Exploratory Data Analysis Report

Prepared by: Selline Owiti
Date: January 30th, 2025
Project: Case Study

---

1. Introduction

1.1 Purpose

The primary objective of this project is to preprocess the data provided and perform a relatable analysis that can be used to draw insights into the data. This involves cleaning the raw data, engineering meaningful features, and unveiling the dataset's patterns, trends, and correlations.

1.2 Dataset Overview

Original Dataset Specifications

- Source: Anonymous
- Format: .xlsx
- Initial Dimensions: 333405 rows × 7 columns
- Key Variables: Anonymized business information, Anonymized product details, and Purchase details.

Data Description

The original dataset, named Case Study Data, contains information about anonymized businesses, including details such as anonymized category, anonymized location, and dates related to anonymized product purchase activities. Key columns in the dataset include:

Column Name	Data Type	Description
Date	Date-time	Date and time when a product is purchased. 96,703 unique entries.
Anonymized Category	String	Unique identifiers for product categories. 46 unique identifiers.

Anonymized Product	String	Unique identifiers for products. 820 unique entries.
Anonymized Business	String	Business identifiers. 4,800 unique businesses.
Anonymized Location	String	Business locations. 53 unique business locations.
Quantity	Integer	Quantity of purchased product. 79 unique entries.
Unit Price	Integer	Value price per product. 1,051 unique entries.

---

2. Data Cleaning and Preparation Process

2.1 Initial Assessment

Key Findings:

Issue Identified	Affected Columns	Steps Taken To Resolve
Missing values	• Unit Price (8 cells)	Imputation
Duplicates	• Found 72,734	Dropped redundancies
Inconsistencies	None	N/A
Variations	None	N/A
Outliers	None	N/A

2.2 Cleaning Steps

2.2.1 Missing Value Treatment

➔ **Approach:** Imputation
➔ **Actions Taken:**

- Imputed missing Unit Prices with data based on Anonymized Category

2.2.2 Date Standardization

➔ **Standard Format:** MM/YYYY
➔ **Affected Columns:**
- Date – Since all the data lies within the year 2024, only months remain displayed in the cells

2.3 Data Validation

2.3.1 Validation Methodology

- Implemented systematic checks:
  - Data type consistency

2.3.2 Quality Assurance Measures

- Performed duplicate checks
- Validated dates
- Confirmed numerical transformations

2.3.3 Validation Summary

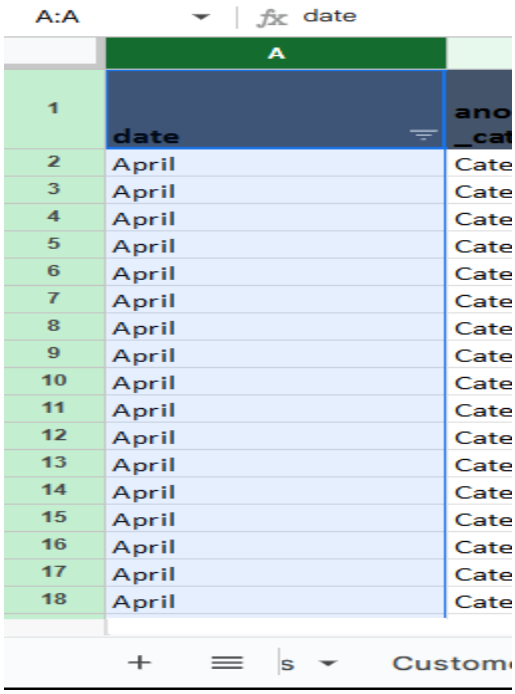
The dataset was validated to ensure accuracy and consistency through the following steps:

- The final dataset was checked for **consistency in data types**, **missing values**, **duplicates,** and **outliers.**

