

The cover page features a white background with abstract green geometric shapes on the left and right sides. The title 'Summer Boot Camp Project 2024' is centered in a green font, with the author's name 'Himanshu Singh (202247650)' below it in a smaller, grey font.

Summer Boot Camp Project 2024

Himanshu Singh (202247650)

INDEX

1. List of Tables
2. List of Figures
3. Problem Statement Objective
4. Importing The Necessary Libraries
5. Basic Exploration
6. Data Wrangling
7. Data Preprocessing
8. Feature Engineering
9. Data visualization
10. Order Analysis
11. Customer Behavior
12. Restaurant Performance
13. Demand Patterns
14. Operational Efficiency
15. Customer Insights

List Of Tables

1. Data Head
2. Data Tail
3. Data Types
4. Data Summary Statistics
5. NULL values
6. Wrong Entries(Delivery time)
7. the average food preparation time for each restaurant
8. the average delivery time compare across different restaurants
9. Correlation between the cost of the order and the rating given

List of Figures

1. Check Outliers
2. Check Outliers (After Handling Missing Values)
3. The Total Number Of orders
4. Countplot Between rating and the day of the week
5. coutplot of day of the week
6. Heatmap between the cost of order and the rating
7. countplot between the cuisine type and the day of the week
8. piechart of rating 4 and higher

Problem Statement 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.

Importing The Necessary Libraries

Reading The Dataset

Basic Exploration

Data Head(Top 5 rows)

	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	25.0	20
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25.0	?
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23.0	28
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25.0	15
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25.0	24

Data Tail(last 5 Rows)

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
1893	1476701	292602	Chipotle Mexican Grill \$1.99 Delivery	Mexican	22.31	Weekend	5	31.0	17
1894	1477421	397537	The Smile	American	12.18	Weekend	5	31.0	19
1895	1477819	35309	Blue Ribbon Sushi	Japanese	25.22	Weekday	Not given	31.0	24
1896	1477513	64151	Jack's Wife Freda	Mediterranean	12.18	Weekday	5	23.0	31
1897	1478056	120353	Blue Ribbon Sushi	Japanese	19.45	Weekend	Not given	28.0	24

Observation

In Delivery_time '?' is there need to check in future

In Rating , In some of the data rating is "Not given"

Data Wrangling

Shape of Data

Number of Rows are 1898 and Number of columns are 9

Data Types of Each Feature

```
order_id          int64
customer_id       int64
restaurant_name    object
cuisine_type       object
cost_of_the_order  float64
day_of_the_week    object
rating            object
food_preparation_time float64
delivery_time      object
dtype: object
```

Observation

delivery time is appearing as "object" but it should be in "int"

rating is appearing as "object" but it should be in "int"

Statistical Summary

	order_id	customer_id	cost_of_the_order	food_preparation_time
count	1.898000e+03	1898.000000	1898.000000	1896.000000
mean	1.477496e+06	171168.478398	80.722007	27.371835
std	5.480497e+02	113698.139743	2798.141333	4.634211
min	1.476547e+06	1311.000000	0.000000	20.000000
25%	1.477021e+06	77787.750000	12.080000	23.000000
50%	1.477496e+06	128600.000000	14.160000	27.000000
75%	1.477970e+06	270525.000000	22.310000	31.000000
max	1.478444e+06	405334.000000	121920.000000	35.000000

Observation

cost of the order is 0.00 at minimum

The Null Values

Check For Missing Values in Columns

```
order_id          0
customer_id       0
restaurant_name    0
cuisine_type      3
cost_of_the_order  0
day_of_the_week   0
rating            0
food_preparation_time 2
delivery_time     0
dtype: int64
```

Observation

3 NULL value are in Cuisine type and 2 NULL values are in Food delivery Time

To fetch all the rows having atleast one NULL value

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
11	1478437	221206	Empanada Mama (closed)	NaN	8.10	Weekend	5	23.0	22
51	1477883	91817	Blue Ribbon Fried Chicken	NaN	29.39	Weekend	Not given	27.0	28
95	1477027	164016	Blue Ribbon Fried Chicken	NaN	16.39	Weekend	Not given	27.0	22
140	1477376	370372	Blue Ribbon Fried Chicken	American	11.59	Weekday	Not given	NaN	24
188	1477872	300670	Shake Shack	American	13.39	Weekend	Not given	NaN	22

