Enhancing Operational Efficiency and Customer Satisfaction: A Data-Driven Approach for MS Medical Store

A Mid-Term Report for the BDM Capstone Project

Submitted by

Name: Aryan Bhardwaj

Roll number: 22F2000384

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

1 Executive Summary

MS Medical Store, founded by **Mr. Prashant Sharma (Sonu Sharma)** around **15 years** ago, located near **Nirphad Eye Hospital in Mathura, Uttar Pradesh**, is a well-established pharmacy that provides essential healthcare products, particularly specialised eye-related medicines such as drops, gels, and ointments. Despite its small size, with only one employee besides the owner, it is known for its high-quality products and personalized customer service

The store offers products such as lubricating eye drops, antibiotic eye drops, anti-inflammatory eye gels, and eye-related ointments/tablets, with prices ranging from **₹0.91 to ₹665**. Operating out of a **single room** that doubles as both the inventory and customer counter, the store faces operational challenges in **inventory management**, **service efficiency**, and **revenue optimisation**. These challenges include difficulties in maintaining optimal stock levels, long customer wait times during peak hours due to limited staffing, and ineffective marketing strategies, particularly for high-margin products.

This report addresses these challenges through a detailed analysis of sales data collected over a **four-month period** (April 1, 2024 - August 5, 2024) from the store's **Marg ERP software**. The dataset, consisting of 4,500 sales transactions involving 169 unique products, was cleaned and analyzed using advanced data analysis techniques to uncover patterns in sales trends and product performance.

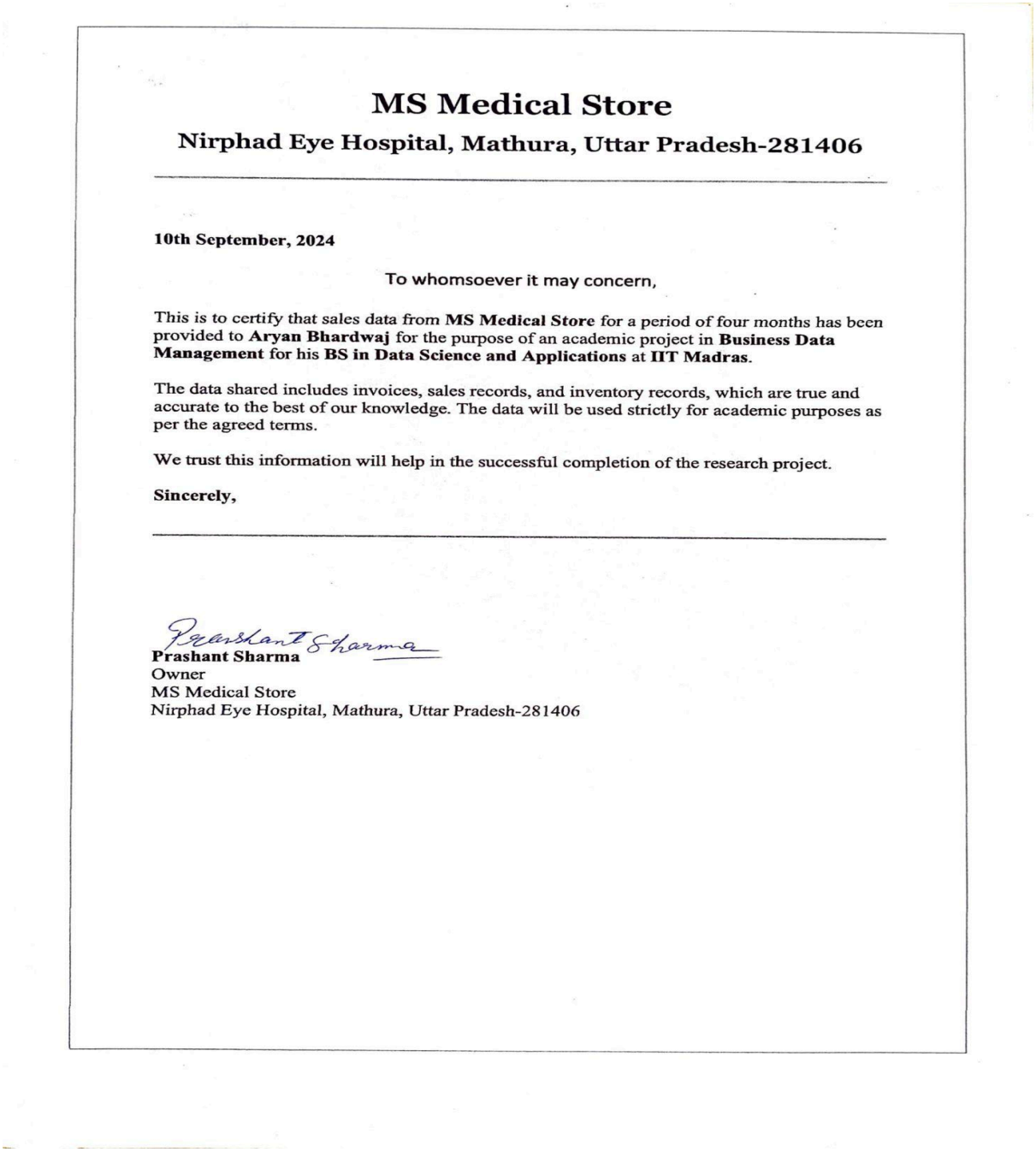
The first analysis conducted was the **Hourly Sales Analysis**, which aimed to identify peak sales hours and optimise resource allocation, particularly staffing. By examining transaction timestamps, we observed that the store experiences the highest sales between **10 AM and 12 PM**, aligning with doctors' visiting hours at the nearby hospital. A secondary peak occurs in the late afternoon, confirming that mid-morning to early afternoon is the most critical period for sales. This analysis helps the store in optimizing staffing levels to meet customer demand during these busy periods.

