Analysis on Customer Demand
Foodhub, DSML Course

28/05/2023

By Chun Wing Ip

Contents / Agenda

- Executive Summary

- Business Problem Overview and Solution Approach

- Data Overview

- EDA - Univariate Analysis

- EDA - Multivariate Analysis

- Appendix

Executive Summary

- Conclusions

    This research found that our company has lived up to the expectation of customers. More than half of the customers rated 4 and 5 for the orders.[1] Also, 35% of the customers made a second purchase or more.[2] The following summarizes customer demand in different aspects.

    American, Japanese, Italian, Chinese and Mexican food received the most orders.[3] The top 5 restaurants with the highest number of orders received are Shake Shack, The Meatball Shop, Blue Ribbon Sushi, Blue Ribbon Fried Chicken and Parm.[4]

    Popular cuisine types and popular restaurants received higher ratings. 4 out of the top 5 popular cuisine types received the highest number of 5-rating.[5] Also, the top 10 popular restaurants has the highest number of 5-ratings.[6]

1.   Refer to ppt slide 18.
2.   Refer to figure 1 in Appendix.
3.   Refer to ppt slide 15.
4.   Refer to ppt slide 14.
5.   Refer to figure 2 and figure 3 in Appendix.
6.   Refer to figure 4 and figure 5 in Appendix.

Executive Summary

- Conclusions

Orders in the weekend are more than a double of orders in the weekday.[1] Ironically, the delivery in the weekend is much faster than the delivery in the weekday.[2]

The required time to prepare the food and deliver the orders are satisfactory. Only 10.5% of orders took more than 60 minutes.[3] Also, the mean order delivery time is about 24 minutes with a standard deviation of 5 minutes.[4] The variation is small. In addition, the food preparation time has a uniform distribution across all orders.[5]

Italian, Thai and Spanish food required food preparation time more than the average required time of all other types.[6] Nonetheless, some types of cuisine might require a longer preparation time. Also, there is no evidence in the data to show that customer rating is correlated to any time variables.[7]

1. Refer to ppt slide 17.
2. Refer to ppt slide 27.
3. Refer to ppt slide 36.
4. Refer to figure 6 in Appendix.
5. Refer to ppt slide 20.
6. Refer to ppt slide 27.
7. Refer to ppt slide 30 to 31.

Executive Summary

- Conclusions (Cont'd)

  There is a segment of customers looking for affordable meals. 2 of the top 5 cheapest cuisine types are also the top 5 popular cuisine types.[1]

  Our most profitable segment of customers is those ordering more expensive meals. 29.2% of orders cost greater than 20 dollars.[2] It accounted for about 60% of the net revenue[3]. However, out of 178 restaurants, only 24 restaurants belong to the Top 5 expensive cuisine types from which this segment might order their meal.[4]

  However, the research has a limitation. Analysis on rating and other numerical variables do not have a strong relationship because the option 'Not given' has a similar distribution to those of other ratings.[5] An exception to that is that the higher the cost of order, the higher the rating.[6]

1. Refer to figure 2 and figure 7 in Appendix.
2. Refer to ppt slide 24.
3. Refer to figure 8 in Appendix.
4. Refer to figure 7 and figure 9 in Appendix.
5. Refer to figure 10, figure 11 and figure 12 in Appendix.
6. Refer to ppt slide 32.

Executive Summary

- Actionable insights & recommendations

  Overall, the company can serve customers better by improving delivery time in the weekday. Although the research shows little relationship between rating and delivery time, shortening the delivery time in weekday can match the delivery time in weekend.

  Currently, the most profitable segment of the company is underserved with a limited selection of restaurants. The company should improve restaurant mix by adding more quality restaurants with a higher cost of orders. This is further supported by the fact that higher cost of the orders yields better rating because the cost of orders might suggest restaurant's quality.

  To keep quality restaurants in the ecosystem, the company can give promotional advertising offer to the restaurants with more good ratings.

  Another way to increase revenue is to boost orders in the weekday. It is less than half of the orders received in the weekend.

Executive Summary

- Actionable insights & recommendations (Cont'd)

  The company can improve upon the 35% retention rate. It is generally more costly to acquire new customers so having a base of loyalty customers is important in terms of profitability. One way is to reward frequent customers.

