

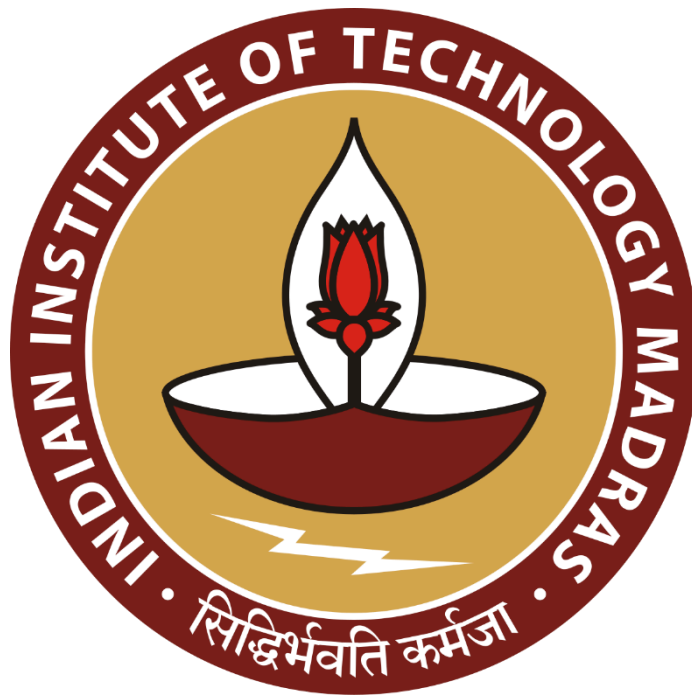
A Data-Driven Analysis of Sales and Customer Behavior

A Final report for the BDM capstone Project

Submitted by

Name: Kapishankar Ashtankar

Roll number: 21F3000371



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Contents

1	Executive Summary and Title	2
2	Proof of Originality	3
3	Metadata and Descriptive Statistics	3
	3.1 Metadata	3
	3.2 Descriptive Statistics	4
4	Detailed Explanation of Analysis Process	8
	4.1 Data Cleaning and Preprocessing	8
	4.2 Analysis Process	8
5	Results and Findings	10
	5.1 Item Sales Analysis	10
	5.2 Time-Based Sales Trends	12
	5.3 Customer Behavior Patterns	14
	5.4 Forecasting Future Demand	16
6	Interpretation of Results and Recommendations	17
	6.1 Interpretation of Results	18
	6.2 Recommendations	18

1. Executive Summary and Title

Balaji Fast Food is a local fast-food vendor offering a variety of snacks and beverages. With increasing customer footfall and fluctuating daily operations, the business faces challenges in identifying high-performing items, managing inventory, and optimizing staff deployment. This project was initiated to solve two key problems: (1) identifying the most frequently sold and revenue-generating food items, and forecasting future demand for effective inventory planning, and (2) analyzing historical sales data to discover time-based trends and patterns that influence business performance.

The dataset, collected from Kaggle, contains 1,000 records with fields such as `order_id`, `date`, `item_name`, `item_price`, `quantity`, `transaction_amount`, `transaction_type`, `time_of_sale`, and `received_by`. Data cleaning included handling missing values, converting data types, and formatting time-related fields. Descriptive statistics were calculated for quantitative variables, and analytical techniques such as group-wise aggregations, time-series analysis, and Prophet-based forecasting were used to derive insights.

The findings revealed that Cold Coffee, Sugarcane Juice, and Frankie are the top-selling items by both quantity and revenue. While Cold Coffee and Sugarcane Juice dominate in terms of volume, items like Frankie and Sandwich yield higher revenue per transaction. Sales patterns across time reveal peak demand during night hours, and notable monthly peaks in October 2022 and January 2023, likely due to festive seasons. Gender-based analysis showed that male staff processed more orders, though female staff excelled in high-value midnight transactions.

Based on the results, key recommendations include: stocking high-selling items in larger quantities, running targeted promotions on profitable but under-ordered items like Sandwiches, scheduling more staff during night hours, and using predictive models to plan resources ahead of demand spikes. These data-driven strategies can significantly enhance inventory control, staff utilization, and long-term business decision-making.

2. Proof of Originality

This project is based entirely on secondary data obtained from the open-source platform Kaggle, which is widely recognized for hosting datasets across domains including business analytics, machine learning, and data science. The dataset used in this analysis provides comprehensive sales transaction records from a fast-food outlet, including details such as order date, item name, quantity sold, item price, item type, transaction amount, time of sale and mode of payment.

The data was not collected through any primary method such as surveys, interviews, or on-site observations. Instead, it was sourced directly from Kaggle and used exclusively for academic purposes.

Dataset Title: Fast Food Sales Report

Dataset Platform: Kaggle

Dataset Creator: Rajat Surana

Dataset Link: <https://www.kaggle.com/datasets/rajatsurana979/fast-food-sales-report>

Dataset Source: <https://www.kaggle.com>

3. Metadata and Descriptive Statistics

3.1 Metadata:

Below is the metadata describing each variable in the dataset used for analyzing sales trends and customer behavior at a balaji fast-food outlet.

