FoodHub

Problem Statement – FoodHub
Python Foundations

06/10/2022

Contents / Agenda

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

Executive Summary

I performed exploratory data analysis on the given data and find meaningful insights which are as follows:

Conclusions:

- The top 5 restaurants which are taking highest number of orders and generating revenues are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken, Parm.
- Majority of the customers don't give the feedback i.e., rating to the restaurant.
- Majority orders received by restaurants are on weekend.

Recommendations:

- The restaurants should focus on American cuisine to increase the revenue since majority orders are done on weekends and the most popular cuisine on weekends is American cuisine.
- The cost of the order should be in between 10-20 orders since the greatest number of orders are done in this cost range.

Business Problem Overview and Solution Approach

Business Problem

- 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.

Business Problem Overview and Solution Approach

Objective

- The food aggregator company has stored the data of the different orders made by the registered customers in their online portal. They want to analyze the data to get a fair idea about the demand of different restaurants which will help them in enhancing their customer experience. Suppose you are hired as a Data Scientist in this company and the Data Science team has shared some of the key questions that need to be answered. Perform the data analysis to find answers to these questions that will help the company to improve the business.

Solution Approach

- The approach used is the exploratory data analysis to find the hidden insights in python, first data is cleaned, then univariate analysis is done, after that multivariate analysis is performed to find the meaningful information.
- Charts and graphs are used to properly convey the meaningful insights to the desired audience.

Data Overview

Data Description

The data contains the different data related to a food order. 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: Cost of the order

Data Overview

Data Description(Continue)

- 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)
- rating: Rating given by the customer out of 5
- 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.
- 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.

Data Overview

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

- 1898 rows and 9 columns.

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1898 entries, 0 to 1897
Data columns (total 9 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   order_id               1898 non-null   int64  
 1   customer_id            1898 non-null   int64  
 2   restaurant_name        1898 non-null   object 
 3   cuisine_type           1898 non-null   object 
 4   cost_of_the_order      1898 non-null   float64
 5   day_of_the_week        1898 non-null   object 
 6   rating                 1898 non-null   object 
 7   food_preparation_time  1898 non-null   int64  
 8   delivery_time          1898 non-null   int64  
dtypes: float64(1), int64(4), object(4)
memory usage: 133.6+ KB
```

Question 3: Are there any missing values in the data? If yes, treat them using an appropriate method?

- No Missing values were founded.

Data Overview

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?

- Minimum Time: 20 minutes, Average Time: 27.37 minutes, Maximum Time: 35 minutes.
- Statistical Summary:

	order_id	customer_id	cost_of_the_order	food_preparation_time	delivery_time
count	1.898000e+03	1898.000000	1898.000000	1898.000000	1898.000000
mean	1.477496e+06	171168.478398	16.498851	27.371970	24.161749
std	5.480497e+02	113698.139743	7.483812	4.632481	4.972637
min	1.476547e+06	1311.000000	4.470000	20.000000	15.000000
25%	1.477021e+06	77787.750000	12.080000	23.000000	20.000000
50%	1.477496e+06	128600.000000	14.140000	27.000000	25.000000
75%	1.477970e+06	270525.000000	22.297500	31.000000	28.000000
max	1.478444e+06	405334.000000	35.410000	35.000000	33.000000

Question 5: How many orders are not rated?

- 736 orders are not rated.

Univariate Analysis

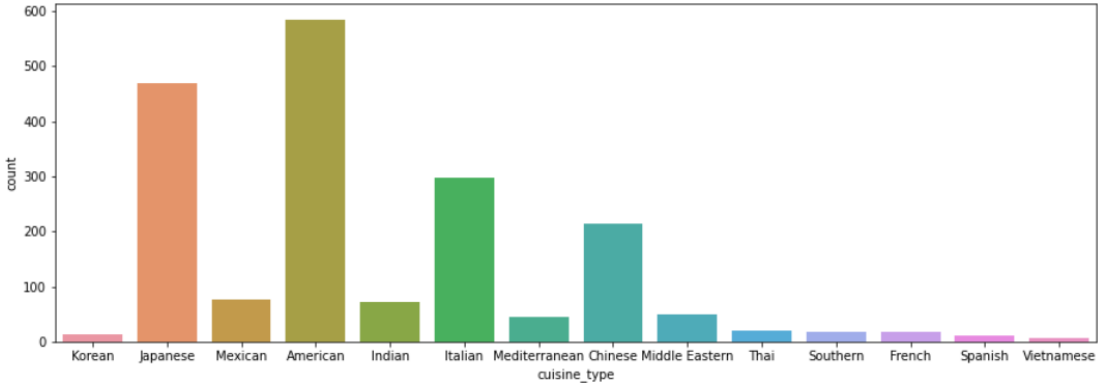
Univariate analysis is the simplest form of analyzing data. “Uni” means “one”, so in other words your data has only one variable. It doesn’t deal with causes or relationships (unlike regression) and its major purpose is to describe; It takes data, summarizes that data and finds patterns in the data.

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

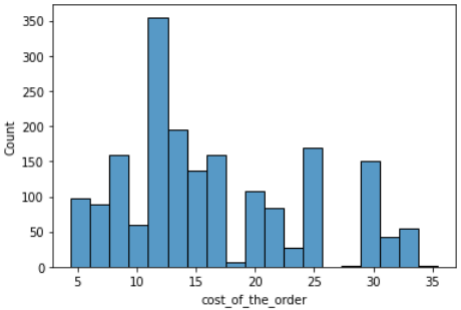
- Order ID: It has 1898 unique values
- Customer ID: It has 1200 unique values
- Restaurant name: It has 178 unique values

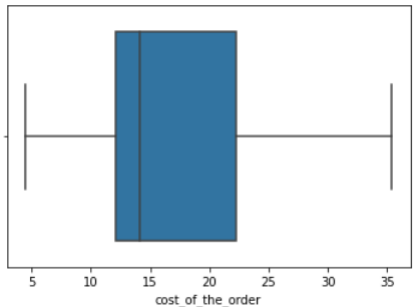
Univariate Analysis

- Cuisine type: It has 14 unique values. Distribution of values can be seen in graph.



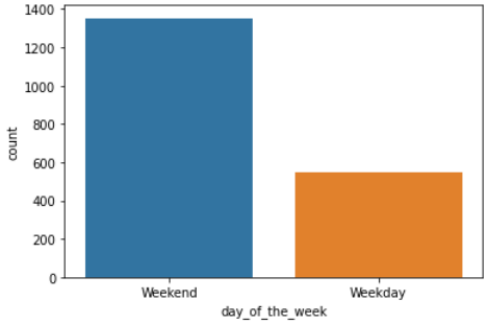
- Cost of the order: The distribution of the variable and statistical summary can be seen in graphs.



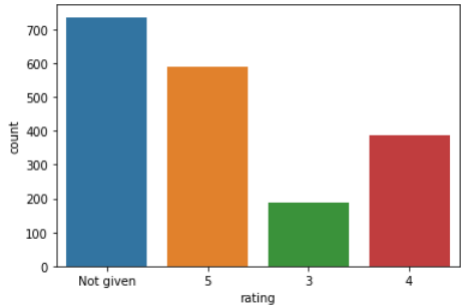


Univariate Analysis

- Day of the week: It has 2 unique values (Weekday and Weekend). Distribution of values can be seen in graph.

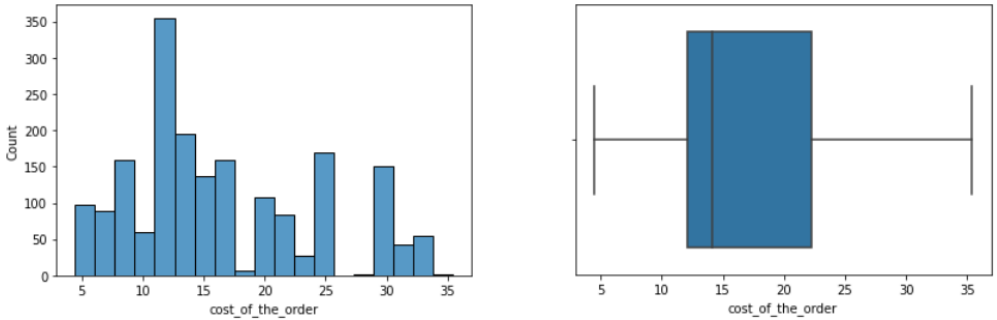


- Rating: It has 4 unique values(Not Given,5,4,3). Distribution of values can be seen in graph.

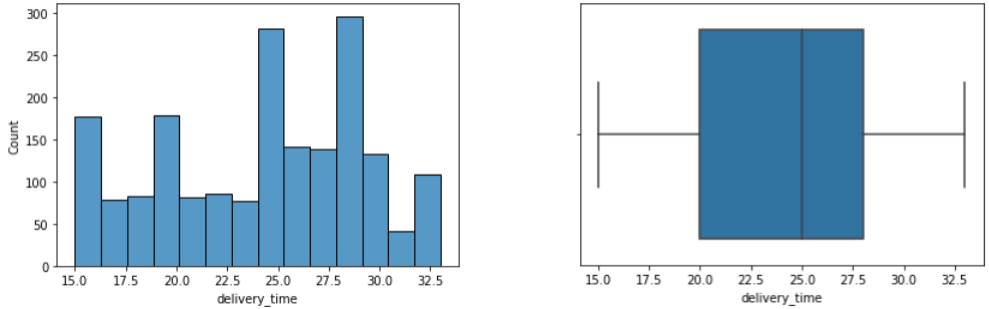


Univariate Analysis

- Food preparation time: The distribution of the variable and statistical summary can be seen in graphs.



- Delivery time: The distribution of the variable and statistical summary can be seen in graphs.



Univariate Analysis

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

- The top 5 restaurants which are taking highest number of orders 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?

- American cuisine is the most popular cuisine on weekends.

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

- The number of total orders that cost above 20 dollars is: 555.
- Percentage of orders above 20 dollars: 29.24 %.

Question 10: What is the mean order delivery time?

- 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.

- The top 3 customers are shown in the image with their customer IDs and number of orders:

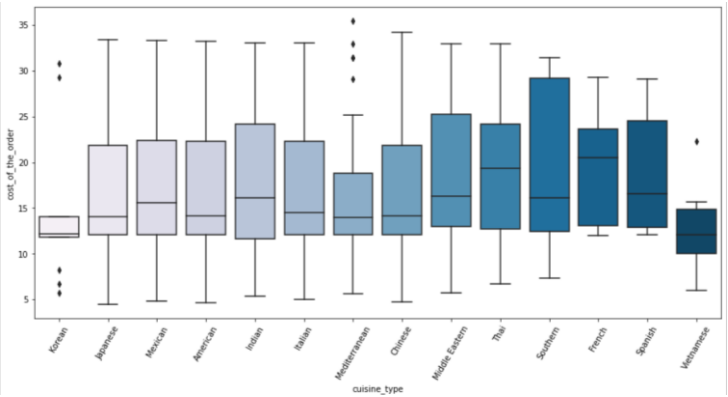
```
52832    13
47440    10
83287     9
```

Multivariate Analysis

Multivariate Analysis is defined as a process of involving multiple dependent variables resulting in one outcome. This explains that most of the problems in the real world are Multivariate.

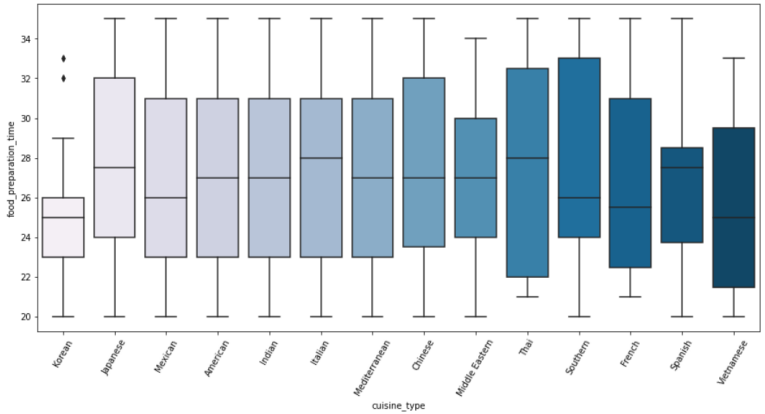
Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset. (It is a good idea to explore relations between numerical variables as well as relations between numerical and categorical variables)

- Cuisine V/s Cost of the order

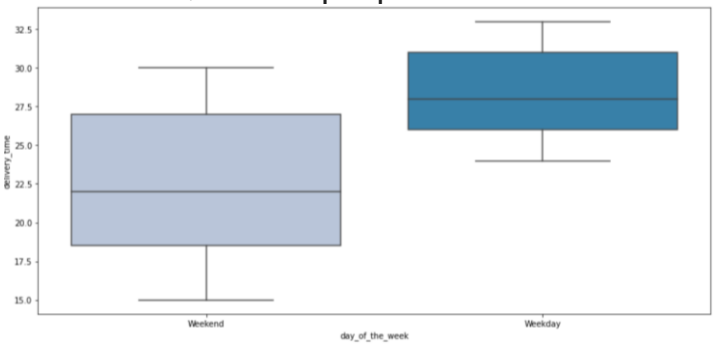


Multivariate Analysis

- Cuisine V/s Food preparation time



- Cuisine V/s Food preparation time

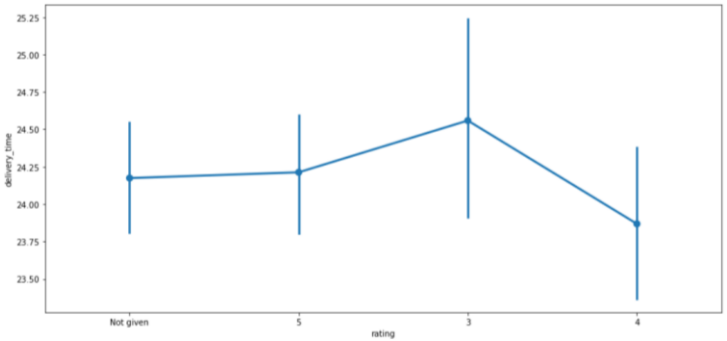


Multivariate Analysis

- Revenue generated by each restaurant

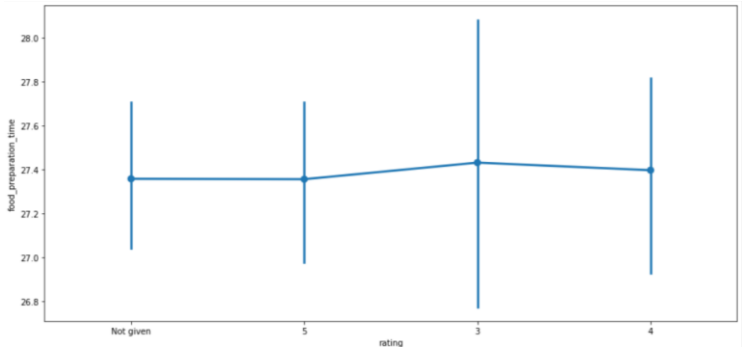
```
restaurant_name
Shake Shack                      3579.53
The Meatball Shop                2145.21
Blue Ribbon Sushi                1903.95
Blue Ribbon Fried Chicken        1662.29
Parm                             1112.76
RedFarm Broadway                  965.13
RedFarm Hudson                    921.21
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
```

- Rating V/s Delivery time

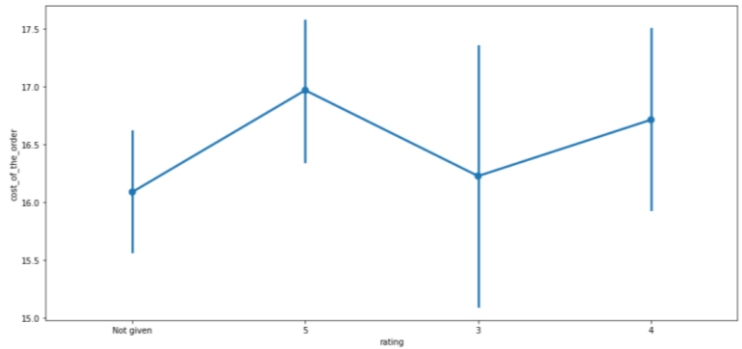


Multivariate Analysis

- Rating vs Food preparation time

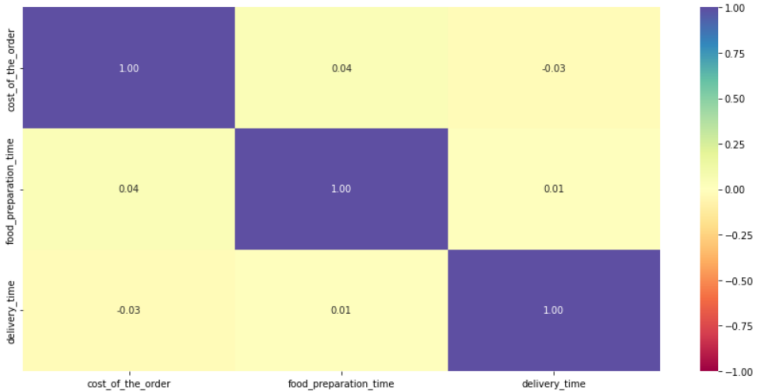


- Rating vs Cost of the order



Multivariate Analysis

- Correlation among variables



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.

- Following restaurant fulfills the given conditions:

	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

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.

- The net revenue 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 must be prepared and then delivered.)

- The number of total orders that take more than 60 minutes to get delivered from the time the order is placed: 200.
- Percentage of orders that take more than 60 minutes to get delivered from the time the order is placed: 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?

- The mean delivery time on weekdays is around 28 minutes.
- The mean delivery time on weekends is around 22 minutes.

Great Learning

Happy Learning !