

DEMAND_orr

April 6, 2024

```
[22]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import sklearn
from sklearn.model_selection import train_test_split
import warnings
warnings.filterwarnings("ignore")
```

```
[3]: train1=pd.read_csv("/home/lenovo/Downloads/Documents/new/train.csv")
meal=pd.read_csv("/home/lenovo/Downloads/Documents/new/meal_info.csv")
fu=pd.read_csv("/home/lenovo/Downloads/Documents/new/fulfilment_center_info.
↪CSV")
```

```
[4]: meal.head()
```

```
[4]:
```

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

```
[5]: fu.head()
```

```
[5]:
```

	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

```
[6]: train1.head()
```

```
[6]:
```

	id	week	center_id	meal_id	checkout_price	base_price	\
0	1379560	1	55	1885	136.83	152.29	

1	1466964	1	55	1993	136.83	135.83
2	1346989	1	55	2539	134.86	135.86
3	1338232	1	55	2139	339.50	437.53
4	1448490	1	55	2631	243.50	242.50

	emailer_for_promotion	homepage_featured	num_orders
0	0	0	177
1	0	0	270
2	0	0	189
3	0	0	54
4	0	0	40

```
[7]: train1.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
```

```
[8]: trainfinal=pd.merge(train1,meal,on="meal_id",how="outer")
trainfinal
```

```
[8]:
```

	id	week	center_id	meal_id	checkout_price	base_price	\
0	1379560	1	55	1885	136.83	152.29	
1	1040403	1	24	1885	136.83	136.83	
2	1103215	1	11	1885	136.83	136.83	
3	1034383	1	83	1885	121.31	120.31	
4	1118999	1	32	1885	114.52	113.52	
...	
456543	1458599	145	65	2104	582.03	582.03	
456544	1417199	145	110	2104	582.03	581.03	
456545	1434028	145	68	2104	582.03	581.03	
456546	1205895	145	51	2104	641.23	640.23	
456547	1443704	145	61	2104	581.03	582.03	

	emailer_for_promotion	homepage_featured	num_orders	category \
0	0	0	177	Beverages
1	0	0	1498	Beverages
2	0	0	959	Beverages
3	0	1	1025	Beverages
4	0	1	445	Beverages
...
456543	0	0	134	Fish
456544	0	0	28	Fish
456545	0	0	40	Fish
456546	0	0	69	Fish
456547	0	0	27	Fish

	cuisine
0	Thai
1	Thai
2	Thai
3	Thai
4	Thai
...	...
456543	Continental
456544	Continental
456545	Continental
456546	Continental
456547	Continental

[456548 rows x 11 columns]

```
[9]: trainfinal1=pd.merge(trainfinal,fu,on="center_id",how="outer")
trainfinal1
```

	id	week	center_id	meal_id	checkout_price	base_price \
0	1379560	1	55	1885	136.83	152.29
1	1018704	2	55	1885	135.83	152.29
2	1196273	3	55	1885	132.92	133.92
3	1116527	4	55	1885	135.86	134.86
4	1343872	5	55	1885	146.50	147.50
...
456543	1437396	141	61	2104	583.03	630.53
456544	1060716	142	61	2104	581.03	582.03
456545	1029231	143	61	2104	583.03	581.03
456546	1141147	144	61	2104	582.03	581.03
456547	1443704	145	61	2104	581.03	582.03

	emailer_for_promotion	homepage_featured	num_orders	category \
0	0	0	177	Beverages
1	0	0	323	Beverages

2	0	0	96	Beverages
3	0	0	163	Beverages
4	0	0	215	Beverages
...
456543	0	1	13	Fish
456544	0	0	42	Fish
456545	0	0	40	Fish
456546	0	0	53	Fish
456547	0	0	27	Fish

	cuisine	city_code	region_code	center_type	op_area
0	Thai	647	56	TYPE_C	2.0
1	Thai	647	56	TYPE_C	2.0
2	Thai	647	56	TYPE_C	2.0
3	Thai	647	56	TYPE_C	2.0
4	Thai	647	56	TYPE_C	2.0
...
456543	Continental	473	77	TYPE_A	4.5
456544	Continental	473	77	TYPE_A	4.5
456545	Continental	473	77	TYPE_A	4.5
456546	Continental	473	77	TYPE_A	4.5
456547	Continental	473	77	TYPE_A	4.5

