

INDIAN INSTITUTE OF TECHNOLOGY, MADRAS
BUSINESS DATA MANAGEMENT CAPSTONE PROJECT
FINAL SUBMISSION

ON

“Promotional offer analysis and stock forecasting at a
Supermarket”

Submitted by

YUKTI SETHI

Date of final report submission: 23-05-2023



TABLE OF CONTENT

<u>S.No.</u>	<u>Topic</u>	<u>Page No.</u>
1.	Executive Summary	1
2.	Problem Statement	1
3.	Detailed explanation of analysis process/method	1
4.	Results and Findings	1
5.	Interpretation of Results	17
6.	Recommendations	19

TABLE OF FIGURES

<u>S.No.</u>	<u>Figure</u>	<u>Page No.</u>
1.	Buying price regression analysis	6
2.	Pareto chart: SKU-wise discount pareto analysis	7
3.	Bar chart: Category-wise SKU distribution	8
4.	Category-wise discount pareto analysis	8
5.	Pareto chart: Category-wise volume pareto	9
6.	Line chart: Weekly discount percentage per item category.	10
7.	Area chart: Day-wise volume analysis for one month	11
8.	Scatter chart: Revenue vs Quantity sold for each category	12
9.	Pareto chart: Reorder point analysis for each category	13
10.	Clustered column chart: Revenue vs discount % for each category	14
11.	Scatter chart: Average discount % vs profit % for each category	15
12.	Clustered column + line chart: Quantity sold vs discount vs profit analysis	15
13.	Scatter chart: Change in profit % vs discount %	17

1. Executive Summary

This final report is a continuation of the mid-term report and describes the business data management capstone project for #####Super Store, a privately-owned grocery store located in Okhla, New Delhi. The project aims to address two business problems faced by the store: the challenge of tracking the effectiveness of discounts on sales and the issue of poor stock management resulting from inadequate sales planning. The report discusses the segmentation of SKUs into categories and an in-depth analysis of sales data to draw rational inferences that will enhance the store's productivity and profitability. By analysing the current baseline discount strategy, the primary focus is on exploring various combinations of three key features: discount, profit, and quantity sold. This analysis enables us to propose a restructuring of the promotional offers for products, specifically targeting low-selling, high-discount, and low-profit items. To address the stock management challenge, a reorder point analysis is conducted to forecast stock levels and prioritize stock commitments, especially for high-volume products during peak customer influx. Along with that, the relationship between revenue and quantity sold is also well-utilized. Based on these interpretations, the report concludes with specific recommendations to improve the store's operations and address the identified challenges.

Primary data of #####Super Store for April 2022: 

2. Problem Statement

- #####Super Store is struggling to track the effectiveness of its promotional offers and discounts, resulting in a decrease in profit margins as well as customer loyalty.
- #####Store is facing problems in the timely procurement of stock for the customer due to poor stock management, thereby leading to a loss of customers.

3. Detailed Explanation of Analysis Process/Method

1. Regression analysis

While imputing the missing values of the buying price of certain SKUs using the corresponding net amount given in the data, a linear predictive equation of the regression line is established by building a regression model using Machine Learning, and the performance is evaluated on the test set to ensure that the predictions are accurate enough to be used as missing value imputations. This linear relationship essentially speaks for the variation in the buying price with the selling price (with discount).

2. SKU-wise discount Pareto analysis

The contribution of each item's discount to the total discount is analysed to identify the items that contribute the most to the overall discounts. These items account for the key-focus areas for promotional offer analysis of discounts. The discount % = Total discount / Total revenue without discount (revenue generated if the discount would not have been levied on the respective items).

3. Grouping SKUs into categories and cleaning

There are about 5744 SKUs and thus, they need to be grouped into categories for better visualisations and insights. Since the name of the product does not include any information about the category it belongs to, the SKUs are manually put into 40 categories by using the sort and filter tools in Excel. The categories are made based on the functionality associated with the products being sold. Further analysis is majorly done on categories. While grouping the SKUs, it is found that certain SKUs like 'TOTAL CHK PARATHA KEEMA-110', have an incorrectly entered buying price value of 0.01 Rs. while having an MRP of 110 Rs. which is indeed incorrect. In these cases, the equation drawn by the linear regression model is used to replace the incorrect buying price values with the predicted buying price.

4. Category-wise discount Pareto analysis

The same discount analysis, earlier done on individual SKUs is now done on grouped categories and their contribution to the overall discount is analysed to identify the key-focus area in terms of the high-discount items and low-discount items and then proceed to check the impact of discount on sales of these 2 extremes. The pareto chart will be better visualised with 40 categories as opposed to 5744 unique SKUs.

5. Category-wise volume Pareto analysis

The average estimate of the quantity sold for each product is analysed to prioritise the stock commitment of these high-demand products. The 2 extreme categories of items found in step 4 are focused to notice the sales volume and the impact of discounts on them. This calls for identifying low-volume products in order to correlate them with discounts at the same time leading to lesser emphasis on stock prioritising unless they are high-revenue-generating premium products.

6. Weekly discount analysis

The discount percentage of all the categories is analysed weekly to know if there is dynamic pricing in discounts with respect to time. This is done to analyse the baseline discounting strategy of the store with respect to variation in sales demand over time so that it can be correlated with sales and profit variation over time.

7. Daily sales trend

The volume of sales on each day is studied using the moving average that takes the average of the quantity sold in the past 4 days in order to smooth out the fluctuations in sales data and identify spikes or dips in sales. This helps in prioritising the availability of items based on the daily trend and forecasting stock demand by projecting the trendline in future. This analysis will also help in the allocation of discounts at the right time when the sales are low and thus will attract more customers for those products identified to be low-selling.

8. Revenue vs Volume sold per category

The total revenue generated and total volume sold for each category will give insights into 4 types of categories:

a) High-revenue low-volume products

These products are premium products indicating the store to focus on targeting those customers willing to pay a premium price for the same.

b) High-revenue high-volume products

These products are considered essential assets for the store and since both the sales as well as revenue are high, the stock as well as customer-specific promotional offers need to be focused on this category.

c) Low-revenue low-volume products

These products are considered invaluable and may need to be evaluated for discontinuing or repackaging.

d) Low-revenue high-volume products.

These are the loss-leaders and the store needs to manage the inventory effectively, if not urgently to prevent loss in customer sales.

9. Reorder Point analysis for stock forecasting

Procuring the right product at the right time is a crucial step in the business. The metrics defined by reorder point analysis are essentially used to keep track of inventory and determine an optimum level of in-hand stock to prevent stockouts and loss in customer sales.

- a) Lead time: It is the time interval between the placement of the order with the vendor and the delivery of the order to the store. It can vary from product to product. However, the lead time of each SKU is not available in the dataset. As per the information obtained from the manager, an average lead time of 14 days has been considered for all the products for analysis.
- b) Average daily sales: For a particular SKU, the total quantity bought in April 2022 is divided by 30 to obtain the average daily demand for that product.
- c) Lead time demand: It is the volume of products that are expected to be bought by the customers during the lead time. $\text{Lead time demand} = \text{Lead time} \times \text{Average daily sales}$.
- d) Safety stock as the name suggests, is the amount of stock kept to consider unexpected hikes in sales and other anomalies. $\text{Safety stock} = \text{Maximum sale in a single day} \times \text{Maximum Lead time} - \text{Lead time demand}$. Since the lead time demand has already been taken on a worst-case average lead time of 14 days, the safety stock has been calculated as follows: $\text{Safety stock} = \text{Maximum daily sale} - \text{Average daily sale}$.

