Walmart_store_Sales

August 31, 2020

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
df_sales=pd.read_csv("Walmart_Store_sales.csv")
```

[3]:
```
df_sales.shape
```

[3]:
```
(6435, 8)
```

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

[4]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	\
0	1	05-02-2010	1643690.90	0	42.31	2.572	
1	1	12-02-2010	1641957.44	1	38.51	2.548	
2	1	19-02-2010	1611968.17	0	39.93	2.514	
3	1	26-02-2010	1409727.59	0	46.63	2.561	
4	1	05-03-2010	1554806.68	0	46.50	2.625	

	CPI	Unemployment
0	211.096358	8.106
1	211.242170	8.106
2	211.289143	8.106
3	211.319643	8.106
4	211.350143	8.106

[5]:
```
df_sales.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6435 entries, 0 to 6434
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Store         6435 non-null   int64
 1   Date          6435 non-null   object
 2   Weekly_Sales  6435 non-null   float64
 3   Holiday_Flag  6435 non-null   int64
```

```
 4   Temperature   6435 non-null   float64
 5   Fuel_Price    6435 non-null   float64
 6   CPI           6435 non-null   float64
 7   Unemployment  6435 non-null   float64
dtypes: float64(5), int64(2), object(1)
memory usage: 402.3+ KB
```

[6]:
```
df_sales.isnull().sum()
```

[6]:
```
Store           0
Date            0
Weekly_Sales    0
Holiday_Flag    0
Temperature     0
Fuel_Price      0
CPI             0
Unemployment    0
dtype: int64
```

[7]:
```
df_sales['Date']=pd.to_datetime(df_sales['Date'])
```

[8]:
```
df_sales.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6435 entries, 0 to 6434
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Store         6435 non-null   int64
 1   Date          6435 non-null   datetime64[ns]
 2   Weekly_Sales  6435 non-null   float64
 3   Holiday_Flag  6435 non-null   int64
 4   Temperature   6435 non-null   float64
 5   Fuel_Price    6435 non-null   float64
 6   CPI           6435 non-null   float64
 7   Unemployment  6435 non-null   float64
dtypes: datetime64[ns](1), float64(5), int64(2)
memory usage: 402.3 KB
```

[9]:
```
df_sales.max()
 ↪            ## Store No. 45 has maximum sales.
```

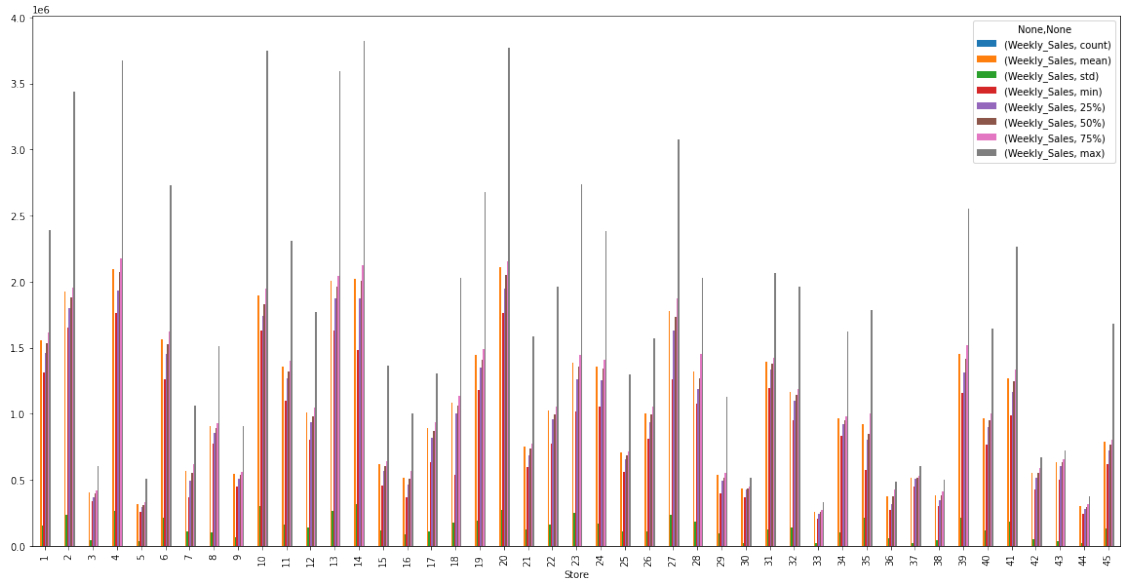
[9]:
```
Store                          45
Date           2012-12-10 00:00:00
Weekly_Sales           3.81869e+06
Holiday_Flag                     1
Temperature                 100.14
```

