Food Hub Project (low code)

Python Foundations

Date: 04/18/2024

Project Submission by Anmol Verma

Contents / Agenda

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

Executive Summary

Conclusions

- There are 1898 unique orders placed from 1200 unique customers, 178 unique restaurants across 14 cuisine types
- American, Japanese and Italian are the most popular cuisines. French food and Thai food have the highest cost per order, while Korean and Vietnamese have the lowest average cost per order. Improving the cuisine mix can help the company drive better revenue per order
- Majority of the orders are placed over the weekend (71%). Average delivery time during weekdays is higher than weekends.
- Average rating is 4.3. Higher cost orders and lower delivery time emerge as two possible drivers for higher rating.
- 10.5% of the orders took over 60 minutes to deliver from the time the order is placed
- Top 5 restaurants contributed to one third of orders and revenue generated; Shake Shack, The Meatball shop and Blue Ribbon have the highest revenue generated among the restaurants
- 4 restaurants have an average rating count of >50 and rating of >4 and are eligible for promotional offers

Recommendations

- Driving better cuisine mix in favor of higher cost per order cuisines like French and Thai food can provide a revenue uplift and also potentially improve rating
- Average delivery time during weekdays is more than weekends; we could explore alternative transport options for our delivery fleet to improve customer experience during weekdays
- 10.5% of the orders took over 60 minutes to deliver – we should better understand what drove the high total time here and work with our partner restaurants and delivery personnel to improve delivery time
- We should prioritize our efforts toward top 5 restaurants to achieve business objectives faster.

Business Problem Overview and Solution Approach

About the business: FoodHub is a food aggregator company that offers its customers access to multiple restaurants through a single smartphone app.

Business Problem: Foodhub collects data related to the different orders made by registered customers on their online portal. It wants to analyze the data to better understand the demand from different restaurants that will help them enhance customer experience

Solution Approach/Methodology: Our objective is to extract actionable insights from the data. To accomplish this task, we will:

- Perform data analyses, both univariate and multivariate analyses to dive deeper into the data and answer the questions shared by the data team
- Focus on cuisine type and feedback ratings to drive our recommendations

Data Overview

The data is collected by a food aggregator company called FoodHub and contains details for food orders made by registered customers on their online platform.

Variable	Description
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 (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.

```
 #   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
```

Key observations

- There are **1898 rows** and **9 columns** in this dataset
- The datatypes of the different columns in the dataset include **integer, floats and string**
- Inspecting the first five rows, we discover that the rating column has some 'not given' values. Thus, the **data has some missing values**, which we have replaced with NaN
- The minimum, average and maximum time it takes for the food to prepare once an order is placed are **20 minutes, 27.4 and 35 minutes** respectively
- 736 orders in the dataset are not rated

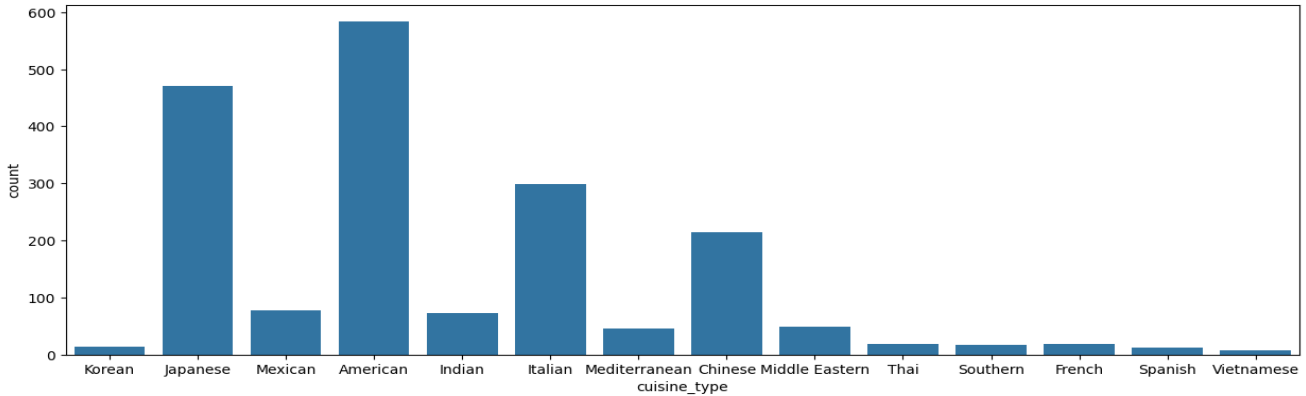
	order_id	customer_id	cost_of_the_order	food_preparation_time	delivery_time	Revenue	total_time
count	1.898000e+03	1898.000000	1898.000000	1898.000000	1898.000000	1898.000000	1898.000000
mean	1.477496e+06	171168.478398	16.498851	27.371970	24.161749	3.248842	51.533720
std	5.480497e+02	113698.139743	7.483812	4.632481	4.972637	2.295598	6.833603
min	1.476547e+06	1311.000000	4.470000	20.000000	15.000000	0.000000	35.000000
25%	1.477021e+06	77787.750000	12.080000	23.000000	20.000000	1.812000	47.000000
50%	1.477496e+06	128600.000000	14.140000	27.000000	25.000000	2.121000	52.000000
75%	1.477970e+06	270525.000000	22.297500	31.000000	28.000000	5.574375	56.000000
max	1.478444e+06	405334.000000	35.410000	35.000000	33.000000	8.852500	68.000000