Finally, $\text{Reorder Point} = \text{Lead time demand} \times \text{Average daily sales} + \text{Safety stock}$

The reorder point for all the 5744 SKUs is calculated individually. After this the SKUs are grouped into categories and the category-wise average reorder point is calculated to forecast the stock of categories at the same time minimising the costs of holding excess inventory.

10. Revenue vs Discount Item-wise Analysis

The categories of items that have high discounts and low revenue need to be analysed. These items are the items which, most likely are resulting in a loss in profit margin. Items having high revenue and high discount indicate a good promotional offer analysis strategy. This strategy can be applied to other low-revenue generating items to maximise profitability. However, profit is a good measure to conclude suitably because there can be premium costly items that may increase revenue.

11. Profit vs Discount Analysis

The average percentage of discount per item is calculated by dividing the total discount by the total volume of sales. The profit percentage for each category sold in April 2022 is

calculated ($\text{Profit \%} = (\text{Revenue} - \text{Purchase Amount}) / \text{Purchase Amount} * 100$) and the impact of discount on the profit margin in the categories is carefully observed to make decisions. The products with high discounts and low profits imply that the discounts offered in that category potentially hurt the profitability of the store. This is done to maintain a healthy balance between sales and profitability.

12. Category-wise sales volume, profit %, and discount analysis

An increase in sales should be more than the decrease in profit on increasing the discount on a particular item. The category-wise analysis between these 3 variables together will give a clear direction on those SKUs having low sales and high discount and profit percentage to discontinue those discounts as they are not enticing enough to drive customer sales, thereby indicating high-priced products which are not meeting customer needs. Similarly, products with a higher discount percentage compared to the profit margin are analysed to notice the categories on which the business is sacrificing the profit margins to boost sales. While this may be a short-term strategy to increase sales, it may not be sustainable in the long run and the business may want to reconsider the discounting and pricing strategies for the same.

13. Change in profit % vs Discount % per category

Categories that show a significant decrease in profit percentage despite higher discount percentages may indicate that the discounts are eroding profitability. This suggests the need to evaluate the pricing and discount strategies within those categories to ensure sustainable profitability.

4. Results and Findings

1. Regression analysis



Figure 1 Buying price regression analysis

Buying Price = $0.80 \times \text{Net Amount} - 1.64$ is the linear equation drawn from the analysis and shows the linear dependence of the buying price on the selling price. The relationship is 92% accurate which is a good confidence interval to impute the missing buying price values. The mean squared error of training and test data match thereby indicating that the model is correctly fit. Impact on profit margins: A coefficient of 0.80 suggests that a significant portion of the net amount is allocated to the buying price. It may indirectly suggest that higher discounts may lead to lower profit margins. This is because the buying price, which includes costs and expenses associated with the products, is influenced by the net amount after discounts. As the net amount decreases due to higher discounts, the buying price may be relatively higher, potentially impacting profit margins. (Figure 1)

[Click here: regression analysis.](#)

2. SKU-wise discount Pareto analysis

The Pareto shows the percentage contribution of each SKU towards the total discount and it is found that the SKU 'LAFZ PERF POCKET MIX 18ML' has the highest percentage contribution of about 77% followed by 'GITS DAHI VADA 200G' with 75% contribution. On the other hand, there are about 2000 SKUs that do not contribute to the overall discount percentage. Since the chart given below contains a huge number of SKUs, it is wise to divide them into categories and the further analysis in categorised groups (Figure 2).

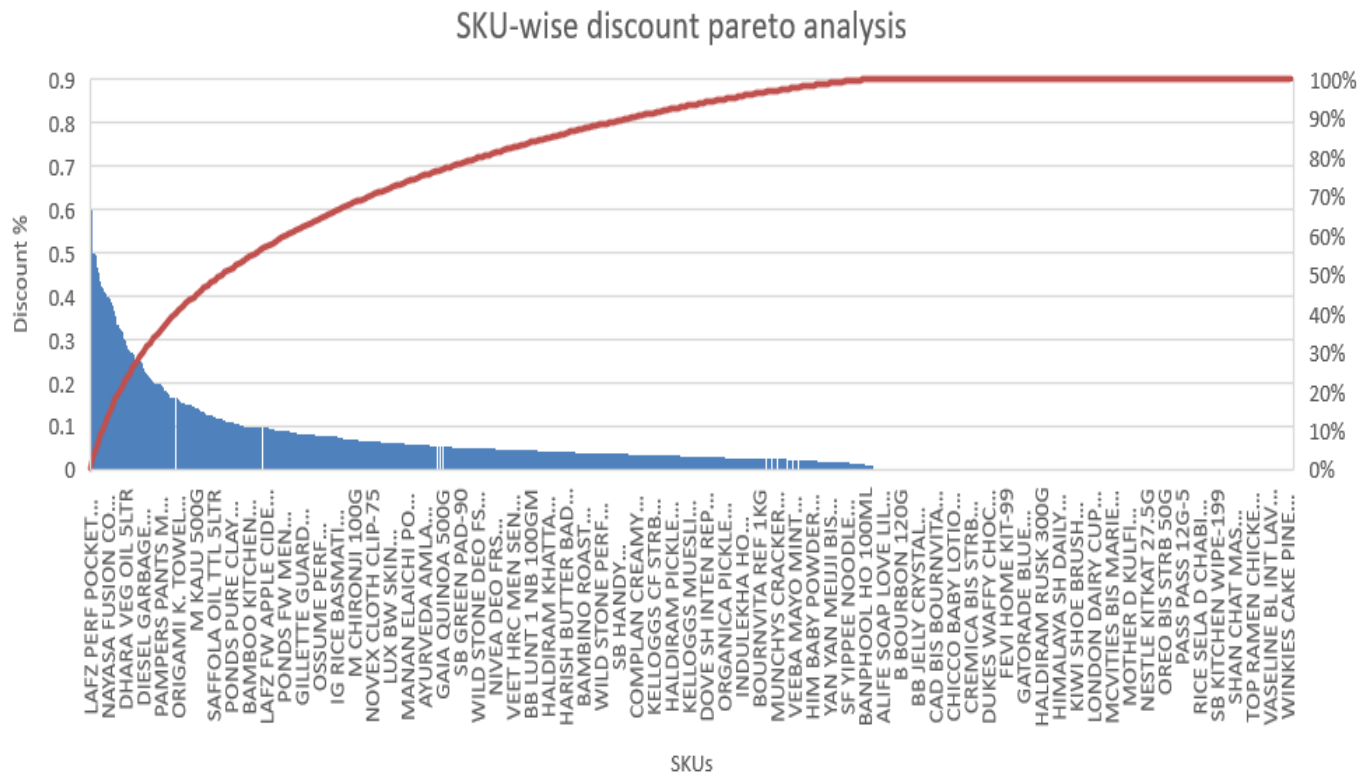


Figure 2 Pareto chart: SKU-wise discount pareto analysis

3. Category-wise SKU distribution

5744 SKUs present in the sales data are manually assigned categories and are grouped according to the function they perform. ‘KITHCN TOOLS MARBLE-80’ and ‘KITCHEN BREAD TONGS-84’, both the SKUs are broadly put in the ‘Utensils’ category, and so on. As seen from the chart below, the ‘Premixes&Masala’ category is associated with the highest number of SKUs while the category ‘Stationary’ is associated with the lowest count. However, these SKUs are limited to those present in one-month sales data and thus exclude those which might have not been bought at all during the month (Figure 3).

