Analyzing stock sales of Retail Clothes seller

BDM CAPSTONE PROJECT- MID TERM SUBMISSION

Prepared by: -

Name- Anmol Sarraf

Roll number: - 22f3002374

Table of contents: -

1. Executive summary: -

This project focuses on "New Dolly Dress Lok" shop located at Bara bazaar, Madhubani which deals with B2C primary and sells traditional Indian Readymade clothes like *Kurta, Salwar suit, Jeans, Pant, shirt etc.* This Business is facing problem regarding overstocking and understocking as there is no record-maintained stock purchased and sold. It also wants to minimize it Revenue loss and move toward online business.

Data was collected from sales made in a day. Which item was sold, how much and price was collected. Along with customer feedback was also collected who didn't buy anything after visiting and stock count was taken on 31st March 2025. First it was logged in excel sheet later it was imported into python using panda's library. In total 226 successful sales transactions were made, and 58 customers' feedback was taken over a period of 3 months.

Descriptive statistics were applied on datasets to get summary of dataset. Kurta was the highest selling item with moderate variability and Left skewness, but Salwaar suit generated maximum revenue also with moderate variability and Right skewness. Most customers didn't buy anything because shop does not have their desired colour and size, indicating Stock issues.

Applying weekly-trend analysis on data revealed that stock issues in high-demand products created a domino effect which then led to sales decline and revenue loss for the business. Most customers reported stock issues especially in June. Visualizations such as Bar charts, Pie charts and pivot summaries helped to highlight patterns and to find out stock issues.

2. Proof of originality: -

Me in Interacting with the Owner and shop view from front

Some more Proofs: - click here

3. Metadata: -

This workbook had 3 worksheets:

- **Sales data:** - This sheet contains all sales made on a day from 1 April to 31 May 2025.
- **Customer data:** This sheet contains a review of customers' opinions about why they didn't buy anything.
- **Stocks data:** - This sheet contains the count of stocks in Shop. This counting was done on 31 March 2025.

3.1. Sales data: - This sheet contains 227 rows and 6 columns. Data was collected manually from all sales in a day.

The description of the following metadata is explained as follows: -

- **Date** *(String)*: - It is the date on which the purchase was made.
- **Item** *(String)*: - Name of item bought
- **Quantity (***Integer***)**: - Total Number of counts
- **Price per item (***Float***)**: - Cost of items in rupees.
- **Total Price (***Float***)**: - This is Product of Quantity and Price per item. This is the final price paid.
- **Week (***String***)**: - it is used to categorize the dates in the week. It contains a week number.

3.2. Customer data: - This Sheet contains 59 rows and 5 columns. This data was collected manually by asking for the customer's feedback.

The description of the following metadata is explained as follows: -

- **Date (***String***)**: - Date on which feedback is given.
- **Looking For (***String***)**: - Which item customer is looking for.
- **Reason for not buying (***String***)**: - Why did customers buy anything.
- **Feedback (***String***)**: - Final opinion given by customer.

3.3. Stocks data: - This sheet contains 10 rows and 2 columns. This was collected by manually counting all the stocks. It was collected on 31st March 2025.

The description of the following metadata is explained as follows: -

- **Items (***String***)**: - Name of the item
- **Count (***Integer***)**: - Count of the item

Link to Data:- click here

4. Descriptive Statistics

Item	Total Count	Qty max	Qty min	Price Mean	Price Median	Price standard deviation	Price range
Coat-pant	9	1	1	₹ 4,333.33	₹ 4,000.00	₹ 1,118.03	(3000, 6000)
Frock	10	2	1	₹ 716.67	₹ 500.00	₹ 541.91	(400, 1800)
Jeans	61	5	1	₹ 1,076.39	₹ 1,100.00	₹ 277.36	(400, 2000)
Pant	14	2	1	₹ 740.00	₹ 700.00	₹ 291.36	(400, 1200)
Salwaar suit	73	6	1	₹ 1,480.23	₹ 1,500.00	₹ 427.50	(700, 3000)
Shirt	50	5	1	₹ 796.77	₹ 800.00	₹ 195.76	(500, 1500)
T-shirt	42	4	1	₹ 537.04	₹ 500.00	₹ 266.24	(300, 1500)
kurta	130	17	1	₹ 882.41	₹ 725.00	₹ 426.94	(300, 2500)
undergarments	23	4	2	₹ 140.00	₹ 100.00	₹ 51.64	(100, 200)
Total	412						

Table-3.1. Descriptive Stats

The Above table displays a Descriptive summary of sales dataset. The table has been calculated using python Panda's library in juypter notebook. This table helps us to understand the distribution of sales and identify trends. From the table we can draw some insight.

