

# FoodHub Data Analysis

## Project Python Foundations

February 12, 2025

Submitted By:  
Alex Kyeremateng Botwe

# Contents / Agenda

	Topics	Page No.
	Executive Summary	5
	Business Problem Overview and Solution Approach	9
	Data Overview	13
	EDA - Univariate Analysis	18
	EDA - Multivariate Analysis	29
	Appendix	45

# List of Tables

No.	Name of Table	Page No.
1	Top 5 rows of the Dataset	12
2	Information on the Dataset	13
3	Statistical information of the Dataset	15

# List of Figures

No.	Name of Figure	Page No.
1	Univariate Analysis of Cuisine Type	20
2	Univariate Analysis of Cost of the Order	22
3	Univariate Analysis of Day of the Week	23
4	Univariate Analysis of Ratings	24
5	Univariate Analysis of Food Preparation Time	25
6	Univariate Analysis of Delivery Time	26
7	Multivariate Analysis of Cuisine vs Cost of the Order	29
8	Multivariate Analysis of Cuisine vs Food Preparation Time	31
9	Multivariate Analysis of Day of the Week vs Delivery time	33
10	Multivariate Analysis of Rating vs Delivery Time	35
11	Multivariate Analysis of Rating vs Food Preparation Time	37
12	Multivariate Analysis of Rating vs Cost of the Order	39
13	Correlation Plot	41

# Executive Summary

- A data set of about 2000 different food orders and delivery services for both weekdays and weekends by registered customers within New York was analyzed. The main objective of this analysis was to get a fair idea about the demands of different restaurants which will help us in enhancing our customer experience. The food delivery market which has been divided among many different companies with inefficient ordering and delivery processes have limited options for customers seeking a wide variety of cuisines from various restaurants across New York. The analysis identified the key elements influencing order services and specific ways in which they impact the process.
- **Business Insights:**

The following inferences were drawn:

- **The Top Performing Restaurants:** Shake Shack was leading with the highest order volume of 219 and revenue of \$3,579.53, followed by The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken and Parm.

## Executive Summary (continued)

- **Most Popular Cuisine on Weekends:** American cuisine had the highest orders with 415 orders followed by Japanese cuisines with 335 orders and Italian cuisines with 207 orders.
- **High Value Orders:** The percentage of orders that exceeded \$20 was 29.24% which relates to 555 order counts. This shows that the company's revenue can thus be increased by upselling
- **Customer Behavior:** The top 5 frequent customers placed orders between 7 to 13. Targeting these customers with loyalty programs could further increase their engagement.
- **Revenue and Ratings:** Restaurants with higher revenue tend to have strong customer ratings, particularly those with over 50 reviews and an average rating of 4.
- **Delivery Performance:** The average delivery time across all orders is 24.16 minutes, with weekdays having a slightly longer average delivery of 28 minutes than weekends which is 22 minutes on the average. It is important to note that 10.54% of orders exceeded 60 minutes.
- **Rate of Orders:** The rate of orders is higher on the weekends than the weekdays

# Executive Summary (continued)

Q17:

- **Business Recommendations:**

- Boost promotions and maintain strong customer relations with top restaurants by introducing special deals like free deliveries for higher orders.
- Encourage customer feedback by providing stimulus packages for reviews and provide advertisements for restaurants with lower ratings
- Tackle the issue of delays during peak periods by providing alternate routes and performing real time tracking of delivery staff.
- Introduce a loyalty program which offer rewards and promotions to loyal customers.
- Data should be procured for customer-specific information such as location preferences, weather conditions, restaurant staffing which may impact delivery times and build a machine learning model to accurately predict delivery times, to optimize deliveries in respective areas

# Executive Summary (continued)

- **Conclusion:**

Due to the inefficiencies in the fragmented food delivery companies in New York each seeking to provide efficient services to their customers and also increasing their revenue, I believe that FoodHub should leverage on top performing restaurants in our customer database, optimize our delivery processes, enhance customer engagements and drive high value purchases. These activities can significantly increase our revenue, customer satisfaction and operational efficiency.

These targeted initiatives will also position FoodHub for sustained business growth and be competitive in the fragmented food delivery market in New York.



# Business Problem Overview and Solution Approach

- **Problem Statement:**

The food delivery platform focuses on optimizing revenue, customer satisfaction and operational efficiency, however there are several challenges which hinder growth:

- **Uneven Performance Across Restaurants:** While top restaurants drive significant revenue, others underperform due to low visibility and customer engagement.
- **Reduced Revenue:** 29.24% of orders exceed \$20, but a lack of upselling strategies limit further revenue growth.
- **Inconsistent Delivery Efficiencies:** Average delivery times are longer on weekdays as compared to weekends and over 10% of orders take more than 60 minutes thus affecting customer retention.
- **Customer Loyalty:** Though a group of customers order frequently, there is no structured loyalty program to inspire repeat orders

# Business Problem Overview and Solution Approach (continued)

- **Underutilized Customer Feedback:** Highly rated restaurants generate strong revenue, but lower rated or less reviewed restaurants struggle with visibility and customer trust.

- **Solution Approach and Methodology:**

To address these challenges, a data-driven strategy focusing on key business levers was adopted:

- **Data Analysis and Performance Benchmarking:**
  - i. Identify high-performing restaurants and cuisines based on order volume, revenue and customer ratings.
  - ii. Determine bottlenecks in delivery times and inefficiencies affecting operations

# Business Problem Overview and Solution Approach (continued)

## ➤ Customer Segmentation and Targeted Marketing:

- i. Analyze customer behavior and create target loyalty programs for frequent customers.
- ii. Personalize promotions based on cuisine preferences to drive engagement.

## ➤ Operational Optimization:

- i. Work with delivery partners to streamline routes and minimize delays.
- ii. Implement real-time tracking and staffing adjustments to reduce delivery inefficiencies.

## ➤ Revenue Enhancement Strategies:

- i. Introduce bundle deals and upselling mechanisms to increase high-value orders.
- ii. Offer exclusive discounts and premium services for top performing restaurants.

# Business Problem Overview and Solution Approach (continued)

➤ **Customer Engagement and feedback Utilization:**

- i. Encourage customer feedback through incentives and increased visibility for under reviewed restaurants.
- ii. Prioritize high-rated restaurants for promotional campaigns to boost customer trust.