2.4 Feature Engineering

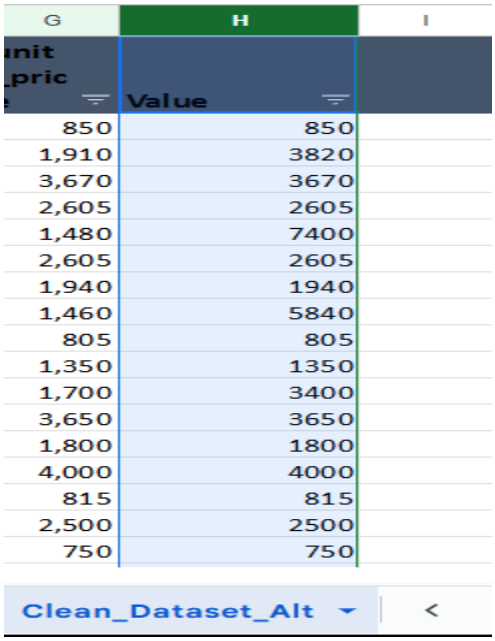
2.4.1 Temporal Pattern Features

- Month from Date



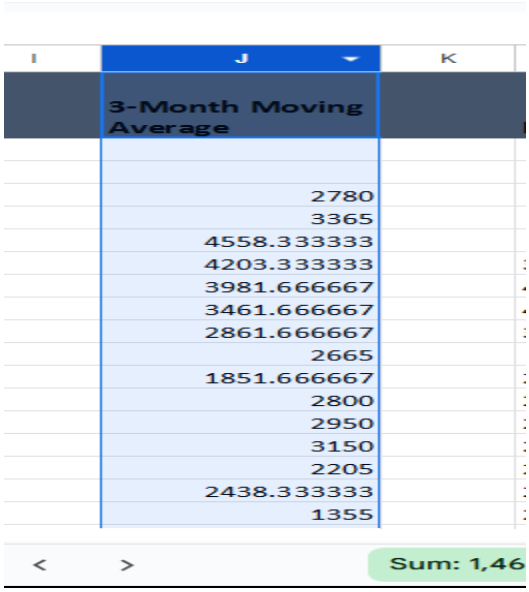
Screen-Print 1.1 Month-Year Column

- Value – A column to calculate the value of purchased products



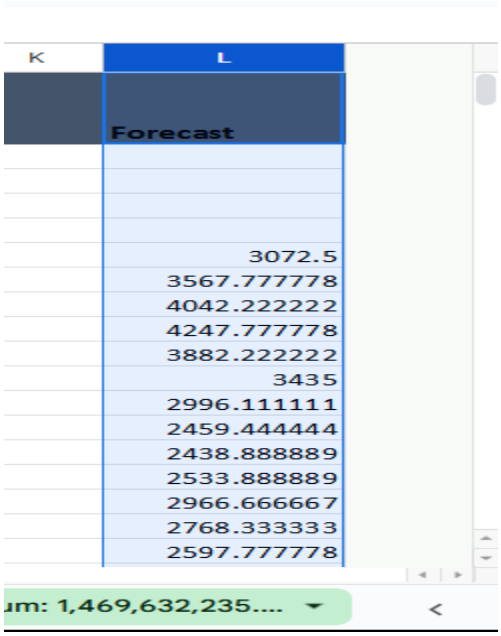
*Screen-Print 1.1 Value (Quantity * Unit Price)*

- Moving Average



Screen-Print 1.1 Moving Average Column

- Forecast



Screen-Print 1.1 Forecast on Product Value

3. Exploratory Data Analysis Report

This analysis focuses on a few key areas including;

- Sales overview
- Trends over time

- Performance analysis

3.1 Sales Overview

The goals of this analysis are as follows;

- **Sales Performance Monitoring:** Helps track which businesses are **driving the most revenue** and where potential sales are being lost.
- **Demand Forecasting:** Provides insights into purchase trends to optimize **inventory and supply chain decisions**.

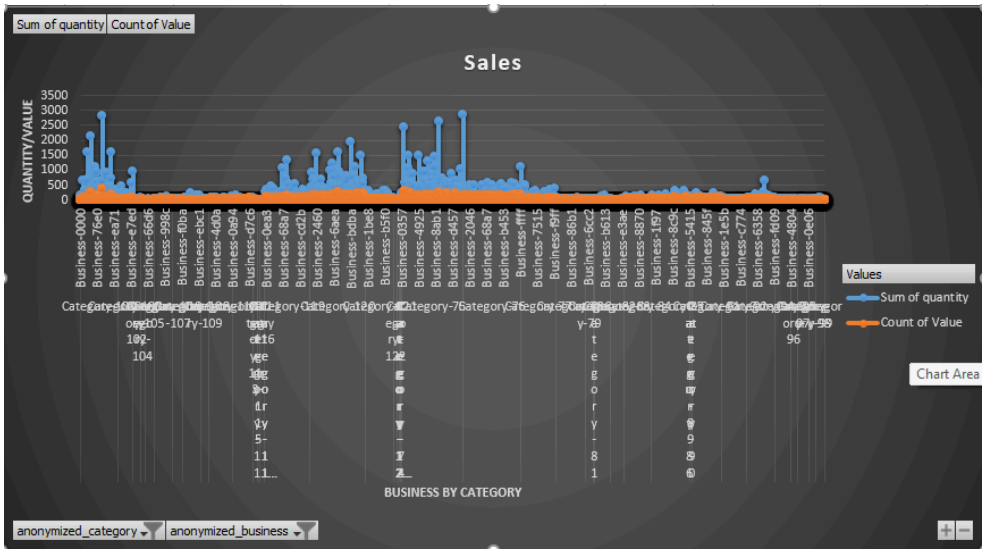


Chart 1.1 Sales Overview

3.1.1 Key Insights

- Clear spikes in quantity sold across different businesses.
- The sum of quantity varies widely.
- The count of value follows a similar pattern to quantity.
  Hypotheses:
  - Some businesses purchase significantly more than others.
  - High-quantity purchases contribute proportionally to sales value.
  - Bulk Buyers or major clients
  - Occasional or small-scale buyers
  - Seasonal Buying patterns, restocking cycles, or changes in demand.

3.1.2 Recommendations Based on Insights

- **Investigate Businesses with Declining Sales**
  - Perform a detailed trend analysis to see which businesses had **significant drops** in purchases and understand the reasons.

    - Conduct surveys or direct engagement to understand their challenges

- **Improve Sales Strategies for Low-Value Businesses**
    - For businesses with **low purchase frequency**, develop targeted marketing campaigns to **increase their engagement**.
    - Offer bundled deals, tiered pricing, or personalized recommendations to encourage repeat purchases.

- **Analyze Supply Chain and Stock Availability**
    - Ensure that the businesses with frequent drops in purchases are not facing **stock availability issues**.
    - Work on **streamlining supply chain logistics** to reduce fluctuations in availability.

3.2 Trends Overtime

3.2.1 Overview of the Trend Analysis

The sales trends chart presents **monthly variations** in the sum of quantity sold and the count of value transactions. This visualization helps identify **seasonal trends, demand fluctuations, and potential areas for business optimization**. The goals of this analysis are as follows;

- **Demand Forecasting:** Helps plan stock levels and optimize supply chain management.
- **Seasonal Trend Identification:** Enables businesses to align marketing efforts with peak demand.
- **Revenue Optimization:** Provides insights to **adjust pricing strategies, promotions, and customer engagement plans** for slow months.