  To probe the relationship of rating and other variables, the company might want to incentivize people to give ratings and feedback for the service. These data will be more reflective of service quality in the future.

Business Problem Overview and Solution Approach

- Problem Statement

  Our company, Foodhub, has a storage of transaction information on an online portal. In order to understand customer experience through customer demand in the process of food ordering on our application, the management team requested our department to conduct an analysis on the order data.

- Methodology

  The Data Science department performed an exploratory data analysis (EDA) on the acquired data. Firstly, the structure and information of the data were reviewed. Data were then summarized in descriptive statistics. Subsequently, univariate analysis was done to show distribution of each variable in the data. Following it was Multivariate analysis which was done to reveal the relationships between two or more variables. Last but not least, conclusion and key insights were drawn from the EDA.

Data Overview

- The data has 9 columns. Data in each row corresponds to one order placed by a customer.

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24

Data Overview

- Q1: There are 1898 rows and 9 columns in the data.

- Q2: There are 9 columns of data. 1 of them is float, 4 of them are integer and another 4 are object. With that said, 5 are numerical variables and 4 are categorical variables. There are 1898 rows of data, indexing from 0 to 1897. There are no missing values in the data. However, it is important to note that rating is a categorical variable here. This is because no entry is labeled 'Not given' in that column.

```
<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
```

Data Overview

- Q3: There are no missing values in the data.

```
order_id                0
customer_id             0
restaurant_name         0
cuisine_type            0
cost_of_the_order       0
day_of_the_week         0
rating                  0
food_preparation_time   0
delivery_time           0
dtype: int64
```

Data Overview

- Q4: The minimum time it takes for food to be prepared once an order is placed is 20. The average time it takes for food to be prepared once an order is placed is 27.37. The maximum time it takes for food to be prepared once an order is placed is 35.

|	count	mean	std	min	25%	50%	75%	max
order_id	1898.0	1.477496e+06	548.049724	1476547.00	1477021.25	1477495.50	1.477970e+06	1478444.00
customer_id	1898.0	1.711685e+05	113698.139743	1311.00	77787.75	128600.00	2.705250e+05	405334.00
cost_of_the_order	1898.0	1.649885e+01	7.483812	4.47	12.08	14.14	2.229750e+01	35.41
food_preparation_time	1898.0	2.737197e+01	4.632481	20.00	23.00	27.00	3.100000e+01	35.00
delivery_time	1898.0	2.416175e+01	4.972637	15.00	20.00	25.00	2.800000e+01	33.00

Univariate Analysis

- Q6 (cont'd)

