Food_Hub_Data_Analysis

NAVYA SHARMA

Food Hub Analysis

SUMMER BOOTCAMP PROJECT 2024

List of tables

List of figures

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.

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 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

Loading the necessary libraries.

1. Display the top 5 rows.

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_tin
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25.0
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
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25.0
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25.0

- Table 1 : Top 5 rows

- **Based on the above result we can observe that the following columns of 'rating' and 'delivery_time' have some missing values as:**
- **rating** have values as **'Not given'** whish needs to be checked.
- **delivery_time** have values as **'?'**.

2. Display the last 5 rows.

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

- Table 2 : Last 5 rows

- **Based on the above result we can say that the following column of 'resturant_name' and rating have some incorrect and missing values as:**
- **rating** have the missing values represented as **'Not given'**
- **restaurant_name** can have the incorrect value on index 1893 as **Chipotle Mexican Grill $1.99 Delivery** as it contains some numerical value and strings both in single value which may be inappropriate as a name.

3. Check the shape of dataset.

- It shows that our dataset have **1898 rows** and **9 columns**.

4. Check the datatypes of each feature.

The datasets have following types of data in given column

- **1. order_id have integer type of data but its ordinal do it's ok similarly..**
- **2. customer_id have integer type of data.**
- **3. restaurant_name have object type of data (user-defined).**
- **4. cuisine_type have object type of data.**
- **5. cost_of_the_order have float type of data (decimal).**
- **6. day_of_the_week have object type of data.**
- **7. rating have the object type of data, which indicates that there are some string values in rating column which can disturb our data analysis.**
- **8. food_preparation_time have float (decimal) type of data.**
- **9. delivery_time have object type of data but it should be having numerical type of data as there is no need for string values in delivery_time.**

5. Check the Statistical summary.

Results of checking the statistical summary is:

- **customer_id** have count of the **1898**,average of **80.722007**,Standard Deviation of **113698.139743**,minimum of **1311**,25 percentile of **77787.75**,50 percentile of **128600**,75 percentile of **270525**,maximum of **405334**.
- **cost_of_the_order** have count of the **1898**,average of **80.722007**,Standard Deviation of **2798.141333**,minimum of **0**,25 percent of **12.08**,50 percent of **14.16**,75 percent of **22.31**,maximum of **121920**.

- **food_preparation_time** have count of the **1896**,average of **27.371835**,Standard Deviation of **4.634211**,minimum of **20**,25 percent of **23**,50 percent of **27**,75 percent of **31**,maximum of **35**.
- Based on the above observation we found that there is a outlier in **'cost_of_order'** as its min being **0** which

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
```

- Table 3 : Null values

In the given dataset we have null values in following columns only:

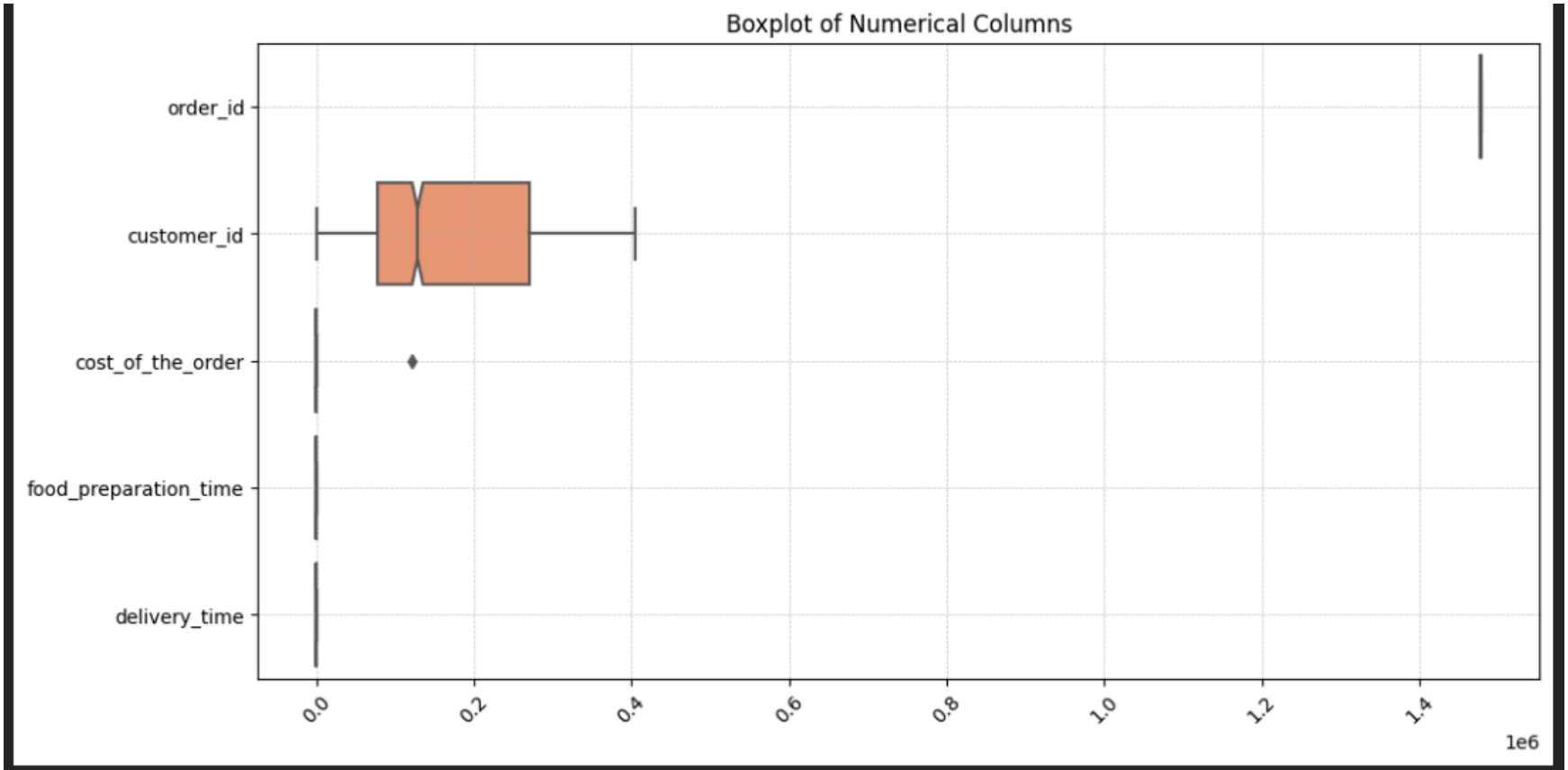
- **cuisine_type** have **3** missing values.
- **food_preparation_time** have **2** missing values.

7. Check the duplicate values

In the given dataset there are 0 duplicate values.

8. Check the anomalies or wrong entries.

- **We will use IQR(Interquartile Range) for detecting outlier in our dataset.**




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

- Figure 1 : Boxplot

The detected outliers for the following columns are:

- **order_id** have **0** outlier.
- **customer_id** have **0** outlier.
- **cost_of_the_order** have **1** outlier.
- **food_preparation_time** have **0** outlier.

9. Checking the outliers and thier authenticity

- **delivery_time** have object type of data so we need to convert it to **float** type to correctly analyse the data(which we have done)
- **customer_id** have integer type of data which is useless in real life we don't need mathematical calculation on this values so it can be converted to **object** type of data if needed.

10. Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.

i. Dropping the duplicate values in case if there is any.

- There are 0 dupliccate values.

ii. Unecessary columns

- All the columns are required in this datasets.

iii. Null Value imputation

- **delivery_time** have value as **'?'** on two indexes **1** and **180**.

- Here we can see that the there are many unique values in this column but '?' makes it a object type so we will replace it with nan and the with mean.

- **We will fill null values with mode here so that it won't affect our datasets.**

