

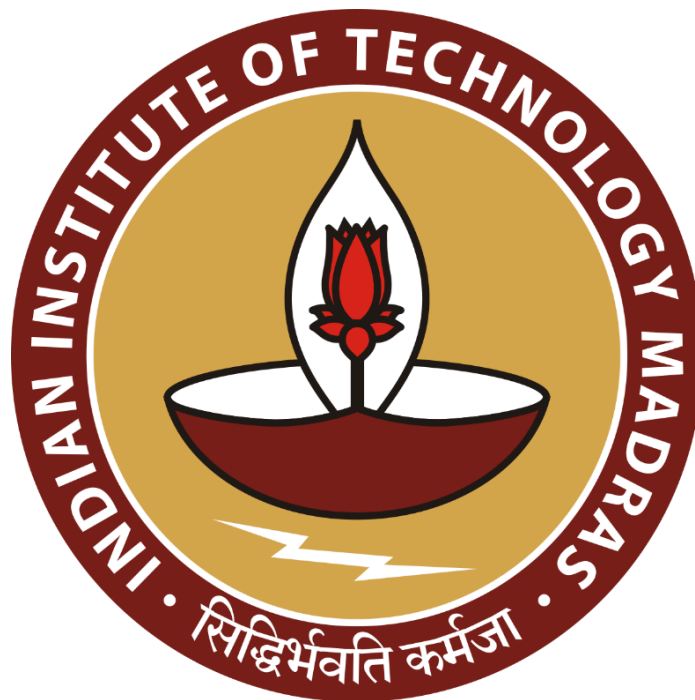
Uncovering Growth Opportunities for Kumar Computers

Final report for the BDM capstone Project

Submitted by

Name: Rajarshi Roy

Roll Number: 21F3000348



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Contents

1	Executive Summary	3
2	Detailed Explanation of Analysis Process/Method	4
3	Results and Findings	8
4	Interpretation of Results and Recommendation	17

1. Executive Summary

The report presents a detailed analysis aimed at uncovering growth opportunities for Kumar Computers, focusing on the company's sales and purchase data for the FY 2022-2023. The analysis began with a comprehensive data cleaning process. Raw sales and purchase data were meticulously organized which was crucial in setting the foundation for the subsequent analyses.

The first major analysis conducted was a time series analysis, aimed at identifying sales trends over the year and forecasting future sales. This analysis involved plotting a time series graph using the monthly sales data, and calculating key metrics such as the 3-month Simple Moving Average. The sales forecast generated through this analysis closely matched the actual sales trends, underscoring the model's effectiveness in predicting future sales patterns.

Next, Pareto analysis was conducted to identify the most impactful products contributing to the majority of sales and purchases. This analysis revealed that the top 20% of products accounted for approximately 80% of the total sales. The analysis also highlighted some discrepancies between sales and purchase data, indicating potential issues with overstocking or understocking.

The report examined the best-selling products to assess the efficiency of inventory management. The analysis identified the top 20 most sold and purchased products, showing a strong correlation between the number of pieces purchased and sold. The analysis revealed some discrepancies, where certain products had slightly higher purchases than sales, indicating a risk of overstocking, while others had more sales than purchases, posing a risk of stockouts.

The report analyzed the company's monthly expenses to understand the overall cost structure. This expense analysis provided valuable insights into where the company's financial resources are being allocated and highlights areas that may require more focused cost management.

Overall, the report provides a thorough examination of Kumar Computers' sales dynamics, inventory management, and cost structure, offering a clear understanding of the factors driving the company's performance over the fiscal year. The analyses conducted have uncovered key insights into sales trends, product performance, and expense distribution, which are crucial for informed decision-making and strategic planning.

2. Detailed Explanation of Analysis Process/Method

A required amount of data was cleaned previously for midterm report to find insights on sales and purchase data. For further analysis, the data needed more cleaning and modification according to the needs of the analysis methods.

Raw Purchase & Sales Data

The raw purchase & sales data was organized with proper column labels and correct categorization of data. The collected data was for FY 22-23 and had the following columns:

Date	Particulars	Vch Type(Sales)	Vch No	Debit	Credit
------	-------------	-----------------	--------	-------	--------

The columns “Pieces” and “Cost per Item” had a blank heading but the column values were indicative of the names. A new column was created by finding the product of the two columns i.e “Pieces” and “Cost per Item”. Then the data was sorted in descending order to get an idea of which were the most sold and most bought products by the company.

Analysis Tools Used

Data was cleaned in Google Sheets using filters. Pandas was used to import the cleaned data into Google Colab. Analysis dependent dataframes were created for the required analysis to be performed. Matplotlib and Seaborn were used for the visualizations.

Analysis Methods

A. Time Series Analysis

A time-series analysis is done not just for visualizing the sales data but is also essential for forecasting because it will allow Kumar Computers to leverage historical data to predict future trends and patterns. This insight is critical for making informed decisions for better inventory management and budgeting.

Data Modification for Time Series Analysis

To perform the time-series analysis, the “Date” and “Credit” column was used. From the “Date” column, the month was extracted and made into a new column. The final cleaned sheet then had “Date”, “Month” and “Credit” as the columns. This was done to club the sales amount for each month from which a time-series graph could be created.

Process of Time-Series Analysis of Sales Data

Step 1: Plotting the Initial Time-Series graph

1. **Date Formatting:** The month data was converted into a datetime object to facilitate time series analysis.
2. **Categorical Ordering:** The months were ordered sequentially to ensure accurate plotting on a time scale.
3. **Calculation of Key Metrics:** Key metrics such as the total sales for each month, simple moving average (SMA) over a 3-month period were calculated to analyze trends.
4. **Plot Creation:** Using pandas and matplotlib, a time series plot was created, featuring the actual sales data and the 3-month SMA.
5. **Visualization:** The plot was customized to include appropriate labels, legends, and grid lines for better interpretability.

