

FoodHub Data Analysis

FoodHub (PGP_AIML_BA_UTA_JAN24_B)

Mar 8, 2024

Contents / Agenda

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

Executive Summary

FoodHub's data analysis reveals the insights into popular cuisines and their correlation with ratings, FoodHub can tailor promotions and enhance partnerships to drive sales and optimize customer satisfaction. It also reveals the top restaurants by order frequency and revenue, emphasizing the potential for strategic partnerships and targeted promotions to drive sales. Insights into customer preferences and delivery efficiency highlight opportunities for growth through expanded offerings and optimized operations.

Conclusions:

Top Restaurants and Cuisines:

- Top 5 restaurants in terms of orders placed and revenue are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken and Parm.
- American, Japanese, Italian, and Chinese cuisines emerge as the most popular choices among customers.

Order Characteristics:

- French cuisine has the highest average order cost, while Korean and Vietnamese cuisines have the lowest.
- Average food preparation time across cuisines ranges from 25 to 28 minutes.
- Delivery time tends to be longer on weekdays compared to weekends.

Growth Opportunities:

- With only 1200 unique customers and 178 restaurants, there is significant room for growth in both customer base and restaurant partnerships.
- Increased orders on weekends present an opportunity for targeted promotions to boost weekday sales and optimize weekend pricing for higher revenue.
- There are more than 700 customers who did not give rating, reflects need to have strong customer feedback system.

Customer Engagement:

- Targeting top customers with personalized promotions based on their preferred cuisines can potentially increase repeat orders.
- Improving the customer feedback process can help gather more data and improve service quality.

Executive Summary

Recommendations:

FoodHub can optimize its operations, attract more customers, and enhance revenue generation while improving overall customer satisfaction and service quality by implementing below recommendations.

Customer Acquisition Strategy:

- Develop a strategy to increase the customer base, especially given the current count of 1200 customers.
- Expand outreach efforts to attract new customers, leveraging various digital channels and promotional campaigns.

Operational Efficiency:

- Restaurants should focus on reducing delivery time on weekdays to improve efficiency.
- Enhance food preparation and delivery efficiency to ensure more orders are delivered within 60 minutes.
- Provide an option to give 'rating' on previously ordered food delivery service, before placing new order, along with post food delivery, to increase chance of getting 'ratings'.
- As Average of the orders are delivered within 60 mins, working with Restaurants for faster food preparation time, including new cuisine type(Variety of Menu) to save food preparation time and optimize faster delivery.

Expansion and Marketing:

- Increase the number of restaurant partnerships, especially with establishments offering American cuisine, to meet high demand.
- Run promotions and concentrate on marketing efforts to increase the customer base and revenue.
- Add additional restaurants offering popular cuisines to diversify offerings and attract more customers.

Promotional Strategies:

- Run deals and promotions for weekdays to increase sales, while charging premium prices on weekends to maximize revenue.
- Offer coupons and incentives for popular restaurants on weekends to capitalize on increased orders.
- Based on the number of repetitive orders data, business would have to give discounts and coupons to existing customers on the basis of previous orders , to increase repetitive sales.

Business Problem Overview and Solution Approach

Business Problem Overview :

- FoodHub, a leading Food Aggregator Company, operates in New York City's online food delivery market.
- Customers utilize FoodHub's smartphone app to place food orders from their preferred restaurants and rate their delivery experience.
- FoodHub's revenue model involves earning a fixed margin on delivery orders collected from partnered restaurants.
- The operational process includes assigning delivery personnel to collect orders from restaurants and deliver them to customers' locations.
- The company aims to optimize operations and enhance customer experience amidst fierce competition in the online food delivery market.

Solution Approach :

- To address the operations optimization and customer experience challenges and capitalize on opportunities for improvement, FoodHub intends to leverage historical customer data for univariate and multivariate data analysis
- Identify popular restaurants and cuisine types to optimize resource allocation.
- Assess food preparation time and delivery operations for efficiency and customer satisfaction.
- Perform peak order analysis on weekdays vs weekends
- Reward repeat customers
- Analyze order cost, ratings and popular cuisine types and restaurants for revenue analysis and optimization

Data Overview

The provided dataset for the Food Hub project contains information about food orders placed by their customer. Below data fields allow for comprehensive analysis of customer behavior, restaurant popularity, service quality, and operational efficiency within the online food delivery platform.

Order ID:

- Unique identifier for each food order placed through the Food Hub platform.
- Helps in tracking individual orders and identifying any patterns or trends.

Customer ID:

- Identification number assigned to each customer who places food orders.
- Allows for analyzing customer behavior, preferences, and repeat orders.

Restaurant Name:

- Name of the restaurant from which the food order is placed.
- Provides insights into the popularity of different restaurants among customers.

Cuisine Type:

- Type of cuisine ordered by the customer.
- Helps in understanding the demand for various types of food and preferences among customers.

Cost of the Order:

- Monetary value of the food order.
- Indicates the revenue generated from each order and allows for analyzing spending patterns of customers.

Day of the Week:

- Indicates whether the order was placed on a weekday or weekend.
- Helps in identifying any differences in ordering behavior between weekdays and weekends.

Rating:

- Rating given by the customer for their overall experience with the food delivery.
- Reflects customer satisfaction levels and can be used to assess service quality.

Food Preparation Time:

- Time taken by the restaurant to prepare the food order. This is calculated by taking the difference between the timestamps of the restaurant's order confirmation and the delivery person's pick-up confirmation.
- Provides insights into the efficiency of restaurants in fulfilling orders and helps in managing delivery expectations.

Delivery Time:

- Time taken by the delivery person to deliver the food package to the customer. This is calculated by taking the difference between the timestamps of the delivery person's pick-up confirmation and drop-off information.
- Allows for evaluating delivery performance and efficiency of the delivery process.

Data Overview

Question 1: How many rows and columns are present in the data?