The second key analysis was the **Revenue Share Analysis**, designed to determine which products contribute the most to the store's revenue. By calculating the total revenue for each product, it was found that **RHOOS DROP, MIXEE LP, and RHOOS GEL** together account for nearly 40% of the store's total revenue. A **bar graph** and **treemap** were used to visually represent the top 20 revenue-generating products. These insights are crucial for the store's inventory management, ensuring that high-demand products are always in stock, while lower-performing items can be managed more efficiently.

The tools used for this analysis included **Python** (with libraries such as Pandas, NumPy, and Seaborn) and **Excel**. The insights gained from the hourly sales and revenue share analyses will allow MS Medical Store to improve its operational efficiency by optimizing stock levels, enhancing customer service during peak hours, and strategically promoting high-revenue products to maximize profitability.

2 Proof of Originality

2.1 Letter of Authorization:



MS Medical Store

Nirphad Eye Hospital, Mathura, Uttar Pradesh-281406

10th September, 2024

To whomsoever it may concern,

This is to certify that sales data from **MS Medical Store** for a period of four months has been provided to **Aryan Bhardwaj** for the purpose of an academic project in **Business Data Management** for his **BS in Data Science and Applications** at **IIT Madras**.

The data shared includes invoices, sales records, and inventory records, which are true and accurate to the best of our knowledge. The data will be used strictly for academic purposes as per the agreed terms.

We trust this information will help in the successful completion of the research project.

Sincerely,

Prashant Sharma

Prashant Sharma
Owner
MS Medical Store
Nirphad Eye Hospital, Mathura, Uttar Pradesh-281406

2.2 Letter of Authorization:

A. **Store Name**: MS Medical Store
B. **Location**: Near Nirphad Eye Hospital, Chhatikara, Mathura, Uttar Pradesh-281406
C. **Hours of Operation**: The store operates daily from **9:00 AM to 8:00 PM**, catering to customers throughout the day and facing peak traffic during patient visits to the hospital.
D. **E-Mail**: mailatprashantagra@gmail.com

2.3 **Images of the Store**:

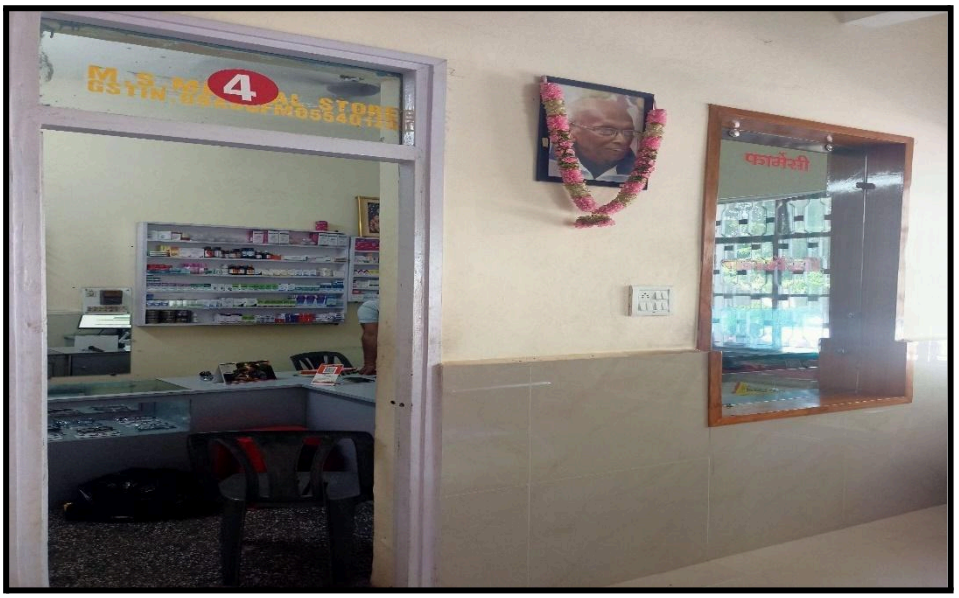


Fig 2.3.1 The storefront to authenticate its location.