Step 2: Forecasting the Sales Data

1. **Data Creation:** A DataFrame was recreated with monthly data and corresponding 'SUM of Credit' values.
2. **Seasonal Factor:** A simplified seasonal factor was computed using a rolling mean of the ratio of actual values to the overall mean.
3. **Forecasting :** The next 6 months were forecasted using the overall mean and the smoothed seasonal factors.
4. **Plot Setup:** A time series plot was prepared using matplotlib to display both the original data and the forecasted values.
5. **Superimposing Forecast:** The forecasted values were plotted as red markers and a dashed line on the same original graph acting as the baseline.

```

import pandas as pd
import matplotlib.pyplot as plt

data = {
    'Month': ['Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec', 'Jan', 'Feb', 'Mar'],
    'SUM of Credit': [220441.00, 186155.00, 61174.00, 89426.00, 817895.01, 152670.00, 560363.00,
                    122680.00, 504188.00, 285758.00, 139390.00, 220630.00]
}

df = pd.DataFrame(data)

month_order = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']
df['Month'] = pd.Categorical(df['Month'], categories=month_order, ordered=True)
df = df.sort_values('Month')
df.set_index('Month', inplace=True)

df['3-Month SMA'] = df['SUM of Credit'].rolling(window=3).mean()
df['Monthly Change (%)'] = df['SUM of Credit'].pct_change() * 100

plt.figure(figsize=(10, 6))
plt.plot(df.index, df['SUM of Credit'], marker='o', label='SUM of Credit')
plt.plot(df.index, df['3-Month SMA'], marker='o', linestyle='--', label='3-Month SMA')
plt.title('Time Series Plot of SUM of Credit')
plt.xlabel('Month')
plt.ylabel('SUM of Credit')
plt.legend()
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

```

Figure 1: Python Code for Time Series Analysis

```

import pandas as pd
import matplotlib.pyplot as plt

# Load the specific sheet 'Purchase Categorized' from the Excel file
file_path = "/kaggle/input/sales-categorized/KC-2022-23 SALES REGISTER.xlsx"
purchase_categorized_data = pd.read_excel(file_path, sheet_name='Sales Categorized')

# Extract the first row as the sales data for each category
category_sales_data = purchase_categorized_data.iloc[0, 1:-1] # Exclude the first column and the last one

# Convert the data to numeric and sort it
category_sales_data = pd.to_numeric(category_sales_data, errors='coerce').sort_values(ascending=False).dropna()

# Calculate cumulative percentage
cumulative_percentage = category_sales_data.cumsum() / category_sales_data.sum() * 100

# Create the Pareto chart
fig, ax = plt.subplots(figsize=(10, 6))

category_sales_data.plot(kind='bar', ax=ax, color='C0')
cumulative_percentage.plot(kind='line', marker='o', ax=ax, color='C1', secondary_y=True)
ax.set_xticklabels(category_sales_data.index, rotation=90)

ax.set_xlabel('Category')
ax.set_ylabel('Sales')
ax.right_ax.set_ylabel('Cumulative Percentage')
ax.right_ax.axhline(y=80, color='C2', linestyle='--')

# Find the category closest to 80% cumulative percentage
closest_category_index = (cumulative_percentage - 80).abs().idxmin()

# Find the x position of this category in the bar chart
x_position = list(category_sales_data.index).index(closest_category_index)
ax.axvline(x=x_position, color='C2', linestyle='--')

plt.title('Pareto Chart for Sales Categories')
plt.show()

```

Figure 2: Python Code for Pareto Analysis

B. Pareto Analysis

This analysis was done to identify the most impactful products contributing to the sales and revenue. By applying the 80/20 rule or the Pareto principle, the company can focus on the top 20% of products that contribute to 80% of the sales, enabling them to optimize inventory and prioritize key customer needs more effectively.

Data Modification for Pareto Analysis

In order to perform the Pareto analysis, the individual sales and purchase data for each product was needed. “Particulars”, “Vch Type(Sales)” were renamed to “Products” and “Sum of Sales”. For both sales and purchase data the list of products were then classified into different categories based on the type of product.

Process of Pareto Analysis of Sales and Purchase Data

Step 1: Data Preparation and Preprocessing

1. **Product Filtering:** From the original sales data, all the products were extracted using filters methods in Google Sheets.
2. **Amount validation:** The sales data was converted to numeric values to handle any potential non-numeric entries, and then sorted in descending order to prioritize the highest-selling categories.

3. **Data Import:** The sales data is imported from an Excel file named 'Final Products Categorized.xlsx' using pandas.

Step 2: Cumulative Percentage Calculation

1. **Cumulative Sum Calculation:** The cumulative sales for each category were calculated, to later plot the cumulative contribution of each category to the total sales.
2. **Percentage Conversion:** These cumulative sales were then converted into percentages of the total sales, providing a cumulative percentage for each category.

Step 3: Plotting the Pareto Chart

1. **Chart Setup:** A Pareto chart was plotted using matplotlib, with two y-axes: one for the sales data and the other for the cumulative percentage.
2. **Axis Labeling:** Labels are set for both x and y axes, including the secondary y-axis for cumulative percentage.
3. **Line Plot Overlay:** The cumulative percentage was plotted as a line chart on a secondary y-axis, overlaying the bar chart to visualize the cumulative impact of categories.

