Amazon Sales Data Analysis Project Report

Objective

To analyze Amazon sales data to derive business insights, identify trends, and support data-driven decision-making using Python.

Tools & Technologies Used

- Python
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Jupyter Notebook

Steps Performed

1. Data Loading

Loaded the Amazon sales dataset using Pandas for analysis.

2. Data Cleaning

- Removed irrelevant columns ('New', 'PendingS')
- Dropped missing values
- Converted data types appropriately (e.g., 'Date' to datetime, 'ship-postal-code' to integer)

3. Data Exploration

Performed initial exploration using:

- .head()
- .info()
- .describe()

4. Univariate Analysis

- Analyzed distribution of categorical variables like 'Size', 'Courier Status', 'B2B', 'Category'

5. Data Visualization

Created multiple plots to visualize the data:

- Countplots for Size Distribution and Courier Status
- Pie Chart for B2B vs B2C
- Bar Chart for Product Category Distribution
- Scatter Plot for Category vs Size
- Monthly Sales Trend Line Plot
- State-wise Sales Bar Chart
- Order Status Countplot

6. Business Insights Generated

- Most Selling Size: Medium (M)
- Most Sold Product Category: Identified through mode
- B2B Contribution: Only 1% of customers are B2B
- Monthly Sales Trend: Seasonal patterns identified
- Top Selling States: Highlighted states with maximum orders

7. Recommendations

- Focus inventory on M-size T-shirts
- Create targeted offers for top-selling states
- Improve delivery logistics to minimize cancellations and delays
- Focus marketing strategies on B2C customers (99% of the dataset)

Conclusion

The Amazon Sales Data Analysis Project enabled effective exploration of the dataset, data cleaning, visualization, and generation of valuable business insights. The analysis supports data-driven strategies for improving sales, marketing, and operational efficiency.