FoodHub Data Analysis

Statistical Methods For Decision Making

Prepared by:
Soumalya Sen
Soumalyasen25@gmail.com September 10,2023

Contents / Agenda

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

Executive Summary

- FoodHub should integrate with restaurants serving American, Japanese, Italian, and Chinese cuisines as these cuisines are very popular among FoodHub customers (accounting for ~80% of the orders).
- FoodHub should provide promotional offers to top-rated popular restaurants that serve most of the orders as the top 5 restaurants account for ~80% of the orders.
- Order volumes are higher (by ~60%) on the weekends compared to the weekdays. As such, more delivery executives should be employed during the weekends to ensure timely delivery of the orders. Weekend promotional offers can also be rolled out to the customers to increase the food orders during weekends.
- Delivery time over the weekends is less compared to the weekdays despite the higher number of orders. This could possibly be due to the dip in traffic volume over the weekends, but further analysis is needed to verify the same.
- The customer rating is a very important factor to gauge customer satisfaction and ~39% of the orders were not rated. The company should investigate the reason behind the low count of ratings. They can redesign the rating page in the app and make it more interactive to lure the customers to rate the order. Customer feedback comments should also be analyzed for further insights.
- Around 11% of the total orders have more than 60 minutes of total delivery time. FoodHub should try to minimize such instances in order to avoid customer dissatisfaction. They can provide some reward to the punctual delivery person.

Business Problem Overview

- 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 the 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.

Solution Approach

- Understand the demand of restaurants in the FoodHub portal
- Cuisine preference of the New York customers
- Get an idea about the cost of the ordered food
- Understand the volume of the orders over weekdays and weekends
- Estimate the revenue generated by the company
- Help the company to take decision on promotional offers
- Order rating analysis

Data Overview

Displaying the first five rows of the dataset:

	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

Checking the shape of the dataset:

```
(1898, 9)
```

Observations: I can see from the shape that the table has 1898 rows and 9 columns.

Datatypes of the different columns in the dataset:

```
<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
```

Observations: The three different datatypes are integer, object (string or mixed type) and float.

check any missing values in the data:

```
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
```

Observations: There are no missing values in the rows for each column in the dataset.

Check the statistical summary of the data

	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

Observation:

- There are 1898 orders.
- On average an order takes 27 minutes to be prepared and 24 minutes to deliver totalling approximately an hour waiting time 50% of orders will typically be delivered within this timeframe an hour.
- Shake Schack is the restaurant with the highest number of orders American cuisine is the most popular cuisine with the highest number of orders.
- Most orders are made on a weekend Most orders do not have a rating. An individual order cost 16.5 dollars on average but can cost up to 35 dollars.
-

Exploratory Data Analysis (EDA): Univariate Analysis

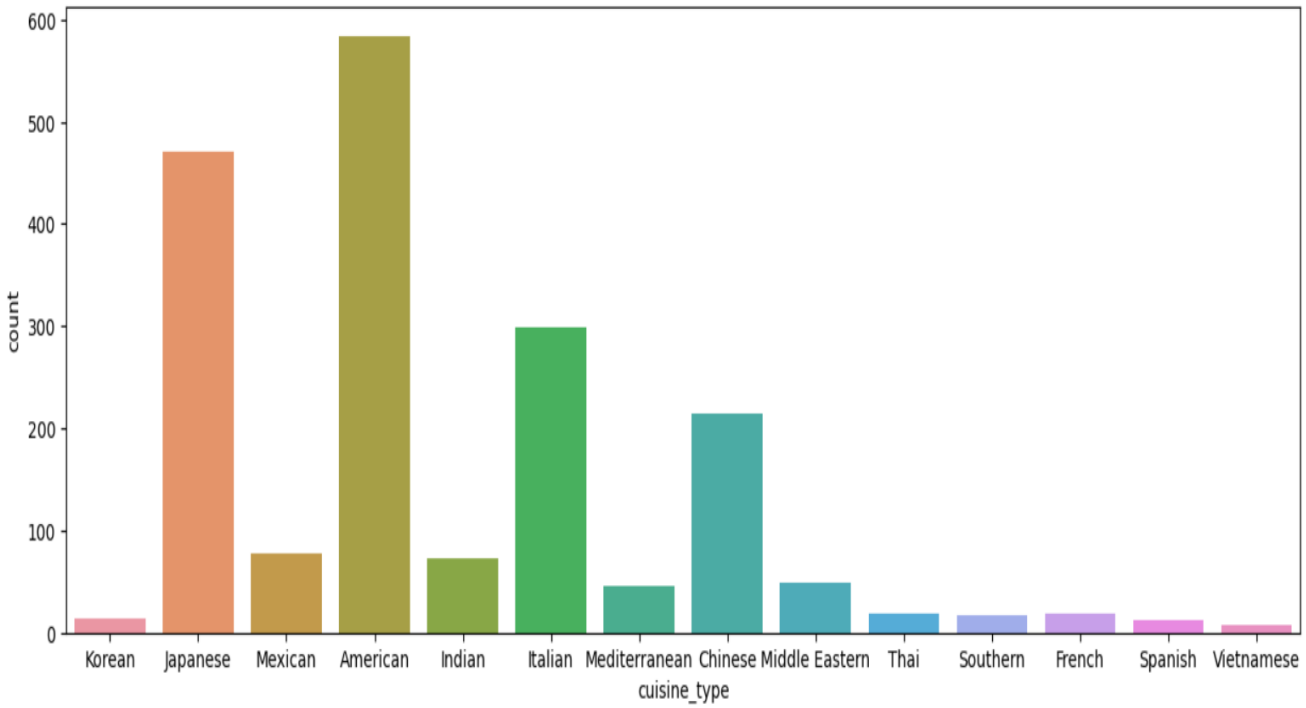
check unique Restaurant Name by value count function

```
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>
```

There are 178 unique restaurants The Shake Shack is the most popular restaurant followed by Meatball Shop and Blue Ribbon Sushi respectively

Check unique cuisine type by unique function for the variables and observations on distributions.