Univariate Analysis: Customer ID, Unique ID and Cuisine type

- There are 1898 orders unique placed from 1200 unique customers
- There are 178 unique restaurants in the dataset from which orders are placed
- There are 14 unique cuisine types ordered on the FoodHub App

1

Cuisine Type

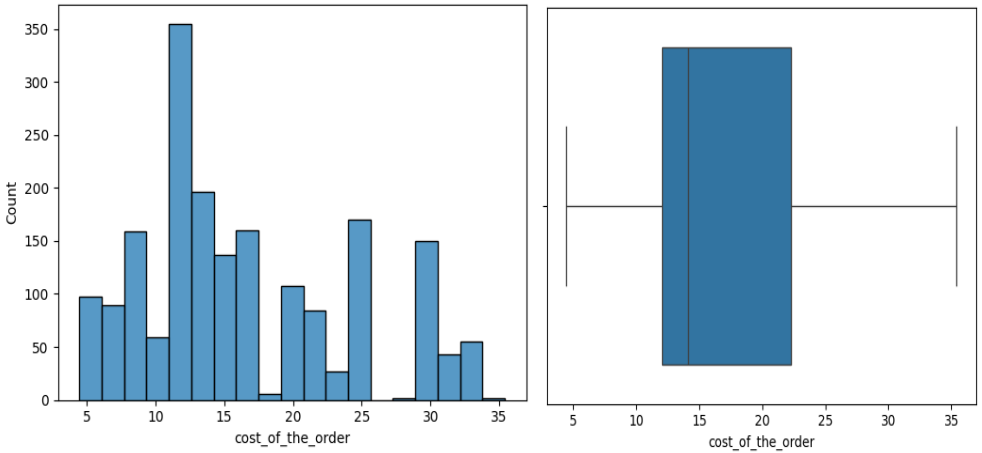


Key observations

- American, Japanese and Italian are the most popular cuisines
- Vietnamese, Korean and Spanish are among the least popular cuisines

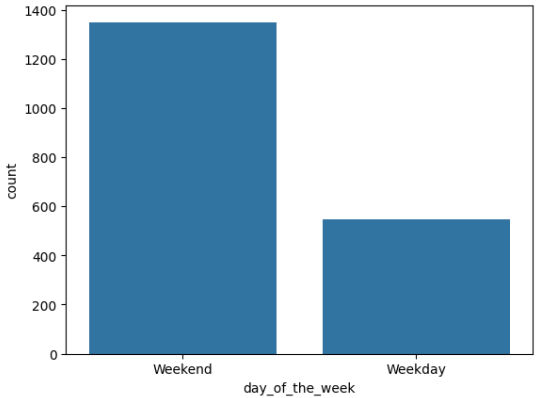
Univariate Analysis: Cuisine Type and Cost of order

2 Cost of the order



- The average cost of the order is $16.4, with 25% of the orders costing <$12 and 75% costing less than $22.3
- There are no outliers

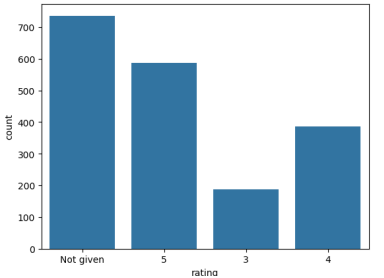
3 Day of the week

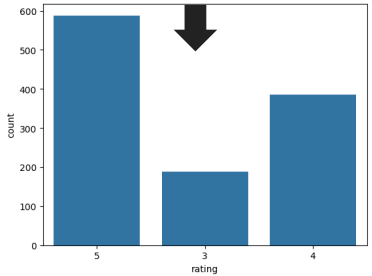


- Out of the 1898 orders, 1351 orders (71%) are placed over the weekend and 547 (29%) orders are placed during the weekdays.

