Project 1 - FoodHub

Python Foundations - Low Code

Date 15/04/2024

Submission by: Habiba Mohamed

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- Data Overview
- EDA - Univariate Analysis
- EDA - Multivariate Analysis
- Appendix

Executive Summary

Question 17:

Conclusions

- Having 14 different cuisine types; most orders are in the weekends, the American cuisine is on the top of the list, Japanese is the 2nd most order cuisine
- Weekends are the dominated days where people order the most
- Mean delivery time is 24.16 mins
- The majority of FoodHub customers approximately 65% (784 customers) have only placed ONE single order. Followed by the second highest population approximately 22% (267 customers) which has only placed TWO orders.

Executive Summary

Question 17:

Recommendations:

- Promote American and Japanese Cuisine: Focus marketing efforts on these cuisines during weekends, as they are the most ordered types.
- Weekend Specials: Offer special deals or promotions for American and Japanese dishes during weekends to capitalize on peak ordering times.
- Encourage Repeat Orders: Implement loyalty programs or offer incentives to encourage customers to place more than one order, targeting the majority who have only placed one or two orders.
- Improve Delivery Time: Streamline delivery processes to reduce the mean delivery time of 24.16 minutes, enhancing customer satisfaction and loyalty.

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 a Data Scientist at Foodhub 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.

Data Overview

- Data overview given

- 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_of_the_order: 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

Answers of Questions from 1 to 5

Question 1: How many rows and columns are present in the data?
Answer 1: we have 1898 Row and 9 columns, below is the first five rows

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

Question 2: What are the datatypes of the different columns in the dataset?

Answer 2: dtypes: float64(1), int64(4), object(4)

Question 3: Are there any missing values in the data? If yes, treat them using an appropriate method

Answer 3: No missing values, all values and filled

Question 4: 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?

Answer 4 :

The statistical summary of the data which is the time it takes for the food to be prepared once an order is placed is:

Minimum: 20 mins

Average: 27.371970 mins

Maximum: 35 mins

Question 5: How many orders are not rated?

Answer 5: rating not given = 736

Univariate Analysis

Answers from 6 to 11:

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

Order ID:

The number of the unique Order ID is 1898 (which means all numbers are unique)

Customer ID:

The number of the unique Customer ID is 1200

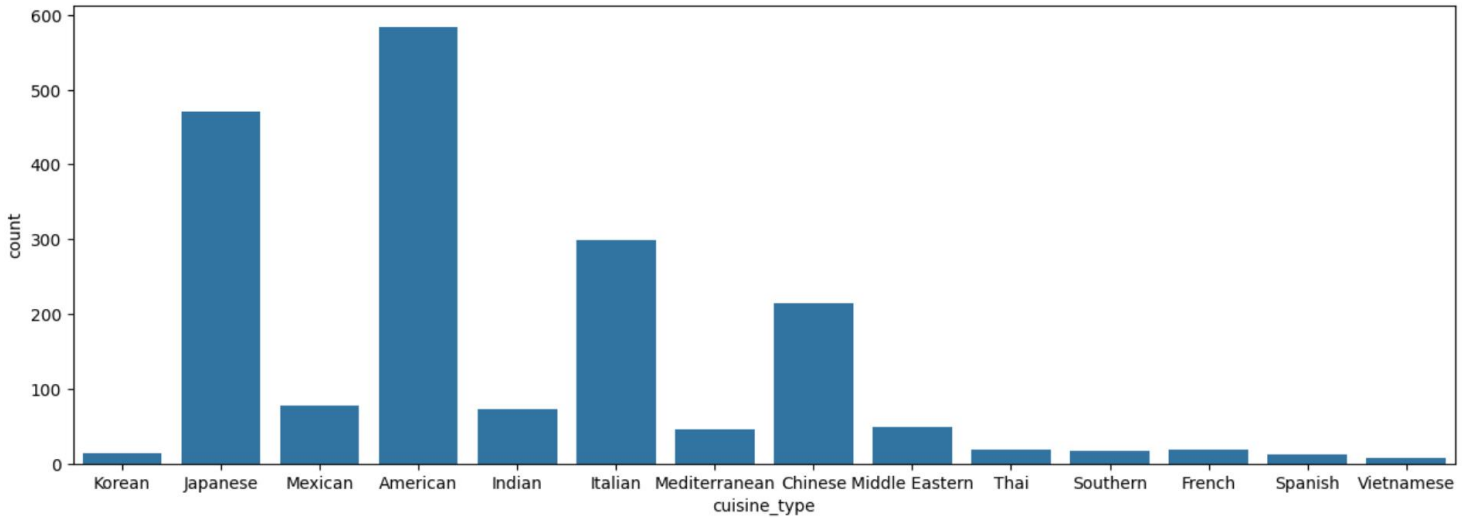
Restaurant Name:

The number of the unique Restaurant Name is 178

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

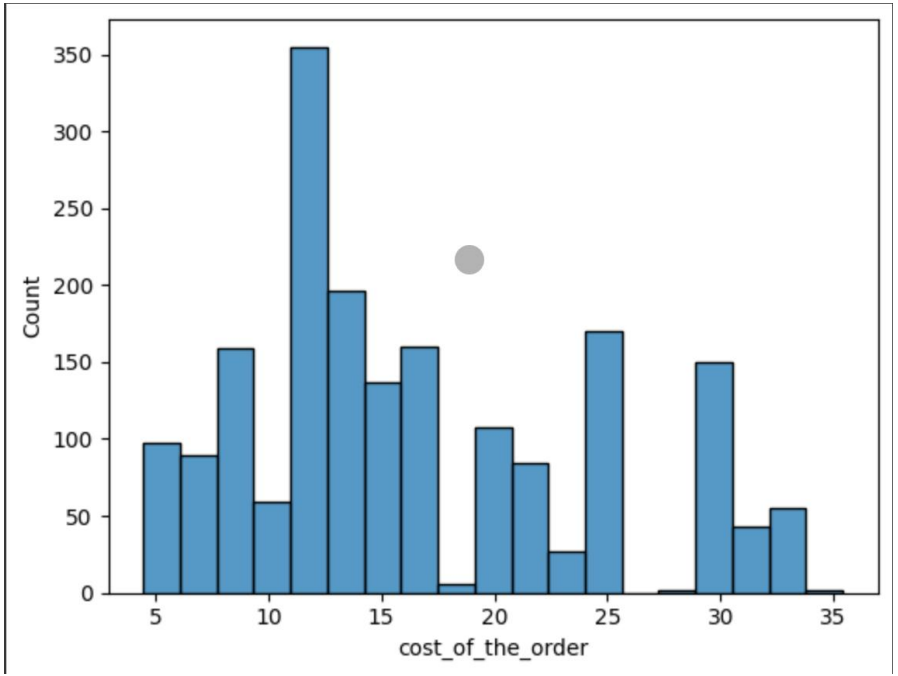
Cuisine type:

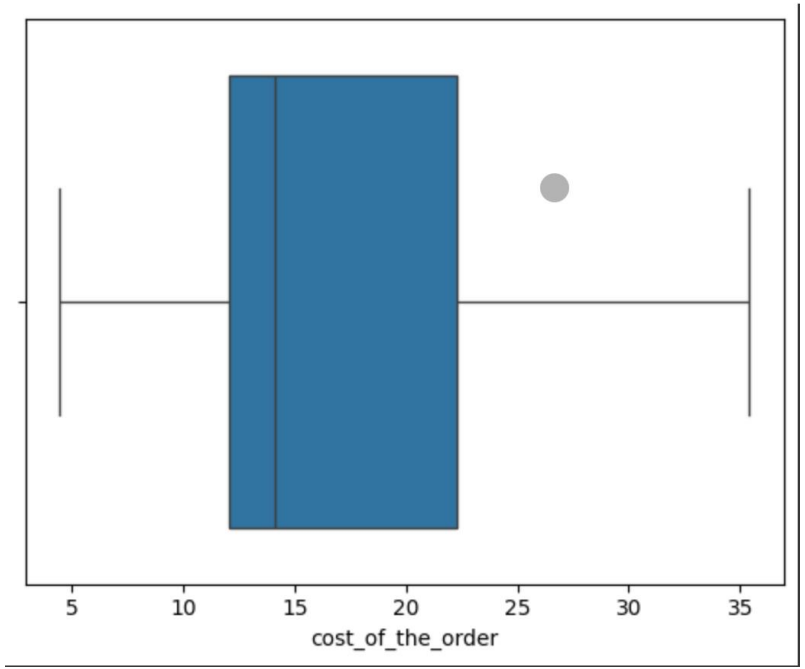
- The number of the unique Cuisine Type is 14
- *The countplot of cuisine type is:*



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

Cost of order:

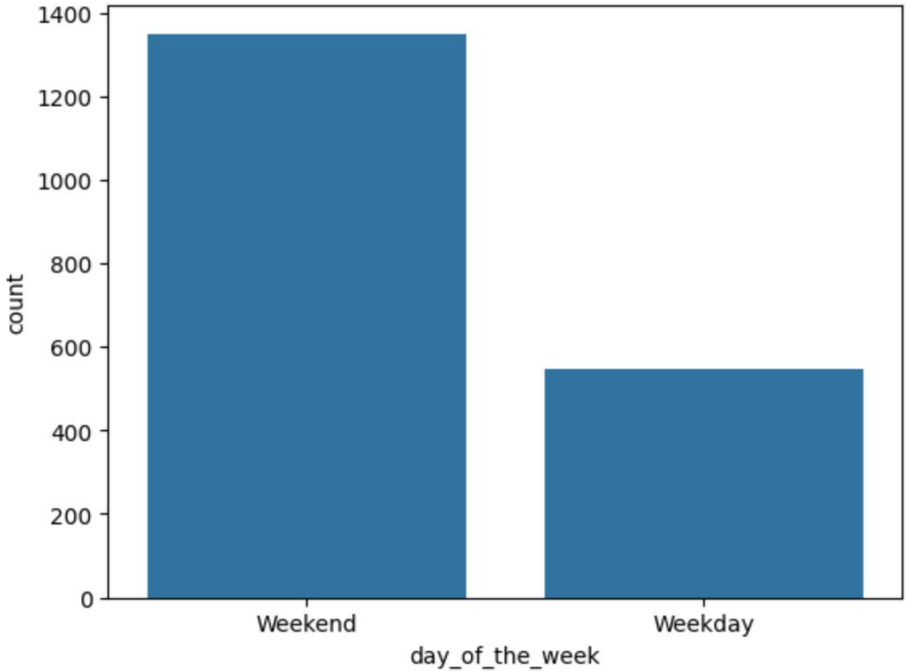




Question 6: Explore all the variables and provide observations on their distributions. (Generally, histograms, boxplots, countplots, etc. are used for univariate exploration.)