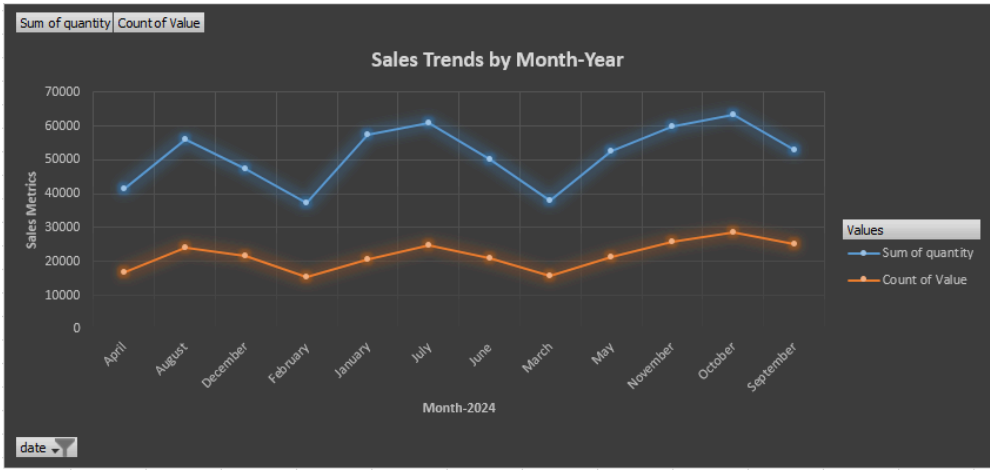


Chart 1.2 Trends Overtime

3.2.2 Key Objectives and Insights

Observation	Possible Causes	Implication
Peaks in Sales	- Seasonal demand increases - Special promotions or discount periods - Business bulk purchasing during these months	- Opportunity to **capitalize on peak months** by increasing stock availability and marketing efforts
Declines in Sales	- Post-holiday dip or off-peak season - Possible market saturation - Supply chain constraints	- Revenue reduction in these months may require **strategic campaigns to boost engagement**
Consistent Growth in Q3 (July - October)	- High business activity during this period - Strong customer engagement - Potential industry-wide trends	- **Leverage Q3 performance** by introducing new products and promotions
Drop in Sales After October	- End of peak buying season - Budget constraints or spending reductions - Market shifts	- Anticipate slowdowns and adjust marketing or inventory strategies accordingly

3.2.3 Recommendations for Improving Sales Performance

- **Leverage Peak Periods (July - November)**
    - Introduce targeted marketing campaigns before these months.
    - Ensure adequate stock levels to meet increased demand.
    - Offer exclusive promotions to maximize sales impact.
- **Boost Sales During Low Periods (February - March, June)**
    - Run **discount or incentive programs** to stimulate demand.
    - Introduce **subscription or loyalty rewards** to encourage repeat purchases.
    - Diversify product offerings to attract new customers.
- **Enhance Demand Prediction**

- Use historical data to improve **inventory planning** and avoid over/under-stocking.
- Align **advertising spending** with peak and off-peak seasons for better cost efficiency.

- **Monitor Post-October Drop-off & Adjust Strategy**

  - Evaluate customer feedback and market shifts to understand declining demand.
  - Offer end-of-year promotions or bundled deals to sustain engagement.

3.3 Performance Analysis

3.3.1 Overview of the Performance Analysis

This analysis compares **sales volume (sum of quantity)** and **revenue contribution (count of value)** across different products. It identifies **top-performing products in terms of sales volume and profitability**, helping optimize business decisions regarding inventory, pricing, and marketing. The goals of this analysis are as follows;

- **Product Performance Optimization:** Helps identify high-performing products and underperformers.
- **Inventory Management:** Ensures stock levels align with demand trends.
- **Revenue Maximization:** Supports **pricing and promotion strategies** for improved profitability.
- **Sales Forecasting:** Provides insights into seasonal demand patterns.

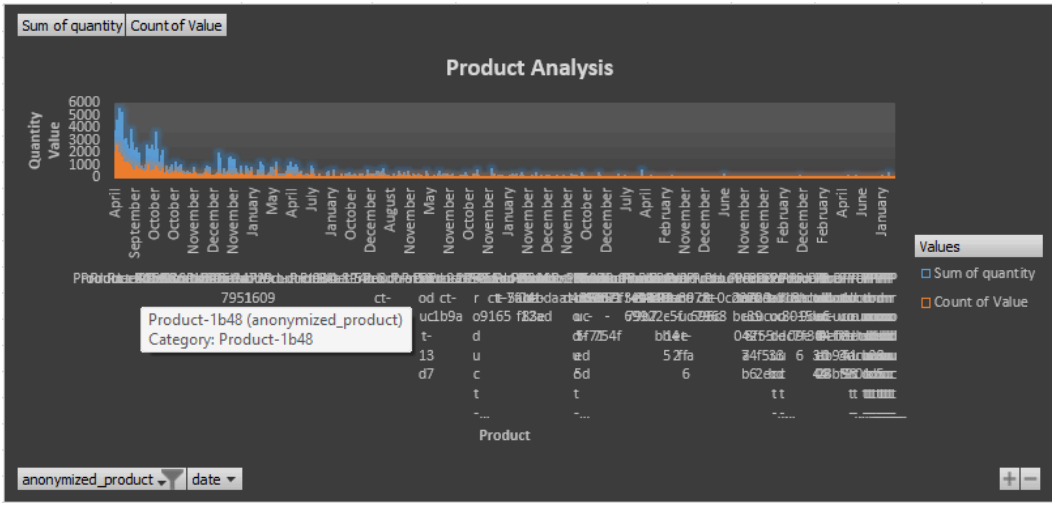


Chart 1.3 Performance analysis based on product value and purchase frequency

3.3.2 Key Objectives and Insights

Observation	Possible Causes	Implications

Top-selling product in terms of quantity and value	- High customer demand - Effective pricing strategy - Strong market positioning	- Prioritize inventory management and marketing for these products
Top-performing products but generate lower value per unit	- Lower price point - Higher sales volume but lower profitability	- Evaluate **price adjustments** or bundling strategies to increase revenue
Seasonal fluctuations in sales for top products	- Demand varies by season - Market shifts or external factors	- Forecast demand and adjust stock levels accordingly
Long tail of products with low sales	- Poor visibility or marketing - Niche or underperforming products	- Consider **discontinuing or revamping** slow-moving products