```
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
```

- Table 4 : Checking Null Values

- Based on the above operations we can now see that the this datasets have **0 null-values**.

- For outliers we have one in **'cost_of_the_order'** and we will replace it with mean.

1. Order Analysis

i. What is the total number of orders in the dataset?

- Total numbers of **orders** in the dataset is **1898**.

ii. What is the average cost of an order?

- Average cost of an **order** according to dataset is **80.7645598313126**

iii. How many unique customers have placed orders?

- **1200** unique **customers** have placed orders.

iv. Which restaurant has received the highest number of orders?

- **Shake Shack** is the restaurant which have recieved highest number of orders.

2. Customer Behavior

i. What is the average rating given by customers?

- We already converted **'rating'** into **numeric** data type for further analysis.

- Average **rating** given by customers is **4.598524762908324**.

ii. How does the rating vary between weekdays and weekends?

It can be seen that on both weekdays and weekdays the average rating is 4.57 and 4.60 which means there is not much difference

```
day_of_the_week
Weekday    4.570384
Weekend    4.609919
Name: rating, dtype: float64
```

- Table 4 : Variation of rating between weekends and weekdays

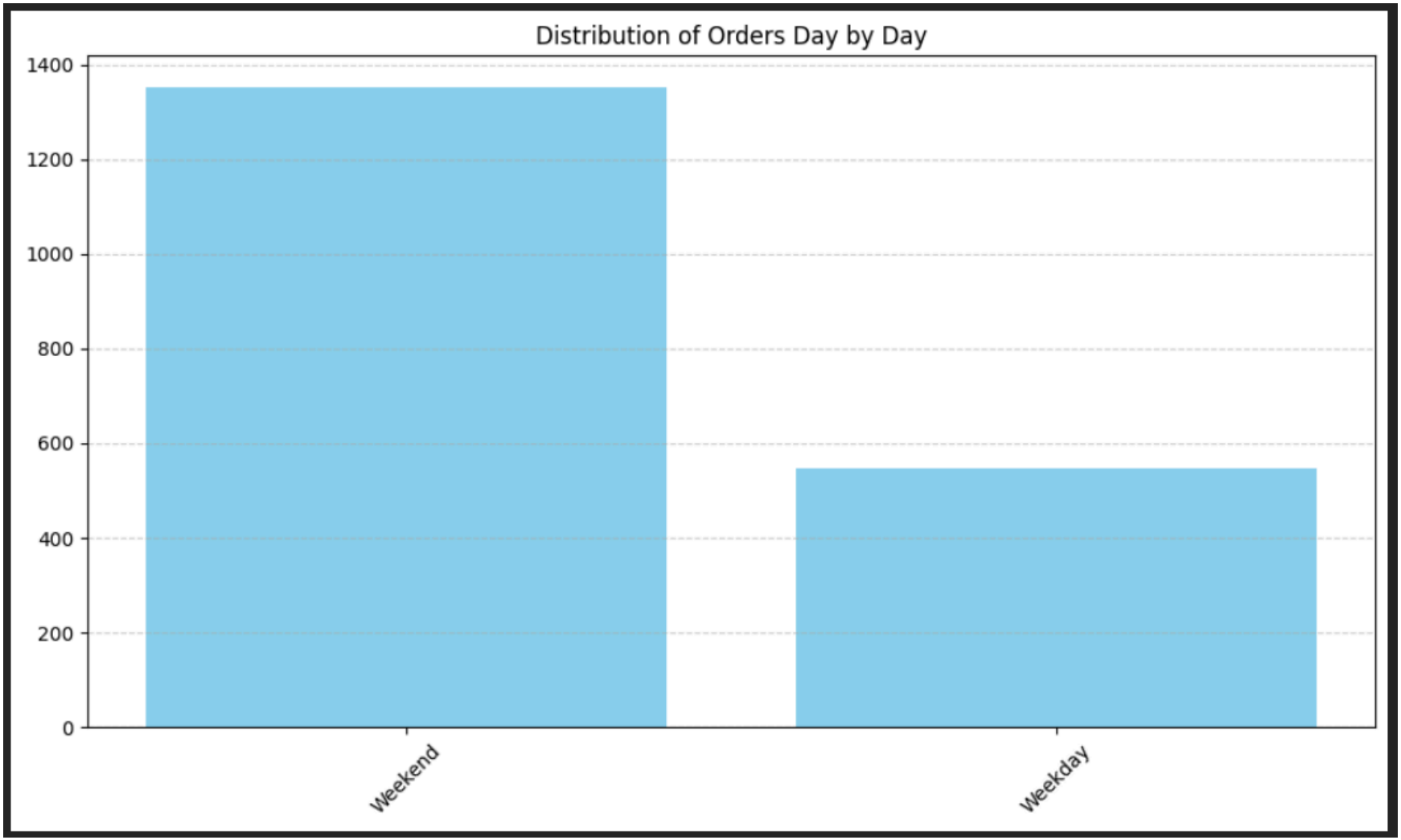
iii. Which cuisine type is ordered the most?

- Most ordered **'cuisine_type'** is **American**.

iv. What is the distribution of orders across different days of the week?

