

Exploratory Data Analysis and Visualization of Amazon Sale Report Dataset using Tableau

¹Mohammed Maaz Ali, ²P.Ricky Aaron, ³Sulaiman Ahmed, ⁴Diana Moses

^{1,2,3}UG Scholar, Department of AI&DS, Methodist College of Engineering and Technology, Hyderabad, India

⁴Professor, Department of CSE, Methodist College of Engineering and Technology, Hyderabad, India

Abstract - This analysis examines an Amazon India e-commerce dataset covering transactions recorded across April, May, and June 2022, spanning major shipping destinations across various Indian states and cities. The dataset captures key attributes including Order ID, Date, Status, Fulfillment, Product Category, Size, Quantity, Amount, and Shipping Locations, forming a structured foundation for evaluating multi-dimensional retail and logistics performance.

The "Set" category emerges as the highest revenue-generating product segment, followed closely by Kurtas, while individual items like Western Dresses and Tops contribute the least to overall revenue. Geographically, Maharashtra leads in both total revenue and order quantity, with major urban shipping hubs like Bengaluru, Hyderabad, and Mumbai pulling in the heaviest distribution volumes.

Logistical behavior analysis reveals that Merchant Fulfilled Network (MFN) and Fulfillment by Amazon (FBA) channels dictate the flow of order delivery, while status monitoring shows a strong majority of successfully shipped orders alongside a trackable segment of cancellations and returns. Time-based tracking reveals notable fluctuations in transaction velocity, highlighted by a gradual decline in sales amounts over daily cycles.

These findings highlight key regional performance disparities, category demand shifts, and operational volume trends, offering a data-driven foundation for improving inventory planning, catalog management, and supply chain strategies in the e-commerce sector.

Keywords: E-Commerce Analytics, Amazon India Dataset, Retail Data Analysis, Sales Performance Analysis, Consumer Purchasing Behavior, Product Category Analysis, Order Management, Supply Chain Analytics, Logistics Performance, Fulfillment by Amazon (FBA), Merchant Fulfilled Network (MFN), Inventory Management.

I. INTRODUCTION

In today's fast-paced world, Amazon has completely changed how we shop globally. The Amazon Sale Report is more than just a bunch of numbers; it's a deep look at how this retail giant is performing. It shows us everything from total revenue and sales growth to the huge volume of orders and how they're winning over new customers. By looking at this report, we can see how Prime memberships are growing, how many independent sellers are joining in, and which products are trending. It also gives us a glimpse into a logistics network so massive it's hard to imagine. Looking at these figures is like seeing the engine of the world's biggest online store—it's a clear picture of how digital shopping has transformed our lives.

To understand how big this change is, we just have to look at the records Amazon has broken. In 2023, the company brought in over \$574 billion in revenue, making it the clear leader in today's economy. This massive business serves more than 200 million loyal Prime members around the world, supported by a team of over 1.5 million employees. Amazon hasn't just joined the retail market; it has redefined it, creating a system that shapes how we all shop every day.

If we look closely at the data from the Amazon.in Sale Report, we can see this global success happening right here on a local level. In just three months between April and June 2022, there were nearly 130,000 individual sales. In those ninety days, shoppers across India bought over 116,000 pieces of clothing, generating about 78.6 million INR in revenue. This isn't just a list of sales; it shows exactly what people want to buy. The data shows that ethnic wear, like sets and kurtas, is incredibly popular. It also shows that while big states like Maharashtra and Karnataka lead in sales, thousands of orders are also heading to smaller cities. This proves that Amazon has successfully connected millions of people, even in remote areas, to a huge marketplace they couldn't reach before.

This huge volume of orders is only possible because Amazon has completely reinvented how products are delivered. In this dataset, nearly 70% of all orders were handled through Amazon's own fulfillment network (FBA),

showing how much our expectations for fast shipping have changed. We now expect same-day or next-day delivery as the standard. This logistical strength is most obvious during big sales events. Amazon Prime Day has become a global shopping event, bringing in \$12.9 billion in sales in 2023 alone. By consistently outperforming traditional sales like Black Friday, Amazon has shown it has the power to create its own peak shopping seasons, changing how we shop and speeding up the move to online retail.

However, the real success of this system is how it helps small businesses grow. Behind all those shipped packages are thousands of small and medium-sized companies that rely on Amazon to survive and thrive. In a big shift for the economy, about 60% of everything sold on the platform now comes from these third-party sellers. These independent business owners made over \$140 billion in 2023, using Amazon's reach to turn local businesses into international success stories. The ethnic wear sellers in this report are a perfect example, selling tens of thousands of items by using Amazon's world-class delivery system.

In the end, the data in these pages shows a major shift in how we buy and sell things. Every row in this spreadsheet and every package delivered represents a part of a global trade network. This Amazon Sale Report is more than just a financial summary; it's a record of how modern shopping has evolved, showing the rise of a platform that has forever changed the way the world connects.