```
Fuel_Price                   4.468
CPI                        227.233
Unemployment                14.313
dtype: object
```

[10]:
```
df1=df_sales[['Weekly_Sales','Store']].groupby('Store')
```

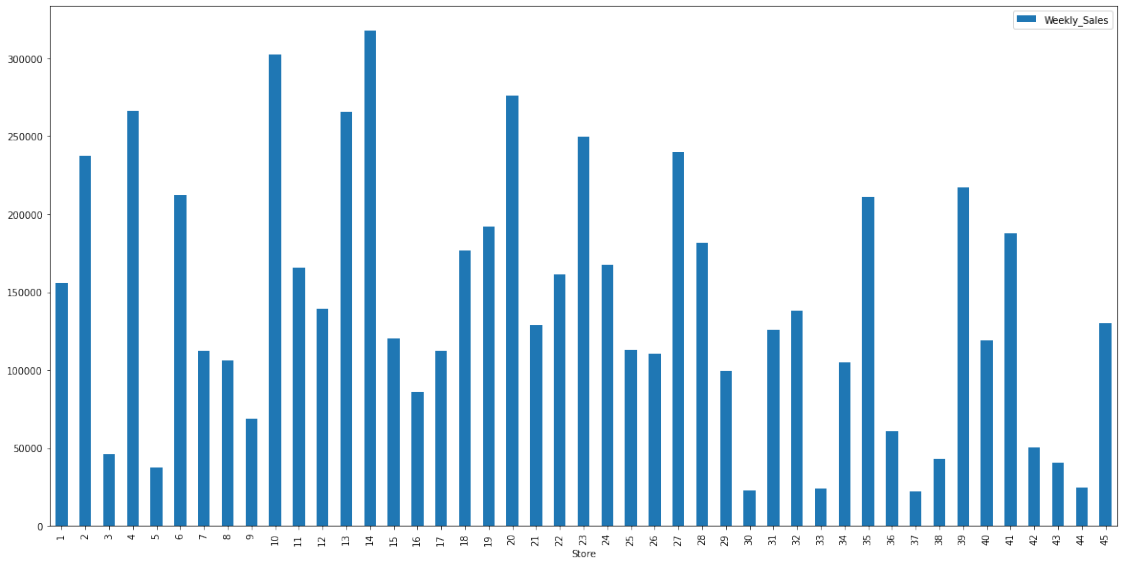
[11]:
```
df1.describe().plot(kind='bar',figsize=(20,10))
```

[11]: <AxesSubplot:xlabel='Store'>



[12]:
```
df1.std().plot(kind='bar',figsize=(20,10))
                    ## store no. 14 has maximum variation of sales.
```

[12]: <AxesSubplot:xlabel='Store'>



[13]:
```
df1['Weekly_Sales'].std().max()
```

[13]: 317569.9494755081

[14]:
```
df2=df1['Weekly_Sales'].std().max()/df1['Weekly_Sales'].mean().max()
df2
```

[14]: 0.1506729774108235

[15]:
```
start_date='2012-07-01'
end_date='2012-09-30'
```

[16]:
```
mask=(df_sales['Date']>start_date) & (df_sales['Date']<=end_date)
```

[17]:
```
df_7=df_sales.loc[mask]
```

[18]:
```
df_7
```

[18]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	\
109	1	2012-09-03	1675431.16	0	58.76	3.669	
122	1	2012-08-06	1697230.96	0	78.30	3.452	
127	1	2012-07-13	1527014.04	0	77.12	3.256	
128	1	2012-07-20	1497954.76	0	80.42	3.311	
129	1	2012-07-27	1439123.71	0	82.66	3.407	
…	…	…	…	…	…	…	
6426	45	2012-08-31	734297.87	0	75.09	3.867	
6427	45	2012-07-09	766512.66	1	75.70	3.911	
6428	45	2012-09-14	702238.27	0	67.87	3.948	

```
6429     45 2012-09-21     723086.20             0        65.32       4.038
6430     45 2012-09-28     713173.95             0        64.88       3.997

             CPI  Unemployment
109   221.059189         7.348
122   221.749484         7.143
127   221.924158         6.908
128   221.932727         6.908
129   221.941295         6.908
…          …             …
6426  191.461281         8.684
6427  191.577676         8.684
6428  191.699850         8.684
6429  191.856704         8.684
6430  192.013558         8.684

[540 rows x 8 columns]
```

[19]:
```
df_7.max()
```

[19]:
```
Store                          45
Date            2012-09-28 00:00:00
Weekly_Sales            2.28354e+06
Holiday_Flag                      1
Temperature                    97.6
Fuel_Price                     4.25
CPI                         226.763
Unemployment                 12.187
dtype: object
```

[20]:
```
df_hol=df_sales[['Weekly_Sales','Holiday_Flag']].groupby('Holiday_Flag')
df_hol
```

[20]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f96d447c650>
```

[21]:
```
df_hol.describe()
```

[21]:
```
             Weekly_Sales                                                \
                    count          mean            std        min
Holiday_Flag
0                  5985.0  1.041256e+06  558957.436147  209986.25
1                   450.0  1.122888e+06  627684.931884  215359.21


                       25%         50%          75%         max
Holiday_Flag
0             551378.3900   956211.20  1414343.530  3818686.45
```

```
1             575865.4825  1018538.04  1555213.175  3004702.33
```

[22]:
```
df_sales.tail()
```

[22]:
```
      Store       Date  Weekly_Sales  Holiday_Flag  Temperature  Fuel_Price  \
6430     45 2012-09-28     713173.95             0        64.88       3.997
6431     45 2012-05-10     733455.07             0        64.89       3.985
6432     45 2012-12-10     734464.36             0        54.47       4.000
6433     45 2012-10-19     718125.53             0        56.47       3.969
6434     45 2012-10-26     760281.43             0        58.85       3.882

             CPI  Unemployment
6430  192.013558         8.684
6431  192.170412         8.667
6432  192.327265         8.667
6433  192.330854         8.667
6434  192.308899         8.667
```

[23]:
```
start_date_a='2010-02-05'
end_date_a='2010-03-31'
```

[24]:
```
mask_a=(df_sales['Date']>start_date_a) & (df_sales['Date']<=end_date_a)
```

[25]:
```
df_1_2010=df_sales.loc[mask_a].max()
df_1_2010
 ↪                          ## Q1-Year2010
```

[25]:
```
Store                          45
Date           2010-03-26 00:00:00
Weekly_Sales             2.41605e+06
Holiday_Flag                     0
Temperature                  97.66
Fuel_Price                     3.2
CPI                        215.145
Unemployment                14.313
dtype: object
```

[26]:
```
start_date_b='2010-04-01'
end_date_b='2010-06-30'
```

[27]:
```
mask_b=(df_sales['Date']>start_date_b) & (df_sales['Date']<=end_date_b)
```

[28]:
```
df_2_2010=df_sales.loc[mask_b].max()
df_2_2010
 ↪                  ## Q2-Year2010
```

[28]:
```
Store                          45
Date            2010-06-25 00:00:00
Weekly_Sales            2.62347e+06
Holiday_Flag                      0
Temperature                   93.66
Fuel_Price                    3.145
CPI                         215.507
Unemployment                14.313
dtype: object
```

[29]:
```
start_date_c='2010-07-01'
end_date_c='2010-09-30'
```

[30]:
```
mask_c=(df_sales['Date']>start_date_c) & (df_sales['Date']<=end_date_c)
```

[31]:
```
df_3_2010=df_sales.loc[mask_c].max()
df_3_2010
 ↪                                  ## Q3-Year2010
```

[31]:
```
Store                          45
Date            2010-09-24 00:00:00
Weekly_Sales            2.37012e+06
Holiday_Flag                      0
Temperature                  100.14
Fuel_Price                    3.159
CPI                          215.29
Unemployment                14.313
dtype: object
```

[32]:
```
start_date_d='2010-10-01'
end_date_d='2010-12-31'
```

[33]:
```
mask_d=(df_sales['Date']>start_date_d) & (df_sales['Date']<=end_date_d)
```

[34]:
```
df_4_2010=df_sales.loc[mask_d].max()
df_4_2010
 ↪                                ## Q4-Year2010
```

