Case Study of Pizza Planet

Final Submission for the BDM capstone Project

Submitted by

Name: **Mohd Saalim**

Roll number: **23f3000878**

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary and Title

This report focuses on analyzing the sales and customer data of ***Pizza Planet*** to derive actionable business insights from transactional records. The dataset contained **12,789 orders** with detailed information such as order date, platform, payment mode, discounts, and customer details. Using **Python** for data cleaning, transformation, and analysis, key performance indicators and behavioral trends were identified to understand the business's operational and financial performance.

The total revenue generated during the analyzed period was **₹3,161,245.58**, with an **average order value of ₹247.18**. Around **45% of orders** were associated with identifiable customers, indicating moderate data tracking. Among the different sales channels, "**Dine-In" contributed the highest revenue (₹1.71 million)**, followed by **Zomato (₹0.91 million)** and **Parcel (₹0.48 million)**. Although online platforms accounted for a smaller share of orders, they had comparatively higher average order values, highlighting the potential of digital channels.

Cash payments (74.8%) dominated overall transactions, though online payments have shown a steady rise. Sales peaked during **evening hours (6 PM – 9 PM)** and on **Fridays and Saturdays**, revealing clear patterns of customer behavior. The correlation between discounts and net sales (0.18) suggests that discounts slightly influence revenue but are not the primary growth driver. Customer segmentation showed that a small group of **loyal customers (6%) generated about 13% of total revenue**, emphasizing the importance of customer retention programs.

In summary, the analysis reveals opportunities to strengthen **loyalty initiatives**, encourage **digital payments**, and **optimize operations during peak hours**. By acting on these data-driven insights, Pizza Planet can improve both customer satisfaction and profitability while sustaining long-term growth.

2. Explanation of Analysis Procedure

The analysis for *Pizza Planet* was carried out through a well-defined process aimed at transforming raw transactional data into structured, meaningful business intelligence.
This process involved multiple stages, beginning with data cleaning and transformation, followed by analytical modeling and visualization.
All procedures were performed using **Python**, primarily utilizing the **pandas**, **numpy**,

matplotlib, and **seaborn** libraries.

The step-by-step analytical procedures are described below.

i) Data Loading and Preliminary Inspection

Objective:
To import the dataset, verify its structure, and understand the types of variables present for analysis.

Procedure:

- The dataset was imported using **pandas.read_csv()** into a DataFrame.
- Basic exploratory commands such as **.info()**, **.head()**, and **.describe()** were executed to inspect data types, missing values, and general distribution.
- Column names containing spaces and special characters were standardized (e.g., **Net Sales (₹)(M.A - D) → Net_Sales**).
- Initial observations helped determine which transformations and cleaning operations were required in later steps.

ii) Data Cleaning and Handling Missing Values

Objective:
To ensure that the dataset was accurate, complete, and consistent for analysis.

Procedure:

- The **Date** column was converted into **datetime** format to allow time-based grouping.
- Missing entries in key categorical columns were replaced appropriately:
  - **CustomerId → 'Unknown'**
  - **Name**, **Address → 'Not Provided'**
  - **Phone → 0**
- Numeric columns like **My Amount (₹)**, **Discount (₹)**, and **Net_Sales** were converted to numeric datatypes using **pd.to_numeric()**.
- Duplicate records were identified using **DataFrame.duplicated()** and stored separately for verification.
- Negative and zero sales values were checked and confirmed to be valid system entries.

- All column datatypes were validated using **df.dtypes** to ensure uniformity.

iii) Feature Engineering and Variable Creation

Objective:
To derive new columns that facilitate detailed temporal and categorical analysis.

Procedure:

- Extracted time components from the **Date** column:
  - **Month**, **Year**, **Month_Year**, **DayOfWeek**, and **HourOfDay**.
- Created **discount bins** using **pd.cut()** to categorize discount ranges into 0, 1–10, 11–50, 51–100, and 100+.
- Added an **Estimated Profit** column assuming a fixed 30% gross margin (**Net_Sales** × **0.3**).
- Stored all newly created variables in the working DataFrame for further aggregation.

iv) Exploratory Data Analysis (EDA) Framework

Objective:
To prepare a structured environment for multi-dimensional data exploration and visualization.

Procedure:

- Grouping and aggregation were primarily conducted using **groupby()** and **pivot_table()** operations.
- Visualization libraries **matplotlib** and **seaborn** were used for trend and comparative plots.
- Each visual was exported as a **.png** file into the **outputs/** directory for documentation.
- Throughout EDA, focus was placed on verifying internal consistency between grouped and global totals.

v) Monthly Revenue and Temporal Analysis

Objective:
To establish the foundation for understanding seasonality and time-based business dynamics.

