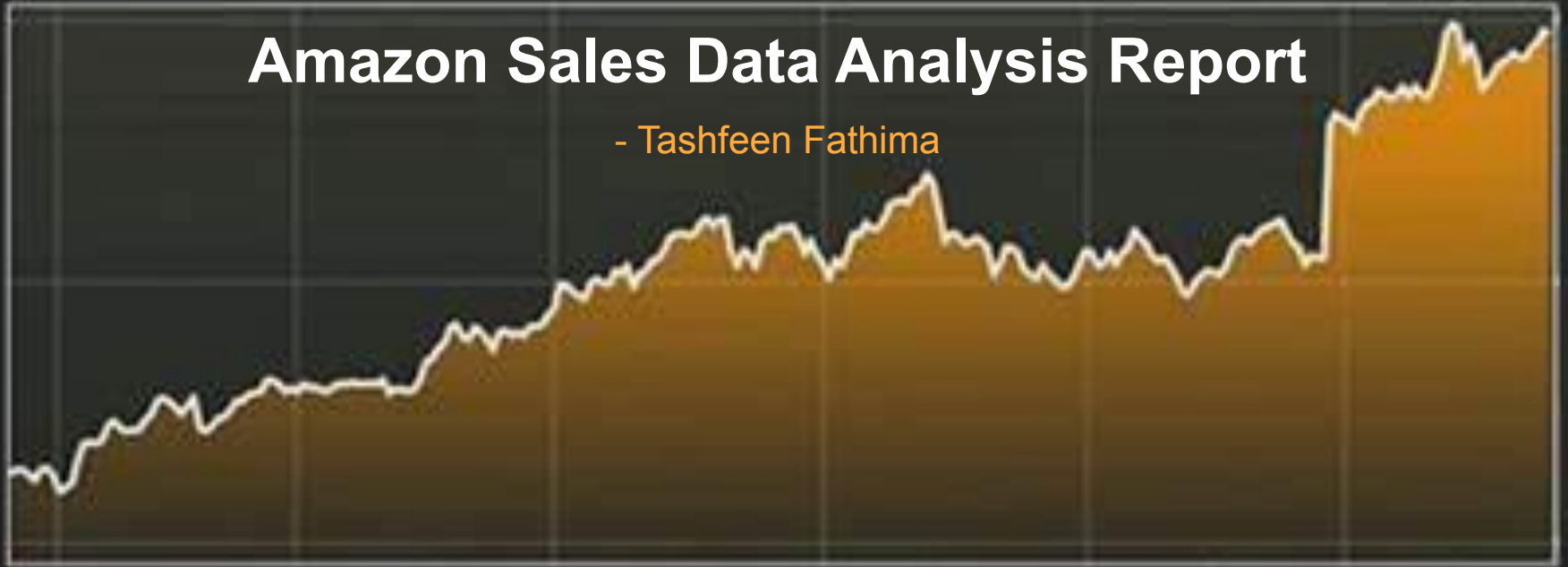




Amazon Sales Data Analysis Report

- Tashfeen Fathima



Problem Statement:

Analyze and Provide Insights on Amazon Sales Report.

Problem Description:

The provided dataset contains information about sales transactions on Amazon, including details such as order ID, date, status, fulfilment method, sales channel, product category, size, quantity, amount, shipping details, and more.

The objective is to conduct a comprehensive analysis of the data and extract actionable insights to support business decision-making.

Datasets:

https://drive.google.com/file/d/1YrjYKtS1WHmINL6eafRsrDzrZaw2_WvX/view?usp=sharing.

Introduction:

This report provides an in-depth analysis of Amazon's sales data to uncover trends, insights, and actionable recommendations for improving business strategies and operations.

Objective:

The aim of this project is to understand overall sales performance, analyze product distribution across categories, investigate fulfillment methods and their impact on sales, segment customers based on purchasing patterns, explore geographical sales distribution.

Architecture:

The analysis follows a systematic approach using the following tools and processes:

- **Tools and Libraries:** Python, Pandas, NumPy, Matplotlib, Seaborn.
- **Data Processing Workflow:**
 - **Data Loading:** Import datasets and inspect initial structure.
 - **Data Cleaning:** Handle missing values, duplicates, and anomalies.
 - **Data Analysis:** Perform exploratory data analysis (EDA).
 - **Visualization:** Create visualizations to derive meaningful insights.
 - **Output Generation:** Generate summary reports and recommendations.

Dataset Information:

Dataset Overview:

- **Number of Rows and Columns:** 128,976 rows × 21 columns.
- **Index Column:** The dataset uses a default integer index, explicitly labeled as index.

