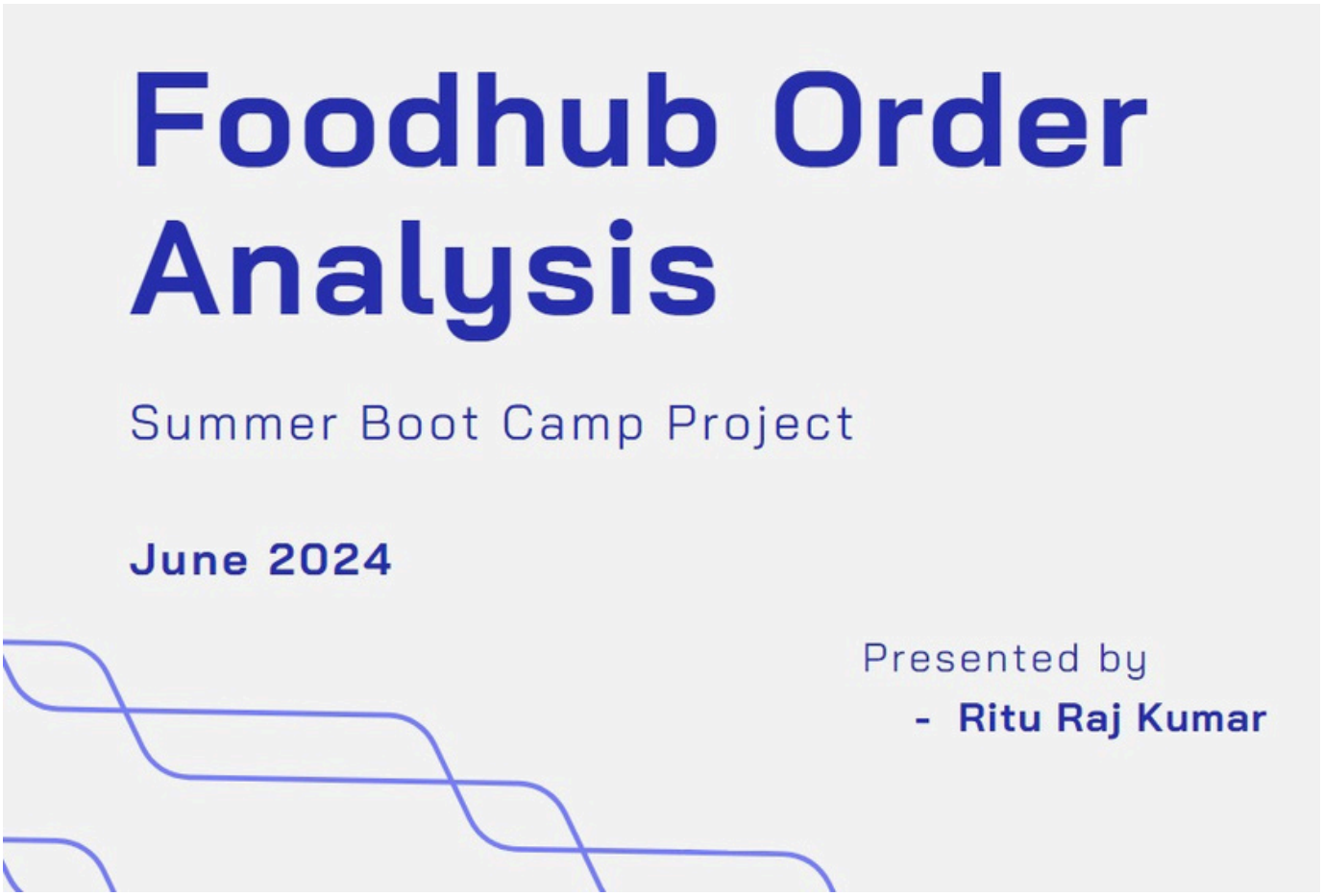


Sl No.	Content	Page
1	Cover Page	1
2	Index	1
3	Problem Statement	2
4	Data Infomation	3
5	Data Cleaning	6
6	Order Analysis	7
7	Restaurant Performance	10
8	Customer Insights	18
9	Conclusion	20

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 nd answers to these questions that will help the company to improve the business.

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 con rmation and the delivery person's pick-up con rmation.

• 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 con rmation and drop-off information

Basic Things

```
# Load the dataset
df = pd.read_csv('2-foodhub_order_New.csv')
```

1. Display the First 5 Rows

	0	1	2	3	4
order_id	1477147	1477685	1477070	1477334	1478249
customer_id	337525	358141	66393	106968	76942
restaurant_name	Hangawi	Blue Ribbon Sushi Izakaya	Cafe Habana	Blue Ribbon Fried Chicken	Dirty Bird to Go
cuisine_type	Korean	Japanese	Mexican	American	American
cost_of_the_order	30.75	12.08	12.23	29.2	11.59
day_of_the_week	Weekend	Weekend	Weekday	Weekend	Weekday
rating	Not given	Not given	5	3	4
food_preparation_time	25.0	25.0	23.0	25.0	25.0
delivery_time	20	?	28	15	24

Observations:

Rating - Not Given dilvery time has error values (?).

2. Display the Last 5 Rows

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

Observations:

Resturant name has $

3. Check the Shape of the Dataset:

Observations:

Shape of the data set is Rows: 1898 & Col: 9

4. Check the Data Types of Each Feature:

Observations:

Here delivery_time is an object type that we need to change

5. Check the Statistical summary

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
count	1.898000e+03	1898.000000	1898	1895	1898.000000	1898	1898	1896.000000	1898
unique	NaN	NaN	178	14	NaN	2	4	NaN	20
top	NaN	NaN	Shake Shack	American	NaN	Weekend	Not given	NaN	24
freq	NaN	NaN	219	582	NaN	1351	736	NaN	161
mean	1.477496e+06	171168.478398	NaN	NaN	80.722007	NaN	NaN	27.371835	NaN
std	5.480497e+02	113698.139743	NaN	NaN	2798.141333	NaN	NaN	4.634211	NaN
min	1.476547e+06	1311.000000	NaN	NaN	0.000000	NaN	NaN	20.000000	NaN
25%	1.477021e+06	77787.750000	NaN	NaN	12.080000	NaN	NaN	23.000000	NaN
50%	1.477496e+06	128600.000000	NaN	NaN	14.160000	NaN	NaN	27.000000	NaN
75%	1.477970e+06	270525.000000	NaN	NaN	22.310000	NaN	NaN	31.000000	NaN
max	1.478444e+06	405334.000000	NaN	NaN	121920.000000	NaN	NaN	35.000000	NaN

Observations:

Issuee in delivery_time, cost_of_the_order - Holds a outlier

```
<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           1895 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  1896 non-null   float64
 8   delivery_time          1898 non-null   object 
dtypes: float64(2), int64(2), object(5)
memory usage: 133.6+ KB
```

Observations:

Rating - Object

- delivery - (it should be int)

- time - object (it should be int)

6. Check the null values

```
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
```

Observations:

There are 3 Null values in cuisine_type

There are 2 Null values in food_preparation_time

7. Check the duplicate values

```
0
```

Observations:

