
Indian Institute of Technology, Madras
B.S. in Data Science and Application
Business Data Management
Capstone Project

# End-Term Report

Title

## Cafe Cognition:

Data-Driven Approach and Insights for Cafeteria

Submitted By

Name: Dharmik Patel

Roll number: 22F1000881

## Declaration Statement

I am working on a Project titled “Cafe Cognition: Data-Driven Approach and Insights for Cafeteria”. I extend my appreciation to The Café Town, for providing the necessary resources that enabled me to conduct my project.

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

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

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

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

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

Signature of Candidate: (**Digital Signature**)

Name: Dharmik Patel

Date: 12/12/2023

# Contents



# List of Figures



# List of Tables

# 1. Executive Summary

The Café Town, situated in Sector 16 of Gandhinagar City, contends with robust competition given its proximity to numerous educational institutions. The business is handled by two cooks and one manager. To establish a competitive edge, this project employs a data-driven methodology, utilizing a detailed six-month sales dataset analyzed with Microsoft Excel, and Python libraries like Matplotlib, and Pandas. This dataset includes comprehensive details of each order, providing a granular understanding of sales dynamics. Apart from the six-month detailed sales data, the project also utilizes one-year less detailed sales data, which does not have as extensive detail as the six-month data but helps in analyzing monthly sales and helps us spot the annual market demographic related to sales.

The analysis centers on critical aspects, including sales data analysis, revenue and sales trends evaluation, social media assessment, and product segmentation. By extracting insights from these areas, the project seeks to boost profitability and establish a unique selling proposition. This involves tailoring offerings based on observed revenue trends, and sales distribution.

The envisioned outcomes aim to position The Café Town as a leader in Sector 16, Gandhinagar City. Leveraging the analyzed data, the goal is to gain a strategic advantage over competitors by offering customized products that align with the preferences of the younger demographic, particularly college students. Through this approach, the project aspires not only to survive but to thrive in the competitive landscape, securing a notable and lucrative position for The Café Town within the local market. Additionally, the incorporation of a comprehensive one-year dataset will further contribute to our strategy, providing insights into annual sales trends and opportunities for sustained success.

## Business Details

Owner Name: Mr. Ajay Shukla
Address: The Café Town, Plot No. 440, Nr. ADC Bank, GH 5 Cir, Sector 16, Gandhinagar, Gujarat 382016
Contact: +91 96621 33922

# 2. Analysis Process

## Data Overview and Cleaning

For this project, two datasets have been provided by the business. The first dataset, named 'Data-1,' spans from January 1, 2023, to July 1, 2023. It contains detailed information about each order during this period, including Timestamp, Payment Type/Method, Invoice Number, Item Name, Item Category, Price, Quantity, Variation, Discount, and Final Total. This dataset facilitated an in-depth product-based analysis of the business. Example rows from Data-1 can be seen in Figure-1.

The second dataset, referred to as 'Data-2,' covers the timeframe from August 1, 2022, to July 31, 2023. It comprises only four main columns: Invoice, Timestamp, Payment Type/Method, and Final Total. The extended timeline of this dataset over 12 months is valuable for identifying time-based trends and the overall revenue of the business.

| Date | Timestamp | Invoice | Payment Ty | Order Type | Item Name | Price | Qty. | Sub Total | Discount | Tax | Final Total | Status | Variatio | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2023-07-04 | 2023-07-04 15:33:42 | 22361 | Cash | Dine In | Aloo Tikki Burger | 50.00 | 1.00 | 50.00 | 0.00 | 0.00 | 50.00 | Success | | Burger |
| 2023-07-04 | 2023-07-04 15:33:42 | 22361 | Cash | Dine In | Orange | 70.00 | 1.00 | 70.00 | 0.00 | 0.00 | 70.00 | Success | | Juice (300Ml) |

Figure-1: Data-1

| Invoice | Date | Payment Type | My Amoun | Total (₹) |
|---|---|---|---|---|
| 23390 | 2023-07-31 23:56:47 | Cash | 10.00 | 10.00 |

Figure-2: Data-2

During the data cleaning process, we addressed raw data provided by the business. Extraneous columns such as 'dinner' and 'biller,' which mostly contained empty entries, were removed. Additionally, the date column in the dataset was initially in string format; we converted it to the date-time format in the Pandas data frame for efficient time-based analysis. Certain columns, like 'Variation' (Small, Medium, and Large), were omitted due to a significant number of empty entries. Despite the potential importance of these columns, the lack of available data entries rendered them unusable for analysis. All the values in the ‘Order Type’ column were ‘Dine In’, so we also neglected the use of that column.

In the case of Data-2, the 'Date' column contained both date and time information. To facilitate a separate analysis of date and time, we split this column into two distinct columns.

These data preparation steps lay the foundation for a comprehensive analysis of product-based insights from Data-1 and time-based trends using the extended timeline in Data-2.

*(Datasets: Link)*

## The Analysis

After completing the data cleaning process, the datasets were ready for analysis. It was imperative to gather basic statistics to gain insights into the business. Utilizing both Data-2 and Data-1, we extracted key statistical information such as total sales, the day with the highest revenue, the top-selling product and its associated revenue, the months contributing the most to revenue, and other relevant metrics. These findings are summarized in Table-1.
To derive this information, Python pandas data frames were employed. Various pandas functions were utilized to navigate and analyze the datasets, enabling the extraction of meaningful insights. The analysis aimed to provide a comprehensive understanding of the business's performance, offering a foundation for further in-depth examination.