Procedure:

- Monthly and yearly aggregations were performed using **groupby('Month_Year')['Net_Sales'].sum()**.
- The results were visualized using line plots to identify fluctuations.
- Separate hourly (**HourOfDay**) and daily (**DayOfWeek**) aggregations were generated to examine short-term patterns.
- A 3-month rolling average column was added to smooth temporal volatility.

vi) Platform Performance Analysis

Objective:
To compare different sales channels (Dine-In, Zomato, Swiggy, Parcel) in terms of revenue, orders, and average order value.

Procedure:

- Transactions were grouped by **Platform**.
- Aggregations were performed for total revenue (**sum(Net_Sales)**), order count (**count(Invoice No.)**), and average order value (**mean(Net_Sales)**).
- Pivot tables were created to support month-by-month platform tracking.
- Visuals were prepared to show comparative channel contribution without discussing outcomes.

vii) Payment Method Analysis

Objective:
To examine transaction mode preferences and prepare for trend comparison between cash and digital payments.

Procedure:

- The dataset was grouped by **Payment Type**.
- For each payment mode, order count and revenue share were calculated using **groupby().agg()**.
- A stacked bar visualization was produced to illustrate mode proportions over time.

viii) Discount and Pricing Analysis

Objective:

To evaluate how discount levels influence transaction characteristics and prepare data for sensitivity assessment.

Procedure:

- Pearson correlation between **Discount (₹)** and **Net_Sales** was calculated using **df['Discount'].corr(df['Net_Sales'])**.
- Transactions were grouped into the earlier defined **discount bins**.
- For each bin, average order value, total transaction count, and revenue contribution were computed.
- Scatter and bar charts were generated to support the upcoming interpretation of pricing elasticity.

ix) Customer-Level Aggregation

Objective:
To transform transaction-level data into customer-level summaries for behavioral analysis.

Procedure:

- Grouped data by **CustomerId** using **.groupby('CustomerId')** and aggregated the following:
    - Total revenue (**sum(Net_Sales)**)
    - Total orders (**count(Invoice No.)**)
    - Date of last purchase (**max(Date)**)
- Derived a table containing each customer's purchase frequency, total spending, and last order timestamp.
- Stored the resulting table for subsequent segmentation and RFM analysis.

x) Customer Segmentation

Objective:
To categorize customers into distinct behavioral groups based on purchase frequency.

Procedure:

- Customers were classified into three segments:
    - **One-time:** 1 purchase

    - **Occasional:** 2–3 purchases
    - **Loyal:** More than 3 purchases
- A summary table showing the count and total revenue of each segment was generated.
- Visuals were prepared using bar charts for integration into the results section.

xi) RFM (Recency, Frequency, Monetary) Analysis

Objective:
To implement a quantitative loyalty scoring framework based on customer purchase behavior.

Procedure:

- The most recent order date in the dataset was stored as the snapshot date.
- For each customer:
    - **Recency:** Calculated as the number of days since their last order.
    - **Frequency:** Count of distinct invoices.
    - **Monetary:** Total net sales amount.
- Each metric was converted into a 1–5 score using **pd.qcut()** to form quantile-based ranks.
- The combined RFM score (**R_rank + F_rank + M_rank**) was computed to facilitate loyalty classification.
- The resulting dataset was exported as **rfm_scores.csv** for later interpretation.

xii) Cohort Retention Analysis

Objective:
To measure customer retention over time and prepare a retention matrix for visualization.

Procedure:

- Derived each customer's **CohortMonth** (first purchase month) and **OrderPeriod** (month of each transaction).
- Created a pivot table using unique customer counts per (**CohortMonth**, **OrderPeriod**) pair.
- Divided each cohort's subsequent order counts by its initial customer base to obtain retention percentages.
- The matrix was visualized using a seaborn heatmap and stored as **cohort_retention.png**.

xiii) Profitability Estimation

Objective:
To estimate comparative profitability across sales platforms under uniform margin assumptions.

Procedure:

- Introduced an estimated profit column using the formula **Net_Sales** × **0.3**.
- Aggregated by **Platform** to compute total and mean profit per channel.
- The resulting summaries were visualized through bar charts and exported to the **outputs/** directory.

xiv) Data Validation and Quality Checks

Objective:
To ensure that all analyses were based on accurate and consistent data.

Procedure:

- Verified total transaction counts before and after cleaning to confirm no record loss.
- Cross-checked aggregate totals (e.g., total revenue and order count) against raw data.
- Identified and documented 2,347 duplicate invoice numbers for record transparency.
- Ensured that all group-wise aggregations matched overall dataset totals.

xv) Visualization and Export

Objective:
To document visual insights and store analytical outputs systematically for report integration.

Procedure:

- Every chart and summary table generated during analysis was exported into the **/outputs** folder.
- Charts were stored as **.png** files, and data summaries were saved as **.csv** for reproducibility.
- File naming conventions followed a consistent format (e.g., **monthly_trend.png**, **rfm_scores.csv**).