In This data set there is no null values

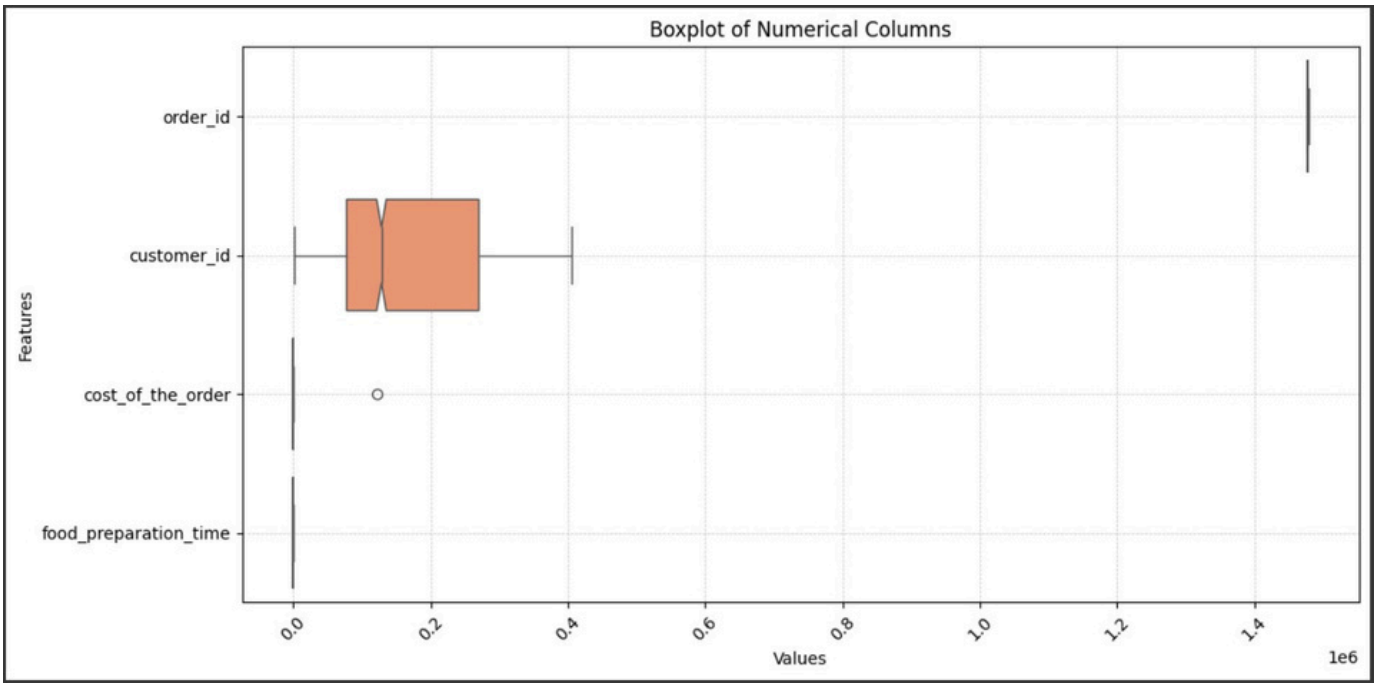
8. Check the anomalies or wrong entries.

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

Observations:

In columnn of "rating" there are value - 5,4,3 which are considerable but also some exceptions like "Not given" values

9. Check the outliers and their authenticity.




```
Outliers detected using IQR method:
order_id                0
customer_id             0
cost_of_the_order       1
food_preparation_time   0
dtype: int64
```

Observations:

In "cost_of_the_order", "customer_id " and "order_id" outliers present that we need correct

10. Data Cleaning

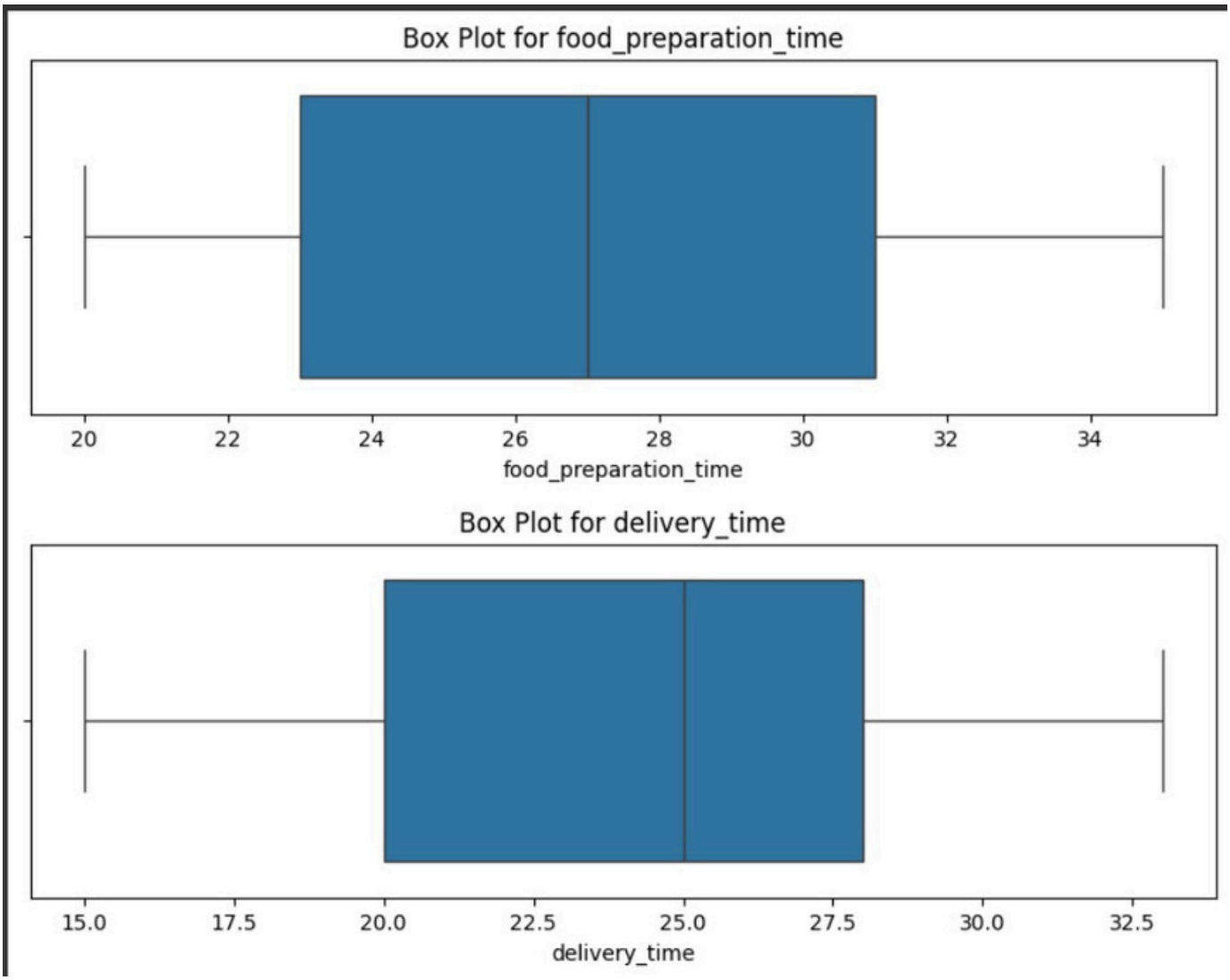
Checking for duplicates

order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time

Checking for Invalid values

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25.0	?
180	1476808	84700	Pepe Giallo	Italian	14.60	Weekday	3	32.0	?

MISSING VAlUES



Values - Corrected

Observations:

Here as we can see all the data is managed outliers were remvoed and whole data were managed