Variable Name	Description	Data type	Range / Values
order_id	Unique identifier for each sales transaction.	Integer	1-1000
date	Date on which the transaction occurred	Date	01-04-2022 - 30-03-2023
item_name	Name of the food or beverage item sold	String	Eg. Cold coffee, Vadapav, Frankie etc
item_type	Category of the item	Categorical	Fast Food / Beverage

item_price	Price per unit of the item sold	Float	20.00 - 60.00
quantity	Number of units sold in the transaction	Integer	1 - 15
transaction_amount	Total amount paid for the transaction (item_price × quantity)	Float	20.00 - 900.00
transaction_type	Mode of payment	String	Cash / Online / Others
received_by	Gender of the staff who processed the order	String	Mr., Mrs.
time_of_sale	Time slot during which the sale occurred	String	Morning / Afternoon / Evening / Night / Midnight

Table 1: Metadata

3.2 Descriptive Statistics:

The following table presents the summary statistics for the numerical variables in the dataset:

	item_price	Quantity	Transaction_amount (₹)
Mean	33.31500	8.16200	275.23000
Std. Dev.	14.92174	4.41307	204.40297
Min	20.00000	1.00000	20.00000
25%	20.00000	4.00000	120.00000
50% (Median)	25.00000	8.00000	240.00000
75%	50.00000	12.00000	360.00000
Max	60.00000	15.00000	900.00000

Table 2: Descriptive Statistics

Key Insights and Interpretation:

1. Item Price:
 - The average item price is ₹33.32, with most prices falling between ₹20 and ₹50.
 - The low standard deviation (₹14.92) indicates that pricing is fairly consistent, helping customers predict cost.
2. Quantity:
 - Customers typically buy around 8 items per transaction.
 - The interquartile range (4 to 12 items) shows variability in order size, which can impact inventory planning.
3. Transaction Amount:
 - The mean transaction value is ₹275.23, but the high standard deviation (₹204.40) indicates wide variation in how much customers spend.
 - The maximum value of ₹900 implies some bulk or large combo purchases, while minimum is ₹20, likely for single-item orders.

These statistics reveal that:

- Most transactions are mid-range (₹240–₹360), helping the business set pricing and combo strategies.
- Understanding item quantities and purchase ranges assists in stock control, pricing optimization, and demand forecasting.
- High-value transactions, although less frequent, could be targeted through promotions or loyalty programs.

Let's uncover the description of categorical variables.

1. Item_type

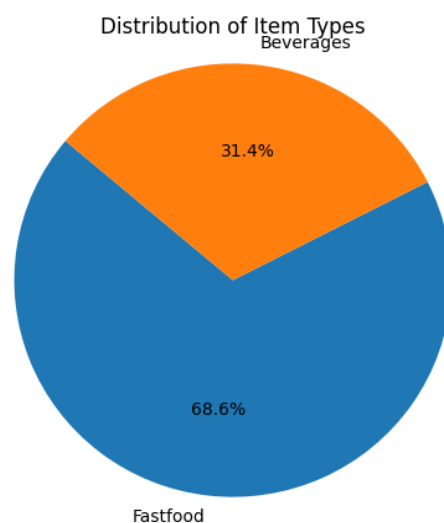


Figure 1: Distribution of item types

Figure 1 is a pie chart illustrating the distribution of item types within the dataset. The chart shows that 68.6% of the data corresponds to Fast Food items, while 31.4% represents Beverages. This visualization effectively demonstrates the proportion of each category relative to the whole, making it easy to compare their sizes at a glance. Pie charts are particularly suitable for categorical data with a small number of categories, as they clearly display how each segment contributes to the total.

2. transaction_type

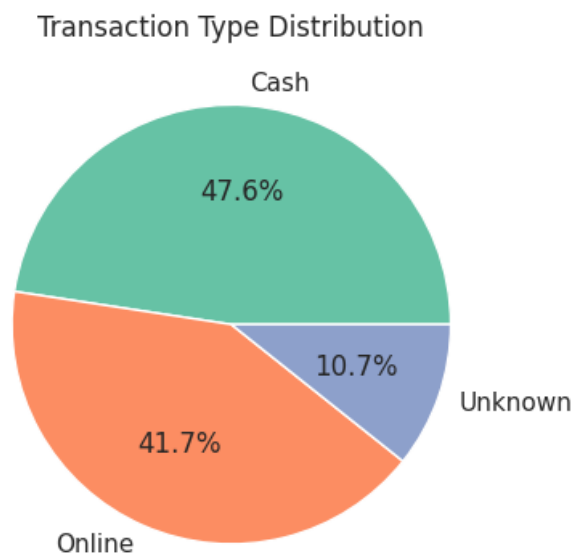


Figure 2: Distribution of Transaction type

Figure 2 is a pie chart depicting the distribution of transaction types within the dataset. The chart shows that Cash transactions account for 47.6%, making them the most common. Online transactions represent 41.7%, while Unknown transaction types make up 10.7%. This visualization clearly highlights that cash and online transactions together constitute the majority of the data, with a small portion remaining unidentified. Pie charts like this are effective for illustrating the proportion of different categories in categorical data.