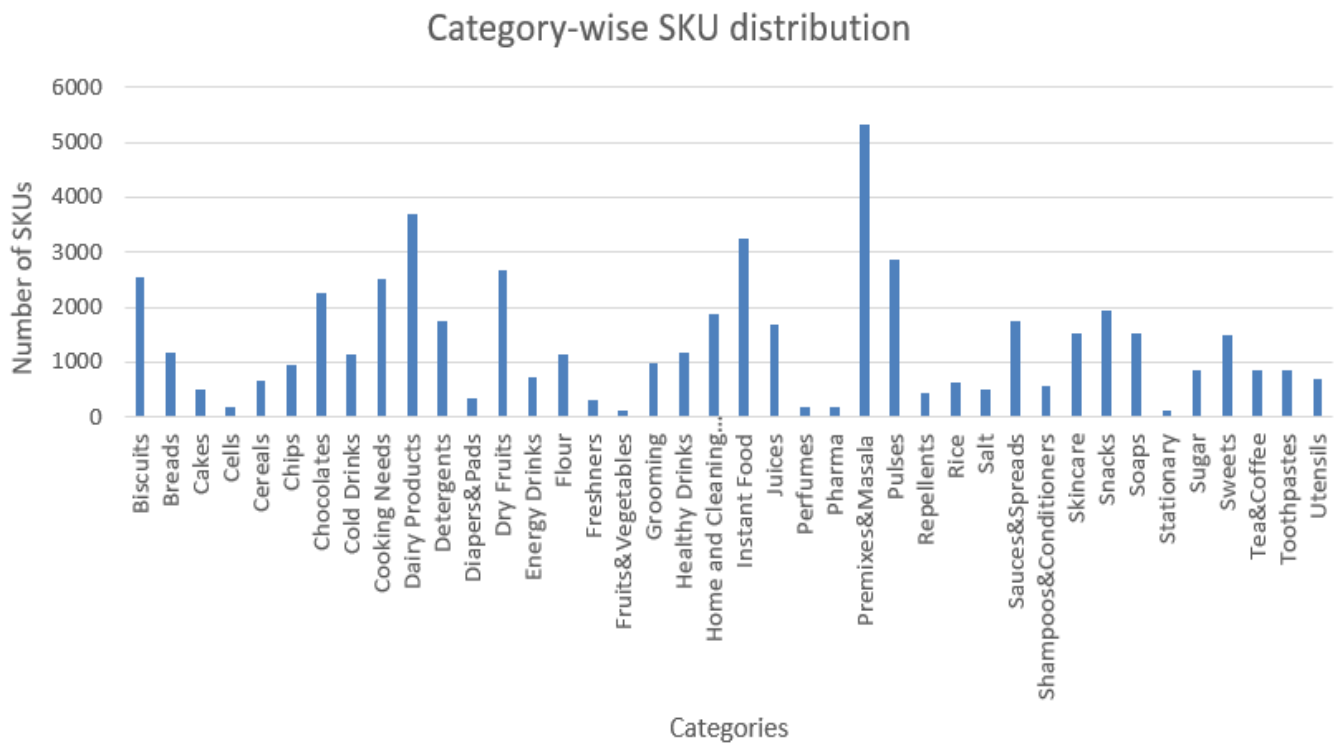


Figure 3 Bar chart: Category-wise SKU distribution

Note: Further analysis is done on categories instead of SKUs for better visualisation and inferences.

4. Category-wise discount Pareto analysis

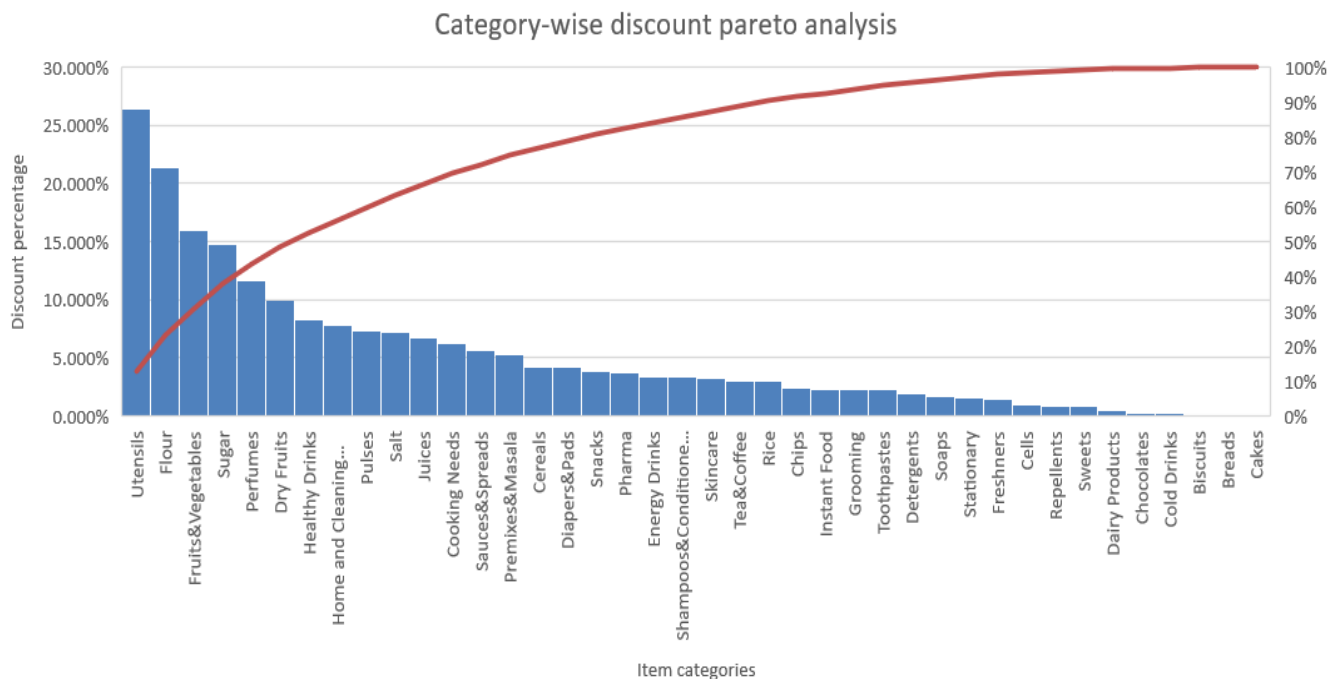


Figure 4 Category-wise discount Pareto analysis

Utensils and Flour category items are contributing to about 80% of the total discount offered by the store and are thus driving a significant portion of the overall discount given by the store. On the other hand, biscuits, breads, and cakes are not associated with any discount. The sales of these 2 extremes will be analysed further to track the effectiveness of discounts.