| | |
|---|---|
| **Total Revenue (01-08-2022 to 31-07-2023)** | 23,59,619 INR |
| **Most Revenue Generated in a Day** | 3,230 INR (24-10-2023) |
| **Most Revenue Generated in a Month** | 3,29,131 INR (September 2022) |
| **Most Revenue Generating Product** | Cold Coffee 300 Ml (65,548 INR) |
| **Most Selling Product** | Tea ( 2,598 items sold) |

Table-1: Important Statistics

### Classwise Sales Distribution

Following the initial statistical analysis, our focus shifted to identifying the most revenue-generating and top-selling products, as well as the predominant item categories. This step is crucial for pinpointing the strengths and weaknesses of the business. To enhance our understanding, we categorized each order into one of three classes: 'Beverage,' 'Food,' and 'Combo' (comprising multiple items, including both Food and Beverages). This classification enables us to discern the most revenue-generating class and identify potential growth opportunities. The insights derived from this categorization were visually represented using a pie chart. This visualization offers a clear depiction of the distribution of revenue and sales among the different product classes, aiding in strategic decision-making for the business.

### Revenue and Sales Distribution among Categories

To understand the distribution of revenue among item categories, we turned to Data-1 for relevant insights. Our objective was to assess the revenue generated by each category and make meaningful comparisons between them. For this visualization, a bar chart with separate bars for each category emerged as the most effective choice. This approach allows for a clear comparison of revenue across different categories.

Similarly, we adopted the same methodology to generate a bar chart illustrating the quantities sold for each category. By plotting these two charts, we could discern any disparities between the categories that are most sold and those that generate the highest revenue. This comparative analysis aids in identifying potential areas for improvement or strategic focus within the business. The visual representation through bar charts provides a straightforward and insightful means of interpreting the data.

## Payment Method Distribution

Within Data-2, a column specifies the payment method chosen by the customers. Leveraging this information, we categorized orders into two distinct segments: Cash and Online paid orders. Given the exclusive nature of these two categories in the dataset, a pie chart emerged as the most suitable visualization method. Matplotlib was employed to generate the pie chart, facilitating a clear and concise representation of the distribution of payment methods.

## Most Selling and Least Selling Items

A crucial strategy for any business is to identify its top-selling star products for effective marketing. Similarly, recognizing the least-selling products is important for analysis, helping to understand whether there is a lack of demand or other issues. To highlight these products along with their precise revenue and sales data, tables have been prepared instead of using visualization methods. The tables offer a concise overview of the sales and revenue information for the top 5 best-selling and the bottom 5 least-selling products. The results were achieved by utilizing a pivot table from Excel on Data-1.

## Purchase Power Analysis

To understand the purchasing power of customers for the business, we employed the pandas 'group by' function to group order totals from Data-1. Considering the connected distribution of this data, we opted for a Histogram as an appropriate representation, selecting an appropriate bar width. The Histogram provides a visual depiction of the distribution of order totals, offering insights into the purchasing behavior of customers.

This analysis proves valuable in understanding the target market of the business, especially in identifying price points that attract customers. By examining the distribution of order totals, the business can strategically capitalize on products with appealing prices, catering to the preferences and purchasing power of its customer base. The Histogram serves as a useful tool for visualizing and interpreting this important aspect of customer behavior.

## Time-Based Business Analysis

To conduct a critical analysis of the business, tracing the revenue over time is essential. Leveraging Data-1, which provides detailed order information from August 2022 to August 2023, we employed line charts for this purpose. Line charts prove efficient and visually effective for time-based tracking. Using Excel, we created a pivot table to compile monthly sales data and

subsequently plotted a line chart. To observe the revenue trend more closely, a regression line was added to the chart for enhanced analysis.
For a more granular breakdown at the day level, we calculated the average daily revenue for each month. Opting for a bar chart over a line chart in this case allows for a straightforward comparison of monthly averages. This approach aids the business in setting daily revenue targets to achieve overall monthly objectives.
The timestamp data in Data-2 facilitated a deeper dive into time-based analysis. To achieve this, we utilized pandas functionality to calculate the percentage of revenue generated in each hour over the entire dataset. Subsequently, we employed Matplotlib to create a line chart illustrating the daily sales portions. This approach was chosen to provide a comprehensive view of revenue distribution across the 24-hour period. The analysis aims to identify peak hours and periods of high customer activity, aiding in the understanding of optimal times for implementing promotional offers to boost business during specific time frames. These insights contribute to strategic decision-making and the optimization of business operations.

Based on the comprehensive analysis conducted, we will construct a SWOT (Strengths, Weaknesses, Opportunities, Threats) table to summarize the findings and provide a holistic view of the business. This SWOT analysis will serve as the foundation for strategic recommendations aimed at uplifting the business by leveraging its strengths, addressing weaknesses, capitalizing on opportunities, and mitigating potential threats.

To access all the analyses, the following link contains the essential Python interactive Python notebook (ipynb) files and Excel analysis files: Link.

# 3. Results and Findings

## Classwise Sales Distribution

Café Town offers an extensive range of food and beverages, primarily categorized as fast food. The menu is strategically classified into three distinct classes: Food, Beverage, and Combo. A crucial aspect in assessing the business's strengths and weaknesses involves comprehending the distribution of revenue among these three classes. Figure-3 presents a pie chart illustrating the revenue distribution sourced from Data-1.

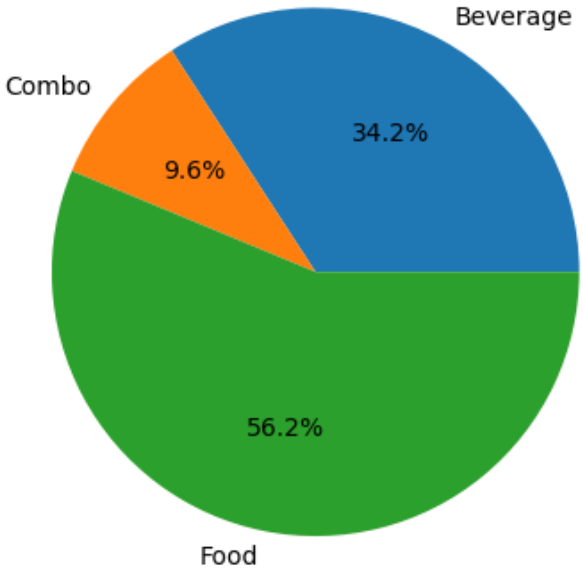


Figure-3: Revenue Distribution by Product Class

### Observations:

- The analysis reveals that food items constitute the primary revenue stream, accounting for over 56% of the total sales. Consequently, food items emerge as a notable strength for the business.
- Beverages are sold more frequently; however, due to their comparatively lower pricing, they generate relatively lesser revenue than food items.
- Combos, incorporating both food and beverages, hold a pivotal position in our business model, particularly given our primary focus on the economical segment of the market. The noteworthy contribution of 9.6% to the total revenue underscores their significance. Strategic adjustments present an opportunity for enhancement in this regard.
- Given our café-centric business model, the primary Unique Selling Proposition (USP) lies in the realm of beverages, particularly coffee. Therefore, from a revenue perspective,

there exists a noteworthy opportunity for substantial growth within both the beverage and combo classes.

## Revenue and Sales Distribution among Categories

Figure-4 displays a bar chart representing the revenue generated by each Item Category sourced from Data-1. It is imperative to conduct a comprehensive analysis to identify the most substantial revenue-generating categories. Equally significant is the examination of categories exhibiting suboptimal performance.

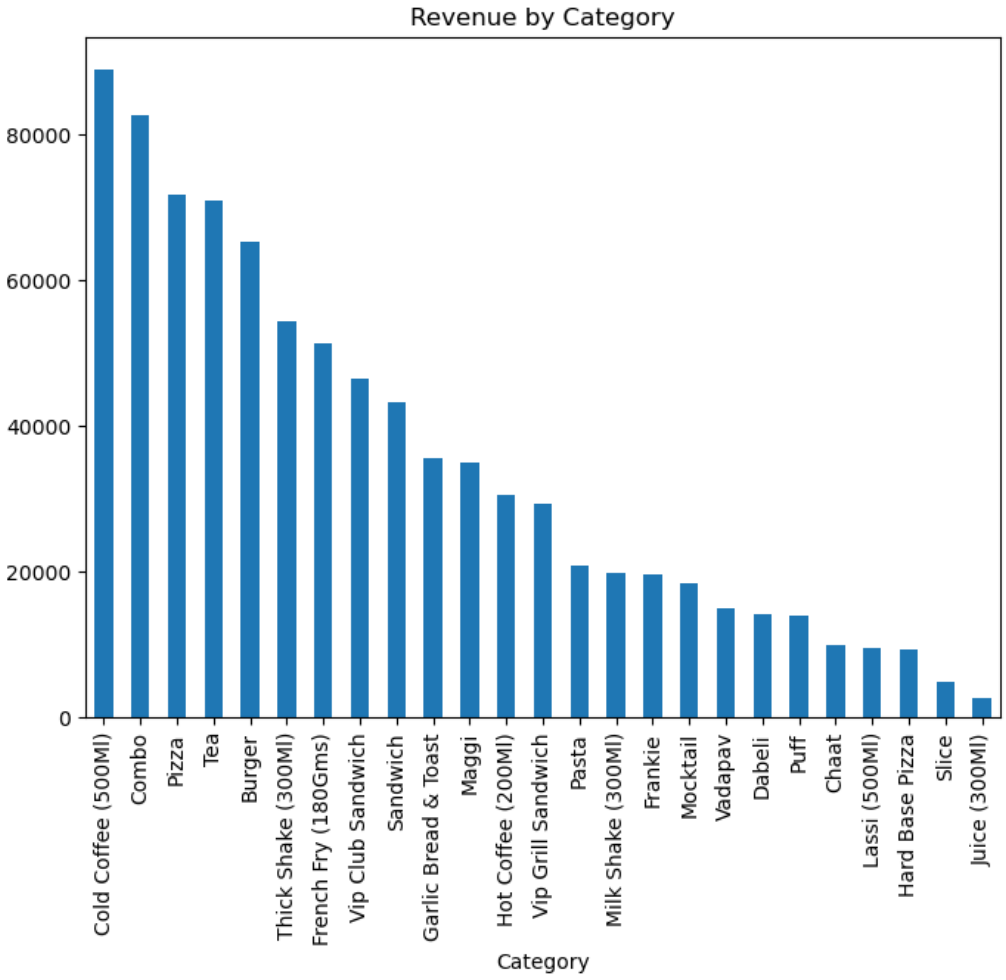


Figure-4: Revenue Distribution by Product Category

### Observations:

- While beverages, in general, contribute less to the overall revenue in comparison to food products, still it can be observed prominence of cold coffee is the most revenue-generating category.

- Given the nature of the establishment as a café, it is ideal for the signature offering to center around coffee-based items, and this is indeed the case for our business. This identification stands as a recognized strength of business
- Revenue by Combo category is very significant and can also be considered a strength of our business.
- Certain categories, including Chaat, Dabeli, Pasta, and Maggi, typically exhibit favorable performance within the context of this business. However, based on the observed bar chart, it appears that these categories are not realizing their full potential. This discrepancy suggests a substantial growth opportunity within these specific categories.

Figure-5 displays the sales data for each category. This information serves as a valuable tool for discerning distinctions between categories that contribute significantly to revenue and those that lead in terms of sales volume.

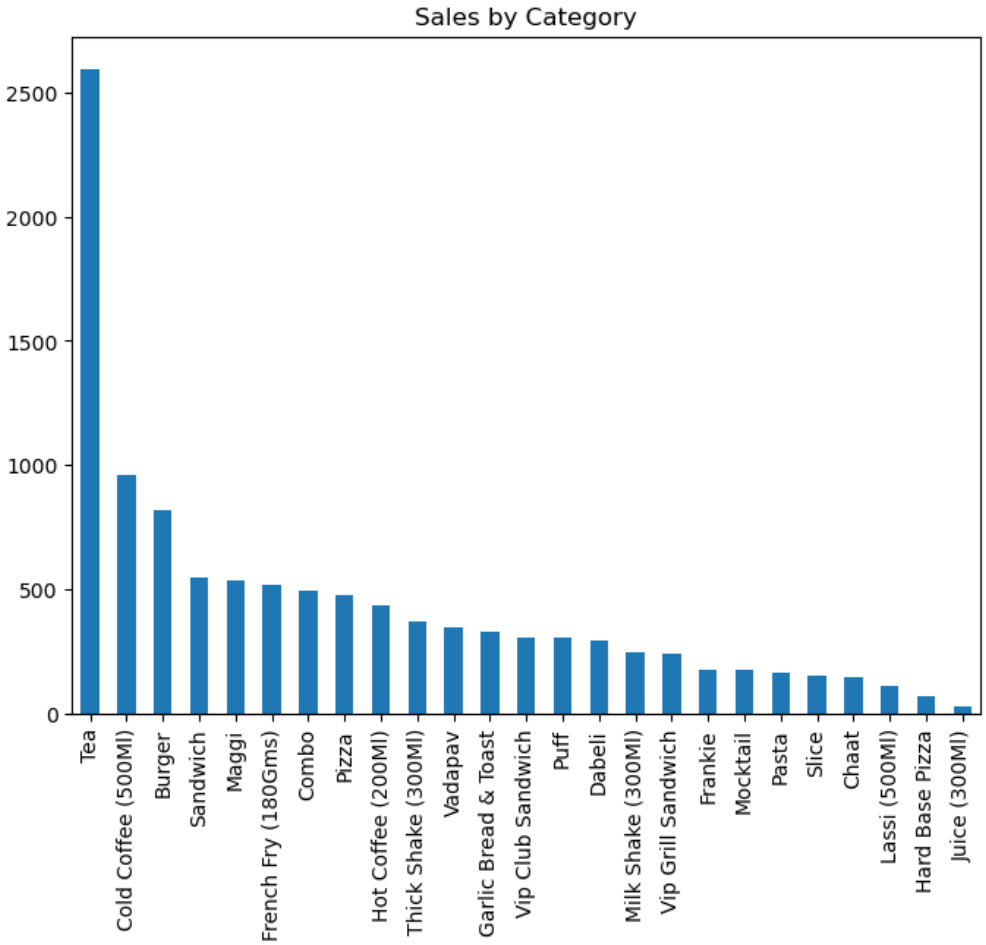


Figure-5: Sales Volume Distribution by Product Category

## Observations:

- Tea category generates exceptionally high sales but the revenue generated by it is not that significant.

- Cold Coffee is generating high sales volume and the highest revenue as well, it can be considered as the signature offering of the business.
- The category labeled as 'Combo,' while not attaining exceptionally high sales volume, demonstrates a noteworthy capacity for revenue generation within the business.

## Most Selling and Least Selling Items

Data-1 facilitated the identification of the high-performing and low-performing products in our business. The business offers a menu comprising more than 100 products, and a majority of them may not require any attention. The analysis will concentrate on the top five best-selling products and the five least-selling products. Table 2a presents details regarding the top five most sold products, while Table 2b provides information on the least sold products.

| Item | Sales | Revenue |
|---|---|---|
| Masala Tea | 1612 | 32,229 INR |
| Cold Coffee 300 Ml | 1093 | 65,548 INR |
| Water 500ml | 921 | 9,210 INR |
| Aloo Tikki Burger | 685 | 34,205 INR |
| Combo-1 | 538 | 64,022 INR |

Table-2a: Details of Top-5 most sold Items

| Item | Sales | Revenue |
|---|---|---|
| Mexican Pizza (Large) | 3 | 420 INR |
| Cheese Capsicum Garlic Toast | 3 | 360 INR |
| Combo-3 | 2 | 280 INR |
| Extra Butter | 1 | 10 INR |
| Cold Bournvita With Ice Cream | 1 | 90 INR |

Table-2b: Details of 5 least sold items

### Observations:

- Cold Coffee and Masala tea are the only products with over 1000 items sold in the six months of business. Among these two, cold coffee is the one generating significant revenue.

- Surprisingly, items categorized as Combo appear in both the top-selling and least-selling categories. Notably, Combo-3 experienced only two sales over a span of six months.

## Payment Method Distribution

From Data-2, we obtained information regarding the payment methods utilized by customers in the business. The pie chart in Figure-6 illustrates the distribution of payment methods, distinguishing between cash and online payments. It reveals that 62.3% of total orders were conducted using cash, while 37.7% were made through online payment methods. The online payment method encompasses transactions made through UPI services offered by various brands, including Paytm, PhonePe, Google Pay, and others.

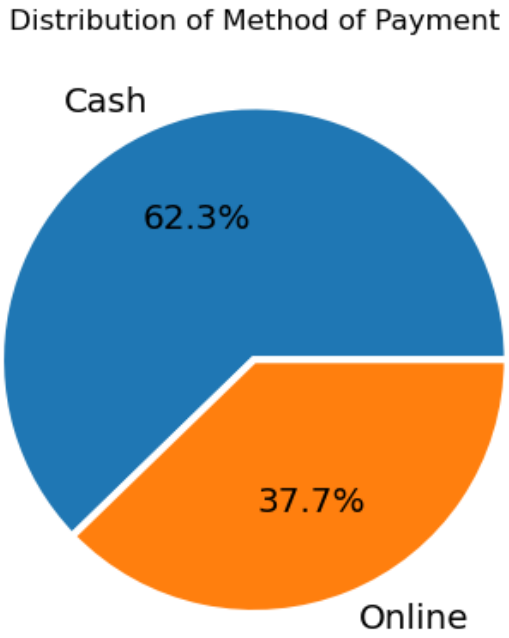


Figure-6: Payment Method Distribution

### Observations:

- The pie chart in Figure 4 illustrates that the cash payment method remains dominant in the business. However, there exists an opportunity to transition towards the cashless payment method, which offers several advantages. UPI-based payments, in particular, facilitate seamless record-keeping and are more convenient compared to traditional cash payments.

## Purchase Power Analysis

A business must comprehend the financial demographics of its target audience to offer products at affordable prices. This understanding aids in determining the appropriate price range for items on the menu and identifies which price points can be overlooked. The histogram in Figure-7 illustrates the distribution of order amounts, derived from the analysis of Data-1.

### Observations:

- It is evident that the majority of orders fall below the 200 INR threshold. This indicates a target market that seeks economical purchases. This observation aligns with the understanding that a significant portion of customers comprises students who may not have an income.

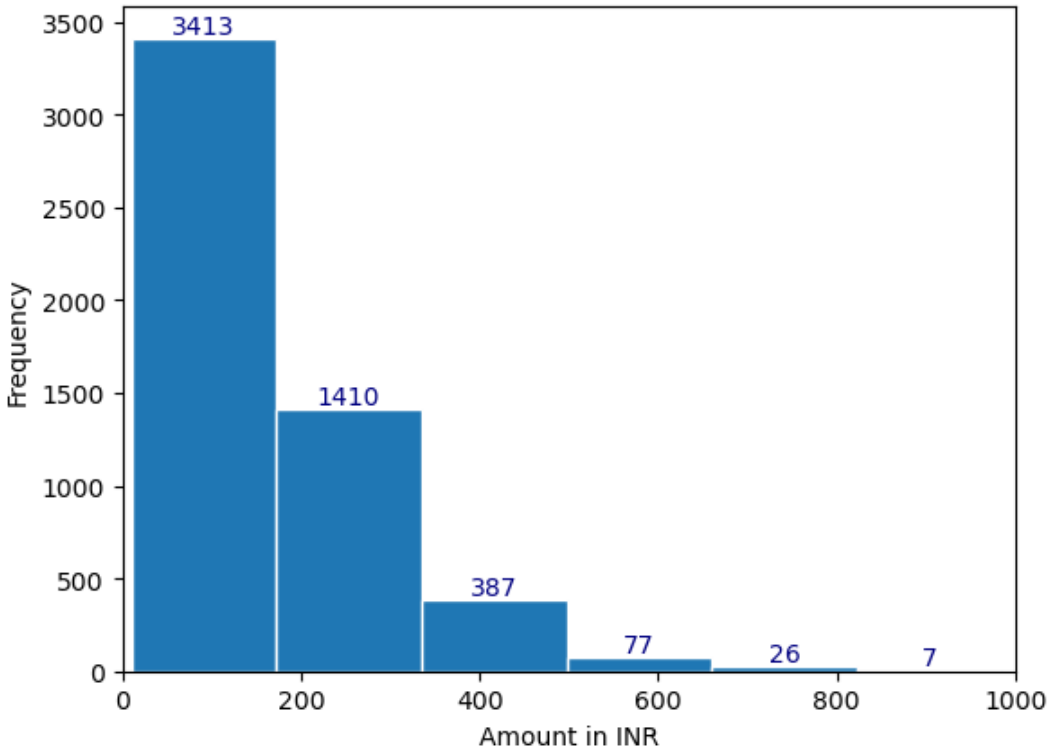


Figure-7: Order Amount Distribution

- Given the relatively modest revenue per order for the business, it becomes imperative to augment customer attraction. In this context, the inclusion of cost-effective products that can effectively cater to customer satisfaction emerges as a crucial strategy.
- Some products should be identified that do not have significant sales, and new products can replace them.

## Monthly Revenue Trend

To assess the business performance, it is crucial to analyze the sales over a substantial time frame. Figure-8 illustrates the line chart depicting monthly revenue trends extracted from Data-2, spanning from. The chart also consists of a regression line, providing a visual representation of the revenue trajectory.

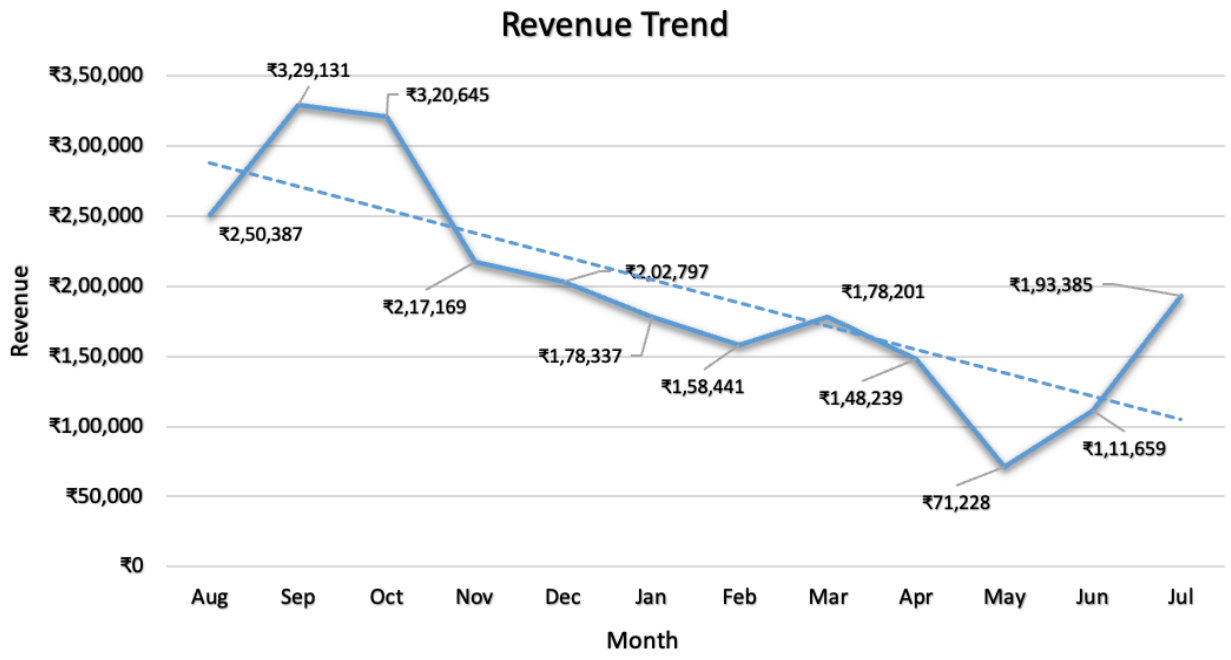


Figure-8: Revenue Trend Over Time

## Observations:

- It is evident from the regression line that the revenue is decreasing over the given period. An ascent was noted starting from May 2023, although it was not substantial.
- The observed decline poses a considerable threat to the business. Factoring in all associated costs, including salaries, a revenue below 1.5 lakh INR would result in a financial loss for the owner.
- The lowest point was observed in May 2023 when the revenue for that month was only around 71,000 INR.
- Concluding, the business demonstrates relatively strong performance in the second half of the year, which is from July to December. During this timeframe, the average monthly revenue reaches 2,52,252 INR.
- In contrast, during the first half of the year, spanning from January to June, the business exhibits comparatively lower performance compared to the second half. Throughout these months, the business yields an average revenue of 1,41,017 INR.
- After discussing the revenue decline with the owner, it became apparent that one contributing factor to the decrease in May was the summer vacation of university students in the area. However, it was also conveyed that this decrease in revenue was still considerably lower than that observed in the year 2022.
- In discussions with the owner, it became evident that high competition constituted a primary factor contributing to the decline in revenue. Adjustments are necessary to ensure the profitability of the business.

## Average Daily Revenue Trend

To further analyze the time-based revenue patterns, the same dataset was utilized to compute the average daily revenue for all months, as illustrated in the bar chart in Figure-9. This breakdown of monthly revenue into daily targets enables the establishment of daily business targets, offering a more comprehensible and convenient approach for business operators.

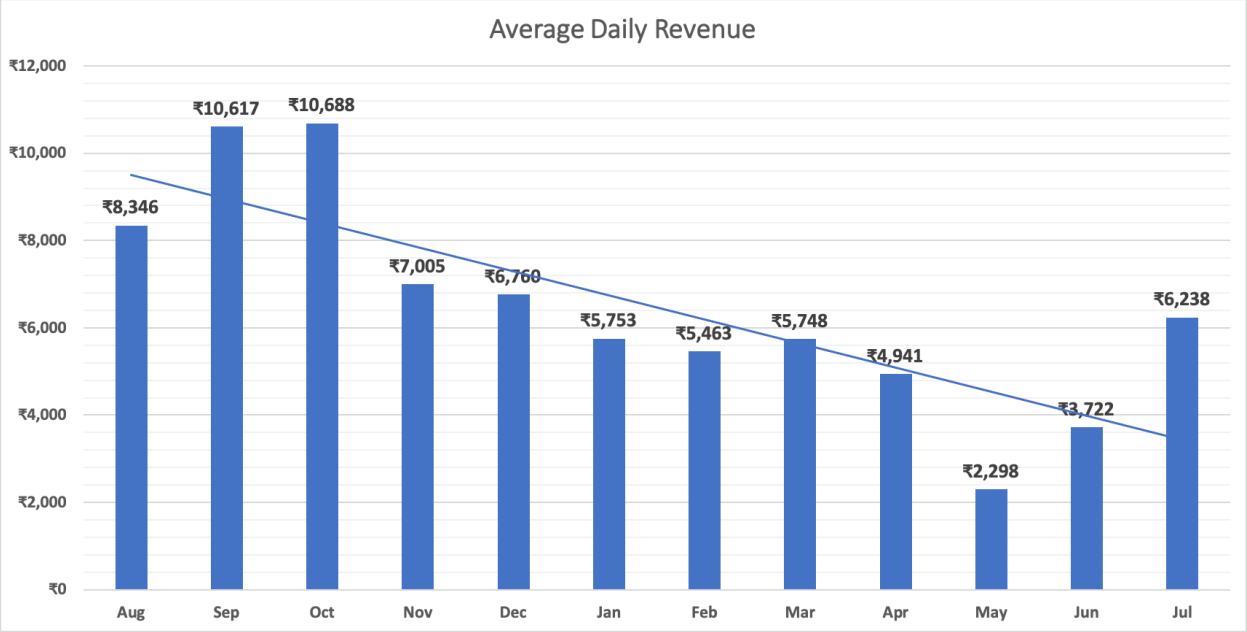


Figure-9: Average Daily Revenue Distribution

### Observations:

- During low periods, the business decreased to below 3000 INR per day. Considering labor charges, costs, and bills, this is not favorable.
- In peak periods, the daily average revenue reached approximately 10,000 INR, a commendable figure. If sustained for a few more months, it could potentially offset the business's lower revenue periods.
- Aside from improving extreme lows, there is a possibility of increasing revenue in November and December as well. During these months, the average revenue is around 7000 INR, which is not bad. However, aiming for a range of 9000 to 10000 INR could further enhance performance.
- Observing the business performance in the second half of the year (July to December), a positive trend can be seen. The mean of the average daily revenue during this period amounts to 8,276 INR. In contrast, the mean for the first half of the year is 4,654 INR, indicating a noteworthy difference in the mean average revenue between the two halves of the year.

## Hourly Revenue Trend

For any business, accurate identification of rush hours and the distribution of sales throughout the day is crucial. Utilizing Data-2, one year of data was analyzed to observe daily sales trends, as depicted in the line chart in Figure-10.

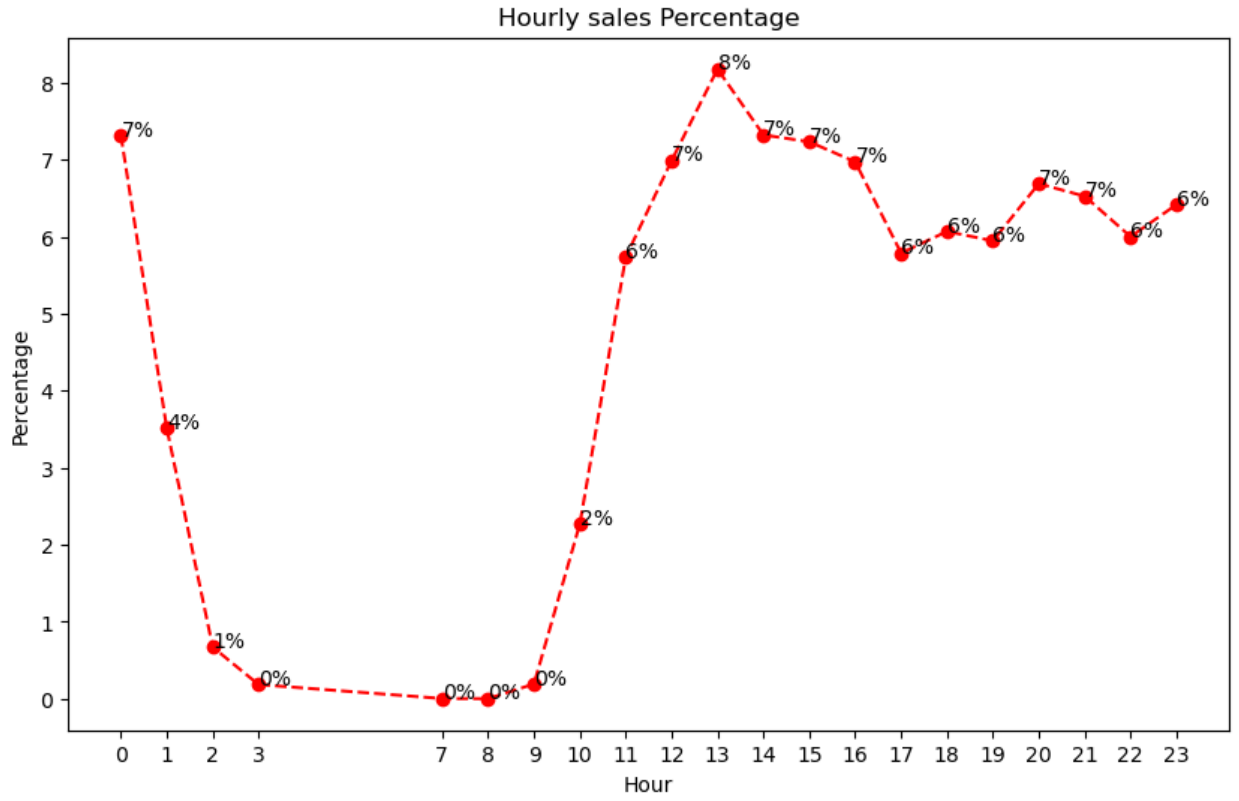


Figure-10: Hourly Revenue Distribution

### Observations:

- The business operates from 9:30 AM to 2 AM the following morning, with sales reaching their peak at approximately 1 PM, typically during the lunchtime period.
- Sales remain consistently stable thereafter, comprising approximately 6%-7% of daily total sales. This consistency signifies minimal fluctuations in sales.
- The workload for employees appears to be fairly consistent, given the uniform sales pattern throughout the day.
- During the 3 PM to 7 PM timeframe, revenue upliftment to align with peak hours could be achieved through the implementation of time-based offers aimed at attracting customers.

# 4. Interpretation of Results and Recommendation

For interpreting the results of the findings generated from this data-driven analysis, the SWOT analysis method was selected to convey the essential insights. Based on all the results found in the business analysis, points were prepared highlighting strengths, weaknesses, opportunities, and threats to the business. This analysis is referred to as the SWOT (Strengths-Weaknesses-Opportunities-Threats) analysis.

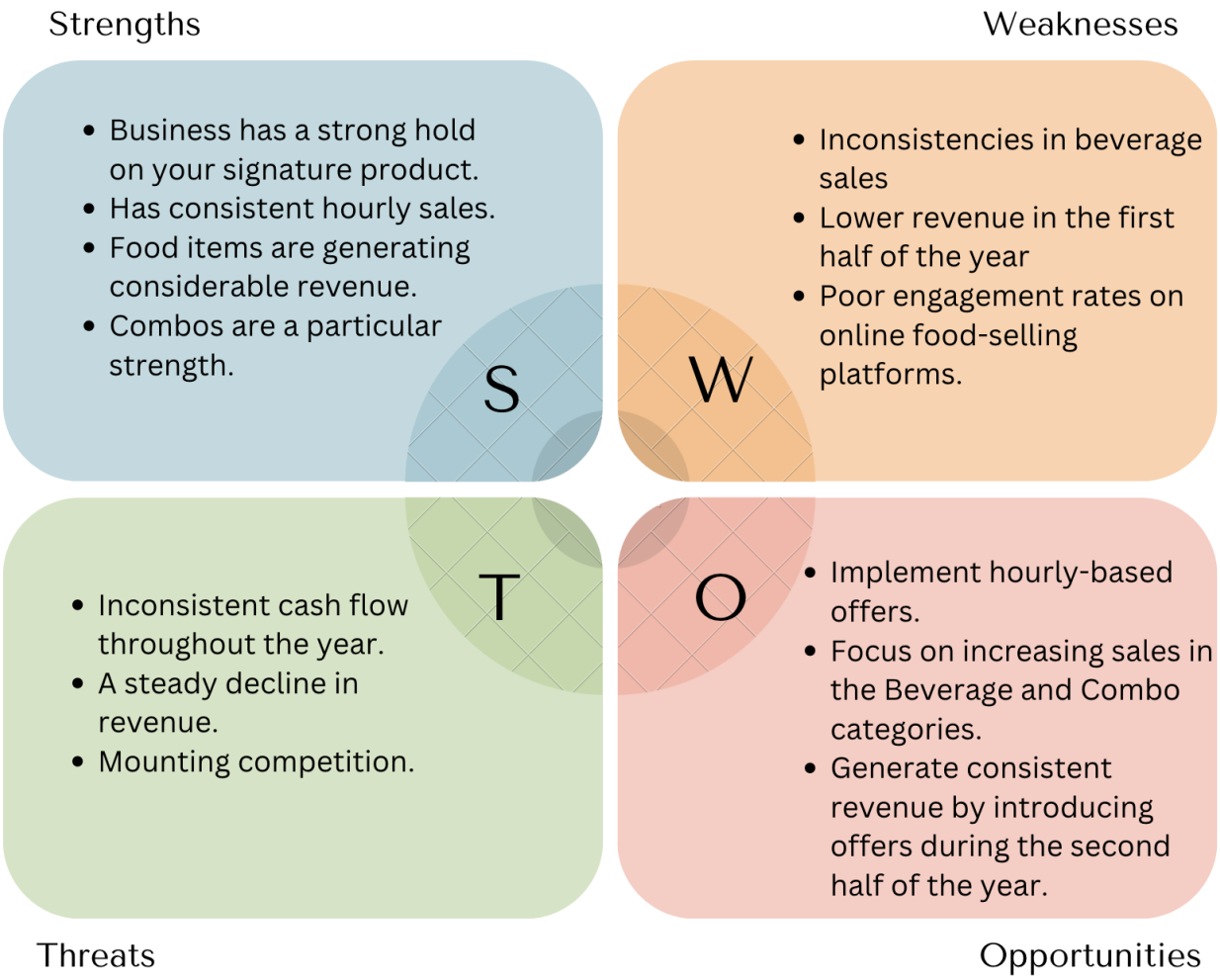


Figure-11: SWOT Analysis

Based on the results and SWOT analysis, the following recommendations were provided to the business owner.

1. Considering that the Combo category is among the most revenue-generating segments of the business, it is suggested to replace Combo-3 with a more lucrative combination of products that have higher sales and revenue.

2. Tea generates substantial sales but comparatively lower revenue. On the other hand, 'Oil Fry Vadapav,' which complements tea, is not contributing significantly to sales or revenue. Individually, Tea is priced at 20 INR, and Vadapav is priced at 25 INR, with a combined cost of 45 INR. Proposing a bundled price of 30 to 35 INR for both items could be considered, ensuring a maintained profit margin.
3. The business currently lacks emphasis on online food delivery platforms like Zomato and Swiggy. Leveraging these platforms could yield substantial revenue and enhance the business's accessibility throughout the city. The recommendation to the business owner is to amplify their online presence and systematically document relevant data.
4. Given the minimal revenue generated during the first half of the year (January to June), it is imperative to attract more customers during this period. To achieve this, we recommend implementing an enticing 'coupon-based offer' to draw a larger crowd.
    - Under this scheme, any customer with a bill exceeding 150 INR will receive a coupon, which can be redeemed on their subsequent order. Upon redemption, the customer will be eligible for a 15-20% discount on a minimum purchase of 250 INR. This offer will be exclusively available from January to June, and the coupon will be valid until the end of the respective month.
5. A significant portion of our customers consists of students. There should be an offer exclusively for students, as many food places have such offers.
    - For example, To avail of a 15% discount on orders above 100 INR, students must present their institute identity card.
    - This will help us in pulling the target market directly and it will give us the edge over the competition in the area.
6. For maintaining and tracking records, the business can subscribe to systems like PetPooja. These systems not only simplify record-keeping but also enable the generation and analysis of business data anytime, anywhere. Additionally, such a system can serve as a Customer Relationship Management (CRM) tool by keeping track of customer mobile numbers, allowing for direct marketing messages through WhatsApp Messenger. Engaging in this type of marketing activity can provide a competitive edge. Moreover, it makes it easy to get feedback from customers directly.
7. To retain customers, implementing a customer loyalty program is essential. Adopting systems like Reelo can help establish an effective loyalty program. In this system, customers earn points in the system's wallet each time they place an order at our business. Once a certain number of points is accumulated, customers can redeem them for various items or avail of special offers. This system, widely adopted by numerous restaurant businesses in urban areas, has proven to be successful.