Main

1. Order Analysis

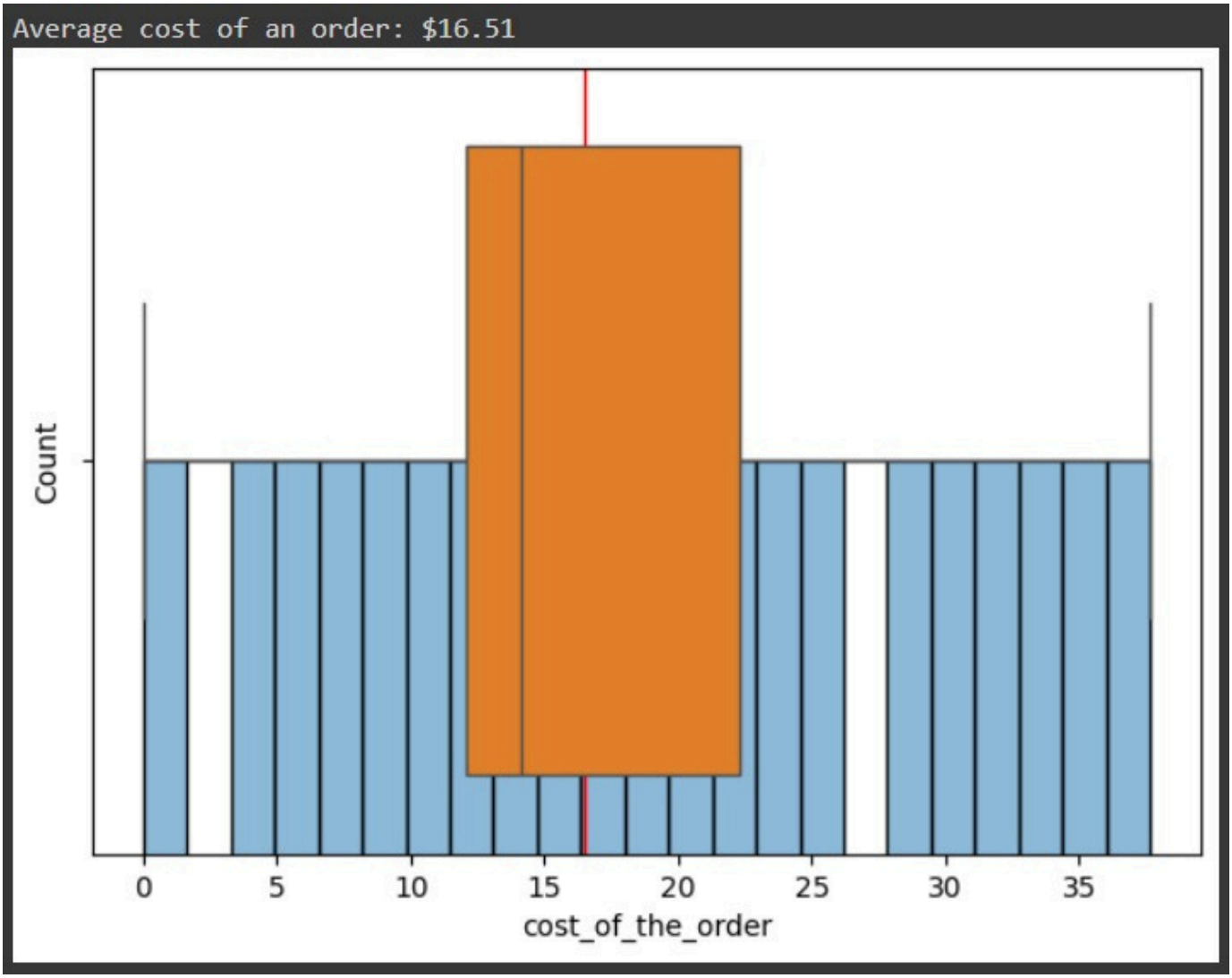
- Total number of orders in the dataset

```
Total number of orders: 1898
(1898, 9)
```

Observations:

Total number of orders placed: 1898

- The average cost of an order



Observations:

The Average FoodHub order cost is **$16.51** and the *majority* of the order cost ranges from **11 - 13 USD**.

- Number Of unique customers have placed orders

```
Number of unique customers: 1200
array([13, 10,  9,  8,  7,  6,  5,  4,  3,  2,  1])
```

Observations:

- No Of unique customers is **1200**
- Among these customers there is an order count range of **1 - 13**.

• Restaurant with highest number of orders

```
Restaurant with the highest number of orders:
restaurant_name
Shake Shack    219
Name: order_id, dtype: int64
```

Observations:

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

2. Customer Behavior

• The average rating given by customers?

```
4.34
```

Observations:

Average rating given by customers is 4.34

• How does the rating vary between weekdays and weekends

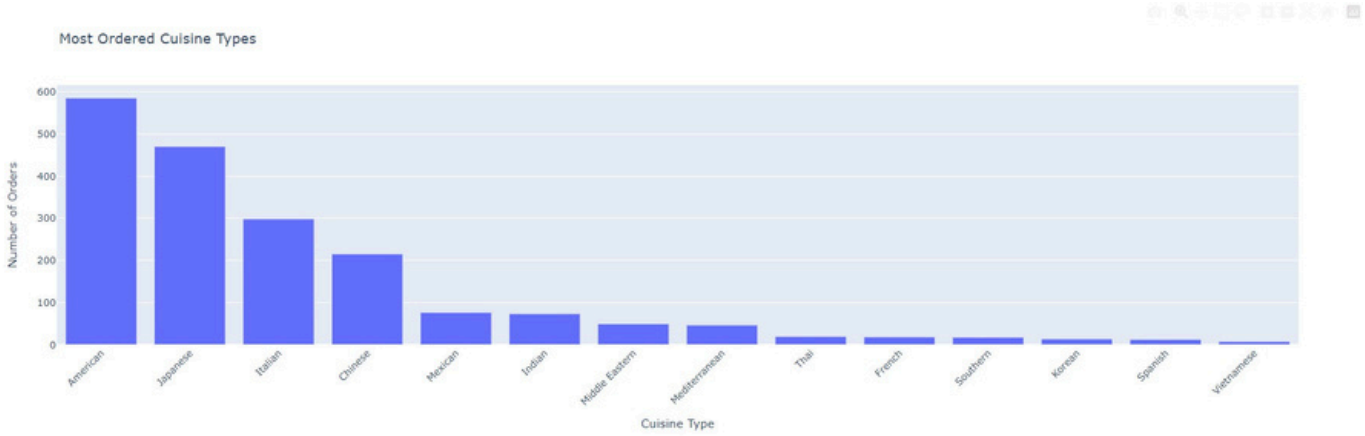
```
weekday rating: 4.31
weekend rating: 4.36
```

Observations:

Average **weekday** rating by Customer: 4.31

Average **weekend** rating by Customer: 4.36

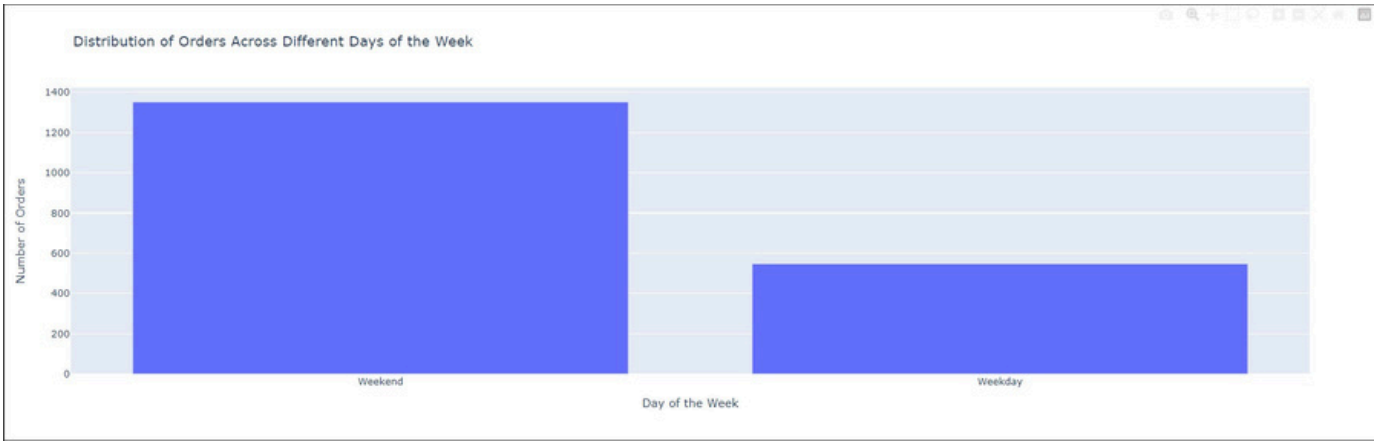
- Cuisine type is ordered the most?



Observations:

American is the most ordered cuisine type :585

- The distribution of orders across different days of the week



Observations:

Number Order Placed On **Weekend** is ***Higher*** than **Weekday** : 1351 > 547

3. Restaurant Performance

• The average food preparation time for each restaurant?

```
restaurant_name
Haru Gramercy Park        20.0
67 Burger                 20.0
Frank Restaurant          20.0
DespaÌ±a                  20.5
Sarabeth's West           21.0
                          ...
Taro Sushi                35.0
Cipriani Le Specialita    35.0
Kambi Ramen House         35.0
Klong                     35.0
Sushi Choshi              35.0
```

Observations:

Restaurants show varying average food preparation times, impacting customer wait times and operational e ciency.

