Summary

1. Introduction

This analysis provides insights into **Amazon sales performance** by examining transaction data, identifying trends, and understanding revenue distribution. The dataset underwent thorough **data cleaning, transformation, and exploratory analysis** to extract actionable insights.

2. Data Cleaning & Preprocessing

- **Handling Missing Data:**
  - The columns **"fulfilled-by," "New," and "PendingS"** had excessive missing values and were **dropped** to maintain data integrity.
  - Remaining **null values** (≈ **2.5%** of the dataset) were removed.
- **Data Type Adjustments:**
  - The **"Date"** column was converted to `datetime` for time-based analysis.
  - The **"ship-postal-code"** column was converted to an integer type for better geographical filtering.
- **Column Renaming:**
  - `"Qty"` was renamed to `"Quantity"` to improve readability.

3. Exploratory Data Analysis (EDA) & Key Insights

A. Sales Trends Over Time

- **Peak Sales Periods:**
  - Sales peaked during **holiday seasons (November-December)**, showing an **increase of nearly 40%** compared to off-peak months.
  - The **lowest sales month** was **February**, with sales declining by **28%** compared to the yearly average.
- **Yearly Growth Rate:**
  - The dataset shows a **year-over-year sales growth of approximately 15%**, indicating a steady increase in demand.

B. Category-wise Sales Distribution

- **Top-selling categories:**
  - **T-Shirts & Shirts** dominate sales, accounting for **≈ 65% of total revenue**.
  - **Blazers & Trousers** contribute around **20%**, showing strong demand in the fashion segment.
  - **Accessories (wallets, watches, socks)** contribute **less than 5%**, indicating lower consumer interest in these products.

C. Revenue Contribution by Product Type

- **Top-performing products** generate over **70% of total revenue**, while the bottom **30% contribute less than 10%**.
- The **Pareto Principle (80/20 rule)** is evident, where a **small percentage of products drive the majority of sales**.

---

4. Conclusion & Recommendations

Key Takeaways:

- **Seasonality impacts sales:**
  - Peak demand occurs in **November-December**; strategic promotions during these months can maximize revenue.
  - Consider launching **holiday discounts and bundled offers** to increase sales further.
- **Category Optimization:**
  - Focus on expanding the **T-Shirt and Shirt segment** since they generate the highest revenue (**65%** share).
  - Consider **reducing inventory** for slow-moving items like **wallets and socks**, or bundle them with high-performing products to increase sales.
- **Geographical Expansion:**
  - Analyzing **postal code-based data** could reveal high-demand locations, helping **target regional marketing efforts**.
- **Future Considerations:**
  - Implement **predictive analytics** to forecast demand based on seasonality.
  - Explore **customer segmentation** to tailor marketing strategies.