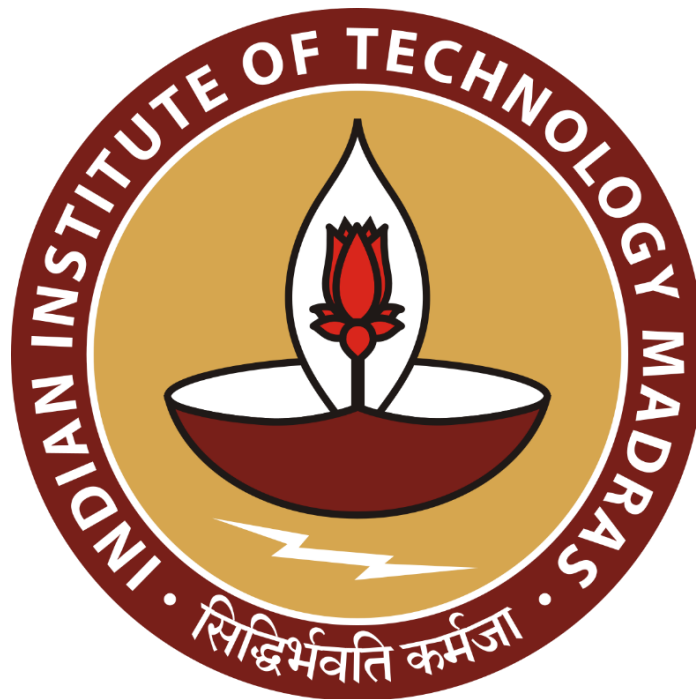# Descriptive Insights: Analyzing Trends in Medical Store

A Final Report for the BDM Capstone Project

Submitted by

Name : ABHIVYAKTI JHA

Roll number:23ds3000137

IITM Online BS Degree Program

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

CONTENTS

EXECUTIVE SUMMARY

Sinha Medicals is a small-scale business located in Labor Colony, near the Police Ground, Rajnandgaon, Chhattisgarh. It operates as a B2C business in the pharmaceutical segment, serving local customers.

The medical shop faces recurring losses, underperforming inventory, and inefficiencies in sales and purchase management. Data analysis reveals that 17.93% of medicines generate zero revenue, indicating significant dead stock. The purchase and sales registers also highlight revenue leakage. Ineffective inventory management, including overstocking low-demand products, ties up capital and impacts profitability.

To address these issues, the objective is to implement a robust inventory management system to phase out dead stock, optimize demand-based procurement, and minimize overstocking. Utilizing tools like Excel for data preprocessing will streamline stock-level analysis and improve profitability.

DETAILED ANALYSIS OF ANALYSIS PROCESS

Process of Data Analysis of Inventory and Sales records:

The data analysis method includes several important processes, each adding to our general understanding of sales income and inventory optimization. In the process, the problem is defined, data is collected and arranged, cleaned and transformed, analysis methods are applied, and finally, some significant insights are drawn.

1. **Data Aquisation approach:**

    Data collecting is a time-consuming process that necessitates careful work to establish credibility with businesses, which calls for building relationship and trust with all parties involved. After working with more than 10 companies, it was observed that our first meetings were frequently brief to a few of times. One early mistake as identified that approaching companies with a straightforward request for data without first establishing a connection. This unintentionally resulted in resistance and unwillingness to share data, underscoring the need of first develop a relationship.

    Finally , successfully data are acquired from a medical shop owner, establishing one of the most important milestones in the jouney of analysis: *data acquisition*. Since Sinha

Medicals is situated near my home and my family buy our medicine from that shop so they trust me with their data.

2. **DATA COLLECTION**

Firstly sales data , inventory data, sales return and purchase return data are collected in an Excel format over the financial year 2022-23.Since they maintained their sales inventory & sales return data in a software named "Pharma Inventory ERP Software" so the data collected by me was in text format which was then converted into excel format. The purchase return data was in raw format so it was also entered into excel.

3. **DATA PREPROCESSING**

The obtained data was carefully cleaned before analysis because it is a combination of clean and mixed information. This was done after the data was converted into an Excel file.Sales data is a cumulative set of data from March 2023 to February 2024, however it has to be organized monthly in order to be analyzed every on a monthly basis.

Details of data cleaning and preparation:

Sales Data cleaning:

Sales data consist of columns named as date, bill receipt ID, customer name, medicines names, and total amount. So for analysis purposes, only data from the column date , name & total amount was taken into consideration. Sales data for the financial year 2023-24 provided by owner consist of following table:

DATE	R- ID	NAME	MEDICINES	AMOUNT
Item purchase date	The id present on bill receipt	The name of customer	Medicines bought by customers	Revenue generated from each customer

The sales sheet has five columns and 22718 rows. The sales sheet needed minor modifications because it combined clean and mixed data. The sales data was initially separated into the months of April 2023 through March 2024 in order to create uniformity. Excel's filter feature was used to evenly distribute the data. The data was not in the right shape because it was taken straight from the software. Numerous cells were combined with nearby cells. Therefore, cleaning was done using the filter and sort function once again after the data had been split on a monthly basis.

Python is used to analyze the data and find that some values are missing once it has been properly formatted and structured. To get the analysis correct, all of the missing data are substituted with the month's mean value using the Simple Imputer module of Sklearn.

