INDIAN INSTITUTE OF TECHNOLOGY, MADRAS

BS, DATA SCIENCE AND PROGRAMMING

BUSINESS DATA MANAGEMENT - CAPSTONE FINAL REPORT

Optimizing Operations and Business Performance for an Italian Restaurant

BY HANANI BATHINA
21f1006169@ds.study.iitm.ac.in

Executive Summary:

The Flying Spaghetti Monster is a highly regarded Italian restaurant, located in the bustling Banjara Hills neighborhood of Hyderabad. This report aims to present a detailed analysis of the restaurant's sales data, spanning a seven-year period, from 2017-2023. Through this comprehensive analysis, the report sheds light on the top-selling products, customer demand

trends, peak demand hours, restaurant occupancy patterns, highest revenue-generating streams, and products with declining popularity, amongst other insights.

The sales data was gathered from the restaurant's point of sale (POS) system, capturing all transactions and item-level details. A PostgreSQL database with 12 tables and 60,000+ rows of orders, payments, and products was used to conduct the data analysis. The analysis was conducted using a range of tools and techniques, including SQL, Python, Matplotlib, and Excel. The methodology used for data analysis and visualization is explained in great detail in the report, along with the various charts and tables that provide a visual representation of the results.

The project objectives are clearly outlined, with a focus on identifying lost opportunities and revenue and providing alternative solutions. The report provides a detailed analysis of the restaurant's occupancy patterns, which is used to determine workforce needs and identify lost revenue due to delivery services via commission. The potential of an in-house delivery service is also explored, along with an assessment of whether the restaurant is primarily takeaway or dine-in. Additionally, the report analyzes the menu to explain the loss in popularity of certain items.

Based on the insights obtained from the data analysis, the report offers several recommendations for the Flying Spaghetti Monster Italian restaurant in Banjara Hills, Hyderabad. To optimize staffing decisions, the report recommends a comprehensive table that will allow the restaurant to allocate the right number of employees at the right times, improving service efficiency and customer satisfaction.

Furthermore, the report suggests marking up prices in case of delivery service-based orders instead of upstarting a delivery service of their own or bearing the delivery service’s commission fee. This can not only save on costs but also lead to increased profitability for the restaurant. The report advises against converting the restaurant into a cloud kitchen, given the potential loss of revenue from dine-in customers.

The pricing strategy is reviewed to ensure maximum revenue gain by correcting the discrepancy between the products' increasing popularity and the lack of a simultaneous increase in their prices. By adjusting the prices of products with increasing popularity, the restaurant can increase profitability and better capture the value of its offerings.

All the graphs, jupyter notebooks, SQL codes, and data are available in the annexure section, allowing for transparency and ease of access. Overall, this report provides a trove of insights and recommendations, which could be instrumental in driving increased profitability and growth for the Flying Spaghetti Monster restaurant

Detailed Explanation of Analysis Process/Method:

Objective 1 - Workforce Management:

To begin the analysis, we extracted dine-in orders using the entity_type_id and determined the earliest and latest billing times for each day over the past six years to establish the restaurant's operating hours. After identifying the busiest hour, which was found to be 8 pm, we constructed month-level time series for each hour to observe general trends in the number of orders over the past 6 years and ascertain the impact of COVID-19 on demand.

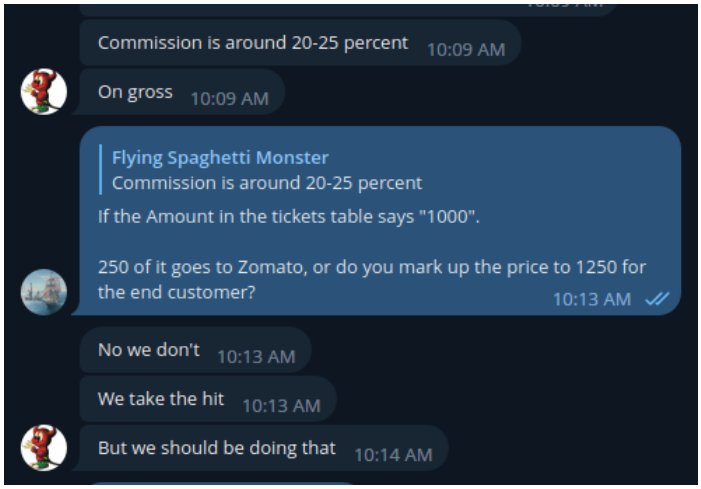
We then checked each hour's month-level time series for trends and found none, so we calculated descriptive statistics such as the mean, mode, max, and standard deviation. Since having excess staff is preferable to losing customers in the restaurant industry, we chose the maximum number of orders for each hour. However, we also included the 75th percentile of the interquartile range to account for outliers.

The resulting comprehensive table in the next section will guide the restaurateur in effectively managing their workforce for each hour.

Objective 2 - Delivery Services Commission:

To begin the analysis, we extracted all orders placed through delivery services such as Swiggy, Zomato, and others using the previously mentioned entity_type_id. We calculated the total revenue generated from these orders and their contribution to the total revenue. We also calculated the revenue lost to delivery services due to their commission percentage, which is an important metric to consider since the restaurant is not marking up its prices and is bearing the commission cost.

We generated month-wise and year-wise plots of the delivery orders over the past 6 years and observed the impact of COVID-19, which led to a spike in the number of delivery orders from March 2020 to March 2021.



(Figure: 1.2.1)

Objective 3 - Takeaway-Heavy or Dine-in Heavy:

The analysis began with extracting dine-in and takeaway orders from the cumulative orders using entity_type_id as the key identifier. The contribution of dine-in and takeaway orders to the overall revenue was then calculated. Month-level and year-level time series were generated for both categories to observe general revenue trends and identify the better performer between the two. Despite the fact that the restaurant gains much more revenue from takeaway and deliveries, the revenue gained through dine-ins is still significant at almost 40%, and thus warrants analysis of the restaurant occupancy.

Assuming an average table occupancy of 1 hour per customer, the restaurant's total capacity was estimated, and the actual number of occupied tables in a day over the past 6 years was calculated. These two pieces of information were then used to plot the occupancy pattern over the same period.

Objective 4 - Popularity Trendlines of Products:

To begin our analysis of the restaurant's products, we first calculated the total number of products and the number of times each product has been ordered. We then determined the contribution of each product to the total revenue to identify the top-performing products. The products were sorted in descending order based on their revenue contribution.

Next, we identified the top five products from the previously sorted list that have experienced the largest reduction in popularity over the past 6 years. To quantify this reduction, we calculated the average total revenue of each of these products in 2017-18 and compared it to the revenue in 2022-23.

These products that were once high-revenue performers in 2017-2018 but are not performing as well in 2022-2023 require more attention from the restaurant and maybe a possible source of improvement for this flailing restaurant.

