Project Title: Evaluating Sales Decline and Management Enhancements in Pizza Cafe.

Project submitted by: Ashutosh Narayan.

Roll no: 21f1003987

BDM CAPSTONE PROJECT FINAL SUBMISSION

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

CONTENT PAGE NO.

Title & Executive Summary

Pizz The Cafe is a pizza cafe offering a variety of pizza types to its customers. As a B2C (business-to-consumer) operation, Pizz The Cafe provides its services directly to its customers.

The cafe offers a diverse range of products, making inventory and staff management a complex and challenging task. This complexity presents significant opportunities for improvement, as optimizing these areas can lead to more efficient operations, better stock control, and more effective staffing. By addressing these challenges, the cafe can enhance overall performance and better meet customer needs.

To address these challenges, I have analyzed the sales data using various metrics and variables. By employing different charts and visualization techniques, I was able to identify trends and patterns within the dataset. This analysis has provided valuable insights and helped formulate strategies to tackle the revenue decline and improve operational efficiency.

For this project, I have sourced my data from GitHub. Here is a detailed explanation of the data source.

Detail of repository

For my project, I used secondary data from GitHub, sourced through extensive manual searches on Kaggle and GitHub. I selected datasets that closely align with my project's objectives.This is the link to the dataset used in the github project-
LINK- https://github.com/PatelYash07/Pizza-sales
I can assure you that there will be very little similarity between my project and the one mentioned above.

Now let’s get a basic overview of the dataset.

Metadata and Descriptive Statistics

Metadata:

Orders data - It contains comprehensive information about the orders, including the specific date and time when each order was placed, providing a detailed record of when the transactions occurred. Following are the keys for the dataset.

1. **order_id** : Distinct ID assigned to each order made by the customer.
2. **date**: Date of order.
3. **time** : Time when order was placed.

Pizzas data : It provides information on the size of the pizza ordered and the price per pizza, detailing both the dimensions of the pizza and the associated cost.Following are the keys for the dataset.

1. **pizza_id :** Unique identifier for each pizza, defined by its type and size, which distinguishes it from other pizzas.
2. **pizza_type_id**: Foreign key that ties each Pizzas dataset to Pizza types dataset.
3. **size** : Size of the pizza, which can be Small, Medium, Large, Extra Large, or Double Extra Large.
4. **price :** The pizza prices, originally in dollars, were converted to Indian Rupees (INR) for a more relevant local audience.

Pizza types data : The dataset provides information on different pizzas, including a unique identifier for each, the pizza's name, its category (such as Classic, Chicken, Supreme, or Veggie), and a comma-separated list of ingredients.Following are the keys for the dataset.

1. **pizza_type_id :** Unique identifier assigned to each type of pizza.
2. **name** : Name of the pizza as shown in the menu.
3. **category** :Category that the pizza fall under in the menu (Classic, Chicken, Supreme, or Veggie)
4. **ingredients** :A list of pizza ingredients separated by commas.

Order Details data : The dataset includes unique identifiers for each pizza within an order, links to the main order and pizza datasets, and records the quantity of each pizza ordered by type and size. Following are the keys for the dataset.

1. **order_details_id** : Unique identifier for each pizza within an order.

2. **order_id** : Foreign key that links the details of each order to the corresponding Orders dataset.
3. **pizza_id** : Foreign key that links the details of each order to the corresponding Pizzas dataset.
4. **quantity** : Quantity ordered for each pizza of the same type and size

Descriptive Statistics :

After processing and cleansing the data, here's a brief summary of the dataset using descriptive statistics for every table.

- **Orders data**

Statistic	Order ID	Date	Time
Count	21,350	21,350	21,350
Unique	21,350	358	16,382
Most Frequent	1	27/11/2015 (115 times)	18:49:37 (6 times)
Frequency	1	-	-
Min	-	01/01/2015	09:52:21
Max	-	31/12/2015	23:05:52
Range	-	364 days	13 hours, 13 minutes, 31 seconds
Mean	-	28/06/2015, 23:41:47	16:35:03
Median	-	28/06/2015	16:48:53
Mode	-	27/11/2015	18:49:37
Standard Deviation	-	104 days, 14 hours, 36 minutes	3 hours, 8 minutes

Pizzas data

Statistic	Pizza ID	Pizza Type ID	Size	Price
Count	96	96	96	96
Unique	96	32	5	27
Most Frequent	-	the_greek	S	1743.0
Frequency	-	5	32	-
Minimum Price	-	-	-	819.0
Maximum Price	-	-	-	3019.8
Range	-	-	-	2200.8
Mean Price	-	-	-	1381.01
Median Price	-	-	-	1365.0
Mode Price	-	-	-	1743.0
Standard Deviation	-	-	-	343.58

Pizza types data

Statistic	Pizza Type ID	Name	Category	Ingredients
Count	32	32	32	32
Unique	32	32	4	32
Most Frequent	All values are unique	All values are unique	Supreme	All values are unique
Frequency	1	1	9	1

Order Details data

Statistic	Order Details ID	Order ID	Pizza ID	Quantity
Count	48,620	48,620	48,620	48,620
Unique	48,620	21,350	91	4
Most Frequent	All values Unique	10,760	big_meat_s	1
Frequency	1	21	1,811	47,693
Minimum	-	-	-	1
Maximum	-	-	-	4
Mean	-	-	-	1.02
Median	-	-	-	1
Standard Deviation	-	-	-	0.14

Having covered all the basic variables from the dataset, the next section will delve into the analysis process. We will provide a detailed explanation of the methods used and discuss why these processes are essential for comprehensive data analysis.

Detailed Explanation of Analysis Processes and Methods

Since the dataset comprises various tables linked by foreign keys, I merged these tables using Excel, leveraging its efficiency in handling complex relationships between tables. For the analysis, I utilized both Python and Excel. Python was employed for advanced data manipulation and visualization, particularly using the Plotly library, while Excel facilitated the effective merging and organization of the data, ensuring accuracy and clarity in the analysis.
Here's a refined version of your statement:

During the pre-analysis process, certain variables indicated significant fluctuations, prompting me to delve deeper into them. I identified and analyzed these specific variables further to gain a better assessment and understanding of their impact.