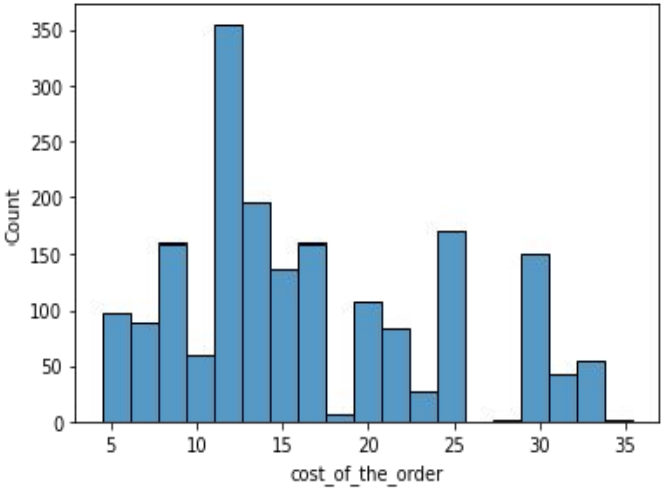
Out[106]: <AxesSubplot:xlabel='cuisine_type', ylabel='count'>

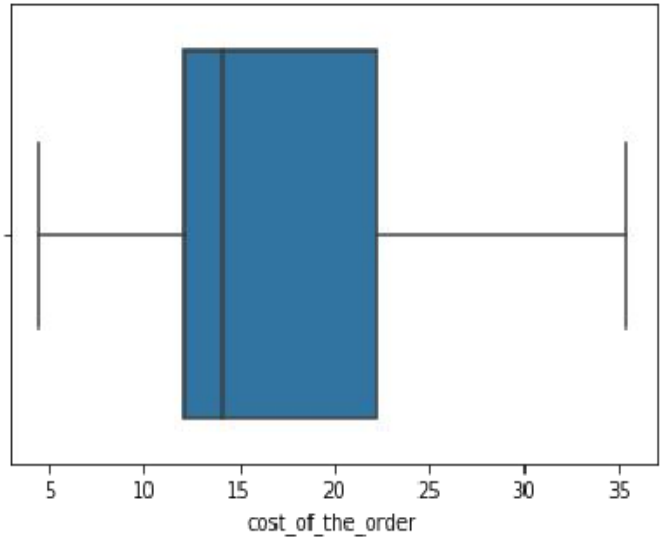


Observation:

- There are 14 unique cuisines in the dataset.
- American cuisine is most ordered followed by Japanese, Italian & Chinese
- 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.

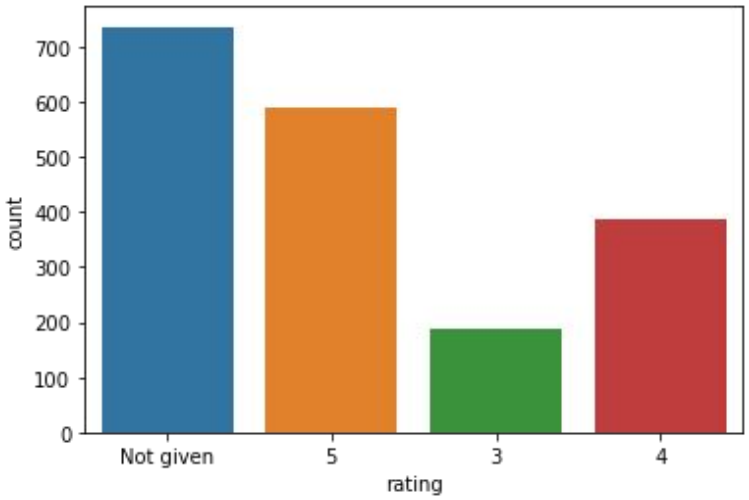
Plot Histogram & Boxplot for the cost of order





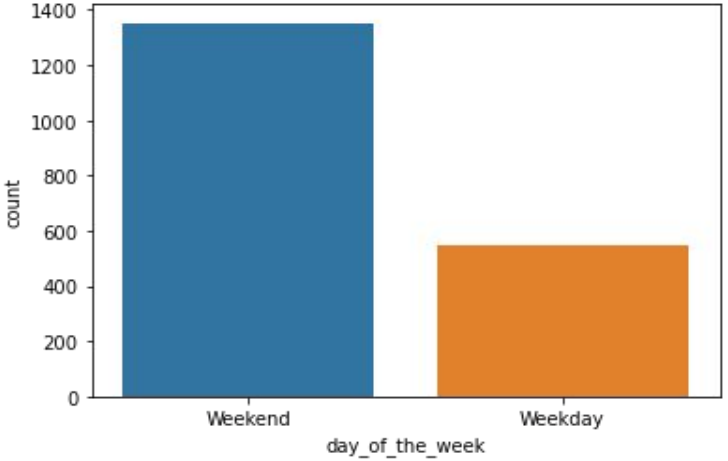
- The average cost of the order is greater than the median cost indicating that the distribution for the cost of the order is right-skewed.
- The mode of the distribution indicates that a large chunk of people prefer to order food that costs around 10-12 dollars.
- There are few orders that cost greater than 30 dollars. These orders might be for some expensive meals.

