## 1. Introduction

This project involves a comprehensive analysis of Amazon sales data to uncover key insights related to sales performance, regional trends, fulfillment efficiency, and product category distribution. The dataset contains detailed transactional data, and our objective is to preprocess, clean, and analyze the data to build an interactive, insightful dashboard in Microsoft Excel.

With over 128,000 rows and numerous attributes, the dataset provides a strong foundation for visual analytics and business intelligence. By leveraging Excel's advanced features, including Power Query, Pivot Tables, Conditional Formatting, Data Validation, Charts, and Hyperlinks, a clean and user-friendly dashboard was created to support decision-making and present summarized insights in a professional format.

## 2. Source of Dataset

The dataset was sourced from the Tableau Public website, which offers a collection of publicly available datasets for analysis and visualization. Specifically, the dataset relates to Amazon's online order transactions, providing comprehensive sales data.

## 3. Dataset Preprocessing

To ensure a clean, standardized, and analysis-ready dataset, extensive preprocessing was performed using Excel Power Query. The raw dataset contained 128,044 rows and numerous attributes like order details, sales amounts, locations, and fulfillment statuses. Here’s a step-by-step overview of how the dataset was transformed:

### a. Data Type Corrections

- Converted the Date column from the original MM/DD/YY format to the more standard and readable DD/MM/YYYY format.
- Ensured numeric columns like Qty (Quantity) and Amount were recognized as numbers, allowing for proper aggregation and visualization.

### b. Text Standardization

- Ship-State Column: Multiple entries for the same state had inconsistent spellings (e.g., "ORISSA", "ODDISA", etc.). These were standardized to a single format — "ODISHA".
- All state names were capitalized properly for consistency and accurate grouping.
- Ship-City Column: Transformed city names to Proper Case using Power Query so that names like "mumbai" became "Mumbai", improving visual clarity.

### c. Handling Missing Values

- Amount: Null values were replaced using the average (mean) of the column, assuming that missing entries were due to recording issues and not genuine zero sales.
- Qty (Quantity): Where quantity was missing but a valid amount was recorded, a quantity of 1 was assumed and filled accordingly.
- Currency: Since all transactions were India-based and consistently in INR, missing currency values were replaced with "INR".

- Other Fields: Fields like Courier Status, Fulfilment, and Sales Channel that had null or blank values were replaced with "UNKNOWN" to maintain data consistency and allow for better grouping in pivot tables and charts.

**d. General Cleaning**

- Removed duplicate rows (if any) and ensured no row had partial data in critical fields like Order ID or Date.
- Trimmed extra white spaces across all text fields to avoid grouping errors.
- Ensured all column names were appropriately labeled and formatted.

**4. Analysis on Dataset**

**Objective 1: Category-wise Total Sales Analysis**

**i. General Description:**
The aim was to identify which product categories contribute the most to total revenue. Understanding this allows businesses to focus on high-performing categories and make informed inventory and marketing decisions.

**ii. Specific Requirements:**

- Aggregate the total sales (Amount) for each product Category.
- Compare the overall revenue across all listed product types.
- Identify which category shows the strongest performance in terms of revenue generation.

**iii. Analysis Results:**
The analysis showed that:

- The **"Set"** category dominated with a total amount of ₹41,304,940.03, making it the highest revenue-generating segment.
- **"Kurta"** followed with ₹23347226.70.
- Other categories like **"Top"** and **"Saree"** performed moderately, while certain niche categories generated minimal revenue.

This insight highlights where the platform should focus its promotions, inventory investments, and supply chain enhancements.

**iv. Visualization:**
A **Bar Chart** titled *Category vs Total Amount* was created in the main dashboard to visually emphasize revenue differences between categories. Data labels and colour formatting were used to enhance readability.

**Objective 2: State-wise Sales Performance**

**i. General Description:**
Understanding geographic performance allows businesses to target regional demand, customize marketing campaigns, and optimize logistic strategies. This objective focused on the sales distribution across Indian states.

**ii. Specific Requirements:**

- Group the dataset by Ship-State and sum the total Amount.
- Standardize state names to avoid duplicates due to spelling inconsistencies (e.g., "Orrisa" vs. "Odisha").
- Visually highlight top-performing and underperforming states.

**iii. Analysis Results:**

- **Telangana**, **Karnataka**, and **Maharashtra** emerged as top performers.
- Some states showed significantly lower sales, which could be due to factors like low penetration, poor logistics, or demographic factors.
- The pattern indicated a concentration of orders in urbanized and e-commerce-friendly regions.

**iv. Visualization:**
A **Map Chart** titled *State-wise Sales Performance* was created using colour gradients. States with higher sales were shown in darker red, while low-performing ones were in lighter shades. This intuitive map offered immediate geographic insights.

**Objective 3: Month-wise Sales Trend**

**i. General Description:**
Tracking how sales fluctuate over months is essential for demand forecasting, seasonal inventory planning, and marketing strategy.