Univariate Analysis: Day of the week, rating and restaurant type

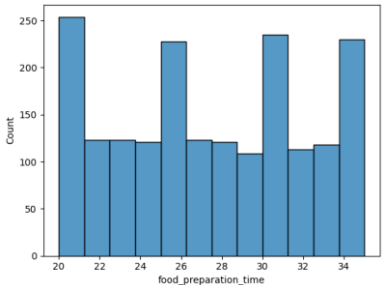
4 Rating

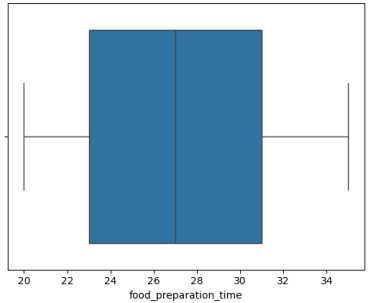




- Out of the 1898 orders, 736 orders were not rated and created missing values which we treated by replacing them with NaN
- After treatment, orders were rated either 3, 4 or 5
- Among rated orders, 588 (51%) orders were rated 5, 386 orders were rated 4 (33%) and, remaining 188 orders (16%) were rated 3. Average rating was 4.3

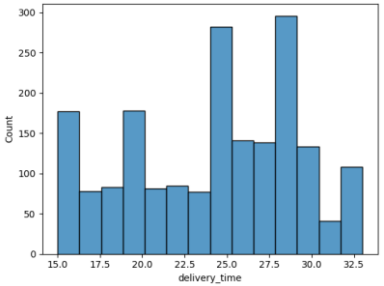
5 Food preparation time

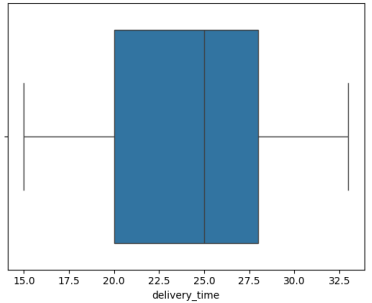




- The average food preparation time is 27.4 minutes
- 25% of the orders were prepared in <23 and 75% in <31 minutes
- There are no outliers

6 Food delivery time





- The average delivery time is ~24.2 minutes
- 25% of the orders were prepared in <20 minutes and 75% in <28 minutes
- There are no outliers

Other observations

Top 5 restaurants

Top 5 restaurants	# of orders placed
Shake Shack	219
The Meatball Shop	132
Blue Ribbon Sushi	119
Blue Ribbon Fried Chicken	96
Parm	68

Top 5 restaurants represented 33% of the total orders

Most Popular cuisine over weekends: American food

Cuisine Type	Orders placed over the weekend
American	415
Japanese	335
Italian	207
Chinese	163
Mexican	53
Indian	49
Mediterranean	32
Middle Eastern	32
Thai	15
French	13
Korean	11
Southern	11
Spanish	11
Vietnamese	4

