Task 1:

<u>**Problem statements**</u>:

1) Monthly revenue analysis for Q2 2022

The aim of this analysis is to understand the monthly revenue trends during the second quarter of 2022 (Q2) by analysing sales data from 'Amazon Sale Report.csv.' This analysis will guide inventory management and marketing strategies. By examining monthly revenue patterns, the company can identify peak sales periods and adjust inventory levels accordingly to meet customer demand. This analysis will help ensure that inventory levels are optimised, reducing instances of stockouts and overstock situations. Additionally, if a decrease in monthly revenue is identified, targeted marketing campaigns can be developed to stimulate sales and improve overall revenue. The insights gained from this analysis can enhance overall sales and customer satisfaction.

2) Product cancellations analysis

The aim of this analysis is to identify the reason behind the cancellations and returns of the products of Amazon. It is conducted by studying and analysing “Amazon Sale Report.csv” to gain further insights in order to understand the trends that cause orders cancellation.

The problem statement revolves around analysing the reason for cancellations and returns of ordered products in an e-commerce company. Amazon faces the challenge of dealing with cancellations and returns and seeks to understand and diminish their causes. Factors like product quality, inaccurate descriptions, poor customer service, and shipping issues are all possible causes of the issue. Improving product quality by merchants, ensuring accurate descriptions, and enhancing customer service are the key strategies. Analysing cancellation rates of different products and customers is also crucial in addressing the issue. By identifying trends, Amazon can further implement these strategies to reduce cancellations and returns, increase customer satisfaction, and boost revenue. This comprehensive analysis provides actionable insights to help the e-commerce company succeed in a competitive market by retaining satisfied customers effectively.

3) Customer Purchase Behaviour Analysis

The goal of this analysis is to understand customer purchase behaviour by examining historical transaction data. By visualising trends and patterns from the joined dataset between International sale Report.csv and Sale Report.csv, we aim to identify key insights into purchasing habits, including frequency of purchases, popular product categories, colour and size preferences, and their respective stock levels. These insights will help in making informed decisions related to inventory management, marketing strategies, and customer engagement.

Task 2:

Data Description

Dataset 1: Amazon sale data

Amazon Sale Report.csv is referred to as “amazon”.

```
> amazon <- read.csv("R dataset/Assignment 3 dataset/Amazon Sale Report.csv")
> str(amazon)
'data.frame':	128975 obs. of  24 variables:
 $ index             : int  0 1 2 3 4 5 6 7 8 9 ...
 $ Order.ID          : chr  "405-8078784-5731545" "171-9198151-1101146" "404-0687676-7273146" "403-9615377-8133951" ...
 $ Date              : chr  "04-30-22" "04-30-22" "04-30-22" "04-30-22" ...
 $ Status            : chr  "Cancelled" "Shipped - Delivered to Buyer" "Shipped" "Cancelled" ...
 $ Fulfilment        : chr  "Merchant" "Merchant" "Amazon" "Merchant" ...
 $ Sales.Channel     : chr  "Amazon.in" "Amazon.in" "Amazon.in" "Amazon.in" ...
 $ ship.service.level: chr  "Standard" "Standard" "Expedited" "Standard" ...
 $ Style             : chr  "SET389" "JNE3781" "JNE3371" "J0341" ...
 $ SKU               : chr  "AN201-RED-M" "AN201-RED-M" "AN201-RED-XL" "AN201-RED-XL" ...
 $ Category          : chr  "Set" "kurta" "kurta" "Western Dress" ...
 $ Size              : chr  "S" "3XL" "XL" "L" ...
 $ ASIN              : chr  "B09KXVBD7Z" "B09K3WFS32" "B07WV4JV4D" "B099NRCT7B" ...
 $ Courier.Status    : chr  "" "Shipped" "Shipped" "" ...
 $ Qty               : int  0 1 1 0 1 1 1 1 0 1 ...
 $ currency          : chr  "INR" "INR" "INR" "INR" ...
 $ Amount            : num  648 406 329 753 574 ...
 $ ship.city         : chr  "MUMBAI" "BENGALURU" "NAVI MUMBAI" "PUDUCHERRY" ...
 $ ship.state        : chr  "MAHARASHTRA" "KARNATAKA" "MAHARASHTRA" "PUDUCHERRY" ...
 $ ship.postal.code  : int  400081 560085 410210 605008 600073 201102 160036 500032 500008 600041 ...
 $ ship.country      : chr  "IN" "IN" "IN" "IN" ...
 $ promotion.ids     : chr  "" "Amazon PLCC Free-Financing Universal Merchant AAT-WNKTBO3K27EJC,Amazon PLCC Free-Financing U
niversal Merchant A"| __truncated__ "IN Core Free Shipping 2015/04/08 23-48-5-108" "" ...
 $ B2B               : logi  FALSE FALSE TRUE FALSE FALSE FALSE ...
 $ fulfilled.by      : chr  "Easy Ship" "Easy Ship" "" "Easy Ship" ...
 $ Unnamed..22       : logi  NA NA NA NA NA NA ...
```

