

# FoodHub Data Analysis

## Python Foundations PGP-DSBA-UTA

October 11, 2023



# Contents / Agenda

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

# Executive Summary

## Objectives:

- To provide an overview of the data within available context provided.
- To get a fair idea about the demand of different restaurants.
- To explore ways to enhance customer experience.

## Conclusions, actionable items:

- The data shows that orders during the weekend are more popular in the weekend. The delivery time is shorter during the weekend, but other reasons customers prefer to order during the weekend also play a role not spelled out in this data set.
- There may be room for growth in promoting cuisine types Vietnamese and Korean, as their associated ratings are comparable to the popular cuisines. It could be due to lack of familiarity with the majority of food hub. A special discount campaign could help introduce this cuisine type.

# Business Problem Overview and Solution Approach

- How can Foodhub enhance customer experience, while the rating function in the app is underutilized by its customers.
- Introduce discount campaign to under optimized cuisine types, as an incentive to increase orders.
- Explore ways to encourage customers to utilize the rating app. After 5 orders + rating given, a discount for next order will be given.

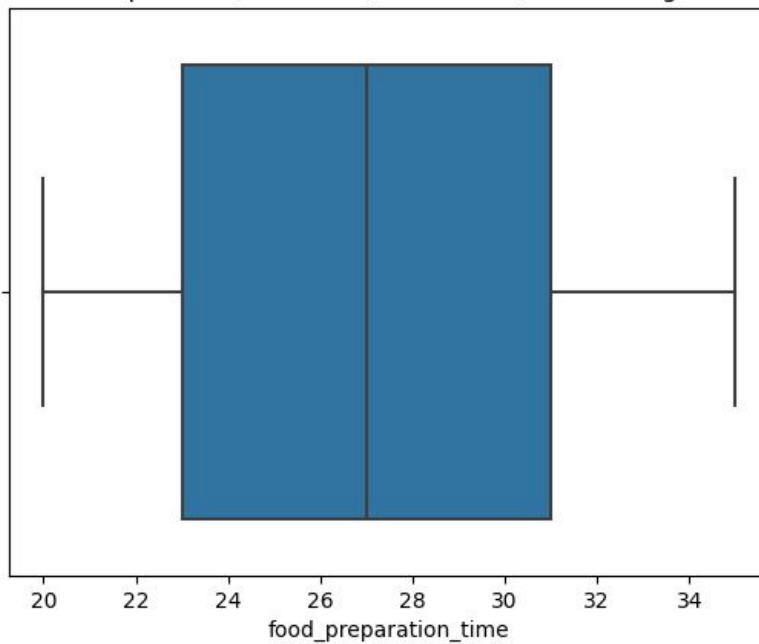
**Note:** *You can use more than one slide if needed*

## Data Overview FoodHub

- The dataset shows us 9 columns, and 1898 rows. Rows reflect the number of orders. The columns include **order\_id**, **customer\_id**, **restaurant\_name**, **cuisine\_type**, **cost\_of\_the\_order**, **day\_of\_the\_week**, **rating**, **food\_preparation\_time**, and **delivery\_time**. (answer to question 1)
- There are three data types in this dataset: Integers (datatype: **int64**) in the columns Order ID, Customer ID, Food preparation time, and Delivery time; Text (datatype: **object**) in the columns Restaurant name, Cuisine Type, Day of the week, and Rating. Decimals (object: **float64**) in the column Cost of order. (answer to question 2)
- It turns out that 736 orders, or 38.77% of total orders were not rated. The column Rating has datatype object, since a text value “not given” was entered. This percentage of missing ratings is significant and should question the overall usefulness of relying on the results of rated orders: 558 orders rated 5, 386 orders rated 4, 188 orders rated 3. (answer to question 3 and 5) See appendix for graph.