- In most items **Price mean** and **Price median** are close, which indicates stable prices. However, in case of Coat-Pant mean (₹ 4,333.33) is higher than median (₹ 4,000.00) with standard deviation of ₹1118.03 indicates that item was sold at higher than median suggesting higher-priced transactions. Salwaar suit and jeans were sold at less than median prices indicating low-priced transactions.
- The **Price standard deviation** column reflects the price **variability** of each item. Coat-pants and Frock show high variability due to high standard deviation, unique design and materials used whereas Undergarments, Shirt and Pants show low variability indicating more stable price. The rest of the items show moderate variability.
- On comparing mean and median of prices, most of items show **Right skewness** as mean price is higher than median prices like Coat-pant, Frock whereas Salwaar suit show **Left skewness** as mean price is lower than median prices. Shirt shows minimal skewness as it remains at nearly equal mean and median.
- The **mode** of Qty Count column is Kurta, which varies around ₹2500 to ₹300 with average of ₹885.85 is also most sold during that period. Seventeen Kurtas was a **max** sale in a day. A total number of 224 successful transactions were made.

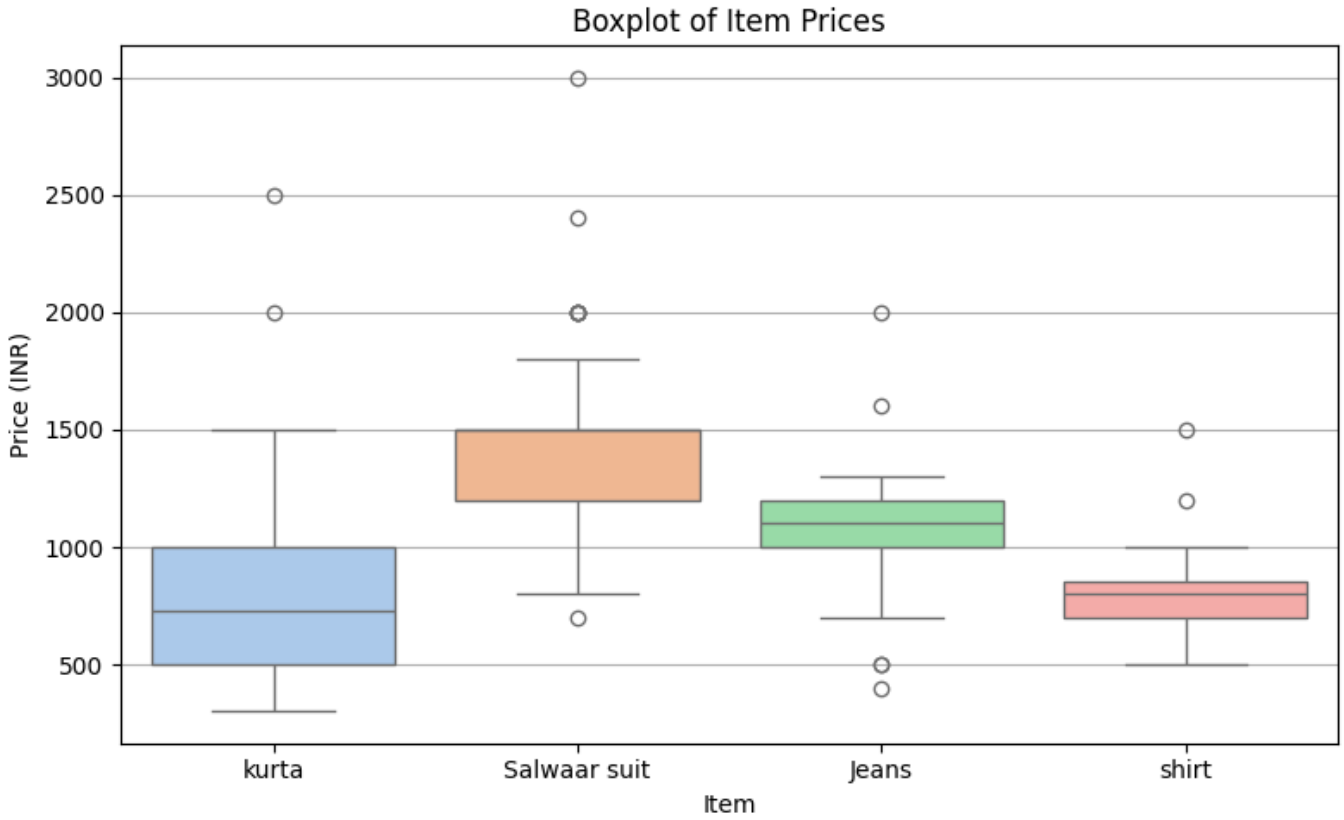

Fig 3.1: Boxplot comparing the price distribution of four key clothing items. Kurta and Jeans show moderate price consistency, while Salwaar Suit exhibits more pricing variability and frequent outliers. Shirt displays the most stable pricing.

The boxplot provides a comparative view of price distribution across four key items which were sold more. **Salwaar suit** shows wider spread means **high variability** in prices with some high and low outliers due to clothes used, different design patterns. Kurta and jeans have tight interquartile ranges indicating stable pricing with some outliers. Shirt displays most consistent pricing with no major outliers

5. Detailed Explanation of Analysis Process/Method

5.1. Objective: -

- The main Objective of this analysis was to understand sales patterns, customer purchases patterns, Item demand and missed sales opportunities over period of 13 weeks.
- While discussing with owner, I got know about people buying patterns and operational challenges and strong competition from online businesses. So, aim was to identify issues which cause revenue loss.
- As there are no records of purchases took place, Owner does not have record which also checks the remaining stocks. So, Objective was to study issue of stock out.

5.2. Data Collection