plot bar graph for 'rating' column



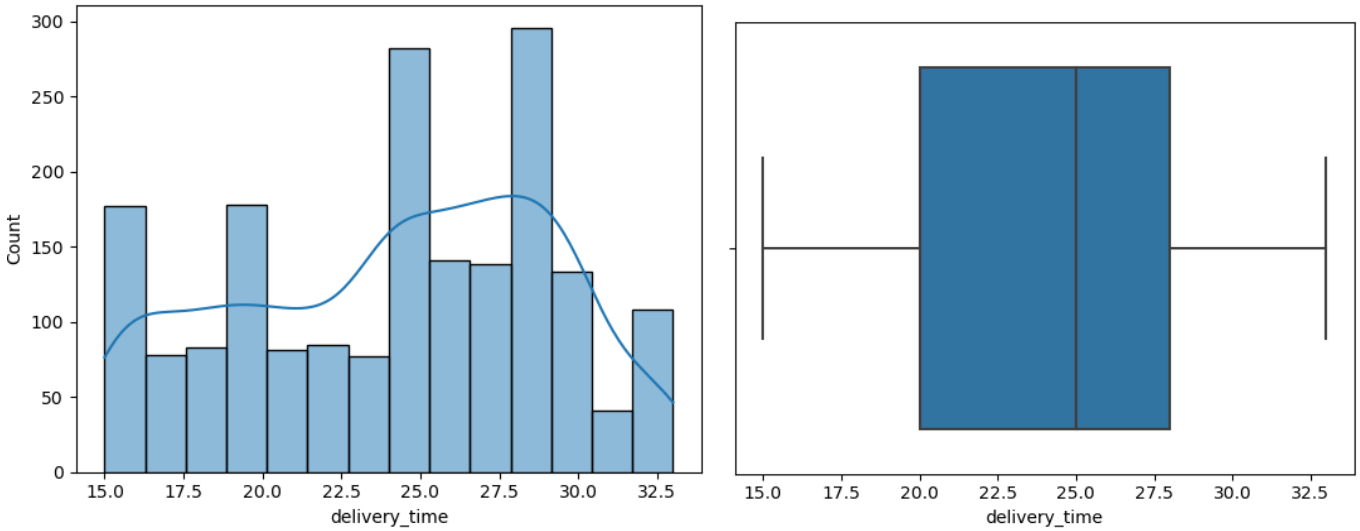
- The distribution of 'rating' shows that the most frequent rating category is 'not given', followed by a rating of 5.
- Only around 200 orders have been rated 3.

plot a bar graph for 'day_of_the_week' column



- The 'day_of_the_week' columns consists of 2 unique values - Weekday and Weekend
- The distribution shows that around 1300 orders are placed on weekends.

Plot the histogram for the delivery time Plot the boxplot for the delivery time



- The most occurrences of delivery time is bin 24-25 minutes.
- In the boxplot it isobserved that the data slightly skewed to the left.
- The delivery time between 24-31 minutes had an increased in counts to 120-160 times from approximately 90 counts.

The top 5 restaurants in terms of the number of orders received

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

Observations:

Shake Shack has by far the highest number of orders with 219 orders followed by The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Rice and Parm Would be interesting to know the cuisine served by these 5 restaurants.

The most popular cuisine on weekends

14

Observations:
Again this aligns with our previous observation that American, Japanese, Italian and Chines cuisine are the most popular The above 5 cuisines are particularly more popular on weekends

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 %
```

Observations:
Orders costing above 20 dollars make up about 29% of total orders.

The mean order delivery time