[456548 rows x 15 columns]

```
[10]: from sklearn.preprocessing import LabelEncoder
lb1=LabelEncoder()
trainfinal1['category']=lb1.fit_transform(trainfinal1['category'])
trainfinal1['cuisine']=lb1.fit_transform(trainfinal1['cuisine'])
trainfinal1['center_type']=lb1.fit_transform(trainfinal1['center_type'])
trainfinal1
```

```
[10]:
```

	id	week	center_id	meal_id	checkout_price	base_price	\
0	1379560	1	55	1885	136.83	152.29	
1	1018704	2	55	1885	135.83	152.29	
2	1196273	3	55	1885	132.92	133.92	
3	1116527	4	55	1885	135.86	134.86	
4	1343872	5	55	1885	146.50	147.50	
...
456543	1437396	141	61	2104	583.03	630.53	
456544	1060716	142	61	2104	581.03	582.03	
456545	1029231	143	61	2104	583.03	581.03	
456546	1141147	144	61	2104	582.03	581.03	
456547	1443704	145	61	2104	581.03	582.03	

	emailer_for_promotion	homepage_featured	num_orders	category	\
0	0	0	177	0	

1		0	0	323	0
2		0	0	96	0
3		0	0	163	0
4		0	0	215	0
...	
456543		0	1	13	4
456544		0	0	42	4
456545		0	0	40	4
456546		0	0	53	4
456547		0	0	27	4

	cuisine	city_code	region_code	center_type	op_area
0	3	647	56	2	2.0
1	3	647	56	2	2.0
2	3	647	56	2	2.0
3	3	647	56	2	2.0
4	3	647	56	2	2.0
...
456543	0	473	77	0	4.5
456544	0	473	77	0	4.5
456545	0	473	77	0	4.5
456546	0	473	77	0	4.5
456547	0	473	77	0	4.5

[456548 rows x 15 columns]

```
[11]: trainfinal1=trainfinal1.drop(["op_area"],axis=1)
trainfinal1
```

```
[11]:
```

	id	week	center_id	meal_id	checkout_price	base_price	\
0	1379560	1	55	1885	136.83	152.29	
1	1018704	2	55	1885	135.83	152.29	
2	1196273	3	55	1885	132.92	133.92	
3	1116527	4	55	1885	135.86	134.86	
4	1343872	5	55	1885	146.50	147.50	
...	
456543	1437396	141	61	2104	583.03	630.53	
456544	1060716	142	61	2104	581.03	582.03	
456545	1029231	143	61	2104	583.03	581.03	
456546	1141147	144	61	2104	582.03	581.03	
456547	1443704	145	61	2104	581.03	582.03	

	emailer_for_promotion	homepage_featured	num_orders	category	\
0		0	177		0
1		0	323		0
2		0	96		0
3		0	163		0

4	0	0	215	0
...
456543	0	1	13	4
456544	0	0	42	4
456545	0	0	40	4
456546	0	0	53	4
456547	0	0	27	4

	cuisine	city_code	region_code	center_type
0	3	647	56	2
1	3	647	56	2
2	3	647	56	2
3	3	647	56	2
4	3	647	56	2
...
456543	0	473	77	0
456544	0	473	77	0
456545	0	473	77	0
456546	0	473	77	0
456547	0	473	77	0

[456548 rows x 14 columns]

```
[12]: trainfinal1=trainfinal1.drop(["center_id","meal_id"],axis=1)
```

```
[13]: trainfinal1.dropna()
```

```
[13]:
```

	id	week	checkout_price	base_price	emailer_for_promotion	\
0	1379560	1	136.83	152.29		0
1	1018704	2	135.83	152.29		0
2	1196273	3	132.92	133.92		0
3	1116527	4	135.86	134.86		0
4	1343872	5	146.50	147.50		0
...
456543	1437396	141	583.03	630.53		0
456544	1060716	142	581.03	582.03		0
456545	1029231	143	583.03	581.03		0
456546	1141147	144	582.03	581.03		0
456547	1443704	145	581.03	582.03		0

	homepage_featured	num_orders	category	cuisine	city_code	\
0	0	177	0	3	647	
1	0	323	0	3	647	
2	0	96	0	3	647	
3	0	163	0	3	647	
4	0	215	0	3	647	
...