```
Weekend    1351
Weekday     547
Name: day_of_the_week, dtype: int64
```

- Table 5 : Distribution of orders across weekdayas and weekends

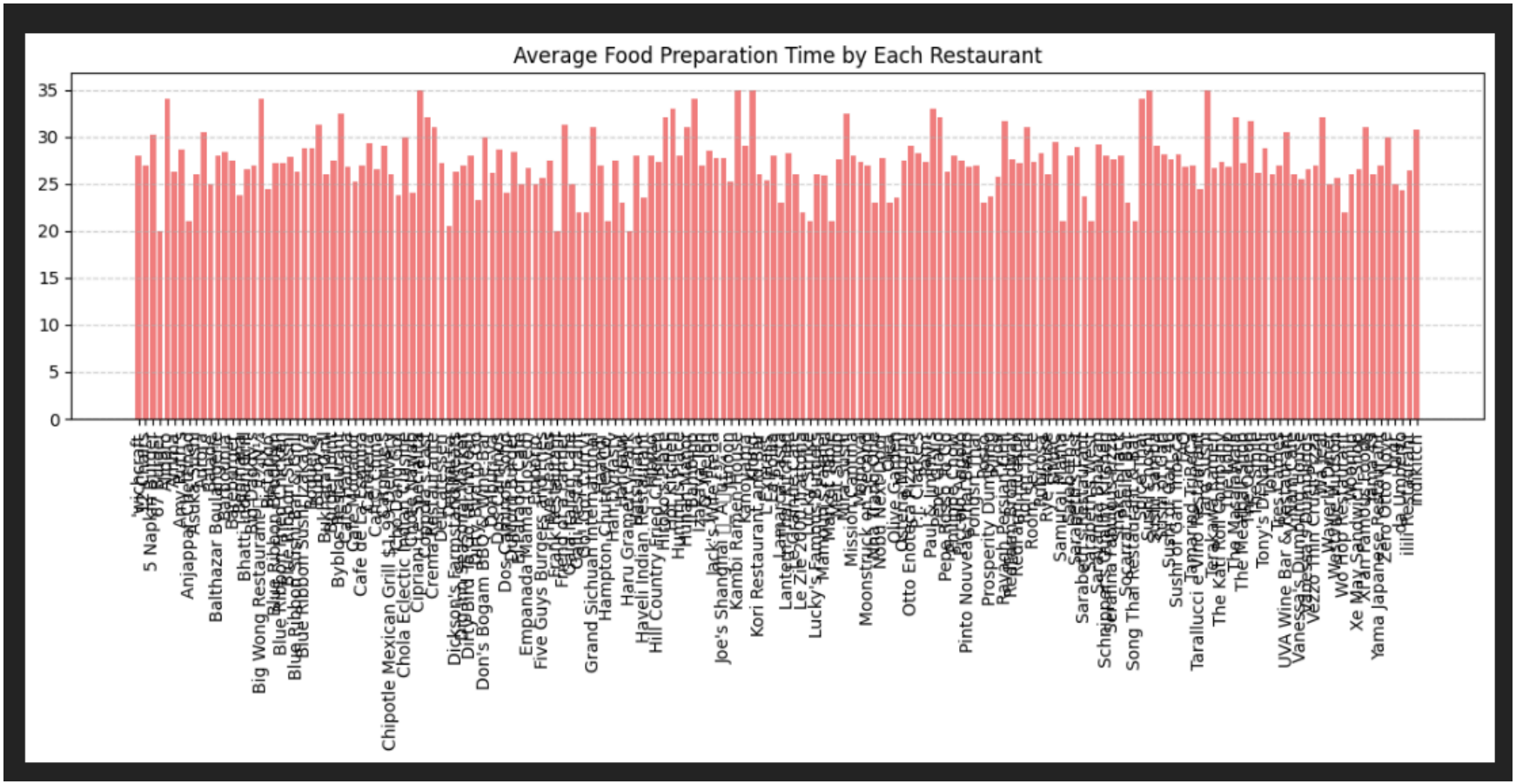


- Figure 2 : Distribution of orders day by day

- We have **1351** orders **distributed** on **Weekend**.
- We have **547** orders **distributed** on **Weekday**.

3. Restaurant Performance

i. What is the average food preparation time for each restaurant?



- Figure 3 : Average food preparation time

Average food_time_preparation for each restaurants are as follows:

- **Haru Gramercy** Park has received an **average** of **20 minutes** .
- **67 Burger** has also received an **average** of **20 minutes**.
- **Frank Restaurant** has received an **average** of **20 minutes**.
- **Despaña** has received an **average** of **20.5 minutes**.
- **Sarabeth's** West has received an average of 21 minutes.
- **Lucky Strike** has received an average of 21 minutes.
- **Song Thai Restaurant & Bar** has received an average of 21 minutes.
- **Hampton Chutney Co.** has received an average of 21 minutes.
- **Samurai Mama** has received an average of 21 minutes.
- **Market Table** has received an average of 21 minutes.
- **Anjappar Chettinad** has received an average of 21 minutes.
- **Go! Go! Curry!** has received an average of 22 minutes.

- **Le Zie 2000 Trattoria** has received an average of 22 minutes.
- **Galli Restaurant** has received an average of 22 minutes.
- **Wo Hop Restaurant** has received an average of 22 minutes.
- **Hangawi** has received an average of 23 minutes.
- **Nha Trang One** has received an average of 23 minutes.
- **Lamarca Pasta** has received an average of 23 minutes.
- **Olea** has received an average of 23 minutes.
- **Socarrat Paella Bar** has received an average of 23 minutes.
- **Posto** has received an average of 23 minutes.
- **Dirty Bird** to Go has received an average of 23.25 minutes.
- **Olive Garden** has received an average of 23.5 minutes.
- **Haveli Indian Restaurant** has received an average of 23.5 minutes.
- **Prosperity Dumpling** has received an average of 23.67 minutes.
- **Sarabeth's Restaurant** has received an average of 23.71 minutes.
- **Benihana** has received an average of 23.8 minutes.
- **Cho Dang Gol** has received an average of 23.83 minutes.
- **Chote Nawab** has received an average of 24 minutes.
- **Dos Caminos Soho** has received an average of 24 minutes.
- **da Umberto** has received an average of 24.33 minutes.
- **Tarallucci e Vino Restaurant** has received an average of 24.5 minutes.
- **Bistango** has received an average of 24.5 minutes.
- **El Parador Cafe** has received an average of 25 minutes.
- **Emporio** has received an average of 25 minutes.
- **Gaia Italian Cafe** has received an average of 25 minutes.
- **Waverly Diner** has received an average of 25 minutes.
- **brgr** has received an average of 25 minutes.
- **Balade** has received an average of 25 minutes.
- **Junoon** has received an average of 25.2 minutes.
- **Cafe Mogador** has received an average of 25.28 minutes.
- **L'Express** has received an average of 25.33 minutes.
- **Vanessa's Dumpling House** has received an average of 25.56 minutes.
- **Five Guys Burgers and Fries** has received an average of 25.62 minutes.
- **Westville Hudson** has received an average of 25.67 minutes.
- **Pylos** has received an average of 25.8 minutes.
- **Mamoun's Falafel** has received an average of 25.86 minutes.
- **Lucky's Famous Burgers** has received an average of 26 minutes.

- **Le Grainne Cafe** has received an average of 26 minutes.
- **Rye House** has received an average of 26 minutes.
- **Asuka Sushi** has received an average of 26 minutes.
- **Yama 49** has received an average of 26 minutes.