```
The mean delivery time for this dataset is delivery_time    24.16
dtype: float64 minutes
```

Observations: Food will typically be delivered to the customer within 24 minutes once ready

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

customer_id	order_id
52832	13
47440	10
83287	9

Observations: The 3 customer id's that had the most frequent orders are id 52832:13 orders, 47440:10 orders, and 83287: 9 orders. These customers would receive the 20% discount vouchers.

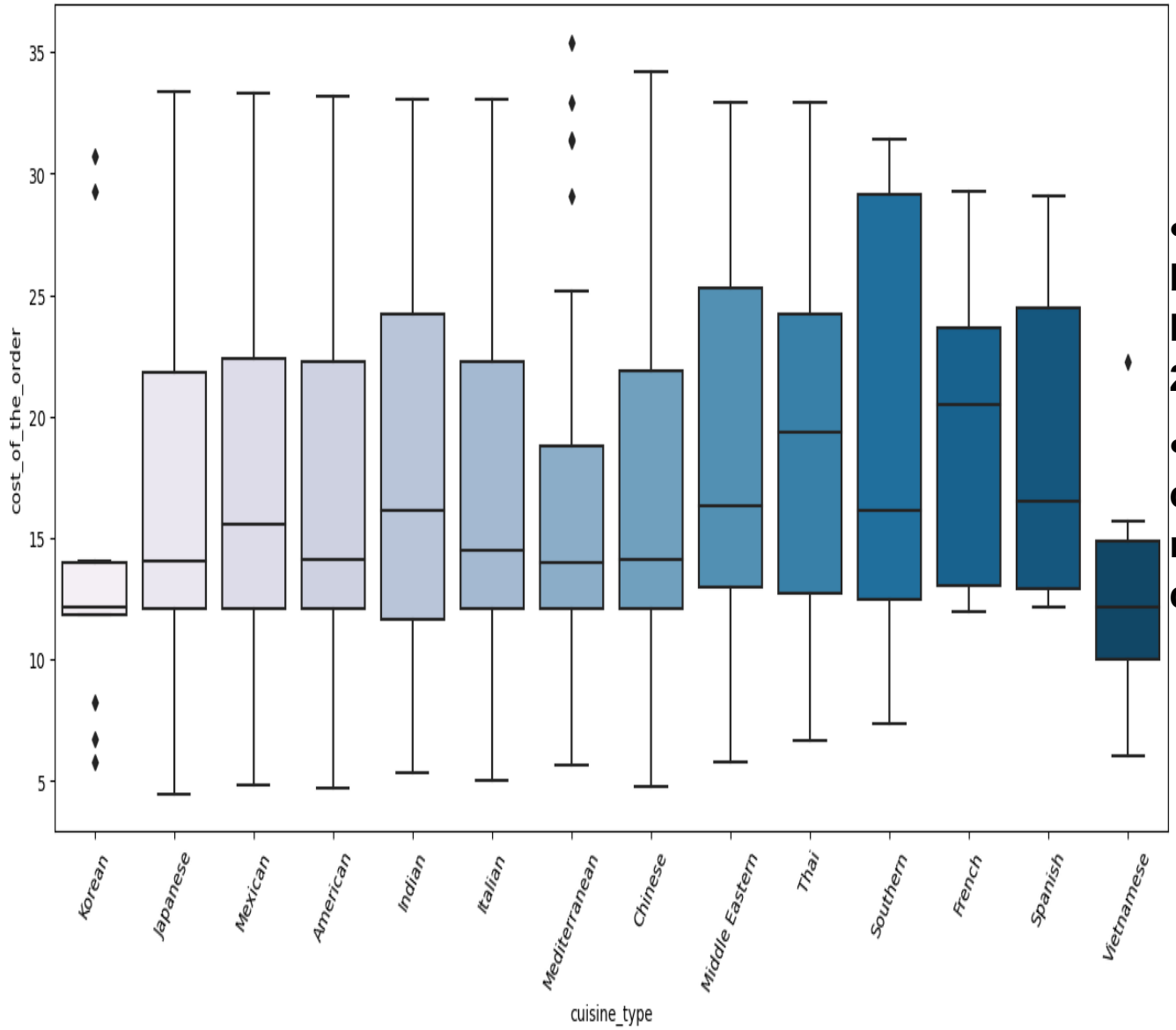
Univariate Analysis - Key Questions

- Delivery time ranges from 15 to 33 minutes, with an average of around 24 minutes and a standard deviation of 5 minutes. The spread is not too high for delivery time either
- There are 736 orders that are not rated
- Top 5 popular restaurants that have received the highest number of orders 'Shake Shack', 'The Meatball Shop', 'Blue Ribbon Sushi', 'Blue Ribbon Fried Chicken' and 'Parm'
- Almost 33% of the orders in the dataset are from these restaurants
- The most popular cuisine type on weekends is American
- There are a total of 555 orders that cost above 20 dollars. The percentage of such orders in the dataset is around 29.24%
- The mean delivery time is around 24.16 minutes
- Customer with ID 52832 has ordered 13 times. Other customers with Customer IDs 52832, 47440, 83287, 250494 and 259341 can avail discount offer of 20%

