FoodHub's Restaurant Demand EDA

FoodHub project for MIT Machine Learning

04/03/2023

Contents / Agenda

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

The top four cuisine types are Japanese, American, Italian and Chinese. The top five restaurants with orders received were Shake Shack(American), Meatball shop(Italian), Blue Ribbon Sushi (Japanese), Blue Fried Chicken(American) and Parm(Italian). The busiest time, or the time with the highest deliveries, was on the weekends. The weekends also had a delivery time of 22 minutes, 2 mins below the mean delivery time of 24. The average cost of an order is $16.40, with a standard deviation of $7.40.

Recommendations would be to continue to satisfy the customer. To do that would be to improve food preparation and delivery time. Telling drivers which restaurants will be busy during the week, or right before the weekend should let them know where to park to receive orders during peak time. Giving incentives to drivers in those areas, like higher peak pay based on order volume, should satisfy available orders. Incentives could also be given to restaurants that keep their food preparation time below a certain threshold. The bottom line is to provide quick and efficient service and to process more orders, therefore increasing revenue.

During the weekdays, the mean delivery time was 28 minutes. We could offer drivers incentives, like $10 per 5 orders delivered during low order days to guarantee pick ups/ drop offs to lower this delivery time.

The problem is knowing the demand of the market, and putting drivers in the proper position. While food preparation time can be kept to a quantified number, batching orders to a delivery driver, having them accept it, and then pick it up burns valuable time. The solution to this problem is to communicate with our drivers the demand of the markets from last year, using metrics from last year and the current popularity of the market.

With 1898 rows and 9 columns of data, three different data types: floats, int64's and objects, the data was very clean, an not missing any values. The data team asked what the minimum, average and maximum amount of time it took for food to be prepared. Which was 20 minutes, 27.37 minutes, and 35 minutes. The data team also asked how many orders were not rated which came to be 736.

There were 1898 unique customer ID's, with 178 unique Restaurant names, and 14 cuisine types(See Figure 1).

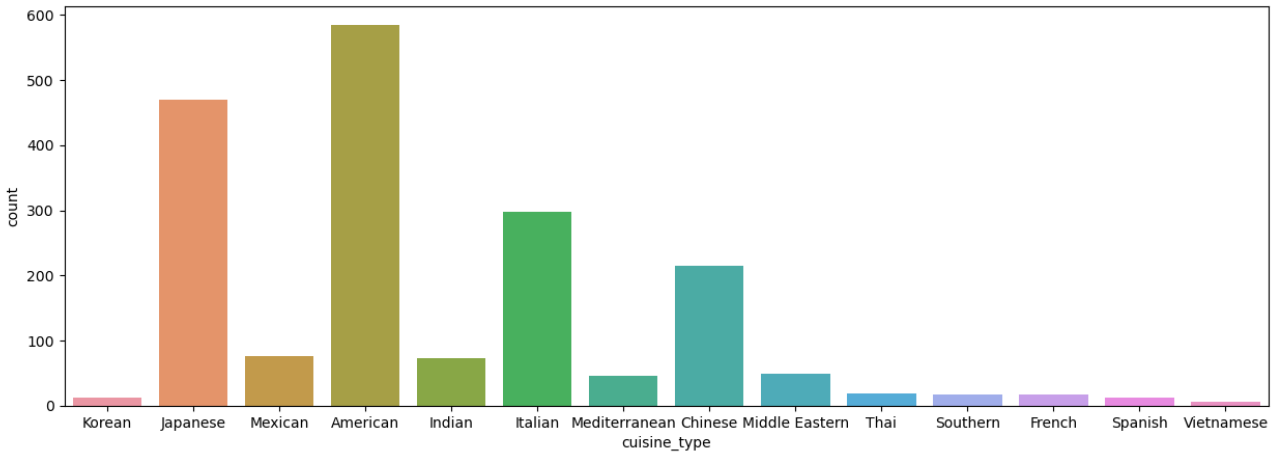

Figure 1

The cost of the order had an average of $16.4, with outliers ranging from $5 to $35. (See Figure 2 and 3)