Fig 2.3.2 The working environment of the store

2.4 **Video Submission:** Click Here

2.5 **Dataset:** Click Here

2.6 **Google Colab file for Analysis:** Click Here

3 Meta Data and Descriptive Analysis

3.1 Data Collection Overview/Meta Data:

Data Source: The data was collected from **Marg ERP software**, which the store uses to manage sales, inventory, and customer transactions.

Data Period: The data covers four months of sales transactions from **April 1, 2024**, to **August 5, 2024**, for the financial year **2024-2025**. This period is significant as it reflects the most recent sales data available for analysis, with the previous financial year's data locked and securely stored and cannot be collected.

Data Type: The dataset comprises sales invoices, along with details of the items sold, item type, discounts applied, and timestamps for each transaction.

Number of Records: The dataset consists of approximately **4500** sales invoices with **169** unique Products, each detailing one or more product sales within a given transaction.

Structure: The data contains the following columns:

Date: The date the transaction occurred (e.g., 01-04-2024).
Bill Number: A unique identifier for each sales transaction (e.g., MS000001).
Payment Method: Method of transaction (e.g., CASH).
Item Name: The name of the product sold (e.g., RHOOS GEL).
Item Type: Category or type of the item (e.g., GEL, DROP, TAB).
Item Package: Package unit size (e.g., 1*10, meaning 1 package with 10 tablets).
Item Code: SKU or product code (e.g., 23EH18P).
Quantity: The quantity of items sold in the transaction (e.g., 1.00).
Unit Cost: The unit cost price of an item to the business (e.g., 122.53).
Unit Price: The unit selling price before discounts (e.g., 155.00).
Total MRP: Maximum Retail Price of the item/items purchased (e.g., 155.00).
Discount: The discount applied to the item/items in the transaction (e.g., 4.65).
Net Amount: The total amount payable after discounts (e.g., 150.35).
Timestamp: The exact time the transaction was recorded (e.g., 2024-04-01 09:03:43).

Samples of Data Collected and Cleaned:

7	SALES STATEMENT FROM 01-04-2024 TO 05-08-2024							
8	BILL NO.	PARTY NAME	AMOUNT	DISCOUNT	NET AMT	TAX PAYABLE	DR/CR	NET AMOUNT
9	01-04-2024							
10	MS000001	CASH	155.00	4.65	150.35	0.00	0.00	150.00
11	RHOOS GEL	1 23EH18P 1	155.00	4.65	150.35	0.00		155.00
12								
13	MS000002	CASH	512.50	32.29	480.21	0.00	0.00	480.00
14	MIXEE LP	DROP 1 D03H018C 1	160.00	10.08	149.92	0.00		160.00
15	OLOCET KT	DROP 1 D02095 1	140.00	8.82	131.18	0.00		140.00
16	RHOOS DROP	1 D03G016A 1	130.00	8.19	121.81	0.00		130.00
17	FEXTRO	120*TAB 1*10 F28117 0:5	82.50	5.20	77.30	0.00		165.00
18								
19	MS000003	CASH	272.25	2.72	269.53	0.00	0.00	270.00
20	PCORT DROP	1 HCLE30032 1	32.25	0.32	31.93	0.00		32.25
21	LUBETEARS	DROP 1 BE240-28 1	120.00	1.20	118.80	0.00		120.00
22	MIXEE KT	DROP 1 DO3H020B 1	120.00	1.20	118.80	0.00		120.00
23								
24	MS000004	CASH	155.00	4.65	150.35	0.00	0.00	150.00
25	RHOOS GEL	1 23EH18P 1	155.00	4.65	150.35	0.00		155.00
26								
27	MS000005	CASH	130.00	10.40	119.60	0.00	0.00	120.00
28	RHOOS DROP	1 D03G016A 1	130.00	10.40	119.60	0.00		130.00
29								