Following is the link of python code for above imputation:

Link to collab file for preprocessing

https://colab.research.google.com/drive/1mbfGf82erezJ1zKdqKshmJgeedCUw9DC?usp=sharing

I started the analytical process by carefully grouping the sales data into three columns: revenue generated, customer name, and date of purchase. However, nothing useful can be inferred from the customer names because it is a healthcare shop and the customers fluctuate with the seasons. In the final form, the analysis was reduced to the revenue generated and the date. The following is the final sheet of sales data:

DATE	REVENUE
The purchase date	Revenue generated by each purchase

The following summarize data obtained from above operations

Month	Revenue(Rs)
April	1081124
May	1099072
June	1013966
July	1327851
August	1265641
September	1604677
October	1206557
November	1221094
December	1316587
January	1118877
February	381886
March	1330781

Table 1 –Revenue per month

Following mathematical key features can be done to gain some insights and optimizes operation

➢ **Total Revenue**

Total Revenue=∑Amount generated per month

Item name	Packing	Op.balance	Purchase	Sale return	Total stock	Closing(Rs)	Sale value(Rs)
The name of the medicine	Quantity of medicine	Balance stock of medicine	New stocks of medicines	Medicine return to manufacturer	Total stocks in store	The amount at which each stock got sold	The amount received at the end of stock sales

> **Average Sales Growth Rate (Over Time)**:

$$GR(PM) = \frac{Sales\ in\ current\ period - sales\ in\ previous\ periods}{sales\ in\ previous\ period} * 100$$

$$Average\ growth\ rate(\%) = \frac{\sum Growth\ rate\ per\ month}{12}$$

> **Customer Metrics**

> Average spend per customer $= \frac{Total\ revenue}{Total\ Customer}$

4. INVENTORY SALES DATA

The data collected has two different forms of inventory data sheets. The first one is inventory records by medicines names and the second one is inventory records by manufacturer names.

Medicines Inventory:

The medicine inventory consists of 8225 rows and 11 columns. Rows consist of medicines names and column consist of its features. The datasheet collected has data in the following manners:

The owner provided mixed data, so some minor cleaning and modifications were necessary. Using filters, sorting, and duplicates, all of the cells were first brought to the proper structure. All of the missing values are then ignored by utilizing the Python sklearn package. Because of the large size of the data, medications that generate more than Rs10000 in revenue are divided into various segments based on their intended usage for additional analysis. There are 265 medicines which give revenue> Rs 10000 and 1700 medicines which give revenue>Rs 1000.

Manufacturer Inventory Data:

In this datasheet provided by owner stocks records are maintained based on manufacturer. There are 7846 rows and 8 columns. For initial all the same name

manufacturer data was merged to get the overall count of medicines from that particular manufacturer. By using python package all the missing values are replaced by mean value of that particular medicine to get analysis.By using pivot tables count, average,mean etc data analysis was done.

After preprocessing the rows and columns are reduce to 765 and 7 respectively.

Item name	Packing	Op.balance	Purchase	Sale return	Total stock	Closing(Rs)	Sale value(Rs)
The name of medicine	Quantity of medicine	Balance stock of medicine	New stocks of medicines	Medicine return to manufacturer	Total stocks in store	The amount at which each stock got sold	The amount received at the end of stock sales

An analysis has been done based on revenue bins for sales segmentation which is shown in further findings and results section.

Following mathematical key features is calculated to get better insights of inventory

➢ Dead Stock percentage

$$\text{Dead Stock}\% = \frac{Number\ of\ Unsold\ items}{Total\ items}$$

5. **SALES RETURN DATA:**

The owner maintains a Sales Return Register that records the stocks returned to manufacturers due to reasons such as expiration and overstocking. However, the register only includes data on the amount of returns and the manufacturer involved. According to the owner, only 55% of the returned merchandise is reimbursed or exchanged, while the remaining 45% results in a financial loss. Given the relatively small size of the dataset compared to other data registers, only minor adjustments are needed to ensure accuracy. Addressing these inefficiencies and refining data tracking could help mitigate the impact of losses on profitability.

RESULTS AND FINDINGS

Descriptive analysis of overall sales data for 2023-24

Descriptive statistics	Total revenue	Customers
Sum	13968113	22721
Mean	1164009.417	1893.417
Standard Error	84156.98038	122.4099
Median	1213825.5	2013.5
Standard Deviation	291528.3317	424.0402
Sample Variance	84988768164	179810.1
Kurtosis	5.012255412	8.25017
Skewness	-1.68756219	-2.73242
Range	1222791	1598
Minimum	381886	637
Maximum	1604677	2235
Count	12	12

Table 2- Descriptive statistics of overall sales data for financial year 2023-24

Some observations from descriptive analysis:

- Both datasets(revenue & customers) show left-skewed distributions, as evident from negative skewness values. This implies a few low values disproportionately affect the averages.

- The high kurtosis values for both revenue and customer count indicate that the data have outliers or extreme values. Due to February sales this outlier is seen in the analysis.

- Revenue and customer counts are relatively variable, as seen from the standard deviations and ranges.

- The median values being close to or higher than the means suggest that outliers do not significantly affect the distributions but still show some asymmetry.

Some insights from the sales data:

1. 22718 customers brought their medicines from the store, averaging approximately 1900 customers per month and 60 customers per day.

2. Since the overall revenue of the store is Rs13968113 annually the average monthly revenue they generate is Rs 1164000 which is an average Rs 38800 per day

