In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
```

In [2]:

```
train = pd.read_csv('train.csv')
meal_info = pd.read_csv('meal_info.csv')
full_filment = pd.read_csv('fulfilment_center_info.csv')
test = pd.read_csv('test.csv')
```

In [3]:

```
train.head()
```

Out[3]:

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

In [4]:

```
meal_info.head()
```

Out[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

In [5]:

```
full_filment.head()
```

Out[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

In [6]:

```
test.head()
```

Out[6]:

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

In [7]:

```
train.shape,test.shape,full_filment.shape,meal_info.shape
```

Out[7]:

```
((456548, 9), (32573, 8), (77, 5), (51, 3))
```

In [8]:

```
train=train.merge(meal_info,on='meal_id',how='left')
train=train.merge(full_filment,on='center_id',how='left')
```

In [9]:

```
test=test.merge(meal_info,on='meal_id',how='left')
test=test.merge(full_filment,on='center_id',how='left')
```

In [10]:

```
train.shape,test.shape
```

Out[10]:

```
((456548, 15), (32573, 14))
```

In [11]:

```
train.head()
```

Out[11]:

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

In [12]:

```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 456548 entries, 0 to 456547
Data columns (total 15 columns):
id                      456548 non-null int64
week                    456548 non-null int64
center_id               456548 non-null int64
meal_id                 456548 non-null int64
checkout_price          456548 non-null float64
base_price              456548 non-null float64
emailer_for_promotion   456548 non-null int64
homepage_featured       456548 non-null int64
num_orders              456548 non-null int64
category                456548 non-null object
cuisine                 456548 non-null object
city_code               456548 non-null int64
region_code             456548 non-null int64
center_type             456548 non-null object
op_area                 456548 non-null float64
dtypes: float64(3), int64(9), object(3)
memory usage: 55.7+ MB
```

In [13]:

```
train.isnull().sum()
```

Out[13]:

```
id                        0
week                      0
center_id                 0
meal_id                   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
op_area                   0
dtype: int64
```

In [14]:

```
train.nunique()
```

Out[14]:

```
id                      456548
week                       145
center_id                   77
meal_id                     51
checkout_price            1992
base_price                1907
emailer_for_promotion        2
homepage_featured            2
num_orders                1250
category                    14
cuisine                      4
city_code                   51
region_code                  8
center_type                  3
op_area                     30
dtype: int64
```

In [15]:

```
train['category']=train['category'].astype('category')
train['cuisine']=train['cuisine'].astype('category')
train['center_type']=train['center_type'].astype('category')
```

In [16]:

```
test['category']=test['category'].astype('category')
test['cuisine']=test['cuisine'].astype('category')
test['center_type']=test['center_type'].astype('category')
```

In [17]:

```
train_dummies = pd.get_dummies(data=train, columns=['category','cuisine','center_type'],drc
```

In [18]:

```
train_dummies.head()
```

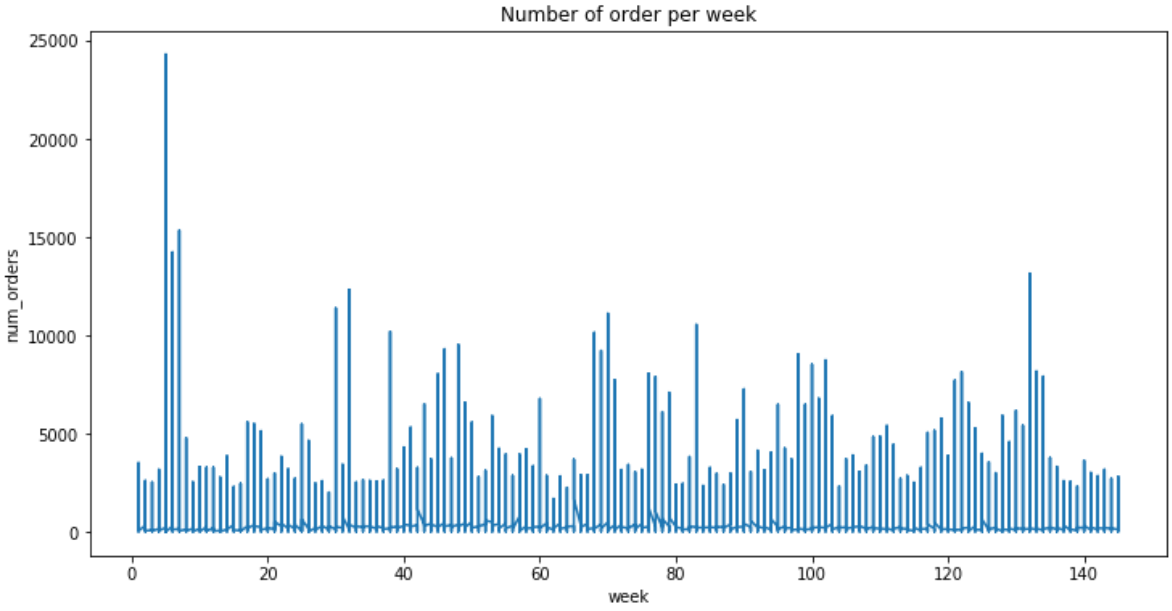
Out[18]:

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

5 rows × 30 columns

In [19]:

```
from matplotlib import pyplot as plt
plt.figure(figsize=(12,6))
plt.title('Number of order per week')
plt.plot(train.week,train.num_orders)
plt.xlabel('week')
plt.ylabel('num_orders')
plt.show()
```



In [20]:

```
train[train.columns[1:]].corr()['num_orders'][:-1]
```

Out[20]:

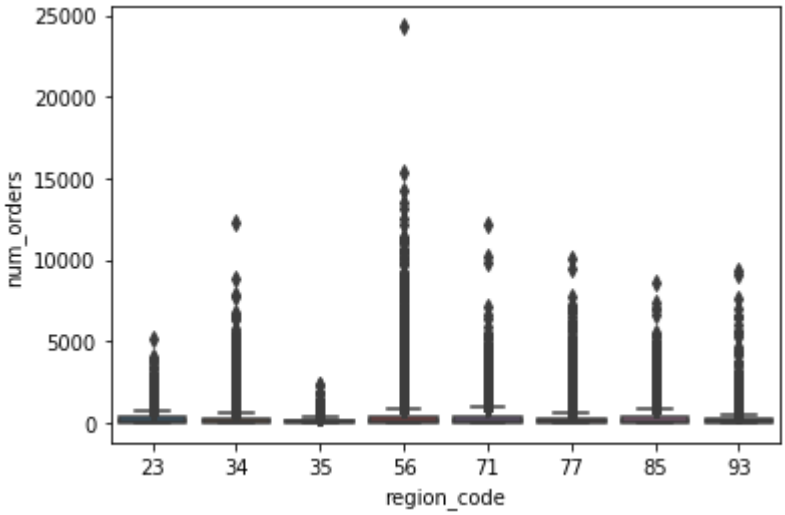
```
week                   -0.017210
center_id              -0.053035
meal_id                 0.010597
checkout_price         -0.282108
base_price             -0.222306
emailer_for_promotion   0.277147
homepage_featured       0.294490
num_orders              1.000000
city_code               0.041596
region_code             0.029744
Name: num_orders, dtype: float64
```

In [21]:

```
sns.boxplot(x="region_code",y="num_orders",data=train)
```

Out[21]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2941d1bdeb8>
```

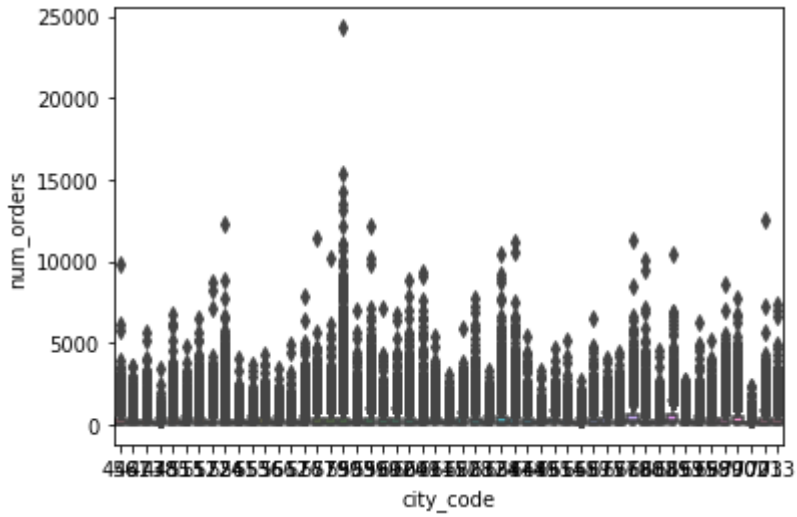


In [22]:

```
sns.boxplot(x="city_code",y="num_orders",data=train)
```

Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2941d0f8a20>
```

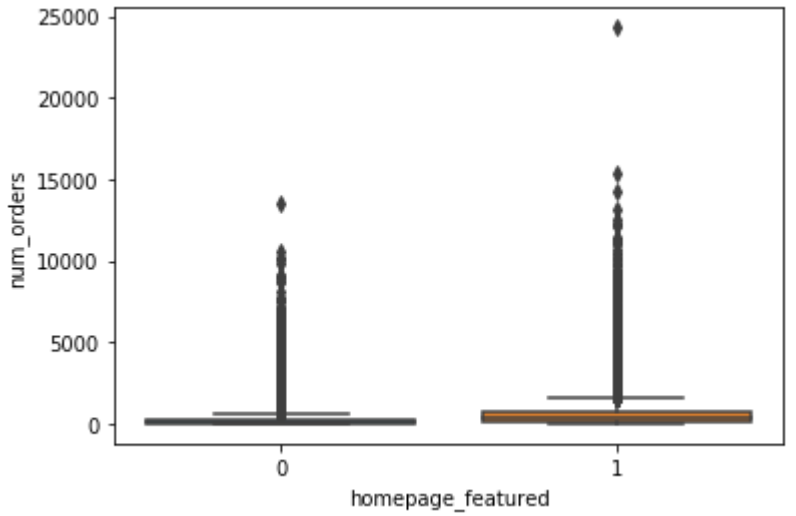


In [23]:

```
sns.boxplot(x="homepage_featured",y="num_orders",data=train)
```

Out[23]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2941e8f3fd0>
```

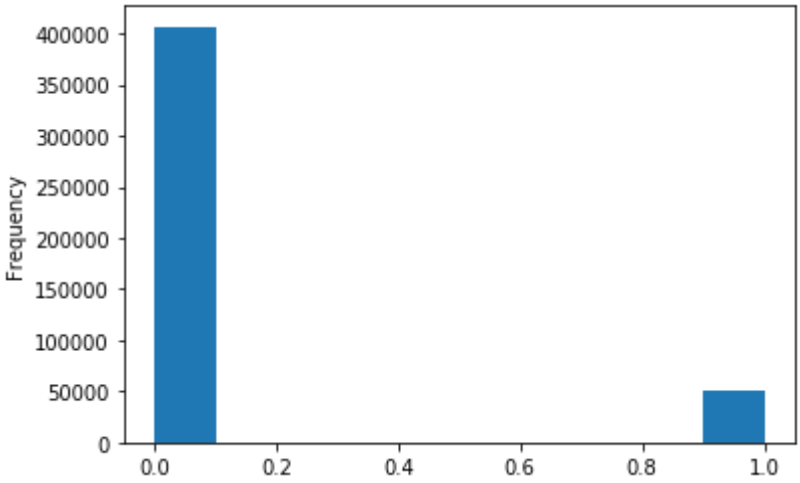


In [24]:

```
train["homepage_featured"].plot.hist()
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2941f12bbe0>
```

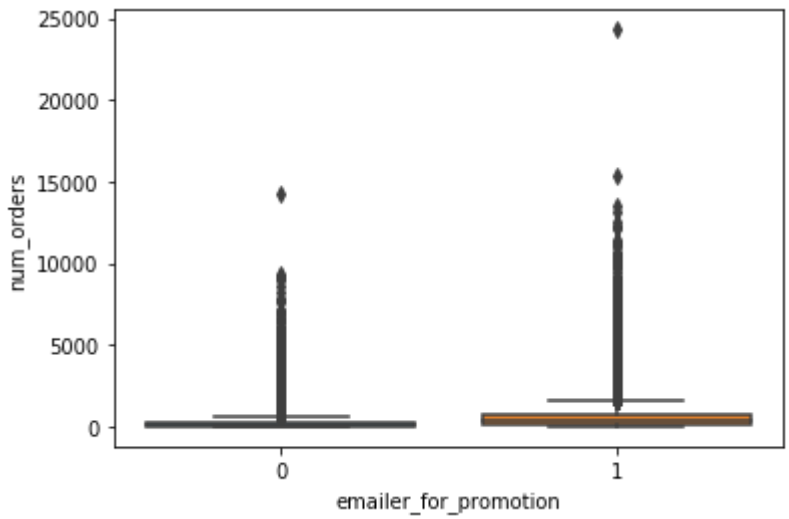


In [25]:

```
sns.boxplot(x="emailer_for_promotion",y="num_orders",data=train)
```

Out[25]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2941f516208>
```

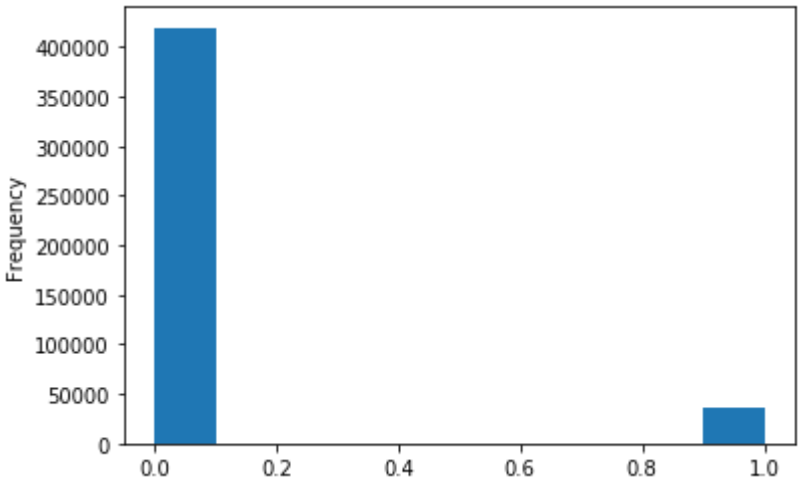


In [26]:

```
train["emailer_for_promotion"].plot.hist()
```

Out[26]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2941bbb1668>
```



In [27]:

```
train_dummies['month'] = train_dummies['week']/4
train_dummies['week_from_yr_start'] = train_dummies['week']/52
train_dummies['quarter'] = train_dummies['week']/13
```

In [28]:

```
train_dummies.drop(['id','week','num_orders','region_code'],inplace=True,axis=1)
```

In [29]:

```
train_dummies.head()
```

Out[29]:

	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	c
0	55	1885	136.83	152.29	0	0	
1	55	1993	136.83	135.83	0	0	
2	55	2539	134.86	135.86	0	0	
3	55	2139	339.50	437.53	0	0	
4	55	2631	243.50	242.50	0	0	

5 rows × 29 columns

In [30]:

```
x=train_dummies
y=train['num_orders']
```

In [31]:

```
y.head()
```

Out[31]:

```
0    177
1    270
2    189
3     54
4     40
Name: num_orders, dtype: int64
```

In [32]:

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0
m = RandomForestRegressor(n_jobs=-1,n_estimators=50)
m.fit(x_train,y_train)
pri = m.predict(x_test)
print("Random forest = ",r2_score(y_test,pri)*100)
```

```
Random forest =  85.6990755273199
```

In [33]:

```
y_test.head()
```

Out[33]:

```
242838    122
314826    216
57041      95
87123     177
154611     80
Name: num_orders, dtype: int64
```

In [34]:

```
test_dummies = pd.get_dummies(data=test, columns=['category','cuisine','center_type'],drop_
```

In [35]:

```
test_dummies['month'] = test_dummies['week']/4
test_dummies['week_from_yr_start'] = test_dummies['week']/52
test_dummies['quarter'] = test_dummies['week']/13
```

In [36]:

```
test_dummies.drop(['id','week','region_code'],inplace=True,axis=1)
test_dummies.head()
```

Out[36]:

	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	c
0	55	1885	158.11	159.11	0	0	
1	55	1993	160.11	159.11	0	0	
2	55	2539	157.14	159.14	0	0	
3	55	2631	162.02	162.02	0	0	
4	55	1248	163.93	163.93	0	0	

5 rows × 29 columns

In [37]:

```
pri = m.predict(test_dummies)
```

In [38]:

```
df1=pd.read_csv("test.csv")
dict={'id':df1['id'],'num_orders':pri.reshape(-1,1)[:,-1]}
df2=pd.DataFrame(dict)
df2.to_csv('Output.csv',index=False)
```