Haru Gramercy Park: 20.0 minutes

67 Burger: 20.0 minutes

Frank Restaurant: 20.0 minutes

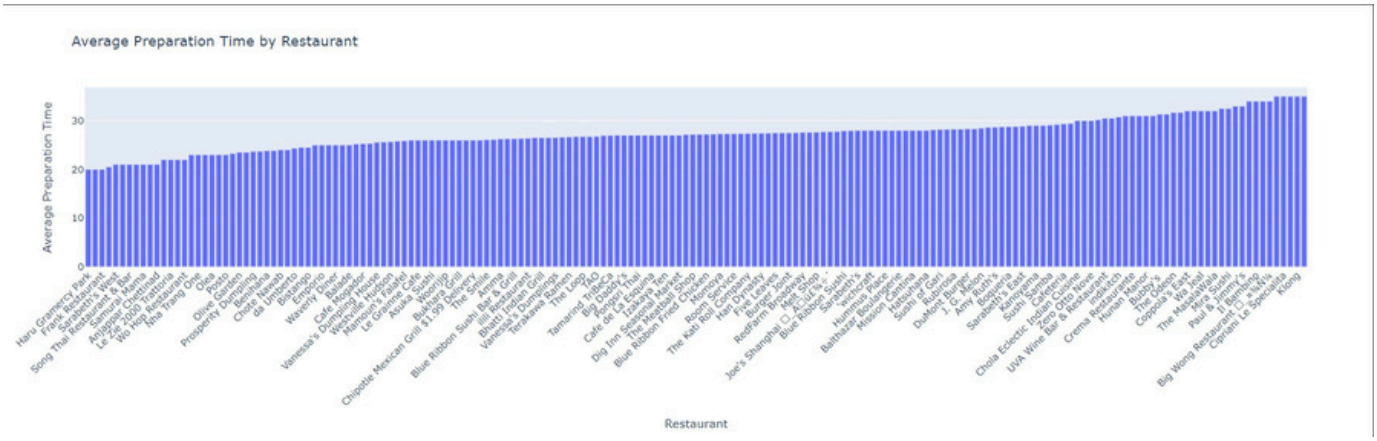
Despaña: 20.5 minutes

Sarabeth’s West: 21.0 minutes

Taro Sushi: 35.0 minutes

Cipriani Le Specialita: 35.0 minutes

• Resturant has the shortest average food preparation time

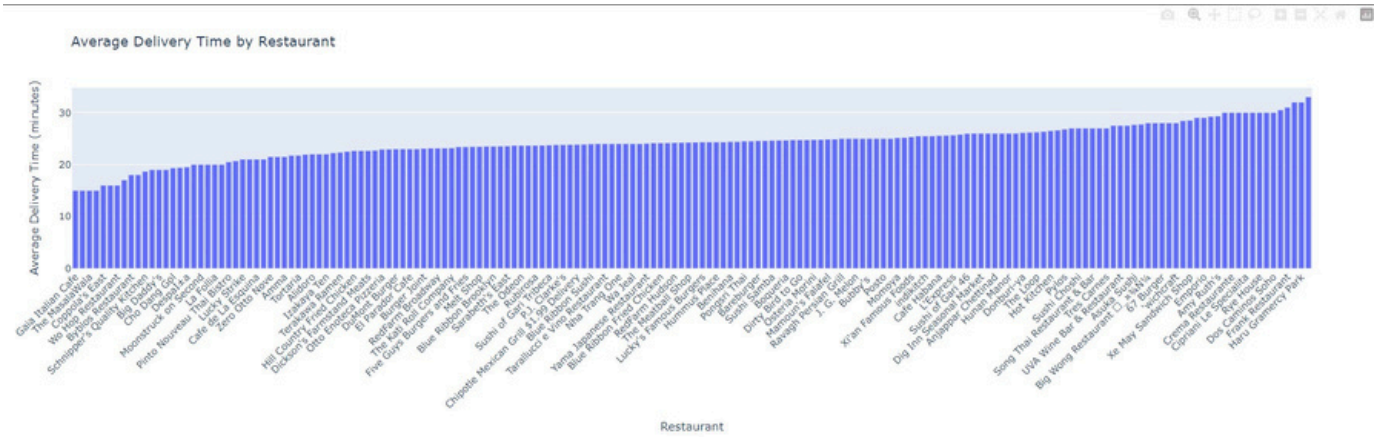


Observations:

Restaurants with the Shortest Average Preparation Time:

- Haru Gramercy Park: 20.00 minutes
- 67 Burger: 20.00 minutes
- Frank Restaurant: 20.00 minutes
- Despaİ±a: 20.50 minutes
- Sarabeth's West: 21.00 minutes

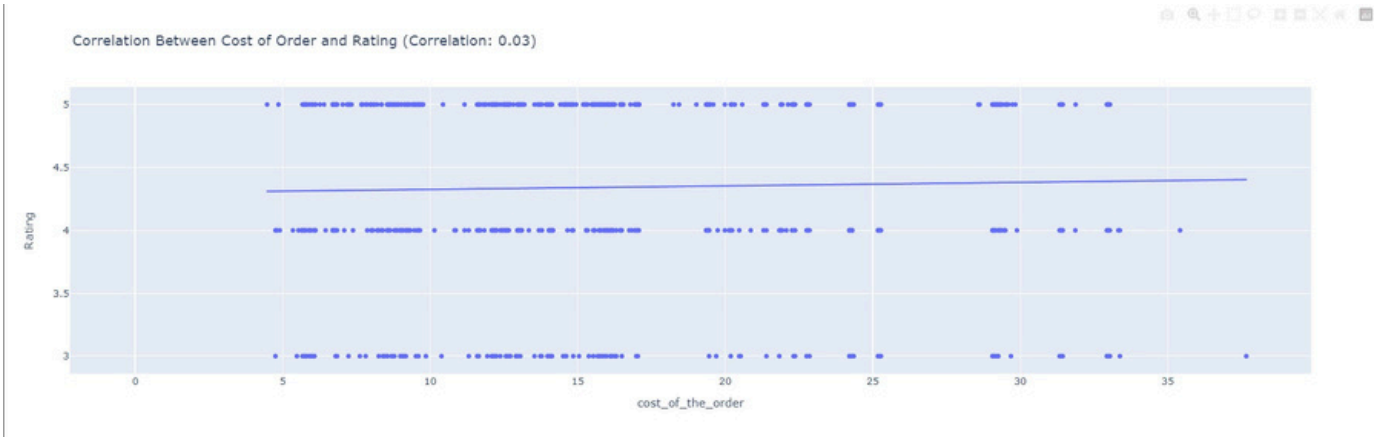
• The average delivery time compare across different restaurants?



Observations:

Restaurants have varied average delivery times, with some delivering in 15 minutes and others taking over 30 minutes.

