FoodHub Data Analysis

Project Python Foundations: PGP Data Science & Business Analytics

Sai Hardiker
November 9th, 2022

Contents / Agenda

- Executive Summary

- Business Problem Overview and Solution Approach

- Data Overview

- EDA - Univariate Analysis

- EDA - Multivariate Analysis

- Appendix

Executive Summary

- Conclusions & Actionable Insights:

  - More orders were placed on the weekends

  - More than 1/3 of total orders are from repeat customers

  - Ratings in order of highest occurrence are Not given, 5, 4, and 3

  - Shake Shack was the most ordered from restaurant because it is a popular and well-known chain

  - Small businesses/lesser know restaurants have very few orders

  - Orders take longer to be delivered on weekdays

  - More than 10% orders took longer than 1 hour

Executive Summary

- Business Recommendations:

  - Add an incentive for customers to provide ratings on their order to gather more data such as a coupon for rating 3 consecutive orders & add a reason tab while rating that can easily identify issues

  - Most popular cuisines, American, Japanese, Italian, and Chinese, have right skewed order costs meaning they have mostly lower prices, so including more options, specifically higher end resteraunts with more expensive items, will help gain more revenue

  - Adding more famous chains will make people more likely to order from them as they are familiar and trusted options, like Shake Shack

  - To increase customer base, increase advertising for FoodHub and a promotional offer such as a discount on the first 3 orders to gain more loyal customers who will order multiple times

  - Have more delivery runners ready in a wider range of locations on weekdays to expedite delivery times

  - Lower delivery prices on weekdays because it will take longer and increase them on weekends for surplus of orders

  - Encourage customers to add ratings and reviews for small businesses to make them more appealing & give promotional offers for first order from restaurants with fewer orders

  - Offer a discount on a customer's next order if it takes longer than an hour to be delivered

# Business Problem Overview and Solution Approach

- The company FoodHub has a databank containing information on orders made by costumers on their website and they wish to analyze this data to gauge the different demands in the restaurant ordering industry in order to be able to better serve their customers.

- I will be conducting an exploratory data analysis on the measures provided by the company, surveying different measures such as days of the week, cuisine type, cost, rating, and delivery time to find patterns in customer behavior. By using graphs and measures of central tendencies, I will be able to gain some insight on the types of orders that gain the highest customer satisfaction and ensure future engagement in the business.

Data Overview

- The data has 9 columns and 1898 rows, meaning there are 1898 sets of the 9 different data measures, with no missing values in the data.

- Datatypes:

  - 4 integer types of data (order id, customer id, food preparation time, and delivery time)

  - 1 float integer type (order cost)

  - 4 object data types (restaurant name, cuisine type, day of the week, and rating)

- Food preparation time:

  - Minimum: 20 minutes

  - Maximum: 35 minutes

  - Average: 27.37 minutes

- 736 of the total orders (38.78%) were not rated by customers
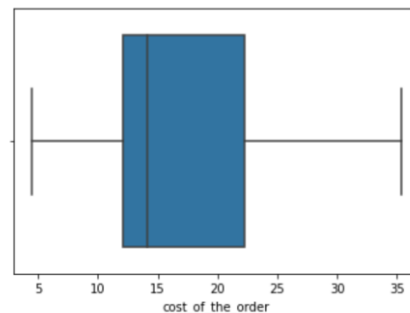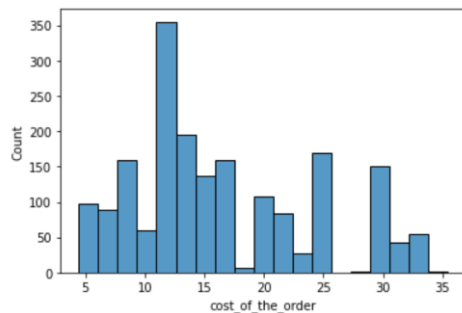
Univariate Analysis

- Order ID: 6-11

  - 1898 unique order ID's, meaning no repeats or missing values

- Customer ID:

  - 1200 unique customer ID's, meaning all orders were placed by a total of 1200 customers & 698 orders from repeat customers

- Restaurant name & Cuisine type:

 - 178 total restaurants encompassing 14 different cuisines received orders from FoodHub
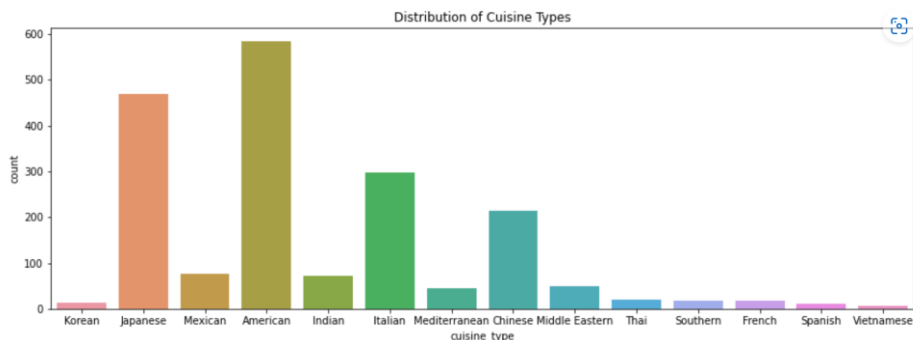
Through this countplot of the cuisine types, we can see the most popular ones are Japanese, American, Italian, and Chinese.

Univariate Analysis

- Order cost:

  - In the boxplot and histogram below, we can tell the distribution of order cost is slightly skewed to the right, so most of the orders are less expensive, and the average cost is around $15 with a max around $35 and a min about $5 with seemingly no outliers.
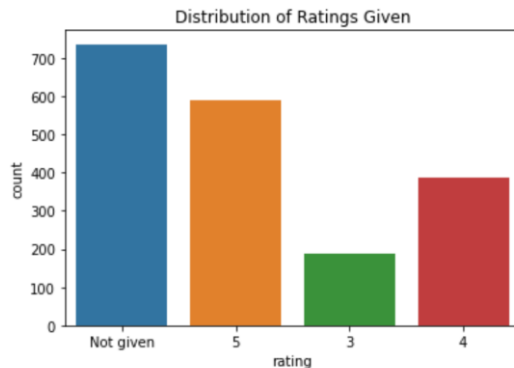

- Day of the week:

  - Looking at the distibution below, we can gather that more orders are placed on the weekend than on weekdays
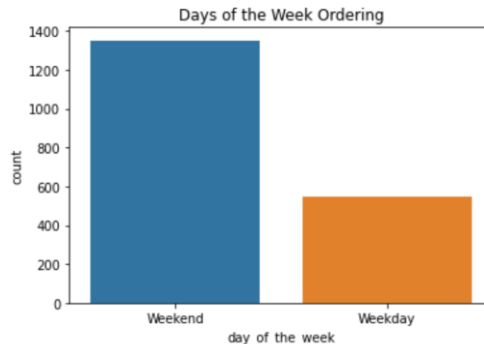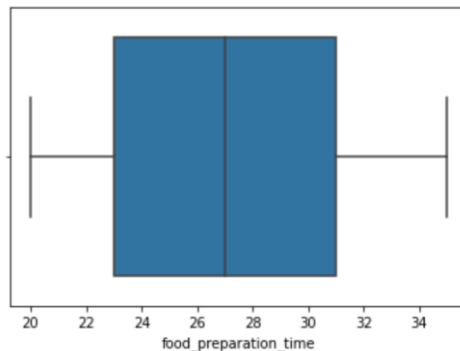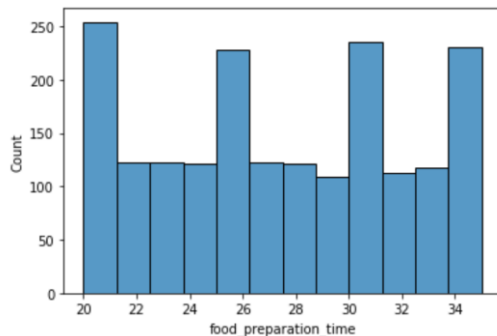
Univariate Analysis

- Rating:

  - While most orders went unrated, many of them were also rated 5 star, some were rated 4, and the rating 3 stars had the least frequency
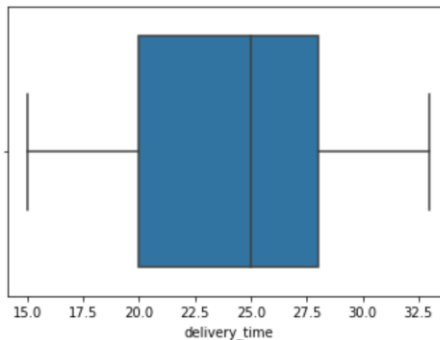

- Food Preparation Time:

  - We see peak occurences in the food preparation times around 21, 25, 31, and 34 minutes with the average being 27 minutes.