Check for percentage wise missing values in columns

The Duplicate values

Check for the duplicate values

As there are zero duplicate values

The Anomalies or Wrong Entries

From above Observation we found wrong entry in delivery time

Check for all unique values in delivery time

Shows the columns which have entry '?' in delivery Time

Replacing all '?' with the NULL

Now there is not any wrong entry in the "Delivery time"

From The Above Observattion "Not given" is mention in the "Rating"

Check for all unique values

```
array(['Not given', '5', '3', '4'], dtype=object)
```

Shows the columns which have entry 'Not given' in rating

	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	25.0	20
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25.0	NaN
6	1477894	157711	The Meatball Shop	Italian	6.07	Weekend	Not given	28.0	21
10	1477895	143926	Big Wong Restaurant Big Wong Restaurant	Chinese	5.92	Weekday	Not given	34.0	28
14	1478198	62667	Lucky's Famous Burgers	American	12.13	Weekday	Not given	23.0	30
...
1887	1476873	237616	Shake Shack	American	5.82	Weekend	Not given	26.0	30
1891	1476981	138586	Shake Shack	American	5.82	Weekend	Not given	22.0	28
1892	1477473	97838	Han Dynasty	Chinese	29.15	Weekend	Not given	29.0	21
1895	1477819	35309	Blue Ribbon Sushi	Japanese	25.22	Weekday	Not given	31.0	24
1897	1478056	120353	Blue Ribbon Sushi	Japanese	19.45	Weekend	Not given	28.0	24

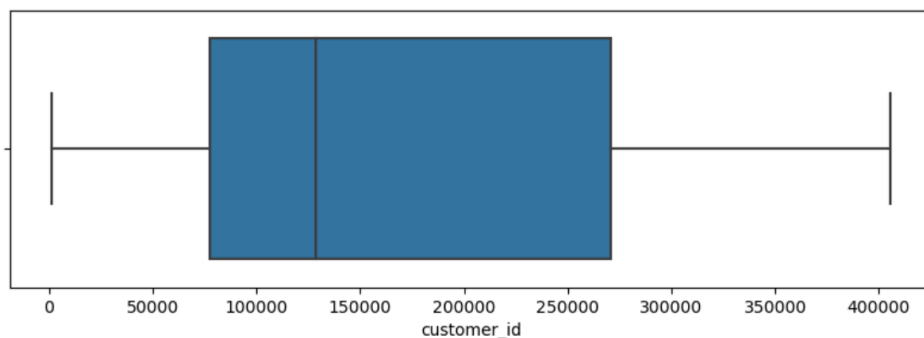
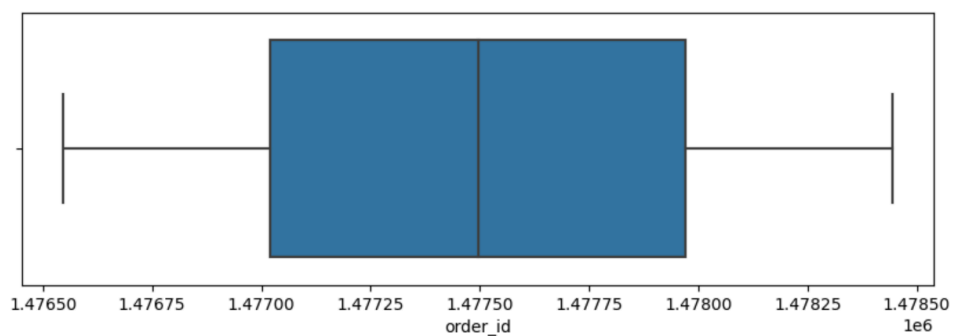
736 rows × 9 columns

