Contents

1. Description: FoodHub Data Analysis

1.1 Context

The number of restaurants in New York is increasing day by day. Lots of students and busy professionals rely on those restaurants due to their hectic lifestyles. Online food delivery service is a great option for them. It provides them with good food from their favorite restaurants. A food aggregator company FoodHub offers access to multiple restaurants through a single smartphone app.

The app allows restaurants to receive a direct online order from a customer. The app assigns a delivery person from the company to pick up the order after it is confirmed by the restaurant. The delivery person then uses the map to reach the restaurant and waits for the food package. Once the food package is handed over to the delivery person, he/she confirms the pick-up in the app and travels to the customer's location to deliver the food. The delivery person confirms the drop-off in the app after delivering the food package to the customer. The customer can rate the order in the app. The food aggregator earns money by collecting a fixed margin of the delivery order from the restaurants.

1.2 Objective

The food aggregator company has stored the data of the different orders made by the registered customers in their online portal. They want to analyze the data to get a fair idea about the demand of different restaurants which will help them in enhancing their customer experience. Suppose you are hired as a Data Scientist in this company and the Data Science team has shared some of the key questions that need to be answered. Perform the data analysis to find answers to these questions that will help the company to improve the business.

1.3 Data Description

The data contains the different data related to a food order. The detailed data dictionary is given below.

Data Dictionary

- order_id: Unique ID of the order
- customer_id: ID of the customer who ordered the food
- restaurant_name: Name of the restaurant
- cuisine_type: Cuisine ordered by the customer
- cost: Cost of the order
- day_of_the_week: Indicates whether the order is placed on a weekday or weekend (The weekday is from Monday to Friday and the weekend is Saturday and Sunday)
- rating: Rating given by the customer out of 5
- food_preparation_time: Time (in minutes) taken by the restaurant to prepare the food. This is calculated by taking the difference between the timestamps of the restaurant's order confirmation and the delivery person's pick-up confirmation.
- delivery_time: Time (in minutes) taken by the delivery person to deliver the food package. This is calculated by taking the difference between the timestamps of the delivery person's pick-up confirmation and drop-off information

2. Data Overview

2.1 : Question No -1: How many rows and columns are present in the data

- The DataFrame has 1898 rows and 9 columns.

2.2 : Question No -2 : What are the datatypes of the different columns in the dataset

- PFA,There are three types of datatypes named int,float and object

```
<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
```

2.3: Question No -3: Statistical summary - PFA,SS attached

	count	mean	std	min	25%	50%	75%	max
order_id	1898.0	1.477496e+06	548.049724	1476547.00	1477021.25	1477495.50	1.477970e+06	1478444.00
customer_id	1898.0	1.711685e+05	113698.139743	1311.00	77787.75	128600.00	2.705250e+05	405334.00
cost_of_the_order	1898.0	1.649885e+01	7.483812	4.47	12.08	14.14	2.229750e+01	35.41
food_preparation_time	1898.0	2.737197e+01	4.632481	20.00	23.00	27.00	3.100000e+01	35.00
delivery_time	1898.0	2.416175e+01	4.972637	15.00	20.00	25.00	2.800000e+01	33.00

2.4 : Question No -4 :Are there any missing values in the data?

- **There are no missing values in the data**

2.5 : Question No -5 : How many orders are not rated?

rating	count
Not given	736
5	588
4	386
3	188

dtype: int64

3. Univariate Analysis:

Question 6: Explore all the variables and provide observations on their distributions. (Generally, histograms, boxplots, countplots, etc. are used for univariate exploration