Univariate Analysis

- Delivery time:

    - The delivery times are slightly skewed to the left, with the average being 25 minutes with a min of 15 min and max around 32.5 min
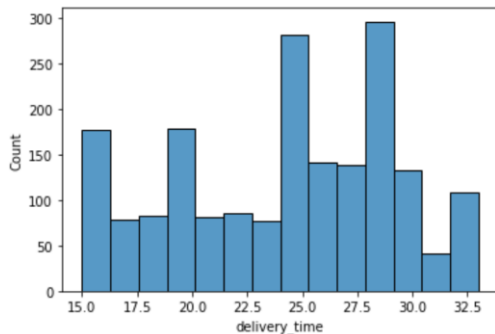
- Top 5 restaurants ordered from:

    - Shake Shack (219 orders, 11.54%)

    - The Meatball Shop (132 orders, 6.95%)

    - Blue Ribbon Sushi (119 orders, 6.27%)

    - Blue Ribbon Fried Chicken (96 orders, 5.06%)

    - Parm (68 orders, 3.58%)

Univariate Analysis

- Most popular cuisine on weekends:

  - American food had 415 orders on the weekends, making up roughly 30% of all weekend orders.

- Orders more than $20:

  - There were 555 orders totaling more than $20, which is 29.24% of all orders

- Mean delivery time:

  - The average time it took orders to be delivered in this dataset is roughly 24.16 minutes

- Discount vouchers:

  - The ID's of the top 3 most frequent customers who qualify for the 20% discount voucher are:

    - 52832 (13 orders)

    - 47440 (10 orders)

    - 83287 (9 orders)

Multivariate Analysis

- Cuisine vs. Cost of order:

    - From the boxplot, we can see that Korean, Mediterranean, and Vietnamese cuisines have outliers in the cost of orders. This could be for a few reasons such as specialty items on the menu that few order, some may have ordered for multiple people, or some may have ordered a small item/appetizer.

    - Many of the other cuisines seem to be right skewed, meaning more of the data points are higher amounts, which is understandable since food has become a lot more expensive, especially when ordering through a delivery service.

Cuisine Type vs. Cost of the Order
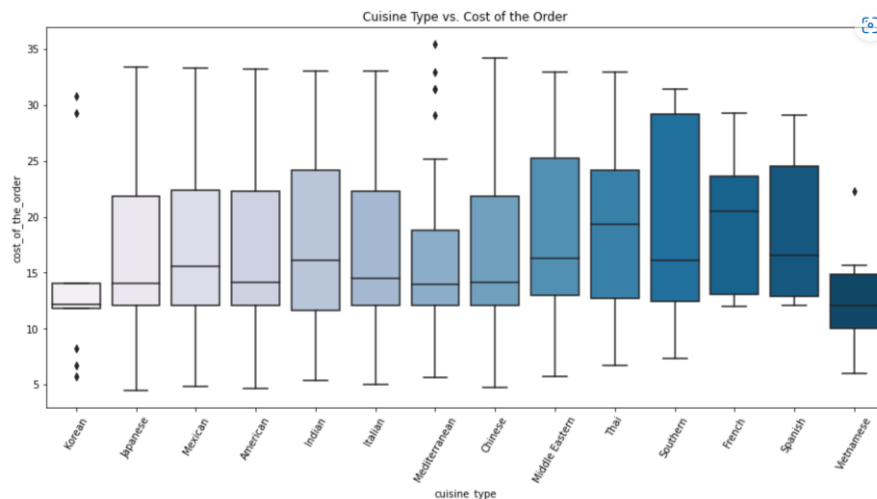
Multivariate Analysis

- Cuisine vs. Food preparation time:

  - In this graph, we see a variety of skewness, including no skew in Japanese, American, Indian, Mediterranean, and Middle Eastern Cuisine. This means those types of restaurants usually have the orders ready around the same time as expected.

  - Again, we see outliers in Korean cuisine meaning there were probably some issues that led to a longer food preparation time or specialty items that take longer in general.

  - We also see some right skewed plots such as Southern and French cuisines which take longer to prepare and then some left skewness in Italian and Spanish cuisine which could be because some of the classic dishes are prepared faster than others.

Cuisine vs. Food Preparation Time
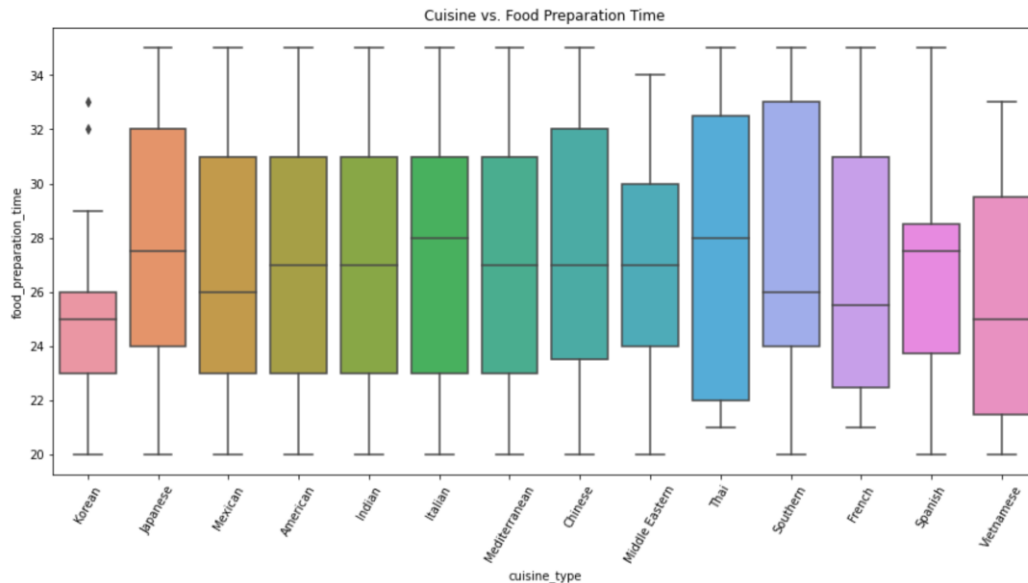
Multivariate Analysis

- Day of the week vs. Delivery time:

  - In the graph below, we can clearly see that orders placed on the weekdays take longer to be delivered than those on the weekends. This is probably because there is more traffic on the streets on weekdays as people are moving out and about. However, on the weekends, more people stay at home, so it is not as busy at restaurants and on delivery routes.

Day of the Week vs. Delivery Time

Multivariate Analysis

- Restaurant revenue:

  - There is a great variance between the restaurants generating the highest and lowest revenues. The most popular restaurant, as mentioned before, is Shake Shack, which generated $3579.53 of revenue. As we go down the list, we see that number growing smaller at a somewhat fast pace. The restaurant with the lowest revenue, Human Manor, generated only $5.72.

  - This could be for a few reasons. First, that Shake Shack is a very popular chain that many people are familiar with and that is why it receives most of the orders. Similarly, other small business may not have built up a reputation yet, so people are opting out of trying something new/taking a risk.

```
restaurant_name
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
Name: cost_of_the_order, dtype: float64
```

```
restaurant_name
Frank Restaurant                 12.08
Samurai Mama                     11.64
Lamarca Pasta                     9.22
Balade                            9.22
Gaia Italian Cafe                 8.78
Chola Eclectic Indian Cuisine     8.73
Woorijip                          8.25
'wichcraft                        8.10
La Follia                         8.05
Market Table                      6.79
Wa Jeal                           6.74
Cipriani Le Specialita            5.92
Big Wong Restaurant �_¤¾Ñ¾        5.92
Hunan Manor                       5.72
Name: cost_of_the_order, dtype: float64
```

Top 14 grossing restaurants   Bottom 14 grossing restaurants
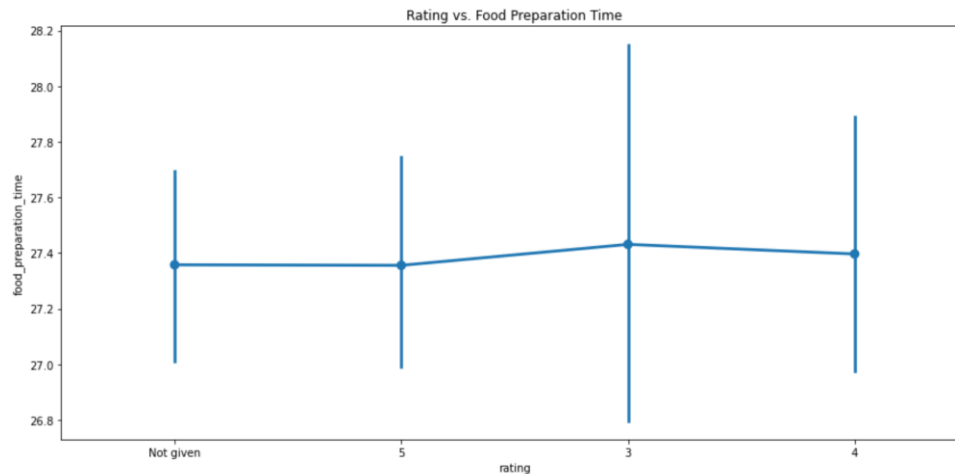
Multivariate Analysis

- Rating vs. Delivery time:

  - We have roughly the same delivery times of orders with 5-star ratings and those with no rating.

  - Orders given a 4-star rating have the shortest delivery times.

  - 3-star rated orders, the lowest available in the dataset, typically had the longest delivery time. This is possibly why customers rated the order lower and not 4 or even 5 stars. However, the error bars on 3-star ratings are particularly longer, meaning there could be a variety of delivery times/reasons for which the order was rated so low that perhaps were related more to the actual food than delivery.
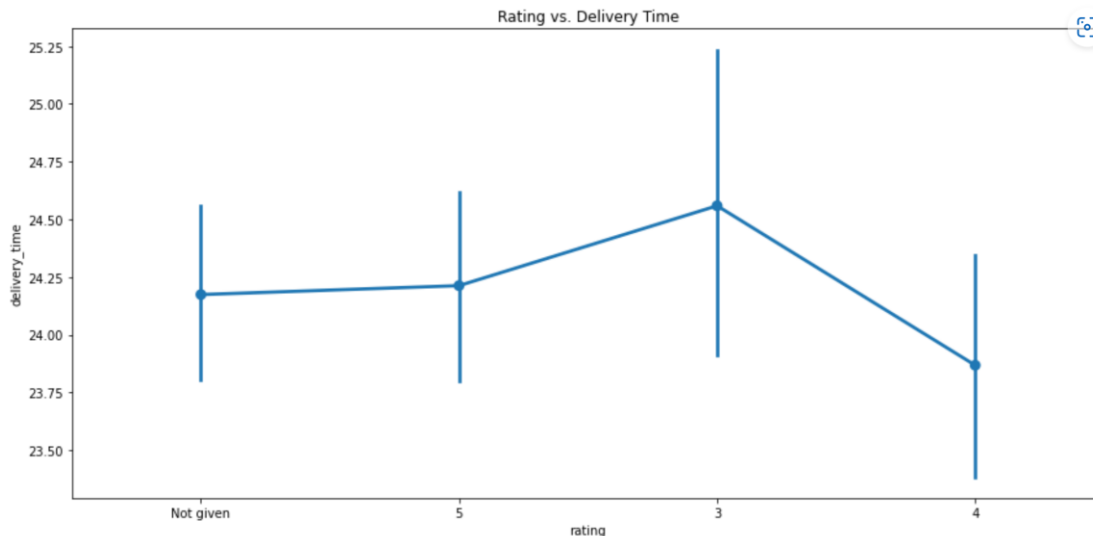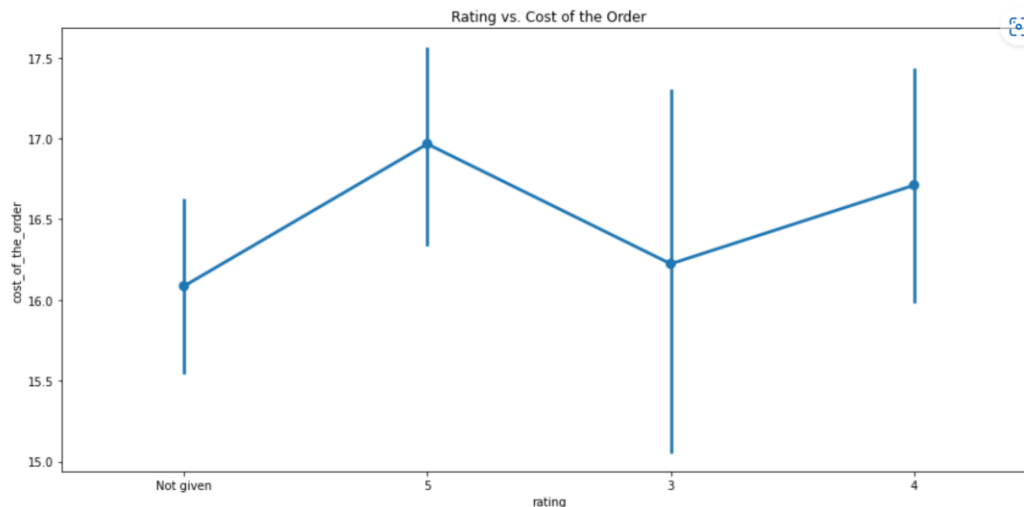
Multivariate Analysis

- Rating vs. Food preparation time:

  - From this graph, we can conclude that customer ratings were not usually affected by food preparation times.

  - However, the length of the error bars on 3-star orders again lead us to think that there was a large variance of food prep times that could've led to this rating. Whether it took too long for the food to be made or not long enough and as a result the quality was lower.


Rating vs. Food Preparation Time

Multivariate Analysis

- Rating vs. Cost of order:

    - We can see that orders rated 5-stars were on the more expensive side and as a result of the quality and other factors such as delivery time, customer satisfaction was greater to the extent that they took the initiative to give the restaurant a rating for a job well done.

    - 4-star rated orders follow shortly behind while 3-star rated orders have a large variety of costs. This furthermore emphasizes the insight that customers gave a 3-star rating when there was some sort of issue with their order/delivery.

Rating vs. Cost of the Order

Multivariate Analysis

- Correlations:

  - There are little to no correlations between order cost, food prep time, and delivery time.

  - This tells us that orders, no matter high or low cost, took around the same amount of time to prepared and delivered.
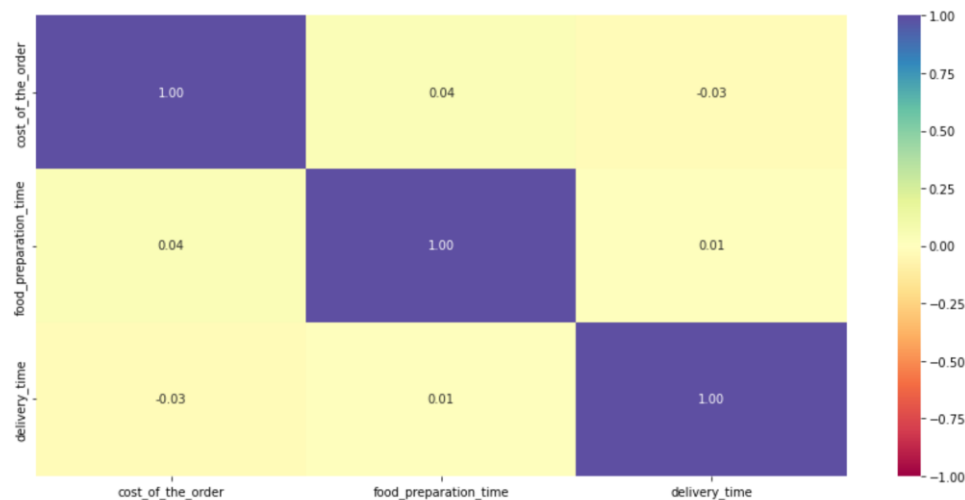
Multivariate Analysis

- Promotional Offer:

  - To find restaurants that qualify for the promotional offer in advertisement of their restaurant, we filtered the restaurants to see which had a rating count of more than 50 and an average rating greater than 4.

  - These were the resulting restaurants that qualify:

	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

Multivariate Analysis

- Net revenue:

    - The net revenue generated by FoodHub was around $6166.30

- Orders taking longer than 1 hour:

    - The number of total orders with delivery time greater than 60 minutes was 200

        - 10.54% of all orders took longer than 60 minutes

- Varying delivery times on weekends/weekdays:

    - The mean delivery time is around 28 minutes on weekdays and 22 minutes on weekends. Again, because there is more traffic around the city on weekdays and more people at home on weekends

APPENDIX

Happy Learning !