- For sales data, Entries of Items sold, how much Quantity and what was price paid was written on a notebook after I specifically asked owner to maintain a record of sales made in a day. As it is my father's shop sometime, I also look after the shop. Later I logged that data manually in my excel sheet.
- For customers feedback data, I asked customers who didn't bought anything and maintain a record of what they are looking for, why they didn't purchase.
- For Stocks data, I counted all the items manually in shop and noted in excel sheet on 31st March 2025.
- I have 3 worksheets of data. All section has been described in metadata section

5.3. Data cleaning/Pre-processing

- Dataset used in this project was created in a structured format from beginning. Each sales entry was recorded first in notebook and later in excel in a proper way. It has clearly defined fields such as name, quantity, price per item and total price. As a result, **no major data cleaning was required** to handle issues like missing values, format issues or duplicates.
- However, some minor preprocessing step involved creating a categorical level for each item like Kurta, Shirt to make it easy to classify. After the excel sheet was ready, then it was imported into python using panda's library where more statistical analysis was done.
- Some excel formula to classify week number from date field, allowing weekly-trend analysis.

5.4. Methods used and why they were used: -

- **Sales vs Stock analysis: -** I will compare total sales made vs total stocks available using Grouped bar chart. This will help me to find total number stock sold and compare it with stocks available to find which stock is understock and which is overstocked.
- **Weekly revenue trend analysis** - This allows me to plot **weekly revenue chart** which helped me in identifying peak sales weeks and low-performing weeks to find trends in revenue generation. It was also used to identify highest selling items and low selling items from monthly sales chart to handle overstocking and understocking.
- **Descriptive Statistics**: - I used this which includes mean, median, total revenue, quantity sold, Standard deviation to compare item-wise performance. It helped me quickly identify top-selling items, Average prices and maximum and minimum quality sold. Using standard deviation, I was able to classify **variability and skewness** of each item.

Here datasets are relatively very small and categorical. The goal was to find patterns and interpret the data and give insight. It was not intended to predict anything like we do in regression analysis. It is a very simple method and easy to understand.