3. Overall trend: The revenue shows a slight downward trend peaks

4. Seasonal peaks & dips: September and February saw the largest and lowest sales, respectively. Upon speaking with the owner, I learned that the store was closed from February 3 to February 18 due to personal emergencies. Customers obtained their temporary medications from nearby stores because the store was closed.Peak revenue can be due to several factors, such as seasonal shifts and increased demand for health supplements, immunity boosters, and flu medications.
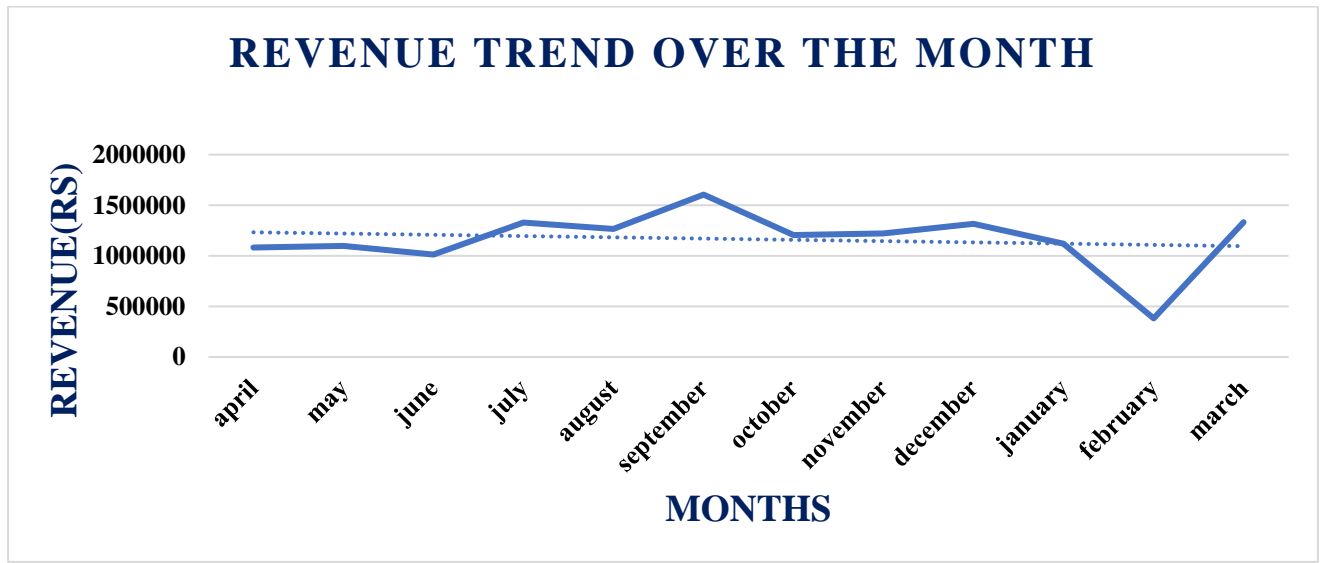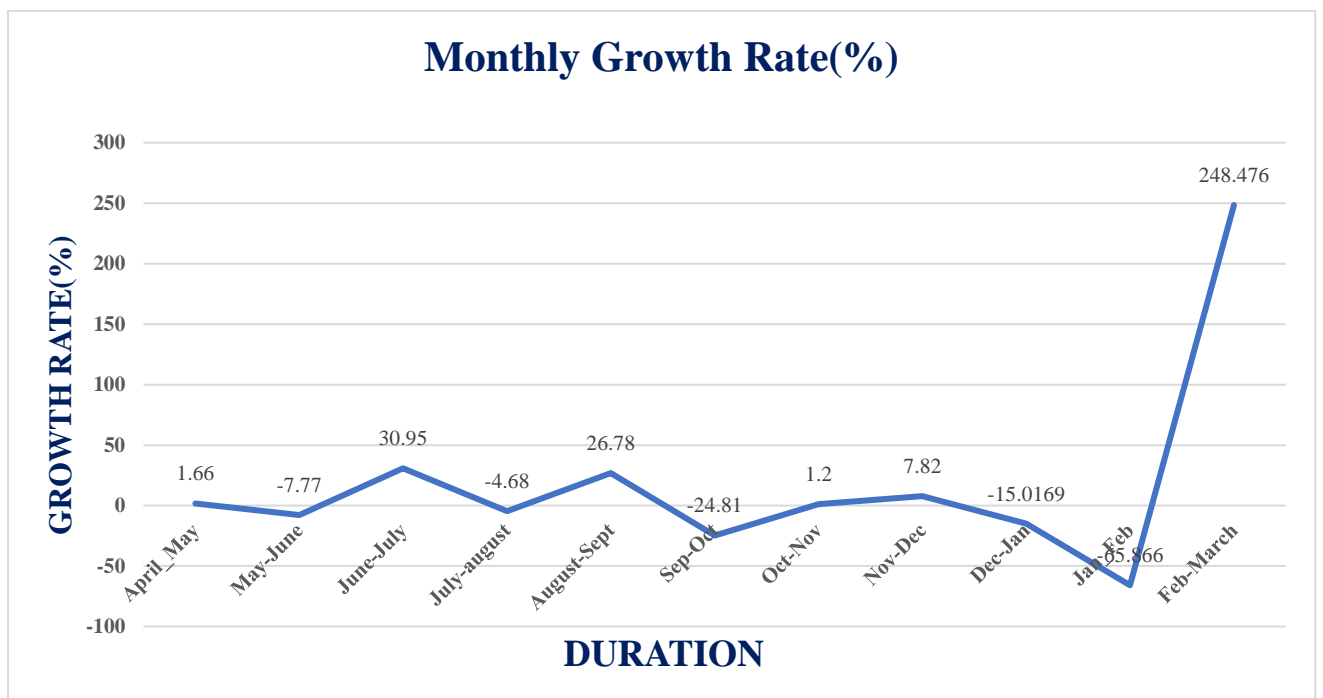

Figure 1 – Revenue trend over the month

Figure -2 Growth rate for overall months

The above figure 2 illustrates the sharp increase in growth from February to March, which can be attributed to the store's non-operation on certain days in February.

The data indicates that the highest growth rate was observed in March, largely due to exceptionally low sales in February, which created a significant base effect. An upward trend in growth between July and September can be attributed to seasonal weather changes, leading to an increase in cases of flu and infections, thereby boosting sales. From October to January, growth rates show a gradual rise followed by a decline. This trend is likely driven by the peak winter season, which typically accelerates demand for medicines related to cold and respiratory ailments.

Inventory data Insights:

There are total of 8225 medicines in the inventory data which is preprocessed and hence reduced to 8127.Since the data is very large and categorising each medicines iwas a very tedious task Hence the medicines are divided in to two ranges:

After preprocessing, the inventory data's total of 8225 medications was lowered to 8127. Since classifying each medication was a tedious operation due to the size of the data, the medications are separated into two ranges

• The medications that bring in over Rs 1000 in revenue
• The medications that help generate income between Rs 1000 and Rs 10,000
There were 265 medications with sales exceeding Rs 10,000 and 1700 medications with sales exceeding Rs 1000. Drugs with sales over Rs10,000 are divided into many categories and used to analyze which market sector has the highest sales margin.