# Data Overview

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24

**Table 1: Top 5 rows of the Dataset**

**Q1:**

There are 1898 rows and 9 columns present in the data frame

# Data Overview (continued)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1898 entries, 0 to 1897
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   order_id              1898 non-null   int64
1   customer_id           1898 non-null   int64
2   restaurant_name       1898 non-null   object
3   cuisine_type          1898 non-null   object
4   cost_of_the_order     1898 non-null   float64
5   day_of_the_week       1898 non-null   object
6   rating                1898 non-null   object
7   food_preparation_time 1898 non-null   int64
8   delivery_time         1898 non-null   int64
dtypes: float64(1), int64(4), object(4)
memory usage: 133.6+ KB
```

**Table 2: Information on the Dataset**

**Q2:**

There are 3 data types namely: **int64(4)** - order\_id, customer\_id, food\_preparation\_time, delivery\_time

**float64(1)** - cost\_of\_the\_order

**object(4)** - restaurant\_name, cuisine\_type, day\_of\_the\_week, rating

## Data Overview (continued)

Q3:

```
order_id          0
customer_id       0
restaurant_name    0
cuisine_type      0
cost_of_the_order  0
day_of_the_week   0
rating            0
food_preparation_time 0
delivery_time     0
dtype: int64
```

There are no missing values in the data

# Data Overview (continued)

	order_id	customer_id	cost_of_the_order	food_preparation_time	delivery_time
<b>count</b>	1.898000e+03	1898.000000	1898.000000	1898.000000	1898.000000
<b>mean</b>	1.477496e+06	171168.478398	16.498851	27.371970	24.161749
<b>std</b>	5.480497e+02	113698.139743	7.483812	4.632481	4.972637
<b>min</b>	1.476547e+06	1311.000000	4.470000	20.000000	15.000000
<b>25%</b>	1.477021e+06	77787.750000	12.080000	23.000000	20.000000
<b>50%</b>	1.477496e+06	128600.000000	14.140000	27.000000	25.000000
<b>75%</b>	1.477970e+06	270525.000000	22.297500	31.000000	28.000000
<b>max</b>	1.478444e+06	405334.000000	35.410000	35.000000	33.000000

**Table 3: Statistical information of the data set**

**Q4:**

From the statistical summary

- Minimum time it takes for food to be prepared once an order is placed is 20 minutes



## Data Overview (continued)

- i. Average time it takes for food to be prepared once an order is placed is 27 minutes 37seconds
- ii. Maximum time it takes for food to be prepared once an order is placed is 35minutes

Q5:

The number of orders not rated are 736