**ii. Specific Requirements:**

- Extract the month from the Date column.
- Calculate each month's contribution to total annual sales in percentage terms.
- Identify peak months and low-demand periods.

**iii. Analysis Results:**

- **April** recorded the highest sales with **22.36%**, followed by **May** (18.95%) and **June** (15.93%).
- A sharp drop was seen in **July (0.88%)**, followed by gradual recovery in August and September.
- The trend hints at seasonality in sales—possibly linked to festivals, end-of-season sales, or marketing campaigns.

**iv. Visualization:**
A **Line Chart** titled *Month-wise Percentage Sales Chart* was used. The smooth curve made it easy to track month-to-month changes in percentage contribution. The use of markers and labels improved visual appeal.

**Objective 4: Fulfillment Channel Breakdown**

**i. General Description:**
Fulfillment method plays a key role in delivery speed, customer satisfaction, and operational costs. This objective studied whether orders were fulfilled by Amazon or individual merchants.

**ii. Specific Requirements:**

- Group orders by the Fulfilled-By field.
- Calculate the percentage share of each fulfillment method.

**iii. Analysis Results:**

- **69.61%** of orders were fulfilled directly by **Amazon**, ensuring faster and standardized service.
- **30.39%** were handled by **Merchants**, possibly introducing variability in service quality and timelines.
- This split emphasizes Amazon's strong in-house fulfillment network.

**iv. Visualization:**
A **Pie Chart** titled *Order Fulfillment Breakdown* was added to the dashboard. It provided a clean and proportional view of fulfillment source distribution.

**Objective 5: Order Status Distribution**

**i. General Description:**
Analyzing order status helps assess operational efficiency, customer satisfaction, and potential issues like product damage, cancellations, or delays.

**ii. Specific Requirements:**

- Group orders by Status field.
- Count the frequency and percentage share of each unique status.
- Identify bottlenecks and issues affecting order completion.

**iii. Analysis Results:**

- **Shipped (66.82%)** was the highest status, followed by **Delivered to Buyer (25.75%)**.
- Minor statuses like “Cancelled,” “In Transit,” “Returned,” and “Pending” had relatively lower values.

**iv. Visualization:**
A **Bar Chart** titled *Order Volume by Status* was implemented. Each bar represented a status with its corresponding percentage, making discrepancies visually evident.

**Objective 6: Sales Channel Distribution**

**i. General Description:**
This objective assessed whether sales occurred through Amazon’s platform or external (non-Amazon) channels.

**ii. Specific Requirements:**

- Classify and compare the number of orders based on Sales Channel.

**iii. Analysis Results:**

- A staggering **99.91%** of orders were through **Amazon.in**, confirming the dominance of the platform.
- Non-Amazon orders were almost negligible, contributing only **0.09%**.

**iv. Visualization:**
A **Bubble Doughnut Chart** was created to reflect the disproportionate distribution between Amazon and Non-Amazon sales.

**Additional Functionalities Implemented**

- **Order Lookup Tool:**
  Created using XLOOKUP, this sheet allows users to enter any **Order ID** and instantly view order details (Status, Date, Fulfilment, City, State, Amount, etc.).
  Formula used:
  =XLOOKUP($A$6, 'Amazon Sale Report(Cleaned)'!B:B, 'Amazon Sale Report(Cleaned)'!D:V, "Not Found")
- **Size-Based Geographic Analysis:**
  Another map-based visualization was made to analyze sales distribution across sizes. A slicer allowed users to filter by size, dynamically updating the map's colour scale for deeper insights.

**5. Conclusion**

This project successfully transformed a large and complex Amazon sales dataset into an insightful, interactive analytical dashboard using Microsoft Excel. We uncovered critical trends in product category performance, geographic sales distribution, monthly patterns, fulfillment strategies, and order statuses. The visuals and slicers enabled dynamic exploration, supporting better business decisions.

Key takeaways include:

- "Set" products are bestsellers.
- Amazon dominates fulfillment and sales channels.
- Southern states lead in revenue generation.
- A major spike in sales occurs around April–June.

This project demonstrates the power of combining Power Query, Excel functions, and visualization tools to derive real business intelligence from raw data.

