

DATA VISUALIZATION FOR AMAZON SALES REPORT

-J. RITHIKA¹, A. T. VAISHNAVI², DR DIANA MOSES³

**UG SCHOLAR, Department of CSE, Methodist College of Engineering and
Technology, Hyderabad, India^{1, 2}**

**Professor, Department of CSE, Methodist College of
Engineering and Technology, Hyderabad, India³**

ABSTRACT

The Amazon Sales Report Dataset offers a structured overview of e-commerce transactions, including product details, pricing, quantities sold, order dates, shipping status, and regional customer data. This dataset enables comprehensive analysis of sales performance, customer behavior, and market trends. It supports business insights such as identifying best-selling products, seasonal sales patterns, and delivery efficiency. Additionally, it is well-suited for data science applications, including sales forecasting, customer segmentation, and inventory optimization. With its rich and varied attributes, the dataset serves as a valuable resource for driving data-informed strategies in the online retail domain.

INTRODUCTION

The rapid growth of Amazon and its continued dominance in various industries has sparked extensive analysis of its business strategies, operational challenges, and future prospects. In 2018, Amazon had already become a global leader in e-commerce, but its ability to sustain this success in an increasingly competitive and regulated environment depended on its ability to adapt to both internal and external factors. A key focus in this context is Amazon's approach to leveraging data, analytics, and advanced technologies to better understand consumer behavior and optimize business functions.

This collection of research examines several critical aspects of Amazon's strategy and operations. It explores how the company uses a data-driven approach to improve marketing, pricing, and website optimization, enhancing its online retail environment. The growing significance of the "long tail" in Amazon's sales distribution is also discussed, highlighting the importance of niche products in driving growth. Additionally, insights into product rank and visibility on Amazon demonstrate how product positioning plays a crucial role in consumer purchasing decisions. The studies also explore the role of regulatory scrutiny, particularly in relation to consumer safety and the evolving nature of e-commerce laws, as well as the influence of sales tax policies on Amazon's operations in urban versus rural areas.

Furthermore, the research delves into the use of co-purchase networks to predict product success, the impact of customer ratings on sales, and the importance of web design in enhancing user experience (UX) and boosting sales. By focusing on these critical areas, the research provides a comprehensive understanding of how Amazon's strategies have shaped its success and the factors that will continue to drive its future growth in the e-commerce space.

LITERATURE REVIEW

Wells et al., The case study is evaluating Amazon's position in 2018 and considering its future trajectory. While Amazon had become a dominant player in several industries, its future success depended on navigating both operational challenges and external factors, such as regulatory scrutiny and increased competition [1]. Panagiotelis, Anastasios, Michael S. Smith, and Peter J. Danaher., highlights the importance of a comprehensive, data-driven approach to understanding online consumer behavior. By modeling the relationships between visit behavior, purchase incidence, and sales volume, retailers can make more informed decisions about marketing, pricing, and website optimization to drive growth in online retail environments [2]. Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith., consists that the *long tail* is becoming an increasingly significant feature of online retail platforms, especially Amazon. As the company continues to expand its product offerings and attract a diverse customer base, the sales distribution curve has shifted to accommodate a large number of niche products. While the head of the curve still captures the largest portion of sales, the long tail is an important source of growth and revenue. This transformation illustrates how the internet and e-commerce platforms are reshaping traditional retail dynamics, creating new opportunities for both consumers and retailers alike [3]. He, Sherry, and Brett Hollenbeck., provides empirical evidence of the significant relationship between product rank and sales on Amazon. It underscores the importance of ranking in driving consumer purchasing behavior and offers insights that can help sellers optimize their listings to maximize sales. The findings are also relevant for anyone interested in understanding how the Amazon marketplace works and how product visibility influences overall success [4]. Prasad, Utpal, et al., aim to utilize the co-purchase network to predict Amazon product sales ranks more effectively. Amazon sales ranks are a key indicator of product visibility and success, and understanding the factors that influence this rank can be valuable for both sellers and researchers [5]. Etumnu, Chinonso E., et al., highlights the ongoing efforts by regulatory bodies like the CPSC to ensure that e-commerce platforms like Amazon maintain consumer safety. It may influence how Amazon collaborates with regulators, strengthens its policies for product safety, and addresses concerns related to third-party sales [6]. Afonso, Whitney B., is important because it provides actionable insights for sellers on platforms like Amazon, especially within the grocery sector,

where consumer trust and convenience are paramount. By understanding the impact of rating distribution on sales, sellers can optimize their strategies for garnering positive feedback and managing negative reviews, ultimately enhancing their sales performance [7]. Rao, Y. V., is significant because it highlights the varying impacts of sales tax policies on different geographic regions, particularly the divide between urban and rural areas in terms of the benefits derived from e-commerce. It provides valuable insights for local policymakers who are grappling with the challenges of adapting to the rapid growth of online shopping and seeking ways to capture sales tax revenue from digital transactions [8]. Santana, Francisca Nunes., analytics play a critical role in the success of e-commerce companies, with Amazon serving as a prime example of how data-driven strategies can enhance business performance. By leveraging big data, machine learning, and other analytics tools, Amazon has been able to optimize key business functions, from customer experience and pricing to inventory management and logistics. The paper suggests that e-commerce companies looking to remain competitive should invest in analytics technologies and develop data-driven strategies to drive growth and profitability [9]. Intelligence, Amazon Artificial General., is that web design is intricately linked to sales performance on e-commerce platforms like Amazon. By enhancing UX, simplifying navigation, personalizing recommendations, and optimizing the checkout process, companies can significantly boost sales [10].

MATERIALS USED

1.Context

This sales report dataset is useful for e-commerce business owners, sales teams, supply chain managers, and financial analysts to track and optimize performance. It helps sellers analyze product demand, order status, and revenue trends, ensuring efficient inventory management and fulfillment. Marketing teams can use it to measure the impact of promotions and regional sales performance. Logistics teams benefit by identifying shipping delays, cancellations, and optimizing delivery routes. Overall, this dataset enables data-driven decision-making to enhance sales, customer experience, and profitability. Each order in the system is identified by a unique Order ID, which serves as the primary reference for tracking and management. The Date field indicates when the order was placed. The Status reflects the current stage of the order, such as *Shipped*, *Delivered*, or *Cancelled*. The Fulfilment field specifies whether the order was handled by Amazon or a Merchant, while the Sales Channel shows the platform through which the order was made, such as *Amazon.in*. Lastly, the Ship Service Level describes the type of shipping selected by the customer, typically categorized as *Standard* or *Expedited*. Each product is described by its Style, which refers to the product's name or design. The SKU (Stock Keeping Unit) is a unique identifier

assigned to each product, used for inventory tracking and management. The Category defines the type of product, such as *Kurta*, *Dress*, or *Set*, helping to organize and classify items for easier navigation and analysis. The Qty field indicates the quantity of units sold for a particular product in a given transaction. The Currency specifies the type of currency used for the transaction, such as *INR* (Indian Rupee). The Amount represents the total sale value, calculated based on the quantity sold and the unit price, and is expressed in the specified currency. The Ship City indicates the city where the order is being delivered, while the Ship State refers to the corresponding state or region within that country. The Ship Postal Code identifies the specific delivery area using a postal or ZIP code. Lastly, the Ship Country denotes the country of delivery, ensuring accurate international or domestic shipping. The Promotion IDs field captures any discounts or promotional codes applied to the order, providing insight into marketing-driven sales. The B2B field indicates whether the order was a business-to-business (B2B) transaction, which can differ in pricing, volume, or terms from regular consumer sales. The Fulfilled By field specifies the shipping method used—such as *Easy Ship*—helping to identify how the order was processed and delivered.

2.Data Description

The Qty (Quantity Sold) refers to the number of items sold in each order, providing insight into sales volume. The Amount represents the sales revenue generated from each transaction. Total Orders indicates the number of orders placed, while Cancelled Orders reflects the count of transactions that were later canceled. Revenue Per Order shows the total sales amount generated per individual order, offering a measure of order value. If available, the Shipping Cost field captures the expense incurred in delivering the product to the customer. Each order is tracked using a unique Order ID, which serves as the primary identifier. The Date indicates when the order was placed, while the Status reflects its current state, such as *Shipped*, *Delivered*, or *Cancelled*. The Fulfilment Type specifies whether the order was handled by Amazon or a Merchant, and the Sales Channel identifies the platform through which the order was placed, such as *Amazon.in*. The Ship Service Level details the type of shipping selected, like *Standard* or *Expedited*. Each product falls under a specific Category, such as *Kurta*, *Top*, or *Dress*, and is identified by a unique SKU (Stock Keeping Unit). The shipping destination is captured through the Ship City, State, and Country fields. Any Promotion IDs represent discounts or promotional offers applied to the order. Lastly, the B2B field indicates whether the transaction was a *Business-to-Business* (B2B) or a standard *Business-to-Consumer* (B2C) order.