```
rating
Not given    736
5            588
4            386
3            188
Name: count, dtype: int64
```

## Q6: EDA - Univariate Analysis

➤ **Order ID:**

- The total number of unique orders placed both on weekdays and weekends were 1898

➤ **Customer ID:**

- The total number of customers who placed orders on weekdays and weekends were 1200. This indicates that some customers placed more than one order.

Q7:

➤ **Restaurant Name:**

- There was a total of 178 restaurant in New York that were considered in this analysis. Below are the top 5 restaurants in terms of number of orders received:
- **Shake Shack** stands out as the top restaurant with 219 orders
- **The Meatball Shop** ranks second with 132 orders
- **Blue Ribbon Sushi** is the third with 119 orders
- **Blue Ribbon Fried Chicken** is fourth with 96 orders
- **Parm** is the fifth with 68 orders

## ➤ Cuisine Type:

- We identify 14 different cuisines across all restaurants.

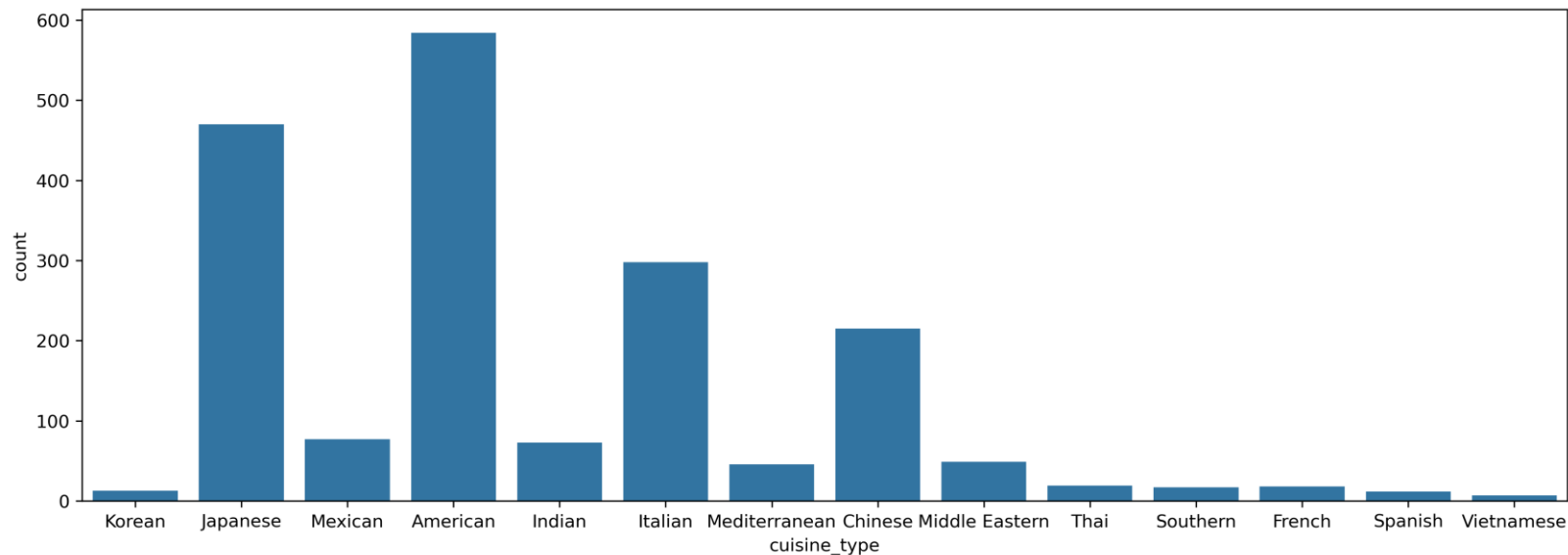


Fig 1: Univariate Analysis of Cuisine Type

## Observations:

- The most popular cuisine on weekends are American with 415 orders followed by Japanese with 335 orders then Italian with 207 orders and finally Chinese with 163 orders.
- The least popular cuisines are Vietnamese, Korean, Spanish, French, Thai and Southern
- The distribution is uneven with certain cuisines significantly outperforming others

➤ **Cost of the Order:**

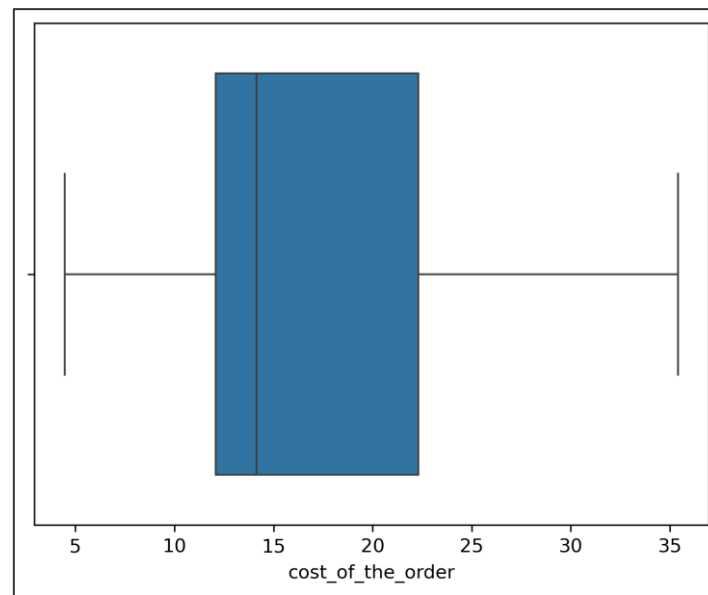
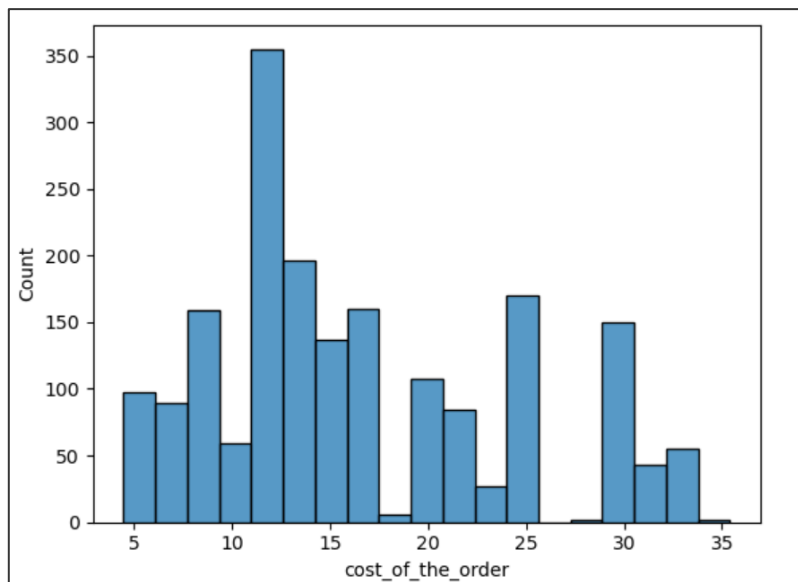


Fig 2: Univariate Analysis of Cost of the Order

## Observations:

- The distribution is right skewed meaning, although the majority of orders tend to be on the lower end of the price range, there are a few significantly high-priced orders.