**6. Future Scope**

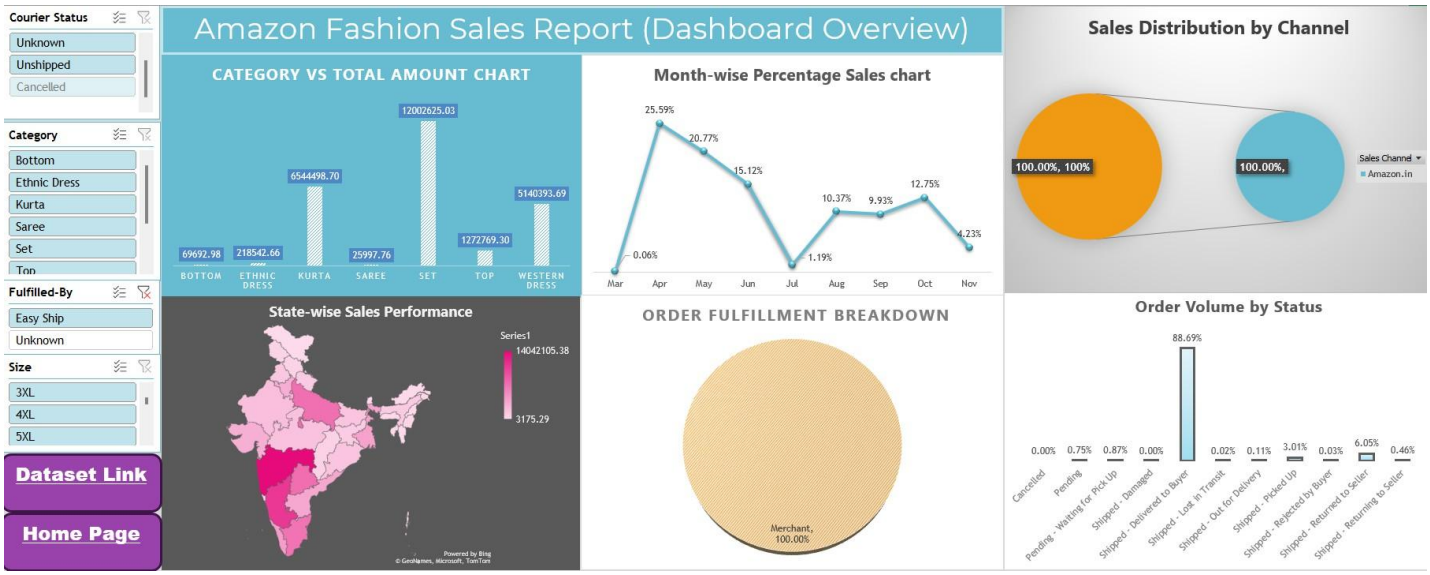
To further enhance the impact of this analysis, several extensions can be considered:

- **Predictive Modelling**: Use regression or time-series analysis to forecast monthly sales.
- **Customer Segmentation**: Apply RFM or clustering techniques to group buyers based on behavior.

- **Return & Damage Rate Analysis**: Dive deeper into why such a large share of orders are marked "Shipped – Damaged".
- **Shipping Time Analysis**: Evaluate delivery speed and performance by fulfillment type or region.
- **Power BI Integration**: Extend dashboards with advanced features like drill-throughs, bookmarks, and dynamic storytelling.
- **Automation**: Connect Excel with Power Automate to send alerts when sales drop or cancellations spike.

## 7. References

- **Dataset Source**: Tableau Public – Amazon Sales Dataset
- **Tools Used**:
  - Microsoft Excel (Dashboards, Charts, Formulas, Slicers)
  - Power Query (Data Cleaning & Preprocessing)
  - Excel Map Charts and Custom Bar Graphs
- **Formulas Used**: XLOOKUP, IFERROR, Data Type Conversion functions
- **Inspiration**: Amazon Marketplace Business Model




| Overall Sales Summary | |
|---|---|
| Total Orders | 128043 |
| Total Sales (Revenue) | 83151691.12 |
| Average Order Value | 649.404428 |
| Total Unique Ship States | 35 |

Home Page

| Order Status Breakdown | |
|---|---|
| Cancelled | 18238 |
| Pending | 656 |
| Pending - Waiting for Pick Up | 282 |
| Shipped | 77458 |
| Shipped - Damaged | 1 |
| Shipped - Delivered to Buyer | 28722 |
| Shipped - Lost in Transit | 5 |
| Shipped - Out for Delivery | 35 |
| Shipped - Picked Up | 975 |
| Shipped - Rejected by Buyer | 11 |
| Shipped - Returned to Seller | 1961 |
| Shipped - Returning to Seller | 150 |
| Shipping | 8 |
| Total Order | 128502 |

Chart Area

| Top and Bottom 3 Amounts (Cateogery wise) | | |
|---|---|---|
| Row Labels | Sum of Qty | Sum of Amount |
| Set | 45289 | 41304940.03 |
| Kurta | 45045 | 23347226.70 |
| Western Dress | 13943 | 11731880.69 |
| Top | 9903 | 5643928.30 |
| Ethnic Dress | 1053 | 833985.66 |
| Bottom | 398 | 163627.98 |
| Saree | 152 | 129765.76 |
| Dupatta | 3 | 915.00 |
| Grand Total | 115786 | 83156270.12 |

| Status | Fulfilment | Sales Channel | Ship-Service-Leve | Style | SKU | Category | Size | ASIN | Courier Status | Qty | Currency | Amount | Ship-City | Ship-State | Ship-Postal-Code | Ship-Country | B2B | Fulfilled-By |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cancelled | Merchant | Amazon.in | Standard | SET389 | SET389-KR-NP-S | Set | S | B09KXVBD7Z | Unknown | 1 | INR | 647.62 | Mumbai | MAHARASHTRA | 400081 | IN | Easy Ship | |

Enter OrderId Here
405-8078784-5731545

Home Page