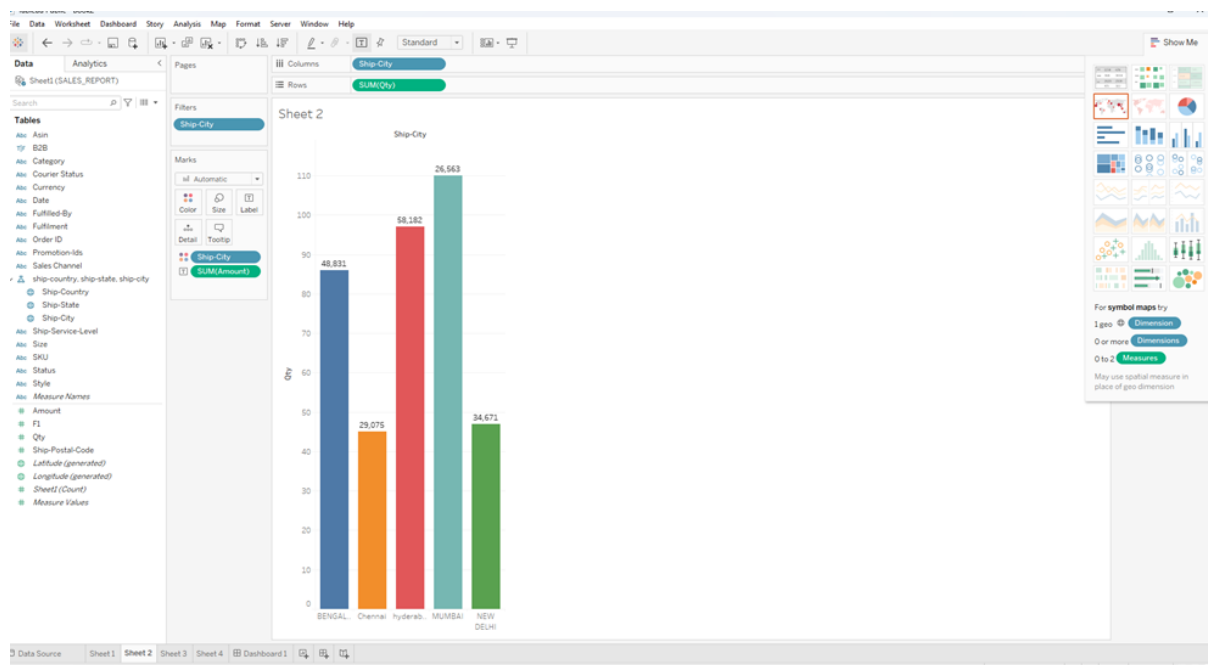
3.Data Aggregation

The sales performance data reveals that Kurta is the highest-selling product category, closely followed by Western Dress. The average order value typically ranges between ₹400 and ₹750, varying based on the product category. A higher number of shipped orders compared to cancelled orders reflects efficient order fulfillment overall. In terms of order fulfillment and shipping, Amazon-fulfilled orders generally offer faster delivery speeds than those fulfilled by merchants, with standard shipping being the most commonly chosen method. Most deliveries are directed to major urban centers such as Mumbai, Bengaluru, and Chennai. On the other hand, cancellation trends show that merchant-fulfilled orders are more likely to be cancelled, particularly for high-priced products, and some regions report more cancellations, possibly due to delivery challenges. Promotional activity plays a notable role in sales, as orders with active promotions tend to see higher volumes, although many transactions still occur at full price without any promotion IDs. Lastly, the majority of transactions are B2C (Business-to-Customer), with only a small portion of orders marked as B2B (Business-to-Business), highlighting the consumer-focused nature of the sales.

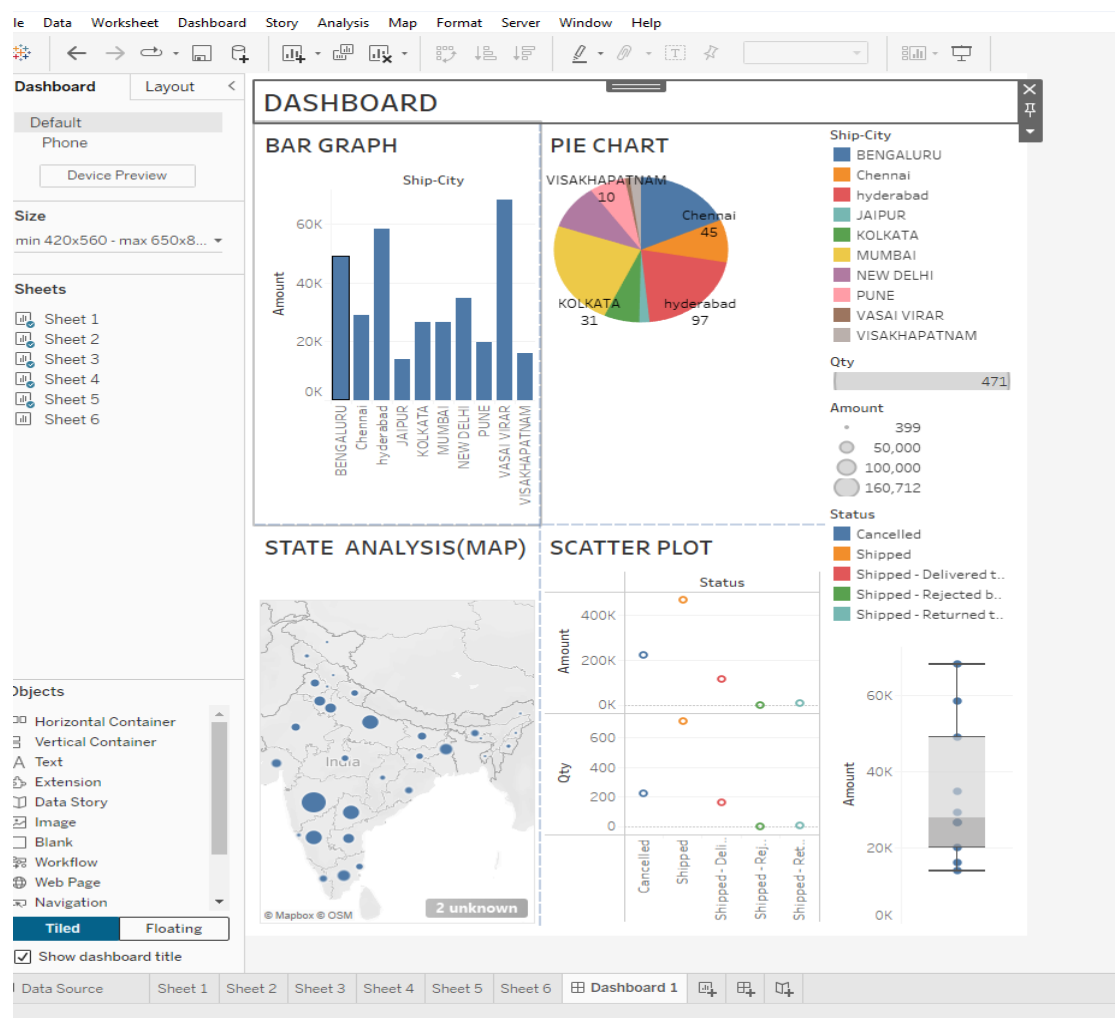
4.Stakeholders

The stakeholders of this sales report dataset include e-commerce sellers, sales & marketing teams, supply chain managers, financial analysts, and customer support teams. Sellers and business owners use the data to track sales performance, revenue trends, and customer demand. Marketing teams analyze sales patterns and promotional effectiveness, while logistics managers focus on fulfillment efficiency and reducing cancellations. Financial analysts monitor revenue, profit margins, and cost variations to ensure profitability. Lastly, customer support teams leverage this data to resolve order-related issues and enhance customer satisfaction.