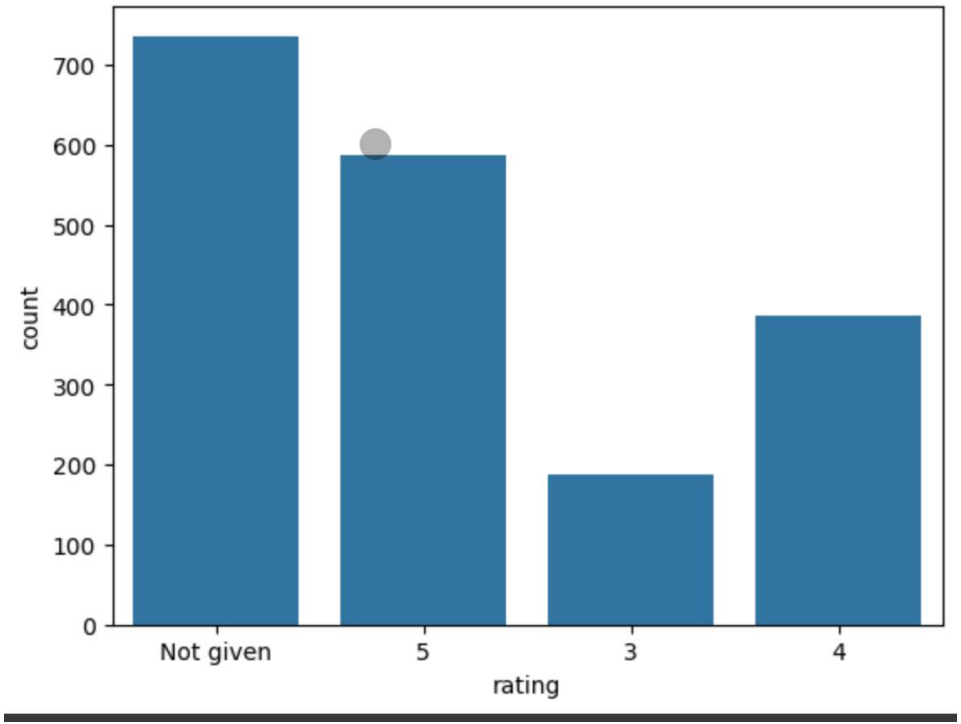
Day of week

The number of unique value for the ‘day_of _the_week” = 2



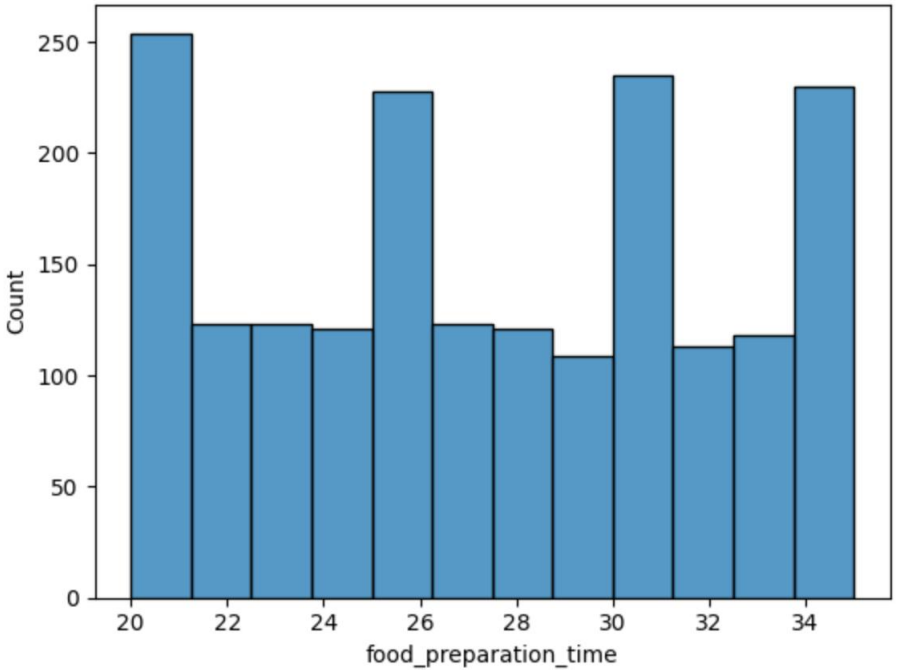
Question 6: Explore all the variables and provide observations on their distributions. (Generally, histograms, boxplots, countplots, etc. are used for univariate exploration.)

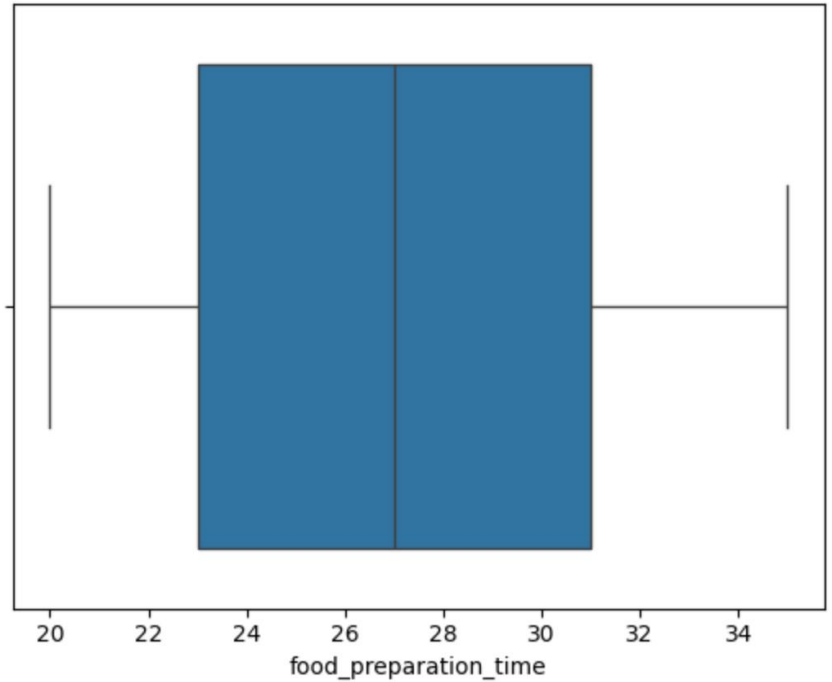
The number of unique value for the ‘rating” = 4



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

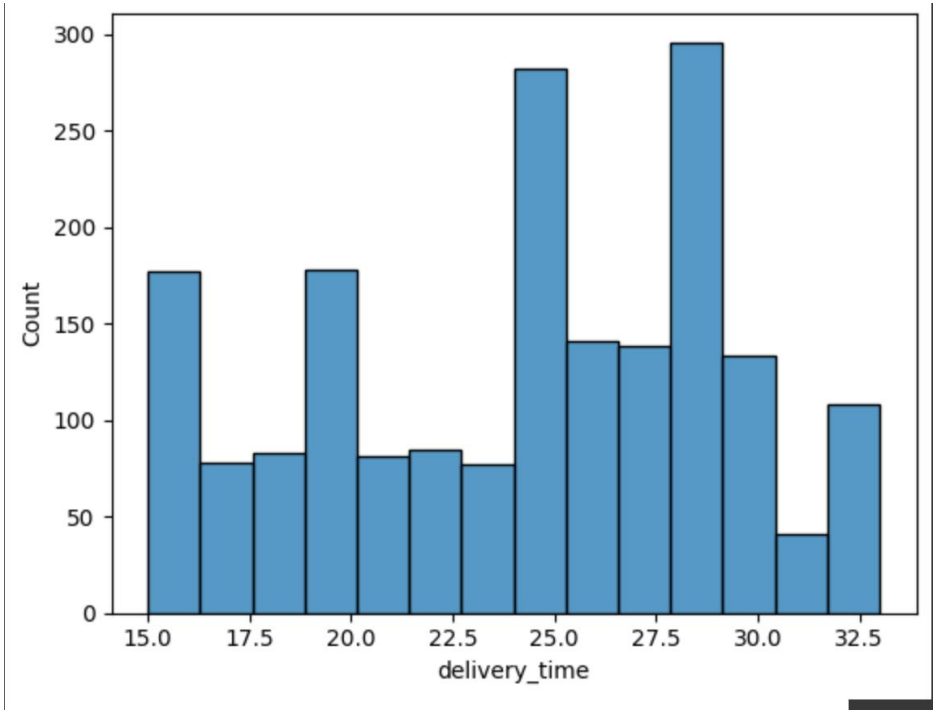
Food preparation time

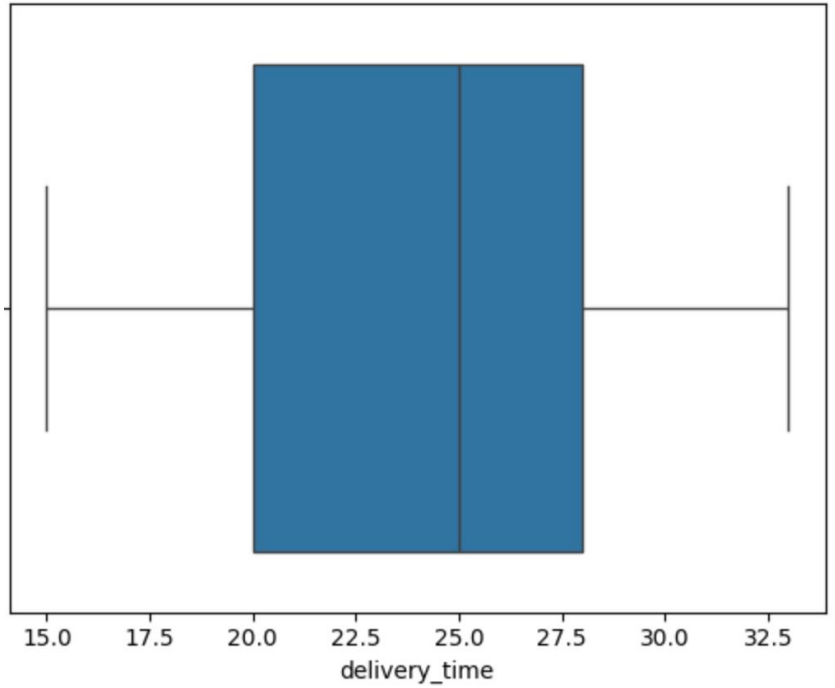




Question 6: Explore all the variables and provide observations on their distributions. (Generally, histograms, boxplots, countplots, etc. are used for univariate exploration.)