5. Category-wise volume Pareto analysis

Pulses, Dairy Products, Premixes&Masala, and Dry Fruits are high-volume products selling 10000, 7000, and 5000 units respectively, contributing to most of the sales volume. The products following the pareto rule 80:20 are the products that need to decrease the discounts at the weekends and during the start and end of the month to gain maximum profits.

Similarly, a safety stock of these high-selling products should be kept on hand to procure the customers. Stationary, Fruits and Vegetables, and Perfumes are the categories which, on the other hand, need to increase the marketing campaigns and promotional offers to sell more as they only have a sales volume of 116, 243, and 211 units, respectively. Coming to the high discount items: Utensils which have the highest discount contribute to only 10% of the total volume sold. Flour on the other hand has an increased contribution of about 45% which is a good sign. Cakes and Breads on the other hand have a sales volume contribution of 10-15% which is a low percentage (Figure 5).

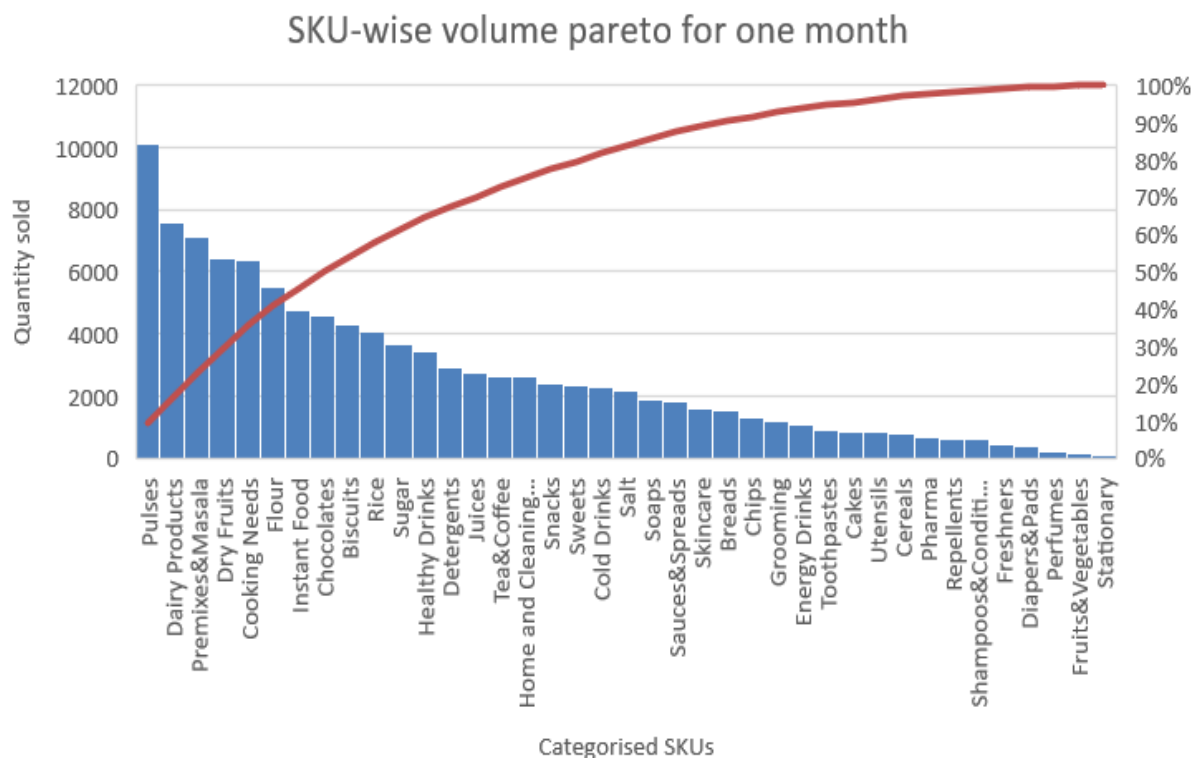


Figure 5 Pareto chart: Category-wise volume pareto

6. Weekly discount analysis

The weekly discount percentage (discount % = Total discount offered / Total revenue generated without discount) of various categories shows that the discount is almost the same in all the weeks because all the coloured lines coincide. Thus, the store has a constant baseline discounting strategy with respect to the time of the month. Week 3, however, shows a higher discount strategy on pulses which might be due to reasons like shelf-life. This calls for timely stock management of such products as well as an effective discounting strategy with respect to weekly sales. On the other hand, the products like 'Fruits and vegetables' and 'Sugar' are also associated with a high discount percentage. The impact of such high discounts needs to be analysed on the profit margins as it is most likely that these products are eating away at the profit margins (Figure 6).

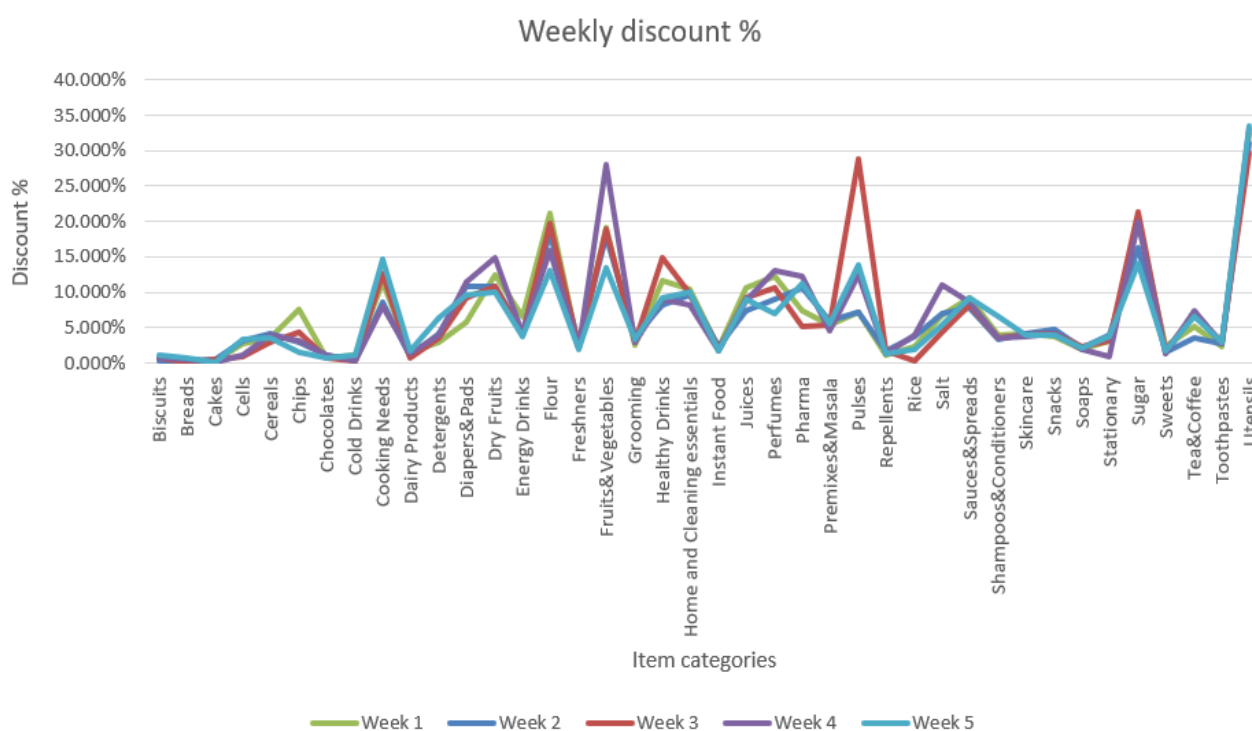


Figure 6 Line chart: Weekly discount percentage per item category

7. Daily Volume trend

The trend in local customer sales and moving average of quantity of products sold in the past 4 days shows that there is a significant increase in sales volume at the start and end of the month. The stock availability should be prioritised at this time or the store should at least have a huge safety stock at hand on these days which see a high influx of customers and increasing sales. (Figure 7)