456543	1	13	4	0	473
456544	0	42	4	0	473
456545	0	40	4	0	473
456546	0	53	4	0	473
456547	0	27	4	0	473

	region_code	center_type
0	56	2
1	56	2
2	56	2
3	56	2
4	56	2
...
456543	77	0
456544	77	0
456545	77	0
456546	77	0
456547	77	0

[456548 rows x 12 columns]

```
[14]: trainfinal1.isnull().sum()
```

```
[14]: id                0
      week              0
      checkout_price    0
      base_price        0
      emailer_for_promotion  0
      homepage_featured  0
      num_orders        0
      category          0
      cuisine           0
      city_code         0
      region_code       0
      center_type       0
      dtype: int64
```

```
[15]: print(trainfinal1.info())
      print(trainfinal1.describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 456548 entries, 0 to 456547
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    456548 non-null  int64
1   week                  456548 non-null  int64
```

```

2  checkout_price      456548 non-null float64
3  base_price          456548 non-null float64
4  emailer_for_promotion 456548 non-null int64
5  homepage_featured    456548 non-null int64
6  num_orders           456548 non-null int64
7  category             456548 non-null int64
8  cuisine              456548 non-null int64
9  city_code            456548 non-null int64
10 region_code          456548 non-null int64
11 center_type          456548 non-null int64

```

dtypes: float64(2), int64(10)

memory usage: 41.8 MB

None

	id	week	checkout_price	base_price \
count	4.565480e+05	456548.000000	456548.000000	456548.000000
mean	1.250096e+06	74.768771	332.238933	354.156627
std	1.443548e+05	41.524956	152.939723	160.715914
min	1.000000e+06	1.000000	2.970000	55.350000
25%	1.124999e+06	39.000000	228.950000	243.500000
50%	1.250184e+06	76.000000	296.820000	310.460000
75%	1.375140e+06	111.000000	445.230000	458.870000
max	1.499999e+06	145.000000	866.270000	866.270000

	emailer_for_promotion	homepage_featured	num_orders	category \
count	456548.000000	456548.000000	456548.000000	456548.000000
mean	0.081152	0.10920	261.872760	5.257443
std	0.273069	0.31189	395.922798	4.459965
min	0.000000	0.00000	13.000000	0.000000
25%	0.000000	0.00000	54.000000	0.000000
50%	0.000000	0.00000	136.000000	5.000000
75%	0.000000	0.00000	324.000000	9.000000
max	1.000000	1.00000	24299.000000	13.000000

	cuisine	city_code	region_code	center_type
count	456548.000000	456548.000000	456548.000000	456548.000000
mean	1.561961	601.553399	56.614566	0.642342
std	1.101974	66.195914	17.641306	0.816105
min	0.000000	456.000000	23.000000	0.000000
25%	1.000000	553.000000	34.000000	0.000000
50%	2.000000	596.000000	56.000000	0.000000
75%	3.000000	651.000000	77.000000	1.000000
max	3.000000	713.000000	93.000000	2.000000

```
[16]: trainfinal1.shape
```

```
[16]: (456548, 12)
```



```
[17]: #DATA Analysis
```

```
[18]: promoapplied=[1]
promodf=trainfinal1.loc[trainfinal1['emailer_for_promotion'].isin(promoapplied)]
print('Promotion Voucher Meal Count:', promodf['id'].count())
print('Overall Orders for promotional orders:', promodf['num_orders'].sum())
```