Orders that cost more than $20

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

Mean Order Delivery time

The mean delivery time for this dataset is ~24.2 minutes

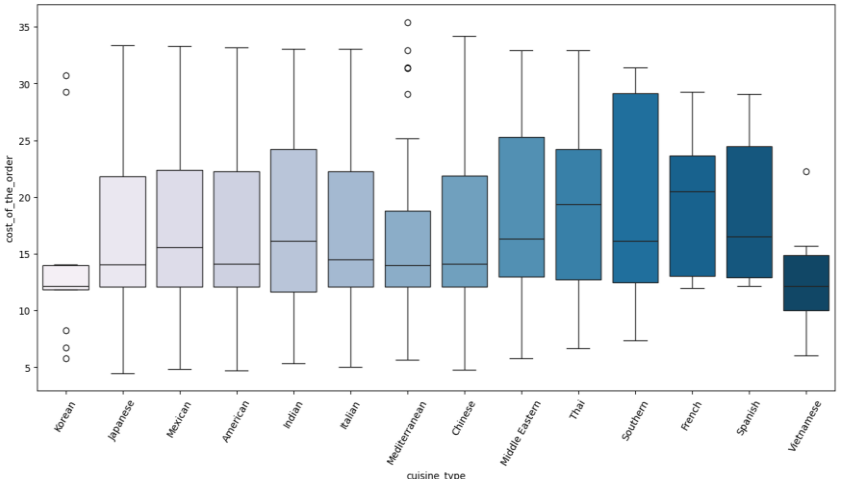
The top 3 most frequent customers

Customer ID	# of orders placed
52832	13
47440	10
83287	9

To these customers, the company has decided to give a 20% discount

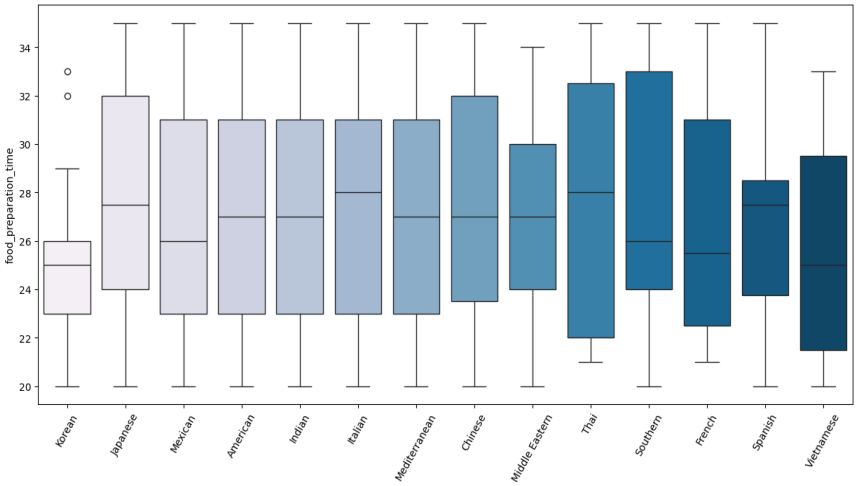
Multivariate Analysis: Cuisine type vs cost of the order and food preparation time

1 Cuisine type vs cost of the order



- Korean and Vietnamese food have the lowest average cost of order; they also have a narrow range of cost of the order
- French food has the highest cost per order, followed by Thai
- Southern, Middle Eastern and Indian food have a relatively wider range of cost per order
- Korean food, Vietnamese food and Mediterranean food have outliers

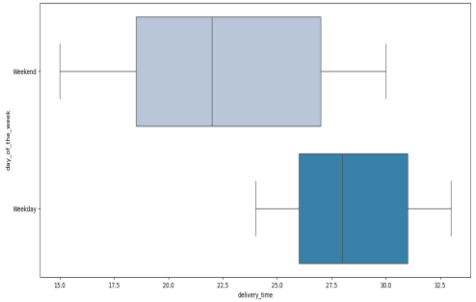
2 Cuisine type vs food preparation time



- Food preparation time is on average the highest for Italian and Thai food while Korean and Vietnamese food have lower food preparation time.
- Thai food preparation time also has widest range in terms of food preparation time while Korean food has the narrowest.
- Only Korean food has outliers

Multivariate Analysis

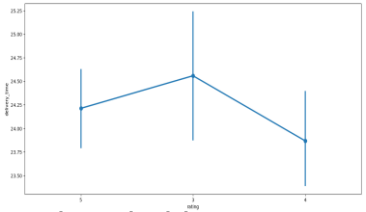
3 Relationship between day of the week and delivery time



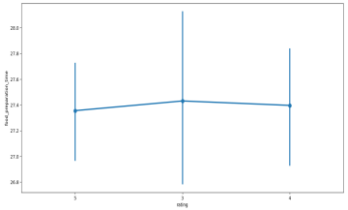
- Average delivery time during weekdays is 28.3 and during weekends is 22.4 minutes
- Mean delivery time is higher during weekdays vs weekends, potentially due to higher demand from office workers