Metrics and Variables

In my analysis, I have used a variety of variables and metrics to gain comprehensive insights. Here I have listed some of them.

Percentages: I used percentages to represent the proportion of sales or revenue within different categories and time periods. This method was particularly useful given the revenue figures are in lakhs and crores, as it provided a clearer comparison of relative performance. Additionally, I incorporated percentages in heatmaps to visualize data distribution and trends more effectively, highlighting variations and intensities across different periods and categories.

Averages: I calculated averages for monthly and daily revenue to normalize performance data and identify trends, deviations, and key metrics over time.

Counts: I tracked counts of pizzas sold, operational days, and peak days to gain insights into sales volume, operational efficiency, and peak sales periods. By analyzing the number of pizzas sold, I could assess overall sales performance and identify high-demand periods. Tracking operational days allowed me to understand the café's availability and its impact on sales, while monitoring peak days helped pinpoint when sales were at their highest. This comprehensive approach enabled me to evaluate the effectiveness of operational strategies and identify key periods that significantly influence sales, thereby providing valuable information for optimizing business practices.

Peak Analysis: I used peak analysis to identify high-sales days, weeks, and hours. This analysis was crucial for uncovering patterns and variations in sales performance. By pinpointing these peak periods, I could better understand seasonal variations and periods of high demand. This approach provided valuable insights into sales trends and helped in planning for seasonal fluctuations effectively.

Interquartile Range (IQR): I analyzed the Interquartile Range (IQR) to assess the variability in daily sales across different months. By classifying months based on their sales variability, I was able to identify those with high or low performance spreads. This analysis highlighted months with consistent sales patterns versus those with significant fluctuations, providing insights into periods of stability and variability in sales performance.

Correlation Coefficients: I found a strong positive correlation between quantity sold and revenue, showing that more sales lead to higher revenue. Additionally, a moderate negative correlation between unique orders and average quantity per order indicates that higher order frequency usually involves smaller quantities per order.These correlations help to further analyze those variables.

Outlier Detection: I used outlier detection to identify exceptional days with unusually high or low performance.

<u>Visualization Techniques:</u>

I have used a variety of graphs and charts to visualize my dataset effectively. These visualizations are listed as follows:

Bubble Chart:

In my analysis, I used a bubble chart to effectively illustrate and compare monthly revenue and the quantity of pizzas sold. Given the large amount of data, the bubble chart simplified this comparison by varying the size of the bubbles to represent the total quantity sold in each month. Additionally, the color variations in the chart made it easy to distinguish months that performed above or below the average revenue threshold.To visualize the patterns more clearly I used a non zero base.

Tables:
I used tables in my analysis to present data in a clear and organized manner, allowing for easy comparison and identification of trends. By sorting the data, I could highlight key insights, such as patterns in peak sales days and monthly revenue trends. This approach made it straightforward to analyze performance across different time periods and helped simplify complex data for better interpretation.

Heatmaps:
In this analysis, I utilized heatmaps as a key visualization tool, especially because there are numerous pizza names (32 in total). This made it difficult to compare trends effectively using traditional charts. Heatmaps allowed me to simplify these comparisons by representing data through color variations. The use of color gradients enabled me to distinguish performance patterns easily, such as identifying top-performing pizzas or spotting trends in revenue and sales across different locations. By providing a visual comparison, heatmaps made it more intuitive to analyze trends and identify key insights.

Box Plot:
I used the box plot to visualize daily average revenue and assess variability across months. It highlighted trends, such as the decline in revenue towards year-end, and classified months by interquartile range (IQR) into those with consistent or variable sales. Outliers revealed exceptional or underperforming days, offering insights into days influenced by promotions or events. This made the box plot an effective tool for understanding revenue distribution and identifying patterns.

Having provided a thorough explanation of the analysis tools and methodologies used, I will now transition to discussing the results and findings. This section will offer a detailed overview of the outcomes from my analysis, outlining key insights .

Result and Findings

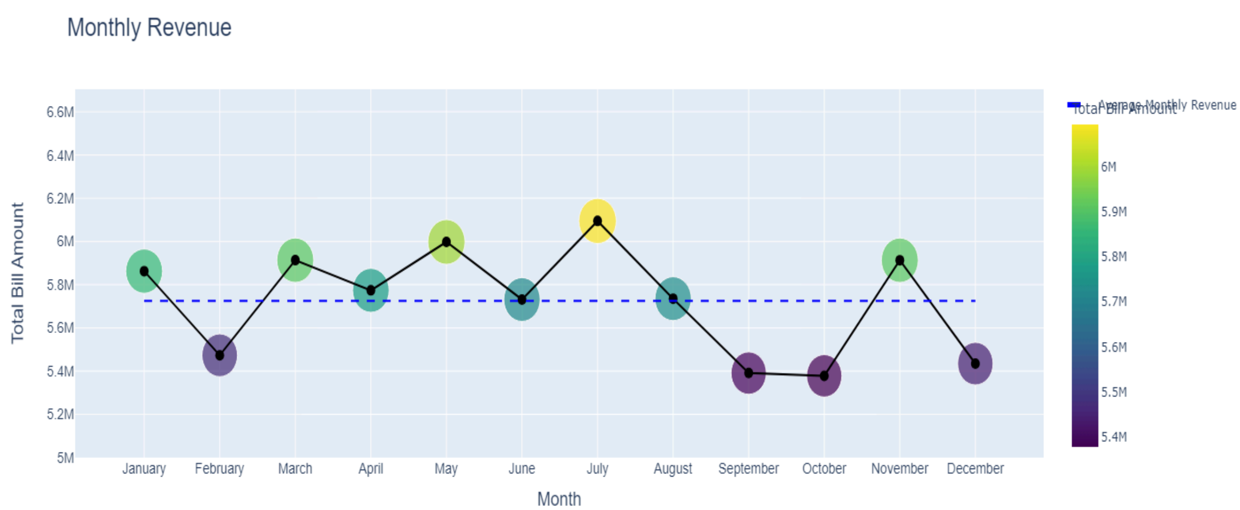
1.Exploring Revenue Trends and Patterns Over Time