## Data overview continued

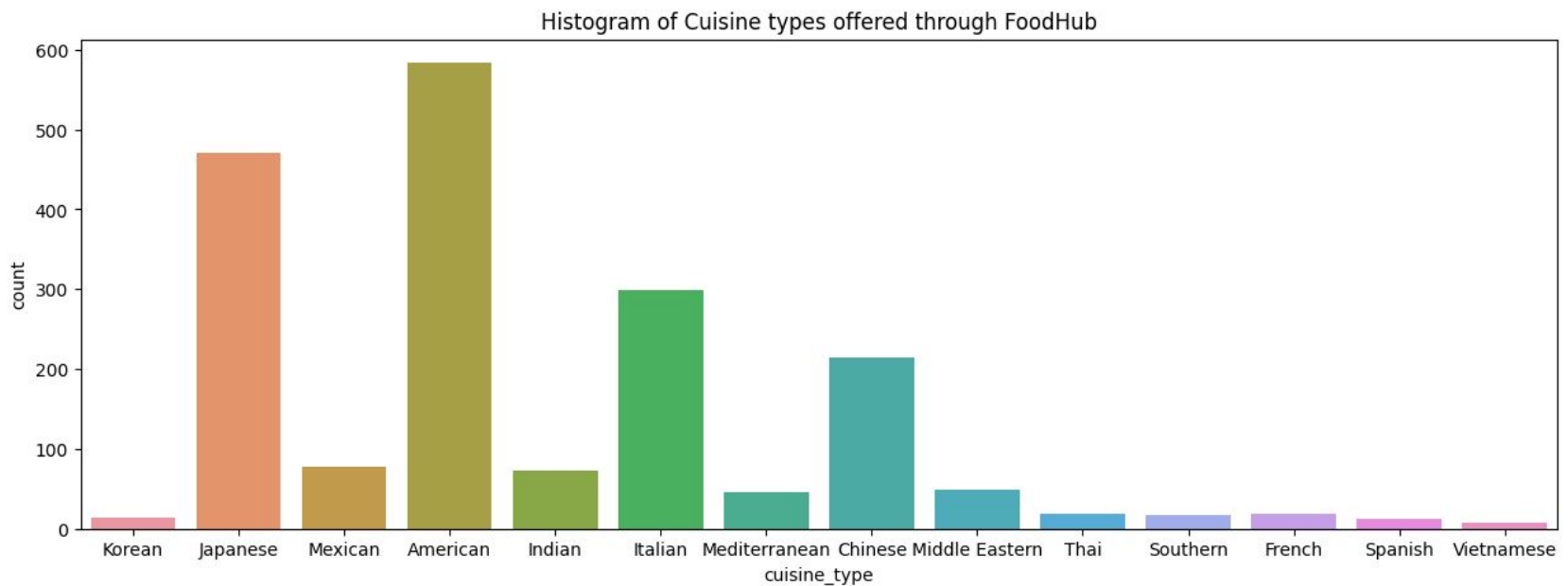
Food Preparation, minimum, maximum, and average time



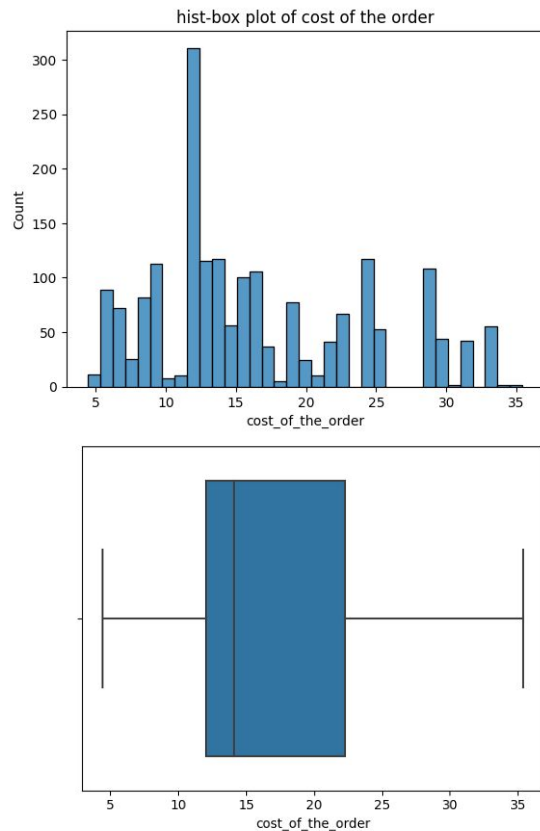
- The minimum amount of time it takes for food to be prepared once an order is placed is 20 minutes, the maximum amount is 35 minutes. The average is 27 minutes. (question 4)

# Exploratory Data Analysis: Univariate Analysis

- 1898 unique orders (**orders\_id**) were placed by 1200 unique customers (**customer\_id**) to 178 unique restaurants (**restaurant\_name**) representing 14 different cuisines (**cuisine\_type**).
- This countplot shows American cuisine as most popular, followed by Japanese and Italian cuisine.