The following tables show the category and its value in terms of medicines:

Medicines category	Medicines
Antibiotics	40
Antidiabetic	25
Cardiovascular	21
Pain relief	25
Respiratory and cough	25
Vit & supplements	25
Hormonal	14
Neurological	16

Gastro	15
General medicines	59

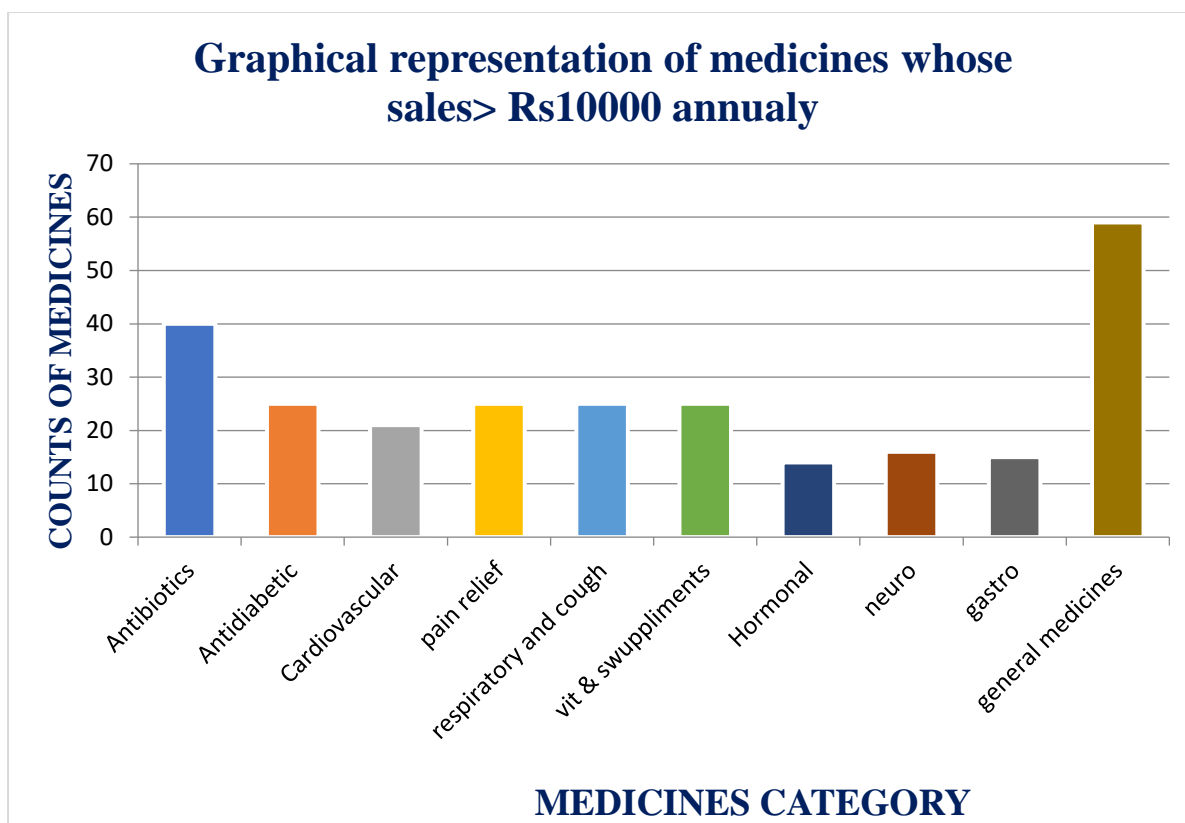*Table-3 Medicines category wise and its counts*

Figure 3- Medicines contribution in revenue whose sales greater than Rs 10000

Dead Stocks based on inventory data analysis

After the preprocessing, it was analyzed that there is total of 1475 medicines that are not giving any revenue which means their sales are zero.

The following chart represents the products in the horizontal axis and their sales values in the vertical axis.

➢ **Overall Sales Performance**:

All the products in the chart represent non-moving items with negligible sales (close to zero).

The graph in figure4 underscores a critical concern with these products, despite their presence in the company's portfolio

➢ **Product Distribution**:

The list includes a mix of:

➢ Vaccines (*Infanrix Hexa Inj*, *Biovac-A Vaccine*),

➢ Nutritional supplements (*Ensure Choco*, *PediaSure Vanilla*),

➢ Therapeutic drugs (*Tendia Tab*, *Neo Mecrazole-20 Tab*).