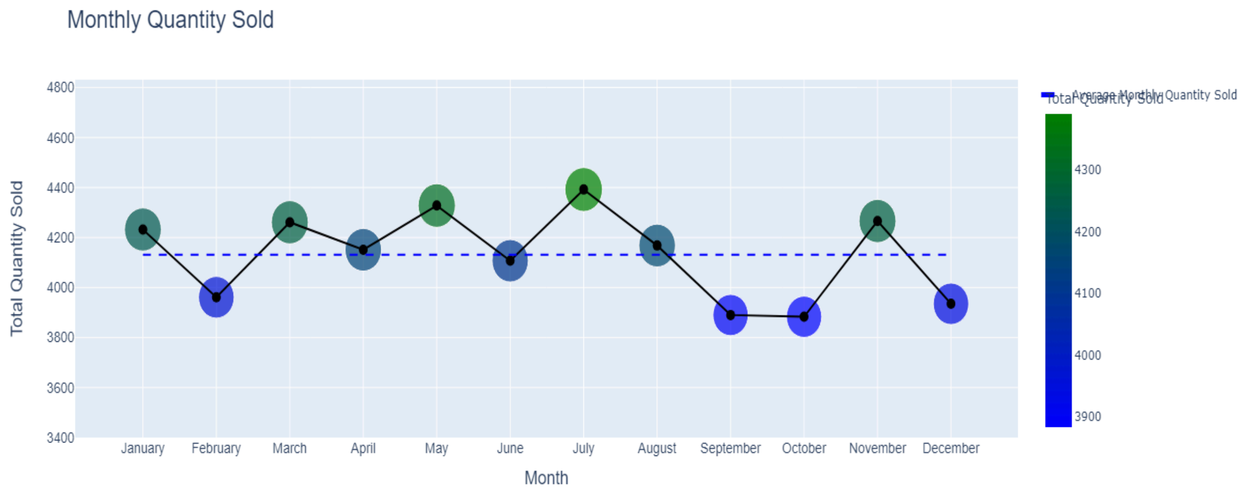
Monthly Revenue comparison using Bubble Chart

Throughout the year, a total of 49,574 pizzas were sold, generating an overall revenue of Rs 6,87,00,244. Notably, revenue exceeded Rs 50,00,000 in every month. To illustrate this effectively, I used this threshold to compare and analyze the monthly revenue. The bubble chart clearly shows that the four months represented by the violet bubble performed comparatively poorly, with their revenues falling below the average monthly revenue of Rs 57,25,020.35, indicated by the blue dotted line.

July emerged as the top revenue-generating month, with a total of Rs 60,94,863.60, while October had the lowest revenue, totaling Rs 53,78,318.40. A similar trend is observed in the quantity sold per month. Where the average quantity sold per month is 4131.
It is to be noted that the June month quantity sold is 4107 which is lower than the monthly average but revenue is Rs 5731336.8 which is slightly above than the average monthly revenue.

The high correlation coefficient of 0.997 between total quantity sold and total revenue demonstrates a very strong positive relationship. This implies that as the number of pizzas sold increases, the total revenue rises nearly proportionally. Therefore, I will primarily use revenue for most of my comparisons going forward.
So we have February ,September,October and December as the low revenue generating month.





Daily Revenue comparison for every Month

Analysis using Table and Pie Chart

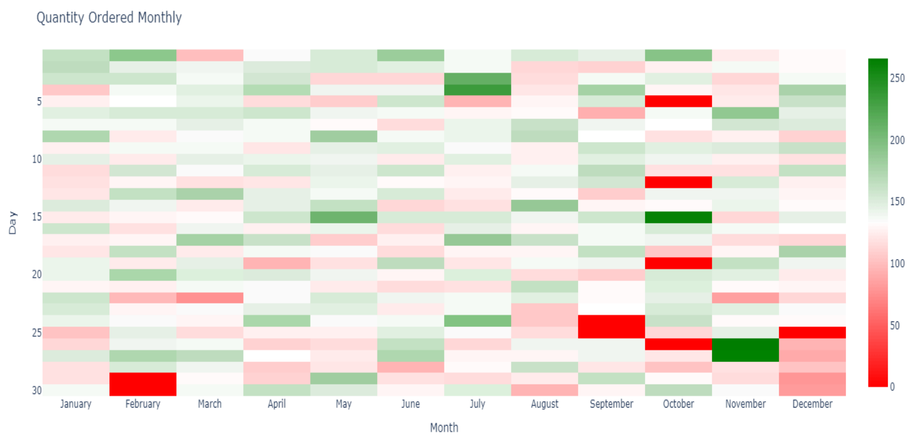
The average daily revenue for the cafe over the past year is Rs 1,91,900. This figure is determined by dividing the total annual revenue by the number of days the cafe was actually open for business. Out of the 365 days in the year, the café was operational for 358 days and closed for 7 days, as illustrated in the accompanying pie chart. This calculation provides a clear picture of the café's daily revenue performance while accounting for operational days.

In the table, the daily average revenue for each month is organized from lowest to highest. Interestingly, months that previously showed lower total revenue are performing better when evaluated based on daily average revenue. For example, October, which recorded the lowest total revenue, actually has the highest daily average revenue. Conversely, December, which was among the four months with the lowest total revenue, follows a similar trend and exhibits the lowest daily average revenue. Notably, five months—December, August, January, March, and June—have daily average revenues below the annual daily average of Rs 1,91,990.

Month	Avg Daily Revenue
December	181163.22
August	185012.0323
January	189117.329
March	190753.4323
June	191044.56
April	192463.04
September	192540.15
May	193478.4194
February	195478.8
July	196608.5032
November	197106.98
October	199196.9778

Analysis using Heatmap

In the heatmap below, I have plotted the quantity sold each day for every month. September (2 days), October (4 days), and December (1 day) feature dark red squares, indicating days with zero orders. Notably, the last quarter of the year shows a greater number of dark red shaded regions, indicating a higher frequency of days with no sales. Additionally, the final week of December displays a high density of red shading, signifying notably lower total sales thus low total revenue during that period.
The four non-working days in October may account for its lowest total revenue despite having the highest average daily revenue, suggesting strong sales on operational days.
While the shop was operational 100% of the time in the other quarters, the last quarter had only 92.4% operational days. This reduction could be a contributing factor to the lowest revenue in that quarter.
The red squares at the end of February, representing the 29th to 31st, are understandable given that these dates are not present in a standard February.