## Univariate Analysis continued



A detailed histogram plot shows us volatility in distribution not visible on the boxplot: Number of orders in the USD10-11 range for instance hover around 10, while number of orders for around USD12 exceed more than 300.

The boxplot of cost of order shows us the visualization of the mean (~\$14), IQR (~\$11 to 22.50), minimum (~\$5) and maximum (~\$35) cost of the order, left whisker covering 1st quartile, and right whisker covering 4th quartile of the distribution.

The distribution is skewed towards the right.



## Univariate Analysis Continued

- Ordering on weekend occurs more than twice compared to ordering on a weekday. See countplot in appendix.
- The rating 3 is the least given, while no rating is the most frequent value. See countplot in appendix
- This histogram shows that there are four peaks in the food preparation time distribution: 20-21 minutes; 25-26 minutes; 30-31 minutes; and 34-35 minutes. The boxplot shows the average food delivery time is approx 27 minutes, with IQR of 23 to 31 minutes. SHOW COMBO HIST / BOX PLOT MULTI PEAK IN HISTO PLOT
- This histogram and boxplot shows that the delivery time is slightly skewed towards the left. The average delivery time is 25 minutes, with IQR between 20 to 28 minutes approximately. Highest distribution is 25 and 28 column.

## Univariate analysis continued part ii

**Question 7:** Which are the top 5 restaurants in terms of the number of orders received? [1 mark]

- The top 5 restaurants in terms of number of orders received are: 1. Shake Shack (219 orders) 2. The Meatball Shop (132 orders) 3. Blue Ribbon Sushi (119 orders) 4. Blue Ribbon Fried Chicken (96 orders) 5. Parm (68 orders).

**Question 8:** Which is the most popular cuisine on weekends? [1 mark]

- The most popular cuisine on weekends is American Cuisine with 415 orders.

**Question 9:** What percentage of the orders cost more than 20 dollars? [2 marks]

- 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? [1 mark]

- 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. [1 mark]