Step 4: Identifying the 80% Contribution Threshold

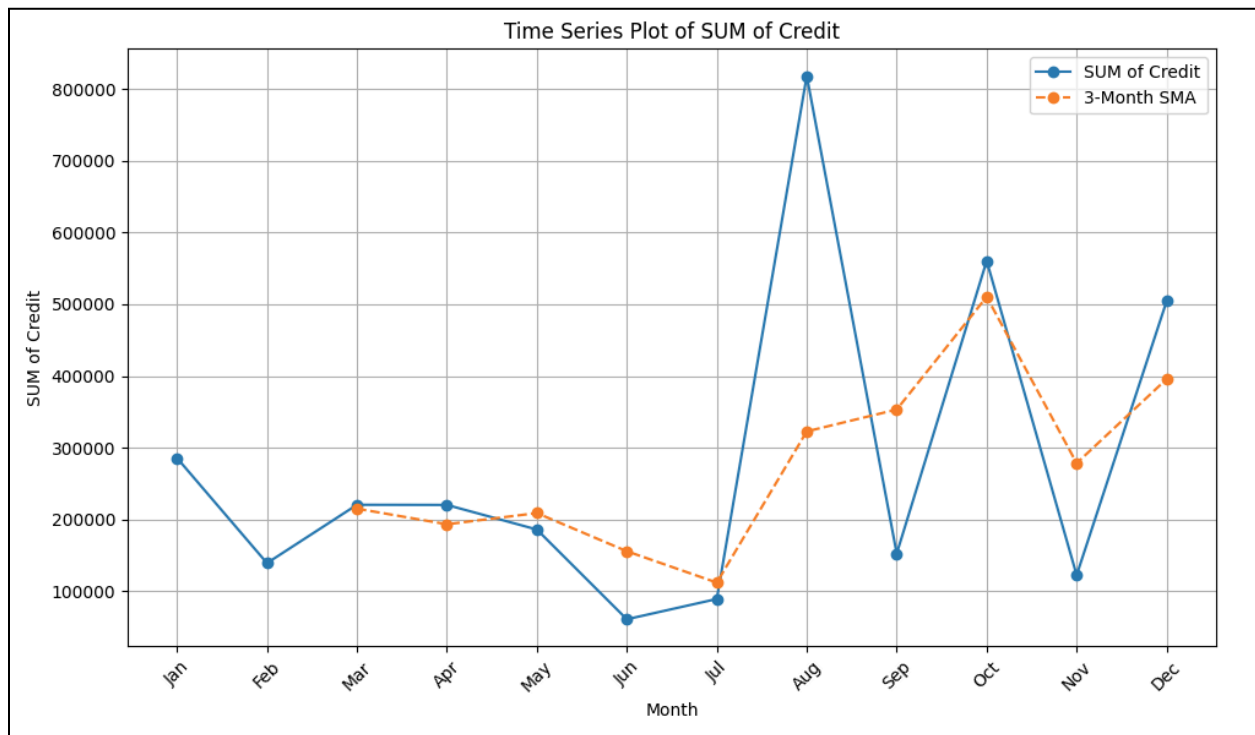
1. **80% Line:** A horizontal dotted line was drawn at the 80% cumulative percentage level, representing the typical threshold used in Pareto analysis to identify the most impactful categories.
2. **Category Identification:** The closest category to the 80% threshold was identified by finding the point where the cumulative percentage is closest to 80%. This category was found to be “Printers” which marked the transition from the high-impact to lower-impact categories.
3. **Vertical Line:** A vertical dotted line was drawn at the category “Printer”, visually marking the point on the bar chart where the cumulative percentage reaches approximately 80%.

3. Results and Findings

Time-Series Analysis

A. Plot of Sum of Credit

The time-series plot of the sales data for Kumar Computers, represented by the "SUM of Credit," provides valuable insights into the company's sales trends over a 12-month period. The data is further enhanced by the inclusion of a 3-Month Simple Moving Average (SMA), which smooths out short-term fluctuations and highlights the underlying trend.



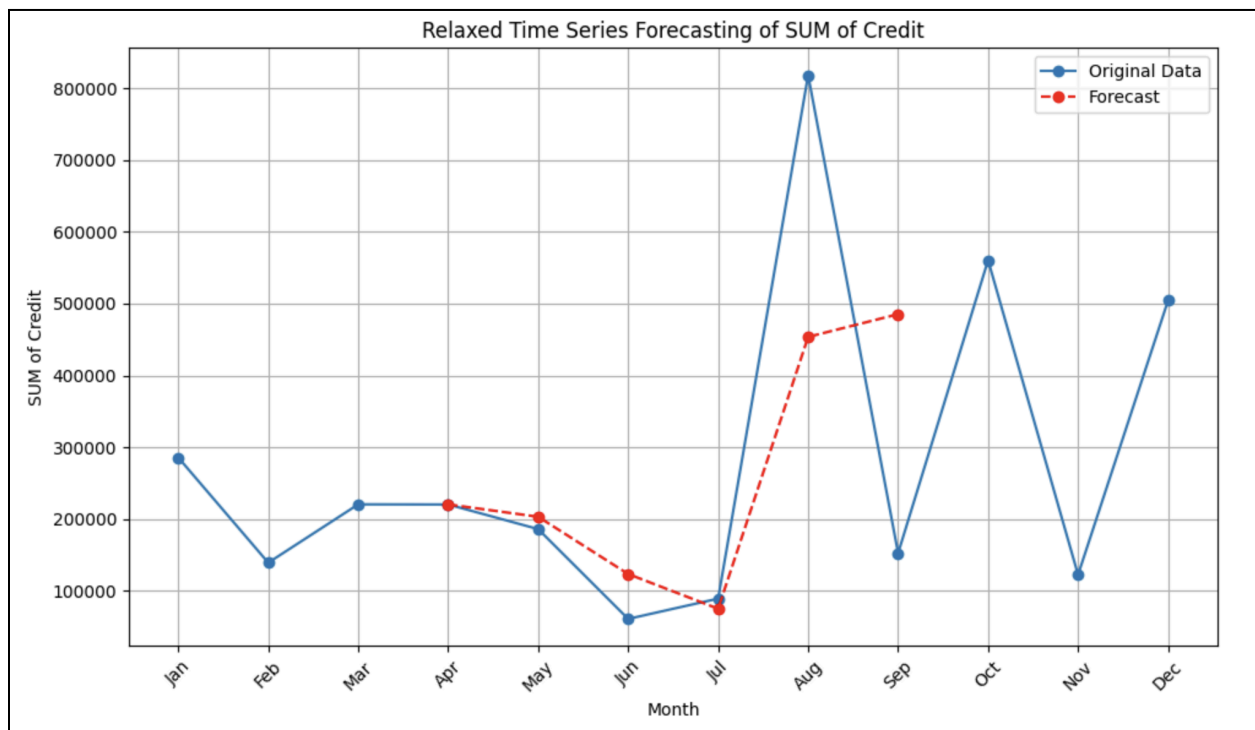
Graph 1: Time Series Plot of Sum of Credit

Observations from Plot of Sum of Credit:

1. The data shows a significant spike in sales in August, with the "SUM of Credit" reaching its highest value of over 800,000. This indicates a seasonal demand, possibly due to back-to-school sales, annual contracts, or other cyclical factors.
2. Conversely, June and September exhibit notable drops in sales, with June being the lowest at around 61,000. This suggests periods of low activity, which might be due to the end of fiscal periods or simply a low demand phase.

