**Sales Performance Data Analysis Report**

**Iyappan Krishnan - 27-May-2025**

**Introduction**

This report presents an analysis of sales performance data. The objective of this project is to explore the provided sales dataset, identify key trends and patterns, and derive insights that can inform business decisions. The analysis includes data loading, exploration, and visualization to understand sales distribution, customer behaviour, and product performance.

**Dataset Description**

The dataset used for this analysis is a CSV file named sales\_data\_sample.csv. It contains detailed sales records, including information about orders, products, customers, and sales figures.

* **Source:** The dataset is a sample sales data file.
* **Structure:** The dataset is in a tabular format with 2823 rows and 25 columns. The columns include various attributes related to sales transactions such as SALES, QUANTITYORDERED, PRICEEACH, ORDERDATE, PRODUCTLINE, DEALSIZE, and customer information like CUSTOMERNAME, STATE, and TERRITORY.
* **Key Features:** Important features include sales figures (SALES), quantity ordered (QUANTITYORDERED), pricing details (PRICEEACH, MSRP), order date (ORDERDATE), product information (PRODUCTLINE, PRODUCTCODE), customer details (CUSTOMERNAME, DEALSIZE, STATE, TERRITORY), and order status (STATUS).

**Methodology**

The following steps were taken to analyse the sales data:

1. **Data Loading:** The sales data was loaded from the sales\_data\_sample.csv file into a pandas DataFrame. Encoding issues were addressed by specifying encoding='latin-1'.
2. **Data Exploration:** The dataset's characteristics were examined by checking its shape, data types, descriptive statistics, missing values, and unique values in categorical columns.
3. **Data Analysis:** Key variables were analyzed for their distribution and correlations. This involved calculating summary statistics, generating visualizations (histograms, scatter plots), and computing correlation coefficients between numerical features. Sales trends over time (monthly and yearly), top-performing product lines and products, and region-wise sales distribution were also investigated.
4. **Data Visualization:** Various plots were created using matplotlib and seaborn to visualize the data. These included histograms, box plots, scatter plots, line plots for time trends, and bar plots for categorical analyses (product lines, dealsize, regions). Correlation heatmaps were used to visualize relationships between numerical features.

**Dashboard Explanation**

(Since a dashboard was not explicitly created in the provided code, this section will explain the key visualizations generated in the analysis that could be part of a dashboard.)

The visualizations created provide key insights into the sales data:

* **Distribution Plots (Histograms, Box Plots):** These plots show the distribution of numerical variables like SALES, QUANTITYORDERED, PRICEEACH, and MSRP. They help understand the range and spread of these values and identify potential outliers.
* **Scatter Plots:** Scatter plots visualize the relationship between SALES and QUANTITYORDERED or PRICEEACH. Adding DEALSIZE as a hue helps understand how different deal sizes influence these relationships.
* **Time Series Plots (Monthly and Yearly Sales Trends):** These line plots show how total sales evolve over time, revealing seasonality, growth, or decline patterns.
* **Bar Plots (Sales by Product Line, Top Products, Region-wise Sales):** These plots illustrate sales performance across different categories, highlighting the most successful product lines, individual products, and geographical regions.
* **Correlation Heatmap:** This visualization shows the correlation matrix of numerical features, indicating the strength and direction of linear relationships between variables.

These visualizations collectively provide a comprehensive overview of sales performance and can form the basis of a sales dashboard.

**Findings & Insights**

Based on the data analysis, the following key findings and insights were identified:

* The dataset contains a substantial number of sales records with detailed information.
* Several columns have missing values, which might require data cleaning steps like imputation or removal depending on the analysis requirements.
* Euro Shopping Channel is a significant customer in terms of order volume.
* The majority of orders have a "Shipped" status, indicating efficient order fulfillment.
* Classic Cars is the most popular product line based on the frequency of orders.
* Positive correlations between SALES and both QUANTITYORDERED and PRICEEACH suggest that selling more units or units at higher prices directly contributes to higher sales revenue.
* Analysis of monthly and yearly sales trends can help in forecasting and identifying peak sales periods.
* Identifying top-performing product lines and individual products can inform inventory management, marketing strategies, and product development.
* Understanding region-wise sales distribution can help in allocating resources and tailoring sales strategies to specific territories.

**Conclusion**

This analysis provided valuable insights into the sales performance data, covering data loading, exploration, and visualization of key metrics. We identified trends in sales over time, examined the performance of different product lines and products, and explored relationships between sales and other variables.