Fig 3.1.1 Data before cleaning

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Date	Bill Number	Party Name	Item Name	Item Type	Item Package	Item Code	Quantity	Unit Cost	Unit Price	Total MRP	Discount	Net Amount	Timestamp
2	01-04-2024	MS000001	CASH	RHOOS GEL	GEL	1	23EH18P	1.00	122.53	155.00	155.00	4.65	150.35	2024-04-01 09:03:43
3	01-04-2024	MS000002	CASH	MIXEE LP	DROP	1	D02045	1.00	126.74	160.00	160.00	10.08	149.92	2024-04-01 10:46:58
4	01-04-2024	MS000002	CASH	OLOCET KT	DROP	1	D02095	1.00	103.74	140.00	140.00	8.82	131.18	2024-04-01 10:46:58
5	01-04-2024	MS000002	CASH	RHOOS DROP	DROP	1	D03G016A	1.00	95.09	130.00	130.00	8.19	121.81	2024-04-01 10:46:58
6	01-04-2024	MS000002	CASH	FEXTRO	TAB	1*10	F28117	5.00	12.92	16.50	82.50	5.20	77.30	2024-04-01 10:46:58
7	01-04-2024	MS000003	CASH	PCORT DROP	DROP	1	HCLE30032	1.00	25.22	32.25	32.25	0.32	31.93	2024-04-01 11:11:51
8	01-04-2024	MS000003	CASH	LUBETEARS	DROP	1	BE240-28	1.00	84.41	120.00	120.00	1.20	118.80	2024-04-01 11:11:51
9	01-04-2024	MS000003	CASH	MIXEE KT	DROP	1	DO3H020B	1.00	84.47	120.00	120.00	1.20	118.80	2024-04-01 11:11:51
10	01-04-2024	MS000004	CASH	RHOOS GEL	GEL	1	23EH18P	1.00	122.53	155.00	155.00	4.65	150.35	2024-04-01 11:21:00
11	01-04-2024	MS000005	CASH	RHOOS DROP	DROP	1	D03G016A	1.00	95.09	130.00	130.00	10.40	119.60	2024-04-01 11:31:08
12	01-04-2024	MS000006	CASH	RHOOS GEL	GEL	1	23EH18P	1.00	122.53	155.00	155.00	4.65	150.35	2024-04-01 11:36:57
13	01-04-2024	MS000007	CASH	MIXEE DROP	DROP	1	D02092	1.00	62.23	85.00	85.00	2.13	82.87	2024-04-01 11:47:20
14	01-04-2024	MS000007	CASH	RHOOS DROP	DROP	1	D03G016A	1.00	95.09	130.00	130.00	3.25	126.75	2024-04-01 11:47:20
15	01-04-2024	MS000008	CASH	PCORT DROP	DROP	1	HCLE30032	1.00	25.22	32.25	32.25	1.35	30.90	2024-04-01 12:07:07
16	01-04-2024	MS000008	CASH	LUBETEARS	DROP	1	BE240-28	1.00	84.41	120.00	120.00	5.04	114.96	2024-04-01 12:07:07
17	01-04-2024	MS000008	CASH	MOXSEF-KT	DROP	1	BE2312-85	1.00	92.01	130.00	130.00	5.46	124.54	2024-04-01 12:07:07
18	01-04-2024	MS000009	CASH	RHOOS GEL	GEL	1	23EH18P	1.00	122.53	155.00	155.00	4.65	150.35	2024-04-01 12:12:34
19	01-04-2024	MS000010	CASH	RHOOS DROP	DROP	1	D03G016A	1.00	95.09	130.00	130.00	9.10	120.90	2024-04-01 12:25:39
20	01-04-2024	MS000010	CASH	CHLOROZOX H	TAB	1	S138	1.00	71.71	100.00	100.00	7.00	93.00	2024-04-01 12:25:39
21	01-04-2024	MS000010	CASH	LCX 5MG TAB	TAB	1*10	TG23-03431	10.00	3.50	5.00	50.00	3.50	46.50	2024-04-01 12:25:39
22	01-04-2024	MS000011	CASH	RHOOS DROP	DROP	1	D03G016A	1.00	95.09	130.00	130.00	5.20	124.80	2024-04-01 12:36:41
23	01-04-2024	MS000011	CASH	MIXEE KT	DROP	1	DO3H020B	1.00	84.47	120.00	120.00	4.80	115.20	2024-04-01 12:36:41

Fig 3.1.2 Data after cleaning

3.2 Descriptive Analysis:

- **Quantity**:
  - Average of **2.69** units sold per transaction with an STD of **4.17**.
  - Most transactions involve 1 unit, with a few large orders (up to **35 units**).
- **Unit Cost**:
  - The average cost is **₹83.51**, ranging from **₹0.70 to ₹502.98** and STD **44.90**.
  - Costs vary widely, most items are priced between **₹60 and ₹120**.
- **Unit Price**:
  - The average price is **₹110.49**, ranging from **₹0.91 to ₹665** and STD **58.52**.
  - Significant markup from unit cost, with most items priced between **₹90 and ₹155**.
- **Discount**:
  - Average discount of **₹3.66**, with most between **₹1.2 and ₹5.1** and STD **3.33**.
  - The maximum discount is **₹65.84**, indicating promotions are applied but modest.
- **Net Amount**:
  - The average net amount is **₹124.76**, ranging from **₹4.97 to ₹882.20** and STD **53.04**.
  - Most transactions are close to the listed MRP with minor discounts applied.
- **Total MRP**:
  - The average total MRP is **₹128.42**, ranging from **₹4.59 to ₹885.74** and STD **54.75**.
- **Item Name**:
  - The dataset containing **169** unique items
- **Item Type**:
  - The dataset contains **11** categories of products such as **Gel**, **Ointment**, **Drop**, **Tablet**, etc
- **Key Insights**:
  - Small, frequent transactions dominate.
  - Prices and costs vary widely, but most are mid-range.
  - Discounts are common but generally modest.