• Checking is there a correlation between the cost of the order and the rating given

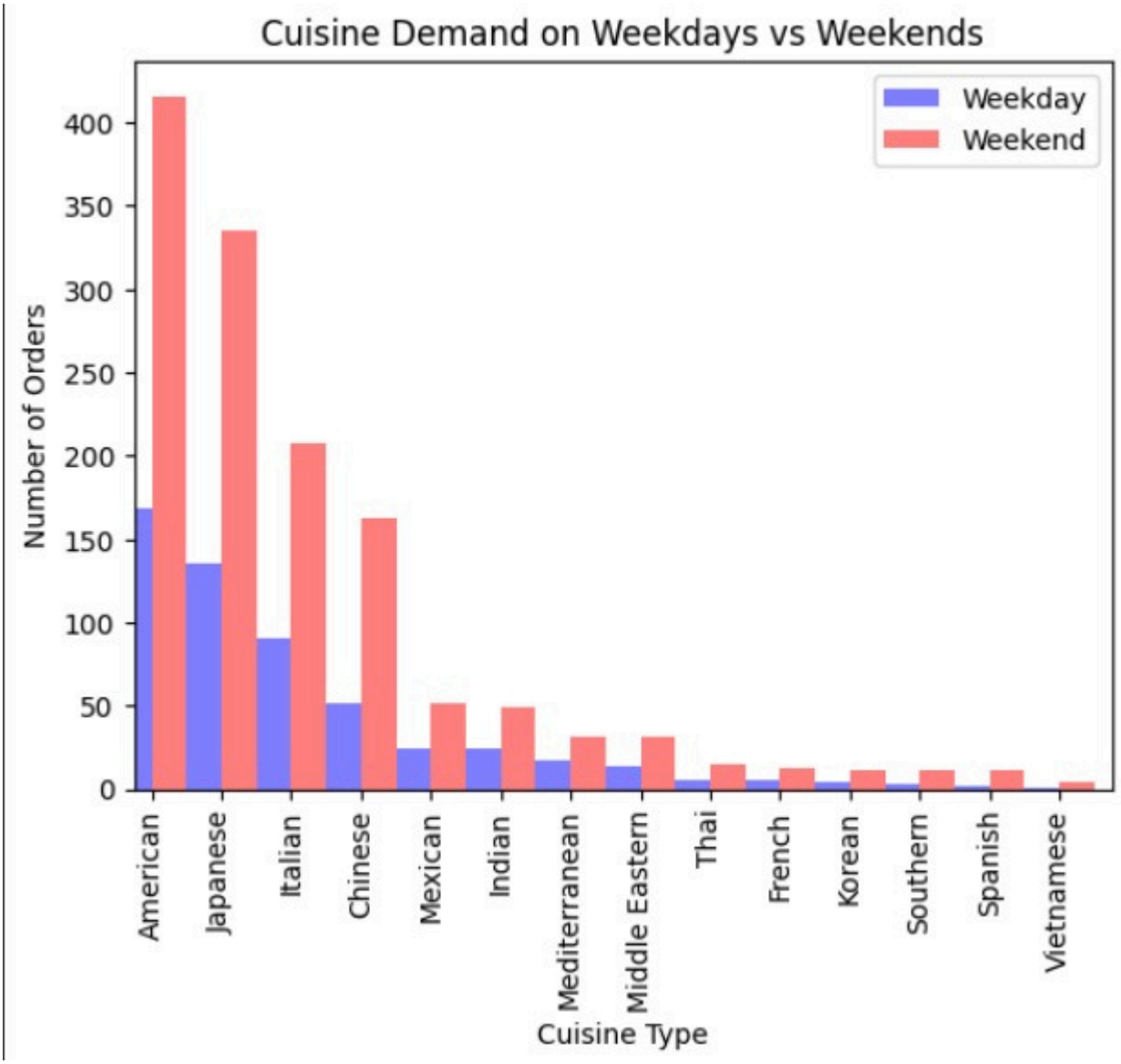


Observations:

The scatter plot indicates a very weak positive correlation (0.03) between the cost of the order and the rating given, suggesting minimal relationship.

4. Demand Patterns

• Howing the demand for different cuisine types vary on weekdays versus weekends?



Observations:

The bar chart reveals that demand for all cuisine types is generally higher on weekends compared to weekdays, with American and Japanese cuisines being the most popular.

• Which day of the week has the highest average order cost?

```
Weekend ($16.58)
```

Observations:

Day with the highest average order cost: ***Weekend*** **$16.58**

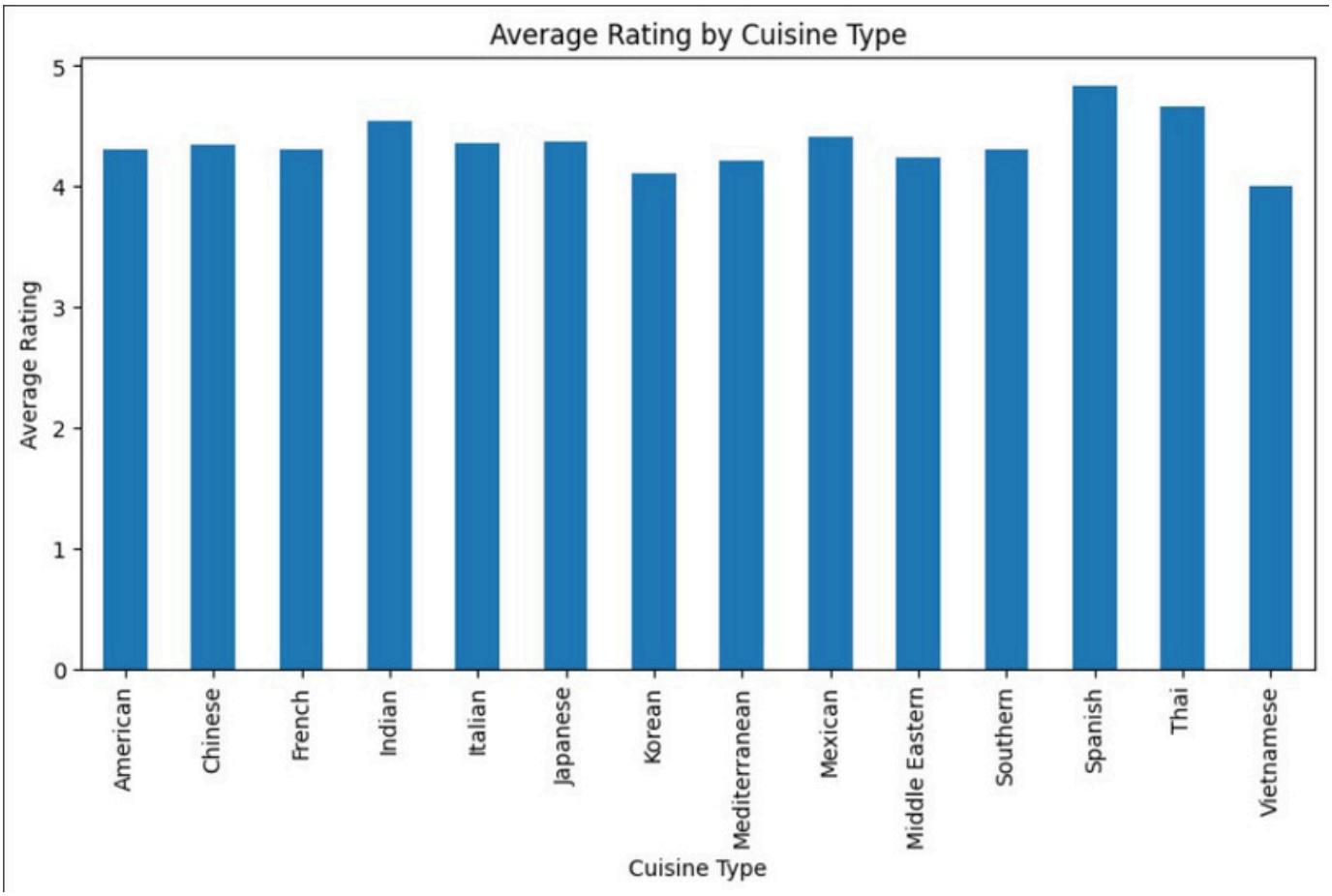
• What is the most common day for orders to be placed?

```
Weekend
```

Observations:

Most common day for orders: **Weekend**

• The average rating vary by cuisine type?



Observations:

Average ratings are generally high across all cuisine types, with Spanish and Indian cuisines receiving the highest average ratings, while Vietnamese cuisine has the lowest.