3. Recieverd_by

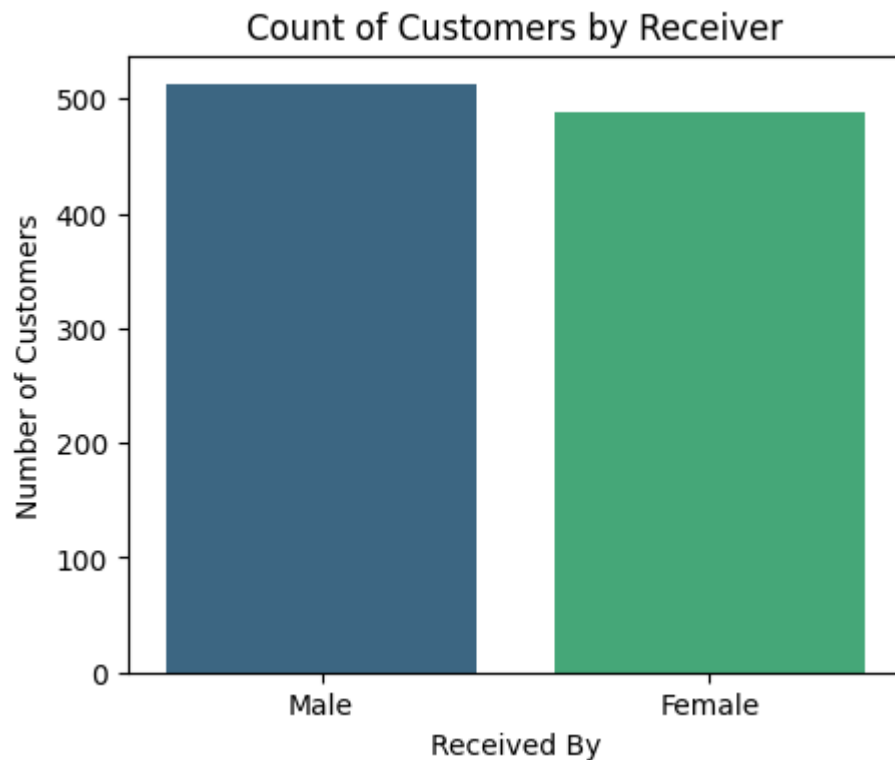


Figure 3: count of customers by receiver

Figure 3 shows that male staff received 512 customers while female staff received 488, indicating a nearly balanced customer handling distribution. Male staff managed slightly more interactions, but the gap is minimal, suggesting fair and effective staff deployment. A bar graph is ideal for this comparison because it clearly displays differences between two categories, making it easy to interpret customer distribution across genders at a glance.

4. Detailed Explanation of Analysis Process

4.1 Data Cleaning and Preprocessing:

The initial step in the analysis was focused on cleaning and structuring the raw data to ensure reliability and accuracy. The dataset was sourced from Kaggle and includes fields such as `order_id`, `date`, `item_name`, `item_price`, `quantity`, `transaction_amount`, `transaction_type`, and `time_of_sale`.

- The date column was converted to standard datetime format to enable time-based operations and forecasting.
- All numerical columns were checked for incorrect or outlier values using `describe()` functions.
- Missing values in the `transaction_type` column (107 missing entries) were handled appropriately — filled based on surrounding values.
- Duplicates and inconsistent entries were also inspected and dropped to avoid skewing the results.

Accurate and clean data reduces the risk of misleading insights and ensures the validity of statistical and predictive models. It's the foundation for dependable time series and EDA.

4.2 Analysis Process:

Following data preparation, the analysis proceeded through several structured stages.

- Descriptive Statistics were calculated for key numerical variables like item price, quantity, and transaction amount to understand the central tendencies and data spread.
- Item Sales Analysis used aggregation techniques such as `groupby()` to compute the frequency, quantity sold, and revenue generated by each item.
- Time-Based Sales Trends were explored by resampling the dataset by day, week, and month using Python libraries such as `pandas` and `matplotlib`. This approach enabled the identification of temporal patterns and seasonal effects.

- Customer Behavior Analysis examined how sales varied across time-of-day slots and by the gender of the staff processing orders. This involved creating pivot tables and heatmaps for insightful visualizations.
- Demand Forecasting was performed using Prophet, a time series forecasting model developed by Facebook, suitable for handling trend and seasonality in business time series data. The model fits historical daily sales data (y) over time (ds) and generates forecasts for future periods, including upper and lower confidence intervals.

Each method was chosen for its suitability in uncovering actionable insights. Aggregation and group-by operations were ideal for summarizing large volumes of transactional data, while time series models like Prophet are robust for sales forecasting due to their ability to model nonlinear growth and seasonal effects. These approaches are directly aligned with our business problem of understanding item popularity and predicting future demand to improve decision-making, inventory control, and customer satisfaction.

5. Results and Findings

This section presents the key results obtained through descriptive, exploratory, and time-series analyses of the fast food sales dataset. Various visualizations and statistical techniques were used to uncover patterns, trends, and anomalies in customer purchasing behavior and item-wise performance.

5.1 Item Sales Analysis

Understanding which items drive the most revenue and customer interest is vital for any fast-food business aiming to optimize inventory, pricing, and promotional strategies. This section presents a detailed analysis of sales performance at the item level using historical transaction data. The analysis focuses on identifying:

- The top-selling items by quantity
- The highest revenue-generating items
- The most frequently ordered items

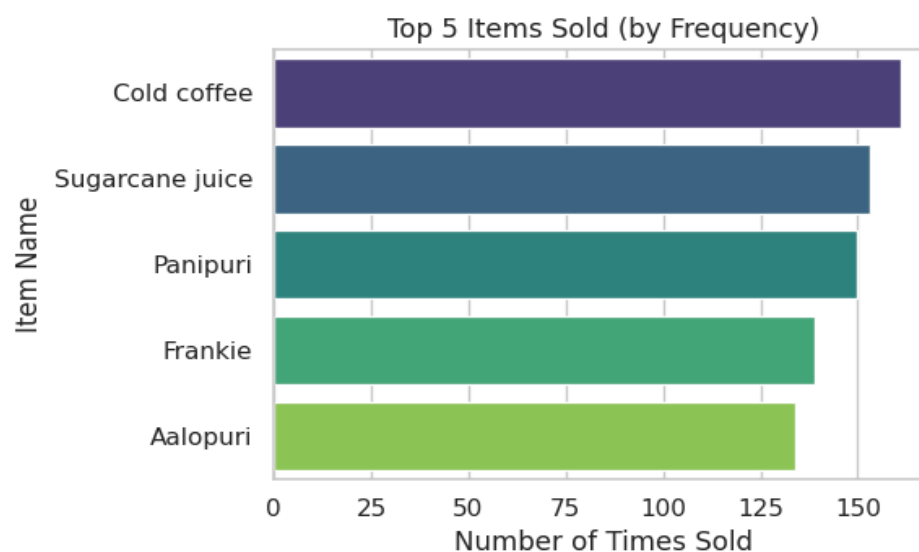


Figure 4: Items sold by frequency

Figure 4 presents a horizontal bar graph showcasing the five most frequently sold items. Cold Coffee leads in sales, followed closely by Sugarcane Juice and Panipuri, with Frankie and Aalopuri slightly behind. This indicates a strong customer preference for cold beverages and popular street snacks.