Delivery time





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

These are the top 5 restaurants in terms of number of orders

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

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

American Cuisine is on the lead then comes the rest of list after.

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

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

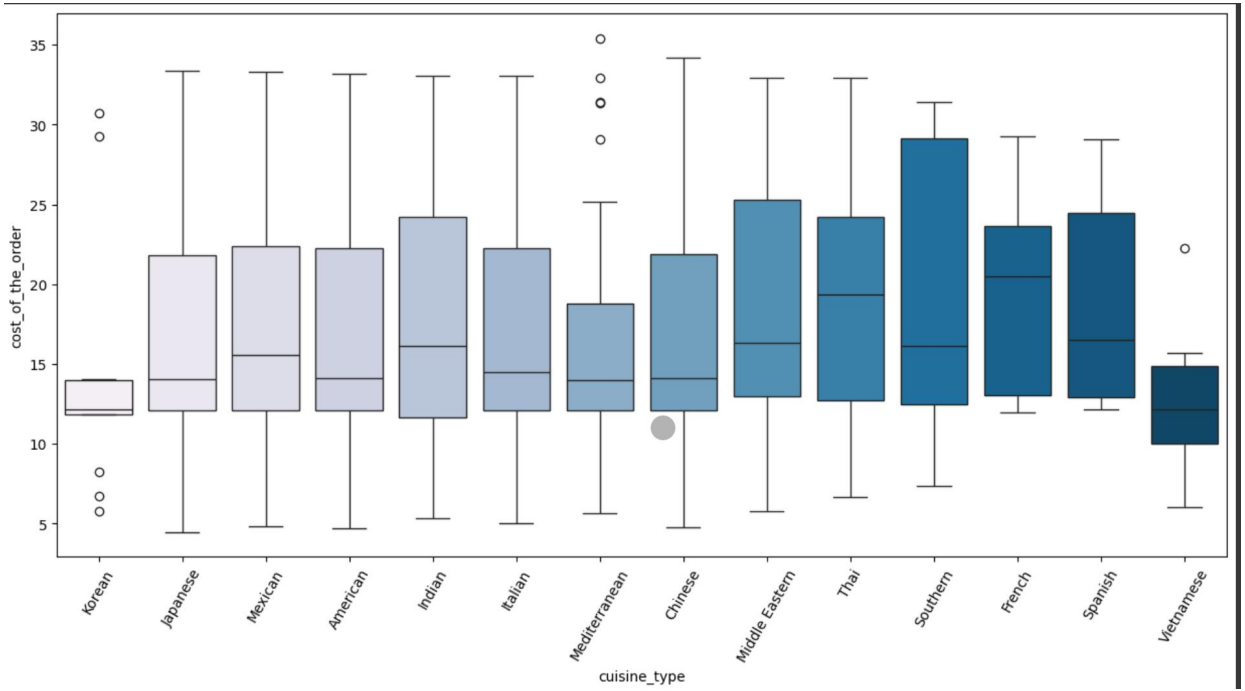
Question 11: The company has decided to give 20% discount vouchers to the top 3 most frequent customers. Find the IDs of these customers and the number of orders they placed.

customer_id	order
52832	13
47440	10
83287	9

Multivariate Analysis: from Question 12 to 16

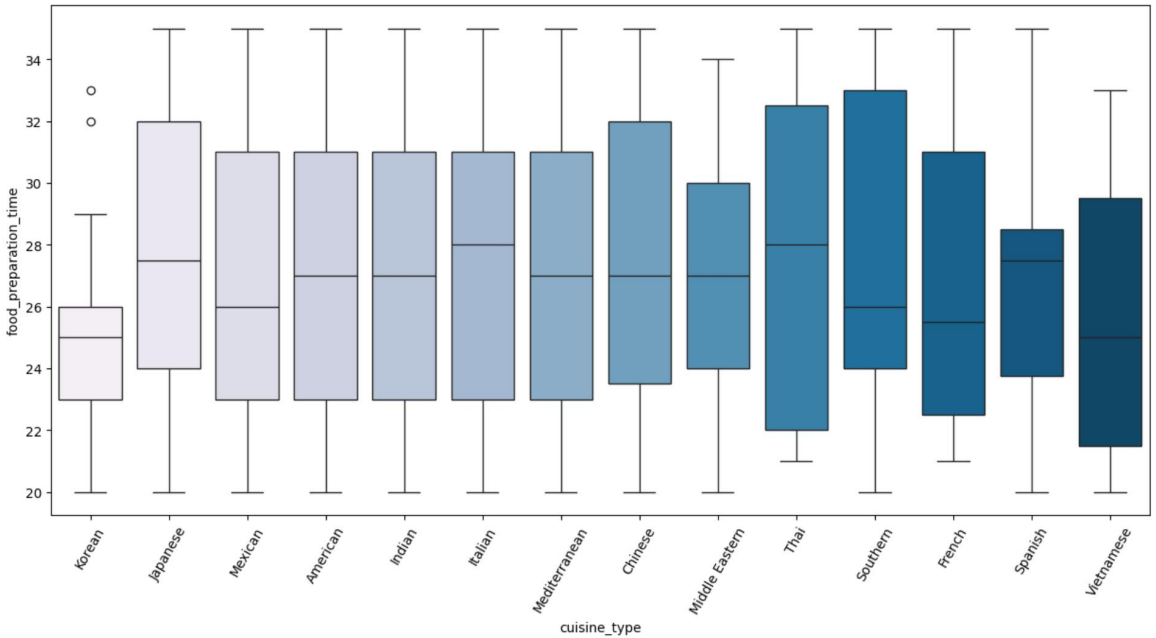
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

Cuisine vs Cost of the order:



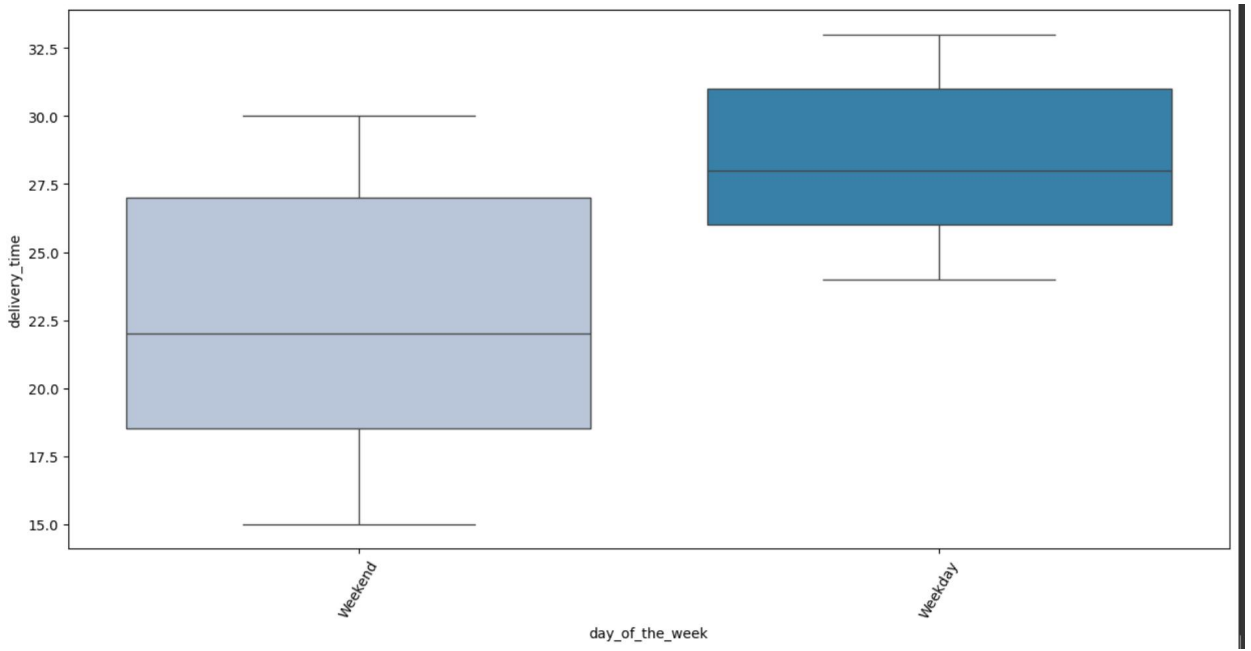
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

Cuisine vs Food Preparation time:



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)

Day of the Week vs Delivery time:



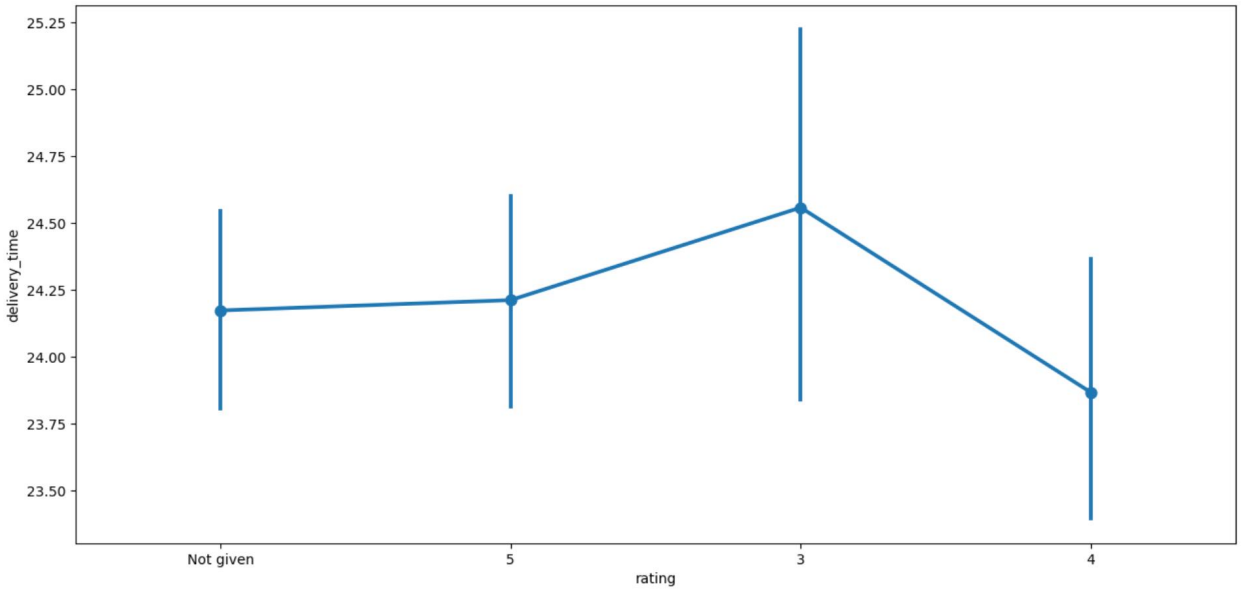
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