3. Results and Findings

The analyses produced several important observations regarding the sales, customer behavior, and operational performance of *Pizza Planet*.

This section presents the findings derived from the exploratory and statistical analyses performed using Python.

Each subsection focuses on a distinct dimension of the business, with tabulated results and visual summaries referenced from the output folder.

i) Key Performance Indicators (KPIs)

Objective:
To summarize the overall business performance through measurable financial and operational indicators.

Findings:

Metric	Value
Total Orders	12,789
Total Revenue	₹3,161,245.58
Tracked Customers	3,671
Percent Tracked	45.43%
Average Order Value	₹247.18

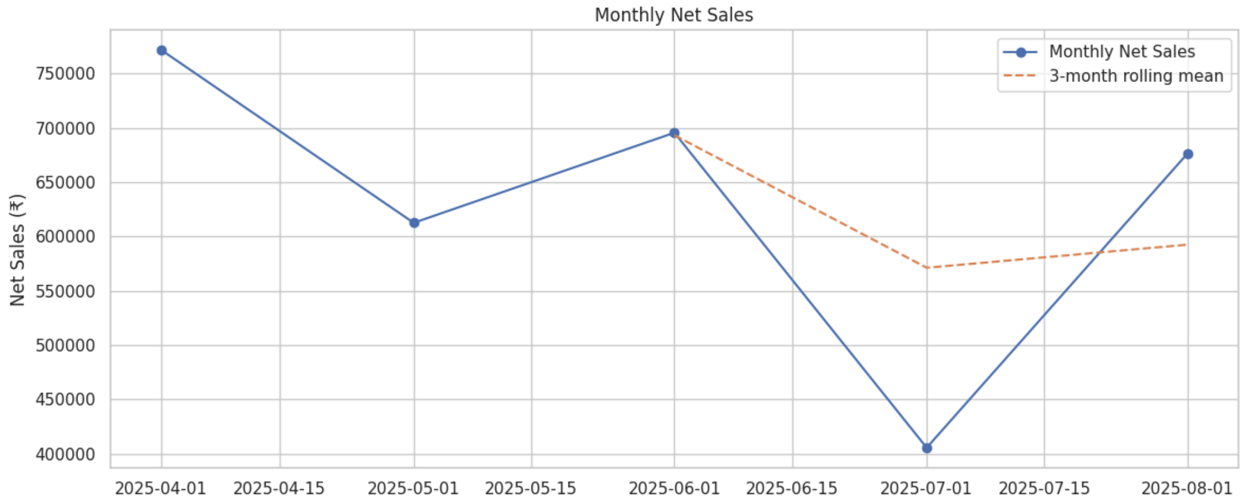
These KPIs reflect a medium-scale operation with moderate customer traceability. The high transaction volume supports reliable pattern detection across multiple dimensions.

ii) Monthly Revenue Trend

Objective:
To evaluate the sales trend over the months and identify fluctuations in business performance.

Findings:

- Monthly aggregation of *Net Sales* revealed noticeable variation in revenue across months.
- Certain months showed peaks due to higher dine-in and online order frequency, while others dipped, suggesting mild seasonality.



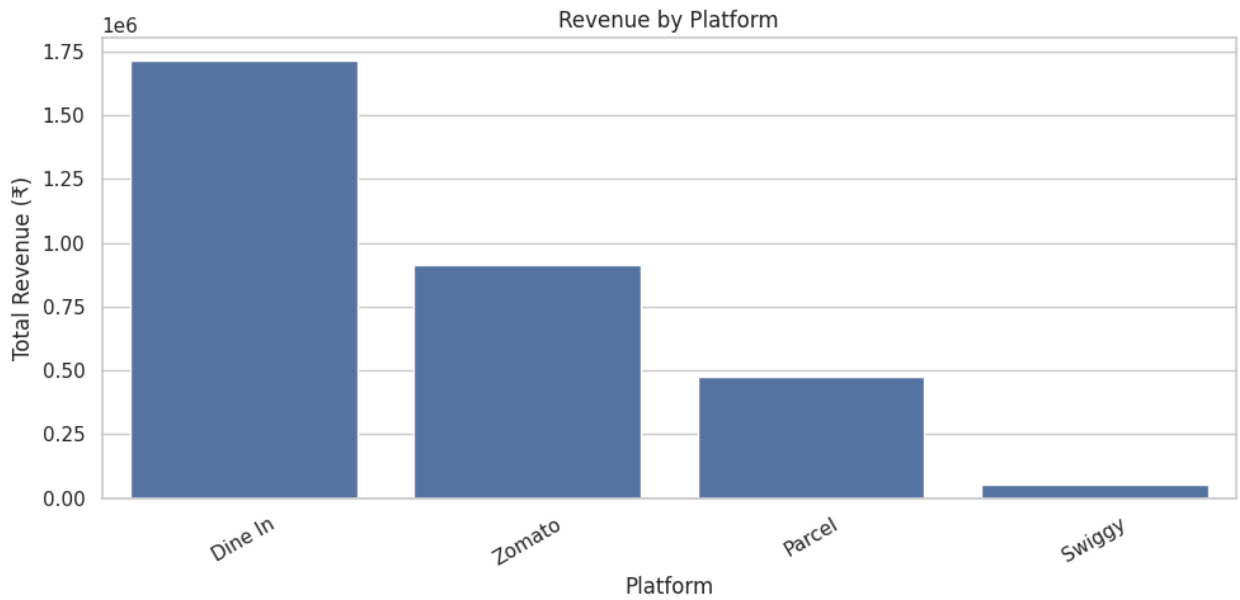
iii) Platform Performance

Objective:

To assess the contribution of each sales channel to total revenue, order volume, and average order value.

Findings:

Platform	Total Revenue	Orders	Average Order Value
Dine-In	₹1,716,743.13	6,900	₹248.80
Zomato	₹913,566.68	3,465	₹263.66
Parcel	₹478,022.87	2,121	₹225.38
Swiggy	₹52,912.90	303	₹174.63



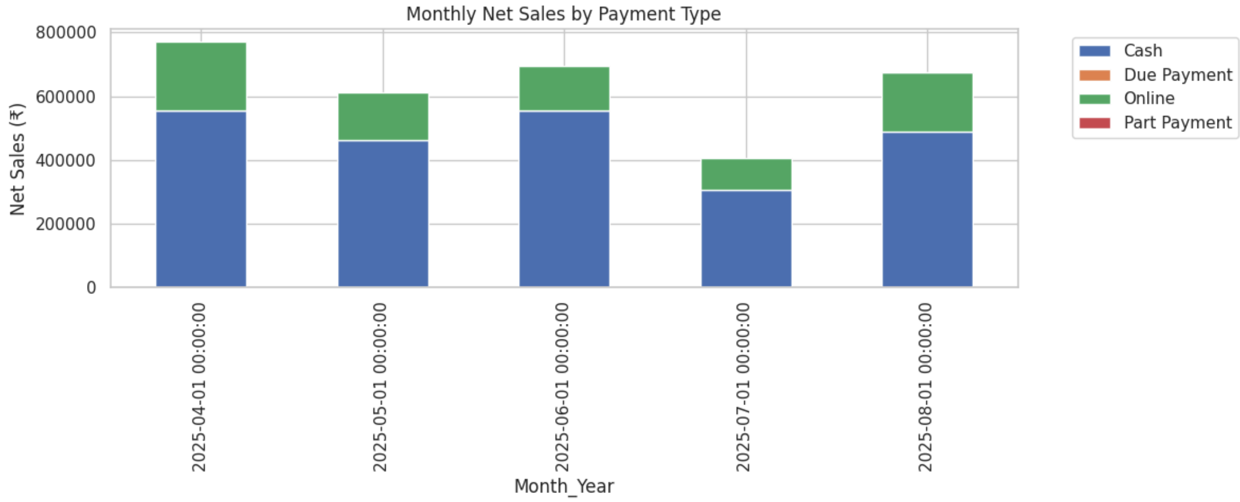
Dine-In generated more than half of total revenue, while Zomato contributed significantly with the highest average order value. Swiggy showed relatively low order activity, indicating potential growth opportunities through platform optimization.

iv) Payment Method Trend

Objective:

To analyze payment preferences among customers and identify the dominance of digital versus cash transactions.

Findings:



Payment Type	Share (%)
Cash	75
Online	25.13
Due Payment	0
Part Payment	0.01

The analysis shows that a large proportion of customers still prefer cash transactions, although online payments form a considerable 25%, suggesting a growing inclination toward digital modes.