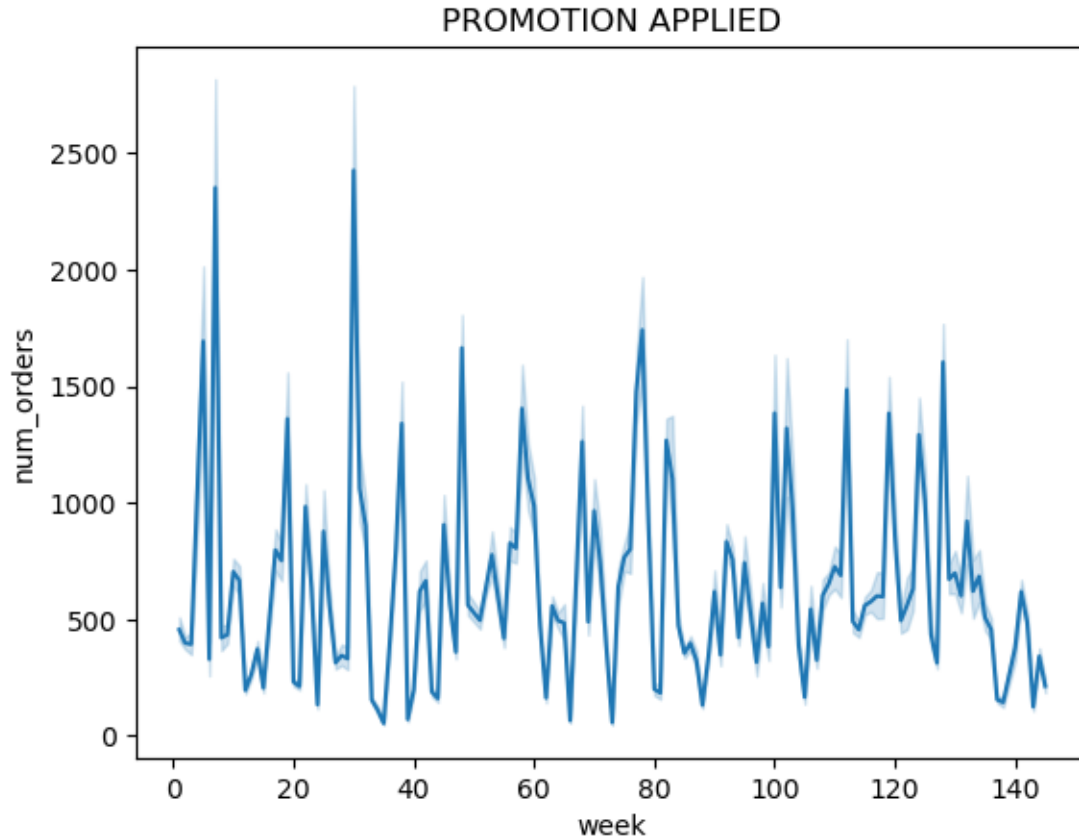
Promotion Voucher Meal Count: 37050
Overall Orders for promotional orders: 23382164

```
[19]: hromoapplied=[1]
hromodf=trainfinal1.loc[trainfinal1['homepage_featured'].isin(promoapplied)]
print('Promotion Voucher Meal Count:', promodf['id'].count())
print('Overall Orders for promotional orders:', promodf['num_orders'].sum())
```

Promotion Voucher Meal Count: 37050
Overall Orders for promotional orders: 23382164