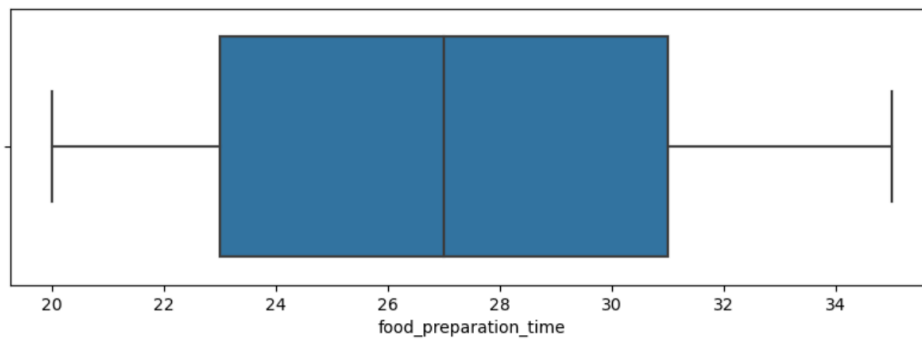
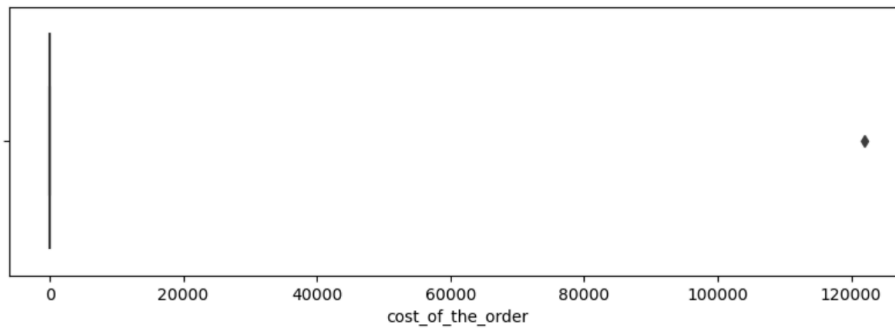
Replacing all 'Not given' with the NULL

Now there is not any wrong entry in the rating

Outliers

First Check for Outliers





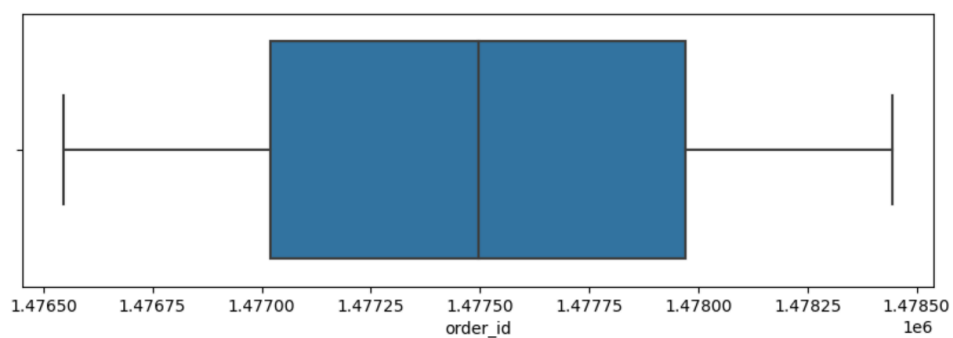
Outliers in 'order_id', 'customer_id', 'cost_of_the_order', 'food_preparation_time'

