

Team ID: PNT2022TMID23576

## PROJECT NAME: DemandEst - AI powered Food Demand Forecaster

TEAM LEADER:

```
train.head()

  id week center_id meal_id checkout_price base_price emailer_for_promotion homepage_featured num_orders
0  1378560    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

[ ] test.head()

  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

[ ] 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
```

```
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

[ ] train[\"num_orders\"].describe()

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
```

## TEAM MEMBER 1:

The screenshot displays a Google Colab notebook interface with two tabs: 'Data\_preprocessing\_and\_Model' and 'Model evaluation last 3 pdfs - sh'. The notebook is open to the 'Data\_preprocessing\_and\_Model' tab, showing a series of code cells and their outputs.

**Code Cell 1:** `train.head()`

**Output 1:** A table showing the first 5 rows of the 'train' dataset.

	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

**Code Cell 2:** `test.head()`

**Output 2:** A table showing the first 5 rows of the 'test' dataset.

	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

**Code Cell 3:** `train.info()`

**Output 3:** Information about the 'train' dataset.

```
<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
```

**Code Cell 4:** `train['num_orders'].describe()`

**Output 4:** Summary statistics for the 'num\_orders' column.

	count	456548.000000
mean	261.872760	
std	395.922798	
min	13.000000	
25%	54.000000	
50%	136.000000	
75%	324.000000	
max	24299.000000	

The notebook interface includes a search bar at the top, a toolbar with various icons, and a status bar at the bottom showing the time as 14:28 on 20-11-2022.

## TEAM MEMBER 2:

The first screenshot shows a Google Colab notebook with the following code and output:

```
train.head()
```

	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

```
test.head()
```

	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

```
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
```

The second screenshot shows the same notebook with the following code and output:

```
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
```

```
train['num_orders'].describe()
```

	count	mean	std	min	25%	50%	75%	max
num_orders	456548.000000	261.872760	395.922798	13.000000	54.000000	136.000000	224.000000	24299.000000

Name: num\_orders, dtype: float64

## TEAM MEMBER 3:

train.head()

	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

[ ] test.head()

	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

[ ] 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
```

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
```

[ ] train['num\_orders'].describe()

	count	mean	std	min	25%	50%	75%	max
num_orders	456548.000000	261.872760	395.922798	13.000000	54.000000	136.000000	324.000000	24299.000000

Name: num\_orders, dtype: float64

## TEAM MEMBER 4:

The first screenshot shows a Google Colab notebook with the following code and output:

```
train.head()
```

	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

```
test.head()
```

	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

```
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
```

The second screenshot shows the same notebook with the following code and output:

```
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
```

```
train['num_orders'].describe()
```

	count	mean	std	min	25%	50%	75%	max
num_orders	456548.000000	261.872760	395.922798	13.000000	54.000000	136.000000	224.000000	24299.000000

Name: num\_orders, dtype: float64