Observations -

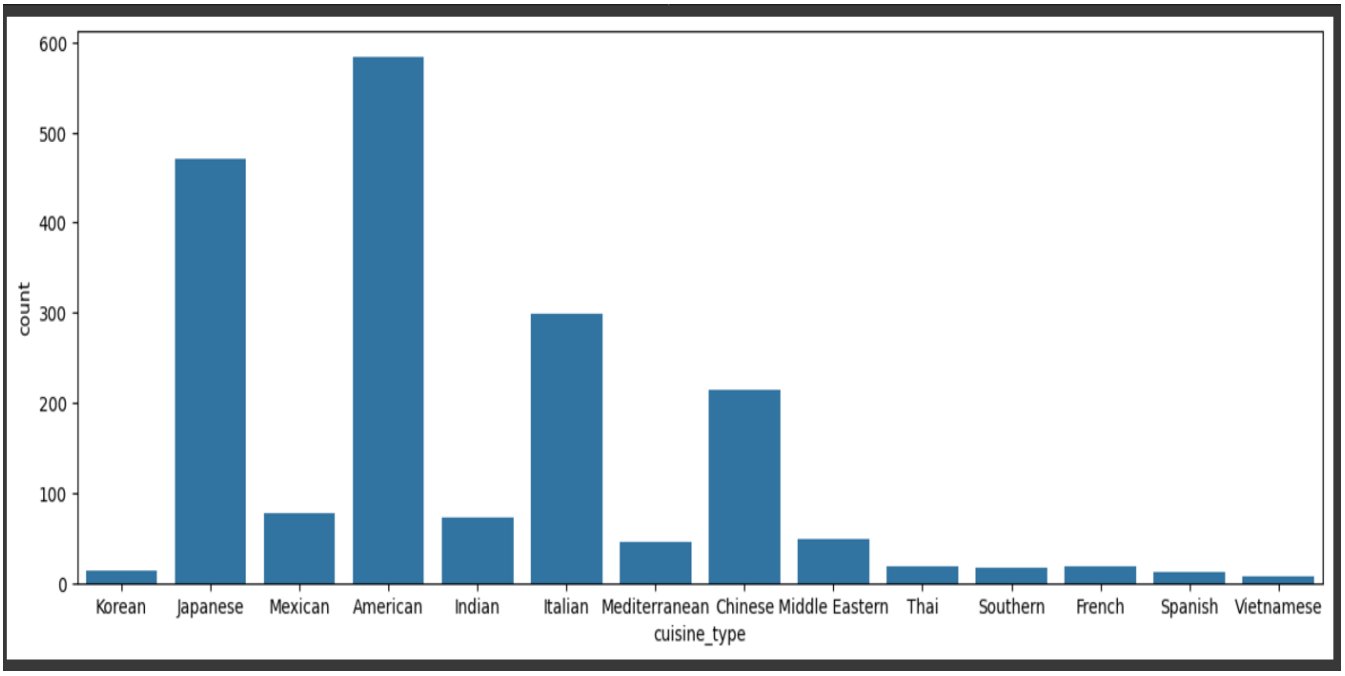
- **There are 1898 unique orders in the dataset.**
- **There are 1200 unique customers in the dataset.**
- **There are 178 unique restaurants in the dataset.**
- **Shake Shack received maximum number of orders**

3.1 CountPlot of Column

A) Cuisine Type

Observations -

- **There are 14 unique cuisines in the dataset**

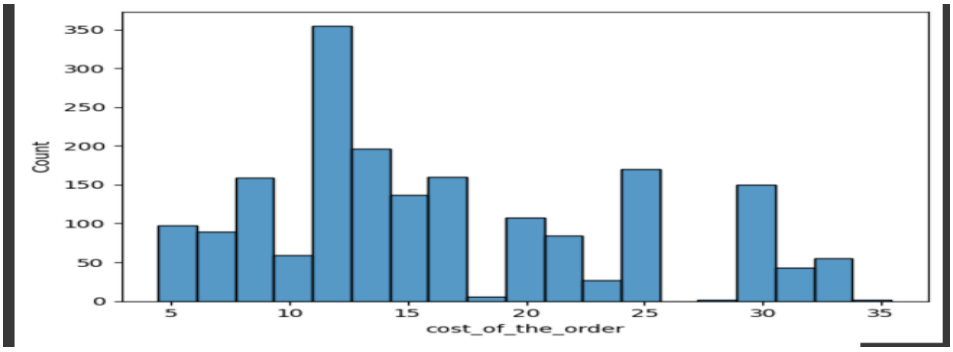


Observation -

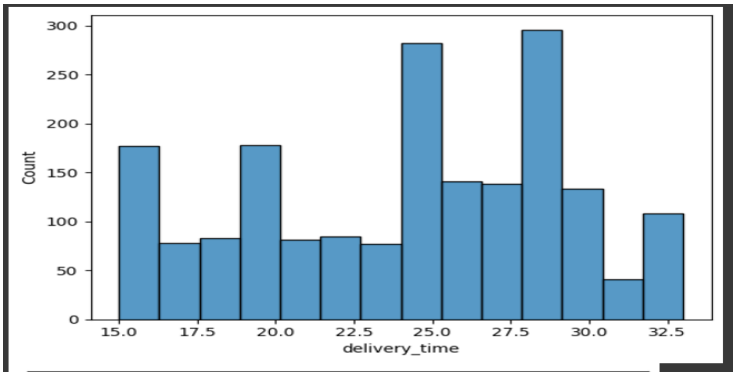
- **There are 14 unique cuisines in the dataset.**
- **The distribution of cuisine types show that cuisine types are not equally distributed.**
- **The most frequent cuisine type is American followed by Japanese and Italian.**

