BUSINESS REPORT

A. Understanding the structure of the data

1. How many rows and columns are present in the data?

Approach Used; The code snippet utilizes the shape attribute of the DataFrame to determine the dimensions of the dataset, the number of rows and columns. By accessing this attribute we obtain a tuple that represents these dimensions.

Insights; In total there are 1898 rows and 9 columns, in this dataset

Inferences; The dataset contains 1898 records, where each record corresponds to a food order. The 9 columns within the dataset provide information about aspects of each order including order details, customer information, restaurant details, cuisine type, cost, day of the week rating, food preparation time and delivery time.

Output; Number of rows: 1898 Number of columns: 9

2. What are the datatypes of the different columns in the dataset?

Approach Used: info () function is used to display the data types of the different columns in the dataset.

Insights: The dataset consists of columns with various data types:

- 1 column with data type **float64**
- 4 columns with data type **int64**
- 4 columns with data type **object** (likely representing strings or categorical values)

Inferences: This observation provides a comprehensive overview of the dataset's structure, including column names, data types, and the number of non-null entries. Understanding the data types is crucial for performing appropriate analyses and conversions as needed. The **object** data type columns may require further investigation to ensure proper handling of categorical data and text values during analysis.

Output; <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

3. Check the statistical summary of the data. What is the minimum, average, and maximum time it takes for food to be prepared once an order is placed?

Approach Used: The **describe()** function is employed to generate a detailed statistical summary of the dataset's numerical columns.

Insights:

- The dataset contains a total of 1898 records.
- The **cost_of_the_order** column shows a mean cost of approximately $16.50, with a minimum cost of $4.47 and a maximum cost of $35.41.
- The **food_preparation_time** column has an average preparation time of approximately 27.37 minutes, with a minimum time of 20 minutes and a maximum time of 35 minutes.
- The **delivery_time** column shows an average delivery time of approximately 24.16 minutes, with a minimum time of 15 minutes and a maximum time of 33 minutes.

Inferences:

- The range of order costs varies considerably, suggesting diversity in customer spending.
- Both food preparation and delivery times appear to be within reasonable ranges, indicating efficient operations.

Output; order_id
count 1.898000e+03
mean 1.477496e+06
std 5.480497e+02
min 1.476547e+06
25% 1.477021e+06
50% 1.477496e+06
75% 1.477970e+06
max 1.478444e+06

customer_id
count 1898.000000
mean 171168.478398
std 113698.139743
min 1311.000000
25% 77787.750000
50% 128600.000000
75% 270525.000000

```
max    405334.000000

restaurant_name
count          1898
unique          178
top       Shake Shack
freq            219
Name: restaurant_name, dtype: object

cuisine_type
count         1898
unique          14
top       American
freq           584
Name: cuisine_type, dtype: object

cost_of_the_order
count   1898.000000
mean      16.498851
std        7.483812
min        4.470000
25%       12.080000
50%       14.140000
75%       22.297500
max       35.410000
Name: cost_of_the_order, dtype: float64

day_of_the_week
count       1898
unique         2
top      Weekend
freq        1351
Name: day_of_the_week, dtype: object

rating
count         1898
unique           4
top       Not given
freq           736
Name: rating, dtype: object

food_preparation_time
count    1898.000000
mean       27.371970
std         4.632481
min        20.000000
```

```
25%       23.000000
50%       27.000000
75%       31.000000
max       35.000000
Name: food_preparation_time, dtype: float64
delivery_time
count    1898.000000
mean       24.161749
std         4.972637
min        15.000000
25%        20.000000
50%        25.000000
75%        28.000000
max        33.000000
Name: delivery_time, dtype: float64
```

4. How many orders are not rated

Approach Used:

To calculate the number of orders that are not rated, I filtered the DataFrame to select rows where the 'rating' column has the value 'Not given', indicating that the order was not rated. Then, I used the **.count()** function to calculate the total count of such orders.

Insights Gained:

There are 736 orders in the dataset that have not been rated by customers.

Inferences:

The high number of orders that are not rated could indicate several things. It might be due to customers forgetting to provide a rating, being indifferent about giving a rating, or possibly a user interface issue that needs investigation. Encouraging customers to provide ratings can lead to better insights into the quality of service and food, and it can also help the company identify areas for improvement.

Output; Number of orders not rated: 736

B. Description of criterion

5. Explore all the variables and provide observations on their distributions. (Generally, distinct number of values, histograms, boxplots, countplots, etc. are used for univariate exploration.)

Approach used; The approach involves using Seaborn to create histograms and countplots for numerical and categorical variables. Histograms with KDE are generated for numerical variables, and countplots are created for categorical variables. The order of x-axis categories is based on value counts. These visualizations reveal insights about order distribution, customer behavior, cuisine popularity, and more, aiding decision-making.

Insights;

Variable: order_id

- All order IDs are unique, indicating that each row corresponds to a unique order.
- No missing values are present in this column.

Variable: customer_id

- There are a total of 1200 unique customer IDs, suggesting that the dataset contains orders from 1200 different customers.
- The customer with ID 52832 has the highest number of orders (13 orders), followed by customers with IDs 47440 and 83287 (10 and 9 orders respectively).
- No missing values are present in this column.

Variable: restaurant_name

- There are a total of 178 unique restaurant names.
- Shake Shack is the most frequent restaurant, with 219 orders, followed by The Meatball Shop (132 orders) and Blue Ribbon Sushi (119 orders).
- No missing values are present in this column.

Variable: cuisine_type

- There are 14 unique cuisine types in the dataset.
- American cuisine has the highest number of orders (584), followed by Japanese (470 orders) and Italian (298 orders).
- No missing values are present in this column.

Variable: cost_of_the_order

- There are 312 unique values in the cost_of_the_order column, indicating variation in the order costs.
- The most common order cost is $12.18, with 86 orders at this price.
- No missing values are present in this column.

Variable: day_of_the_week

- Most of the orders (1351 out of 1898) were placed on weekends, while the rest (547 orders) were placed on weekdays.
- No missing values are present in this column.

Variable: rating

- The majority of ratings are not given (736 orders), followed by a rating of 5 (588 orders), 4 (386 orders), and 3 (188 orders).
- No missing values are present in this column.

Variable: food_preparation_time

- The food preparation time varies across a range of values from 20 to 35 minutes.
- The most common preparation time is 21 minutes, with 135 orders having this preparation time.
- No missing values are present in this column.

Variable: delivery_time

- Delivery time varies across a range of values from 15 to 33 minutes.

- The most common delivery time is 24 minutes, with 162 orders having this delivery time.
- No missing values are present in this column.

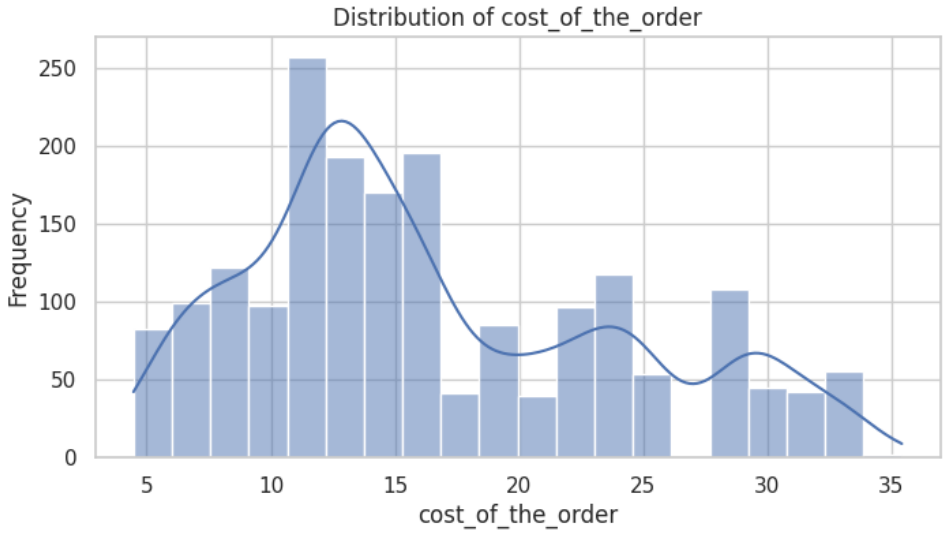
Overall Insights:

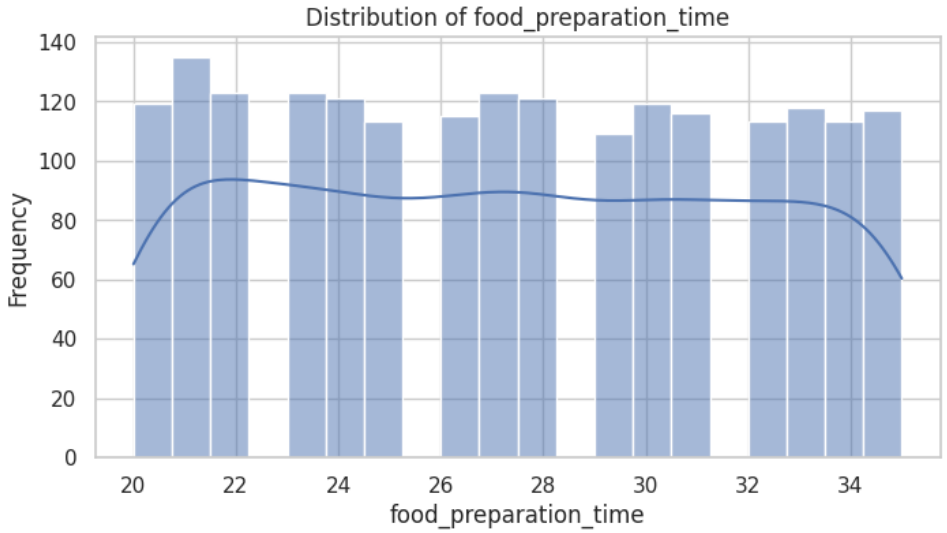
- The dataset contains orders from 1200 unique customers and 178 unique restaurants with a variety of cuisine types.
- Most orders are placed on weekends, and the majority of ratings are not provided.
- Food preparation time and delivery time show variations across different time ranges.