- **Woorijip** has received an average of 26 minutes.
- **Kori Restaurant and Bar** has received an average of 26 minutes.
- **Bukhara Grill has received an average of 26 minutes.
- **V-Nam Cafe** has received an average of 26 minutes.
- **Chipotle Mexican Grill $1.99 Delivery** has received an average of 26 minutes.
- **Tortaria** has received an average of 26 minutes.
- **The Smile** has received an average of 26.11 minutes.
- **Donburi-ya** has received an average of 26.14 minutes.
- **Amma** has received an average of 26.25 minutes.
- **Pepe Rosso To Go** has received an average of 26.29 minutes.
- **Blue Ribbon Sushi Bar & Grill** has received an average of 26.3 minutes.
- **Dickson's Farmstand Meats** has received an average of 26.33 minutes.
- **ilili Restaurant** has received an average of 26.39 minutes.
- **Café China** has received an average of 26.5 minutes.
- **Bhatti Indian Grill** has received an average of 26.5 minutes.
- **Xe May Sandwich Shop** has received an average of 26.5 minutes.
- **Vanessa's Dumplings** has received an average of 26.54 minutes.
- **Empanada Mama (closed)** has received an average of 26.62 minutes.
- **Terakawa Ramen** has received an average of 26.67 minutes.
- **Cafe Habana** has received an average of 26.75 minutes.
- **The Loop** has received an average of 26.75 minutes.
- **Pinto Nouveau Thai Bistro** has received an average of 26.75 minutes.
- **TAO** has received an average of 26.78 minutes.
- **Yama Japanese Restaurant** has received an average of 26.94 minutes.
- **Tamarind TriBeCa** has received an average of 26.96 minutes.
- **12 Chairs** has received an average of 27 minutes.
- **Big Daddy's** has received an average of 27 minutes.
- **Moonstruck on Second** has received an average of 27 minutes.
- **Pongsri Thai** has received an average of 27 minutes.
- **Vezzo Thin Crust Pizza** has received an average of 27 minutes.
- **Cafe de La Esquina** has received an average of 27 minutes.
- **Tres Carnes** has received an average of 27 minutes.

- **Izakaya Ten** has received an average of 27 minutes.
- **Haandi** has received an average of 27 minutes.
- **Dig Inn Seasonal Market** has received an average of 27 minutes.
- **Blue Ribbon Fried Chicken** has received an average of 27.16 minutes.
- **RedFarm Hudson** has received an average of 27.16 minutes.
- **The Meatball Shop** has received an average of 27.18 minutes.
- **Delicatessen** has received an average of 27.2 minutes.
- **Blue Ribbon Brooklyn** has received an average of 27.25 minutes.
- **Momoya** has received an average of 27.3 minutes.
- **Parm** has received an average of 27.31 minutes.
- **Room Service** has received an average of 27.33 minutes.
- **Hill Country Fried Chicken** has received an average of 27.36 minutes.
- **The Kati Roll Company** has received an average of 27.38 minutes.
- **Bareburger** has received an average of 27.41 minutes.
- **Han Dynasty** has received an average of 27.41 minutes.
- **Osteria Morini** has received an average of 27.43 minutes.
- **Five Leaves** has received an average of 27.5 minutes.
- **Piccolo Angolo** has received an average of 27.5 minutes.
- **Burger Joint** has received an average of 27.5 minutes.
- **Sushi of Gari 46** has received an average of 27.54 minutes.
- **RedFarm Broadway** has received an average of 27.59 minutes.
- **Serafina Fabulous Pizza** has received an average of 27.6 minutes.
- **Melt Shop** has received an average of 27.64 minutes.
- **Jack's Wife Freda** has received an average of 27.72 minutes.
- **Joe's Shanghai** has received an average of 27.75 minutes.
- **Nobu Next Door** has received an average of 27.76 minutes.
- **Shake Shack** has received an average of **27.9** minutes.

ii. Which restaurant has the shortest average food preparation time?