3.3.3 Recommendations for Improved Performance

- **Leverage Top Products for Growth**
    - Ensure adequate stock levels and **targeted marketing** for high-demand periods.
    - Consider **upselling or bundling strategies** to boost revenue per unit.
- **Optimize Pricing & Promotions for Mid-Performing Products**
    - Analyze price elasticity to determine **optimal pricing strategies**.
    - Introduce promotions or limited-time discounts during slow months.
- **Improve Visibility & Sales of Low-Performing Products**
    - Reevaluate product placement, **advertising, and distribution strategies**.
    - Consider removing low-demand products to **streamline inventory** and reduce holding costs.
- **Enhance Demand Forecasting & Inventory Management**
    - Use historical sales data to **predict demand and prevent stockouts or excess inventory**.
    - Align procurement with **seasonal sales trends**.

4. Advanced Analysis

4.1 Customer Segmentation

This analysis aims at the following goals;

- **Customer Retention:** Identifies valuable customers to prioritize.
- **Revenue Optimization:** Tailor's marketing efforts to increase customer spending.
- **Personalized Customer Engagement:** Designs campaigns specific to each segment's needs.

4.1.1 Overview of Customer Segmentation

This analysis categorizes customers into three groups based on their **purchasing value:**

- **High-Value Customers (7 Businesses)** – Contribute the highest revenue.
- **Medium-Value Customers (256 Businesses)** – Moderate revenue contribution.
- **Low-Value Customers (Remaining Businesses)** – Generate the least revenue.

By understanding these groups, businesses can **tailor engagement strategies, optimize resource allocation, and maximize revenue potential.**

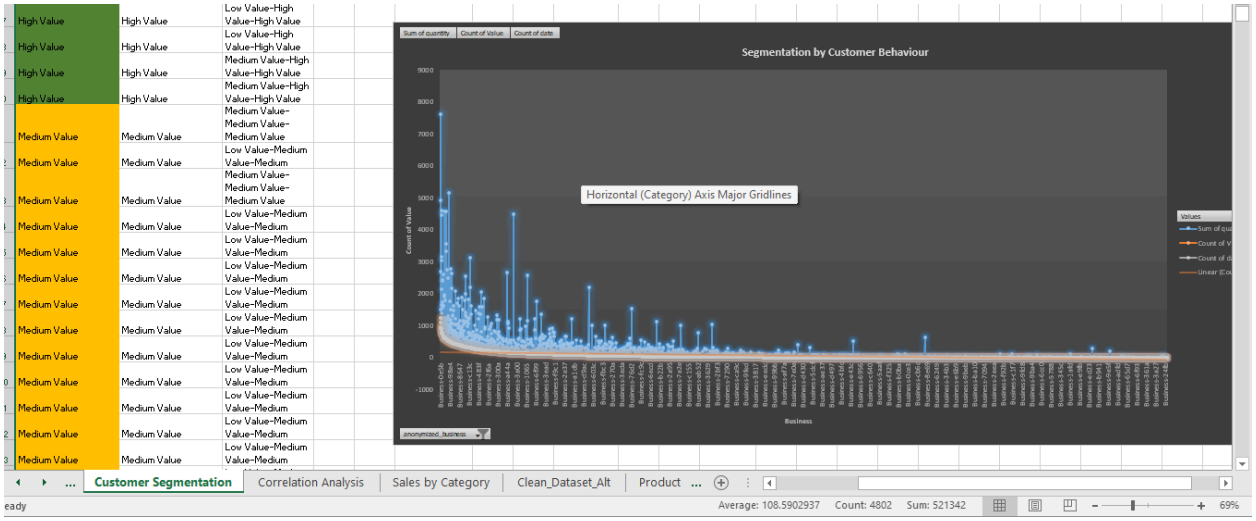


Chart 2.1 Customer Segmentation

4.1.2 Key Observations and Insights

Customer Segment	Observations	Possible Causes	Implications
High-Value (7 Businesses)	A small percentage (less than 3%) contributes the majority of revenue.	Strong brand loyalty, bulk purchases, or high demand for premium products.	Retaining these customers is **critical for business stability** and long-term profitability.
Medium-Value (256 Businesses)	A large segment with moderate revenue impact.	Price sensitivity, smaller or infrequent purchases.	Potential to **convert into high-value customers** with targeted engagement.
Low-Value (Remaining Businesses)	The majority of customers generate low revenue individually but significantly in total.	Small-scale businesses, lack of frequent orders, or poor brand engagement.	Requires strategies to **increase purchase frequency and order size**.

4.1.3 Engagement and Retention Strategies for Each Segment

- **High-Value Customers (VIP Businesses)**

Goal: Strengthen loyalty and increase lifetime value.

- ✔ Provide **exclusive deals**, early access to new products, and premium customer service.
- ✔ Assign **dedicated account managers** to ensure satisfaction.
- ✔ Offer **personalized discounts or loyalty rewards** for repeat purchases.
- ✔ Conduct **feedback sessions** to maintain strong relationships.

- **Medium-Value Customers**

Goal: Convert to high-value customers.

- ✔ Implement **targeted upselling and cross-selling strategies**.
- ✔ Offer **tiered loyalty programs** to encourage more purchases.
- ✔ Provide **personalized recommendations** based on past buying patterns.
- ✔ Create **special promotions or volume-based discounts**.