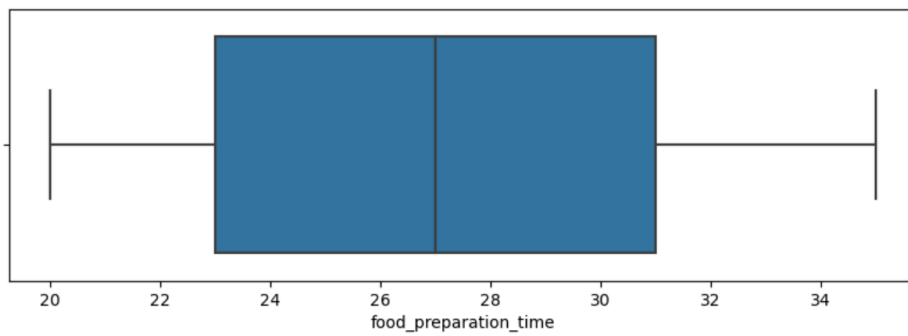
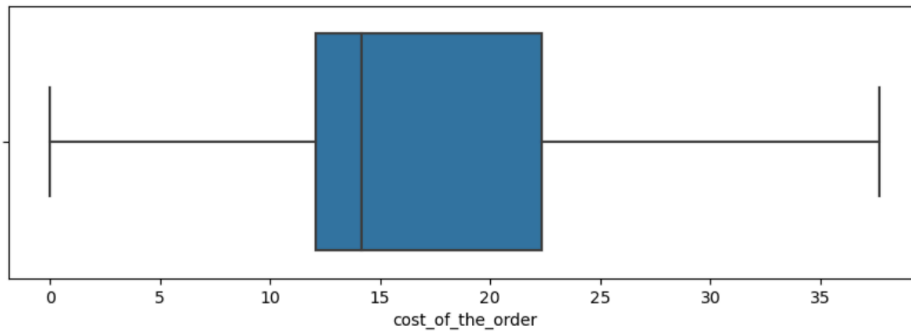
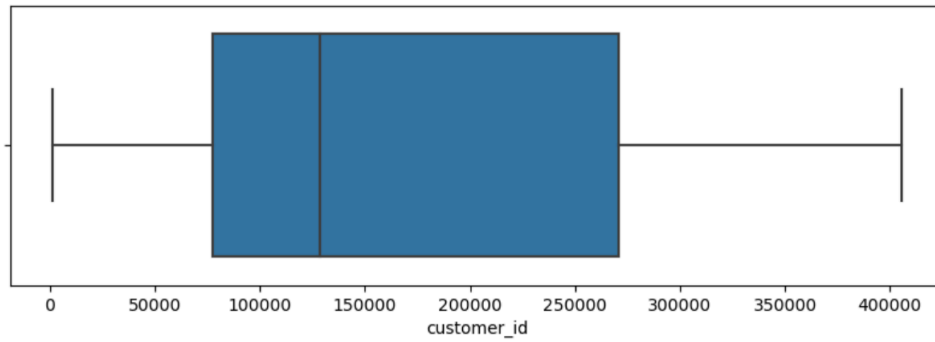
Removing outliers

Since, outliers is present in Cost of the order so REPLACING NULL values in Numerical Columns using Median

REPLACING NULL values in categorical columns using Mode

Boxplots after Outliers treatment

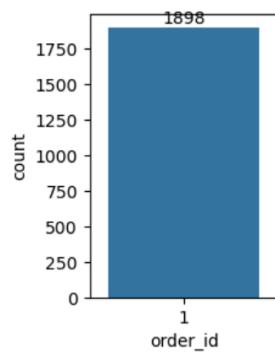




Check for missing values in any column

1. Order Analysis

a) What is the total number of orders in the dataset ?



=>The Total number of orders are 1898

b) What is the average cost of an order?

=>The Average Cost of Order is 16.50

c) How many unique customers have placed orders?

=>The total Number of Unique customers have placed orders are 1200

d) Which restaurant has received the highest number of orders

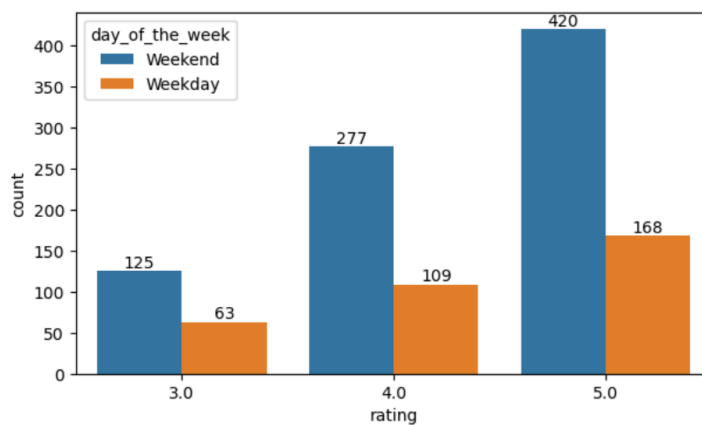
=>Restaurant Shake Shack has received the highest number of orders : 219

2. Customer Behavior

a) What is the average rating given by customers?