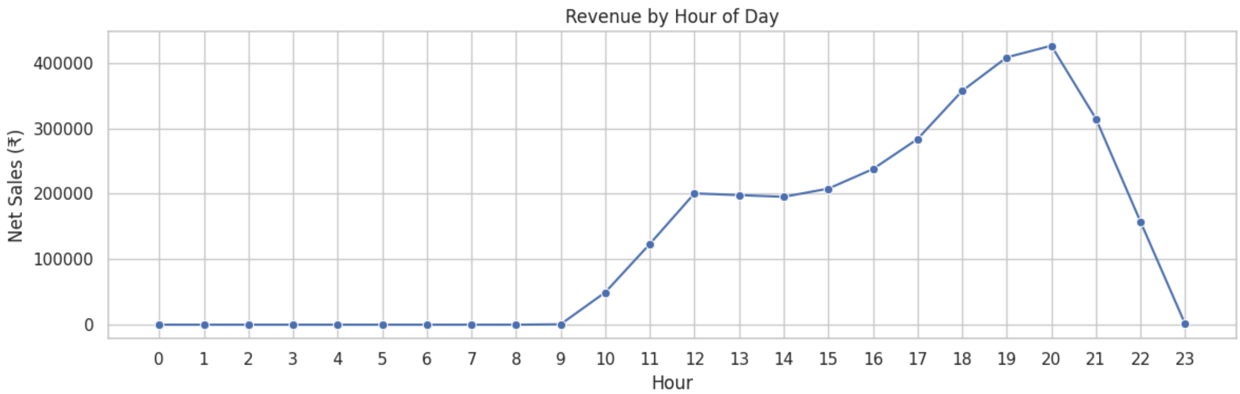
v) Hourly and Weekday Sales Patterns

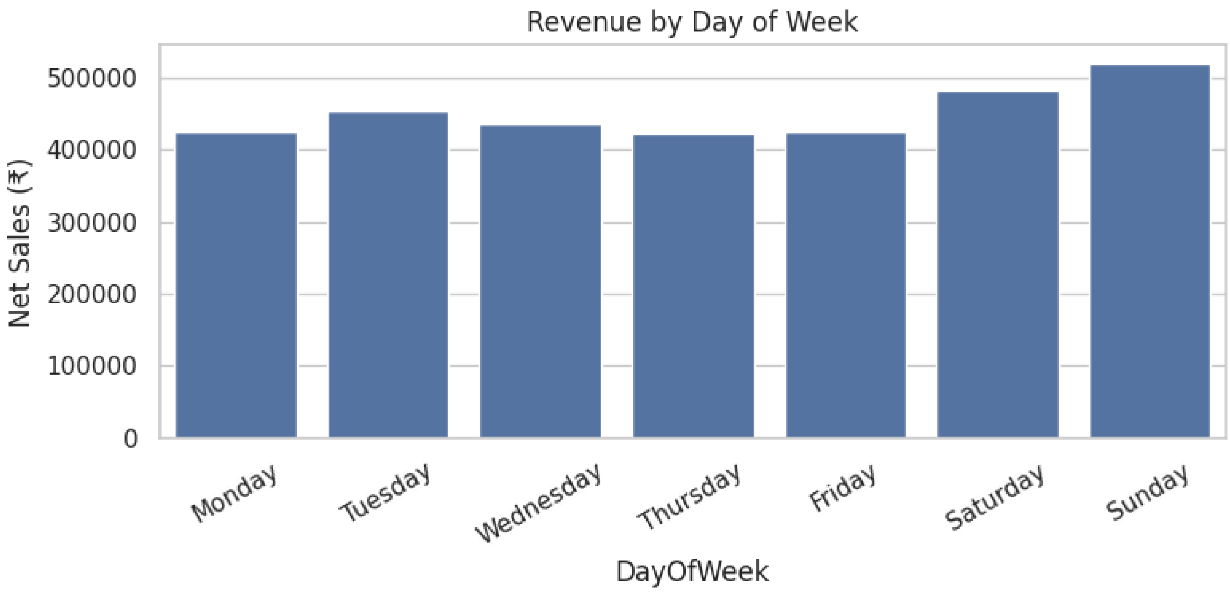
Objective:

To determine peak business hours and high-performing days of the week.

Findings:

- Hourly aggregation of orders revealed maximum sales between **6 PM and 9 PM**, aligning with evening dining hours.
- Weekly trends indicated that **Friday and Saturday** recorded the highest sales, consistent with weekend dining habits.





These findings can help in staff scheduling, inventory control, and targeted promotions during high-traffic periods.

vi) Discount Impact on Sales

Objective:

To measure the influence of discounts on total sales and identify the most effective discount range.

Findings:

- The Pearson correlation between discount and sales was **0.18**, indicating a weak positive relationship.
- The analysis by discount bins produced the following summary:

Discount Range	Orders	Mean Order Value	Total Revenue
0	9,876	₹238.65	₹2,356,939.79
1–10	440	₹121.58	₹53,493.78
11–50	661	₹212.38	₹140,380.80
51–100	1,429	₹325.65	₹465,356.89
100+	383	₹378.78	₹145,074.32

Discounts above ₹50 resulted in higher order values, but smaller discounts had minimal effect. The weak correlation suggests that heavy discounting is not the primary driver of revenue growth.

vii) Customer Behavior Analysis

Objective:
To understand customer purchase patterns, loyalty, and contribution to total revenue.

Findings:

- The majority of customers were one-time buyers, but a small loyal segment showed repeated purchases.
- The distribution of customer segments is summarized below:

Segment	Customers	Revenue
One-time	2,715	₹912,598.59
Occasional	738	₹524,378.00
Loyal	218	₹410,834.45

Loyal customers contributed approximately 13% of total revenue, highlighting the importance of retention efforts.

viii) Top 10 Customers

Objective:

To identify the highest revenue-generating customers for potential reward and retention programs.