[34]:
```
Store                          45
Date            2010-12-31 00:00:00
Weekly_Sales            3.81869e+06
Holiday_Flag                      1
Temperature                   93.52
Fuel_Price                    3.336
CPI                         215.556
Unemployment                14.313
dtype: object
```

[39]:
```
start_date_e='2011-01-01'
end_date_e='2011-03-31'
```

[40]:
```
mask_e=(df_sales['Date']>start_date_e) & (df_sales['Date']<=end_date_e)
```

[41]:
```
df_1_2011=df_sales.loc[mask_e].max()
df_1_2011
```

[41]:
```
Store                          45
Date           2011-03-25 00:00:00
Weekly_Sales             2.3165e+06
Holiday_Flag                    0
Temperature                  99.2
Fuel_Price                  4.069
CPI                        222.398
Unemployment                14.021
dtype: object
```

[42]:
```
df_sales['Store'].value_counts()
```

[42]:
```
43    143
41    143
33    143
29    143
25    143
21    143
17    143
13    143
9     143
5     143
1     143
44    143
40    143
36    143
32    143
28    143
24    143
20    143
16    143
12    143
8     143
37    143
45    143
39    143
2     143
35    143
31    143
```

```
27    143
23    143
19    143
15    143
11    143
7     143
3     143
42    143
38    143
34    143
30    143
26    143
22    143
18    143
14    143
10    143
6     143
4     143
Name: Store, dtype: int64
```

[50]:
```
df_store=df_sales[df_sales['Store']=='1']
df_store
```

[50]:
```
Empty DataFrame
Columns: [Store, Date, Weekly_Sales, Holiday_Flag, Temperature, Fuel_Price, CPI,
Unemployment]
Index: []
```

[57]:
```
df_new=df_sales.loc[df_sales['Store'] == 1]
df_new
```

[57]:
```
     Store       Date  Weekly_Sales  Holiday_Flag  Temperature  Fuel_Price  \
0        1 2010-05-02    1643690.90             0        42.31       2.572
1        1 2010-12-02    1641957.44             1        38.51       2.548
2        1 2010-02-19    1611968.17             0        39.93       2.514
3        1 2010-02-26    1409727.59             0        46.63       2.561
4        1 2010-05-03    1554806.68             0        46.50       2.625
..     …        …            …             …          …          …
138      1 2012-09-28    1437059.26             0        76.08       3.666
139      1 2012-05-10    1670785.97             0        68.55       3.617
140      1 2012-12-10    1573072.81             0        62.99       3.601
141      1 2012-10-19    1508068.77             0        67.97       3.594
142      1 2012-10-26    1493659.74             0        69.16       3.506

            CPI  Unemployment
0    211.096358         8.106
1    211.242170         8.106
```

```
2    211.289143         8.106
3    211.319643         8.106
4    211.350143         8.106
..          …           …
138  222.981658         6.908
139  223.181477         6.573
140  223.381296         6.573
141  223.425723         6.573
142  223.444251         6.573

[143 rows x 8 columns]
```

[58]:
```
df_new.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 143 entries, 0 to 142
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Store         143 non-null    int64
 1   Date          143 non-null    datetime64[ns]
 2   Weekly_Sales  143 non-null    float64
 3   Holiday_Flag  143 non-null    int64
 4   Temperature   143 non-null    float64
 5   Fuel_Price    143 non-null    float64
 6   CPI           143 non-null    float64
 7   Unemployment  143 non-null    float64
dtypes: datetime64[ns](1), float64(5), int64(2)
memory usage: 10.1 KB
```

[62]:
```
df_new_s=pd.get_dummies(df_new['Date'])
df_new_s
```

[62]:
```
     2010-01-10  2010-02-04  2010-02-07  2010-02-19  2010-02-26  2010-03-09  \
0             0           0           0           0           0           0
1             0           0           0           0           0           0
2             0           0           0           1           0           0
3             0           0           0           0           1           0
4             0           0           0           0           0           0
..          …           …           …           …           …           …
138           0           0           0           0           0           0
139           0           0           0           0           0           0
140           0           0           0           0           0           0
141           0           0           0           0           0           0
142           0           0           0           0           0           0

     2010-03-12  2010-03-19  2010-03-26  2010-04-06  …  2012-09-03  \
```

```
0              0           0           0           0  …           0
1              0           0           0           0  …           0
2              0           0           0           0  …           0
3              0           0           0           0  …           0
4              0           0           0           0  …           0
..           …           …           …         …  …         …
138            0           0           0           0  …           0
139            0           0           0           0  …           0
140            0           0           0           0  …           0
141            0           0           0           0  …           0
142            0           0           0           0  …           0

     2012-09-14  2012-09-21  2012-09-28  2012-10-02  2012-10-08  2012-10-19  \
0             0           0           0           0           0           0
1             0           0           0           0           0           0
2             0           0           0           0           0           0
3             0           0           0           0           0           0
4             0           0           0           0           0           0
..          …           …           …         …         …         …
138           0           0           1           0           0           0
139           0           0           0           0           0           0
140           0           0           0           0           0           0
141           0           0           0           0           0           1
142           0           0           0           0           0           0

     2012-10-26  2012-11-05  2012-12-10
0             0           0           0
1             0           0           0
2             0           0           0
3             0           0           0
4             0           0           0
..          …           …         …
138           0           0           0
139           0           0           0
140           0           0           1
141           0           0           0
142           1           0           0

[143 rows x 143 columns]
```