	Quantity	Unit Cost	Unit Price	Total MRP	Discount	Net Amount
count	9823.000000	9823.000000	9823.000000	9823.000000	9823.000000	9823.000000
mean	2.691438	83.506219	110.488850	128.423046	3.658570	124.764476
min	1.000000	0.697973	0.906667	4.970000	0.000000	4.590000
25%	1.000000	62.278583	90.000000	100.000000	1.200000	99.000000
50%	1.000000	94.141905	130.000000	130.000000	3.250000	124.800000
75%	1.000000	118.815567	155.000000	155.000000	5.100000	150.350000
max	35.000000	502.979712	665.000000	885.740000	65.840000	882.200000
std	4.179974	44.909395	58.529933	54.759061	3.330362	53.046756

4 Detailed Explanation of Analysis Process/Method:

1. Hourly Sales Analysis:

This analysis aimed to identify the specific hours when the store experiences the highest sales. Understanding these patterns will optimize resource allocation, especially in staffing, ensuring that the appropriate number of staff members is available during peak hours. Additionally, it will assist the owner in determining the cost-effectiveness of hiring a full-time or part-time employee based on hourly revenue, maximising gains during busy periods.

First, the **Timestamp** column in the dataset was converted to a datetime format, which enabled me to extract the hour from the timestamps. This step was essential for grouping sales transactions by the time of day.

After extracting the **Hour** from each transaction's timestamp, I grouped the data by hour, with total sales (**Net Amount**) aggregating each hour using the Pivot Table (Hours in Rows and Net amount in Values). This grouping allowed me to understand the revenue generated at each hour of the day.

A line plot with shaded areas was created to show total sales for each hour, making it easy to observe peak and low sales periods throughout the day.

```python
import pandas as pd

# Loading the dataset
sales_data = pd.read_excel('BDM Data for Analysis.xlsx', sheet_name='Sales
Cleaned')
# Converting the 'Timestamp' column to datetime
sales_data['Timestamp'] = pd.to_datetime(sales_data['Timestamp'])

# Extracting hour from the 'Timestamp'
sales_data['Hour'] = sales_data['Timestamp'].dt.hour

# Grouping by 'Hour' to calculate total sales in each hour
hourly_sales = sales_data.groupby('Hour')['Net Amount'].sum().reset_index()
```

Fig 4.1.1 Python Script for Hourly Sales Analysis

2. Revenue Share Analysis:

The *revenue share analysis* helps identify the products that contribute the most to the store's revenue. By understanding which items bring in the highest earnings, the store can prioritise stocking and manage inventory needs in peak hours. To begin, I grouped the sales data by **Item Name** and calculated the total revenue for each product which was nothing but **Net Amount** in my case.

$$\textit{Revenue (for each item)} = \textit{Quantity Sold} \times \textit{Price per Unit} - \textit{Discount}$$

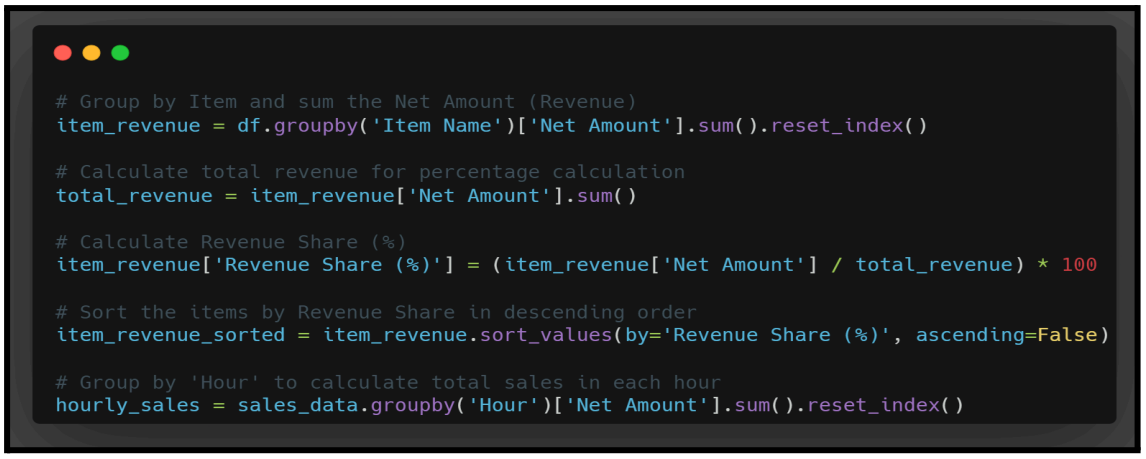
After that, I determined the percentage of total revenue each item contributed to assess its importance.

$$\textit{Revenue Share (\%)} = \frac{\textit{Item Revenue}}{\textit{Total Revenue}} \times 100$$

To simplify the findings, I focused on the top 20 items that generated the most revenue. The remaining products were combined into a single category labeled **"Others"** to avoid cluttering the results. This approach allowed for a clearer comparison of the store's key items versus the less significant ones. I visualized the data using a **Bar Plot**, where the top 20 products were displayed along with the **"Others"** group to showcase the revenue distribution.