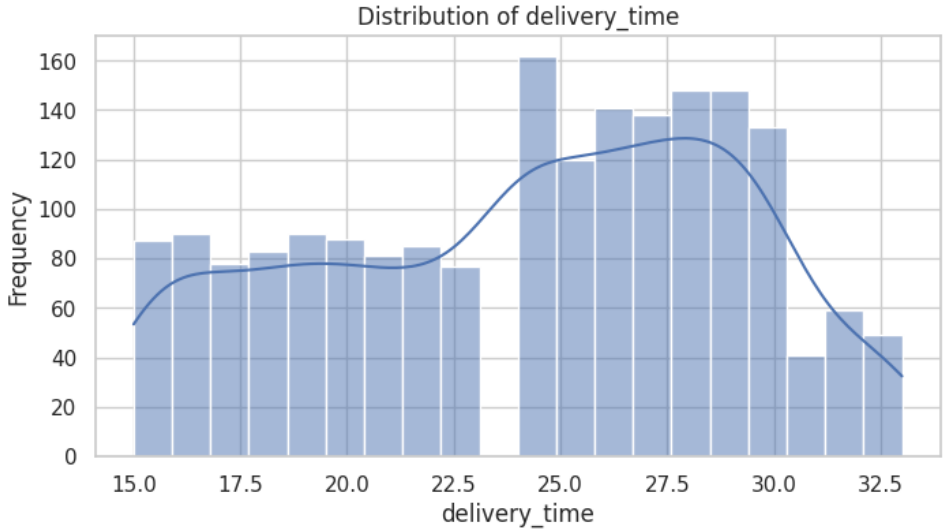
Inferences;

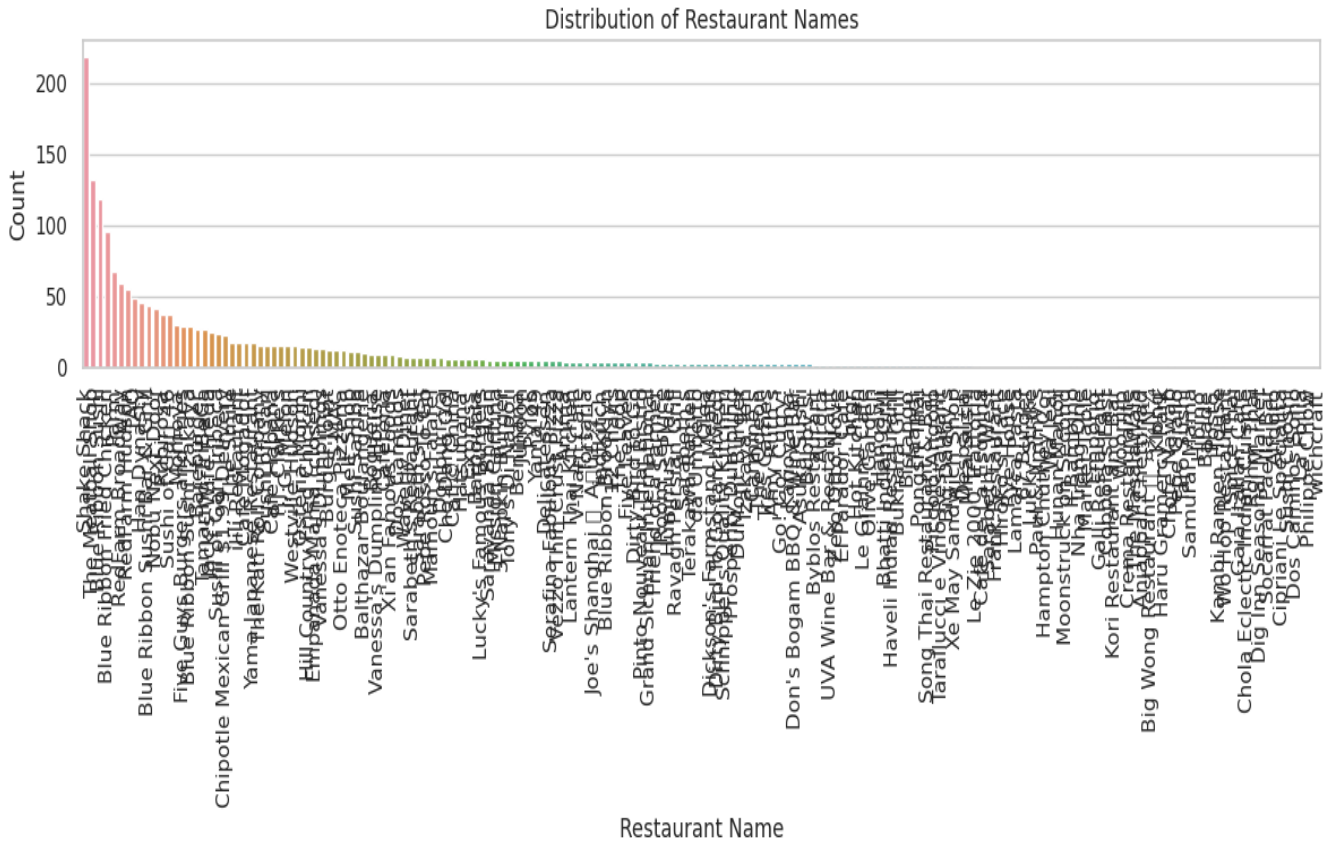
- The business tends to have higher activity on weekends, which could be due to increased customer demand.
- The frequency of ratings not being given suggests that customers might not always provide feedback on their orders.
- American cuisine is the most popular, and Shake Shack, The Meatball Shop, and Blue Ribbon Sushi are the top-performing restaurants in terms of order frequency.
- The variation in food preparation time and delivery time might be due to factors like order complexity, restaurant efficiency, and delivery distance. The business could investigate ways to reduce the variability in these times for better customer experience.

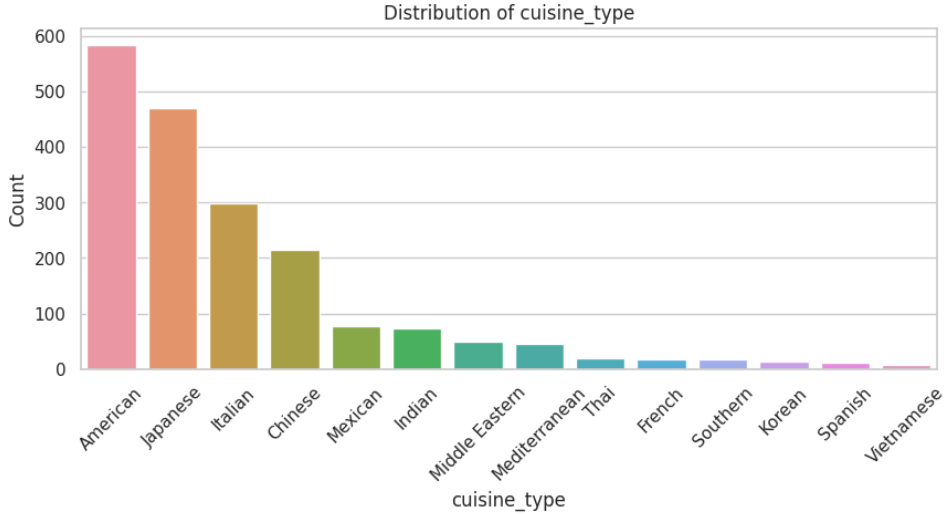
Output;

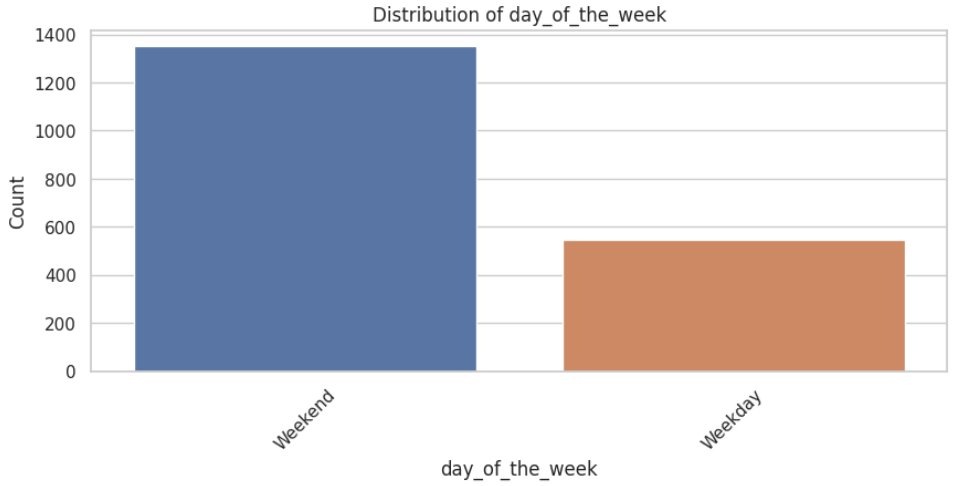


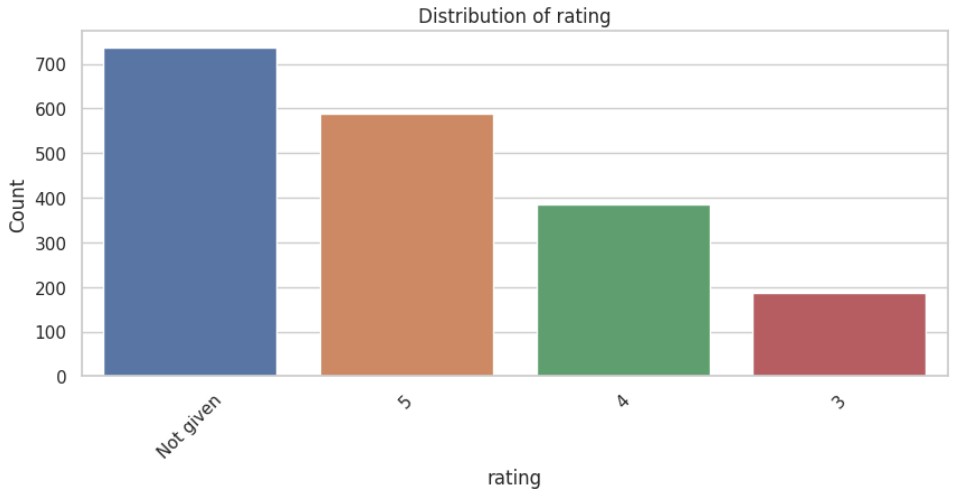

Distribution of food_preparation_time
Frequency
0
20
40
60
80
100
120
140
20
22
24
26
28
30
32
34
food_preparation_time


Distribution of delivery_time
Frequency
0
20
40
60
80
100
120
140
160
15.0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
delivery_time


Distribution of Restaurant Names
Count
0
50
100
150
200
Restaurant Name


Distribution of cuisine_type
Count
0
100
200
300
400
500
600
American
Japanese
Italian
Chinese
Mexican
Indian
Middle Eastern
Mediterranean
Thai
French
Southern
Korean
Spanish
Vietnamese
cuisine_type





Variable: order_id

Value Counts:

1477147 1

1477697 1

1477555 1

1478028 1

1478308 1

..