This indicates that the underperformance spans across multiple product categories and is not isolated to a single segment.
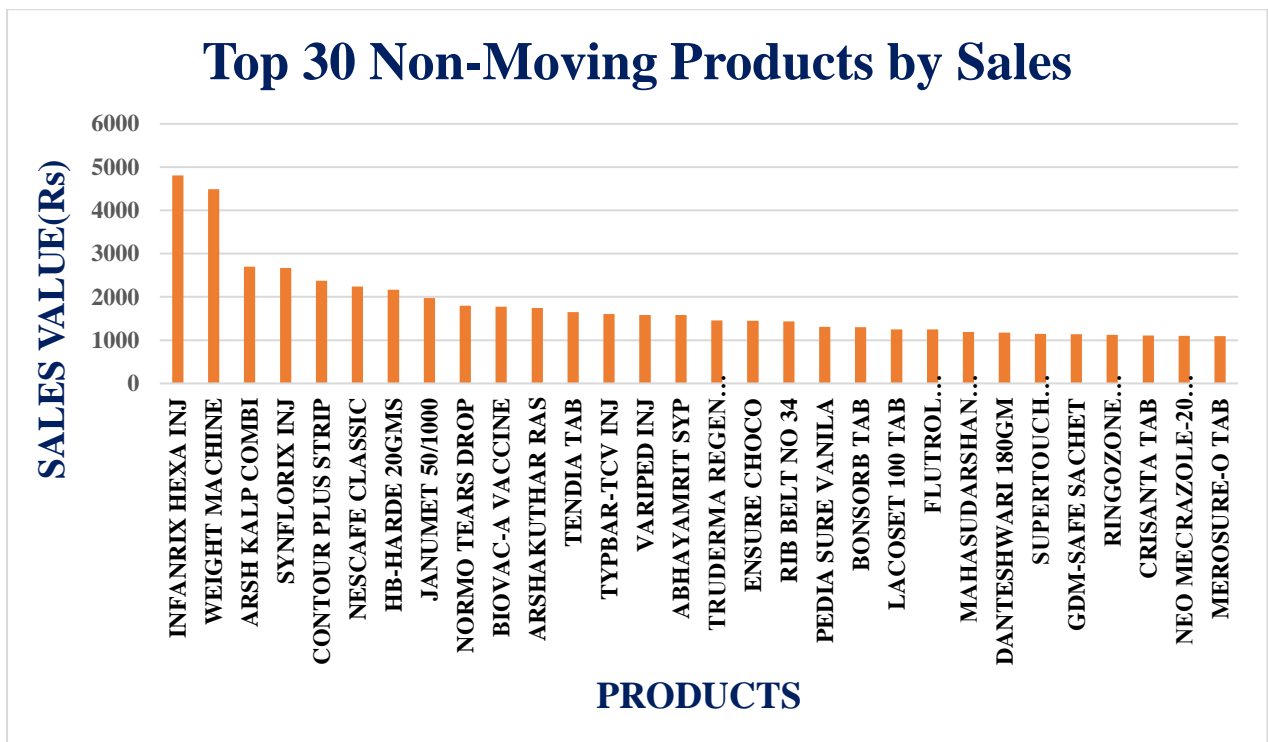

Figure 4-Top 30 nonmoving products(dead stocks)

	Total inventory	Dead stock in inventory
No of medicines	*8225*	*1475(17.94%)*
Inventory values(Rs)	*Rs 16564940*	*Rs 2441005*

Table 4- comparison between total stocks and dead stocks

The average amount spent by customers per visit:

The below graph shows the following insights:

1. Customer spending averages per visit peak in September at Rs. 802, followed by a decline in the subsequent months.
2. January marks the lowest spending at Rs. 511, indicating a potential post-holiday slowdown.
3. Spending gradually increases again in March to Rs. 656, showing signs of recovery.
4. February although has the lowest sales of all the months but customer average spend increases this is due to changes in weather conditions.
5. Spending is relatively stable from April to August, with a slight upward trend.
6. Seasonal factors, promotions, or events might influence spending patterns, particularly the September peak and January dip
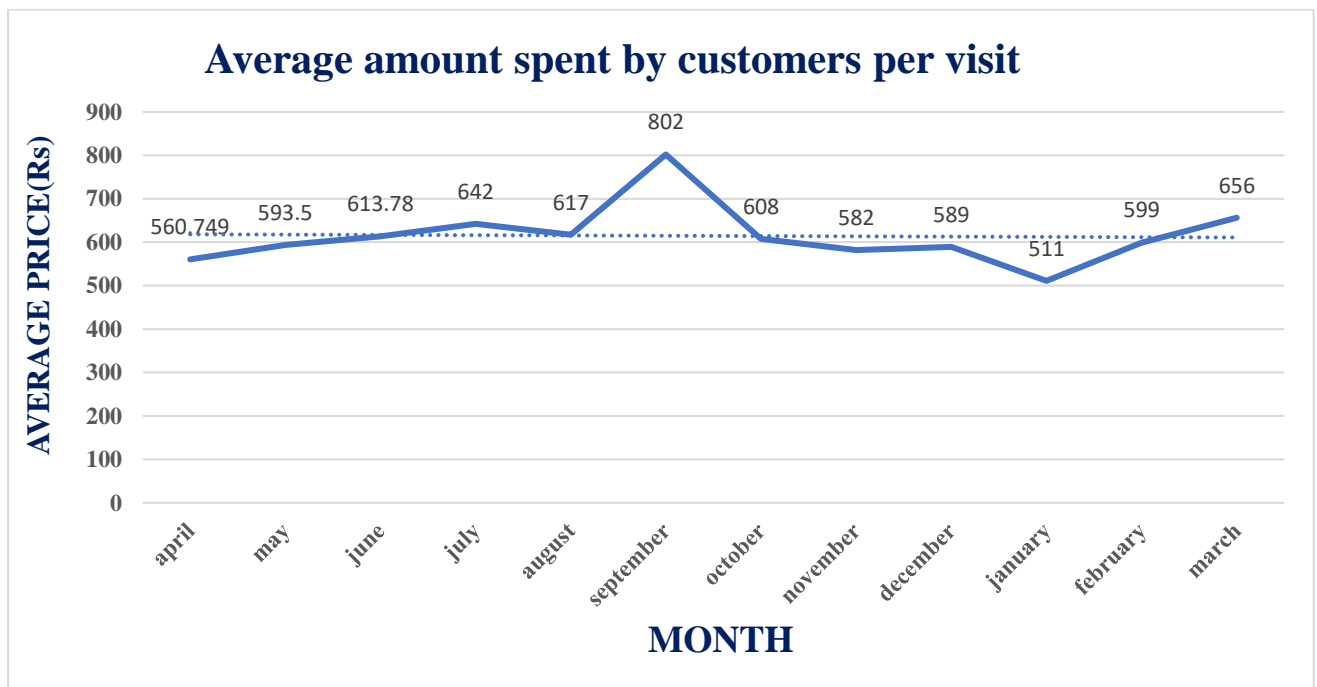

Figure 5 – Average amount spent by customers per visit

There are data in which based on the manufacturer medicines are categorize.There are total of 756 manufacturers providing 7237 medicines , whose medicines are sell by the store.