  "cuisine_type"

```
In [14]: df["cuisine_type"].value_counts()

Out[14]: American          584
         Japanese          470
         Italian           298
         Chinese           215
         Mexican            77
         Indian             73
         Middle Eastern     49
         Mediterranean      46
         Thai               19
         French             18
         Southern           17
         Korean             13
         Spanish            12
         Vietnamese          7
         Name: cuisine_type, dtype: int64
```
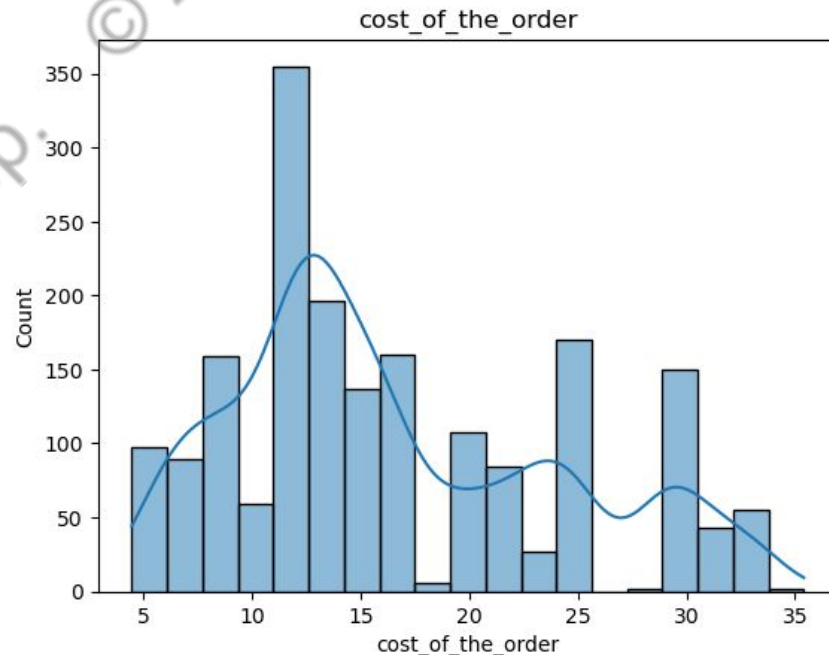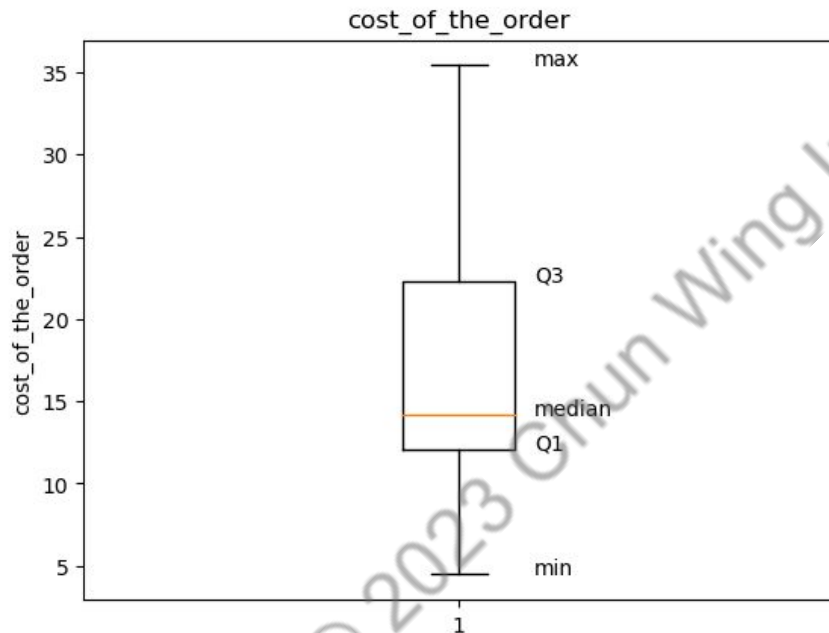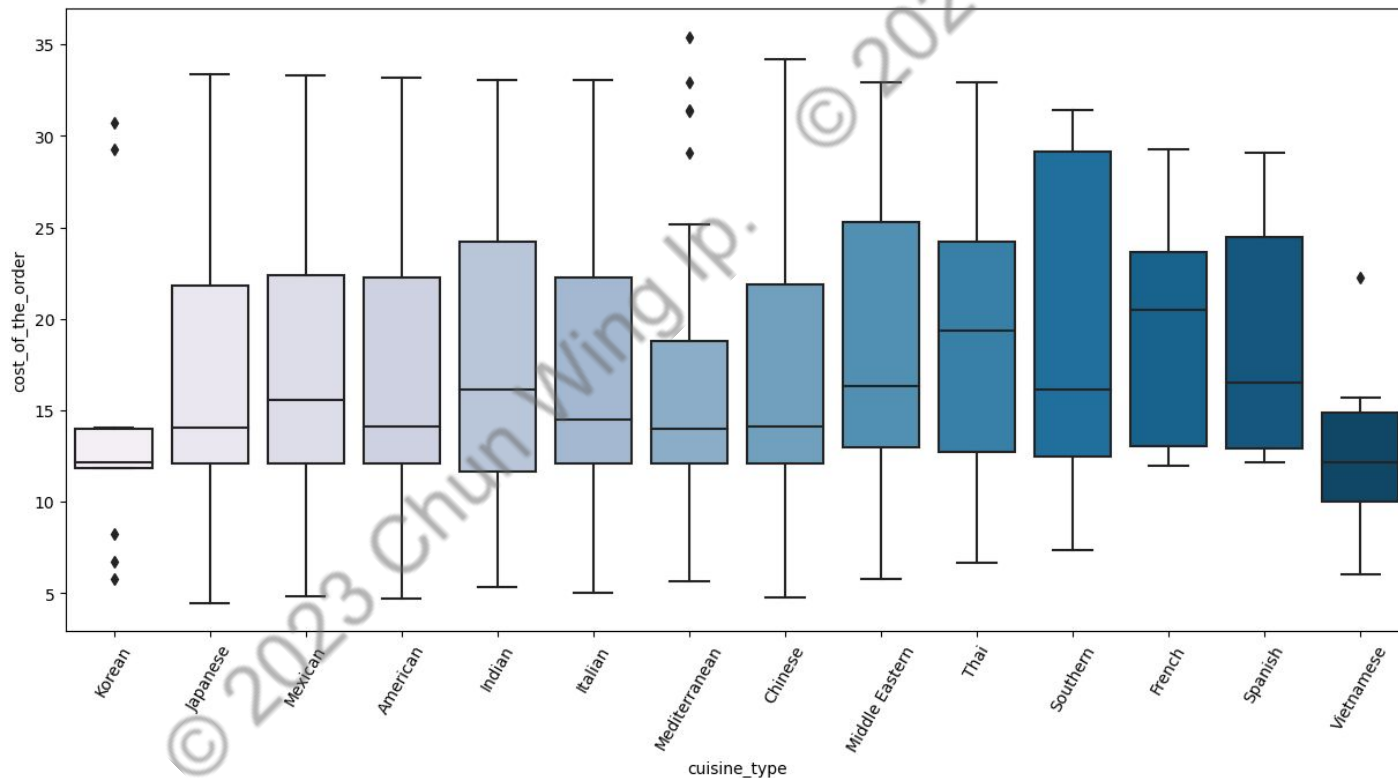