These insights are valuable for business decisions. High-selling items should be prioritized in inventory and promotions, while Frankie and Aalopuri could benefit from targeted marketing to boost sales. This sales gap suggests an opportunity to increase visibility of lower selling items through bundling strategies or repositioning on the menu.

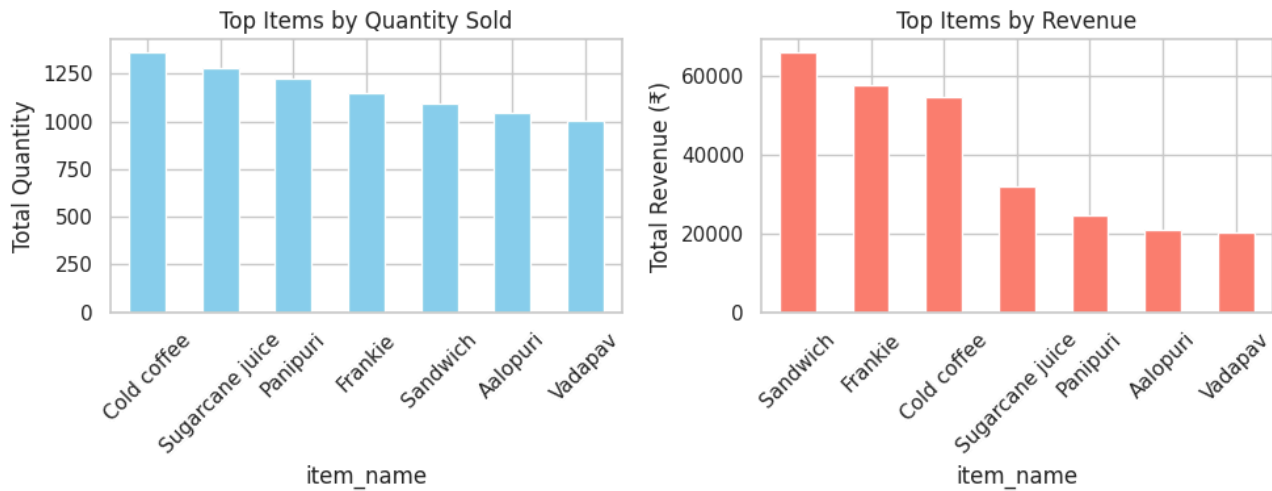


Figure 5: Top items by quantity sold and top items by Revenue

Figure 4 shows the top 5 items by frequency sold, where Cold Coffee, Sugarcane Juice, and Panipuri emerge as the most frequently purchased items. Figure 5 (left), which highlights the total quantity sold, reinforces this pattern—Cold Coffee leads again, followed closely by Sugarcane Juice and Panipuri. This indicates that these items are consistent crowd favorites and are likely purchased in high volumes on a regular basis. Other items like Frankie, Aalopuri, Sandwich, and Vadapav also appear but in slightly lower quantities.

However, Figure 5 (right), which focuses on total revenue, presents a different picture. Sandwich ranks highest in revenue generation, despite not being the most frequently purchased. This suggests that sandwiches are priced higher, leading to greater earnings per sale. Frankie also performs well in terms of revenue, indicating strong profitability. Interestingly, while Sugarcane Juice and Panipuri are sold in large quantities, they generate comparatively less revenue, likely due to their lower price points.

Considering all three graphs, the top three overall selling items are Cold Coffee, Sugarcane Juice, and Frankie. Cold Coffee and Sugarcane Juice dominate in terms of customer demand and sales volume, while Frankie stands out for its high revenue and steady sales.

5.2 Time-Based Sales Trends

Time-based sales analysis is essential for understanding customer behavior, identifying peak business hours, and recognizing fluctuations in demand. By evaluating daily, weekly, and monthly sales performance businesses can enhance inventory management, staff allocation, and promotional planning. These insights support accurate forecasting and more efficient operations.