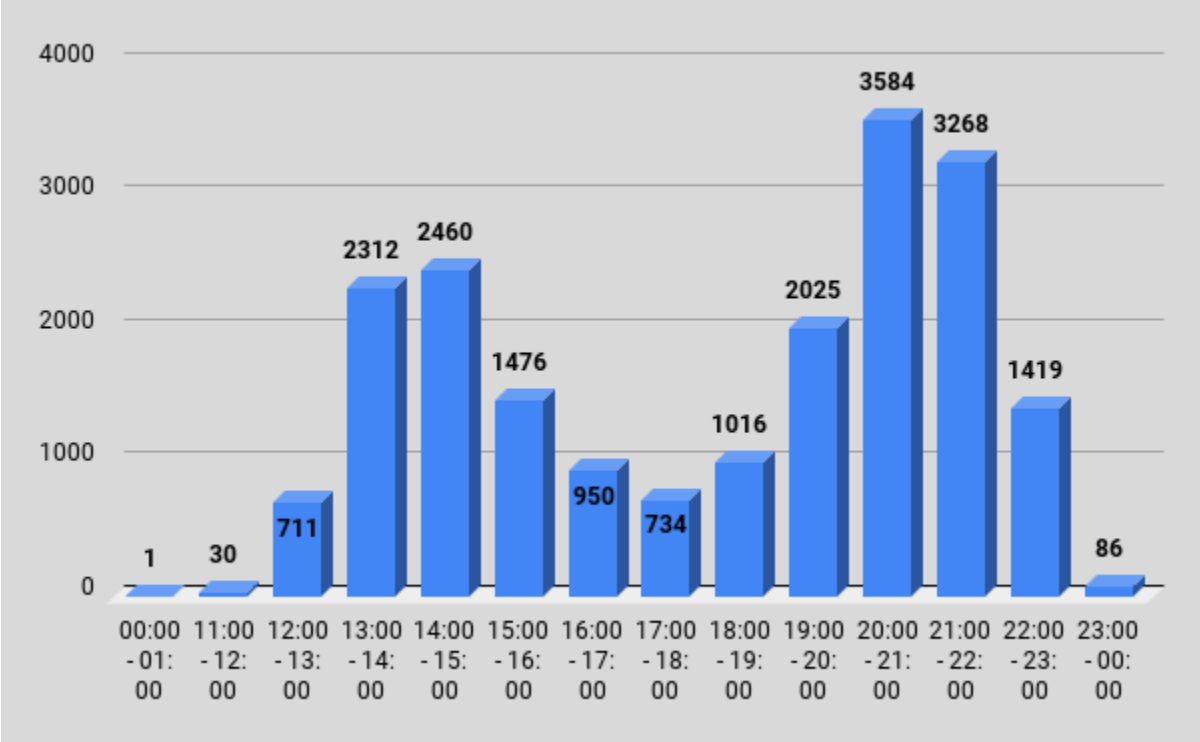
Furthermore, as an additional piece of information, we generated trendlines to observe the popularity of various food items over time, which will provide valuable insight into customer preferences and identify areas for improvement.

Results and Findings:

Objective 1 - Workforce Management:

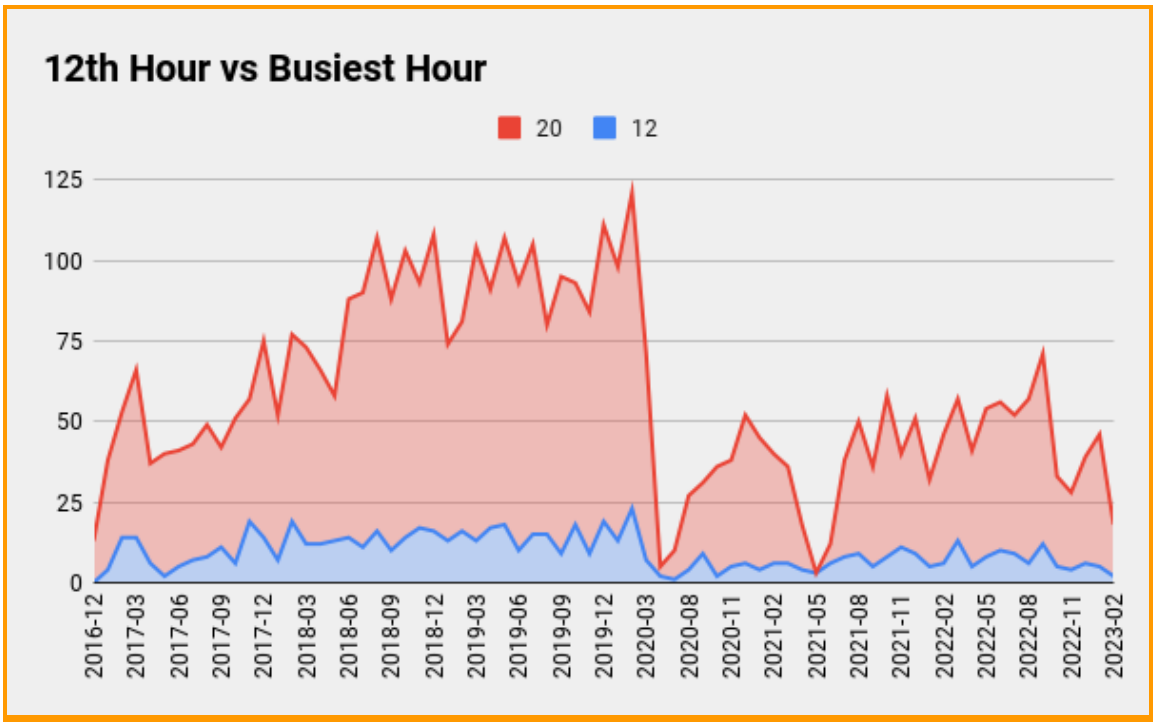
1. The Earliest Billed Hour: 11:04:17

2. The Latest Billed Hour: 00:11:09

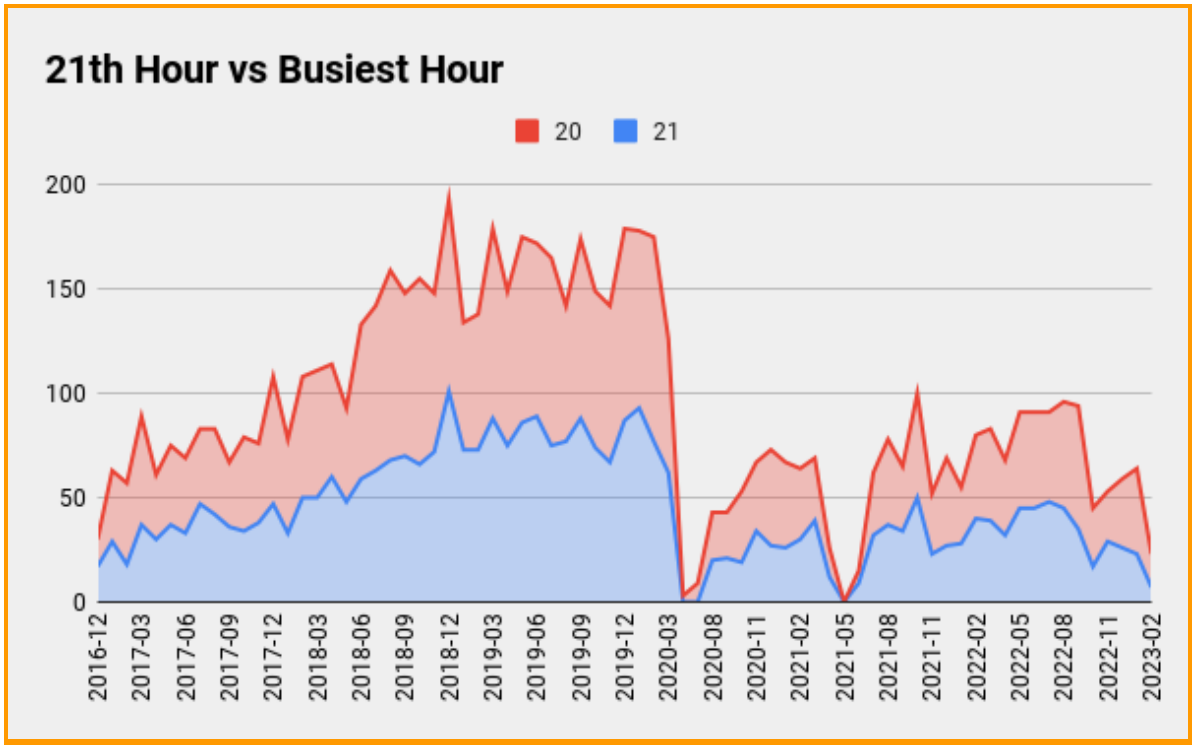


(Figure 2.1.1)

This graph displays the total number of dine-in orders received by the restaurant over the course of its entire history, broken down by operating hour. The X-axis represents the hours of operation, while the Y-axis represents the total number of dine-in orders received during those hours. Upon analysis, it becomes evident that the busiest hour for dine-in orders is from 8:00 PM to 9:00 PM, while the least busy hour is from midnight to 1:00 AM. These findings can be useful in optimizing staffing levels and scheduling, particularly during peak and off-peak hours.



