

Foundations of Data Science

Project - FoodHub

Course Name - Data science and machine learning.

Date - 6 July 2023

Submitted By:
Niharika Srivastava

Contents / Agenda

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

Executive Summary

Data digging using various python libraries has led us to following conclusions:

1. American, Japanese and Italian cuisine is the most popular and ordered cuisine amongst customer, and restaurant which offer this cuisine generate better revenue.
2. Customers place more orders on weekends, could be due work off day. During the weekdays they are eating in office cafeterias.
3. Shake shack, The meatball shop, blue ribbon fried chicken and blue ribbon sushi are the most liked and highly rated restaurants.
4. Average price of orders from these restaurants fall between 15.99 and 17.3, which is close to the overall average of the cost.
5. Cost of the order shows a positive correlation with Rating. In order to generate better revenue Restaurants should aim to better the ratings given to them.
6. Food preparation time shows no correlation with the rating.
7. Delivery cost of the order time shows a positive correlation with the rating but is not the driving force behind it.
8. Delivery on weekends are faster may be because more delivery person working on weekend as part time job ex, students.

Executive Summary

Recommendations:-

1. Food Hub company should invite onboard more Vendors/Restaurants which offer American cuisine, Italian and Japanese cuisine.
2. Food hub should promote restaurants which offer more than 1 cuisine.
3. Food hub App should have a bell notification for orders that are left unrated.
4. Food Hub should also provide customers with better digital tools to reduce the hassle of giving ratings to the order.
5. To improve the rating Restaurants can offer complimentary food, extra ketchup packets, better cutlery, tissues etc.
6. Company should come up with Slash deals and limited time offer on weekdays to lure customers.
7. Company should employ more delivery people to reduce the delivery time on weekdays.
8. Company should do more advertisement of less ordered cuisine and can offer Promo codes on them.
9. Better digital campaigns.
10. more data can be gathered to understand the impact of other variables on the cost and ratings of the orders for example time of the day when order is placed, location of restaurant etc. .

Executive Summary

Potential Benefits of implementing the recommendations are :-

Food Hub Company.

- Better Turnover.
- Improvement of retention rate of customers.

Restaurants

- Good ratings.
- Better revenue generation.

Customers

- Good services.
- Better experience.

Business Problem Overview and Solution Approach

Background :-

FoodHub is a food aggregator company that offer access to multiple restaurants through a single smartphone app. This App allows Restaurants to receive order, Assigns delivery person to pickup the order from the restaurant and deliver it to the customers location, Requests Customers to rate the order.

From the time the order is confirmed till the time it is delivered, The time stamp of various steps of the process to deliver the order is recorded in the app for each and every order placed through it.

FoodHub earns money by collecting a fixed margin of the delivery order from the restaurants.

Objective and Solution Approach :-

FoodHub wants to analyse the data to get a fair idea about the demand of different restaurants which will help them in enhancing their customer experience. As a data scientist our Approach to fulfil our objective would be.

- Reading and understanding the data.
- Data Pre-processing & Exploratory Data Analysis
- Understanding the impact of various factors on the Demand of restaurants.
- Identifying factors that significantly influence the demand and ratings and are driving force behind it.
- Drawing Insights and Recommendations.

Data overview

- The data of the different orders was obtained from the online portal of the food aggregator company.
- Data comprises of 9 columns and 1898 rows.
- The detailed data dictionary is provided in next slide.
- There seemed to be a few outliers in some of the attributes.

Data Overview

First, by looking at all attributes. We can divide them into a few categories as follows

Data Category	Attributes.
Key	Order_id, Customer_id.
Restaurant info	Restaurant_name, Cuisine_name
Time	Day_of_the_week, food_preparation_time, delivery_time
Review info	Ratings
Objective Variable	Price.

Data Overview

Given below is overall description and information of all the 9 variables present of the data.

- Key** - This category is generally determining the data. Each order has unique order_id and each customer has unique customer_id. The relationship and possible correlation of such value are described in the summary below.

Column	Data Type	Description of data	Possible correlation
Order_id	Int64	Unique ID of the order	Null
Customer_id	int64	Unique ID of the customer	One host id could be linked with multiple Order_id

Data Overview

- 2. Restaurant Info** - This category includes the information of the restaurant name and the type of cuisine it serves.

Column	Data Type	Description of data	Possible correlation
Restaurant_name	object	Name of the restaurant	<ul style="list-style-type: none"> Each restaurant is linked to only 1 cuisine type. Except – Meatball Shop It can influence the cost of the order.
Cuisine_name	object	Cuisine ordered	<ul style="list-style-type: none"> Each cuisine_name could be linked to many Restaurants. It can influence the cost of the order.

Data Overview

3. **Time** - This category includes the information about the day and time when the order was made and when the order was delivered.

Column	Data Type	Description of data	Possible correlation
Day_of_the_week	Object	Indicates whether the order is placed on a weekday or weekend	<ul style="list-style-type: none"> It can influence the food preparation and delivery time
food_preparation_time	Int64	Time (in minutes) taken by the restaurant to prepare the food.	<ul style="list-style-type: none"> It can possibly influence rating of the order
Delivery_time	int64	Time (in minutes) taken by the delivery person to deliver the food package.	<ul style="list-style-type: none"> It can possibly influence rating of the order

Data Overview

4. **Review info** - This category includes the information about the ratings given by the customers.

Column	Data Type	Description of data	Possible correlation
Ratings	Object	Rating given by the customer out of 5	<ul style="list-style-type: none"> It can possibly influence the cost of the order.

5. **Cost** - It indicates the cost of order. This should be target variable or variable of interest for any Business Planning . It could be influenced by the rating, cuisine and restaurant and could impact demand .

Column	Data Type	Description of data	Possible correlation
Cost	int64	Cost of the Order	<ul style="list-style-type: none"> It can impact the demand.

Data Overview:-

Summary of the data - Here is our summary of data with basic analysis, after we have gained a general understanding of the set.

Variable	Mean	S.D.	Min	Max	Top
Order_id	1.477496e+06	5.480497e+02	1.476547e+06	1.478444e+06	NA
Customer_id	171168.4	113698.1	1311	405334	NA
Restaurant_name	NA	NA	NA	NA	Shake Shack
cuisine_type	NA	NA	NA	NA	American
cost	16.5	7.5	4.6	35.4	NA
Day_of_the_week	NA	NA	NA	NA	Weekend
rating	NA	NA	NA	NA	Not given
food_preparation_time	27.4	4.6	27.3	35	NA
Delivery_time	24.2	4.5	24.9	33	NA

Data Overview:-

Summary of the data –

- So the data given has total 178 Restaurants offering 14 different cuisine.
- Services are offered by the restaurants on both Weekend and weekdays.
- Based on restaurants services provided, customers rate them on the scale of 1 to 5.
- 736 Orders are not rated..
- Min, Max and average time taken to prepare the food are 27.3 min, 35 mins and 27.4 mins respectively.
- There were no missing or anomalous values in the data.

Exploratory Data Analysis

EDA steps that are performed are:

Univariate Analysis

- Boxplot
- Histogram
- Count Plot
- Descriptive Statistics

Bivariate Analysis

- Conditional Boxplot
- Aggregation methods
- Point plot
- Heat Map

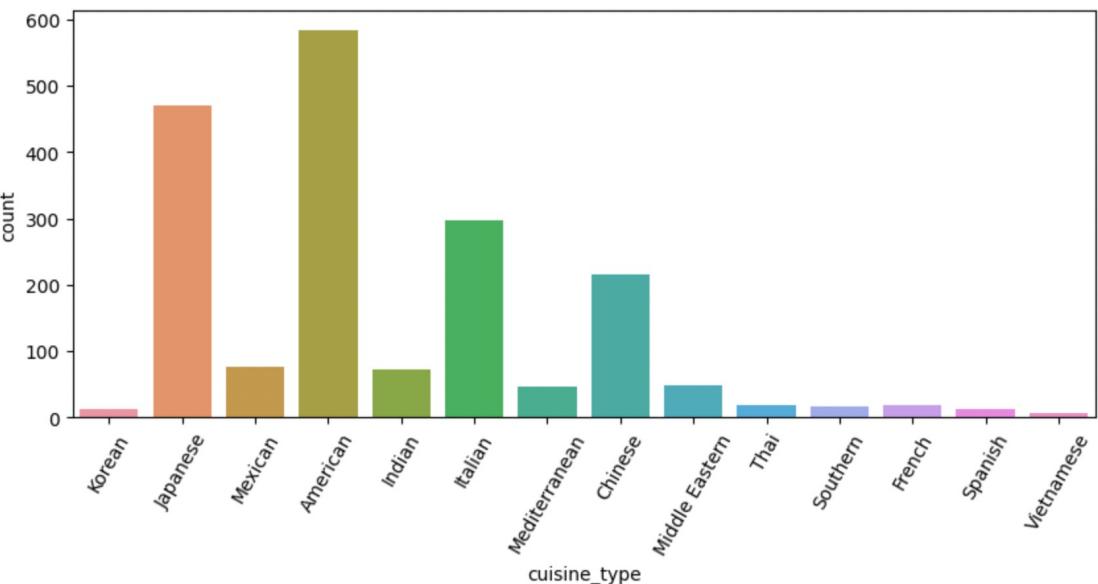
Univariate Analysis – Order_ID, Customer_Id, Restaurant name.

Lets us begin the Exploratory data analysis with Order ID Customer_id, Restaurant name using nunique() function of statistics.

- We have the data of 1898 orders,
- Placed by 1200 Customers.,
- Across 178 Restaurants.
- Top 5 Restaurants with maximum number of orders are :-
 - Shake Shack
 - The Meatball Shop
 - Blue Ribbon Sushi
 - Blue Ribbon Fried Chicken
 - Parm
- Top 3 most frequent customer id are.
 - 52832,
 - 47440,
 - 83287.

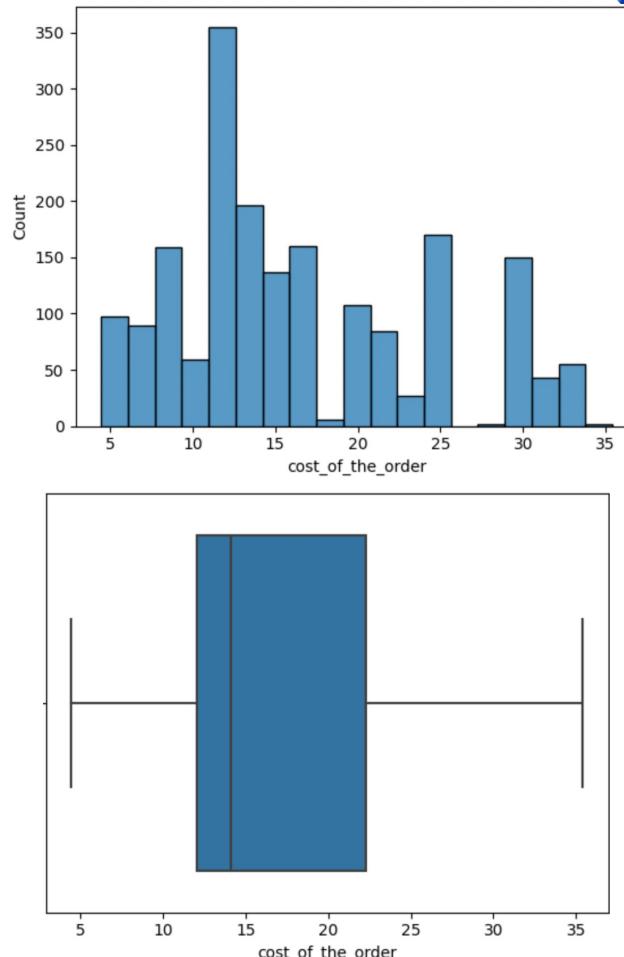
Univariate Analysis – Cuisine_type

- We observed there are 14 different type of cuisines available across all the restaurants.
- Most popular cuisines amongst customer is American Cuisine, followed by Japanese, Italian and Chinese.
- Total 584 orders were made for American cuisine alone, which is 30% of the total orders.
- American cuisine remains favorite and most ordered on both weekend and weekdays.



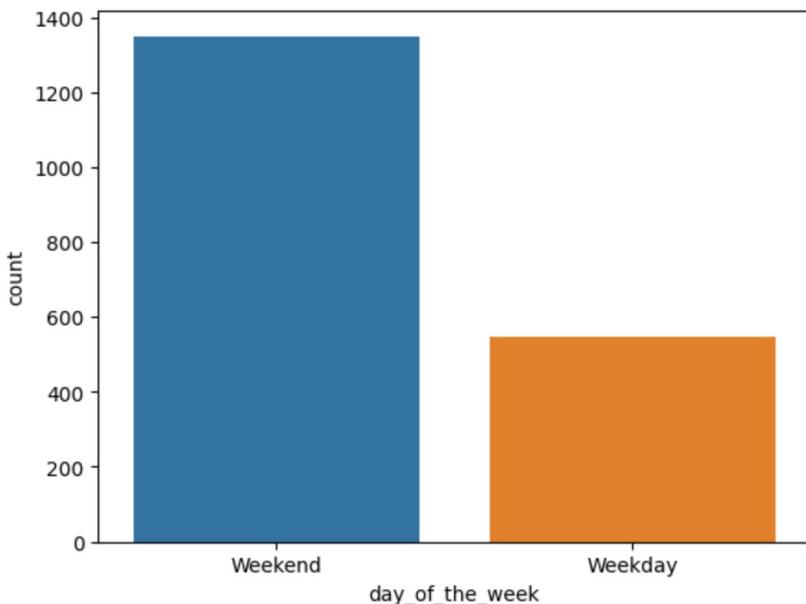
Univariate Analysis – Cost of the order

- From Histogram we can see that Most Orders cost between \$10 to \$15.
- Median cost of all the orders is around 14.
- We also observed that mean(16) is greater than median(14), which implies data is right skewed.
- From the boxplot we also observed, 25% of the orders cost above around \$22. and 25% cost below \$12 whereas 50% of orders cost between \$12 and \$22
- Price of the order ranges from \$4.6 to \$35
- There are no outliers on the upper or lower bound of the box plot.



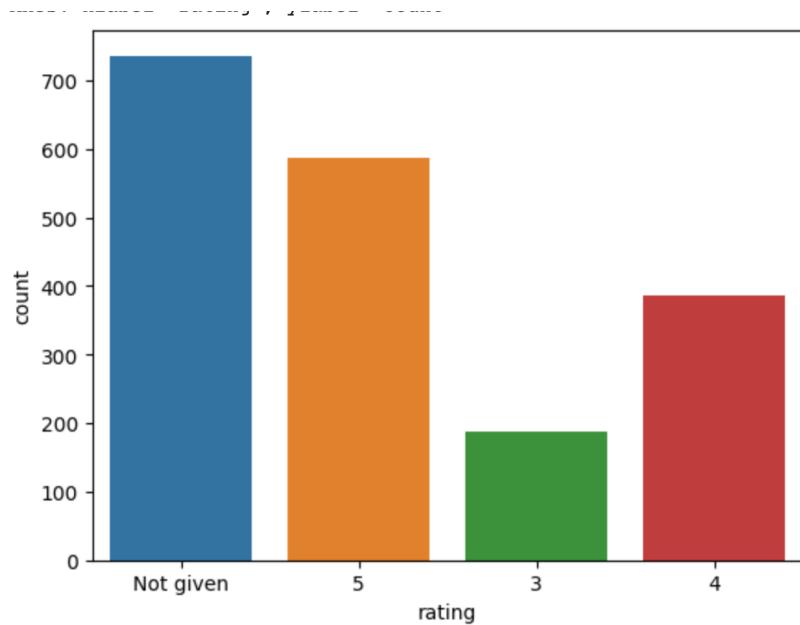
Univariate Analysis – Day of the Week.

- From the given count plot of the day of the week, we saw that customers place more order on the weekend than on weekdays.
- It could be due to people being outside for their jobs during the weekdays and eating in the office canteens or nearby restaurants.
- American cuisine being most popular during the weekends.



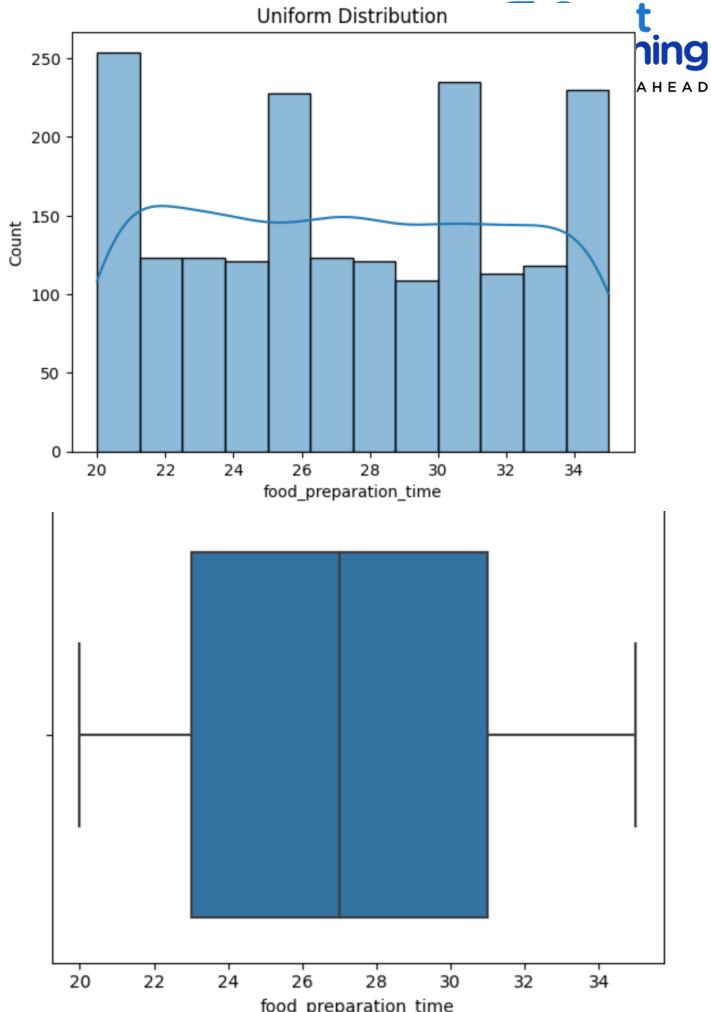
Univariate Analysis – Ratings.

- From the given count plot of the Ratings, we observed there are 4 unique ratings, 5 , 3, 4, Not given.
- More than 700 orders have not been given any Ratings.



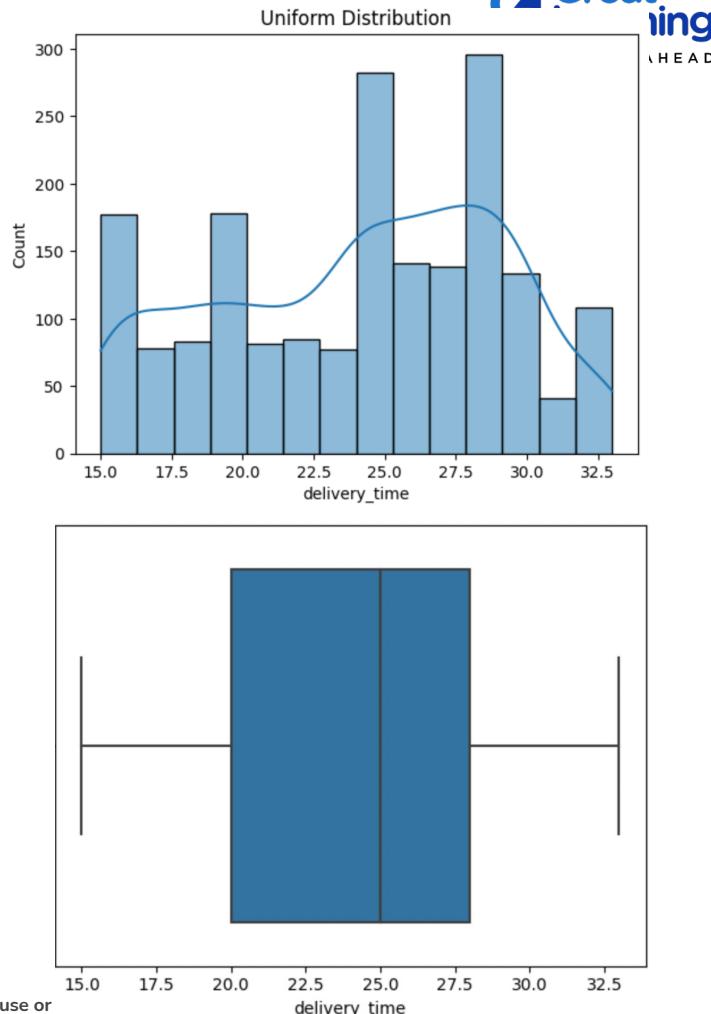
Univariate Analysis – food Preparation time

- From Histogram we can see that food preparation Data is uniformly distributed.
- Median food preparation time of all the orders is around 27 minutes and is almost similar to mean 27.4 minute.
- There are no outliers on the upper or lower bound of the box plot.



Univariate Analysis – food delivery time

- From Histogram we can see that Data for food delivery time is not uniformly distributed.
- Median delivery time of all the orders is around 25 minutes and is more than mean of 24.1 minute.
- Data is left skewed.
- Variance in delivery time could be due to the distance of the restaurant from the delivery place.
- 25% of the orders gets delivered between 15 to 20 minutes.
- There are no outliers on the upper or lower bound of the box plot.

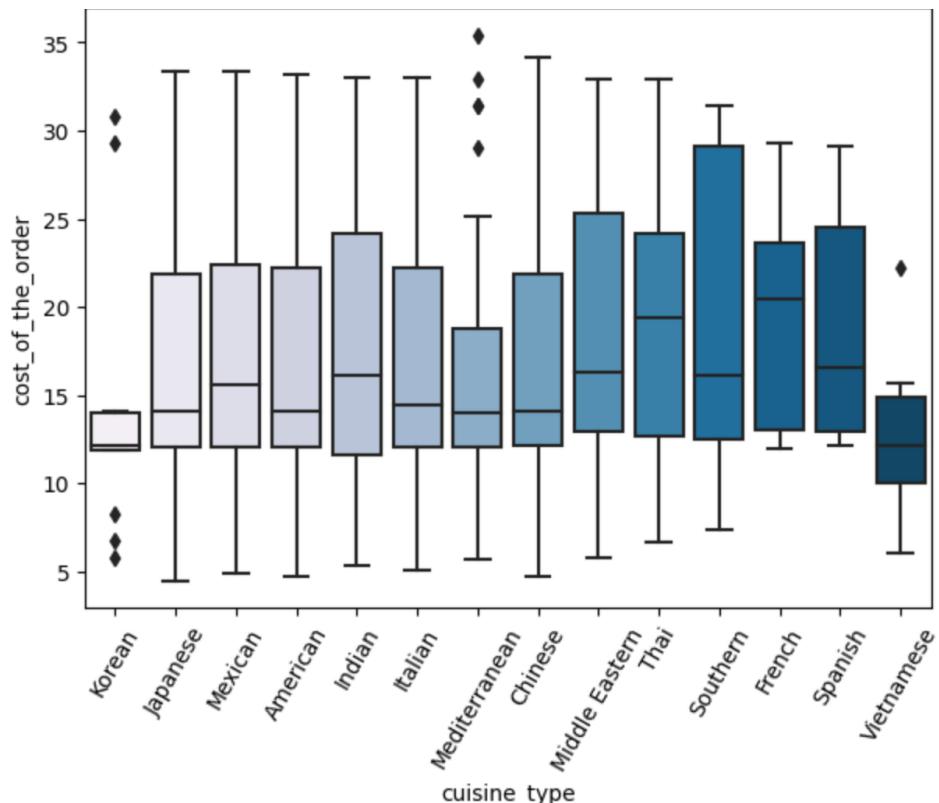


Questions :-

- Q - Which are the top 5 restaurants in terms of the number of orders received?
- A Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken, Parm.
- Q - Which is the most popular cuisine on weekends?
- A American
- Q - What percentage of the orders cost more than 20 dollars?
- A 29.24%
- Q - What is the mean order delivery time?
- A 24.16 minutes
- Q - 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.
- A 52832 13,
47440 10,
83287 9.

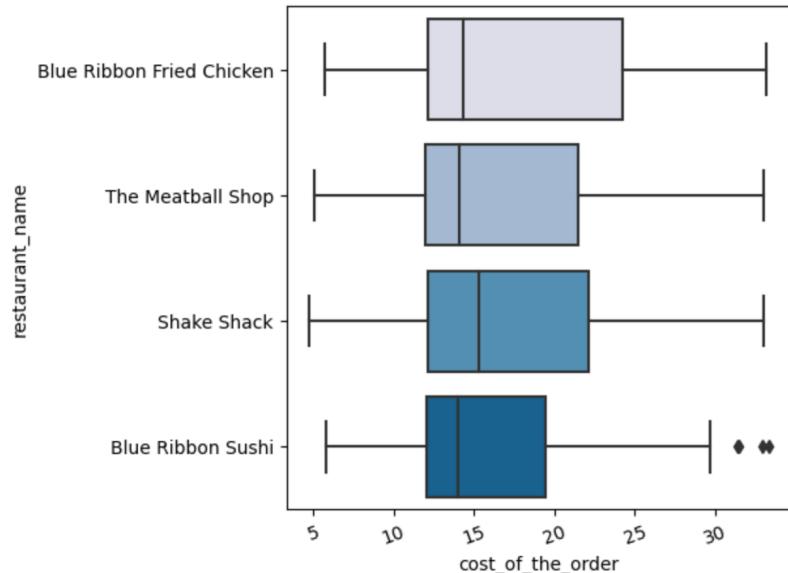
Multivariate Analysis - Cuisine vs Cost of the order

- As observed from the conditional box plot, amongst the 14 cuisines that are available to be ordered, French cuisine has the highest Mean and Median cost of \$19.8 and \$20.47.
- Whereas we also saw in the previous slides that has only been ordered 18 times, typically not very demanded cuisine.
- American cuisine on the other hand which is very popular amongst customer has median price of \$14.120.
- Highest price variation is observed in middle Eastern cuisine.



Multivariate Analysis - Restaurant vs Cost of the order

- As seen in Table 1 and Table 2, restaurants with more than 50 orders and average rating above 4 are Blue ribbon fried chicken, meatball shop, shake shack, Blue ribbon sushi. So we have done our analysis for these 4 restaurants.
- Of these 4 restaurants, The meatball shop has highest rating(4.5) with median price less than \$15.
- We also observed that meatball shop is the only restaurant which offers both American and Italian Cuisine.
- 3 of these 4 restaurants offer American cuisine.



	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

Table 1. - Avg Rating of top Restaurants

	restaurant_name	cost_of_the_order
0	Blue Ribbon Fried Chicken	17.315521
1	Shake Shack	16.344886
2	The Meatball Shop	16.251591
3	Blue Ribbon Sushi	15.999580

Table 2. - Avg Cost of top Restaurants

Multivariate Analysis - Restaurant vs Cost of the order

```
df.groupby(['restaurant_name'])['cost_of_the_order'].sum().sort_values(ascending = False).head(5)
```

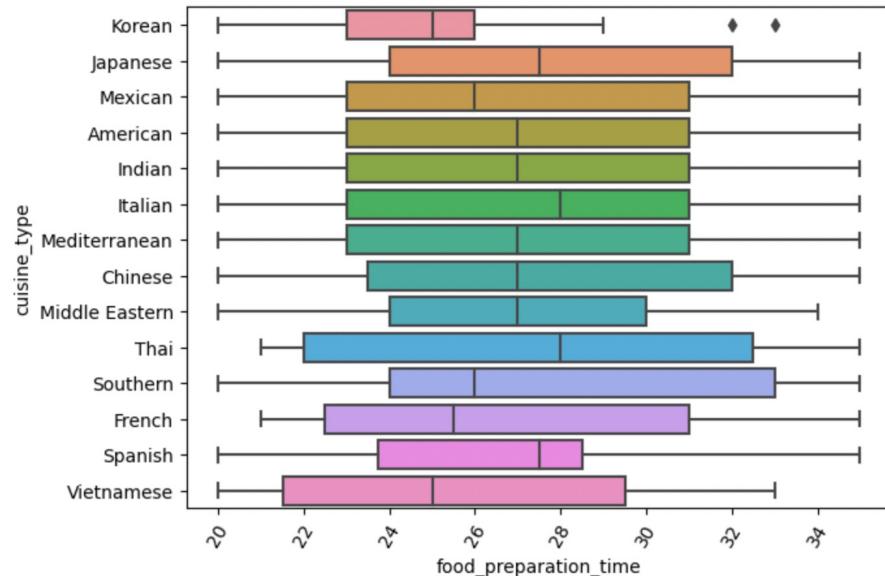
restaurant_name	cost_of_the_order
Shake Shack	3579.53
The Meatball Shop	2145.21
Blue Ribbon Sushi	1903.95
Blue Ribbon Fried Chicken	1662.29
Parm	1112.76

Name: cost_of_the_order, dtype: float64

- Shake Shack, The meatball Shop are also the highest revenue generating restaurants.

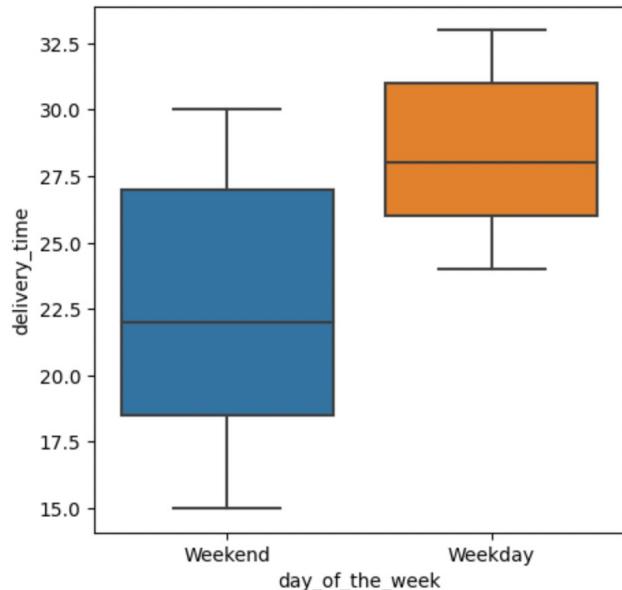
Multivariate Analysis - Cuisine vs Food Preparation time

- Thai and Italian cuisine has highest median food Preparation time.
- Orders for these cuisines might take longest to reach to Customer after order is placed.



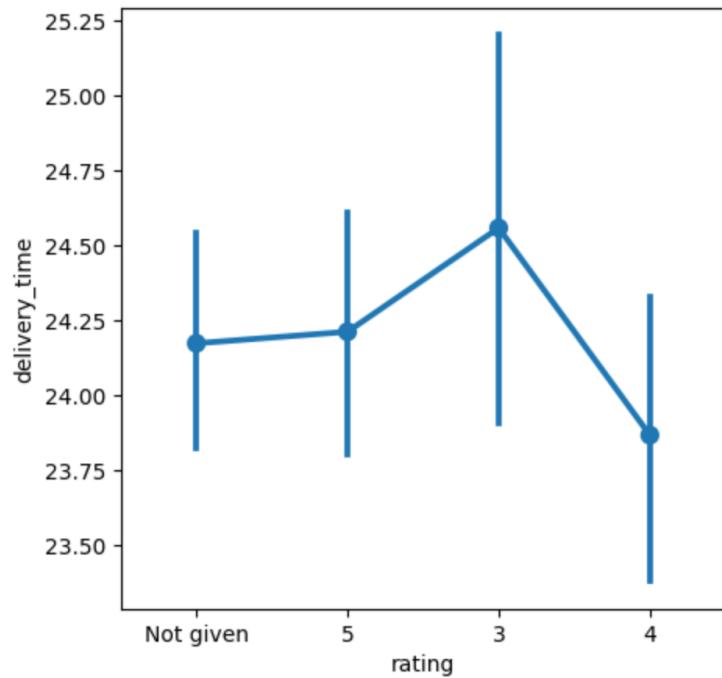
Multivariate Analysis – Day of the Week vs Delivery time

- Median Delivery time is longest on weekday.
- We need more data on this for further analysis, time of the day when the order is made should also be considered.
- Apparently orders placed during rush hours of the weekdays take more time to reach in contrast to orders placed on weekends when City Roads are usually not jam packed.
- Also there are more volunteers delivering the orders on weekends when order number is also more, which is in turn reducing the delivery time.



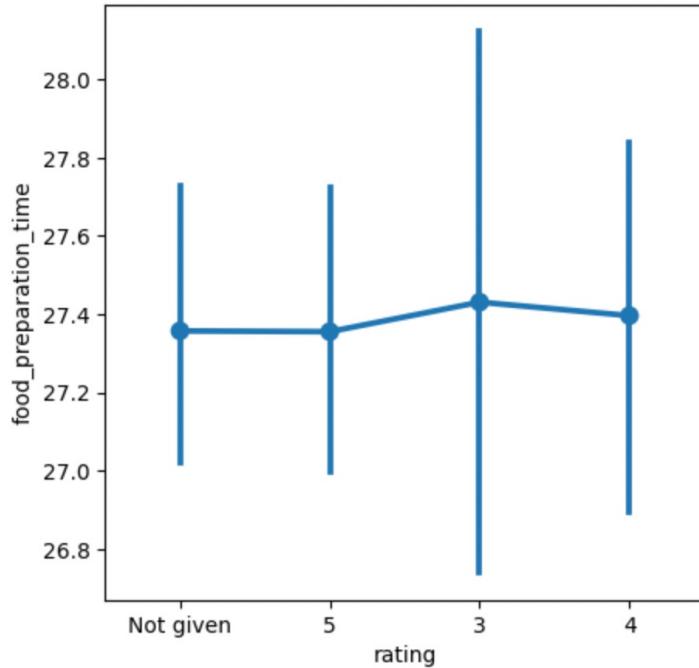
Multivariate Analysis – Rating vs Delivery time.

- Restaurants with Rating 3 have longest food delivery time.
- Apparently restaurants with rating 5 have longer food delivery time when compared to restaurants with rating 4.
- In above case there could be some factor other than just delivery time which is deciding the rating.
- Hence delivery time doesn't seem to be the sole determining factor for the ratings given



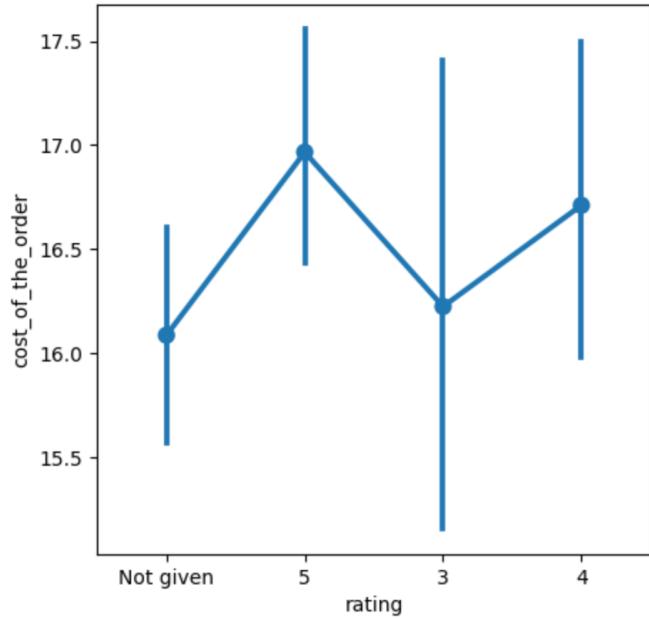
Multivariate Analysis – Rating vs food Preparation time.

- Food Preparation time does not seem to affecting the ratings.
- It can not be the deciding factor behind rating given to the restaurants.



Multivariate Analysis – Rating vs cost of the order.

- Cost of the order seems to be affected by the ratings given to the order.
- Cost of the order is increasing with the increase in ratings.
- Evidently, restaurants with high rating can ask for higher price.
- Food company should reach back to customers via app and should ask them to rate their order, this would help restaurants to increase their average rating and generate better revenue.



Questions :-

Q 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

A

restaurant_name	rating
The Meatball Shop	4.511905
Blue Ribbon Fried Chicken	4.328125
Shake Shack	4.278195
Blue Ribbon Sushi	4.219178

Q 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

A The net revenue is around 6166.3 dollars

Q The company wants to analyse 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)

A Percentage of orders taking above 60 mins: 10.54 %

Q The company wants to analyse the delivery time of the orders on weekdays and weekends. How does the mean delivery time vary during weekdays and weekends?

A The mean delivery time on weekdays is around 28 minutes
The mean delivery time on weekend is around 22 minutes

APPENDIX



Happy Learning !