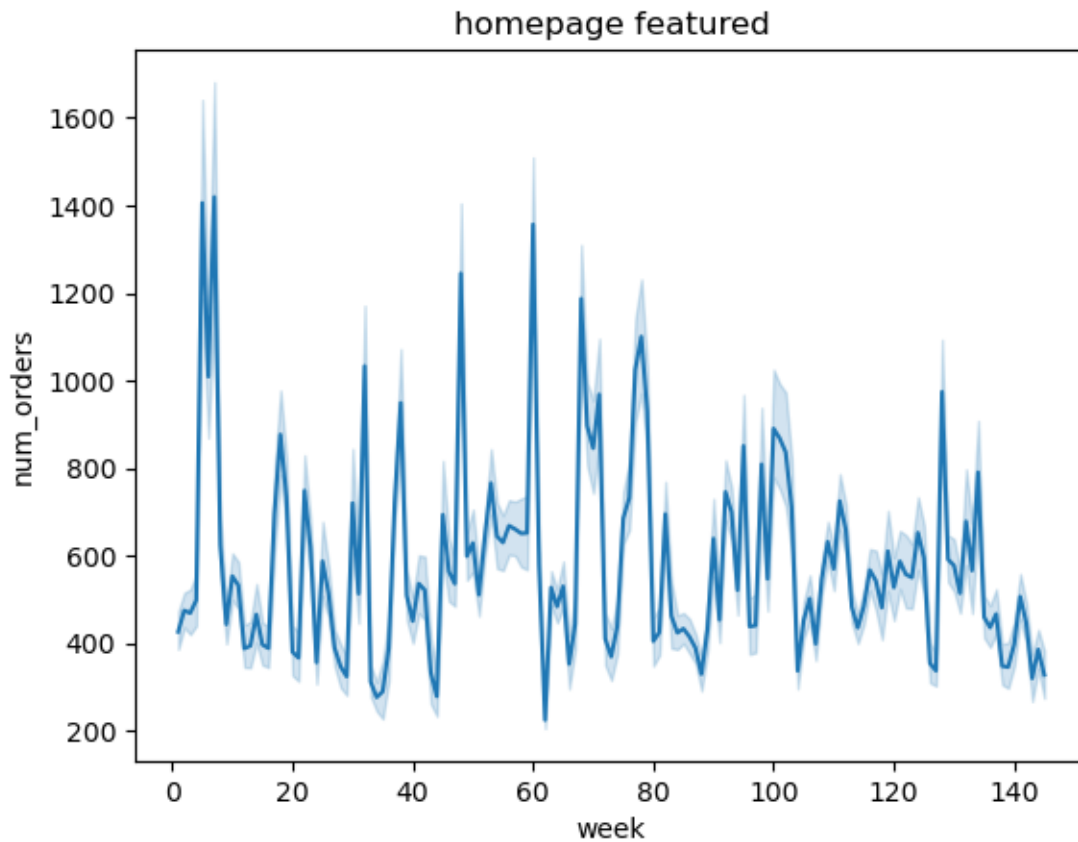
```
[23]: df=promodf
fig=sns.lineplot(df,x='week',y='num_orders').set(title='PROMOTION APPLIED')
fig
```

```
[23]: [Text(0.5, 1.0, 'PROMOTION APPLIED')]
```



```
[24]: df1=hromodf
fig=sns.lineplot(df1,x='week',y='num_orders').set(title='homepage featured')
fig
```

```
[24]: [Text(0.5, 1.0, 'homepage featured')]
```



```
[25]: #SPLIT DATA
```

```
[26]: trainfinal1
```

```
[26]:
```

	id	week	checkout_price	base_price	emailer_for_promotion	\
0	1379560	1	136.83	152.29		0
1	1018704	2	135.83	152.29		0
2	1196273	3	132.92	133.92		0
3	1116527	4	135.86	134.86		0
4	1343872	5	146.50	147.50		0
...

456543	1437396	141	583.03	630.53	0
456544	1060716	142	581.03	582.03	0
456545	1029231	143	583.03	581.03	0
456546	1141147	144	582.03	581.03	0
456547	1443704	145	581.03	582.03	0

	homepage_featured	num_orders	category	cuisine	city_code	\
0	0	177	0	3	647	
1	0	323	0	3	647	
2	0	96	0	3	647	
3	0	163	0	3	647	
4	0	215	0	3	647	
...	
456543	1	13	4	0	473	
456544	0	42	4	0	473	
456545	0	40	4	0	473	
456546	0	53	4	0	473	
456547	0	27	4	0	473	

	region_code	center_type
0	56	2
1	56	2
2	56	2
3	56	2
4	56	2
...
456543	77	0
456544	77	0
456545	77	0
456546	77	0
456547	77	0