=> The Average Rating given by Customers is 4.3

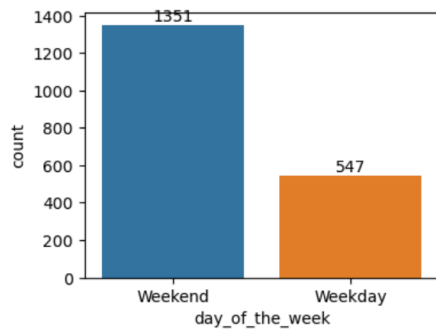
b) How does the rating vary between weekdays and weekends?



c) Which cuisine type is ordered the most?

=> American cuisine type has ordered the most 585 times.

d) What is the distribution of orders across different days of the week?



=> The distribution of orders across different days of the week (Weekend and Weekday)

3. Restaurant Performance

a) What is the average food preparation time for each restaurant?

	restaurant_name	food_preparation_time
39	Cipriani Le Specialita	35.0
140	Sushi Choshi	35.0
83	Kambi Ramen House	35.0
85	Klong	35.0
148	Taro Sushi	35.0
...
96	Market Table	21.0
43	Despalza	20.5
3	67 Burger	20.0
58	Frank Restaurant	20.0
68	Haru Gramercy Park	20.0

178 rows × 2 columns

=> The average food preparation time for each restaurant

b) Which restaurant has the shortest average food preparation time?

	restaurant_name	food_preparation_time
68	Haru Gramercy Park	20.0
3	67 Burger	20.0
58	Frank Restaurant	20.0
43	Despalza	20.5
132	Sarabeth's West	21.0
93	Lucky Strike	21.0
138	Song Thai Restaurant & Bar	21.0
65	Hampton Chutney Co.	21.0
128	Samurai Mama	21.0
96	Market Table	21.0

c) How does the average delivery time compare across different restaurants?

	restaurant_name	delivery_time
132	Sarabeth's West	33.0
148	Taro Sushi	32.0
68	Haru Gramercy Park	32.0
58	Frank Restaurant	31.0
64	Haandi	30.5
...
61	Galli Restaurant	16.0
152	The MasalaWala	15.0
110	Paul & Jimmy's	15.0
71	Hibino	15.0
60	Gaia Italian Cafe	15.0

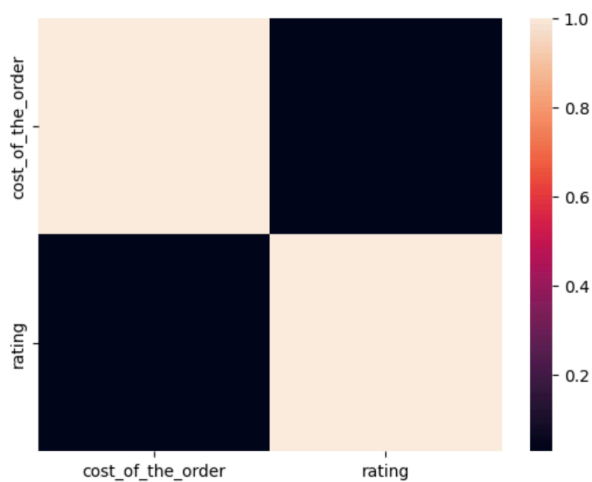
178 rows × 2 columns

d) Is there a correlation between the cost of the order and the rating given?

	cost_of_the_order	rating
cost_of_the_order	1.000000	0.028631
rating	0.028631	1.000000

