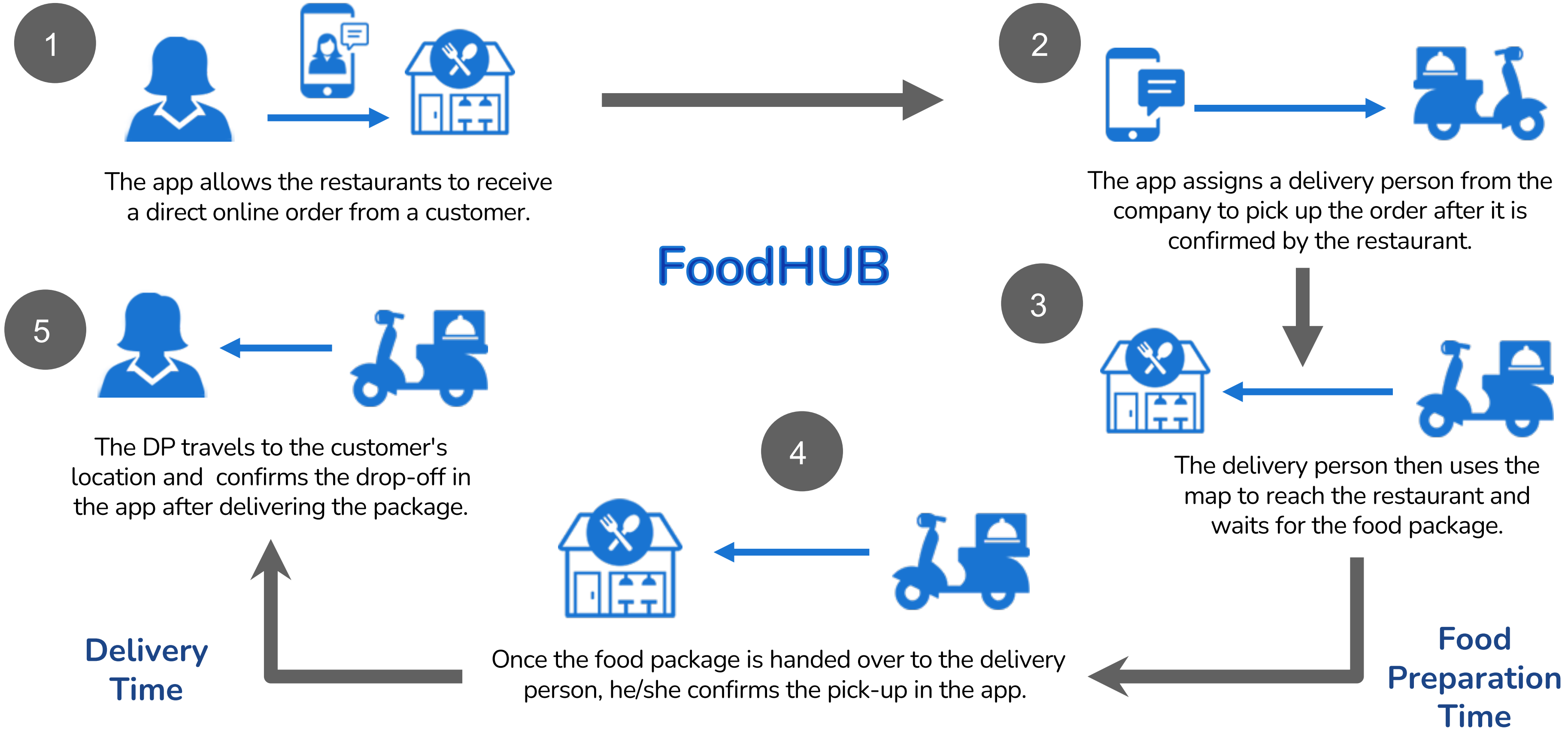
Foodhub Orders Analysis

PGP DSBA – May Cohort

Date: 22nd July 22

Contents / Agenda

Background and Business Overview



Objective

Analyzing the data to get a fair idea about the demand of different restaurants which will help them in enhancing their customer experience.