- **Vietnamese appears to be the least popular of all the cuisines.**

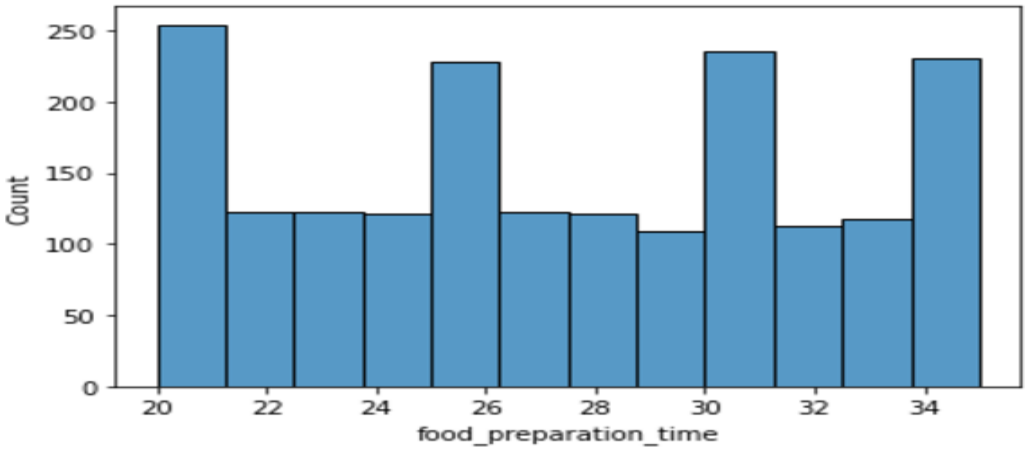
B) Cost of the order



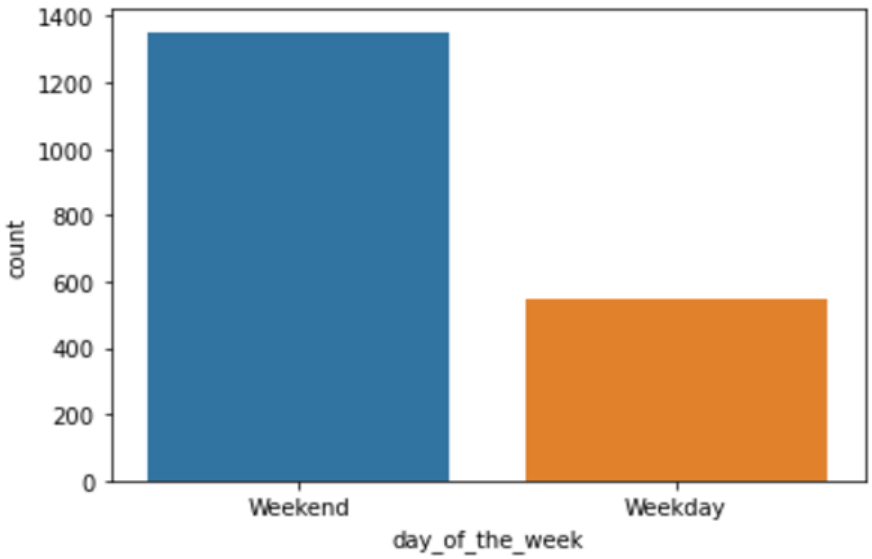
C) Delivery Time



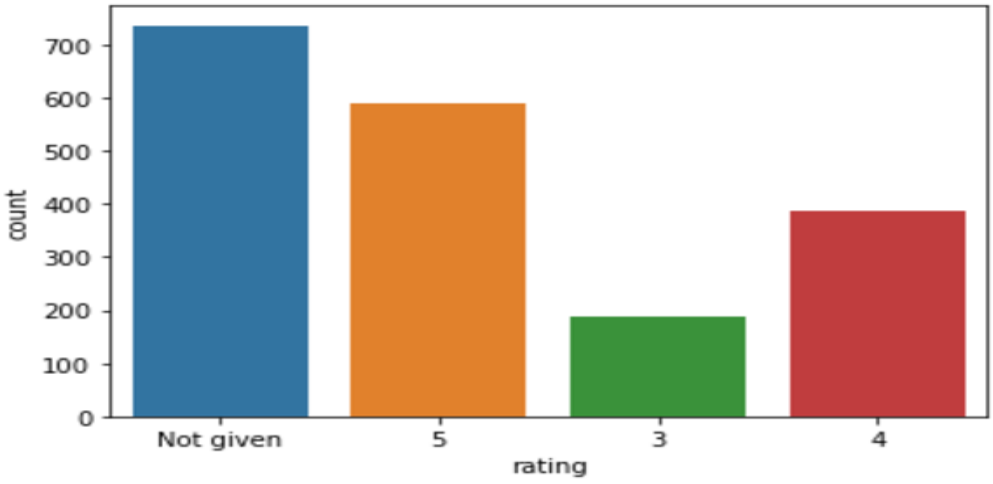
D) Food Preparation Time



E) Days of the week



F) Rating



Question 7: Which are the top 5 restaurants in terms of the number of orders received?

restaurant_name	count
Shake Shack	219
The Meatball Shop	132
Blue Ribbon Sushi	119
Blue Ribbon Fried Chicken	96
Parm	68

dtype: int64

Question 8: Which is the most popular cuisine on weekends?

cuisine_type	count
American	415
Japanese	335
Italian	207
Chinese	163
Mexican	53
Indian	49
Mediterranean	32
Middle Eastern	32
Thai	15
French	13
Korean	11
Southern	11
Spanish	11
Vietnamese	4

dtype: int64

Question 9: What percentage of the orders cost more than 20 dollars?