6. Results and Findings

Following section summarizes the key patterns observed from sales data and customer feedback data collected over the period of 13 weeks.

6.1. Sales vs. Stock Analysis

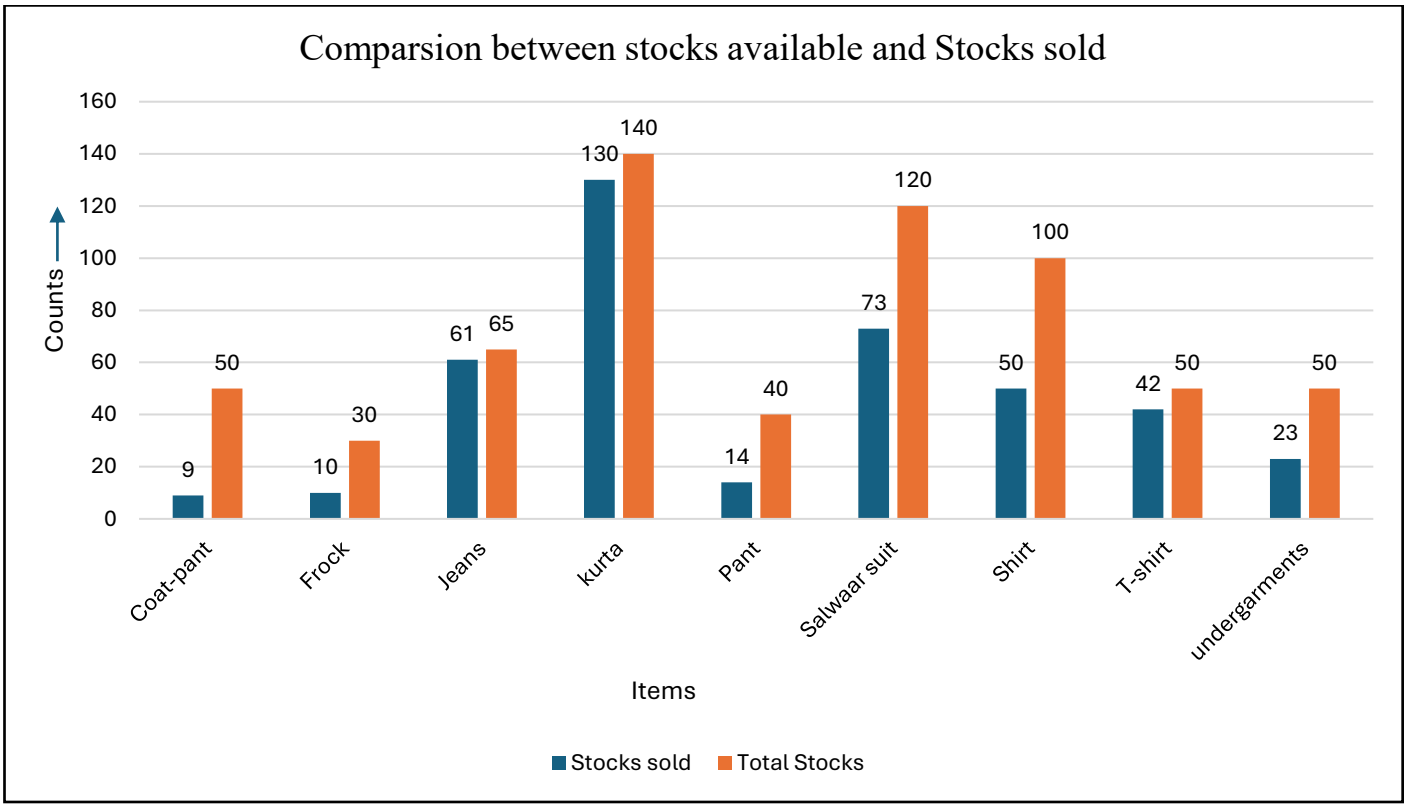

Fig-6.1 A comparison between stocks and sales

As Fig-1 shows comparison between total stocks available and quantities sold for each product category. From the chart, we can see that some items have significant gaps between sold and available, which indicates poor stock alignment.