Figure 1: Structure of Amazon Sale Report

Before cleaning, we represented the dataset’s structure by using str(amazon) as shown in Figure 1. This function provides a compact and comprehensive summary of the dataset, including the types and contents of each variable. It is shown that amazon sale data consists of 128975 rows and 24 columns. Also, it shows that the data consists of Amazon sales made in India based on the currency column (INR). This dataset consisted of mainly int, chr, num and logi data types.

```
> summary(amazon)
     index          Order.ID             Date              Status           Fulfilment        Sales.Channel     
 Min.   :     0   Length:128975      Length:128975      Length:128975      Length:128975      Length:128975     
 1st Qu.: 32244   Class :character   Class :character   Class :character   Class :character   Class :character  
 Median : 64487   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character  
 Mean   : 64487                                                                                                 
 3rd Qu.: 96731                                                                                                 
 Max.   :128974                                                                                                 

 ship.service.level    Style               SKU              Category             Size               ASIN          
 Length:128975      Length:128975      Length:128975      Length:128975      Length:128975      Length:128975     
 Class :character   Class :character   Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character  


 Courier.Status          Qty            currency             Amount          ship.city          ship.state       
 Length:128975      Min.   : 0.0000   Length:128975      Min.   :   0.0   Length:128975      Length:128975     
 Class :character   1st Qu.: 1.0000   Class :character   1st Qu.: 449.0   Class :character   Class :character  
 Mode  :character   Median : 1.0000   Mode  :character   Median : 605.0   Mode  :character   Mode  :character  
                    Mean   : 0.9044                      Mean   : 648.6                                        
                    3rd Qu.: 1.0000                      3rd Qu.: 788.0                                        
                    Max.   :15.0000                      Max.   :5584.0                                        
                                                         NA's   :7795                                          
 ship.postal.code ship.country       promotion.ids         B2B          fulfilled.by       Unnamed..22    
 Min.   :110001   Length:128975      Length:128975      Mode :logical   Length:128975      Mode :logical  
 1st Qu.:382421   Class :character   Class :character   FALSE:128104    Class :character   FALSE:79925    
 Median :500033   Mode  :character   Mode  :character   TRUE :871       Mode  :character   NA's :49050    
 Mean   :463966                                                                                         
 3rd Qu.:600024                                                                                         
 Max.   :989898                                                                                         
 NA's   :33                                                                                             
```

Figure 2: Summary of Amazon Sale Report

To gain a deeper understanding of the dataset, we also used the summary(amazon) function as shown in Figure 2. This function provides a five-number summary for each numerical variable, which includes the minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum values. The “Qty” column has a minimum value of 0, a first quartile value of 1, a median value of 1, a mean of 0.90, a third quartile value of 1, a maximum value of 15. This suggests that the majority of orders consist of one item (median and mean values close to 1), with some orders occasionally including more items, up to a maximum of 15. The presence of a minimum value of 0 indicates that there are some instances where no items were ordered, which might represent cases of cancelled orders or other anomalies in the data. The non-numeric variables provide information on the length, class, and mode of the data. Additionally, the summary indicates if there are any missing values in these columns. We also realise that there is a column “Unnamed..22” which is a categorical variable and consists of either “False” or NA only.

<u>Dataset 2: International sale data:</u>

```
> str(international_sale)
'data.frame':	37432 obs. of  10 variables:
 $ index    : int  0 1 2 3 4 5 6 7 8 9 ...
 $ DATE     : chr  "2006-05-21" "2006-05-21" "2006-05-21" "2006-05-21" ...
 $ Months   : chr  "21-Jun" "21-Jun" "21-Jun" "21-Jun" ...
 $ CUSTOMER : chr  "REVATHY LOGANATHAN" "REVATHY LOGANATHAN" "REVATHY LOGANATHAN" "REVATHY LOGANATHAN" ...
 $ Style    : chr  "MEN5004" "MEN5004" "MEN5004" "MEN5009" ...
 $ SKU      : chr  "MEN5004-KR-L" "MEN5004-KR-XL" "MEN5004-KR-XXL" "MEN5009-KR-L" ...
 $ Size     : chr  "L" "XL" "XXL" "L" ...
 $ PCS      : chr  "1" "1" "1" "1" ...
 $ RATE     : chr  "616.56" "616.56" "616.56" "616.56" ...
 $ GROSS.AMT: chr  "617" "617" "617" "617" ...
> 
```