Quarter	Quarterly Revenue
1	1,72,49,400
2	1,75,03,059
3	1,72,21,361
4	1,67,26,424

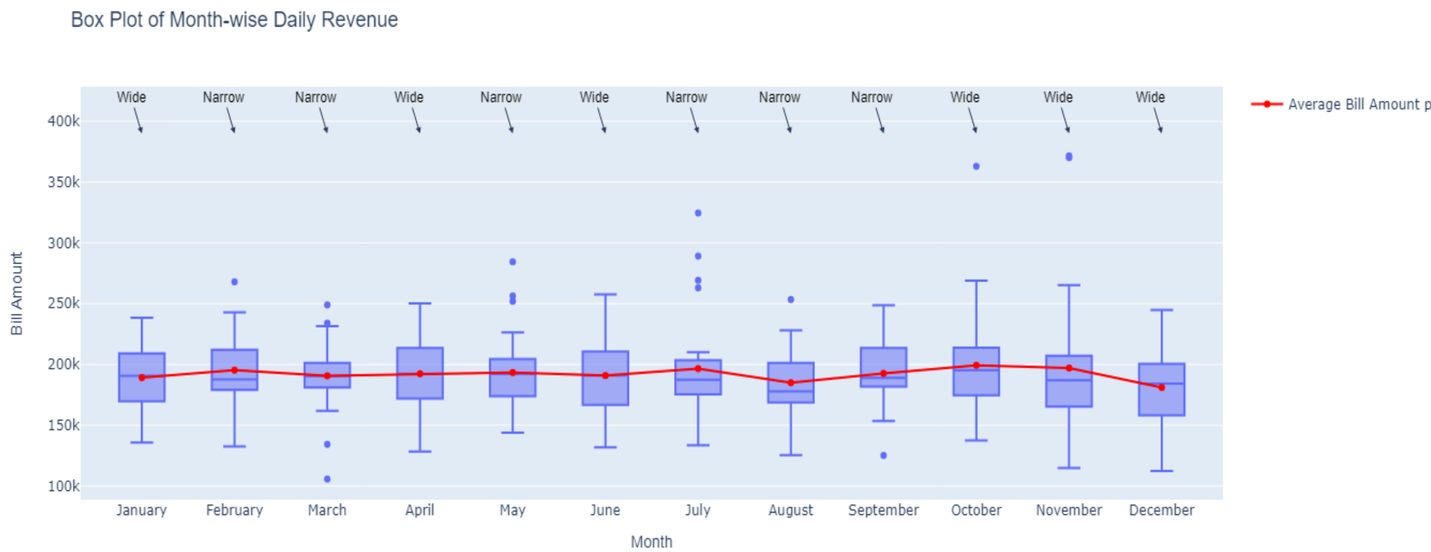
Analysis using Box Plot

In the box plot the red line represents the daily average revenue for each month.It shows a slight decline in daily total price towards the end of the year.

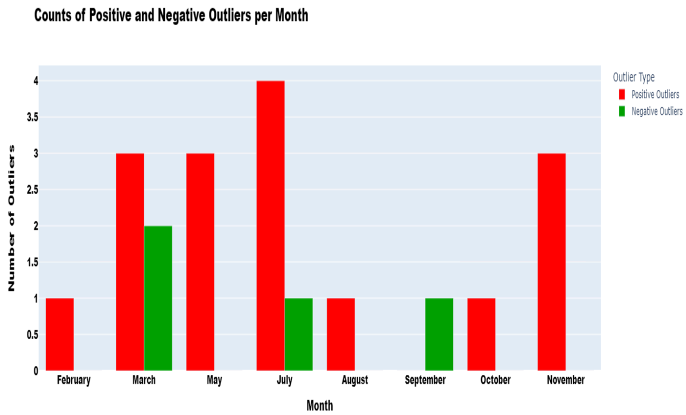
The spread of the boxes (interquartile range) and the position of the whiskers in the box plot illustrate the variability in daily total sales for each month.The classification of these spreads as 'Wide' or 'Narrow' is based on a median IQR threshold. If a month's IQR exceeds this threshold, it is classified as 'Wide'; if it is below the threshold, it is classified as 'Narrow'.Months such as February, March, May, July, August, and September exhibit narrower spreads, which indicates higher variability in daily sales.

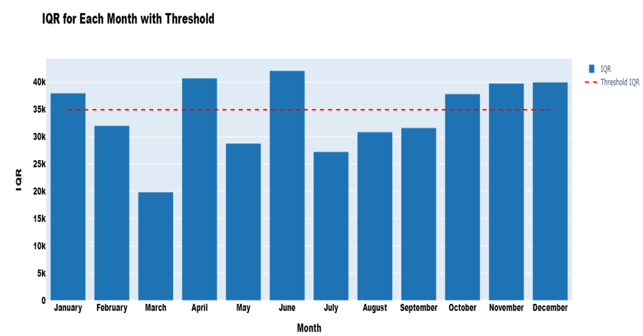
Wider boxes in months like January,April,June,October,November and December imply more consistent daily sales.
As shown in the bar graph ,June exhibits the highest interquartile range (IQR), indicating considerable variability in daily revenues. This means that while some days may experience substantial earnings, others might perform less well. In contrast, March has the lowest IQR, suggesting that daily revenues are

closely grouped around the median. This implies that business performance in March is stable and predictable, with minimal fluctuations in daily revenue.