We will majorly be focusing on the following variables:

- Variables that affect the number of orders from a particular restaurant or a particular cuisine.
- Variables that affect the delivery time
- Factors that affect the demand of the service and the revenue.
- Exploring ways to encourage the restaurants and the customers to use the platform frequently.

Data Overview

The data contains several data points related to a food order. The detailed data dictionary is given below

Variable	Description
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

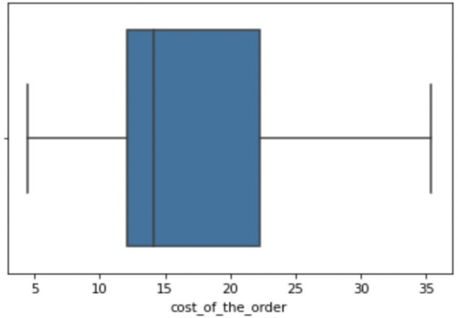
Data Overview

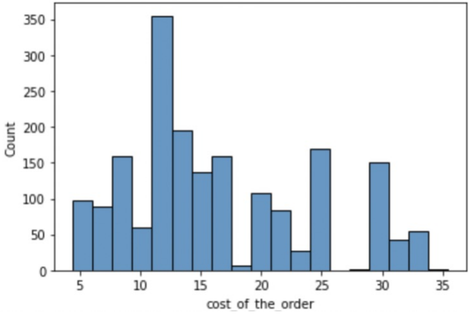
- The total no. of **observations/orders** are: **1898**
- The total no. of **Variables** are: **9**
- Total no. of **unique customers** in the observed dataset: **1200**
- Total no. of **unique restaurants** in the observed dataset: **178**
- Total no. of **unique cuisine types**: **14**
- There are **no missing values** to be found in the received dataset.

Univariate Analysis

There are various factors affecting the food aggregating business of FoodHUB. We shall dive into each variable now to understand their distribution. The observations are written besides the presented graphs.

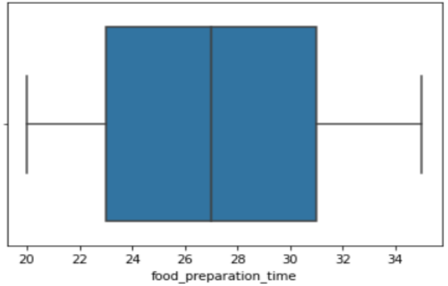
Cost of order

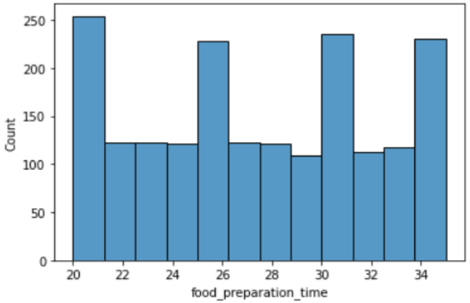




- The orders range from \$5 to \$35
- 50% of the orders fall in the range of \$12 to \$23.
- The highest number of orders have a value of \$11 to \$12.
- **In a separate analysis, it is observed that around 30% of the total orders cost more than \$20.**

Food Preparation time

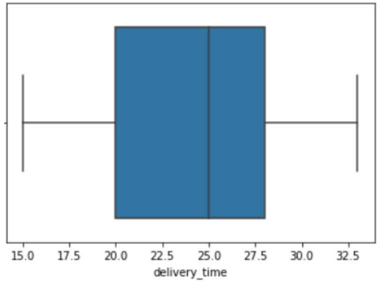


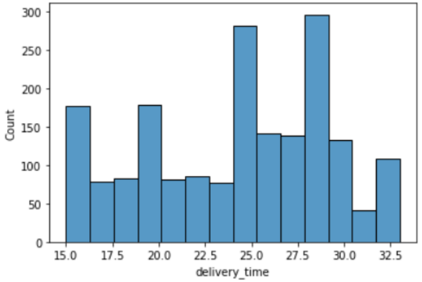


- 50% of orders take 23 to 31 minutes of preparation time.
- The minimum preparation time required is 20 minutes and 35 minutes is maximum time taken.
- Most number of orders are experienced to have 20, 25, 30 and 34 minutes of preparation time.