Items such as Coat-pant, Frock and pant shows low demand compared to stock available suggesting **overstocking** and potentially occupies space which can be used to store Kurta and T-shirt which are **understocked**.

For example, out 50 coat-pant only 9 has been sold and same goes for frock out 10 out of 30 whereas for Kurta 130 has been sold out 150 which indicates high demand and same goes for jeans. Rest of items has shown moderate demand with good amount of supply and demand.

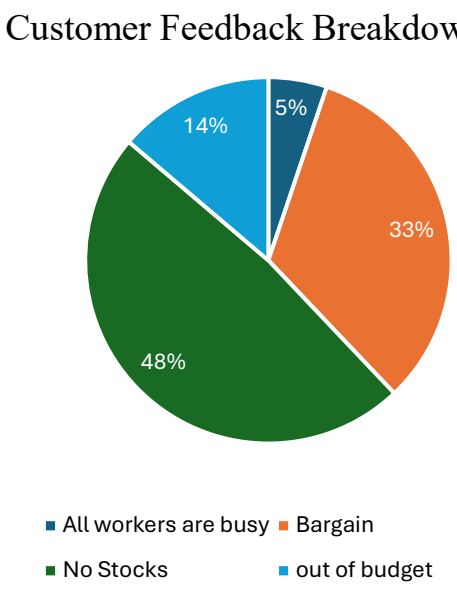

Fig-6.2 pie chart representing percentage distribution of customer feedback

Items	Apr	May	Jun	Grand Total
Coat-pant		1		1
Jeans			1	1
Kurta	1	2	7	10
Pant	1			1
Salwaar suit	2	3	7	12
shirt			1	1
T-Shirt	1		1	2
Grand Total	**5**	**5**	**17**	**28**

Table-6.1 Customer Reported No stocks

Fig-6.2 also shows percentage wise distribution of Customer feedback data. 48% of customer responses were related to "No Stock" followed by bargaining failure (33%). This is Strong sign of **Stock related issues and Revenue loss.**

Table 6.1 shows all the monthly items distribution of customers feedback who said that reason they didn't buy was **"Size/colour not as desired"** meaning they didn't get what they wanted **(No stocks)**. This indicates that understocking. Kurta (10) and Salwaar suit (12) has highest – number of stocks issues reported especially in June.