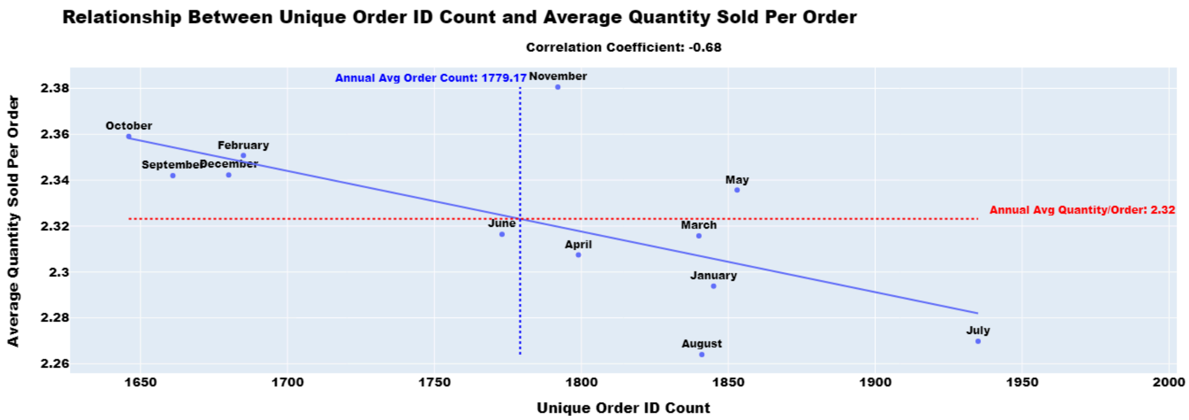
The Box Plot also reveals several months with outliers (dots extending above or below the whiskers). Higher outliers might reflect exceptionally successful days, such as those involving promotional events or holidays, whereas lower outliers could indicate days with underperformance.
In the bar graph below Count of Outliers for each month is presented.The red bar represents overperforming days while the green bar represents the underperforming days.July has the highest overperforming days while March has highest underperformers.
November has three outliers and two of those days have revenue greater than Rs 3,50,000.





One interesting thing came up when I was analyzing average quantity order wrt unique Order Id and there's a moderate negative correlation of -0.68 between the Unique Order ID Count and the Average

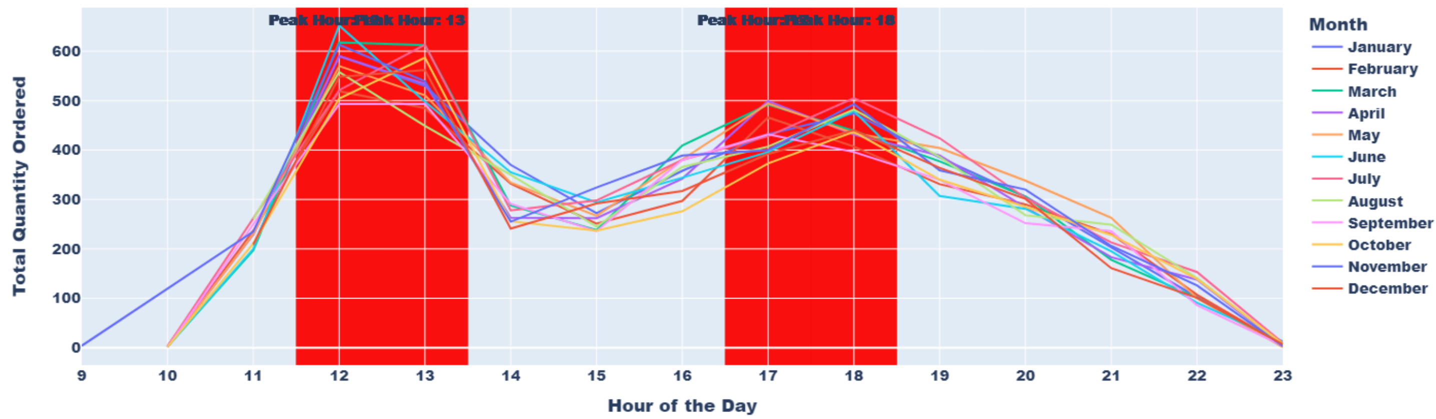
Quantity Sold Per Order. This suggests that as more unique orders are placed, the average quantity sold per order decreases.
October and November show a higher average quantity sold per order but fewer unique orders.
July and August have higher unique order counts but lower average quantities. It suggests customers are ordering more frequently but in smaller amounts, possibly individual orders.



Finding Peak Days and Hours

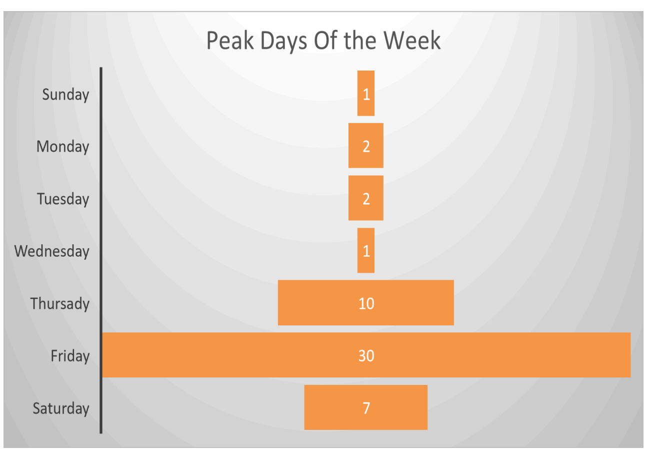
Certain hours of the day experience higher order volumes. The interval from 12:00 to 13:00 is consistently a peak hour each week, with additional peak times around 13:00, 17:00, and 18:00.
In the line chart the red shaded region represents the peak hours.The trend is quite similar across the months.

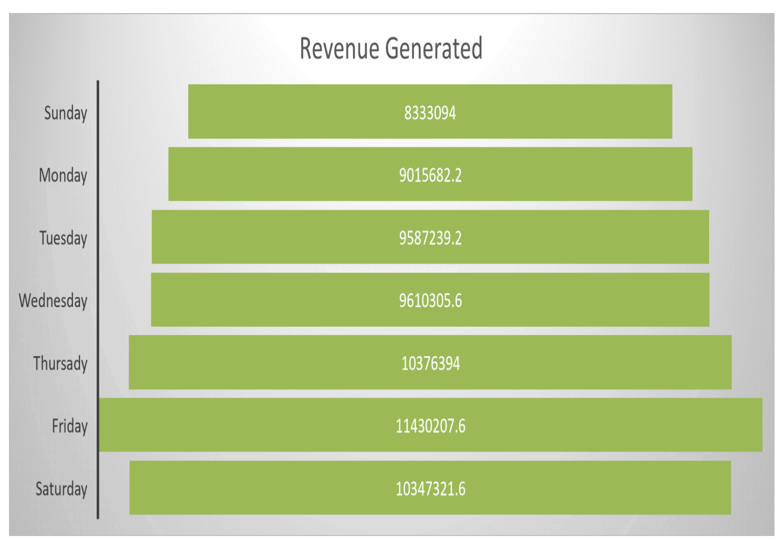
Hour of the Day	Weeks in Top 4 for Most Working Hours
12	53
13	52
18	48
17	37
19	21
16	15



Certain days of the week perform better than others . As shown in the table,Out of 53 weeks, Friday was the peak day for 30 weeks, while Thursday was the peak day for 10 weeks. In contrast, from Sunday to Wednesday, there were fewer weeks where these days were the peak days

It is not surprising that revenue trends follow a similar pattern, with the highest total revenue recorded on Fridays and the lowest on Sundays.