[63]:
```
df_walmart=pd.concat([df_new_s,df_new],axis=1)
df_walmart
```

[63]:
```
     2010-01-10 00:00:00  2010-02-04 00:00:00  2010-02-07 00:00:00  \
0                      0                    0                    0
1                      0                    0                    0
2                      0                    0                    0
```

```
3                      0                      0                      0
4                      0                      0                      0
..                   …                      …                      …
138                    0                      0                      0
139                    0                      0                      0
140                    0                      0                      0
141                    0                      0                      0
142                    0                      0                      0

     2010-02-19 00:00:00  2010-02-26 00:00:00  2010-03-09 00:00:00  \
0                      0                      0                      0
1                      0                      0                      0
2                      1                      0                      0
3                      0                      1                      0
4                      0                      0                      0
..                   …                      …                      …
138                    0                      0                      0
139                    0                      0                      0
140                    0                      0                      0
141                    0                      0                      0
142                    0                      0                      0

     2010-03-12 00:00:00  2010-03-19 00:00:00  2010-03-26 00:00:00  \
0                      0                      0                      0
1                      0                      0                      0
2                      0                      0                      0
3                      0                      0                      0
4                      0                      0                      0
..                   …                      …                      …
138                    0                      0                      0
139                    0                      0                      0
140                    0                      0                      0
141                    0                      0                      0
142                    0                      0                      0

     2010-04-06 00:00:00  …  2012-11-05 00:00:00  2012-12-10 00:00:00  \
0                      0  …                      0                      0
1                      0  …                      0                      0
2                      0  …                      0                      0
3                      0  …                      0                      0
4                      0  …                      0                      0
..                   …  …                      …                      …
138                    0  …                      0                      0
139                    0  …                      0                      0
140                    0  …                      0                      1
141                    0  …                      0                      0
142                    0  …                      0                      0
```

```
     Store       Date  Weekly_Sales  Holiday_Flag  Temperature  Fuel_Price  \
0        1 2010-05-02    1643690.90             0        42.31       2.572
1        1 2010-12-02    1641957.44             1        38.51       2.548
2        1 2010-02-19    1611968.17             0        39.93       2.514
3        1 2010-02-26    1409727.59             0        46.63       2.561
4        1 2010-05-03    1554806.68             0        46.50       2.625
..     …        …           …            …          …         …
138      1 2012-09-28    1437059.26             0        76.08       3.666
139      1 2012-05-10    1670785.97             0        68.55       3.617
140      1 2012-12-10    1573072.81             0        62.99       3.601
141      1 2012-10-19    1508068.77             0        67.97       3.594
142      1 2012-10-26    1493659.74             0        69.16       3.506

            CPI  Unemployment
0    211.096358         8.106
1    211.242170         8.106
2    211.289143         8.106
3    211.319643         8.106
4    211.350143         8.106
..          …         …
138  222.981658         6.908
139  223.181477         6.573
140  223.381296         6.573
141  223.425723         6.573
142  223.444251         6.573

[143 rows x 151 columns]
```

[64]:
```
df_wal_new=df_walmart.drop('Date',axis=1)
df_wal_new
```

[64]:
```
     2010-01-10 00:00:00  2010-02-04 00:00:00  2010-02-07 00:00:00  \
0                      0                    0                    0
1                      0                    0                    0
2                      0                    0                    0
3                      0                    0                    0
4                      0                    0                    0
..                   …                    …                    …
138                    0                    0                    0
139                    0                    0                    0
140                    0                    0                    0
141                    0                    0                    0
142                    0                    0                    0

     2010-02-19 00:00:00  2010-02-26 00:00:00  2010-03-09 00:00:00  \
0                      0                    0                    0
```

```
1                      0                    0                    0
2                      1                    0                    0
3                      0                    1                    0
4                      0                    0                    0
..                   …                    …                    …
138                    0                    0                    0
139                    0                    0                    0
140                    0                    0                    0
141                    0                    0                    0
142                    0                    0                    0

     2010-03-12 00:00:00  2010-03-19 00:00:00  2010-03-26 00:00:00  \
0                      0                    0                    0
1                      0                    0                    0
2                      0                    0                    0
3                      0                    0                    0
4                      0                    0                    0
..                   …                    …                    …
138                    0                    0                    0
139                    0                    0                    0
140                    0                    0                    0
141                    0                    0                    0
142                    0                    0                    0

     2010-04-06 00:00:00  …  2012-10-26 00:00:00  2012-11-05 00:00:00  \
0                      0  …                    0                    0
1                      0  …                    0                    0
2                      0  …                    0                    0
3                      0  …                    0                    0
4                      0  …                    0                    0
..                   …  …                    …                    …
138                    0  …                    0                    0
139                    0  …                    0                    0
140                    0  …                    0                    0
141                    0  …                    0                    0
142                    0  …                    1                    0

     2012-12-10 00:00:00  Store  Weekly_Sales  Holiday_Flag  Temperature  \
0                      0      1    1643690.90             0        42.31
1                      0      1    1641957.44             1        38.51
2                      0      1    1611968.17             0        39.93
3                      0      1    1409727.59             0        46.63
4                      0      1    1554806.68             0        46.50
..                   …    …           …          …           …
138                    0      1    1437059.26             0        76.08
139                    0      1    1670785.97             0        68.55
140                    1      1    1573072.81             0        62.99
```

```
141                   0      1    1508068.77             0        67.97
142                   0      1    1493659.74             0        69.16

     Fuel_Price         CPI  Unemployment
0         2.572  211.096358         8.106
1         2.548  211.242170         8.106
2         2.514  211.289143         8.106
3         2.561  211.319643         8.106
4         2.625  211.350143         8.106
..          …         …           …
138       3.666  222.981658         6.908
139       3.617  223.181477         6.573
140       3.601  223.381296         6.573
141       3.594  223.425723         6.573
142       3.506  223.444251         6.573

[143 rows x 150 columns]
```

[65]:
```
df_wal_new.isnull().sum()
```

[65]:
```
2010-01-10 00:00:00    0
2010-02-04 00:00:00    0
2010-02-07 00:00:00    0
2010-02-19 00:00:00    0
2010-02-26 00:00:00    0
                      ..
Holiday_Flag           0
Temperature            0
Fuel_Price             0
CPI                    0
Unemployment           0
Length: 150, dtype: int64
```

[66]:
```
X=df_wal_new.drop('Weekly_Sales',axis=1)
```

[67]:
```
X.shape
```

[67]:
```
(143, 149)
```

[68]:
```
y=df_wal_new['Weekly_Sales']
```

[69]:
```
y.shape
```

[69]:
```
(143,)
```

[75]:
```
from sklearn.model_selection import train_test_split
```

[76]:
```
from sklearn.linear_model import LinearRegression
```

[77]:
```
lr=LinearRegression()
```

[78]:
```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.
↪25,random_state=42)
```

[79]:
```
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(107, 149)
(36, 149)
(107,)
(36,)
```