Top manufacturer based on revenue:

From the graph(figure 6) following conclusions can be drawn:

1) From the chart it is evident that SUN manufacturer gives the maximum revenue share with total items sold of 73742 and its gross value is 826012.

2) The top five contributors to the revenue share are **SUN, MAPLE, INTAS, LUPI, and CIPLA** having revenue share of **Rs 826012, Rs 668917, Rs 581484, Rs 413946, Rs 342730,** and the items sold are 73742, 62068, 50556, 18456, 25569 respectively.

In the following graph, the top contributors in terms of medicines is shown:

- SUN Pharma leads with the highest count (251), followed by INTAS (243) and MACLE (227), indicating strong competition among the top three.

- A notable gap exists between the top three and mid-tier manufacturers like CIPLA (148) and DRRED (125).
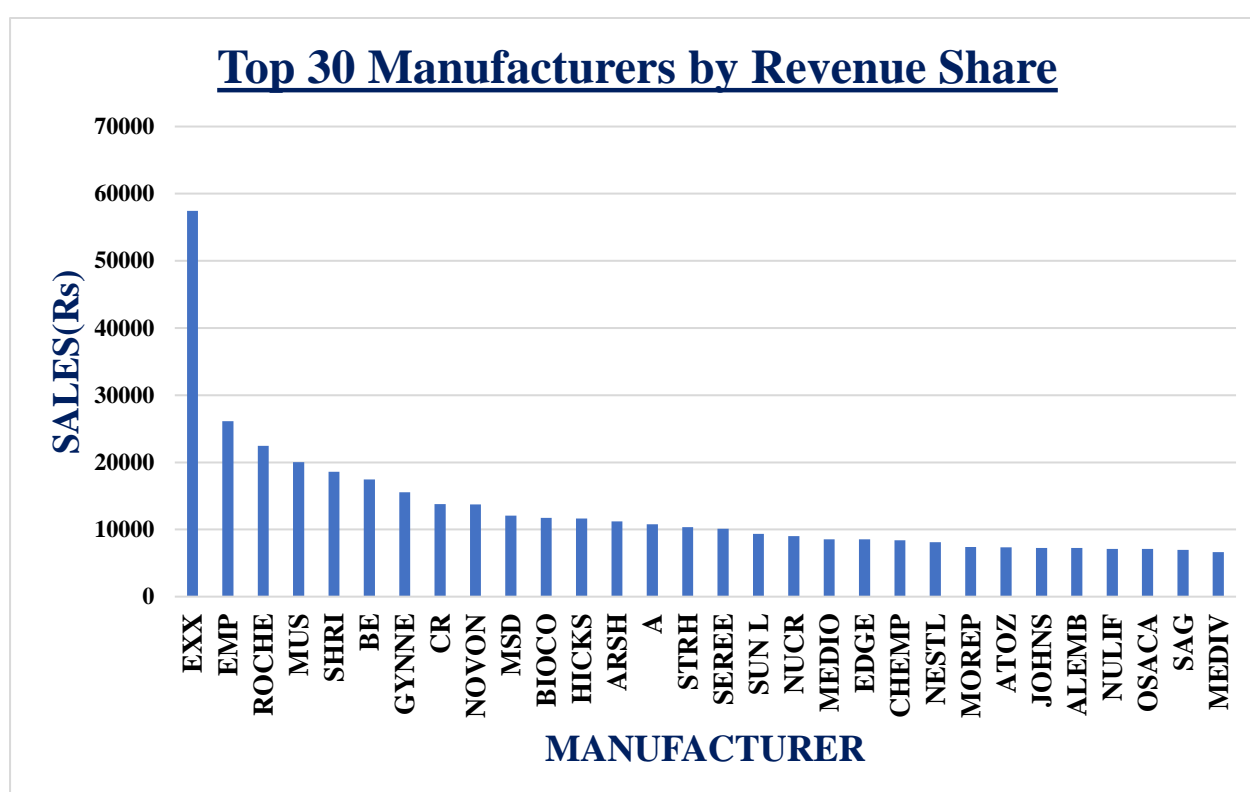

Figure 6 – Top manufactures whose share in revenue is maximum

Mathematical key features calculation for overall manufacturer data:

➢ There are 756 manufacturers with 7237 medications in total. The total number of items sold was 1122730, and the gross sales were Rs 1402491.

From the below graph, it is evident that most of the manufacturers are in the category of Rs 0-50000 revenue generation. 687 manufacturers fall in this category. 33 manufacturers falls on the revenue of Rs 50001-100000 and very few manufacturers bring the higher range of revenues.
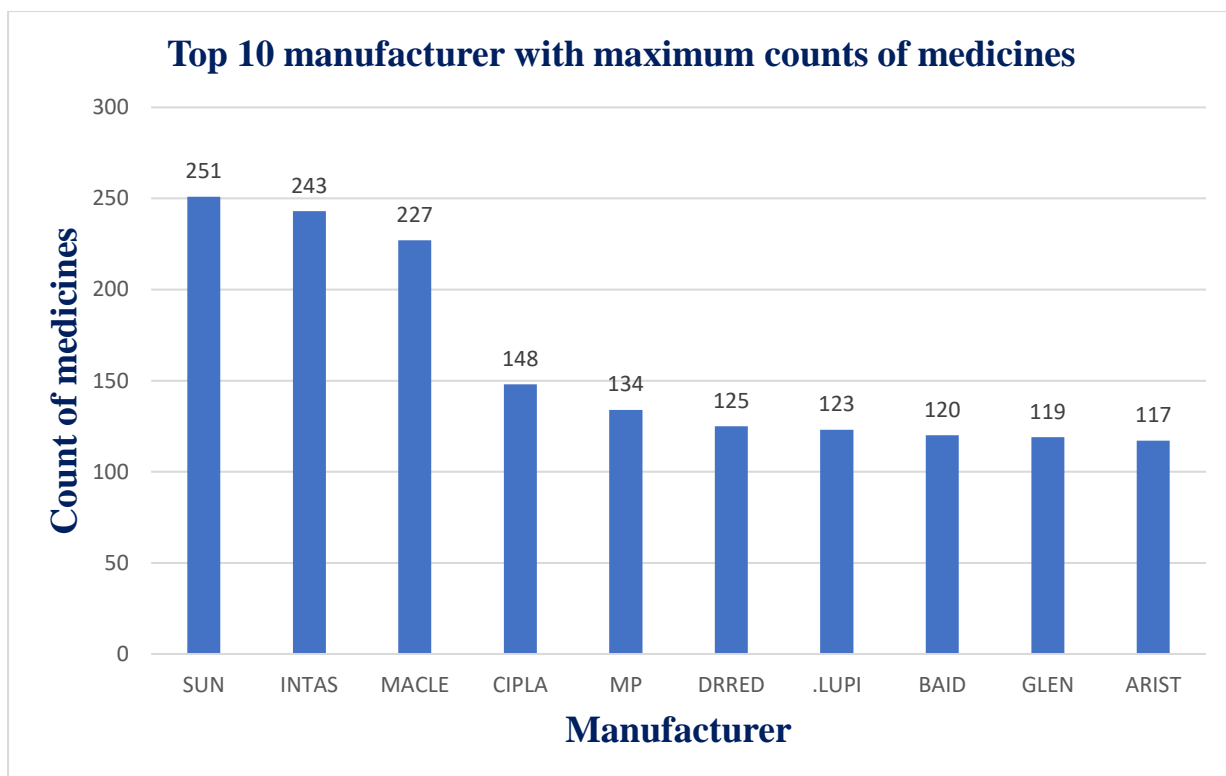
Top 10 manufacturer with maximum counts of medicines