Answer: There are total 1898 rows and 9 columns present in data. This info is obtained using `df.shape()` function to get the answer.

Question 2: What are the datatypes of the different columns in the dataset?

Answer: The dataset comprises three distinct data types: integer (int64), object, and floating-point (float64).

The data types for each column are as follows:

- **Integer (int64):** Columns: 'order_id', 'customer_id', 'food_preparation_time', 'delivery_time'
- **Object:** Columns: 'restaurant_name', 'cuisine_type', 'day_of_the_week', 'rating'
- **Float (float64):** Column: 'cost_of_order'

This information is obtained using the `df.info()` function to provide an overview of the data types for each column in the dataset.

Question 3: Are there any missing values in the data? If yes, treat them using an appropriate method.

Answer: There are no missing values observed in the dataset when checked the sum of null values for each column using `df.isnull().sum()`. However the 736 orders had the value 'Not given' and as per data dictionary "ratings are given by the customers out of 5", so 'rating' column can be treated by replacing 'Not given' with NaN.

Data Overview

Question 4: Check the statistical summary of the data. What is the minimum, average, and maximum time it takes for food to be prepared once an order is placed?

Answer: Statistical summary of the data is checked using `df.describe(include='all').T`.

Food preparation time observations:

- The minimum food preparation time is 20 minutes
- The average food preparation time is 27.37 minutes
- The maximum food preparation time is 35 minutes

Other Observations from dataset:

- The top 'restaurant_name' is Shake Shack.
- Total 178 unique 'restaurant_names' been observed in the dataset.
- The top 'cuisine_type' is American & 14 unique 'Cuisine_type' are observed in the dataset.
- The top 'day_of_the_week' is 'Weekend', means more number of orders been placed on weekend vs weekday.
- There are 4 unique values observed for 'rating' in the dataset.

Note - Code added in Appendix

Question 5: How many orders are not rated?

Answer: There are 736 Orders which are 'Not Given', out of total orders 1898.

This answer is obtained using `df.value_counts()`

EDA - Univariate Analysis

Question 6: Explore all the variables and provide observations on their distributions. (Generally, histograms, box plots, count plots, etc. are used for univariate exploration.)

Answer: Please find all below Observations on various columns in the dataset for univariate analysis -

Order ID:

- There are a total of 1898 unique orders, which matches the no of entries in dataset, that means each order is unique entry.
- This column can be used as primary key in database or table.

Customer ID:

- There are total of 1200 unique customers and there are 1898 unique orders, which means there are some customers who ordered food more than once.
- Top 20 customers who has ordered more than once.

Restaurant Name:

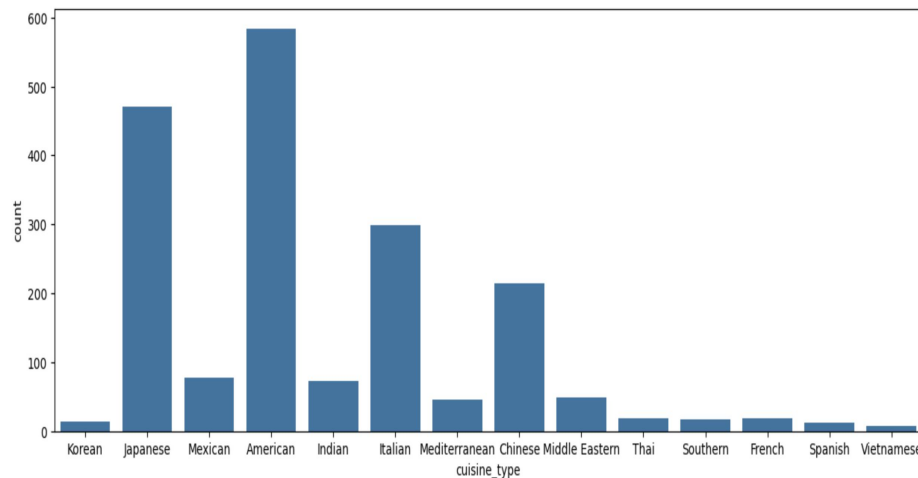
- There are 178 different restaurants from which customers have placed orders.
- There are special characters in the Restaurant Names, for example 'Big Wong Restaurant \x8c_Ñ¼' and 'Joe's Shanghai \x8e_À\x8eü£¼÷'

Note - Code added in Appendix

EDA - Univariate Analysis

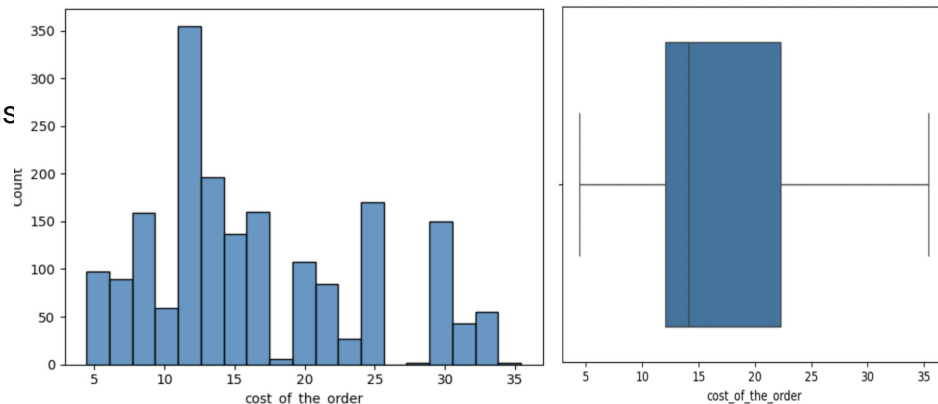
Cuisine Type:

- There are total 14 cuisine types.
- The most popular cuisine type is American as its displaying highest count of orders in graph below.
- The least popular cuisine type is Vietnamese, as it's with least orders.
- There are 4 cosine types which are popular (American, Japanese, Italian, Chinese) amongst all cuisine types.