[456548 rows x 12 columns]

```
[27]: x=trainfinal1.drop(['num_orders'],axis=1)
      y=trainfinal1['num_orders']
```

```
[28]: from sklearn.model_selection import train_test_split

      # Assuming 'x' contains the features and 'y' contains the target variable
      # 'test_size=0.20' specifies that 20% of the data will be used for testing, and
      ↪80% for training
      # 'random_state=0' sets the random seed to ensure reproducibility
      x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20,
      ↪random_state=0)
```

```
[29]: x_train
```

```
[29]:
```

	id	week	checkout_price	base_price	emailer_for_promotion	\
141016	1325459	105	581.03	582.03		0
406574	1485641	125	320.13	320.13		0
189872	1339897	56	534.53	533.53		0
323934	1006827	23	326.89	327.89		0
142275	1357504	40	323.98	324.98		0
...	
152315	1212361	87	485.03	485.03		0
439107	1338177	142	221.22	220.22		0
117952	1186492	23	341.50	408.43		0
435829	1436388	144	321.13	321.13		0
305711	1243490	82	309.49	308.49		0

	homepage_featured	category	cuisine	city_code	region_code	\
141016	0	4	0	522	56	
406574	0	5	3	526	34	
189872	0	6	2	713	85	
323934	0	9	2	526	34	
142275	0	0	0	615	34	
...	
152315	0	1	1	590	56	
439107	0	0	2	676	34	
117952	1	6	2	596	71	
435829	0	5	3	699	85	
305711	0	0	1	638	56	

	center_type
141016	0
406574	2
189872	0
323934	2
142275	1
...	...
152315	0
439107	1
117952	0
435829	2
305711	1

