

Foundations for Data Science: FoodHub Data Analysis

Ayo Adeniran

Monday July 31, 2023

Overview

- 1 Background
- 2 Observations and Analysis from Data
- 3 Exploratory Data Analysis (EDA)
 - Univariate Analysis
 - Multivariate Analysis
- 4 Conclusions
- 5 Recommendations

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.

Context (ctd)

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

- Load the data.
- Analyze the data (to get a fair idea about the demand of different restaurants which will help them in enhancing their customer experience).
- Answer the key questions posed by the Data Science team.
- Perform the data analysis to find answers to these questions that will help the company to improve the business.

Data Dictionary

The detailed data dictionary is given below.

- *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_of_the_order*: Cost of the order
- *day_of_the_week*: Indicates whether the order is placed on a weekday or weekend (Weekday is Mon. - Fri.; Weekend is Sat - Sun.)
- *rating*: Rating given by the customer out of 5
- *food_preparation_time*: Time (in minutes) taken by the restaurant to prepare the food.
- *delivery_time*: Time (in minutes) taken by the delivery person to deliver the food package.

Preliminary observations

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

Answer: The data has 1898 rows and 9 columns.

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

Answer: There are 4 columns with integer entries (the order ID, customer ID, food preparation time and delivery time) , 1 column with float entries (the cost of the food ordered), and 4 columns whose entries are 'object' data types (the restaurant name, cuisine type, day of the week, and rating - typically on a scale of 1 to 5 as well as the cases where no rating was given).

Question 3: Are there any missing values in the data? If yes, treat them using an appropriate method. Answer: There are no missing values in the data.

Preliminary observations (Ctd)

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: According to the data, the minimum time it takes to get the food prepared once an order is placed is 20 minutes. The average time it takes is 27.4 minutes and the maximum time is 35 minutes. Once picked up, the minimum time it takes for the food to be delivered is 15 minutes, on average it takes 24.2 minutes, and the maximum it takes is 33 minutes.

Question 5: How many orders are not rated?

Answer: There are 736 orders that are not rated. So, there are 1162 orders each with a rating of 3, 4, or 5.

Univariate Analysis

Question 6: Explore all the variables and provide observations on their distributions.

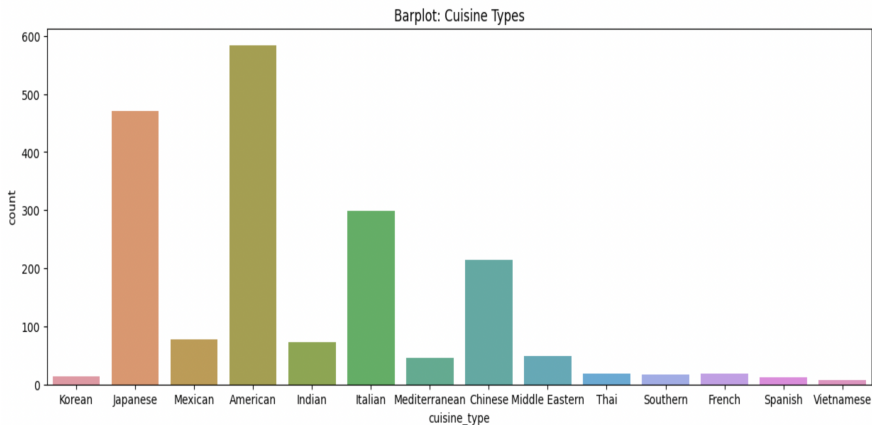
Answer:

While there are 1898 total orders, there are a total of 1200 unique customer IDs in the data set.

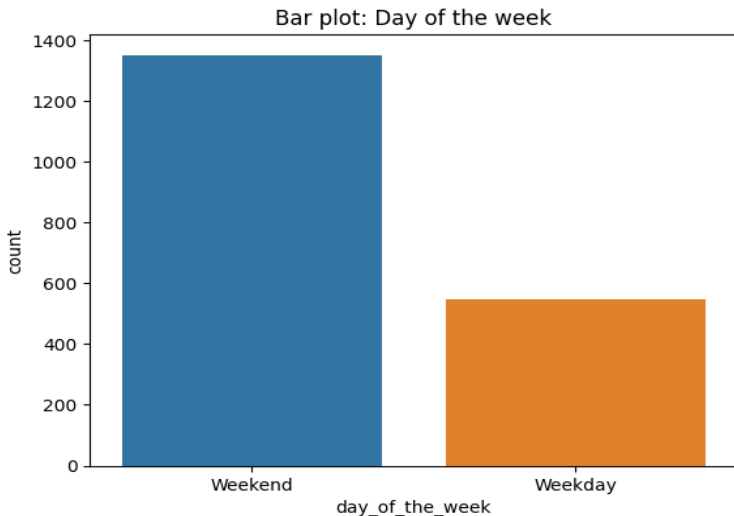
There are a total of 178 different restaurants in the dataset and 14 unique cuisine types ordered across all these different restaurants.

The cuisine type that was ordered the most was American (584), followed by Japanese (470), and Italian (298). The least ordered meals were Korean(13), Spanish(12), and Vietnamese (7).

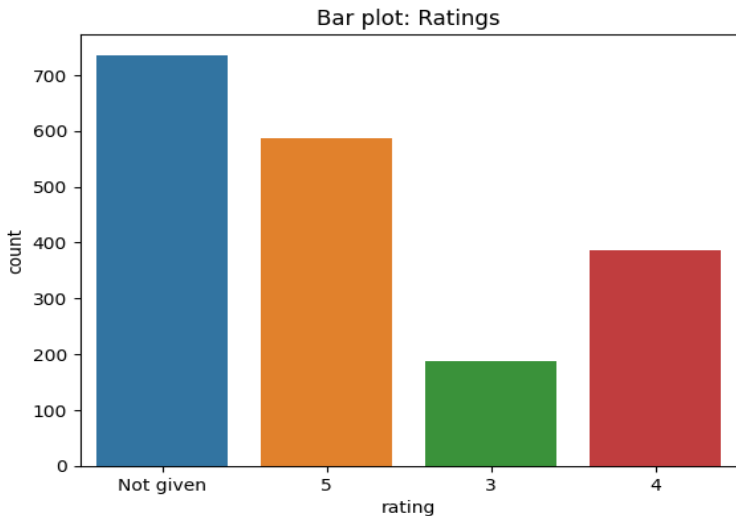
Cuisine Type



Day of Week

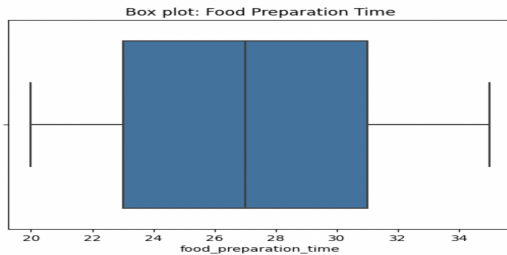
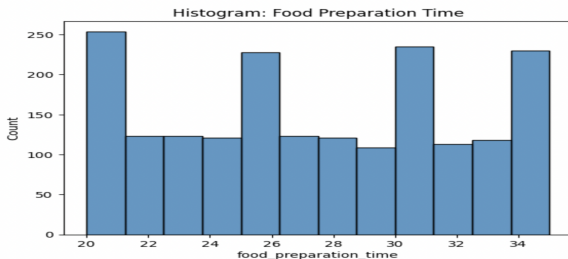


Ratings



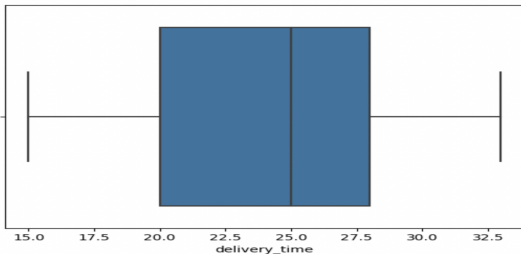
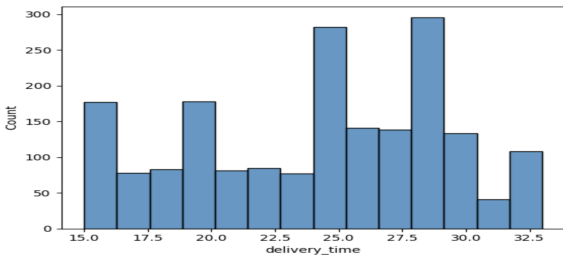
Univariate Analysis

Food preparation time



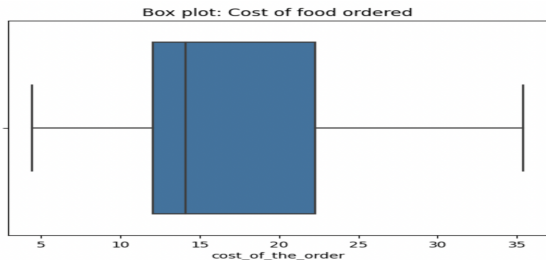
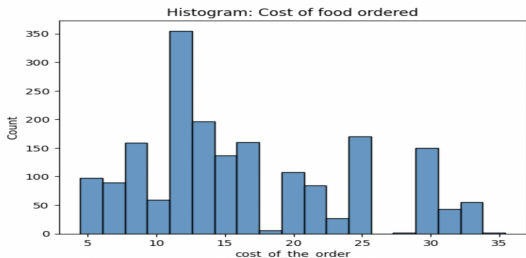
Univariate Analysis

Delivery time



Univariate Analysis

Cost of food ordered



EDA - Multivariate Analysis

The data is divided into two classes – food ordered on the weekend versus food ordered on weekdays. The bar chart shows that the volume of orders placed on weekends is more than twice the orders placed on weekdays.

There were in general 3 ratings left (3, 4, and 5). For the majority of orders, there were no ratings left (this is consistent with reality as many people would generally leave no ratings unless the food or delivery was either very great or very poor).

For those who left ratings, the number of 5-star ratings was roughly equivalent to the total number of 4-star and 3-star ratings combined. There were no 1-star or 2-star ratings. This seems to indicate that majority of clients were satisfied with their order.

EDA ctd

Although on average, it took 27 minutes to get orders prepared, the highest number of meals took 20 minutes to get prepared. Most meals (approx. 75%) took about 20 - 31 minutes to be prepared.

The median delivery time was 25 minutes while the most frequent time for delivery was 28 minutes and 45 seconds. All meals were delivered in the 15 - 33 minutes time range.

The number of orders that have above 60 minutes of total delivery time is 200 (i.e. 10.54 %)

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

Other Questions

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

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

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

There were 14 unique cuisines available on weekends and American cuisine is the most popular cuisine on weekends.

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

The number of total orders that cost above \$20 is 555 (i.e. 29.24%)

Question 10: What is the mean order delivery time? The mean delivery time was 24.16 minutes.

Incentives

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

Customer ID	Number of orders places
52832	13
47440	10
83287	9

Question 13: The company wants to provide a promotional offer in the ads of the restaurants. Here are the restaurants fulfilling the criteria to get the promotional offer.

Restaurant name	No. of Ratings	Ave. Rating
Shake Shack	133	4.511905
The Meatball Shop	84	4.328125
Blue Ribbon Sushi	73	4.278195
Blue Ribbon Fried Chicken	64	4.219178

Other observations

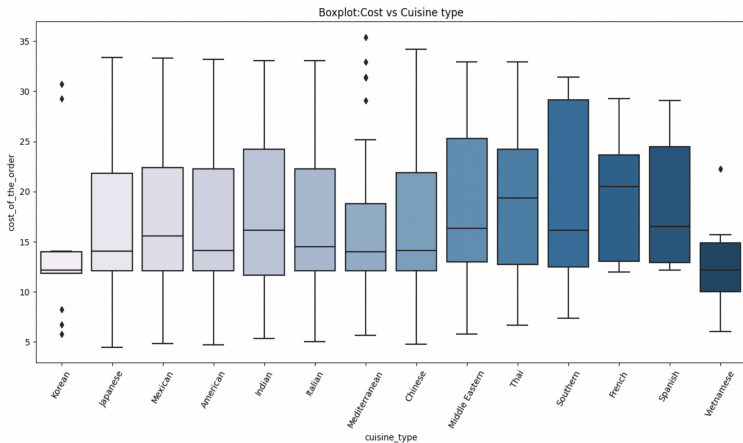
Restaurant name	Revenue generated
Shake Shack	3579.53
The Meatball Shop	2145.21
Blue Ribbon Sushi	1903.95
Blue Ribbon Fried Chicken	1662.29
Parm	1112.76
RedFarm Broadway	965.13
RedFarm Hudson	921.21
TAO	834.50
Han Dynasty	755.29
Blue Ribbon Sushi Bar & Grill	666.62
Rubirosa	660.45
Sushi of Gari 46	640.87
Nobu Next Door	623.67
Five Guys Burgers and Fries	506.47

Observations about Revenue

- The total revenue generated by all the restaurants was roughly about \$17,000.
- The top 5 restaurants on the list above generated \$3000 more than the remaining 9 restaurants.
- The most revenue was generated by Shake Shack and was 7 times the least revenue which was generated by Five Guys Burgers and Fries.

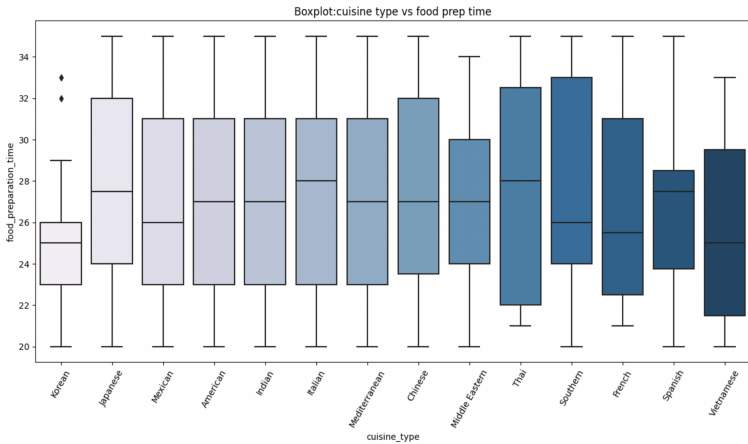
Multivariate Analysis

Cost vs Cuisine Type



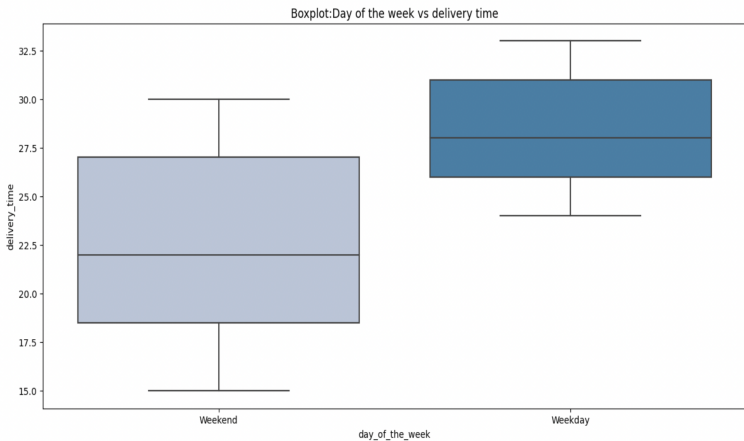
Multivariate Analysis

Cuisine Type vs Food preparation Time



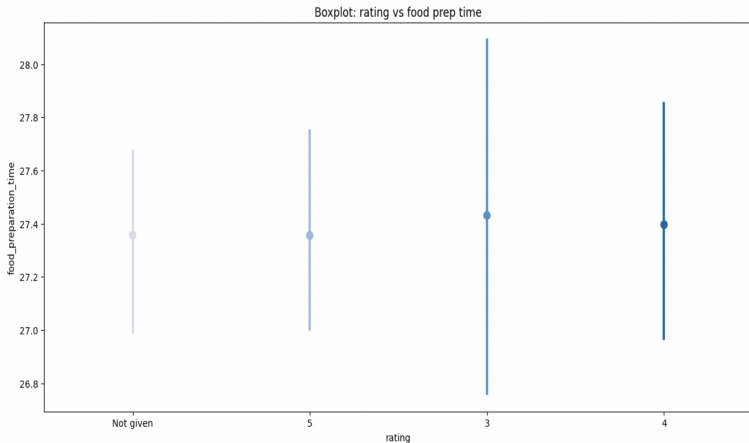
Multivariate Analysis

Day of week vs Delivery time



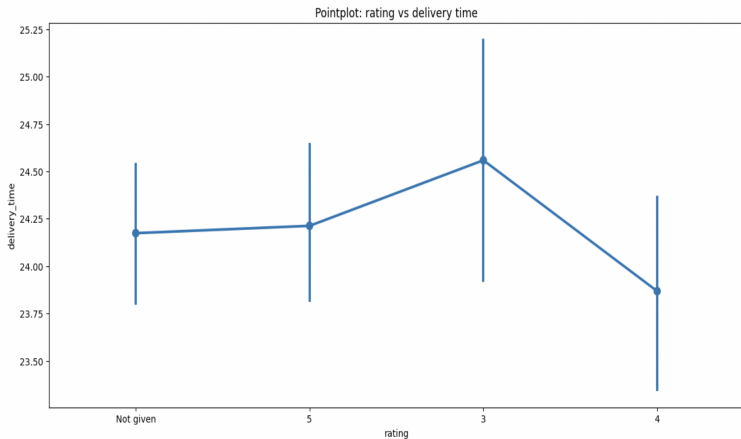
Multivariate Analysis

Rating vs Food Prep time



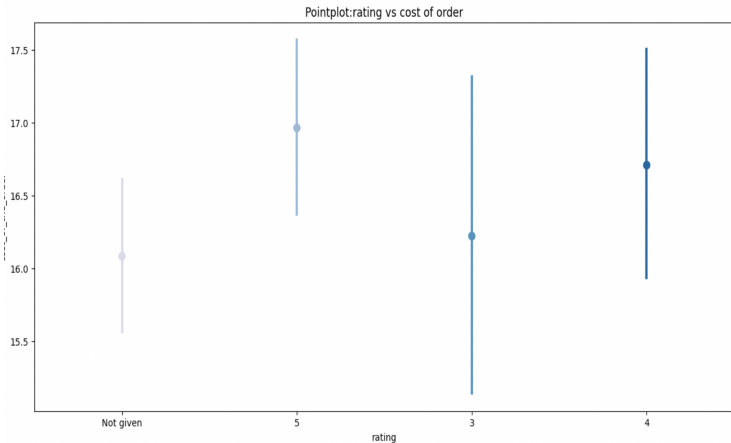
Multivariate Analysis

Rating vs Delivery time



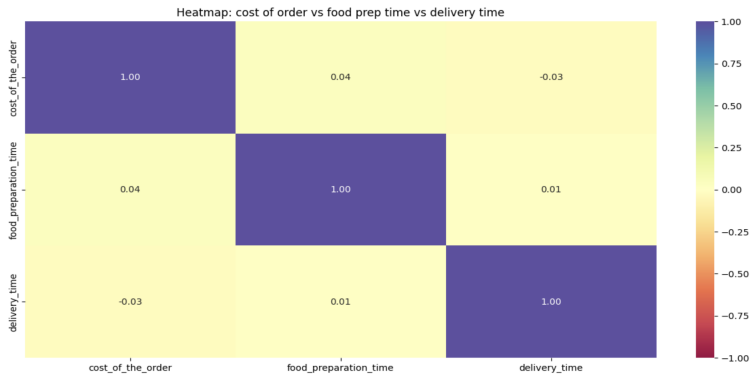
Multivariate Analysis

Rating vs Cost of order



Multivariate Analysis

Cost of order vs Food Prep time vs Delivery time



Conclusions

- As we would expect, those who had to wait a little bit longer than others for their food to be delivered (roughly about 24.5 minutes) seemed to give a lower rating of 3 on average.
- There wasn't much significant variation between ratings given and food preparation time.
- On average, those who paid the most (\$17) seemed to give the worst ranking (3) whereas those who paid slightly lower seemed to give better reviews.
- It takes longer (approx. 33% more time) for food to be delivered on weekdays than during weekends.
- About 10.54% of meals ordered through the app take over an hour to be made and delivered.
- The net revenue for FoodHub is roughly \$6200.

Conclusions (ctd)

- From the summary earlier, we saw that on average, meals ordered costs \$16.50. The cuisines people ordered the most (350 orders) cost them roughly about \$11.67 per order.
- From the boxplot, we can see that the cost of the orders was slightly skewed to the right – this means that people were more likely to order meals that were considered to be cheaper (roughly \$15 or less) than meals that were considered expensive (greater than \$15).
- There is a slightly positive correlation between cost of orders and food prep time, but this is not very significant. Interestingly, there is a slightly negative correlation between cost of orders and delivery time. Food prep time and delivery time were positively correlated but this relationship is very negligible.

Conclusions (ctd)

- The median time for food preparation was highest for Thai and Italian cuisines. It was lowest for Korean cuisine, even though there were a few outliers that took over 32 minutes.
- The median cost of French cuisine was the highest of all the various cuisine types whereas Korean and Vietnamese cuisines generally had the lowest median in terms of cost of orders.
- Considering the ratings given in light of cost of food ordered, it seems that slightly higher prices did not deter customers from giving better ratings (4 or 5), as one might have initially thought.

Recommendations

- In the future, it would be good to know the period under review (i.e. collect date and time data) so as to be able to analyze the net revenue for the company with respect to time.
- Foodhub should negotiate an increase in the percentage per order for the 5 restaurants with the highest revenue in order to increase its overall revenue.
- The company should consider giving more incentives for customers to leave more ratings as well as higher ratings.
- The company needs to investigate why 10.5% of meals take an hour or more to be made and delivered and see how this number can be reduced and/or minimized.