Cost of the Order:

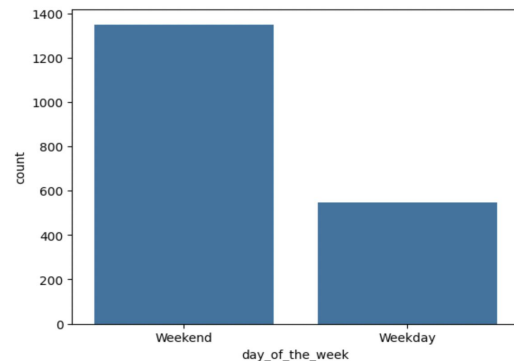
- Average 'cost of the order' is approx 14 dollars.
- The 'cost of the order' data is right skewed, means most of the orders are of low cost & few orders are of high cost.
- Approx 25% of orders are between 12 to 22 dollars.
- There are less orders between 22 to 35 dollars.



EDA - Univariate Analysis

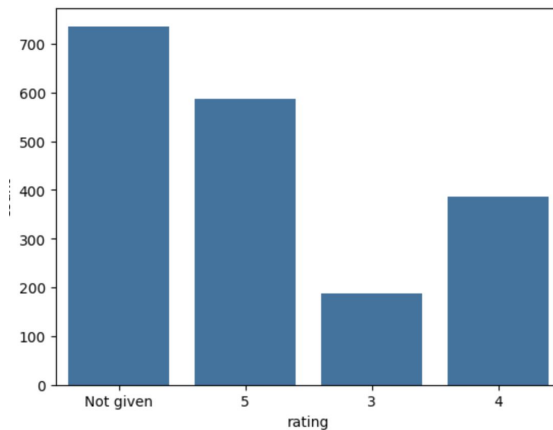
Day of the Week:

- The unique values in 'day_of_the_week' columns are 'Weekend' & 'Weekday'.
- The Order count is more on Weekend compared to Weekdays.



Rating:

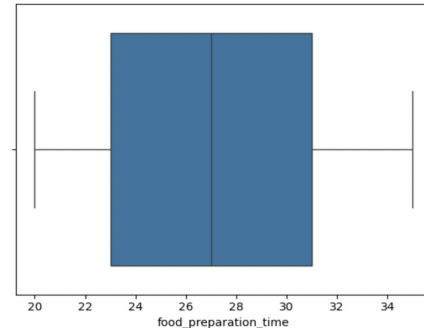
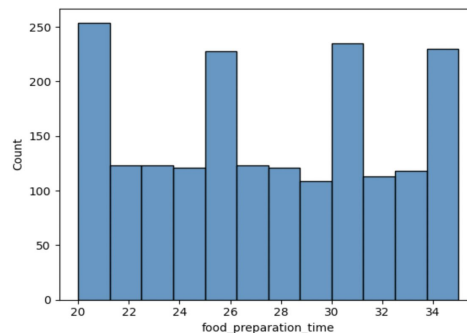
- There are 4 unique values displayed in rating column - Not given, 5, 3, 4
- Highest rating given is 5.(approx 500 orders)
- Lowest rating given is 3 (approx 200 orders)
- Most of the Orders has not been rated (more than 700 orders), 'Not given' bar is the highest bar.
- Approx 400 orders got 4 rating.



EDA - Univariate Analysis

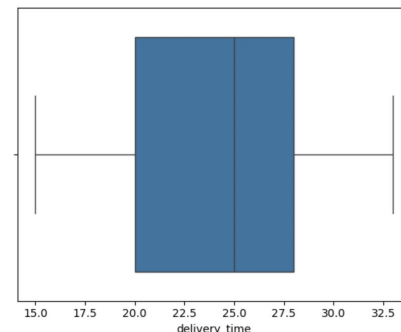
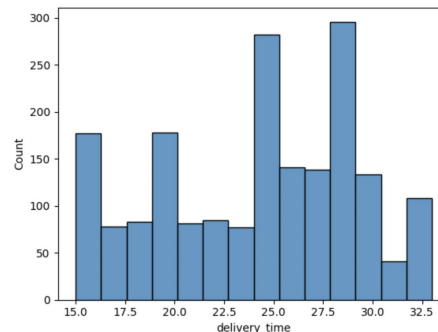
Food Preparation Time:

- Minimum food preparation time is approx 20 min.
- Average food preparation time is 27 min.
- Max food preparation time is approx 35 min.
- Approx 120 orders took same food preparation time.
- Approx 250 orders took 20 min as 'food preparation time'



Delivery Time:

- The Minimum delivery time taken is approx 15 min.
- The Maximum delivery time taken is approx 32.5 min.
- Average delivery time is approx 25 min.
- Approx 50% orders took 25 min of 'delivery time'.
- Approx 75% of orders took more than 28 min of 'delivery time'
- delivery time data is left skewed



EDA - Multivariate Analysis

Question 7: Get top 5 restaurants with highest number of orders.

Answer - Top 5 restaurants with highest number of orders respectively are as listed below.

(This information is obtained using function `.value_counts().head(5)`)

Restaurant Name	Orders
Shake Shack	219
The Meatball Shop	132
Blue Ribbon Sushi	119
Blue Ribbon Fried Chicken	96
Parm	68

EDA - Multivariate Analysis

Question 8: Which is the most popular cuisine on weekends?

Answer - The most Popular cuisine type is 'American'

(This information is obtained using `df_popular_cuisine = df.groupby(['cuisine_type','day_of_the_week'])['day_of_the_week'].apply(lambda x: (x=='Weekend').sum())`
`most_popular_cuisine = df_popular_cuisine[df_popular_cuisine == df_popular_cuisine.max()].index[0][0]`)

Question 9 : What percentage of the orders cost more than 20 dollars ?