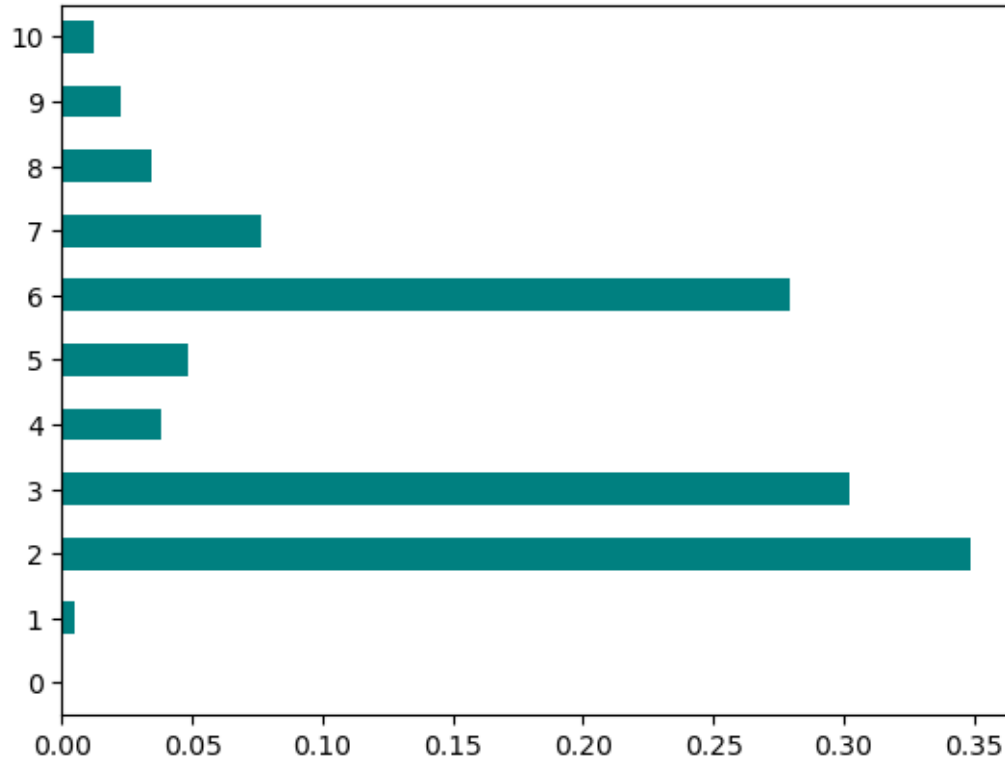
[365238 rows x 11 columns]

```
[30]: from sklearn.feature_selection import mutual_info_classif
%matplotlib inline
imp=mutual_info_classif(x_train,y_train)
imp
```

```
[30]: array([0.          , 0.004913   , 0.34859344, 0.30248776, 0.037814   ,  
          0.04827926, 0.27964634, 0.07668279, 0.03469004, 0.02268267,  
          0.01255353])
```

```
[31]: futureim=pd.Series(imp)  
futureim.plot(kind='barh',color="teal")
```

```
[31]: <Axes: >
```



```
[32]: from sklearn.linear_model import LinearRegression  
from sklearn.ensemble import RandomForestRegressor  
from sklearn.tree import DecisionTreeRegressor  
model=LinearRegression()  
model2=RandomForestRegressor(max_depth=3,random_state=0)  
model3=DecisionTreeRegressor(max_depth=3,random_state=0)
```

```
[33]: model.fit(x_train,y_train)
```

```
[33]: LinearRegression()
```

```
[34]: li=model.predict(x_test)  
li
```

```
[34]: array([289.60194036, 243.5403869 , 388.72269403, ..., 257.96884317,
          -4.40359529, 288.94314784])
```

```
[35]: from sklearn.metrics import r2_score, mean_squared_error
      from math import sqrt
      print('R2 Score:' ,r2_score(y_test,li))
      print('Mean Square Error :',mean_squared_error(y_test,li))
      print('RMSE :',sqrt(mean_squared_error(y_test,li)))
```

```
R2 Score: 0.21950351705753968
Mean Square Error : 123586.56533992218
RMSE : 351.5488093279825
```

```
[36]: model2.fit(x_train,y_train)
```

```
[36]: RandomForestRegressor(max_depth=3, random_state=0)
```

```
[37]: ri=model2.predict(x_test)
      ri
```

```
[37]: array([124.63783825, 163.67333426, 319.51315838, ..., 319.51315838,
          163.67333426, 317.98077467])
```

```
[38]: print('R2 Score:' ,r2_score(y_test,ri))
      print('Mean Square Error :',mean_squared_error(y_test,ri))
      print('RMSE :',sqrt(mean_squared_error(y_test,ri)))
```

```
R2 Score: 0.2659522976483112
Mean Square Error : 116231.70163086394
RMSE : 340.92770733817446
```

```
[39]: model3.fit(x_train, y_train)
```

```
[39]: DecisionTreeRegressor(max_depth=3, random_state=0)
```

```
[40]: vi=model3.predict(x_test)
      vi
```

```
[40]: array([124.69017341, 163.71525952, 319.60183796, ..., 319.60183796,
          163.71525952, 319.60183796])
```

```
[41]: print('R2 Score:' ,r2_score(y_test,ri))
      print('Mean Square Error :',mean_squared_error(y_test,ri))
      print('RMSE :',sqrt(mean_squared_error(y_test,ri)))
```

```
R2 Score: 0.2659522976483112
Mean Square Error : 116231.70163086394
RMSE : 340.92770733817446
```

```
[42]: print(x_train)
      print(y_train)
      print(x_test)
```

	id	week	checkout_price	base_price	emailer_for_promotion	\
141016	1325459	105	581.03	582.03		0
406574	1485641	125	320.13	320.13		0
189872	1339897	56	534.53	533.53		0
323934	1006827	23	326.89	327.89		0
142275	1357504	40	323.98	324.98		0
...
152315	1212361	87	485.03	485.03		0
439107	1338177	142	221.22	220.22		0
117952	1186492	23	341.50	408.43		0
435829	1436388	144	321.13	321.13		0
305711	1243490	82	309.49	308.49		0

	homepage_featured	category	cuisine	city_code	region_code	\
141016	0	4	0	522	56	
406574	0	5	3	526	34	
189872	0	6	2	713	85	
323934	0	9	2	526	34	
142275	0	0	0	615	34	
...
152315	0	1	1	590	56	
439107	0	0	2	676	34	
117952	1	6	2	596	71	
435829	0	5	3	699	85	
305711	0	0	1	638	56	

	center_type
141016	0
406574	2
189872	0
323934	2
142275	1
...	...
152315	0
439107	1
117952	0
435829	2
305711	1