```
restaurant_name          'wichcraft
food_preparation_time          20.0
dtype: object
```

- Based on above operations we can say that the restaurants having **shortest average food preparation time** is **'wichcraft'** with **20 minutes**.

iii. How does the average delivery time compare across different restaurants?

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

178 rows × 2 columns

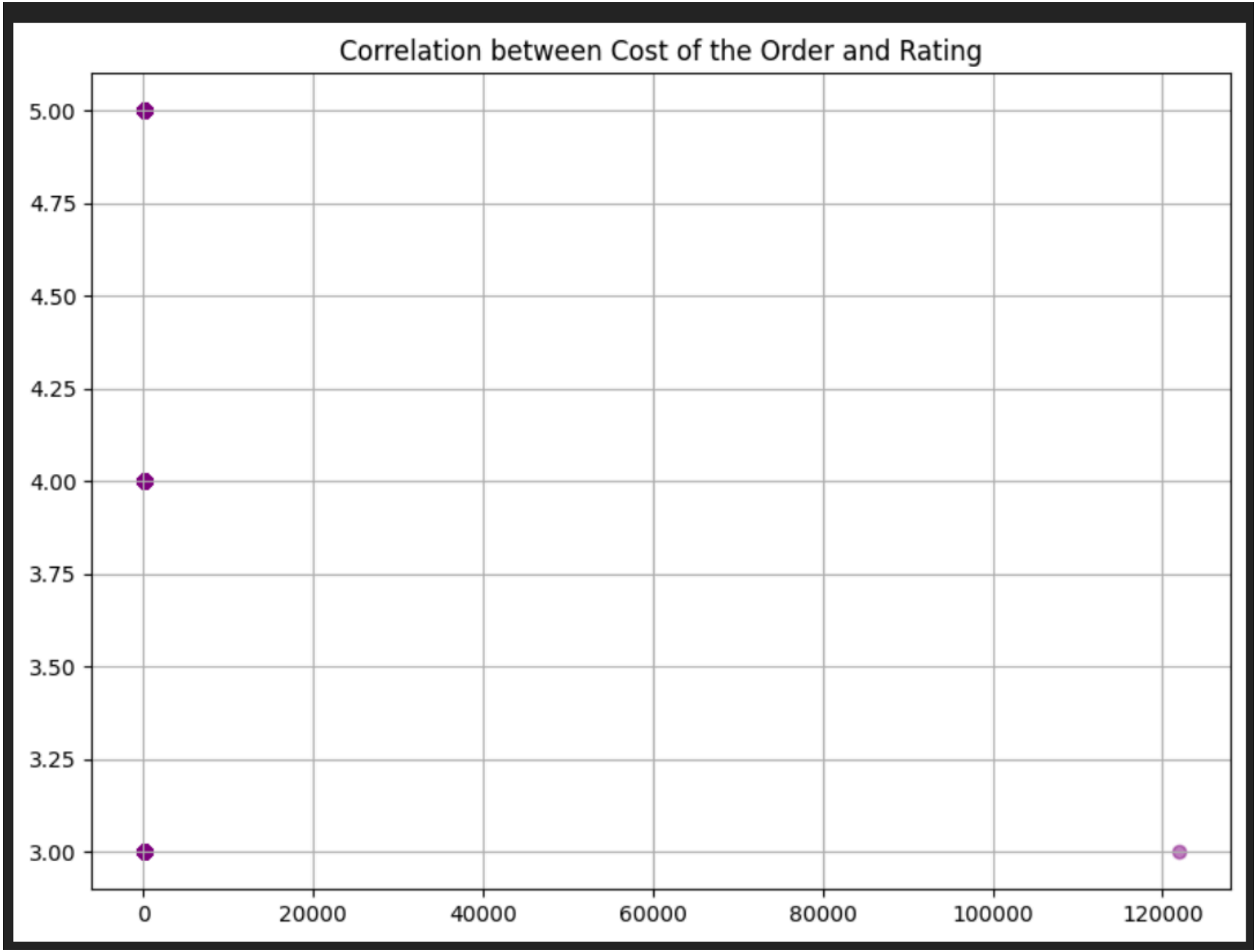
- Table 6 : Comparing average delivery time across restaurants

- **Restaurants with the shortest average delivery times (15 minutes) include Gaia Italian Cafe, Paul & Jimmy's, The MasalaWala, and Hibino. On the other end, Sarabeth's West has the longest average delivery time at 33 minutes**.

iv. Is there a correlation between the cost of the order and the rating given?

Plotting the above results to get a better visualisation