The revenue generated by the restaurants:

```
restaurant_name
Shake Shack                      3579.53
The Meatball Shop                2145.21
Blue Ribbon Sushi                1903.95
Blue Ribbon Fried Chicken        1662.29
Parm                             1112.76
RedFarm Broadway                  965.13
RedFarm Hudson                    921.21
TAO                               834.50
Han Dynasty                       755.29
Blue Ribbon Sushi Bar & Grill     666.62
Rubirosa                          660.45
Sushi of Gari 46                  640.87
Nobu Next Door                    623.67
Five Guys Burgers and Fries       506.47
Name: cost_of_the_order, dtype: float64
```

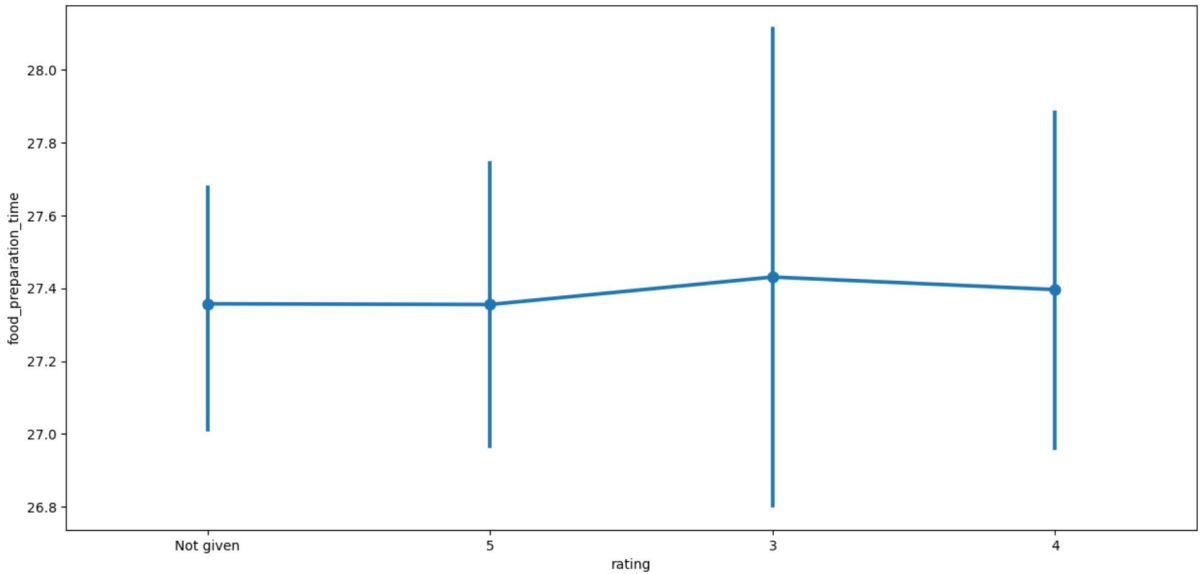
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

Rating vs Delivery time:



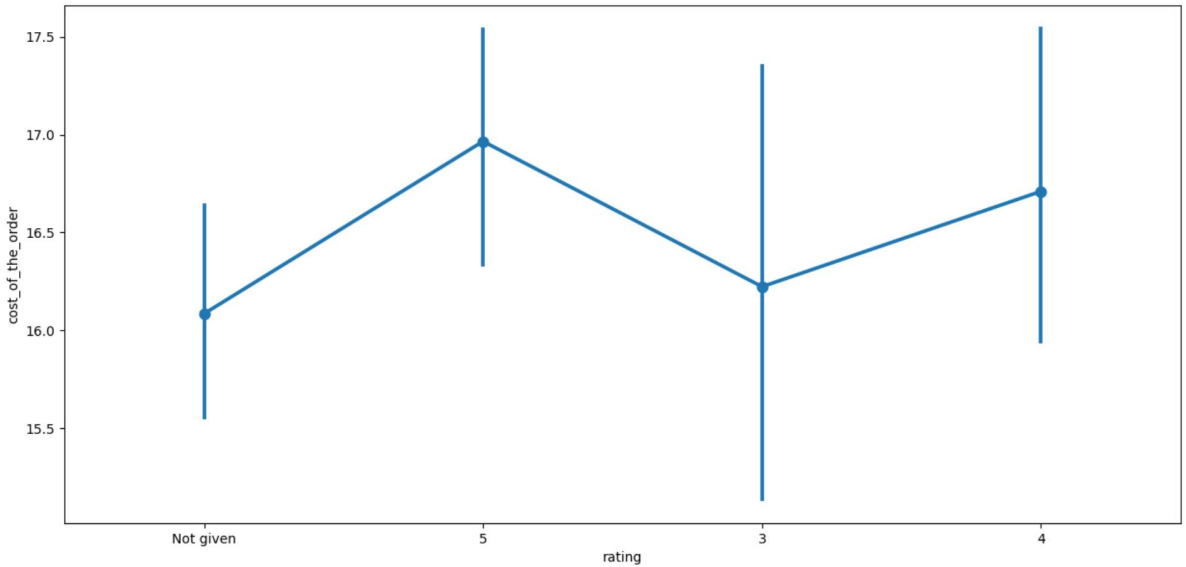
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