- Between certain months like the sharp rise in August followed by the drop in September, the company experiences irregular sales cycles. This volatility presents challenges for efficient inventory management.
- The orange dotted line provides a smoother view of the sales trend, mitigating the impact of short-term fluctuations. The SMA shows a gradual increase starting from July, peaking in October, before a slight dip in November, and then rising again in December.
- The SMA indicates a relatively stable period of growth from April to July, with only minor fluctuations. This could reflect steady business operations and consistent customer demand during these months.

B. Forecasting of Sales



Graph 2: Time Series Forecasting of Sales

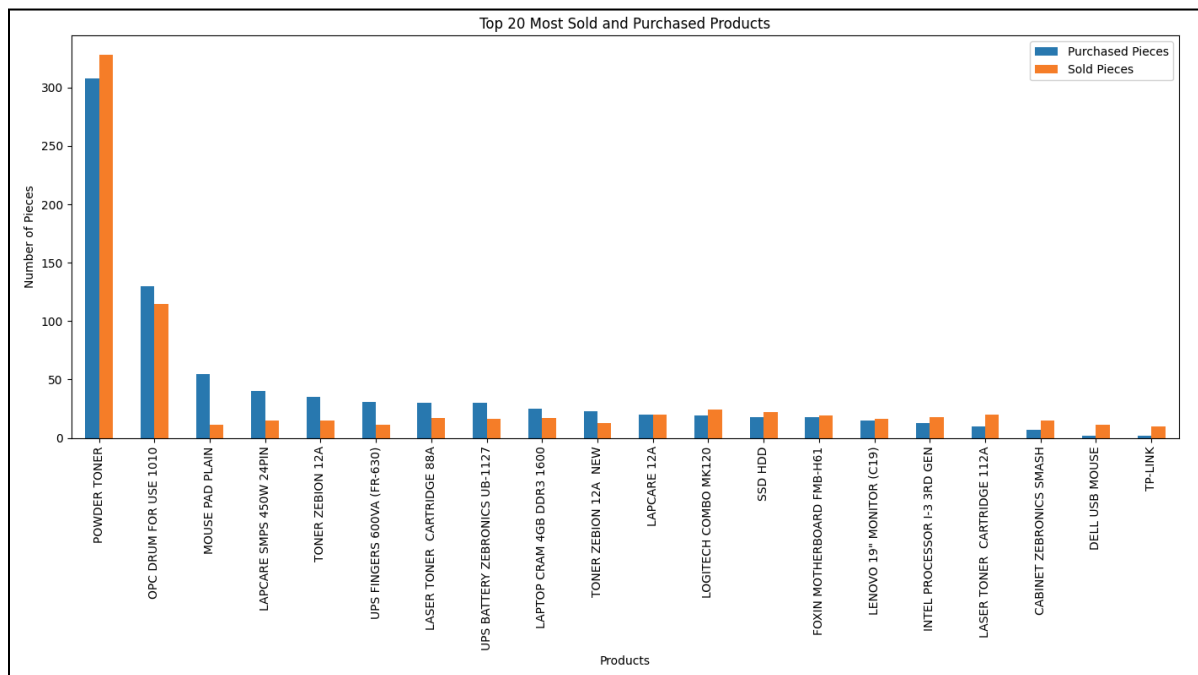
Observations

- The forecast shows a relatively consistent trend through April and May, followed by a decline in June and July, and then a significant increase in August and September. This is very reminiscent of the actual sales which means that the forecast is correctly capturing the sale trends.

2. Unlike the sharp peaks and valleys in the original data, the forecast reflects a more smoothed pattern, which indicates that the model is generalizing the seasonality and trends rather than reacting to every fluctuation. This avoids overfitting and overstocking of products based on just one season of data.
3. The forecast predicts a significant increase in sales for August and September, which may be reflective of an attempt to capture some level of seasonality based on the historical spike in August.
4. The increase in forecast in the month of September suggests that the model recognizes August and October as high-demand months, so instead of lowering the stock, it almost flattens out with a slight increase to avoid understocking for the month of October.
5. The lower forecasts for June and July suggest that these months might be quieter, which could be an opportunity to reduce costs, run promotions, or prepare for the upcoming peak.