(Figure: 2.1.2)



(Figure 2.1.3)

This figure displays the month-wise total number of dine-in orders received at the 12th hour vs the 20th hour, and the 21st hour vs the 20th hour, with months on the X-axis and the total number

of orders on the Y-axis. A comprehensive comparison of other hours with the 20th hour is available through the link provided in the annexure.

Upon careful analysis, we can observe that the number of dine-in orders does not exhibit any discernible trend over time. As a result, we recommend utilizing descriptive statistics to construct a table that allows the restaurateur to maintain optimal staffing levels while avoiding any idle resources.

	Mean	Standard Deviation	Min	25%	50%	75%	Max
12	9.45	5.2	0	5	9	13	23
13	31.32	14.86	0	21	29	44	59
14	33.5	15.78	1	21	33	44	72
15	20.191	9.064	1	14	19	27	41
16	12.986	6.97	1	8	12	17	31
17	10.041	5.57	0	6	9	14	24
18	13.9	7.21	0	10	14	18	36
19	27.71	12.96	3	20	27	34	58
20	49.06	24.71	0	33	44	65	98
21	44.76	24.69	0	28	39	66	101
22	19.43	11.96	0	11	18	28	46
23	1.178	3.31	0	0	0	1	17

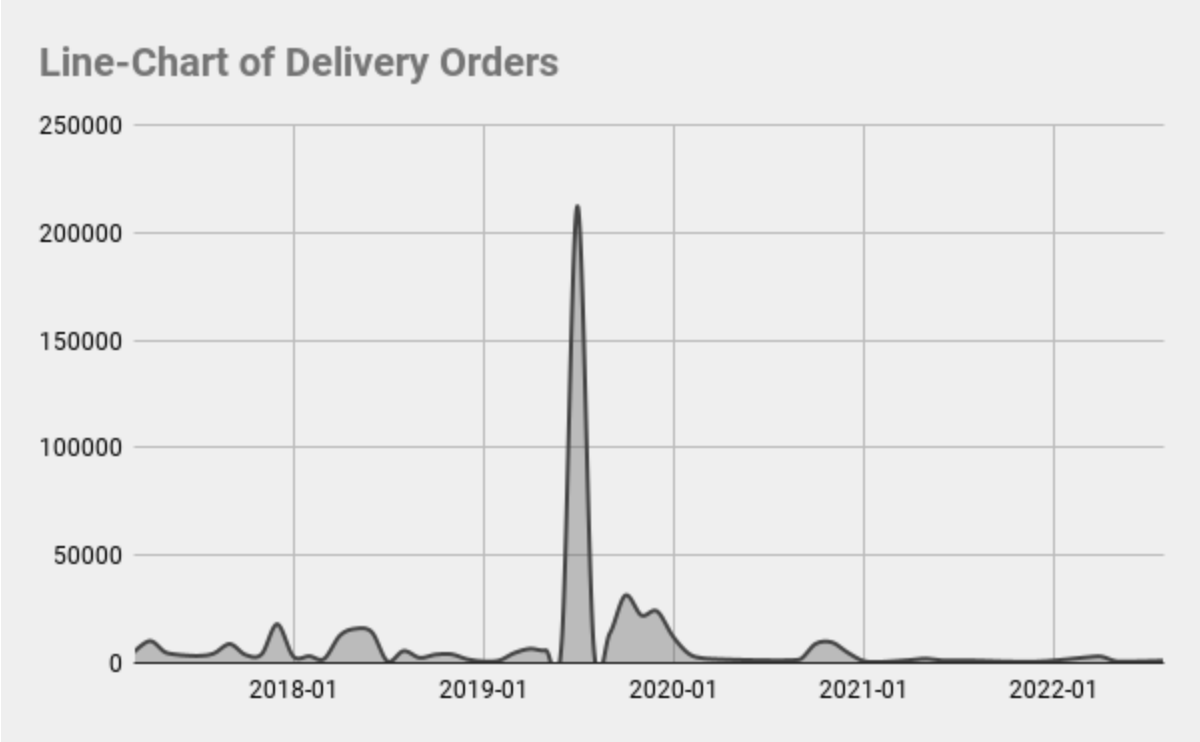
(Figure: 2.1.4)

The following table displays the descriptive statistics of the number of dine-in orders for each operating hour from 12 to 23. To enhance the readability of the table, the Max and 75IQR columns are color-coded with green indicating lower values and red indicating higher values. The rationale behind selecting the Max and 75 IQR metrics will be elucidated in the next section.

Objective 2 - Delivery Services Commission:

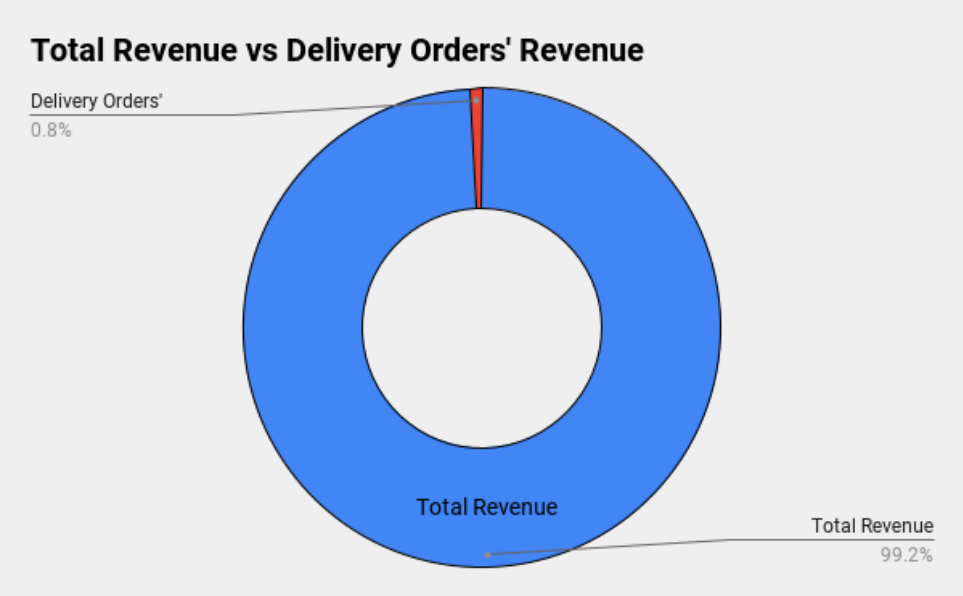
```
Total Revenue from Delivery Orders: Total:  5,27,530.0

Total Commission = 0.25 * 527530 = 1,31,966.0
```