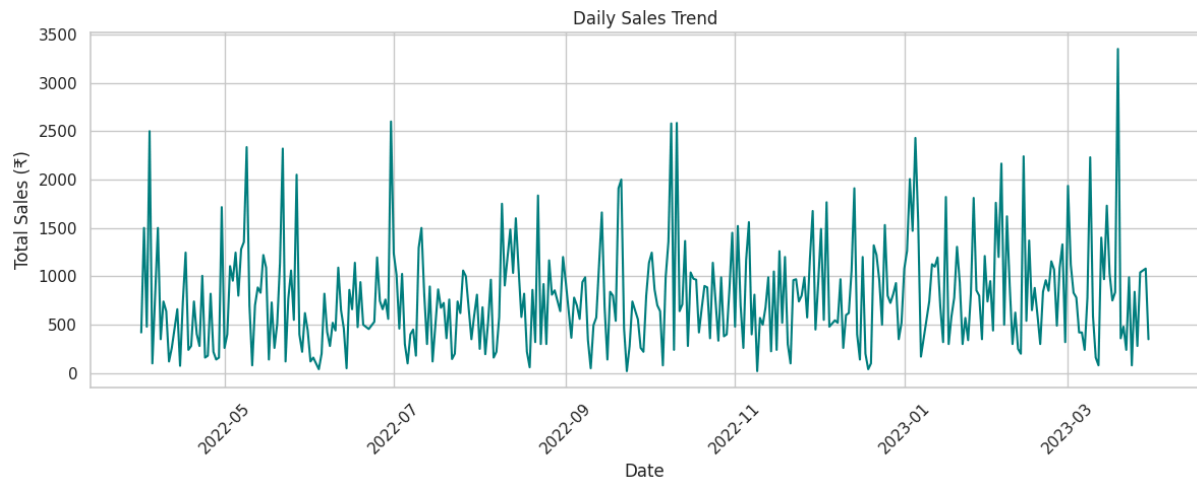


Figure 6: Daily sales trend

Figure 6, the daily sales trend line graph, exhibits frequent peaks and dips, reflecting the natural variability in day-to-day operations. Such fluctuations are often driven by external factors like weekdays vs. weekends, weather changes, holidays, or promotional events. Despite the volatility, specific spikes indicate high-demand days worth exploring further.

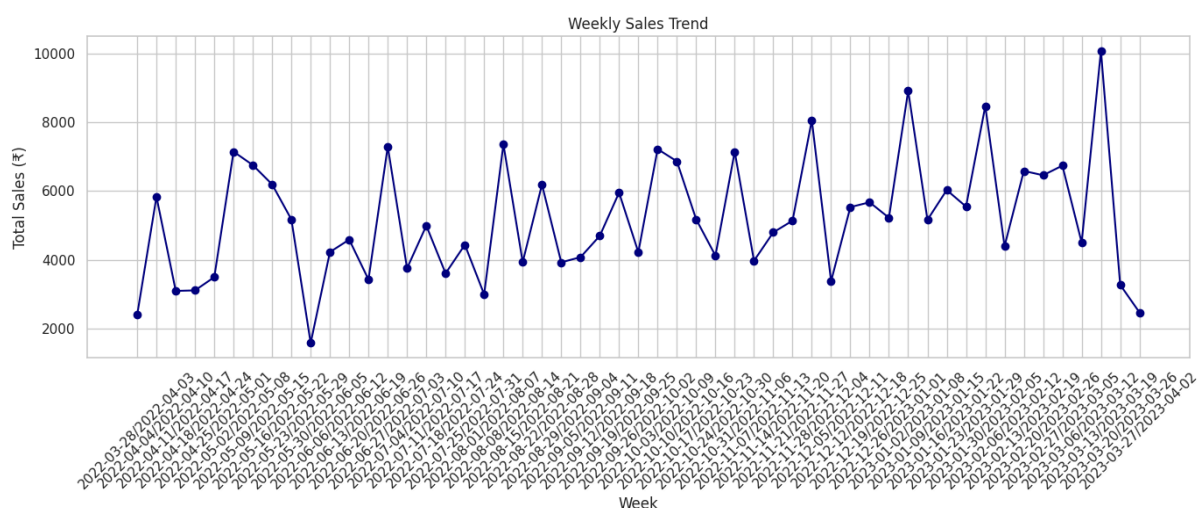


Figure 7: Weekly sales trend

Figure 7, which shows the weekly sales trend, smooths out daily noise and provides a clearer mid-term view. It reveals a moderately consistent upward trend, with some weeks performing

significantly better—potentially due to festivals or special promotions. This visualization helps in understanding customer behavior at a broader scale and aids in scheduling resources.