Analysis of best Sold & Purchased products

A. Top 20 Products Based on Item

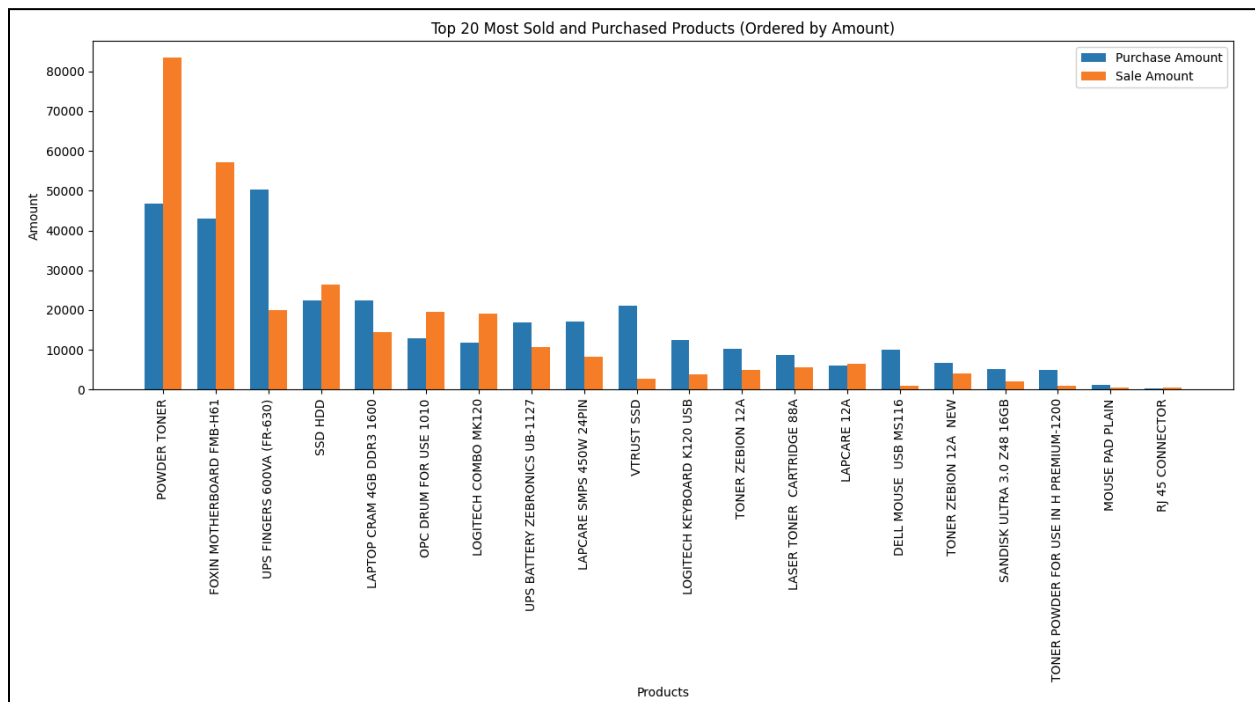


Graph 3: Top 20 Most Sold and Purchased Products

Observations

1. Most of the products in the top 20 show a strong correlation between the number of pieces purchased and sold. This indicates efficient inventory management where the company is closely matching its purchases with sales demand without overstocking or understocking. This minimizes excess inventory while still meeting customer demand.
2. A few products show slight discrepancies between purchases and sales. For instance, “Mouse pad plain”, “Lapcare SMPS 450W 24pin”, “Toner Zeboin 12A” have slightly more purchases than sales, indicating a potential risk of overstocking. Conversely, some items with more sales than purchases like “Laser Toner Cartridge 112A”, “Cabinet Zebronics SMASH” and “Dell USB Mouse” could risk stock outs if demand continues to exceed supply.

B. Top 20 Products based on Amount

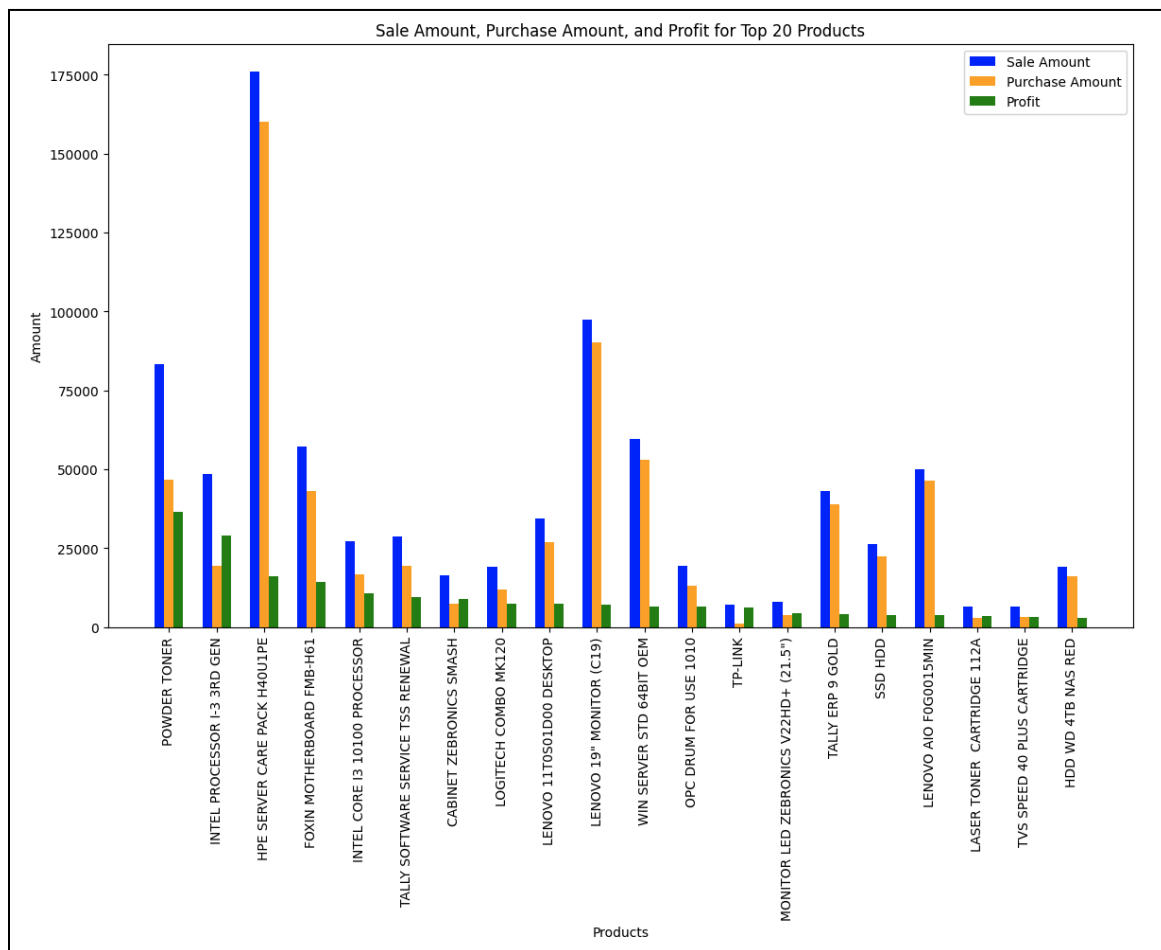


Graph 4: Top 20 Most Sold and Purchased Products by Amount

Observations

1. The top 20 items based on the amount sold or purchased include higher-value items even if they are sold in smaller quantities. These products tend to have a higher price per unit, which drives up their total sales or purchase value.
2. The top 20 items based on the number of pieces sold or purchased may include lower-value items that sell in large volumes. These are typically staple products or those with a lower price point that customers purchase more frequently.
3. Some products with higher prices naturally appear on the top 20 list by amount, even if they are not sold in large quantities. Conversely, products with lower prices may not make it to the top by amount, but they do by quantity due to higher sales volume.