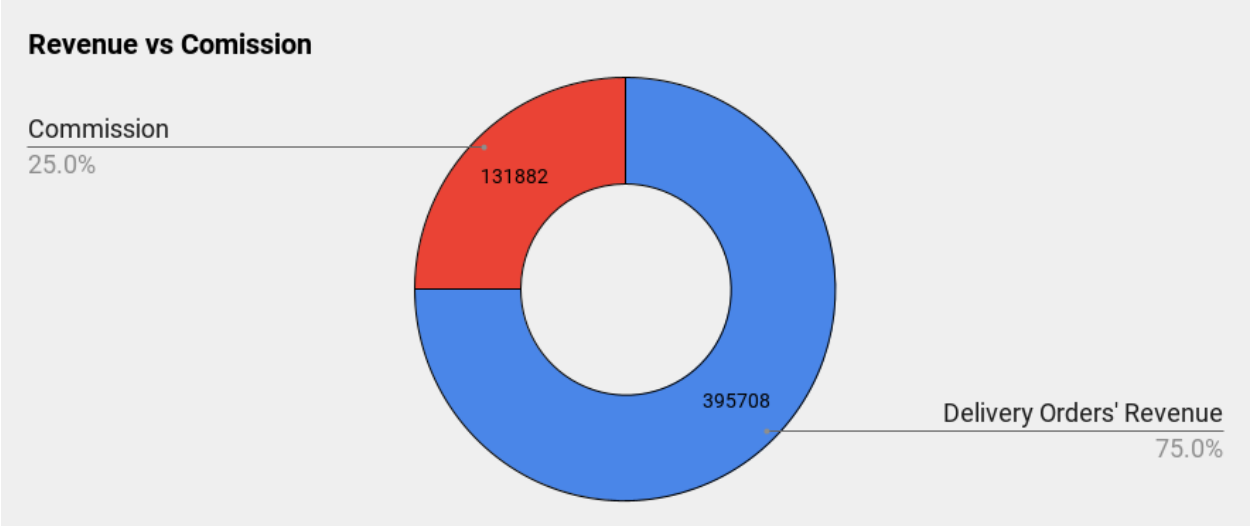
(Figure: 2.2.1)

This is a line chart that displays the revenue obtained from delivery orders over the course of several months. On the X-axis, we have the months, and on the Y-axis, we have the total revenue. The line chart depicts a noticeable upward trend in the revenue obtained from delivery orders, particularly during the Covid-19 pandemic. This increase in revenue can be attributed to the shift in consumer behavior towards ordering food online, as more people opted for home delivery during the pandemic.



(Figure: 2.2.2)

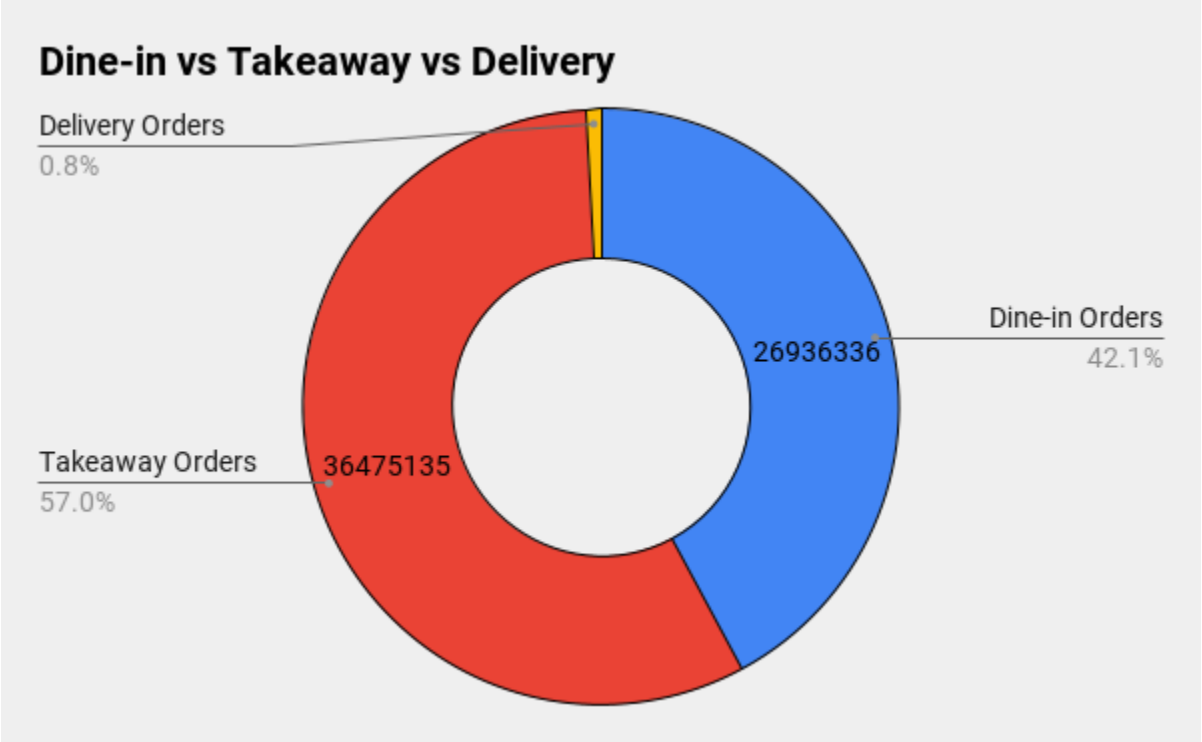
This pie chart displays a comparison between the total revenue earned by the restaurant and the revenue contributed solely by the delivery orders. The blue slice represents the total revenue, which is a whopping 99.2% of the revenue earned. In contrast, the red slice represents the revenue obtained solely from delivery orders, which is a mere 0.8%. As we can observe, the revenue earned from delivery orders is quite minimal and practically insignificant when compared to the revenue generated from the dine-in and takeaway orders.



(Figure: 2.2.3)

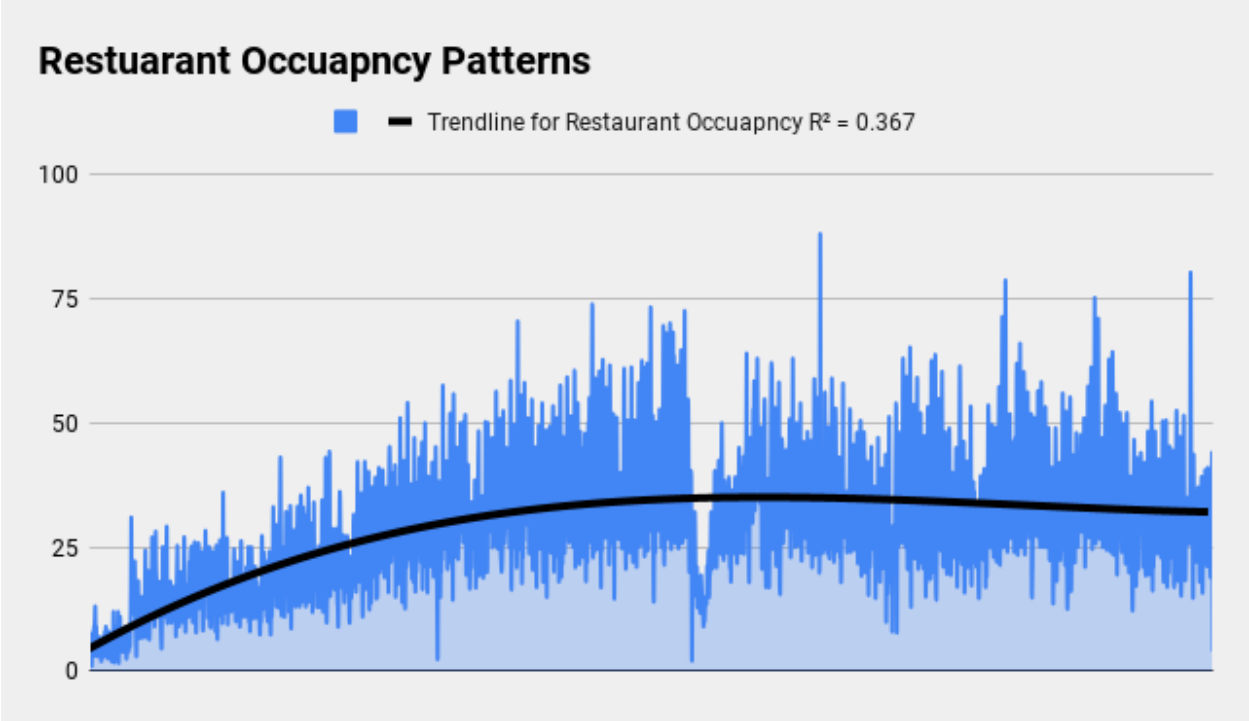
This pie chart depicts the comparison between the revenue earned from delivery orders and the commission lost to delivery services. The red slice represents the revenue earned from delivery orders, while the green slice shows the amount lost in commission paid to delivery services. It is notable that the restaurant is not increasing the prices of the delivery orders, but is bearing the burden of the commission, which is leading to a substantial loss in revenue. This loss is of great significance and must be addressed in order to ensure the profitability and sustainability of the restaurant's operations.