### ➤ Day of the Week:

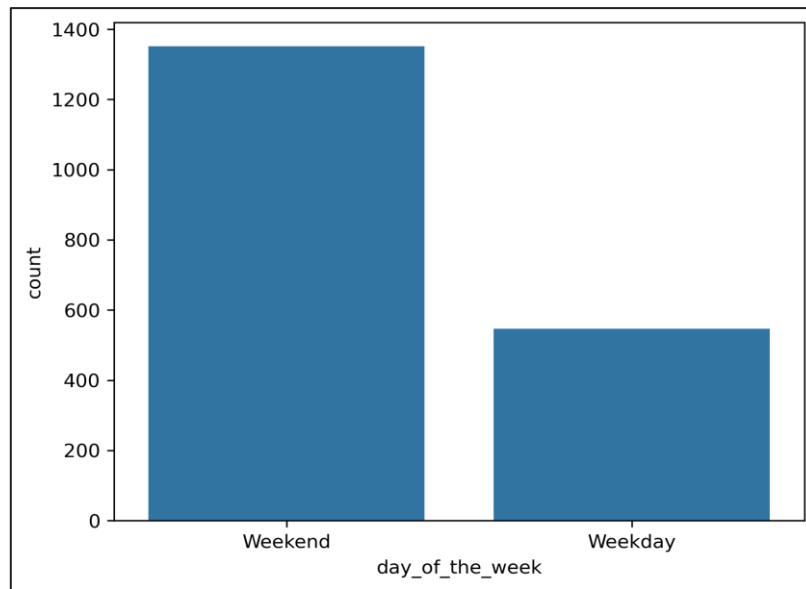


Fig 3: Univariate Analysis of Day of the Week

## Observations:

- A higher number of orders were place on weekends than the weekdays

## ➤ Rating:

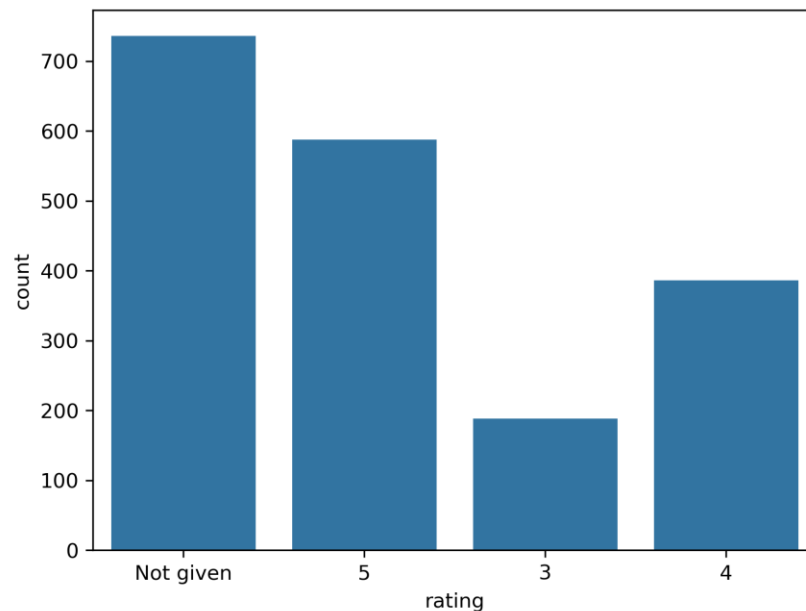


Fig 4: Univariate Analysis of Ratings



## Observations:

- The distribution is moderately right skewed as the top-rated restaurant dominate with rating of 5. Also a larger number of customers did not give ratings.

### ➤ Food Preparation Time:

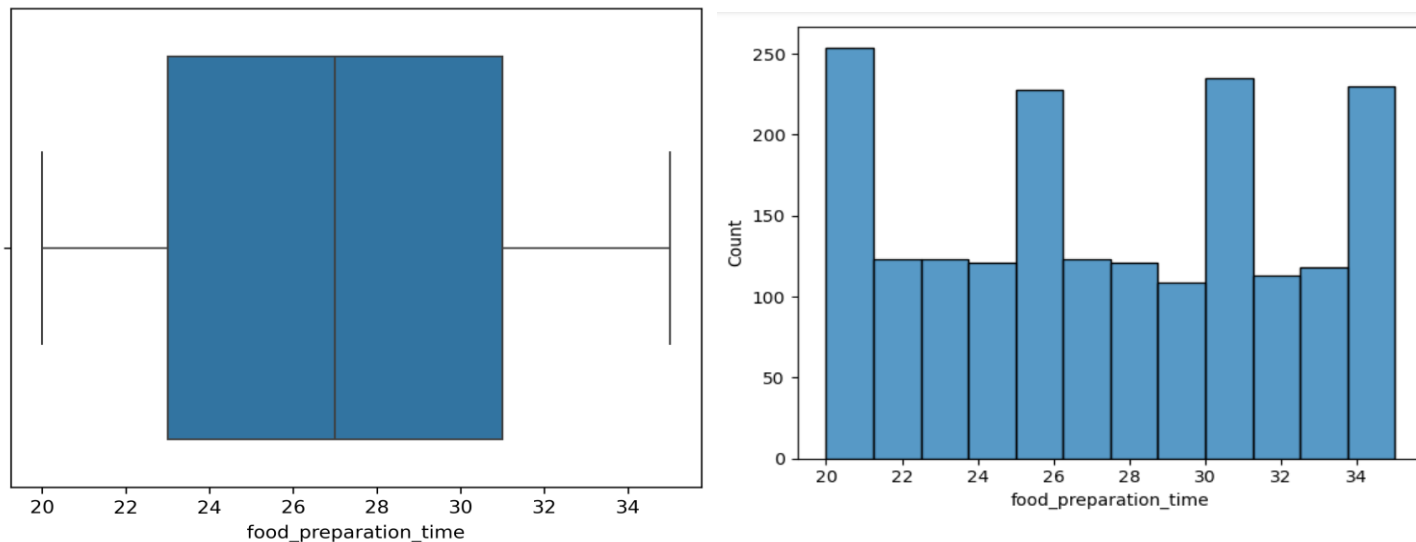


Fig 5: Univariate Analysis of Food Preparation Time

## Observations:

- This is a no skewed distribution suggesting that the data is symmetrically distributed and thus we can infer that the food preparation time is approximately the same

### ➤ Delivery Time:

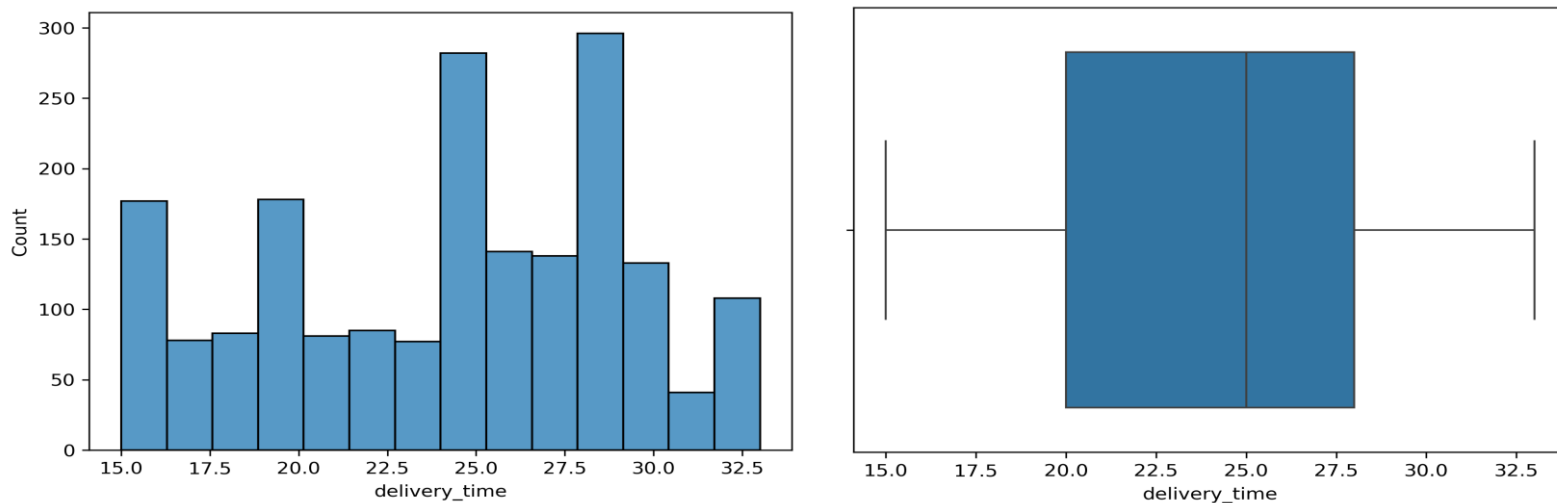


Fig 6: Univariate Analysis of Delivery Time

## Observations:

- Mean delivery time is 24.16 minutes and Median which is 25.00 min are close, indicating a fairly symmetric distribution with a slightly left skew.
- A peak around 20-28 minutes, confirming most deliveries fall in this range
- Most deliveries therefore occur within a predictable range of 20-28 minutes, making it a reliable timeframe for setting customer expectations. However, a few deliveries take longer
- The total number of orders that cost above 20 dollars is 555
- The percentage orders above 20 dollars is 29.24%

**Q8:** The most popular cuisine on weekends are American.

**Q9:** 29.24% of orders exceed \$20.

**Q10:** The mean order delivery time across all orders is 24.16 minutes.

**Q11:** The top most frequent customers who were selected to receive 20% discount vouchers placed between 7 to 13 orders. Below are their IDs and the respective orders they placed:

```
customer_id
52832      13
47440      10
83287       9
250494      8
259341      7
Name: count, dtype: int64
```

## Q12: EDA - Multivariate Analysis

### ➤ Cuisine vs Cost of the Order:

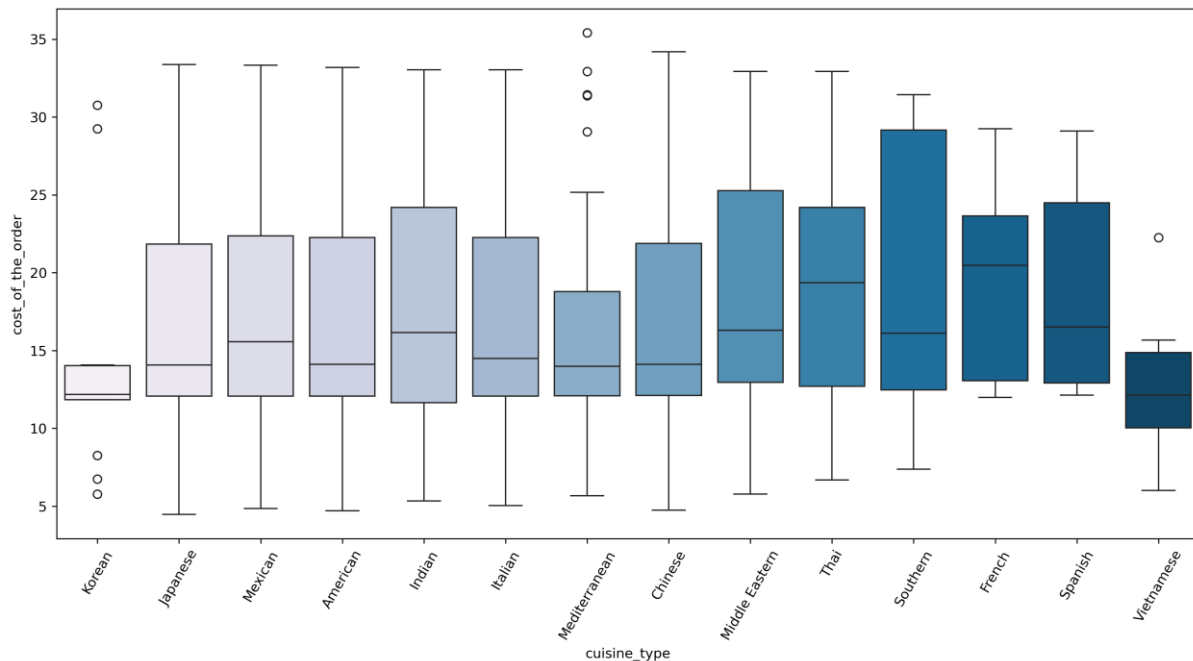


Fig 7: Multivariate Analysis of Cuisine vs Cost of the Order

Analysis were made on the average order cost across different cuisine types to identify trends in pricing.

### **Observations:**

- This helps determine which cuisines have higher spending per order and which rely more on frequent, lower-value orders.
- Some cuisines on the average tend to be more expensive, while others tend to be more affordable.
- Certain cuisines show inconsistency in their pricing which indicates that there might be a mixture of both budget and premium options.
- There are visible outliers which indicates that some expensive orders with respect to specific cuisines.
- French, Southern, and Thai cuisines have the highest average order costs, while Korean and Vietnamese cuisines tend to be more affordable.

➤ Cuisine vs Food Preparation Time:

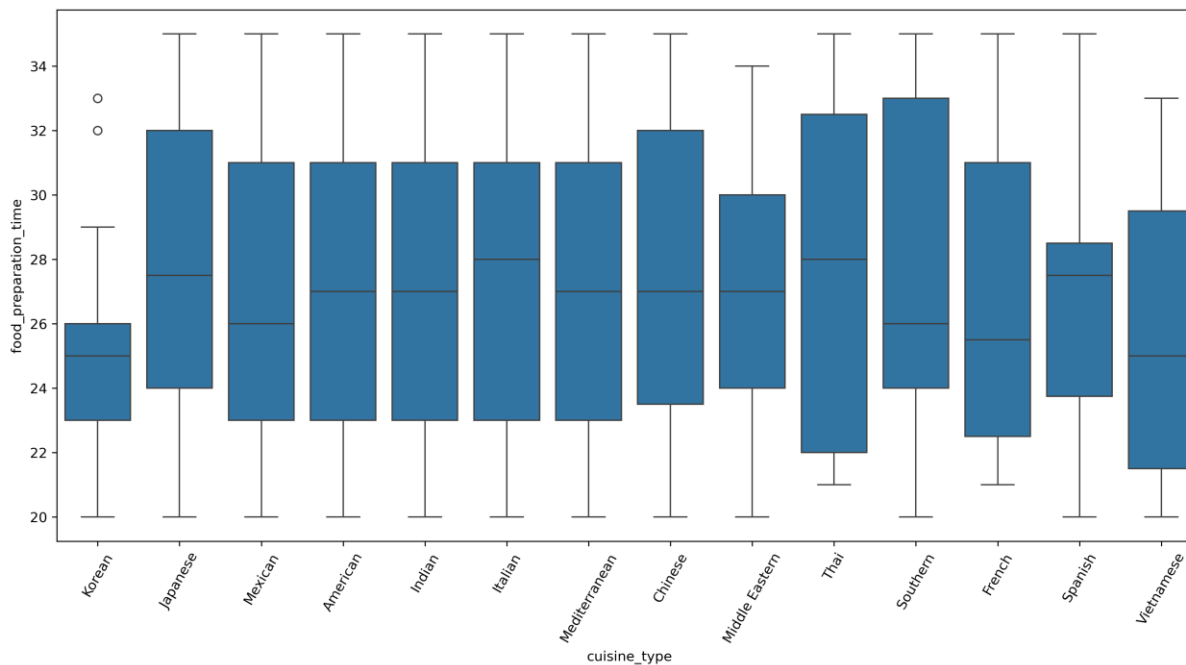


Fig 8: Multivariate Analysis of Cuisine vs Food Preparation Time

Analysis were carried out on the food preparation time across different cuisine types.

### **Observations:**

- Cuisines, like Middle Eastern and Southern, have higher median preparation times, indicating that they take longer on average to prepare.
- Other cuisines, like Vietnamese and Korean, have lower median preparation times, meaning they take less time to prepare.
- There is visible inconsistency in certain cuisines preparation times, probable due to the complexity in preparation
- The few outliers for example in the Korean cuisine indicates that these take extremely longer periods to prepare.



➤ **Day of the Week vs Delivery Time:**

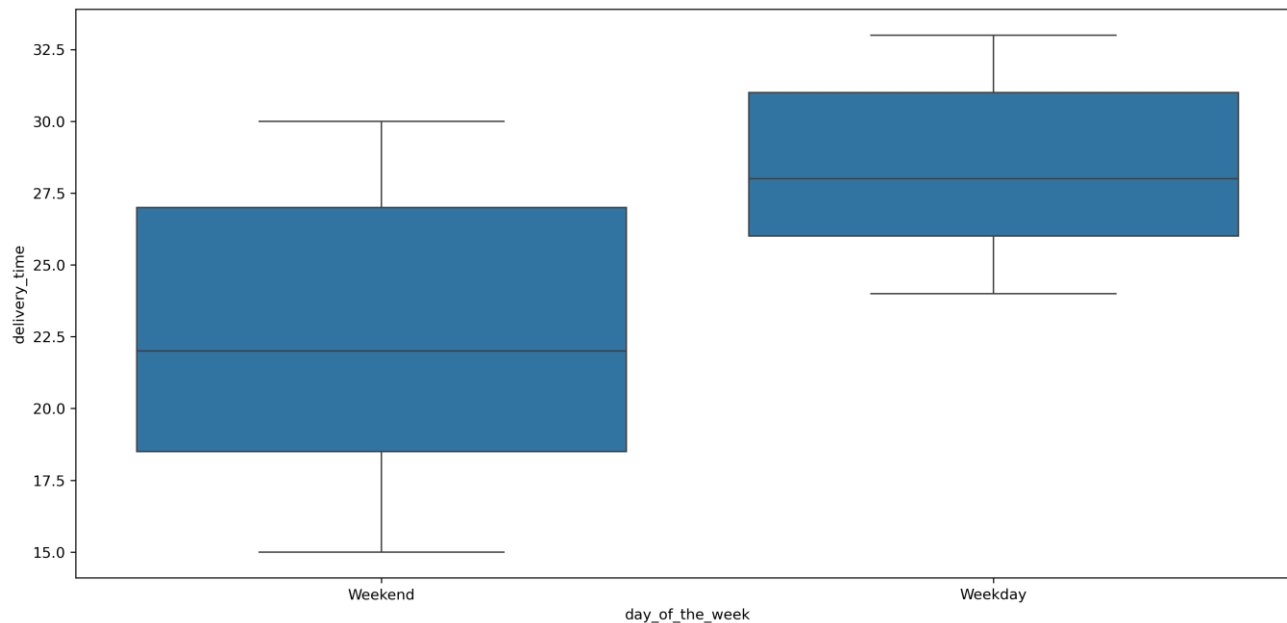


Fig 9: Multivariate Analysis of Day of the Week vs Delivery time

The boxplot shows how delivery time varies on weekdays and weekends:

**Observations:**

- Weekends tend to have longer delivery times, with a higher median and more variability.
- Weekdays have relatively similar delivery times, but some days show slightly lower or higher medians.
