

TEAM ID: PNT2022TMID32006

PROJECT NAME: DemandEst - AI powered Food DemandForecaster

Exploratory Data Analysis

In [96]: `train.head()`

Out[96]:

| | 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 [97]: `test.head()`

Out[97]:

| | id | week | center_id | meal_id | checkout_price | base_price | emailer_for_promotion | homepage_featured |
|---|---------|------|-----------|---------|----------------|------------|-----------------------|-------------------|
| 0 | 1028232 | 146 | 55 | 1885 | 158.11 | 159.11 | 0 | 0 |
| 1 | 1127204 | 146 | 55 | 1993 | 160.11 | 159.11 | 0 | 0 |
| 2 | 1212707 | 146 | 55 | 2539 | 157.14 | 159.14 | 0 | 0 |
| 3 | 1082698 | 146 | 55 | 2631 | 162.02 | 162.02 | 0 | 0 |
| 4 | 1400926 | 146 | 55 | 1248 | 163.93 | 163.93 | 0 | 0 |

In [98]: `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 [99]: `test.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32573 entries, 0 to 32572
Data columns (total 8 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   id                   32573 non-null  int64
1   week                 32573 non-null  int64
2   center_id            32573 non-null  int64
3   meal_id              32573 non-null  int64
```

```
3 meal_id 32573 non-null int64
4 checkout_price 32573 non-null float64
5 base_price 32573 non-null float64
6 emailer_for_promotion 32573 non-null int64
7 homepage_featured 32573 non-null int64
dtypes: float64(2), int64(6)
memory usage: 2.0 MB
```

```
In [100]: train['num_orders'].describe()
```

```
Out[100]: count    456548.000000
           mean      261.872760
           std       395.922798
           min        13.000000
           25%       54.000000
           50%      136.000000
           75%      324.000000
           max     24299.000000
           Name: num_orders, dtype: float64
```

```
In [101]: train.describe()
```

| Out[101]: | 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 000000 | 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% | 1.244996e+06 | 39.000000 | 43.000000 | 1558.000000 | 228.950000 | 243.500000 | 0.000000 | 0.00000 | 54.000000 |