Figure 7- Top 10 manufacturer with maximum medicines counts

MFG	Count of MFG	Sum of Sold items	Sum of Gross(Rs)	Average price per item(rs)	revenue contribution by mgf(%)	medical efficiency	sales per item sold(Rs)
SUN	251	73742	826012.46	11.20	5.88	3290.88	293.79
INTAS	243	50556	581484.57	11.50	4.14	2392.94	208.04
MACLE	227	62068	668914.84	10.77	4.76	2946.76	273.42
CIPLA	148	25569	342730.57	13.40	2.44	2315.74	172.76
MP	134	20363	202807.61	9.95	1.44	1513.48	151.96
DRRED	125	23298	318955.96	13.69	2.27	2551.64	186.3
.LUPI	123	18456	413946.29	22.42	2.95	3365.41	150.04
BAID	120	1157	91510.46	79.09	0.65	762.58	9.64
GLEN	119	9932	277675.58	27.95	1.97	2333.40	83.46
ARIST	117	31263	306804.29	9.81	2.18	2622.25	267.20

Table 5- Mathematical calculation of top 10 manufacturers (with maximum medicine counts)

Sales segmentation by bin analysis:

There are 166 manufacturers in total whose count of sold goods is less than 10, but some of them nonetheless generate gross revenue higher than 10,000, according to manufacturer-based data. Filtering this manufacturer reveals that 138 manufacturers in total contribute less than Rs 1,000 in revenue from their drugs.This manufacturer's medications can therefore be considered dead stock.
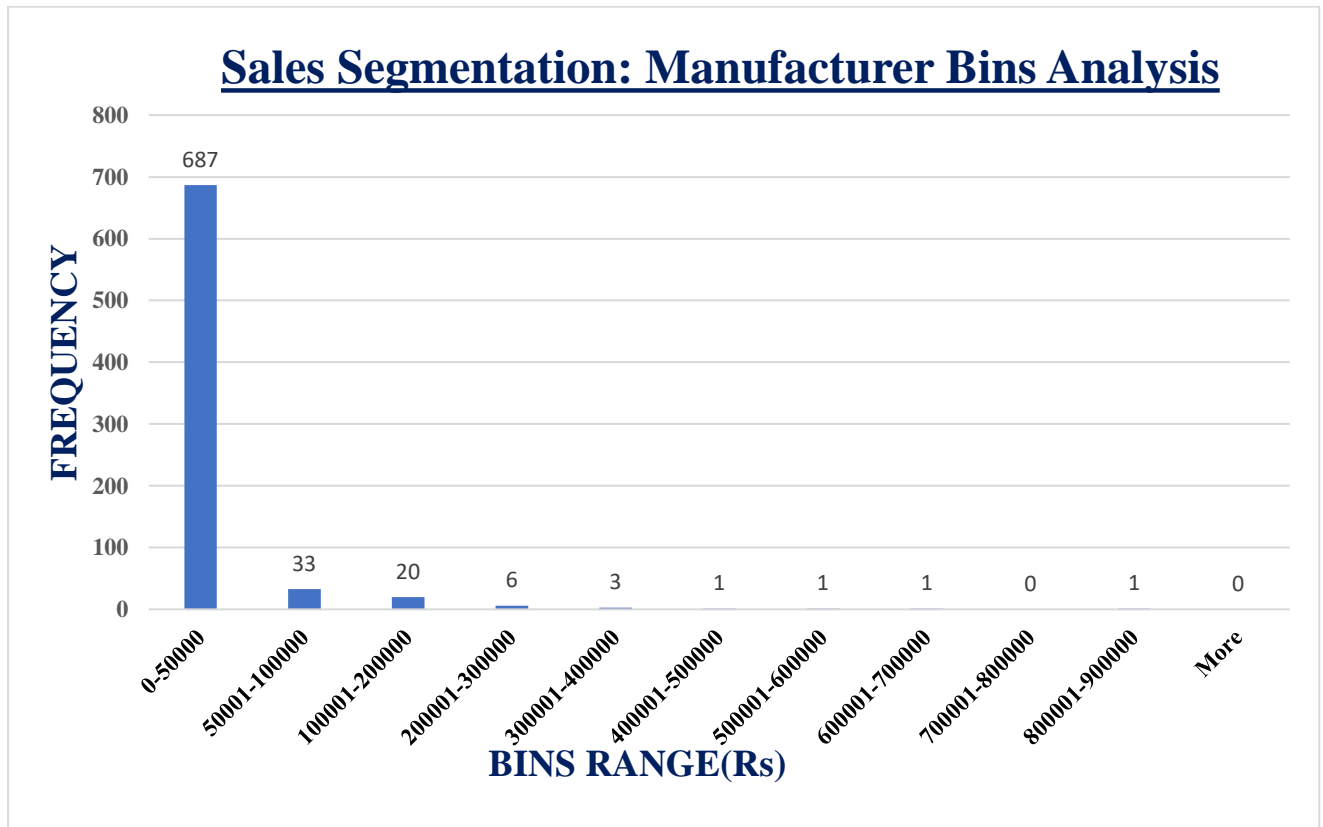