=> there is least correlation between the cost of order and the rating

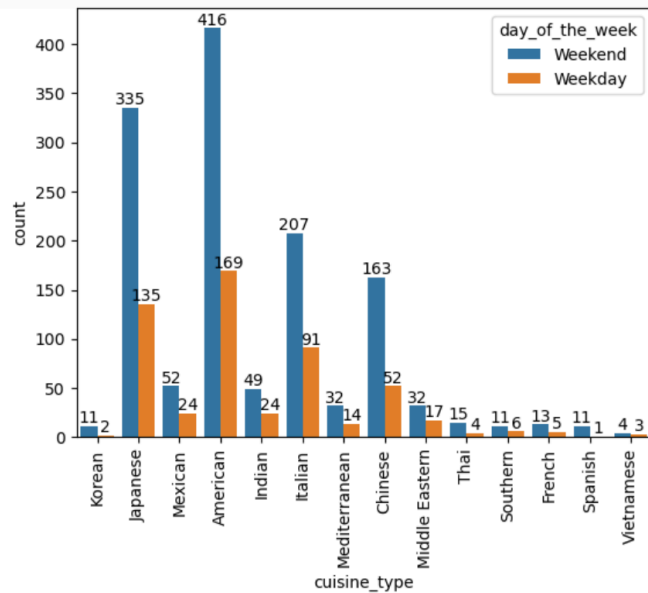
HeatMap between the cost of order and the rating



4. Demand Patterns

a) How does the demand for different cuisine types vary on weekdays versus weekends?

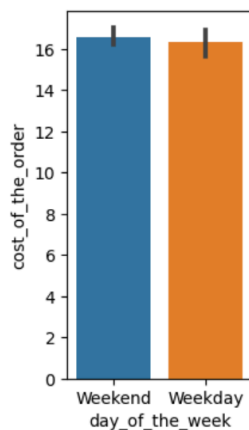
=>countplot between the cuisine type and the day of the week



b) Which day of the week has the highest average order cost?

Average cost of the order in day of the week

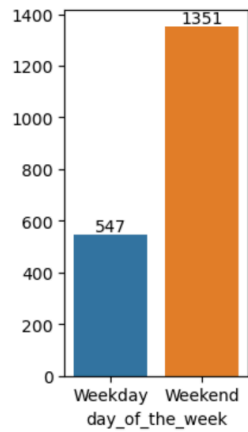
The barplot between the day of the week and the cost of the order



c) What is the most common day for orders to be placed?

=> The most common day for orders to be placed is Weekend

The Barplot between day of the week and the order Id



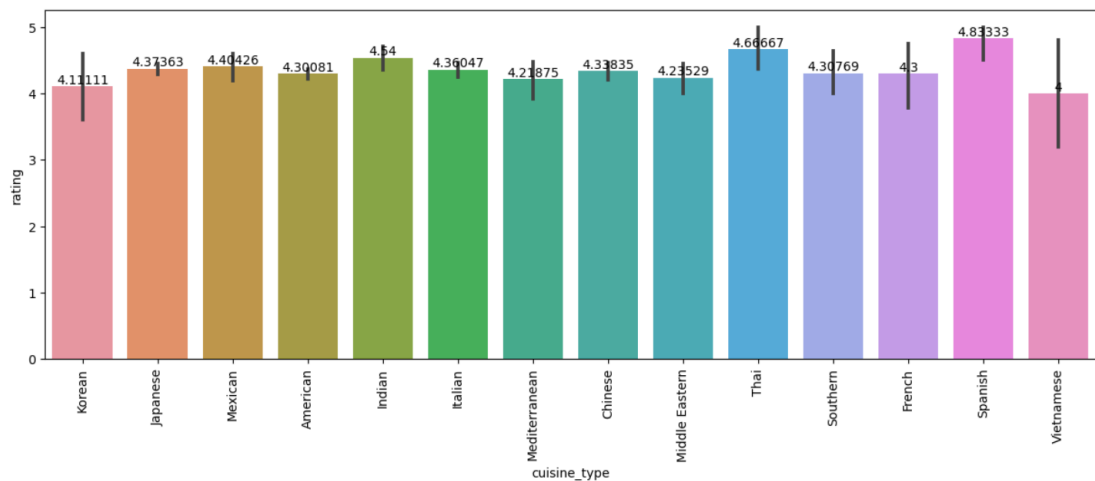
d) How does the average rating vary by cuisine type?

```

cuisine_type
American    4.300813
Chinese     4.338346
French      4.300000
Indian      4.540000
Italian     4.360465
Japanese    4.373626
Korean      4.111111
Mediterranean 4.218750
Mexican     4.404255
Middle Eastern 4.235294
Southern    4.307692
Spanish     4.833333
Thai        4.666667
Vietnamese  4.000000
Name: rating, dtype: float64

