

KWANZA TUKULE DATA ANALYSIS REPORT

By Anastacia Nyamoita (anastacianyamoita@gmail.com)

1. i) Data Cleaning and Preparation

- Duplicate Records:** Identified and removed 3,524 duplicate records to maintain data accuracy.
- Missing Values:** Detected 8 missing values in the "UNIT PRICE" column, which were dropped due to their minimal impact.
- Data Type Conversion:** Converted "UNIT PRICE" from 'Object' to 'Float' for numerical analysis.

ii) Feature Engineering: Create the following columns: "Month-Year".

```
Dataset with Month-Year Column:
      DATE      Month-Year
0 2024-08-18 21:32:00 August-2024
1 2024-08-18 21:32:00 August-2024
2 2024-08-18 21:32:00 August-2024
3 2024-08-18 21:32:00 August-2024
4 2024-08-18 21:32:00 August-2024
```

2. Exploratory Data Analysis (EDA)

Sales Overview

i) By Business

The top-performing businesses based on total sales value:

Rank	Business ID	Total Sales Value	Percentage Contribution (%)
1	Business-978e	28,037,358.0	1.58%
2	Business-fe7d	26,997,121.0	1.53%
3	Business-6068	16,464,195.0	0.93%
4	Business-07de	16,258,068.0	0.92%
5	Business-7a03	13,968,451.0	0.79%

Top Performing Products

Most Frequently Purchased Products

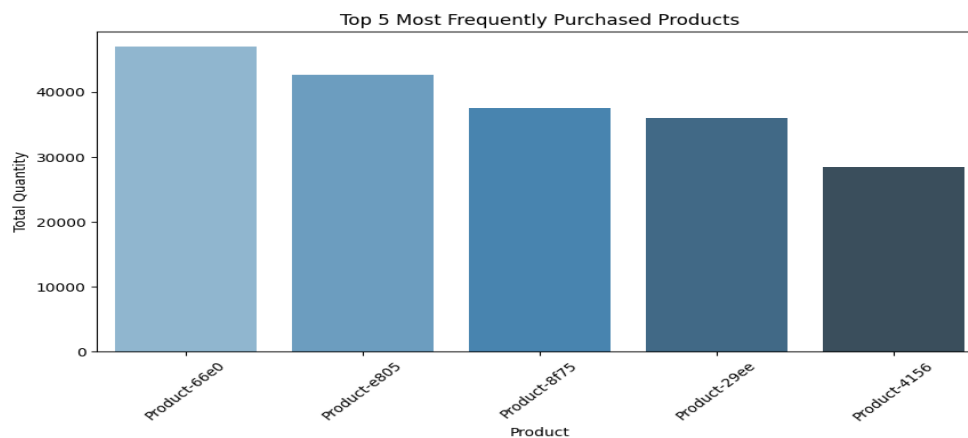
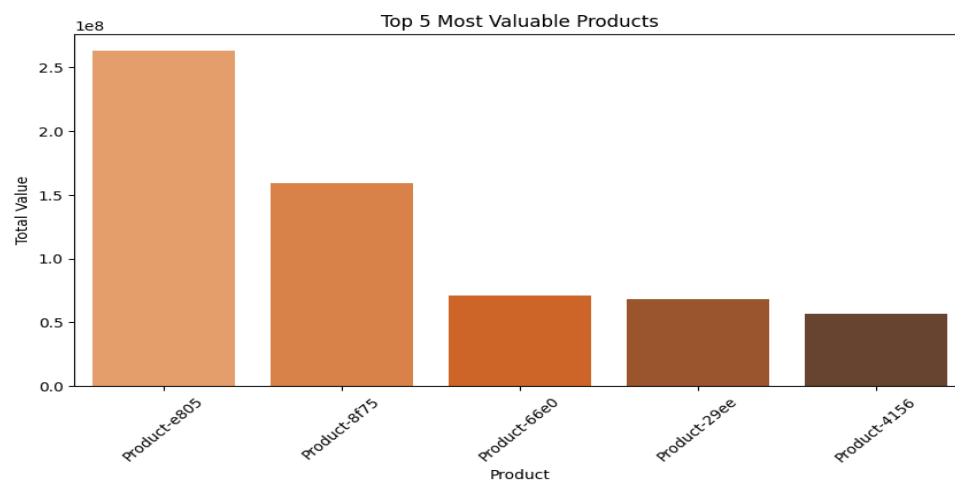
Rank	Product ID	Total Quantity	Total Value
1	Product-66e0	46,957	70,704,225.0
2	Product-e805	42,602	262,787,281.0
3	Product-8f75	37,566	158,797,460.0
4	Product-29ee	35,940	68,248,274.0
5	Product-4156	28,487	56,956,007.0