II. LITERATURE SURVEY

Rongrong Zhang, in the research paper titled "Research on the application of machine learning in business analytics: Cases of Amazon and eBay", investigates how machine learning is utilized by e-commerce giants Amazon and eBay, focusing on its role in market prediction, customer behavior analysis, and operation optimization [1]. The analysis highlights that machine learning is crucial for Amazon in areas such as personalized recommendation systems, sales and demand forecasting, fraud detection, supply chain optimization, and marketing. Similarly, eBay leverages machine learning for risk management, search and recommendation optimization, customer service, and commodity pricing and sales forecasting. Zhang argues that machine learning helps these companies make more accurate decisions, improve operational efficiency, enhance user experience, and drive data-driven decision-making, although the research acknowledges limitations like data privacy concerns and restricted access to proprietary data and algorithms.

Monika Mishra et al., presented a paper on big data predictive analysis of Amazon product reviews. The authors

analyzed over 5GB of Amazon product reviews from 2005 to 2015, using Hadoop to store and analyze the massive dataset [2]. Their descriptive analysis revealed shopping patterns and customer sentiments based on reviews and ratings across factors like year, month, and product category. For instance, they found that the rating of 5 was the most frequent (64.57%), indicating high customer satisfaction, and that "Books" was the most popular product category based on ratings and reviews. Furthermore, the review count was found to have significantly increased from 2010 to 2014, with the highest counts occurring during the holiday months (November, December, and January). Beyond descriptive insights, the researchers implemented a recommendation model using Microsoft Azure Machine Learning Studio for predictive analysis, focusing on item recommendation and rating prediction. This model, evaluated with metrics like Normalized Discounted Cumulative Gain (NDCG), achieved encouraging results, such as an NDCG of 0.98 for item recommendation from rated items.

Jongwook Woo and Monika Mishra wrote a paper that applies several machine learning (ML) models to a massive Amazon e-commerce dataset to analyze and predict ratings and recommend products [3]. Since the Amazon product review dataset is large, with about 7 million records and 15 attributes, the authors presented a Big Data architecture suitable for storing and computation, which is not possible with traditional architecture. They developed various models in Oracle Big Data and Azure Cloud Computing services to predict review ratings and recommendations, using both traditional and Big Data algorithms. Their goal was to predict product ratings on Amazon in Big Data architecture, adopting the Spark ML platform as a Big Data solution for predictive computation, and Azure ML for a small sample dataset. The comparative analysis in the paper showed that for massive datasets (greater than 500 MB), traditional systems like Azure ML either cause a memory error or take more than a day, while distributed parallel computing systems such as Hadoop and Spark can predict the ratings in an hour.

The paper, authored by Sarah S. Alrumiah and Mohammed Hadwan, aims to study the values of implementing Big Data Analytics (BDA) in e-commerce for both vendors and consumers by analyzing fifteen selected papers [4]. The authors highlight that a primary challenge in e-commerce, due to the information revolution, is processing and analyzing the huge amount of data to gain benefits. BDA capabilities are utilized in e-commerce as a key growth direction to boost vendors' revenues and attract customers. The research found that Electronic vendors (E-vendors) leverage BDA for competitive advantages, such as understanding consumer behavior and increasing income through improved customer loyalty, while recommendation

systems derived from BDA personalize the searching and shopping experience for customers. However, the authors also note negative effects, including shopping addiction for customers and the expense of BDA tools and professionals for e-vendors, concluding that despite BDA enhancing the electronic shopping experience for both parties, the rapid growth of data remains a challenge.

Tanjim Ul Haque et al, proposed a supervised learning model to polarize a large, unlabeled Amazon product review dataset, using a combination of feature extraction approaches[5]. The authors utilized a pool-based active learning approach to label the large-scale dataset, which included reviews from Electronics, Cell Phone and Accessories, and Musical Instruments products. The methodology also involved data pre-processing steps like tokenization, removing stop words, and POS tagging. Feature extraction was performed using a combination of the Bag of Words approach, which focused on nouns and adjectives, and the TF-IDF & Chi Square approach. The model was evaluated using metrics such as accuracy, precision, recall, and F-measure, with the highest accuracy of 94.02% achieved on the Musical Instruments dataset using a Linear Support Vector Machine with 10-fold cross-validation. The authors concluded that their proposed model showed greater effectiveness and achieved better results than some related works in comparative analysis.

Ekaterina Shubina and Violeta Cvetkoska investigated the factors driving frequent purchases on Amazon using a dataset of 602 online survey responses and quantitative analysis via Power BI[6]. Their study, titled "Understanding the Drivers of Frequent Purchases on Amazon: Factors Influencing Online Shopping Behaviour," found that the product category plays a critical role, with groceries and gourmet food, followed by clothing and fashion, emerging as the most significant predictors of repeat buying, associated with a 2.51-fold increase in the likelihood of purchasing a few times a month compared to the average. Additionally, the strategic use of platform features significantly enhances the likelihood of frequent purchases, with the utilization of filter search functions being associated with a 1.42-fold increase in the probability of multiple monthly purchases, and customers who occasionally rely on product reviews exhibiting a comparable 1.41-fold increase. Shoppers who sometimes use the "save for later" feature showed a higher propensity (1.28-fold increase) to purchase items regularly, and individuals who browse the platform a few times a week are also more likely (1.27 times) to make multiple monthly purchases than the average. The authors conclude that both the product category and the way customers interact with Amazon's features substantially influence online purchase frequency.