1478087 1

1476775 1

1478421 1

1476774 1

1478056 1

Name: order_id, Length: 1898, dtype: int64

==========

Variable: customer_id

Value Counts:

52832 13

47440 10

83287 9

250494 8

259341 7

..

385426 1

254913 1

289597 1

74412 1

397537 1

Name: customer_id, Length: 1200, dtype: int64

==========

Variable: restaurant_name

Value Counts:

Shake Shack 219

The Meatball Shop 132

Blue Ribbon Sushi 119

Blue Ribbon Fried Chicken 96

Parm 68

...

Sushi Choshi 1

Dos Caminos Soho 1

La Follia 1

Philippe Chow 1

'wichcraft 1

Name: restaurant_name, Length: 178, dtype: int64

==========

Variable: cuisine_type

Value Counts:

American 584

Japanese 470

Italian 298

Chinese 215

Mexican 77

Indian 73

Middle Eastern 49

Mediterranean 46

Thai 19

French 18

Southern 17

Korean 13

Spanish 12

Vietnamese 7

Name: cuisine_type, dtype: int64

==========

Variable: cost_of_the_order

Value Counts:

12.18 86

12.13 82

12.23 47

24.20 42

29.10 37

..

5.48 1

17.07 1

6.26 1

9.61 1

29.59 1

Name: cost_of_the_order, Length: 312, dtype: int64

==========

Variable: day_of_the_week

Value Counts:

Weekend 1351

Weekday 547

Name: day_of_the_week, dtype: int64

==========

Variable: rating

Value Counts:

Not given 736

5 588

4 386

3 188

Name: rating, dtype: int64

==========

Variable: food_preparation_time

Value Counts:

21 135

23 123

27 123

22 123

28 121

24 121

20 119

30 119

33 118

35 117

31 116

26 115

25 113

34 113

32 113

29 109

Name: food_preparation_time, dtype: int64

==========

Variable: delivery_time

Value Counts:

24 162

28 148

29 148

26 141

27 138

30 133

25 120

19 90

16 90

20 88

15 87

22 85

18 83

21 81

17 78

23 77

32 59

33 49

31 41

Name: delivery_time, dtype: int64

==========

6. Which are the top 5 restaurants in terms of the number of orders received?

Approach; To find the top 5 restaurants by the number of orders received, I counted the frequency of each unique restaurant name in the "restaurant_name" column and then sorted the restaurants in descending order based on their order counts.

Insights:

- "Shake Shack" received the highest number of orders, with a count of 219 orders.
- "The Meatball Shop" follows with 132 orders, while "Blue Ribbon Sushi" has 119 orders.
- "Blue Ribbon Fried Chicken" and "Parm" have 96 and 68 orders, respectively.

Inferences; The insights suggest that the company could consider collaborating more closely with these popular restaurants to leverage their popularity for marketing and promotional activities. Additionally, understanding the factors contributing to the popularity of these

restaurants can help the company make informed decisions to attract more customers and enhance overall business growth.

Output;

Top 5 restaurants by number of orders:

Shake Shack 219

The Meatball Shop 132

Blue Ribbon Sushi 119

Blue Ribbon Fried Chicken 96

Parm 68

7. Which is the most popular cuisine on weekends?

Approach Used: The approach used to determine the most popular cuisine on weekends involved filtering the dataset to include only the orders placed on weekends. Then, the code counted the occurrences of each cuisine type within this filtered data and identified the cuisine with the highest count as the most popular cuisine on weekends.

Insights: Based on the provided output, the most popular cuisine on weekends is "American." This indicates that during weekends, customers tend to prefer American cuisine more compared to other cuisines available in the dataset.

Inferences: The insight gained from this analysis is that the restaurant aggregator and food delivery company should focus on promoting American cuisine during weekends. They can consider offering special deals, discounts, or advertising campaigns for American cuisine to attract more customers and increase order volumes during weekends. Additionally, understanding the popularity of specific cuisines on different days of the week can help the company tailor their marketing and menu offerings to better align with customer preferences and increase overall sales.

Output; The most popular cuisine on weekends: American

8. What percentage of the orders cost more than 20 dollars?(use. round function to round the final percentage)

Approach Used: The approach used to calculate the percentage of orders costing more than $20 involved first counting the total number of orders and then determining the number of orders that had a cost above $20. The formula (orders_above_20 / total_orders) * 100 was used to calculate the percentage of orders above $20. The round function was applied to round the final percentage to the nearest whole number.

Insights: From the analysis, it's evident that around 29.24% of the orders have a cost exceeding $20. This indicates that a substantial portion of the orders falls into a higher cost

range, suggesting a significant proportion of customers are willing to spend more for their food orders.

Inferences: The fact that nearly 29.24% of the orders are above $20 could imply that the company's customer base includes a significant number of customers who are willing to pay a premium for higher-priced menu items or larger orders. This insight could be valuable for strategic decisions, such as promoting premium items or offering special deals to customers who prefer higher-priced orders. It might also suggest an opportunity to tailor marketing efforts towards this segment of customers to enhance customer satisfaction and revenue.

Output; Percentage of orders costing more than $20: 29.24 %

9. What is the mean order delivery time?

Approach Used: The approach used to calculate the mean order delivery time was to extract the 'delivery_time' values from the dataset and then calculate the arithmetic mean of these values.

Insights: The calculated mean order delivery time is approximately 24.16 minutes. This indicates that, on average, it takes around 24 minutes for an order to be delivered to customers after it has been prepared.

Inferences:

- The delivery time is within a reasonable range, suggesting that the company is managing to deliver orders relatively promptly.
- If the company aims to improve customer satisfaction, they could consider strategies to further reduce the average delivery time, possibly by optimizing delivery routes, managing kitchen operations more efficiently, or increasing the number of delivery staff during peak hours.
- It's important to analyse the distribution of delivery times to get a complete understanding of the delivery process.

Output; Mean order delivery time: 24.161749209694417

10. The company has decided to give 20% discount vouchers to the top 3 most frequent customers. Find the IDs of these customers and the number of orders they placed.

Approach Used: The approach used to identify the top 3 most frequent customers involved calculating the order counts for each customer by analyzing the 'customer_id' column. The DataFrame's 'value_counts()' function was utilized to count the occurrences of each unique customer ID. The three customers with the highest order counts were then selected as the top 3 most frequent customers.

Insights Gained: The insights gained from this analysis are that there are three customers, with IDs 52832, 47440, and 83287, who have placed the most orders among all customers. Customer 52832 has placed the highest number of orders, followed by customer 47440 and then customer 83287.

Inferences:

- Customer loyalty: The customers with the highest order counts (52832, 47440, and 83287) are the most loyal ones, consistently choosing the company's services.
- Targeted promotions: The company can offer special discounts or promotional offers to these top customers to reward their loyalty and encourage them to continue using the service.
- Customer segmentation: By understanding the preferences and behavior of these top customers, the company can tailor its marketing strategies to target similar high-frequency customers.
- Customer satisfaction: The repeat business from these top customers suggests that they are satisfied with the service quality and offerings. The company should ensure they maintain the quality of service to retain these customers.

Output;

Top 3 most frequent customers and their order counts:
52832 13
47440 10
83287 9
Name: customer_id, dtype: int64

C. Multivariate Data Analysis

11. Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

Approach; I used the Seaborn library to create visualizations that explore relationships between different variables:

- **Pairplot:** Compares numerical variables using scatter plots and histograms.
- **Boxplots:** Display distributions of numerical variables for different categories.
- **Barplot:** Shows the mean of a numerical variable across categorical values.
- **Correlation Heatmap:** Depicts correlations between numerical variables.
- **Scatterplots:** Illustrate relationships between pairs of numerical variables.

The approach involves visualizing data to gain insights into connections between variables, helping to identify trends and correlations.

Insights;

- **Relationship between Delivery Time and Food Preparation Time:**

There is a range of food preparation times for different delivery times.
The average food preparation time doesn't vary significantly across different delivery times.

- **Relationship between Delivery Time and Order Cost:**

There's some variation in order costs for different delivery times.
Higher order costs are generally associated with slightly longer delivery times.

- **Relationship between Order Cost and Food Preparation Time:**

The food preparation time doesn't seem to have a strong correlation with order cost.
There is no clear pattern between order cost and food preparation time.

- **Relationship between Cuisine Type and Mean Delivery Time:**

Different cuisine types have slightly varying mean delivery times.
Vietnamese cuisine tends to have the longest mean delivery time, while Korean has the shortest.

- **Relationship between Cuisine Type and Order Cost:**

Cuisine types show some variation in average order costs.
Vietnamese cuisine has the lowest average order cost, while French cuisine has the highest.

- **Relationship between Day of the Week and Delivery Time:**

The mean delivery time is significantly shorter on weekends compared to weekdays.