- Outliers on certain days indicate occasional extreme delays.
- Shake Shack dominates in terms of cost of order followed by The Meatball Shop and then Blue Ribbon Sushi.
- The restaurant with the least cost of orders are Five Guys Burgers and Fries, Nobu Next Door and Sushi of Gari 46

➤ **Rating vs Delivery Time:**

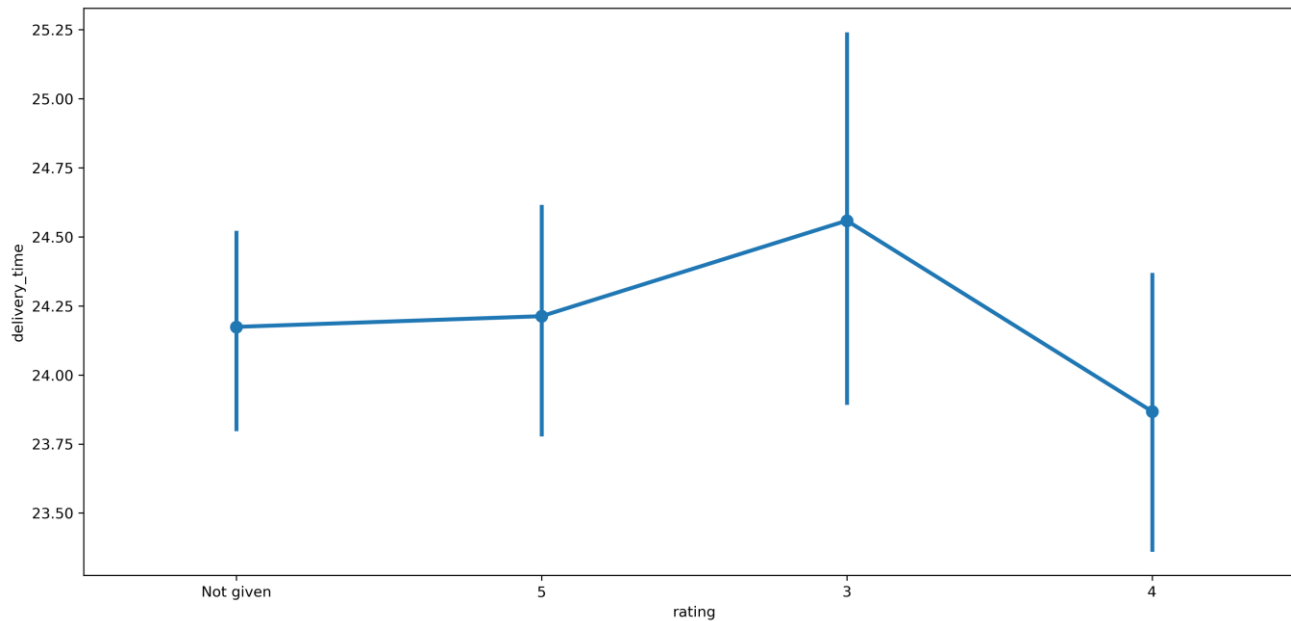


Fig 10: Multivariate Analysis of Rating vs Delivery Time

## Observations:

- Orders with lower ratings typically have longer delivery time. For example, orders rated 1 or 2 have a higher median delivery time, indicating longer wait time and this can contribute to customer dissatisfaction.
- Delivery times for purchases with higher ratings of 4 to 5 are shorter and more reliable, suggesting that quicker deliveries frequently result in better customer experiences
- The mid-range ratings of 3 show a great deal of variation. Even orders with average ratings had lengthy delivery periods, indicating that there are other factors influencing customer satisfaction.

➤ Rating vs Food preparation Time:

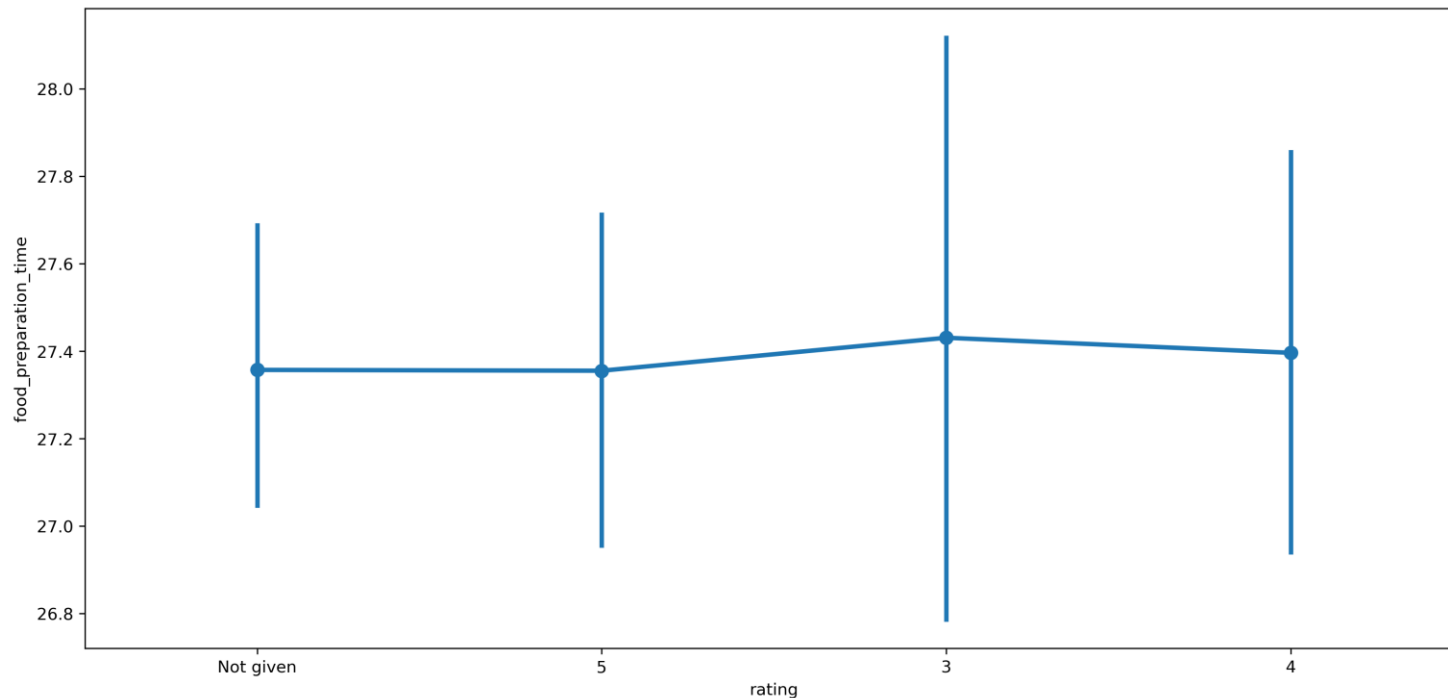
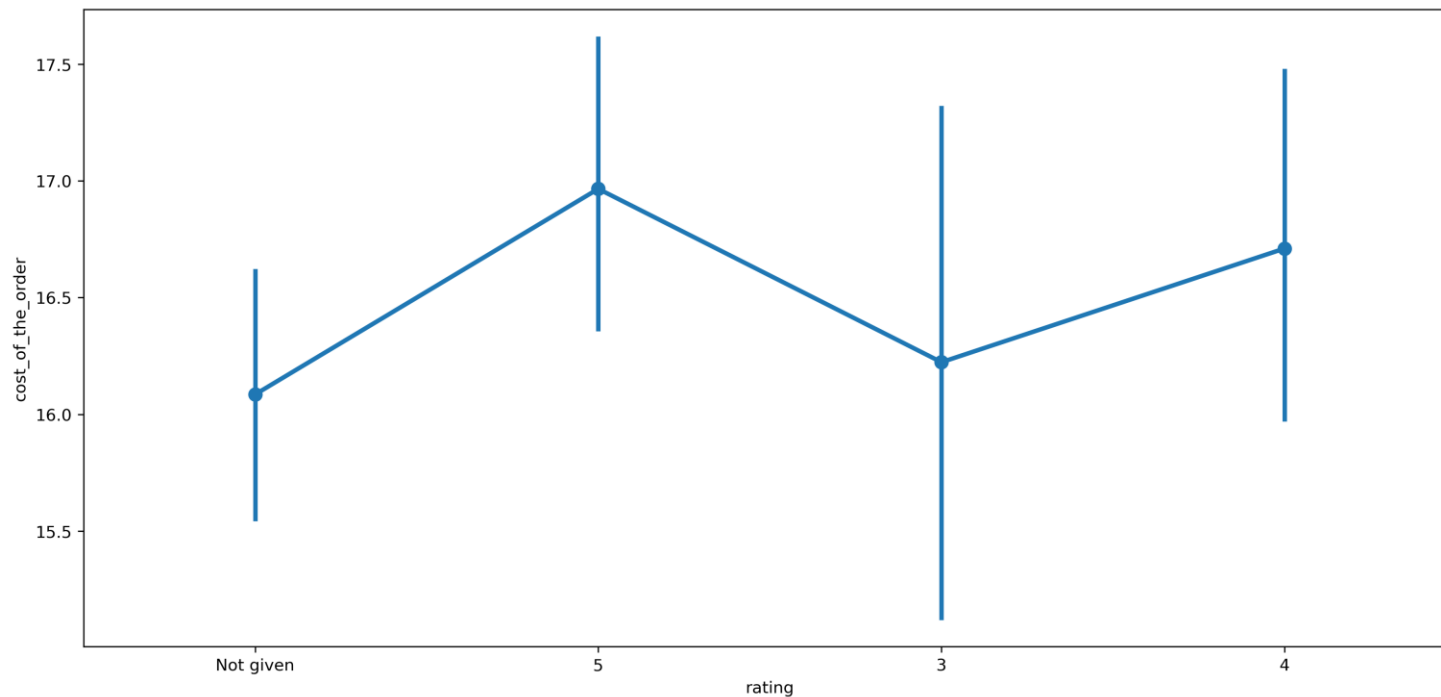


Fig 11: Multivariate Analysis of Rating vs Food Preparation Time

## Observations:

- Orders with higher ratings of 4-5 typically take a little less time to prepare, indicating that faster food preparation increases customer satisfaction.
- Average preparation hours are higher for orders with lower ratings, indicating that long prep times might be a contributor to negative reviews.
- Moderate preparation times are associated with ratings of 3, but the pattern indicates that customers may accommodate delays if the food quality meets their expectations.
- Overall, the pattern indicates that meal preparation time decreases slightly as ratings increase, supporting the notion that higher ratings are a direct result of quicker service.

➤ **Rating vs Cost of the Order:**



**Fig 12: Multivariate Analysis of Rating vs Cost of the Order**

## Observations:

- The average cost of orders with better ratings of 4-5 are higher, indicating that customers may be happier with meals that cost more.