Most Valuable Products

Rank	Product ID	Total Sales Value
1	Product-e805	262,787,281.0
2	Product-8f75	158,797,460.0
3	Product-66e0	70,704,225.0
4	Product-29ee	68,248,274.0
5	Product-4156	56,956,007.0

- **Top Performing Products:**

- **Most Frequently Purchased:** "Product-66e0" was the most frequently purchased product, followed by "Product-e805" and "Product-8f75." (See table provided)
- **Most Valuable:** "Product-e805" generated the highest total sales value, followed by "Product-8f75" and "Product-66e0." (See table provided)
- **Graph / Visualizations of the two provided below;**

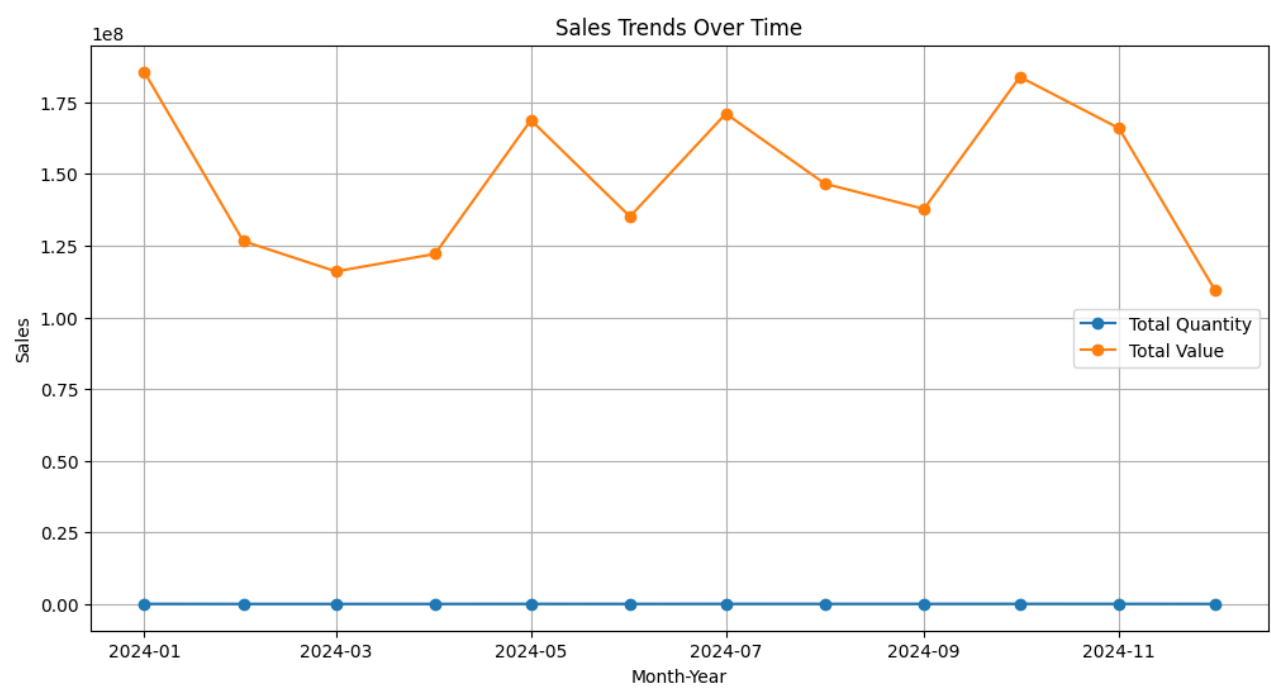


SALES TRENDS OVER TIME

Month-Year	Sales Value
January-2024	185,626,186.0
February-2024	126,579,702.0
March-2024	116,000,676.0
April-2024	122,110,750.0
May-2024	168,781,502.0
June-2024	135,138,614.0
July-2024	171,040,881.0
August-2024	146,581,003.0
November-2024	165,933,104.0
December-2024	109,549,774.0

Key Observation as shown in the graph below of Sales Trend Over Time;

- Sales show significant dips in **March, November, and December 2024**.