[365238 rows x 11 columns]

141016	40
406574	42
189872	26

```

323934    323
142275    14
...
152315    42
439107    352
117952    55
435829    28
305711    68

```

Name: num_orders, Length: 365238, dtype: int64

	id	week	checkout_price	base_price	emailer_for_promotion	\
242838	1365542	61	281.33	280.33		0
314826	1495000	133	377.33	377.33		0
57041	1327310	62	262.93	260.93		0
87123	1324739	49	620.86	620.86		0
154611	1010902	41	165.87	167.87		0
...
231834	1123843	130	336.62	336.62		0
161380	1339156	7	319.13	320.13		0
173848	1113658	71	280.36	281.36		0
243709	1478282	31	640.23	640.23		0
82834	1457273	63	279.36	361.81		0

	homepage_featured	category	cuisine	city_code	region_code	\
242838	0	13	3	658	34	
314826	0	10	2	700	56	
57041	0	8	1	649	34	
87123	0	11	0	590	56	
154611	0	0	2	526	34	
...
231834	0	6	2	590	56	
161380	0	10	2	577	56	
173848	0	10	2	526	34	
243709	0	7	0	658	34	
82834	0	9	2	703	56	

	center_type
242838	1
314826	0
57041	0
87123	2
154611	2
...	...
231834	0
161380	0
173848	2
243709	1
82834	0

[91310 rows x 11 columns]

```
[43]: print(y_test)
```

```
242838    310
314826    593
57041     175
87123     53
154611     95
```

```
...
```

```
231834    28
161380    95
173848   122
243709    55
82834     541
```

Name: num_orders, Length: 91310, dtype: int64

```
[44]: print(ri)
```

```
[124.63783825 163.67333426 319.51315838 ... 319.51315838 163.67333426
 317.98077467]
```

```
[45]: from sklearn.ensemble import GradientBoostingRegressor
```

```
[46]: gbm=GradientBoostingRegressor(n_estimators=500, learning_rate=0.1, max_depth=3,
    ↪random_state=42)
```

```
[47]: gbm.fit(x_train,y_train)
```

```
[47]: GradientBoostingRegressor(n_estimators=500, random_state=42)
```

```
[49]: hi=gbm.predict(x_test)
hi
```

```
[49]: array([119.29089357, 420.08936444, 503.14927439, ..., 399.79509758,
 84.57086285, 331.86973118])
```

```
[50]: print('R2 Score: ',r2_score(y_test,hi))
print('Mean Square Error :',mean_squared_error(y_test,hi))
print('RMSE :',sqrt(mean_squared_error(y_test,hi)))
```

```
R2 Score: 0.6513140100616891
Mean Square Error : 55212.169203078134
RMSE : 234.9726988462237
```

```
[51]: new=pd.read_csv("/home/lenovo/Downloads/test_QoiM09B.csv")
new
```

```

[51]:
      id  week  center_id  meal_id  checkout_price  base_price  \
0    1028232   146        55    1885         158.11      159.11
1    1127204   146        55    1993         160.11      159.11
2    1212707   146        55    2539         157.14      159.14
3    1082698   146        55    2631         162.02      162.02
4    1400926   146        55    1248         163.93      163.93
...
32568  1250239   155        61    1543         482.09      484.09
32569  1039516   155        61    2304         483.09      483.09
32570  1158107   155        61    2664         322.07      323.07
32571  1444235   155        61    2569         322.07      323.07
32572  1291286   155        61    2490         276.45      276.45

      emailer_for_promotion  homepage_featured
0                        0                    0
1                        0                    0
2                        0                    0
3                        0                    0
4                        0                    0
...
32568                    0                    0
32569                    0                    0
32570                    0                    0
32571                    0                    0
32572                    0                    0

[32573 rows x 8 columns]

```