FoodHub Project

Python Foundations – Business Analysis

Bowen Wilder
2024-05-31

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Contents / Agenda

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

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Executive Summary

- Please mention conclusions, actionable insights & recommendations
- Please add the answer for question 17

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

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Problem Overview and Solution Approach

- Please define the problem
- Please mention the solution approach / methodology

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

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Shape of the Dataset

- There are 1,898 ROWS in the dataset
- There are 9 COLUMNS in the dataset
    - Sample

	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

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Dataset Column Datatypes

- Integers:
  - order_id,
  - customer_id,
  - food_preparation_time
- Strings:
  - restaurant_name,
  - cuisine_type,
  - day_of_the_week,
  - rating
- Floats:
  - cost_of_the_order

```
 #   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
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Dataset Missing Values

- There are no missing values in the dataset

```
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
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Dataset Statistical Summary

- Food Preparation Minimum time: 20 minutes
- Food Preparation Mean time: 27.37 minutes
- Food Preparation Maximum time: 35 minutes

	count	mean	std	min	25%	50%	75%	max
order_id	1898.0	1.477496e+06	548.049724	1476547.00	1477021.25	1477495.50	1.477970e+06	1478444.00
customer_id	1898.0	1.711685e+05	113698.139743	1311.00	77787.75	128600.00	2.705250e+05	405334.00
cost_of_the_order	1898.0	1.649885e+01	7.483812	4.47	12.08	14.14	2.229750e+01	35.41
food_preparation_time	1898.0	2.737197e+01	4.632481	20.00	23.00	27.00	3.100000e+01	35.00
delivery_time	1898.0	2.416175e+01	4.972637	15.00	20.00	25.00	2.800000e+01	33.00

- order_id and customer_id are categorical values therefore their statistical summary is not very helpful or necessary

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Unrated Orders

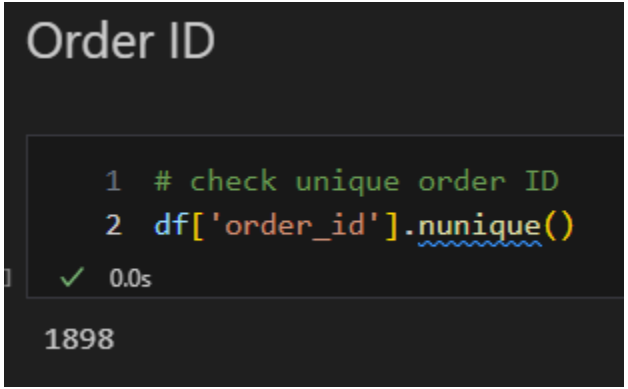
- 736 order ratings were not provided by customers that used the service

```
rating
Not given    736
5            588
4            386
3            188
Name: count, dtype: int64
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Order ID

- FoodHub completed 1,898 unique orders. One for every row in the dataset.



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Customer ID

- FoodHub captured 1,200 unique customer ids.
- 416 Customers ordered more than once from the service

```
1  # check unique customer ID
2  df['customer_id'].nunique()
```

✓ 0.0s

```
1200
```

```
1  # check unique customer ID
2  df['customer_id'].value_counts().head(420)
```

✓ 0.0s

```
customer_id
52832     13
47440     10
83287      9
250494     8
259341     7
          ..
155124     2
198194     1
298794     1
62929      1
75169      1
Name: count, Length: 420, dtype: int64
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Restaurant Name

- There are 178 unique restaurant names in the dataset

```
1  # check unique Restaurant Name
2  df['restaurant_name'].nunique()
```

✓ 0.0s

```
178
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

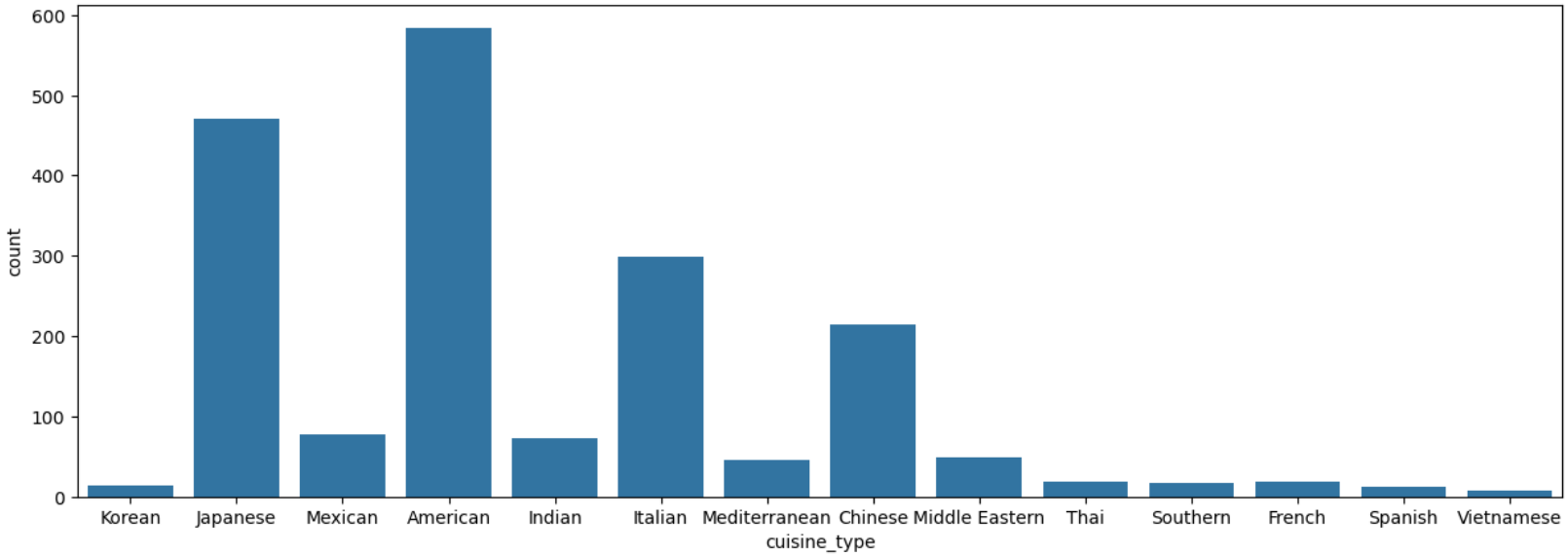
Univariate Analysis: Cuisine Type

- There are 14 unique cuisine types

```
1  # Check unique cuisine type
2  df['cuisine_type'].nunique()
```

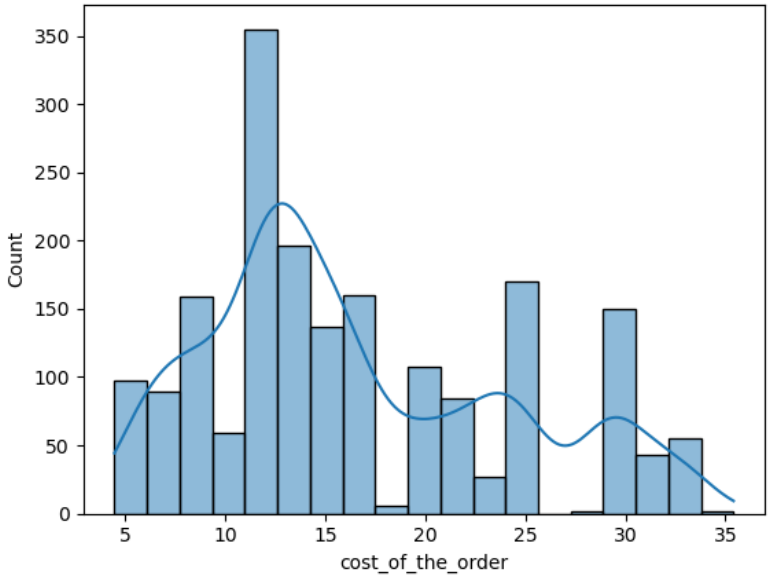
✓ 0.0s

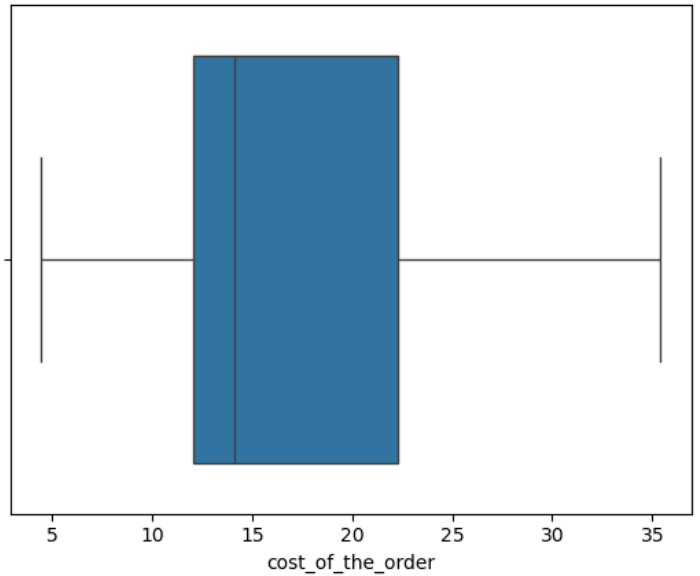
```
14
```



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Cost of the Order





Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

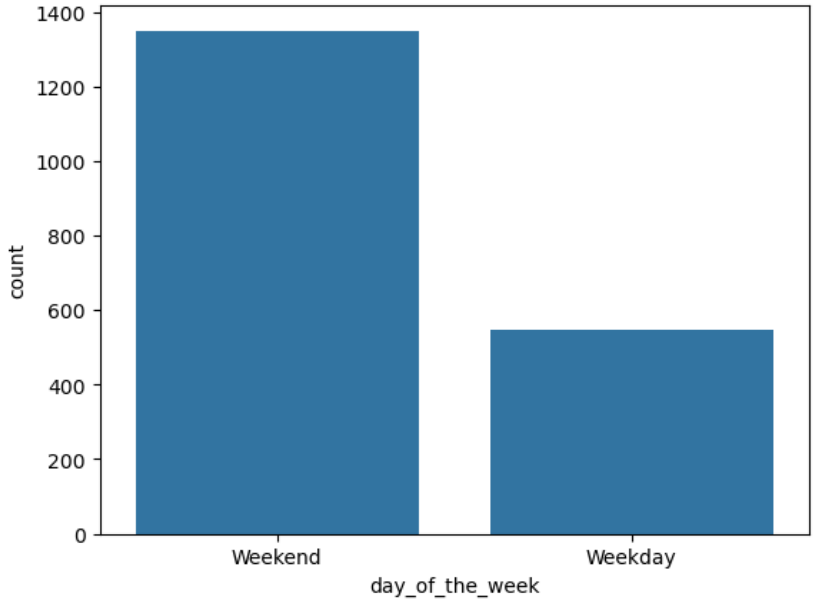
Univariate Analysis: Day of the Week

- There are only 2 unique values in the day_of_the_week column

```
1  # Check the unique values
2  df['day_of_the_week'].nunique()
```

✓ 0.0s

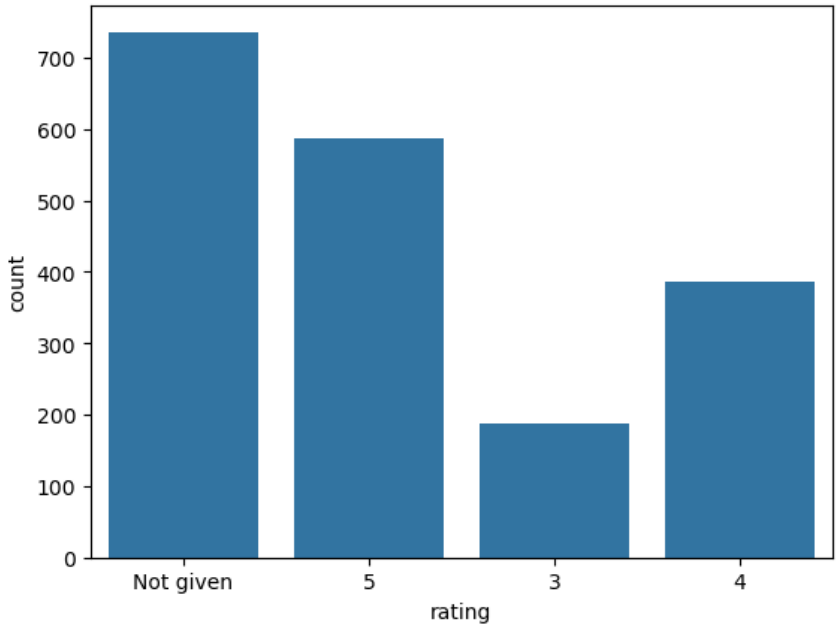
```
2
```



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

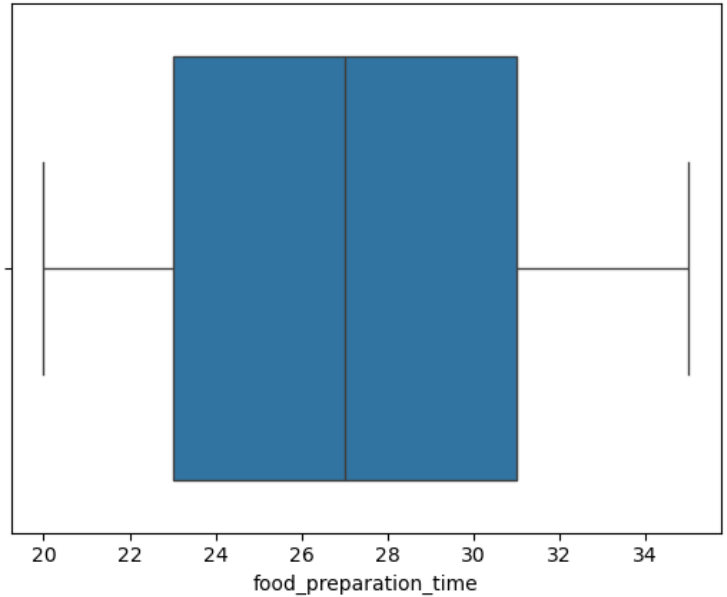
Univariate Analysis: Rating

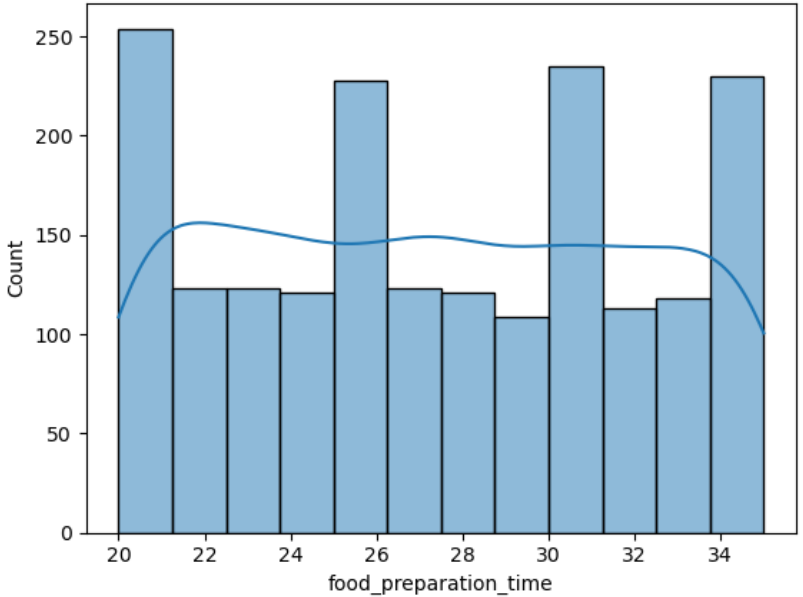
- Many customers chose not to leave a rating at all



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

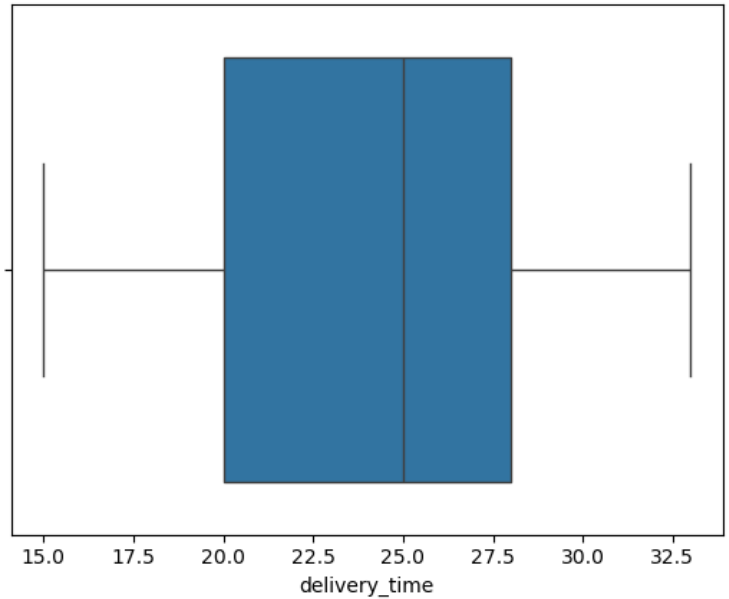
Univariate Analysis: Food Preparation Time

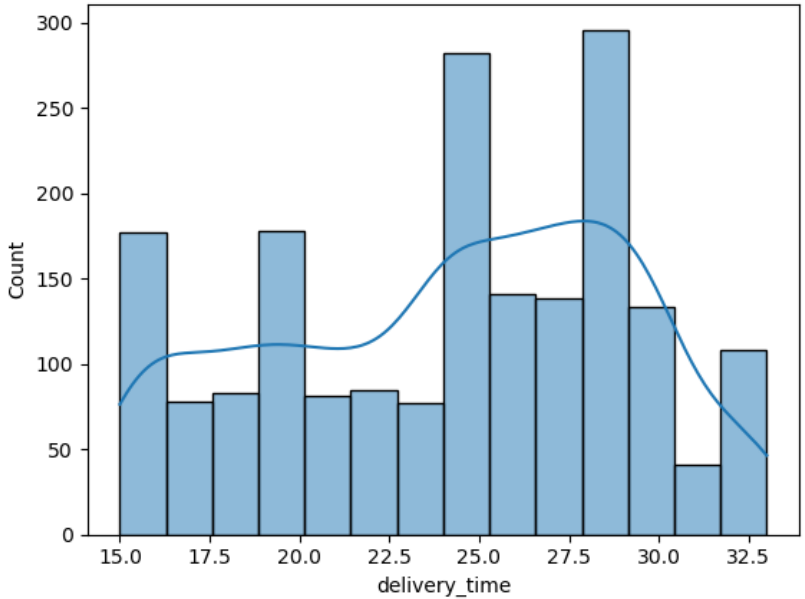




Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Delivery Time





Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Top 5 Restaurants by Number of Orders Received

```
1  # Get top 5 restaurants with highest number
2  df['restaurant_name'].value_counts().head(5)
```

✓ 0.0s

```
restaurant_name
Shake Shack                219
The Meatball Shop          132
Blue Ribbon Sushi          119
Blue Ribbon Fried Chicken   96
Parm                        68
Name: count, dtype: int64
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Most Popular Weekend Cuisines

- American cuisine is most popular on the weekends

```
1  # Get most popular cuisine on weekends
2  df_weekend = df[df['day_of_the_week'] == 'Weekend']
3  df_weekend['cuisine_type'].value_counts() ## Complet
```

✓ 0.0s

```
cuisine_type
American          415
Japanese          335
Italian           207
Chinese           163
Mexican            53
Indian             49
Mediterranean      32
Middle Eastern     32
Thai               15
French             13
Korean             11
Southern           11
Spanish            11
Vietnamese          4
Name: count, dtype: int64
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Percentage of Orders Greater than $20

- The number of total orders that cost above 20 dollars is: 555
- Percentage of orders above 20 dollars: 29.24 %

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Delivery Order Mean Time

- The mean delivery time for this dataset is 24.16 minutes

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univariate Analysis: Most Frequent Customers Discount

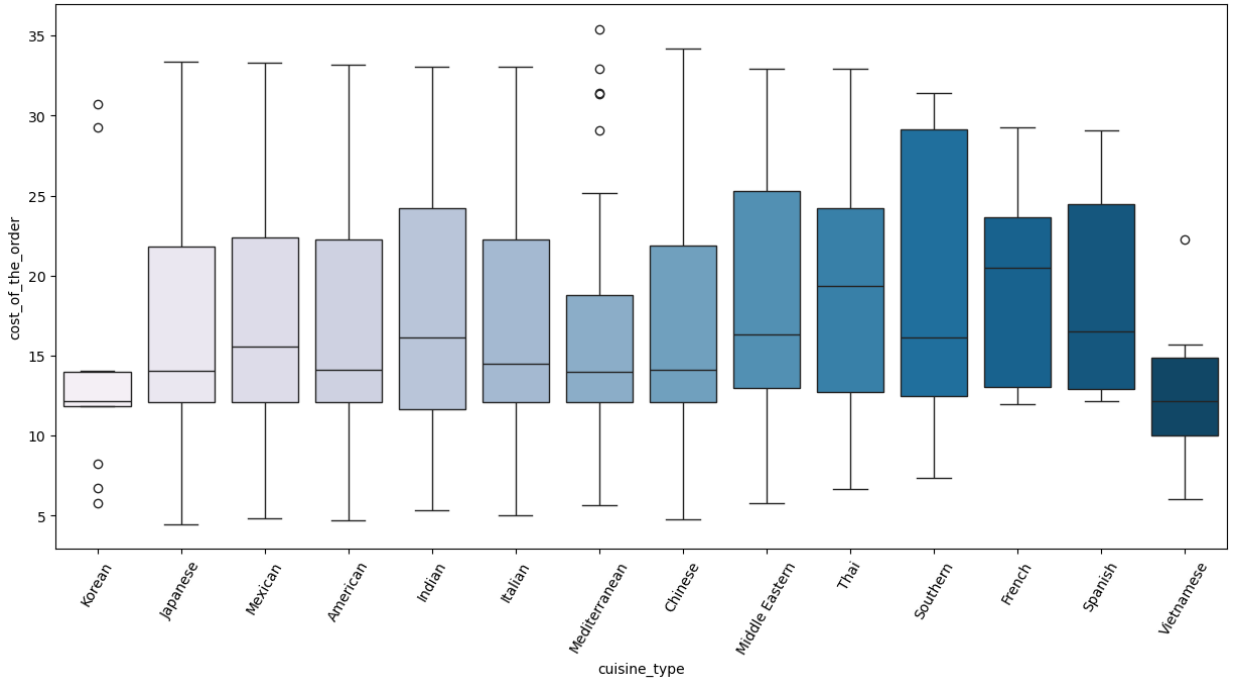
```
1  # Get the counts of  each customer_id
2  df['customer_id'].value_counts().head(3)
```

✓ 0.0s

```
customer_id
52832    13
47440    10
83287     9
Name: count, dtype: int64
```

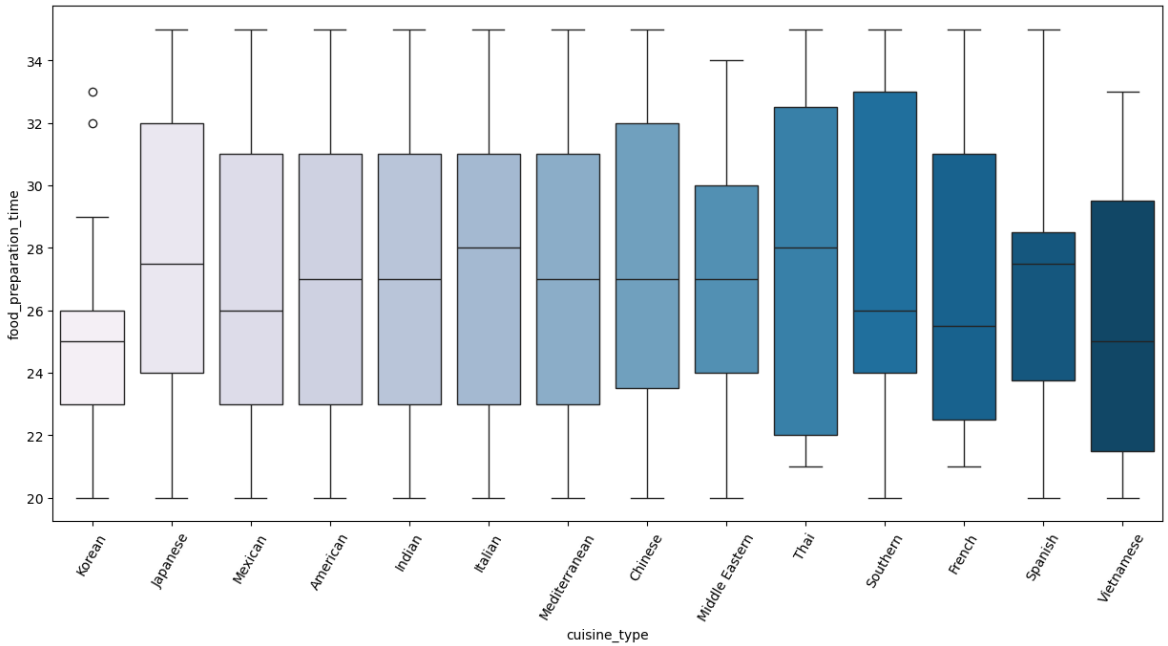
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Cuisine v. Cost of Order



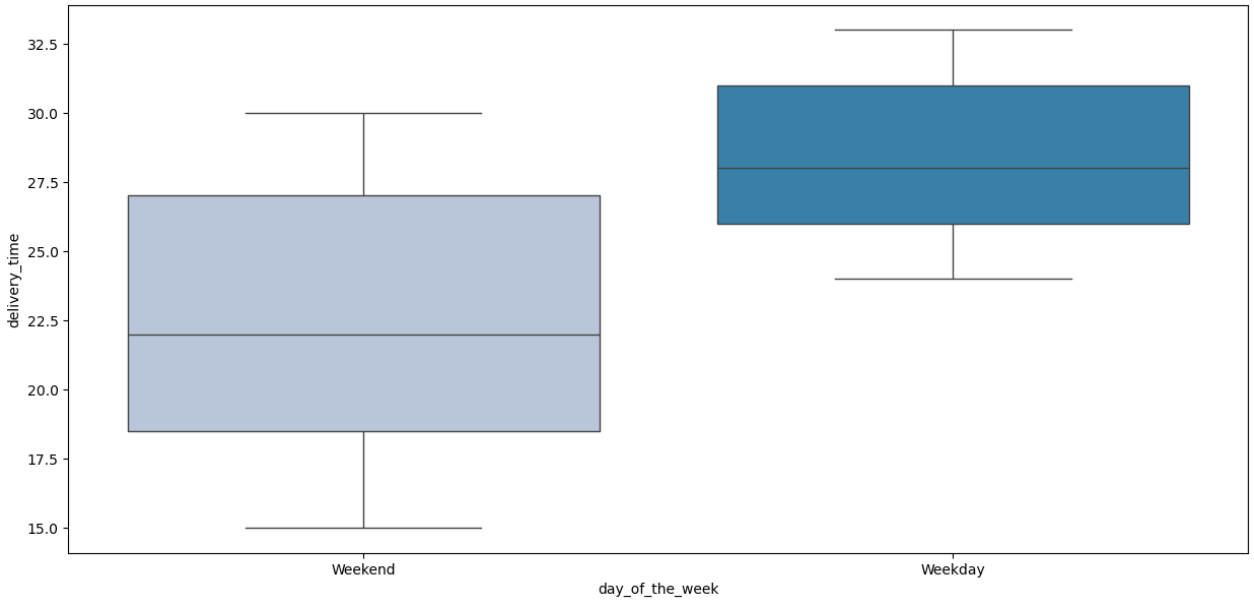
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Cuisine v. Food Preparation Time



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Day of Week v. Delivery Time



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

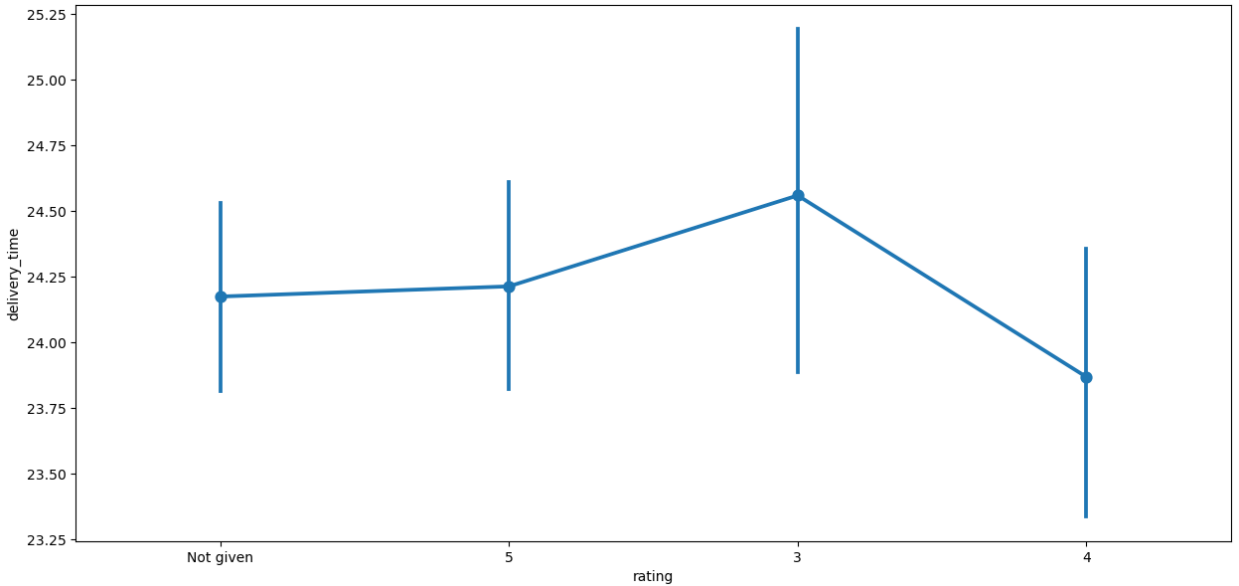
Multivariate Analysis: Revenue by Restaurant

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

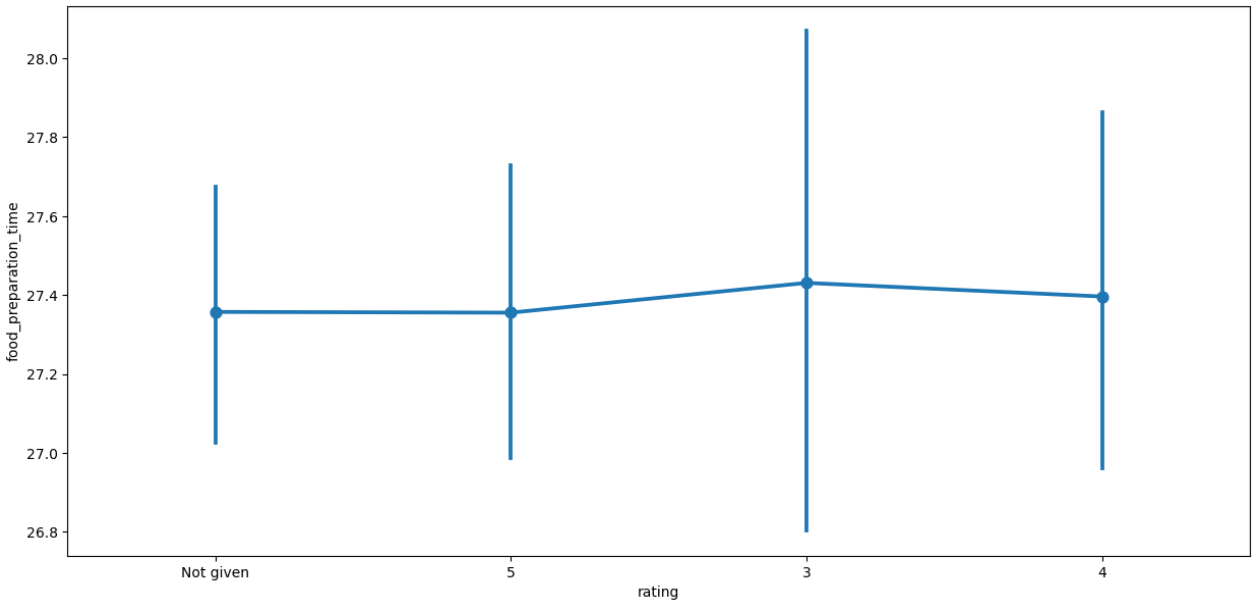
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Rating v. Delivery Time



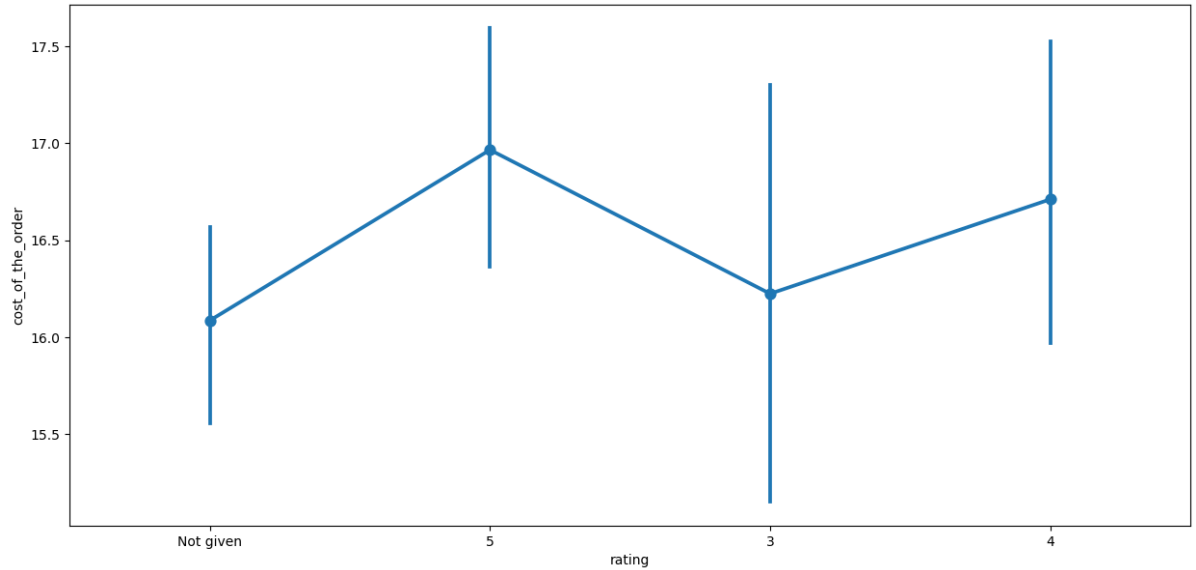
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Rating v. Food Preparation Time



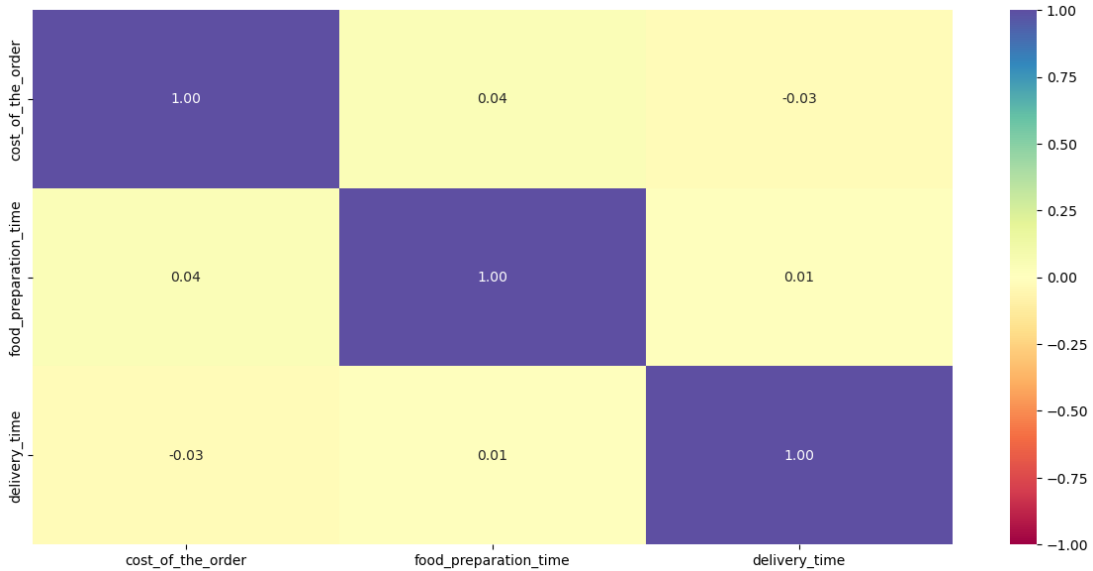
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Rating v. Cost of Order



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Numeric Variable Correlations



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Promotional Offer

- FoodHub 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

	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

	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

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Net Revenue

- The net revenue is around 31,314.82 dollars

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Total Delivery Time

- The number of orders where the total time is greater than 60 min is: 1898
- Percentage of orders that have more than 60 minutes of total delivery time is: 100.0 %

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multivariate Analysis: Delivery Time (Weekend v. Weekday)

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

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Conclusions

Customer Preferences and Patterns:

- The most popular cuisine type is American, followed by Japanese and Italian. These three cuisines account for the majority of the orders.
- Less popular cuisines include Vietnamese, Spanish, and Korean.

Order Cost Analysis:

- The average cost of an order is approximately $16.50, with a standard deviation of $7.48.
- Most orders (75%) cost between $4.47 and $22.30, with the highest cost recorded at $35.41.

Time Analysis:

- The average food preparation time is approximately 27.37 minutes, with a standard deviation of 4.63 minutes.
- The average delivery time is approximately 24.16 minutes, with a standard deviation of 4.97 minutes.
- There is a notable difference in delivery times between weekdays (average 28.34 minutes) and weekends (average 22.47 minutes).

Rating Insights:

- A significant number of orders (736) have no rating provided.
- Of the orders that have ratings, the majority are rated 5 or 4.
- The highest average ratings are observed in cuisines like Chinese and Japanese.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Reccomendations

Data-Driven Decisions:

- Continuously monitor and analyze customer preferences and operational metrics to adapt to changing trends and demands.
- Use the data to inform strategic decisions, such as menu adjustments, pricing strategies, and marketing initiatives.

Optimize Resources:

- Allocate resources effectively based on the data, ensuring adequate staffing during peak times and efficient use of delivery personnel.
- Implement technology solutions, such as route optimization software, to enhance delivery efficiency.

Customer Engagement:

- Engage with customers through personalized offers based on their order history and preferences.
- Implement loyalty programs to retain customers and encourage repeat orders.

Continuous Monitoring:

- Set benchmarks and goals for improvement in these areas, and review progress regularly to ensure continuous operational excellence.
- Regularly track key performance indicators (KPIs) such as preparation time, delivery time, and customer ratings.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.