3. ADVANCED ANALYSIS.

. Advanced Analysis

- **Customer Segmentation:** Businesses were segmented into three groups based on their purchasing behaviour (Total Quantity, Total Value, and Transaction Frequency):
 - **High Value:** Businesses with high total sales value and frequent transactions.
 - **Medium Value:** Businesses with moderate sales value and transaction frequency.
 - **Low Value:** Businesses with low sales value and infrequent transactions.
 - (See provided segmentation table)

Business ID	Total Quantity	Total Sales Value	Transaction Frequency
Business-00008	1,044	45.08	Medium Value
Business-00051	2,645	512.01	Low Value
Business-00292	6,773	40.06	Medium Value
Business-003d9	8,221	761.03	Medium Value
Business-00721	12,722	5056.01	Medium Value

Sales Forecasting:

Sales forecasting was conducted using both ARIMA and Exponential Smoothing models and both had the same outputs:

Month-Year	Forecasted Sales Value
January-2025	126,506,000.0
February-2025	121,406,700.0
March-2025	122,940,200.0

Forecast Insights:

- Projected **decline in sales** in early 2025, necessitating proactive intervention.
- Promotional efforts and demand stimulation strategies should be explored.

ANOMALY DETECTION.

Attached Graph Interpretation.

Blue Line (Sales Value): Represents the monthly sales performance. Significant dips are observed in March, November, and December.

Month	Sales Value
November-2024	110,000,000.0

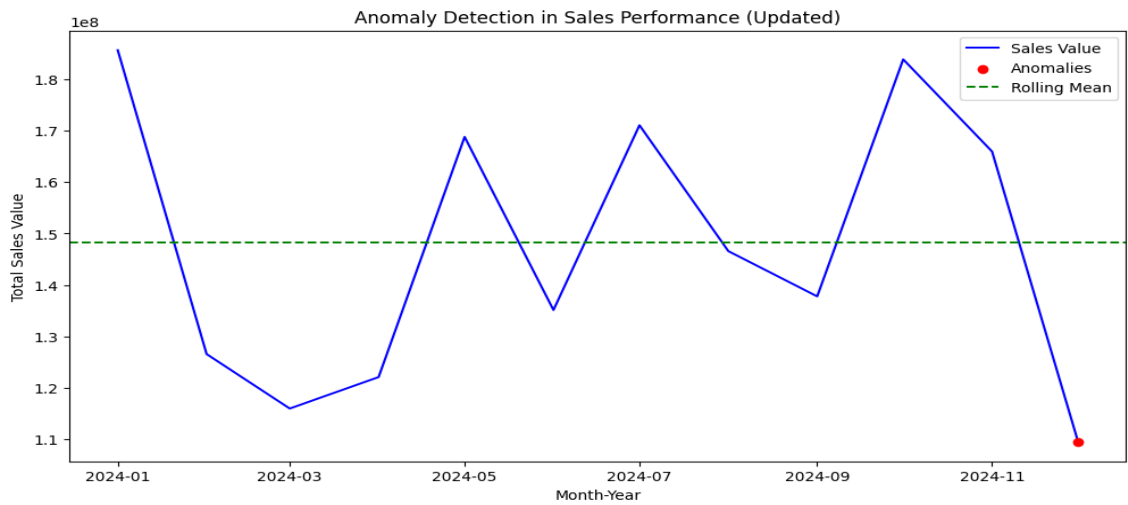
December-2024	109,549,774.0
---------------	---------------

Green Dashed Line (Rolling Mean): Shows the smoothed trend (approximately 1.5e8) that helps detect deviations from typical sales performance.

Red Dots (Anomalies): Two anomalies are detected: November 2024: Sales value drops to 110,000,000. December 2024: Sales value decreases further to 109,549,774.

Interpretation with December Data: Two Significant Anomalies:

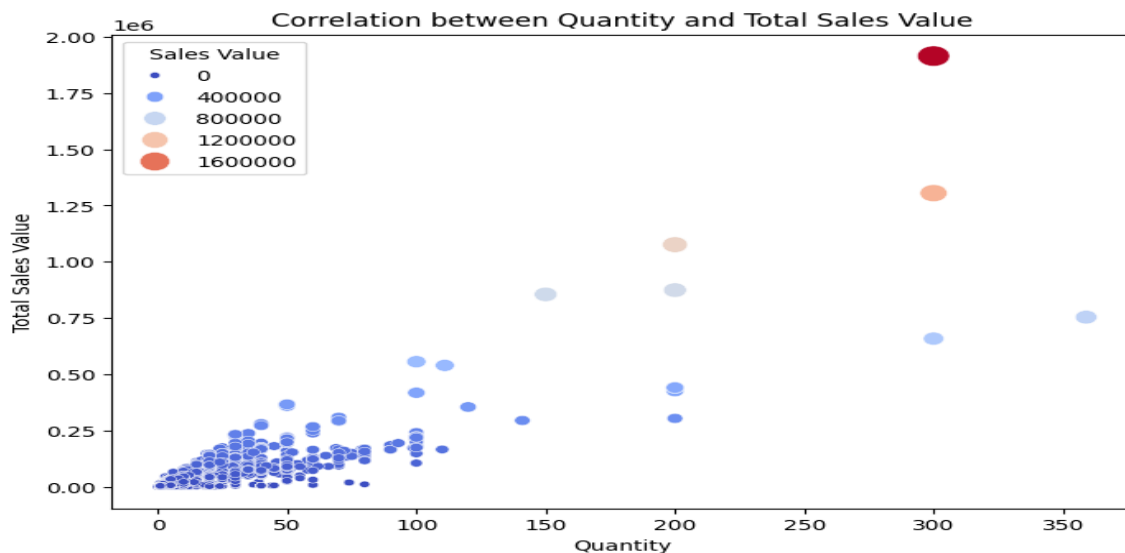
1. November 2024: The first anomaly is a sharp decline from earlier values.
2. December 2024: A continued drop below the rolling mean and anomaly detection threshold.