In addition to the bar plot, a **Treemap** was also created so explain the revenue share per item more efficiently and interactively. The size and colour of each box in the treemap represented the proportion of revenue contributed by each product, with larger, darker boxes indicating higher revenue shares.

These visualizations can provide the store with a better understanding of which items are most valuable and can help in making more informed decisions about inventory and marketing strategies.



```python
# Group by Item and sum the Net Amount (Revenue)
item_revenue = df.groupby('Item Name')['Net Amount'].sum().reset_index()

# Calculate total revenue for percentage calculation
total_revenue = item_revenue['Net Amount'].sum()

# Calculate Revenue Share (%)
item_revenue['Revenue Share (%)'] = (item_revenue['Net Amount'] / total_revenue) * 100

# Sort the items by Revenue Share in descending order
item_revenue_sorted = item_revenue.sort_values(by='Revenue Share (%)', ascending=False)

# Group by 'Hour' to calculate total sales in each hour
hourly_sales = sales_data.groupby('Hour')['Net Amount'].sum().reset_index()
```

Fig 4.2.1 Python Script for Revenue Share Analysis

5 Results and Findings

The hourly sales analysis was conducted using both a line plot (Figure 5.1) and a heatmap to gain deeper insights into sales patterns across different hours of the day.

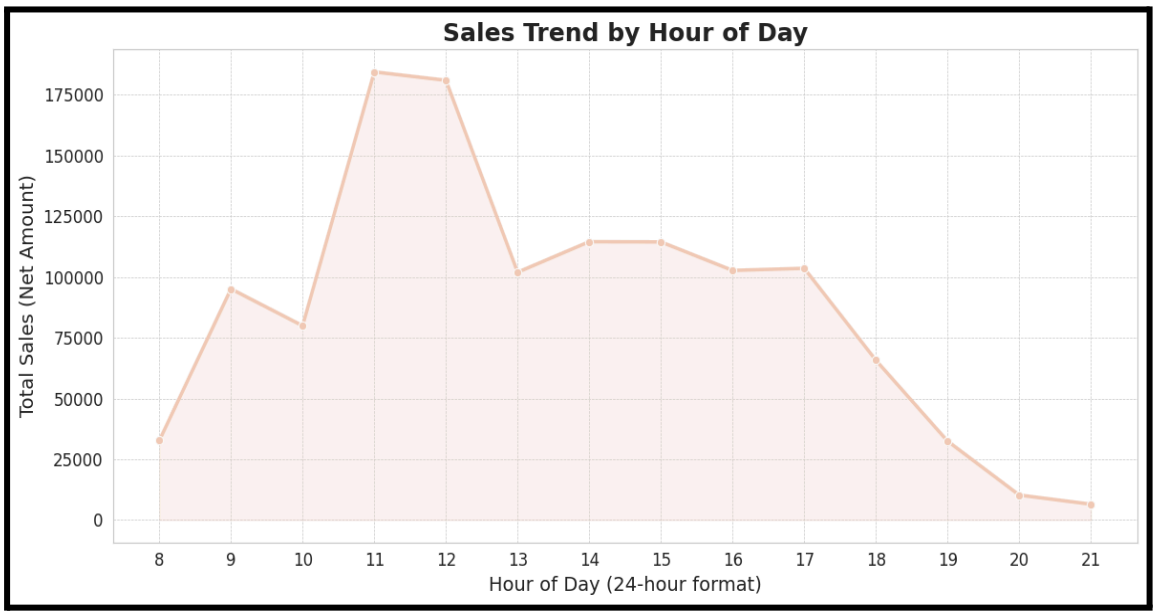


Fig 5.1 Sales Trend by Hour of Day

The line plot in **Fig 5.1** clearly illustrates a sales pattern that starts building from 8 AM, reaching its first peak around 11 AM. The busiest hours for the store are between **10 AM and 12 PM**, which correlates with feedback from the store's owner about nearby hospital doctors' visiting hours in the morning.

After the mid-morning peak, sales experience a sharp drop between **12 PM and 2 PM**, possibly reflecting a midday break or the conclusion of morning doctor appointments.