```
Correlation coefficient between cost of order and rating: -0.055418689178310894
Interpretation: There is a negative correlation.
```

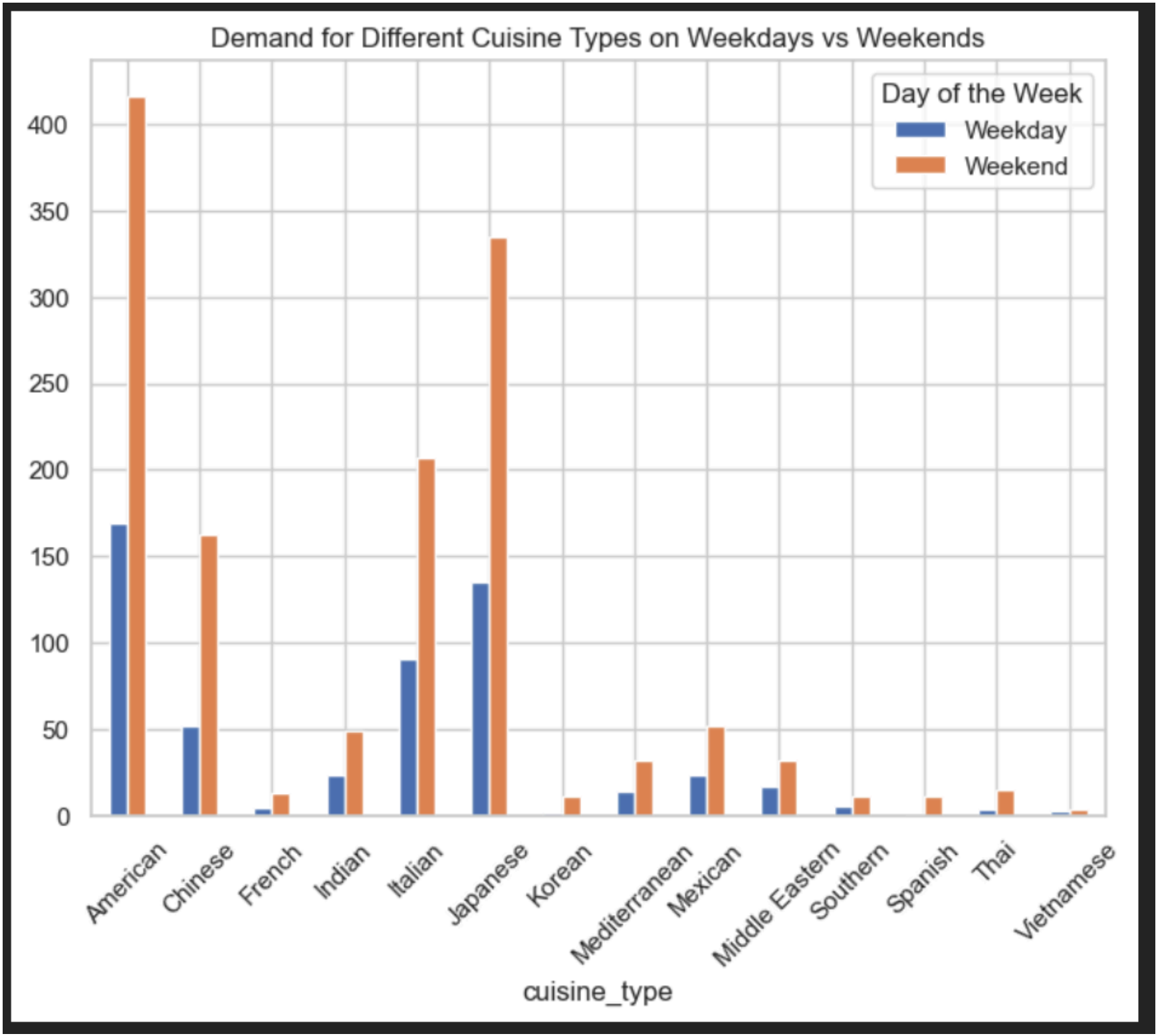


- Figure 4 : Correlation between cost of the order and rating

- **Here we can see that the correlation coefficient of 'cost_of_the_order and 'rating' is -0.055418689178310894 which further shows that the both have strong negative correlation between them.**

4. Demand Patterns

i. How does the demand for different cuisine types vary on weekdays versus weekends?



- Figure 5 : Demand for different cuisine types on weekdays and weekends

- Based on above results we can say that most ordereder food on **weekday** is **American**, **italian**, **japanese**, following **chinese**.
- Most ordered food on **weekend** is **American**, **Japenese**, **Italian**, **Chinese**.
- **Japenese** was ordered more than **Italian** on **weekend**.

ii. Which day of the week has the highest average order cost?