Ben Chen et al determined the most influential factors in determining an Amazon product's sales rank by analyzing Amazon product and review data [7]. The authors found that the most significant factors were the number of products Amazon showed that other customers also bought, the number of products Amazon showed that customers also viewed, and the product's price. These findings were consistent across the Digital Music category, the Office Products category, and the Holsters subcategory under Cell Phones and Accessories. Based on their analysis, the authors suggest that those wanting to sell on Amazon should design a webpage that attracts attention, set a lower price than competitors, encourage buyers to leave 5-star reviews, and incentivize people to write thoughtful and helpful reviews.

HAO, ONGQI et al., focused on developing a short-term e-commerce sales forecasting system to address the difficulties online retailers face in predicting sales trends due to volatility and irregular demand [8]. Following the CRISP-DM methodology, Amazon sales dataset was analyzed and three models—Bidirectional Long Short-Term Memory (BiLSTM), Temporal Convolutional Network (TCN), and XGBoost were developed. The models were evaluated using R2, RMSE, MAE, and MAPE. The optimized BiLSTM model achieved the most balanced and accurate predictions with the lowest Mean Absolute Percentage Error (MAPE) of 12.91% and an R^2 of 0.8481, while the TCN model had the highest explanatory power (R^2 of 0.9336) but tended to overestimate peaks, and XGBoost performed poorly with an R^2 of 0.6586. To ensure practical utility, the author successfully deployed the best models within an interactive Streamlit dashboard, which allows users to upload data, visualize 7-day forecasts, and receive actionable business insights and AI-driven recommendations on inventory planning and operational strategy.

The discussion, "The Process of Providing Security Protection in the Amazon E-Commerce System," by Indrajit Ghosal and Dr. K. Balaji, elucidates the significant financial growth of Amazon, which the authors attribute to the implementation of an efficient data protection system in the global market[9]. Ghosal and Balaji examined various data on Amazon's net revenue and net sales using a deductive research approach, an explanatory research design, and a secondary data collection method, with thematic analysis for data interpretation. The paper asserts that Amazon's financial growth is driven by its efficient data and network security system, particularly the Amazon Web Services (AWS) cloud infrastructure, which was created to meet high security standards for customers' data privacy and access management. Furthermore, the authors discuss that the company's focus on securing customer data has cultivated loyalty and trust, which, along with expansion into various business sectors, has

contributed to Amazon's notable growth in the retail and e-commerce world.

Bessachi Houda et al, conducted a strategic analysis of artificial intelligence (AI) applications in digital marketing, using Amazon as a leading case study[10]. The research examined how AI applications, such as smart personalization, big data analysis, dynamic pricing, chatbots, and sentiment analysis, improve marketing performance and operational efficiency. The findings indicated that AI is a fundamental strategic pillar for reshaping customer experience, enhancing operational efficiency, and increasing customer loyalty. By employing key performance indicators (KPIs) and analytical tools like SWOT analysis and the balanced scorecard, the authors demonstrated that AI applications have enhanced Amazon's marketing performance metrics, including conversion rates, delivery times, and advertising effectiveness. The study concluded that integrating AI into the marketing strategy strengthens an organization's ability to adapt to rapid market changes and achieve sustainable competitive advantage, provided ethical and privacy considerations are addressed.

III. MATERIALS AND METHODS

The dataset, "**Amazon Sale Report**", is a comprehensive log of **e-commerce transactions**, containing **128,975 rows and 24 columns**. It provides a detailed snapshot of online retail operations in India, specifically focused on **apparel sales** such as "kurtas" and clothing "Sets." The dataset tracks the complete lifecycle of customer orders, capturing information from the moment of purchase to final delivery, making it highly suitable for **sales performance analysis, logistical planning, and customer demand forecasting**.

The columns within the dataset can be grouped into several key operational areas. Product and Order details are tracked via unique identifiers like "**Order ID**", "**SKU**", "**ASIN**", and attributes such as "**Category**", "**Size**", and "**Style**". Logistics and Status tracking are well-documented through variables like **Status** (e.g., Shipped, Cancelled), "**Courier Status**", "**Fulfilment**" (Amazon vs. Merchant), and "ship-service-level" (Standard vs. Expedited). Financial and Geographical metrics include the "**Qty**" ordered, the "**Amount**" paid (transacted in **INR**), and detailed shipping destinations broken down by "**ship-city**", "**ship-state**", and "**ship-postal-code**". Additionally, the dataset flags whether an order was a **B2B** (business-to-business) purchase and lists any "**promotion-ids**" applied during checkout.