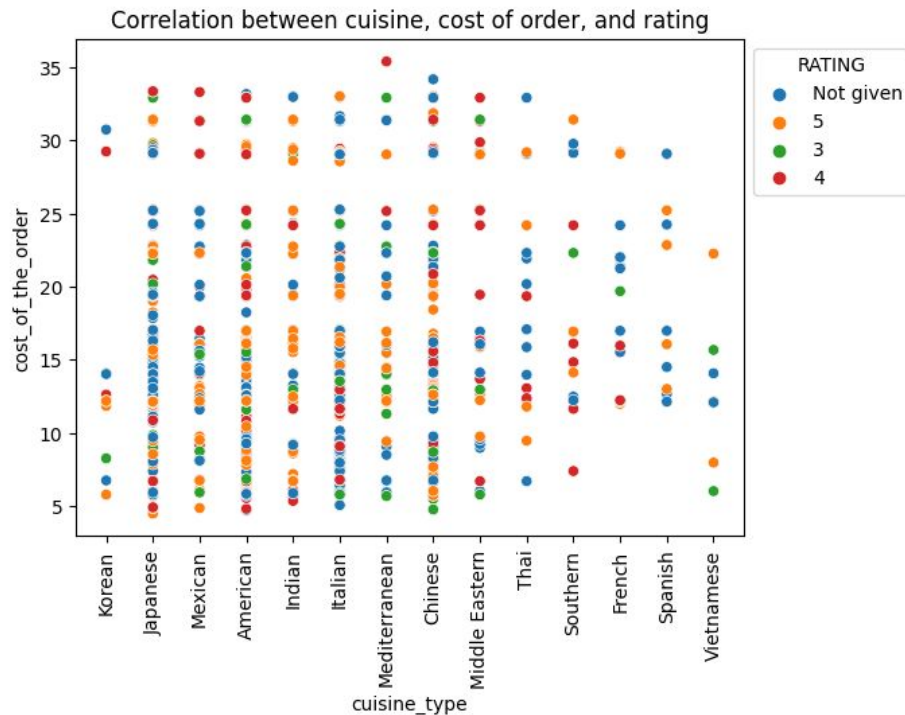
Customer id: 52832 13 orders

Customer id: 47440 10 orders

Customer id: 83287 9 orders

# 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) [10 marks]

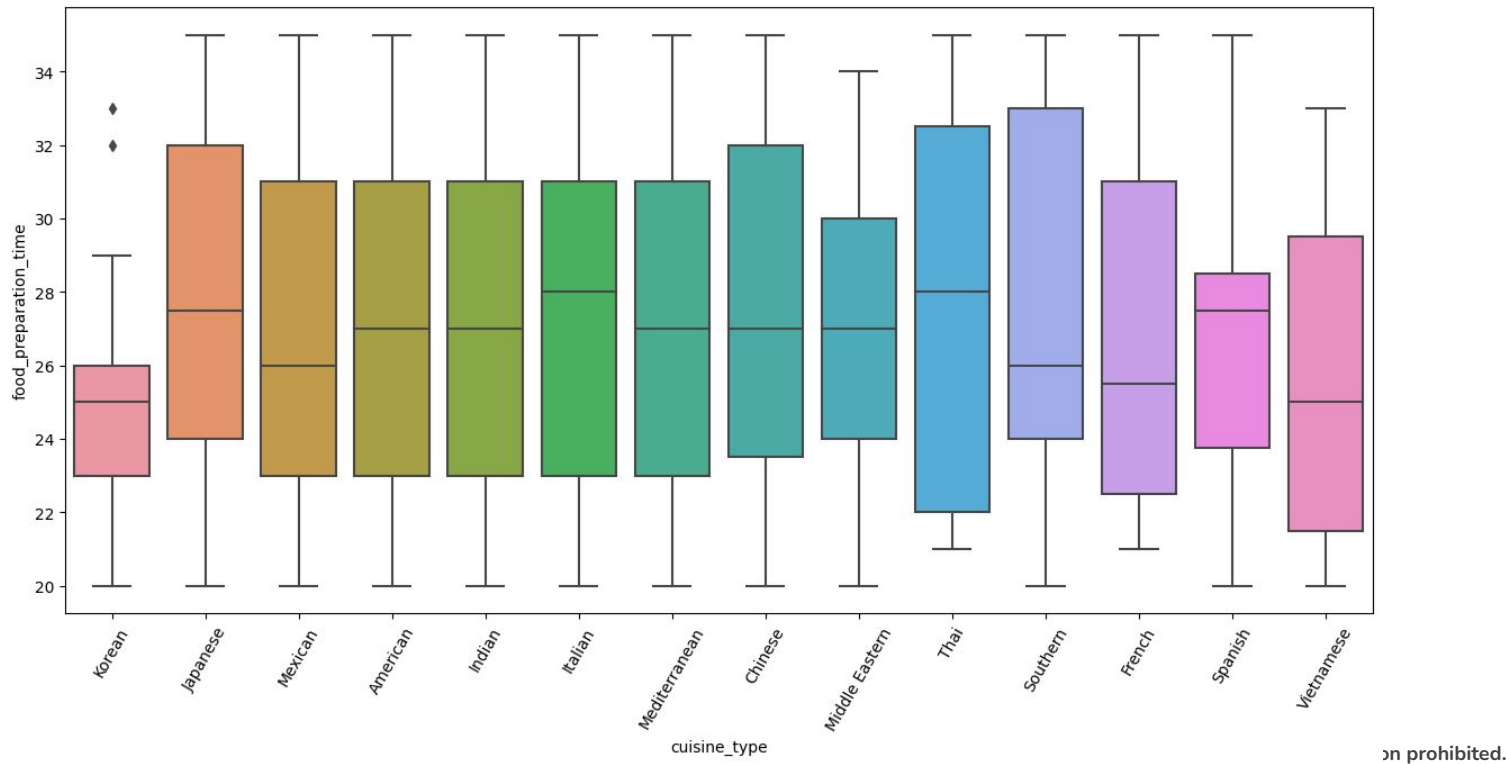


In this scatterplot we see visualized the popularity of various cuisines, and their associated ratings. We can not determine however how many individual restaurants fall within each cuisine type. For instance, there may only be two Vietnamese restaurants in the area resulting in the relative few orders.