5. Operational E ciency

- Average Delivery Time for All Orders

```
24.16 minutes
```

Observations:

Average delivery time for all orders: **24.16** minutes

- Restaurant with Longest Average Delivery Time

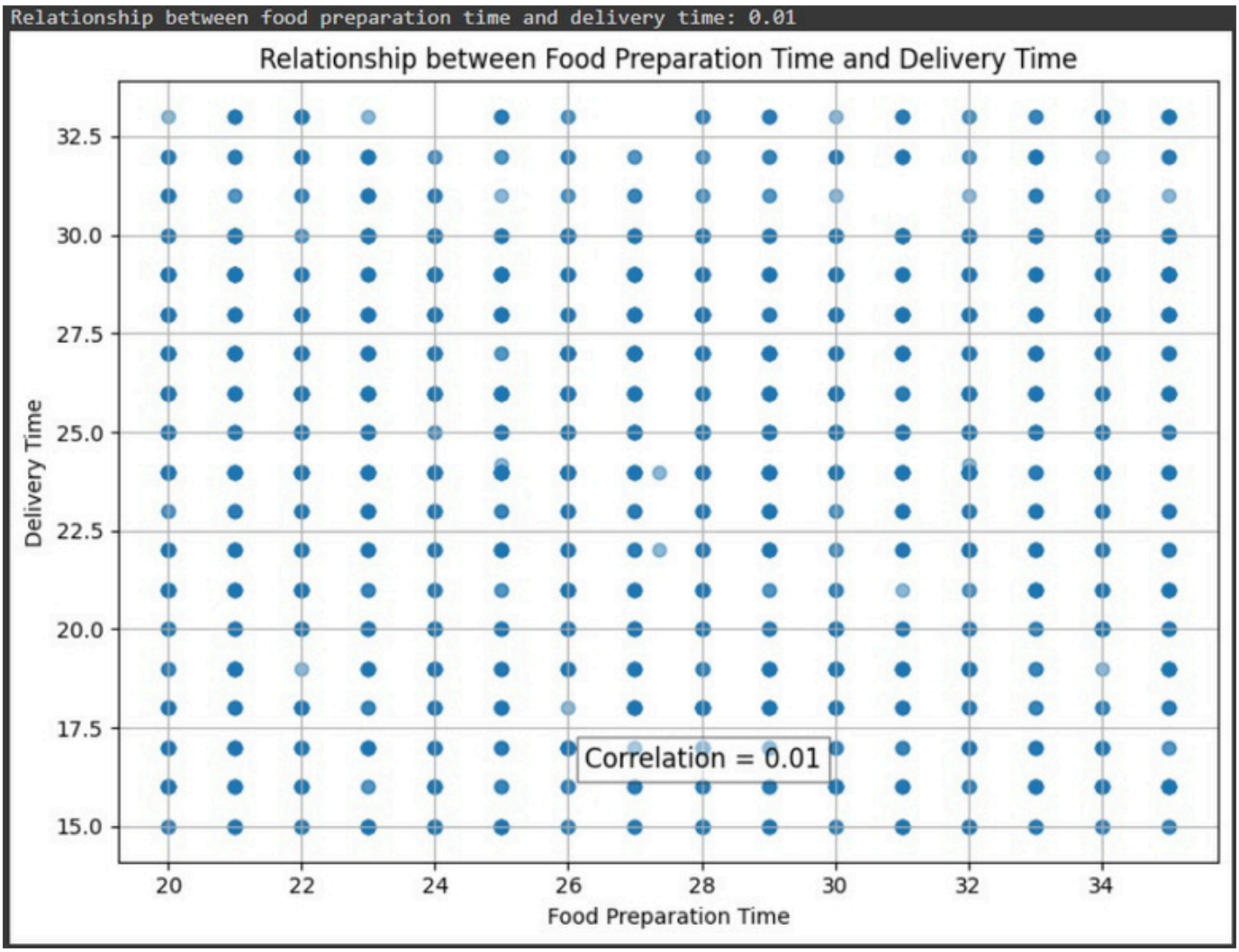
```
Sarabeth's West (33.00 minutes)
```

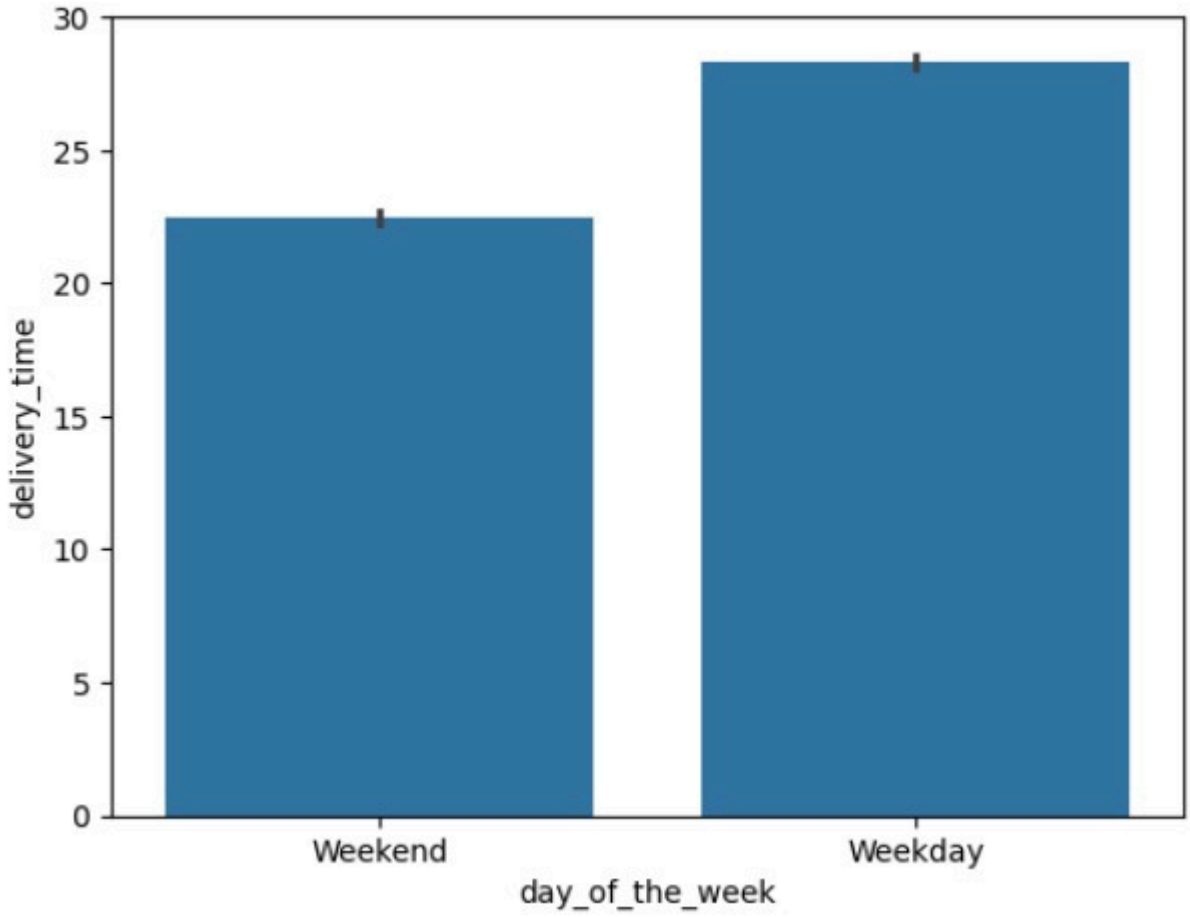
Observations:

The restaurant with the longest average delivery time is Sarabeth's West, taking an average of 33.00 minutes for delivery.

- Relationship Between Food Preparation Time and Delivery Time

Relationship between food preparation time and delivery time: 0.01



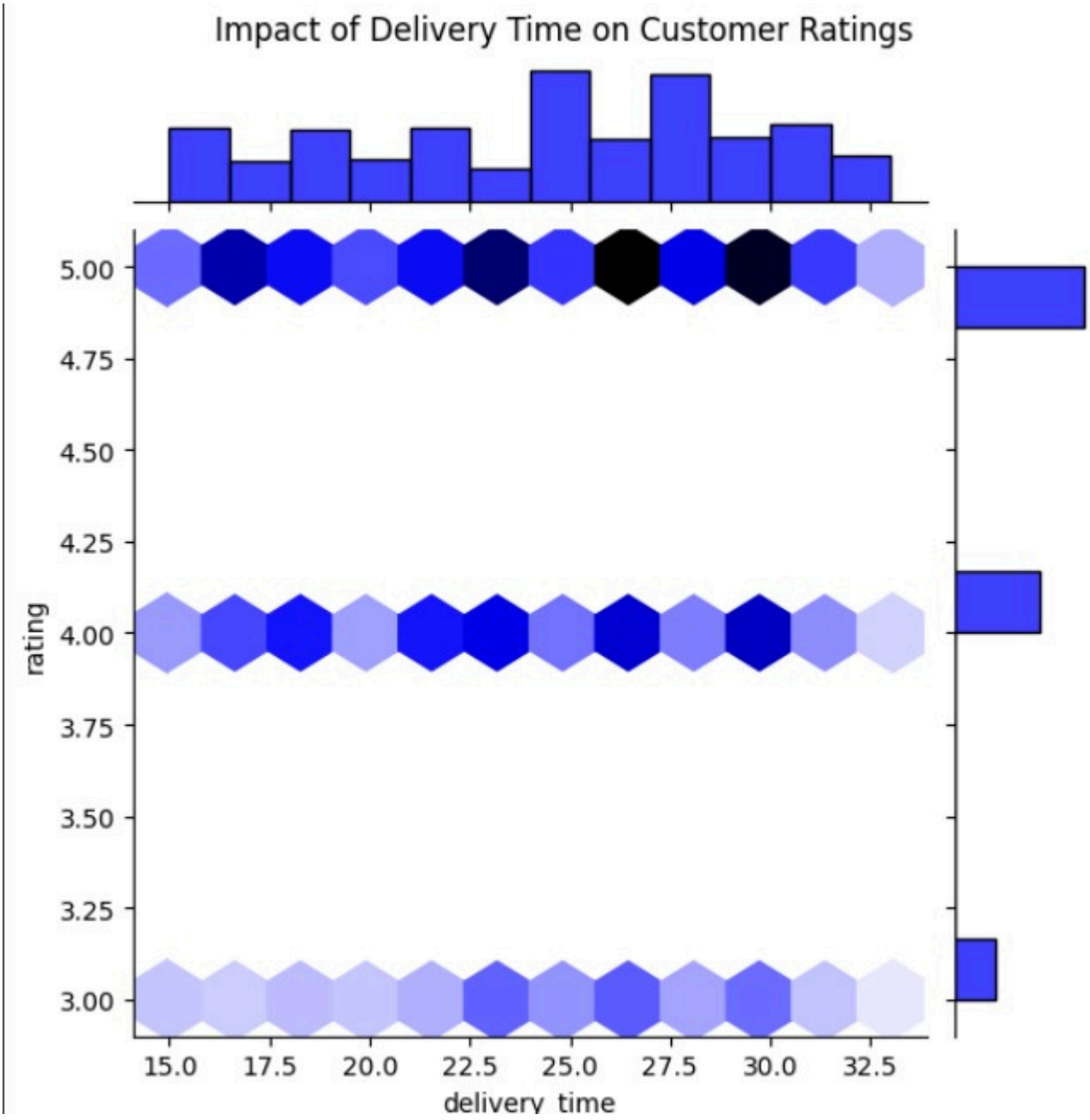


Observations:

On average delivery times are higher during the weekday than on the weekend.

1. Weekday - 28 minutes
2. Weekend - 22 minutes

- Impact of Delivery Time on Customer Ratings

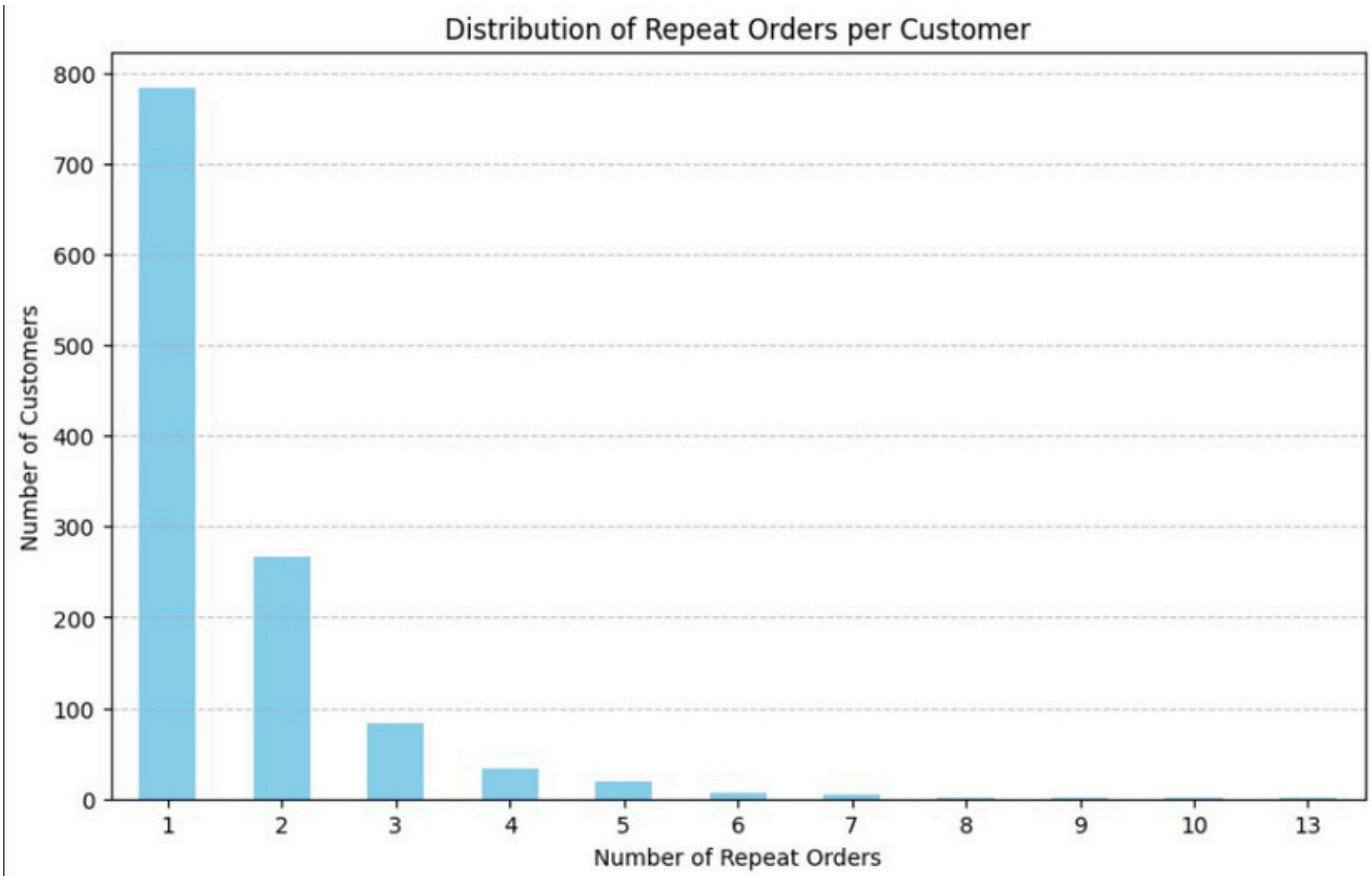


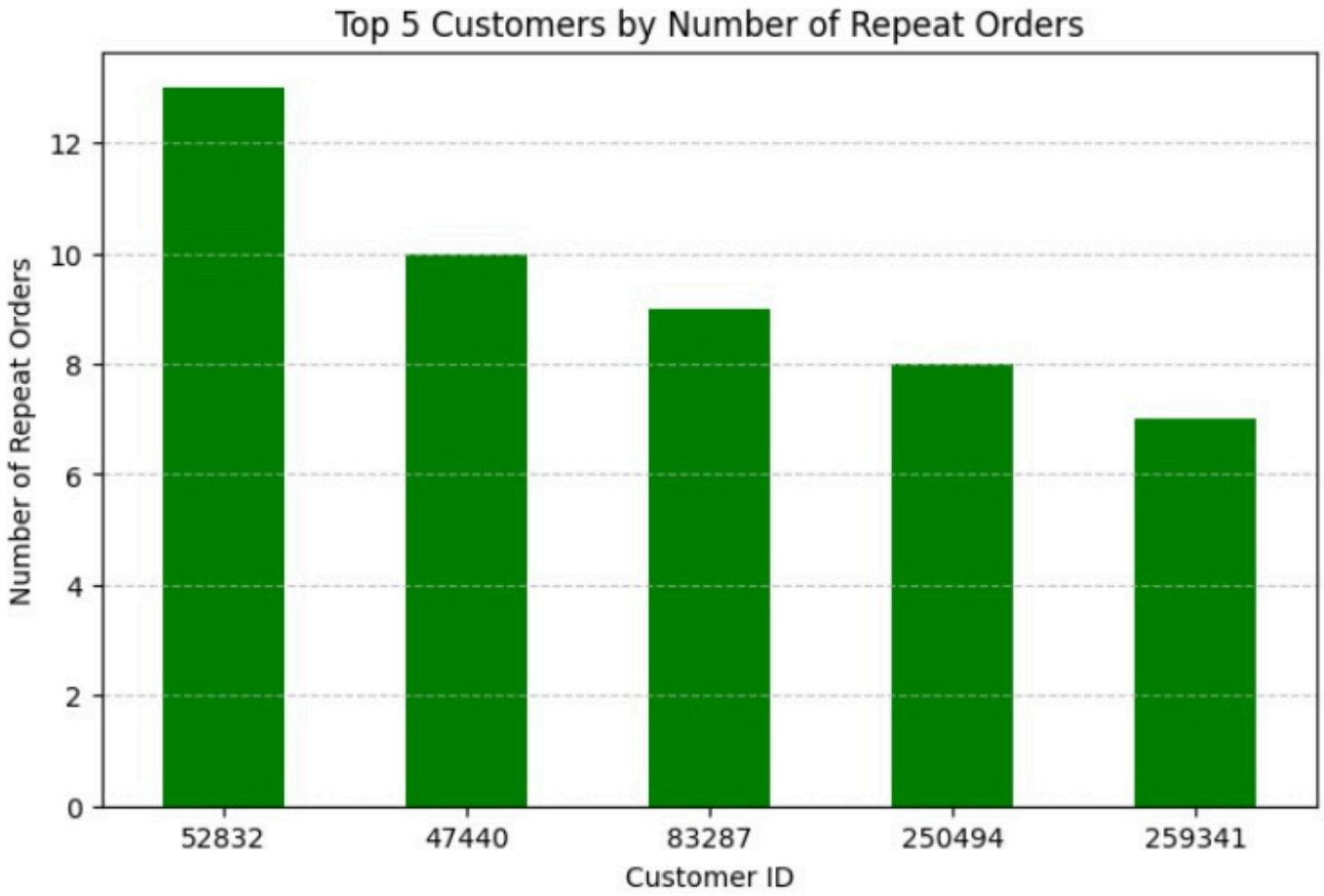
Observations:

Customer ratings are evenly spread across various delivery times, indicating that delivery time has little to no effect on the ratings given.

6. Customer Insights

• Repeat Order Rate

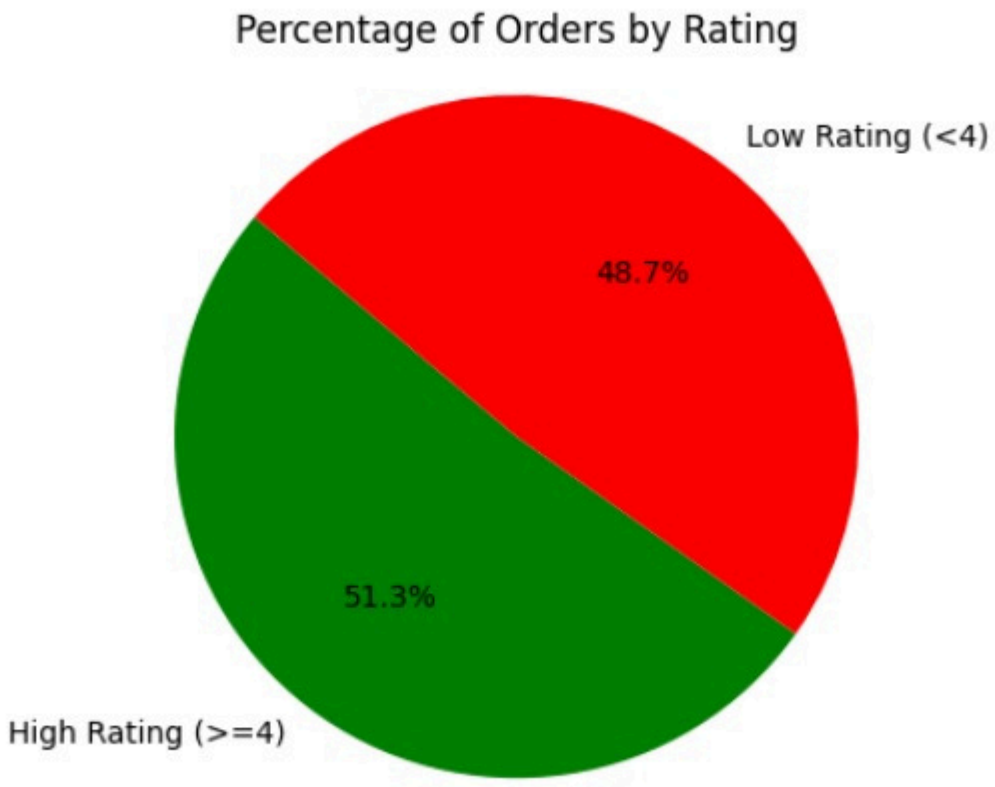




Observations:

Most customers (approximately 700+) place only one order, with the number of customers dropping signi cantly as the number of repeat orders increases. Notably, the top 5 customers have placed between 7 and 13 repeat orders, demonstrating strong loyalty within a small segment of the customer base.

- Percentage of Orders Receiving a Rating of 4 or Higher




```
Percentage of orders receiving a rating of 4 or higher: 51.32%
```

Observations:

Percentage of orders receiving a rating of 4 or higher: 51.32%

Conclusion

- Interestingly, even though our weekends are busier, our delivery times are quicker compared to weekdays—talk about e ciency!
- However, we've noticed that a good chunk of our orders (38%) aren't rated. We're keen to hear from you so we can keep improving.
- Only three standout restaurants have received over 100 orders, and the crowd favorites are clear: American, Japanese, Italian, and Chinese cuisines are the top picks among our customers.
- At Foodhub, weekends are bustling with activity, as that's when most of our orders come through. While many customers place just a single order, we've got one loyal fan who has ordered an impressive numbers.

The analysis provides valuable insights into customer behavior, restaurant performance, and operational e ciency. Recommendations for the company based on these insights might include:

1. Enhancing Customer Experience: Focus on improving delivery times, especially for restaurants with longer delivery durations.

2. Promotional Strategies: Implement targeted promotions on days with lower average orders to balance demand.

3. Cuisine Offering: Increase the availability of popular cuisine types, particularly on weekends when demand is higher.

4. Rating Improvement: Address factors leading to lower ratings, such as long delivery times or high costs, to improve overall customer satisfaction.

Business Implication

1. Operational Efficiency:
By understanding outliers in preparation and delivery times, restaurants can investigate and address potential inefficiencies or exceptional cases that cause delays.
Standardizing preparation and delivery processes could reduce variability and improve customer satisfaction.
2. Menu Optimization:
Analysis of 'cuisine_type' and associated preparation times can guide menu adjustments. For instance, if certain cuisines consistently take longer to prepare, restaurants might streamline those processes or adjust menu offerings to balance preparation times.
3. Customer Experience:Filling missing values and ensuring data accuracy allows for better prediction models, which can improve customer experience through more accurate delivery time estimates and personalized service
4. Strategic Decisions:The insights gained from clean and processed data support strategic decision-making, such as identifying popular dishes, peak order times, and areas for improvement in service delivery.
This foundational work is essential for any subsequent steps in data analysis or machine learning, enabling more accurate and actionable insights for improving restaurant operations and customer satisfaction.