However, this dip is temporary, as sales pick up again after 2 PM, maintaining steady activity through the afternoon until around 6 PM. After this, sales gradually decline as the store approaches closing time.

This trend confirms that the store's busiest period spans from mid-morning to early afternoon, with two distinct peaks, one in the late morning and a second smaller one in the early evening, before narrowing off post 6 PM.

The heatmap in **Fig 5.2** adds more information to this analysis by showing sales activity across different days of the week along with the hourly patterns.

Monday, **Wednesday** and **Friday** stand out with significantly higher sales during mid-morning and early afternoon hours. This observation aligns with the line plot findings, indicating that the 10 AM to 12 PM window is indeed critical for sales across most days, especially on Mondays, Wednesdays and Fridays.

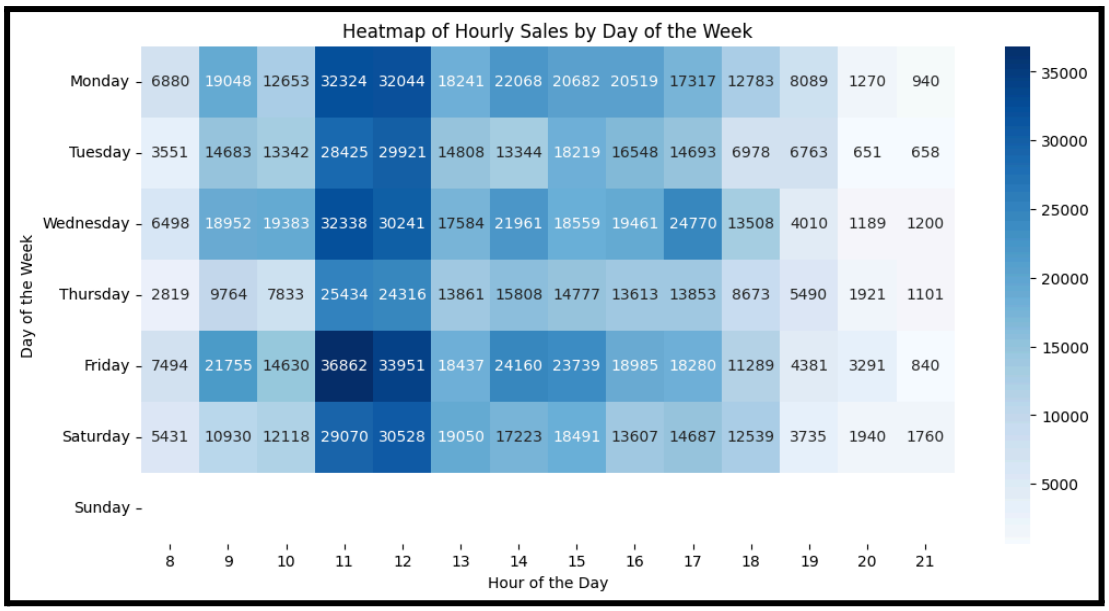


Fig 5.2 Heatmap of Hourly Sales by Day of the Week

In particular, **Fridays** stand out with the highest concentration of sales between **10 AM and 1 PM**, making it the most profitable day and time slot of the week. On the other hand, Tuesday and Thursday have comparatively lower sales activity during this same period, hinting that customer traffic and purchasing behaviour may vary by day, not just by the hour.

Moreover, the heatmap shows a sharp decline in sales after 8 PM on all days, suggesting limited customer engagement in late evening hours which is natural in this kind of business.