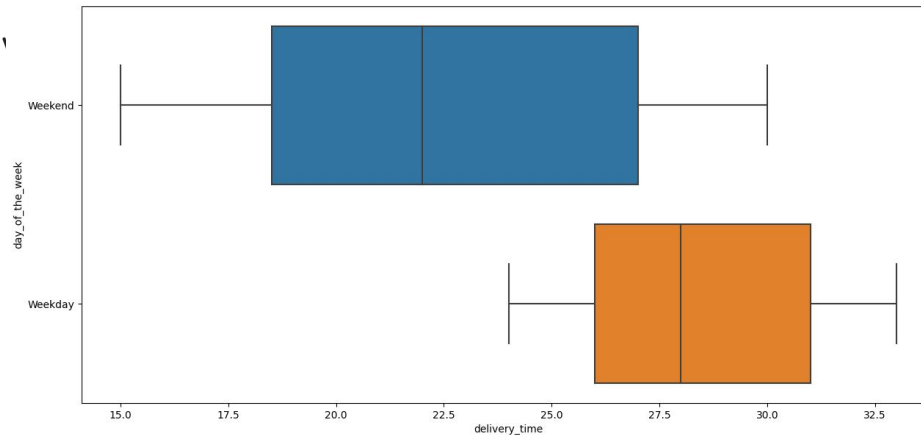
# Multivariate analysis continued

Relationship between food preparation time and cuisine type:



## Multivariate analysis continued

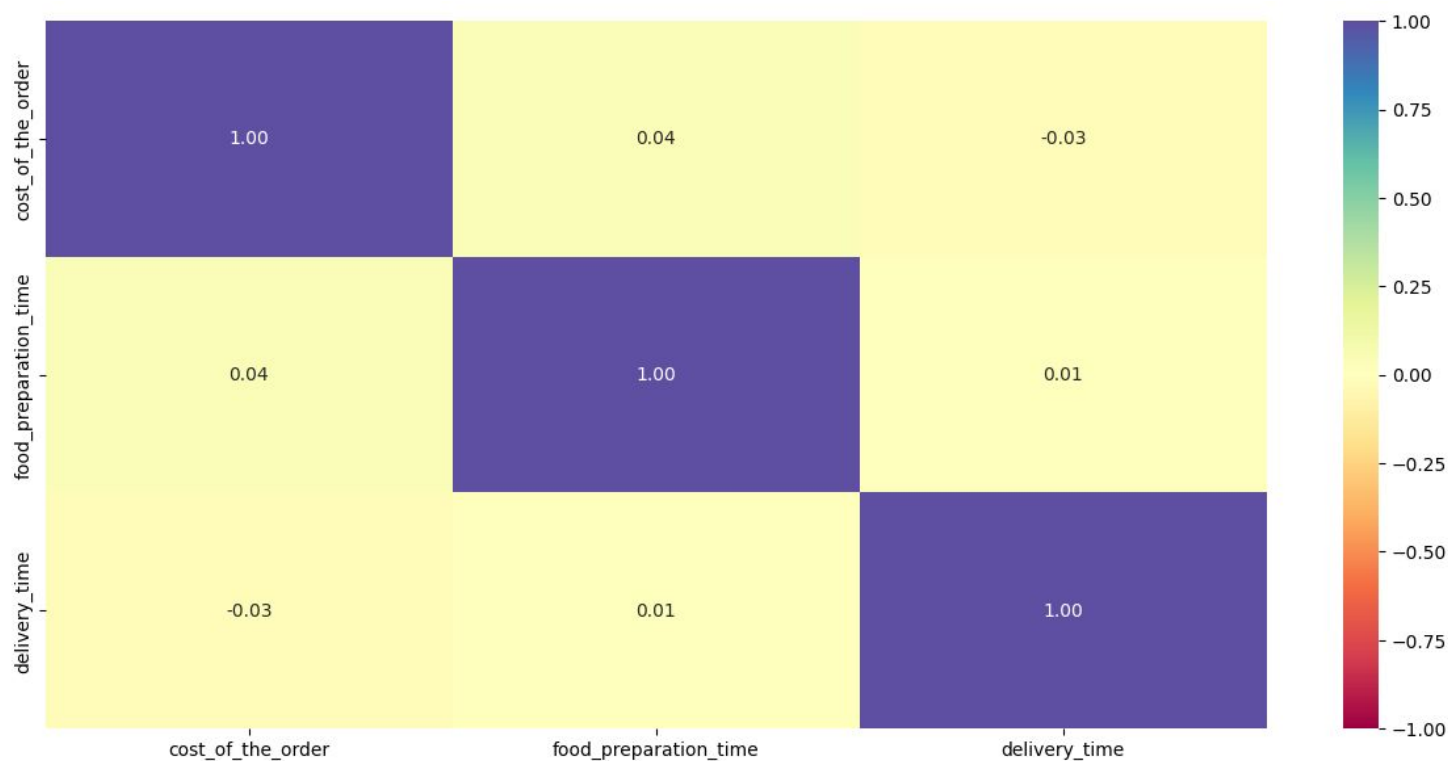
- Relationship between day of the week and delivery time: Delivery time during weekend is shorter.