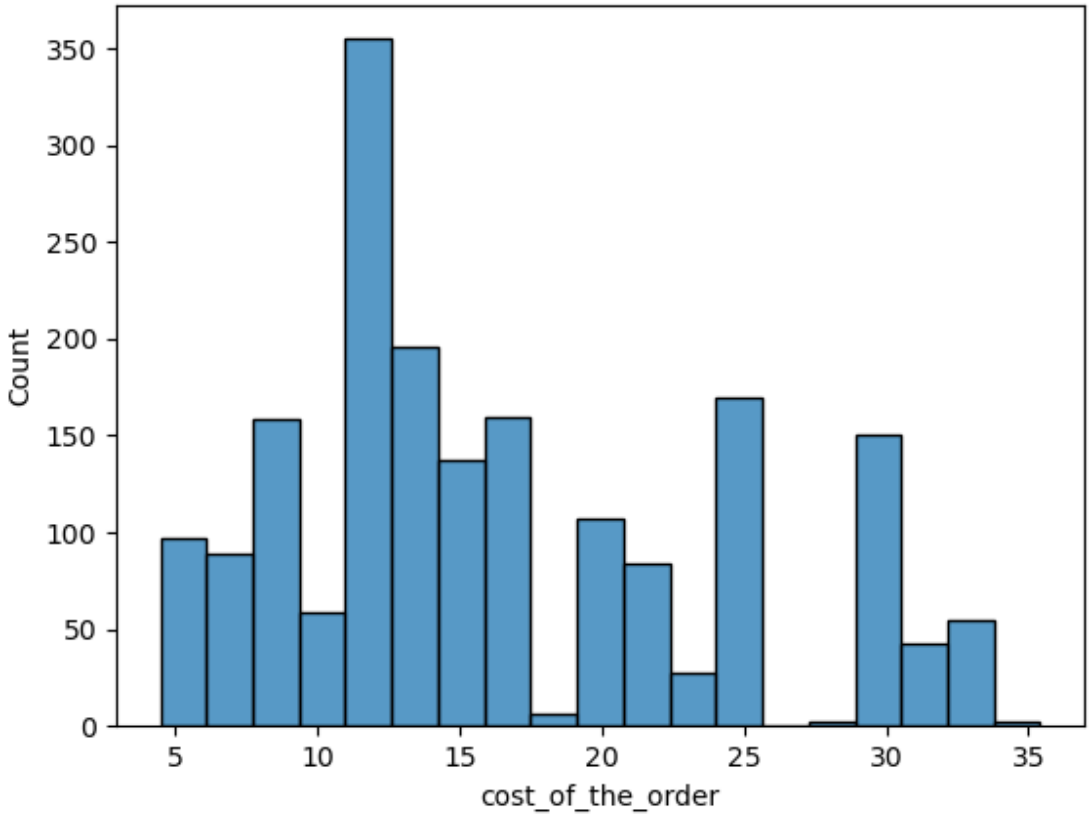

Figure 2

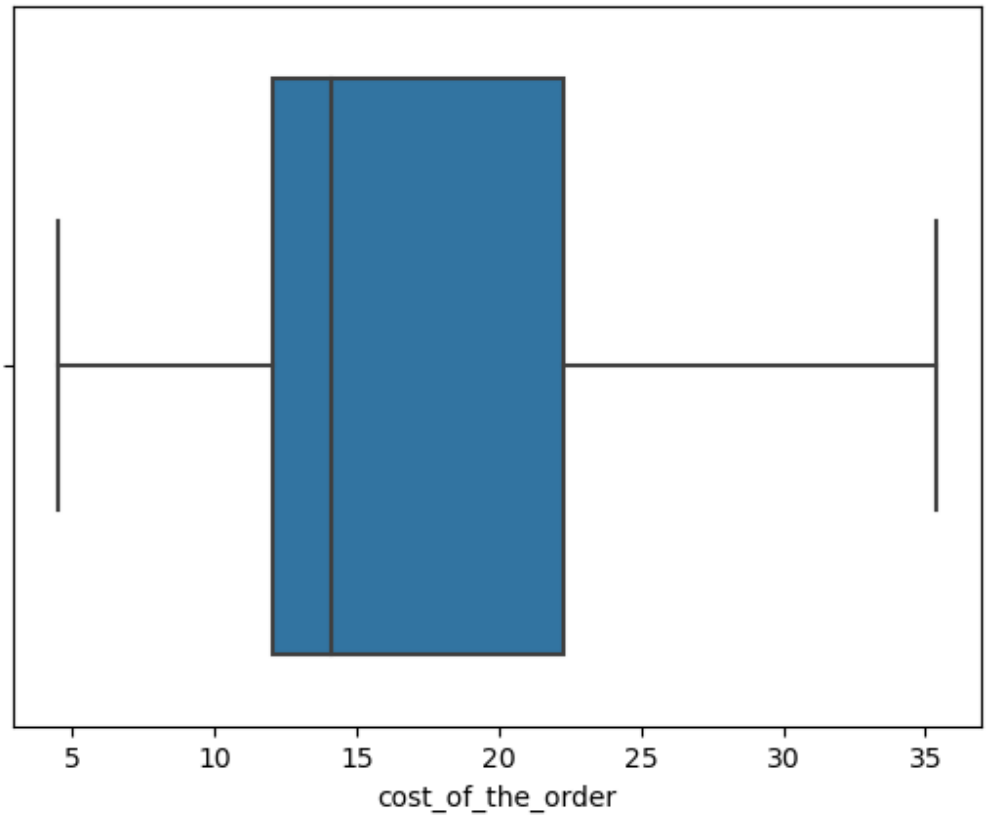

Figure 3

The days of the week that we did our univariate analysis on were bunched together and were two categorical variables, the weekday and the weekend. The weekends having the majority of orders being placed. (See Figure 4)