- **Low-Value Customers**

Goal: Increase purchase frequency and engagement.

- ✔ Run **email marketing campaigns** highlighting best-selling products.
- ✔ Offer **seasonal discounts or bundle deals** to boost spending.
- ✔ Provide **educational content or product demonstrations** to encourage usage.
- ✔ Engage via **social media and targeted ads** to increase brand awareness.

4.1.4 Recommendations for Business Growth

- **Focus on Retention for High-Value Customers** – Losing even one could significantly impact revenue.
- **Develop Incentive Programs for Medium-Value Customers** – Convert them into VIP clients.
- **Improve Accessibility & Engagement for Low-Value Customers** – Encourage frequent purchases.
- **Monitor Customer Behavior Trends** – Identify opportunities for personalized offers.

4.2 Forecasting

4.2.1 Overview of Forecast Study

This forecast study aims at;

- **Revenue Growth Strategy** – Helps anticipate and maximize sales opportunities.
- **Inventory & Supply Chain Optimization** – Reduces overstocking or shortages.
- **Data-Driven Decision-Making** – Informs marketing, pricing, and investment strategies.
- **Budgeting & Resource Allocation** – Aligns financial planning with expected revenue.

This forecast analysis suggests that **sales trends are following the same patterns as before but with higher yields.** This indicates:

- **Sustained market demand** for products.
- **Potential business growth** if trends continue.
- **Seasonal or periodic sales cycles** affect purchasing behavior.

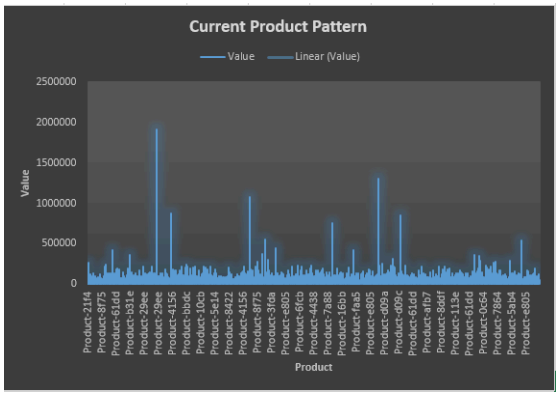


Chart 2.2.a Current Product Value Trend

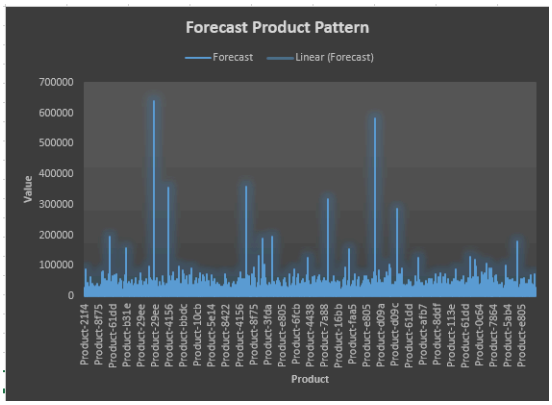


Chart 2.2.b Forecast Product Value Trend

4.2.2 Key Observations and Insights

Observation	Implication	Strategic Action
Same Sales Trend Patterns	Indicates a **predictable sales cycle** based on historical data.	Use historical trends to optimize **inventory management, pricing strategies, and marketing campaigns**.
Higher Yields (Increased Sales & Value)	Suggests **growth in demand**, improved pricing, or higher customer retention.	Identify factors driving growth and scale successful strategies.
Possible Seasonal Patterns	Indicates peaks and dips in sales at specific times.	Prepare for **high-demand periods** with inventory and marketing boosts, while optimizing costs in slow months.

4.2.3 Report on Forecast Results

Sales Forecast Analysis

Trends, Insights, and Strategic Recommendations

- **Executive Summary**
    - Provide a brief overview of sales trends and key findings.
    - Highlight the expected **increase in sales volume and revenue**.
- **Data Analysis & Methodology**
    - Explain how historical sales data was used to make projections.
    - Describe any forecasting techniques used (e.g., moving averages, regression analysis).
- **Key Findings**
    - **Overall growth trend:** Sales are expected to increase.
    - **Seasonality impact:** Identify peak and low sales months.
    - **Product performance:** Which products drive the most growth?

4.2.4 Recommendations for Business Strategy

- **Leverage Predictive Analytics** – Use AI/ML tools to refine forecasts.
- **Align Sales & Marketing Strategies** – Increase engagement during peak sales periods.
- **Ensure Scalable Operations** – Strengthen the supply chain to meet demand.
- **Monitor External Market Factors** – Consider economic or competitive shifts affecting sales.

4.3 Anomaly Detection

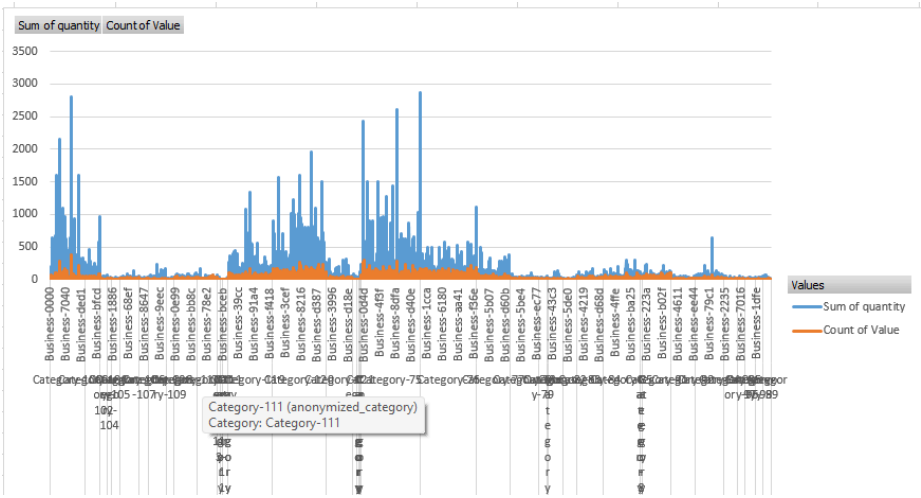


Chart 2.3 Anomaly Detection

4.3.1 Key Observations on Spikes and Drops

- Some businesses have extremely high purchase volumes.
- Other businesses have very low quantities.
- Some sharp declines after high spikes.