6.2. Weekly sales Trend

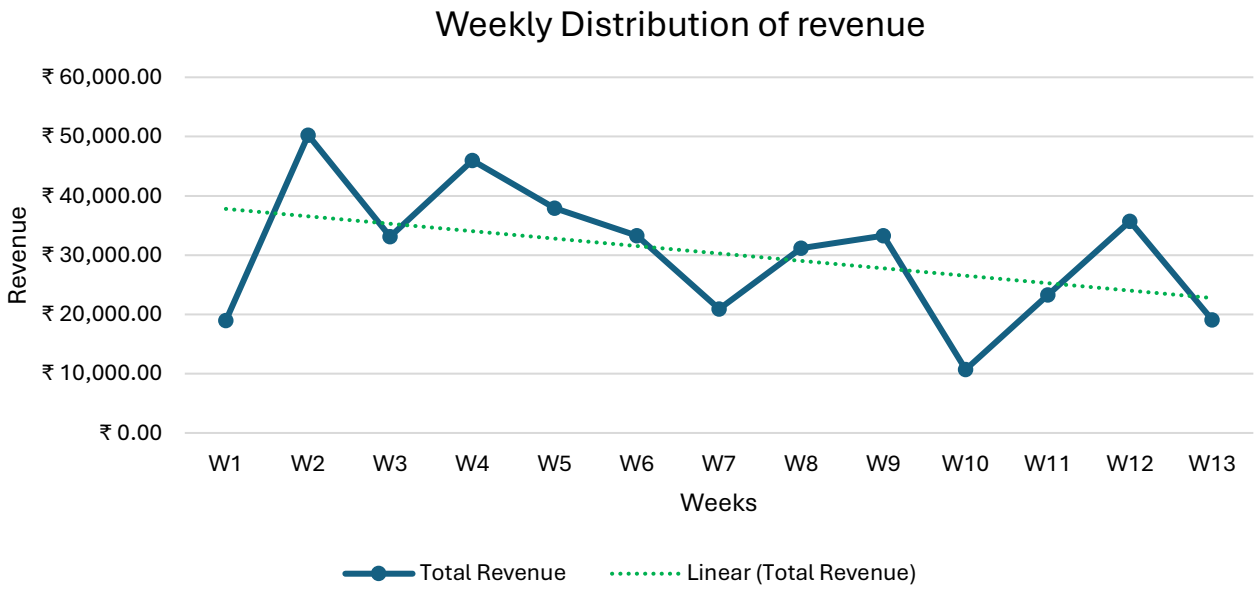

Fig-6.2. Weekly trend in revenue generated

- Above figure shows distribution of revenue across weeks and linear trendline which shows a **downward trend** in generated revenue indicating **Potential Revenue loss**. The graph also shows high fluctuations week to week indicating inconsistency.
- Week 2 with approx. Rs 50,000 were highest revenue generator and week 10 with Rs 10,000 was lowest.

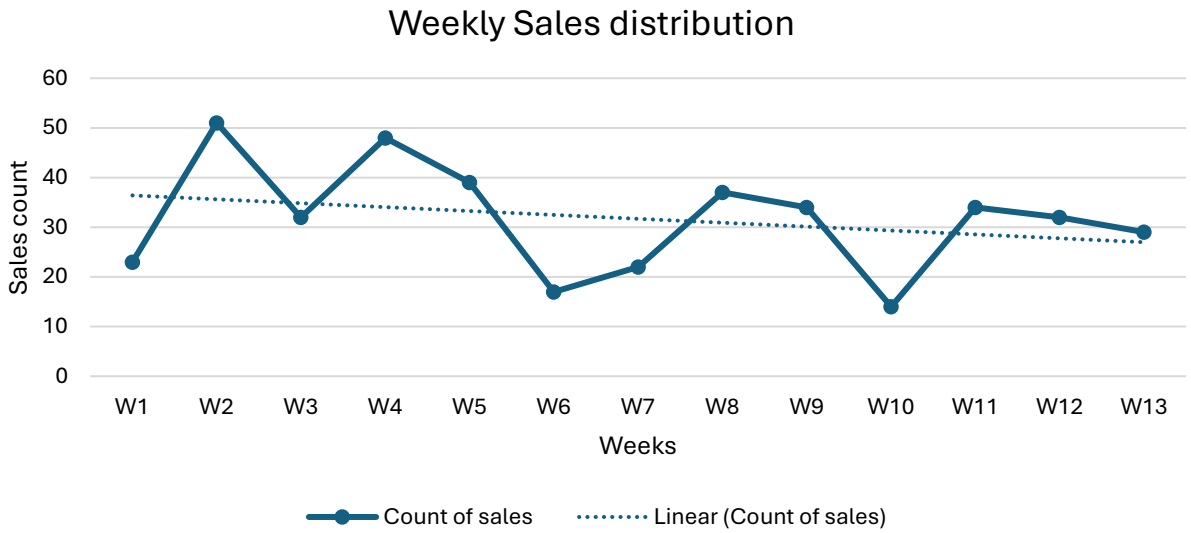

Fig-6.3. Weekly trend in sales

- Fig 6.3 indicates the weekly count of sales. Like revenue trend, it also shows a downward trend in number of sales over the period of 13 weeks.
- Sales are peaking in week 2 with 51 sales and lowest sales in week 10 with 12 sales
- The Similar patters observed in fig 6.2 (Revenue graph) and fig 6.3 (sales graph) confirms that **decline in sales is main reason behind revenue loss**.