(Calculated orders above 20 & then took percentage of orders above 20 dollars)

Answer - The number of total orders that cost above 20 dollars are 555.

Percentage of orders above 20 dollars is 29.24%

Note - Code added in Appendix

Question 10: What is the mean order delivery time?

Answer - The mean delivery time for this dataset is 24.16 minutes.

(This information obtained using function `.mean()` of delivery time of all orders.)

EDA - Multivariate Analysis

Question 11: The company has decided to give 20% discount vouchers to the top 3 most frequent customers. Find the IDs of these customers and the number of orders they placed.

Answer-

Below are top 3 most frequent 'customer_id's & number of orders they placed who will get 20% discount vouchers by company. (Obtained this data using `df['customer_id'].value_counts().head(3)`)

Customer IDs	# of Orders
52832	13
47440	10
83287	9

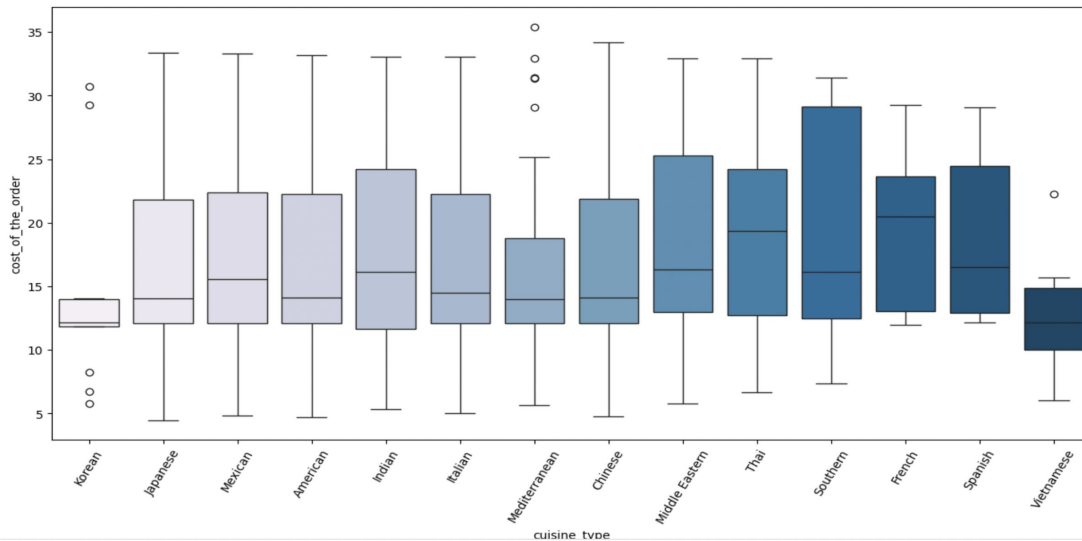
EDA - Multivariate Analysis

Q12 - Perform a multivariate analysis to explore relationships between the important variables in the dataset.^D

1. Cuisine Type Vs Cost of Order -

Observations:

- Mediterranean, Korean, Vietnamese cuisine type has some outliers.
- Vietnamese cuisine is least ordered cuisine type & cost of the order can go as low as \$ 5
- Average order cost for Vietnamese cuisine has the lowest cost of order compared to other cuisine type
- The top five cuisines in terms of high cost of order are Thai, Middle Eastern, Spanish, Southern, and French

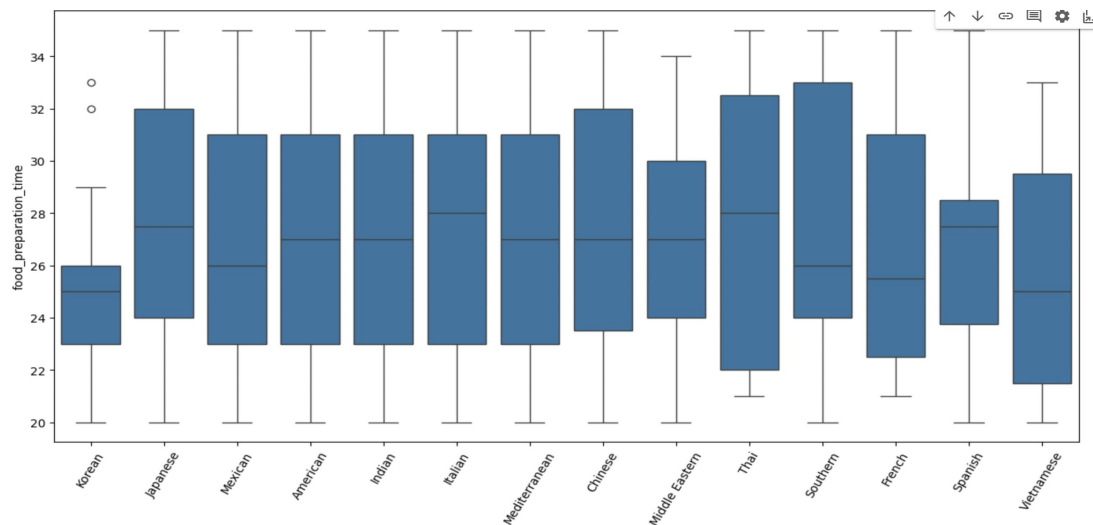


EDA - Multivariate Analysis

2. Cuisine type Vs Food Preparation Time -

Observations:

- Mexican, American, Indian, Italian, mediterranean have same IQR.
- 'Korean' cuisine has less orders, also it shows outliers.
- Mean is between 25 , min is 20 and max is 35.
- Food Preparations Times across all cuisine types are almost same.(approx. 20 to 35 min)

