Report of Task 1 (EDA)

1. Introduction:

This report aims to present a detailed analysis of the sales and revenue data based on the provided datasets, which contain information about customers, products, and transactions over the period from 2022 to 2024. By merging and analyzing these datasets, we can gain valuable insights into the performance of sales across different regions, identify the top-selling products, evaluate the relationship between product price and quantity sold, and determine which regions and product categories contribute the most to overall revenue. The analysis also explores trends over time to identify seasonal patterns that can inform future sales strategies.

2. Data Overview:

The datasets used in this analysis are as follows:

- **Customers:** This table contains information about each customer, including their unique ID, name, the region they are from, and their signup date. This dataset is crucial for understanding customer demographics and regional distribution.
- **Products:** The product dataset contains details about each product sold, including product ID, product name, category, and price. Understanding product categories and their pricing is essential for identifying key revenue drivers.
- **Transactions:** The transactions dataset includes details of each sale, including the transaction ID, customer ID, product ID, transaction date, quantity sold, and the total value of the sale. This is the most important dataset for tracking revenue generation and sales volume.

3. Data Preprocessing:

Before diving into the analysis, the data underwent several preprocessing steps to ensure accuracy and consistency:

- **Missing Values:** We checked for any missing values in the datasets and found none. This ensured that the analysis could be performed without any data gaps.
- **Duplicate Rows:** We identified and removed any duplicate rows to ensure that the analysis was based on unique records only.
- **Date Conversion:** The transaction dates were converted into datetime format for better time-based analysis, enabling us to perform operations like resampling and aggregating by month.
- **Data Merging:** The three datasets (Customers, Products, and Transactions) were merged based on their respective IDs (customer ID and product ID). This resulted in a comprehensive dataset containing all the necessary information for analysis.

4. Revenue by Region:

One of the primary objectives was to determine the total revenue generated by each region. By grouping the merged dataset by region and summing the total value of transactions, we identified which regions contributed the most to overall revenue. A bar chart was created to visually represent these values, making it easy to compare regional performance. This analysis revealed that certain regions consistently outperformed others in terms of revenue, suggesting a stronger market presence in those areas.

5. Top 10 Selling Products:

Identifying the top-selling products is essential for understanding consumer preferences and demand patterns. By summing the quantities sold for each product, we were able to rank products by sales volume. A bar chart was used to visualize the top 10 selling products, providing a clear overview of the products that are most popular with customers. This information is useful for inventory management, marketing strategies, and identifying products that could benefit from further promotion.

6. Price vs Quantity Sold:

To understand the relationship between product price and sales volume, a scatter plot was created. This plot allows us to examine whether there is a correlation between the price of a product and the quantity sold. The analysis showed that while some high-priced products sold in large quantities, others did not, suggesting that factors other than price, such as product category or promotional discounts, could also influence purchasing behavior.

7. Correlation Analysis:

A correlation heatmap was generated to analyze the relationships between key variables such as price, quantity sold, and total sales value. This helped identify any strong correlations that could provide insights into sales performance. For example, a high correlation between price and sales value indicated that more expensive products contributed significantly to total revenue, while quantity sold did not always correlate strongly with revenue, highlighting the impact of pricing strategies.

8. Monthly Revenue Trend:

The monthly revenue trend was analyzed to identify any seasonal patterns in sales. By resampling the transaction data to a monthly frequency and summing the total revenue for each month, a line plot was created to visualize the trend. The analysis revealed certain months with higher revenue, which may be attributed to seasonal demand, holidays, or promotional activities. Understanding these trends helps in forecasting future sales and planning marketing campaigns accordingly.

9. Revenue by Product Category:

Revenue by product category was calculated by grouping the data by product category and summing the total sales value. A bar chart was created to visualize the contribution of each category to overall revenue. This analysis highlighted the categories that generate the most revenue and provided insights into the relative success of different product lines. Categories with higher revenue potential could be prioritized in marketing and inventory decisions.

10. Revenue by Region and Product:

To gain deeper insights into regional sales performance, a stacked bar chart was created to show the revenue generated by each region for different product categories. This visualization revealed regional preferences for specific categories and helped in identifying which products were driving revenue in each region. For example, certain regions showed a strong preference for specific product categories, allowing for targeted marketing and sales strategies.

11. Conclusion:

The analysis of sales and revenue data provided valuable insights into the performance of the company's products and regional markets. By examining revenue by region, product, and category, as well as analyzing trends over time, we identified key drivers of sales and areas for improvement. The top-selling products and regions were identified, and insights into the relationship between price and quantity sold were obtained. Additionally, the seasonal trends in revenue highlighted the importance of timing in sales strategies. This analysis can inform future business decisions, such as adjusting pricing strategies, optimizing inventory, and launching targeted marketing campaigns to further increase sales and revenue.