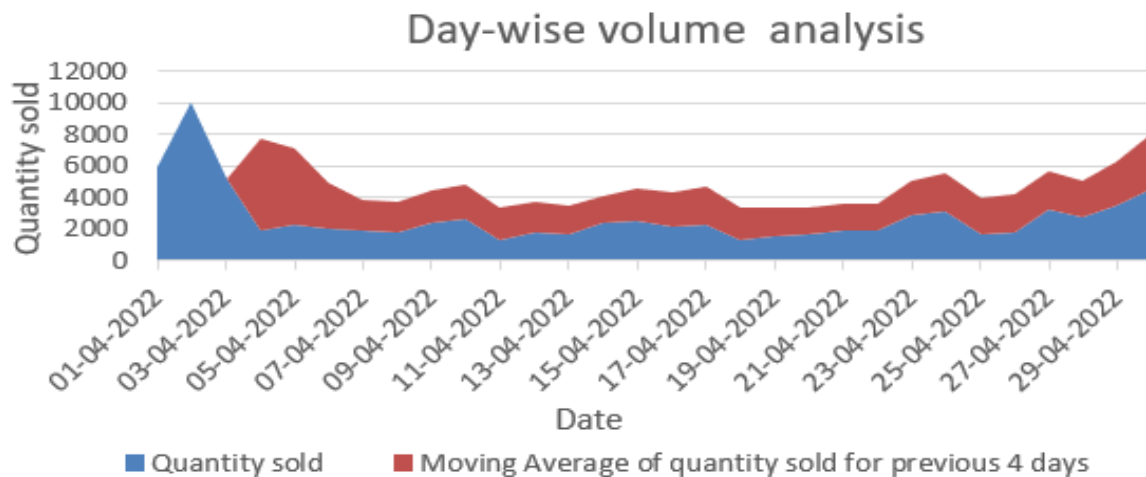


Figure 7 Area chart: Day-wise volume analysis for one month

8. Revenue vs Volume sold per category

With a revenue of Rs. 7 lakhs in April 2022, about 3382 units of products in the 'Dry Fruits' category are sold. This corresponds to high-revenue low-volume products which are premium products and require customer-specific promotional offers to target those customers willing to pay a premium price for the same. Similarly, 'Dairy Products' constitute high-revenue high-volume products which are critical for the store to prioritise stock commitment for these products. 'Stationary' is a low-revenue low-volume category proving to be invaluable and thus should be considered to either discontinue or increase the promotional offers for the same. 'Biscuits', on the other hand, are low-revenue high-volume 'loss-leader' categories that require a special focus on stock management. All the categories are divided into quadrants: I= Low-revenue-high-volume items, II=Low-revenue-low-volume items, III=High-revenue-low-volume items, IV= High-revenue-high-volume items (Figure 8)

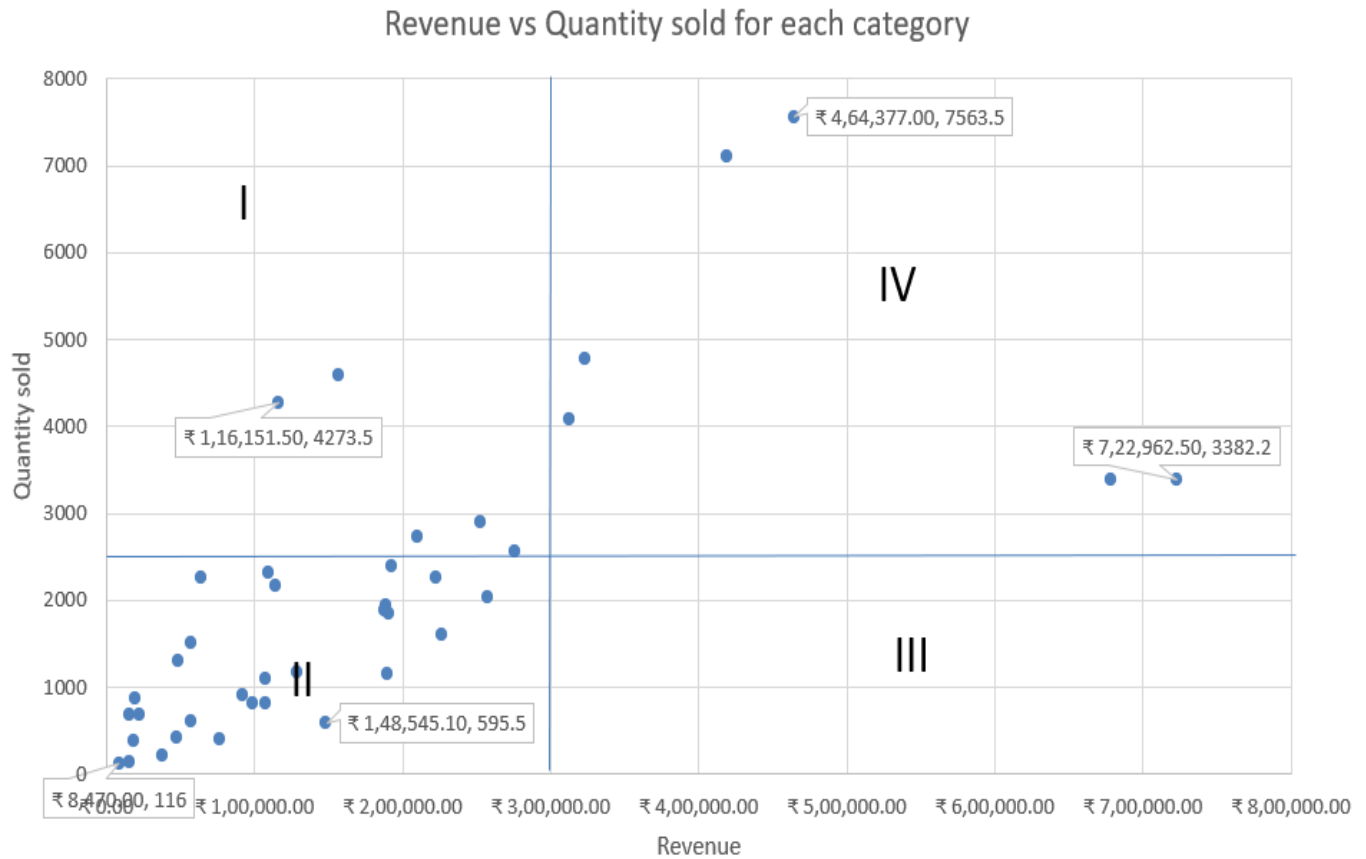


Figure 8 Scatter chart: Revenue vs Quantity sold for each category

9. Reorder Point Analysis

The Pareto shows the reorder points for each category. Sugar has the highest reorder point of 1700 units followed by Dairy Products (600 units) and Pulses (400 units). Thus, the current stock of Sugar must not be allowed to fall below 1700 before placing the new order. These categories of products require high inventory monitoring as they are critical in procuring customers on time. The reorder point also speaks for sales happening in 14 days (lead time). In other words, a maximum of 1700 units of sugar category SKUs are expected to be sold in 14 days keeping the outliers (like sudden NGO sale hikes observed earlier) separate. (Figure 9)

