



# Food Hub Data Analysis

BY- Gauri Shivhare

202200319788

TO- Mr. Nishant Chaturvedi

## Data

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.

## Data Description

The data contains the different data related to a food order. The detailed data dictionary is given below. Data Dictionary • 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 • 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

1- Display the top 5 rows.

[27]:	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	25.0	20
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25.0	?
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23.0	28
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25.0	15
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25.0	24

## 2- Display the last 5 rows

[14]:	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
1893	1476701	292602	Chipotle Mexican Grill \$1.99 Delivery	Mexican	22.31	Weekend	5	31.0	17
1894	1477421	397537	The Smile	American	12.18	Weekend	5	31.0	19
1895	1477819	35309	Blue Ribbon Sushi	Japanese	25.22	Weekday	Not given	31.0	24
1896	1477513	64151	Jack's Wife Freda	Mediterranean	12.18	Weekday	5	23.0	31
1897	1478056	120353	Blue Ribbon Sushi	Japanese	19.45	Weekend	Not given	28.0	24

## 4- Check the datatypes of each feature.

```
i]: (1898, 9)
```

## 5- Check the Statistical summary

[16]:	order_id	customer_id	cost_of_the_order	food_preparation_time
count	1.898000e+03	1898.000000	1898.000000	1896.000000
mean	1.477496e+06	171168.478398	80.722007	27.371835
std	5.480497e+02	113698.139743	2798.141333	4.634211
min	1.476547e+06	1311.000000	0.000000	20.000000
25%	1.477021e+06	77787.750000	12.080000	23.000000
50%	1.477496e+06	128600.000000	14.160000	27.000000
75%	1.477970e+06	270525.000000	22.310000	31.000000
max	1.478444e+06	405334.000000	121920.000000	35.000000

## 6- Check the null values

[18]:

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
1893	False	False	False	False	False	False	False	False	False
1894	False	False	False	False	False	False	False	False	False
1895	False	False	False	False	False	False	False	False	False
1896	False	False	False	False	False	False	False	False	False
1897	False	False	False	False	False	False	False	False	False

1898 rows × 9 columns

7- Check the duplicate values

```
[19]: order_id          0
      customer_id       0
      restaurant_name   0
      cuisine_type       3
      cost_of_the_order  0
      day_of_the_week    0
      rating             0
      food_preparation_time 2
      delivery_time      0
      dtype: int64
```

8- Check the anomalies or wrong entries.

5

9- Check the outliers and their authenticity.

10- Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.

Box plot showing the distribution of four variables: order\_id, customer\_id, cost\_of\_the\_order, and food\_preparation\_time. The x-axis represents values from 0.0 to 1.4, with a multiplier of 1e6 at the bottom right. The y-axis lists the variables. order\_id has a very narrow distribution near 1.5e6. customer\_id has a wider distribution from approximately 0.05e6 to 0.4e6. cost\_of\_the\_order and food\_preparation\_time have very narrow distributions near 0.0e6, with a single outlier for cost\_of\_the\_order at approximately 0.15e6.