Univariate Analysis

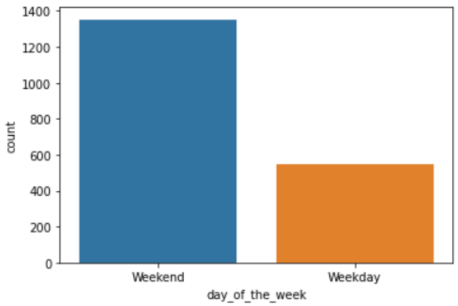
Delivery time





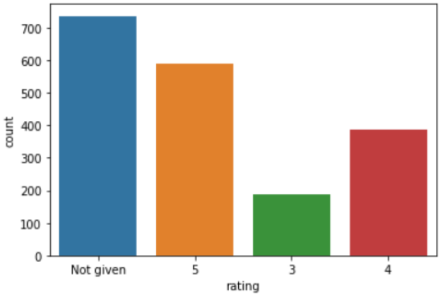
- 50% of the delivery time fall in the range of 20 to 28 minutes with the median being 25 minutes.
- More than 200 number of orders experience the delivery time 24 and 28 minutes.
- **In a separate analysis, it is observed that the average delivery time is 24.16 minutes.**

Orders depending on the day of the week



- Orders placed during the weekend are significantly higher than weekdays.
- Further investigation regarding this point shall help us capitalize on the orders on weekdays as well.

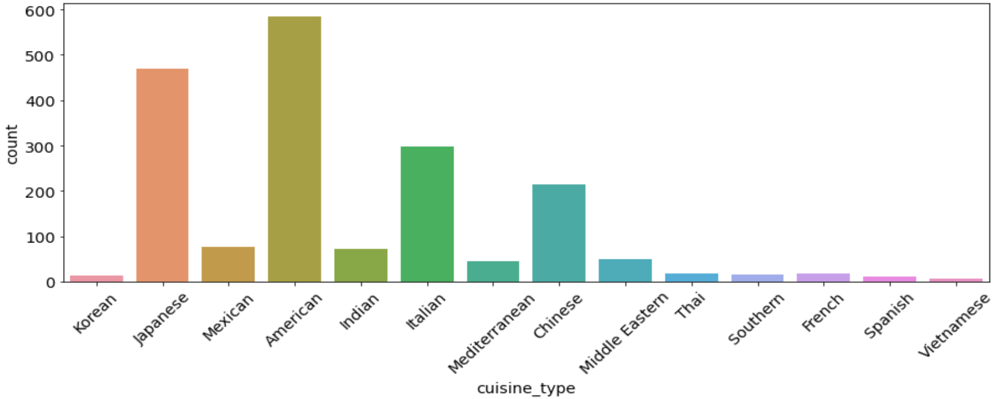
Ratings given to each order



- Most of the orders are not given a rating. It is important to devise a way to incentivize ratings.
- The highest number of rating given is 5 stars, although it is necessary to understand whether the rating is given to the delivery or the restaurant for its food.

Univariate Analysis – Cuisines

Distribution with respect to Cuisine Types




```
#To find out the number of restaurants serving the top 5 cuisines
df_cuisine_rest=df.groupby('cuisine_type')['restaurant_name'].nunique()
df_cuisine_rest
```

```
cuisine_type
American          41
Chinese           16
French             3
Indian            14
Italian           31
Japanese          29
Korean             5
Mediterranean      5
Mexican           11
Middle Eastern     7
Southern           2
Spanish            3
Thai               9
Vietnamese         3
Name: restaurant_name, dtype: int64
```

- The most ordered cuisine types are American, Japanese, Italian and Chinese (in that order) from 178 total number of restaurants.
- But as observed in the snapshot on the right, **the most ordered cuisines seem related to the number of restaurants serving them.**
- **The top 5 cuisines ordered are also the top 5 cuisines served by the network of restaurants on the platform.** We will have to investigate further if there is an actual relationship between them.