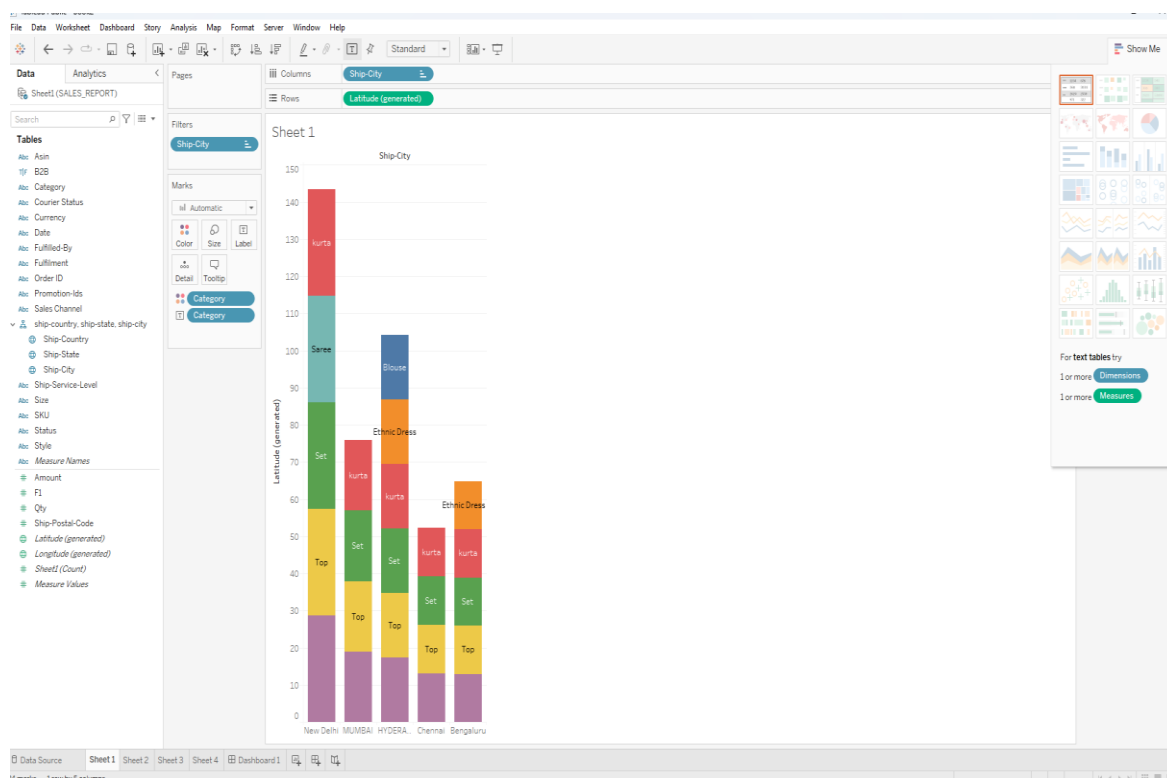
DATA VISUALIZATION



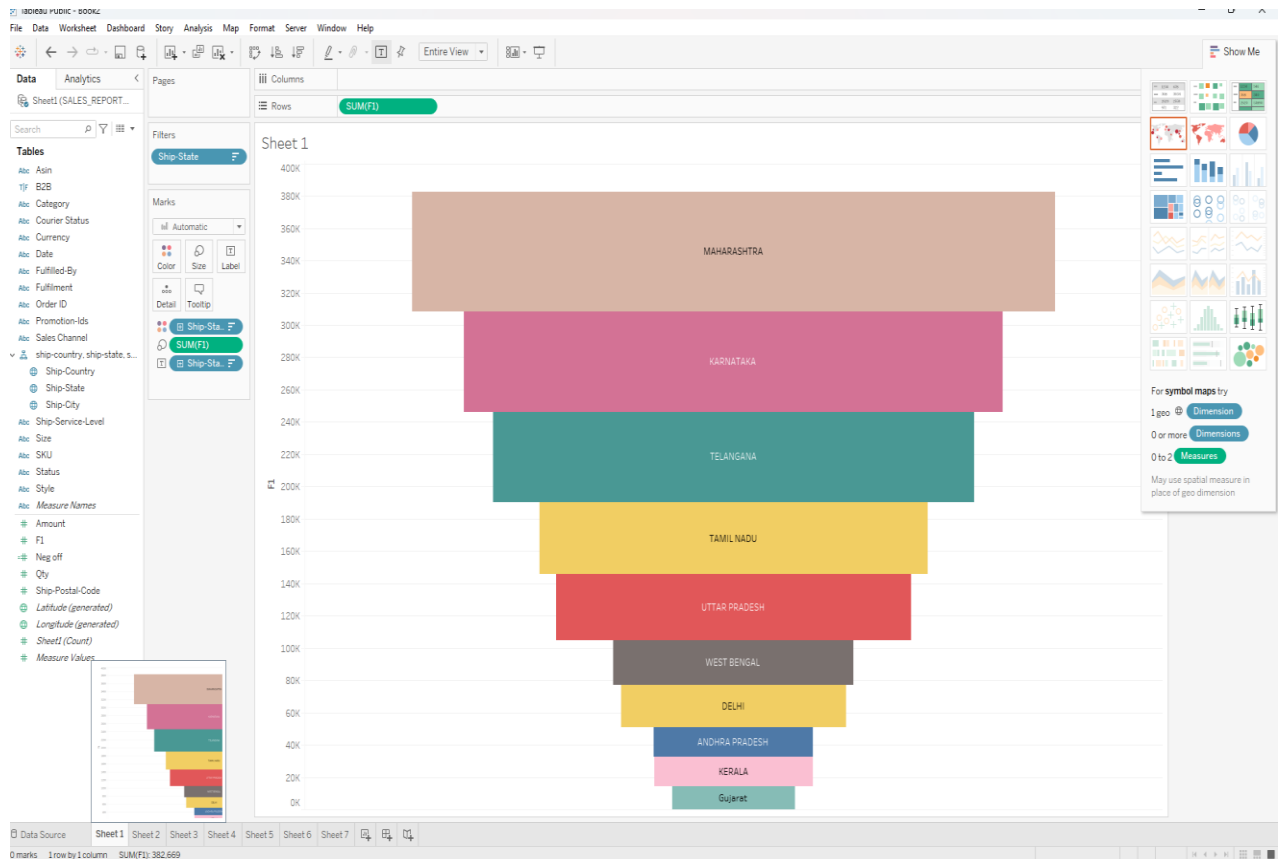
The bar chart visualizes the total quantity (Qty) of sales across different cities using data from a sales report. The cities include BENGALURU, Chennai, Hyderabad, MUMBAI, and NEW DELHI. Hyderabad recorded the highest quantity at 58,182 units, followed by BENGALURU (48,831), MUMBAI (26,563), NEW DELHI (34,671), and Chennai (29,075). The color differentiation enhances visual clarity, and the bars are labeled with the respective sales quantities. The chart is filtered by "Ship-City," and it uses SUM(Qty) on the Rows shelf and Ship-City on Columns. It provides a quick comparative view of city-wise sales performance.



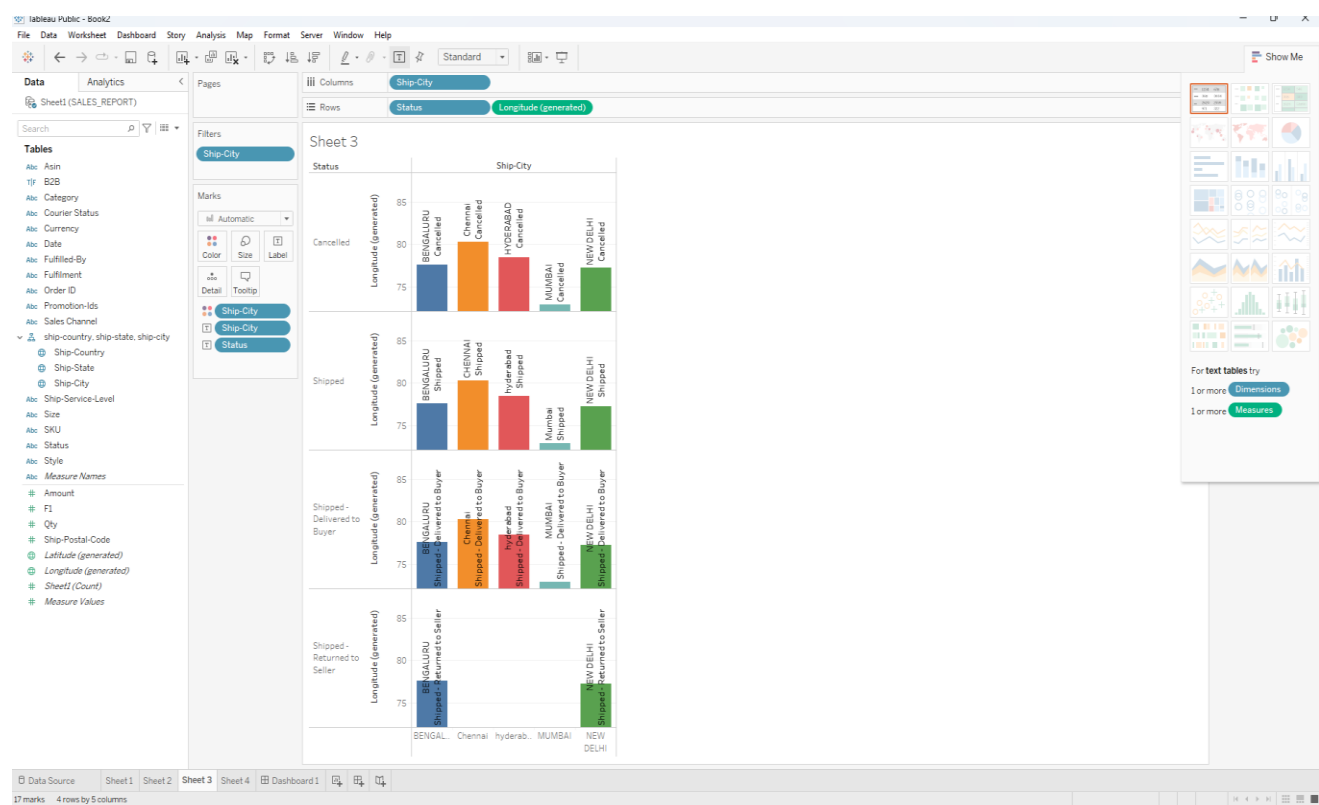
The dashboard provides a comprehensive view of sales performance across various Indian cities. The bar graph shows sales amounts by city, with VISAKHAPATNAM and HYDERABAD leading. The pie chart visualizes order quantities, highlighting Hyderabad with the highest number (97), followed by Chennai and Kolkata. The map offers a state-level distribution of sales, using bubble sizes to indicate volume. The scatter plot analyzes sales quantity and amount across different shipping statuses such as Shipped, Cancelled, and Returned. Filters on the right allow users to interact with the data by city and status. The dashboard supports dynamic insights into sales trends and distribution.



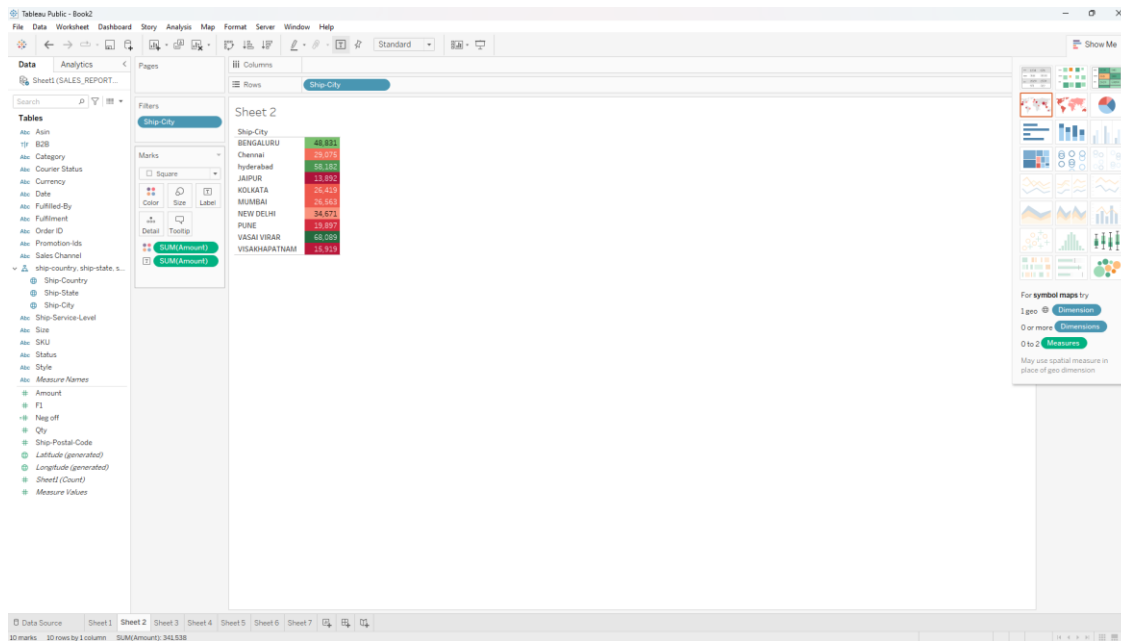
The stacked bar chart visualizes product category distribution across five Indian cities: New Delhi, Mumbai, Hyderabad, Chennai, and Bengaluru. Each bar represents a city and is segmented by categories such as Saree, Kurta, Set, Top, Blouse, and Ethnic Dress. New Delhi shows the highest product variety and quantity, with significant contributions from Sarees and Sets. Hyderabad and Mumbai also display diverse product sales, whereas Bengaluru and Chennai have lower quantities. The chart uses color to distinguish categories and leverages the latitude field to align bars. It effectively highlights category-wise performance and inventory distribution across key locations.



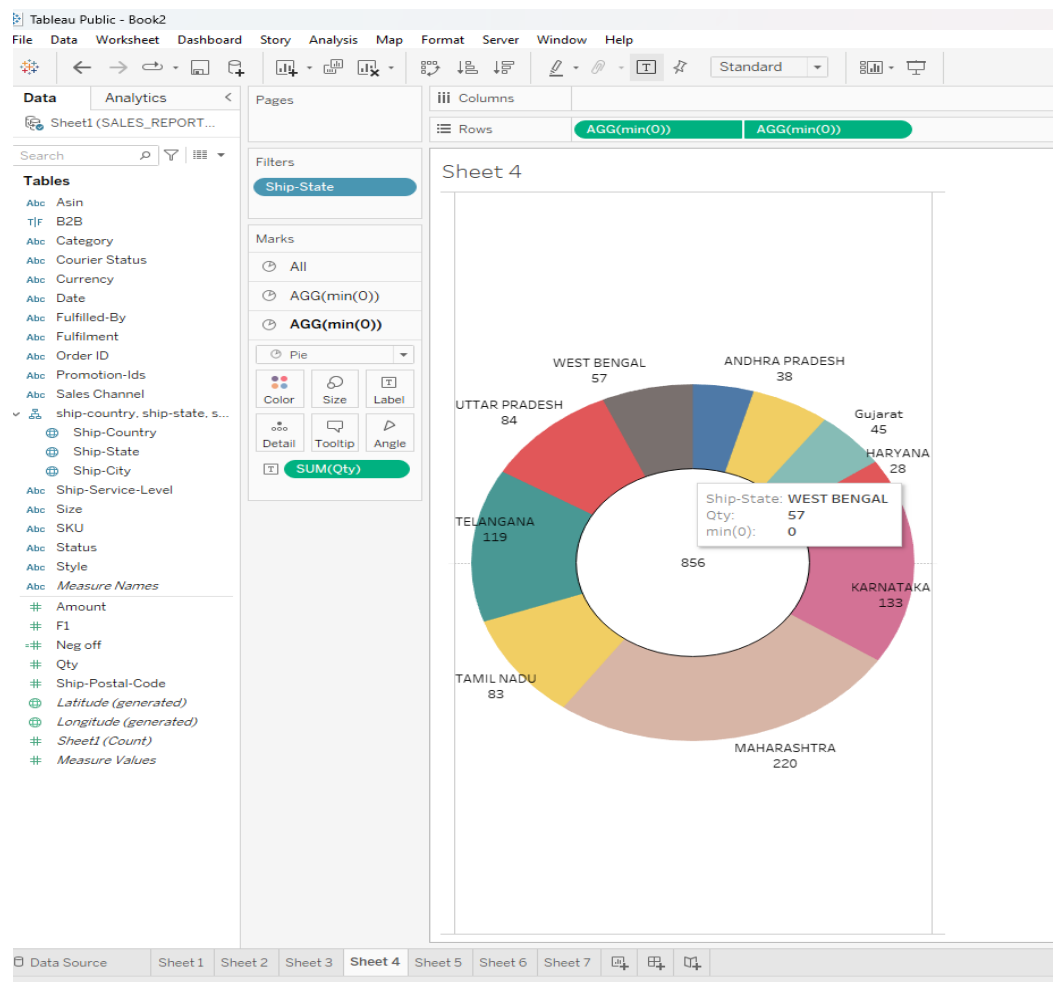
The funnel chart displays the distribution of a metric (F1) across various Indian states based on shipping data. Maharashtra leads with the highest value, followed by Karnataka, Telangana, and Tamil Nadu. Uttar Pradesh, West Bengal, and Delhi occupy mid-tier positions, while Andhra Pradesh, Kerala, and Gujarat show relatively lower figures. Each state is color-coded for clarity, and the descending funnel shape effectively illustrates the comparative performance of states. The chart helps in quickly identifying top-performing regions in terms of shipping-related metrics, enabling strategic decision-making for regional sales and logistics optimization. A Ship-State filter allows for interactive exploration.



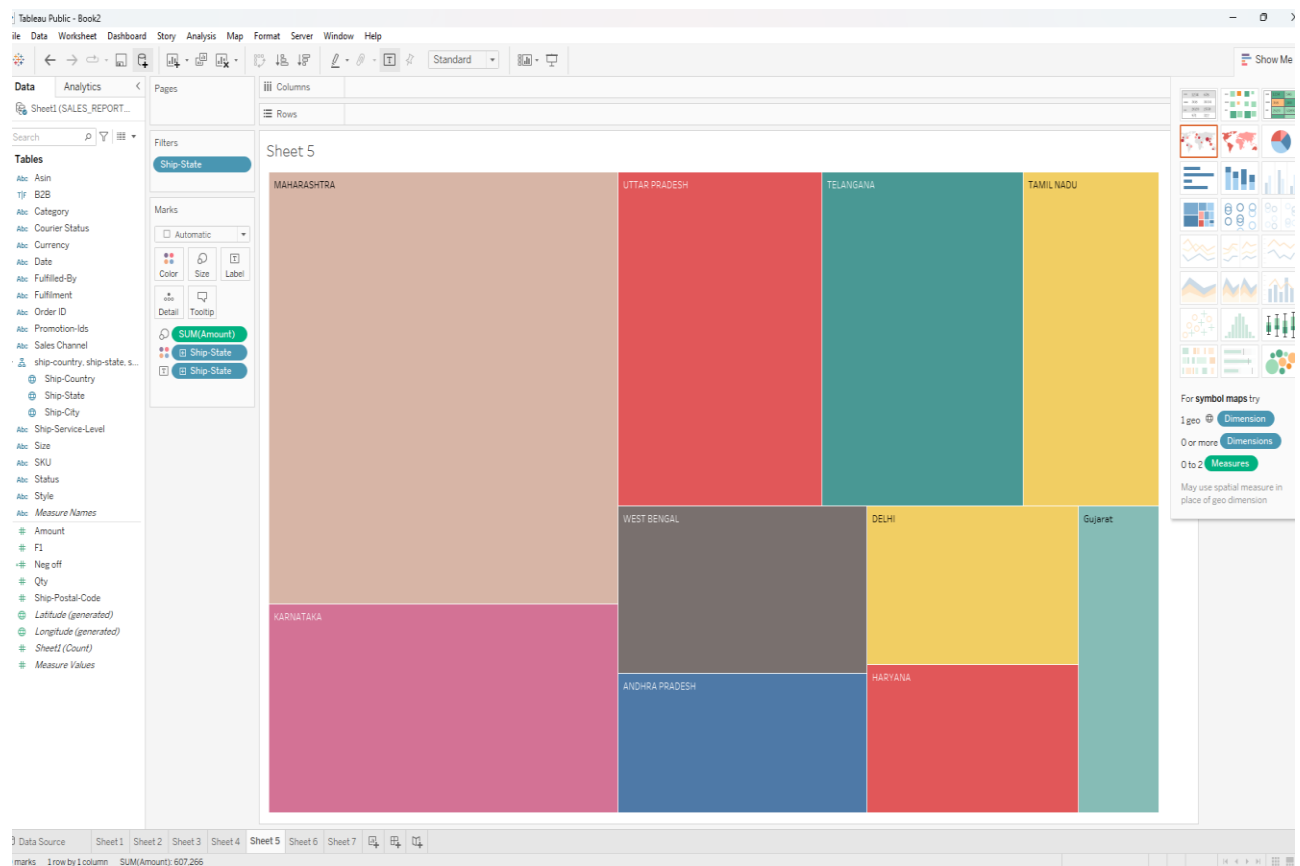
The visualization presents a bar chart segmented by shipping status across various cities, with each panel representing a different status category—Cancelled, Shipped, Shipped-Delivered to Buyer, and Shipped-Returned to Seller. Cities like Bengaluru, Chennai, Hyderabad, Mumbai, and New Delhi are shown with varying order volumes under each status. Bengaluru appears consistently across all categories, indicating high shipping activity. The bars use color-coding for better distinction and are plotted against longitude to reflect geographical spread. This view enables quick analysis of order fulfillment trends, highlighting performance issues or success in logistics across different cities and shipping outcomes.



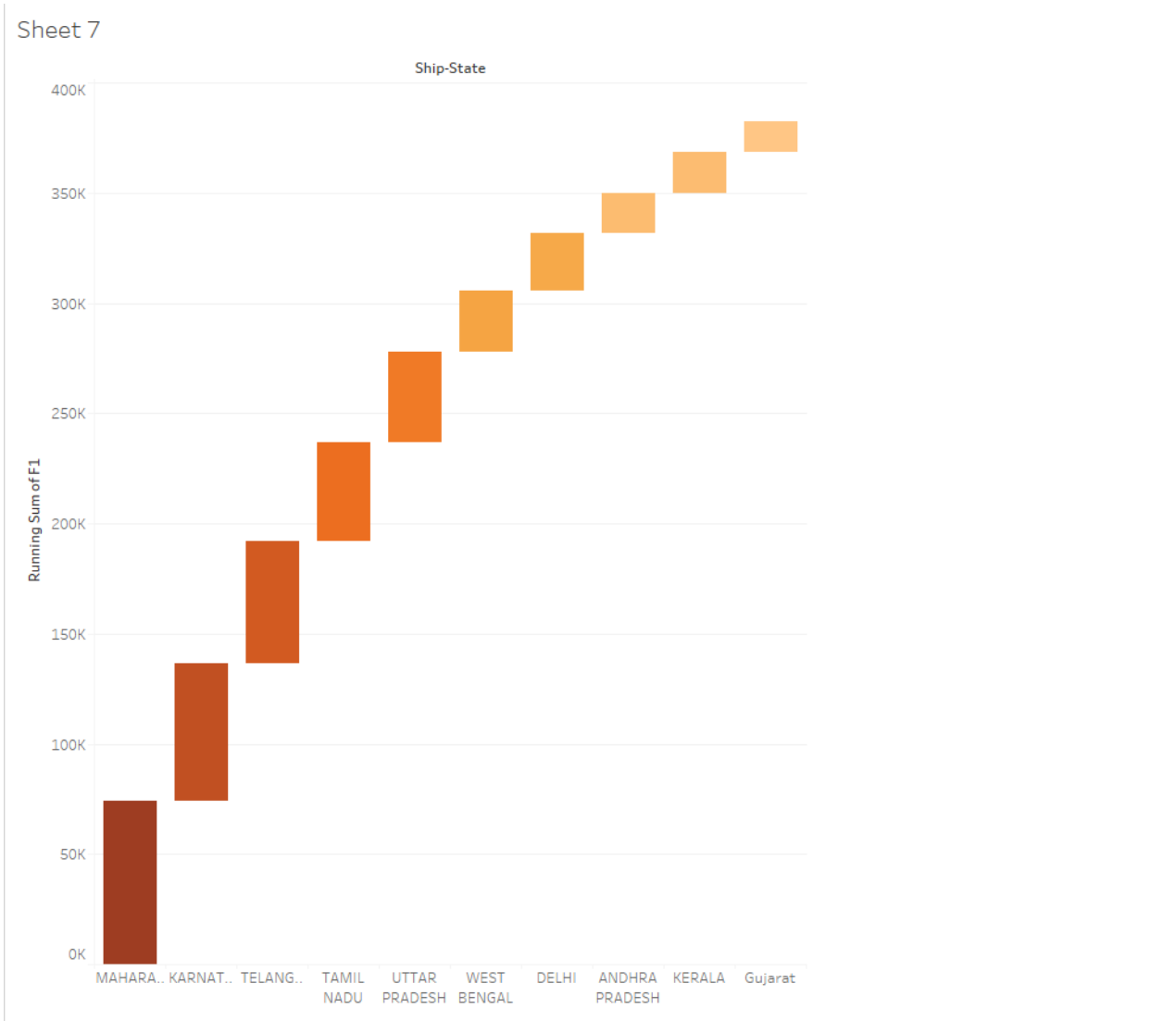
The visualization displays a heat map of total sales amounts across various cities, using color gradients to represent value intensity. Vasai Virar recorded the highest sales with ₹68,089, followed by Hyderabad and Bengaluru with ₹58,182 and ₹48,831, respectively. Cities like Jaipur, Visakhapatnam, and Pune show comparatively lower sales figures, highlighted in darker red shades. The use of a diverging color scale—from green (high) to red (low)—helps easily identify performance disparities among cities. This representation enables quick assessment of city-wise revenue contribution, highlighting areas of strong and weak sales performance for better strategic planning and targeted business actions.



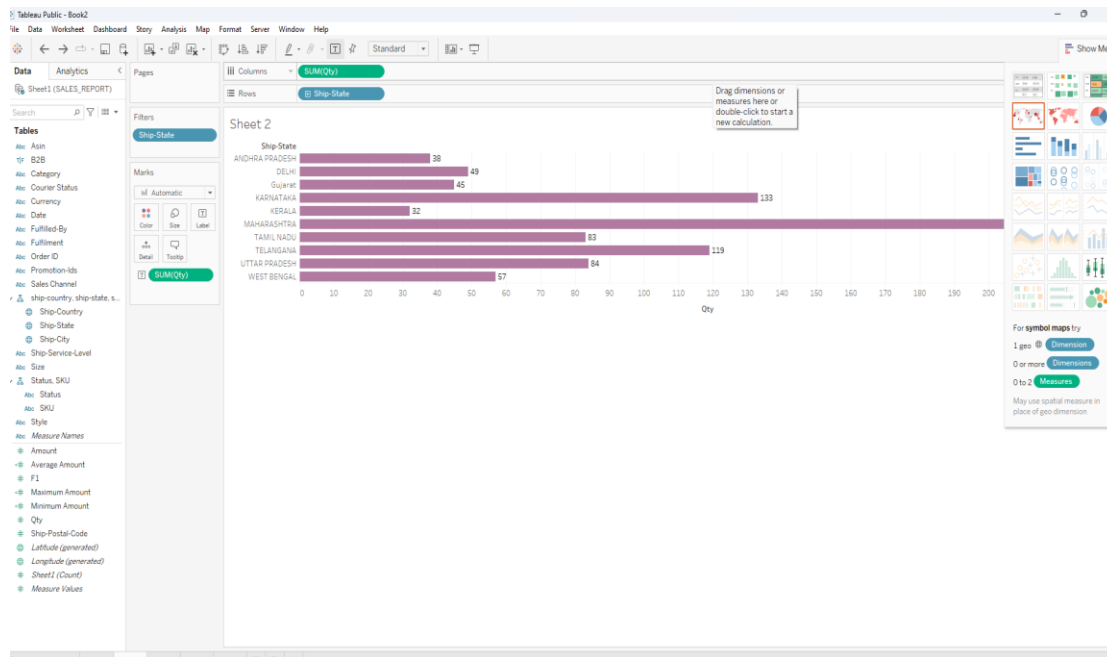
The visualization presents a donut chart illustrating the quantity of orders (Qty) distributed across different Indian states. Maharashtra leads with the highest quantity of 220, followed by Karnataka (133) and Telangana (119), indicating strong order volumes in these regions. Other states like Andhra Pradesh, Haryana, and Gujarat display significantly lower quantities, with Haryana at the lowest (28). The chart effectively segments each state's contribution, with labels for easy comparison and a central total of 856 units. This layout helps highlight regional performance, allowing businesses to identify high-performing areas and potential regions for improvement in distribution or marketing efforts.



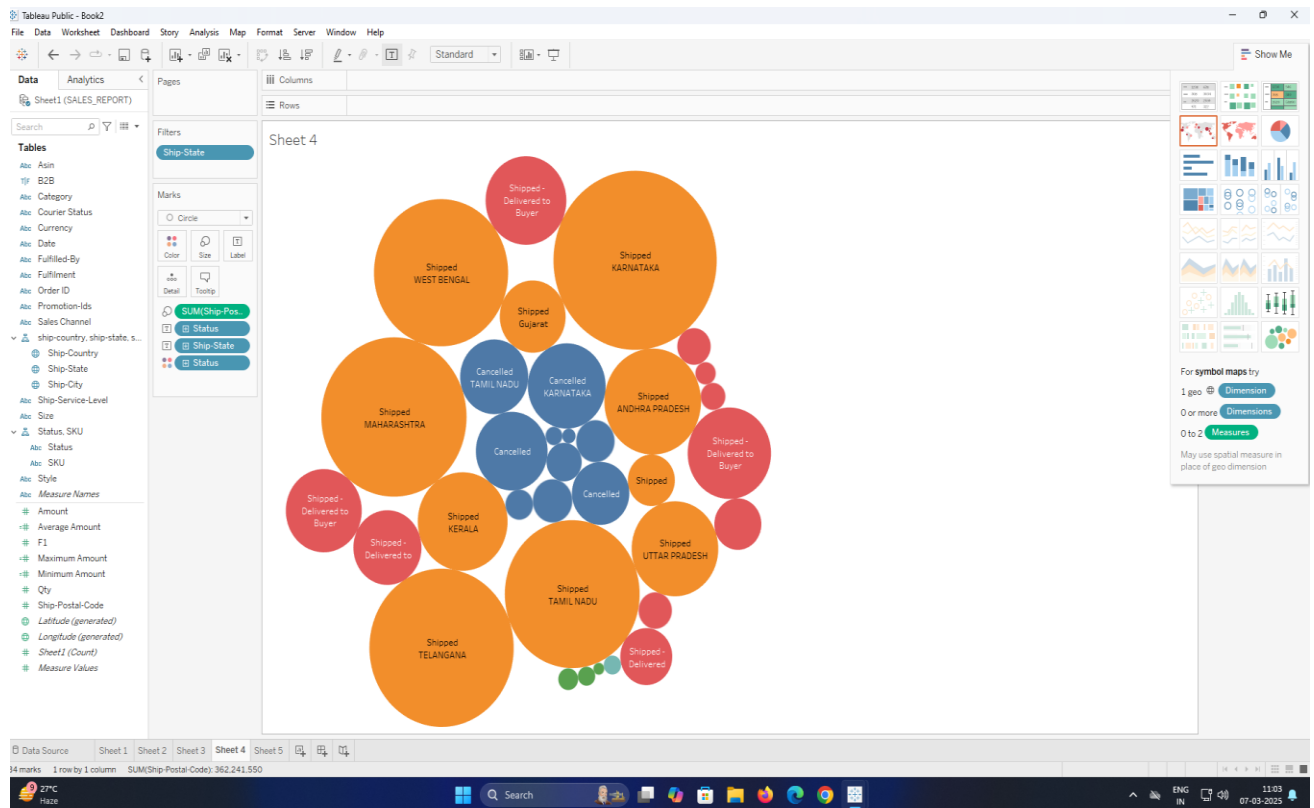
The treemap visualization in Sheet 5 displays the total sales amount across different Indian states. Each rectangle's size corresponds to the sales volume (SUM(Amount)), making it easy to compare contributions at a glance. Maharashtra holds the largest share, indicating it is the highest revenue-generating state, followed by Karnataka, Uttar Pradesh, Telangana, and Tamil Nadu. States like Gujarat, Haryana, and Delhi contribute smaller amounts in comparison. This layout is effective for identifying top-performing regions and supports strategic decision-making for resource allocation, regional marketing, or sales efforts. The color differentiation enhances visual clarity and highlights distribution patterns across the states.



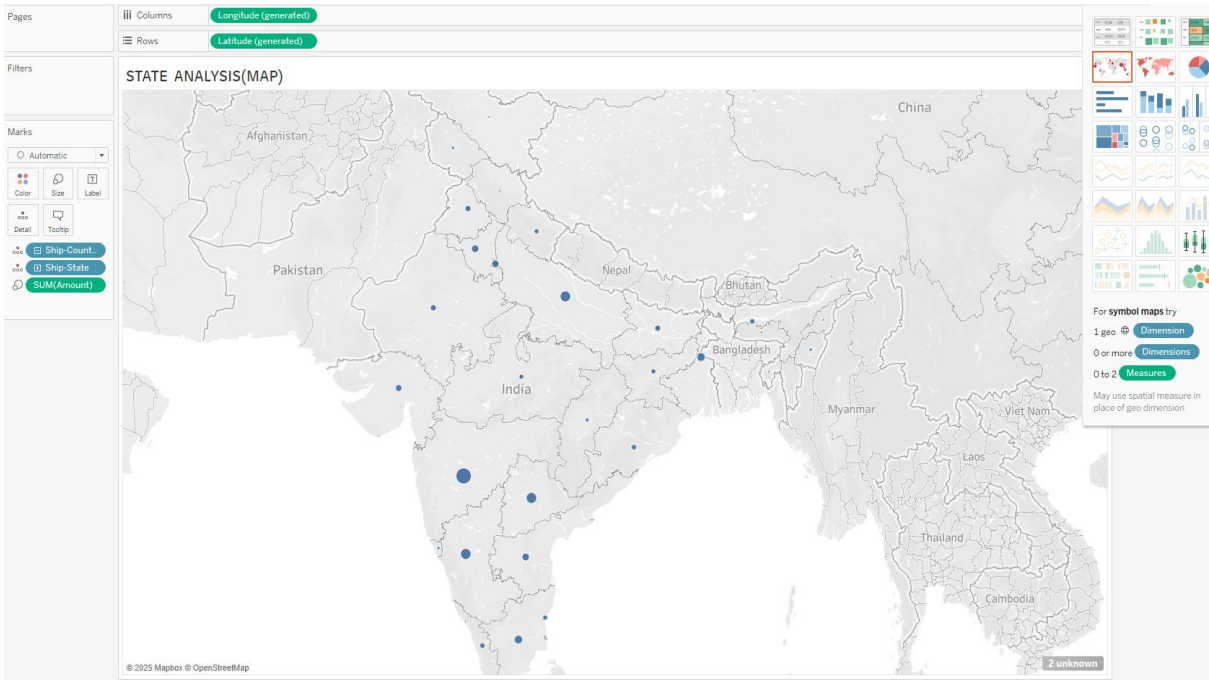
The screenshot represents a bar chart displaying the running sum of the field **F1** across various Indian states, organized by the ship-state dimension. The chart highlights a cumulative trend, where each state’s contribution adds to the total sum, effectively showing the progressive buildup of values. Maharashtra and Karnataka lead with the highest initial values, followed by Telangana and Tamil Nadu. As the chart progresses, the increments become smaller, indicating that states like Gujarat, Kerala, and Andhra Pradesh contribute the least to the cumulative total. The use of gradient colors enhances the visual understanding of distribution and progression across states.