4 Relationship between rating and other variables like food preparation time, delivery time and cost of order

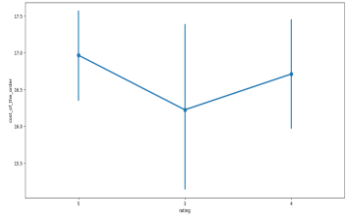
Rating and delivery time



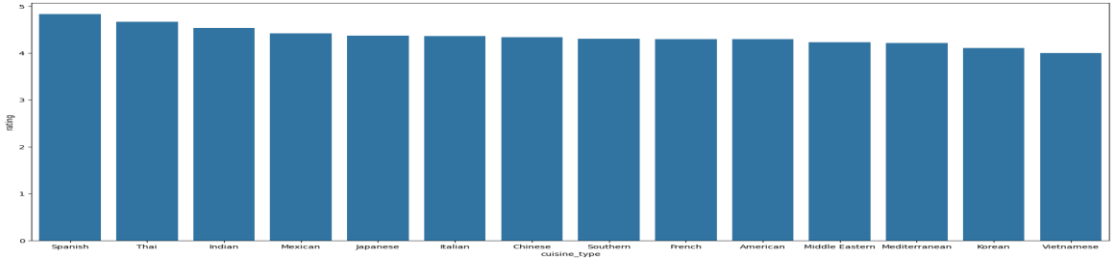
Rating and food preparation



Rating and cost of the order



Rating and cuisine type

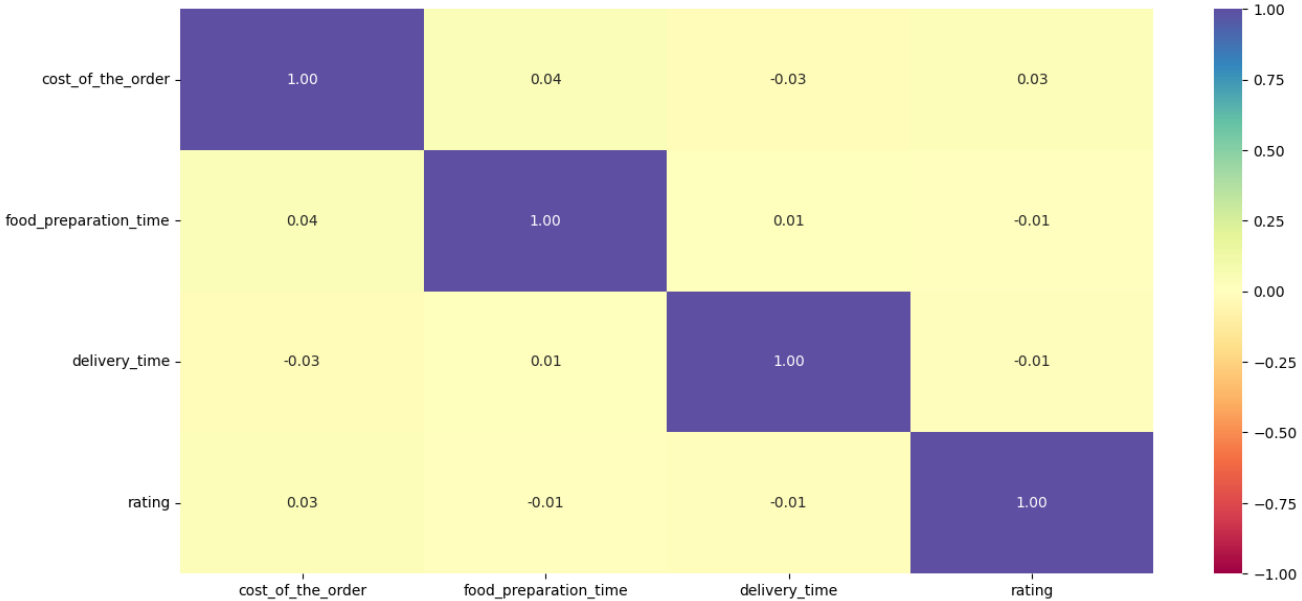


- Interestingly, delivery time is not the lowest for the highest-rated orders but it is the highest at ~24.5 minutes for the lowest-rated orders; thus, reducing delivery time can potentially improve rating.
- Food preparation time is similar on average across rating
- Higher cost orders have a higher rating of 4 or 5 while lower cost orders have a lower rating of 3. One possible explanation is expensive restaurants are preparing better food, which is attracting a higher rating
- Ratings for Vietnamese, Korean & Mediterranean food is lower, while that for Spanish, Thai & Indian food is higher.

Multivariate Analysis: Correlation Matrix

5

Relationship between rating and other variables like food preparation time, delivery time and cost of order



- The heatmap shows the correlation between cost of the order, food preparation time, delivery time and rating.
- There is low correlation between all the different variables

Restaurants Data

6

Top restaurants by revenue generated

Restaurant Name	Amount
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

- Shake Shack, The Meatball shop and Blue Ribbon have the highest revenue generated among the restaurants;
- Top 5 restaurants contribute to 33% of revenues

7

Restaurants fulfilling criteria to get promotional offers

	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

These restaurants can have a rating count of >50 and average rating of >4

8

Total revenue generated by company across all orders

If the company charges the restaurant 25% on the orders that cost >$20 and 15% on orders that cost >$5, then the total revenue generated for the restaurant is **$6166.3.** Below is a snapshot of the first five rows of the table

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time	Revenue
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20	7.6875
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23	1.8120
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28	1.8345
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15	7.3000
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24	1.7385

Other observations

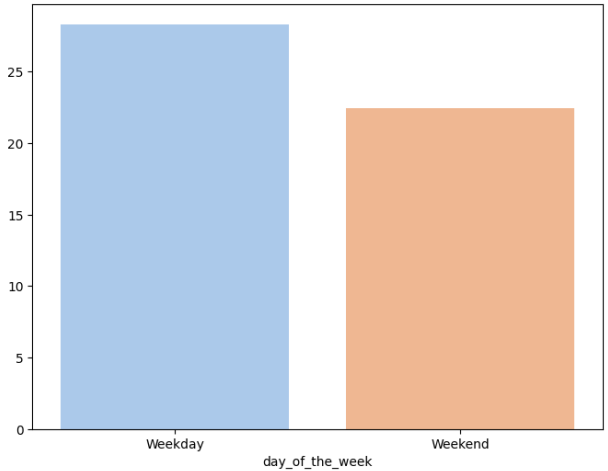
9

Orders that took >60 minutes to deliver from the time order is placed

After calculating the total time to deliver orders (food preparation time + delivery time), we find that **200 orders** took more than 60 minutes to get delivered; this translates to **10.54% of the orders**

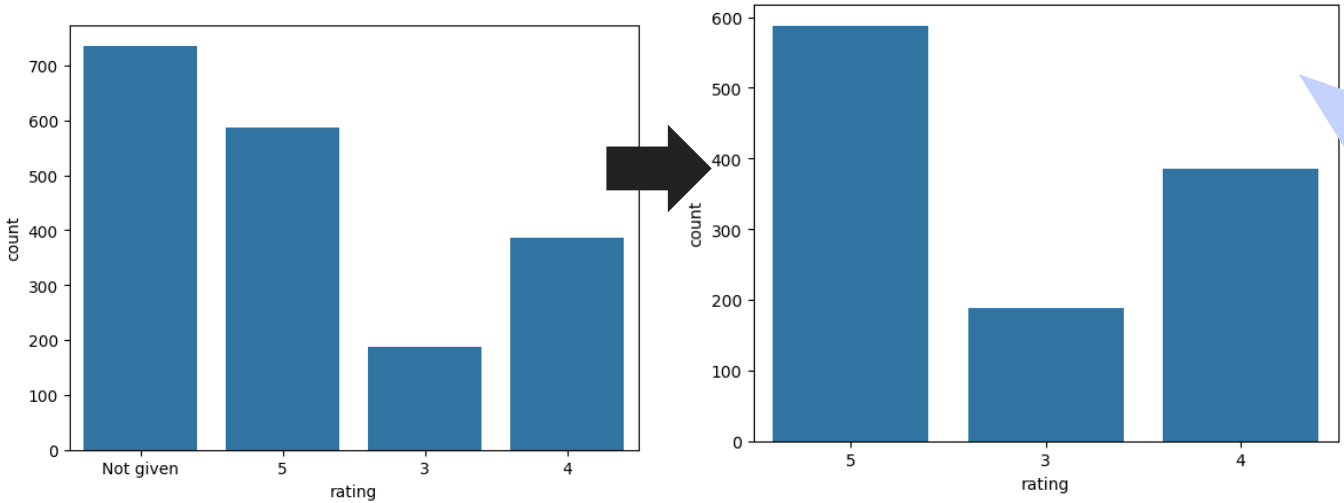
Delivery time by weekdays and weekends

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



APPENDIX

Missing value treatment and other assumptions



Replacement with NaN because replacing with a mean rating would have not been representative of true customer experience. Given how crucial this data is, we decided to leave it as it is

Additionally, code hasn't been shown but code output has been shown in select places to answer the specific questions asked by the Data Science Team.

Great Learning

Happy Learning !