Objective 3 - Takeaway-Heavy or Dine-in Heavy:



(Figure: 2.3.1)

This pie-chart depicts the individual revenue contributions of delivery, takeaway, and dine-in orders. In order to analyze the restaurant occupancy, delivery and takeaway orders have been considered as a single entity. As we observe, dine-in orders contribute to 42.1% of the total revenue, while non-dine-in orders contribute to 57.8%. It is evident that both dine-in and non-dine-in orders are significant and play a crucial role in the restaurant's revenue generation

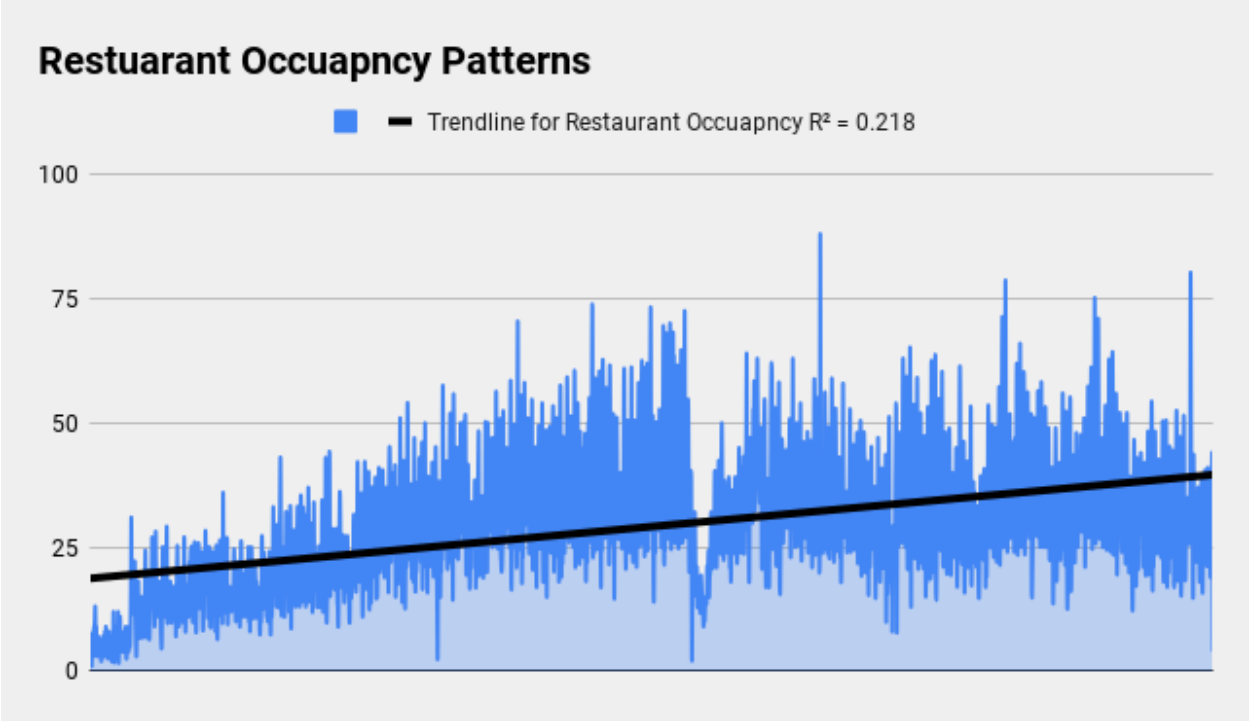


(Figure: 2.3.2)

This graph depicts the percentage of total restaurant capacity that was utilized every day over the past six years through a smoothed line-chart. The X-axis displays the time, and the Y-axis represents the percentage of restaurant capacity utilized. Upon analysis, we can see a clear upward trend in the initial years of the restaurant, which was followed by a decline in occupancy due to the Covid-19 pandemic. However, the restaurant has managed to regain its pre-Covid occupancy levels and even exceeded them slightly.

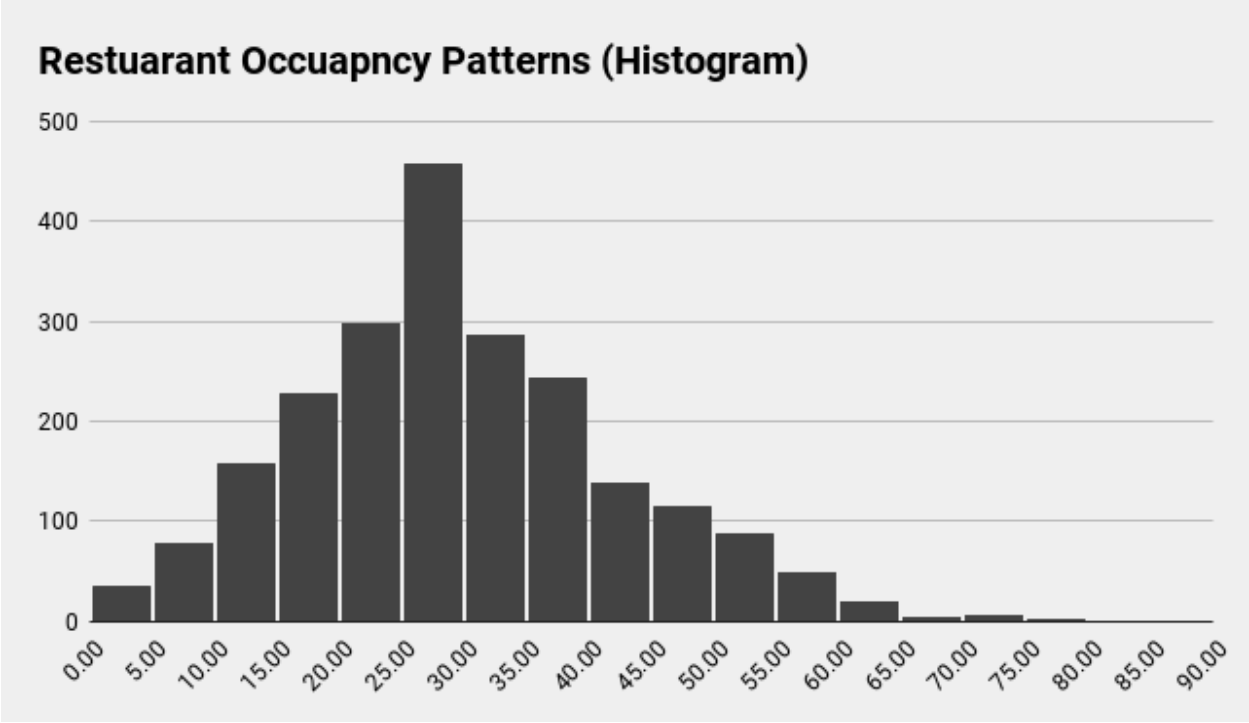
It can be inferred that the restaurant has reached its natural threshold in terms of occupancy, and unless a significant overhaul is made, such as expanding the restaurant's capacity or enhancing its advertising, there will likely be no substantial increase in occupancy. The data was fit to a polynomial of degree 3 with an R-squared score of 0.367.