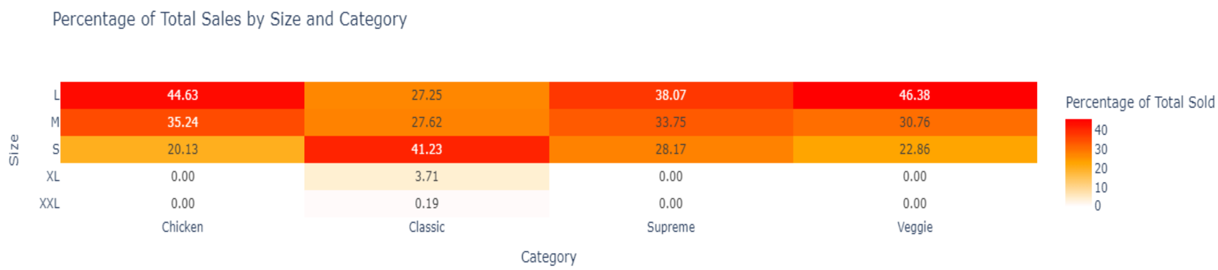




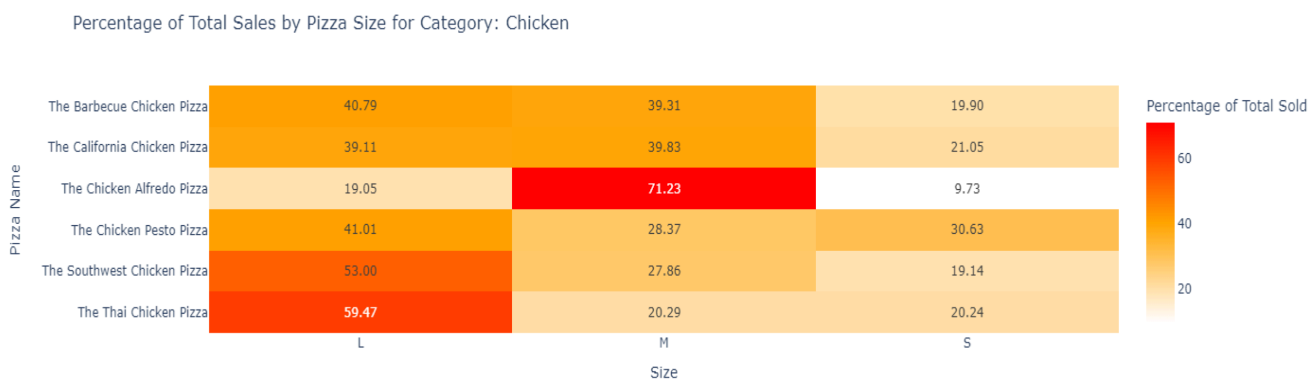
2. <u>Exploring Products and Categories</u>

Analysis on the basis of variation in Size.

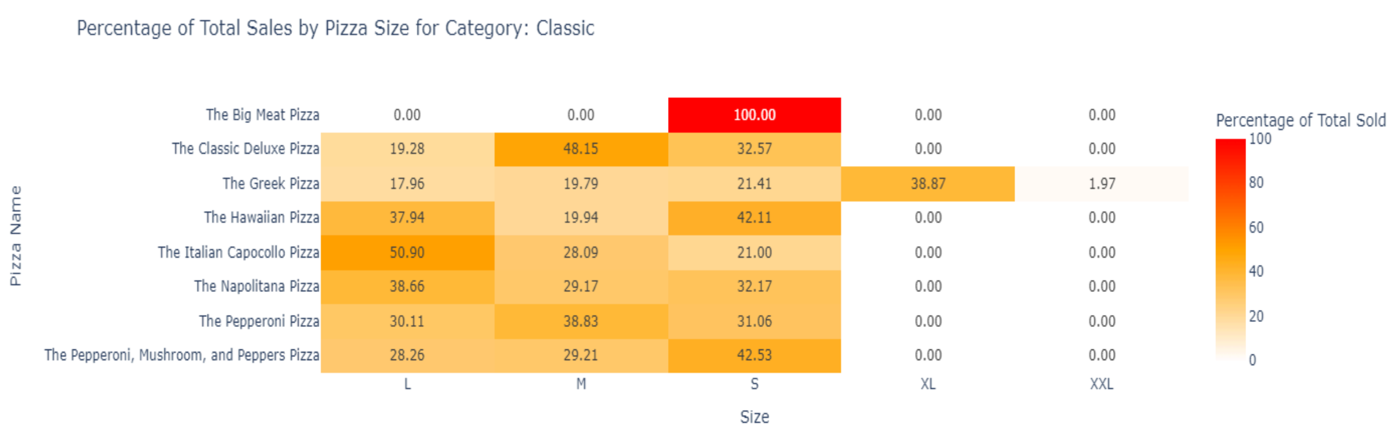
The pizzas are divided into four categories Chicken,Classic,Supreme and Veggie.The categories Chicken ,Supreme and Veggie are selling more Pizzas in L size while the Classic Category is selling its most pizza in S size thi category also sells pizza in XL and XXL size.