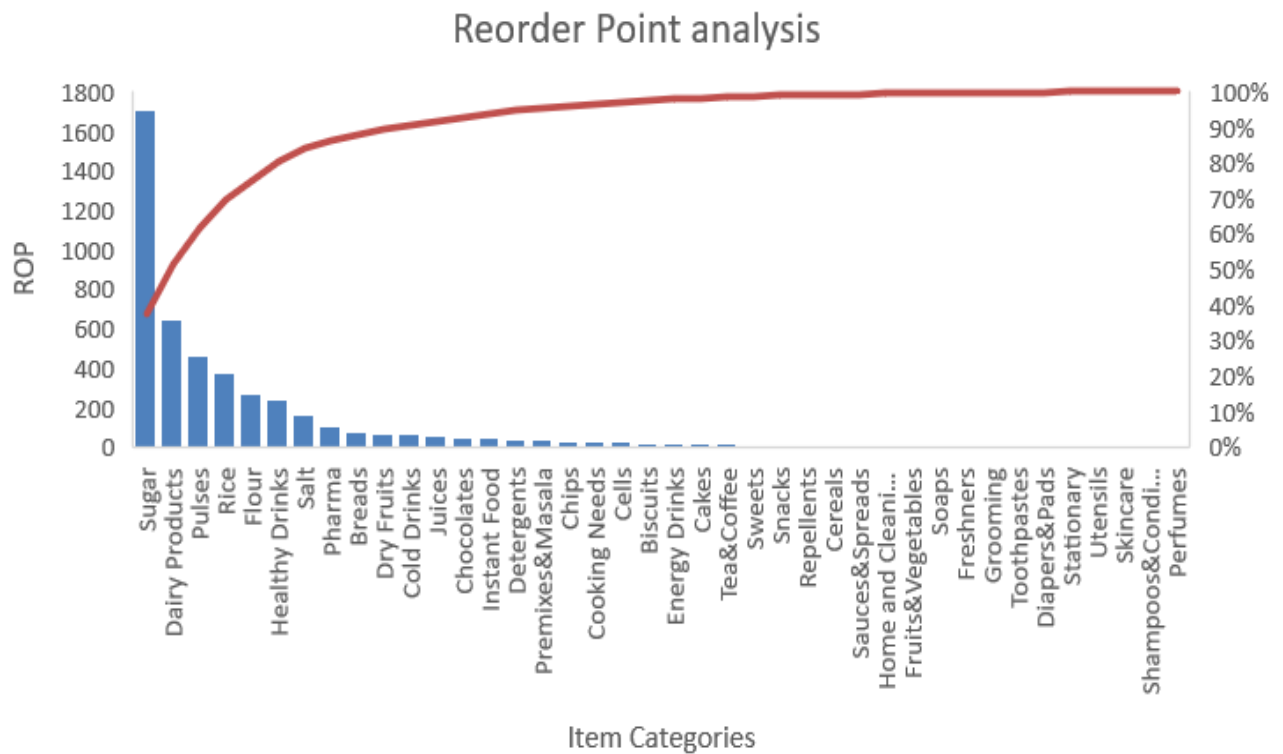


Figure 9 Pareto chart: Reorder point analysis for each category

10. Revenue vs Discount Item-wise Analysis

The item categories in the left region are the ones having high discounts on the average MRP of the SKUs present inside that category. However, categories like Fruits and Vegetables, despite having a large discount of about 20% are not making much revenue. Similarly, perfumes have a discount of 12% and is yet making a low revenue compared to other categories. The 'Utensils' category is having the highest discount and is yet not making much revenue. Thus, the pricing strategies for these products needs to be evaluated. Cakes, Breads, and cold drinks have a low discount of 0.32% and a low revenue of Rs. 19,245 in the entire month. This may indicate that the price point of these products is not aligned with the perceived value or customer expectations. (Figure 10)

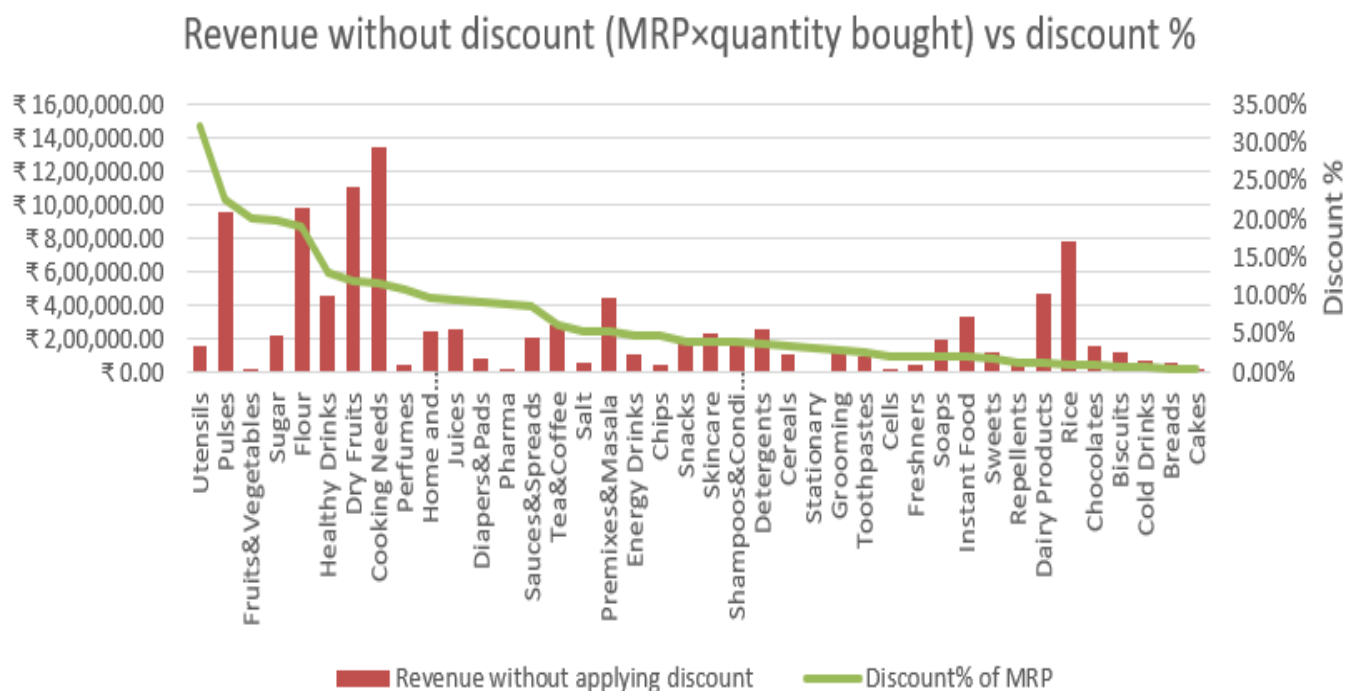


Figure 10 Clustered column chart: Revenue vs discount % for each category

11. Average discount% vs Profit% analysis

Products in the Pulses category, despite having a large discount of 29%, are earning a profit of just 8%. This indicates that levying the discounts to increase sales of pulses is eating away at the profit margin. It may be possible that the store wants to desperately increase sales to sell the products before they reach the shelf life. This, in turn, directs the stock monitoring mechanism to be focused on such products. Products in the ‘Pharma’ category have a high profit of 73.27% despite having a low discount of 9.93% and thus these have strong demand and should be priced the way they are. However, ‘Breads’ category is associated with low profit and low discounts implying that the cost of producing or acquiring these items is relatively high, and the revenue generated does not compensate for the expenses incurred. (Figure 11)