[80]:
```
lr.fit(X_train,y_train)
```

[80]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

[81]:
```
lr.coef_
```

[81]:
```
array([-4.55820995e+04,  7.52658688e+04,  1.25746667e+04,  5.29188159e+04,
       -1.34950396e+05,  6.08502570e+04, -1.69138445e+03, -5.55558455e+04,
       -1.30102562e+05,  1.35725801e+05,  1.81898940e-12, -1.16415322e-10,
       -1.45519152e-11,  8.99977689e+04,  4.36557457e-11,  2.20627332e+04,
        2.56455310e+03, -8.89286275e+04,  5.82076609e-11, -5.82076609e-11,
       -5.82076609e-11, -4.90398075e+04,  1.07475681e+05, -2.52004718e+04,
       -8.88771215e+04, -1.05981976e+05, -8.77967924e+03,  6.54836185e-11,
        4.48936586e+04, -2.17356678e+04, -4.36557457e-11,  6.58365453e+04,
       -4.69557138e+04, -1.28701930e+05,  1.30967237e-10,  1.25666919e+05,
        4.80213203e-10, -1.59518143e+05, -1.21314511e+05,  5.82076609e-11,
       -6.25424621e+04,  3.65483499e+05,  6.03696121e+03, -8.14575509e+04,
       -3.36502404e+04,  5.09317033e-11,  8.44121956e+05, -2.59202503e+05,
       -5.40306683e+04,  0.00000000e+00, -1.94459727e+05, -2.40269755e+05,
       -8.73114914e-11, -2.91038305e-11,  9.94013806e+03,  1.44310651e+05,
       -5.82076609e-11, -4.36557457e-11,  3.93392051e+04,  1.90363213e+04,
        3.31005496e+04, -2.91038305e-11,  1.37769825e+05,  3.44734972e+04,
        4.21970239e+04, -2.91038305e-11, -1.45519152e-10,  8.73114914e-11,
       -1.07910008e+05, -4.60469178e+04, -6.54836185e-11,  4.19877036e+04,
       -5.74842033e+04, -1.12006064e+05,  1.05986129e+05, -2.15910088e+04,
       -8.67307137e+04, -1.29523270e+05,  8.28770529e+04,  5.31634132e+04,
        5.64293895e+04, -1.39564211e+04, -3.88278407e+04,  2.13897413e+05,
        1.69135541e+04, -1.26988207e+05, -1.03873116e+05,  9.13281555e+04,
       -2.69038956e+04, -3.67861816e+04, -8.83866148e+04, -1.18757897e+04,
       -3.70419874e+03,  4.44541625e+04,  1.45519152e-11,  2.91038305e-11,
```

```
 5.21220503e+04, -2.91038305e-11, -2.91038305e-11, -1.61959771e+05,
 7.92804680e+04, -1.36645798e+05, -1.89815414e+05, -2.65004721e+05,
 1.13884428e+05,  2.12894019e+05, -4.21350914e+04,  5.62338170e+04,
 9.63742051e+04,  0.00000000e+00, -5.50298464e+04,  8.65326584e+04,
 0.00000000e+00,  5.36153749e+04, -5.09991377e+04, -1.02242307e+05,
 8.08080105e+04,  0.00000000e+00,  8.13941709e+03,  0.00000000e+00,
 0.00000000e+00,  2.26795960e+05,  8.93618896e+04,  0.00000000e+00,
 0.00000000e+00,  4.46452831e+04, -2.62644639e+04, -4.84901937e+04,
-1.03176414e+05,  1.53208603e+05,  0.00000000e+00, -6.16945424e+04,
 3.20006611e+04,  9.50810299e+04, -4.63474704e+04, -6.95686514e+04,
-1.24998856e+05,  1.29210092e+05,  0.00000000e+00, -8.37157015e+04,
-9.48127596e+04,  5.51493964e+04, -2.96774510e+04,  0.00000000e+00,
 7.35793082e+04, -2.21631496e+03,  8.27871502e+03,  2.91288068e+03,
-3.29988037e+04])
```

[82]:
```
lr.intercept_
```

[82]:
```
1278762.3604722081
```

[83]:
```
y_pred=lr.predict(X_test)
```

[84]:
```
y_pred
```

[84]:
```
array([1552349.7175653 , 1472464.02929566, 1479206.96620687,
       1569361.48587466, 1541299.84762051, 1507560.38111026,
       1537773.29943651, 1542994.11215553, 1518979.7141412 ,
       1480883.47571113, 1569168.18039863, 1471255.15995856,
       1622643.93259301, 1539086.98985624, 1593802.87651338,
       1525264.79154786, 1467747.95723933, 1499000.84808679,
       1528338.30427658, 1482400.04807416, 1513524.90374581,
       1563140.16206358, 1568079.26011634, 1564184.83789038,
       1531257.42370184, 1549206.68393886, 1511592.49158027,
       1545584.78787479, 1466462.6934815 , 1543101.29233741,
       1558273.31854838, 1577625.53645179, 1510731.96538203,
       1479985.40096819, 1510496.79337154, 1544264.45864923])
```

[85]:
```
from sklearn.metrics import r2_score,mean_squared_error
```

[86]:
```
print(r2_score(y_pred=y_pred,y_true=y_test))
```

```
0.013425867152094706
```

[87]:
```
print(mean_squared_error(y_test,y_pred))
```

```
33815377831.906292
```

[107]:
```
from sklearn.linear_model import Ridge
```

[161]:
```
rdg=Ridge(alpha=7)
```

[162]:
```
rdg.fit(X_train,y_train)
```

[162]:
```
Ridge(alpha=7, copy_X=True, fit_intercept=True, max_iter=None, normalize=False,
      random_state=None, solver='auto', tol=0.001)
```

[163]:
```
y_pred_rdg=rdg.predict(X_test)
```

[164]:
```
r2_score(y_test,y_pred_rdg)
```

[164]:
```
0.03463779415346946
```

[165]:
```
mean_squared_error(y_test,y_pred_rdg)
```

[165]:
```
33088327220.89571
```

[166]:
```
from sklearn.linear_model import Lasso
```

[203]:
```
ls=Lasso(alpha=9)
```

[204]:
```
ls.fit(X_train,y_train)
```

```
/usr/local/lib/python3.7/site-
packages/sklearn/linear_model/_coordinate_descent.py:476: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations.
Duality gap: 6938734680.090748, tolerance: 212221212.13193247
  positive)
```

[204]:
```
Lasso(alpha=9, copy_X=True, fit_intercept=True, max_iter=1000, normalize=False,
      positive=False, precompute=False, random_state=None, selection='cyclic',
      tol=0.0001, warm_start=False)
```