- Orders with lower ratings of 1-2 seems to have lower cost, which suggest that less expensive meals are identified as having lesser quality or value.
- Price alone may not be the determining factor for evaluations, as ratings of 3 indicate a modest priced range that suggests mixed experiences.
- Although the trend indicates a little increase in price with higher ratings, the difference is not substantial, indicating that other factors such as the quality of the food or the service also play a considerable role in ratings.



➤ Correlation among Variables:

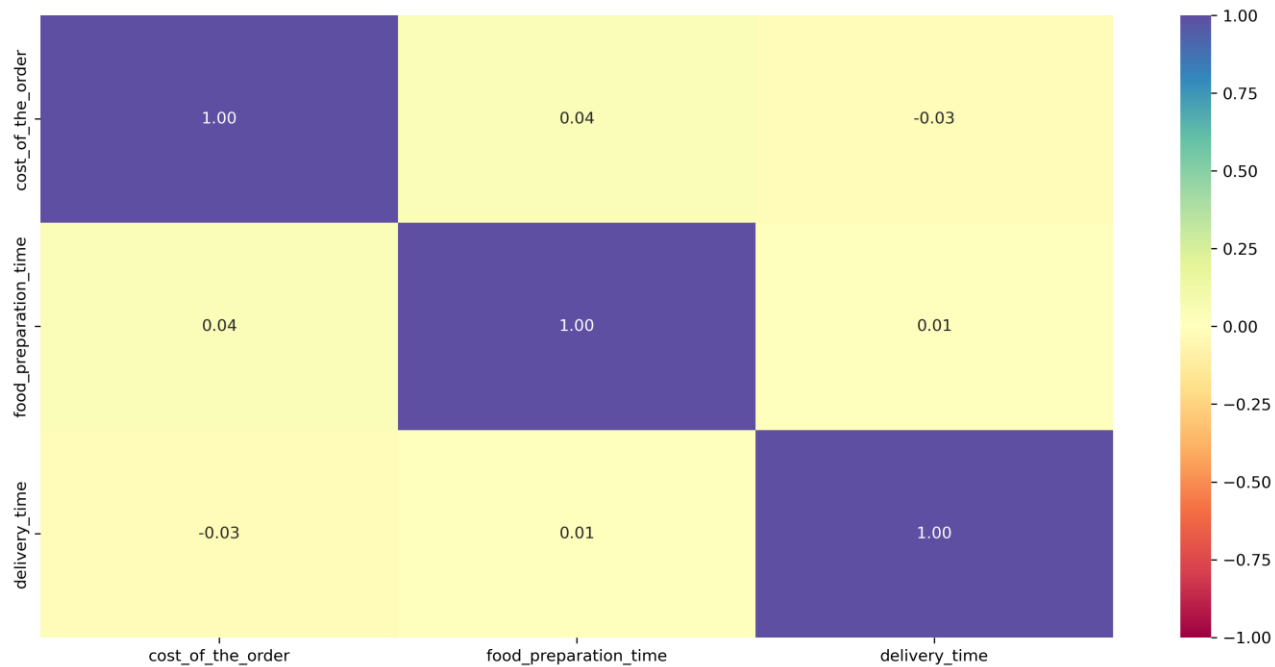


Fig 13: Correlation Plot

## Observations:

- Food preparation time and delivery time show almost no correlation – The food preparation times generally do not have any impact on delivery times.
- Cost of the order has almost no correlation with both preparation and delivery times – This suggests the cost of orders do not necessarily have any impact on prepare time or the delivery time.
- Food preparation time and cost of the order have almost no correlation – The price of an order does not strongly impact how long it takes to prepare.

### Q13: Restaurants Eligible for the Promotional Offer:

- The following restaurants meet the criteria of having more than 50 ratings and an average rating greater than 4:
  - Shake Shack – 133 ratings, 4.28 average rating
  - The Meatball Shop – 84 ratings, 4.51 average rating
  - Blue Ribbon Sushi – 73 ratings, 4.22 average rating
  - Blue Ribbon Fried Chicken – 64 ratings, 4.33 average rating

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time	Revenue	total_time
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20	7.6875	45
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23	1.8120	48
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28	1.8345	51
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15	7.3000	40
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24	1.7385	49

**Table 4: Net Revenue**

**Q14:** The company has earned a total revenue of \$6,166.30 from commissions across all orders

**Q15:** The percentage of orders with total delivery time over 60 minutes is 10.54%

**Q16:** The mean delivery time on weekdays is around 28 minutes

The mean delivery time on weekends is around 22 minutes