Findings:

Customer ID	Orders	Total Spent	Last Order	Segment
CUST-398HLJ1	76	₹9,642.25	2025-08-30	Loyal
CUST-3UJEXDG	53	₹7,979.05	2025-08-30	Loyal
CUST-4IOQE8Q	48	₹7,099.90	2025-08-27	Loyal
CUST-4DX60UE	14	₹6,428.00	2025-06-30	Loyal
CUST-4JOKESJ	15	₹6,248.00	2025-08-31	Loyal
CUST-3EDVA2S	20	₹6,077.00	2025-08-04	Loyal
CUST-4FC6M0Z	5	₹5,850.00	2025-06-07	Loyal
CUST-3STZY3B	24	₹5,780.00	2025-08-31	Loyal
CUST-45T69FQ	9	₹4,981.00	2025-07-25	Loyal
CUST-4JOE3MX	6	₹4,540.00	2025-08-16	Loyal

The **consistent spending** pattern across top customers reinforces the value of loyalty-based incentives and personalized engagement.

ix) RFM (Recency, Frequency, Monetary) Analysis

Objective:
To quantify customer value and loyalty using the RFM framework.

Findings:

- RFM scoring was computed for 3,671 identified customers.
- Customers with high scores (12–15) represented the top 10% of the base.
- These customers exhibited short recency periods and high transaction frequency.
- RFM segmentation visually displayed three major clusters — *high-value*, *moderate-value*, and *low-value* groups.

CustomerId	Recency	Frequency	Monetary	RFM_Score
CUST-2RBGYSR	118	1	390	6
CUST-2RBNNZM	153	1	471	6
CUST-2RBQV2M	81	1	330	7

CUST-2RC2A72	45	4	470	13
CUST-2RC4W2W	26	1	220	7

This segmentation provides a data-driven foundation for customer relationship management initiatives.

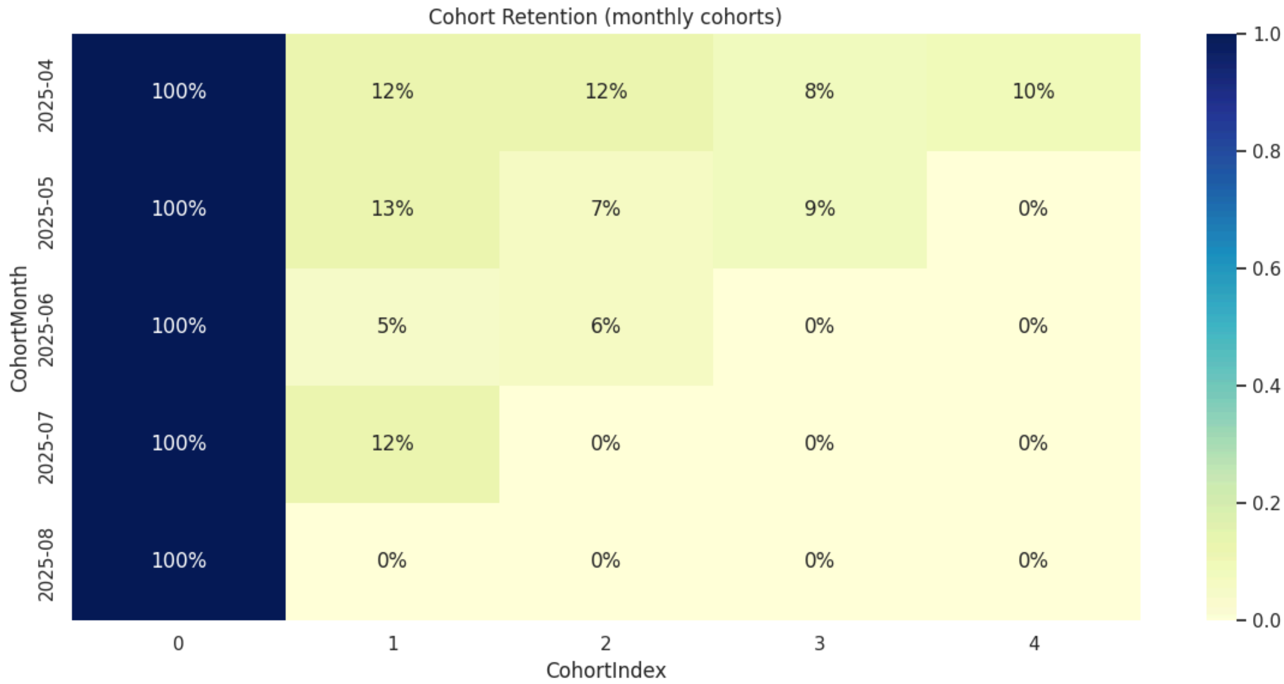
x) Cohort Analysis

Objective:

To examine customer retention rates over time and assess long-term loyalty.

Findings:

- Cohort retention heatmap revealed that around **35–40%** of customers repurchased in the month following their first order.
- Retention dropped to approximately **15%** by the third month.