[205]:
```
y_pred_ls=ls.predict(X_test)
```

[211]:
```
r2_score(y_test,y_pred_ls)
```

[211]:
```
0.002013180421515548
```

[213]:
```
from statsmodels.api import OLS
```

[214]:
```
ols=OLS(y_train,X_train).fit()
```

[215]:
```
ols.summary()
```

```
/usr/local/lib/python3.7/site-
packages/statsmodels/regression/linear_model.py:1698: RuntimeWarning: divide by
zero encountered in true_divide
```

```
  return 1 - (np.divide(self.nobs - self.k_constant, self.df_resid)
/usr/local/lib/python3.7/site-
packages/statsmodels/regression/linear_model.py:1699: RuntimeWarning: invalid
value encountered in double_scalars
  * (1 - self.rsquared))
/usr/local/lib/python3.7/site-
packages/statsmodels/regression/linear_model.py:1620: RuntimeWarning: divide by
zero encountered in double_scalars
  return np.dot(wresid, wresid) / self.df_resid
/usr/local/lib/python3.7/site-packages/statsmodels/base/model.py:1446:
RuntimeWarning: invalid value encountered in multiply
  cov_p = self.normalized_cov_params * scale
/usr/local/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py:903:
RuntimeWarning: invalid value encountered in greater
  return (a < x) & (x < b)
/usr/local/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py:903:
RuntimeWarning: invalid value encountered in less
  return (a < x) & (x < b)
/usr/local/lib/python3.7/site-
packages/scipy/stats/_distn_infrastructure.py:1912: RuntimeWarning: invalid
value encountered in less_equal
  cond2 = cond0 & (x <= _a)
```

[215]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:           Weekly_Sales   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                    nan
Method:                 Least Squares   F-statistic:                       nan
Date:                Mon, 31 Aug 2020   Prob (F-statistic):                nan
Time:                        11:53:34   Log-Likelihood:                1739.6
No. Observations:                 107   AIC:                            -3265.
Df Residuals:                       0   BIC:                            -2979.
Df Model:                         106
Covariance Type:            nonrobust
==============================================================================
=======
                           coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-------
2010-01-10 00:00:00  -4.55e+04        inf         -0        nan         nan
nan
2010-02-04 00:00:00  8.426e+04        inf          0        nan         nan
nan
2010-02-07 00:00:00  1.692e+04        inf          0        nan         nan
```

```
nan
2010-02-19 00:00:00  4.779e+04        inf          0        nan         nan
nan
2010-02-26 00:00:00 -1.401e+05        inf         -0        nan         nan
nan
2010-03-09 00:00:00  6.167e+04        inf          0        nan         nan
nan
2010-03-12 00:00:00  3732.7386        inf          0        nan         nan
nan
2010-03-19 00:00:00 -5.774e+04        inf         -0        nan         nan
nan
2010-03-26 00:00:00 -1.307e+05        inf         -0        nan         nan
nan
2010-04-06 00:00:00  1.404e+05        inf          0        nan         nan
nan
2010-04-16 00:00:00  1.756e-11        inf          0        nan         nan
nan
2010-04-23 00:00:00  1.088e-10        inf          0        nan         nan
nan
2010-04-30 00:00:00  1.001e-11        inf          0        nan         nan
nan
2010-05-02 00:00:00  8.677e+04        inf          0        nan         nan
nan
2010-05-03 00:00:00  3.486e-11        inf          0        nan         nan
nan
2010-05-11 00:00:00  2.418e+04        inf          0        nan         nan
nan
2010-05-14 00:00:00  1.529e+04        inf          0        nan         nan
nan
2010-05-21 00:00:00 -7.839e+04        inf         -0        nan         nan
nan
2010-05-28 00:00:00 -1.558e-10        inf         -0        nan         nan
nan
2010-06-08 00:00:00  1.403e-11        inf          0        nan         nan
nan
2010-06-18 00:00:00  1.677e-10        inf          0        nan         nan
nan
2010-06-25 00:00:00 -4.675e+04        inf         -0        nan         nan
nan
2010-07-05 00:00:00  1.201e+05        inf          0        nan         nan
nan
2010-07-16 00:00:00  -2.15e+04        inf         -0        nan         nan
nan
2010-07-23 00:00:00 -8.619e+04        inf         -0        nan         nan
nan
2010-07-30 00:00:00 -1.031e+05        inf         -0        nan         nan
nan
```

```
2010-08-10 00:00:00 -7453.4837        inf         -0        nan         nan
nan
2010-08-13 00:00:00  6.685e-11        inf          0        nan         nan
nan
2010-08-20 00:00:00  4.654e+04        inf          0        nan         nan
nan
2010-08-27 00:00:00  -2.07e+04        inf         -0        nan         nan
nan
2010-09-04 00:00:00 -1.099e-11        inf         -0        nan         nan
nan
2010-09-07 00:00:00  7.023e+04        inf          0        nan         nan
nan
2010-09-17 00:00:00  -4.61e+04        inf         -0        nan         nan
nan
2010-09-24 00:00:00 -1.272e+05        inf         -0        nan         nan
nan
2010-10-09 00:00:00 -1.255e-11        inf         -0        nan         nan
nan
2010-10-12 00:00:00   1.35e+05        inf          0        nan         nan
nan
2010-10-15 00:00:00 -7.413e-11        inf         -0        nan         nan
nan
2010-10-22 00:00:00 -1.575e+05        inf         -0        nan         nan
nan
2010-10-29 00:00:00 -1.197e+05        inf         -0        nan         nan
nan
2010-11-06 00:00:00  5.695e-11        inf          0        nan         nan
nan
2010-11-19 00:00:00 -5.744e+04        inf         -0        nan         nan
nan
2010-11-26 00:00:00  3.705e+05        inf          0        nan         nan
nan
2010-12-02 00:00:00  3647.3982        inf          0        nan         nan
nan
2010-12-03 00:00:00 -8.598e+04        inf         -0        nan         nan
nan
2010-12-11 00:00:00 -3.105e+04        inf         -0        nan         nan
nan
2010-12-17 00:00:00  1.026e-10        inf          0        nan         nan
nan
2010-12-24 00:00:00   8.54e+05        inf          0        nan         nan
nan
2010-12-31 00:00:00 -2.461e+05        inf         -0        nan         nan
nan
2011-01-04 00:00:00 -4.317e+04        inf         -0        nan         nan
nan
2011-01-07 00:00:00  3.342e-11        inf          0        nan         nan
```