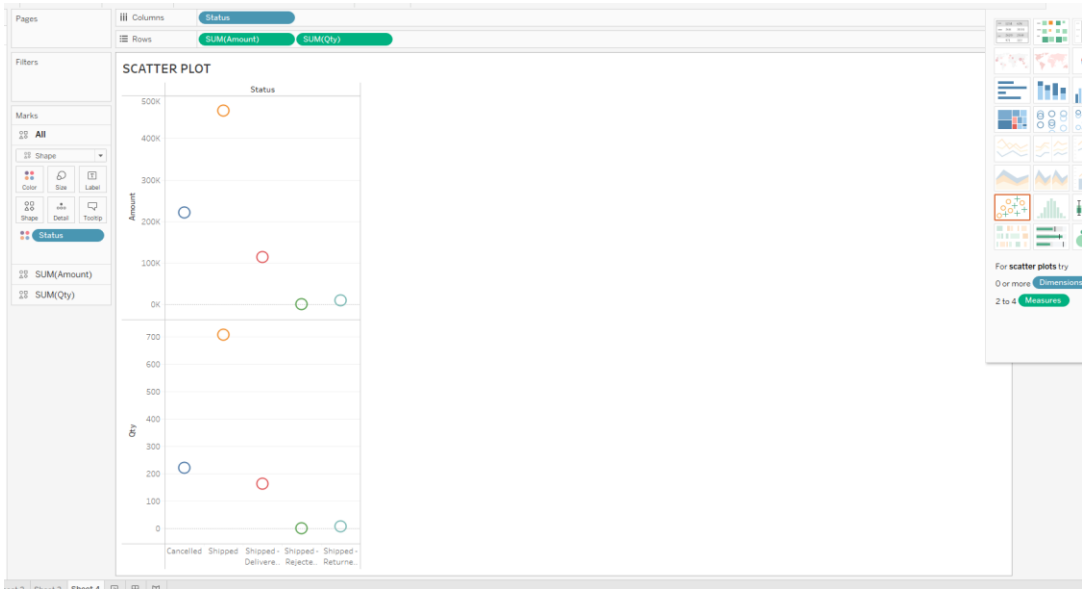
The image represents a horizontal bar chart illustrating the sum of product quantities(Qty) shipped across various Indian states. Each bar represents a state's total shipped quantity, with Maharashtra leading at 133 units, followed by Telangana (119) and Uttar Pradesh (84). States such as Kerala, Gujarat, and Andhra Pradesh show comparatively lower quantities, with Kerala at the bottom with 32 units. The chart is sorted for easy comparison, and the use of labels on bars enhances clarity. The filter on ship-state allows dynamic exploration of specific states, making this visualization effective for analyzing regional shipment distribution.



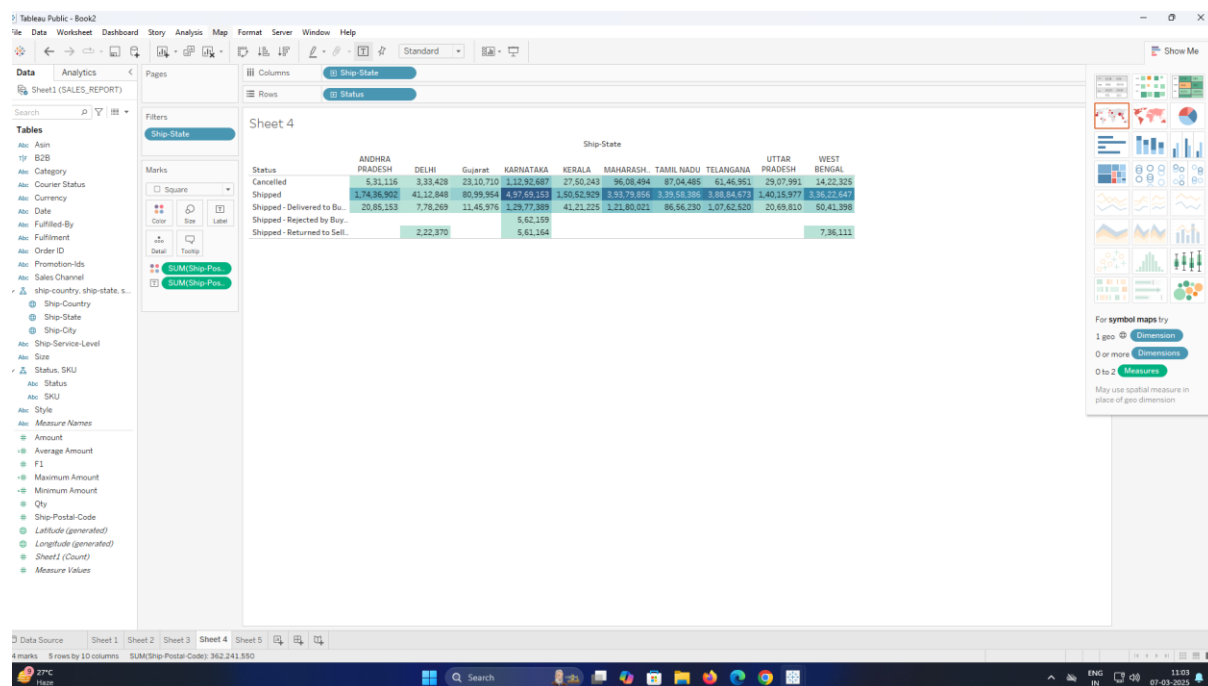
The bubble chart visualizes the shipping status of orders across various Indian states. Each bubble represents a state and is color-coded by status: orange for "Shipped," blue for "Cancelled," red for "Delivered to Buyer," and green for other statuses. The size of each bubble corresponds to the total shipping postal codes, indicating volume. Karnataka, Maharashtra, and West Bengal have the largest volumes of shipped orders, while Tamil Nadu and Karnataka also show notable cancellations. The chart helps identify regional trends in order fulfillment performance. It offers a clear overview of delivery success and issues across states.



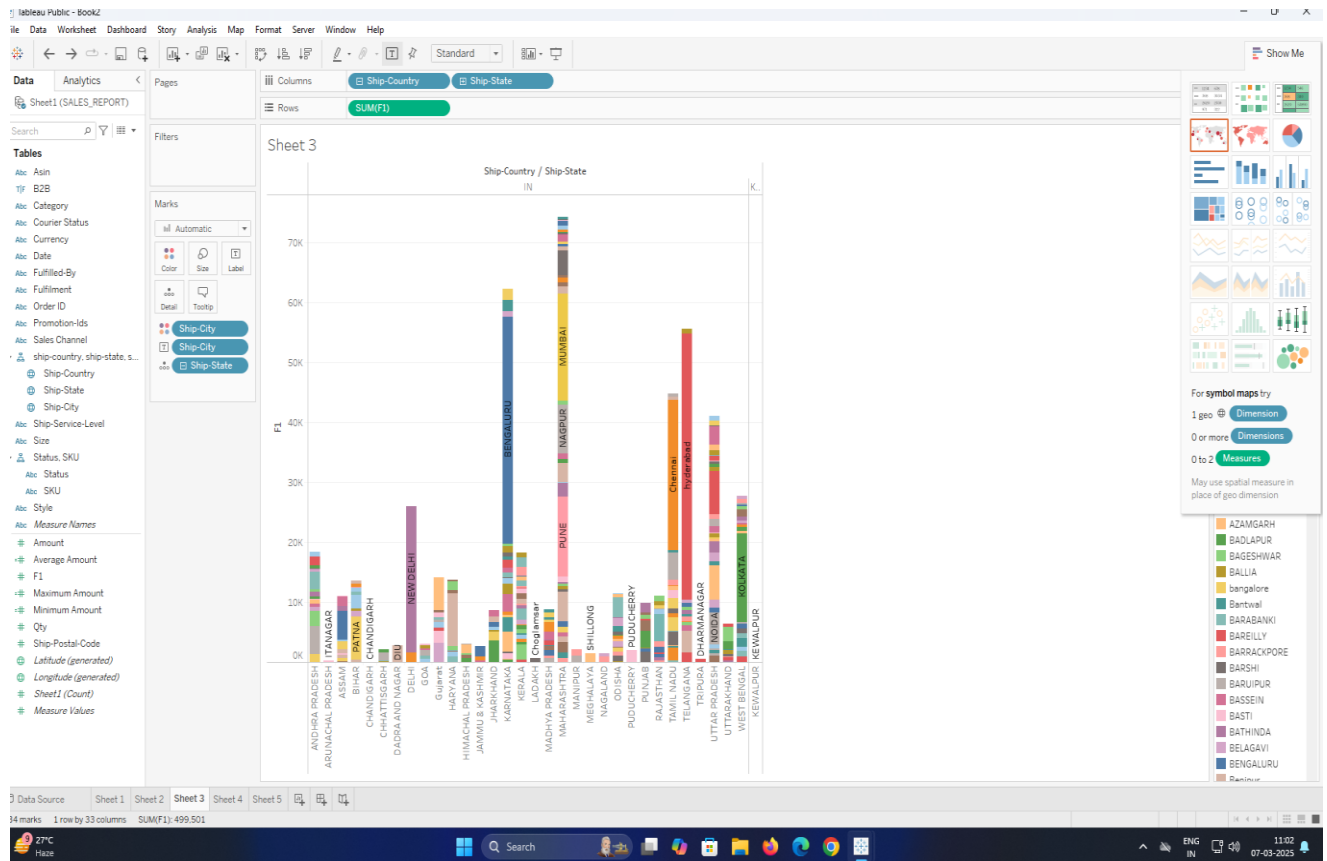
This map visualize titled as "STATE ANALYSIS (MAP)" displays the distribution of shipping activity across Indian states using geographic coordinates. Each blue circle represents a state, and its size corresponds to the total shipment amount, highlighting regional shipping volumes. Southern and western states such as Maharashtra, Tamil Nadu, and Karnataka show larger bubbles, indicating higher transaction values. The map provides an intuitive view of where most shipments and revenue originate, helping to identify high-performing regions. Smaller circles in northern and northeastern regions reflect lower activity. The visualization aids in strategic decision-making based on state-wise sales performance.



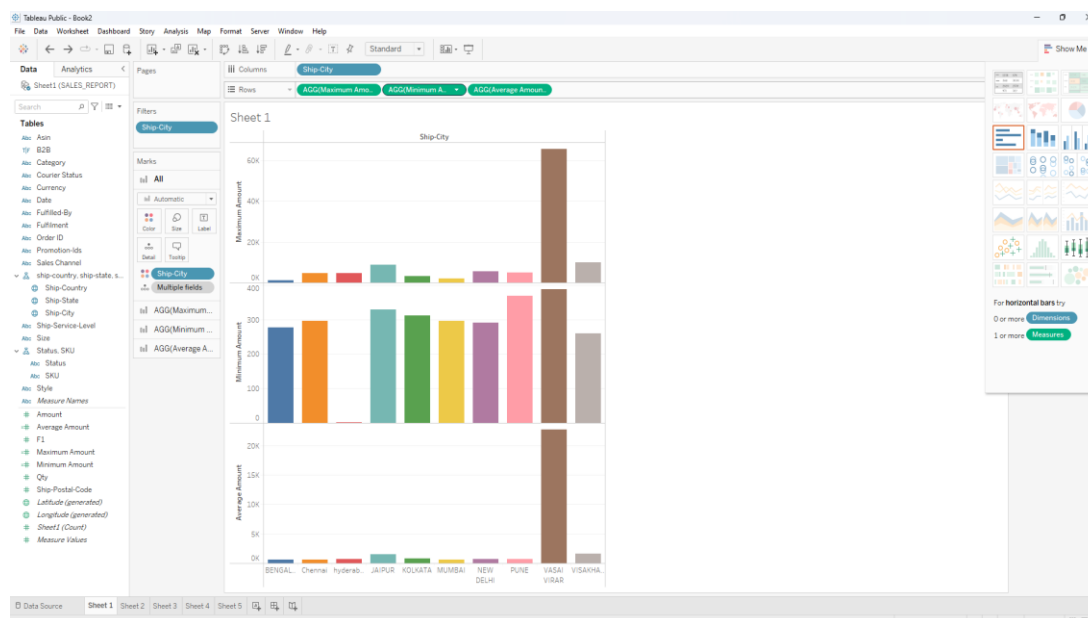
This scatter plot analyzes shipping status by comparing total **Amount** and **Quantity (Qty)** across different categories. Each status—such as "Shipped," "Cancelled," "Delivered," "Returned," and "Rejected"—is plotted with corresponding values. The "Shipped" status shows the highest total amount, indicating it drives most revenue, followed by "Delivered to Buyer." In terms of quantity, "Cancelled" and "Shipped" statuses have the highest volumes. Returned and rejected orders have minimal impact in both amount and quantity. This visualization provides insights into order performance and helps identify which statuses contribute most to sales and which reflect issues in order fulfillment.



This Tableau heatmap presents a detailed breakdown of shipping status across Indian states. States like **Karnataka**, **Tamil Nadu**, and **Maharashtra** show the highest values in the "Shipped" category, indicating strong logistics performance. Meanwhile, **Andhra Pradesh** and **Delhi** display notable cancellation figures, signaling potential fulfillment issues. Metrics such as "Shipped - Delivered to Buyer" and "Shipped - Returned to Seller" offer insight into post-shipping outcomes, with **West Bengal** and **Uttar Pradesh** performing well in completed deliveries. The color gradient helps identify high and low-performing states by status, enabling quick comparison and deeper understanding of regional shipping trends.



This stacked bar chart visualizes shipping quantity (F1) across various Indian states and cities. Major hubs like **Mumbai**, **Bengaluru**, **Hyderabad**, **Chennai**, and **Thiruvananthapuram** show significantly higher shipping activity, indicating their importance in the logistics network. Each bar is color-segmented by city, allowing a clear comparison of intra-state city contributions. States like **Maharashtra**, **Karnataka**, and **Tamil Nadu** stand out due to the combined impact of multiple high-performing cities. Lesser activity is noted in northeastern and smaller states. The chart helps identify top-performing shipping cities and states, supporting strategic decisions in distribution and logistics optimization.



This bar chart compares the **maximum, minimum, and average transaction amounts** across different Indian shipping cities. **Vasai Virar** stands out with the highest maximum and average amounts, indicating potentially larger or more premium transactions. Cities like **Pune, Hyderabad, and Chennai** show moderate values across all three metrics, reflecting consistent transaction behavior. In contrast, cities such as **Mumbai, New Delhi, and Kolkata** maintain balanced minimum and average values with relatively lower maximums. This visualization helps identify outliers, assess spending patterns, and evaluate city-level performance, which can be vital for strategic pricing, marketing, or logistics planning.

RESULTS AND DISCUSSIONS

The analysis of Amazon's sales report reveals that the highest shipping volumes and revenues are concentrated in urban regions such as Karnataka, Maharashtra, and West Bengal, with cities like Bengaluru, Mumbai, and Chennai emerging as major e-commerce hubs. "Shipped" and "Delivered" orders account for the majority of transactions, indicating an efficient fulfillment process, particularly for Amazon-fulfilled orders. In contrast, merchant-fulfilled orders experience higher cancellation rates, pointing to potential gaps in third-party logistics and customer service. Product-wise, Kurtas and Western Dresses lead in sales, reinforcing consumer preferences for specific fashion categories. The average order value typically ranges from ₹400 to ₹750, with cities like Vasai Virar and Pune showing the highest maximum and average

transaction amounts, suggesting pockets of premium purchasing behavior. While promotional codes help increase order volumes, a significant number of transactions still occur without any discounts, indicating strong underlying product demand. Standard shipping remains the preferred delivery method, further supporting the reliability and cost-effectiveness of Amazon's logistics model. Business-to-consumer (B2C) transactions dominate the dataset, confirming Amazon's focus on retail customers over bulk business orders. Overall, the results demonstrate that data-driven fulfillment, regional targeting, and optimized delivery play critical roles in Amazon's success. The findings also highlight areas for improvement, particularly around merchant reliability and cancellation mitigation, offering actionable insights for enhancing operational efficiency and customer satisfaction.

CONCLUSION

In conclusion, the comprehensive analysis of the Amazon order dataset revealed key insights across product performance, regional demand, and operational efficiency. Sales peaks aligned with promotional periods, and categories like "Kurta" and "Top" drove significant revenue. Amazon-fulfilled orders outperformed merchant-fulfilled ones in delivery success, with higher cancellations linked to standard shipping by third parties. Maharashtra and Karnataka emerged as top-ordering states, indicating strong market presence. While most purchases were single-unit and moderately priced, occasional bulk orders influenced revenue trends. Promotional campaigns, particularly financing offers, boosted order volume, though their long-term profitability impact warrants further analysis.

REFERENCE

- [1]. Wells, John R., Galen Danskin, and Gabriel Ellsworth. "Amazon. com, 2018." *Harvard Business School Case Study* 1 (2018): 716-402.
- [2]. Panagiotelis, Anastasios, Michael S. Smith, and Peter J. Danaher. "From Amazon to Apple: Modeling online retail sales, purchase incidence, and visit behavior." *Journal of Business & Economic Statistics* 32.1 (2014): 14-29.
- [3]. Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith. "The longer tail: The changing shape of Amazon's sales distribution curve." *Available at SSRN 1679991* (2010).
- [4]. He, Sherry, and Brett Hollenbeck. "Sales and rank on amazon. com." *Available at SSRN 3728281* (2020).
- [5]. Prasad, Utpal, et al. "Analysis of the co-purchase network of products to predict amazon sales-rank." *Big Data Analytics: 5th International Conference, BDA 2017, Hyderabad, India, December 12-15, 2017, Proceedings* 5. Springer International Publishing, 2017.
- Consumer Products Safety Commission. "In re Amazon. com." (2021).
- [6]. Etumnu, Chinonso E., et al. "Does the distribution of ratings affect online grocery sales? Evidence from Amazon." *Agribusiness* 36.4 (2020): 501-521.
- [7]. Afonso, Whitney B. "The impact of the amazon tax on local sales tax revenue in urban and rural jurisdictions." *Public Budgeting & Finance* 39.2 (2019): 68-90.
- [8]. Rao, Y. V. "USAGE OF ANALYTICS IN E-COMMERCE INDUSTRY TO IMPROVE THE BUSINESS OF THE COMPANIES: A CASE STUDY OF AMAZON." *International Journal of Management Research and Reviews* 14.6 (2024): 1-23.
- [9]. Santana, Francisca Nunes. *How is Web Design Related to Sales? the Relationship Between Sales and Web Design within Amazon Com*. MS thesis. Universidade NOVA de Lisboa (Portugal), 2023.
- [10]. Intelligence, Amazon Artificial General. "The Amazon Nova family of models: Technical report and model card." (2024).