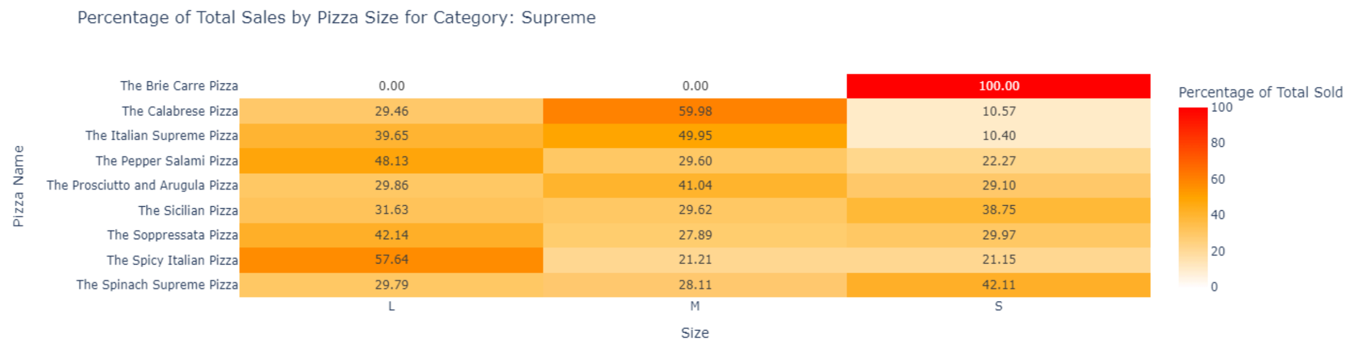
The pizza in the Chicken category is being mostly sold in L size but 71% of Chicken Alfredo pizza is getting sold in M size and only 9% of them are being sold in S size.The S size pizza in this category has relatively less demand.More than half of The Thai Chicken and The SouthwestChicken Pizza are being sold in L size.



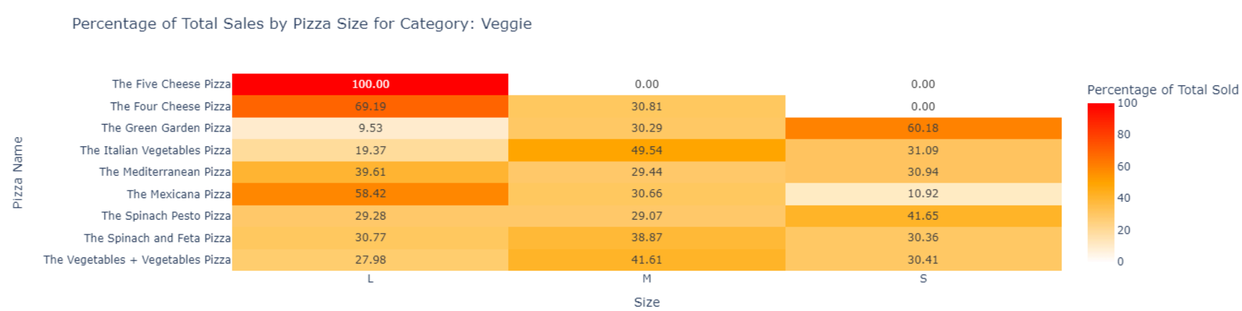
The Classic Category has similar variation in the L and M size although two pizzas stand out in terms of size the first one being The Big Meat Pizza which is only sold in S size and The Greek Pizza which is being mostly sold in XL size.



The Supreme pizza exhibits the most balanced sales distribution across sizes, with each size representing about 33% ± 5% of total sales. However, there are notable anomalies: The Spicy Italian and Soppressata pizzas have over half of their sales in the L size, while the Calabrese pizza shows a similar trend with the M size. Additionally, the Brie Carre pizza has all its sales concentrated in the S size.



In the Veggie category, the L size leads in sales. The Four Cheese Pizza, Five Cheese Pizza, and Mexican Pizza each have more than 50% of their sales in the L size, with the Five Cheese Pizza having all its sales in this size. Conversely, the Green Garden Pizza stands out with 60% of its sales in the S size.



Analysis on the basis of monthly performance.

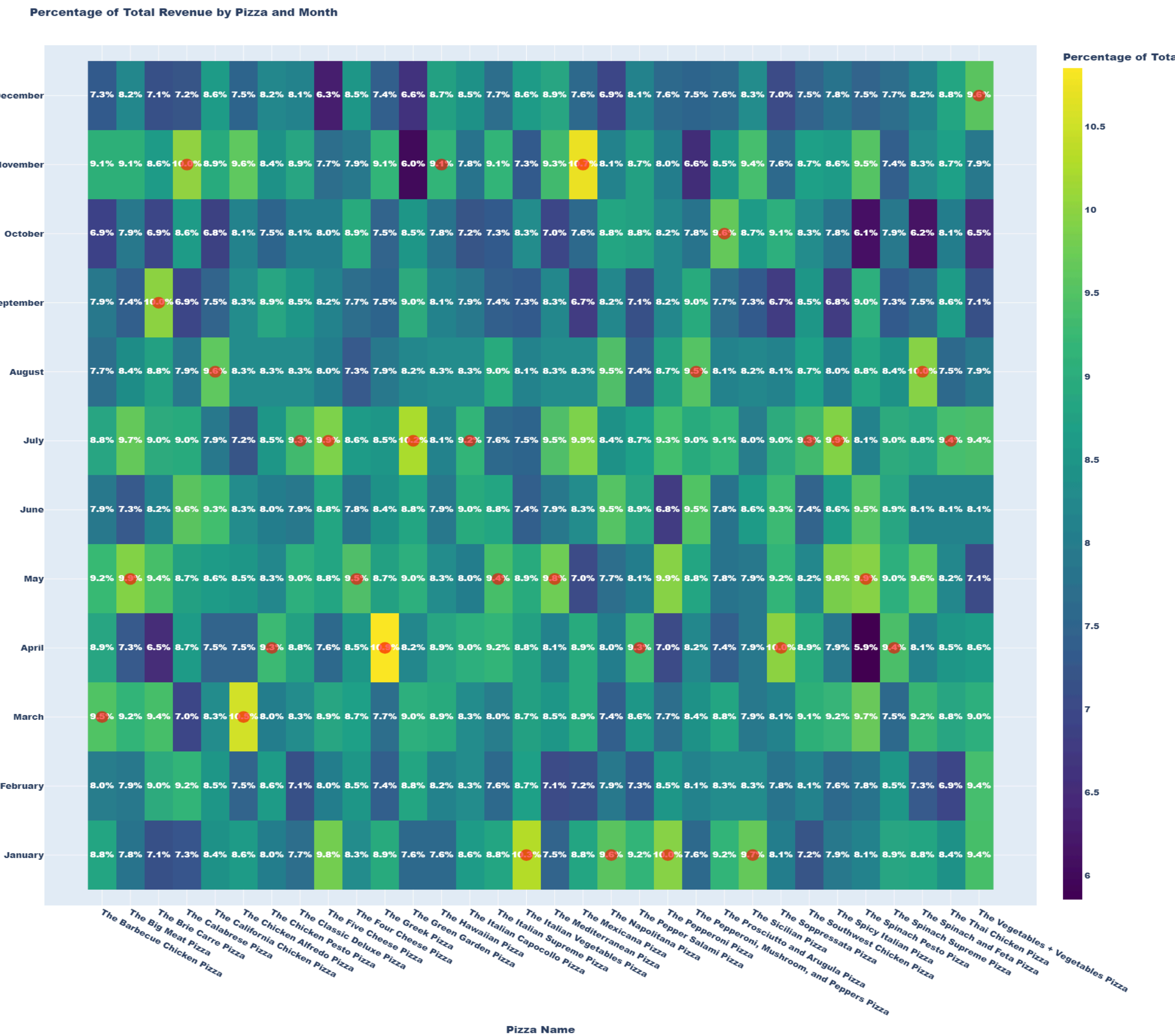
Out of 32 pizzas The Barbecue Chicken Pizza and The Thai Chicken Pizza each rank among the top 4 selling pizzas for 12 months. This indicates these pizzas are highly popular and maintain strong performance throughout the year.