Possible Reasons for Spikes and Drops

- **Seasonality:** Certain businesses may purchase in bulk only at specific times
- **Promotional Campaigns:** Discounts, promotions, or bulk-buying incentives may have led to occasional spikes.
- **Stock Management Practices:** Businesses might follow different inventory replenishment strategies, leading to sharp increases when restocking.
- **Supply Chain Disruptions:** Drops could indicate supply chain issues, stockouts, or budget constraints.
- **Customer Preferences:** Certain product categories might be in higher demand for some businesses but not others.

4.4 Correlation Analysis: Quantity vs Value

This correlation study helps in;

- **Optimizing Product Pricing** – Adjust prices based on demand elasticity.
- **Enhancing Revenue Strategies** – Drive profit growth with bundled offerings.
- **Improving Customer Targeting** – Focus marketing on different customer segments.
- **Balancing Inventory Management** – Reduce overstock or stockouts based on demand patterns.
- **Refining Sales Forecasting** – Align future sales predictions with demand drivers.

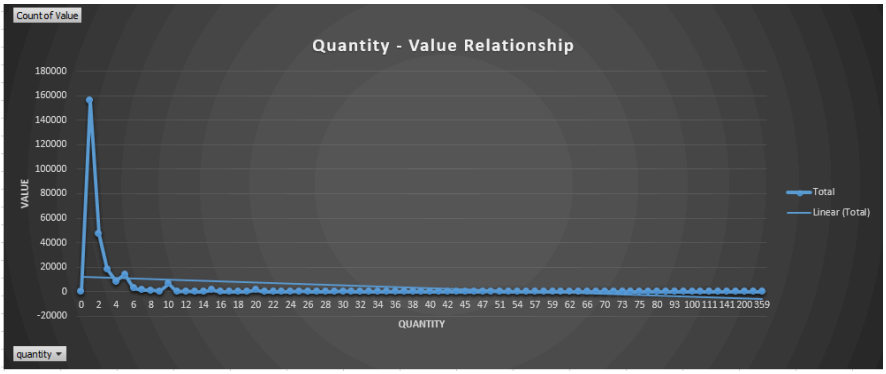


Chart 2.4 Correlation Analysis

4.4.1 Key Observations and Insights

From the visuals, **products with lower value are purchased in higher quantities, while high-value products have lower purchase volumes.** This pattern suggests:

- **Price Sensitivity** – Customers prefer buying more of cheaper products.
- **Demand Elasticity** – High-value products may have a niche market with selective buyers.
- **Bulk Purchase Trends** – Businesses may prefer stocking low-cost items in large quantities.

4.4.2 Factors Driving Sales Performance

Several factors influence the observed correlation between **Quantity & Value:**

Factor	Impact on Sales Performance
Price Elasticity	Lower-cost products have **higher demand**, while expensive products see lower but stable sales.
Market Segmentation	Different customer groups may prioritize affordability over premium features.
Consumer Purchasing Power	Buyers may opt for bulk purchases of affordable items instead of fewer expensive products.
Product Necessity vs. Luxury	Essentials (low-cost) sell more frequently, while premium items are **occasional purchases**.
Competitor Pricing	If competitors offer similar high-value products at lower prices, it can impact demand.
Promotions & Discounts	Lower prices from discounts can increase sales for both product categories.

4.4.3 Strategic Recommendations

Based on these insights, the following strategies can **optimize sales and revenue:**

Strategy	Implementation
Bundle High & Low-Value Products	Encourage sales of premium products by bundling them with frequently purchased low-cost products.
Tiered Pricing & Discounts	Offer bulk discounts for low-cost items and loyalty incentives for high-value purchases.
Targeted Marketing	Promote high-value products to businesses/customers with higher spending capacity.
Demand-Based Inventory Management	Ensure stock availability for high-demand low-value products while optimizing premium product supply.
Upselling & Cross-Selling	Suggest complementary purchases to drive additional revenue (e.g., accessories for high-value items).

5. Strategic Insights and Recommendations

5.1 Product Strategy

5.1.1 Prioritizing a Product Category for Marketing Campaigns

- **Recommended Product Category: High-Value Products**
  📌 Justification:

  ✔ **Higher Profit Margins**: Despite lower sales volume, high-value products contribute significantly to revenue.
  ✔ **Customer Loyalty Potential**: Businesses purchasing premium products often show repeat buying behavior.
  ✔ **Market Expansion Opportunity**: Targeting niche customers (e.g., high-spending businesses) can enhance long-term growth.

- **Marketing Strategies for High-Value Products:**
  - **Targeted Digital Ads & Retargeting** – Focus on high-spending customers with personalized recommendations.
  - **Exclusive Offers & Loyalty Incentives** – Provide discounts, early access, or VIP services for repeat buyers.
  - **Bundle with High-Quantity, Low-Value Items** – Encourage sales by offering complimentary add-ons.

5.2 Customer Retention

5.2.1 Identifying and Re-Engaging Inactive Businesses

📌 **Approach:** Use historical purchasing data to identify businesses that have reduced or stopped purchases over time.