Bar Chart of Cuisine_Type
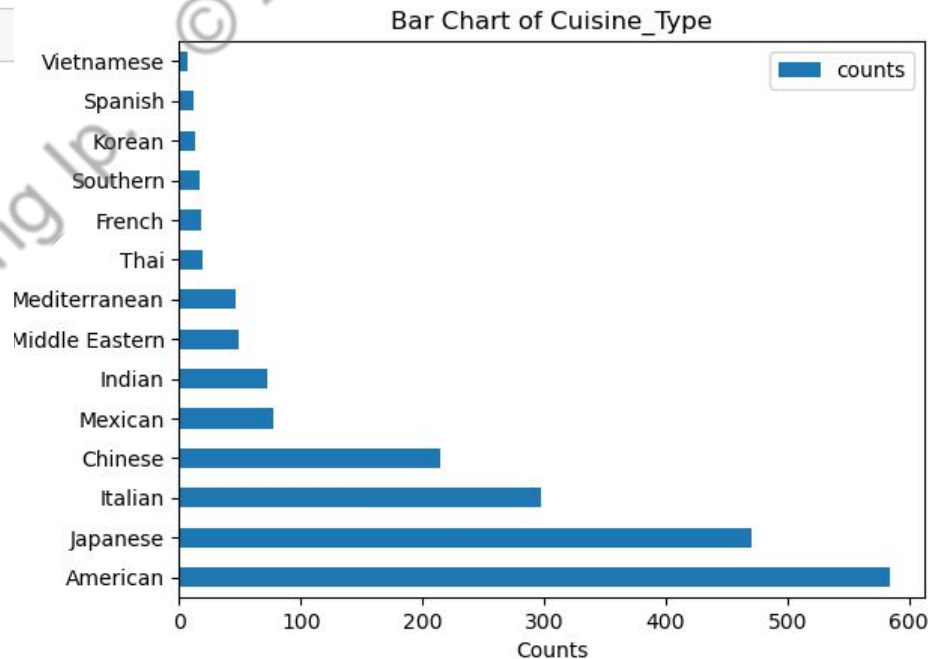
Univariate Analysis

- Q6 (cont'd)

  "cost_of_the_order"

Multivariate Analysis

- Q12: The relationship between cuisine_type and cost_of_the_order

Multivariate Analysis

- Q12: Correlations among variables