Best performers – Cuisines

Highest Grossing Cuisine Types

Cuisine	Revenue ($)
American	9530
Japanese	7663
Italian	4892
Chinese	3505
Indian	*1235*

Top 5 cuisines on weekdays

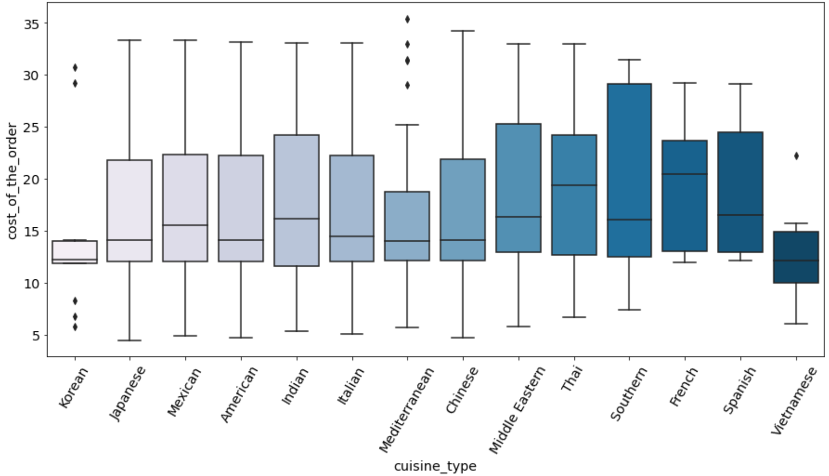
Cuisine	No. of Orders
American	169
Japanese	135
Italian	91
Chinese	52
Mexican	24

Top 5 cuisines on weekends

Cuisine	No. of Orders
American	415
Japanese	335
Italian	207
Chinese	163
Mexican	53

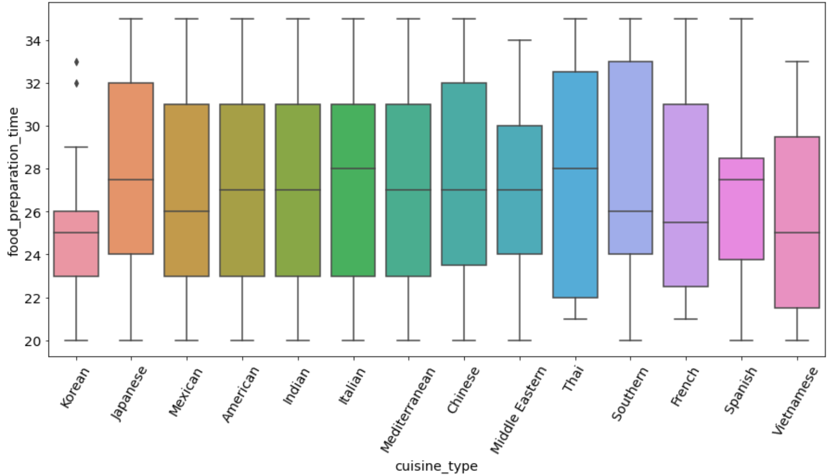
Multivariate Analysis – Cuisine centric

Cuisine vs Cost of Order



- The cost of order between American, Italian, Japanese and Mexican seems to be similar where 50% of the data seems to fall around 12 to 23 dollars per order.
- There are maximum no. of outlying cost of orders in the Korean cuisine. Other cuisine which experience outliers are Mediterranean and Vietnamese.

