Sprint I
Dataset Collection

Date	1 November 2022
Team ID	PNT2022TMID27576
Project Name	DemandEst - AI powered Food Demand Forecaster

Screenshots:

Dataset has been collected and explored.

In [1]:
```
import pandas as pd
```

In [5]:
```
train =  pd.read_csv("../Dataset/train.csv")
test = pd.read_csv("../Dataset/test.csv")
fulfilment_center = pd.read_csv("../Dataset/fulfilment_center_info.csv")
meal_info = pd.read_csv("../Dataset/meal_info.csv")
```

In [6]:
```
train.head()
```

Out[6]:

	id	week	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	num_orders
0	1379560	1	55	1885	136.83	152.29	0	0	177
1	1466964	1	55	1993	136.83	135.83	0	0	270
2	1346989	1	55	2539	134.86	135.86	0	0	189
3	1338232	1	55	2139	339.50	437.53	0	0	54
4	1448490	1	55	2631	243.50	242.50	0	0	40

In [7]:
```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 456548 entries, 0 to 456547
Data columns (total 9 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   id                    456548 non-null  int64
 1   week                  456548 non-null  int64
 2   center_id             456548 non-null  int64
 3   meal_id               456548 non-null  int64
 4   checkout_price        456548 non-null  float64
 5   base_price            456548 non-null  float64
 6   emailer_for_promotion  456548 non-null  int64
 7   homepage_featured     456548 non-null  int64
 8   num_orders            456548 non-null  int64
dtypes: float64(2), int64(7)
memory usage: 31.3 MB
```

In [9]:
```
train.describe()
```

Out[9]:

	id	week	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	num_orders
count	4.565480e+05	456548.000000	456548.000000	456548.000000	456548.000000	456548.000000	456548.000000	456548.00000	456548.000000
mean	1.250096e+06	74.768771	82.105796	2024.337458	332.238933	354.156627	0.081152	0.10920	261.872760
std	1.443548e+05	41.524956	45.975046	547.420920	152.939723	160.715914	0.273069	0.31189	395.922798
min	1.000000e+06	1.000000	10.000000	1062.000000	2.970000	55.350000	0.000000	0.00000	13.000000
25%	1.124999e+06	39.000000	43.000000	1558.000000	228.950000	243.500000	0.000000	0.00000	54.000000
50%	1.250184e+06	76.000000	76.000000	1993.000000	296.820000	310.460000	0.000000	0.00000	136.000000
75%	1.375140e+06	111.000000	110.000000	2539.000000	445.230000	458.870000	0.000000	0.00000	324.000000
max	1.499999e+06	145.000000	186.000000	2956.000000	866.270000	866.270000	1.000000	1.00000	24299.000000

In [10]:
```
fulfilment_center.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77 entries, 0 to 76
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   center_id    77 non-null     int64  
 1   city_code    77 non-null     int64  
 2   region_code  77 non-null     int64  
 3   center_type  77 non-null     object 
 4   op_area      77 non-null     float64
dtypes: float64(1), int64(3), object(1)
memory usage: 3.1+ KB
```

In [12]:
```
meal_info.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   meal_id   51 non-null     int64 
 1   category  51 non-null     object
 2   cuisine   51 non-null     object
dtypes: int64(1), object(2)
memory usage: 1.3+ KB
```

In [13]:
```
fulfilment_center.head()
```

Out[13]:

	center_id	city_code	region_code	center_type	op_area
0	11	679	56	TYPE_A	3.7
1	13	590	56	TYPE_B	6.7
2	124	590	56	TYPE_C	4.0
3	66	648	34	TYPE_A	4.1
4	94	632	34	TYPE_C	3.6

In [14]:
```
meal_info.head()
```

Out[14]:

	meal_id	category	cuisine
0	1885	Beverages	Thai
1	1993	Beverages	Thai
2	2539	Beverages	Thai
3	1248	Beverages	Indian
4	2631	Beverages	Indian