6.3. Item-wise Quantity and Revenue comparison

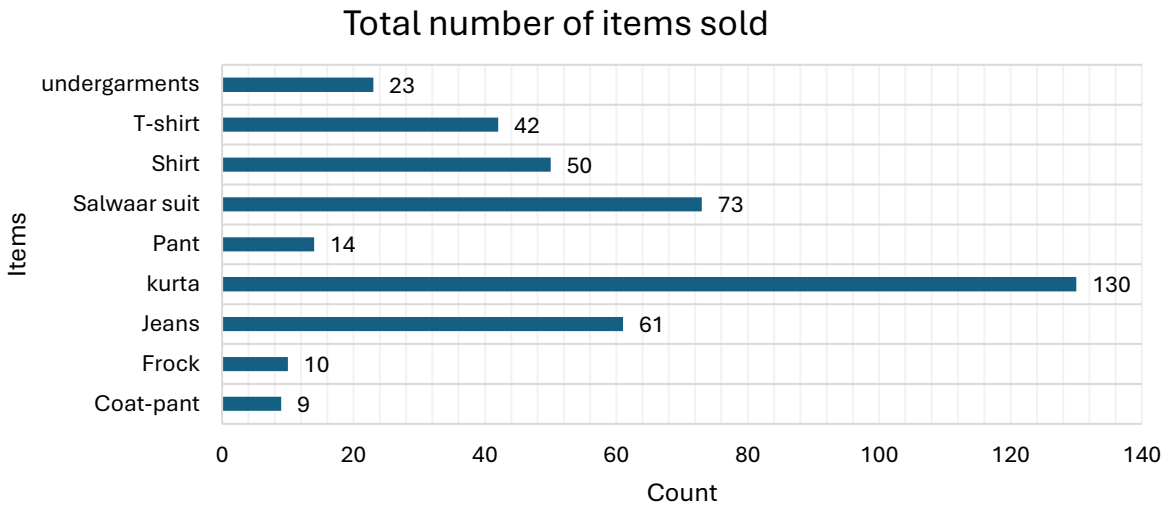

Fig-6.4 item wise sales count

- The above fig illustrates the total number of items sold in 13 weeks. Kurta and Salwaar Suit were the most sold items during the 13-week period. These categories accounted for most sales, indicating strong and consistent demand

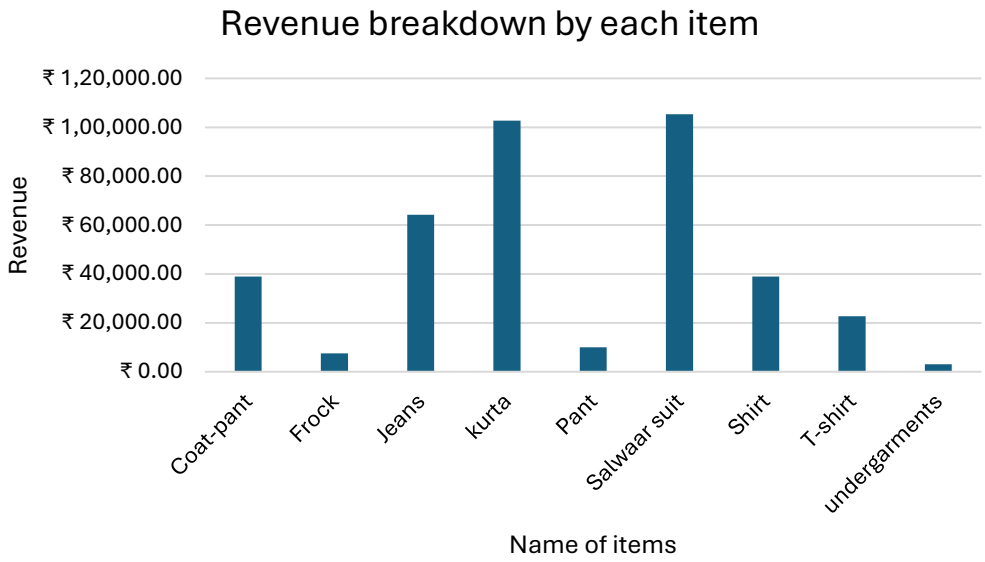

Fig 6.5 item wise revenue graph

- Fig 6.5 illustrates revenue generated by each item. Kurta and Salwaar Suit generated most revenue.
- Comparing fig-6.4, fig-6.5 indicates that Kurta is the most sold item, but Salwaar suit generated more revenue than Kurta.
- Shirts and T-shirts were sold in moderate numbers but generated very little revenue.
- Coat-pants had lower units sold but still generated significant revenue, likely due to high prices.