- **Re-Engagement Strategies:**
  - **Personalized Outreach** – Send tailored emails, discounts, or product recommendations based on past purchases.
  - **Feedback Surveys** – Understand why purchase frequency declined and address customer concerns.
  - **Subscription or Auto-Reorder Options** – Offer recurring purchase plans for convenience.
  - **Reactivation Discounts** – Provide limited-time offers to encourage reorders.
  - **Dedicated Account Management** – Assign sales reps to high-value customers for personalized engagement.

5.3 Operational Efficiency

5.3.1 Inventory and Supply Chain Optimization

📌 **Challenges Identified:**

✔ **Seasonal Demand Fluctuations**: High sales volume in specific months.
✔ **Stock Imbalance**: Overstocking low-performing products, understocking high-demand items.
✔ **Inefficient Procurement Cycles**: Delays in restocking best-selling products.

- **Optimization Strategies:**
  - **Demand-Based Inventory Management** – Use predictive analytics to adjust stock levels dynamically.
  - **Supplier Collaboration & Forecasting** – Work with suppliers to ensure timely restocking based on demand trends.
  - **Automated Restocking Alerts** – Implement real-time tracking to prevent stockouts.
  - **Dynamic Pricing Strategy** – Adjust pricing based on demand seasonality.
  - **Warehouse & Logistics Optimization** – Reduce delivery lead times through route planning and better storage management.

5.3 Summary

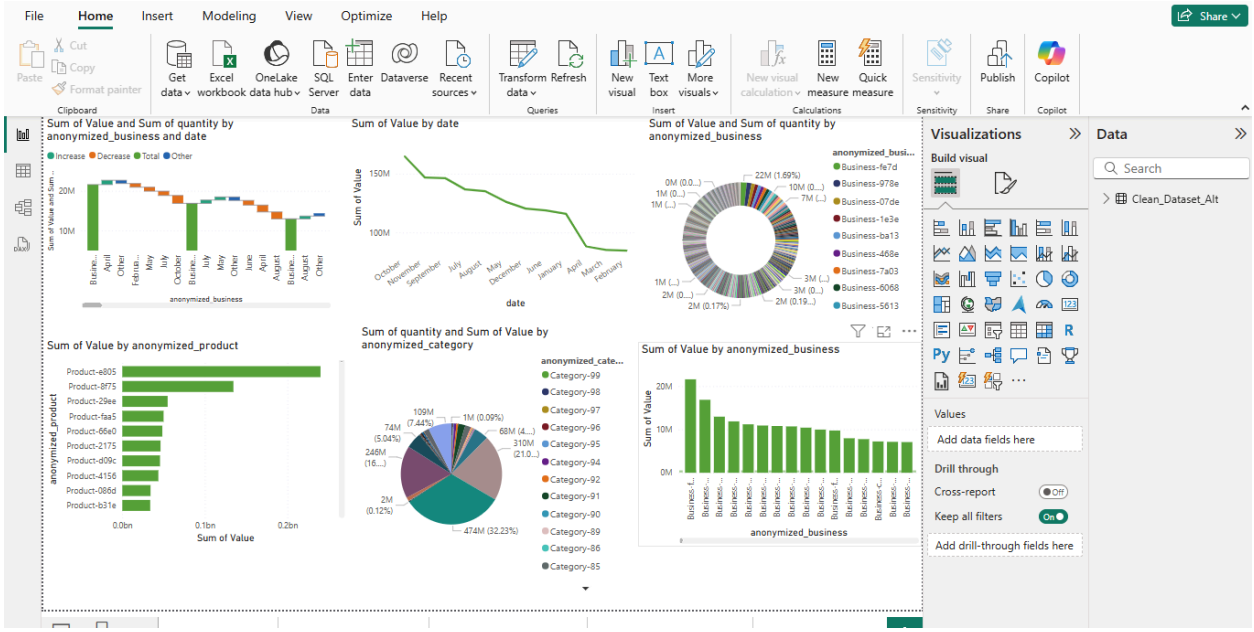
- Prioritize **high-value products** in marketing campaigns.
- Identify & re-engage **inactive businesses** with personalized strategies.
- Improve **inventory & supply chain management** to align with demand patterns.

---

6. Dashboard and Reporting

6.1 Dashboard Overview

This report summarizes key insights derived from the **Power BI interactive dashboard** developed to analyze sales performance, customer segmentation, and business trends. The dashboard provides a visual representation of total sales, product performance, customer categories, and time-series trends to support data-driven decision-making.



Screen-print 2.1

6.1.1 Sales Overview

- **Total Quantity and Value by Anonymized Category**: Highlights the best-performing product categories based on revenue and sales volume.
- **Top-performing Products and Businesses**: Identifies products and businesses contributing the most to overall sales.

6.1.2 Sales Trends Overtime

- The time-series analysis reveals **seasonal patterns**, with peak sales occurring during specific periods.
- A few **unusual spikes and drops** were noted, possibly due to promotional campaigns, external economic factors, or shifts in demand.
- **Recommendation**: Businesses should align marketing strategies with seasonal peaks and address low-performing months with targeted offers.

6.1.3 Customer Segmentation

- Businesses were classified into three groups based on purchasing behavior:
  - **High Value:** 7 businesses contributing significantly to sales.
  - **Medium Value:** 256 businesses with moderate transactions.
  - **Low Value:** The remaining businesses with minimal sales.
- **Recommendation**: Prioritize engagement with high-value customers through personalized marketing while nurturing medium-value customers with loyalty incentives.

6.1.4 Performance Analysis

- Some high-value products have **low sales volume**, while low-cost products are purchased in large quantities.
- **Recommendation**: Optimize pricing and promotional strategies for high-margin products to increase profitability.