Yearly Profit of Top 20 Products

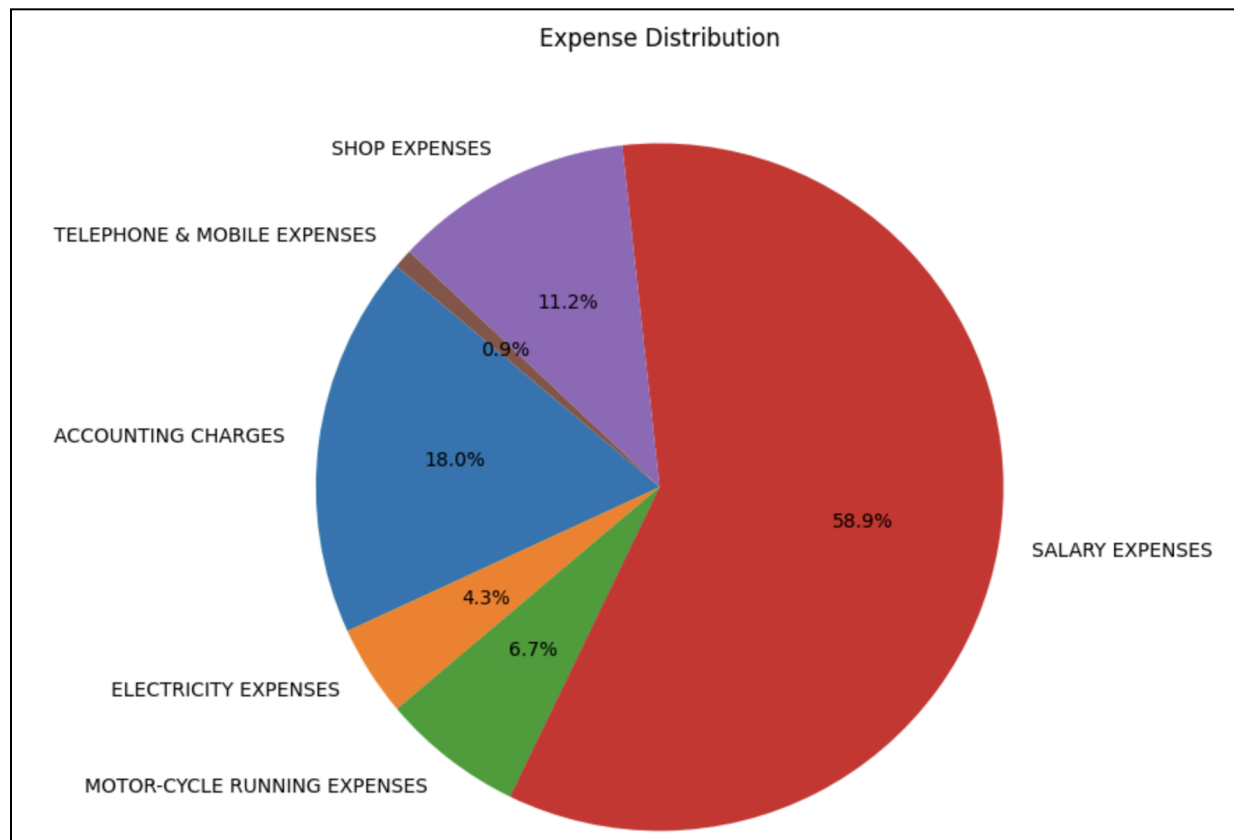


Graph 5: Profits of Top Selling Products

Observations

1. "Powder Toner" demonstrates significant profitability, with sales amounts far exceeding purchase costs. Hardware products like "Opc Drum For Use 1010", "Intel Core I3 10100 Processor" show balanced sales and purchase amounts, resulting in moderate profits. These items have stable demand but lower profit margins, requiring careful pricing and cost control.
2. Software products like "Win Server Std 64bit Oem", "Tally Erp 9 Gold" exhibit low profit margins, where sales and purchase amounts are closely aligned. These products may need higher sales volumes or cost reductions to enhance their profitability.
3. High-sales, low-profit products highlight potential areas for cost optimization or pricing adjustments. Understanding demand trends can help focus efforts on maximizing profitability for these items.