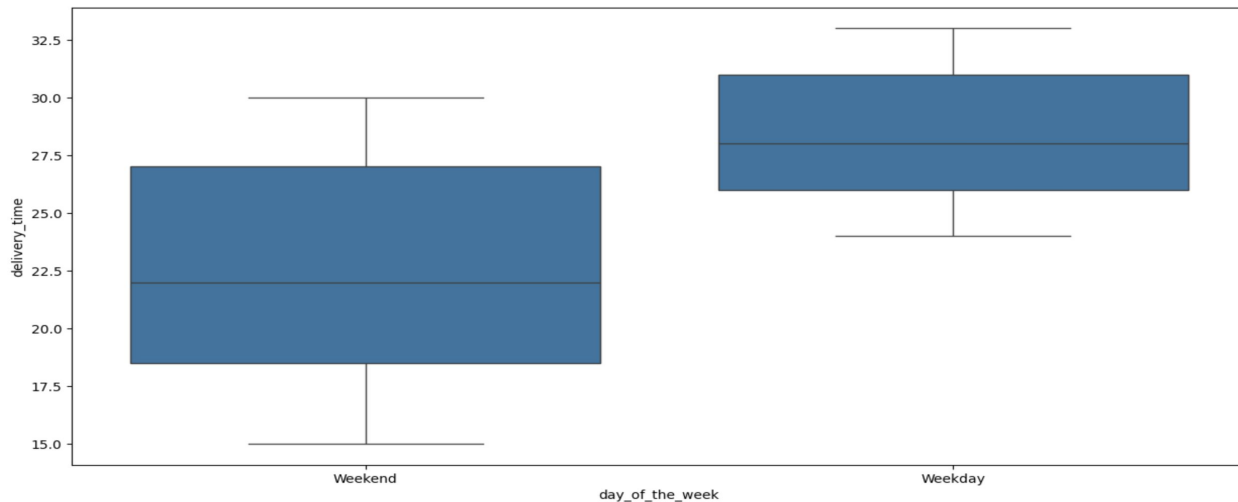


EDA - Multivariate Analysis

3.Day of the Week vs Delivery time -

Observations -

- Orders were More on Weekends compared to Weekdays.
- The average time to deliver food is more on weekdays than on weekends
- 50% of orders on weekend are delivered approximately between 18 and 27 minutes.
- Min delivery time on weekday is greater than average delivery time on weekend.



EDA - Multivariate Analysis

4. Observations on the revenue generated by the restaurants:

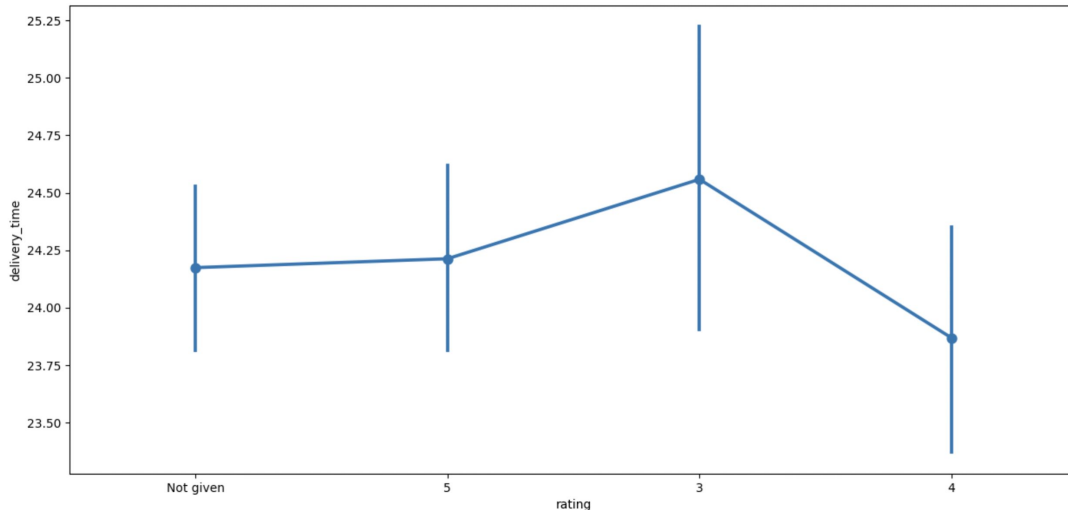
- Maximum revenue is earned by 'Shake Shack' i.e '3579.53'
- Minimum revenue is earned by 'Five Guys burgers and Fries' i.e 506.47
- Top 5 restaurants which generated revenue higher than \$1000 are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken, Parm.

Restaurant Name	Revenue
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

EDA - Multivariate Analysis

5. Rating Vs Delivery Time - Observations -

- 5 rating is been given by customers for the orders which took delivery time of approx 23 to 24 min.
- 4 rating is been given by customers for the orders which took delivery time of approx 24 min.
- 3 rating is been given by the customers for the orders which took delivery time of approx 26 min.
- When delivery time was between 23- 24 min few customers did not gave any rating hence marked as 'Not given'.

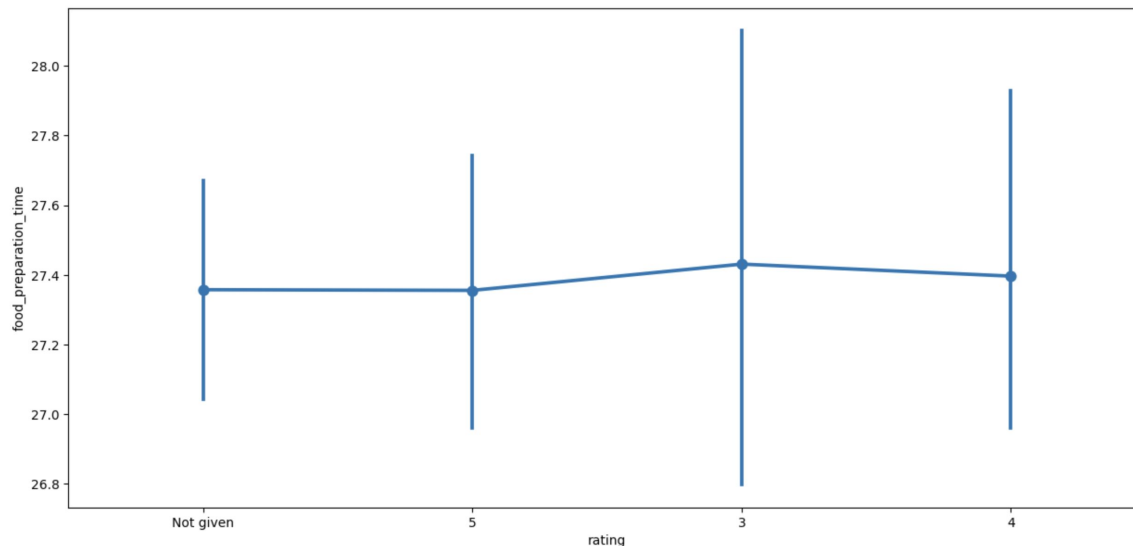


EDA - Multivariate Analysis

6. Rating vs Food preparation time -

Observations:

- Orders which took food preparation time between 26 to 30 min got rating '3', which is max given rating from all orders.
- Orders which took food preparation time between 27 to 27.8 min got rating '4'
- Average orders with food preparation time 27 min, got 3 rating.

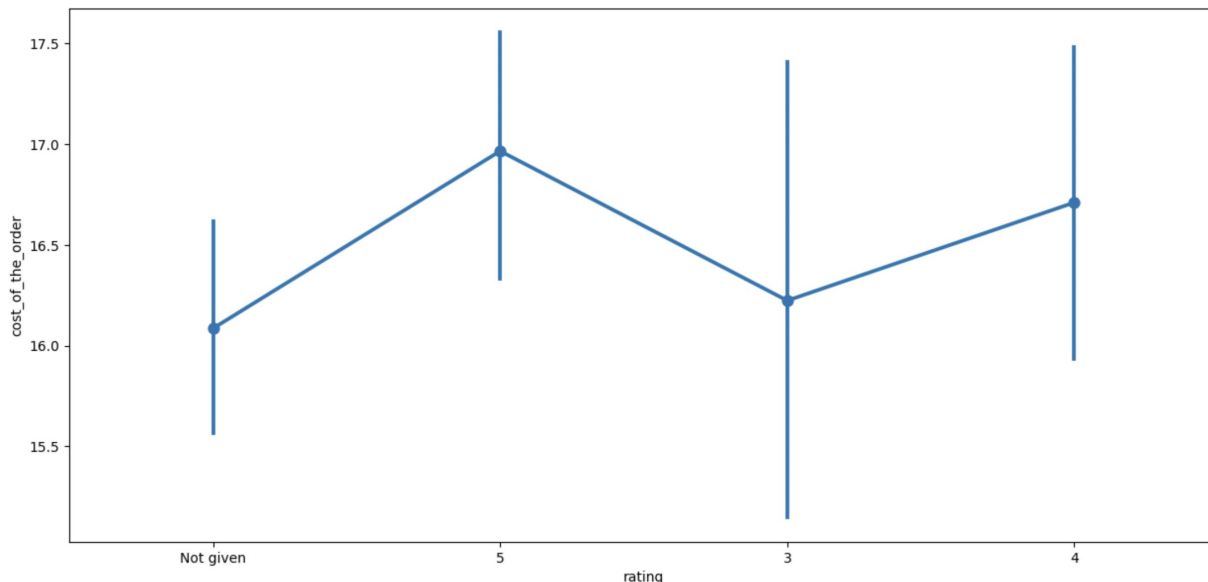


EDA - Multivariate Analysis

7. Rating Vs Cost of order

Observations:

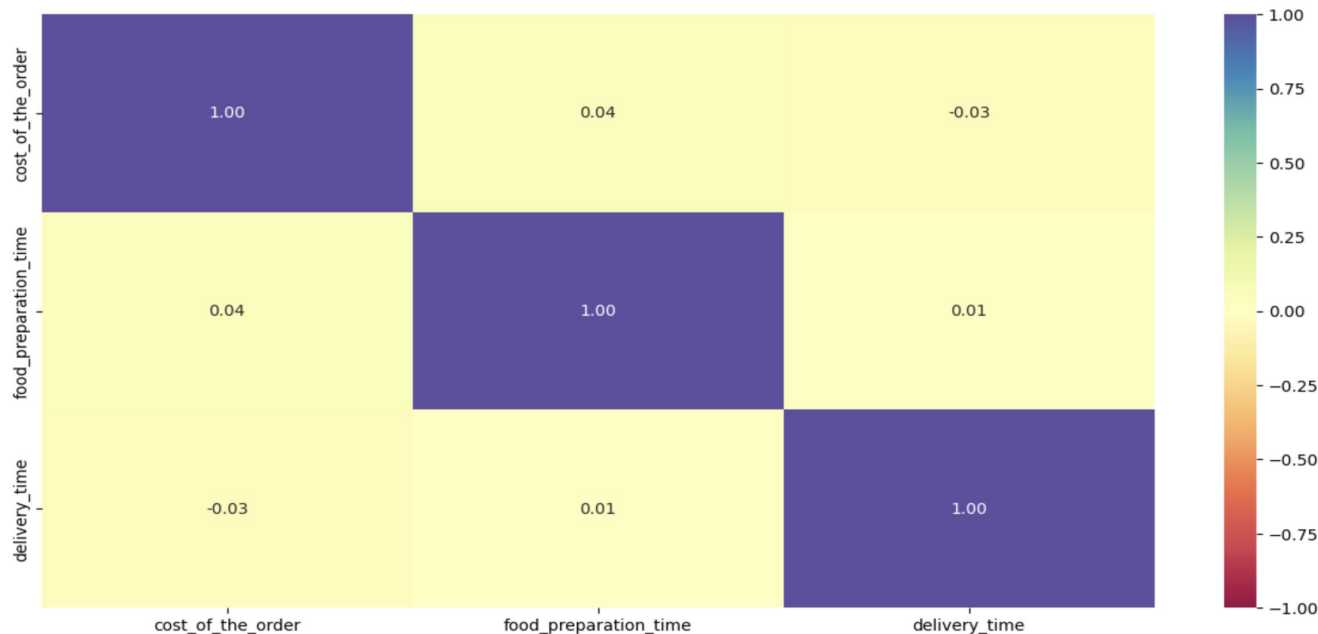
- Orders with cost between 13 to 17 \$ got rating '3'
- Orders with cost between 16 to 18\$ got rating '5'
- Orders with cost between 15 to 18 \$ got rating '4'



EDA - Multivariate Analysis

8. Correlation among Variables - Observations -

- As per displayed heatmap below, There is no major Correlation between variables as most of the correlation are below 0.05



EDA - Multivariate Analysis

Question 13: The company 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. Find the restaurants fulfilling the criteria to get the promotional offer.

Answer - Four restaurants qualify for getting promotional offer as listed below.

Note - Code added in Appendix

	restaurant_name	count	mean
16	Blue Ribbon Fried Chicken	64	4.328125
17	Blue Ribbon Sushi	73	4.219178
117	Shake Shack	133	4.278195
132	The Meatball Shop	84	4.511905

EDA - Multivariate Analysis

Question 14: The company charges the restaurant 25% on the orders having cost greater than 20 dollars and 15% on the orders having cost greater than 5 dollars. Find the net revenue generated by the company across all orders.

Answer- The net revenue generated by the company across all orders is \$6166.3.

Note - Code added in the Appendix

Question 15: The company wants to analyze the total time required to deliver the food. What percentage of orders take more than 60 minutes to get delivered from the time the order is placed? (The food has to be prepared and then delivered.)

Answer -

10.54 % of orders take more than 60 minutes to get delivered from the time the order is placed

The number of total orders that cost above 60 dollars is 200.

Note - Code added in the Appendix

EDA - Multivariate Analysis

Question 16: The company wants to analyze the delivery time of the orders on weekdays and weekends. How does the mean delivery time vary during weekdays and weekends?

Answer-

The mean delivery time on weekdays is around 28 minutes.

The mean delivery time on weekends is around 22 minutes.

Note - Code added in the Appendix

Conclusions and Recommendations

Question 17: What are your conclusions from the analysis? What recommendations would you like to share to help improve the business?

Conclusion -

- The top 5 restaurants, based on the number of orders received and revenue generated, are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken, and Parm.
- American, Japanese, Italian, and Chinese cuisines are the most popular among customers.
- The average food preparation time across all cuisines is between 25 - 28 minutes.
- Delivery times are longer on weekdays compared to weekends.
- There is a positive correlation between food preparation time and the cost of the order, indicating that higher-cost orders typically require more preparation time.
- A positive correlation is also observed between ratings and the cost of the order, suggesting that more expensive orders are likely to receive higher ratings.
- With only 1200 unique customers and 178 restaurants, there is significant room for growth in both customer base and restaurant partnerships.
- Increased orders on weekends present an opportunity for targeted promotions to boost weekday sales and optimize weekend pricing for higher revenue.
- Targeting top customers with personalized promotions based on their preferred cuisine types can potentially increase repeat orders.
- Expanding the selection of American cuisine restaurants could satisfy the high demand for this cuisine type.
- Only 10.54% of orders are delivered within 60 minutes, highlighting an area for improvement in food preparation and delivery efficiency.
- Approximately 38.78% of customers have not provided ratings, business would definitely need a good feedback mechanisms to gather more comprehensive customer insights.

Conclusions and Recommendations

Recommendations:

- Business would need a strong online marketing strategy to increase number of customers, as current data is showing total 1200 customers only.
- Restaurant can reduce delivery time on weekdays to improve efficiency
- Based on the number of repetitive orders data, business would have to give discounts and coupons to existing customers on the basis of previous orders , to increase repetitive sales.
- As American Cuisine Type is the most popular cuisine, business expansion strategy to tie up with more American restaurants / famous restaurant chains, to offer wider choice of food menu would increase sale, Also, adding number of restaurants on various locations,can increase new customers & helps all customers to order food from most nearby restaurant.
- As around 40% customers has not given ratings, restaurants can provide an option to give 'rating' on previously ordered food delivery service, before placing new order, along with post food delivery, to increase chance of getting 'ratings'.
- As Average of the orders are delivered within 60 mins, working with Restaurants for faster food preparation time, including new cuisine type(Variety of Menu) to save food preparation time and optimize faster delivery.
- From data it shows - there are more orders on weekends, the Food Hub can run some deals for weekdays to increase sales and charge premium on weekends to increase revenue.

APPENDIX



```
#Question 4:Check the statistical summary of the data.  
#What is the minimum, average, and maximum time it takes for food to be prepared once an order is placed?  
df.describe(include='all').T  
#Observations - The minimum Food_preparation_time is '20 min'  
#The average food preparation time is '27.37 min'  
# The maximum food preparation time is '35min'  
# The top 'Restaurant_name' is Shake Shack. Total '178'unique 'restaurant_names' been displayed in dataset.  
#The top 'cuisine_type' is American & 14 unique 'Cuisine_type' of food is been ordered by customers.  
#The top 'day_of_the_week'is 'Weekend'. 2 different types of 'day_of_the_week' been displayed.  
#There are total 4 types of 'rating' been displayed.
```



	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



APPENDIX

```
# checking 'restaurant_name' column to find unique Restaurant Names.
df['restaurant_name'].unique()

'Tamarind TriBeCa', 'The Meatball Shop', 'Barbounia',
'Anjappar Chettinad', 'Bukhara Grill',
'Big Wong Restaurant \x8c\x94N', 'Empanada Mama (closed)', 'Pylos',
' Lucky's Famous Burgers', 'Shake Shack', 'Sushi of Gari',
'RedFarm Hudson', 'Blue Ribbon Sushi',
'Five Guys Burgers and Fries', 'Tortaria', 'Cafe Mogador',
'Otto Enoteca Pizzeria', 'Vezzo Thin Crust Pizza',
'Sushi of Gari 46', 'The Kati Roll Company', 'Klong',
'5 Napkin Burger', 'TAO', 'Parm', 'Sushi Samba',
'Haru Gramercy Park', 'Chipotle Mexican Grill $1.99 Delivery',
'RedFarm Broadway', 'Cafeteria', 'DuMont Burger',
'Sarabeth's East', 'Hill Country Fried Chicken', 'Bistango',
'Jack's Wife Freda', 'Mamoun's Falafel', 'Prosperity Dumpling',
'Blue Ribbon Sushi Bar & Grill', 'Westville Hudson',
'Blue Ribbon Brooklyn', 'Nobu Next Door', 'Osteria Morini',
'Haandi', 'Benihana', 'Han Dynasty', 'Chote Nawab',
'Mission Cantina', 'Xi'an Famous Foods', 'Rubirosa',
'Joe's Shanghai \x8e\x8e\x8e\x8e', 'Bareburger', 'The Odeon',
'Pongsri Thai', 'Yama Japanese Restaurant', 'Momoya',
'Balthazar Boulangerie', 'Café China', 'Boqueria',
'Song Thai Restaurant & Bar', 'Five Leaves',
'Pinto Nouveau Thai Bistro', 'Amy Ruth's', 'Pepe Giallo',
'Indikitch', 'Yama 49', 'Piccolo Angolo', 'Pepe Rosso To Go',
'L'Express', 'Amma', 'Delicatessen', 'S'MAC',
'Vanessa's Dumplings', 'Bhatti Indian Grill', 'Taro Sushi',
'Donburi-ya', 'Hatsuhana', 'Samurai Mama', 'Waverly Diner',
'Tarallucci e Vino Restaurant', 'P.J. Clarke's',
'Lantern Thai Kitchen', 'ilili Restaurant', 'The Smile',
'Vanessa's Dumpling House', 'Bubby's ', 'Woorijip',
'Dirty Bird To Go (archived)', 'Haveli Indian Restaurant',
'Dos Caminos', 'da Umberto', 'Sushi of Gari Tribeca',
'Burger Joint', 'Room Service', 'Sarabeth's Restaurant',
'Xe May Sandwich Shop', 'Hibino', 'Mira Sushi', 'Melt Shop',
'J. G. Melon', 'Hummus Place', 'Saravanaa Bhavan',
'Friend of a Farmer', 'The Loop', 'Balade', 'Posto',
```

APPENDIX

```
[ ] # Question 9: What percentage of the orders cost more than 20 dollars?
# Get orders that cost above 20 dollars
df_greater_than_20 = df[df['cost_of_the_order']>20] ## Write the appropriate column name to get the orders having cost above $20

# Calculate the number of total orders where the cost is above 20 dollars
print('The number of total orders that cost above 20 dollars is:', df_greater_than_20.shape[0])

# Calculate percentage of such orders in the dataset
percentage = (df_greater_than_20.shape[0] / df.shape[0]) * 100

print("Percentage of orders above 20 dollars:", round(percentage, 2), '%')
```

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

APPENDIX



```
#Question 13: The company 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.
#Find the restaurants fulfilling the criteria to get the promotional offer.
# Filter the rated restaurants
df_rated = df[df['rating'] != 'Not given'].copy()

# Convert rating column from object to integer
df_rated['rating'] = df_rated['rating'].astype('int')

# Create a dataframe that contains the restaurant names with their rating counts
df_rating_count = df_rated.groupby(['restaurant_name'])['rating'].count().sort_values(ascending = False).reset_index()
df_rating_count.head()
```

	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

APPENDIX

```
#function to determine the revenue
def compute_rev(x):
    if x > 20:
        return x*0.25
    elif x > 5:
        return x*0.15
    else:
        return x*0

df['Revenue'] = df['cost_of_the_order'].apply(compute_rev) ## Write the appropriate column name to compute the revenue
df.head()
```

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time	Revenue
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20	7.6875
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23	1.8120
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28	1.8345
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15	7.3000
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24	1.7385

```
# get the total revenue and print it
total_rev = df['Revenue'].sum() ## Write the appropriate function to get the total revenue
print('The net revenue is around', round(total_rev, 2), 'dollars')
```

The net revenue is around 6166.3 dollars

APPENDIX

```
# Calculate total delivery time and add a new column to the dataframe df to store the total delivery time
df['total_time'] = df['food_preparation_time'] + df['delivery_time']

## Write the code below to find the percentage of orders that have more than 60 minutes of total delivery time (see Question 9 for reference)
df_greater_than_60=df[df['total_time']>60]
print('the number of total orders that cost above 60 dollars is:',df_greater_than_60.shape[0])
percentage = (df_greater_than_60.shape[0]/ df.shape[0])*100
print("percentage of orders above 20 dollars:",round(percentage,2), '%')
```

the number of total orders that cost above 60 dollars is: 200
percentage of orders above 20 dollars: 10.54 %

APPENDIX

```
[ ] # Get the mean delivery time on weekdays and print it
    print('The mean delivery time on weekdays is around',
          round(df[df['day_of_the_week'] == 'Weekday']['delivery_time'].mean()),
          'minutes')

## Write the code below to get the mean delivery time on weekends and print it
print('The mean delivery time on weekends is around',
      round(df[df['day_of_the_week'] == 'Weekend']['delivery_time'].mean()),
      'minutes')
```

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