**PROJECT REPORT**

**Introduction**

The purpose of this analysis is to perform data cleaning, visualization, and correlation analysis on a dataset containing sales data. The dataset includes columns for Date, Sales, Expenses, Profit, and Region. The analysis includes generating various plots to visualize the data and calculating correlations to understand the relationships between different variables.

**Data Cleaning**

The dataset was loaded and checked for missing values. Any missing values were handled appropriately to ensure a clean dataset for analysis. The Date column was converted to a datetime object, and numerical columns (Sales, Expenses, Profit) were ensured to have the correct data types.

**Data Visualization**

From the comprehensive analysis of the sales data through various visualizations and correlation analysis, we can draw several insights about the company's performance and trends.

This report summarizes the findings from six different visualizations, including time series plots, distribution analyses, regional comparisons, and correlation studies.

**Sales Trends Over Time:**

The line plot of sales over time (Image 6) reveals significant variability in daily sales from early 2020 to early 2024. Sales fluctuate roughly between 1000 and 5000 units, showing no clear long-term upward or downward trend. However, the data suggests the presence of cyclical or seasonal patterns that warrant further investigation.

To gain a clearer picture of these patterns, we examined the monthly average sales (Image 1). This time series plot confirms the presence of seasonal fluctuations, with notable peaks and troughs occurring at regular intervals. The highest peak in monthly average sales appears around January 2022, reaching approximately 3600 units. Conversely, the lowest point is observed around July 2021, dipping to about 2400 units. This cyclical behavior suggests that the business may be influenced by seasonal factors, possibly related to holiday shopping periods or annual events that impact consumer behavior.

**Sales Distribution:**

The histogram of sales (Image 5) provides insight into the overall distribution of sales figures. The data shows a roughly normal distribution with a slight right skew. The majority of sales fall between 1500 and 4500 units, with the highest frequency occurring in the 3000-3500 range. This distribution indicates that while the business has a stable core of sales activity, there are occasional high-performing outliers that push the distribution slightly to the right.

**Regional Performance:**

The box plot analysis of sales by region (Image 4) reveals significant differences in sales performance across four regions:

1. Greater Accra stands out as the top-performing region, boasting the highest median sales and the largest interquartile range. This suggests not only higher overall sales but also greater variability, possibly indicating more diverse market conditions or a larger customer base.
2. Volta, in contrast, shows the lowest median sales but maintains a similar range to other regions. This could point to untapped potential or the need for targeted strategies to boost performance in this area.
3. Central and Upper West regions demonstrate similar sales distributions, positioning themselves between Greater Accra and Volta in terms of performance.

All regions exhibit some outliers, particularly on the upper end of the sales range, indicating occasional high-performing days or transactions across all areas.

**Relationship Between Expenses, Sales, and Profit:**

The scatter plot of expenses versus profit (Image 3) and the correlation matrix (Image 2) provides crucial insights into the financial dynamics of the business:

1. There is no clear linear relationship between expenses and profit, as evidenced by the widely dispersed points in the scatter plot. This suggests that factors beyond expenses play significant roles in determining profitability.
2. The correlation matrix reveals extremely weak correlations between sales, expenses, and profit:

* Sales and Expenses: Very weak positive correlation (0.022)
* Sales and Profit: Very weak negative correlation (-0.017)
* Expenses and Profit: Very weak positive correlation (0.0049)

These negligible correlations indicate that the relationships between these financial metrics are complex and likely influenced by numerous other factors not captured in this analysis. The slight negative correlation between sales and profit, though weak, is particularly interesting and may warrant further investigation into pricing strategies or cost structures.

Conclusions and Recommendations:

1. The business demonstrates strong seasonal patterns in sales, suggesting the need for season-specific strategies to capitalize on peak periods and mitigate troughs.
2. Regional performance varies significantly, with Greater Accra leading in sales. Management should investigate the factors behind this success and consider applying similar strategies to boost performance in other regions, particularly Volta.

In conclusion, while the business shows stable overall performance with interesting seasonal patterns, there are significant opportunities for optimization across regions and throughout the year. Further in-depth analysis of additional variables and advanced statistical techniques could uncover more actionable insights to drive strategic decision-making and improve overall business performance.

**Image 1**

**A graph showing a line

Description automatically generated with medium confidence**

**Image 2**

**A red and white squares with numbers

Description automatically generated**

**Image 3**

**A graph of blue dots

Description automatically generated**

**Image 4**

**A diagram of different colored squares

Description automatically generated**

**Image 5**

**A graph of sales

Description automatically generated**

**Image 6**

**A blue scribbles on a white background

Description automatically generated**