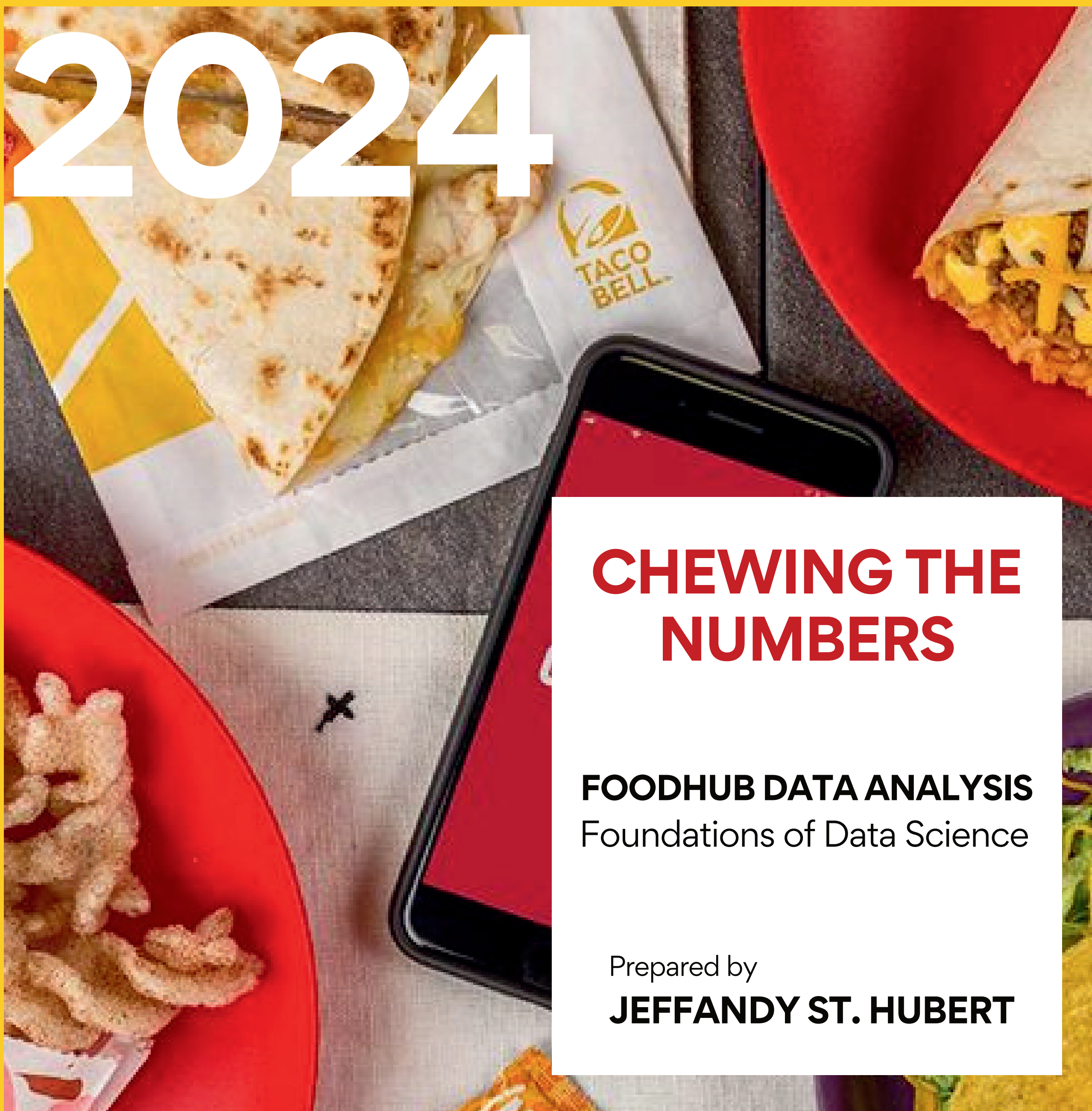
FOODHUB 2024



CHEWING THE NUMBERS

FOODHUB DATA ANALYSIS
Foundations of Data Science

Prepared by
JEFFANDY ST. HUBERT

Massachusetts
Institute of
Technology

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Among the 14 cuisine types, the most popular are:

1. AMERICAN
2. JAPANESE
3. ITALIAN
4. CHINESE
5. MEXICAN

CONCLUSIONS

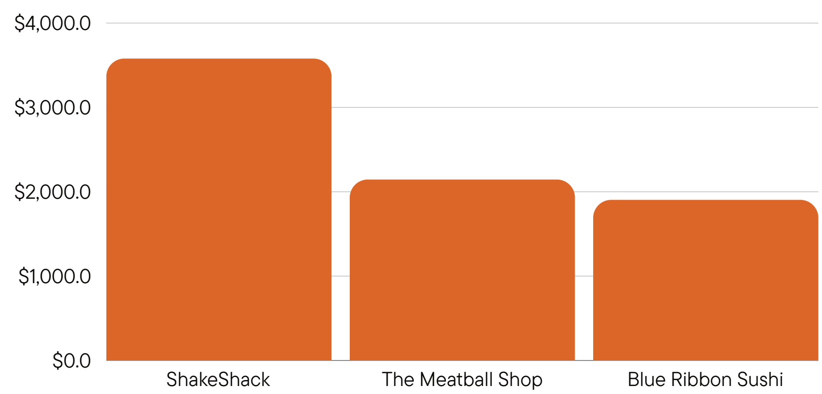
The total amount of orders during the weekend is **double** the total amount of orders during the week.

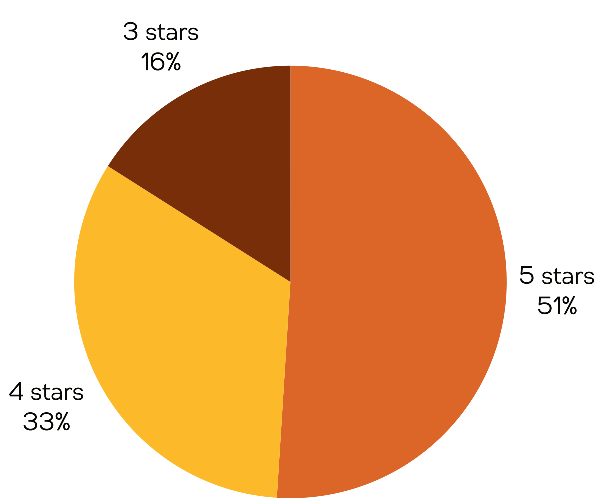
In terms of customer rating:

1. THE MEATBALL SHOP
2. BLUE RIBBON FRIED CHICKEN
3. SHAKE SHACK AND BLUE RIBBON SUSHI

Top Restaurants by Revenue:

Shake Shack	$3,579.53
The Meatball Shop	$2,145
Blue Ribbon Sushi	$1,903.95





Out of the 1,898 total orders in the data, 61% of the customers provided ratings. The distribution of these 1,162 ratings is as follows:

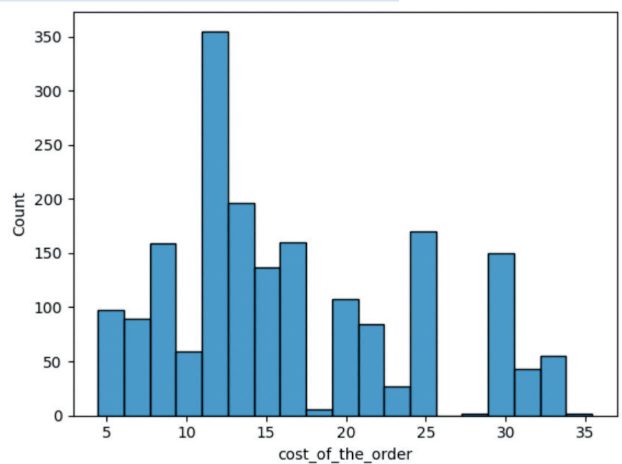
Customer Ratings Distribution:

5-stars:	51%
4-stars:	33%
3-stars:	16%

EXECUTIVE SUMMARY CONT.

RECOMMENDATIONS

REVISE THE PRICING STRUCTURE



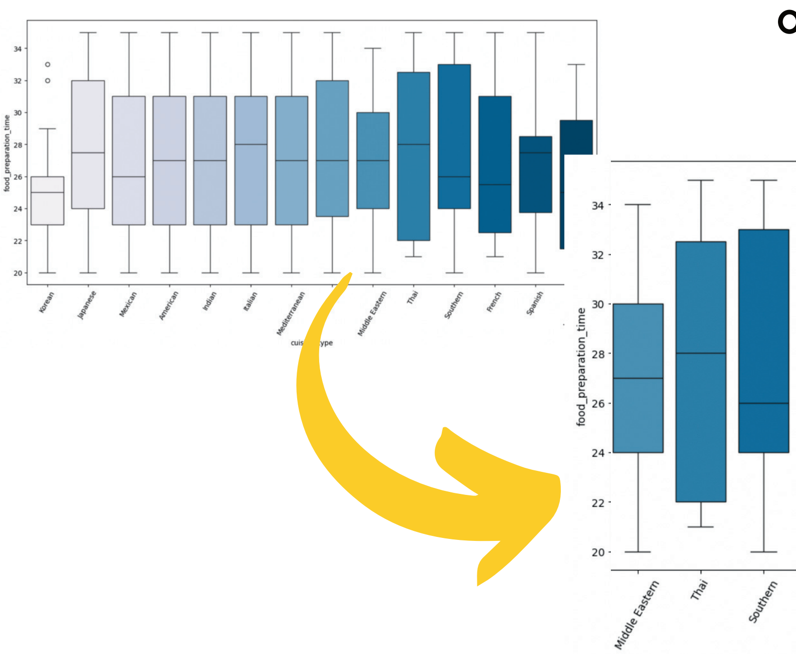
Divide the pricing structure into three segments

<u>Orders under $5</u>: Given the lower order value, consider a 25% flat commission rate that ensures profitability without discouraging restaurants from offering lower-cost items.

<u>Orders between $5 and $20:</u> Given the findings that the bulk of orders fall within this range, a competitive rate that incentivizes more orders could be 20%.

<u>Orders over $20:</u> Given the higher order value, the commission rate should be reduced to 15% to incentivize restaurants to offer higher-cost items and to increase company revenue.

IMPROVE OPERATIONS



Operational Improvements:

- **Preparation Time Management:** Streamline the preparation process, especially for ***time-intensive cuisines like Thai*** to reduce wait times and improve overall service efficiency.
- **Delivery Efficiency**: Focus on improving delivery times to enhance customer experience, by optimizing delivery routes or partnering with more efficient couriers.

BUSINESS PROBLEM AND SOLUTION

PROBLEM

With a large increase in New York food merchants, and food aggregators like Grubhub, Seamless, and Uber Eats, the market is very competitive and Foodhub desires to raise the bar. Fortunately, there is also an increase in students and professionals moving into the city. If restaurant and customer satisfaction increase, Foodhub will be the preferred choice for online food delivery service.

SOLUTION

- Create initiatives for restaurants to earn more.
- Draw more customers to the Foodhub app through promotions and their favorite cuisines.

DATA OVERVIEW

THE SHAPE OF THE DATA

There are 1898 rows and 9 columns in the data.

```
<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
```

```
df.replace('Not given', np.nan, inplace=True) # Replacing 'Not given' with NaN
```

ALTHOUGH THERE WERE NO NULL VALUES, THERE WERE 736 ‘NO RATINGS.’

```
order_id: 0 null values
customer_id: 0 null values
restaurant_name: 0 null values
cuisine_type: 0 null values
cost_of_the_order: 0 null values
day_of_the_week: 0 null values
rating: 736 null values
food_preparation_time: 0 null values
delivery_time: 0 null values
```

DATA OVERVIEW

MIN, MAX AND MEAN OF THE DATA

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
order_id	1898.0	NaN	NaN	NaN	1477495.5	548.049724	1476547.0	1477021.25	1477495.5	1477969.75	1478444.0
customer_id	1898.0	NaN	NaN	NaN	171168.478398	113698.139743	1311.0	77787.75	128600.0	270525.0	405334.0
restaurant_name	1898	178	Shake Shack	219	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cuisine_type	1898	14	American	584	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cost_of_the_order	1898.0	NaN	NaN	NaN	16.498851	7.483812	4.47	12.08	14.14	22.2975	35.41
day_of_the_week	1898	2	Weekend	1351	NaN	NaN	NaN	NaN	NaN	NaN	NaN
rating	1898	4	Not given	736	NaN	NaN	NaN	NaN	NaN	NaN	NaN
food_preparation_time	1898.0	NaN	NaN	NaN	27.37197	4.632481	20.0	23.0	27.0	31.0	35.0
delivery_time	1898.0	NaN	NaN	NaN	24.161749	4.972637	15.0	20.0	25.0	28.0	33.0

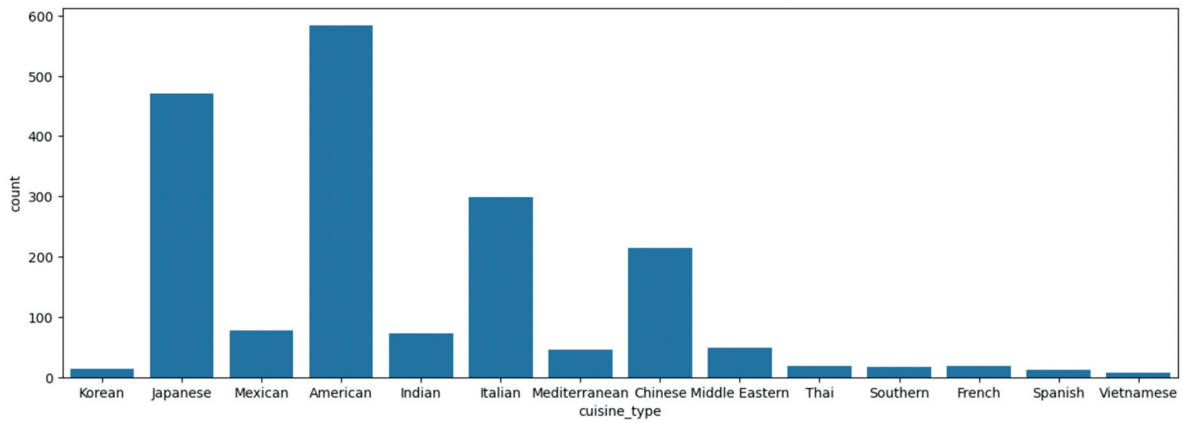
AMOUNT OF ORDERS RATED

```
df.fillna('Not given', inplace=True)
#I know the answer is 736 from above but lets check again
df['rating'].value_counts()
```

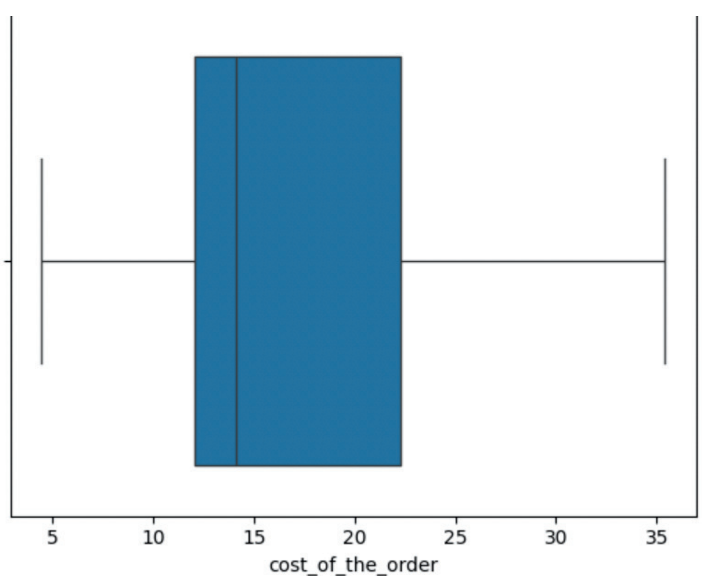
```
rating
Not given    736
5            588
4            386
3            188
Name: count, dtype: int64
```

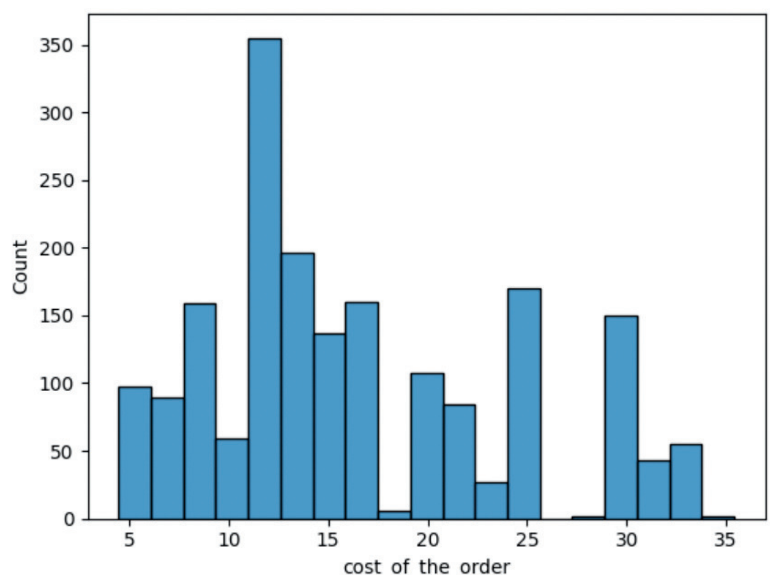
UNIVARIATE ANALYSIS

CUISINE TYPE



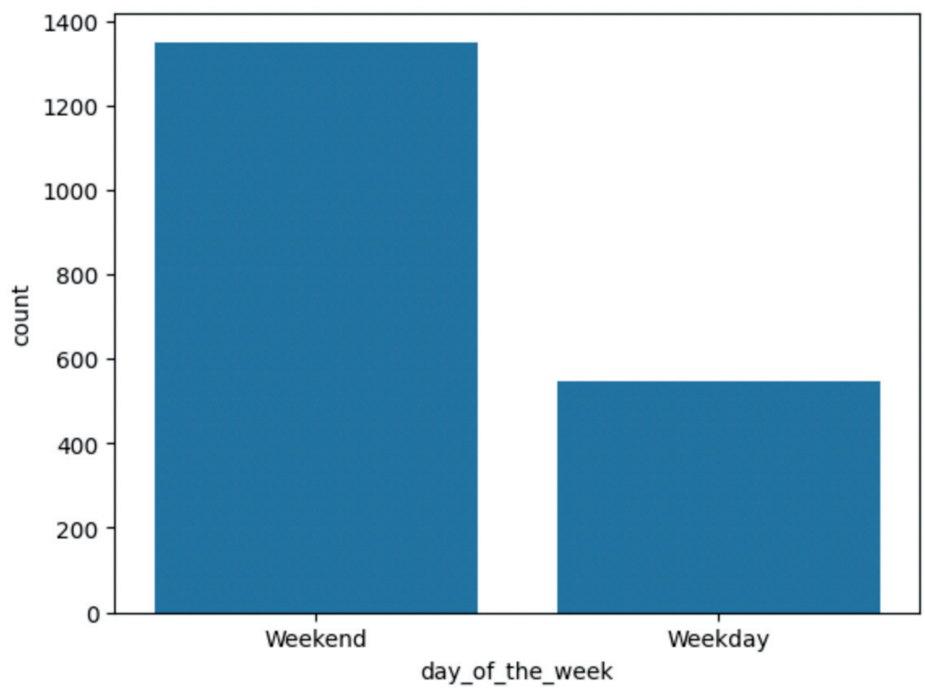
COST



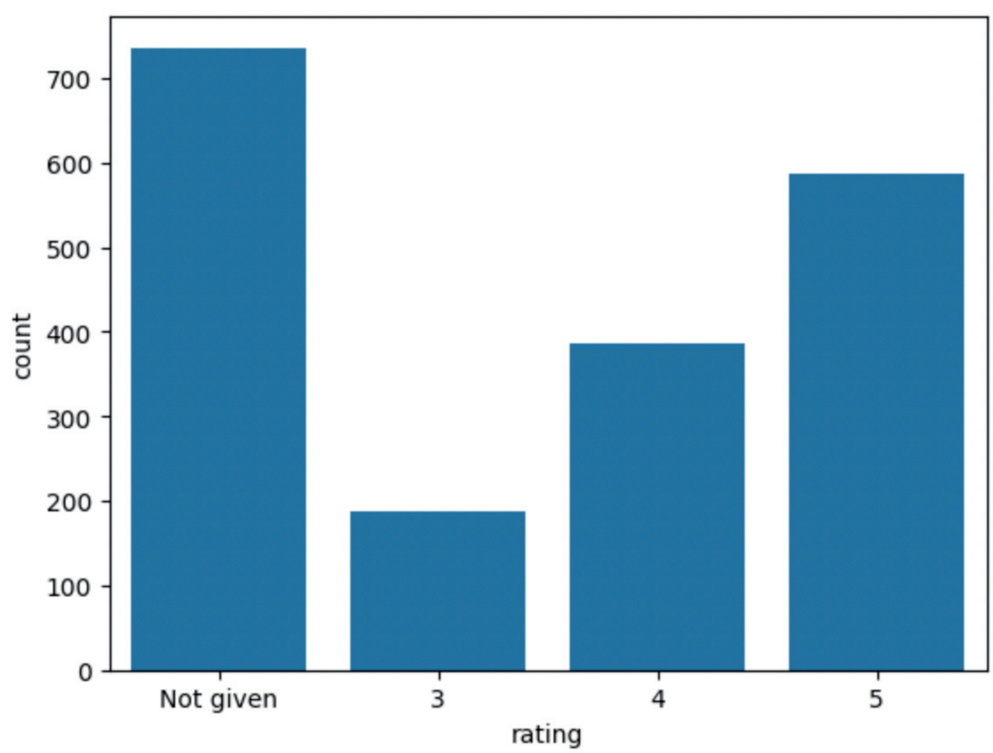


UNIVARIATE ANALYSIS CONT.

DAY OF THE WEEK

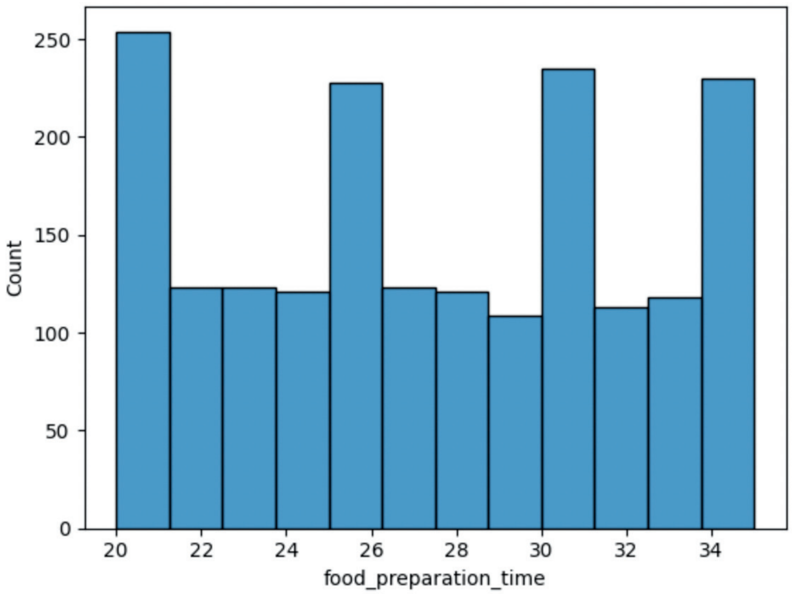


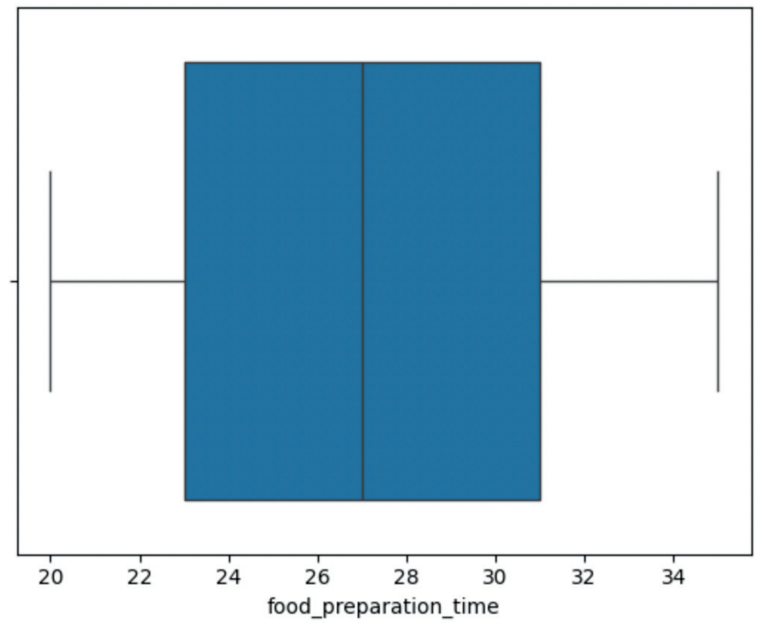
RATING



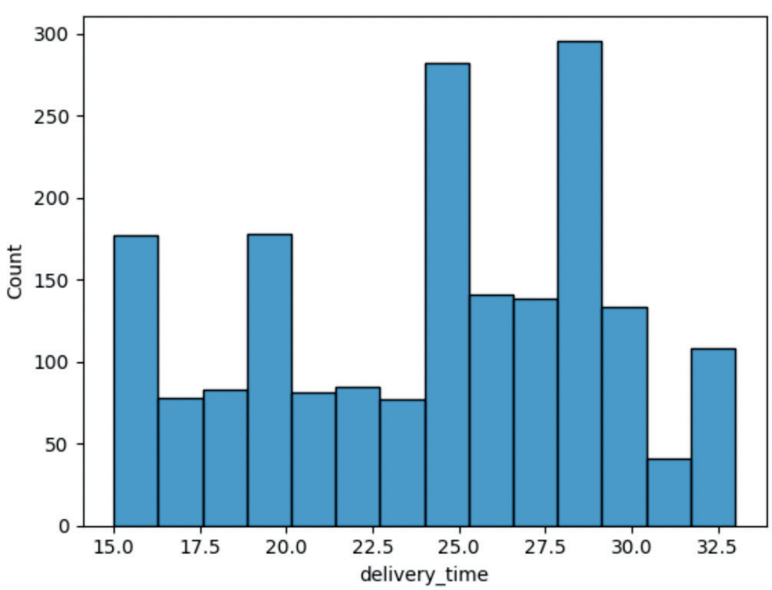
UNIVARIATE ANALYSIS CONT.

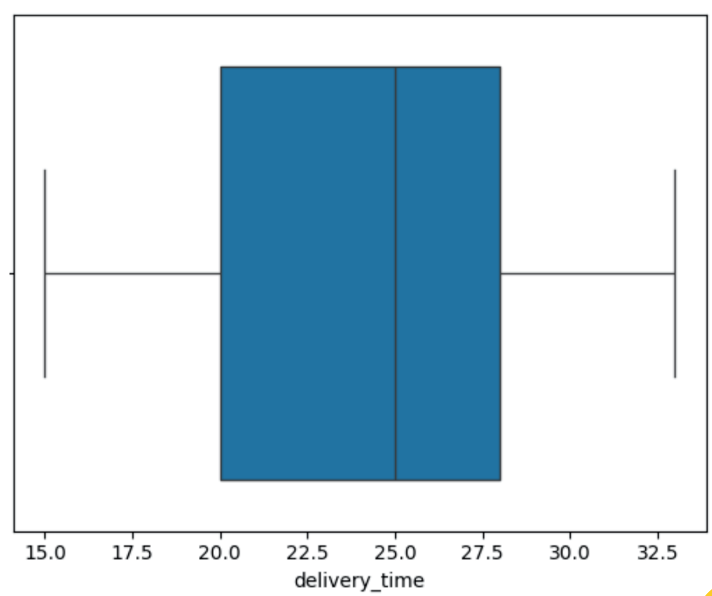
FOOD PREPARATION TIME





DELIVERY TIME

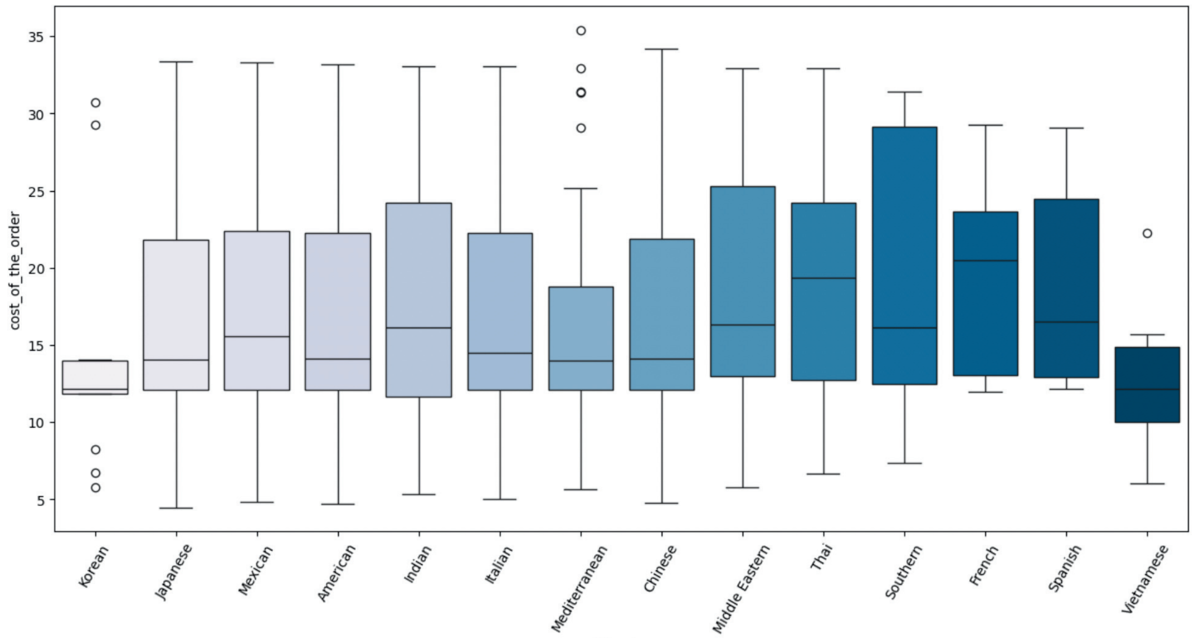




MULTIVARIATE ANALYSIS

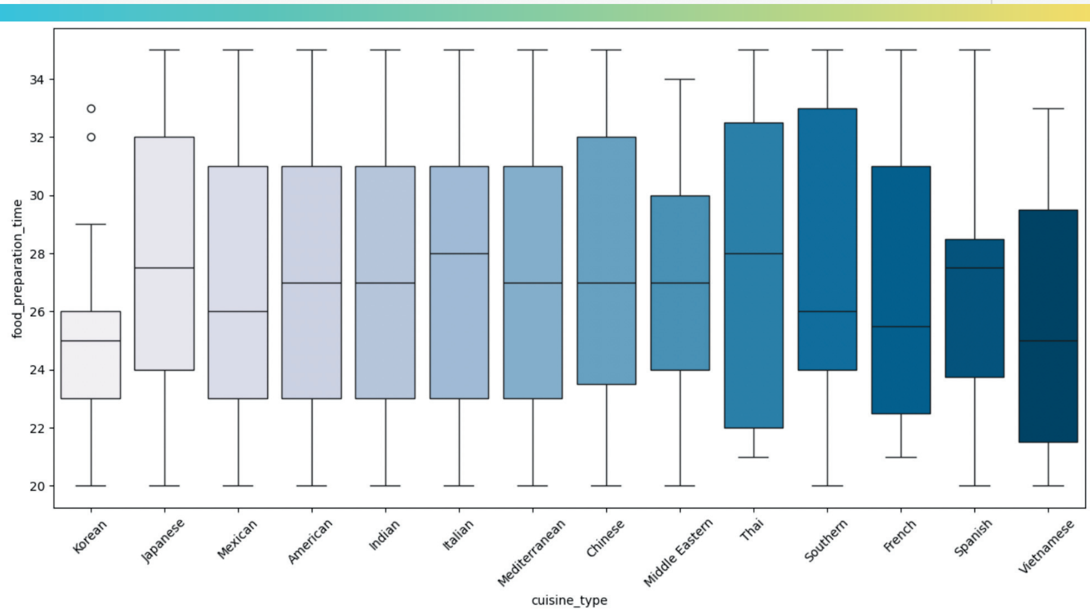
CUISINE VS ORDER COST

```
# Relationship between cost of the order and cuisine type
plt.figure(figsize=(15,7))
sns.boxplot(x = "cuisine_type", y = "cost_of_the_order", data = df, palette = 'PuBu')
plt.xticks(rotation = 60)
plt.show()
```



CUISINE VS FOOD PREPARATION TIME

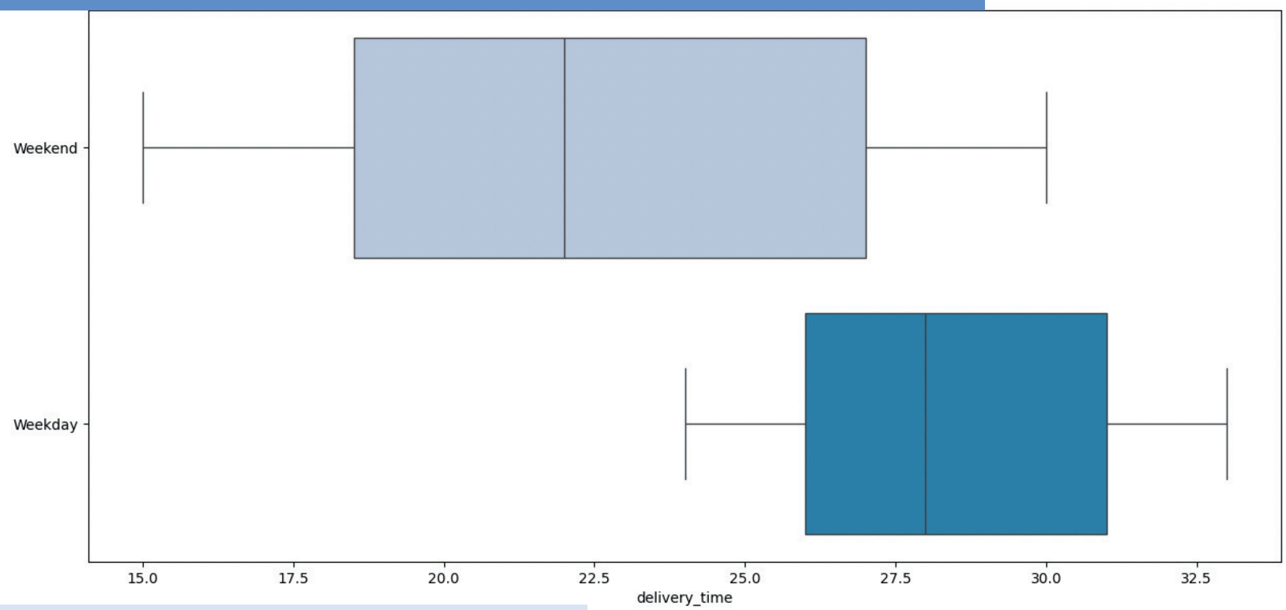
```
[29] # Relationship between food preparation time and cuisine type
     plt.figure(figsize=(15,7))
     sns.boxplot(x = "cuisine_type", y = "food_preparation_time", data = df, palette = 'PuBu')
     plt.xticks(rotation = 45)
     plt.show()
```



MULTIVARIATE ANALYSIS

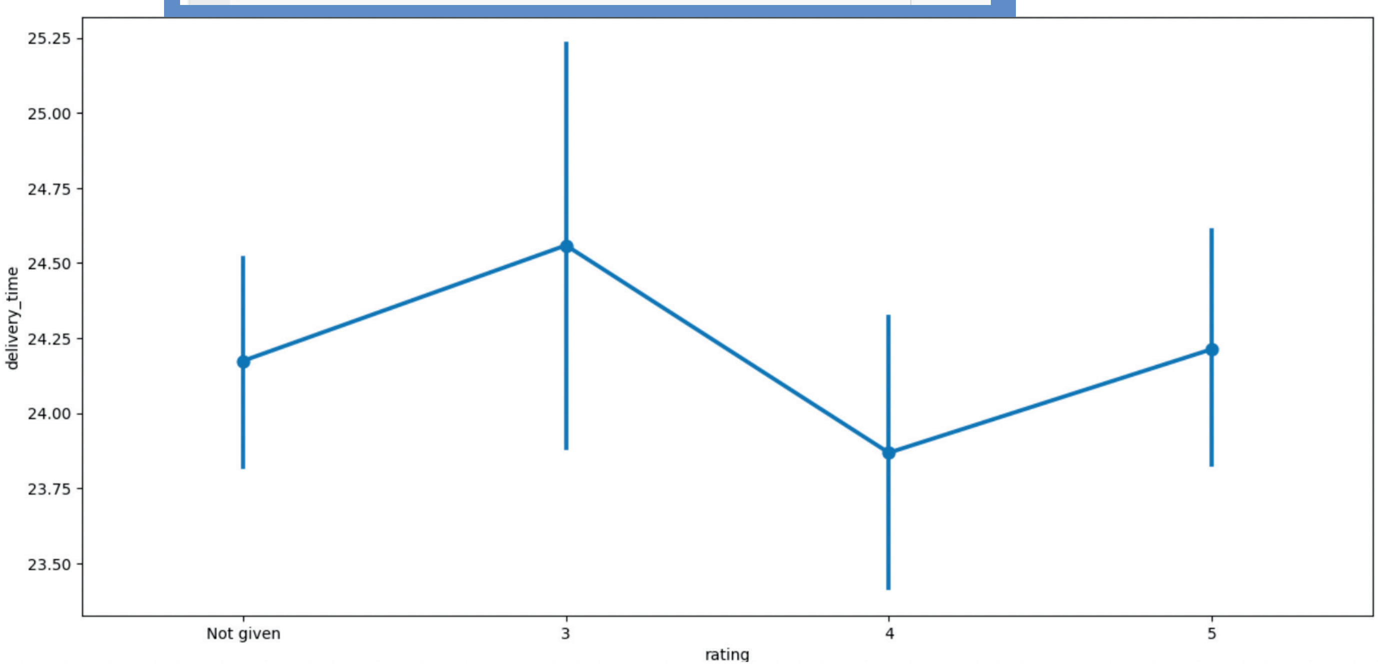
DAY OF THE WEEK VS DELIVERY TIME

```
# Relationship between day of the week and delivery time
plt.figure(figsize=(15,7))
sns.boxplot(x = "delivery_time", y = "day_of_the_week", data = df, palette = 'PuBu')
plt.show()
```



RATING VS DELIVERY TIME

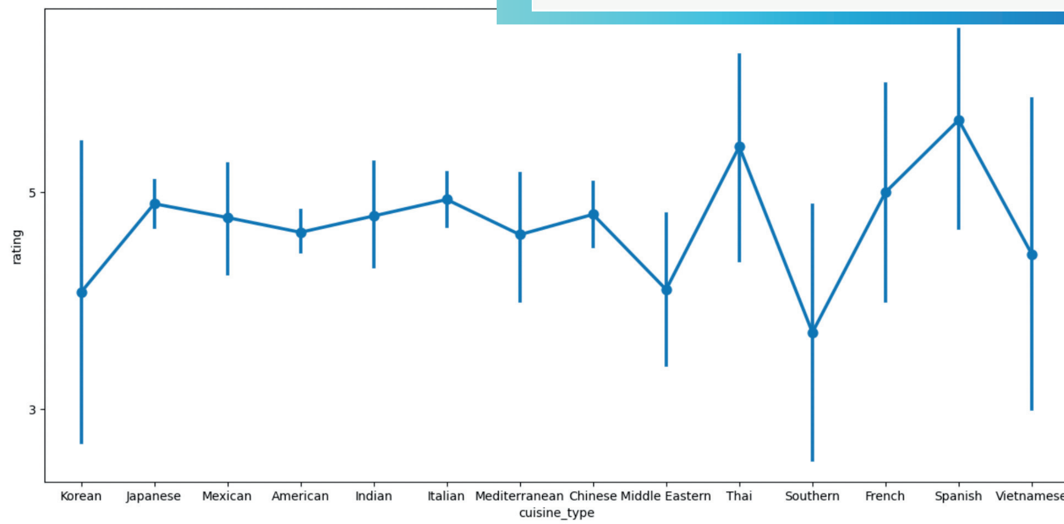
```
# Relationship between rating and delivery time
plt.figure(figsize=(15, 7))
sns.pointplot(x = 'rating', y = 'delivery_time', data = df, order=['Not given', 3, 4, 5])
plt.show()
```



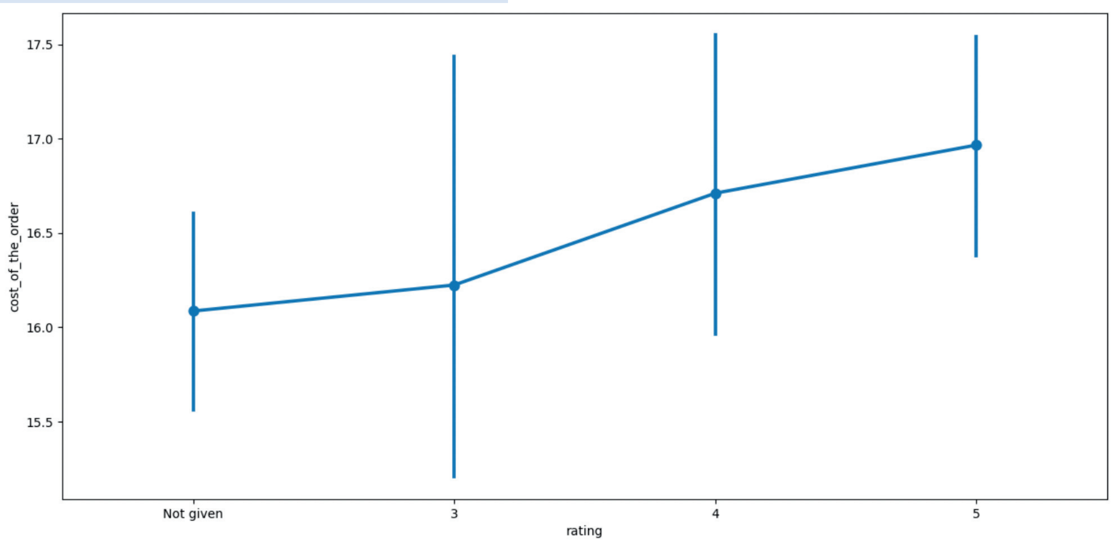
MULTIVARIATE ANALYSIS

RATING VS CUISINE TYPE

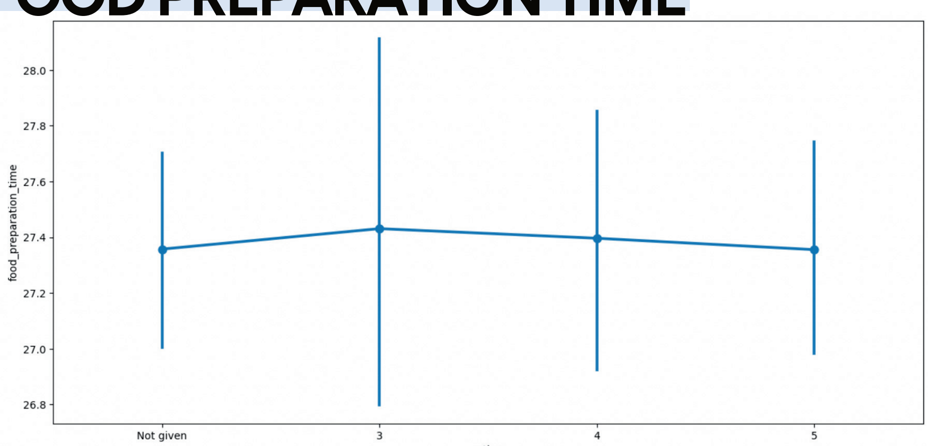
```
[ ] # Relationship between rating and cuisine type
    plt.figure(figsize=(15, 7))
    sns.pointplot(x = 'cuisine_type', y = 'rating', data = df) #order=['Not given', 3, 4, 5])
    plt.show()
```



RATING VS ORDER COST

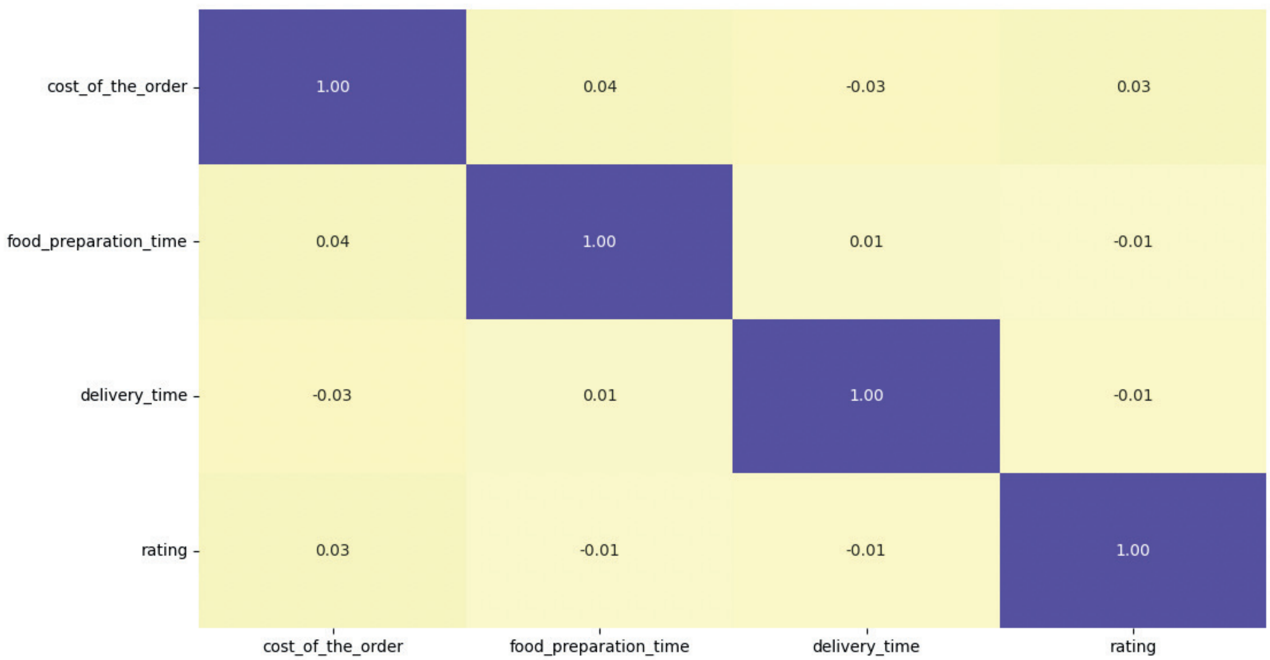


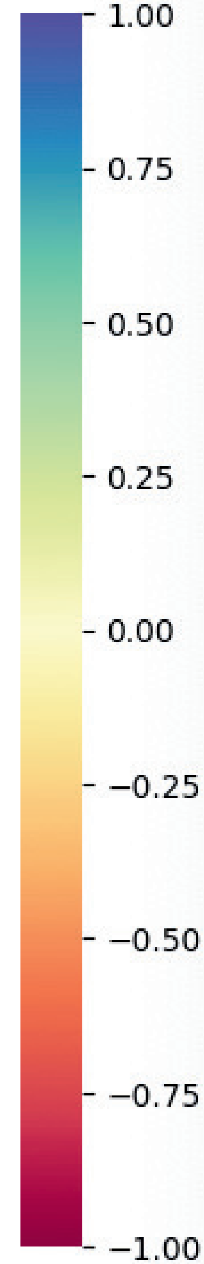
RATING VS FOOD PREPARATION TIME

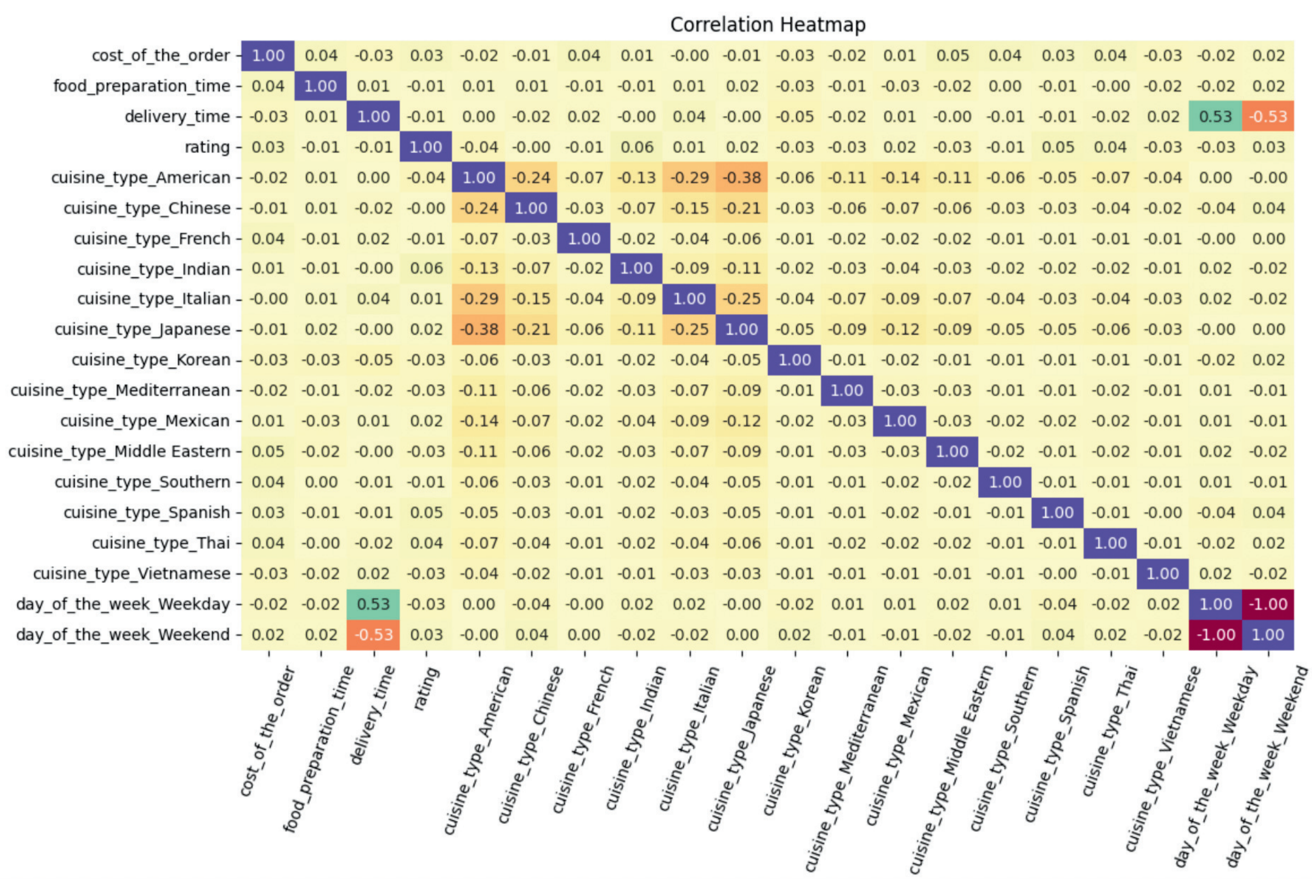


MULTIVARIATE ANALYSIS

HEAT MAP







THANK YOU

MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Massachusetts Institute of Technology

Great Learning

POWER AHEAD