Exploratory Data Analysis (EDA): Multivariate Analysis

A multivariate analysis to explore relationships between the important variables in the dataset.

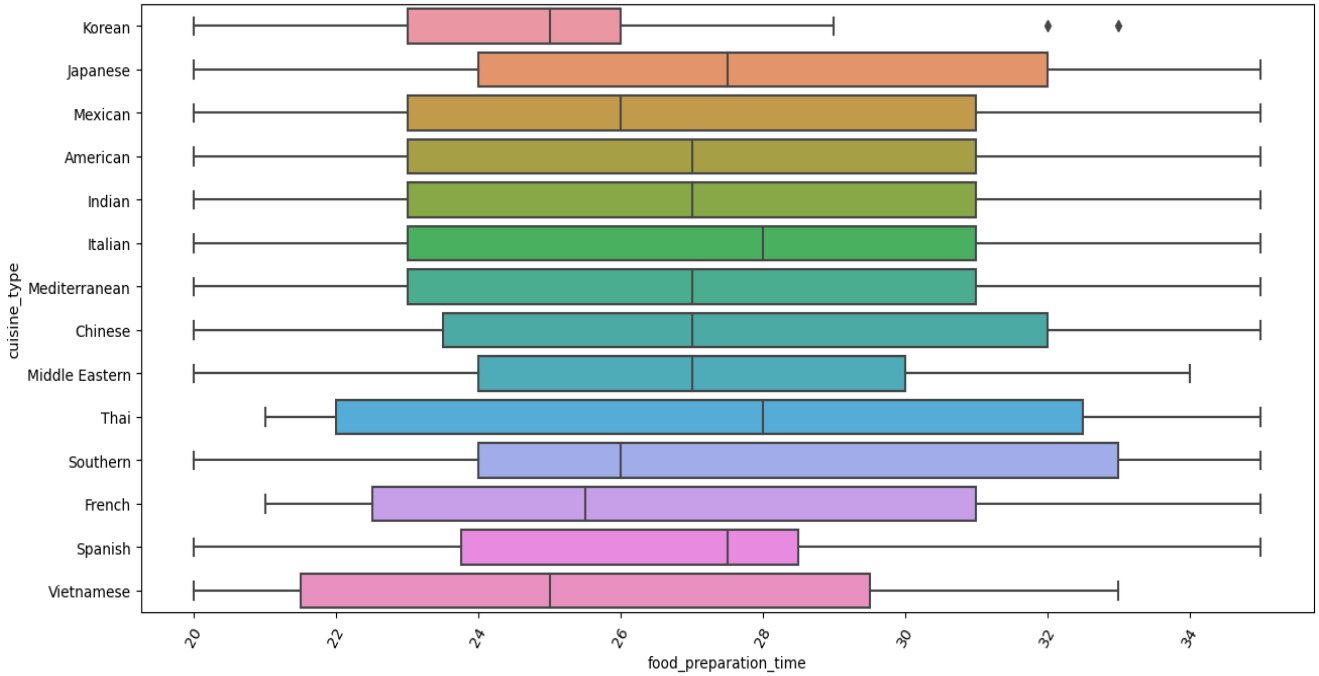
Relationship between cost of the order and cuisine type



- **We could see in these boxplots that most cuisines had a cost from about 10 to 25 dollars per order.**
- **Few outliers were observed but these could be normal since eating habits could differ.**

Relationship between food preparation time and cuisine type

To visualize the relationship between food preparation time and cuisine type using boxplot.

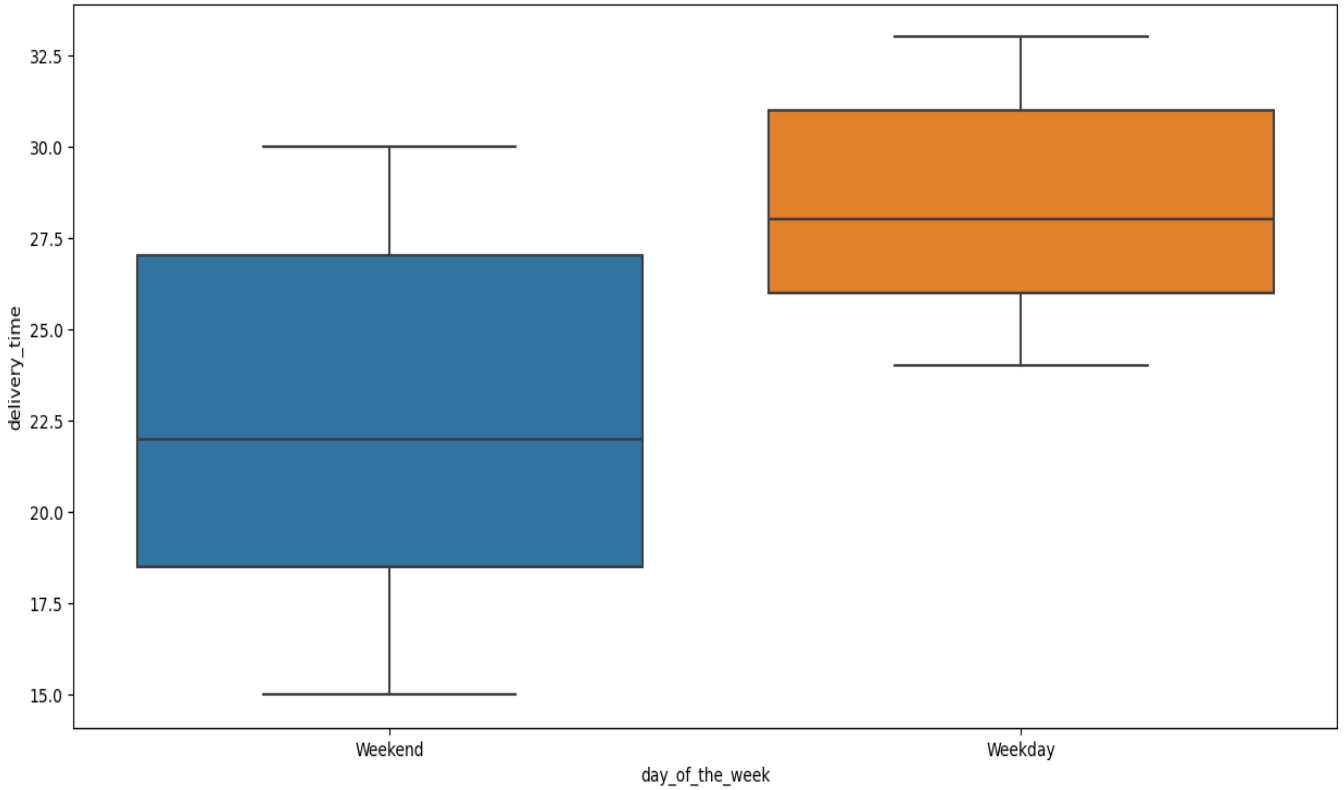


Observations:

Thai and Italian food take on average the longest time to prepare Korean and Vietnamese food take the least time to prepare There are outliers for Korean food so some specific dishes might take a long time on average to prepare American, Chinese, Indian and Middle Eastern food take on average similar time to prepare

Relationship between day of the week and delivery time

Visualize the relationship between day of the week and delivery time using boxplot



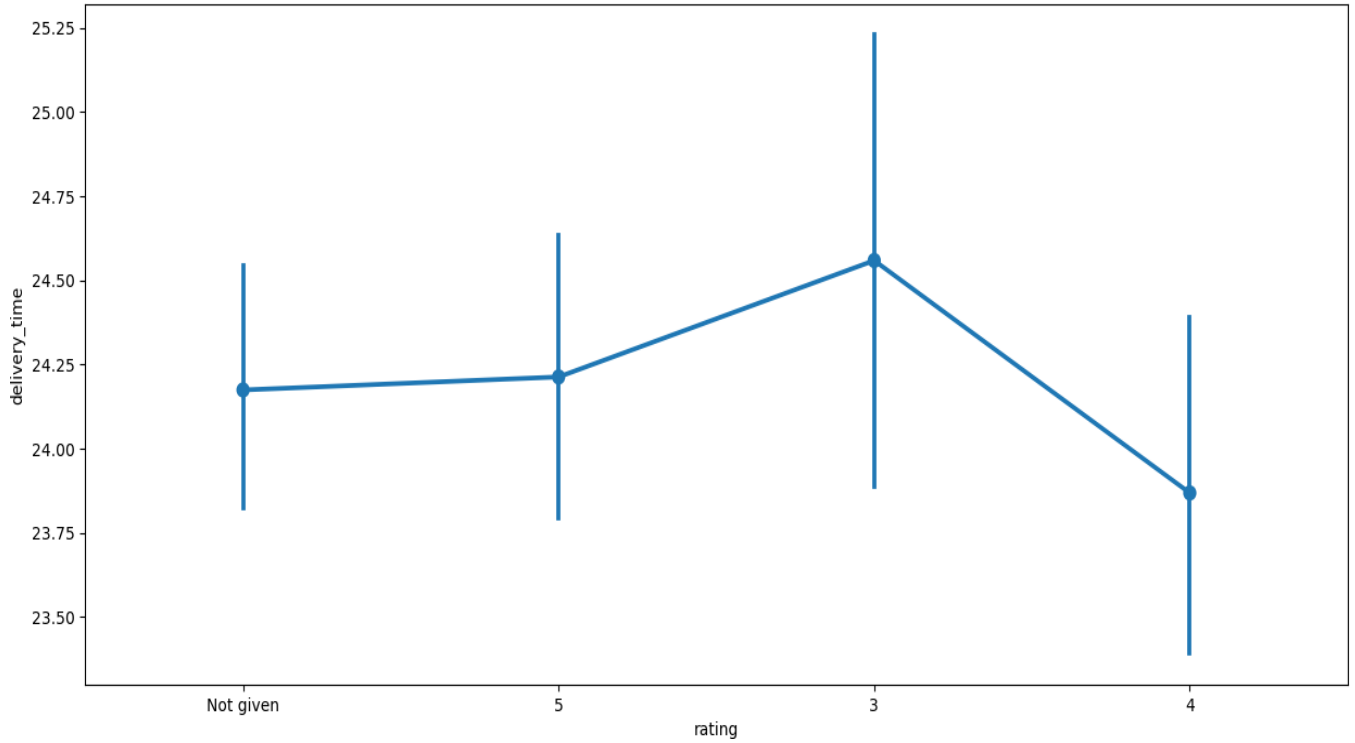
Observations:

The delivery time is generally higher on weekdays than on weekends It take 28 minutes on average to deliver food to a customer on weekdays and 22 minutes on weekends

Observations on the revenue generated by the restaurant

