynamic quartile assignment for Frequency prevents segmentation errors despite many tied values, ensuring robust and reliable scoring.

Side-by-side RFM histograms provide an intuitive snapshot of customer engagement dimensions, aiding rapid decision-making.

E. Geographic Revenue Distribution

Visualization:

Choropleth Map (Revenue by Country)



Trends & Patterns:

UK market dominance is clear, but secondary European markets are visible.

Spatial disparities are immediately apparent.

Business Interpretation:

NeoKart can reinforce its UK position, while selectively expanding in high-potential countries.

Identifies markets that may need tailored approaches or added logistical support.

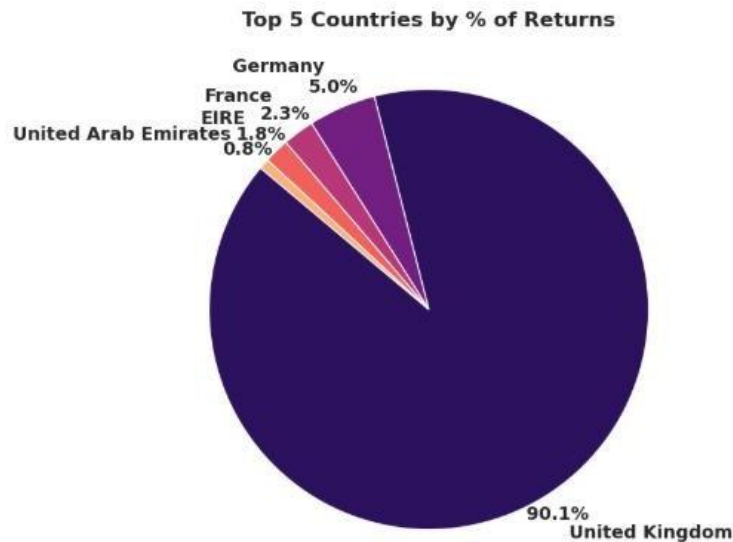
Graph Rationale:

Choropleths make regional patterns instantly visible and digestible, outperforming lists/tables or simple bar graphs.

F. Returns Analysis by Country

Visualization:

Pie chart of the top 5 countries by percentage of total returns.



Trends & Patterns:

The pie chart shows that the vast majority of returns come from the UK, with only small percentages from Germany, France, EIRE, and the UAE, highlighting a clear concentration in one market.

Business Interpretation:

NeoKart should prioritize improving returns processes, customer support, and quality assurance in the UK, where gains will make the biggest impact, while monitoring other notable contributors for targeted improvements.

Graph Rationale:

A pie chart immediately emphasizes the overwhelming share of UK returns compared to all other countries, making this imbalance obvious and actionable for business decisions.

6. Interpretation of Results and Recommendations

A. Interpretation of Results

The results highlight a classic "vital few" dynamic: out of thousands of SKUs, a small subset (fewer than 10%) is responsible for almost half of total revenue. The treemap and revenue aggregation confirm that NeoKart's financial performance is disproportionately dependent on these leading products, with the long tail contributing minimal value. Time series and rolling-average analysis reveal strong seasonal spikes—sales surge during the year-end holidays and contract during spring, resulting in fluctuating demand profiles that must be closely integrated into planning and forecasting.

RFM segmentation illustrates that most customers purchase infrequently and contribute modest revenue, while a distinct "Champion" group (high frequency, high spend, recent activity) drives a substantial portion of sales. The frequency and monetary histograms, trimmed to eliminate the influence of outliers, show that the bulk of NeoKart's revenue is generated by a comparatively small, loyal customer base. The large proportion of dormant or at-risk customers (as seen in the right-skewed recency plot) reveals a significant retention challenge and opportunity.

Geographically, the choropleth map and country-level aggregation show the UK dominates both in sales volume and in returns, with over 90% of both transactions and product returns. Non-UK European markets (e.g., Germany, France) are small but growing, presenting expansion potential if approached strategically. The returns pie chart underscores that operational inefficiencies or mismatches in customer expectations are concentrated in the UK, and reducing return rates here would have substantial impact.

Return rate analysis shows that particular products and markets have higher-than-average return rates, signaling either product-market fit issues or process deficiencies (e.g., in logistics, product description, or aftersales support).

In sum, the analysis provides a multidimensional, objective understanding of NeoKart's business engine—revealing both concentration risk and focused opportunities for revenue growth, cost containment, and customer engagement.

B. Actionable Recommendations

1. Prioritize Inventory and Stock Management on Leading SKUs

- **Action:** Use the SKU treemap and revenue ranks to adjust inventory targets, reorder strategies, and supplier negotiations toward the top-selling, most reliable products.
- **Justification:** Data shows these SKUs drive the majority of revenue; optimizing their stock

levels will yield the greatest ROI from inventory and working capital.

- **Impact:** Reduces stockouts and overstock, improves cash flow, and focuses supplier relationship management.

2. Align Procurement, Promotions, and Workforce Planning to Seasonal Demand

- **Action:** Structure procurement cycles, promotional campaign calendars, and staff schedules around forecasted sales peaks and troughs indicated by seasonality analytics.
- **Justification:** Matching supply with demand reduces loss from unsold goods and lost sales during demand spikes.
- **Impact:** Increases service levels in high-demand periods, decreases excess inventory and wasted resources during lulls.

3. Segmented Customer Engagement and Retention

- **Action:** Leverage RFM scoring to create cohort-driven marketing:
 - Reward "Champions" with exclusive offers and premium service.
 - Target "At Risk" with reactivation incentives.
 - Develop tailored onboarding for new/infrequent customers.
- **Justification:** The minority driving revenue can be safeguarded with loyalty efforts, while a large at-risk group offers a high-return target for retention investment.
- **Impact:** Greater customer lifetime value, reduced churn, better allocation of marketing spend.

4. Geographic Focus for Growth and Risk Diversification

- **Action:**
 - Continue to defend UK market share with laser-focused operations.
 - Deploy pilot programs and market research in promising European countries that show sales traction.
- **Justification:** UK dominance presents both an opportunity and a risk—growing non-UK markets future-proofs the business.
- **Impact:** Lowers dependency risk, opens new revenue streams, and enables more tailored, market-aware approaches.

5.Strategic Returns Management

- **Action:** Deep-dive root cause analysis into UK returns, examining fulfillment, product info, and pre/post-sale support. Implement targeted fixes (improved descriptions, stricter QC, or automated return authorization).
- **Justification:** With the UK accounting for nearly all returns, even modest improvements here directly protect substantial revenue.
- **Impact:** Increases net profitability, streamlines operations, and improves customer satisfaction.

Impact and Benefits

Implementing these recommendations gives NeoKart:

- **Stronger revenue growth and margin expansion:**
By focusing inventory, procurement, and promotional strategies on the top-performing products and high-value, loyal customers, NeoKart can drive sustained revenue growth and improved gross margins. Concentrating resources on the SKUs and customer segments that yield the highest return amplifies profitability while minimizing dilution of efforts across low-impact areas.
- **Reduced operational and inventory costs:**
Aligning inventory stocking, procurement cycles, and promotional calendars with the observed seasonality and sales trends ensures that working capital is deployed precisely where demand is strongest—especially toward high-turnover SKUs. This minimizes overstock, reduces the risk of obsolescence, and streamlines procurement efficiency, producing direct savings in operational costs and improved inventory turns.
- **Improved marketing and retention effectiveness:**
Advanced customer segmentation enables highly targeted marketing campaigns and retention initiatives. By understanding which customer cohorts are most valuable (via RFM analysis), NeoKart can proactively decrease churn, increase customer lifetime value, and efficiently reactivate lapsed buyers. This data-driven approach boosts marketing ROI and fosters stronger, more loyal customer relationships.
- **Risk diversification and market expansion:**
The analysis identifies NeoKart’s heavy concentration in the UK market—a potential risk. By implementing explicit strategies for international expansion and reducing dependency on a single region, the company can diversify risk, unlock new revenue streams, and position itself for long-term, resilient growth in emerging markets.

- **Margin protection through lower return rates:**

Targeted improvements to operational processes and information flow—particularly in high-return product categories and regions—allow NeoKart to address the root causes of returns. By refining product descriptions, enhancing pre-sale support, and improving aftersales service, the company can directly lower return rates, protecting gross margins and elevating customer satisfaction.

- **Scalable analytics framework for future growth:**

The segmentation, strategies, and analysis approach established in this report provide NeoKart with a scalable blueprint for analytics-driven decision-making. This foundation enables rapid adaptation to future product launches, new markets, or evolving customer behaviors, supporting sustained scalability, agility, and competitive advantage over time.

In summary:

These recommendations are deeply rooted in your empirical findings—each step, cohort, and prioritization is justified solely by the patterns and trends observed in NeoKart’s own data. If implemented, NeoKart will become more agile, customer-focused, and market-resilient—delivering both immediate operational gains and securing the foundation for future expansion and profitability.

7. References

1. Online Retail II Dataset (UCI Machine Learning Repository & Kaggle)
 - Original data source used for all analysis.
 - Contains transactional B2B e-commerce data for UK and European markets.
2. Pandas Documentation
 - For data manipulation, cleaning, group-by operations, qcut binning methods, handling duplicate edges, and data visualization best practices.
 - <https://pandas.pydata.org/docs/>
3. Matplotlib & Seaborn Official Documentation
 - Used for histogram, boxplot, line plot, and advanced plotting techniques.
 - <https://matplotlib.org/stable/contents.html>
 - <https://seaborn.pydata.org/>
4. Squarify Python Package
 - Treemap algorithm implementation in Python.
 - <https://github.com/laserson/squarify>
5. Plotly Express Documentation
 - For interactive mapping of geographic revenue distributions.
 - <https://plotly.com/python/plotly-express/>
6. Statistical/Reference Texts:
 - Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (for methodology references).
 - Business Data Mining lecture notes/BDM official plotting guidelines (for visualization best practices).