```
nan
2011-01-14 00:00:00 -1.774e+05        inf         -0        nan         nan
nan
2011-01-21 00:00:00 -2.255e+05        inf         -0        nan         nan
nan
2011-01-28 00:00:00  4.017e-11        inf          0        nan         nan
nan
2011-02-09 00:00:00 -2.135e-12        inf         -0        nan         nan
nan
2011-02-12 00:00:00 -1.099e+04        inf         -0        nan         nan
nan
2011-02-18 00:00:00   1.51e+05        inf          0        nan         nan
nan
2011-02-25 00:00:00 -5.001e-11        inf         -0        nan         nan
nan
2011-03-06 00:00:00 -1.843e-12        inf         -0        nan         nan
nan
2011-03-18 00:00:00   4.93e+04        inf          0        nan         nan
nan
2011-03-25 00:00:00  2.661e+04        inf          0        nan         nan
nan
2011-04-02 00:00:00  4.388e+04        inf          0        nan         nan
nan
2011-04-03 00:00:00  5.867e-11        inf          0        nan         nan
nan
2011-04-11 00:00:00  1.239e+05        inf          0        nan         nan
nan
2011-04-15 00:00:00  4.625e+04        inf          0        nan         nan
nan
2011-04-22 00:00:00  5.429e+04        inf          0        nan         nan
nan
2011-04-29 00:00:00  6.766e-11        inf          0        nan         nan
nan
2011-05-08 00:00:00 -8.871e-11        inf         -0        nan         nan
nan
2011-05-13 00:00:00   -3.6e-11        inf         -0        nan         nan
nan
2011-05-20 00:00:00 -9.438e+04        inf         -0        nan         nan
nan
2011-05-27 00:00:00  -3.52e+04        inf         -0        nan         nan
nan
2011-06-05 00:00:00  7.709e-11        inf          0        nan         nan
nan
2011-06-17 00:00:00  5.111e+04        inf          0        nan         nan
nan
2011-06-24 00:00:00  -4.94e+04        inf         -0        nan         nan
nan
```

```
2011-07-01 00:00:00 -9.646e+04        inf         -0        nan         nan
nan
2011-07-10 00:00:00  9.325e+04        inf          0        nan         nan
nan
2011-07-15 00:00:00 -2.542e+04        inf         -0        nan         nan
nan
2011-07-22 00:00:00 -8.891e+04        inf         -0        nan         nan
nan
2011-07-29 00:00:00 -1.315e+05        inf         -0        nan         nan
nan
2011-08-04 00:00:00  9.358e+04        inf          0        nan         nan
nan
2011-08-07 00:00:00  4.806e+04        inf          0        nan         nan
nan
2011-08-19 00:00:00  5.045e+04        inf          0        nan         nan
nan
2011-08-26 00:00:00 -2.067e+04        inf         -0        nan         nan
nan
2011-09-09 00:00:00 -4.261e+04        inf         -0        nan         nan
nan
2011-09-12 00:00:00  1.921e+05        inf          0        nan         nan
nan
2011-09-16 00:00:00  9716.5740        inf          0        nan         nan
nan
2011-09-23 00:00:00 -1.366e+05        inf         -0        nan         nan
nan
2011-09-30 00:00:00  -1.18e+05        inf         -0        nan         nan
nan
2011-10-06 00:00:00  1.009e+05        inf          0        nan         nan
nan
2011-10-14 00:00:00 -4.172e+04        inf         -0        nan         nan
nan
2011-10-21 00:00:00 -4.972e+04        inf         -0        nan         nan
nan
2011-10-28 00:00:00 -1.021e+05        inf         -0        nan         nan
nan
2011-11-02 00:00:00   103.1875        inf          0        nan         nan
nan
2011-11-03 00:00:00  8027.2258        inf          0        nan         nan
nan
2011-11-11 00:00:00  2.849e+04        inf          0        nan         nan
nan
2011-11-18 00:00:00 -3.741e-11        inf         -0        nan         nan
nan
2011-11-25 00:00:00 -3.048e-13        inf         -0        nan         nan
nan
2011-12-08 00:00:00  4.814e+04        inf          0        nan         nan
```

