Optimizing sales, analyzing the shopping trends and increasing Revenue of a small-scaled super market

Mid Term Report for the BDM Capstone Project

Submitted by

K VENKATA SAI HARINI

(23F1001704)

23F1001704@ds.study.iitm.ac.in

IITM Online BS Degree Program

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

DECLARATION

I, **K VENKATA SAI HARINI (23F1001704),** student of IIT Madra BS Degree Program, pursuing Diploma in Data Science, hereby declare that the **Mid Term Report** entitled "**Optimizing sales, analyzing shopping trends and increasing Revenue of a small-scaled super market**" is an original work carried out by me. I extend my gratitude to **Vaishnavi Super market,** for providing me the necessary resources that enabled me to work smoothly on this project.

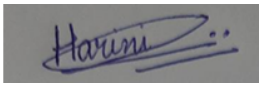
I hereby assert that the data presented and assessed in this report is genuine and precise to utmost extent of my knowledge and capabilities. The data has been gathered from **Primary source** and carefully analyzed to assure its reliability.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of data contained in this report.

I understand that execution of this project is intended for individual completion and is not to be undertaken collectively.

Also, in the event that plagiarism is detected in the report at any stage of the project completion, I am fully aware and prepared to accept disciplinary measures imposed by relevant authority.

I understand that all recommendations made in this project are within the context of the academic project taken towards course fulfilment in the BS Degree Program offered by IIT Madras. The institute does not endorse any of the claims or comments.



Signature of Candidate

K VENKATA SAI HARINI

16/03/2025

INDEX

1. Executive Summary and Title

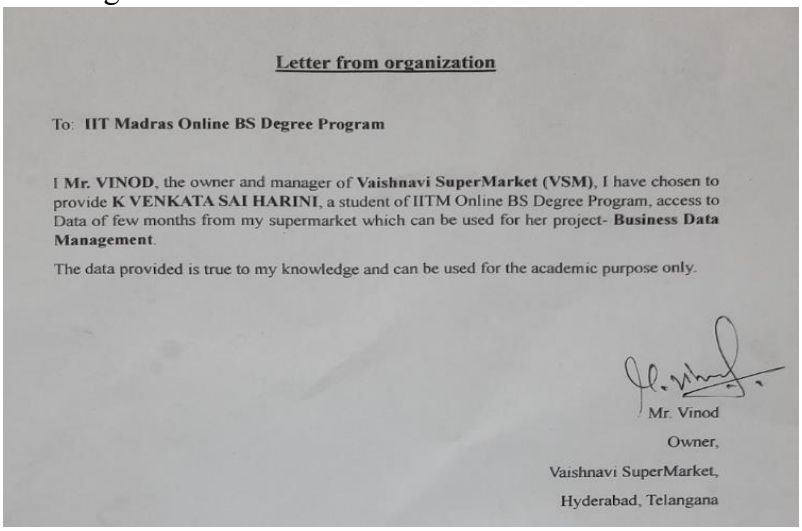
Vaishnavi SuperMarket(VSM) is a small-scaled supermarket located in Malkajgiri district, Hyderabad. VSM is a B2C business, selling groceries and household necessities. The supermarket faced challenges in meeting sales targets, optimizing revenue, and understanding customer purchasing patterns. To address these issues, a data driven approach was adopted to analyze sales trends, customer preferences and pricing strategies.

The analysis was conducted using Microsoft Excel for data structuring and visualization, and Google Colab for advanced findings. Key methodologies included Pivot Tables, and Charts & Graphs, which provided insights into revenue distribution, customer behavior, and seasonal sales trends. Quantitative analysis techniques like statistical analysis and qualitative analysis like Narrative and Thematic analysis have been used. Additionally, data visualization through charts and graphs supported quantitative insights, while qualitative interpretations of customer behavior and shopping patterns helped refine strategic decisions. Comparative analyses were performed on profit margins, payment modes, and category-wise sales performance to optimize pricing strategies. Additionally, sales trends were assessed both for festival and non-festival periods to measure their relative contributions to total revenue.

Findings revealed that while non-festival periods accounted for high sales volume, festival seasons contributed significantly to revenue. This is due to discounts, offers available in nearby larger supermarkets. Shopping trends indicated UPI as most preferred payment mode, and returning customers contribute to majority of total sales, highlighting the existence of good retention strategy. December and January saw a higher sales volumes compared to other months, depicting the impact of festive season. This project provides actionable insights into sales trends, customer behavior and revenue optimization.

2. Proof of Originality

2.1 Letter from the organization



Letter from organization

To: **IIT Madras Online BS Degree Program**

I **Mr. VINOD**, the owner and manager of **Vaishnavi SuperMarket (VSM)**, I have chosen to provide **K VENKATA SAI HARINI**, a student of IITM Online BS Degree Program, access to Data of few months from my supermarket which can be used for her project- **Business Data Management**.

The data provided is true to my knowledge and can be used for the academic purpose only.

Mr. Vinod
Owner,
Vaishnavi SuperMarket,
Hyderabad, Telangana

Fig 2.1: Letter from Organization

2.2 Photograph of the organization

Figure 2.2 Photo of Organization

2.3 Photograph of interaction with the owner

Fig 2.3: Interaction with owner

2.4 Additional Proof of Work (Photos and Videos): Google Drive

3. Metadata and Descriptive Statistics

3.1 Metadata

Datasheet Link: click here

Colab Link: click here

Summary:

A. **Data Collection:** By periodic visits to VSM, the day wise sales of VSM has been collected in the form photographs and drafted in the form of excel sheets for a duration of 4 months i.e from 01/12/2024 to 31/03/2025, to analyse the shopping trends and observe the sales in order to optimize revenue.

B. **Content of Data:** The data contents are as follows:-

1. **Date:** Day of sale in format: DD-MM-YYYY.
2. **Invoice ID:** Unique ID for transaction. Using chronological invoice numbering, the id is in the form of "YYYY" + "MM" + "XXX", which is a 3-digit number starts with 001, so on which resets after every month.
3. **CustomerID:** This is a unique identity given to each customer. Overall 300 customers chosen for this dataset. Format: "VSM" + "XXXX", which ranges from 0001 to 0300.
4. **Product Name:** Name of product, along with brand and qty. There are a total of 47 items chosen for analyses.
5. **Product Category:** All the above products are categorized into 10 categories.

6. **Quantity Sold:** Number of goods that has been sold.
7. **Purchase Price per Unit:** Price at which owner bought products from the vendors.
8. **Selling Price per Unit:** Price at which a particular product is sold to the customer.
9. **Total Revenue:** It is the product of Qty sold and the SP (per unit).
10. **Profit Margin:** It is the extra amount earned on a product. Formula: (SP-CP)/ SP.
11. **Customer Type:** [Returning, New]. If a customer visits the store more than once, only then customer type is 'returning', else 'new' customer.
12. **Payment Mode:** This is the transaction type which is: [Card, UPI, Cash].
13. **Festival Indicator:** If the 'Date' is nearing a festival (on or before 2days), then 'Yes' else 'No'.

The size of dataset is (1663,13). This is just a sample dataset, with four transaction per day. And each transaction contains number of products ranging in (2,5). In order to avoid multiple products in same row, a transaction is extended to further rows, where Invoice ID indicates a unique transaction.

C. **Sample of Data:** The dataset used in analysis is recorded for 4-months, and below is first 30rows of the dataset.

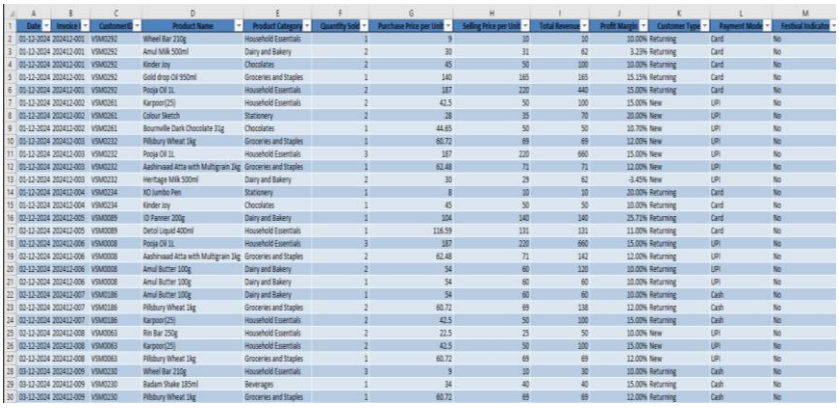


Fig 3.1: Sample Data

D. **Description of Data:** The primary objective of this dataset is to analyze shopping trends, revenue and sales patterns.

SNo.	Column Name	Units in	Data Type
1.	Date	-	Date
2.	Invoice ID	-	Alpha-Numeric
3.	CustomerID	-	Alpha-Numeric
4.	Product Name	-	String
5.	Product Category	-	String
6.	Quantity Sold	units	Numeric
7.	Purchase Price per Unit	₹	Numeric
8.	Selling Price per Unit	₹	Numeric
9.	Total Revenue	₹	Numeric
10.	Profit Margin	%	Numeric
11.	Customer Type	-	String
12.	Payment Mode	-	String
13.	Festival Indicator	-	String

3.2 Descriptive Statistics

The entire dataset is selected and performed Data Analysis-descriptive statistics. It summarizes and describes key characteristics of dataset using measures like- mean, median, std dev, kurtosis and skewness. By using this, we understand data distribution for meaningful business insights. This will help us in strategic decision-making by identifying high-revenue category, customer shopping patterns and impact of external factors like festivals.

a) <u>Key take-aways of below table:</u>

Mean Total revenue is ₹189.04, with high standard deviation of ₹185.45, indicating significant variation in revenue transaction. The profit margin averages is around 11.7%, with minimal variability (std dev= 5.98%), showing consistency in profitability. The skewness values of revenue and profit are positive, meaning a right-skewed distribution. The range of revenue is very large i.e ₹1135, indicating a mix of large and small transactions. Overall, dataset reflects high variability in sales, but a stable profit margin.

Stat Measure	Quantity Sold	Purchase Price per Unit	Selling Price per Unit	Total Revenue	Profit Margin
Mean	1.984957882	83.6704272	95.32912154	189.0385078	0.117003503
Standard Error	0.024199377	1.586830969	1.810019441	4.549034927	0.001467126
Median	2	62	69	131	0.12
Mode	1	42.5	50	100	0.1
Standard Deviation	0.986551333	64.69134451	73.7902105	185.453392	0.059811271
Sample Variance	0.973283532	4184.970054	5444.995166	34392.96059	0.003577388
Kurtosis	-0.727825902	0.119312288	-0.175529116	3.515638183	0.628117586
Skewness	0.625223453	0.947829004	0.874702759	1.775365126	-0.691674561
Range	3	271	280	1135	0.291625616
Minimum	1	4	5	5	-0.034482759
Maximum	4	275	285	1140	0.257142857
Sum	3299	139060.25	158437	314182	194.4598228
Count	1662	1662	1662	1662	1662
Largest(2)	4	275	285	1140	0.257142857
Smallest(2)	1	4	5	5	-0.034482759
	1	1	1	1	1

Fig 3.2: Descriptive Statistics

b) **<u>Descriptive Stats using Pivot Table:</u>** The below descriptive stats is plotted for two columns- Qty Sold(with slider- Product Category) and Revenue(with slider- Payment Mode). The first one suggests- the average qty sold per transaction is stable across months, with march having slightly higher sales. The std dev of nearly 0.99 across all months suggest minor variability in qty sold. The second table suggests- total revenue fluctuates across months, with Jan having highest revenue ₹85756 and February having the lowest ₹69760, mostly due to the festival-season. The high std dev of aprox. ₹185 suggest significant variation in transaction values, indicating both small and huge purchases taking place.

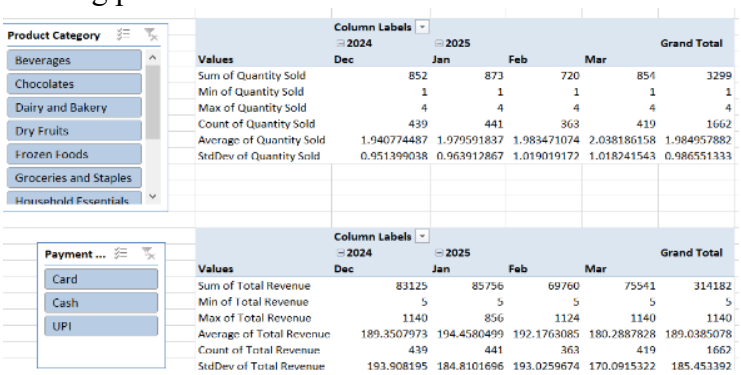


Fig 3.3: Descriptive Stats

c) **From Gcolab:** The below is df.describe() function from the gcolab file.

	Quantity Sold	Purchase Price per Unit	Selling Price per Unit	Total Revenue
count	1662.000000	1662.000000	1662.000000	1662.000000
mean	1.984958	83.670427	95.329122	189.038508
std	0.986551	64.691345	73.790211	185.453392
min	1.000000	4.000000	5.000000	5.000000
25%	1.000000	30.000000	35.000000	50.000000
50%	2.000000	62.000000	69.000000	131.000000
75%	3.000000	126.650000	149.000000	261.500000
max	4.000000	275.000000	285.000000	1140.000000

Fig 3.4: Describe() method

4. Detail explanation of Analysis Process/Method

a) Data Collection:

I collected Sales data from the owner on a regular basis and the final data collected is from December 1, 2024 to March 31, 2024. The owner was consented for the data and project's objectives and methodology, both personally and in written. Authorization for the same had been taken.

Although the data was stored in a software system, there was no export or download functionality. As a workaround, pictures of transaction records were taken daily and manually converted into Excel format. This data then underwent a thorough cleansing and organization process to ensure consistency and usability for analysis. Once structured, descriptive statistics were generated to understand key measures such as mean, median, standard deviation, skewness, and trends within the dataset.

b) Analysis:

To derive meaningful insights, various analyses were performed using Pivot tables and Charts in Excel. The following analysis were conducted-

1. Revenue Contribution Analysis using Pareto Chart: Identified top 20% product categories contributing 80% of revenue using the Pareto principle.
2. Category-wise Revenue and Quantity Sold Analysis: each category's contribution to the total revenue and sales volume was evaluated, which assessed the categories that were driving higher sales.
3. Purchase price and Sales price Comparison: Analysed the Purchase Price and Selling Price (per units') values of each Product which helped to determine products with higher markup and profitability.
4. Category-wise Purchase & Sales Trends: Identified the categories that had higher sales volumes and revenue contributions.
5. Month-wise Revenue Analysis: Growth & Decline patterns of Revenue across months, to look how efficiently revenue translated into sales.
6. Month-wise Quantity Sold Analysis: To check an overview of sales volume fluctuating over period of four months to detect seasonal trend.
7. Monthly Expense tracking: analysed fixed expenses such as electricity, employee salaries and operational costs to assess cost effieciency.
8. Profit Vs Revenue Analysis: compared revenue and profit (per unit) on a month-wise and category-wise basis.
9. Customer Behaviour Analysis: Studied various customer patterns based on-customer type, payment mode, and influential factors like festival season.

10. Shopping Trend Analysis: Performed basic analysis like- Product Preference analysis, monthly Ice-cream sales trends, Festival impact on specific product categories. Used slider-based month-wise and product-wise comparisons.

The above analysis using Pivot tables and charts, provides valuable business insights on Sales Trends, product performance, customer behaviour and profitability. Addition features like sliders and conditional formatting have been used to the above Pivot Tables in order to clearly distinguish and easily draw insights. Further analysis is currently being carried out on gcolab to explore advanced ML models, trend forecasting and deeper customer segmentation. This will enable a more data-driven decision-making.

Few Formulas:

A) General Analysis:

$$\text{Mean} = \frac{\sum \text{Values}}{n} \qquad \text{Standard Deviation} = \sqrt{\frac{\sum (X_i - \mu)^2}{n}} \qquad \text{Variance} = \frac{\sum (X_i - \mu)^2}{n}$$

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \mu}{\sigma} \right)^3 \qquad \text{Profit Margin} = \frac{\text{Profit}}{\text{Revenue}} \times 100 \qquad \text{Contribution to Revenue} = \frac{\text{Product Revenue}}{\text{Total Revenue}} \times 100$$

B) Quantitative Analysis:

$$Y = a + bX \qquad r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \qquad \text{SMA} = \frac{X_1 + X_2 + \cdots + X_n}{n} \qquad Z = \frac{X - \mu}{\sigma}$$

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1} \qquad F = \frac{\text{Variance Between Groups}}{\text{Variance Within Groups}} \quad \text{(ANOVA)}$$

The above formulas were useful for providing valuable insights from the data.

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Quantity Sold	1662	3299	1.984957882	0.973283532
Purchase Price per Unit	1662	139060.25	83.6704272	4184.970054
Selling Price per Unit	1662	158437	95.32912154	5444.995166
Total Revenue	1662	314182	189.0385078	34392.96059
Profit Margin	1662	194.4598228	0.117003503	0.003577388

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	40597950.58	4	10149487.64	1152.72466	0	2.373001786
Within Groups	73123702.33	8305	8804.780534			
Total	113721652.9	8309				

Fig 5.1: Contribution to Revenue

Observation: High variability seen in Total Revenue and Selling Price, low in Quantity Sold and Profit Margin. Sum of squares for both Between Groups and Within Groups is very high, indicating strong differences among group means. P-values is zero, strong evidence to reject the null hypothesis. All the metrics for different groups like Revenue, profit margins, CP, SP all behave differently, so it is better to treat and analyze them separately for pricing strategies.

5. Results and Findings

5.1 Results and Findings

1. Groceries, Household Essentials and Personal Care categories contribute nearly 63% of total revenue, thus forming store's core revenue drivers. Any pricing or stock adjustment in these categories would have a significant impact on overall revenue and profitability.
2. Dry Fruits generate higher revenue despite lower sales volume. This suggests premium pricing and seasonal demand.
3. Beverages, snacks and chocolates have lower revenue contribution, thus bundling or promoting offers could help increase contribution.

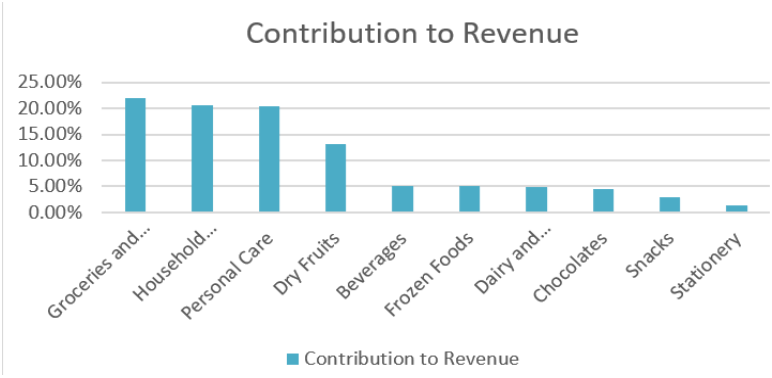


Fig 5.1: Contribution to Revenue

4. High Margin items drive profitability, and could be used for promotions and upselling. Example- Nivea Deodrant, Frozen Fries and Pooja Oil have healthy markups, making them key contributors to store's profitability.
5. Low-cost products have smaller markups, bundling them with higher-margin products could enhance overall sale value. For example- Lays and Colour Papers have lower profit margins.
6. Some products are sold below Purchase price, i.e they are sold at loss. For instance, Pista (₹275 → ₹270) and Heritage milk(₹30 → ₹29) are sold at loss, to attract customers and increase footfall. IN such case, these need to be strategically sold with high marh=gin items.

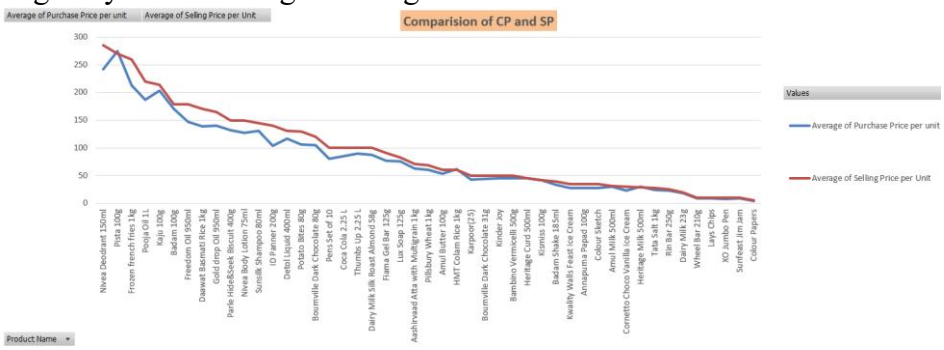


Fig 5.2: SP Vs CP

7. After a Category-wise analysis, it is found that Groceries, Household Essentials, and Personal Care have highest total selling price (₹34,188, ₹33,080, ₹32,372), indicating their major contribution to store sales. Ensuring proper inventory levels and strategic promotions for these items is crucial. While, Snacks and Stationery Have the Lowest Overall Revenue Contribution.

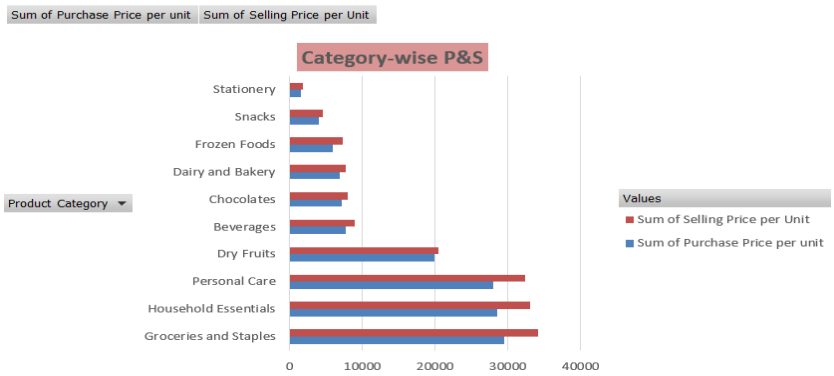


Fig 5.3: Category-wise Analysis

8. Jan → Feb: Revenue ↓18%, Quantity ↓17.5% (almost the same). Feb → Mar: Revenue ↑8%, Quantity ↑18.6% (higher quantity but lower avg. selling price). Dec → Jan: Revenue ↑3.2%, Quantity ↑2.5% (consistent pricing). March had more items sold, but at a lower per-unit price.
9. With a profit of ₹5,427.49 per unit, December was the most profitable month, possibly due to increased holiday spending or premium-priced product sales. At ₹33,682 revenue per unit and ₹3,914.22 profit per unit, February was the weakest month, reinforcing its overall lower sales and revenue trends. With ₹43,574 revenue per unit, January led in revenue generation, possibly due to strong consumer demand post-holiday season.
10. Frozen Foods consistently increased its profit margin, reaching 21.80% in March, making it the most profitable category in terms of percentage growth. Stationery maintained a stable 20% margin across all months, indicating strong markup control and pricing consistency. The profit margin for Dry Fruits dropped from 3.66% in December to just 1.44% in March, suggesting increased costs, competitive pricing pressure, or discounts affecting profitability.

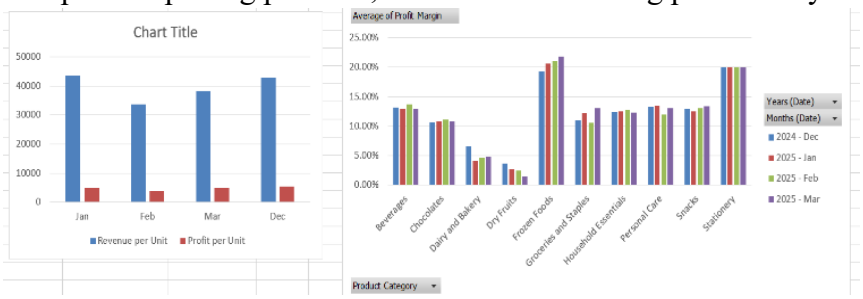


Fig 5.4: Revenue Vs Profit

5.2 Explanation of Trends and Patterns

A) Customer Behaviour

1. Returning customers significantly outnumber New Customer, indicating strong customer retention. New customers make up approx. 34.3% of total customers, while returning customers account to 65.7%.
2. December 2024 has high customer volume(439). January and March have nearly same number of customers, indicating steady footfall. But February has seen lower footfall, suggesting a seasonal dip in shopping activity.
3. The highest number of new customers came in January(158), indicating a good customer acquisition.
4. UPI is the most preferred payment mode, indicating a shift towards digital payments, due to convenience, cash backs and other merchant preferences.

5. While the UPI transactions account to 51.9% of all payments, cash payments to 27.3%, credit card payments being least preferred, account to only 20.8%.
6. Month-wise trends: UPI usage steadily increased from Dcember,2024 to March,2025. Cash payments fluctuated, with a dip in Feb and rise again in March. Card transactions show a steep decline from dec to march, implying customers are moving away from card-based transactions.
7. UPI transactions(862) alone exceeds the combination of Total of Cash and Card, 800 combined. This is a clear dominance of digital payments, reinforcing a growing trend of contactless and instant payment methods.

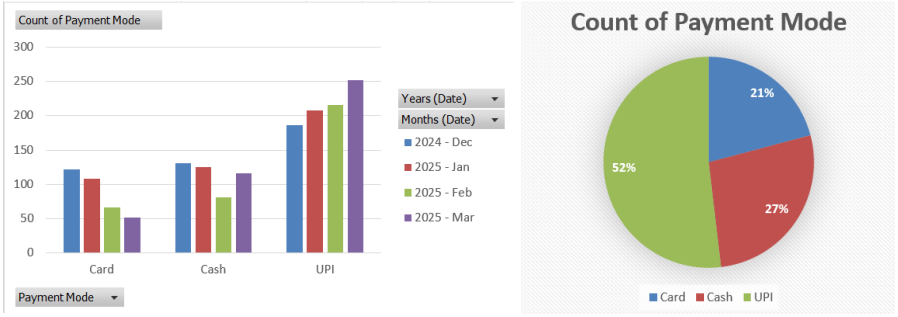


Fig 5.5: Payment Mode

8. Festival season contributes 27.4%, despite only 24.5% of total transactions being from festival days. Meaning that festival days generate higher revenue per transaction compared to that of non-festival days.
9. Revenue per transaction(211.7) is higher than non-festival days (181.7).
10. Festival days contribute nearly 25.2% of total sales volume, despite being less in number.

B) <u>Shopping Trends</u>

1. Higher Sales but Lower Profit margin for Thumbs up. This indicates that Thumbs Up is more preferred(locally) in terms of sales.
2. Equal SP for both the products, yet coca cola has more profitability.
3. Ice-cream: Significant sales increase in March, accounting to 42% of total sales across 4 months. This indicates seasonal preference, likely due to rising temperatures and demand.
4. While Kwality Walls had higher overall sales, Cornetto Choco Vanilla saw a sharp rise in March(57% of its total sales), suggesting growing consumer preference towards this variant in mere future.

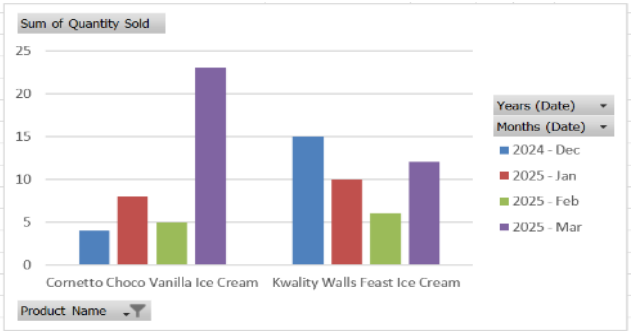
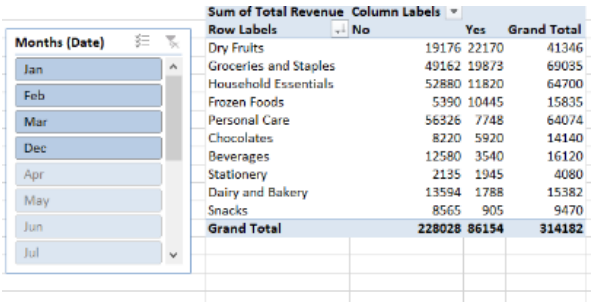


Fig 5.6: Product based Analysis

5. Dry Fruits and Frozen Foods see over half their revenue coming from festival periods, indicating they are heavily seasonal and linked to festive demand. These categories needs to be strategically stocked before festive seasons.

6. Personal Care and Household Essentials sell more on non-festive seasons, though personal care products like soaps and shampoos saw a higher sale in March, during Holi.
7. Beverages contribute just 22% revenue during festivals and Dairy&Bakery only 11.6%, indicating that much more marketing efforts need to be put on these categories.
8. December had seen a strong festival-related revenue share(27%) driven by Frozen Foods and Dry Fruits. Frozen Foods in particular had 86.6% of its December sales in festival periods, highlighting a holiday season impact on its demand.



Sum of Total Revenue	Column Labels		
Row Labels	No	Yes	Grand Total
Dry Fruits	19176	22170	41346
Groceries and Staples	49162	19873	69035
Household Essentials	52880	11820	64700
Frozen Foods	5390	10445	15835
Personal Care	56326	7748	64074
Chocolates	8220	5920	14140
Beverages	12580	3540	16120
Stationery	2135	1945	4080
Dairy and Bakery	13594	1788	15382
Snacks	8565	905	9470
Grand Total	**228028**	**86154**	**314182**

Fig 5.7: Product Category wise Revenue (Slider: Month)

Pricing Strategies:

1. Bundling up Low Margin with High Margin Products can increase profits. For instance, 'Buy 2 Thumbs Up + 1 Chips at ₹X' would be a great way to cross-sell and balance margins.
2. Apply premium pricing to Dry Fruits and Frozen Foods during the festive season. Using scarcity marketing like 'limited stocks left over' to trigger urgency and justify the premium pricing.
3. Launching Combo Offers like 'Dairy with sweets' , or 'beverages with snacks' during the festival season. We could consider Festival-themed packaging.
4. Using Tiered pricing: Economic, standard and premium, where USP could be 'buy more, save more'.