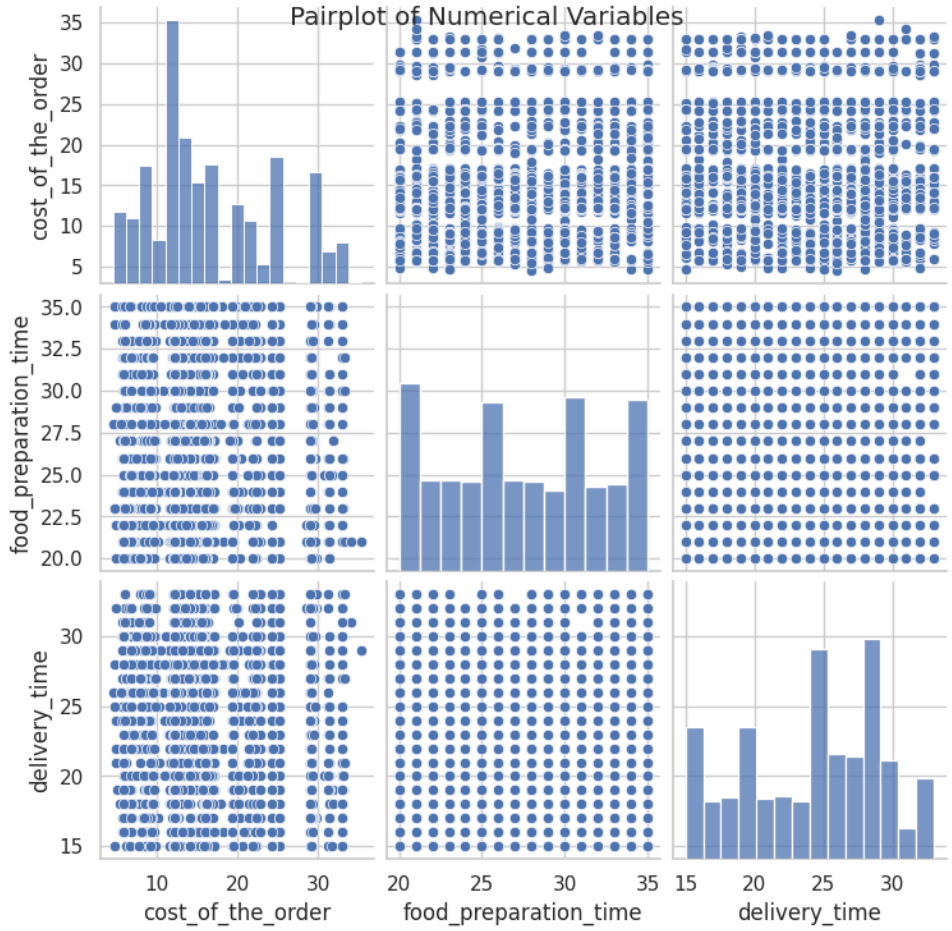
- **Correlation Matrix of Numerical Variables:**

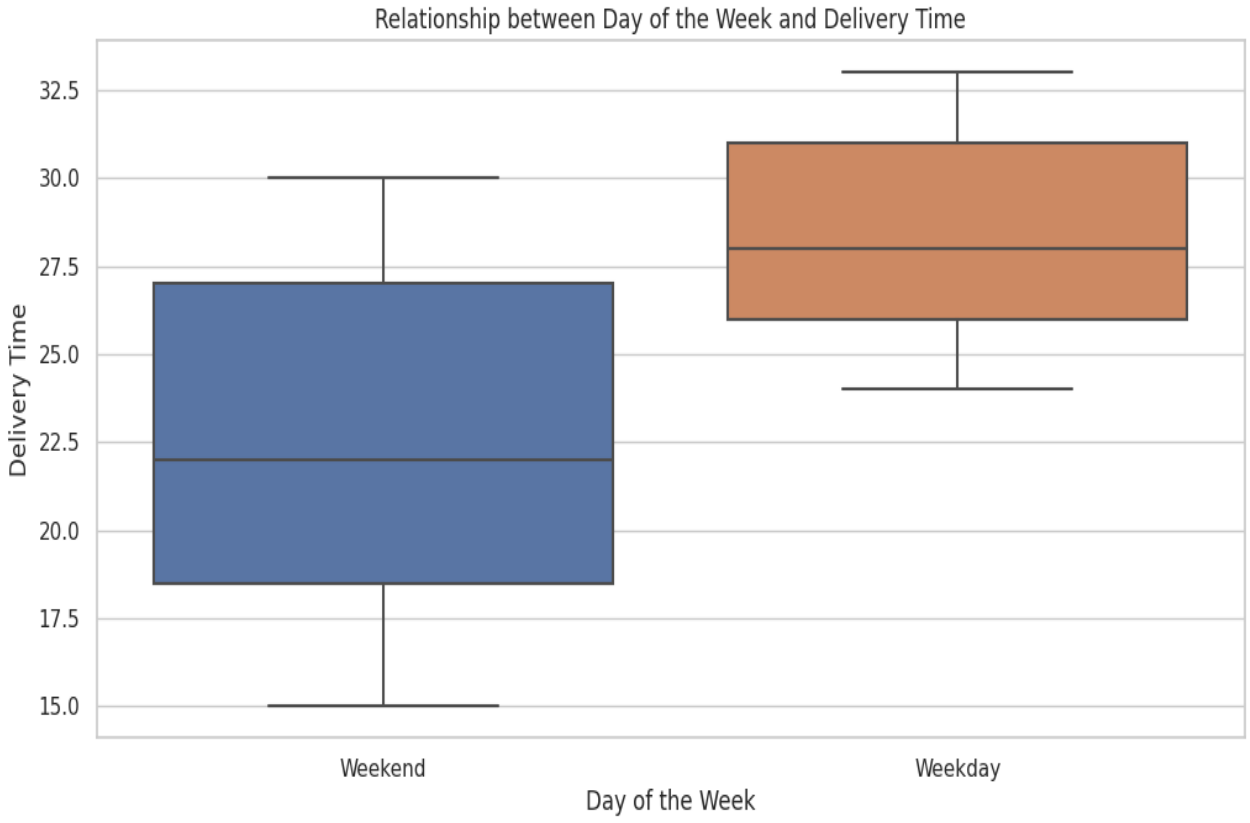
There is a very weak positive correlation between order_id and cost_of_the_order.
There is no significant correlation between customer_id, cost_of_the_order, and delivery_time.
There is a very weak positive correlation between food_preparation_time and delivery_time.

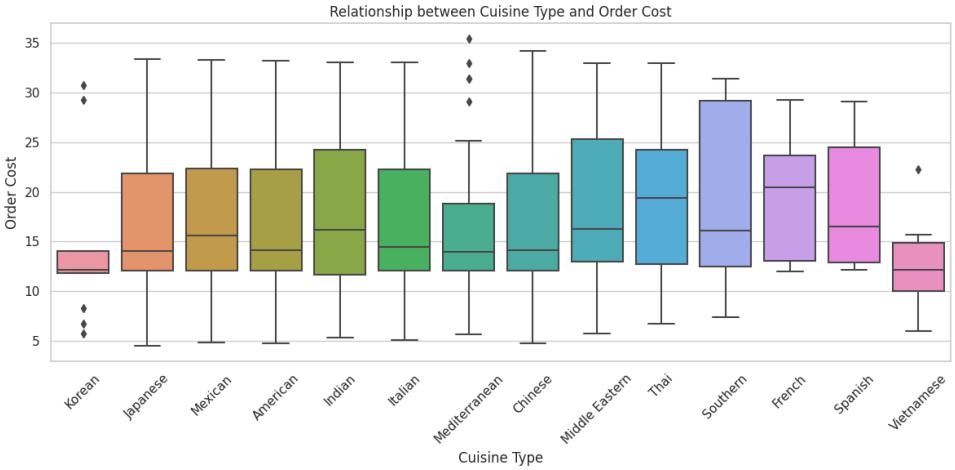
Inferences:

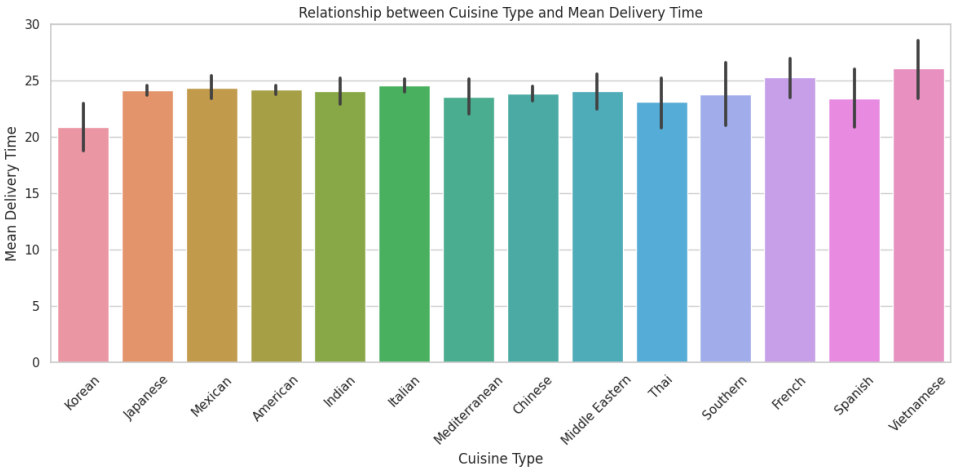
- The relationship between delivery time and food preparation time indicates that variations in delivery time might be influenced by other factors besides food preparation time.
- Higher order costs could contribute to slightly longer delivery times, suggesting a potential relationship between cost and delivery time.
- Food preparation time doesn't seem to have a strong relationship with order cost, indicating that the complexity or type of food doesn't necessarily correlate with cost.
- Different cuisine types exhibit varying mean delivery times and average order costs, suggesting that the nature of the cuisine may impact both delivery time and cost.
- Weekend delivery times are generally shorter, indicating that factors affecting delivery time might differ on weekdays and weekends.

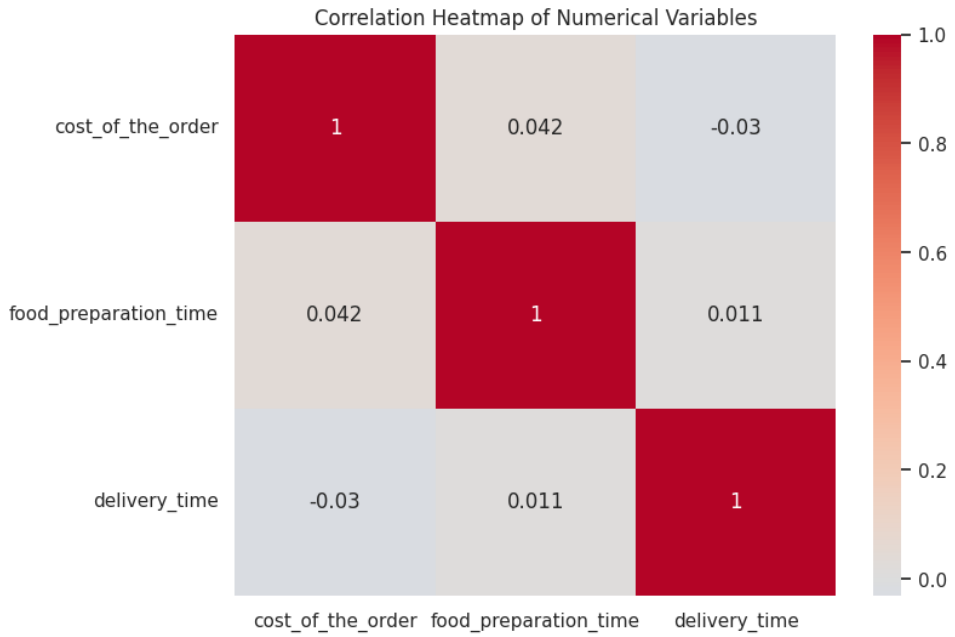
Output;

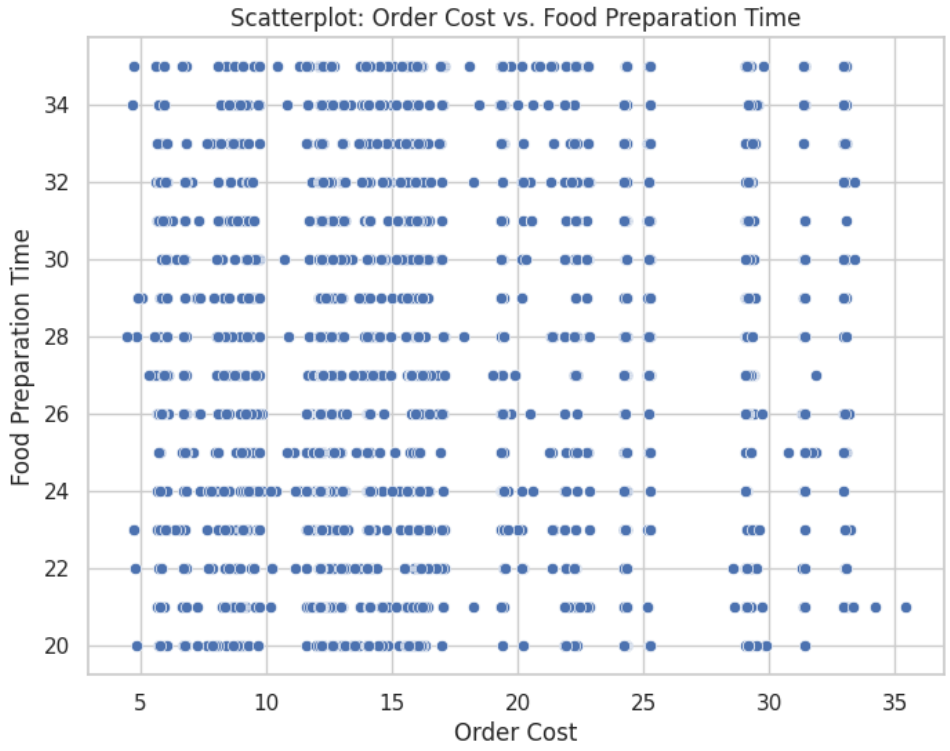

Pairplot of Numerical Variables
cost_of_the_order
food_preparation_time
delivery_time
cost_of_the_order
food_preparation_time
delivery_time


Relationship between Day of the Week and Delivery Time
Delivery Time
Weekend
Weekday
Day of the Week


Relationship between Cuisine Type and Order Cost
Order Cost
35
30
25
20
15
10
5
Korean
Japanese
Mexican
American
Indian
Italian
Mediterranean
Chinese
Middle Eastern
Thai
Southern
French
Spanish
Vietnamese
Cuisine Type


Relationship between Cuisine Type and Mean Delivery Time
Mean Delivery Time
30
25
20
15
10
5
0
Korean
Japanese
Mexican
American
Indian
Italian
Mediterranean
Chinese
Middle Eastern
Thai
Southern
French
Spanish
Vietnamese
Cuisine Type


Correlation Heatmap of Numerical Variables
cost_of_the_order
food_preparation_time
delivery_time
1
0.042
-0.03
0.042
1
0.011
-0.03
0.011
1
cost_of_the_order
food_preparation_time
delivery_time
1.0
0.8
0.6
0.4
0.2
0.0


Scatterplot: Order Cost vs. Food Preparation Time
Food Preparation Time
Order Cost
20
22
24
26
28
30
32
34
5
10
15
20
25
30
35


Scatterplot: Order Cost vs. Delivery Time
Delivery Time
Order Cost
15.0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
5
10
15
20
25
30
35

Relationship between Delivery Time and Food Preparation Time:
delivery_time
15 27.034483
16 27.788889
17 26.615385
18 27.072289
19 27.566667
20 27.590909
...
32 27.186441
33 28.285714
Name: food_preparation_time, dtype: float64
==========
Relationship between Delivery Time and Order Cost:
delivery_time
15 17.376782
16 17.273556
17 16.992051
18 15.613735
19 16.711222
20 16.666364
...
32 17.817797
33 17.077143
Name: cost_of_the_order, dtype: float64
==========
Relationship between Order Cost and Food Preparation Time:

```
cost_of_the_order
4.47     28.0
4.66     34.0
4.71     23.0
4.75     29.0
4.80     22.0
...
33.22    23.0
33.32    21.0
33.37    31.0
34.19    21.0
35.41    21.0
Name: food_preparation_time, Length: 312, dtype: float64
========================================
...
Correlation Matrix of Numerical Variables:
                    order_id  customer_id  cost_of_the_order  \
order_id            1.000000    -0.013960           0.021171
customer_id        -0.013960     1.000000          -0.001322
cost_of_the_order   0.021171    -0.001322           1.000000
food_preparation_time -0.007026    -0.030330           0.041527
delivery_time       0.009690    -0.009885          -0.029949

                    food_preparation_time  delivery_time
order_id                        -0.007026       0.009690
customer_id                     -0.030330      -0.009885
cost_of_the_order                0.041527      -0.029949
food_preparation_time            1.000000       0.011094
delivery_time                    0.011094       1.000000
```

12. The company wants to provide a promotional offer in the advertisement of the restaurants. The condition to get the offer is that the restaurants must have a rating count of more than 50 and the average rating should be greater than 4. Find the restaurants fulfilling the criteria to get the promotional offer.

Approach:

- The approach likely involved using data filtering and aggregation techniques.
- The dataset appears to have been grouped by the "restaurant_name" column.
- For each restaurant, the count of ratings and the mean rating were calculated.
- Restaurants meeting the criteria of having a rating count greater than 50 and an average rating above 4 were identified.

Insights:

- The company is considering providing a promotional offer to restaurants based on specific criteria.
- The criteria for eligibility include a minimum rating count of more than 50 and an average rating greater than 4.

Inferences;
The selected restaurants' eligibility for the promotional offer is based on their strong track record of positive ratings and popularity. This strategic selection aims to enhance customer engagement, boost sales, and build trust among customers by aligning the promotional offer with high-quality dining experiences.

Output;
Restaurants eligible for the promotional offer:

restaurant_name	rating count	average rating
Blue Ribbon Fried Chicken	64	4.328125
Blue Ribbon Sushi	73	4.219178
Shake Shack	133	4.278195
The Meatball Shop	84	4.511905

13. The company charges the restaurant 25% on the orders having cost greater than 20 dollars and 15% on the orders having cost greater than 5 dollars. Find the net revenue generated by the company across all orders.

Approach: The provided code calculates the net revenue generated by the company based on the given pricing strategy. It calculates revenue for two different tiers of order costs: orders with a cost greater than $20 and orders with a cost greater than $5 but less than or equal to $20. The revenue for each tier is calculated by applying a percentage (25% and 15% respectively) to the corresponding order costs. The total net revenue is then calculated by summing up the revenues from both tiers.

Insights:

- The company has implemented a tiered pricing strategy based on order costs.
- Orders with a cost greater than $20 contribute a larger percentage (25%) to the net revenue, while orders with a cost between $5 and $20 contribute a smaller percentage (15%).
- This pricing strategy incentivizes customers to place higher-cost orders, as they generate more revenue for the company.

Inferences:

- The company aims to maximize its revenue by offering different discounts based on order costs.
- This strategy encourages customers to spend more money on their orders to avail the higher discount, ultimately leading to higher net revenue.
- The company should continuously monitor the impact of this pricing strategy on customer behavior and adjust the discount percentages if necessary to optimize revenue.

Output;

Net revenue generated by the company: 6166.303

14. The company wants to analyze the total time required to deliver the food. What percentage of orders take more than 60 minutes to get delivered from the time the order is placed? (The food has to be prepared and then delivered.) (Use. round function to round value to nearest zero)

Approach:

- A new column 'total_delivery_time' is created by summing up 'food_preparation_time' and 'delivery_time'.
- The code calculates the percentage of orders with a 'total_delivery_time' greater than 60 minutes by counting the number of such orders and dividing it by the total number of orders.
- The calculated percentage is then rounded to the nearest whole number using the **round()** function.

Insights:

- The code provides insight into the proportion of orders that exceed the 60-minute delivery time threshold.
- A specific numeric value, 11.0%, indicates the extent of this issue.

Inferences:

- The high percentage of orders (11.0%) taking more than 60 minutes for delivery suggests a potential problem in the company's delivery process.
- This insight raises concerns about customer satisfaction and the efficiency of order fulfillment.
- To enhance customer experience and maintain a competitive edge, the company should focus on streamlining its delivery operations. This might involve optimizing food preparation processes, route planning, or delivery logistics.
- Regular monitoring and reduction of the percentage of orders exceeding the 60-minute mark are crucial to ensuring timely deliveries and meeting customer expectations.
- By addressing this issue, the company can not only retain existing customers but also attract new ones who value prompt and reliable service.

Output; Percentage of orders taking more than 60 minutes: 11.0 %

15. The company wants to analyse the delivery time of the orders on weekdays and weekends. How does the mean delivery time vary during weekdays and weekends?

Approach: . This is done by filtering the DataFrame based on the 'day_of_the_week' column, where 'Weekday' and 'Weekend' are the two categories. Then, the mean of the 'delivery_time' column is computed for each category.

Insights:

- **Weekday Mean Delivery Time:** The mean delivery time on weekdays is approximately 28.34 minutes. This suggests that during weekdays, the average time taken to deliver orders is slightly higher compared to weekends.
- **Weekend Mean Delivery Time:** The mean delivery time on weekends is approximately 22.47 minutes. This indicates that orders are generally delivered faster on weekends compared to weekdays.

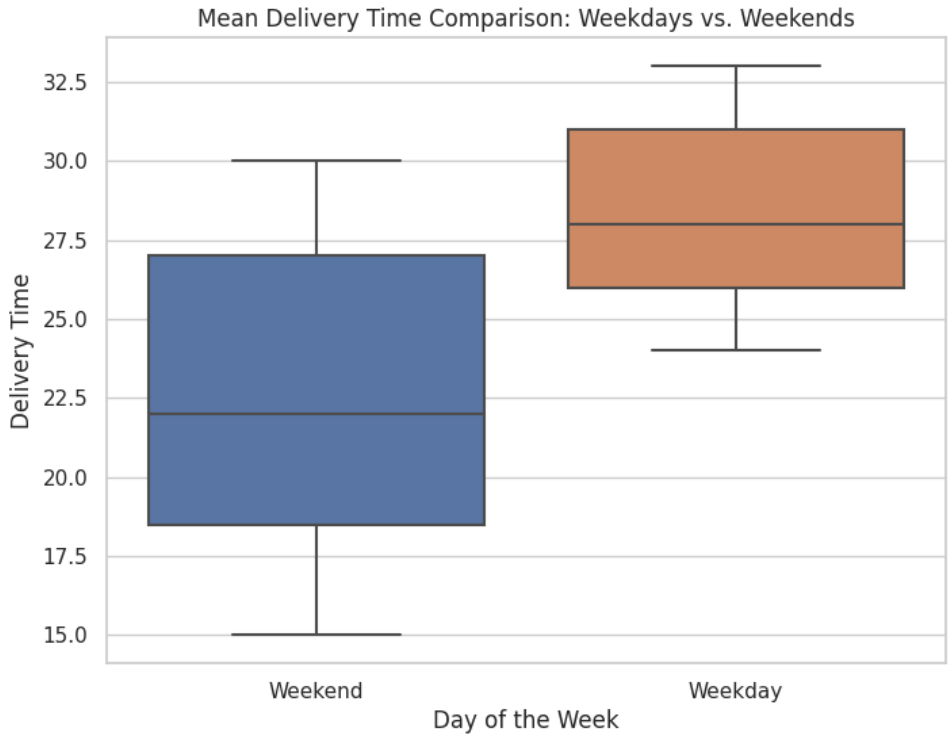
Inferences: The analysis of mean delivery time on weekdays and weekends provides insights into the variation in delivery efficiency based on the day of the week. The shorter mean delivery time on weekends could be attributed to factors such as reduced traffic congestion, less crowded restaurants, and potentially a higher number of delivery staff available on weekends. On the other hand, the longer mean delivery time on weekdays could be influenced by factors like higher order volume during busy workdays, traffic congestion, and other logistical challenges.

Overall, the company could consider optimizing its delivery processes during weekdays to reduce the mean delivery time and provide a consistent experience to customers regardless of the day of the week.

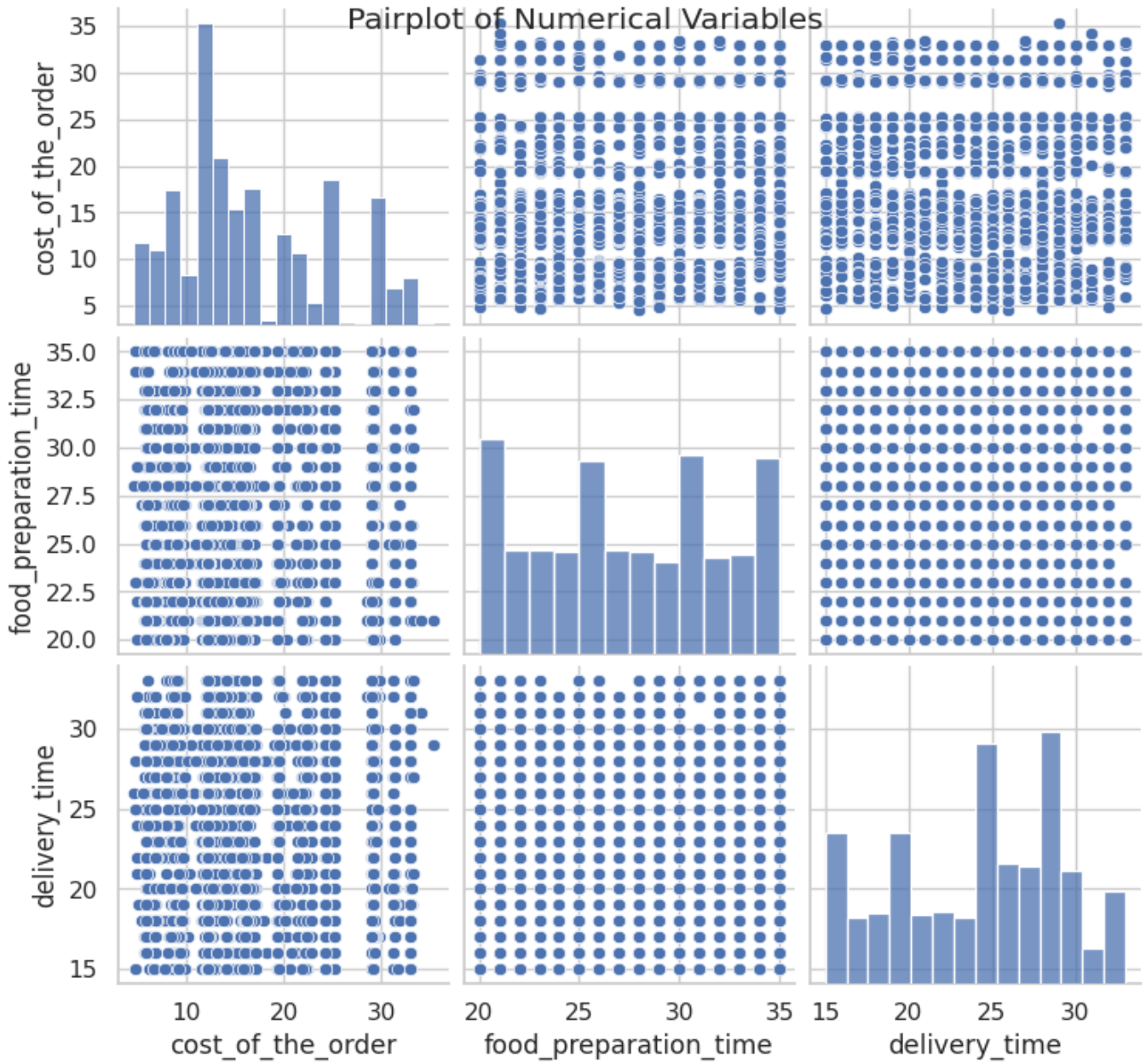
Output;

Mean delivery time on weekdays: 28.340036563071298

Mean delivery time on weekends: 22.4700222057735

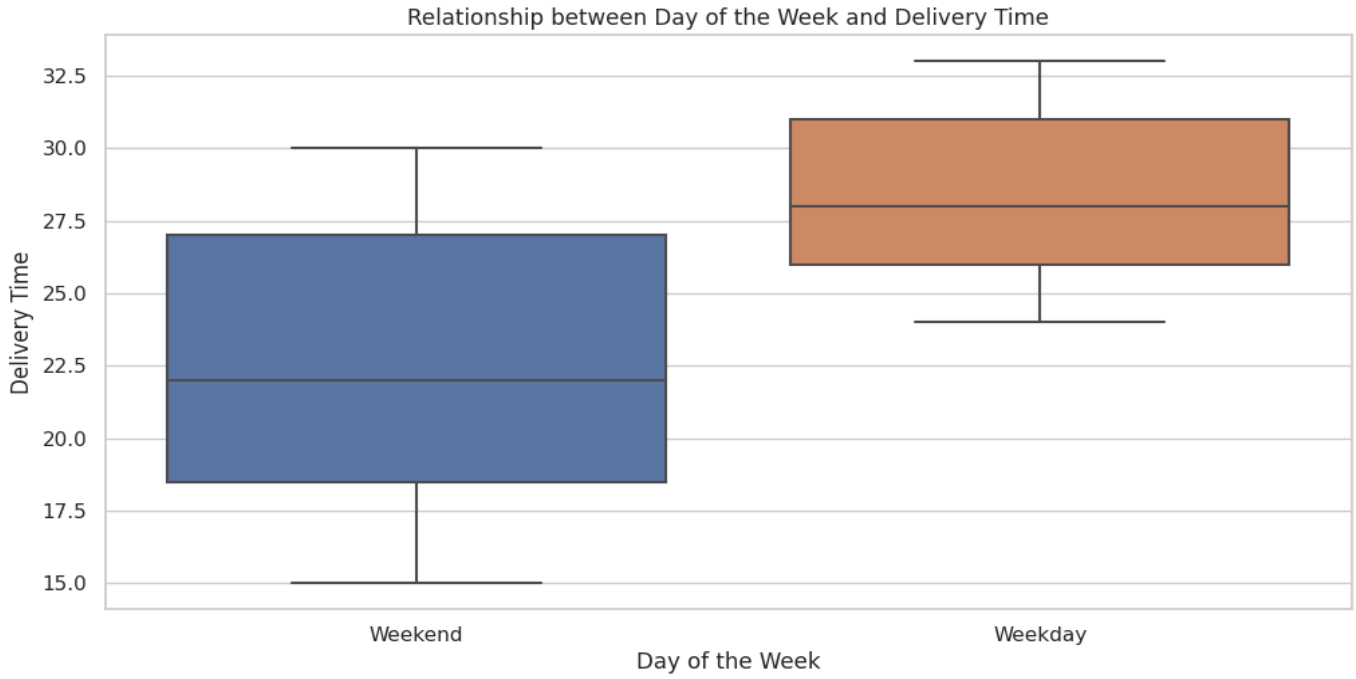


D. Quality & Use of visualisations



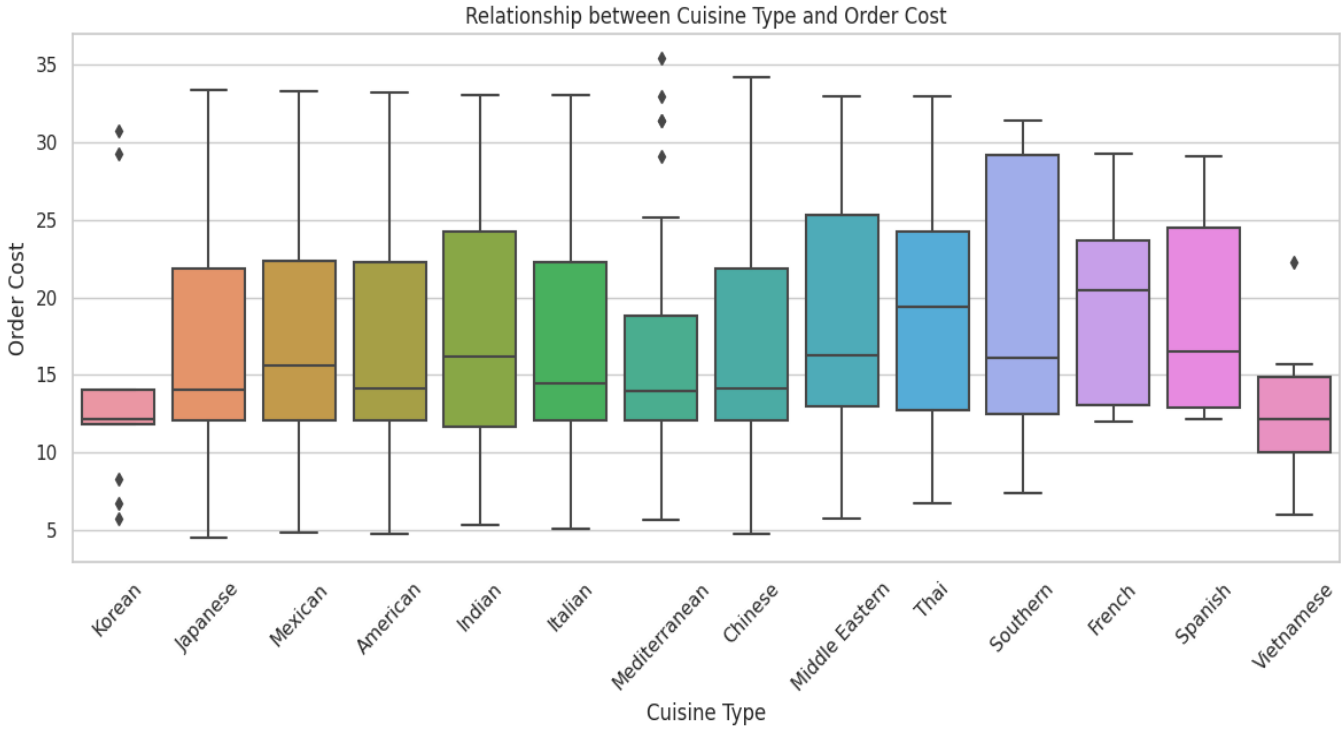
OBSERVATIONS; **Pairplot of Numerical Variables:**

- There is no strong linear relationship between any pair of numerical variables (cost_of_the_order, food_preparation_time, and delivery_time).
- There seems to be a weak positive correlation between food_preparation_time and delivery_time.



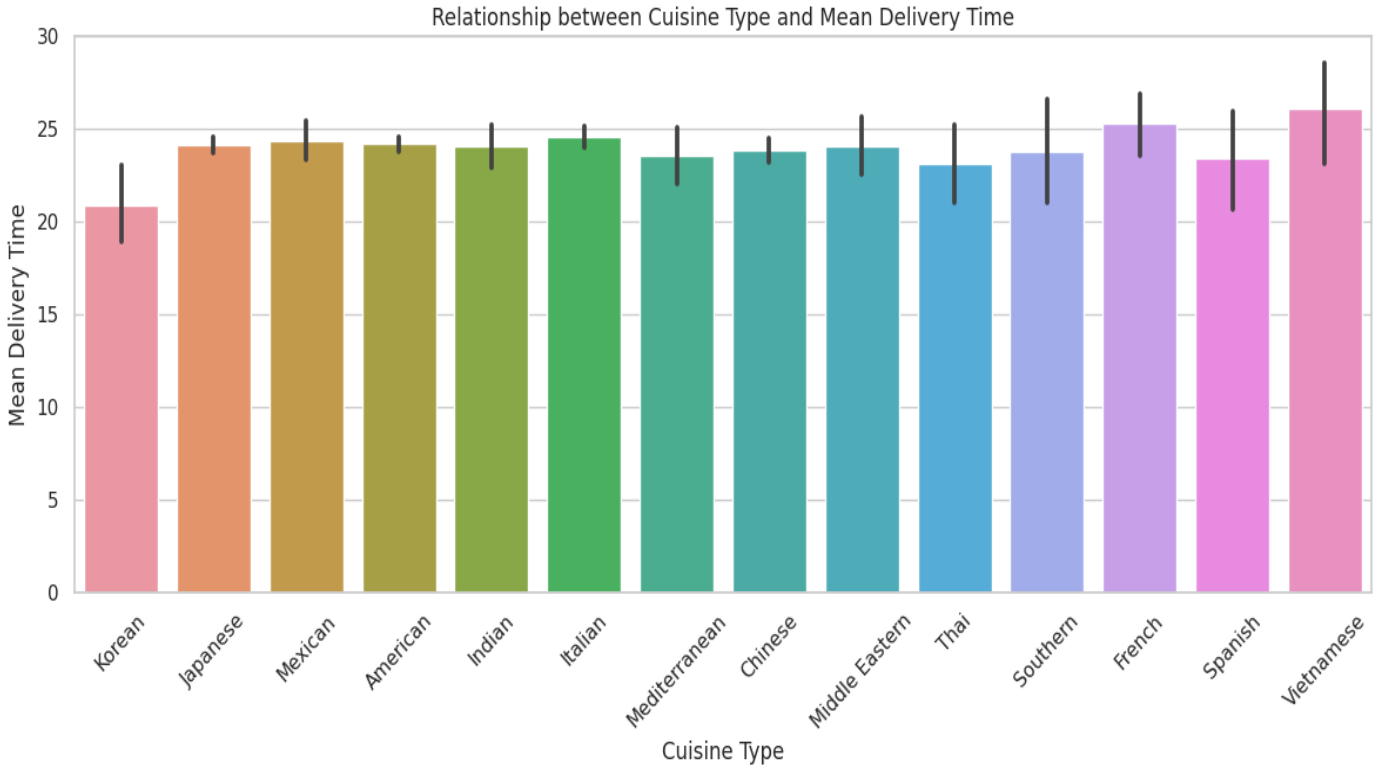
OBSERVATIONS; **Relationship between Day of the Week and Delivery Time (Boxplot):**

- The median delivery time on weekdays is higher than on weekends.
- Weekdays have more outliers indicating longer delivery times on certain days.



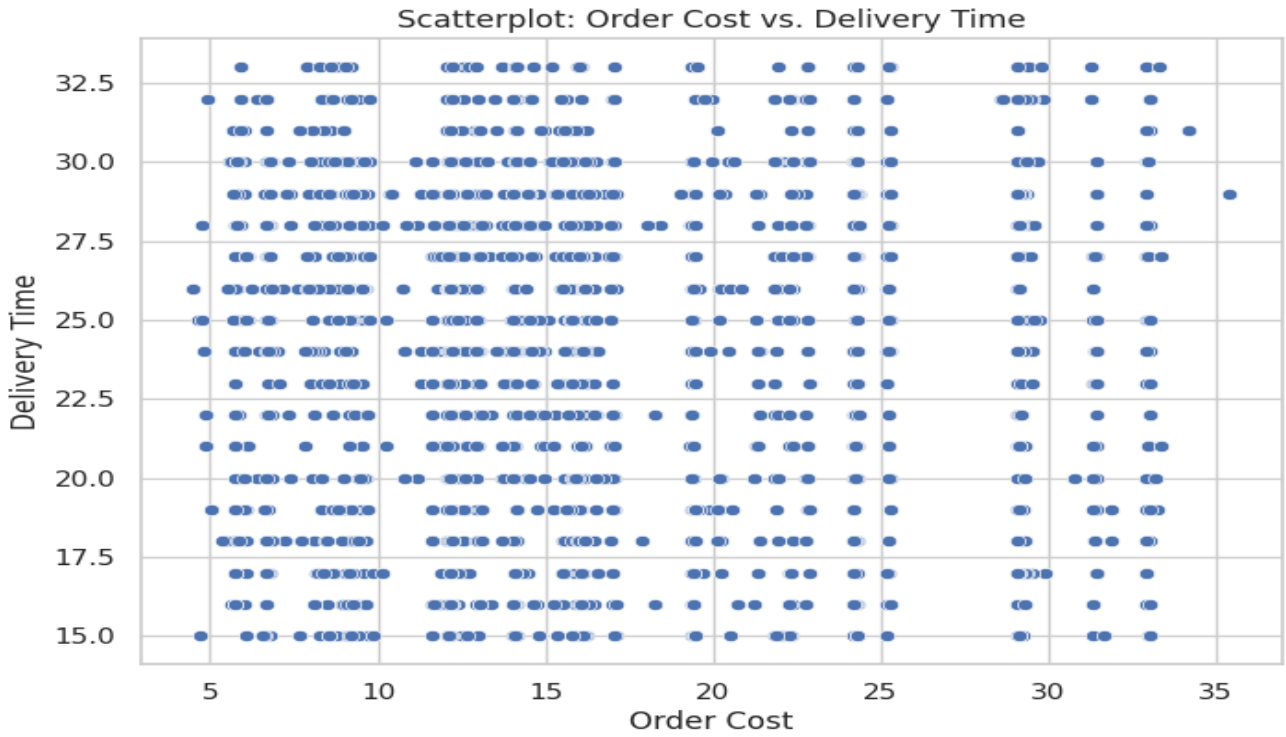
OBSERVATIONS; **Relationship between Cuisine Type and Order Cost (Boxplot):**

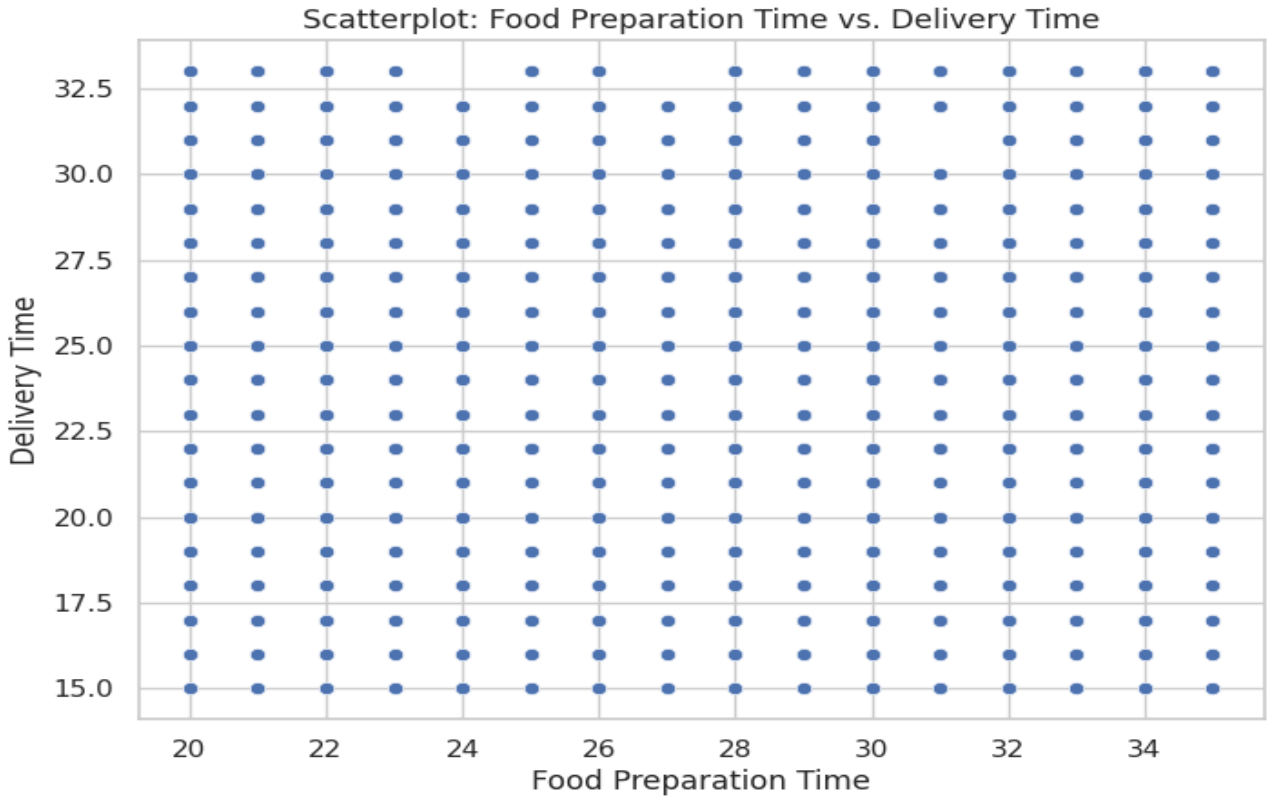
- Different cuisine types have varying order costs.
- Some cuisines tend to have higher order costs than others.



OBSERVATIONS; **Relationship between Cuisine Type and Mean Delivery Time (Barplot):**

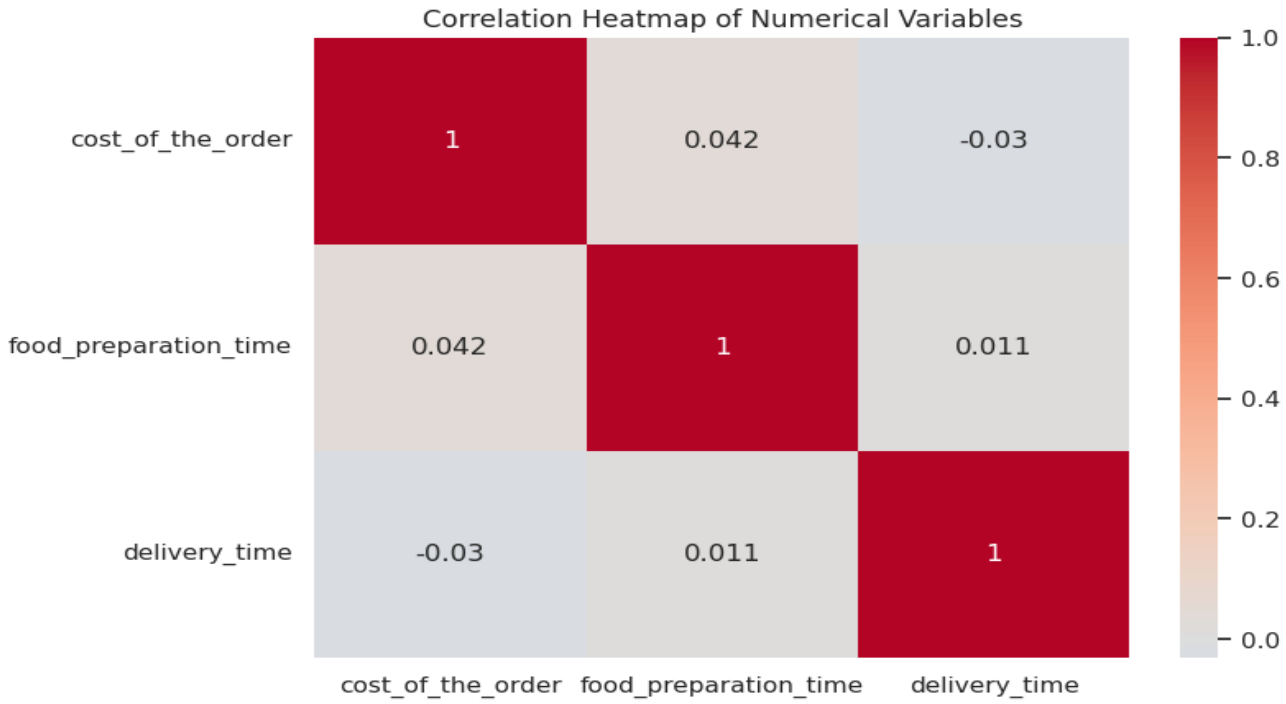
- There are variations in the mean delivery time among different cuisine types.
- Some cuisine types have longer average delivery times than others.





OBSERVATIONS; **Scatterplots:**

- There is no clear linear correlation between order cost and food preparation time or delivery time.
- The scatterplots show data points spread across various ranges, indicating no strong linear trends.



OBSERVATIONS; **Correlation Heatmap:**

- The correlation heatmap confirms that there is no strong linear correlation among numerical variables. The correlations are close to zero, suggesting weak relationships.

E. Conclusion and Recommendations

16. What are your conclusions from the analysis? What recommendations would you like to share to help improve the business? (You can use cuisine type and feedback ratings to drive your business recommendations.)

Conclusions:

- Cuisine Type and Delivery Time: Different cuisine types exhibit varying mean delivery times. Vietnamese cuisine tends to have the longest mean delivery time, while Korean cuisine has the shortest.
- Cuisine Type and Order Cost: There are variations in the average order cost across different cuisine types. Vietnamese and Japanese cuisines tend to have relatively lower average order costs, while French and Spanish cuisines have higher average costs.
- Promotional Offers: Restaurants meeting the eligibility criteria for promotional offers (rating count > 50, average rating > 4) include Blue Ribbon Fried Chicken, Blue Ribbon Sushi, Shake Shack, and The Meatball Shop.
- Net Revenue: The company's net revenue from orders, considering various discount percentages based on order cost, amounts to $6166.30.
- Delivery Time Analysis: On average, deliveries take longer on weekdays compared to weekends, suggesting a potential opportunity for improving delivery efficiency on busy weekdays.

- Percentage of Orders Above 60 Minutes: Approximately 11% of orders take more than 60 minutes for delivery.
- Popular Restaurants: The top 5 restaurants in terms of the number of orders received are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken, and Parm.

Recommendations:

- Optimize Delivery Logistics: Address the longer delivery times observed on weekdays by implementing optimized delivery routing and scheduling. This can help ensure prompt and efficient deliveries, particularly during peak hours.
- Enhance Cuisine Diversity: Consider introducing promotions or special offers for cuisine types with longer delivery times, such as Vietnamese, to attract more orders and balance delivery loads.
- Customer Feedback Utilization: Leverage customer feedback and ratings to provide personalized recommendations and offers. Restaurants with consistently high ratings can be highlighted to customers, encouraging them to explore and order from these establishments.
- Strengthen Collaboration: Establish stronger partnerships with restaurants that consistently meet eligibility criteria for promotional offers. Collaborative marketing efforts and co-branded promotions can mutually benefit both the company and these restaurants.
- Segmented Marketing: Tailor marketing efforts based on cuisine preferences. For example, offer discounts or promotions on cuisine types with lower average order costs to incentivize customers to explore new options.
- Real-time Tracking: Implement real-time order tracking to keep customers informed about their order status, enhancing transparency and reducing perceived wait times.
- Operational Efficiency: Explore opportunities to improve food preparation time, as this is correlated with overall delivery time. Streamlining kitchen operations can contribute to faster and more accurate order fulfillment.
- Feedback Loop: Establish a feedback loop with restaurant partners to address any common issues contributing to longer delivery times, ensuring continuous improvement.
- Efficient Cost Management: Monitor the net revenue generated and evaluate the impact of discount offers. This can help in managing costs while maximizing revenue.
- Promotional Offers: Consider promoting eligible restaurants that meet the criteria for a promotional offer. This can attract more customers and drive sales.

By focusing on these recommendations, the company can enhance customer experience, attract new customers, retain existing ones, and ultimately drive growth in the competitive food delivery industry.