R-squared is a statistical measure that represents the total explained variance of the dependent variable, which in this case is the percentage of total restaurant capacity utilized.



(Figure: 2.3.3)

This graph depicts the same data as the previous one, i.e., the percentage of total restaurant capacity utilized over the past 6 years. However, in this graph, a linear trendline has been fitted on the data, and it has a lower R2 score of 0.218 compared to the previous graph. This indicates that a linear model is not the best fit for this data and suggests the presence of other factors that contribute to the variation in restaurant occupancy over time. Despite this, we can still observe that the trendline shows a gradual increase in restaurant occupancy over time, with a slight dip during the covid-19 pandemic, followed by a recovery to pre-covid levels.

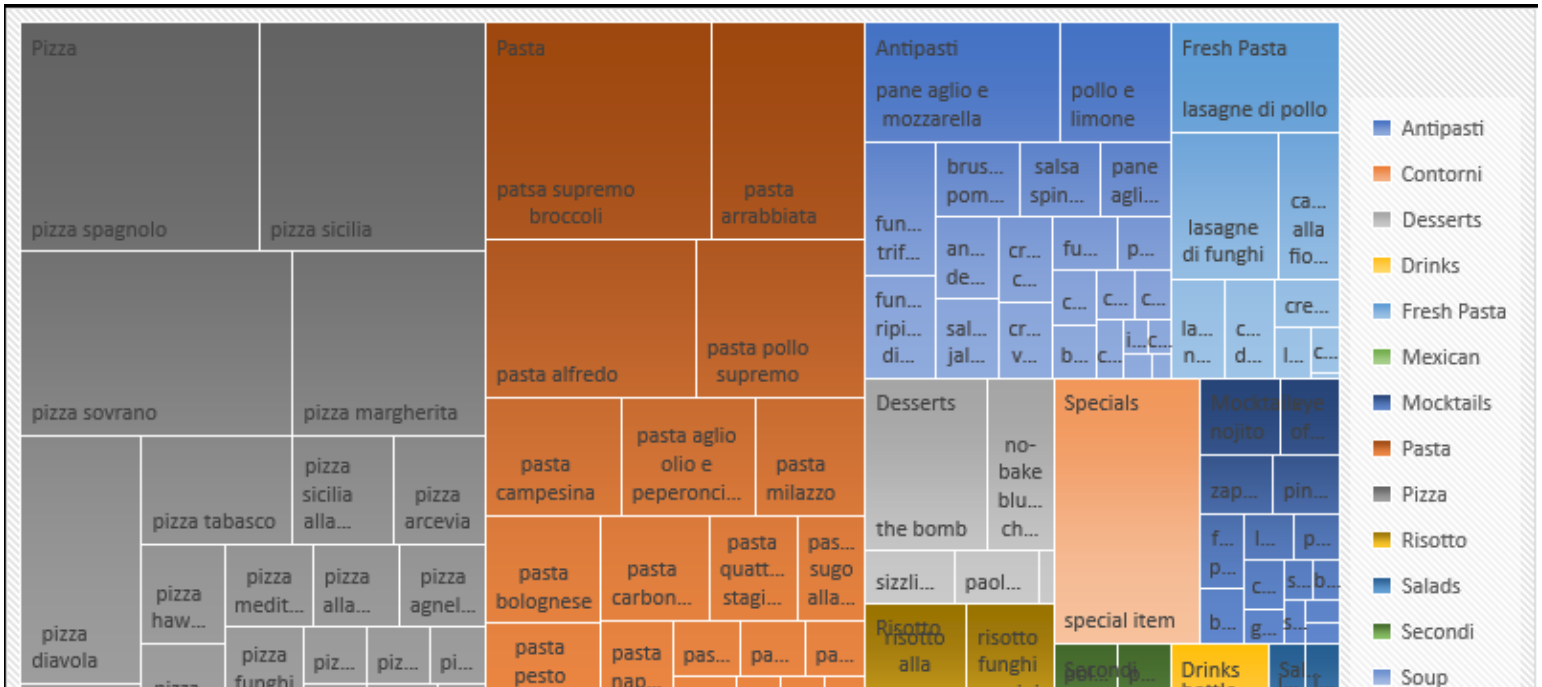


(Figure: 2.3.4)

The present graph depicts a histogram of the distribution of counts of various percentages of the total restaurant capacity that was utilized. The most frequently occurring range is the 25%-30% bucket with a count of 430-460, followed by 20%-25% range with a count of 290-310, and 30%-35% range with a count of 280-300.

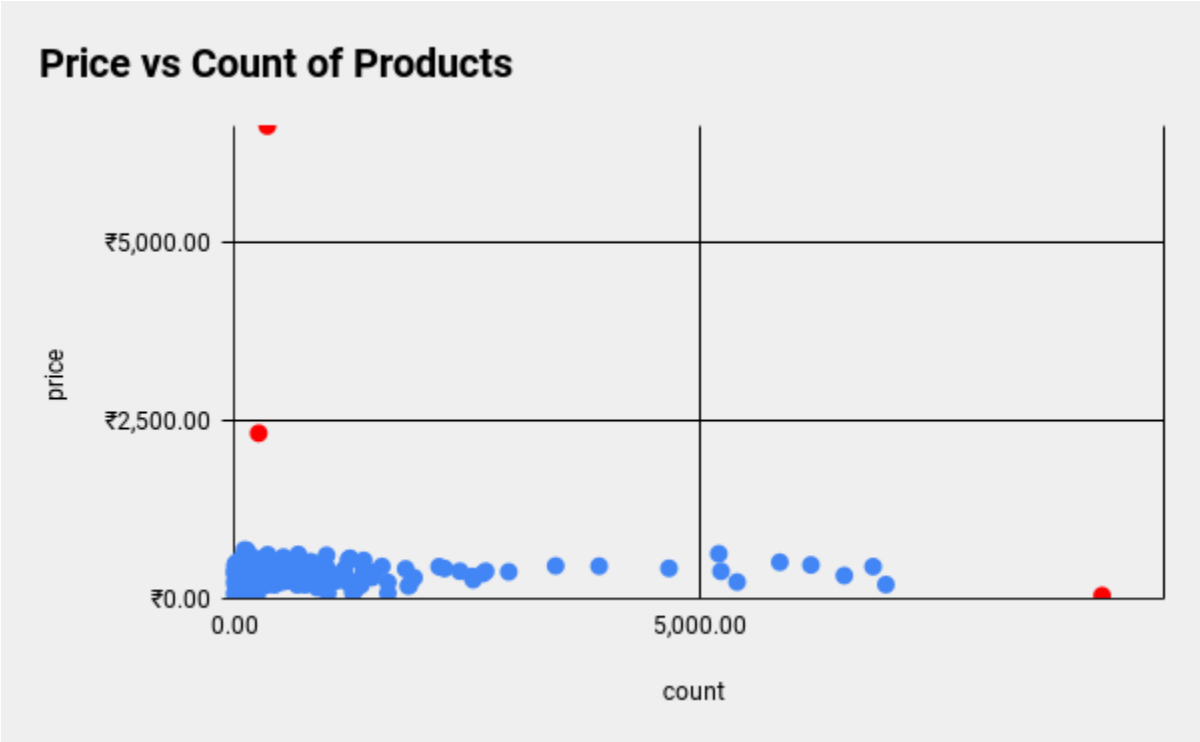
Upon discussions with the restaurant owner, it has been determined that this distribution is in line with the industry standard. It is noteworthy that the graph demonstrates another instance of the normal distribution that is ubiquitous in nature, albeit with slight right-skewness.

Objective 4 - Popularity Trendlines of Products:



by the Pasta and Antipasti categories. Furthermore, an analysis of individual products reveals that Pizza Spagnolo generates the highest revenue, followed closely by Pizza Sicillia.