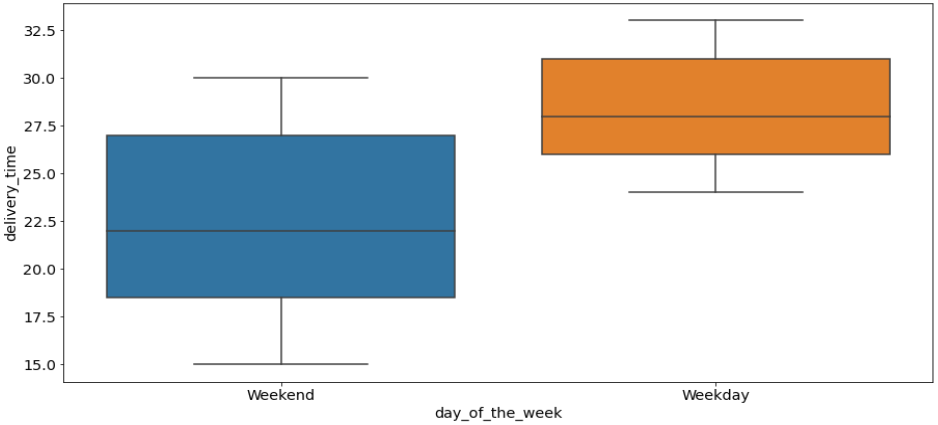
Cuisine vs Food Preparation time



- Korean cuisine takes the least amount of preparation time
- The pattern seems to repeat itself - American, Japanese, Italian and Mexican are observed to have similar preparation time.
- We could further investigate Thai cuisine which has 50% of the orders requiring 22 to 32 minutes of preparation time.

Multivariate Analysis – Delivery time

Day of the week vs Delivery time



- It is understood that it takes longer to deliver on weekdays due to the traffic, *which could also be the reason for low no. of orders on weekdays.*
- This is quite important to investigate as the factors affecting it will have a direct impact on the total revenue.
- *Appointing more drivers or encouraging deliveries from nearby restaurants on weekdays could help us boost the number of orders.*
- It takes around 15 to 30 minutes of delivery time on weekends as compared to 23 to33 minutes on weekdays

*Total delivery time = Food preparation time + Delivery time

Orders taking more than 60 minutes of Total delivery time = 200

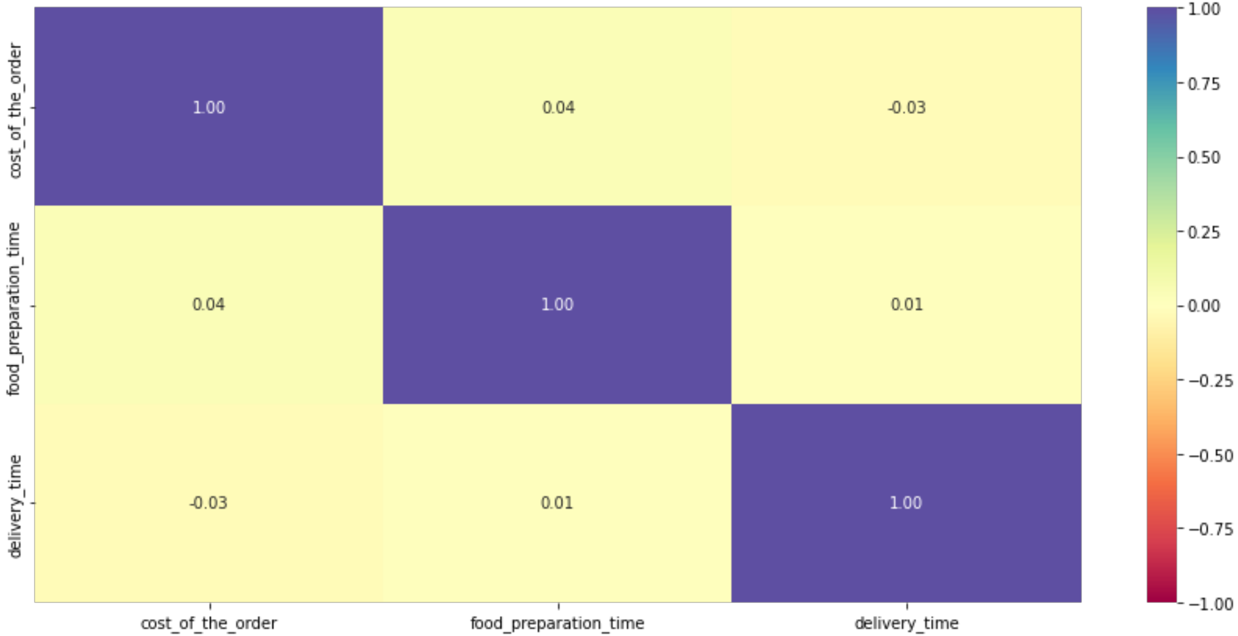
The percentage of orders taking more than 60 mins of Total delivery time = 11 percent