Figure 3: Structure of International sale data

As shown in Figure 3, International sale data consists of 37432 rows of data and 10 variables. All columns except the index have a data type of chr. It records information such as the date of the products sold, the name of customers and detailed information about the type of products they bought. The first several values shown in the date column are from the year 2006, indicating that there are some data points which might be a little outdated. This suggests that the dataset includes historical data that may not reflect the most current trends or conditions.

```
> summary(international_sale)
     index            DATE              Months            CUSTOMER            Style               SKU
 Min.   :    0   Length:37432       Length:37432       Length:37432       Length:37432       Length:37432
 1st Qu.: 9358   Class :character   Class :character   Class :character   Class :character   Class :character
 Median :18716   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
 Mean   :18716
 3rd Qu.:28073
 Max.   :37431
     Size               PCS                RATE             GROSS.AMT
 Length:37432       Length:37432       Length:37432       Length:37432
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

Figure 4: Summary of international sale data

Summary(international_sale) is used to provide a statistical summary for each column in the data frame as shown in Figure 4. For numeric variables such as “index” column, the summary includes a five-number summary, which are the minimum, first quartile, median, mean, third quartile, and maximum values. Summary(international_sale) also provides the total number of missing values in these columns. For example, the index column has a minimum value of 0, a first quartile value of 2318, a median of 4635, a mean of 4635, a third quartile value of 6952, and a maximum value of 9270, with no missing values reported. For non-numeric variables which are char types, the summary provides information about the length of the column, its class, and its mode. For instance, the “SKU” column has a length of 37432 entries, is classified as a character type, and its mode is also character.

Dataset 3: Sale Report

```
> str(sale)
spc_tbl_ [9,271 × 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ index     : num [1:9271] 0 1 2 3 4 5 6 7 8 9 ...
 $ SKU Code  : chr [1:9271] "AN201-RED-L" "AN201-RED-M" "AN201-RED-S" "AN201-RED-XL" ...
 $ Design No.: chr [1:9271] "AN201" "AN201" "AN201" "AN201" ...
 $ Stock     : num [1:9271] 5 5 3 6 3 11 3 16 8 14 ...
 $ Category  : chr [1:9271] "AN : LEGGINGS" "AN : LEGGINGS" "AN : LEGGINGS" "AN : LEGGINGS" ...
 $ Size      : chr [1:9271] "L" "M" "S" "XL" ...
 $ Color     : chr [1:9271] "Red" "Red" "Red" "Red" ...
```

Figure 5: The structure of Sale Report

Figure 5 shows a dataframe before cleaning with a total of 9271 observations and a total of 7 variables. The dataset is well-organised, with each row providing complete information about a specific product variant. All variables have type chr except index and Stock variables which have both are num type. Index: Serves as an identifier for each row. It is numeric and sequential. SKU Code: A unique identifier for each product variant, combining design, colour, and size. Design No.: Represents the base design of the product. It is a string and may have repeated values across different SKUs. Stock: indicates the number of items available for each SKU. Category variable describes the category to which the product belongs to. Size: specifies the size of the product, typical size labels (S, M, L, XL,...). Lastly, Color variable indicates the colour of the product.

```
> summary(sale)
     index          SKU.Code          Design.No.            Stock        
 Min.   :   0   Length:9271        Length:9271        Min.   :   0.00  
 1st Qu.:2318   Class :character   Class :character   1st Qu.:   3.00  
 Median :4635   Mode  :character   Mode  :character   Median :   8.00  
 Mean   :4635                                         Mean   :  26.25  
 3rd Qu.:6952                                         3rd Qu.:  31.00  
 Max.   :9270                                         Max.   :1234.00  
                                                      NA's   :36       
   Category             Size              Color          
 Length:9271        Length:9271        Length:9271       
 Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character  
```

Figure 6: Summary of Sale Report

The index ranges from 0 to 9270, indicating a total of 9271 rows. Each SKU is unique, indicating no duplicates in product variants. The stock values show significant variation: The minimum stock is 0, indicating out-of-stock items. The median stock level is 8, suggesting that half of the products have a stock level of 8 or less. The mean stock level is relatively high at 26.25, indicating some products have very high stock levels. The maximum stock is 1234, an outlier compared to the 3rd quartile value of 31. Categories can be analysed to understand product and customer segmentation. Size and Color attributes allow for detailed analysis of product variants, which can highlight popular choices.

Data cleaning & manipulation

Amazon Sale Data:

```
> #remove Unnamed..22 column
> filtered = amazon%>%select(-Unnamed..22)
> dim(filtered)
[1] 128975     23
> #change date to Date format
> filtered$Date <- parse_date_time(filtered$Date, orders = c("mdy", "m-d-y"))
> filtered$Date <- as.Date(filtered$Date)
> ## remove null values
> colSums(is.na(filtered))
           index         Order.ID             Date            Status
               0                0                0                 0
       Fulfilment    Sales.Channel ship.service.level            Style
               0                0                0                 0
             SKU         Category             Size              ASIN
               0                0                0                 0
   Courier.Status              Qty         currency            Amount
               0                0                0              7795
       ship.city       ship.state ship.postal.code      ship.country
               0                0               33                 0
    promotion.ids              B2B     fulfilled.by
               0                0                0
> amazon = na.omit(filtered)
> dim(amazon)
[1] 121149     23
> write.csv(amazon, file = "C:/Users/User/Desktop/cleaned_amazon_data.csv", row.names = F
ALSE)
> #check if there is any duplicated rows
> any_duplicated <- any(duplicated(amazon))
> print(any_duplicated)
[1] FALSE
```

Figure 7: Dimension of Amazon Sale data before and after

As mentioned earlier in data description, the column “Unnamed..22” is undefined and consists of NA and FALSE only. Therefore, “Unnamed..22” column is first to be removed. By parsing the date-time objects with parse_date_time() and then converting them to pure dates with as.Date(), ensure that the Date column contains only dates. By applying colSums(is.na(amazon)), we can see that the dataset contains a lot of NA values, thus the dataset is cleaned using the na.omit() function to remove NA values from our dataset entirely. By using the dim() function, we observed that only 121149 rows of data remained. Since the output of any(duplicated(amazon)) is FALSE, it implies that there are no duplicated rows in the dataset. Therefore, no action is required to remove duplicated rows.

Data Joining

```
# change the column name of SKU Code in Sale_Report to SKU
# change the column name of Design No. in Sale_Report to Style
names(sale)[names(sale) == "SKU.Code"] <- "SKU"
names(sale)[names(sale) == "Design.No."] <- "Style"
```

Figure 8: Changing of column names in Sale dataset

To analyse customer purchase behaviour, our group combined both International sale dataset with Sale dataset by using the inner_join function by ‘SKU’, ‘Size' and ‘Style’). SKU, Style and Size variables ensure that each product variant is uniquely identified, also ensuring the accuracy of matching products across datasets.

Since the variables of SKU and Style are represented as ‘SKU.Code’ and ‘Design.No.’ in the Sale dataset, our group will change the variable names correctly to match the ones in the International sale dataset shown in Figure 8 above.

```
> joined_report <- inner_join(international_sale, sale, by = c("SKU", "Size", "Style"), relationship = "many-to-many")
> head(joined_report)
  index.x     DATE Months           CUSTOMER   Style           SKU Size  PCS   RATE GROSS.AMT index.y Stock Category
1       0 06-05-21 Jun-21 REVATHY LOGANATHAN MEN5004   MEN5004-KR-L    L 1.00 616.56    617.00    6772     5    KURTA
2       1 06-05-21 Jun-21 REVATHY LOGANATHAN MEN5004  MEN5004-KR-XL   XL 1.00 616.56    617.00    6775    11    KURTA
3       2 06-05-21 Jun-21 REVATHY LOGANATHAN MEN5004 MEN5004-KR-XXL  XXL 1.00 616.56    617.00    6776     3    KURTA
4       3 06-05-21 Jun-21 REVATHY LOGANATHAN MEN5009   MEN5009-KR-L    L 1.00 616.56    617.00    6802    12    KURTA
5       4 06-05-21 Jun-21 REVATHY LOGANATHAN MEN5011   MEN5011-KR-L    L 1.00 616.56    617.00    6814     7    KURTA
6       5 06-05-21 Jun-21 REVATHY LOGANATHAN MEN5025   MEN5025-KR-L    L 1.00 649.03    649.00    6898     4    KURTA
   Color
1   Blue
2   Blue
3   Blue
4 Maroon
5   Blue
6   Blue
>
```

Figure 9: Function for inner_join and the head of of joined dataset

Next is to use the inner join function together with the head of the joined dataset, shown in Figure 9.

Result: A combined dataset that includes only the rows where the 'SKU', 'Style', and 'Size' match in both datasets. The reason for using inner_join is to ensure that only records with matching 'SKU', 'Style', and 'Size' in both datasets are included. This avoids including incomplete or unmatched records, maintaining data integrity. The resulting dataset will include only relevant entries where there is a corresponding sale record and customer purchase record. After performing the inner join, the resulting dataset will have columns from both original datasets, such as: index, SKU Code, Design No., Stock, Category, Size, Color (from Sales Dataset) Customer, Date of Purchase, Pieces Bought, Rate, Gross Amount (from International Dataset).

```
> # check for NA values in joined dataset
> colSums(is.na(joined_report))
  index.x      DATE    Months  CUSTOMER     Style       SKU      Size       PCS      RATE GROSS.AMT   index.y     Stock  Category     Color
        0         0         0         0         0         0         0         0         0         0         0     37440         0         0
```

Figure 10: Showing the column which contains the number of NA values.

Next, our group will perform data cleaning for the joined dataset. Using is.na() to check for NA values. Figure 10 shows that in the Stock variable, there are 37440 rows with NA values.

```
> summary(clean_joined_report)
    index.x             DATE                Months            CUSTOMER             Style
 Min.   : 1524   Min.   :2021-07-16   Length:14025       Length:14025       Length:14025
 1st Qu.: 6384   1st Qu.:2021-10-21   Class :character   Class :character   Class :character
 Median :10009   Median :2021-12-21   Mode  :character   Mode  :character   Mode  :character
 Mean   :10000   Mean   :2021-12-20
 3rd Qu.:13596   3rd Qu.:2022-02-22
 Max.   :17234   Max.   :2022-03-31
                 NA's   :5415
     SKU                Size                PCS                RATE              GROSS.AMT
 Length:14025       Length:14025       Length:14025       Length:14025       Length:14025
 Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character



     Stock           Category             Color
 Min.   :   0.0   Length:14025       Length:14025
 1st Qu.:   3.0   Class :character   Class :character
 Median :  11.0   Mode  :character   Mode  :character
 Mean   :  43.7
 3rd Qu.:  46.0
 Max.   :1234.0
```

Figure 11: Summary of cleaned and joined dataset.

Now we will use na.omit() to remove all NA values from the joined dataset. In the joined dataset, there is already a present on index.x that uniquely identifies each row, making column index.y unnecessary. Dropping the index.y column will make the analysing of problem statement much more efficient. The summary and descriptive statistics of the variables in the cleaned and joined dataset are shown in Figure 11. Given all the summary, we can further our group analysis on problem statement 3.

Task 3

Data Visualization

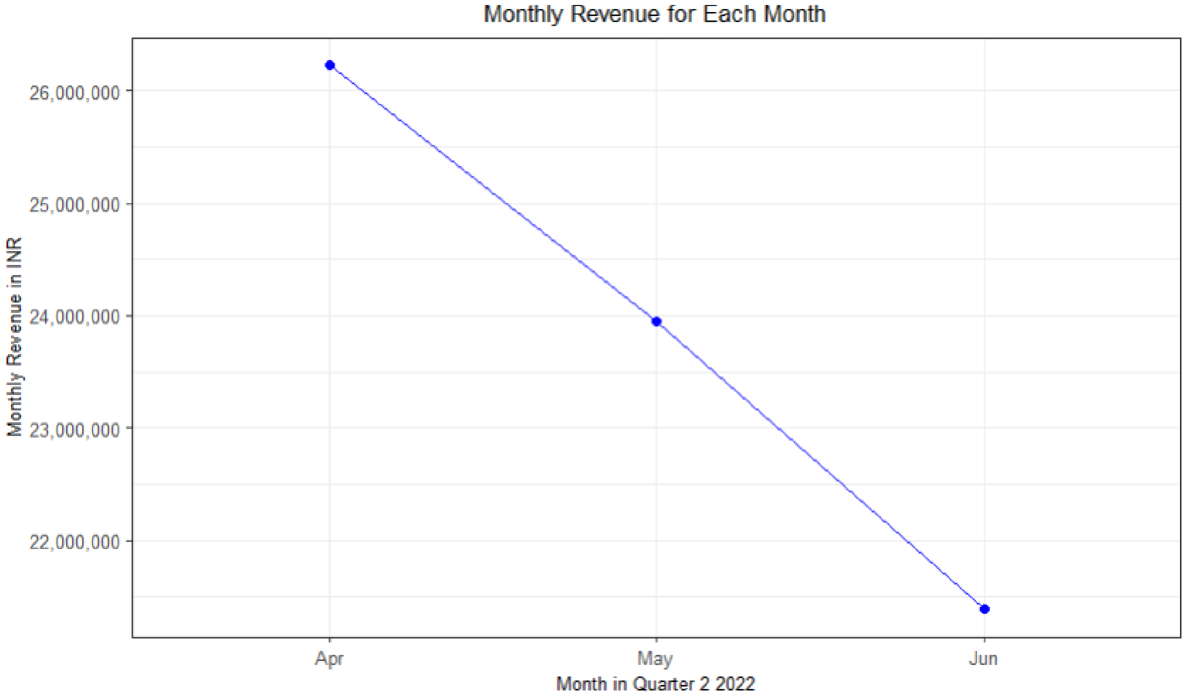

Figure 12: Line Chart for Monthly Revenue from April to June 2022

The line chart displays the monthly revenue in INR for each month in the second quarter of 2022, specifically from April to June. It shows a clear downward trend in revenue over this period. In April, the revenue was the highest, exceeding INR 26 million. However, there is a noticeable decline in revenue in the subsequent months. By May, the revenue decreased to approximately INR 24 million. The downward trend continues into June, where the revenue drops further to just above INR 21 million.

As a decrease in monthly revenue is identified, the company can develop targeted marketing campaigns to stimulate sales and improve overall revenue. For example, the significant revenue drop from April to June suggests that the company may need to implement promotional strategies or discounts to attract more customers during slower periods.

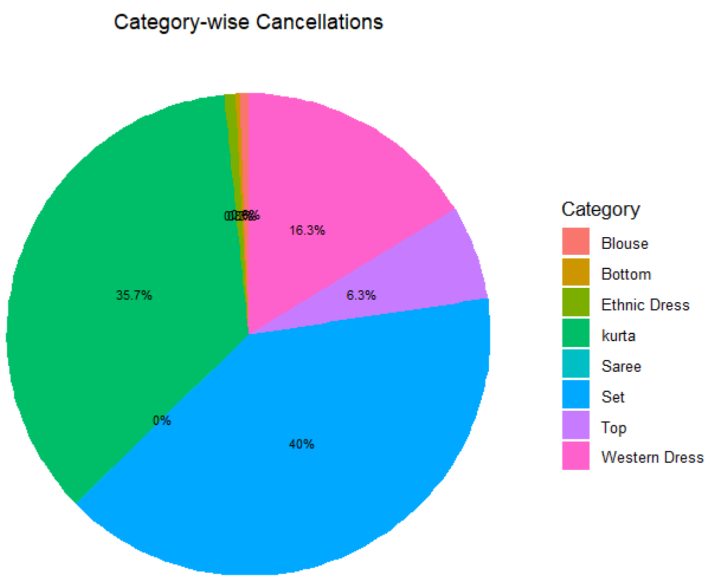

Figure 13: Pie Chart for Amazon Cancelled Sales Distribution by Sales Category

The pie chart illustrates the distribution of order cancellations across various product categories. Set accounts for the largest proportion of cancellations at 40%, closely followed by Kurta with 35.7%. Western Dress experiences a significantly lower cancellation rate at 16.3%. The categories of Top, Blouse, Bottom, and Set collectively represent a minimal share of cancellations, with Top at 6.3% and the remaining categories contributing very small percentages each. This data suggests that Set and Kurta are the most frequently cancelled items, while cancellations for other clothing categories are comparatively rare.

Using the given data, Amazon could identify several factors contributing to the high number of cancellations. Firstly, the wide variety of sizes and colours available for these clothing categories may lead to customers experiencing size and fit issues, resulting in higher return and cancellation rates. Additionally, misleading or inaccurate product descriptions and images can cause customers to cancel orders when they realise the actual product does not match their expectations. Moreover, cancellations may also stem from quality issues, with products failing to meet customer expectations.

To address these issues, Amazon should implement strict quality control measures to maintain high product standards, particularly in categories with high cancellation rates like Western Dress and Kurta. Besides, providing detailed and accurate size guides for all clothing items would help customers choose the right size, reducing cancellations due to fit issues. Additionally, offering exchange policies could decrease cancellation rates, as customers might opt for an exchange rather than cancelling or returning the order.

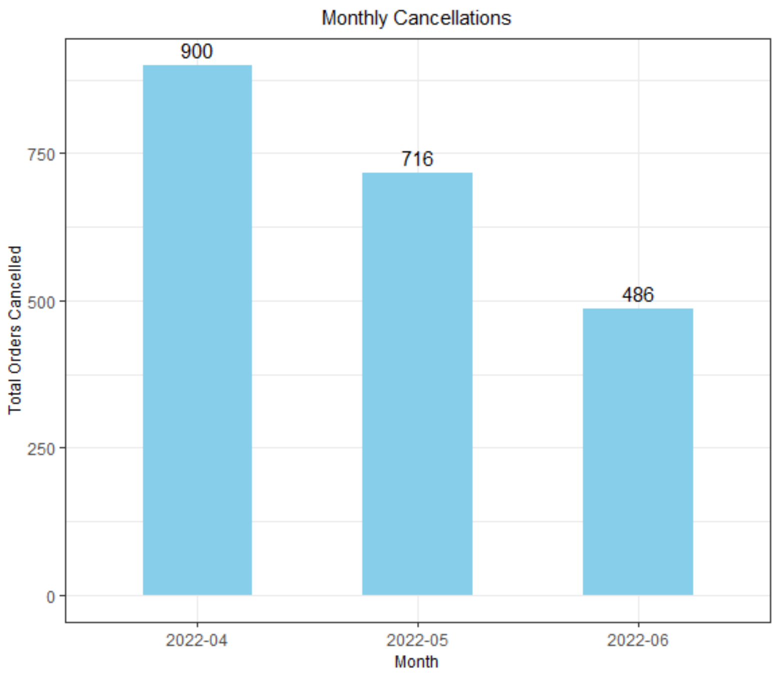

Figure 14: Bar Chart for Monthly Cancellations from April to June 2022

The bar graph shows the total number of order cancellations per month from April to June 2022, with the x-axis indicating the months, and the y-axis representing the total number of cancelled orders.

The highest number of cancellations occurred in April (900) followed by a significant number in May (716). This can be attributed to the holiday of Eid al-Fitr, which is a major festival for Muslims, Indian Muslims in this case. During this period, there is typically a spike in shopping for festive clothing, such as kurta sets and kurtas. After the Eid al-Fitr holiday, the cancellations dropped significantly to 486 in June. This decrease can be associated with the end of the festive shopping season. Customers are less likely to be purchasing specific festive attire, and the urgency for timely delivery is reduced. Overall, there is a clear trend of decreasing cancellations over these three months.

The increase in orders for kurta sets and kurtas ahead of Eid al-Fitr might have led to a higher rate of cancellations. Possible reasons could include issues with size, fit, quality, or delayed deliveries, as customers are particularly sensitive to these factors when shopping for specific festive occasions. Therefore, merchants should anticipate the increase in demand for specific products during festive seasons and prepare accordingly. This can include increasing inventory, staffing, and customer support to handle the surge effectively.

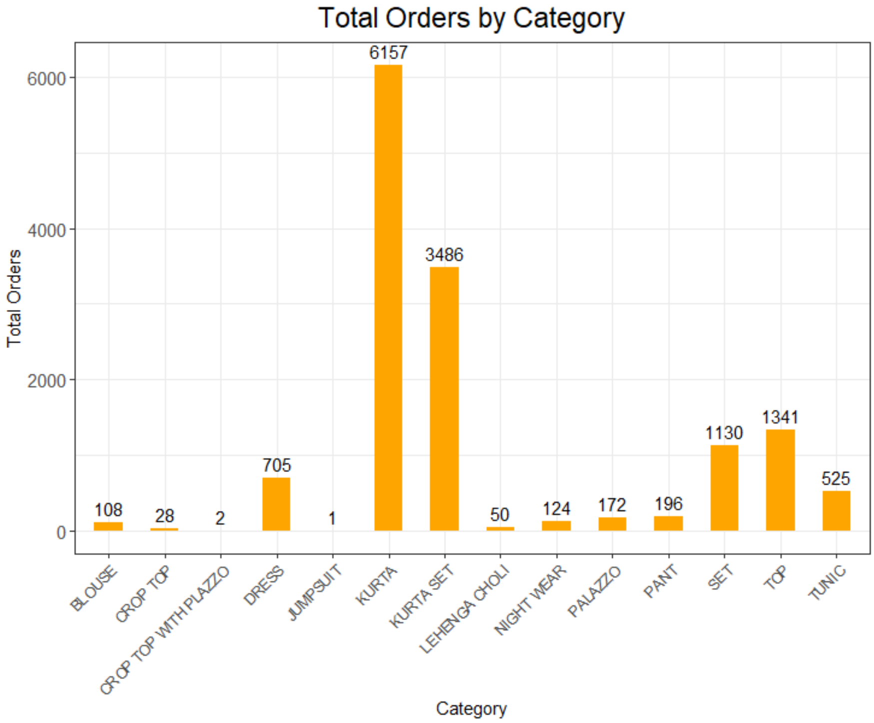

Figure 15: Bar Graph for Total Orders (PCS) of Different Clothes Category

The bar graph presents the total number of pieces of clothing sold by product category. The Kurta category stands out with the highest number of orders, totalling 6,157, followed by the Kurta Set category with 3,486 orders. The Top category also shows significant demand with 1,341 orders, while the Set category has 1,130 orders. These categories dominate the total orders, indicating their high popularity among customers, likely due to cultural preferences and seasonal demands, such as festivals.

In the moderate volume segment, Dress and Tunic categories have 705 and 525 orders, respectively, suggesting they are fairly popular but not as prominent as

Kurtas and Kurta Sets. Niche categories like Nightwear, Palazzo, and Pant show consistent demand with 124, 172, and 196 orders respectively. Conversely, Ankle Leggings, Blouse, Crop Top, and Lehenga Choli exhibit lower order volumes, which may indicate lesser customer interest or a need for better marketing and product improvement. Specifically, Crop Top with Palazzo, Jumpsuit, Suit, and Earring categories show negligible orders, suggesting either very low demand or issues like poor visibility, limited variety, or high competition.

This predictive analysis enables Amazon to maintain inventory levels more effectively, aligning supply with anticipated demand. For instance, during festive seasons, Amazon can strategically increase stocks of popular items like sets and kurtas, ensuring they can meet the expected surge in customer demand effectively.

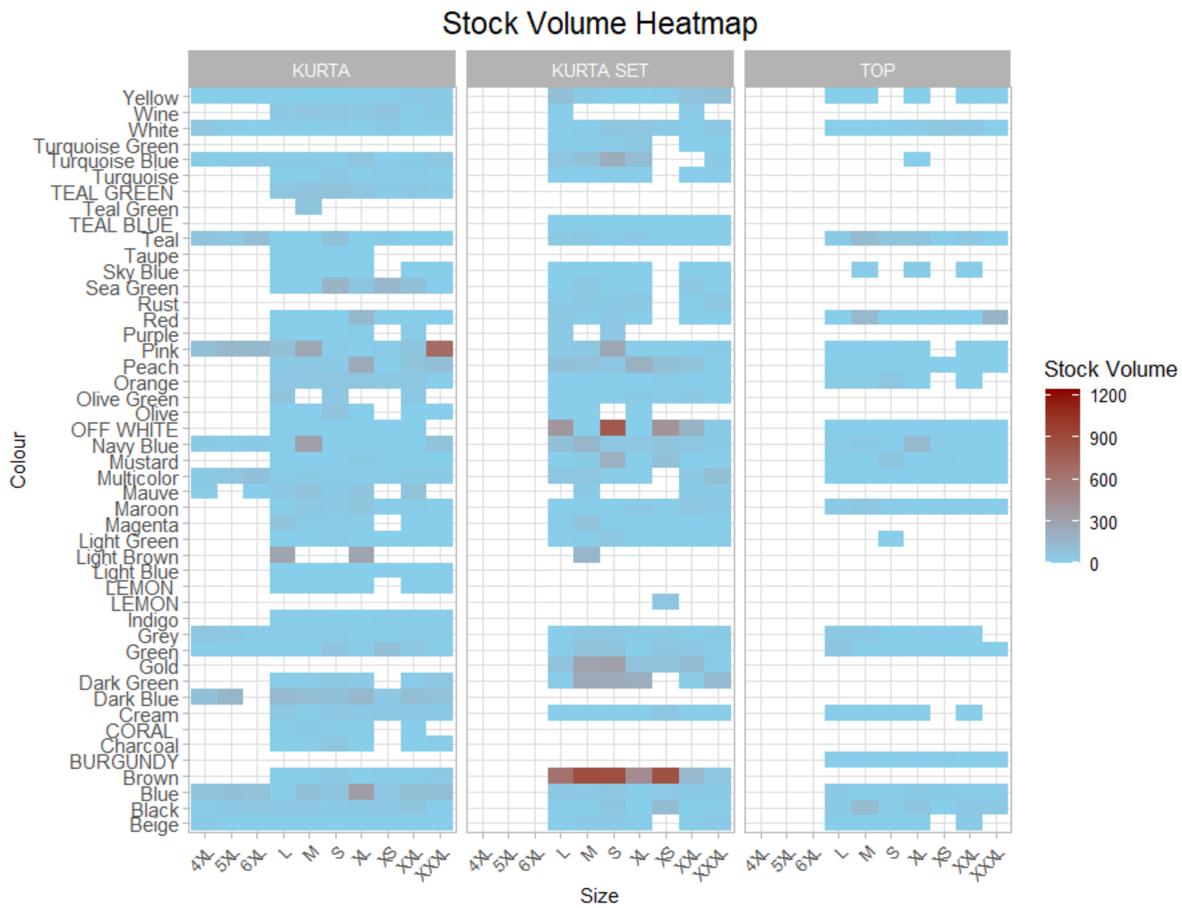

Figure 16: Heatmap showing the relationship between size, colour and stock volume

The heatmap in Figure 21 illustrates stock volumes for various colours and sizes across three product categories: Kurta, Kurta Set, and Top. Each cell in the heatmap represents a specific colour and size combination, with rows and columns labelled accordingly. The colour intensity of each cell indicates stock levels: darker shades of red signify higher stock volumes, while lighter shades of blue represent lower stock volumes.

According to the heatmap, sizes XS to L in brown colour exhibit the highest stock density, followed closely by size XXXL in pink and size S in off-white. These combinations have substantial inventory, suggesting they are in high demand and require consistent restocking. Conversely, sizes such as 4XL, 5XL, 6XL, XXXL, and XXL show limited stock volumes, indicating lower demand for these extreme sizes.

To enhance inventory management, the company should balance stock levels by identifying overstocked and understocked items. Overstocked items, indicated by darker red areas, such as the Kurta in pink (size XXXL), can be targeted for promotional strategies or discounts to clear excess stock. Understocked items, shown in light blue areas, like Kurta Sets in sizes 4XL, 5XL, and 6XL, suggest a need to increase inventory to meet potential demand and avoid stockouts.

By analysing these trends, the company can better forecast future demand, adjust inventory levels based on seasonal trends, and streamline product offerings. For instance, avoiding restocking of items in less demanded sizes might be prudent, as liquidating excess stock in these sizes can be challenging and repurchasing could lead to losses due to slower inventory turnover.