Recommendations:

1. Analyze operational issues, demand changes, or market conditions. Explore whether the anomalies are seasonal or one-off events. Plan Strategic Interventions:
2. Develop strategies to prevent a prolonged decline in sales (e.g., promotions, process improvements).

CORRELATION ANALYSIS



Observations:

Positive Correlation: As the Quantity increases, there's a general upward trend in the Total Sales Value. This indicates a positive correlation between the two variables.

Non-Linearity: The relationship doesn't appear to be perfectly linear. There's some curvature in the data points, suggesting that the increase in Total Sales Value might not be directly proportional to the increase in Quantity.

Variability: There's significant variability in Total Sales Value for a given Quantity. This implies that other factors besides Quantity are influencing the sales value.

Interpretation:

1. The positive correlation suggests that higher quantities sold generally lead to higher total sales revenue. This is intuitive, as selling more units naturally results in more revenue.
2. The non-linearity indicates that the relationship between Quantity and Total Sales Value might not be as simple as a straight line. Factors like pricing, discounts, product mix, and market demand could be influencing the sales value in ways that aren't captured solely by the quantity sold.
3. The variability highlights the presence of other factors that affect total sales revenue. These factors could include:
 4. **Product Profitability:** Some products might have higher profit margins than others, influencing the overall sales value despite similar quantities sold.
 5. **Pricing Strategies:** Discounts, promotions, and price changes can significantly impact the sales value per unit.
 6. **Market Demand:** Seasonal fluctuations, competitor actions, and overall market conditions can affect sales volume and revenue.

4. STRATEGIC INSIGHTS & RECOMMENDATIONS

Product Strategy

- **Marketing Focus:** Invest in targeted campaigns for high-growth categories such as **Product-66e0** (Total Sales: Ksh70,704,225, Total Quantity: 46,957) and **Product-e805** (Total Sales: Ksh262,787,281, Total Quantity: 42,602). These products are high-performers and should be promoted through digital advertising and seasonal discounts.
- **Stock Optimization:** Ensure optimal stock levels for top-selling products like **Product-8f75** (Total Quantity: 37,566) to prevent stockouts, especially during peak demand months like **May** whose Total Sales were KSH168,781,502) and **July** (Total Sales: KSH171,040,881).

Customer Retention

- **Loyalty Programs:** Introduce tiered rewards for businesses with high transaction frequency, such as **Business-978e** (Total Sales: Ksh28,037,358, 1.58% of total revenue) and **Business-fe7d** (Total Sales: Ksh26,997,121, 1.52% contribution). Reward bulk purchases with discounts, cashback, or early access to new stock.
- **Personalized Marketing:** Use purchasing trends to tailor product recommendations. For example, businesses purchasing **Product-4156** (Total Sales: Ksh56,956,007, Total Quantity: 28,487) could receive cross-selling offers for complementary products in the same category.

Inventory Management

- **Just-In-Time Inventory Strategy:** Implement JIT for high-demand products like **Product-e805** and **Product-8f75**, ensuring efficient storage while minimizing holding costs.
- **Regular Stock Audits:** Identify and phase out slow-moving items that are not contributing significantly to revenue. Products with lower transaction frequency should either be bundled into promotions or replaced with better-performing alternatives.