```

=> Barplot between the cuisine Type and teh rating



5. Operational Efficiency

a) What is the average delivery time for all orders?

=> The Average Delivery time for all the orders is 24.16 min

b) Which restaurant has the longest average delivery time?

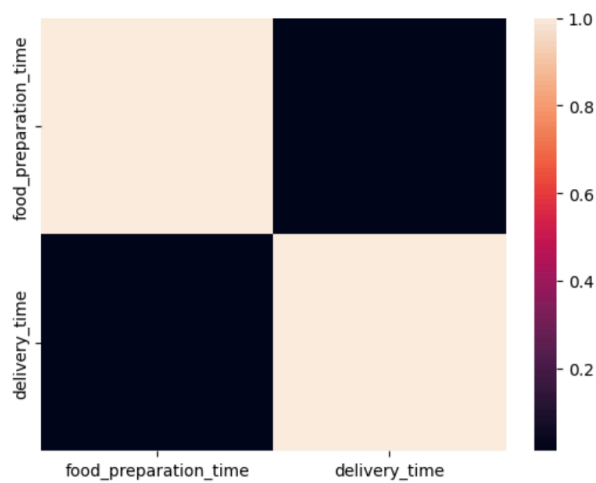
=> The Sarabeth restaurant has the longest average Delivery time of 33 min

c) Is there a relationship between food preparation time and delivery time?

	food_preparation_time	delivery_time
food_preparation_time	1.000000	0.011234
delivery_time	0.011234	1.000000

=> The Correlation between the food preparation time and delivery time os minimal

=> Heatmap between the food preparation time and delivery time

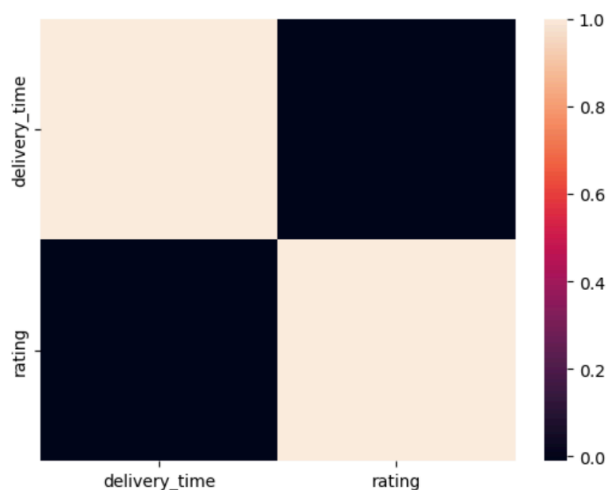


d) How does the delivery time impact customer ratings?

	delivery_time	rating
delivery_time	1.000000	-0.010121
rating	-0.010121	1.000000

=> The Delivery and Rating has the minal correlation in between

Heatmap in between the The delivery and rating



6. Customer Insights

a) What is the repeat order rate (number of customers who have placed more than one order)?

```
52832      12
47440       9
83287       8
250494      7
259341      6
65009       6
276192      6
82041       6
60052       5
97991       5
Name: customer_id, dtype: int64
```

b) What percentage of orders receive a rating of 4 or higher?

=> The percentage of orders receive the rating of 4 or higher is 51.31%

Piechart for the rating of 4 and higher

Percentage of Orders with Rating 4 or Higher