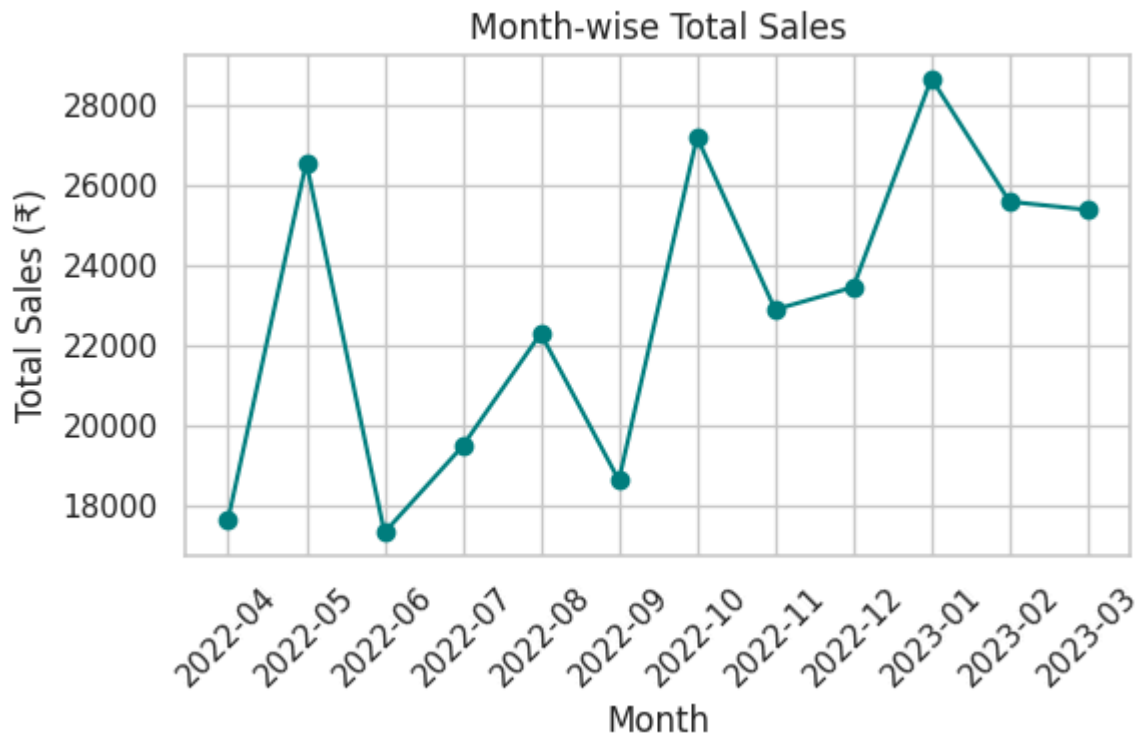


Figure 8: Monthly sales trend

The month-wise chart provides a broader overview in figure 8. It shows a steady growth pattern over time, with sales generally increasing from mid-2022 to early 2023. Notable peaks in October 2022 and January 2023 suggest high-demand periods—possibly festive seasons or successful campaigns. This line graph is particularly useful for strategic planning and long-term decision-making, as it highlights sustained performance trends and seasonal impacts.

5.3 Customer Behavior Patterns

Understanding customer behavior is essential for optimizing sales strategies, enhancing service delivery, and improving overall business performance. This section analyzes how customer purchase patterns vary based on factors such as time of day and gender of the staff attending the customer.

By exploring behavioral patterns, we aim to uncover when customers are most active, what influences their buying decisions, and how internal factors (like staff interaction) or external conditions (like time of day) impact transaction amounts. These insights can support more personalized marketing, staff deployment, and operational planning.

1. Time of Sale vs. Revenue

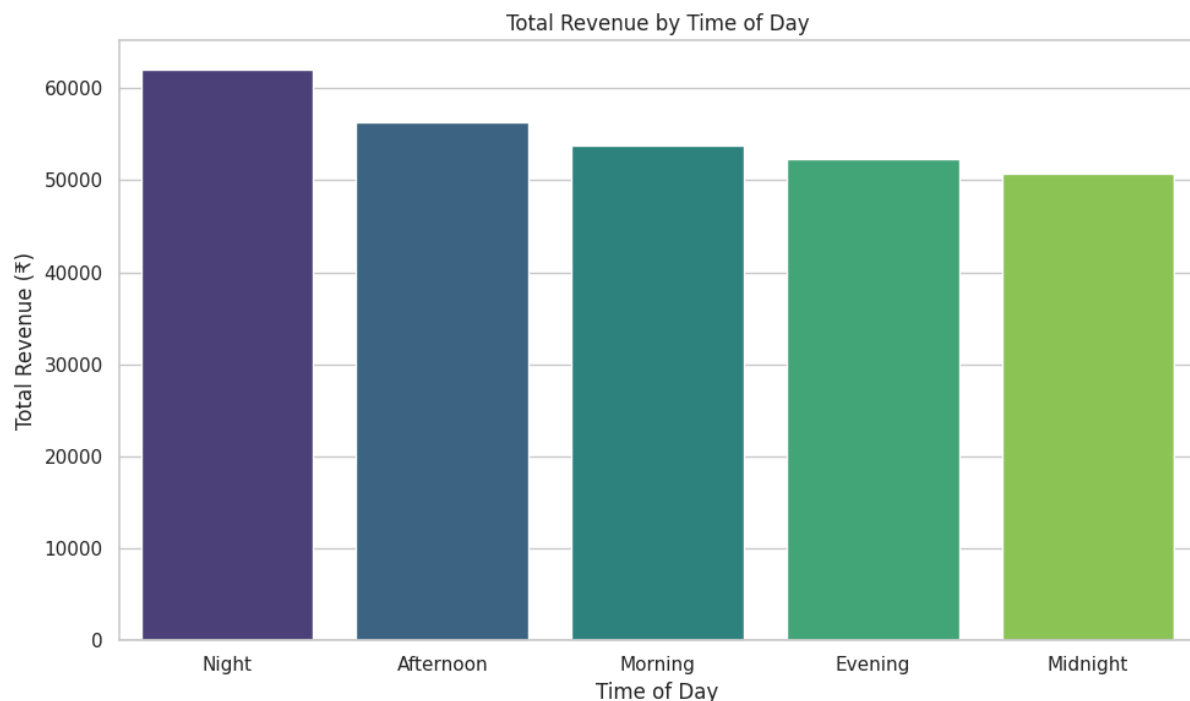


Figure 9: Total revenue by time of day

The chart shows that Night (7pm - 11pm) generates the highest revenue among all time slots, followed by Afternoon and Morning. This indicates that customers prefer purchasing more frequently or high-value items during the night, suggesting it is the busiest and most profitable time. Evening and Midnight also contribute significantly, but slightly less.

This suggests that late-day sales are driven by customer availability and cravings, making this period critical for inventory and staffing. A bar chart is ideal here as it clearly compares discrete time periods and effectively highlights the revenue differences across each time slot.

Beyond time-of-day trends, internal factors like staff gender also influence sales. The following charts explore this further.