```
The day with the highest average order cost is Weekend with an average cost of 106.86
```

- On **weekend** rstaurant have recived **highest average order cost** of **106.86**.

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

```
The most common day for orders to be placed is Weekend with 1351 orders.
```

The most common day for orders to be placed is Weekend with 1351 orders.

iv.How does the average rating vary by cuisine type?

```
cuisine_type
Spanish              4.916667
Thai                 4.842105
Indian               4.684932
Japanese             4.636170
Mexican              4.631579
Italian              4.630872
French               4.611111
Chinese              4.590698
American             4.558974
Southern             4.470588
Middle Eastern       4.469388
Mediterranean        4.456522
Korean               4.384615
Vietnamese           4.285714
Name: rating, dtype: float64
```

- Table 7 : Variation of average rating by cuisine type

Average rating vary by cuisne type are as follows:

- **Spanish** have rating of **4.916667**
- **Thai** have rating of **4.842105**
- **Indian** have rating of **4.684932**
- **Japanese** have rating of **4.636170**
- **Mexican** have rating of **4.631579**
- **Italian** have rating of **4.630872**
- **French** have rating of **4.611111**
- **Chinese** have rating of **4.590698**
- **American** have rating of **4.558974**
- **Southern** have rating of **4.470588**
- **Middle Eastern** have rating of **4.469388**
- **Mediterranean** have rating of **4.456522**
- **Korean** have rating of **4.384615**
- **Vietnamese** have rating of **4.285714**

5. Operational Efficiency

i. What is the average delivery time for all orders?

```
The average delivery time for all orders is 24.16 minutes.
```

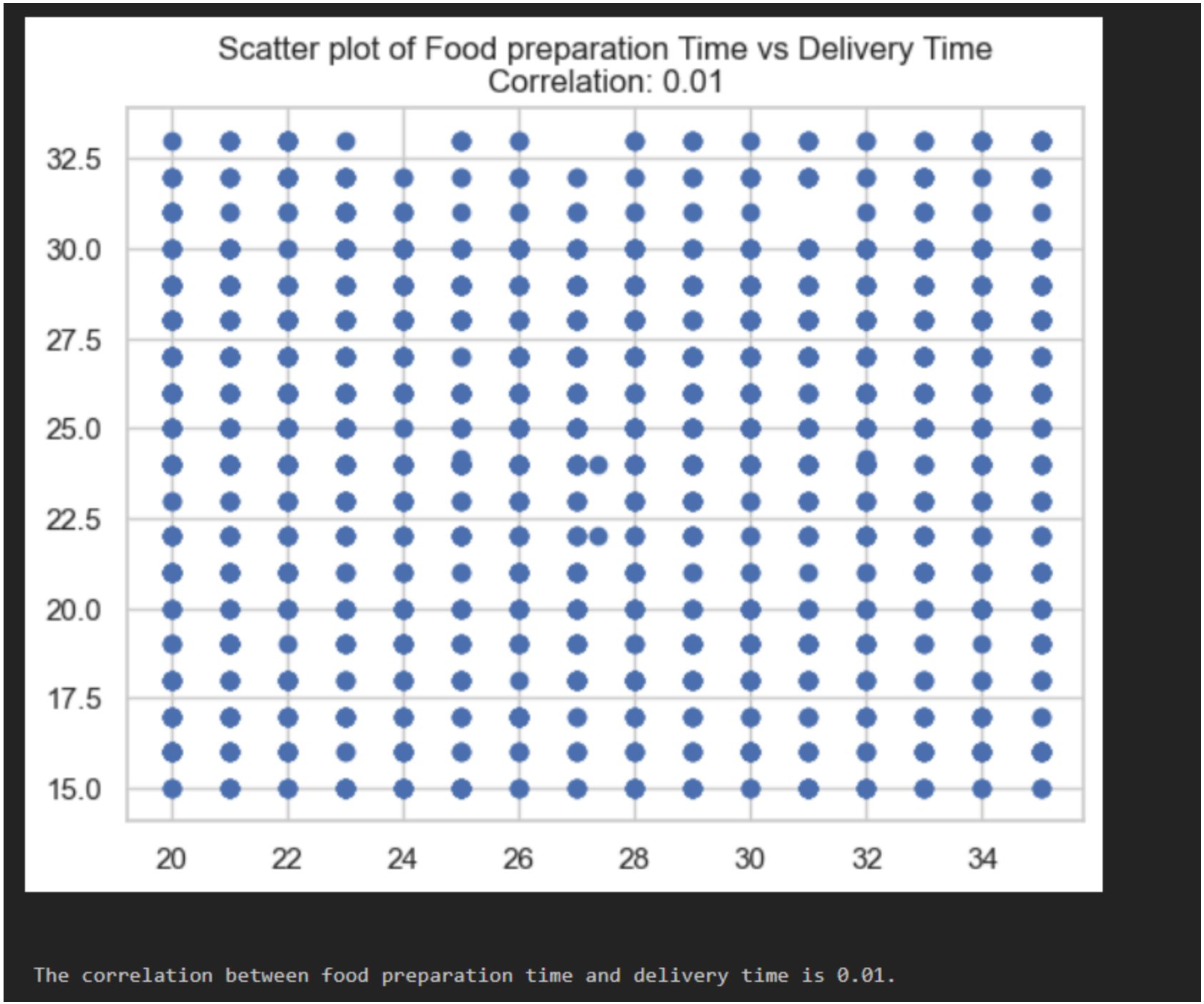
- The average **delivery time** taken for all orders is **24.16 minutes**.

ii. Which restaurant has the longest average delivery time?