```
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
```

Relationship between rating and delivery time

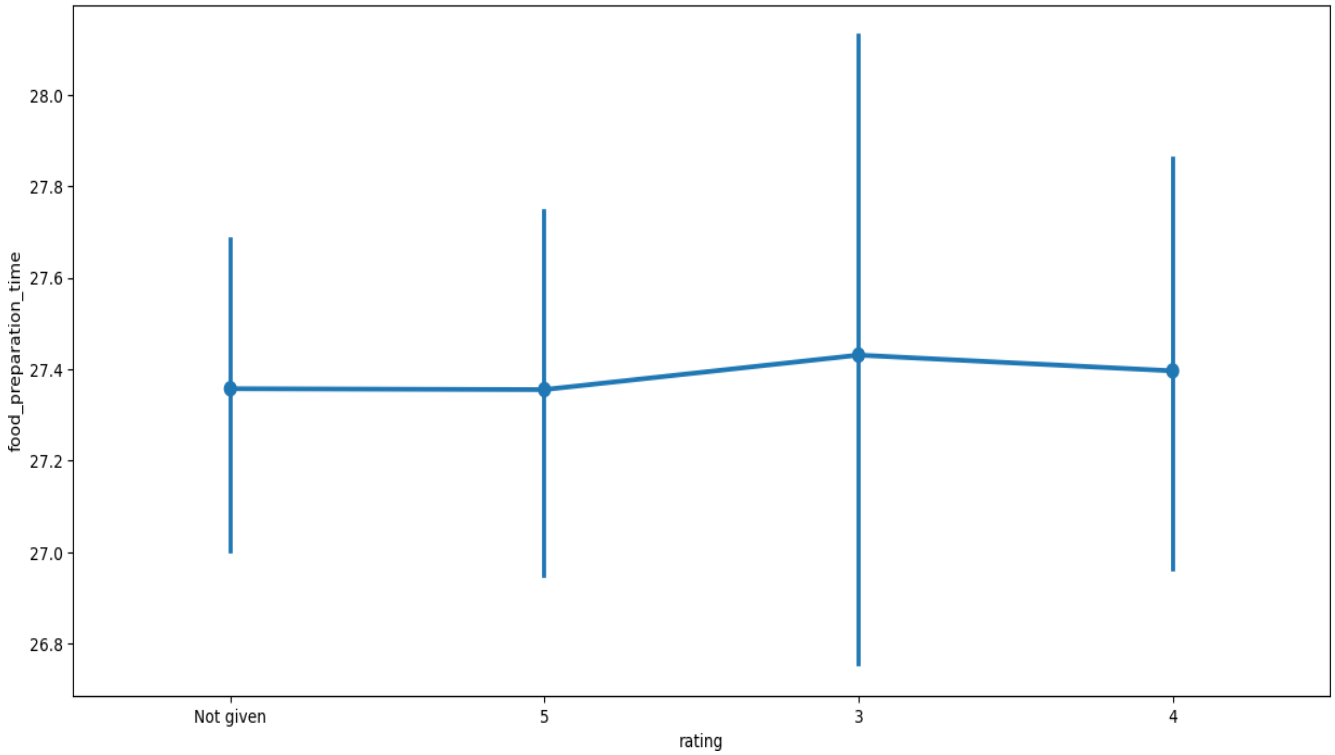


Observations:

The highest delivery times are linked to the lowest rating which is 3 The lowest food preparation times are linked to either no ratings at all or customers giving the highest rating.

Relationship between rating and food preparation time

relationship between rating and food preparation time using point plot

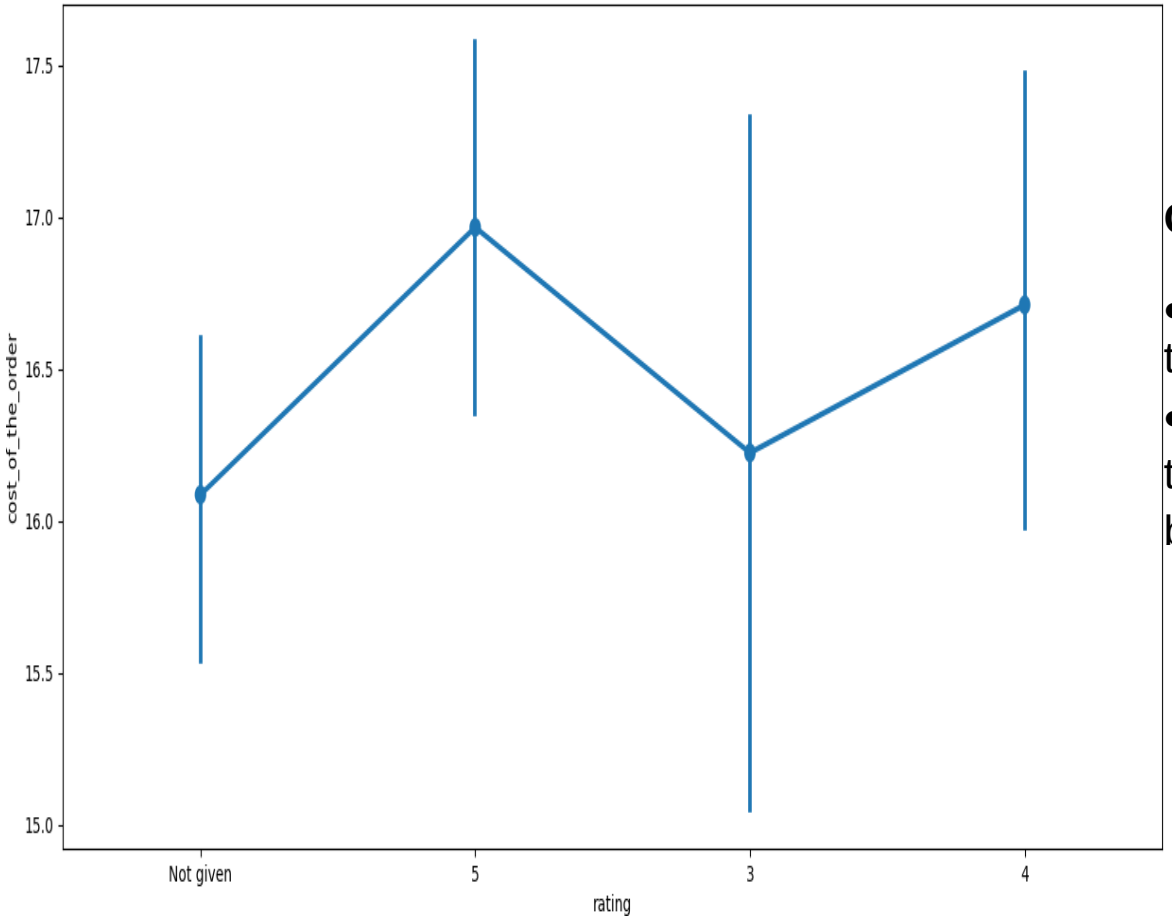


Observations:

The highest food preparation times are linked to the lowest rating which is 3 The lowest food preparation times are linked to the highest rating which is 5

Relationship between rating and cost of the order

visualize the relationship between rating and cost of the order using point plot



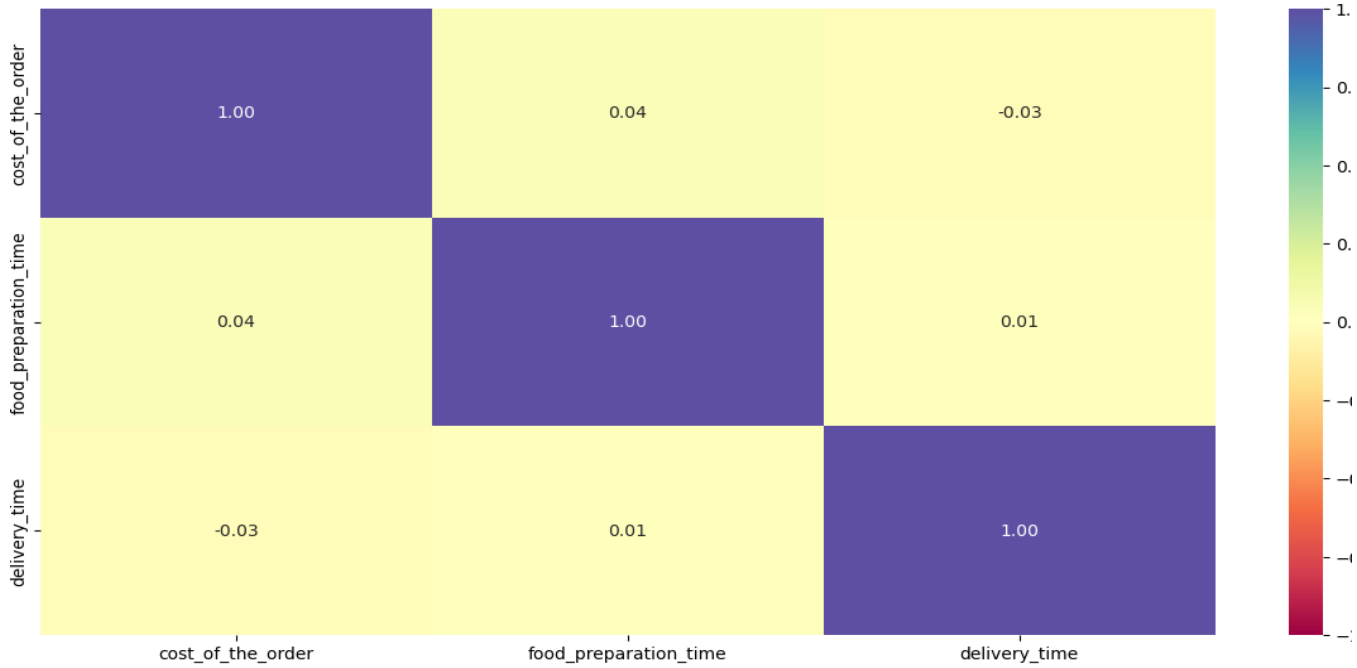
Observations:

- The highest cost ‘cost of order’ is linked to the highest rating which is 5
- The lowest cost ‘cost of order’ is linked to either a low rating of 3 or no rating being given by the customer at all.

Multivariate Analysis - Key Questions

- The restaurants fulfilling the criteria to get the promotional offer are: 'The Meatball Shop', 'Blue Ribbon Fried Chicken', 'Shake Shack' and 'Blue Ribbon Sushi'.
- The net revenue generated on all the orders given in the dataset is around 6166.3 dollars.
- Approximately 10.54 % of the total orders have more than 60 minutes of total delivery time.
- The mean delivery time on weekdays is around 28 minutes whereas the mean delivery time on weekends is around 22 minutes.
- This could be due to the dip of traffic volume in the weekends.

Correlation among variables Plot the heatmap



Observations:

Food preparation time has a correlation with cost of order

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. Check 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

Create a data frame that contains the restaurant names with their rating counts

Group the restaurant names with their ratings and find the mean rating of each restaurant

	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

Observations:

The Meatball Shop, Blue Ribbon Fried Chicken, Shake Shack and Blue Ribbon Sushil all qualify for this offer as they all have average ratings greater than 4 Of all the restaurants The Meatball Shop has the highest average customer rating

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.

Determine the revenue

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time	total_time	Rev
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20	45	7
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23	48	1
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28	51	1
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15	40	7
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24	49	1

Get the total revenue