- The top 14 restaurants with highest cost of order, happen to be also the most popular restaurants with the highest number of orders. For instance, Shake Shack at number 1, has a total cost of order USD 3,579.53 from 219 orders. Average cost of order at Shake Shack is \$16.34

## Multivariable analysis continued

There is neutral correlation between cost of the order, food preparation time, and delivery time.



## Multivariate analysis continued II

**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. [3 marks]

	restaurant_name	rating
0	The Meatball Shop	4.51190 5
1	Blue Ribbon Fried Chicken	4.32812 5
2	Shake Shack	4.27819 5
3	Blue Ribbon Sushi	4.21917 8



## Multivariate Analysis continued part III

**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. [3 marks]

- The net revenue is around 6166.3 dollars

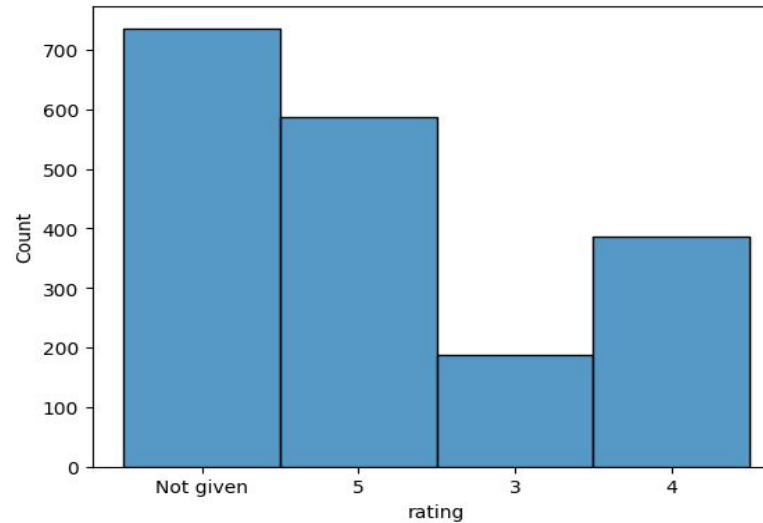
**Question 15:** The company wants to analyze the total time required to deliver the food. What percentage of orders take more than 60 minutes to get delivered from the time the order is placed? (The food has to be prepared and then delivered.) [2 marks]

- The number of orders with total time longer than 60 is: 200
- Percentage of total delivery time longer than 60 minutes: 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? [2 marks]

- The mean delivery time on weekdays is around 28 minutes
- The mean delivery time on weekends is around 22 minutes

Re: Question 5:  
Histogram showing rate  
distribution 1 through 5. Notice  
ratings 1 and 2 are missing in the  
data. More than a third of the  
data for ratings were 'not given'  
by the customer.



# APPENDIX



## Slide Header



- Please add any other pointers (if needed)



**Happy Learning !**