```
("Sarabeth's West", 33.0)
```

- Restaurant having longest average delivery time is **'Sarabeths's West'** with **33 minutes** of average delivery time.

iii.Is there a relationship between food preparation time and delivery time?

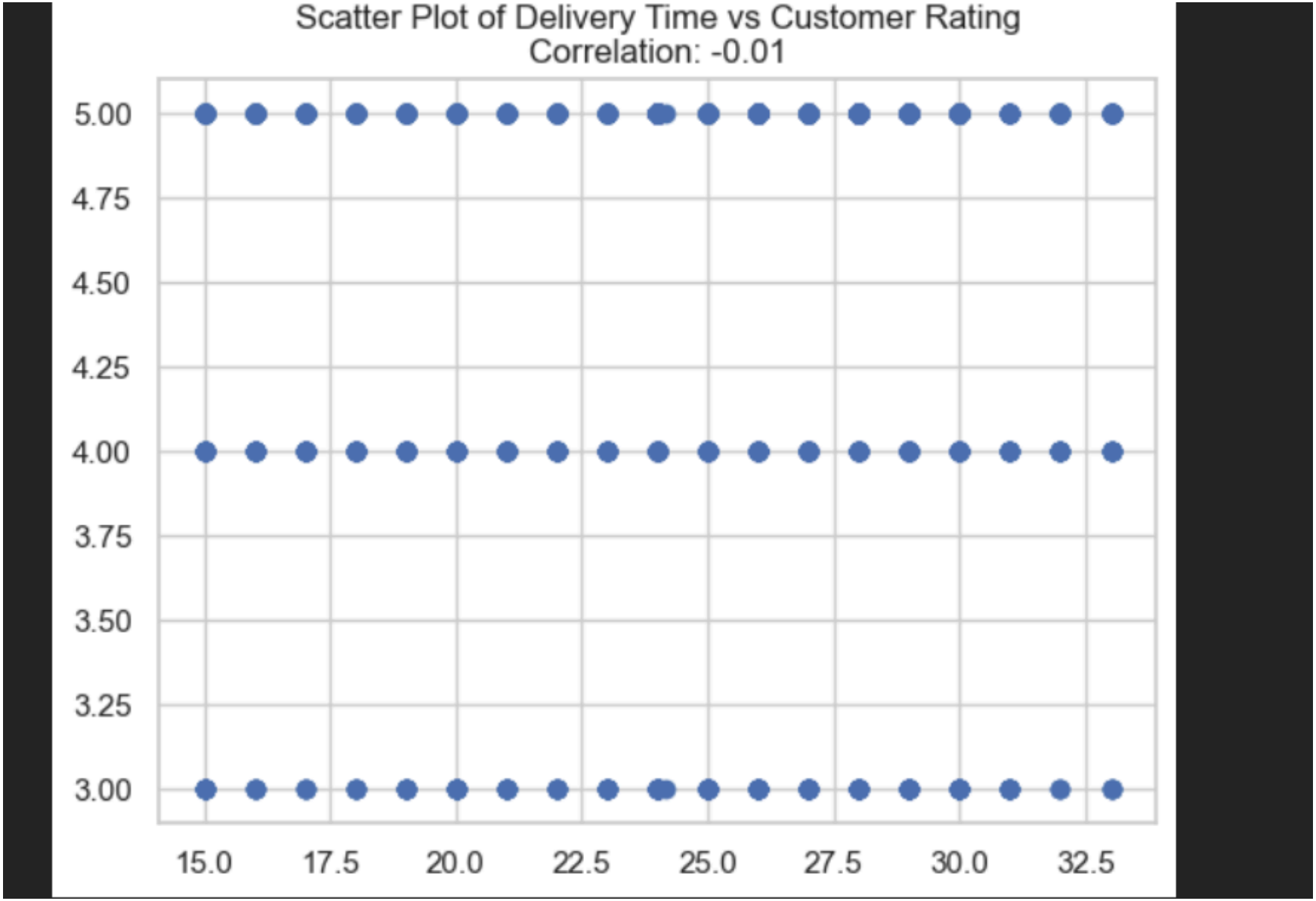


- Figure 6 : Scatter plot between food preparation time and delivery time

- From abbove results we can observe that **correlation coefficient** between **food_preparation_time** and **delivery_time** is **0.01**.

- This shows that there is a **neglible correlation** between the data.

iv. How does the delivery time impact customer ratings?



- Figure 7 : Scatter plot between delivery time and customer ratings

- It doesn't affect the **customers rating** as correlation between them is **-0.01**.
- It shows that the **rating** depends on some other parameters.

6. Customer Insights

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

```
0.3466666666666667
```

```
416
```

- **Repeat order** rate is **0.3466666666666667**
- **Number of customers** who have placed more than one order is **416**.

ii. What percentage of orders receive a rating of 4 or higher?

```
90.09483667017913
```

- The **percentage** of **orders recieving** a **rating** of **4** or higher is **90.09483667017913**.

Thank You!!

In [ ]: