Statistical Methods for Decision Making

PROJECT REPORT

Contents:

Problem Statement

Context

The number of restaurants in New York is increasing day by day. Lots of students and busy professionals rely on those restaurants due to their hectic lifestyles. Online food delivery service is a great option for them. It provides them with good food from their favorite restaurants. A food aggregator company FoodHub offers access to multiple restaurants through a single smartphone app.

The app allows the restaurants to receive a direct online order from a customer. The app assigns a delivery person from the company to pick up the order after it is confirmed by the restaurant. The delivery person then uses the map to reach the restaurant and waits for the food package. Once the food package is handed over to the delivery person, he/she confirms the pick-up in the app and travels to the customer's location to deliver the food. The delivery person confirms the drop-off in the app after delivering the food package to the customer. The customer can rate the order in the app. The food aggregator earns money by collecting a fixed margin of the delivery order from the restaurants.

Objective

To analyze the dataframe to get a fair idea about the demand of different restaurant which help them in enhancing their customer experience. Organize data analysis to find some insights that will help the company to improve the business.

Data Dictionary

Column name	Description	Data Type
Order Id	Unique ID of the order	int64
Customer Id	ID of the customer who ordered the food	int64
Restaurant Name	Name of the restaurant	Object
Cuisine Type	Cuisine ordered by the customer	Object
Cost of the Order	Cost of the Order	Float64
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)	Object
Rating	Rating given by the customer out of 5	Object
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.	Int64
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	Int64

Criteria

Data Overview

Import the required libraries - Overview of the dataset shape, datatypes - Statistical summary and check for missing values - Answer all the key questions asked in this section

Question 1: How many rows and columns are present in the data?
Observation: The DataFrame has 1898 rows and 9 columns

Question 2: What are the datatypes of the different columns in the dataset?
Observation: There is 4 int64 datatype are present, 4 object datatypes are present and also 1 float64 are present in the DataFrame.

Question 3: Are there any missing values in the data? If yes, treat them using an appropriate method.
Observation: There is no missing value or null value present in the DataFrame.

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?

Observation:

1- The average cost of an order is ~16.5, the average preparation time for an ordered food is 27.4 min, and the average delivery time for an order is 24.2 min.

2- In these three variables, none of them seem to have outliers.

3- Cost of the Order has slightly right skewed.

4- The minimum, average, and maximum times it takes for food to be prepared are 20 min, 27.4 min, and 35 min, respectively.

Question 5: How many orders are not rated?
Observation: Total 736 orders are not rated in the DataFrame.

Univariate Analysis

Explore all the variables and provide observations on the distributions of all the relevant variables in the dataset - Observation and Insights - Answer all the key questions asked in this section.

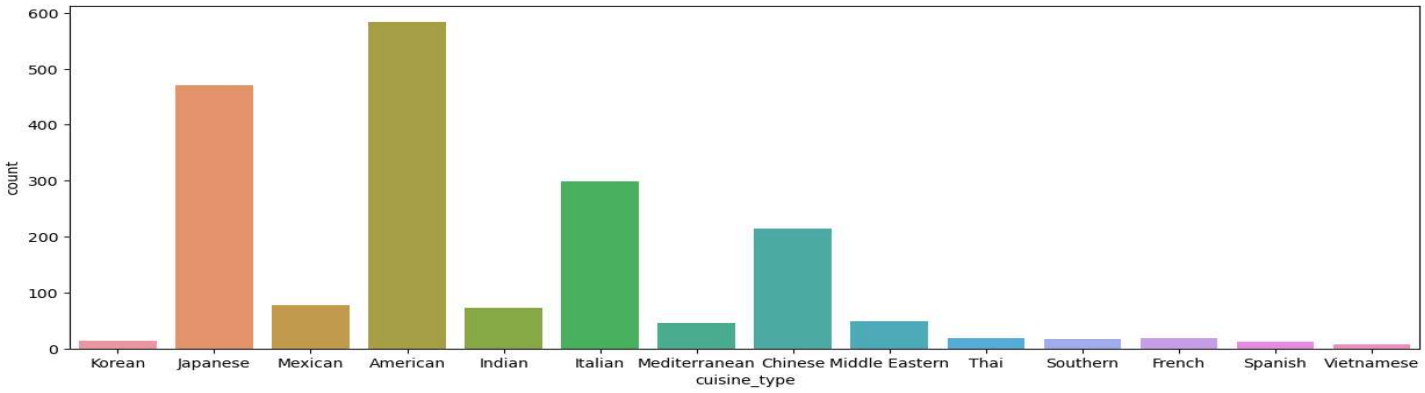
Question 6: Explore all the variables and provide observations on their distributions. (Generally, histograms, boxplots, countplots, etc. are used for univariate exploration.)

6.1: Order ID
Observation: There are 1898 order id in the DataFrame.

6.2: Customer ID
Observation: There are 1200 customer id present in the DataFrame.

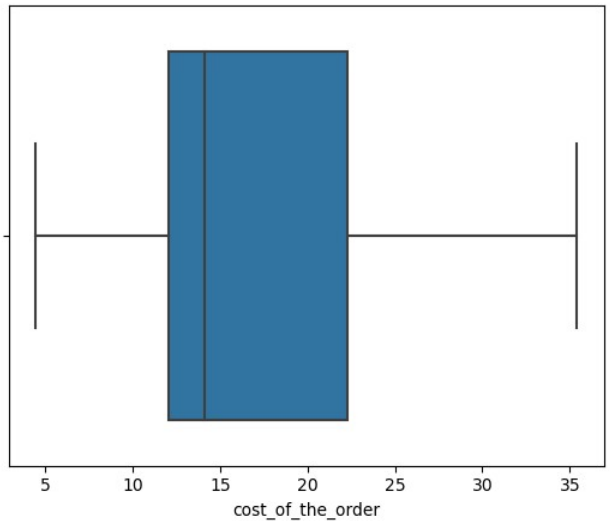
6.3: Restaurant name
Observation: There are 178 unique restaurant name present in the DataFrame.

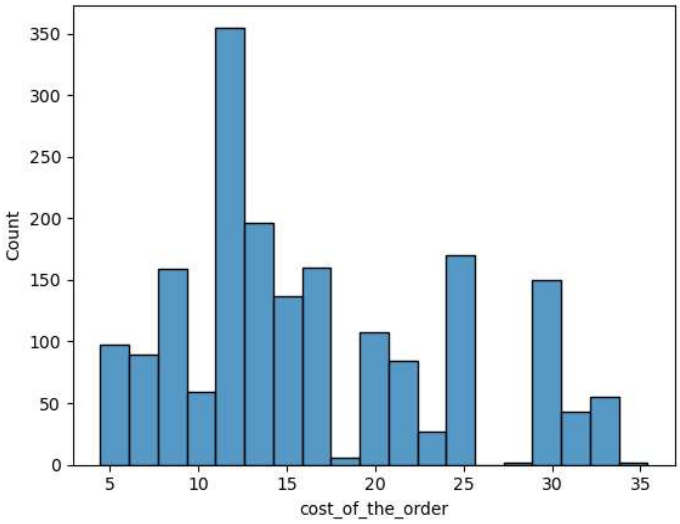
6.4: Cuisine type
Observation: There are 14 unique cuisine type present in the DataFrame.



Observation: From the above plot the most popular cuisine is found to be American. The least popular cuisine is Vietnamese.

6.5: Cost of the order





Observation:

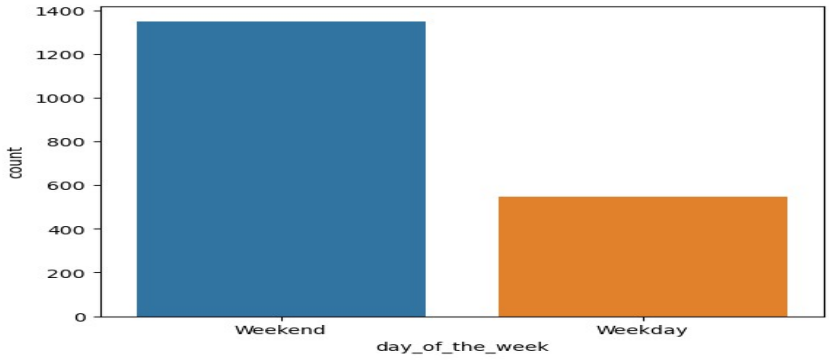
1- The distribution is slightly right-skewed which indicates the mean is greater then the median.

2- The cost of an order ranges from almost 4.5 to 35.4, while the median cost is 14.1.

3- A large portion of order costs are between 10 and 17 approximately.

6.5: Day of the week

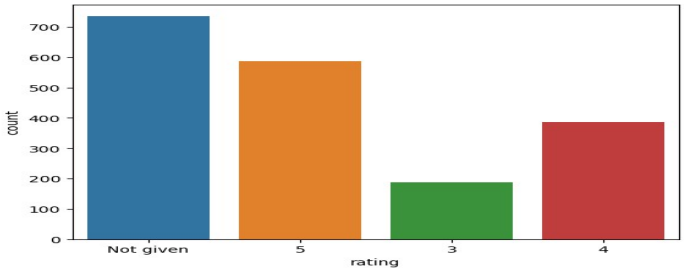
Observation: There are 2 unique values are present in the day of the week column 'Weekday' and 'Weekend'.



Observation: From the above plot we see that most of the order are placed during weekends.

6.6: Rating

Observation: Most of the customer given more than 3 rating or have not given any rating.

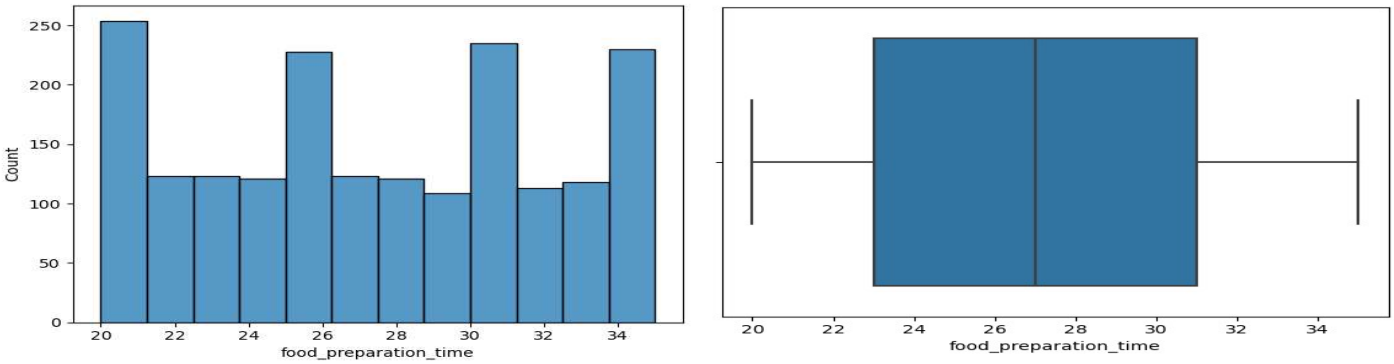


Observation: From the above plot it is very clear that the number of those who did not give rate on the order is more than those who rated 3,4,5.

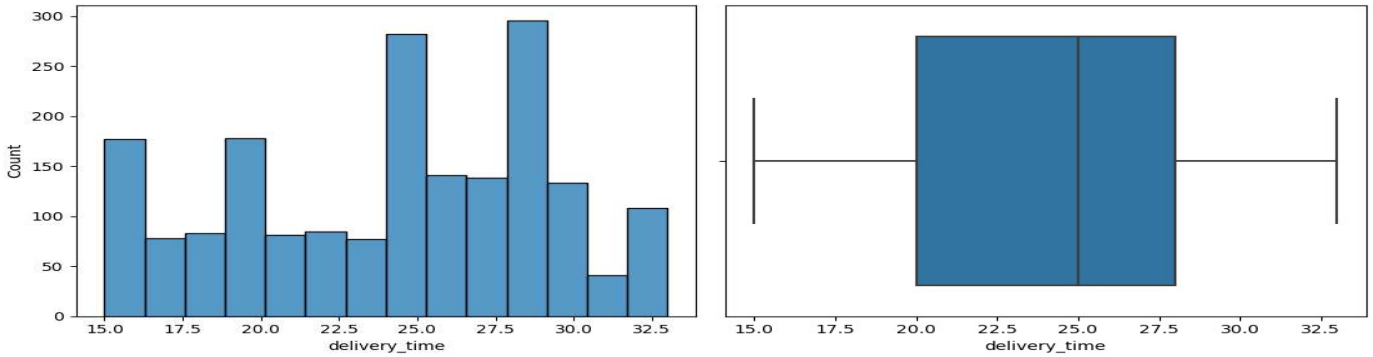
6.7: Food Preparation time

Observation:

1- The distribution is relatively uniform and mean and median are very similar.
2- Each food preparation takes between 20 and 35 min.



6.8: Delivery time



Observation:

1- The distribution is not much skewed and mean and median are close.

2- Orders take between 15 and 33 min to be delivered.

Question 7: Which are the top 5 restaurants in terms of the number of orders received?

Observation: The top 5 reataurants in terms of the number of orders received are 'Shake Shack ','The Meatball Shop','Blue Ribbon Sushi','Blue Ribbon Fried Chicken' and 'Parm'.

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

Observation: The most popular cuisine on weekends is American, with 415 number.

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

Observation: The number of total orders that cost more than 20 dollars is 555 and the percentage of orders above 20 dollars is 29.24%.

Question 10: What is the mean order delivery time?

Observation: The mean delivery time for this dataset is 24.16 minutes.

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.

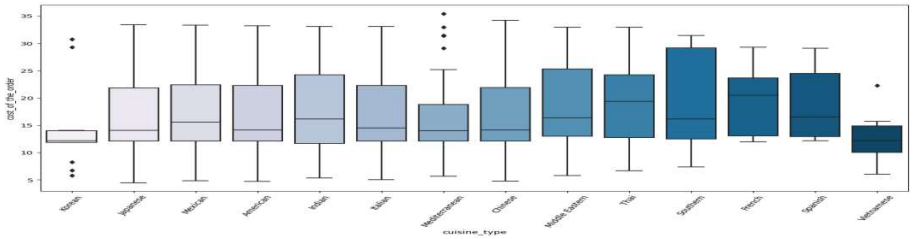
Observation: The top most frequent custoner id is '52832' and the number of orders is '13'.

Multivariate Analysis

Perform bivariate/multivariate analysis to explore relationships between the variables in the dataset - Observation and Insights - Answer all the key questions asked in this section

Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

12.1: Cuisine vs Cost of the order

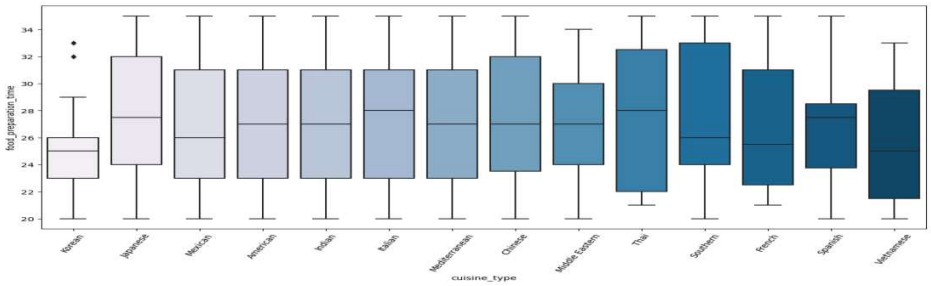


Observation:

1- The highest mean costs belong to the orders for French, Middle Eastern, Spanish, and Thai cuisine types.

2- The lowest mean costs belong to the orders for Korean and Vietnamese cuisine types. The same cuisine types seem to have the smallest dispersion.

12.2: Cuisine vs Food Preparation time

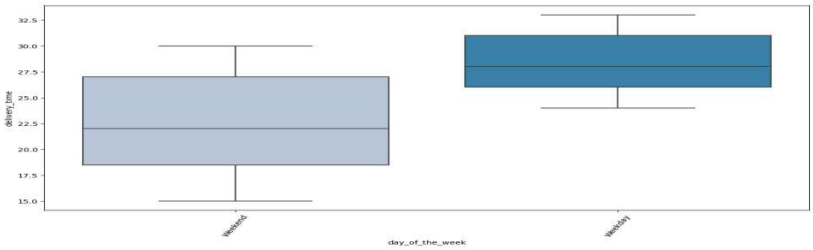


Observation:

1- The mean food preparation times are close, but the highest and lowest among them belong to the Southern and Korean cuisine types, respectively.

2- The minimum and maximum food preparation times are almost similar across cuisine types, except Korean, which has a relatively smaller maximum preparation time.

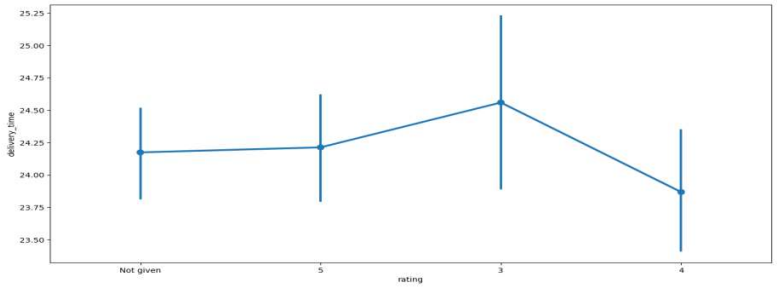
12.3: Day of the Week vs Delivery time



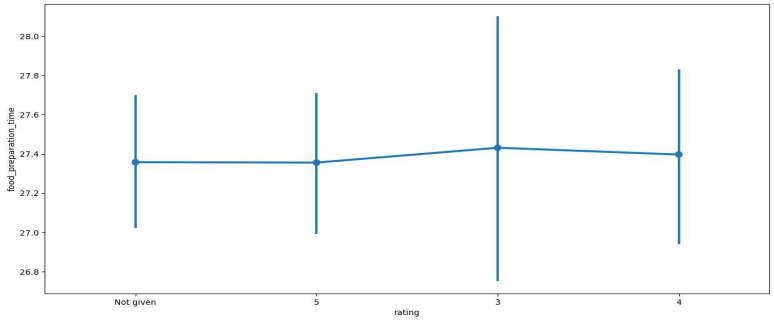
Revenue generated by the restaurant:

The restaurent name 'Shake Shack' is most revenue generated that is 3579.53 on the cost of the order.

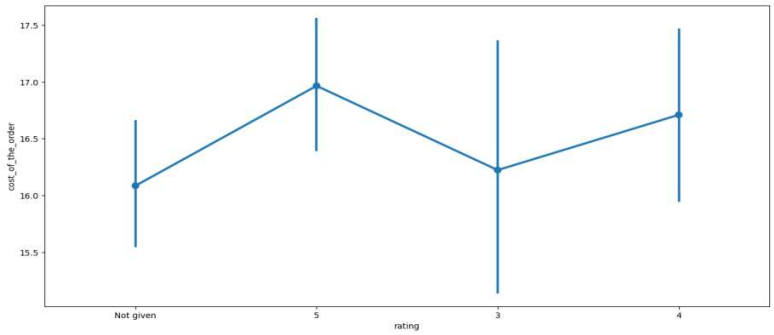
12.4: Rating vs Delivery time



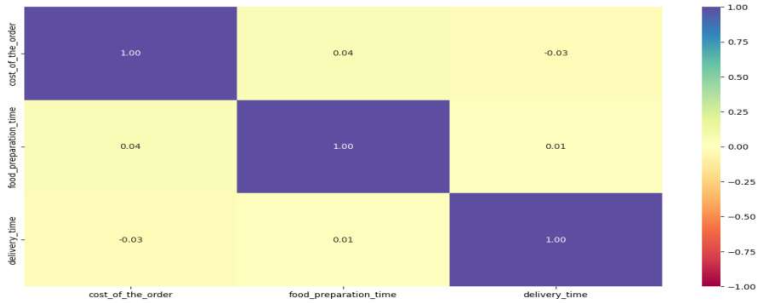
12.5: Rating vs Food preparation time



12.6: Rating vs Cost of the order



12.7: Correlation among variables



Observation: From the heatmap we see that the correlation between 'cost of the order', 'food preparation time' and 'delivery time' are correlated.

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.

Observation: The four restaurants fulfilled the specified criteria are 'Blue Ribbon Fried Chicken', 'Blue Ribbon Sushi', 'Shake Shack', and 'The Meatball Shop'.

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.

Observation: The net revenue on the dataset is around 6166.3 dollars.

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.)

Observation: The number of total time that have 60 min above is 200 and percentage of total time avobe 60min 10.54 %.

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?

Observation: The mean delivery time on weekdays is around 28 minutes and the mean delivery time on weekends is around 22 minutes.

Actionable Insights and Recommendations

Actionable Insights – Business Recommendations.

Question 17: What are your conclusions from the analysis? What recommendations would you like to share to help improve the business? (You can use cuisine type and feedback ratings to drive your business recommendations.)

Conclusions:

From the analysis of the dataframe it is very clear that the food aggregator company need to change something. rating is very low, there are four cuisine types have the most number of restaurant, there is no significant correlation found between some variables like order cost, food preparation time, delivery time, and customer rating.

There is some Recommendations are provided bellow for the above analysis:

Recommendations:

1- The number of orders which is not been rated is large number to the total number of orders (736 out of 1898). It is recommended to the company considers approaches to motivate its customers to rate their orders so more reliable conclusions could be made based on the ratings.

2- American, Japanese, Italian, and Chinese cuisine types are the top four cuisine types. These cuisine types have the most number of restaurants, too. That said, it appears that if the company adds to the number of restaurant choices in other cuisines, their number of orders/customers may increase, too.

3- No significant correlation was found between order cost, food preparation time, delivery time, and customer rating.

4- The highest 3-rated orders were for the Vietnamese cuisine, making it of the lowest average rating among all cuisine types. The company may need to find new and more valueable Vietnamese restaurants with better quality and add them to their restaurant options.

5- The majority of the orders are placed on the weekend and the number of unique customers ordering food on the weekend is almost double that during the weekdays. The company may come up with discounts to increase the number of orders and customers on the weekdays.