Figure 8- Sales contribution by manufacturer(In bins analysis)

INTERPRETATION OF RESULTS AND RECOMMENDATIONS

SALES ANALYSIS:

From the sales data analysis some key features needed to be calculate for getting better overall performance of store.

Some mathematical key calculations based on sales data:

Total Revenue

Total Revenue=∑Amount generated per month
Total Revenue=Rs 13968113

Interpretation: Indicates strong revenue generation overall, but month-to-month fluctuations suggest some instability in sales performance

Average Sales Growth Rate (Over Time):

$$GR(PM) = \frac{Sales\ in\ current\ period - sales\ in\ previous\ periods}{sales\ in\ previous\ period} * 100$$

GR(PM) [Growth rate per month]

Months period	Value(%)
April_May	1.66
May-June	-7.77
June-July	30.95
July-august	-4.68
August-Sept	26.78
Sep-Oct	-24.81
Oct-Nov	1.2
Nov-Dec	7.82
Dec-Jan	-15.0169
Jan_Feb	-65.866
Feb-March	248.476

$$Average\ growth\ rate(\%) = \frac{\sum Growth\ rate\ per\ month}{12}$$

Average Growth Rate=18.071%

This is a healthy growth rate over time, but the large fluctuations in monthly growth rates highlight potential challenges:

Significant Negative Growth Periods:

➢ January-February (-65.87%) and September-October (-24.81%) suggest weak demand or operational issues.

High Growth Spikes:

> ➤ February-March (248.48%) indicates seasonal or event-based sales spikes but could be unsustainable long-term.

Customer Metrics

$$Average\ spend\ per\ customer = \frac{Total\ revenue}{Total\ Customer}$$

Average spend per customer=Rs 620

<u>INVENTORY MANAGEMENT:</u>

Dead Stock percentage based on inventory data

$$Dead\ Stock\% = \frac{Number\ of\ Unsold\ items}{Total\ items}$$

Dead stock=17.94%

Interpretation: About 18% of stock is dead stock (unsold or obsolete items). This is a relatively high proportion, suggesting inefficiencies in stock management

Average price per item:

To determine the average price of items sold by each manufacturer, the following formula is used:

$$Average\ Price\ Per\ Item = \frac{Sum\ of\ Gross\ Values}{Sum\ of\ sold\ items}$$

Average price per item=Rs 12.49

Medicine Efficiency (Revenue per Medicine)

To determine the revenue generated per different medicine offered by each manufacturer, this formula is used:

$$Revenue\ per\ medicine = \frac{Sum\ of\ gross\ values}{Count\ of\ medicines}$$

Revenue per medicine=Rs 1937.84

Sales per Item Sold for Each Manufacturer

To understand how effectively the manufacturer's items are selling, we calculate the Sales per Item Sold ratio:

$$Sales\ per\ item\ sold = \frac{Sum\ of\ sold\ items}{Counts\ of\ medicines}$$

Sales per item sold=Rs 155.13

Deadstock based on manufacturer data:

$$Deadstock = \frac{Manufacturer\ contributes\ almost\ zero\ in\ sales}{total\ manufacturer}$$

Dead stock=18.254%

Interpretation: About 18% of your stock is dead stock (unsold or obsolete items). This is a relatively high proportion, suggesting inefficiencies in stock management.

Stock turnover ratio

Efficiency in managing inventory

$$Stock\ Turnover\ Ratio = \frac{Cost\ of\ Goods\ Sold}{Average\ Inventory}$$

COGS=(Beginning Inventory + Purchases During the Period) – (Ending Inventory)

COGS=Rs16430865 + Rs13985186 - Rs2445678

COGS=Rs 2,79,70379

Average inventory=(beginning inventory + ending inventory)

Average Inventory=Rs 9438271

Stock turnover ratio = 2.96

Indicates the inventory is replenished nearly three times per period. While this is acceptable, it's below the industry standard of **4-6 for pharmacies**, implying:

- Overstocking or slow-moving inventory.
- Inefficient replenishment practices

RECOMMENDATIONS:

Following are some recommendations based on overall analysis:

1) Improves sales growth consistency:There is variation in revenue per month. If these periods coincide with holidays or other external factors, consider adjusting sales strategies, such as targeted marketing campaigns, special offers.

2) Identify slow moving products: Regularly review inventory and identify products with low sales and try not to overstock it.

3) Medicines that contributes high in revenue it should be closely monitored to avoid shortages.

4) Manufacturers with high sales performance should be prioritized for new products or exclusive deals. On the other hand, manufacturers contributing to dead stock (18.25%)

should be evaluated for product relevance and sales performance. If necessary, replace them with more popular brands.

5) Loyalty program: Introduce a customer loyalty program to incentivize repeat purchases. Offer discounts or points for every purchase that can be redeemed later, which can increase the average spend per customer and improve overall sales.

Both opportunities and concerns are identified by the data analysis. The store can boost profitability and operational efficiency by concentrating on cutting dead stock, enhancing inventory turnover, and optimizing the product mix. Furthermore, strengthening client interaction, utilizing periods of high growth, and increasing sales consistency will all help to promote sustainable growth and improved overall performance.