2. Gender-wise Staff Analysis

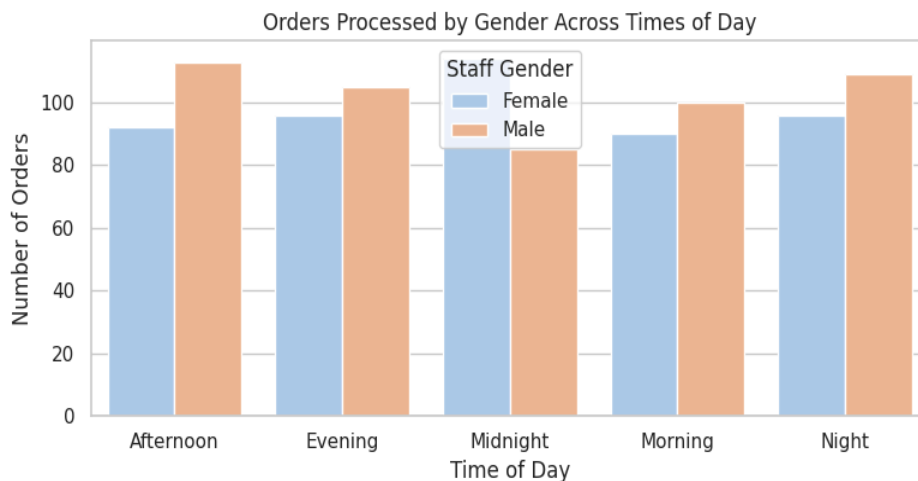


Figure 10: Order processed by gender across times of day

The analysis of orders processed by gender (Figure 10) across different times of day reveals that male staff consistently handle a higher number of orders compared to their female counterparts. This trend is evident across all time slots, with the most significant differences observed in the Afternoon and Night. Interestingly, during Midnight hours, the number of orders processed by female staff nearly matches that of male staff, indicating a more balanced distribution of workload during that period. Let's look at the sales by gender performance.

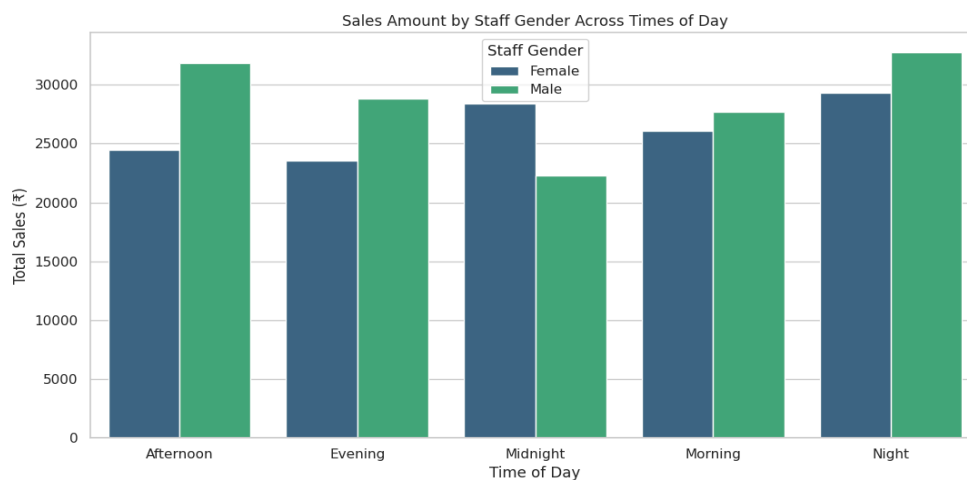


Figure 11: Sales Amount by gender across times of day

When examining the sales amount by gender across the same time slots (Figure 11), it becomes clear that male staff generally generate higher total sales. This trend aligns with their higher order volume in most time frames such as Afternoon, Evening, Morning, and Night. However, a notable exception occurs at Midnight, where female staff outperform male staff in terms of total sales despite processing a similar or slightly lower number of orders.

This suggests that female staff may be more effective at generating higher-value sales during this quieter period.

Together, these insights highlight key operational strengths of each gender group. Male staff appear to excel during high-traffic times, handling a greater volume of transactions, while female staff demonstrate potential for achieving better sales outcomes in lower-traffic periods, possibly through better customer interaction or upselling techniques. These patterns could inform smarter staff scheduling to maximize both efficiency and revenue, such as deploying more male staff during peak hours and utilizing female staff during times that demand more value-focused engagement.

5.4 Forecasting Future Demand

Knowing what customers are likely to order next week or month helps us stay one step ahead. By studying past sales records and applying forecasting models, we can better stock up on raw materials, ensure the right staff is available, and avoid over-ordering slow-moving items. This demand forecast allows smoother operations and smarter planning.

To better understand product performance, a detailed analysis of item-wise sales (Figure 4) was conducted. The following chart highlights the top 3 best-selling items, which significantly contribute to overall sales and customer satisfaction. To predict future demand, a forecast was implemented using Prophet, a robust time series model that helps identify trends and plan inventory more effectively.

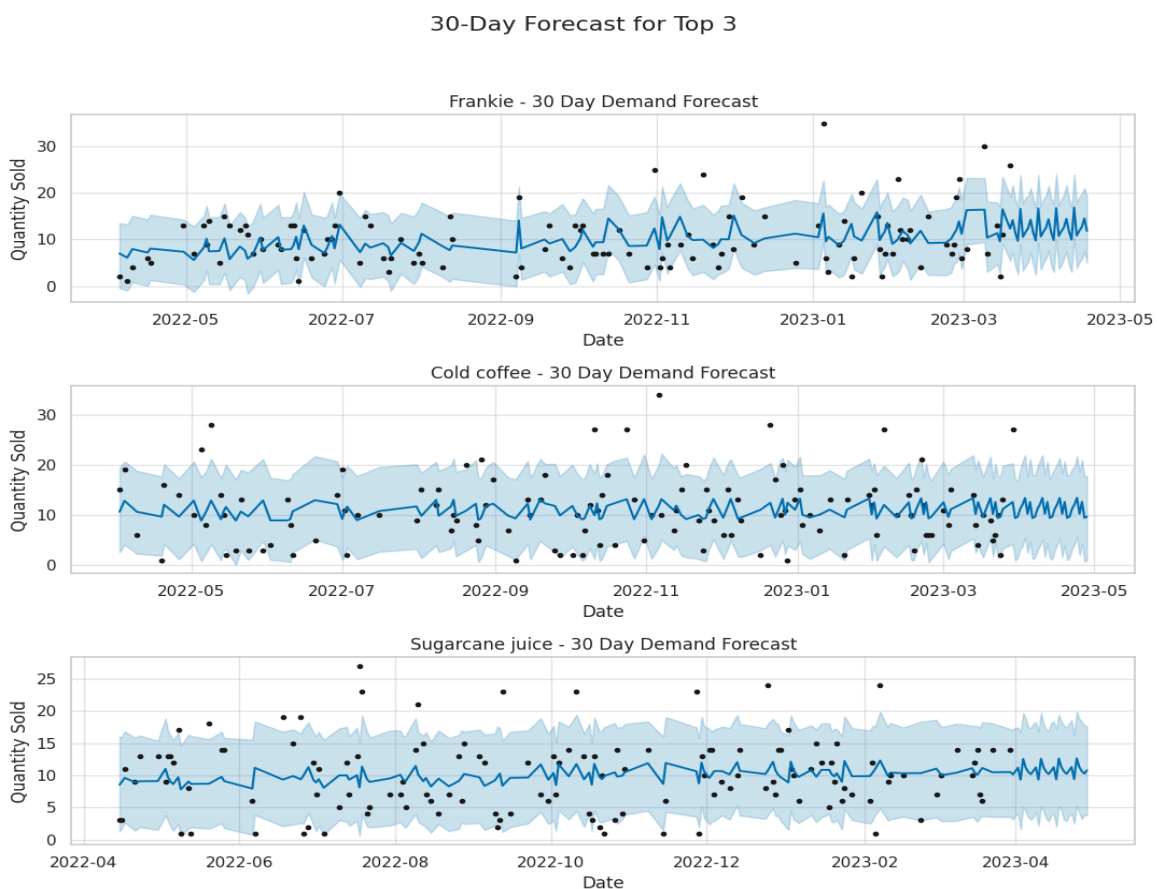


Figure 12: Demand forecast for top selling items

The forecast graph (Figure 12) shows projected demand trends for Frankie, Cold Coffee, and Sugarcane Juice using historical sales data.

In the Prophet forecast graph, black dots represent actual past sales, while the solid blue line shows the predicted demand. The light blue shaded area indicates the confidence range—wider means more uncertainty, narrower means more reliable predictions. These visual elements make it easy to understand both past performance and future expectations at a glance. The line graph is ideal for time series data as it clearly visualizes trends, fluctuations, and seasonal patterns over time.

Frankie shows a moderately increasing trend in demand over time with noticeable fluctuations. The forecast suggests that its popularity is growing, making it a strong candidate for inventory and promotional focus.

Cold Coffee exhibits high variation in daily demand but remains relatively stable on average. The confidence intervals are wide, indicating unpredictable spikes—likely influenced by weather or promotional factors.

Sugarcane Juice maintains a more consistent demand with lower volatility. The forecast suggests steady sales, making it a reliable product for stock planning.

6. Interpretation of Results and Recommendations

6.1 Interpretation of Results

The analysis of historical sales data reveals distinct patterns in customer preferences, item performance, and time-based trends. The top-selling items by quantity were Cold Coffee, Sugarcane Juice, and Panipuri, whereas Sandwiches and Frankies contributed more to revenue due to higher pricing. This indicates a difference between what sells the most and what earns the most, offering valuable insights for menu planning and profit optimization.

Time-based trends show that sales are highest during Night hours, followed by Afternoon and Morning, with notable monthly peaks in October 2022 and January 2023, possibly due to seasonal demand or festive periods. These trends align with customer behavior insights, revealing that male staff handle more orders, but female staff generate higher sales during low-traffic periods like midnight—highlighting opportunities for optimized staff deployment.

Demand forecasting using Prophet for top-selling items showed that Frankie demand is steadily rising, Cold Coffee has fluctuating but high demand, and Sugarcane Juice shows steady, consistent sales. This empowers the business to plan raw material purchases, manage inventory, and run promotions more effectively.

6.2 Recommendations

1. Inventory Prioritization

Focus procurement and stocking efforts on Cold Coffee, Sugarcane Juice, and Frankie, as these items consistently lead in sales volume. Ensuring their availability during peak hours helps prevent stockouts and maintains customer satisfaction.

2. Menu Optimization

Capitalize on high-margin items such as Sandwiches and Frankie by introducing combo offers, upselling strategies, and value-based pricing. Promote these items more prominently on the menu to enhance visibility and maximize profits.

3. Menu Planning Enhancement

Use sales insights to refine the menu—retain high-performing items, consider phasing out low-selling items, or reposition them through bundled deals. Introduce seasonal items based on monthly trends and explore limited-time offers to drive interest and test new products with minimal risk.

4. Staff Scheduling Optimization

Utilize sales-by-time and gender analysis to improve staff deployment. Assign more male staff during high-traffic hours (Afternoon and Night), where they handle larger volumes effectively, and schedule female staff during low-traffic periods like Midnight, where they demonstrate higher average order value performance.

5. Campaign Timing and Promotions

Align marketing and promotional campaigns with historically high-demand months such as October and January. Launch seasonal offers, loyalty rewards, or discounts during these months to increase footfall and boost overall revenue.

Implementing these recommendations will lead to improved inventory turnover, higher profitability, and enhanced customer satisfaction. The business can reduce operational waste, optimize staffing, and ensure that popular items are always available—ultimately leading to better sales forecasting, strategic decision-making, and sustainable growth.