From a data quality and statistical perspective, the dataset provides excellent raw material for analysis but requires some preliminary cleaning. There are substantial **missing values** in

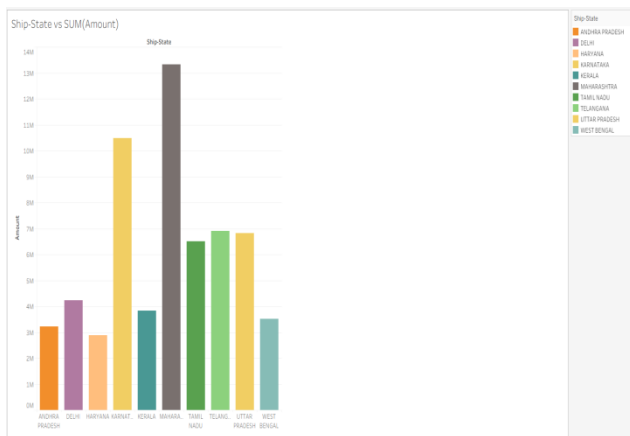
a few columns: "fulfilled-by" is missing about 89,698 entries, and "**promotion-ids**" is missing roughly 49,153 entries. There are also about 7,795 missing entries in the "**Amount**" column, which likely correspond to cancelled or unfulfilled orders. Summary statistics reveal that the **average order value** is approximately ₹648, with the **maximum order** reaching ₹5,584. Geographically, the data highlights strong urban demand, with **Bengaluru** emerging as the city with the highest volume of orders and **Maharashtra** acting as the top-performing state.

Tableau is a powerful and widely used business intelligence and data visualization platform designed to help organizations transform raw data into meaningful and interactive visual insight. It is built for both technical and non-technical users, allowing them to explore, analyze, and present data without requiring advanced programming skill. Tableau is available in multiple versions to suit different needs, including Tableau Desktop for local development, IT Server and Tableau Online for organizational sharing and collaboration, Tableau Public for free public visualization, It is for data cleaning and preparation. The tool works by first connecting to a wide variety of data sources such as Excel sheets, CSV files, Text files, etc... Once the data is connected, the user can clean and prepare it using features like data type formatting, calculated field, joins, and data blending, etc.. which allows combining multiple data sources into a single analysis. Tableau Prep is also available as a dedicated solution for more complex data transformation tasks before the data is brought into the main workspace. One of Tableau's most notable features is its intuitive drag-and-drop interface, which allows users to build a wide range of visualizations simply by dragging data fields onto a canvas. Tableau automatically suggests the most suitable chart type based on the selected data, making it easy to create bar charts, line charts, pie charts, scatter plots, heat maps, tree maps, etc... The geographic mapping feature is particularly useful for visualizing regional and spatial data, helping businesses to identify performance differences across cities and countries. It also supports the creation of an interactive dashboard, where multiple visualisations are combined into a single unified interface. These dashboards can include filters, parameters, and actions that allow users to interact with the data by clicking, hovering, or selecting different options. Dashboards can then be published and shared through Tableau Server, Tableau Online, or Tableau Public, making insights accessible to teams and decision-makers from any device or browser. In terms of advanced features, Tableau offers calculated fields and Level of Detail expression. These capabilities make Tableau suitable for a wide range of use cases including sales and revenue analysis, customer behavior tracking, financial reporting, supply chain management, healthcare analytics, and marketing performance monitoring. The key advantages of Tableau

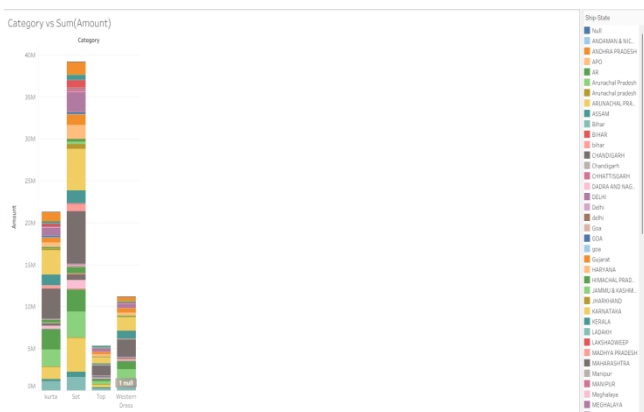
include its user-friendly design, ability to handle large datasets efficiently, and strong real-time collaboration features. Tableau remains one of the most trusted and widely adopted analytics platforms in the industry, valued for its ability to make data analysis accessible, efficient, and visually compelling for businesses of all size.

IV. DATA ANALYTICS AND VISUALIZATION:

[1] Bar Graphs:



The graph visualizes the geographical spread of e-commerce sales across the top ten Indian states, drawing from the three-month Amazon.in Sale Report (April-June 2022) focused on apparel. It confirms Maharashtra's position as the leading commercial powerhouse, recording the maximum sales amount in this dataset, with states like Karnataka also showing significant contribution.



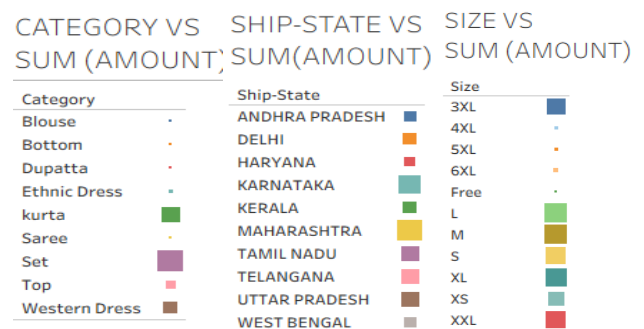
The charts show that ethnic clothing, especially 'SETS', is the top seller in every state. Using Crosstab charts, we can track total revenue by looking at the state, category, and size all at once. Color-coding makes the results easy to read: blue marks the highest sales and red marks the lowest. This helps us quickly identify which specific products and locations are making the most money.

[2] Text Tables:

SHIP-STATE VS SUM(AMOUNT)		CATEGORY VS SUM(AMOUNT)		SIZE VS SUM(AMOUNT)	
Ship-State		Category		Size	
ANDHRA PRADESH	3,219,832	Blouse	458,408	3XL	9,157,148
DELHI	4,235,216	Bottom	150,668	4XL	334,452
HARYANA	2,882,093	Dupatta	915	5XL	425,157
KARNATAKA	10,481,114	Ethnic Dress	791,218	6XL	576,249
KERALA	3,830,228	kurta	21,299,547	Free	205,193
MAHARASHTRA	13,335,534	Saree	123,934	L	13,234,886
TAMIL NADU	6,515,650	Set	39,204,124	M	13,906,754
TELANGANA	6,916,616	Top	5,347,792	S	10,629,210
UTTAR PRADESH	6,816,642	Western Dress	11,216,073	XL	12,464,966
WEST BENGAL	3,507,880			XS	7,022,375
				XXL	10,636,288

The charts visualizes ethnic wear, especially 'SETS', is the top seller in every state. Crosstab charts link the shipment state, product category, and size to total sales. The highest sales are highlighted in blue, and the lowest in red, making it quick and easy to see which exact product/location combinations are generating the most money.

[3] Heat Map:



This heat map visualizes total sales by looking at the state, category, and size together. The bigger the mark on the map, the higher the sales were. This makes it easy to spot the best-selling and worst-selling product combinations.

[4] Highlight Table:

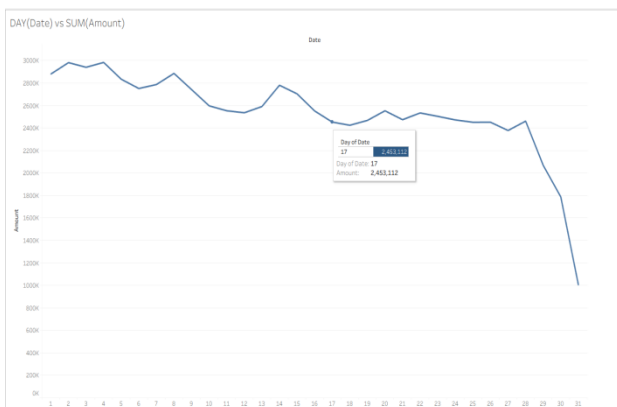
SHIP-STATE VS SUM(AMOUNT)	
Ship-State	
ANDHRA PRADESH	3,219,832
DELHI	4,235,216
HARYANA	2,882,093
KARNATAKA	10,481,114
KERALA	3,830,228
MAHARASHTRA	13,335,534
TAMIL NADU	6,515,650
TELANGANA	6,916,616
UTTAR PRADESH	6,816,642
WEST BENGAL	3,507,880

The Highlight Table effectively combines the data structure of a crosstab with the visual intensity of a heat map. It is used for multi-dimensional analysis, displaying a key metric like SUM (AMOUNT) across variables such as State, Category, and Size. The crucial feature is the color coding: blue instantly signals the configuration with the maximum total sales, while red identifies the lowest-performing combination. This visual synthesis enables users to quickly locate performance extremes and derive granular insights.

[5] Line Charts:

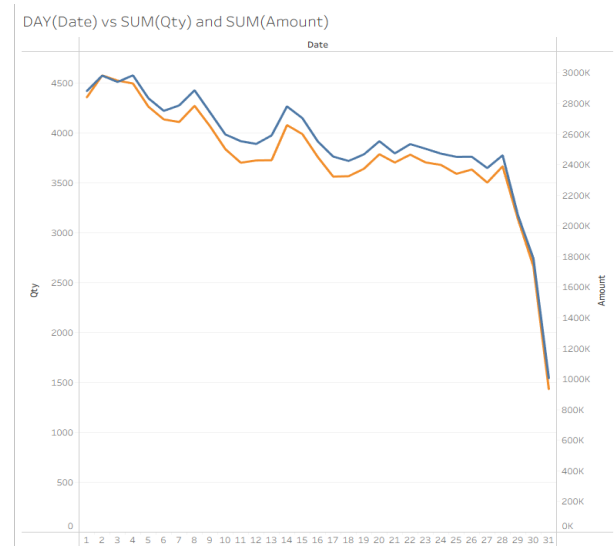


This continuous line graph tracks the total number of units sold (SUM(Qty)) over the three-month period captured in the Amazon Sale Report. It is designed to reveal month-over-month sales velocity and demand trends. The visualization clearly indicates a peak in sales volume occurred in April, which suggests a strong start or promotional activity early in the quarter. Conversely, the lowest recorded quantity was in March, potentially reflecting a pre-quarter baseline or the tail end of a previous reporting cycle. This trend analysis is critical for effective inventory and logistics planning.



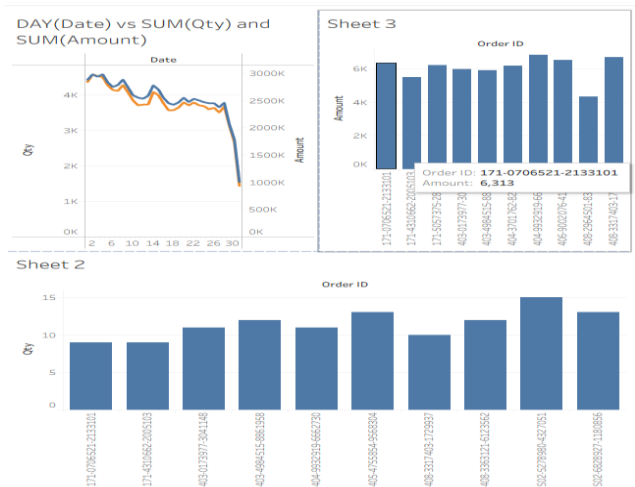
This line chart shows daily sales amounts over the course of a month, showing how revenue fluctuates day by day. The

trend starts out pretty strong, hovering near the 3,000K mark early on, but it gradually slopes downward as the days go by. There's a tooltip highlighting day 17, which brings up a specific revenue amount of 2,453,112. The most noticeable part is right at the very end of the month, where the line completely nosedives down to 1,000K by day 31.



This dual-axis chart the monthly sales performance, combining both the sheer volume of units sold (SUM(Qty)) and the total revenue generated (SUM(Amount)). It clearly shows a massive surge in activity right at the start of the month, likely driven by promotional campaigns or payday shopping habits. However, we see a distinct slowdown, with both sales and quantity hitting their lowest points towards the end of the month. This pattern is crucial for us to review our inventory and marketing spend throughout the cycle.

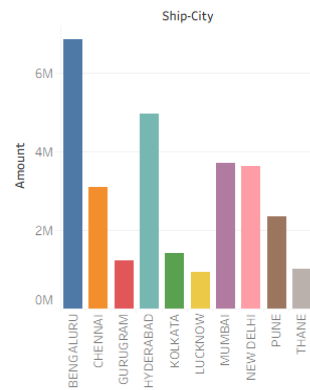
[6] Dash Board:



This dashboard follows sales by showing how order numbers and total cash start to dip as the month goes on, finishing with a big drop-off at the very end. The bar charts

break things down by specific Order IDs to show the biggest sales. For example, one huge order ending in 2133101 brought in more than 6,300. It is a way to see which big orders are carrying the team and where the month is starting to lose steam.

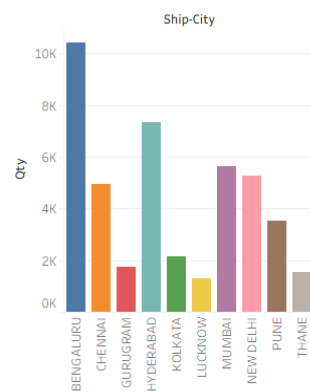
Ship-City vs SUM(Amount)



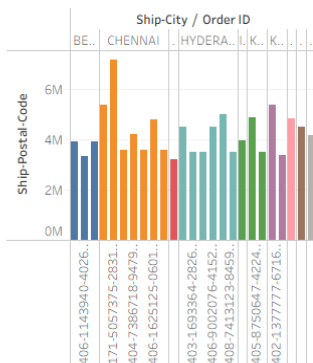
Sales Channel vs Ship-City

Sales Chan..	Ship-City
Amazon.in	BENGALURU
	CHENNAI
	GURUGRAM
	HYDERABAD
	KOLKATA
	LUCKNOW
	MUMBAI
	NEW DELHI
	PUNE
	THANE

Ship-City vs Sum(Qty)

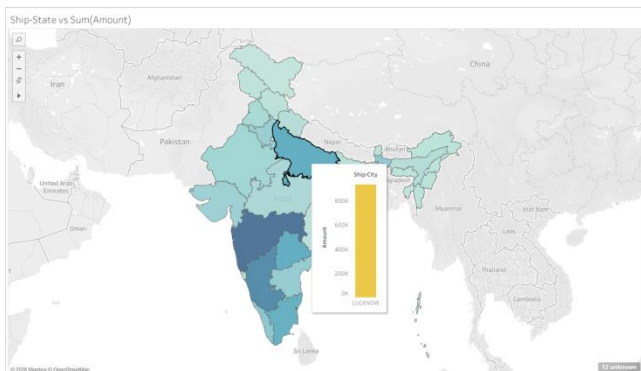


Ship-City and Order ID vs SUM(Ship-Postal-Code)



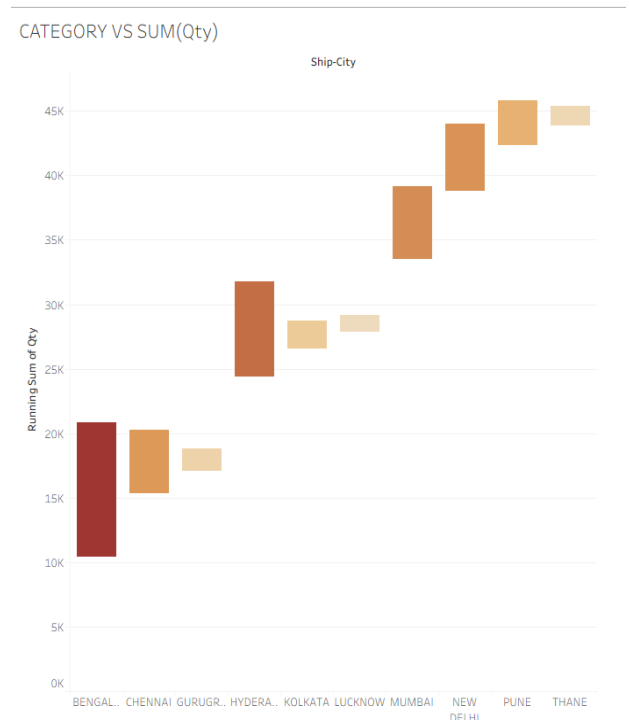
This dashboard visualizes the sales performance by showing a steady decline in both order quantity and total revenue as the month progresses, culminating in a sharp drop-off right at the end. The supporting bar charts break this data down by specific Order IDs, highlighting the top-performing transactions. For example, a single major order (ID ending in 2133101) brought in over 6,300 in amount.

[7] Tooltip:



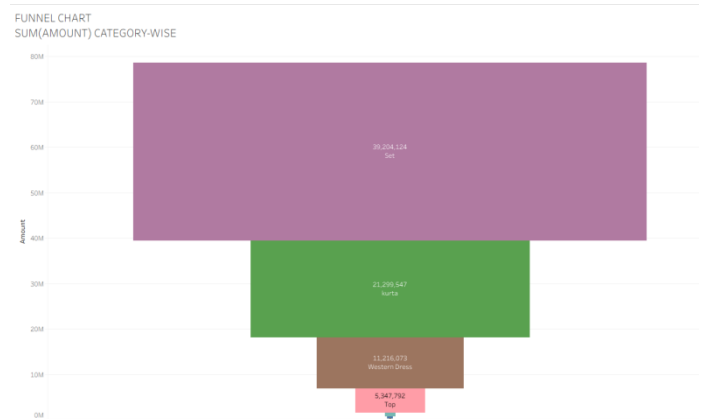
This is a Tableau map tracking order amounts across India, and Maharashtra has the highest volume based on that darker blue shading. When you hover over Uttar Pradesh, a tooltip pops up specifically for Lucknow, showing a yellow bar chart where the total amount is creeping up toward 1,000K.

[8] Water Fall:



This Gantt bar chart in Tableau shows the running total of order quantities across major shipping cities like Bengaluru, Hyderabad, and Mumbai. It builds from left to right, showing a clear picture of how each location contributes to the cumulative volume. The distinct color shading helps differentiate the cities, with large jumps visibly occurring at hubs like Bengaluru, Hyderabad, and Mumbai.

[9] Funnel:



This funnel chart breaks down total sales amounts by product category, showing a clear revenue. The "Set" category is the heavy hitter here, sitting at the top with a massive 39,204,124. As you move down the funnel, it decreases through "kurta" and "Western Dress," finally reaching the smallest segment, "Top," which brings in 5,347,792.

V. DISCUSSIONS

The e-commerce sales data provides valuable insights into customer purchasing behavior, product performance, and regional sales trends across different states in India. Using Tableau visualization, the dataset was analyzed based on shipping state, city, product category, order quantity, and timeline to identify major business patterns.

The analysis reveals that Maharashtra generated the highest sales revenue among all states, indicating that customer demand and purchasing activity were significantly stronger in this region during the observed period. Geographic analysis and the interactive map layout highlight the importance of regional market differences, showing how a few key states drive the vast majority of the overall platform revenue.

Product-wise analysis shows that the "Set" category was the highest selling and most profitable product segment by a wide margin, pulling in over 39 million INR. Kurtas also contributed strongly to overall sales, whereas Tops and Western Dresses generated comparatively lower revenue. This suggests that customers have a strong preference for traditional ethnic wear sets over individual clothing items, an insight that can help improve inventory management and supply chain planning.

Time-based analysis using line charts showed a steady decline in daily sales amounts over the course of the month, ending in a sharp drop-off down to 1,000K by the final day. Tracking these daily fluctuations helps pinpoint specific periods of high consumer spending, while catching sudden drops that might signal tracking issues or logistical bottlenecks.

The interactive dashboards, Gantt charts, funnel visualizations, and detailed tooltips created in Tableau vastly improved data interpretation and enabled efficient comparison of sales performance across different dimensions. Overall, the analysis demonstrates how business intelligence tools can transform raw e-commerce data into meaningful insights that support better operational planning and commercial decision-making.

VI. CONCLUSION

This project successfully demonstrates the application of Tableau in analyzing Amazon e-commerce sales data and converting raw transactional records into meaningful business insights. Through various visualization techniques such as geographic maps, line charts, Gantt charts, and funnel charts, the analysis identified important patterns related to sales performance, consumer preferences, and regional logistics trends.

The study reveals that Maharashtra generated the highest performance in terms of revenue and sales quantity, while the "Set" category emerged as the most profitable and frequently purchased apparel segment. The analysis also showed that customers preferred FBA (Fulfillment by Amazon) networks for order fulfillment, reflecting the growing importance of fast-shipping infrastructure and third-party merchant engagement in the online retail space.

Time-based sales analysis highlighted fluctuations in demand across the observed months of April, May, and June 2022, indicating the impact of daily transaction cycles on volume. These findings can help businesses improve inventory management, supply chain coordination, catalog depth, and operational planning.

This document is highly useful for e-commerce retailers because it provides a clear understanding of consumer behavior, geographic sales trends, and product category performance through visual analytics. E-commerce managers can use these insights to identify high-demand products, optimize stock management for top-performing apparel items, adjust regional distribution strategies, and develop better marketing promotions. The analysis also helps businesses understand geographic demand concentrations and the operational efficiency of integrated logistics networks. By using these insights, e-commerce platforms can improve fulfillment metrics, increase buyer satisfaction, and maximize overall profitability.

Overall, the project proves that data analytics and business intelligence tools like Tableau play a significant role in supporting data-driven decision-making. The visual dashboards and interactive reports make complex retail data easier to understand, allowing stakeholders to quickly identify business opportunities and operational bottlenecks. The analytical framework developed in this study can be further expanded with larger multi-year datasets and predictive machine learning analytics to improve future sales forecasting and strategic commercial planning.

REFERENCES

- [1] HAO, ONGQI. Incorporate Machine Learning in Analyse Amazon Sales Datasets to Improve Operational Strategy. *Diss. Universiti Tunku Abdul Rahman*, 2025.
- [2] Mishra, Monika, et al. "Big data predictive analysis of Amazon product review." *KSII The 14th Asia Pacific International Conference on Information Science and Technology, APIC-IST, KSII, Beijing, China*. 2019.
- [3] Woo, Jongwook, and Monika Mishra. "Predicting the ratings of Amazon products using Big Data." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11.3 (2021): e1400.
- [4] Alrumiah, Sarah S., and Mohammed Hadwan. "Implementing big data analytics in e-commerce: Vendor and customer view." *IEEE Access* 9 (2021): 37281-37286.
- [5] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." *2018 IEEE international conference on innovative research and development (ICIRD)*. *IEEE*, 2018.
- [6] Shubina, Ekaterina, and Violeta Cvetkoska. "Understanding the Drivers of Frequent Purchases on Amazon: Factors Influencing Online Shopping Behaviour." *International Journal of Business Research Management (IJBRM)* 16.2 (2025): 21-30.
- [7] Chen, Ben, et al. "Influential Factors in Increasing an Amazon products Sales Rank." *arXiv preprint arXiv:2411.04305* (2024).
- [8] Zhang, Rongrong. "Research on the application of machine learning in business analytics: Cases of Amazon and eBay." *Applied and Computational Engineering* 51 (2024): 59-65.
- [9] Ghosal, Indrajit, and K. Balaji. "The Process of Providing Security Protection in the Amazon E-Commerce System." *Journal of E-Commerce Security* 4 (2022): 1-7.
- [10] Houda, Bessachi, Bouazza Sabrine, and Habbiche Wahiba. "Strategic Analysis of Artificial Intelligence Application in Marketing: Amazon as a Model." (2025).

Citation of this Article:

Mohammed Maaz Ali, P.Ricky Aaron, Sulaiman Ahmed, & Diana Moses. (2026). Exploratory Data Analysis and Visualization of Amazon Sale Report Dataset using Tableau. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(6), 46-54. Article DOI <https://doi.org/10.47001/IRJIET/2026.106004>