The treemap is a hierarchical representation of data that uses nested rectangles to display the relative proportions of each category and product. This method of visualization is useful for quickly identifying the largest contributors to the restaurant's revenue and observing any potential patterns or trends. The dominance of the Pizza category in terms of revenue highlights the importance of this product line for the restaurant's overall success.



(Figure: 2.4.2)

This scatter chart displays the relationship between the count of products and their corresponding price in a restaurant. The red points represent outliers, which are products with unusually high or low prices, which may suggest a data-entry error or a one-off case. The majority of products are clustered in the price range of 0-800.

The black vertical line serves to partition products into two categories: high-count products that are also reasonably priced. These products are of the highest priority to the restaurant, as they generate the most revenue. Interestingly, all of the high-priority products were found to be pizzas, including "pizza sicilia," "pizza margherita," "patsa supremo broccoli," "pizza sovrano," "pizza spagnolo," "pasta alfredo," and "pizza diavola."

```
pizza spagnolo  :  3288816.0
pizza sicilia  :  3096379.0
pizza sovrano  :  3007065.0
patsa supremo broccoli  :  2940416.0
special item  :  2319442.0
pizza margherita  :  2130552.0
pasta arrabbiata  :  2014426.0
pasta alfredo  :  1996341.0
pizza diavola  :  1794924.0
pasta pollo supremo  :  1593407.0
pane aglio e mozzarella  :  1389985.0
```

(Figure: 2.4.3)

Based on revenue, the top-performing products are presented above in descending order. Consistent with the earlier findings, a majority of these products belong to the pizza category. It is important to note that the list is limited to the top 10 for the sake of presentation.

```
bottled water  :  9335
pane aglio e mozzarella  :  7008
pizza sicilia  :  6870
pizza margherita  :  6560
patsa supremo broccoli  :  6201
pizza sovrano  :  5866
the bomb  :  5406
pasta arrabbiata  :  5233
pizza spagnolo  :  5209
pasta alfredo  :  4673
```

(Figure: 2.4.4)

The above is a descending order list of products based on the number of orders. The majority of the products belong to the pizza and pasta categories, as demonstrated by the previous graphs. Interestingly, bottled water is also among the top products, which is self-explanatory. For clarity, only the top 10 products have been included for presentation purposes.

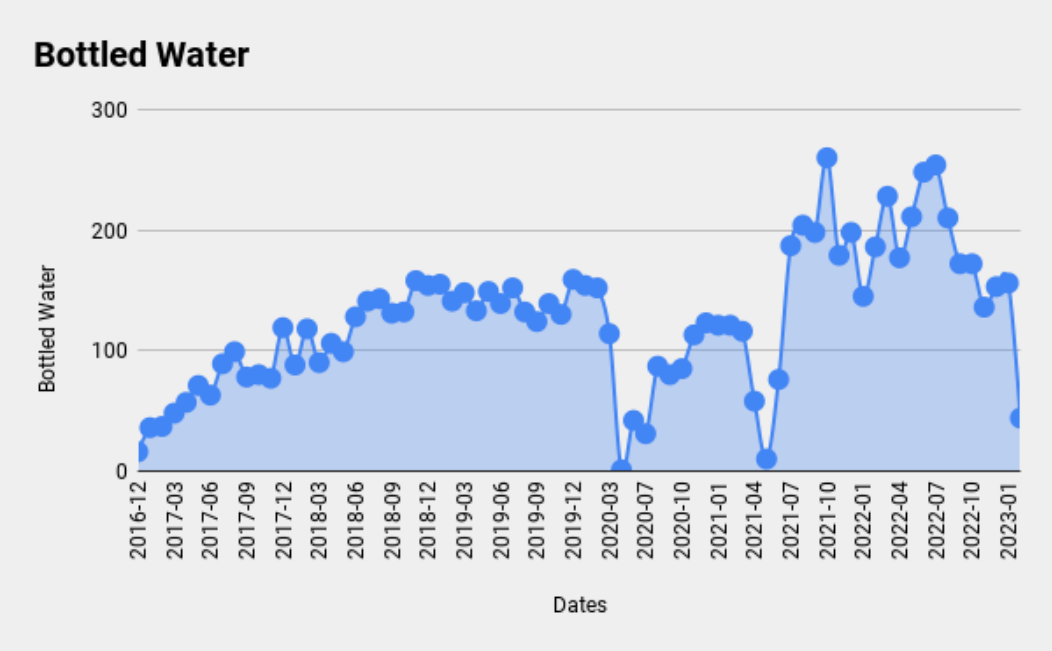
```
pizza alpi  :  -70.39285714285711
pasta pollo al forno  :  -33.33333333333337
bottled water  :  -8.716615946561461
lasagne di agnello  :  0.0
insalata tonno  :  0.0
marilyn mongoose  :  0.0
diet coke  :  0.05065856129685642
lasagne di verdure  :  1.4285714285714448
pizza bismark  :  1.6310160427807432
thums up  :  7.474747474747474
smooth operator  :  8.214285714285694
```

(Figure: 2.4.5)

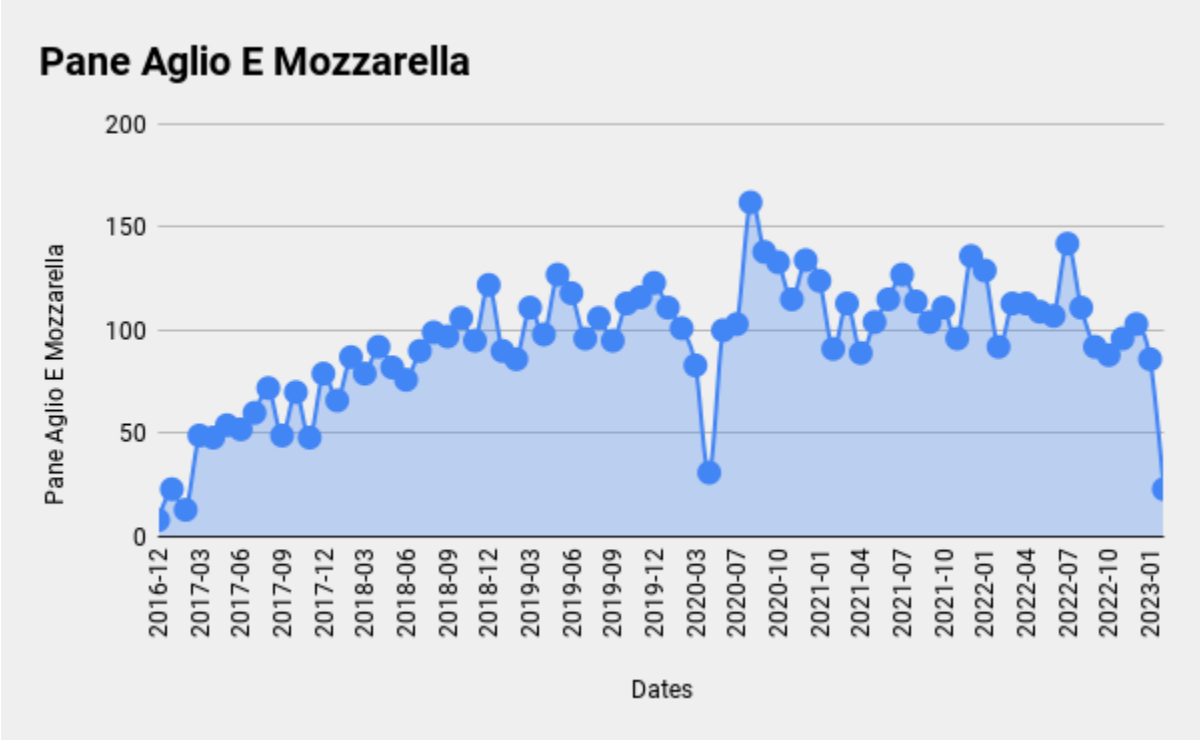
Products with the most significant decrease in popularity were obtained by comparing the average sales between 2017-2018 and 2022-2023. Refer to the Detailed Explanation section for a comprehensive analysis.

```
bottled water
pane aglio e mozzarella
pizza sicilia
pizza margherita
patsa supremo broccoli
```