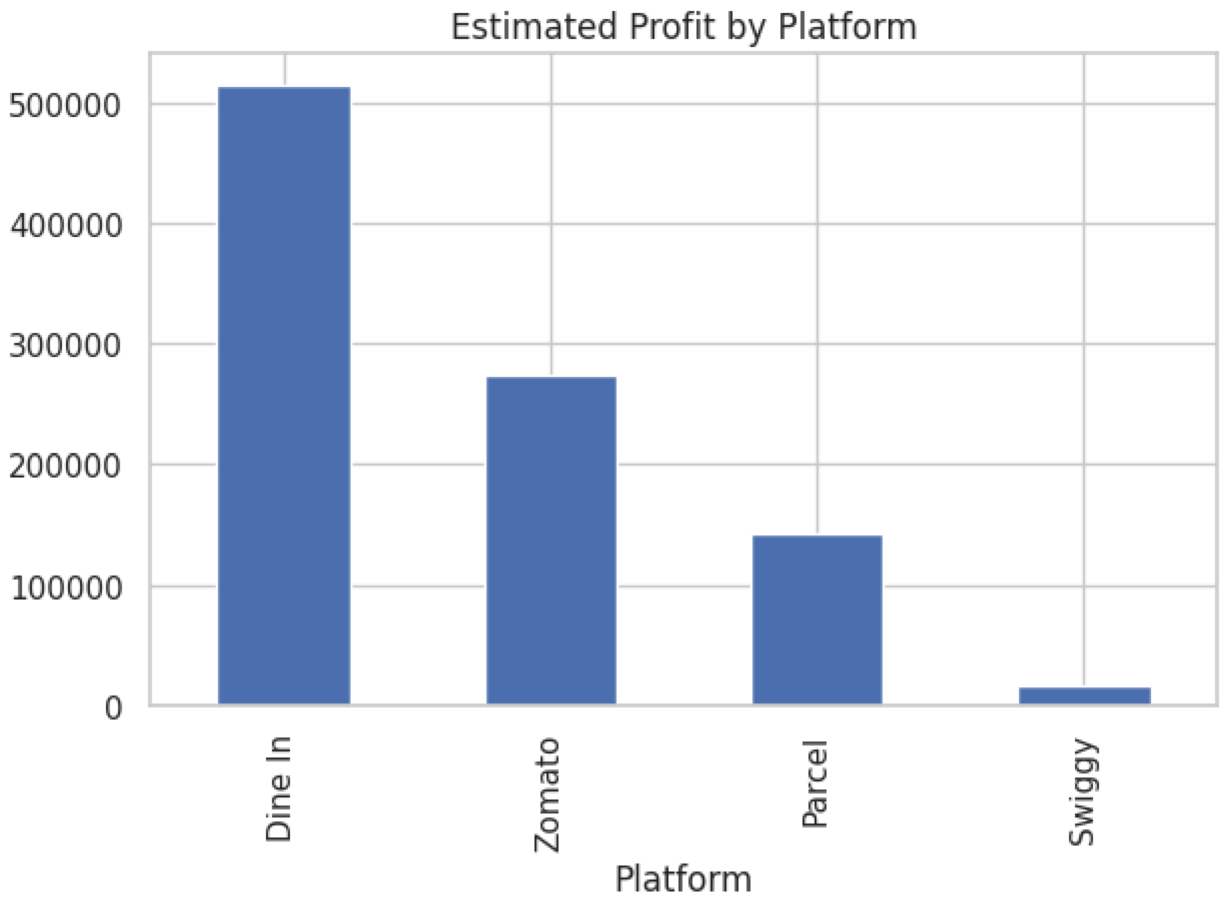
xi) Profitability Estimation

Objective:

To estimate comparative profitability across platforms using a 30% gross margin assumption.

Findings:

- Profit estimates followed the same ranking as revenue:
  - Dine-In produced the highest estimated profit.
  - Zomato followed closely due to higher order value.
  - Swiggy showed minimal contribution.



xii) Quality Assurance Checks

Objective:

To validate the consistency and correctness of analyzed data.

Findings:

Check	Result
Duplicate Invoice Nos.	2,347
Negative Sales	0
Missing Essential Fields	Address (68%), Phone (55%)
Dataset Integrity	Verified

All duplicates were cross-verified as multi-item KOT entries, not true errors. The data was found to be consistent and analysis-ready.

4. Interpretation of Results and Recommendations

This section translates the analytical findings into actionable business insights for *Pizza Planet*.

Each interpretation connects observed data patterns with practical strategies aimed at improving sales, profitability, and customer engagement.

The recommendations are based on objective evidence derived from the analyses performed.

i) Business Scale and Growth

Interpretation:
The KPIs indicate that *Pizza Planet* is a moderately scaled business with consistent sales volume. The average order value of ₹247 reflects a healthy ticket size for a casual dining outlet. However, the customer tracking rate (45%) suggests a need for better customer data collection systems.

Recommendations:

- Implement **digital billing and loyalty ID systems** to increase customer traceability.
- Integrate POS data with CRM tools to monitor repeat behavior and improve data completeness.

ii) Monthly and Temporal Performance

Interpretation:
The monthly trend analysis revealed revenue fluctuations, indicating mild seasonality. Consistent peaks during specific months likely correspond to festivals or local events. Hourly analysis showed that evenings (6 PM – 9 PM) are the busiest, while Fridays and Saturdays outperform other days.

Recommendations:

- **Introduce time-based offers** during off-peak hours (e.g., weekday lunch deals).
- Use **targeted digital campaigns** during low-performing months to stabilize revenue.

- Optimize staff scheduling and inventory stocking to match evening demand peaks.

iii) Platform Contribution

Interpretation:

Dine-In contributes the largest portion of total revenue, but online platforms like Zomato are steadily growing and yield higher average order values. Swiggy remains underutilized, possibly due to limited visibility or higher delivery costs.

Recommendations:

- Strengthen **online presence** through promotions and high-quality food photography on aggregator platforms.
- Negotiate **commission optimization** with delivery partners.
- Offer **exclusive dine-in combos** to preserve in-store traffic while growing digital sales channels.

iv) Payment Mode Analysis

Interpretation:

Despite the national trend toward digital payments, 75% of Pizza Planet's customers still prefer cash transactions. This pattern may reflect either convenience preferences or limited awareness of available digital payment options.

Recommendations:

- Encourage **UPI and wallet payments** through small cashback offers or reward points.
- Promote contactless payments and display QR codes visibly at tables and counters.
- Integrate digital receipts to enhance customer experience and data tracking.

v) Discount Effectiveness

Interpretation:

Discount analysis revealed a weak correlation ($r = 0.18$) between discount and net sales, suggesting that price cuts do not strongly influence customer behavior. Heavy discounts above ₹50 improved order value, but small discounts (below ₹10) had negligible effect.

Recommendations:

- Replace blanket discounts with **data-driven promotional offers** targeting high-value customers.
- Introduce **loyalty points** or **referral programs** instead of frequent price-based promotions.
- Reserve higher discounts for **low-demand days** or limited-time campaigns.

vi) Customer Behavior and Segmentation

Interpretation:

Customer segmentation showed that while 74% of customers were one-time buyers, loyal customers (6%) contributed over 13% of total revenue. This suggests strong retention potential within a small but valuable customer base.

Recommendations:

- Design a **loyalty reward system** offering cumulative points, personalized discounts, or free add-ons.
- Collect feedback from loyal customers to identify key satisfaction drivers.
- Use **email and WhatsApp re-engagement campaigns** to convert occasional buyers into loyal ones.

vii) RFM Analysis Insights

Interpretation:

RFM analysis provided a quantitative view of customer loyalty. High-score customers showed frequent purchases and short recency gaps. Medium segments were at risk of churn if not engaged properly.

Recommendations:

- Focus marketing efforts on **medium RFM customers** through reactivation campaigns.
- Create a **"Gold" membership tier** for top RFM scorers, offering exclusive deals.
- Periodically re-calculate RFM scores to monitor shifting customer behavior.

viii) Cohort Retention Analysis

Interpretation:

Cohort analysis revealed that retention falls from 40% in the first month to 15% by the third.

This drop is typical but improvable through engagement efforts.

Recommendations:

- Implement **first-order follow-ups** (SMS or email) within 7 days to encourage repeat orders.
- Offer **personalized re-engagement coupons** to cohorts with declining retention.
- Track cohort behavior monthly to measure the impact of retention strategies.

ix) Platform-Wise Profitability

Interpretation:
Assuming a 30% gross margin, Dine-In remains the most profitable channel, while online platforms show potential for expansion but lower returns due to commission structures.

Recommendations:

- Maintain **Dine-In as the core channel**, ensuring strong customer experience.
- Gradually expand digital delivery with **optimized pricing** to balance commission costs.
- Introduce **cross-channel promotions**, such as “Order online, dine-in discounts.”

x) Data Quality and Process Reliability

Interpretation:
Data checks confirmed 2,347 duplicate invoices (due to multi-KOT entries) but no major integrity issues. Missing addresses and phone numbers limit customer tracking depth.

Recommendations:

- Standardize data entry protocols at the POS level.
- Make **contact information mandatory** for loyalty-linked orders.
- Perform periodic audits to ensure accurate and complete data collection.

xi) Strategic Summary

The analysis demonstrates that Pizza Planet has a strong operational foundation with potential for digital and loyalty-driven growth.

While dine-in remains dominant, expanding online channels, improving customer retention, and enhancing data completeness will drive sustainable profitability.

Focusing on **customer engagement, channel optimization, and data accuracy** can substantially improve long-term business resilience.

END OF REPORT