Monthly Expenses of the company



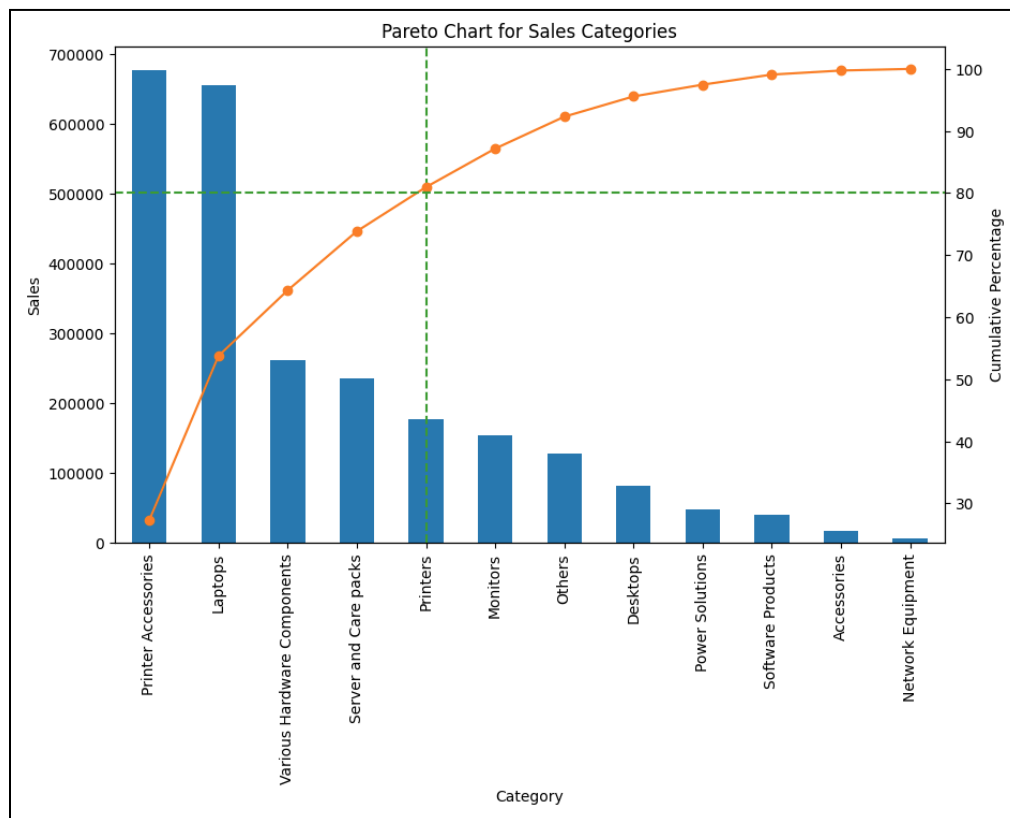
Graph 6: Pie Chart of Monthly Expenses

Observations

1. Salary expenses dominate the expense distribution, accounting for the largest share of 58.9%. It is clearly the most significant expense, reflecting the importance of human resources in the overall cost structure.
2. Accounting charges, contributing 18% of the total expenses indicates a substantial investment in accounting services or related financial management.
3. Motor-cycle running expenses accounts for 6.7%, showing a moderate level of spending on transportation or vehicle-related costs.
4. With 11.1% of the total, shop expenses are notable but not as large as salary or accounting charges. This includes rent, maintenance and other operational costs.
5. Representing 4.2% of the expenses, electricity costs are relatively small but necessary for the operation followed by telephone and mobile expenses which is the smallest category, at 0.9%, indicating minimal spending on communication services.

Pareto Analysis

A. Pareto Analysis of Sales Data

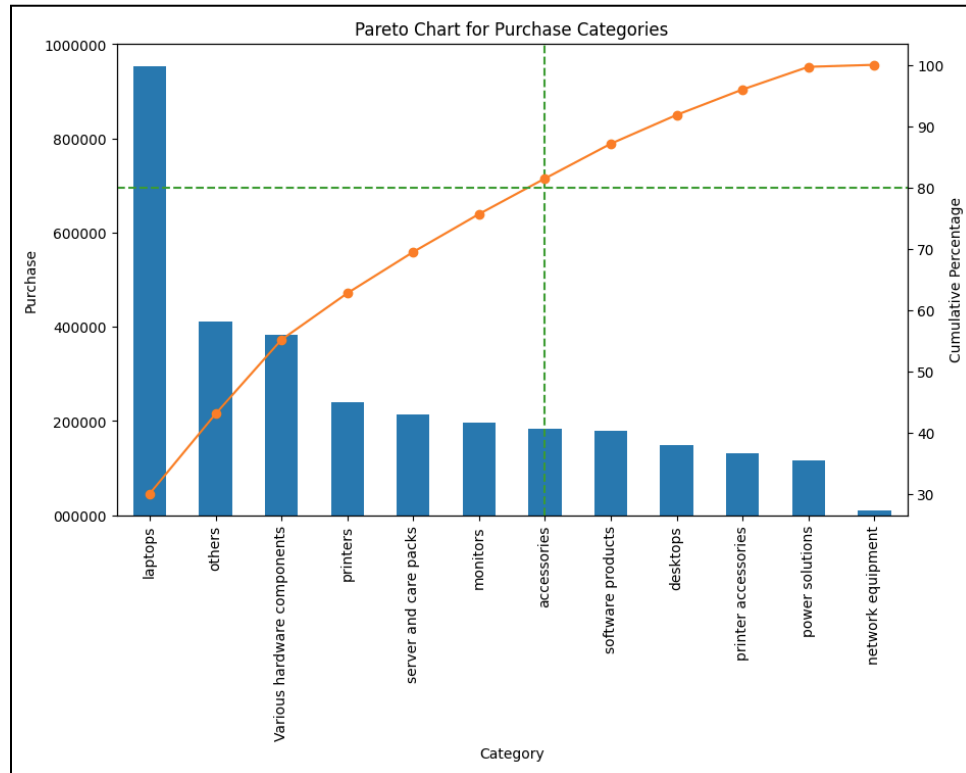


Graph 7: Pareto Chart of Sales Data

Observations:

1. The top three categories: “Printer Accessories”, “Laptops”, and “Various Hardware Components” account for a significant portion of the total sales. Specifically, these three categories alone contribute to approximately 80% of the total sales.
2. This supports the Pareto principle which indicates that these categories are critical to the business and should be the primary focus for inventory management which could potentially yield the most substantial benefits for the business.
3. Categories beyond the top five contribute marginally to the overall sales. Efforts to optimize these areas might not have as significant of an impact compared to the major contributors, though they could still be important for niche markets or specific customer needs.
4. While Printer Accessories and Laptops should be the focus for high-impact strategies, maintaining a strong presence in Hardware Components, Servers, and Printers can ensure a balanced product portfolio.