```
nan
2011-12-16 00:00:00 -2.762e-11        inf         -0        nan         nan
nan
2011-12-23 00:00:00  2.727e-11        inf          0        nan         nan
nan
2011-12-30 00:00:00 -1.857e+05        inf         -0        nan         nan
nan
2012-01-06 00:00:00   7.14e+04        inf          0        nan         nan
nan
2012-01-13 00:00:00 -1.438e+05        inf         -0        nan         nan
nan
2012-01-20 00:00:00  -1.98e+05        inf         -0        nan         nan
nan
2012-01-27 00:00:00 -2.732e+05        inf         -0        nan         nan
nan
2012-02-03 00:00:00  1.083e+05        inf          0        nan         nan
nan
2012-02-17 00:00:00  2.088e+05        inf          0        nan         nan
nan
2012-02-24 00:00:00  -4.78e+04        inf         -0        nan         nan
nan
2012-03-02 00:00:00  4.881e+04        inf          0        nan         nan
nan
2012-03-08 00:00:00  9.271e+04        inf          0        nan         nan
nan
2012-03-16 00:00:00          0        nan        nan        nan         nan
nan
2012-03-23 00:00:00 -6.005e+04        inf         -0        nan         nan
nan
2012-03-30 00:00:00  8.217e+04        inf          0        nan         nan
nan
2012-04-05 00:00:00          0        nan        nan        nan         nan
nan
2012-04-13 00:00:00  5.631e+04        inf          0        nan         nan
nan
2012-04-20 00:00:00 -4.858e+04        inf         -0        nan         nan
nan
2012-04-27 00:00:00 -1.015e+05        inf         -0        nan         nan
nan
2012-05-10 00:00:00  8.889e+04        inf          0        nan         nan
nan
2012-05-18 00:00:00          0        nan        nan        nan         nan
nan
2012-05-25 00:00:00  1635.2875        inf          0        nan         nan
nan
2012-06-01 00:00:00          0        nan        nan        nan         nan
nan
```

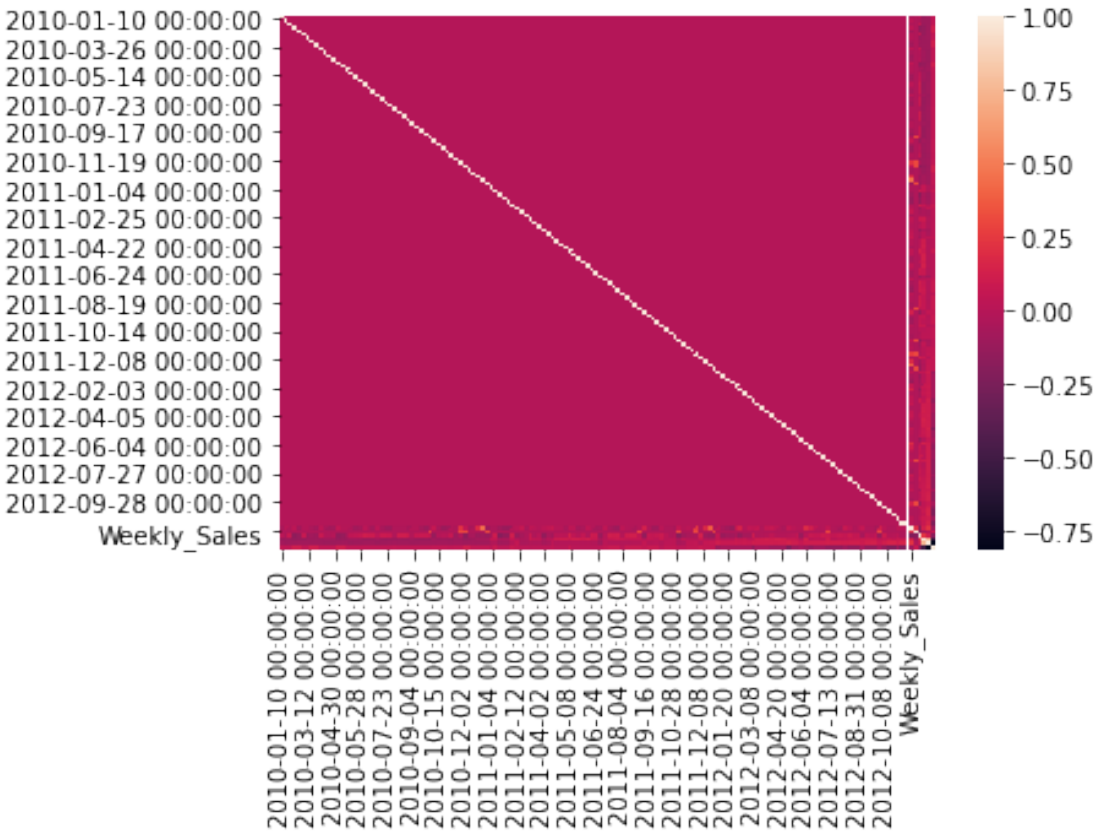
```
2012-06-04 00:00:00          0        nan        nan        nan         nan
nan
2012-06-07 00:00:00    2.2e+05        inf          0        nan         nan
nan
2012-06-15 00:00:00  7.893e+04        inf          0        nan         nan
nan
2012-06-22 00:00:00          0        nan        nan        nan         nan
nan
2012-06-29 00:00:00          0        nan        nan        nan         nan
nan
2012-07-09 00:00:00  4.709e+04        inf          0        nan         nan
nan
2012-07-13 00:00:00 -3.212e+04        inf         -0        nan         nan
nan
2012-07-20 00:00:00 -5.363e+04        inf         -0        nan         nan
nan
2012-07-27 00:00:00 -1.066e+05        inf         -0        nan         nan
nan
2012-08-06 00:00:00  1.442e+05        inf          0        nan         nan
nan
2012-08-17 00:00:00          0        nan        nan        nan         nan
nan
2012-08-24 00:00:00 -6.109e+04        inf         -0        nan         nan
nan
2012-08-31 00:00:00  3.197e+04        inf          0        nan         nan
nan
2012-09-03 00:00:00  8.956e+04        inf          0        nan         nan
nan
2012-09-14 00:00:00 -4.539e+04        inf         -0        nan         nan
nan
2012-09-21 00:00:00 -6.889e+04        inf         -0        nan         nan
nan
2012-09-28 00:00:00 -1.272e+05        inf         -0        nan         nan
nan
2012-10-02 00:00:00   1.25e+05        inf          0        nan         nan
nan
2012-10-08 00:00:00          0        nan        nan        nan         nan
nan
2012-10-19 00:00:00 -7.727e+04        inf         -0        nan         nan
nan
2012-10-26 00:00:00 -9.048e+04        inf         -0        nan         nan
nan
2012-11-05 00:00:00  5.187e+04        inf          0        nan         nan
nan
2012-12-10 00:00:00 -2.225e+04        inf         -0        nan         nan
nan
Store                5398.4091        inf          0        nan         nan
```

```
nan
Holiday_Flag         7.197e+04        inf           0        nan         nan
nan
Temperature         -2093.1356        inf          -0        nan         nan
nan
Fuel_Price          -1.303e+04        inf          -0        nan         nan
nan
CPI                  7884.0385        inf           0        nan         nan
nan
Unemployment         1148.7626        inf           0        nan         nan
nan
==============================================================================
Omnibus:                        0.912   Durbin-Watson:                   0.037
Prob(Omnibus):                  0.634   Jarque-Bera (JB):                0.464
Skew:                          -0.072   Prob(JB):                        0.793
Kurtosis:                       3.288   Cond. No.                     2.34e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The input rank is higher than the number of observations.
[3] The condition number is large, 2.34e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

[218]:
```
sns.heatmap(df_wal_new.corr())
```

[218]:
```
<AxesSubplot:>
```


2010-01-10 00:00:00
2010-03-26 00:00:00
2010-05-14 00:00:00
2010-07-23 00:00:00
2010-09-17 00:00:00
2010-11-19 00:00:00
2011-01-04 00:00:00
2011-02-25 00:00:00
2011-04-22 00:00:00
2011-06-24 00:00:00
2011-08-19 00:00:00
2011-10-14 00:00:00
2011-12-08 00:00:00
2012-02-03 00:00:00
2012-04-05 00:00:00
2012-06-04 00:00:00
2012-07-27 00:00:00
2012-09-28 00:00:00
Weekly_Sales
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
2010-01-10 00:00:00
2010-03-12 00:00:00
2010-04-30 00:00:00
2010-05-28 00:00:00
2010-07-23 00:00:00
2010-09-04 00:00:00
2010-10-15 00:00:00
2010-12-02 00:00:00
2011-01-04 00:00:00
2011-02-12 00:00:00
2011-04-02 00:00:00
2011-05-08 00:00:00
2011-06-24 00:00:00
2011-08-04 00:00:00
2011-09-16 00:00:00
2011-10-28 00:00:00
2011-12-08 00:00:00
2012-01-20 00:00:00
2012-03-08 00:00:00
2012-04-20 00:00:00
2012-06-04 00:00:00
2012-07-13 00:00:00
2012-08-31 00:00:00
2012-10-08 00:00:00
Weekly_Sales

[ ]: