

DATA INTEGRATION AND VISUALIZATION

STUDENT NAME: CALLUM APPS
MODULE CODE: CS3DV20
STUDENT NUMBER: 30010641
ASSIGNMENT REPORT TITLE: COURSEWORK
DATE (WHEN THE WORK COMPLETED): 30TH NOVEMBER 2023
ACTUAL HOURS SPENT FOR THE ASSIGNMENT: 40 HOURS

Contents

Student Name:	1
Introduction	2
EDA & Star Schema	2
amazon database story	3
Analysis of Dashboard.....	5
Recommendation to improvement of sales performance:	5
Conclusion.....	7
Future WORK	7
References.....	8

INTRODUCTION

When delving into data visualization and analysis, the ability to segregate specific data segments for individual examination proves invaluable in uncovering patterns and pinpointing issues within an organisation's sale. In this coursework, the dataset focuses on Amazon's sales, utilising Tableau primarily to detect anomalies in the dataset. This approach serves to provide evidence regarding shifts in performance among certain items and the contributing factors behind these changes. Leveraging the Tableau desktop application, it facilitates an in-depth analysis of trends and the contributions of various services towards Amazon's achievement of specific objectives. Segmenting data offers a host of advantages, particularly in scrutinizing crucial aspects within the dataset. This method facilitates trend identification, anomaly detection, performance evaluation, effectiveness of shipping data, and the identification of seasonal sales trends—all of which are achievable through Tableau. At the heart of effective data analysis lies visualization. It offers vivid representations, converting CSV sales data into diverse visual aids that alternate based on the dataset, effectively illustrating distinct information. Prior to this visualization, the data underwent cleaning via Tableau's data interpreter, ensuring a more lucid and precise analysis. This approach proves pivotal in the realm of e-commerce, providing a significant advantage by spotlighting specific insights for improvements or highlighting trends that drive sales during the current period. Ultimately, it offers a clear pathway for businesses to identify necessary enhancements and capitalize on trends for sustained success. Whilst mostly focusing on quantitative analysis within this report, involving the use of sales figures, performance metrics and other numerical data, Amazon could use qualitative data to further analyse their data.

EDA & STAR SCHEMA

The amazon sales report narrative has been constructed to allow for the investigation and discovery of trends overtime. In the construction of a star schema seen in **figure 1**, it allows for a structured and efficient analysis of the Amazon dataset, this would be due to it containing the fields related to **the dataset, such as date, Category, Colour or OrderID**. Within the dimension tables would lay the descriptive elements relevant to the specific entity. This dimension tables would be the branches stemming from the central data table of the star schema, offering a clear representation of individual aspects within the data that may not be present within other tables, an example of this would be the idea of the Category being offered in the international sales report, facilitating information of how each category performed and most popular categories, allowing for trend identification via product attributes. The use of star schemas would be essential in the analysis of larger datasets because of their simplified structures, employing a central fact table surrounded by dimension tables, allowing for organising and retrieval of data to be efficient and simplistic, enabling quick navigation across various aspects of the dataset. In addition to this, star schemas employ the use of multidimensional query's, allowing for them to retrieve data from a selection of dimensional tables simultaneously, meaning analysts to explore patterns and trends with further effectiveness. Summarily, the use of a star schema to improve decision-making techniques is efficient due to the structured data analysis it provides allowing for insightful interpretations of the dataset, especially for larger datasets like the Amazon Sales Report.



Figure 1 Star Schema

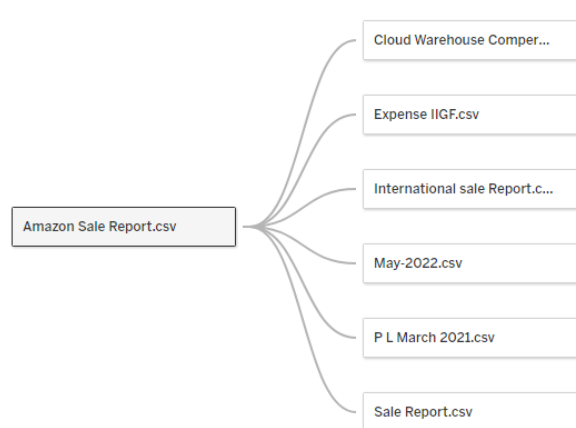


Figure 2 Grouped classes

AMAZON DATABASE STORY

Question:

Examining sales trends within specific product categories; revealing distinctive shifts, emphasising the influential role of certain shipping-services/promotions in shaping sales overtime, as well as methods amazon can use to improve.

Through the use of a storyboard within tableau, it would provide an analytical view of sales data, allowing for the detailed interpretation of various patterns and trends that could make an effect on revenue and sales distribution. The storyboard I have created first focuses on the sales trends between categories, and why some items are not performing as well, which then later shifts to one of the factors that could help these underperformers, such as improving or offering different shipping methods, therefore hoping to increase Amazons sales revenue.

Figure 3 Story page 2 Sales for each category in a month



Within my story board **second page** It explores sales trends within specific product categories and unveils distinctive shifts in Amazon's sales landscape over time. **Figure 3** illustrates the dominance of Kurtas and sets, consistently representing over 30% of Amazon's sales, while other items like Tops and the Western Dress linger at a lower **5-15%** range. These trends emphasise the influential role of specific product categories in shaping Amazon's sales trajectory, pinpointing areas where strategic interventions could enhance overall performance.

Recommendations to enhance Amazon's sales performance are found within analysis of the subject. Firstly, an exploration of the sales trends for underperforming items within product categories such as the **Top** or **Western Dress**. Targeted strategies such as promotions, free expedited shipping or further advertisements can provide them with the needs necessary to boost their sales, better aligning them with the sales of Kurtas or Set's, which may already be receiving these strategies from Amazon, making up a large contribution of their total sales.

The **third page** explores the percentage of revenue that each item would pull in for Amazon through different fulfilments. This further interprets the fact that the sales of Kurtas and sets experienced a much higher spike compared to that of the top over the three months, suggesting that Amazon would be using various methods in order to promote these items allowing them to have a higher share of the market whilst staying at reliable figures across Amazon and Merchant sales.

This analysis would provide further detailing into which product will need the most promotion or services offered in order to increase its market share.

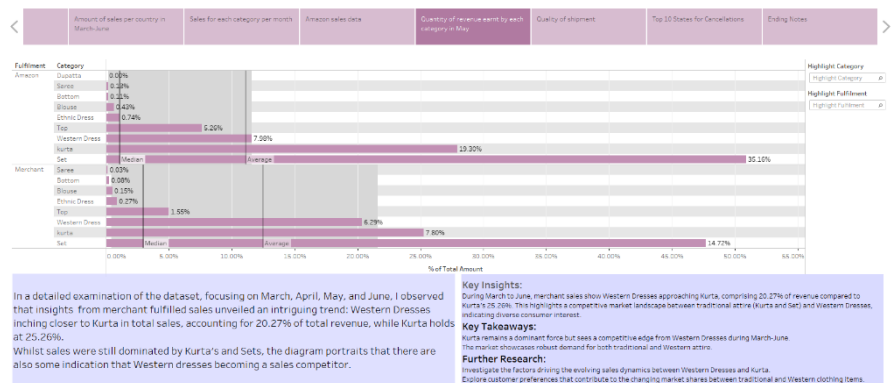


Figure 4 Story page 3 Revenue and merchants

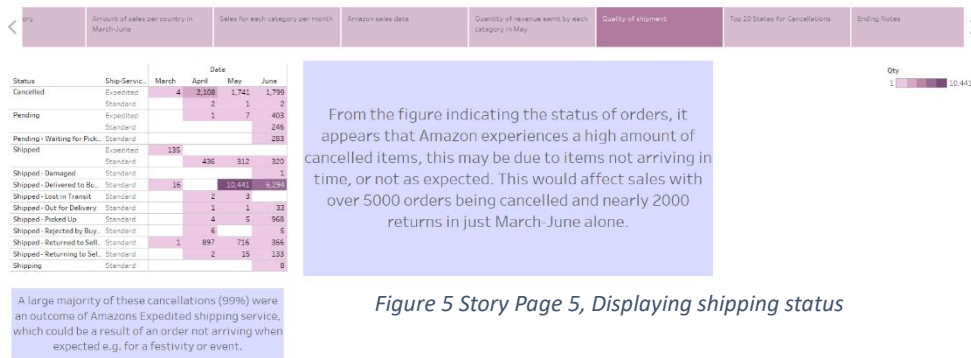


Figure 5 Story Page 5, Displaying shipping status

However, **Figure 5** on **page 5** unmistakably reveals a trend: Amazon is experiencing a significant surge in cancellations, seemingly tied to their expedited shipping with **99%** of cancellations being through this service. This hints at potential challenges linked to the pace of Amazon's shipping or the precision of their delivery date projections. They may be experiencing this as a consequence of offering faster delivery to promote its underperforming items, without having policies setup to overcome this challenge. To delve into this, a comprehensive analysis should pinpoint the specific shipping methods and Amazon brands responsible for fulfilling these items. This will illuminate whether the need for improvement falls on Amazon itself or its network of merchants.

Seen in **figure 6 on page 6**, a higher focus is put onto the countries which are experiencing the highest volume of cancellations, therefore stating which countries Amazon would need to prioritise the fixture of its delivery and customer service department. With **Maharashtra** Experiencing **1007 Cancellations** alone, followed by **Karnataka** at **770**. Overall from the storyboard we can interpret that Amazon's sales can be further bolstered using a range of dynamic methods, targeting specific products, enhancing Amazon's shipping, and more interactive customer service methods.

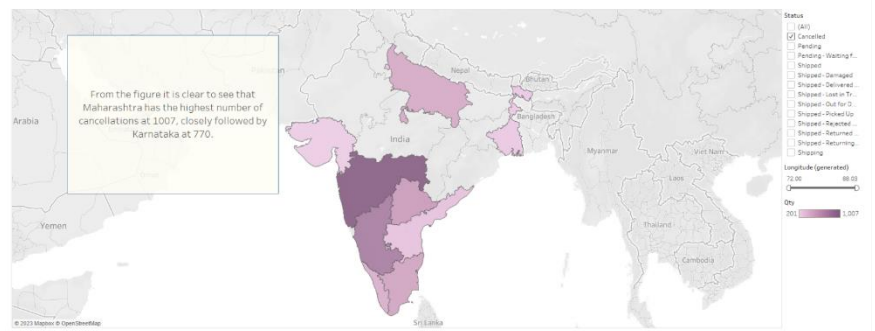


Figure 6 Story Page 6, Map of cancellations

Assumptions Interpreted from story for improvement:

If Amazon were to provide help to their customers before the cancellation is processed, it would allow them to gain an understanding on the issues the customers are having, and resolve them, hence improving the probability of customers returning for further sales.

Implementing more enhanced tracking features would prevent the loss of parcels whilst also allowing for the customer to track when the parcel will arrive, further improving estimations, and allowing for the customer to be in for the delivery, which may also be a factor towards the larger number of cancellations. Additionally, increasing the strength of a parcel would prevent items from arriving damaged, therefore potentially reducing the number of returns. To optimise product success, Amazon could allocate additional resources and time to improve its delivery services. Guaranteeing more reliable deliveries and implementing an enhanced customer service framework could significantly boost sales on their website. This enhancement promises added convenience and ease for customers, potentially driving further sales due to an improved overall shopping experience. Lost, cancelled, or returned items will come at a cost to the company, therefore exemplifying the extent in which this needs to be fixed.

Summary: In prioritising these recommendations, Amazon will be able to increase their sales whilst also incorporating operational strategies along with consumer/customer needs, therefore leading to further sales in the future. Summarily, allowing for the stories end notes to finish with the question: *if faster service was cheaper and more reliable, would there be more sales?*

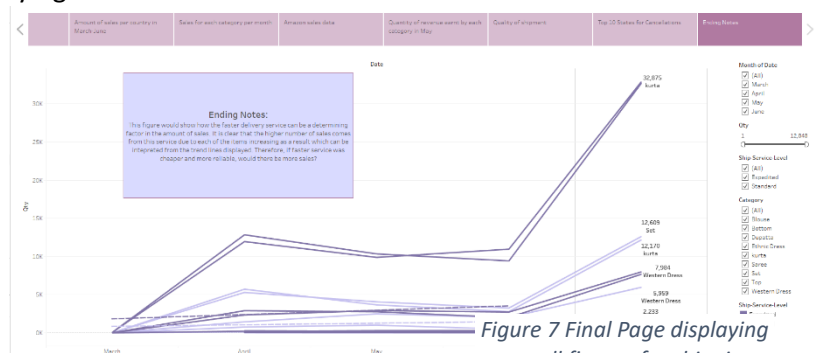
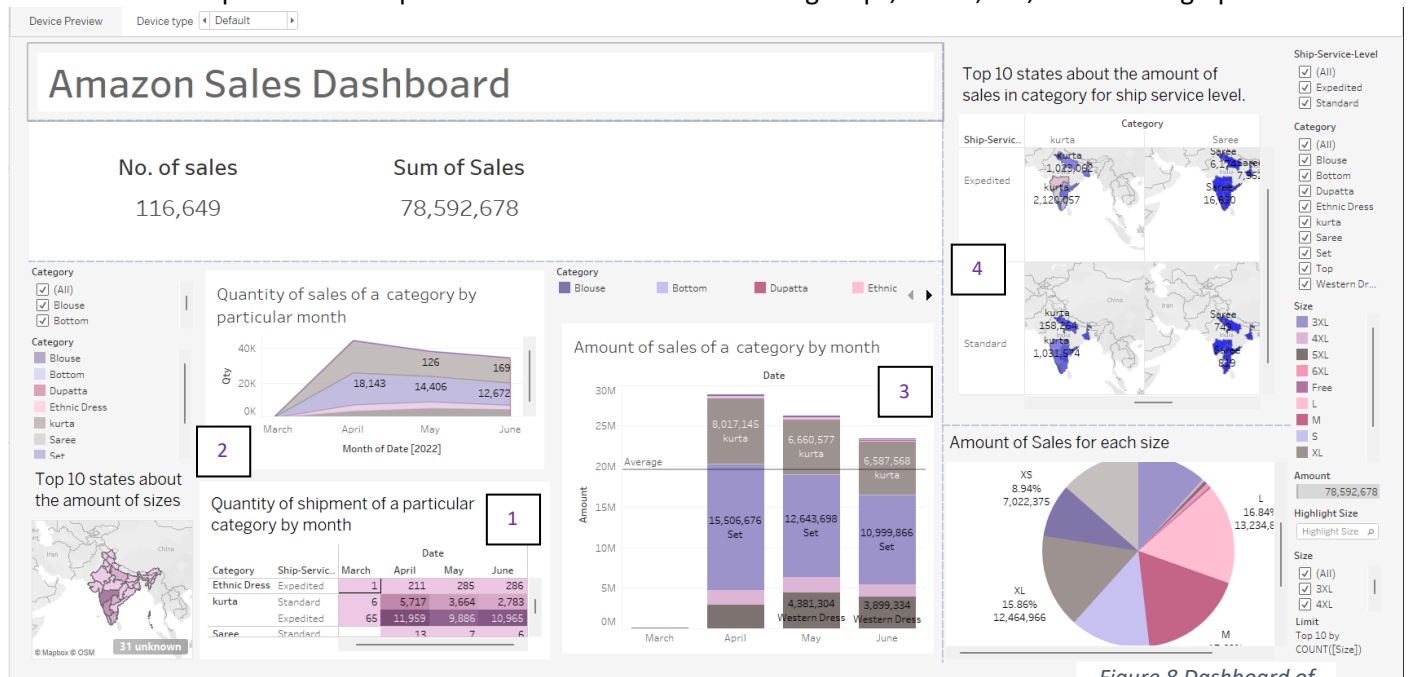


Figure 7 Final Page displaying overall figures for shipping methods Aswell as summary

ANALYSIS OF DASHBOARD

To visualise the dataset, the creation of a dashboard that would be utilised to provide a comprehensive view of how individual product performs and highlights where frequent transactions occur. A dashboard would be a useful tool in the realm of data visualisation due to the extensive amount of information they provide, illustrating various metrics through the use of informative visual aids.

The dashboard displayed in figure 2 features an array of techniques used for visualisation, including bar charts, pie charts, scatter diagrams, and interactive maps used to show data from Order Status to Quality of shipment, whilst also allowing for filtering, providing users with a deep insight into information gathered via the dataset. The dashboard would present the required visualisations for task 3 using maps, tables, bar, and scatter graphs.



(REFERENCED TO NUMBER ON FIGURE 8)

1. QUANTITY OF SHIPMENT OF A PARTICULAR CATEGORY BY MONTH
2. QUANTITY OF SALES BY CATEGORY IN A PARTICULAR MONTH
3. AMOUNT OF SALES IN CATEGORY BY MONTH
4. TOP 10 STATES ABOUT THE AMOUNT OF SALES IN CATEGORY FOR SHIP SERVICE LEVEL

Figure 8 Dashboard of figures showing tasks 1-4 required in question 3 Aswell as extras

RECOMMENDATION TO IMPROVEMENT OF SALES PERFORMANCE:

The dashboard displays a variety of different infographics that determine how its sales are distributed per category per state. This would then allow for an in-depth analysis of Amazons specific transactions providing more of an insight on the areas that may require further promotion or alterations to the sales/shipping methods currently being utilised.

This in-depth analysis will ensure that the patterns being constructed by sales is unravelled, as well as unearthing information about seasonal trends and consumer behaviour when it comes to status of orders. Therefore further enforcing the idea that this dashboard would equip Amazon with the data required to make informed decisions, pinpointing areas that require improvement, in order to increase their sales performance. Furthermore, this dashboard would be a metric that allows for measure overtime, tracking trends, therefore being a great tool in the prediction of future sales/seasonal trends per category.

Further promotion for less frequently bought Items

Key Insight: The dashboard is used to determine which products, within the given categories, are not as commonly purchased across the states. This is evident from the 'Quantity of sales of a category by specific month' report seen in figure 9, which shows that the Kurta and the Set is currently ruling the market due to goods such as the Western Dress, with sales of 3778, cannot be compared to the Set's 18,143 sales.

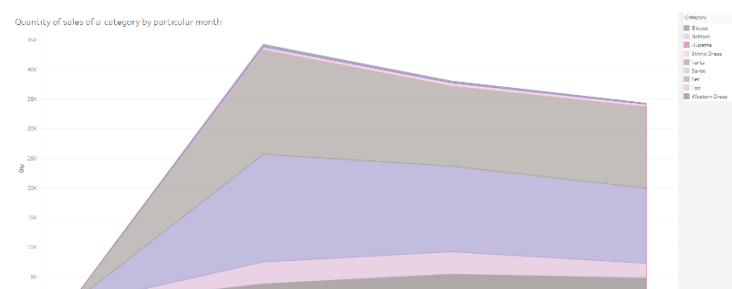


Figure 9 Quantity of sales of a category by particular month

Further Exploration: Whilst sales in Sets seem highly concentrated, this would portray that there underlying factors causing these sales, possibly stemming from increased promotion or seasonal trends. Understanding the specific demographics, regional preferences, promotion strategies, seasonal trends and popular categories would allow for Amazon to verify the factors causing this steep increase in sales and apply it to different categories they have within the Clothing market.

Solution: By using these effective strategies and applying a similar mindset to products in several categories, Amazon might profit from this. Implementing a more intensive marketing effort to such items will be essential for the growth in sales, given the obvious divide in sales distribution among Amazon products. Utilising A/B {5} testing to evaluate each product and identify the top performer would be a useful tactic for Amazon. Their ability to determine which product would profit most from intensive promotions and to maximise promotional efforts for each item would be made possible by the insightful information this study would supply.

Exploiting Sales Concentration in Different Regions

Key Insights: Figure 10 highlights the top 10 states for number of sales within the dataset. This portrays a high number of sales seen within **Maharashtra (11,093)** and **Karnataka (16,630)** for the Saree for example. This would indicate the highest market presence within these areas compared to that of **Andhra Pradesh with a lower presence and sales of 885**.

Further Exploration: Incorporating these sales into other regions would aid in the growth of sales in less saturated markets, even though a larger concentration of sales in this region is beneficial to Amazon's sales. As a counterbalance to the existing robust markets in various locations, Amazon can leverage the data presented in Figure 7 to identify areas with lower sales revenue and focus additional effort there.

Solution: To address this challenge and expand sales in these regions Amazon can execute targeted regional promotion tactics. By dedicating resources to understand the distinctive market dynamics in untapped areas, Amazon can finely tailor promotional strategies to align with the specific preferences of these markets. In doing so Amazon can unlock new opportunities and drive increased sales in these specific areas, further increasing revenue from external areas.

Rasing Shipping Quality and upping Expedited quantities

Key Insights: The graph portrait in figure 13 displays information regarding the quality of Amazons shipping methods, this analysis would provide information such as Amazon still shipping a large portion of their items via standard services, with **5717 Kurtas** being sent using this method just in **April** alone. Using this information would provide us with an analysis stating the ideology that Amazon may not be offering their faster shipping methods in certain states or errors, charging to high of a fee, or due to unreliability, displayed by the number of cancellations in figure 13.

Figure 12 Cancelled orders

Status	Ship-Servic..	March	April	May	June
Cancelled	Expedited	4	2,108	1,741	1,799
	Standard		2	1	2

Therefore, creating a dampen on their sales, affecting sales targets for specific states dramatically.

Further Exploration: Through the implementation of further expedited shipping services Amazon may be able to boost their sales significantly. This would be due to the added customer convenience that comes with the faster shipping methods, with this strategic shift, Amazon would effectively eliminate the need for its customers to search elsewhere for their items when they know that it would be guaranteed to be with them within days of ordering. In order for this to work, Amazon would also need to look into the quality of their shipping, with **2,108 items** being cancelled using their expedited service alone in April, this would go to show that the quality also needs to be assured also.

Solution: Addressing this challenge involves Amazon expanding the availability of expedited shipping services to a wider range of states. By offering this higher-tier shipping option, Amazon can elevate sales through improved customer satisfaction and convenience, however they would have to guarantee this delivery will be delivered efficiency and with care, otherwise customers may sway to other providers, which can be achieved using a revised

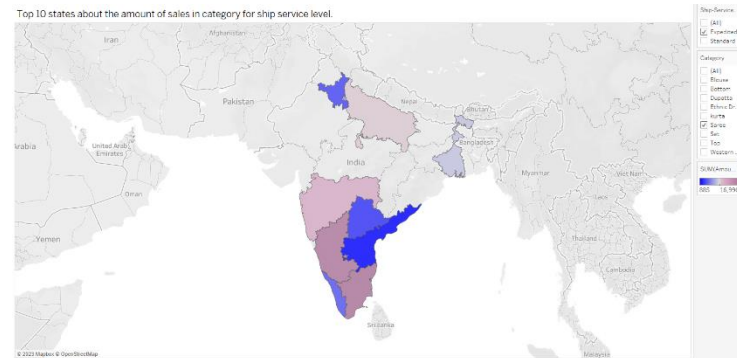


Figure 10 map view of concentration of sales

Quantity of shipment of a particular category by month

Category	Ship-Servic..	Date			
		March	April	May	June
Blouse	Standard	1	124	66	22
	Expedited		276	227	147
Bottom	Standard		36	81	48
	Expedited		131	45	57
Dupatta	Expedited				3
Ethnic Dress	Standard		102	126	42
	Expedited	1	211	285	286
kurta	Standard	6	5,717	3,664	2,783
	Expedited	65	11,959	9,886	10,965
Saree	Standard		13	7	6
	Expedited		55	46	25
Set	Expedited	63		10,344	9,425
	Standard	5	5,295	4,062	3,247
Top	Standard	1	758	1,003	471
	Expedited	8	2,903	2,761	1,998
Western Dress	Standard	4	1,440	2,485	2,030
	Expedited	2	2,338	2,923	2,721

Figure 13 Quantity of shipment by particular month

shipping policy for their employees, money-back guarantees or tracking methods. This expansion would pave the way for Amazon to penetrate new markets, enticing more customers to choose Amazon as their primary supplier.

Ensuring Inventory of specific size ranges

Key insights: Through the analysis of Amazons sales and basing it off of size per sale, it would provide insights into Amazons customers, highlighting specific transaction trends among them regarding sales with the highest demand. Most notable sales were seen from Large, making up for **16.84% of Amazons total sales seen in figure 14**, closely followed by Extra Large being the predecessor most in demand sales. Having knowledge about size distribution and its impact on sales figures, would remain crucial to Amazons inventory management methods.

Further Exploration: Looking further into the relationship between size availability and sales performance would provide insight into sales that would require further inventory management, tailored promotions, or enhanced marketing strategies in order to keep up with demands of the customers. Exploring size availability trends across popular sizes would be convenient to Amazon, allowing for them to have an estimation/prediction on how much stock they will need to keep up with the ever-growing customer demands. Allowing for the fulfilment of each of the order effectively without any arising issues.

Solution: Amazon would need to ensure their inventory is effectively optimised based on the regional demographics presented by Amazons sales, this would be crucial for them in meeting customer demands for popular sizes like Large. Through regular checks and recalibration of stock levels, would ensure the constant replenishment of popular sizes that may be selling out quicker, an example of this would be Large, making up **\$13,234,886** of Amazons revenue for the period. Through specifically tailoring marketing campaigns around specific sizes that need a boost in sales, would benefit Amazons sales whilst a predictive modelling algorithm using the database can be put into place to ensure consistent stock of larger sizes, aligning stock levels with customer needs, whilst also providing an edge over other competitors {4}, which would then allow them to increase sales performance as a result.

CONCLUSION

Throughout this report, a detailed analysis of **Amazons sales report** has taken place utilising Tableau's visualisation tools in the creation of various data visualisation techniques. The created figures would then provide insights of Amazons sales performance, using various data sources, which I could then interpret and describe throughout the report. The segmentation of this data was a crucial element in revealing significant trends and highlighting the dominance for specific item categories over sales, which, through my analysis, I found to be Kurtas and sets. These insights also provided me with information on underperforming categories, allowing me to suggest methods on how Amazon can promote or increase the sales in this sector. Furthermore, the analysis of the shipping data portrait in **figures 12 & 13** revealed interesting statistics, stating Amazons problems with their faster shipping service, emphasising the need for Amazon to look into these issues and resolve them through the use of enhanced delivery methods, better customer service, more accurate delivery predictions or revised policies.

The dashboard was created to provide these insights regarding the sales distribution across Amazons categories and states. The findings from this figure further emphasised the need of regional promotions/tailored promotions to a specific product which are less frequently purchased. Aswell as the requirement of enhancement to their shipping and inventory optimisation to ensure their stock levels can align with customers' needs based on demographics.

Summarily, the analysis also uncovered seasonal trends, consumer behaviour patterns and shipping methods within the dataset, shedding light on the factors driving Amazon's sales success. These insights emphasised the need for adaptive strategies, considering customer preferences, A/B testing {5} for underachieving products, and implementing tailored promotional efforts across different product categories and regions as well as also describing the need in improving Amazon's shipping quality as depicted by the story in question 2. Each of these factors would be essential in the renovation of Amazons services used to increase their sale streams for underperforming items.

Future Application: Within my future work, if I were to use tableau again, I would develop more in-depth visualisations, ensuring that my analysis of the dataset was more thorough, enabling me to enhance my presumptions, as well as possible predictions that can be made off of the dataset. Allowing for myself to come up with more creative and insightful stories.

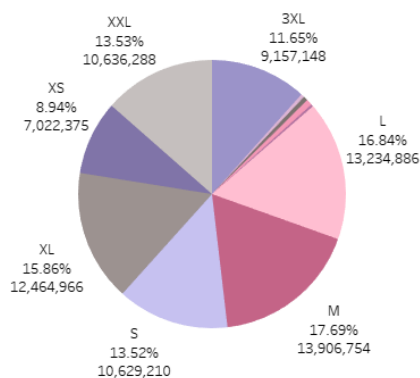


Figure 14 Size distribution pie chart

REFERENCES

1. **Tableau. (n.d.). Create a Story** - Tableau. Retrieved November 15, 2023, from https://help.tableau.com/current/pro/desktop/en-us/story_create.htm
2. **Tableau. (n.d.). Create Dashboards - Tableau**. Retrieved November 15, 2023, from https://help.tableau.com/current/pro/desktop/en-us/dashboards_create.htm
3. **LearnDataModeling. (n.d.). Designing Star Schema**. Retrieved November 17, 2023, from <https://learndatamodeling.com/blog/designing-star-schema/>
4. **Debutify. (n.d.). Does Free Shipping Increase Sales?** Retrieved November 22, 2023, from <https://debutify.com/blog/does-free-shipping-increase-sales/#:~:text=Competitive%20advantage%20over%20competitors,-In%20today%27s%20competitive&text=If%20a%20business%20offers%20free,and%20potentially%20outperform%20their%20competitors.>
5. A Refresher on A/B Testing. *Harvard Business Review*. Retrieved November 22, 2023, from <https://hbr.org/2017/06/a-refresher-on-ab-testing>
6. Making a great dashboard. Retrieved November 22, 2023, from. <https://youtu.be/6oFTdbrugUs?si=1RrlsN4rNvjlxqqJ>