In [82]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import datetime as dt
```

In [ ]:

```
```

In [53]:

```
dfk=pd.read_csv("E:\walmart-sales-dataset-of-45stores.csv")
print(dfk.to_string())
```

```
      Store        Date  Weekly_Sales  Holiday_Flag  Temperature  Fuel_Price         CPI  Unemployment
0         1  05-02-2010    1643690.90             0        42.31       2.572  211.096358         8.106
1         1  12-02-2010    1641957.44             1        38.51       2.548  211.242170         8.106
2         1  19-02-2010    1611968.17             0        39.93       2.514  211.289143         8.106
3         1  26-02-2010    1409727.59             0        46.63       2.561  211.319643         8.106
4         1  05-03-2010    1554806.68             0        46.50       2.625  211.350143         8.106
5         1  12-03-2010    1439541.59             0        57.79       2.667  211.380643         8.106
6         1  19-03-2010    1472515.79             0        54.58       2.720  211.215635         8.106
7         1  26-03-2010    1404429.92             0        51.45       2.732  211.018042         8.106
8         1  02-04-2010    1594968.28             0        62.27       2.719  210.820450         7.808
9         1  09-04-2010    1545418.53             0        65.86       2.770  210.622857         7.808
10        1  16-04-2010    1466058.28             0        66.32       2.808  210.488700         7.808
11        1  23-04-2010    1391256.12             0        64.84       2.795  210.439123         7.808
12        1  30-04-2010    1425100.71             0        67.41       2.780  210.389546         7.808
13        1  07-05-2010    1603955.12             0        72.55       2.835  210.339968         7.808
14        1  14-05-2010    1494251.50             0        74.78       2.854  210.337426         7.808
15        1  21-05-2010    1399662.07             0        76.44       2.826  210.617093         7.808
16        1  28-05-2010    1432069.95             0        80.44       2.759  210.896761         7.808
17        1  04-06-2010    1615524.71             0        80.69       2.705  211.176428         7.808
```

In [54]:

```
for x in dfk.index:
    if dfk.loc[x,"Holiday_Flag"] > 0:
        dfk.loc[x,"Holiday_Flag"] = "Holiday_week"
    elif dfk.loc[x,"Holiday_Flag"] < 1:
        dfk.loc[x,"Holiday_Flag"] = "Non_holiday_week"
print(dfk.to_string())
```

```
      Store        Date  Weekly_Sales      Holiday_Flag  Temperature  Fuel_Price         CPI  Unemployment
0         1  05-02-2010    1643690.90  Non_holiday_week        42.31       2.572  211.096358         8.106
1         1  12-02-2010    1641957.44      Holiday_week        38.51       2.548  211.242170         8.106
2         1  19-02-2010    1611968.17  Non_holiday_week        39.93       2.514  211.289143         8.106
3         1  26-02-2010    1409727.59  Non_holiday_week        46.63       2.561  211.319643         8.106
4         1  05-03-2010    1554806.68  Non_holiday_week        46.50       2.625  211.350143         8.106
5         1  12-03-2010    1439541.59  Non_holiday_week        57.79       2.667  211.380643         8.106
6         1  19-03-2010    1472515.79  Non_holiday_week        54.58       2.720  211.215635         8.106
7         1  26-03-2010    1404429.92  Non_holiday_week        51.45       2.732  211.018042         8.106
8         1  02-04-2010    1594968.28  Non_holiday_week        62.27       2.719  210.820450         7.808
9         1  09-04-2010    1545418.53  Non_holiday_week        65.86       2.770  210.622857         7.808
10        1  16-04-2010    1466058.28  Non_holiday_week        66.32       2.808  210.488700         7.808
11        1  23-04-2010    1391256.12  Non_holiday_week        64.84       2.795  210.439123         7.808
12        1  30-04-2010    1425100.71  Non_holiday_week        67.41       2.780  210.389546         7.808
13        1  07-05-2010    1603955.12  Non_holiday_week        72.55       2.835  210.339968         7.808
14        1  14-05-2010    1494251.50  Non_holiday_week        74.78       2.854  210.337426         7.808
15        1  21-05-2010    1399662.07  Non_holiday_week        76.44       2.826  210.617093         7.808
16        1  28-05-2010    1432069.95  Non_holiday_week        80.44       2.759  210.896761         7.808
17        1  04-06-2010    1615524.71  Non_holiday_week        80.69       2.705  211.176428         7.808
```

In [55]:

```
dfk
```

Out[55]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	Non_holiday_week	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	Holiday_week	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	Non_holiday_week	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	Non_holiday_week	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	Non_holiday_week	46.50	2.625	211.350143	8.106
...	...	...	...	...	...	...	...	...
6430	45	28-09-2012	713173.95	Non_holiday_week	64.88	3.997	192.013558	8.684
6431	45	05-10-2012	733455.07	Non_holiday_week	64.89	3.985	192.170412	8.667
6432	45	12-10-2012	734464.36	Non_holiday_week	54.47	4.000	192.327265	8.667
6433	45	19-10-2012	718125.53	Non_holiday_week	56.47	3.969	192.330854	8.667
6434	45	26-10-2012	760281.43	Non_holiday_week	58.85	3.882	192.308899	8.667

6435 rows × 8 columns

In [181]:

```
desk=dfk.describe()
desk
```

Out[181]:

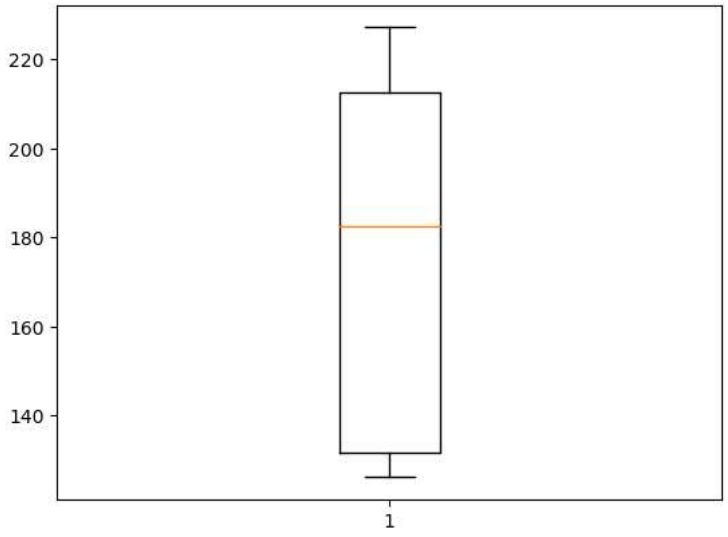
	Store	Weekly_Sales	Temperature	Fuel_Price	CPI	Unemployment
count	6435.000000	6.435000e+03	6435.000000	6435.000000	6435.000000	6435.000000
mean	23.000000	1.046965e+06	60.663782	3.358607	171.578394	7.999151
std	12.988182	5.643666e+05	18.444933	0.459020	39.356712	1.875885
min	1.000000	2.099862e+05	-2.060000	2.472000	126.064000	3.879000
25%	12.000000	5.533501e+05	47.460000	2.933000	131.735000	6.891000
50%	23.000000	9.607460e+05	62.670000	3.445000	182.616521	7.874000
75%	34.000000	1.420159e+06	74.940000	3.735000	212.743293	8.622000
max	45.000000	3.818686e+06	100.140000	4.468000	227.232807	14.313000

In [185]:

```
plt.boxplot(dfk['CPI'])
```

Out[185]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1df8e1c8a00>,
  <matplotlib.lines.Line2D at 0x1df8e1c8cd0>],
 'caps': [<matplotlib.lines.Line2D at 0x1df8e1c8fa0>,
  <matplotlib.lines.Line2D at 0x1df8e1d72b0>],
 'boxes': [<matplotlib.lines.Line2D at 0x1df8e1c8700>],
 'medians': [<matplotlib.lines.Line2D at 0x1df8e1d7580>],
 'fliers': [<matplotlib.lines.Line2D at 0x1df8e1d7850>],
 'means': []}
```

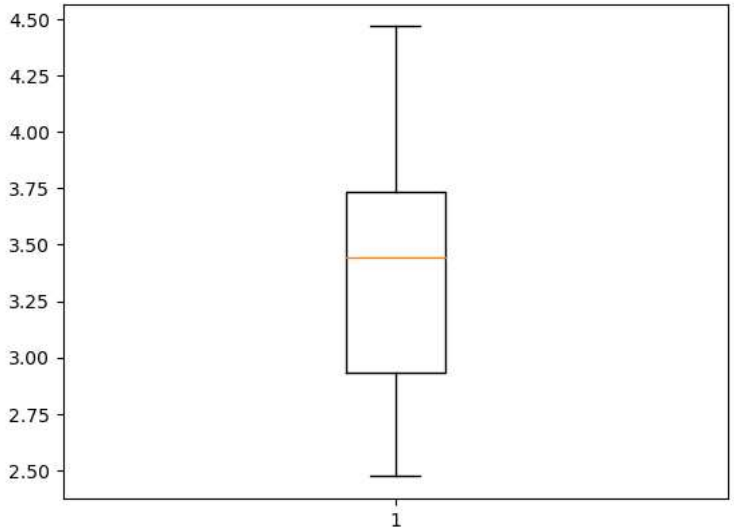


In [186]:

```
plt.boxplot(dfk['Fuel_Price'])
```

Out[186]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1df8e230520>,
  <matplotlib.lines.Line2D at 0x1df8e230820>],
 'caps': [<matplotlib.lines.Line2D at 0x1df8e230af0>,
  <matplotlib.lines.Line2D at 0x1df8e230dc0>],
 'boxes': [<matplotlib.lines.Line2D at 0x1df8e230250>],
 'medians': [<matplotlib.lines.Line2D at 0x1df8e23e0d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x1df8e23e3a0>],
 'means': []}
```

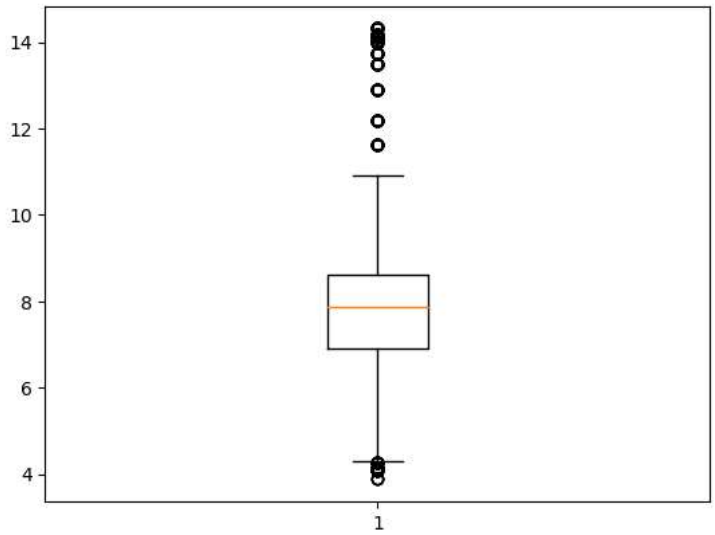


In [188]:

```
plt.boxplot(dfk['Unemployment'])
```

Out[188]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1df8e47b7c0>,
  <matplotlib.lines.Line2D at 0x1df8e47bac0>],
 'caps': [<matplotlib.lines.Line2D at 0x1df8e47bd90>,
  <matplotlib.lines.Line2D at 0x1df8e30a0a0>],
 'boxes': [<matplotlib.lines.Line2D at 0x1df8e47b4f0>],
 'medians': [<matplotlib.lines.Line2D at 0x1df8e30a3a0>],
 'fliers': [<matplotlib.lines.Line2D at 0x1df8e30a670>],
 'means': []}
```

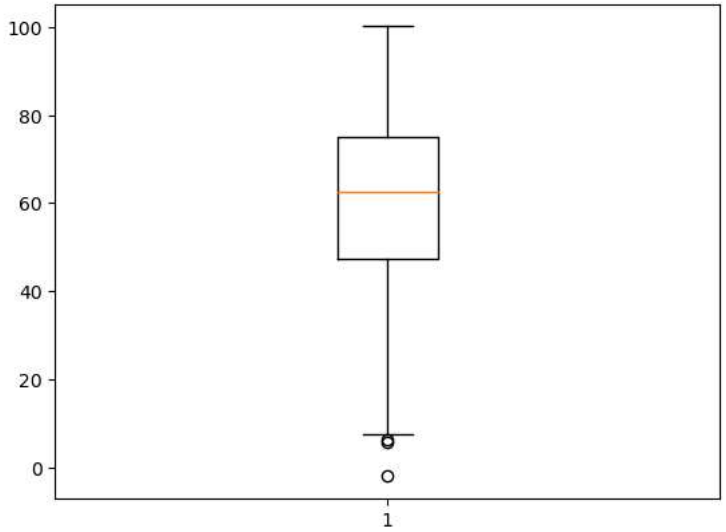


In [189]:

```
plt.boxplot(dfk['Temperature'])
```

Out[189]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1df8e3607c0>,
  <matplotlib.lines.Line2D at 0x1df8e360a90>],
 'caps': [<matplotlib.lines.Line2D at 0x1df8e360d60>,
  <matplotlib.lines.Line2D at 0x1df8e36c070>],
 'boxes': [<matplotlib.lines.Line2D at 0x1df8e3604f0>],
 'medians': [<matplotlib.lines.Line2D at 0x1df8e36c340>],
 'fliers': [<matplotlib.lines.Line2D at 0x1df8e36c610>],
 'means': []}
```



In [ ]:

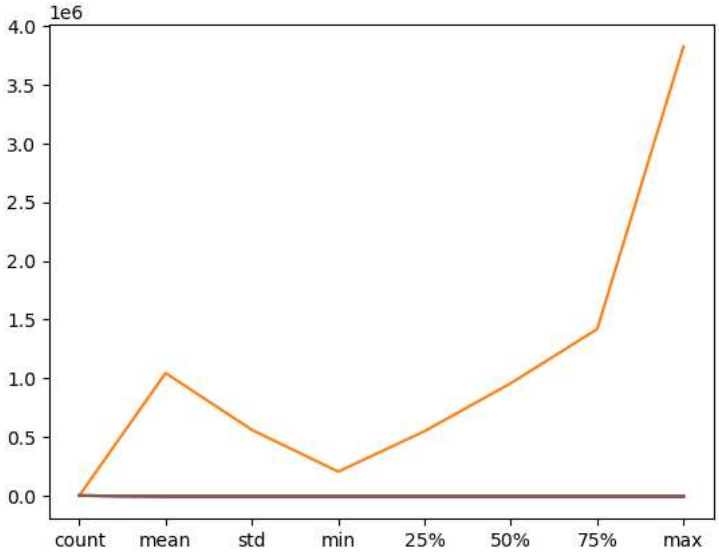
```
```

In [177]:

```
plt.plot(desk)
```

Out[177]:

```
[<matplotlib.lines.Line2D at 0x1df8d0f3490>,
 <matplotlib.lines.Line2D at 0x1df8d0f3430>,
 <matplotlib.lines.Line2D at 0x1df8d0f35e0>,
 <matplotlib.lines.Line2D at 0x1df8d0f3700>,
 <matplotlib.lines.Line2D at 0x1df8d0f3820>,
 <matplotlib.lines.Line2D at 0x1df8d0f3940>]
```



In [172]:

```
# relations items
dfk.corr()
```

Out[172]:

	Store	Weekly_Sales	Temperature	Fuel_Price	CPI	Unemployment
Store	1.000000	-0.335332	-0.022659	0.060023	-0.209492	0.223531
Weekly_Sales	-0.335332	1.000000	-0.063810	0.009464	-0.072634	-0.106176
Temperature	-0.022659	-0.063810	1.000000	0.144982	0.176888	0.101158
Fuel_Price	0.060023	0.009464	0.144982	1.000000	-0.170642	-0.034684
CPI	-0.209492	-0.072634	0.176888	-0.170642	1.000000	-0.302020
Unemployment	0.223531	-0.106176	0.101158	-0.034684	-0.302020	1.000000

In [ ]:

In [58]:

```
dfk.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   object 
 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(1), object(2)
memory usage: 402.3+ KB
```

In [59]:

```
col=dfk.columns
dfk[col].isnull().sum()
```

Out[59]:

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

In [60]:

```
dfk.duplicated().to_string( index=False)
```


```
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse\nFalse
```


In [62]:

```
dfk.drop_duplicates(inplace=True)
```

In [64]:

```
dfk.duplicated()
```

Out[64]:

```
0       False
1       False
2       False
3       False
4       False
        ...
6430    False
6431    False
6432    False
6433    False
6434    False
Length: 6435, dtype: bool
```

In [65]:

```
dfmax1=dfk.groupby('Store')['Weekly_Sales'].sum()
dfmax1
```

Out[65]:

```
Store
1     2.224028e+08
2     2.753824e+08
3     5.758674e+07
4     2.995440e+08
5     4.547569e+07
6     2.237561e+08
7     8.159828e+07
8     1.299512e+08
9     7.778922e+07
10    2.716177e+08
11    1.939628e+08
12    1.442872e+08
13    2.865177e+08
14    2.889999e+08
15    8.913368e+07
16    7.425243e+07
17    1.277821e+08
18    1.551147e+08
19    2.066349e+08
20    3.013978e+08
21    1.081179e+08
22    1.470756e+08
23    1.987506e+08
24    1.940160e+08
25    1.010612e+08
26    1.434164e+08
27    2.538559e+08
28    1.892637e+08
29    7.714155e+07
30    6.271689e+07
31    1.996139e+08
32    1.668192e+08
33    3.716022e+07
34    1.382498e+08
35    1.315207e+08
36    5.341221e+07
37    7.420274e+07
38    5.515963e+07
39    2.074455e+08
40    1.378703e+08
41    1.813419e+08
42    7.956575e+07
43    9.056544e+07
44    4.329309e+07
45    1.123953e+08
Name: Weekly_Sales, dtype: float64
```

In [66]:

```
print("index {} and max_value {}".format(dfmax1.idxmax(),dfmax1.max()))
```

```
index 20 and max_value 301397792.46
```

In [67]:

```
dfmax2=dfk.groupby('Store')['Weekly_Sales'].std()
dfmax2
```

Out[67]:

```
Store
1     155980.767761
2     237683.694682
3      46319.631557
4     266201.442297
5      37737.965745
6     212525.855862
7     112585.469220
8     106280.829881
9      69028.666585
10    302262.062504
11    165833.887863
12    139166.871880
13    265506.995776
14    317569.949476
15    120538.652043
16     85769.680133
17    112162.936087
18    176641.510839
19    191722.638730
20    275900.562742
21    128752.812853
22    161251.350631
23    249788.038068
24    167745.677567
25    112976.788600
26    110431.288141
27    239930.135688
28    181758.967539
29     99120.136596
30     22809.665590
31    125855.942933
32    138017.252087
33     24132.927322
34    104630.164676
35    211243.457791
36     60725.173579
37     21837.461190
38     42768.169450
39    217466.454833
40    119002.112858
41    187907.162766
42     50262.925530
43     40598.413260
44     24762.832015
45    130168.526635
Name: Weekly_Sales, dtype: float64
```

In [68]:

```
print("index {} and max_value {}".format(dfmax2.idxmax(),dfmax2.max()))
```

```
index 14 and max_value 317569.9494755081
```

In [69]:

```
dfk.groupby('Holiday_Flag')['Weekly_Sales'].mean().to_frame().reset_index()
```

Out[69]:

	Holiday_Flag	Weekly_Sales
0	Holiday_week	1.122888e+06
1	Non_holiday_week	1.041256e+06

In [70]:

```
# filter
holiday=dfk[dfk.Holiday_Flag=='Holiday_week']
holiday
```

Out[70]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	12-02-2010	1641957.44	Holiday_week	38.51	2.548	211.242170	8.106
31	1	10-09-2010	1507460.69	Holiday_week	78.69	2.565	211.495190	7.787
42	1	26-11-2010	1955624.11	Holiday_week	64.52	2.735	211.748433	7.838
47	1	31-12-2010	1367320.01	Holiday_week	48.43	2.943	211.404932	7.838
53	1	11-02-2011	1649614.93	Holiday_week	36.39	3.022	212.936705	7.742
...	...	...	...	...	...	...	...	...
6375	45	09-09-2011	746129.56	Holiday_week	71.48	3.738	186.673738	8.625
6386	45	25-11-2011	1170672.94	Holiday_week	48.71	3.492	188.350400	8.523
6391	45	30-12-2011	869403.63	Holiday_week	37.79	3.389	189.062016	8.523
6397	45	10-02-2012	803657.12	Holiday_week	37.00	3.640	189.707605	8.424
6427	45	07-09-2012	766512.66	Holiday_week	75.70	3.911	191.577676	8.684

450 rows × 8 columns

In [71]:

```
# filter
non_holiday=dfk[dfk.Holiday_Flag=='Non_holiday_week']
non_holiday
```

Out[71]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	Non_holiday_week	42.31	2.572	211.096358	8.106
2	1	19-02-2010	1611968.17	Non_holiday_week	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	Non_holiday_week	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	Non_holiday_week	46.50	2.625	211.350143	8.106
5	1	12-03-2010	1439541.59	Non_holiday_week	57.79	2.667	211.380643	8.106
...	...	...	...	...	...	...	...	...
6430	45	28-09-2012	713173.95	Non_holiday_week	64.88	3.997	192.013558	8.684
6431	45	05-10-2012	733455.07	Non_holiday_week	64.89	3.985	192.170412	8.667
6432	45	12-10-2012	734464.36	Non_holiday_week	54.47	4.000	192.327265	8.667
6433	45	19-10-2012	718125.53	Non_holiday_week	56.47	3.969	192.330854	8.667
6434	45	26-10-2012	760281.43	Non_holiday_week	58.85	3.882	192.308899	8.667

5985 rows × 8 columns

In [72]:

```
holidays_higher_sales=holiday[(holiday.Weekly_Sales)>(non_holiday.Weekly_Sales.mean())]
holidays_higher_sales
```

Out[72]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	12-02-2010	1641957.44	Holiday_week	38.51	2.548	211.242170	8.106
31	1	10-09-2010	1507460.69	Holiday_week	78.69	2.565	211.495190	7.787
42	1	26-11-2010	1955624.11	Holiday_week	64.52	2.735	211.748433	7.838
47	1	31-12-2010	1367320.01	Holiday_week	48.43	2.943	211.404932	7.838
53	1	11-02-2011	1649614.93	Holiday_week	36.39	3.022	212.936705	7.742
...	...	...	...	...	...	...	...	...
5819	41	30-12-2011	1264014.16	Holiday_week	34.12	3.119	196.358610	6.759
5825	41	10-02-2012	1238844.56	Holiday_week	22.00	3.103	196.919506	6.589
5855	41	07-09-2012	1392143.82	Holiday_week	67.41	3.596	198.095048	6.432
6334	45	26-11-2010	1182500.16	Holiday_week	46.15	3.039	182.783277	8.724
6386	45	25-11-2011	1170672.94	Holiday_week	48.71	3.492	188.350400	8.523

220 rows × 8 columns

In [73]:

```
dfk['year']=dfk['Date'].apply(lambda x:x[6:])
dfk['month']=dfk['Date'].apply(lambda x:x[3:5])
dfk['day']=dfk['Date'].apply(lambda x:x[0:2])
```

In [79]:

```
dfk
```

Out[79]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	year	month	day
0	1	05-02-2010	1643690.90	Non_holiday_week	42.31	2.572	211.096358	8.106	2010	02	05
1	1	12-02-2010	1641957.44	Holiday_week	38.51	2.548	211.242170	8.106	2010	02	12
2	1	19-02-2010	1611968.17	Non_holiday_week	39.93	2.514	211.289143	8.106	2010	02	19
3	1	26-02-2010	1409727.59	Non_holiday_week	46.63	2.561	211.319643	8.106	2010	02	26
4	1	05-03-2010	1554806.68	Non_holiday_week	46.50	2.625	211.350143	8.106	2010	03	05
...	...	...	...	...	...	...	...	...	...	...	...
6430	45	28-09-2012	713173.95	Non_holiday_week	64.88	3.997	192.013558	8.684	2012	09	28
6431	45	05-10-2012	733455.07	Non_holiday_week	64.89	3.985	192.170412	8.667	2012	10	05
6432	45	12-10-2012	734464.36	Non_holiday_week	54.47	4.000	192.327265	8.667	2012	10	12
6433	45	19-10-2012	718125.53	Non_holiday_week	56.47	3.969	192.330854	8.667	2012	10	19
6434	45	26-10-2012	760281.43	Non_holiday_week	58.85	3.882	192.308899	8.667	2012	10	26

6435 rows × 11 columns

In [ ]:

```
# df=df.drop('Date' , axis=1)
```

In [159]:

```
dfa=dfk.groupby(['month' ,'year'])['Weekly_Sales'].sum().to_frame().reset_index()
dfa
```

Out[159]:

	month	year	Weekly_Sales
0	01	2011	1.637040e+08
1	01	2012	1.688945e+08
2	02	2010	1.903330e+08
3	02	2011	1.863313e+08
4	02	2012	1.920636e+08
5	03	2010	1.819198e+08
6	03	2011	1.793564e+08
7	03	2012	2.315097e+08
8	04	2010	2.314124e+08
9	04	2011	2.265265e+08
10	04	2012	1.889209e+08
11	05	2010	1.867109e+08
12	05	2011	1.816482e+08
13	05	2012	1.887665e+08
14	06	2010	1.922462e+08
15	06	2011	1.897734e+08
16	06	2012	2.406103e+08
17	07	2010	2.325801e+08
18	07	2011	2.299114e+08
19	07	2012	1.875095e+08
20	08	2010	1.876401e+08
21	08	2011	1.885993e+08
22	08	2012	2.368508e+08
23	09	2010	1.772679e+08
24	09	2011	2.208477e+08
25	09	2012	1.806455e+08
26	10	2010	2.171618e+08
27	10	2011	1.832613e+08
28	10	2012	1.843617e+08
29	11	2010	2.028534e+08
30	11	2011	2.101624e+08
31	12	2010	2.887605e+08
32	12	2011	2.880781e+08

In [103]:

```
# filter
year2010=dfa[dfa.year=='2010']
year2010
```

Out[103]:

	month	year	Weekly_Sales
2	02	2010	1.903330e+08
5	03	2010	1.819198e+08
8	04	2010	2,314124e+08
11	05	2010	1.867109e+08
14	06	2010	1.922462e+08
17	07	2010	2.325801e+08
20	08	2010	1.876401e+08
23	09	2010	1.772679e+08
26	10	2010	2.171618e+08
29	11	2010	2.028534e+08
31	12	2010	2.887605e+08

In [104]:

```
# filter
year2011=dfa[dfa.year=='2011']
year2011
```

Out[104]:

	month	year	Weekly_Sales
0	01	2011	1.637040e+08
3	02	2011	1.863313e+08
6	03	2011	1.793564e+08
9	04	2011	2.265265e+08
12	05	2011	1.816482e+08
15	06	2011	1.897734e+08
18	07	2011	2.299114e+08
21	08	2011	1.885993e+08
24	09	2011	2.208477e+08
27	10	2011	1.832613e+08
30	11	2011	2.101624e+08
32	12	2011	2.880781e+08

In [105]:

```
# filter
year2012=dfa[dfa.year=='2012']
year2012
```

Out[105]:

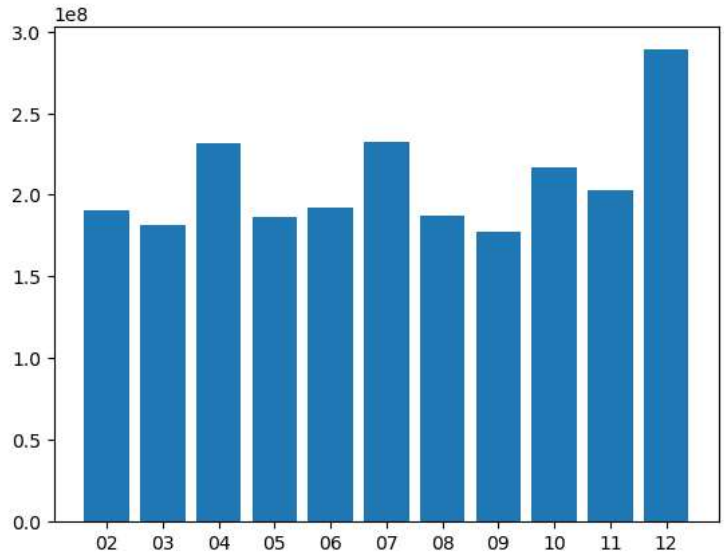
	month	year	Weekly_Sales
1	01	2012	1.688945e+08
4	02	2012	1.920636e+08
7	03	2012	2.315097e+08
10	04	2012	1.889209e+08
13	05	2012	1.887665e+08
16	06	2012	2.406103e+08
19	07	2012	1.875095e+08
22	08	2012	2.368508e+08
25	09	2012	1.806455e+08
28	10	2012	1.843617e+08

In [106]:

```
plt.bar(year2010['month'],year2010['Weekly_Sales'])
#one month notfound 12month is higher sales
```

Out[106]:

```
<BarContainer object of 11 artists>
```

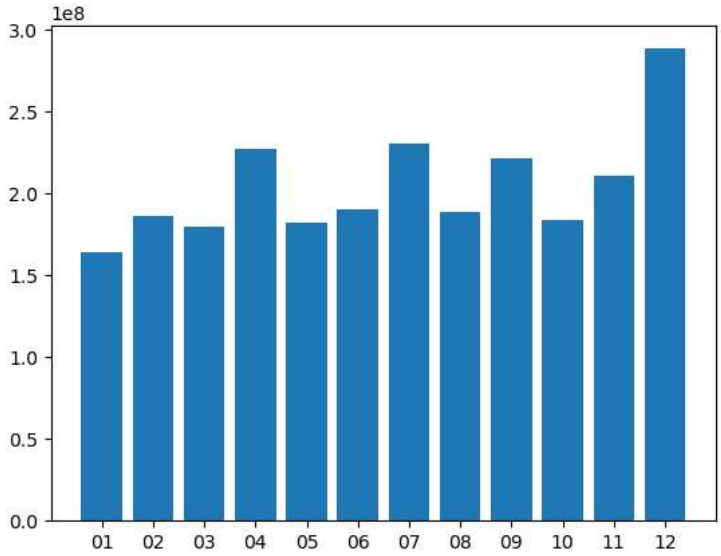


In [107]:

```
plt.bar(year2011['month'],year2011['Weekly_Sales'])
# 12month is higher sales
```

Out[107]:

```
<BarContainer object of 12 artists>
```

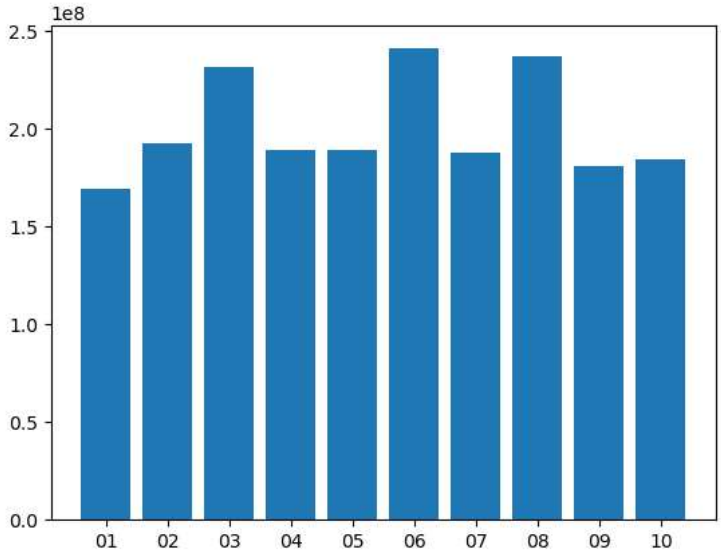


In [108]:

```
plt.bar(year2012['month'],year2012['Weekly_Sales'])
# 11 and12 months not found and 6month is higher sales
```

Out[108]:

```
<BarContainer object of 10 artists>
```



In [109]:

```
dfak=dfk.groupby('year')['Weekly_Sales'].sum().to_frame().reset_index()
dfak
```

Out[109]:

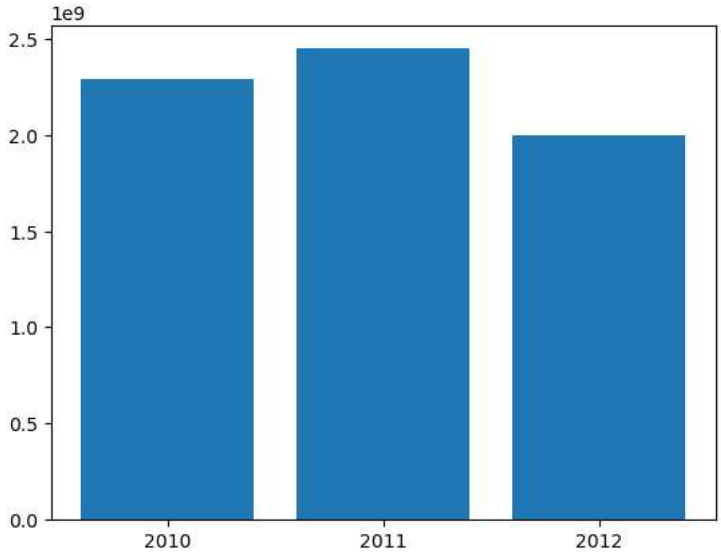
	year	Weekly_Sales
0	2010	2.288886e+09
1	2011	2.448200e+09
2	2012	2.000133e+09

In [98]:

```
plt.bar(dfak['year'],dfak['Weekly_Sales'])
```

Out[98]:

```
<BarContainer object of 3 artists>
```



In [161]:

```
sales_and_temp=dfk.groupby('Store').agg({
    'Weekly_Sales' :lambda sales :sales.sum(),
    'Temperature'  :lambda temp :temp.mean()
}).reset_index()
sales_and_temp
```

Out[161]:

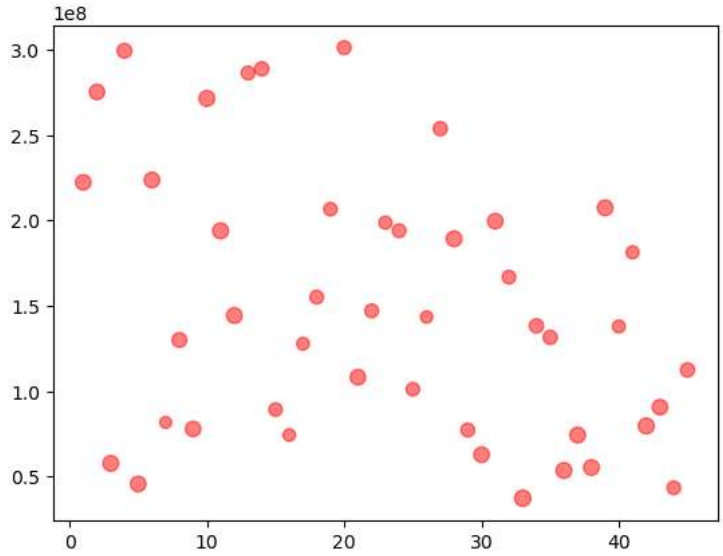
	Store	Weekly_Sales	Temperature
0	1	2.224028e+08	68.306783
1	2	2.753824e+08	68.216364
2	3	5.758674e+07	71.434196
3	4	2.995440e+08	62.253357
4	5	4.547569e+07	69.410140
5	6	2.237561e+08	69.700000
6	7	8.159828e+07	39.720280
7	8	1.299512e+08	62.513986
8	9	7.778922e+07	67.775175
9	10	2.716177e+08	72.241189
10	11	1.939628e+08	72.480769
11	12	1.442872e+08	70.262797
12	13	2.865177e+08	53.697133
13	14	2.889999e+08	57.790979
14	15	8.913368e+07	51.833846
15	16	7.425243e+07	45.030070
16	17	1.277821e+08	46.387203
17	18	1.551147e+08	53.371259
18	19	2.066349e+08	52.295035
19	20	3.013978e+08	55.451399
20	21	1.081179e+08	68.847622
21	22	1.470756e+08	54.897133
22	23	1.987506e+08	48.805105
23	24	1.940160e+08	54.030000
24	25	1.010612e+08	52.138392
25	26	1.434164e+08	43.658252
26	27	2.538559e+08	57.311119
27	28	1.892637e+08	70.262797
28	29	7.714155e+07	54.897133
29	30	6.271689e+07	68.847622
30	31	1.996139e+08	68.847622
31	32	1.668192e+08	52.747552
32	33	3.716022e+07	76.728182
33	34	1.382498e+08	58.495874
34	35	1.315207e+08	57.311119
35	36	5.341221e+07	71.160350
36	37	7.420274e+07	71.160350
37	38	5.515963e+07	70.262797
38	39	2.074455e+08	70.597343
39	40	1.378703e+08	47.674545
40	41	1.813419e+08	48.410350
41	42	7.956575e+07	72.241189
42	43	9.056544e+07	68.877692
43	44	4.329309e+07	53.697133
44	45	1.123953e+08	57.790979

In [169]:

```
plt.scatter(sales_and_temp['Store'],sales_and_temp['Weekly_Sales'],s=sales_and_temp['Temperature'],color='red',alpha=.5)
# temperature is not effect in thr weekly_sales
```

Out[169]:

```
<matplotlib.collections.PathCollection at 0x1df8ccdbfa0>
```



In [ ]:

In [162]:

```
sales_and_cpi=dfk.groupby('Store').agg({
    'Weekly_Sales' :lambda sales :sales.sum(),
    'CPI'  :lambda cpi :cpi.mean()
}).reset_index()
sales_and_cpi
```

Out[162]:

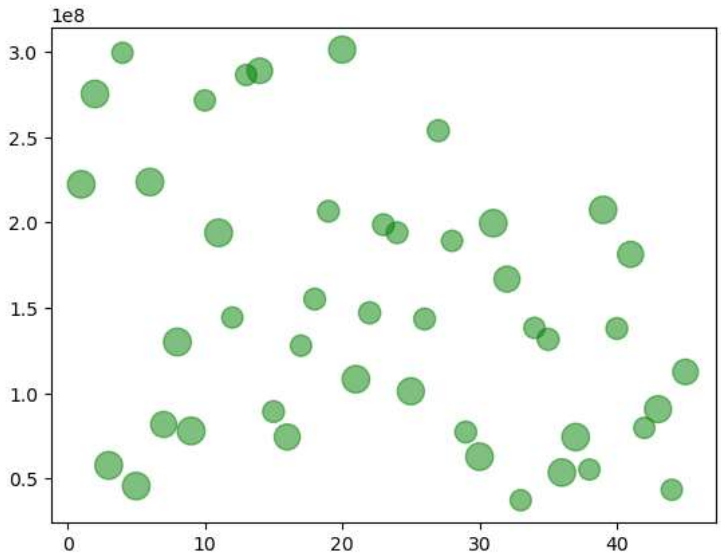
	Store	Weekly_Sales	CPI
0	1	2.224028e+08	215.996892
1	2	2.753824e+08	215.646311
2	3	5.758674e+07	219.391531
3	4	2.995440e+08	128.679669
4	5	4.547569e+07	216.565581
5	6	2.237561e+08	217.553197
6	7	8.159828e+07	193.664243
7	8	1.299512e+08	219.439026
8	9	7.778922e+07	219.626689
9	10	2.716177e+08	128.679669
10	11	1.939628e+08	219.391531
11	12	1.442872e+08	128.679669
12	13	2.865177e+08	128.679669
13	14	2.889999e+08	186.285678
14	15	8.913368e+07	135.092607
15	16	7.425243e+07	193.664243
16	17	1.277821e+08	128.679669
17	18	1.551147e+08	135.092607
18	19	2.066349e+08	135.092607
19	20	3.013978e+08	209.038131
20	21	1.081179e+08	215.646311
21	22	1.470756e+08	139.011284
22	23	1.987506e+08	135.092607
23	24	1.940160e+08	135.092607
24	25	1.010612e+08	209.038131
25	26	1.434164e+08	135.092607
26	27	2.538559e+08	139.011284
27	28	1.892637e+08	128.679669
28	29	7.714155e+07	135.092607
29	30	6.271689e+07	215.646311
30	31	1.996139e+08	215.646311
31	32	1.668192e+08	193.664243
32	33	3.716022e+07	128.679669
33	34	1.382498e+08	128.679669
34	35	1.315207e+08	139.011284
35	36	5.341221e+07	214.729069
36	37	7.420274e+07	214.729069
37	38	5.515963e+07	128.679669
38	39	2.074455e+08	214.729069
39	40	1.378703e+08	135.092607
40	41	1.813419e+08	193.664243
41	42	7.956575e+07	128.679669
42	43	9.056544e+07	207.735162
43	44	4.329309e+07	128.679669
44	45	1.123953e+08	186.285678

In [170]:

```
plt.scatter(sales_and_cpi['Store'],sales_and_cpi['Weekly_Sales'],s=sales_and_cpi['CPI'],color='g',alpha=.5)
# CPI is effect in the weekly_sales
# increase the CPI leads to increase the weeekly_sales
```

Out[170]:

```
<matplotlib.collections.PathCollection at 0x1df8ce3af40>
```



In [ ]:

```
                    # THE FINAL PROJEGT DATA SCIENCE METHODOLOGE
```

In [ ]:

```

```