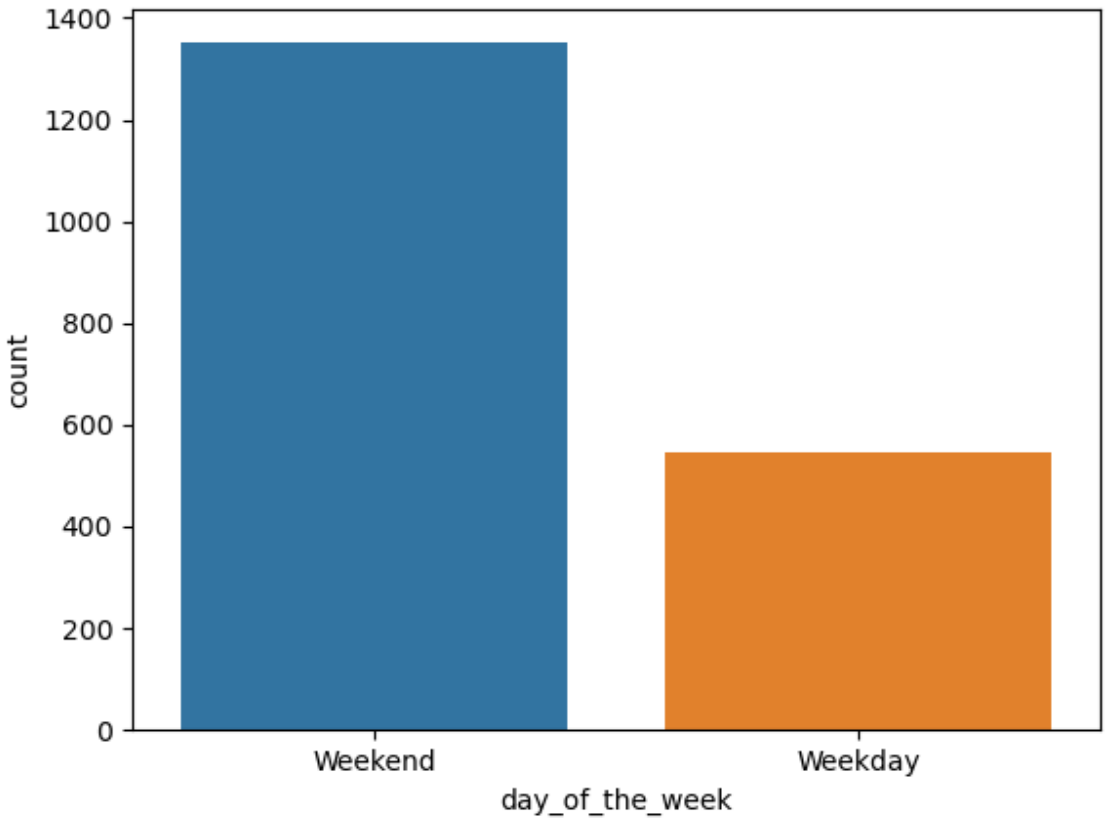

Figure 4

The ratings that were gathered from the univariate analysis were: 'Not given', 5, 3, and 4. Ignoring the 'Not given' ratings from the data we can see that there is a positive correlation, and that the customers are mostly satisfied. A solution to the 'Not given' ratings will be in the

final conclusion. (See Figure 5)