The heatmap below illustrates the performance of each pizza across different months, effectively comparing each pizza's sales performance month-to-month.

In the heatmap, dark blue boxes represent the lowest-selling month for each pizza, with darker violet boxes indicating increasingly poor performance. Conversely, yellow boxes signify better performance, with the intensity of the color reflecting higher sales.

The red dot highlights the best-selling month for each pizza.The distribution of these top-selling months is as follows: January has 4, February has 0, March has 2, April has 5, May has 5, June has 0, July has 7, August has 3, November has 3, and the remaining months each have 1. This distribution indicates that July has the highest number of top-selling pizzas, while February and June have none.

July stands out with the highest concentration of peak sales, making it a crucial month for maximizing revenue. April and May also demonstrate strong performance, with several pizzas reaching their sales peak during these months. On the other hand, February and June show no peaks, indicating a potential opportunity to boost sales through targeted marketing efforts during these quieter months. This distribution highlights the importance of aligning promotional strategies with peak months to fully capitalize on consumer demand.



Ingredient used

Out of 54 unique ingredients, only 4—Tomatoes, Red Onions, Garlic, and Mushrooms—are used across all four categories. Additionally, 38 ingredients are exclusive to individual categories.

After thoroughly analyzing the graphs, charts, and heatmaps, we can confidently offer a set of recommendations and insights to guide the business. These insights are based on a detailed review of sales patterns,timely trends and performance trends, aiming to enhance operational efficiency and improve revenue.

Interpretation of Results and Recommendation

- October has a high daily average revenue, but its overall revenue is lower due to closing days. To maximize potential revenue, the café should ensure it remains open during these high-demand months, as this will help drive revenue even further.

- Wider spreads in daily revenue indicate significant variability, with some days showing exceptional performance while others lag. To address this, we can implement targeted marketing efforts, discounts, or special events on underperforming days to boost sales. Additionally, planning events or promotions to align with known high-performing periods within these months can further enhance revenue.

- Narrow Spread Months like February, March, May, July, August, September indicate that daily revenue is closely clustered around the median, suggesting consistent but potentially limited growth. Months like August have low dail

- For July and August, where we see high order volumes but lower average quantity per order, we should focus on upselling and cross-selling. We can train staff to suggest add-ons or larger sizes and implement promotions like 'Buy one, get one half-off' to increase order sizes. Additionally, we can introduce limited-time offers, such as discounts on a second pizza or additional items, to encourage customers to order more per transaction.This strategy can be implemented to all months having lower average quantities per order.

- To manage peak times (12:00-14:00, 17:00-19:00), we should schedule extra staff to handle higher order volumes and maintain customer satisfaction. During quieter periods, such as early mornings or late evenings, we can reduce staff levels and use part-time or flexible workers. This approach will help reduce staffing costs while enhancing customer satisfaction.

- By training our staff to handle multiple roles, we can quickly reassign them based on current needs. This strategy enhances operational efficiency and reduces the necessity for specialized staff during off-peak periods.

- We can drive more traffic and increase revenue on peak days by offering special promotions, such as a 'Friday Feast' with discounts or free items, and by creating bundle deals, like discounts on combo meals or family packs, to boost average transaction values

- We can optimize inventory and promotions by adjusting strategies based on pizza size within each category. We should focus on strong promotions and robust inventory for Large (L) sizes, balance Medium (M) sizes with targeted deals, and boost Small (S) sizes with specific promotions. By aligning inventory and marketing with size preferences, we can enhance sales and better meet customer demand.
- Barbecue Chicken Pizza and Thai Chicken Pizza consistently rank among the top-selling pizzas throughout the year. These pizzas should be a primary focus for promotions and marketing campaigns as they have demonstrated strong year-round demand.
- In the heatmap,For pizzas with dark blue or violet boxes (low-selling months), we should reassess their place on the menu. We can either work on increasing their visibility or consider replacing them with better-performing options if they consistently underperform.
- We can manage inventory levels more effectively by analyzing category sales patterns each month, allowing for better handling of the 38 ingredients used in specific categories. For ingredients common across all categories, we should source them from cheaper suppliers and buy in bulk to optimize costs.

In conclusion, the data suggests several improvements: maintaining operations during peak times, using promotions to increase sales, optimizing staffing, focusing on popular pizzas, and managing inventory more effectively. Implementing these changes can help improve profitability and overall efficiency.