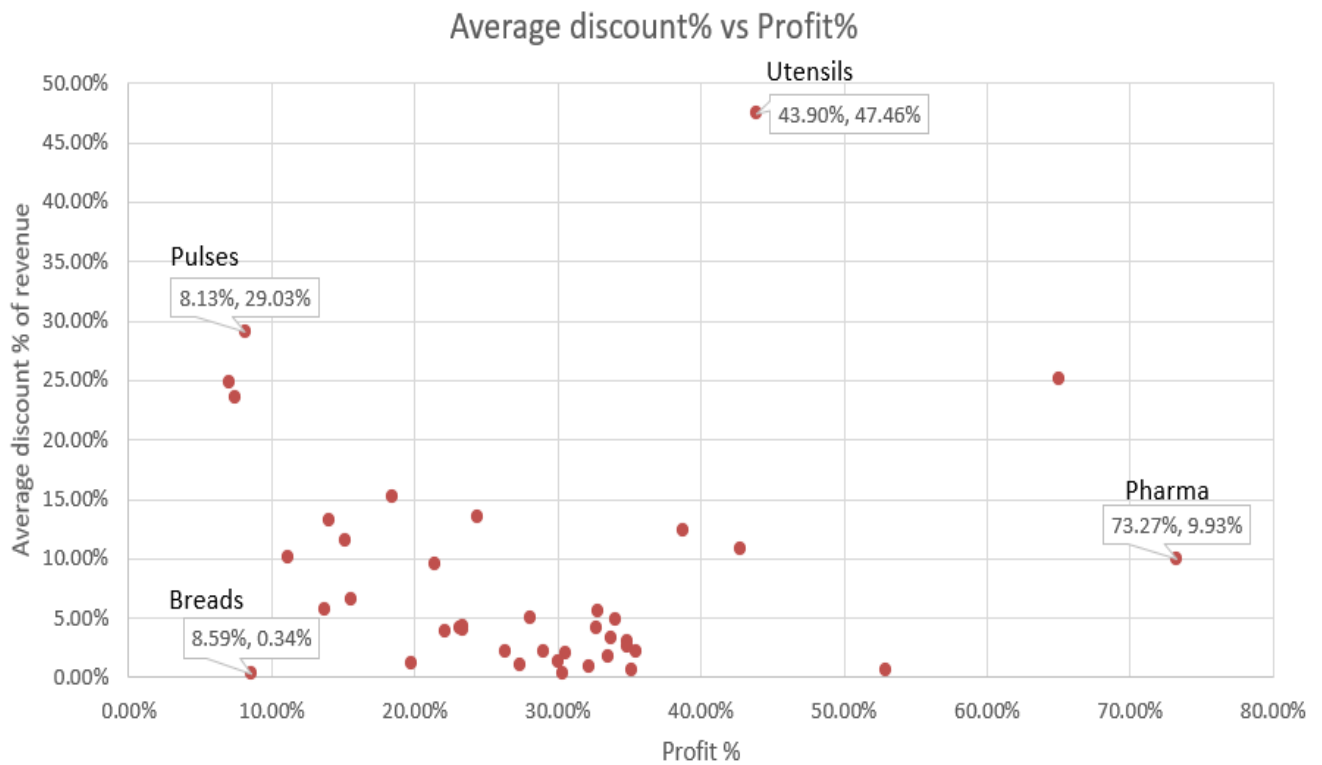


Figure 11 Scatter chart: Average discount % vs profit % for each category

12. Category-wise sales, discount% and profit% analysis

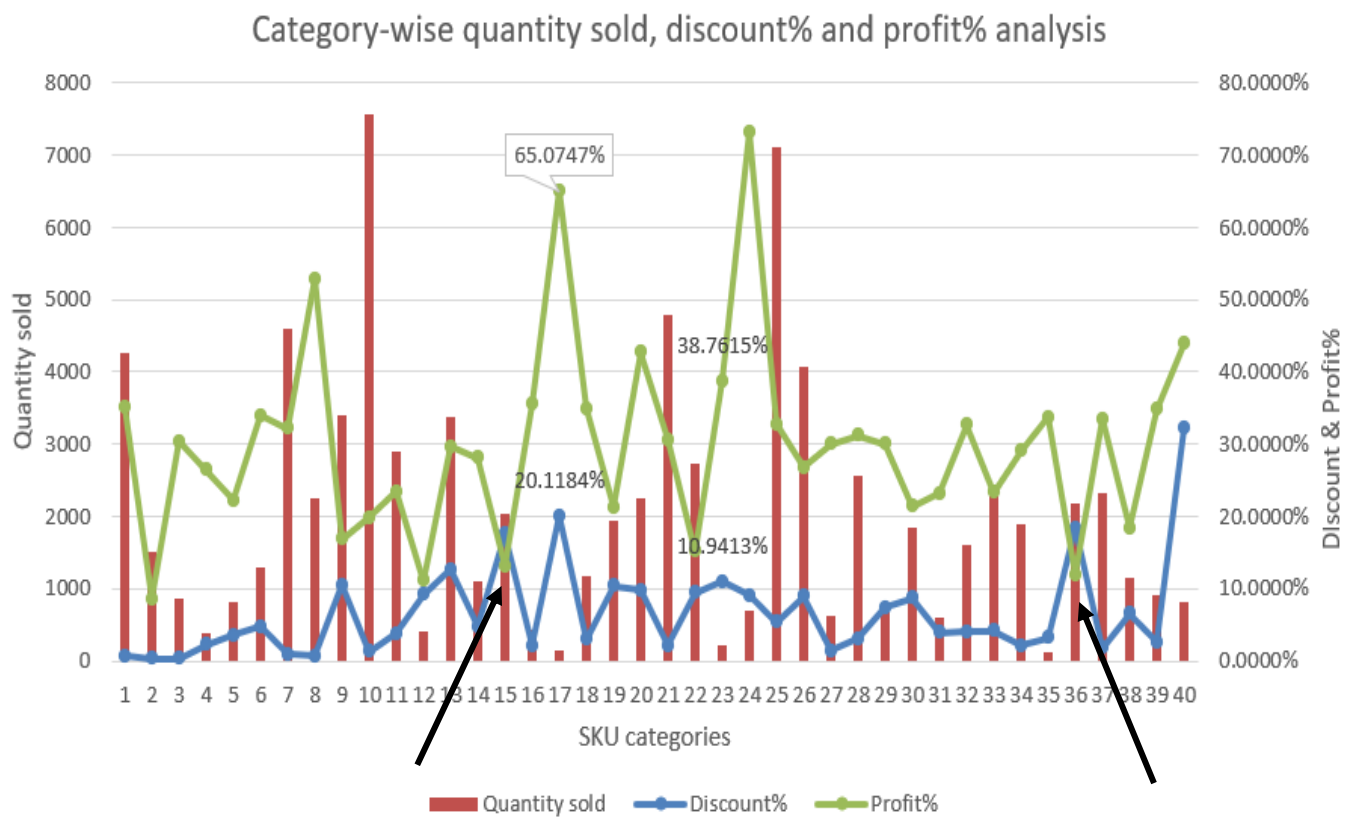


Figure 12 Clustered column + line chart: Quantity sold vs discount vs profit analysis

Flour & Stationary have a higher discount percentage (17%) compared to their profit margin (13%), indicating that even with a high discount, the profit margin remains narrow. This can be due to factors such as high production costs or intense competition within the market. Businesses may need to review their pricing strategy and evaluate their costs to determine if adjustments need to be made to maintain a sustainable profit margin.