Operational Adjustments

- **Investigate Q4 Sales Anomalies:** Sales anomalies in **November 2024 (Ksh165,933,104)** and **December 2024 (Ksh109,549,774)** indicate potential disruptions. Evaluate:
 - ✓ Supply chain issues causing stockouts.
 - ✓ Economic factors like inflation affecting purchasing power.
 - ✓ Competitor price wars drawing customers away.
- **Promotional Campaigns:** Implement time-limited discounts in historically slow months like **March (Total Sales: Ksh116,000,676)** and **November**, where a notable sales decline was detected. Possible promotions:
 - ✓ **Flash Sales** on best-sellers to drive volume sales.

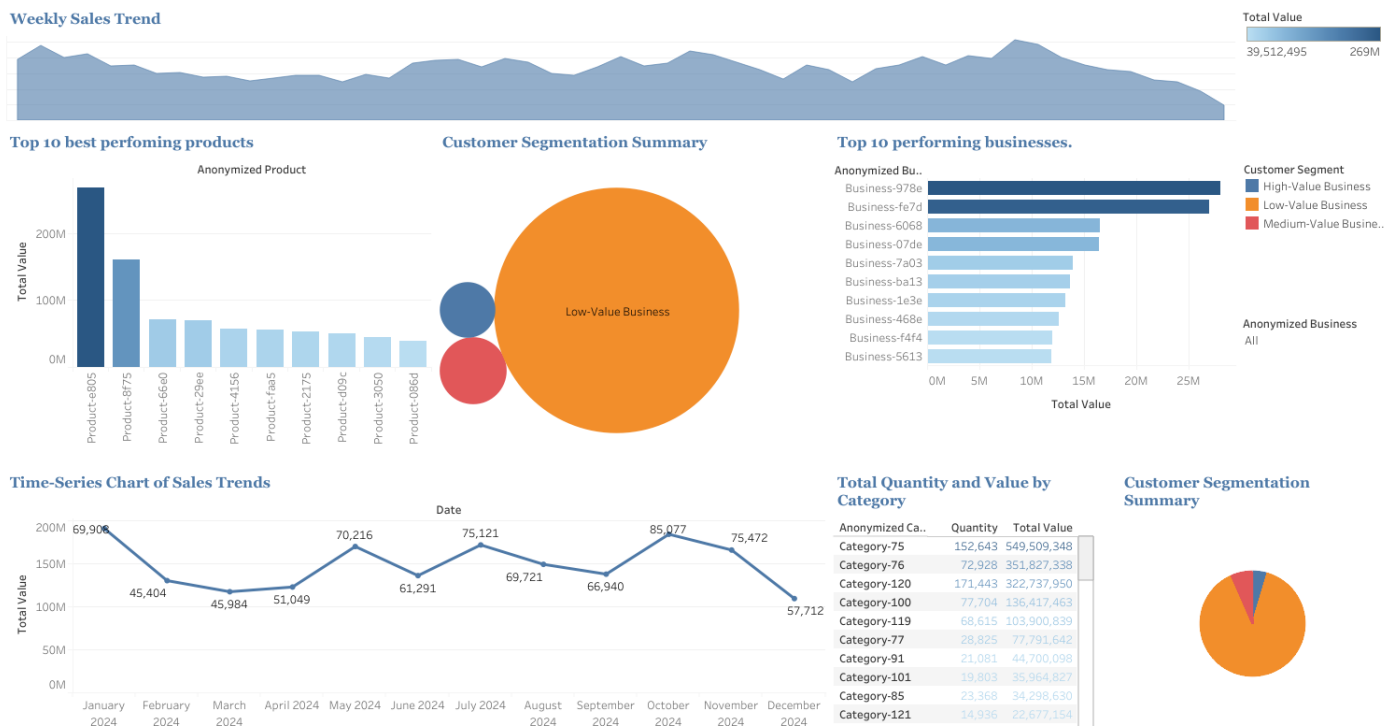
5. DASHBOARD AND REPORTING

- An interactive dashboard was created using Tableau to visualize key insights. The dashboard includes:
 - ✓ Total Quantity and Value by Category
 - ✓ Top-performing products and businesses
 - ✓ Time-series chart of sales trends
 - ✓ Customer segmentation summary

Tableau Link:

https://public.tableau.com/views/SalesRevenueofBusinessesandProducts/Dashboard3?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Dashboard Image below;



6. Bonus Section

- **Predictive Analysis with External Factors:** To enhance predictive analysis, external factors such as economic indicators, competitor actions, and industry trends should be incorporated into the models.
- **Scalability Considerations:** For larger datasets, implementing a cloud-based data warehousing solution and utilizing distributed computing frameworks can improve data storage, processing, and analysis efficiency.

Prepared by: Anastacia Nyamoita

Email: anastacianyamoita@gmail.com