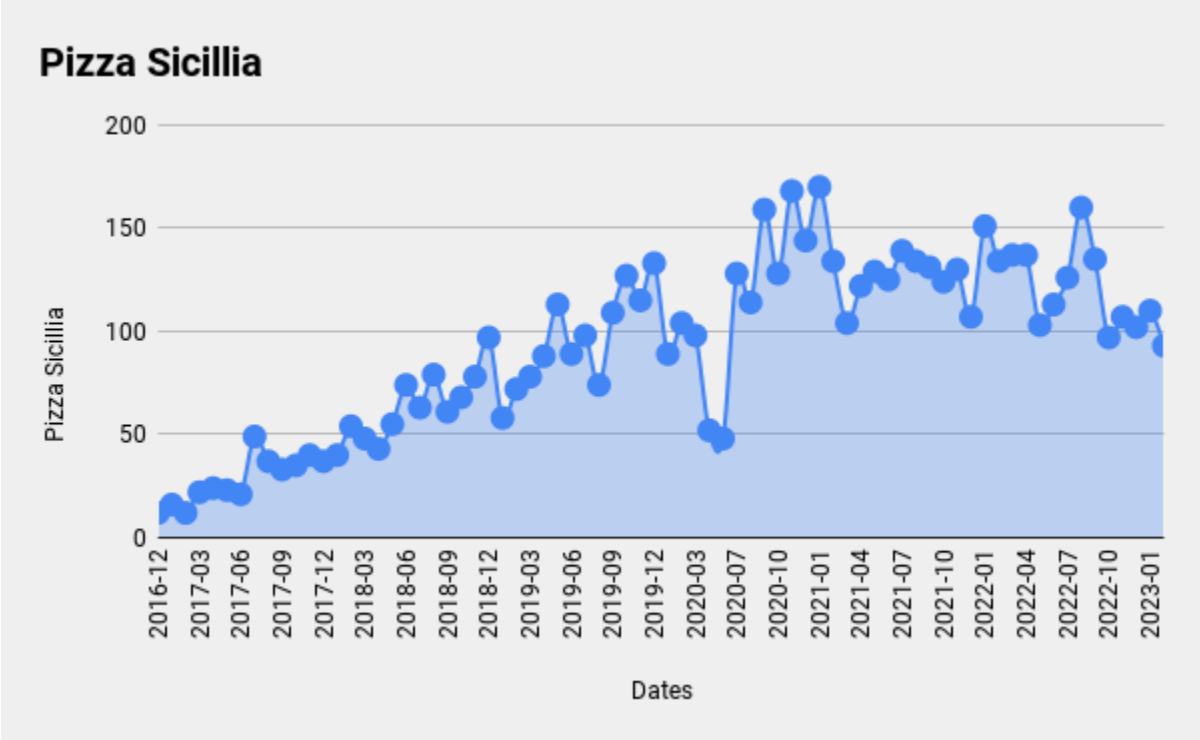
The following is a list of the top 5 products that have exhibited the greatest reduction in popularity. These products have been selected based on their high revenue and count, and are the result of a combination of the previously conducted analyses. Specifically, the products were identified by comparing the average sales figures from 2017-2018 to 2022-2023 and then cross-referenced with the products that appeared in the highest revenue and order count lists. It should be noted that a more comprehensive report detailing the entire product set will be provided to the restaurant. However, for the purposes of this presentation, we have limited the list to only 5 products.



(Figure: 2.4.7)



(Figure: 2.4.8)



(Figure: 2.4.9)

The above three line charts illustrate the count of orders of Bottled Water, Pane Aglio E Mozzarella, and Pizza Sicillia against their respective dates on the X-axis and the count of orders on the Y-axis.

Upon examination of the charts, an overall increase in the number of orders for these products is observed. This finding is supported both visually and by the R2 score of the trendline.

However, despite this increase in popularity, our analysis identified these products as having the most significant reduction in popularity. We attribute this discrepancy to the fact that the rate of increase in popularity for these products has outpaced their rate of price mark-up.

Interpretation of Results and Recommendations:

Objective 1 - Workforce Management:

As there are no clear trends in the sales data, we will use standard descriptive statistics to optimize the restaurant's workforce management. Our goal is to maintain staffing levels that are optimal, with as few idle resources as possible.

Given that the restaurant's business strategy focuses on minimizing the loss of any potential customers, we will use the maximum number of orders received during each hour over the 6-year period as our primary guide for staffing decisions.
However, we recognize that there may be outlier cases, such as corporate events or large group bookings, that could inflate the number of orders received during a particular hour. To address this issue, we will also use the 75th interquartile range (IQR) value, which can help to identify and handle outlier cases while also meeting peak demand.

Overall, we recommend using the maximum number of orders received during each hour, while also considering the 75th IQR value, as a basis for optimizing workforce management. This approach can help to ensure that the restaurant is adequately staffed to meet customer demand, while minimizing the risk of idle resources or overburdened staff.

Hour	75%	Max
12	13	23
13	44	59
14	44	72
15	27	41
16	17	31
17	14	24
18	18	36
19	34	58
20	65	98
21	66	101
22	28	46
23	1	17

(Figure: 3.1.1)
`A comprehensive table that will help the restauranter in staffing decisions.`

Objective 2 - Delivery Services Commission:

Given that delivery services account for only 0.8% of the total revenue, starting an in-house delivery service exclusively for the Banjara Hills branch may not be economically feasible. Instead, we recommend that the prices for orders obtained through delivery services be marked up to compensate for the commission paid to the delivery service providers. However, if the FSM Restaurant group decides to start a delivery service that serves all their branches in Hyderabad, it may be a more viable option to maximize profits while maintaining control over delivery operations.

Objective 3 - Takeaway-Heavy or Dine-in Heavy:

The analysis shows that takeaway and delivery services account for the majority of revenue at 57.8%, with dine-in sales contributing 42.1%. While it may be tempting to consider converting the restaurant to a cloud kitchen, the dine-in revenue is not insignificant and accounts for almost half of the total revenue. Therefore, we recommend maintaining the current seating capacity and not completely eliminating the dine-in option.

Additionally, the restaurant's occupancy pattern has returned to pre-COVID levels, and in some cases, even surpassed them. This is a positive sign, and we do not anticipate significant changes unless the FSM Banjara Hills branch decides to expand its seating capacity. Overall, we recommend maintaining the current balance between dine-in and takeaway/delivery services and continuing to monitor the occupancy patterns.

Objective 4 - Popularity Trendlines of Products:

To identify products that have experienced a reduction in popularity, we combined data on high-revenue and high-order products that also had a decrease in average sales between 2017-2018 and 2022-2023. We then graphed 3-5 of these products and found a general upward trend in their popularity over time. This suggests that the rate of increase in popularity is not aligned with the rate of price increases for these products.

Therefore, we recommend that FSM Banjara Hills consider increasing the prices of these products to better align with their increasing popularity. This could help improve revenue and profitability.

Annexure:

Link to the entire data,

Link to the Google Sheets in which all the graphs were created

Data used in Objective-1,

Jupyter notebook used in the analysis of Objective-1,

Entire set of Graphs generated for Objective-1,

Data used in Objective-2,

Jupyter notebook used for the analysis of Objective 2,

Entire set of graphs generated for Objective-2,

Data used in Objective-3,

Jupyter notebook used in the analysis of Objective-3,

Entire set of graphs generated for Objective-3,

Data used in Objective-4,

Jupyter notebook used in the analysis of Objective-4,

Entire set of Graphs generated for Objective-4