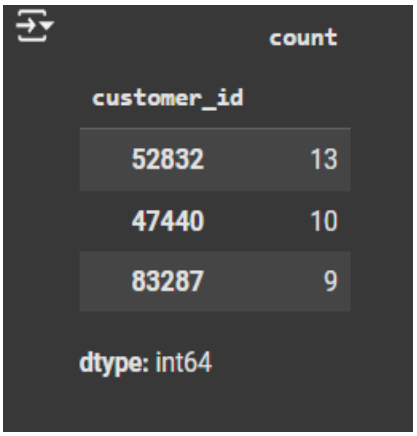
- The number of total orders that cost above 20 dollars is: 555

Percentage of orders above 20 dollars: 29.24 %

Question 10: What is the mean order delivery time?

- The mean delivery time for this dataset is 24.16 minutes

Question 11: 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.



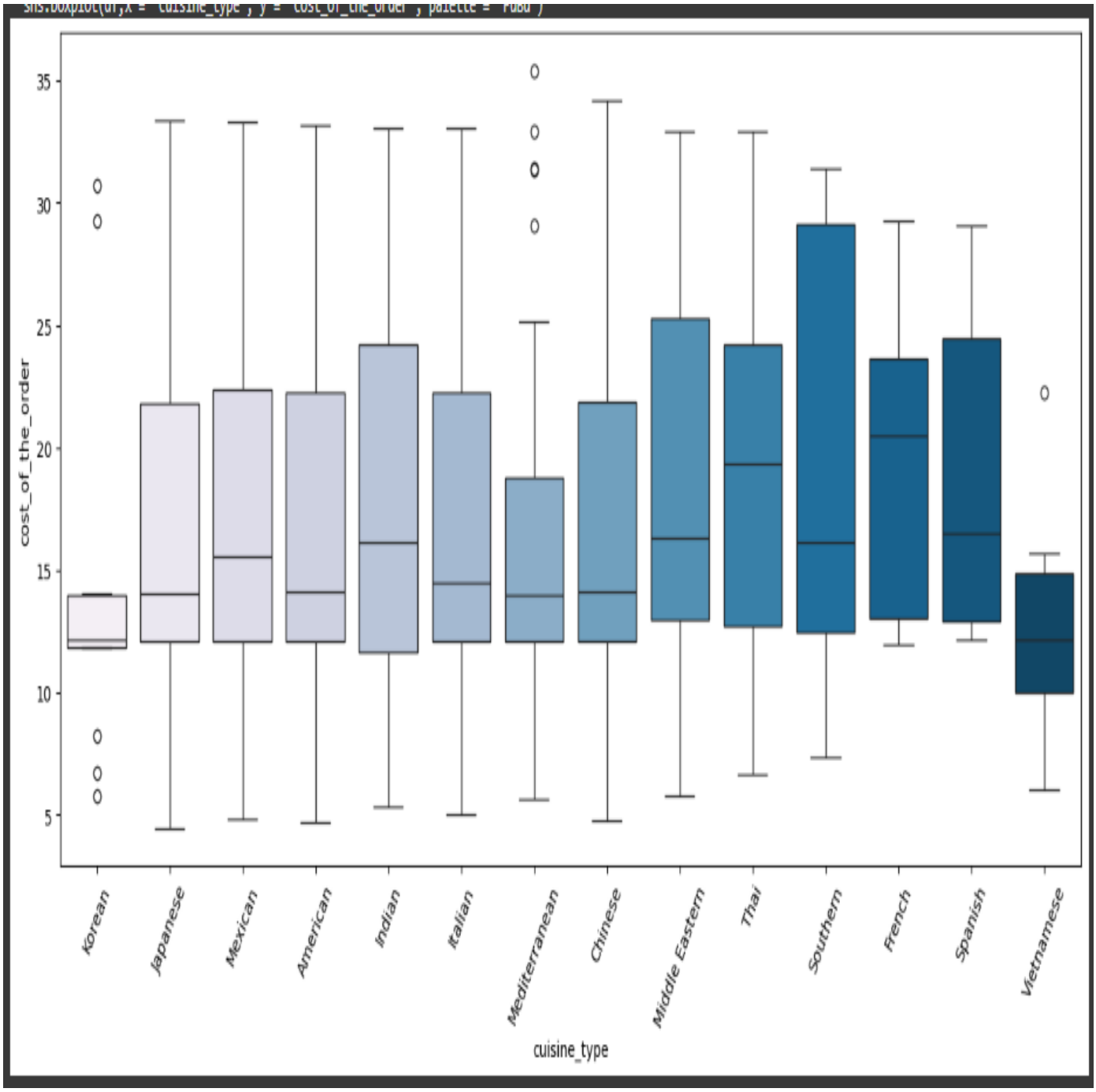
customer_id	count
52832	13
47440	10
83287	9