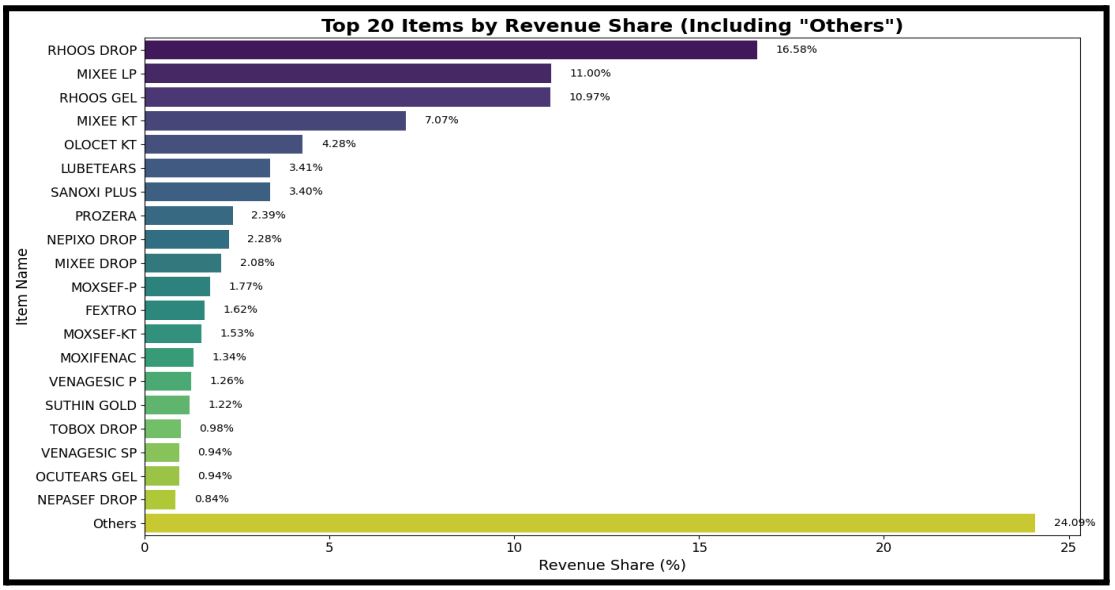


Fig 5.3 Revenue Share Percentage per item

The bar graph (Figure 5.3) reveals the top 20 items contributing to the overall revenue, highlighting a significant concentration of revenue among a few key products.

Notably, **RHOOS DROP** dominates the chart, contributing 16.58% of total revenue. **MIXEE LP** and **RHOOS GEL** follow closely, contributing 11.00% and 10.97% respectively, showing that these three products alone account for nearly 40% of the total revenue.

This suggests that optimizing stock levels and ensuring availability for these high-revenue items could greatly impact overall sales performance.

Additionally, items like **OLOCET KT** (4.28%) and **LUBETEARS** (3.41%) also play an essential role in revenue generation, though their contributions are smaller than the top three items.

The large share allocated to "Others" (24.09%) indicates a substantial number of products that contribute to the remaining revenue, implying that while a few key items dominate, a wide variety of products still make up a considerable portion of sales.

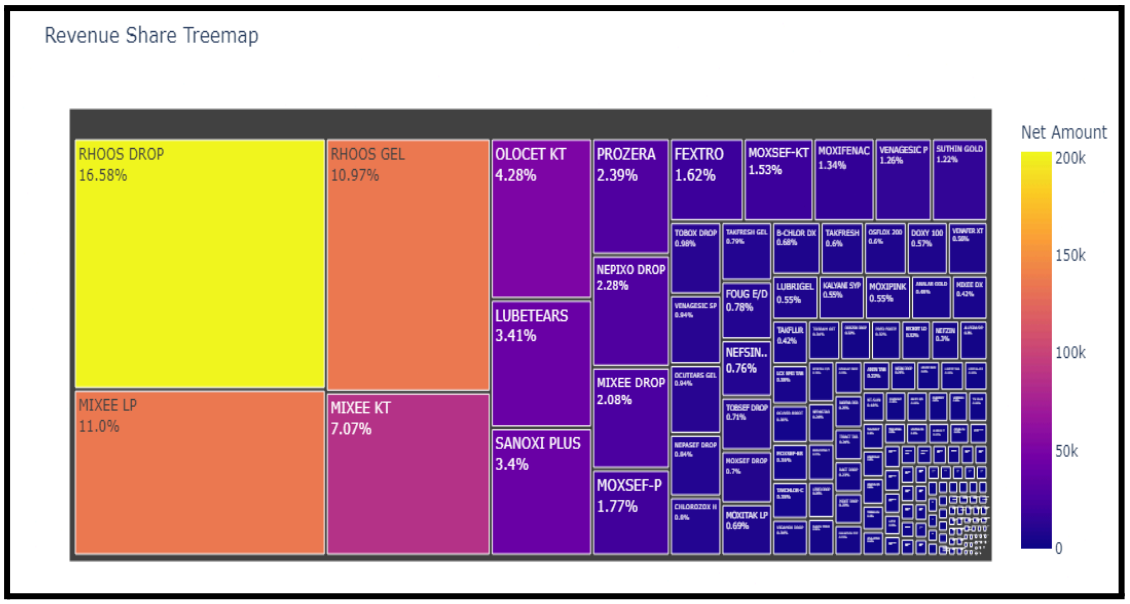


Fig 4.2.2 Revenue Share Treemap

The treemap in **Figure 5.4** offers a different perspective by combining both revenue share percentages and the absolute revenue per item (viewable in the interactive version of the treemap). Each product is represented by a block, where the size corresponds to its revenue share, and the colour gradient, ranging from bright yellow to deep purple, indicates the intensity of revenue contribution, with yellow signifying the highest revenue-generating items.

Similar to the bar graph, **RHOOS DROP**, **MIXEE LP**, and **RHOOS GEL** dominate the treemap, visually occupying the largest blocks and reinforcing the observation that a small number of products drive a significant portion of the store's revenue.

In the end, this is just the tip of the iceberg, I plan to do a more in-depth analysis to find out the way to solve the problems that the business is facing.

Note: *The Python script and the interactive Treemap can be accessed in this* ***Google Colab***