Dataset Description:

- **index:** Default index column, used for identifying rows in the dataset. | **Int64**
- **Order ID:** Unique identifier assigned to each order for tracking and processing purposes. | **Object**
- **Date:** The date on which the order was placed, stored in string format. | **Object**
- **Status:** Indicates the current state of the order, such as processed, shipped, or delivered. | **Object**

Dataset Information(Continuation):

Dataset Description(Continuation):

- **Fulfilment:** Describes the method or channel used to fulfill the order, such as direct shipping or drop-shipping. | **Object**
- **Sales Channel:** The platform through which the product was sold, e.g., online, retail, or marketplace. | **Object**
- **ship-service-level:** Shipping service type, which can refer to expedited, standard, or other delivery options. | **Object**
- **Category:** The product category or type to which the item belongs, such as electronics, apparel, etc. | **Object**
- **Size:** The size of the product, typically referring to dimensions or capacity (e.g., small, medium, large). | **Object**

Dataset Information(Continuation):

Dataset Description(Continuation):

- **Courier Status:** Indicates the current status of the shipment as provided by the courier service, such as in transit or delivered. | **Object**
- **Qty:** The quantity of items sold in the particular order. | **Int64**
- **currency:** The currency in which the transaction was conducted (e.g., USD, INR, EUR). | **Object**
- **Amount:** The total value of the sale, including price, taxes, and other fees, in numerical format. | **Float64**
- **ship-city:** The city where the order is being shipped to. | **Object**
- **ship-state:** The state or region where the order is being shipped to. | **Object**

Dataset Information(Continuation):

Dataset Description(Continuation):

- **ship-postal-code**: The postal code or ZIP code associated with the shipping address. | **Float64**
- **ship-country**: The country where the order is being shipped. | **Object**
- **B2B**: A boolean indicator showing whether the order is a business-to-business (B2B) transaction. | **Bool**
- **fulfilled-by**: Indicates the entity responsible for fulfilling the order, such as a third-party or the seller. | **Object**
- **New**: An empty column, likely left for future use or unpopulated data. | **Float64**
- **PendingS**: Another empty column with no data, possibly used for future processing. | **Float64**

Dataset Information(Continuation):

Missing Values and Data Issues:

After inspecting the dataset for missing values, we identified the following:

- **Missing Values:**
 - **Currency and Amount:** 7,800 rows have missing values in these columns.
 - **Ship-related Columns (Ship City, State, Postal Code, and Country):** 35 rows have missing values.
 - **Fulfilled By:** A large portion of the data (89,713 rows) is missing in this column.
 - **New and PendingS Columns:** Both columns are entirely empty and have been removed from the dataset.

Data Cleaning and Preprocessing:

- **Dropping Unnecessary Columns:** Removed the unnecessary index column to focus on relevant data.
- **Handling Missing Values:** Calculated the percentage of missing values for each column and decided to drop columns where more than 50% of the data was missing.
- **Removing Duplicates:** Eliminated duplicate entries.
- **Fixing Data Types:** The Date column is initially an object converted to datetime for easier manipulation.

Exploratory Data Analysis:

Sales Overview:

- **Sales Trend Over Time**
 - Sales peaked during specific months (shown as a line plot).

Product Analysis:

- **Sales Performance by Category**
 - T-shirts and Shirts dominate in sales volume.

Fulfillment Method Preferences:

- **Fulfillment Distribution**
 - Majority of orders fulfilled by **Amazon**, followed by **Merchant** fulfillment.
- **Cancellation Rate by Fulfillment**
 - **Merchant fulfillment** has a higher cancellation rate.

Exploratory Data Analysis(Continuation):

Customer Segmentation:

- **B2B vs B2C Sales Distribution**
 - Major sales are in **B2C** category.

Geographical Sales Distribution:

- **Top 10 States by Sales**
 - Maharashtra and Karnataka are top contributors to total sales.
- **Top 10 Cities by Sales**
 - Cities like **Mumbai** and **Bengaluru** drive high sales.

Shipment Status & Courier Analysis:

- **Shipment Status**
 - Analysis of successful vs cancelled orders.
- **Courier Status**
 - Breakdown of courier performance.

Key Insights:

Sales Overview

- Sales spike during specific months.

Product Performance

- **Top Categories:** T-shirts & Shirts dominate sales.

Fulfillment Methods

- Majority of orders fulfilled by **Amazon**; higher cancellation rate in **Merchant** fulfillment.

Geographical Distribution

- **Top States:** Maharashtra and Karnataka lead in sales.

Recommendations:

- **Focus on High-Demand Products:** Prioritize T-shirts and Shirts.
- **Optimize Shipping:** Promote **Amazon fulfillment** for fewer cancellations.
- **Target Underperforming Regions:** Focus marketing in states with low sales.

Conclusion:

This analysis highlights key sales patterns and customer preferences. By optimizing product focus, fulfillment methods, and regional marketing, the business can boost revenue, enhance customer satisfaction, and drive growth in e-commerce.

Conclusion:

This analysis highlights key sales patterns and customer preferences. By optimizing product focus, fulfillment methods, and regional marketing, the business can boost revenue, enhance customer satisfaction, and drive growth in e-commerce.

Thank You