# APPENDIX

## A. Data Sources

The analysis was conducted using the [FoodHub](#) Orders Dataset, which includes:

- Cost of the Order, Cuisine Type, Restaurant Name, Customer ID, Order ID.
- Ratings, Day of the week.
- Food Preparation Time, Delivery Time.

## B. Data Preprocessing Steps

The dataset was preprocessed to ensure accuracy and consistency before the analysis was conducted:

- I checked for missing values in all 9 columns.
- No imputations were made because the dataset had no missing values.
- Food\_Preparation\_Time was converted to the right data type.

## C. Analytical Methods:

- Descriptive statistics were used to understand certain key indicators like average ratings, revenue distribution, and delivery times.
- Correlation between the variables, cost\_of\_the\_order, food\_preparation\_time, and delivery\_time were investigated.
- Heatmaps, bar charts, and point plots were used to determine trends and insights
- Restaurants were grouped by revenue, performance and ratings in order to determine the top performing restaurant.



## D. Commissions on Orders:

- There was 25% commission on orders greater \$20
- There was 15% commission on orders greater than \$5
- There were no commission on orders less than or equal to \$5

## E. Glossary

- Net Revenue: The total commission earned by FoodHub from restaurant orders.
- Rating: The number of ratings received by a restaurant.
- Average Rating: The mean rating score of a restaurant's orders.
- Delivery Time: The time taken for the order to reach the customer after preparation.
- Food Preparation Time: The time taken by the restaurant to prepare an order.

## References

Great Learning. (n.d.) *Python - Foundations*. **Great Learning**.

[https://olympus.mygreatlearning.com/courses/124963/modules/items/6398188?pb\\_id=18483](https://olympus.mygreatlearning.com/courses/124963/modules/items/6398188?pb_id=18483)

Reva Institute of Technology & Management. (2024, September 15). *Analyzing FoodHub Data for Restaurant Demand*. **CliffNotes**. <https://www.cliffsnotes.com/study-notes/19714192>



**Happy Learning !**