‘Fruits & Vegetables’ and ‘Perfumes’ have high discounts of 20.11% and 10.94% respectively and high profits (65.87% and 38.7% respectively), yet have a very low sales volume of 143 and 211 units sold in one month. This indicates that the market demand for these products is relatively low or that the selling price, even with the discount, is still too high for consumers. Thus, the pricing strategy, sales channels, and marketing efforts need to be evaluated to increase product visibility and boost sales.

‘Flour’ and ‘Sugar’ have high discount percentages (17% and 18% respectively) (blue line marker higher compared to the relative profit bar (13% and 11% respectively)) but low sales, which indicates that the discounts offered on these products are eroding the profitability of the store and are not effective in increasing the sales relative to the decrease in profit suffered. Thus, re-evaluation of their inclusion in the product portfolio or restructuring their pricing and promotional offers is required. (Figure 12)

13.Change in profit % vs discount % per category

The slope of the trendline between discount% and change in profit% shows, for a discount of 8% on Pulses, the decrease in profit was 31%, thereby indicating a problem with the discounts. The categories below the trendline constitute those categories that have a high decrease in profit on applying a small discount. These categories are Pulses, Fruits and Vegetables, cereals, etc. The discounting strategy for these categories thus needs to be re-evaluated as the loss in profit margin with discount is significantly high. (Figure 13)



Figure 13 Scatter chart: Change in profit % vs discount %

5. Interpretation of Results

- The linear regression equation between the buying price and the net amount shows that buying price is 80% of the net amount accounting for an overall profit of 20% for the store.
- The present baseline discounting strategy is almost the same for all the weeks on average. This shows that it is static and does not cater to the wide variety of products and market demand.
- The discounts on Utensils (highest among all categories = 26.4%) are not leading to an increase in sales volume at the same time keeping a profit the same as the discount applied thereby neutralising the profit as well. This could indicate intense competition or market saturation in the utensil category. Customers might have alternative options available that offer similar or better value at lower prices, resulting in lower demand for utensils even with the discount. It is crucial to assess the competitive landscape and understand the factors influencing customer choices.

- Breads and cakes have no discount and contribute to only 10-15% of the total sales, despite being a day-to-day eating item.
- The sales trend shows that the sales are high on the weekends and during the start and end of the month. This provides incentives to decrease the discounts on high-selling products as well as prioritise the commitment to the stock of such products during this time.
- Dry Fruits are high-revenue low-volume products and thus fall in the premium category of products. These products need to be prioritised in stock but only to about 64 units.
- Dairy Products are high-revenue-high-volume products that are crucial for the store and these products should have a stock that is up-to-date within the confined limits of the shelf life as these products have less shelf life.
- 'Stationary' constitutes low-revenue low-volume products proving to be invaluable and thus should be considered to either discontinue or increase the promotional offers for the same.
- 'Biscuits', on the other hand, are low-revenue high-volume 'loss-leader' categories which require a special focus on stock management.
- Sugar has the highest reorder point of 1700. A maximum of 1700 units of sugar category SKUs are expected to be sold 14 days keeping the outliers (like sudden NGO sale hikes observed earlier) separate. Similarly, the reorder point for all the categories has been estimated taking the same lead time.
- Pulses, Dairy Products and Premixes, and Masalas are the high-selling categories of products accounting for about 60% of the total sales volume. Thus about 450 units of pulses, 639 units of dairy products, and 34 units of premixes and masalas need to be kept on hand when a new order is placed.
- About half of the items do not require a special focus on stock and thus, the store can work with a lesser stock of these items without hampering the revenue of the store. These items include perfumes, fresheners, soaps, repellents, etc.
- Flour, with a discount percentage of 17% but a profit of only 13% is eroding the profitability of the store and thus these discounts need to be discontinued and the competitive market in that area with respect to selling flour needs to be closely monitored. Similarly, sugar has an ineffective discounting strategy.

- Fruits and Vegetables have very low sales despite having a high profit of 65% and a high discount of 20%. This might indicate either a quality issue in such items or the selling price still not meeting the customers' expectations.

6. Recommendations

- The owner should follow the reorder points that have been estimated in this analysis to maintain stock of different products, keeping in mind the shelf life of the corresponding items.
- The owner should establish a website for the store. The attractive discounts appearing on the website will be more effective and available to the customers browsing through the store.
- Since most of the items being sold at a larger discount leading to a double decrease in profit belong to the following categories: Pulses, Fruits and Vegetables, Cereals, and Dairy Products, it is most likely that there might be a quality issue where such edible products are stored or the store is not keeping a high stock of such products leading to the shelf life coming very close to the present date that the store is forced to sell these products at a higher discount and thus desperately willing to sacrifice the profit margins. Thus, the stock of such products needs to be kept no more than what has been estimated by the reorder point, and the quality checking should be strict.
- On interacting with the owner, it is found that there is no provision of any membership for the customers. Adding a membership option with a small fee and giving certain special discounts on high-selling day-to-day items like Pulses, Dairy Products, and Premixes and Masalas to those enrolled as members can lead to an increase in profit.
- The free home delivery given to the customers by the store can also be restricted to those enrolled as a member. This will cut the costs of delivery.
- On visiting the store, it is found that a lot of labourers are present and keeping an eye on the customers. The small store has about 8 labourers who are responsible for managing the display of products. The labour can be reduced to cut these variable costs incurred by the store to increase net gain.
- Products like Bread and cakes should be given special clustered discounted offers in which they can be grouped and sold at a discount percentage. Since they are having a high average profit margin of 19 %, some amount can be given as a discount

percentage to increase the sales and increase the overall revenue, at the same time keeping a reasonable profit.

Conclusion

This project, with the above findings justified by an in-depth analysis, has aimed to solve the problems faced by #####Superstore. The discounts on various products have been effectively tracked to discontinue the same on some of them while arising the need for putting new discounts on some other products. Similarly, the stock of various items has been objectively calculated using reorder point analysis with the limitation of not taking the shelf life into account and keeping the lead time the same for all sorts of products. Business data management is a critical discipline that empowers organizations to harness the full potential of their data assets. By treating data as a valuable resource and implementing robust data management practices, businesses can unlock new opportunities, enhance their competitiveness, and thrive in an increasingly data-centric business landscape.