6.1.5 Forecasting and Predictive Analysis

- A **3-month sales forecast** was generated using time-series models to anticipate future demand.
- **Recommendation**: Inventory management should be adjusted based on the forecast to minimize stockouts or overstocking.

6.1.6 Correlation Analysis

- A strong correlation exists between product price and purchase volume.
- **Recommendation**: Businesses should explore bundling strategies or discounts for high-volume, low-value products.

6.2 Dashboard Functionality

- **Slicers** enable filtering by category, business, and period.
- **Drill-through** provides detailed insights into business and product performance.
- **Tooltips and Labels** enhance clarity and data interpretation.
- **Interactive visuals** ensure dynamic data exploration for deeper insights.

6.3 Strategic Recommendations

1. **Product Strategy**: Focus marketing campaigns on the best-selling and high-margin product categories.
2. **Customer Retention**: Identify businesses with declining purchase frequency and re-engage them through targeted promotions.
3. **Operational Efficiency**: Improve inventory management based on seasonal demand fluctuations and sales forecasts.

6.4 AOB Section

6.4.1 Predictive Analysis

External Factors Influencing Sales;

1. **Economic Conditions:** Inflation, interest rates, and GDP growth impact consumer spending.
2. **Competitor Actions:** New product launches, pricing strategies, and marketing campaigns can affect sales.
3. **Seasonality:** Holidays, weather patterns, and special events influence demand.
4. **Supply Chain Disruptions:** Raw material shortages, logistics delays, and geopolitical issues affect availability.
5. **Market Trends & Consumer Behavior:** Shifts in customer preferences, technology adoption, and brand perception play a role.

Methodology to Incorporate External Factors:

1. **Data Collection:**
   - Use economic data from sources like the World Bank, IMF, and local government reports.
   - Track competitor pricing and promotions using web scraping or market research.
   - Integrate social media sentiment analysis for consumer behavior trends.
2. **Feature Engineering:**
   - Create new variables such as inflation-adjusted prices, sentiment scores, and competitor pricing indexes.
3. **Predictive Modeling:**
   - Implement **Time Series Forecasting (ARIMA, Prophet)** with external regressors.
   - Use **Machine Learning (Random Forest, XGBoost, LSTM)** to capture non-linear patterns.
   - Test model accuracy using cross-validation.
4. **Scenario Analysis:**
   - Simulate different conditions (e.g., price increase, supply chain issues) to measure potential sales impact.

6.4.2 Scalability

If the dataset grows **10x**, optimizations are needed for storage, processing, and analysis:

Data Storage & Management:

- **Cloud Storage (AWS S3, Azure Blob Storage):** Store large datasets efficiently.
- **Database Optimization:** Use a **columnar database** (BigQuery, Redshift) for faster querying.
- **Partitioning & Indexing:** Improve retrieval speed by splitting data into logical partitions (e.g., by date, product category).

Processing & Performance:

- **Distributed Computing (Spark, Dask):** Scale processing across multiple nodes.

- **ETL Optimization:** Streamline Extract, Transform, and Load (ETL) pipelines using parallel processing.
- **Pre-Aggregated Tables:** Store summary tables for frequent queries instead of recalculating from raw data.

Analysis & Visualization:

- **Power BI DirectQuery Mode:** Connect directly to large databases instead of importing data into memory.
- **Data Sampling:** Use representative samples for exploratory analysis before running full queries.
- **AI & Automation:** Implement auto-refreshing dashboards and automated anomaly detection.

7. Conclusion

This report provides an overview of the data analysis and visualization process, from data cleaning and exploratory data analysis (EDA) to advanced insights and dashboard reporting using Power BI.

7.1 Data Cleaning and Preparation

Data cleaning involved handling missing values, correcting data types, and standardizing formats for analysis. The Date column was formatted to Month-Year (MMM-YYYY) for effective trend analysis. Duplicate entries were removed, and inconsistencies in categorical data were resolved to ensure accuracy.

7.2 Exploratory Data Analysis (EDA) Report

EDA revealed key insights into sales performance, customer segmentation, and product trends. The analysis included:

- **Top-performing products** based on quantity and value.
- **Sales trends over time** to identify seasonal patterns and anomalies.
- **Customer segmentation** into High, Medium, and Low-Value groups based on purchasing behavior.

7.3 Advanced Analysis

Deeper analysis provided strategic insights, including:

- **Correlation Analysis:** Products with lower unit values tend to have higher purchase quantities, while high-value products are purchased in smaller volumes.
- **Forecasting:** Sales trends were projected for the next three months using time-series analysis, helping in inventory and marketing planning.
- **Strategic Insights:** Recommended priority product categories, customer retention strategies, and operational improvements.

7.4 Dashboard and Reporting

An interactive Power BI dashboard was created to visualize key insights, featuring:

- **Total Quantity and Value by Category**
- **Top-performing products and businesses**
- **Sales trends over time**
- **Customer segmentation overview**

Slicers and drill-throughs were implemented for enhanced interactivity, allowing users to filter and explore specific insights efficiently. Visuals were formatted for clarity, with labels and tooltips to enhance user experience.

7.5 Conclusion

The analysis provided valuable business insights, supporting data-driven decision-making. Power BI enabled a seamless presentation of findings, aiding in strategic planning and operational efficiency. Future recommendations include incorporating external market factors for predictive analysis and optimizing data processing for scalability as the dataset grows. This report serves as documentation for the entire data analysis and visualization workflow.

Appendix

A. Code Repository

- All processing scripts are available in: Colab Notebook
- Full documentation of functions and methodologies included
- The Decoded Dataset: Decoded Dataset
- Visualizations on an interactive dashboard using Power BI: Kwanza Tukule Interactive Dashboard