B. Pareto Analysis of Purchase Data

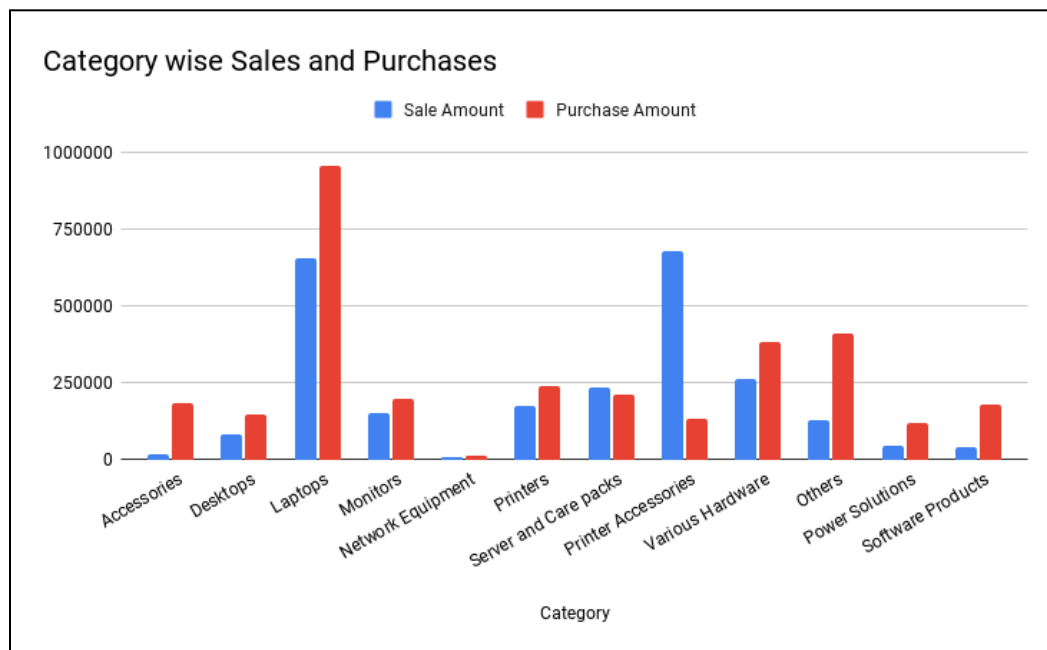


Graph 8: Pareto Chart of Purchase Data

Observations

1. The top sales categories, such as "Printer Accessories" and "Laptops," account for the majority of the sales, yet the Purchase Pareto shows that purchases in these categories are not prioritized as highly because they appear much lower on the purchase chart. This could result in frequent stock shortages, negatively impacting customer satisfaction and sales performance.
2. Certain categories like "Various Hardware Components" and "Others" appear prominently in the Purchase Pareto but do not contribute significantly to sales. This suggests that the company is over-purchasing in categories that do not drive revenue, leading to excess inventory, increased holding costs, and potential obsolescence.
3. The disparity between what is being purchased and what is driving sales indicates an inventory imbalance. The company may be holding onto large amounts of stock in categories that do not sell as well, tying up capital that could be better allocated to more profitable items.
4. The company's failure to align purchases with sales trends results in missed revenue opportunities. If high-demand categories are not adequately stocked, customers may turn to competitors, leading to a loss of market share.

Comparison of Sales and Purchase Categories



Graph 9: Category wise Sales and Purchase

Observations

1. The "Laptops" and "Printer Accessories" categories have the highest sales amount among all categories, highlighting them as key revenue drivers. This suggests that the company should continue to focus on these categories, possibly expanding product offerings or promotions.
2. However the "Printer Accessories" category shows a significant discrepancy between purchase and sales amounts, with purchases being substantially lower than sales indicating a potential understock which shall require inventory increase.
3. The "Monitors" and "Server and Care Packs" categories exhibit a balanced ratio between purchase and sales amounts. This balance suggests efficient inventory management and consistent demand, making them stable contributors to overall profitability.
4. The "Various Hardware" category shows a notable purchase amount although relatively high sales. This indicates a need for better alignment between stock levels and market demand or a strategic review of product offerings in this category.
5. The "Power Solutions" and "Software Products" categories show higher purchase and relatively lower sales. This presents opportunities for growth. Targeted efforts could help boost sales and improve profitability in this segment.

4. Interpretation of Results and Recommendations

Interpretations

Seasonality and Sales Trends

Sales show clear seasonality, with a significant spike in August and drops in June and September, likely due to cyclical events like back-to-school periods or annual contracts. This volatility challenges inventory management, requiring strategic planning to prevent overstocking or understocking.

3-Month SMA Analysis

The 3-Month Simple Moving Average (SMA) smooths out short-term fluctuations, revealing steady growth from April to July and peaks in October and December, likely driven by business budget cycles and holiday demand.

Category-Wise Sales and Purchases

High sales in laptops and printer accessories indicate strong demand, but understocking in printer accessories points to a need for better inventory planning. Overstocking in the "Others" category suggests poor demand forecasting, tying up capital and increasing storage costs. Minimal sales in "Power Solutions" and "Software Products" require strategic reevaluation.

Forecast Accuracy

The sales forecast effectively captures seasonality and trends, avoiding overfitting. It predicts higher sales in August and September, with lower activity in June and July, guiding financial planning and inventory management.

Pareto Analysis

"Printer Accessories," "Laptops," and "Various Hardware Components" contribute to 80% of total sales, emphasizing the need to focus on these categories for the most significant business impact.

Expense Management

Monthly expenses are dominated by salary costs, which account for nearly 59% of the total, followed by significant costs in accounting and motor-cycle running expenses.

Recommendations

Inventory Management

1. Increase stock levels ahead of the August demand spike; reduce inventory during quieter months like June and July to lower carrying costs.
2. According to forecasting models for key categories like "Printer Accessories" and "Laptops" need to maintain optimal inventory levels and minimize understocking or overstocking.
3. Reassess inventory levels for low-performing categories like "Power Solutions" and "Software Products," shifting focus to more profitable items.

Sales and Marketing Strategies

1. Focus marketing efforts on top performers like "Printer Accessories" and "Laptops" with targeted promotions and loyalty programs.
2. Optimize less significant categories by tailoring marketing strategies to niche markets, enhancing sales without substantial investment.

Financial Planning

1. Increase sales during slow months like June and July by running promotions or planning maintenance.
2. Ensure sufficient stock and strong marketing efforts for the forecasted sales increase in August and September to maximize revenue.

Product Portfolio Management

1. Keep a mix of high-demand and niche products to meet diverse customer needs and maximize profitability.
2. Consider discontinuing or revitalizing low-performing categories like "Power Solutions" and "Software Products" through strategic marketing or new product introductions.

Below are the original sales and purchase data along with the cleaned data in subsequent sheets.

The kaggle notebook is also given where the data visualizations were done.

[Sales Data](#)

[Purchase Data](#)

[Kaggle Notebook](#)