Rating vs Food preparation time:



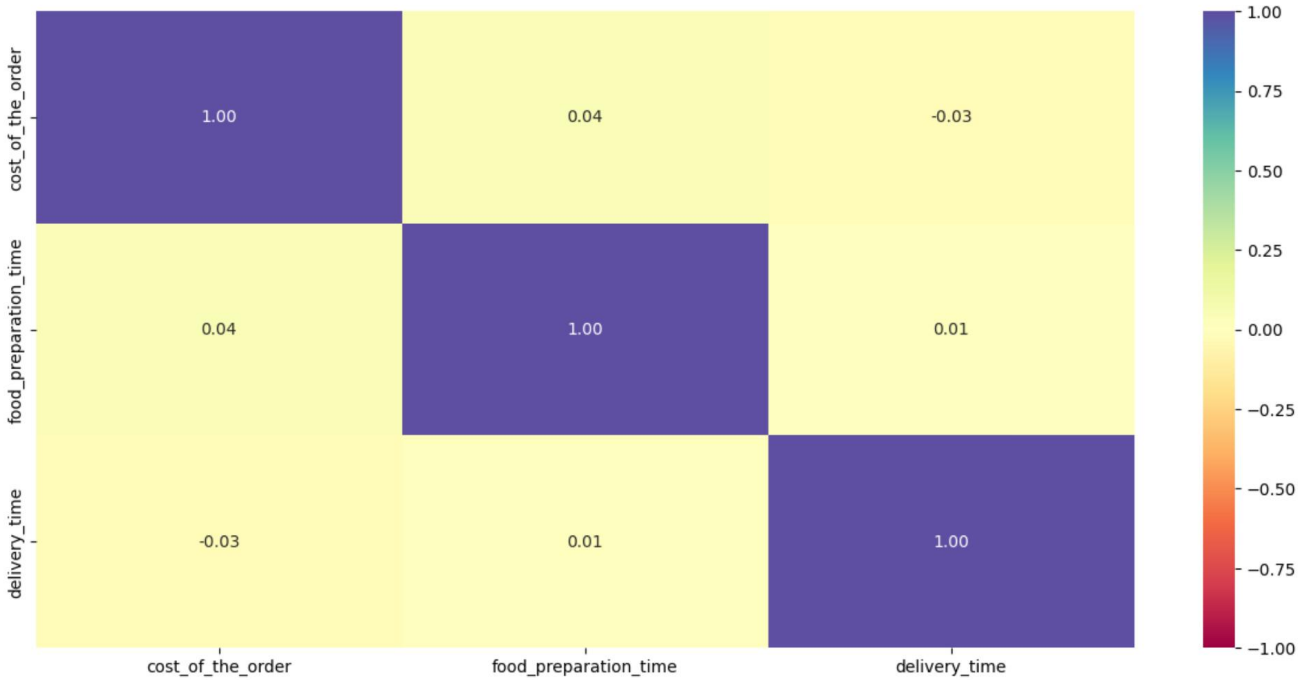
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

Rating vs Cost of the order:



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)

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	Shake Shack	133
1	The Meatball Shop	84
2	Blue Ribbon Sushi	73
3	Blue Ribbon Fried Chicken	64
4	RedFarm Broadway	41

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

- The net revenue on all orders is: 6166.30

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.)

- Percentage 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, and 22 mins on weekends

More recommendations for Question 17

Certainly, here are some additional recommendations that helps customer satisfaction and sales:

- Weekend Promotions: Introduce weekend-exclusive promotions or discounts to incentivize more frequent ordering during peak times.
- Enhance User Experience: Continuously improve the app's user interface and experience to make ordering food easier and more enjoyable for customers.
- Personalized Recommendations: Implement a recommendation system based on past orders and customer preferences to suggest relevant dishes, increasing the likelihood of repeat orders.
- Optimize Delivery Logistics: Invest in technology or partnerships to optimize delivery routes and reduce delivery times further, enhancing overall customer satisfaction.
- Customer Engagement: Engage with customers through targeted email campaigns, social media, or push notifications to keep them informed about new offerings, promotions, and updates.
- Feedback Mechanism: Implement a feedback mechanism within the app to gather insights from customers about their ordering experience, food quality, and delivery service, allowing for continuous improvement.
- Community Building: Foster a sense of community among users by featuring user-generated content, such as reviews and photos, and encouraging social sharing of food experiences.
- By implementing these recommendations, the online food delivery app can better capitalize on its strengths, improve customer satisfaction, and drive increased sales and engagement.

Great Learning

Happy Learning !