dtype: int64

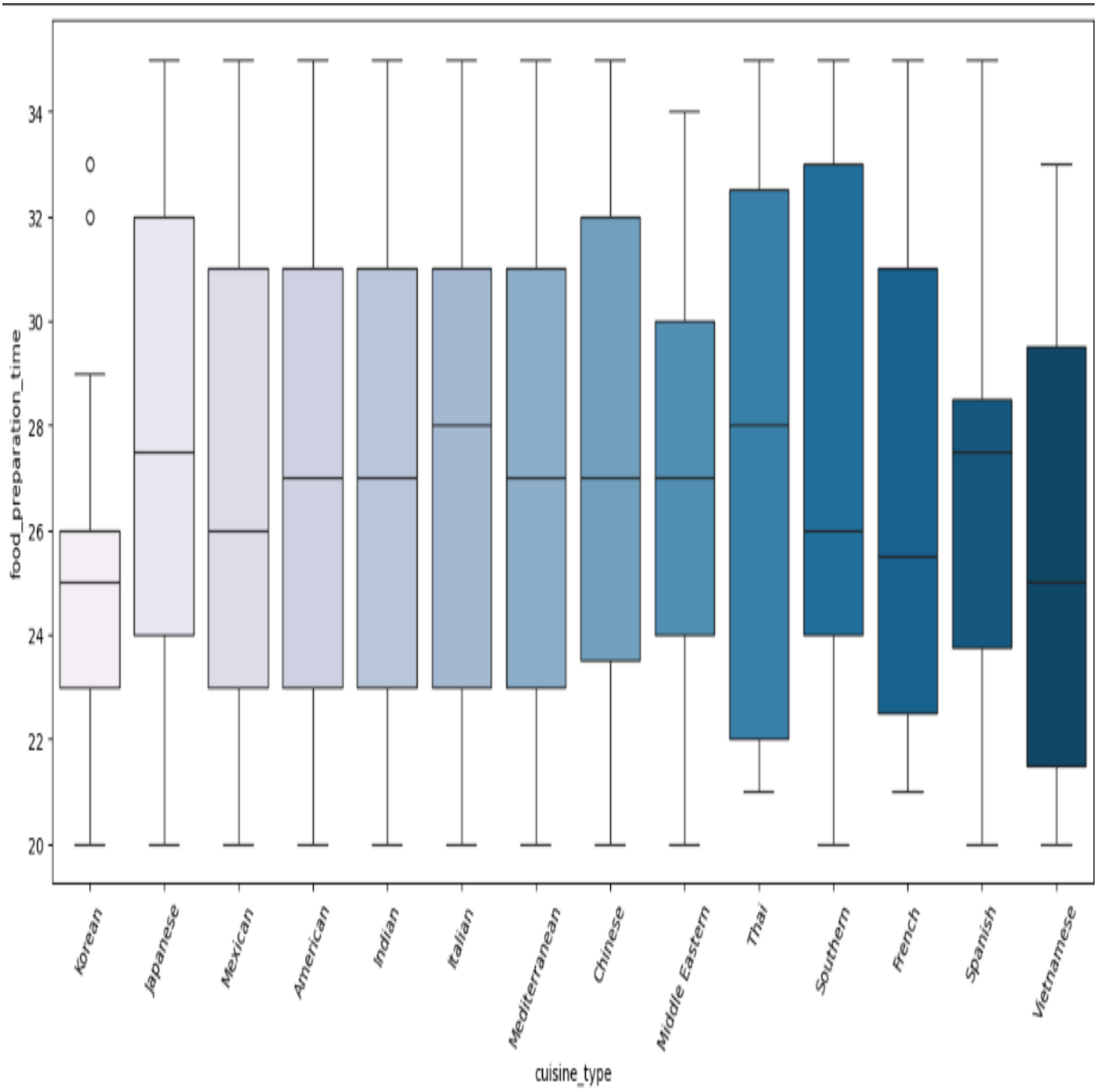
4. Multivariate Analysis

Question 12: 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)

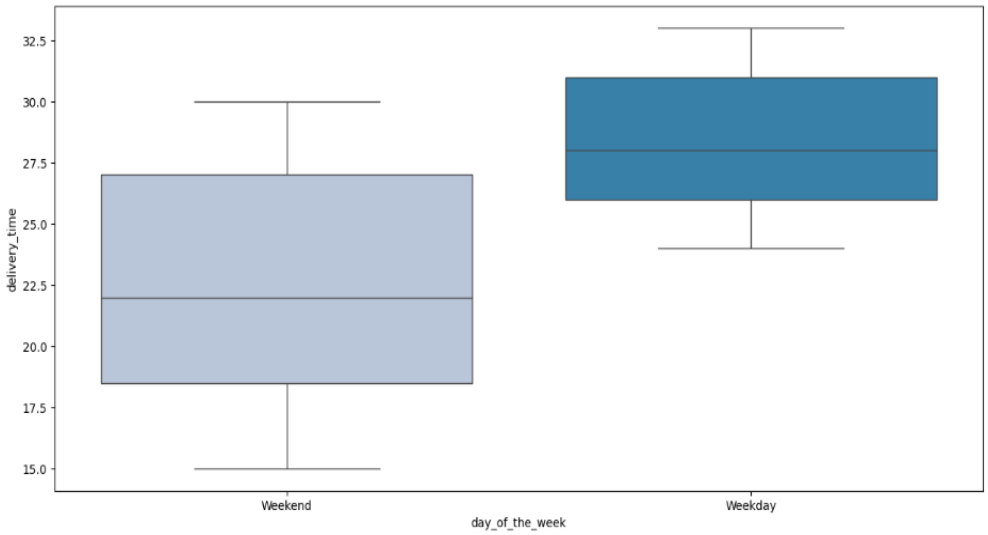
- **4.1** Cuisine vs Cost of the order



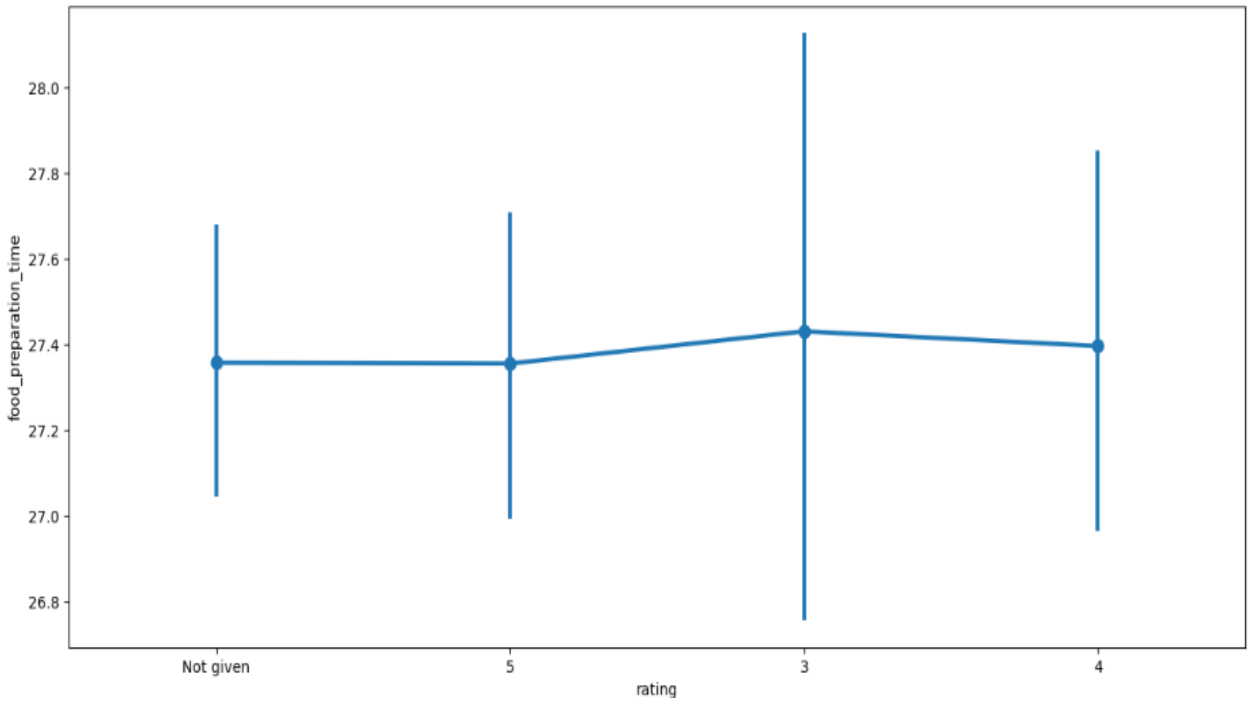
- 4.2 Cuisine vs Food Preparation time



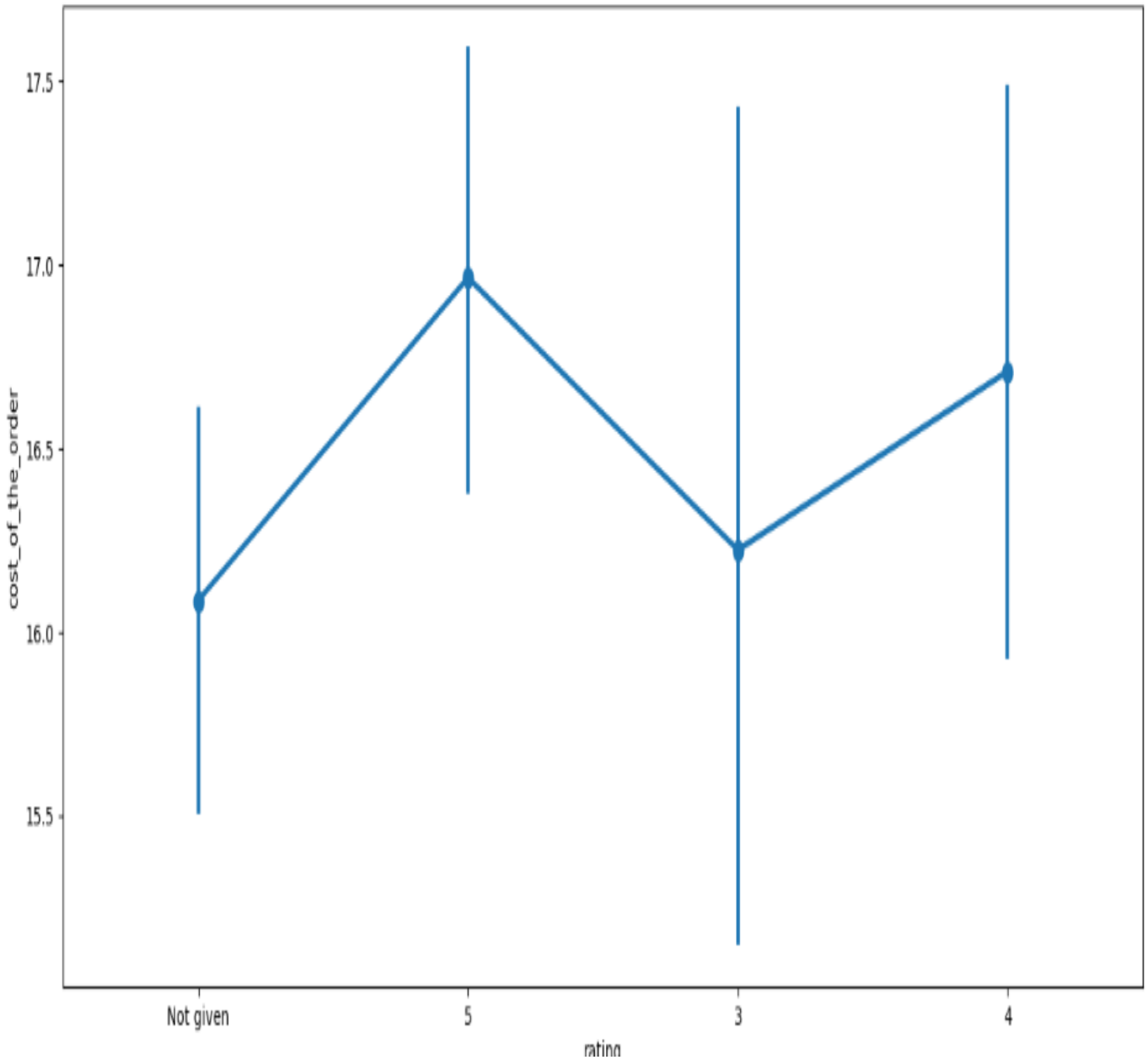
- 4.3 Day of the Week vs Delivery time



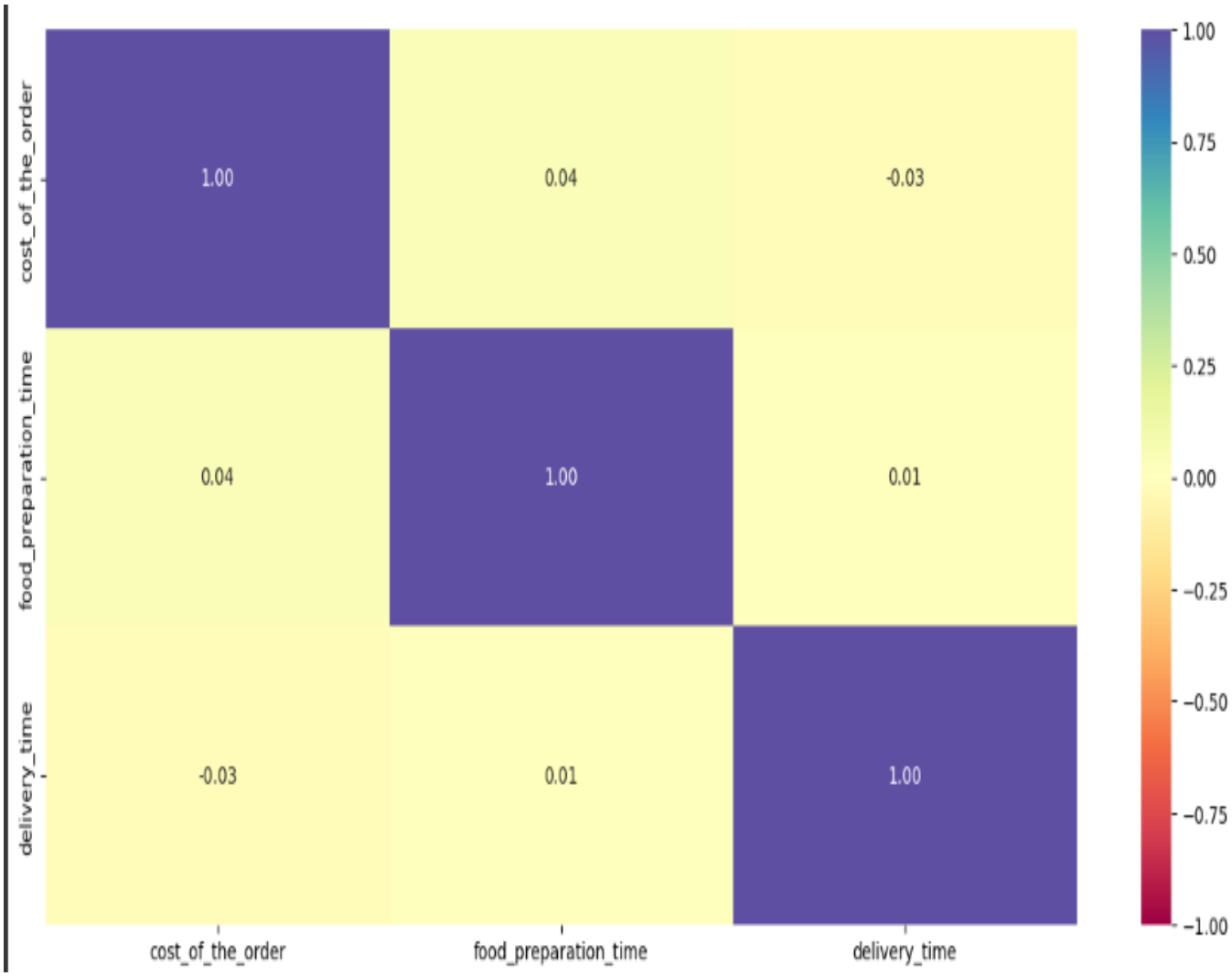
- 4.4 - Point Plot B/W rating and food preparation time



- 4.5 Point B/W Rating and Cost of the order


17.5
17.0
16.5
16.0
15.5
cost_of_the_order
Not given
5
3
4
rating

4.6 Correlation among variables



Question 13: 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.

	restaurant_name	rating
0	The Meatball Shop	4.511905
1	Blue Ribbon Fried Chicken	4.328125
2	Shake Shack	4.278195
3	Blue Ribbon Sushi	4.219178

Question 14: 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.

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

Question 15: 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.

- The percentage of orders that have more than 60 minutes of total delivery time is 10.54 %

Question 16: The company wants to analyze the delivery time of the orders on weekdays and weekends. How does the mean delivery time vary during weekdays and weekends?

- The mean delivery time on weekdays is around 28 minutes

5. Actionable Insights and Recommendations

5.1 **Conclusions**:

- Order count increase on the weekends compared to the weekdays
- Total delivery time over the weekends is less compared to the weekdays
- Approximately 39% of the orders have not been rated
- Shake Shack is the most popular restaurant that has received the highest number of orders.
- Approximately 80% of the orders are for American, Japanese, Italian and Chinese cuisines. Thus, it seems that these cuisines are quite popular among customers of FoodHub

5.2 **Business Recommendations:**

- FoodHub should integrate with restaurants serving American, Japanese, Italian and Chinese cuisines as these cuisines are very popular among FoodHub customers
- FootHub should run special weekends incentives of Delivery partners so that orders demand are easily fulfilled by riders.
- The company should investigate the reason behind the low count of ratings. They can recreate the rating page in the app and make it more interactive to lure the customers to rate the order
- Around 11% of the total orders have more than 60 minutes of total delivery time. FoodHub should find the reason for delays like traffic,terrain issue, Restaurant high preparation time extra and solve this to increase customer satisfaction.
- FoodHub should provide special offers to top-rated popular restaurants that serve most of the orders