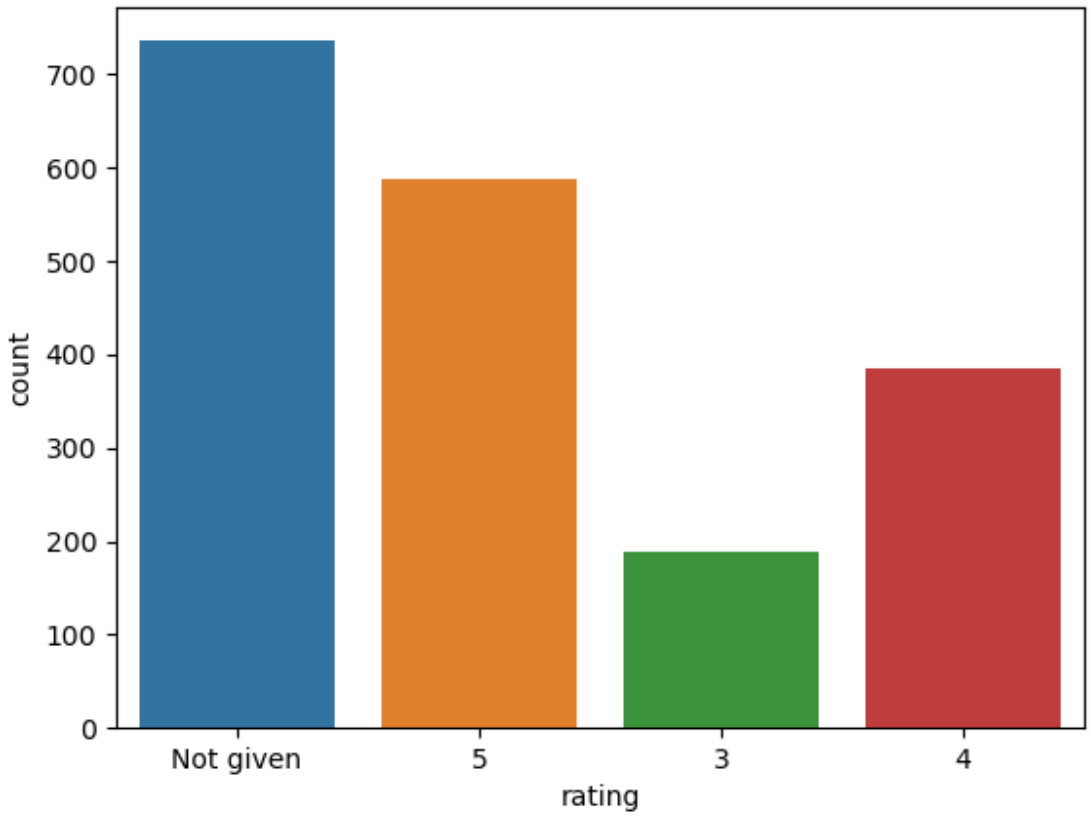

Figure 5

The food preparation time had a range from 20 minutes to 35 minutes with an average of 27.3 minutes. The data seems to be uniform. (See figure 6 and 7)

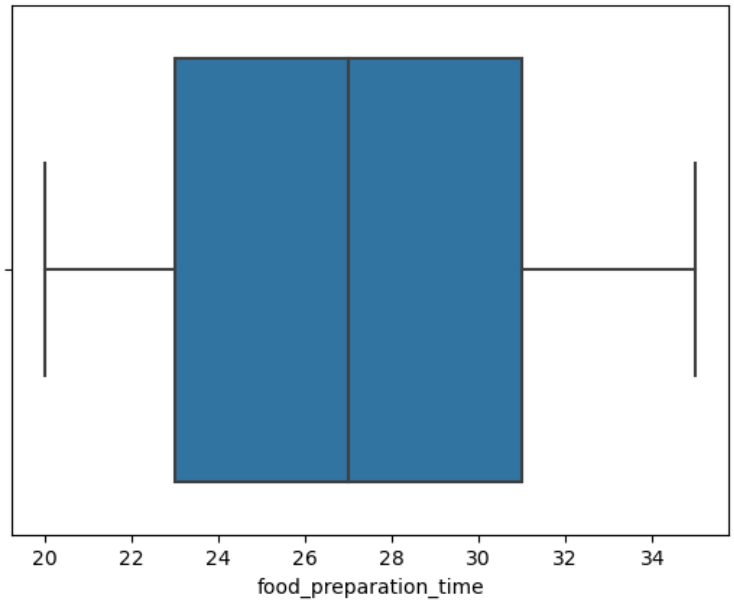

Figure 6

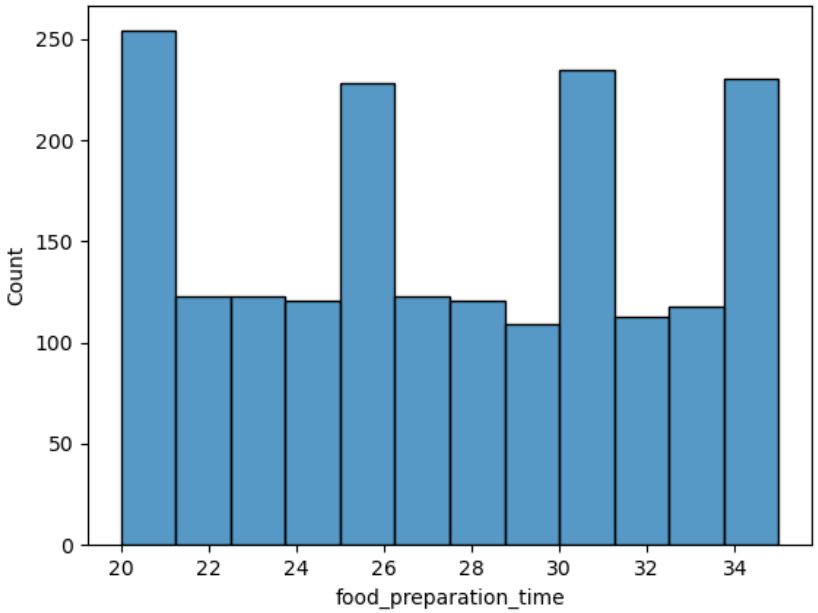

Figure 7

Delivery times ranged from 15 to 33 minutes with an average of 24.16 minutes. The data was uniform. (See Figure 8 and 9)

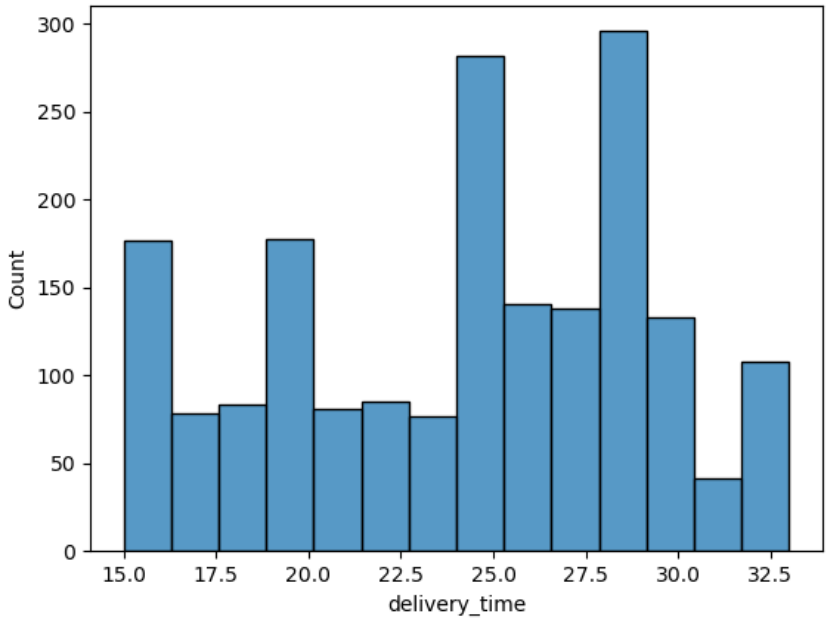

Figure 8

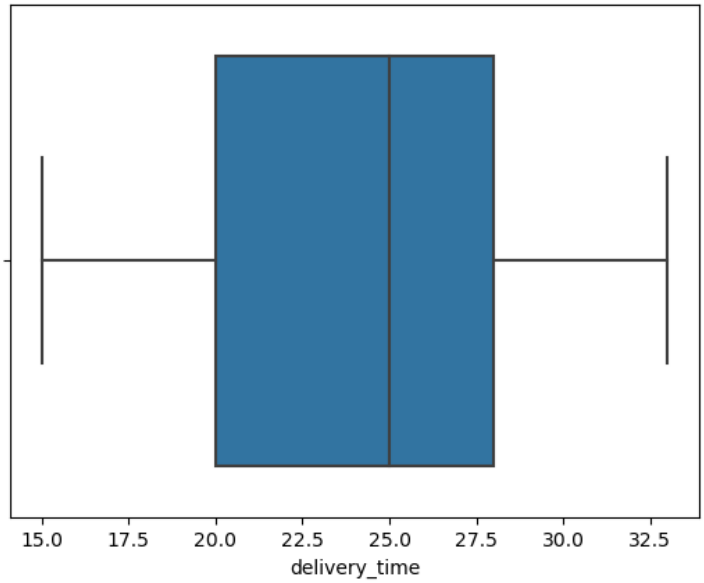

Figure 9

The data team asked what the top 5 restaurants were in terms of the number of orders received. These were Shake Shack(219), The Meatball Shop(132), Blue Ribbon Sushi(119), Blue Ribbon Fried Chicken(96), and Parm(68).

The data team asked what the most popular cuisine on the weekends were. These were: American(415), Japanese(335), Italian(207), Chinese(163), and Mexican(53). This finding helps illustrate why the top restaurants receive as many orders as they do; all of them being on this list as a cuisine type.

The data team asked what percentage of the orders cost more than $20. The number of total orders that cost above 20 dollars was 555, and the percentage was 29.24%. With the average being $16.4, and with a standard deviation of $7.4, it makes sense that ~30% of the data is within this metric.

The data team asked what the mean order delivery time was, which was 24.16 minutes.

FoodHub wanted to know what the top 3 most frequent customers were, in order to give them a 20% discount voucher. The ID's and the number of orders they placed were: 52832(13), 47440(10), and 83287(9).

The cuisine type and cost of the order were compared to get an understanding of whether customers were willing to spend more on a certain type of food.

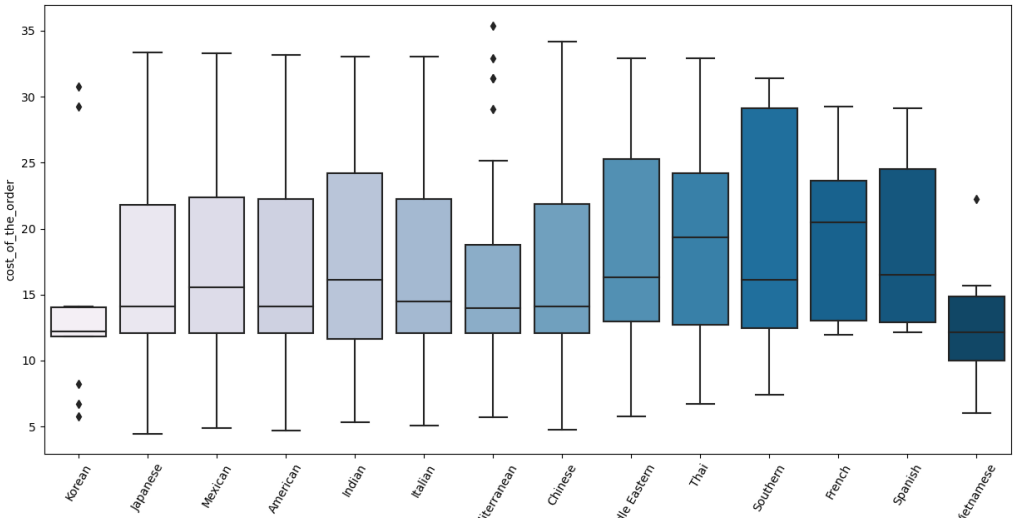

Figure 10

Korean cuisine had the most outliers, and the shortest average, with the Mediterranean in second. Otherwise, there didn't seem to be a strong correlation between the average cost of food and cuisine type.(See Figure 10)

Cuisine vs Food Preparation time was compared next to understand if a certain type of Cuisine took longer to cook than others. (See Figure 11)

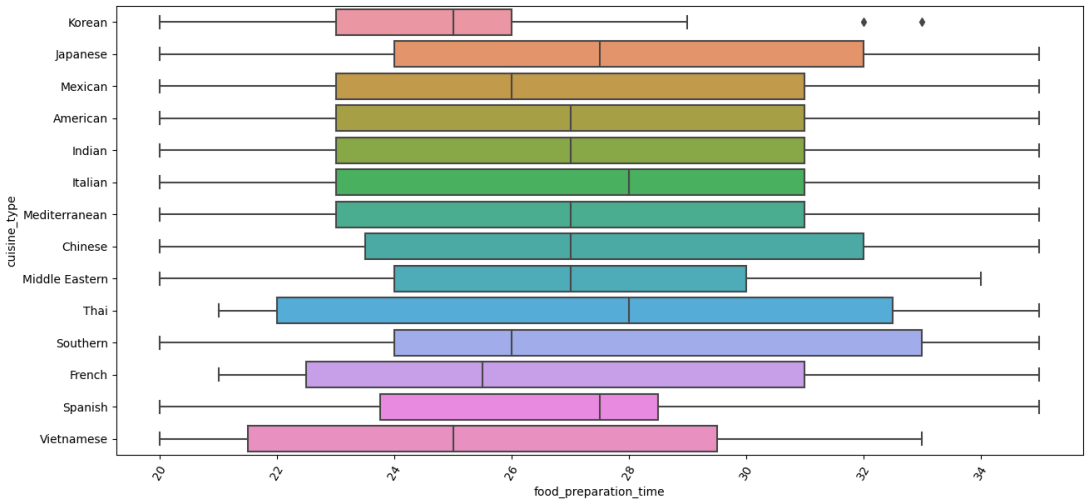

Figure 11

The top foods all have roughly the same mean preparation time, implying that the choice of Cuisine is not based on Food preparation time.

Day of the week vs Delivery time was plotted to show the difference between the delivery times during the week and during the weekend. From the boxplot we can see that the weekend has a much

higher set of data, and therefore a larger range. The weekday boxplot shows us a smaller range, but a much higher delivery time. This may be due to a smaller amount of drivers available to complete orders. (See Figure 12)

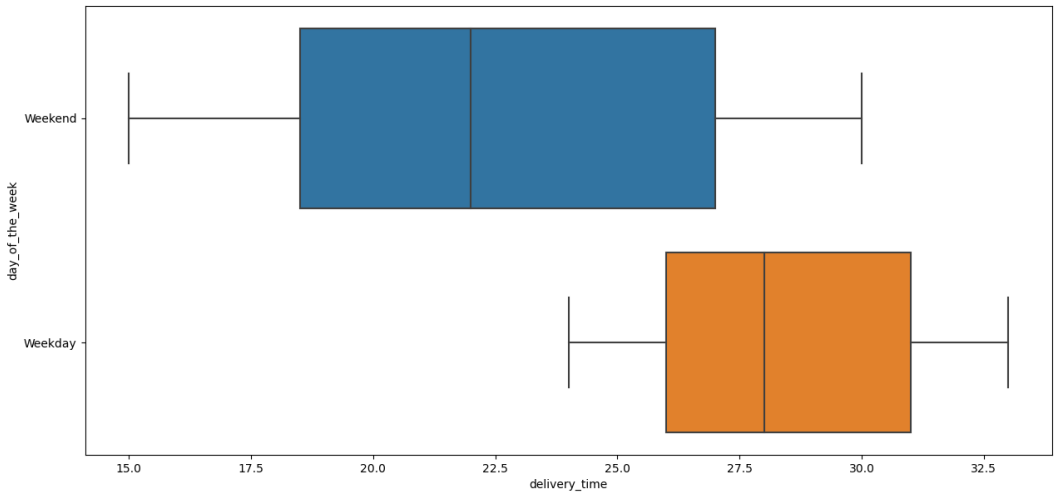

Figure 12

I ran some summaries based on the revenue generated by the restaurants. Those can be seen here:

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

Unsurprisingly, the top 5 restaurants with the highest revenue are the same as the ones with the highest numbers of orders received.

Rating vs Delivery time was analyzed through a pointplot to find a relationship between high or low ratings and how long the order took to reach the customer after it was prepared at the restaurant. From this there seems to be a small correlation between a bad rating(3) and delivery time. While it may not be significant, due to there being other problems possible with an order, for instance a driver forgetting a drink or the restaurant failing to provide quality food, it is noticeable that as time increases so does the probability of a 3 star rating. It's also worth mentioning the amount Not givens in this chart, as with more of those stars being placed properly the data could more properly read. (See Figure 13)

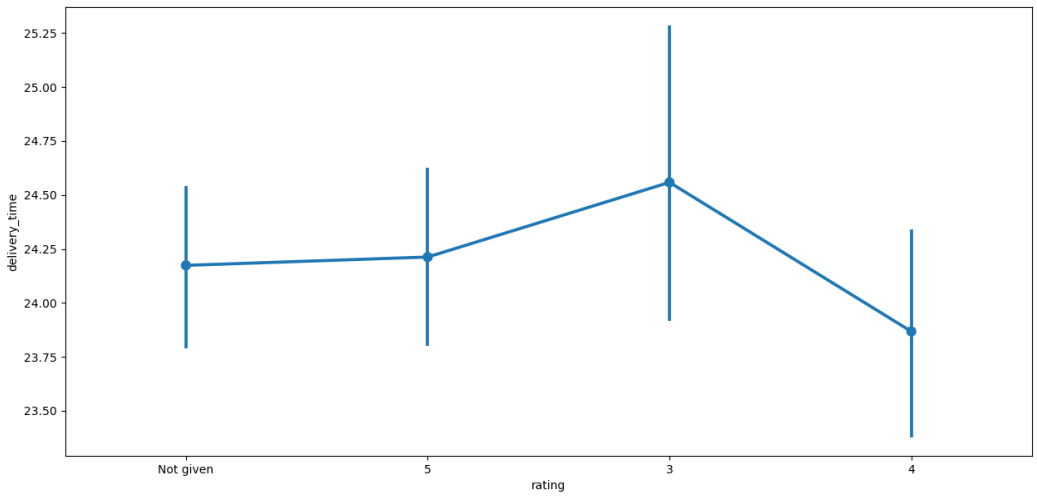

Figure 13

Rating vs Cost of the order was also analyzed through a point plot to measure the relationship between whether a customer would rate higher based on the amount of food they purchased. From the chart we can gather that lower prices of food received a lower rating. This may be due to the drivers not handling orders properly. A driver would treat a higher tipped order(which would be correlated with a higher cost of the order) with more care than a lower cost order; ie, check for drinks, arrive and pick up/drop off on time. (See Figure 14)

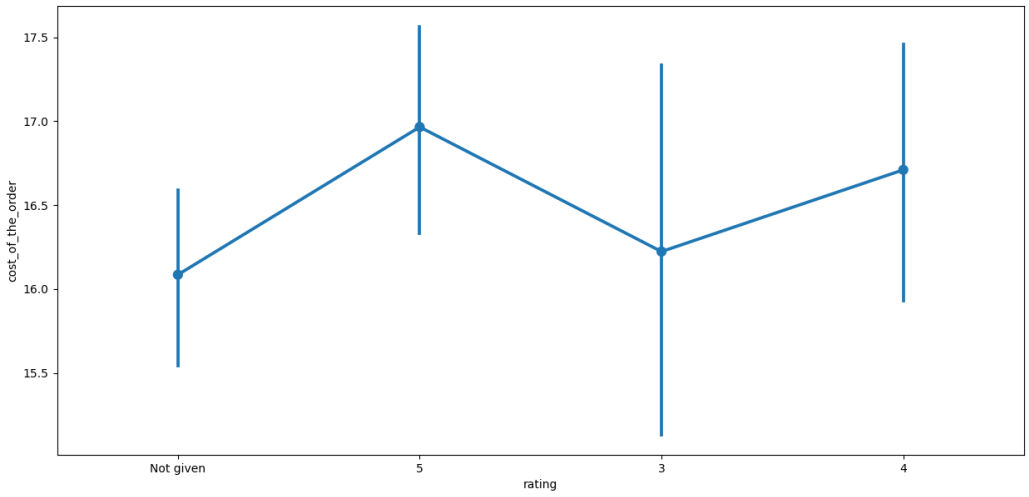

Figure 14

There was not a significant correlation among the variables: Cost of the order, food preparation time, and delivery time, with most of the variables having a correlation lower than .05. (See Figure 15)

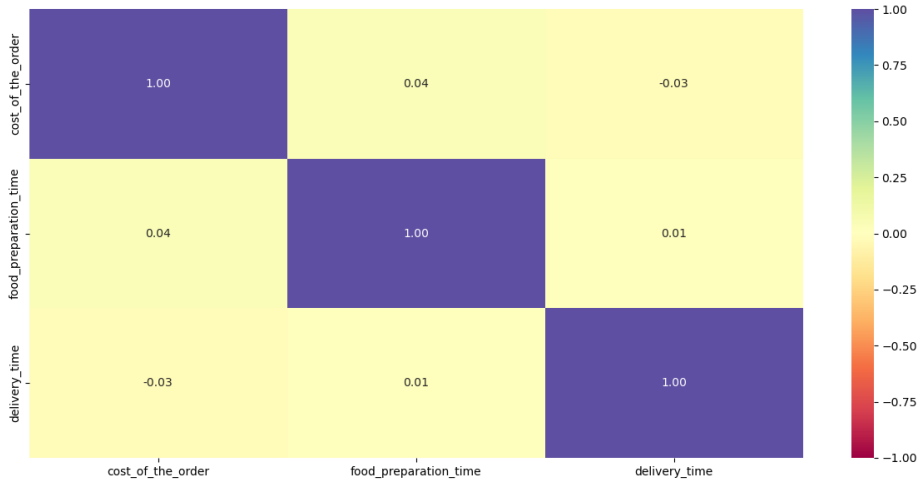

Figure 15

The net revenue was computed to be around $6166.3. Currently we are charging the restaurants 25% on orders having cost greater than $20 and 15% on orders having cost greater than $5. If we increase either percentage by 5% we would see an increase in revenue of ~$900. Something to consider when giving restaurants incentives: we could offer the restaurants that sell smaller orders a reduced % on orders that cost greater than $5 and increase the amount by 1-2% on orders above $20 using a probability function to determine if we would make profits based on the markets for that day.

FoodHub wanted to analyze the total time required to deliver the food based on adding the food preparation time and the delivery time. The total number of orders that take more than 60 minutes is 200, and the percentage of orders above 60 minutes is 10.54%.

FoodHub wanted to analyze the delivery time of the orders on weekdays and weekends. Those means were 28 minutes and 22 minutes respectively. The reason why these two numbers are different is due to the larger volume of orders on the weekends, and the availability of drivers.

To conclude this analysis it's recommended to continue to satisfy the customer. Have a list given out of these top restaurants so that delivery drivers can expect where the largest amount of orders to pick up will be. Focus on incentivizing the restaurant and the driver to decrease total delivery time, and to continue to offer top customers a discount to show they are valued.

While the data didn't know a significant correlation between delivery time and rating, there could still be incentive given to customers to provide a rating based on the service provided to them. This could be a discount on a meal or a free drink from a participating restaurant. It may be possible to provide a rating system similar to Netflix that guesses ratings based off of delivery time, and then to compare that to actual ratings that customers provide later to see if the prediction was correct. With this feedback we would better be able to provide quality EDA for our customers.