```
The net revenue is around 6166.3 dollars
```

Observations:

The company will generate 6166.3 dollars

Percentage of orders take more than 60 minutes to get delivered from the time the order is placed

```
10.54
```

Observations:
Only 10.54 percent of orders have a total delivery time over 60 mins So over 89 percent of orders have a total delivery time under 60 minutes

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 The mean delivery time on weekends is
around 22 minutes
```

Observations:
The mean delivery time is 28 minutes on weekdays and 22 minutes on weekends

Appendix

Conclusions:

The Shake Shack is the most popular restaurant and has generated the highest revenue of all the 14 restaurants at 3579.53 probably due to high demand of their cuisine and good prices for their menu as well as a reasonable preparation and delivery time The delivery time is generally higher on weekdays than on weekends probably due to traffic that riders encounter On average an order takes 27 minutes to be prepared and 24 minutes to deliver totalling approximately an hour waiting time and looking at the ratings I can say that customers are generally happy to wait an hour for quality for Most orders are made on a weekend probably because most people are home and not willing to cook themselves after a long week Most orders do not have a rating. An individual order cost 16.5 dollars on average but can cost up to 35 dollars. Food preparation time has a correlation with cost of order as more expensive orders could take longer to prepare 89 percent of orders have a total delivery time under 60 minutes The Meatball Shop has the highest average customer rating American, Japanese, Italian and Chinese cuisine are the most popular The highest delivery times are linked to the lowest rating and vise versa The highest cost cost of order is linked to the highest rating and vise versa meaning that customers don't mind paying more for quality food.

Recommendations:

American, Japanese, Italian and Chinese cuisine are in high demand so might be worth bringing in more restaurants that make these cuisines There is a significantly higher number of customers willing to order food on weekends rather than weekdays so more drivers and chefs could be deployed to meet this demand and drive-up customer ratings.