Average Delivery time on:

Weekdays = 28 minutes
Weekends = 22 minutes

Correlation analysis

Correlation matrix



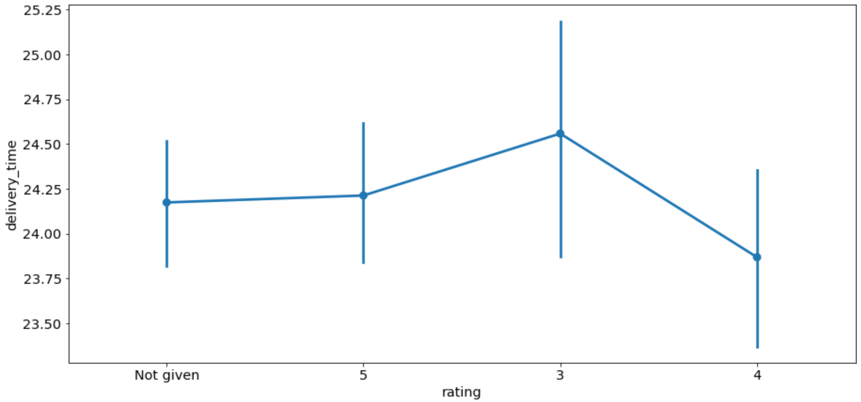
The cost of order and food preparation time has the highest correlation. But we do not see any high coefficients of correlation between other numerical variables.

Multivariate Analysis – Ratings received

Analyzing rating across different delivery parameters.

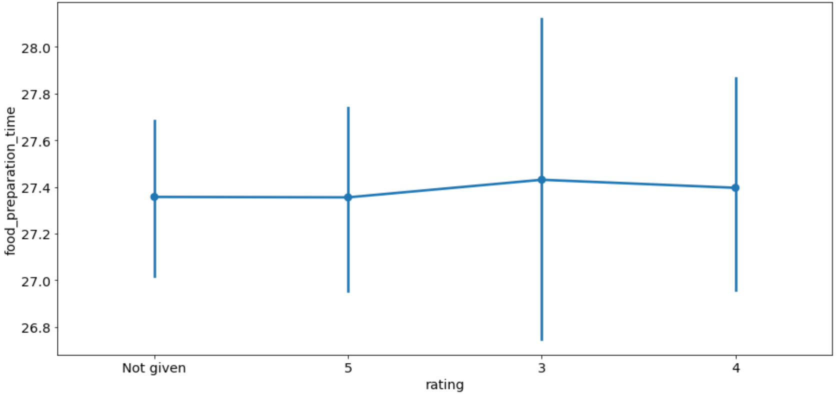
Rating vs Delivery time

*Avg. Order Delivery time = 24.16 minutes



- The 4-star rating is given when the delivery time drops significantly.
- 3-star rating is given when the delivery time is higher.
- Safe to say that the delivery time does impact the ratings given

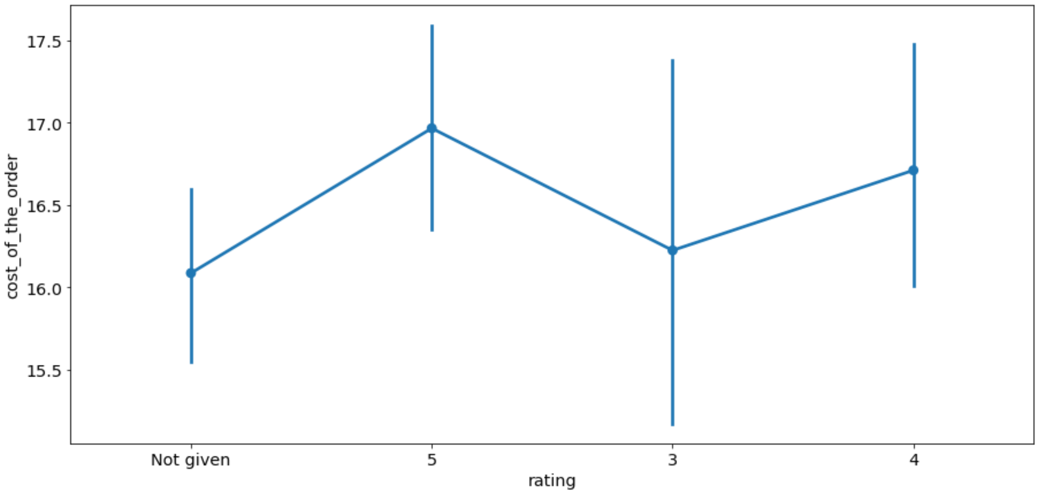
Rating vs food preparation time



- The same trend can be observed in this relationship. 3-star rating have the highest range of food preparation time.

Multivariate Analysis – Ratings received

Rating vs Cost of order



- As we can observe, the 5 star rated orders range \$16.5 to \$175 and above (which is the higher end of the price spectrum), and the 3 star rated orders range from \$15.2 to \$17.3 (lower end of the price spectrum).
- One possible inference could be the quality of food. Cheap food implies lower quality, resulting into lower ratings.
- The ratings are indeed affected by the cost of order significantly. We would require detailed investigation towards this direction.

Multivariate Analysis – Ratings received

Rating vs Day of the week

```
df.groupby('rating')['day_of_the_week'].value_counts()
```

```
rating      day_of_the_week
3           Weekend            125
            Weekday             63
4           Weekend            277
            Weekday            109
5           Weekend            420
            Weekday            168
Not given   Weekend            529
            Weekday            207
Name: day_of_the_week, dtype: int64
```

- The code on the left was run to figure out whether the ratings received had any relationship with the longer delivery time on weekdays.
- **The correlation would have been established if the data suggested that 3-star ratings are more prevalent on weekdays rather than weekends.**
- Seems like the relationship is more dependent on the number of orders. As the number of orders are higher on weekends, so are the rating given.
- **There is no correlation between the longer delivery time on weekdays and the ratings given.**

Best performers - Restaurants

Great Learning
POWER AHEAD

Top 5 restaurants – orders and revenue generated

Restaurant	SHAKE SHACK	THE MEATBALL SHOP	BLUE RIBBON SUSHI BAR & GRILL SOUTH BEACH	BROMBERG BROS. BLUE RIBBON FRIED CHICKEN	Parm FAMOUS ITALIAN
Orders	219	132	119	96	68
Revenue ($)	3579	2145	1903	1662	1112

Receivers of the Promotional Offer

Criteria:

The restaurant must have a rating count of 50 and the average rating shall be greater than 4.

The code used for obtaining the result is as shown below:

```
# 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)
df_rating_count.head()
```

```
[43] # Get the restaurant names that have rating count more than 50
     rest_names = df_rating_count[df_rating_count['rating']>50]['restaurant_name'] ## Complete the code to ge

     # Filter to get the data of restaurants that have rating count more than 50
     df_mean_4 = df_rated[df_rated['restaurant_name'].isin(rest_names)].copy()

     # Group the restaurant names with their ratings and find the mean rating of each restaurant
     df_mean_4.groupby(['restaurant_name'])['rating'].mean().sort_values(ascending = False).reset_index().dro
```

1 — THE MEATBALL SHOP — 4.51 Avg Rating

2 — BROMBERG BROS BLUE RIBBON FRIED CHICKEN — 4.32 Avg Rating

3 — SHAKE SHACK — 4.27 Avg Rating

4 — BLUE RIBBON SUSHI BAR & GRILL SOUTH BEACH — 4.21 Avg Rating

Revenue of FoodHUB

Criteria:

25% on order cost greater than $20 and 15% on order cost greater than $5.

The net revenue for the observed duration is 6166.3 Dollars

The code used for obtaining the result is as shown below:

```
#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 apprpriate column name to compute the revenue
df.head()
```

```
# 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
```

Conclusions and Recommendations

Sr No.	Conclusions	Recommendations
1	The top 5 cuisines served are American, Japanese, Italian and Chinese due to the availability of these cuisines among the network of 178 no. of restaurants present on the platform. *(Slide no. 9)*	If the factor influencing the number of orders for a particular cuisine is directly relatable to the availability of those restaurants on our platform, then its worth looking into increasing our network of restaurants for other cuisines.
2	The number of orders on the weekends are higher than the weekdays. Weekdays also have higher delivery time than weekends. *(Pg 8 and 12)*	We should investigate this further to understand if the number of orders are being impacted by the delivery time. Measures to reduce the delivery time on weekdays must be taken to increase the no. of orders. Also, various promotional offers could be introduced to encourage the no. of orders on weekdays.
3	The number of ratings 'Not given' is the highest among all the ratings received. The next highest rating received is '5', which is encouraging. *(Pg 8)*	• Although it may need further investigation to understand if the '5' rating given is for the delivery or the food quality. Building separate rating input for delivery and food shall help us understand that. • Incentive to rate an order shall help us enrich our data further. A loyalty program might help incentivize the rating process, where each rating given puts the customer closer to the higher slabs of loyalty with greater discounts.
4	Rating is heavily impacted by the delivery time and mildly by the food preparation time *(Pg 13)*. 11% of the orders take more than 60 minutes of delivery time. *(Pg 12)*	Need to investigate further if we need to increase our network of drivers to decrease the total delivery time as it directly impacts the rating and the number of orders (and hence the revenue).

Conclusions and Recommendations

Sr No.	Conclusions	Recommendations
5	Ratings are low when the cost of order is low. *(Pg 15)*	One possible explanation could be the quality of food. Cheap food usually implies lower quality. But we would need to further investigate this. Our resolution of Sr No. 3 would help us understand this better.
6	Indian Cuisine is not among the most ordered cuisine but generates significant revenue to be the top 5 revenue generating restaurants. *(Pg 10)*	We should consider expanding our network to include more Indian restaurants as it seems like the cost per order is high for this cuisine, which shall ultimately result in higher revenue.
7	We do not have any data on the users of this app such as age, gender or location.	We could segregate our studies further depending on the user information. For example, a certain age group, gender or location would prefer certain kind of food/restaurants and understanding the frequency of certain user types using this app could help us take measures to increase and adopt engagement strategies.
8	No available data if the users were referrals or organic users (from advertisements or app store searches).	We need to understand if the users have been using the app through referrals or have been organic users. This would help us encourage one type of customer acquisitions and enhance the number of users.

APPENDIX

Some more case studies for reference from other food aggregating businesses which might help us enhance our services:

- Zomato case study
- Uber Eats SWOT analysis
- How swiggy become the food delivery giant in India (*in the presence of zomato)
- How does Grubhub work? The business model and tech stack

bracket emphasis given by the author

Happy Learning !