In [1]:
```
import matplotlib.pyplot as plt
```

In [2]:
```
import pandas as pd
```

In [3]:
```
import numpy as np
```

In [4]:
```
df = pd.read_csv("/Users/u.v._ray/Downloads/Walmart_Store_sales.csv")
```

In [5]:
```
df
```

Out[5]:

| | 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 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **6430** | 45 | 28-09-2012 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 |
| **6431** | 45 | 05-10-2012 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 |
| **6432** | 45 | 12-10-2012 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 |
| **6433** | 45 | 19-10-2012 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 |
| **6434** | 45 | 26-10-2012 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 |

6435 rows × 8 columns

In [6]:
```
df_store = df.groupby(['Store'])['Weekly_Sales'].agg(['sum','std','mean'])
```

In [7]:
```
df_store['mean/std (1/CV)'] = df_store['mean']/df_store['std']
```

In [8]:
```
df_store['coefficient of variation(CV)'] = df_store['std']/df_store['mean']
```

# Store-wise sum, mean, standard deviation, CV (coefficient of variation) and 1/(CV)

In [9]:
```
df_store
```

Out[9]:

| Store | sum | std | mean | mean/std (1/CV) | coefficient of variation(CV) |
|---|---|---|---|---|---|
| 1 | 2.224028e+08 | 155980.767761 | 1.555264e+06 | 9.970873 | 0.100292 |
| 2 | 2.753824e+08 | 237683.694682 | 1.925751e+06 | 8.102160 | 0.123424 |
| 3 | 5.758674e+07 | 46319.631557 | 4.027044e+05 | 8.694034 | 0.115021 |
| 4 | 2.995440e+08 | 266201.442297 | 2.094713e+06 | 7.868902 | 0.127083 |
| 5 | 4.547569e+07 | 37737.965745 | 3.180118e+05 | 8.426840 | 0.118668 |
| 6 | 2.237561e+08 | 212525.855862 | 1.564728e+06 | 7.362531 | 0.135823 |
| 7 | 8.159828e+07 | 112585.469220 | 5.706173e+05 | 5.068303 | 0.197305 |
| 8 | 1.299512e+08 | 106280.829881 | 9.087495e+05 | 8.550456 | 0.116953 |
| 9 | 7.778922e+07 | 69028.666585 | 5.439806e+05 | 7.880502 | 0.126895 |
| 10 | 2.716177e+08 | 302262.062504 | 1.899425e+06 | 6.284032 | 0.159133 |
| 11 | 1.939628e+08 | 165833.887863 | 1.356383e+06 | 8.179167 | 0.122262 |
| 12 | 1.442872e+08 | 139166.871880 | 1.009002e+06 | 7.250300 | 0.137925 |
| 13 | 2.865177e+08 | 265506.995776 | 2.003620e+06 | 7.546394 | 0.132514 |
| 14 | 2.889999e+08 | 317569.949476 | 2.020978e+06 | 6.363884 | 0.157137 |
| 15 | 8.913368e+07 | 120538.652043 | 6.233125e+05 | 5.171059 | 0.193384 |
| 16 | 7.425243e+07 | 85769.680133 | 5.192477e+05 | 6.053978 | 0.165181 |
| 17 | 1.277821e+08 | 112162.936087 | 8.935814e+05 | 7.966815 | 0.125521 |
| 18 | 1.551147e+08 | 176641.510839 | 1.084718e+06 | 6.140790 | 0.162845 |

| | | | | | |
|---|---|---|---|---|---|
| **19** | 2.066349e+08 | 191722.638730 | 1.444999e+06 | 7.536924 | 0.132680 |
| **20** | 3.013978e+08 | 275900.562742 | 2.107677e+06 | 7.639263 | 0.130903 |
| **21** | 1.081179e+08 | 128752.812853 | 7.560691e+05 | 5.872253 | 0.170292 |
| **22** | 1.470756e+08 | 161251.350631 | 1.028501e+06 | 6.378248 | 0.156783 |
| **23** | 1.987506e+08 | 249788.038068 | 1.389864e+06 | 5.564175 | 0.179721 |
| **24** | 1.940160e+08 | 167745.677567 | 1.356755e+06 | 8.088169 | 0.123637 |
| **25** | 1.010612e+08 | 112976.788600 | 7.067215e+05 | 6.255458 | 0.159860 |
| **26** | 1.434164e+08 | 110431.288141 | 1.002912e+06 | 9.081773 | 0.110111 |
| **27** | 2.538559e+08 | 239930.135688 | 1.775216e+06 | 7.398888 | 0.135155 |
| **28** | 1.892637e+08 | 181758.967539 | 1.323522e+06 | 7.281744 | 0.137330 |
| **29** | 7.714155e+07 | 99120.136596 | 5.394514e+05 | 5.442400 | 0.183742 |
| **30** | 6.271689e+07 | 22809.665590 | 4.385796e+05 | 19.227797 | 0.052008 |
| **31** | 1.996139e+08 | 125855.942933 | 1.395901e+06 | 11.091264 | 0.090161 |
| **32** | 1.668192e+08 | 138017.252087 | 1.166568e+06 | 8.452336 | 0.118310 |
| **33** | 3.716022e+07 | 24132.927322 | 2.598617e+05 | 10.767931 | 0.092868 |
| **34** | 1.382498e+08 | 104630.164676 | 9.667816e+05 | 9.239989 | 0.108225 |
| **35** | 1.315207e+08 | 211243.457791 | 9.197250e+05 | 4.353863 | 0.229681 |
| **36** | 5.341221e+07 | 60725.173579 | 3.735120e+05 | 6.150859 | 0.162579 |
| **37** | 7.420274e+07 | 21837.461190 | 5.189003e+05 | 23.761933 | 0.042084 |
| **38** | 5.515963e+07 | 42768.169450 | 3.857317e+05 | 9.019129 | 0.110875 |
| **39** | 2.074455e+08 | 217466.454833 | 1.450668e+06 | 6.670767 | 0.149908 |
| **40** | 1.378703e+08 | 119002.112858 | 9.641280e+05 | 8.101772 | 0.123430 |

| | | | | | |
|---|---|---|---|---|---|
| **41** | 1.813419e+08 | 187907.162766 | 1.268125e+06 | 6.748681 | 0.148177 |
| **42** | 7.956575e+07 | 50262.925530 | 5.564039e+05 | 11.069866 | 0.090335 |
| **43** | 9.056544e+07 | 40598.413260 | 6.333247e+05 | 15.599741 | 0.064104 |
| **44** | 4.329309e+07 | 24762.832015 | 3.027489e+05 | 12.225939 | 0.081793 |
| **45** | 1.123953e+08 | 130168.526635 | 7.859814e+05 | 6.038183 | 0.165613 |

## Mean / Standard Deviation of Store with highest standard deviation (Store 14) is 6.363884

## CV of the Store with highest standard deviation (Store 14) is 0.157137

In [10]:
```
df_store['sum'].idxmax()
```

Out[10]:
```
20
```

## Hence, store number 20 has the maximum sales

In [11]:
```
df_store['std'].idxmax()
```

Out[11]:
```
14
```

## Hence, store number 14 has the maximum standard deviation

In [12]:
```
df["Date"] = pd.to_datetime(df["Date"], dayfirst = True)
df['Quarter'] = pd.to_datetime(df['Date']).dt.to_period('Q')
```

In [13]:
```
df
```

Out[13]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 |

6435 rows × 9 columns

In [14]:
```
df_2012Q3 = df.loc[df['Quarter'] == '2012Q3']
```

In [15]:
```
df_2012Q3
```

Out[15]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter |
|---|---|---|---|---|---|---|---|---|---|
| **126** | 1 | 2012-07-06 | 1769854.16 | 0 | 81.57 | 3.227 | 221.883779 | 6.908 | 2012Q3 |
| **127** | 1 | 2012-07-13 | 1527014.04 | 0 | 77.12 | 3.256 | 221.924158 | 6.908 | 2012Q3 |
| **128** | 1 | 2012-07-20 | 1497954.76 | 0 | 80.42 | 3.311 | 221.932727 | 6.908 | 2012Q3 |
| **129** | 1 | 2012-07-27 | 1439123.71 | 0 | 82.66 | 3.407 | 221.941295 | 6.908 | 2012Q3 |
| **130** | 1 | 2012-08-03 | 1631135.79 | 0 | 86.11 | 3.417 | 221.949864 | 6.908 | 2012Q3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6426** | 45 | 2012-08-31 | 734297.87 | 0 | 75.09 | 3.867 | 191.461281 | 8.684 | 2012Q3 |
| **6427** | 45 | 2012-09-07 | 766512.66 | 1 | 75.70 | 3.911 | 191.577676 | 8.684 | 2012Q3 |
| **6428** | 45 | 2012-09-14 | 702238.27 | 0 | 67.87 | 3.948 | 191.699850 | 8.684 | 2012Q3 |
| **6429** | 45 | 2012-09-21 | 723086.20 | 0 | 65.32 | 4.038 | 191.856704 | 8.684 | 2012Q3 |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 |

585 rows × 9 columns

In [16]:
```
df_2011Q3 = df.loc[df['Quarter'] == '2011Q3']
```

In [17]:
```
df_2011Q3
```

Out[17]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter |
|---|---|---|---|---|---|---|---|---|---|
| **73** | 1 | 2011-07-01 | 1488538.09 | 0 | 85.55 | 3.524 | 215.184137 | 7.962 | 2011Q3 |
| **74** | 1 | 2011-07-08 | 1534849.64 | 0 | 85.83 | 3.480 | 215.277175 | 7.962 | 2011Q3 |
| **75** | 1 | 2011-07-15 | 1455119.97 | 0 | 88.54 | 3.575 | 215.361109 | 7.962 | 2011Q3 |
| **76** | 1 | 2011-07-22 | 1396926.82 | 0 | 85.77 | 3.651 | 215.422278 | 7.962 | 2011Q3 |
| **77** | 1 | 2011-07-29 | 1352219.79 | 0 | 86.83 | 3.682 | 215.483448 | 7.962 | 2011Q3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6374** | 45 | 2011-09-02 | 726482.39 | 0 | 70.63 | 3.703 | 186.618927 | 8.625 | 2011Q3 |
| **6375** | 45 | 2011-09-09 | 746129.56 | 1 | 71.48 | 3.738 | 186.673738 | 8.625 | 2011Q3 |
| **6376** | 45 | 2011-09-16 | 711367.56 | 0 | 69.17 | 3.742 | 186.802400 | 8.625 | 2011Q3 |
| **6377** | 45 | 2011-09-23 | 714106.42 | 0 | 63.75 | 3.711 | 187.029532 | 8.625 | 2011Q3 |
| **6378** | 45 | 2011-09-30 | 698986.34 | 0 | 70.66 | 3.645 | 187.256664 | 8.625 | 2011Q3 |

630 rows × 9 columns

In [18]:
```
df_store_2012Q3 = df_2012Q3.groupby(['Store'])['Weekly_Sales'].agg('mean')
```

In [19]:
```
df_store_2012Q3
```

Out[19]:
```
Store
1     1.557996e+06
2     1.869489e+06
3     4.075389e+05
4     2.138215e+06
5     3.202916e+05
6     1.551332e+06
```

| | |
|---|---|
| 7 | 6.355990e+05 |
| 8 | 9.037656e+05 |
| 9 | 5.401654e+05 |
| 10 | 1.772097e+06 |
| 11 | 1.347391e+06 |
| 12 | 9.643326e+05 |
| 13 | 2.032405e+06 |
| 14 | 1.629812e+06 |
| 15 | 5.855447e+05 |
| 16 | 5.478109e+05 |
| 17 | 9.584195e+05 |
| 18 | 1.037674e+06 |
| 19 | 1.400273e+06 |
| 20 | 2.068579e+06 |
| 21 | 6.944307e+05 |
| 22 | 9.880877e+05 |
| 23 | 1.433961e+06 |
| 24 | 1.382798e+06 |
| 25 | 7.006986e+05 |
| 26 | 1.051976e+06 |
| 27 | 1.715978e+06 |
| 28 | 1.236977e+06 |
| 29 | 5.131719e+05 |
| 30 | 4.303617e+05 |
| 31 | 1.369747e+06 |
| 32 | 1.184348e+06 |
| 33 | 2.641246e+05 |
| 34 | 9.604612e+05 |
| 35 | 8.709555e+05 |
| 36 | 2.947455e+05 |
| 37 | 5.175437e+05 |
| 38 | 4.311910e+05 |
| 39 | 1.593470e+06 |
| 40 | 9.902458e+05 |
| 41 | 1.391834e+06 |
| 42 | 5.612892e+05 |
| 43 | 6.154286e+05 |
| 44 | 3.393270e+05 |

```
45    7.370206e+05
Name: Weekly_Sales, dtype: float64
```

In [20]:
```
df_store_2011Q3 = df_2011Q3.groupby(['Store'])['Weekly_Sales'].agg('mean')
```

In [21]:
```
df_store_2011Q3
```

Out[21]:
```
Store
1     1.482299e+06
2     1.775147e+06
3     3.744784e+05
4     2.063440e+06
5     3.006473e+05
6     1.492996e+06
7     6.343589e+05
8     8.582549e+05
9     5.211744e+05
10    1.759068e+06
11    1.299483e+06
12    9.418439e+05
13    1.964246e+06
14    1.964976e+06
15    5.993759e+05
16    5.651860e+05
17    9.032377e+05
18    9.361896e+05
19    1.419899e+06
20    2.014044e+06
21    7.404919e+05
22    9.859813e+05
23    1.393429e+06
24    1.360220e+06
25    6.748932e+05
26    1.033906e+06
27    1.769770e+06
28    1.252402e+06
```

```
29    5.058514e+05
30    4.039673e+05
31    1.389767e+06
32    1.160821e+06
33    2.473689e+05
34    9.139889e+05
35    8.143773e+05
36    3.568670e+05
37    5.104592e+05
38    3.866201e+05
39    1.444304e+06
40    9.721564e+05
41    1.289663e+06
42    5.463608e+05
43    6.022217e+05
44    3.001900e+05
45    7.420390e+05
Name: Weekly_Sales, dtype: float64
```

In [22]:
```
## Computing the store wise quarterly growth rate from 2011 Q3 to 2012 Q3
QGR_Q3_12_11 = df_store_2012Q3 - df_store_2011Q3
```

In [23]:
```
QGR_Q3_12_11
```

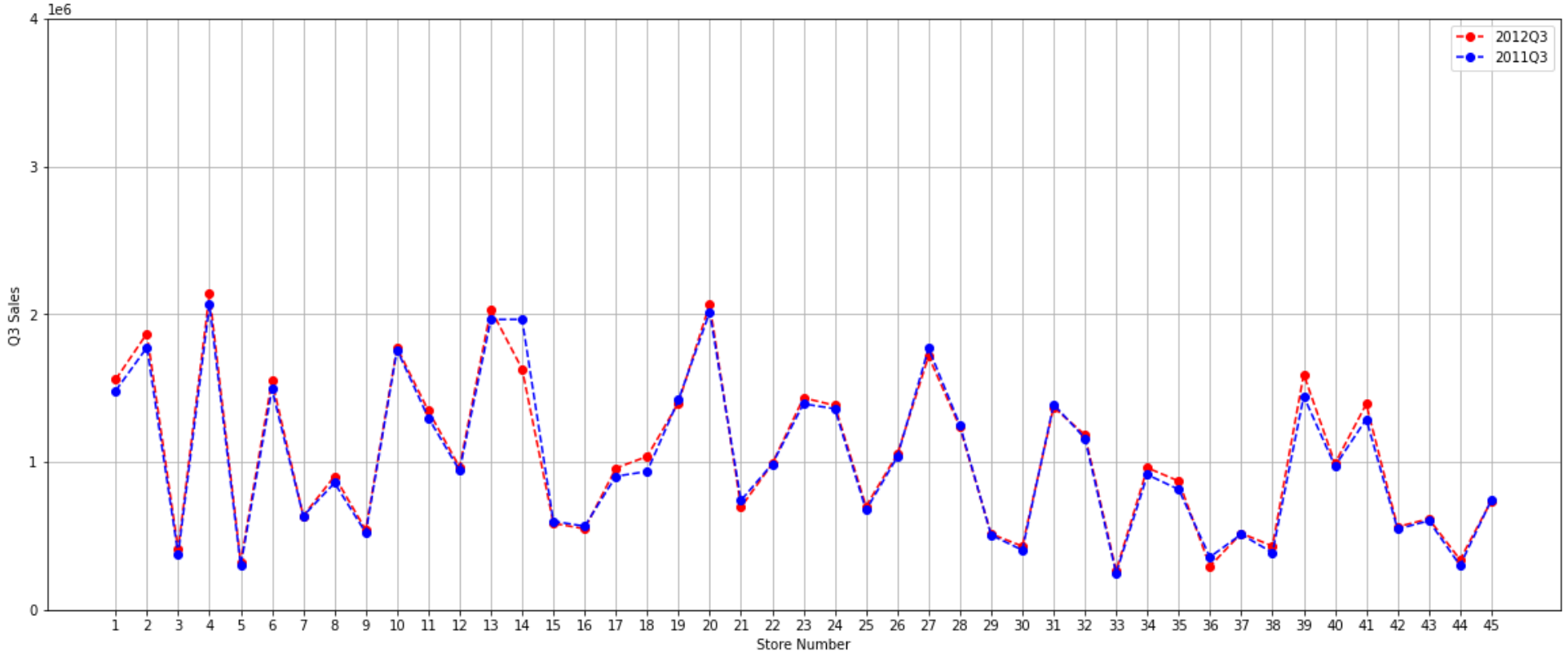
Out[23]:
```
Store
1      75697.241648
2      94341.488242
3      33060.520165
4      74774.381044
5      19644.336758
6      58335.273626
7       1240.179286
8      45510.710165
9      18990.951484
10     13028.531978
11     47907.660714
```

```
12     22488.712418
13     68158.148681
14   -335163.514451
15    -13831.168956
16    -17375.079615
17     55181.784670
18    101484.697967
19    -19625.663571
20     54535.010604
21    -46061.198077
22      2106.410714
23     40531.948846
24     22578.041868
25     25805.430934
26     18070.695055
27    -53792.500659
28    -15424.675330
29      7320.493352
30     26394.371538
31    -20019.281758
32     23527.348626
33     16755.765220
34     46472.296868
35     56578.126484
36    -62121.498516
37      7084.502198
38     44570.891593
39    149166.362088
40     18089.355549
41    102171.620330
42     14928.405714
43     13206.944835
44     39137.038736
45     -5018.397527
Name: Weekly_Sales, dtype: float64
```

In [24]:

```
plt.clf()
plt.figure(figsize = (20,8))
plt.plot(df_store_2012Q3, label = '2012Q3',color='red', marker='o', linestyle='dashed')
plt.ylabel('Q3 Sales')
plt.plot(df_store_2011Q3, label = '2011Q3',color='blue', marker='o', linestyle='dashed')
plt.xlabel('Store Number')
plt.xticks(np.arange(1,46))
plt.yticks(np.arange(0*(10**7),.5*(10**7),0.1*(10**7)))
plt.grid()
plt.legend()
plt.show()
```

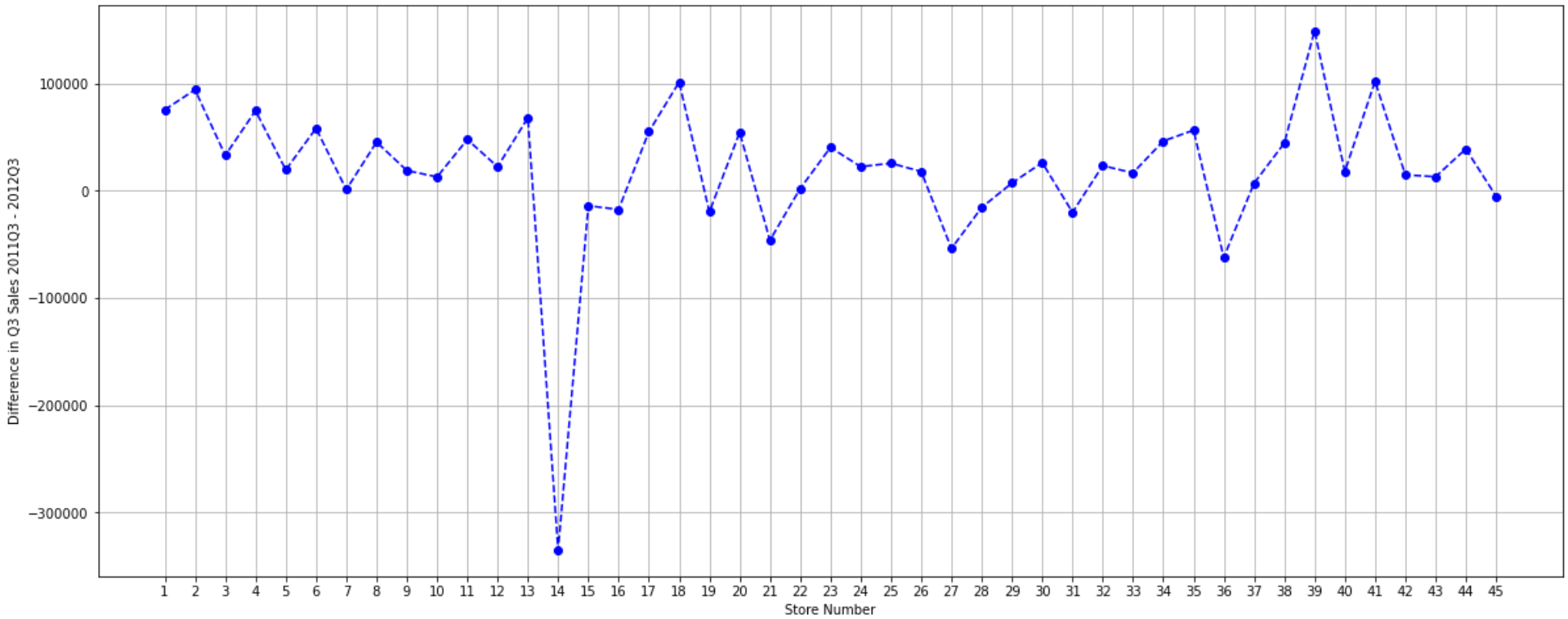
```
<Figure size 432x288 with 0 Axes>
```




In [25]:

```
plt.clf()
plt.figure(figsize = (20,8))
plt.plot(QGR_Q3_12_11, 'bo--')
plt.xticks(np.arange(1,46))
plt.xlabel('Store Number')
plt.ylabel('Difference in Q3 Sales 2011Q3 - 2012Q3')
plt.grid()
plt.show()
```

```
<Figure size 432x288 with 0 Axes>
```




In [26]:
```
(QGR_Q3_12_11).idxmax()
```

Out[26]:
```
39
```

In [27]:
```
(QGR_Q3_12_11).max()
```

Out[27]:
```
149166.36208791193
```

In [28]:
```
(QGR_Q3_12_11).idxmin()
```

Out[28]:
```
14
```

## Hence, the highest quarterly growth on 2012Q3 in median growth rate on sales is achieved by Store 39, The growth value in sales in the 3rd quarter is 149166.36208791193

In [29]:
```
df
```

Out[29]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 |

6435 rows × 9 columns

In [90]:
```
df["Month"] = df["Date"].dt.month
df["Day"] = df["Date"].dt.day
df["Year"] = df["Date"].dt.year
```

In [91]:
```
df
```

Out[91]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 |

6435 rows × 12 columns

In [32]:

```
df_holiday = df.copy().loc[df['Holiday_Flag'] == 1]
```

In [33]:
```
df_holiday
```

Out[33]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 |
| **31** | 1 | 2010-09-10 | 1507460.69 | 1 | 78.69 | 2.565 | 211.495190 | 7.787 | 2010Q3 | 9 | 10 | 2010 |
| **42** | 1 | 2010-11-26 | 1955624.11 | 1 | 64.52 | 2.735 | 211.748433 | 7.838 | 2010Q4 | 11 | 26 | 2010 |
| **47** | 1 | 2010-12-31 | 1367320.01 | 1 | 48.43 | 2.943 | 211.404932 | 7.838 | 2010Q4 | 12 | 31 | 2010 |
| **53** | 1 | 2011-02-11 | 1649614.93 | 1 | 36.39 | 3.022 | 212.936705 | 7.742 | 2011Q1 | 2 | 11 | 2011 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6375** | 45 | 2011-09-09 | 746129.56 | 1 | 71.48 | 3.738 | 186.673738 | 8.625 | 2011Q3 | 9 | 9 | 2011 |
| **6386** | 45 | 2011-11-25 | 1170672.94 | 1 | 48.71 | 3.492 | 188.350400 | 8.523 | 2011Q4 | 11 | 25 | 2011 |
| **6391** | 45 | 2011-12-30 | 869403.63 | 1 | 37.79 | 3.389 | 189.062016 | 8.523 | 2011Q4 | 12 | 30 | 2011 |
| **6397** | 45 | 2012-02-10 | 803657.12 | 1 | 37.00 | 3.640 | 189.707605 | 8.424 | 2012Q1 | 2 | 10 | 2012 |
| **6427** | 45 | 2012-09-07 | 766512.66 | 1 | 75.70 | 3.911 | 191.577676 | 8.684 | 2012Q3 | 9 | 7 | 2012 |

450 rows × 12 columns

In [34]:
```
def holiday(val):
    if val == 2:
        return 'Super Bowl'
    elif val == 9:
        return 'Labour Day'
    elif val == 11:
        return 'Thanksgiving'
    elif val == 12:
        return 'Christmas'
```

In [35]:
```
df_holiday['Holiday'] = df_holiday.apply(lambda x: holiday(x['Month']),axis=1)
```

In [36]:
```
df_holiday
```

Out[36]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | Ho |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | Super |
| **31** | 1 | 2010-09-10 | 1507460.69 | 1 | 78.69 | 2.565 | 211.495190 | 7.787 | 2010Q3 | 9 | 10 | 2010 | Labou |
| **42** | 1 | 2010-11-26 | 1955624.11 | 1 | 64.52 | 2.735 | 211.748433 | 7.838 | 2010Q4 | 11 | 26 | 2010 | Thanks |
| **47** | 1 | 2010-12-31 | 1367320.01 | 1 | 48.43 | 2.943 | 211.404932 | 7.838 | 2010Q4 | 12 | 31 | 2010 | Chri |
| **53** | 1 | 2011-02-11 | 1649614.93 | 1 | 36.39 | 3.022 | 212.936705 | 7.742 | 2011Q1 | 2 | 11 | 2011 | Super |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **6375** | 45 | 2011-09-09 | 746129.56 | 1 | 71.48 | 3.738 | 186.673738 | 8.625 | 2011Q3 | 9 | 9 | 2011 | Labou |
| **6386** | 45 | 2011-11-25 | 1170672.94 | 1 | 48.71 | 3.492 | 188.350400 | 8.523 | 2011Q4 | 11 | 25 | 2011 | Thanks |
| **6391** | 45 | 2011-12-30 | 869403.63 | 1 | 37.79 | 3.389 | 189.062016 | 8.523 | 2011Q4 | 12 | 30 | 2011 | Chri |
| **6397** | 45 | 2012-02-10 | 803657.12 | 1 | 37.00 | 3.640 | 189.707605 | 8.424 | 2012Q1 | 2 | 10 | 2012 | Super |
| **6427** | 45 | 2012-09-07 | 766512.66 | 1 | 75.70 | 3.911 | 191.577676 | 8.684 | 2012Q3 | 9 | 7 | 2012 | Labou |

450 rows × 13 columns

In [37]:
```
df_non_holiday = df.loc[df['Holiday_Flag'] == 0]
```

In [38]:
```
df_non_holiday
```

Out[38]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 |
| **5** | 1 | 2010-03-12 | 1439541.59 | 0 | 57.79 | 2.667 | 211.380643 | 8.106 | 2010Q1 | 3 | 12 | 2010 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 |

5985 rows × 12 columns

In [39]:

```
non_holiday_mean = df_non_holiday['Weekly_Sales'].mean()
```

In [40]:
```
non_holiday_mean
```

Out[40]:
```
1041256.3802088564
```

In [41]:
```
df_hol_gt_mean = df_holiday.copy().loc[df_holiday['Weekly_Sales'] > non_holiday_mean]
```

In [42]:
```
df_hol_gt_mean
```

Out[42]:

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | Ho |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | Super |
| **31** | 1 | 2010-09-10 | 1507460.69 | 1 | 78.69 | 2.565 | 211.495190 | 7.787 | 2010Q3 | 9 | 10 | 2010 | Labou |
| **42** | 1 | 2010-11-26 | 1955624.11 | 1 | 64.52 | 2.735 | 211.748433 | 7.838 | 2010Q4 | 11 | 26 | 2010 | Thanks |
| **47** | 1 | 2010-12-31 | 1367320.01 | 1 | 48.43 | 2.943 | 211.404932 | 7.838 | 2010Q4 | 12 | 31 | 2010 | Chri |
| **53** | 1 | 2011-02-11 | 1649614.93 | 1 | 36.39 | 3.022 | 212.936705 | 7.742 | 2011Q1 | 2 | 11 | 2011 | Super |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **5819** | 41 | 2011-12-30 | 1264014.16 | 1 | 34.12 | 3.119 | 196.358610 | 6.759 | 2011Q4 | 12 | 30 | 2011 | Chri |
| **5825** | 41 | 2012-02-10 | 1238844.56 | 1 | 22.00 | 3.103 | 196.919506 | 6.589 | 2012Q1 | 2 | 10 | 2012 | Super |
| **5855** | 41 | 2012-09-07 | 1392143.82 | 1 | 67.41 | 3.596 | 198.095048 | 6.432 | 2012Q3 | 9 | 7 | 2012 | Labou |
| **6334** | 45 | 2010-11-26 | 1182500.16 | 1 | 46.15 | 3.039 | 182.783277 | 8.724 | 2010Q4 | 11 | 26 | 2010 | Thanks |
| **6386** | 45 | 2011-11-25 | 1170672.94 | 1 | 48.71 | 3.492 | 188.350400 | 8.523 | 2011Q4 | 11 | 25 | 2011 | Thanks |

220 rows × 13 columns

## Following are the Holidays which have higher sales than the mean sales in non-holiday season for all stores together:

In [43]:
```
df_hol_gt_mean[['Store','Date','Holiday','Weekly_Sales']]
```

Out[43]:

| | Store | Date | Holiday | Weekly_Sales |
|---|---|---|---|---|
| **1** | 1 | 2010-02-12 | Super Bowl | 1641957.44 |
| **31** | 1 | 2010-09-10 | Labour Day | 1507460.69 |
| **42** | 1 | 2010-11-26 | Thanksgiving | 1955624.11 |
| **47** | 1 | 2010-12-31 | Christmas | 1367320.01 |
| **53** | 1 | 2011-02-11 | Super Bowl | 1649614.93 |
| **...** | ... | ... | ... | ... |
| **5819** | 41 | 2011-12-30 | Christmas | 1264014.16 |
| **5825** | 41 | 2012-02-10 | Super Bowl | 1238844.56 |
| **5855** | 41 | 2012-09-07 | Labour Day | 1392143.82 |
| **6334** | 45 | 2010-11-26 | Thanksgiving | 1182500.16 |
| **6386** | 45 | 2011-11-25 | Thanksgiving | 1170672.94 |

220 rows × 4 columns

# Monthly and Semester View of Sales in units with Insights

In [203...

```
df
```

Out[203...

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 20 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 20 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 20 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 20 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 | 20 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 | 20 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 | 20 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 | 20 |
| | | 2012- | | | | | | | | | | | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6434** | 45 | 10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 | 20 |

6435 rows × 13 columns

In [204…
```
df['month_year'] = df['Date'].dt.to_period('M')
```

In [205…
```
df
```

Out[205…

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 20 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 20 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 20 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 20 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 | 20 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 | 20 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 | 20 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 | 20 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 | 20 |

6435 rows × 13 columns

In [206...

```
df_monthly = df.groupby(['month_year'])['Weekly_Sales'].agg(['sum'])
df_monthly
```

Out[206...

| month_year | sum |
|---|---|
| **2010-02** | 1.903330e+08 |
| **2010-03** | 1.819198e+08 |
| **2010-04** | 2.314124e+08 |
| **2010-05** | 1.867109e+08 |
| **2010-06** | 1.922462e+08 |
| **2010-07** | 2.325801e+08 |
| **2010-08** | 1.876401e+08 |
| **2010-09** | 1.772679e+08 |

| | |
|---|---|
| **2010-10** | 2.171618e+08 |
| **2010-11** | 2.028534e+08 |
| **2010-12** | 2.887605e+08 |
| **2011-01** | 1.637040e+08 |
| **2011-02** | 1.863313e+08 |
| **2011-03** | 1.793564e+08 |
| **2011-04** | 2.265265e+08 |
| **2011-05** | 1.816482e+08 |
| **2011-06** | 1.897734e+08 |
| **2011-07** | 2.299114e+08 |
| **2011-08** | 1.885993e+08 |
| **2011-09** | 2.208477e+08 |
| **2011-10** | 1.832613e+08 |
| **2011-11** | 2.101624e+08 |
| **2011-12** | 2.880781e+08 |
| **2012-01** | 1.688945e+08 |
| **2012-02** | 1.920636e+08 |
| **2012-03** | 2.315097e+08 |
| **2012-04** | 1.889209e+08 |
| **2012-05** | 1.887665e+08 |
| **2012-06** | 2.406103e+08 |
| **2012-07** | 1.875095e+08 |
| **2012-08** | 2.368508e+08 |

| | |
|---|---|
| **2012-09** | 1.806455e+08 |
| **2012-10** | 1.843617e+08 |

In [207…

```
monthly_index = df_monthly.index.to_series().astype(str)
monthly_index = list(monthly_index)
monthly_index
```

Out[207…
```
['2010-02',
 '2010-03',
 '2010-04',
 '2010-05',
 '2010-06',
 '2010-07',
 '2010-08',
 '2010-09',
 '2010-10',
 '2010-11',
 '2010-12',
 '2011-01',
 '2011-02',
 '2011-03',
 '2011-04',
 '2011-05',
 '2011-06',
 '2011-07',
 '2011-08',
 '2011-09',
 '2011-10',
 '2011-11',
 '2011-12',
 '2012-01',
 '2012-02',
 '2012-03',
 '2012-04',
 '2012-05',
 '2012-06',
 '2012-07',
 '2012-08',
 '2012-09',
 '2012-10']
```

In [208…
```
df_monthly['sum'].values
```

Out[208…
```
array([1.90332983e+08, 1.81919802e+08, 2.31412368e+08, 1.86710934e+08,
       1.92246172e+08, 2.32580126e+08, 1.87640111e+08, 1.77267896e+08,
       2.17161824e+08, 2.02853370e+08, 2.88760533e+08, 1.63703967e+08,
       1.86331328e+08, 1.79356448e+08, 2.26526511e+08, 1.81648158e+08,
       1.89773385e+08, 2.29911399e+08, 1.88599332e+08, 2.20847738e+08,
       1.83261283e+08, 2.10162355e+08, 2.88078102e+08, 1.68894472e+08,
       1.92063580e+08, 2.31509650e+08, 1.88920906e+08, 1.88766479e+08,
       2.40610329e+08, 1.87509452e+08, 2.36850766e+08, 1.80645544e+08,
       1.84361680e+08])
```

In [209…
```
len(monthly_index)
```

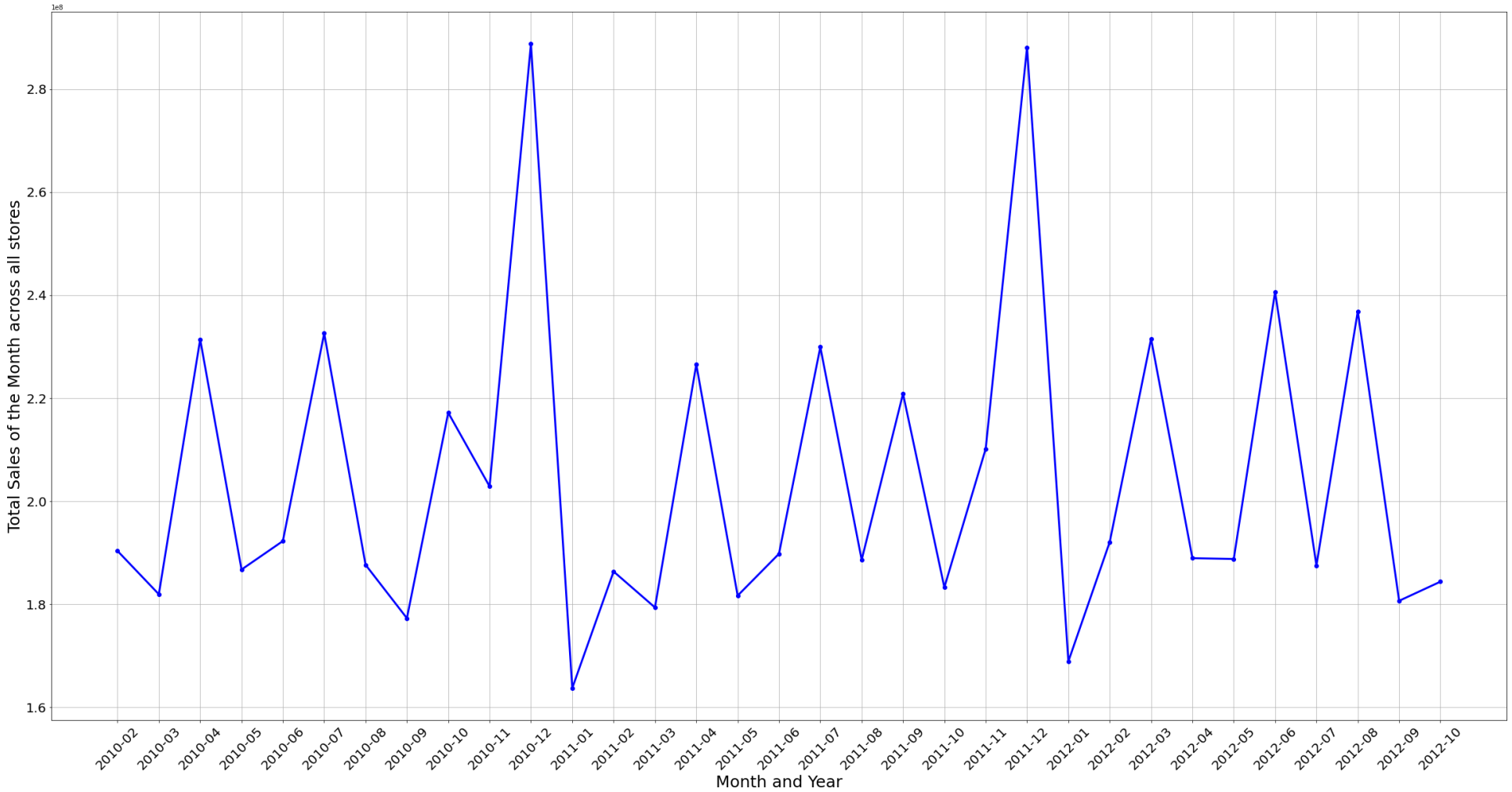
Out[209…
```
33
```

In [210…
```
len(df_monthly['sum'].values)
```

Out[210…
```
33
```

# The following plot shows us the monthly view of sales in units across all stores

In [218…
```
plt.figure(figsize = (40,20))
plt.xlabel("Month and Year", size = 25)
plt.ylabel("Total Sales of the Month across all stores", size = 25)
plt.plot(monthly_index,df_monthly['sum'].values, color = 'b', marker = 'o', linestyle = '-', linewidth = 3)
plt.grid()
_ = plt.xticks(rotation=45,size=20)
_ = plt.yticks(size = 20)
```




In [219…
```
df["semester"] =  (df["Month"].astype(int) - 1) // 6
```

In [227…
```
df
```

Out[227

Out [227...

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 20 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 20 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 20 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 20 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 | 20 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 | 20 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 | 20 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 | 20 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 | 20 |

6435 rows × 13 columns

In [239…
```
df["Semester"]= df["Year"].astype(str) + "S"+ np.where(df.Date.dt.quarter.gt(2),2,1).astype(str)
```

In [240…
```
df
```

Out[240…

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 20 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 20 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 20 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 20 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 20 |
| **…** | … | … | … | … | … | … | … | … | … | … | … | … | |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 | 20 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 | 20 |
| | | 2012- | | | | | | | | | | | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6432** | 45 | 10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 | 20 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 | 20 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 | 20 |

6435 rows × 14 columns

In [232…
```
df['Quarter'][22]
```

Out[232…
```
Period('2010Q3', 'Q-DEC')
```

In [241…
```
df_semester = df.groupby(['Semester'])['Weekly_Sales'].agg(['sum'])
df_semester
```

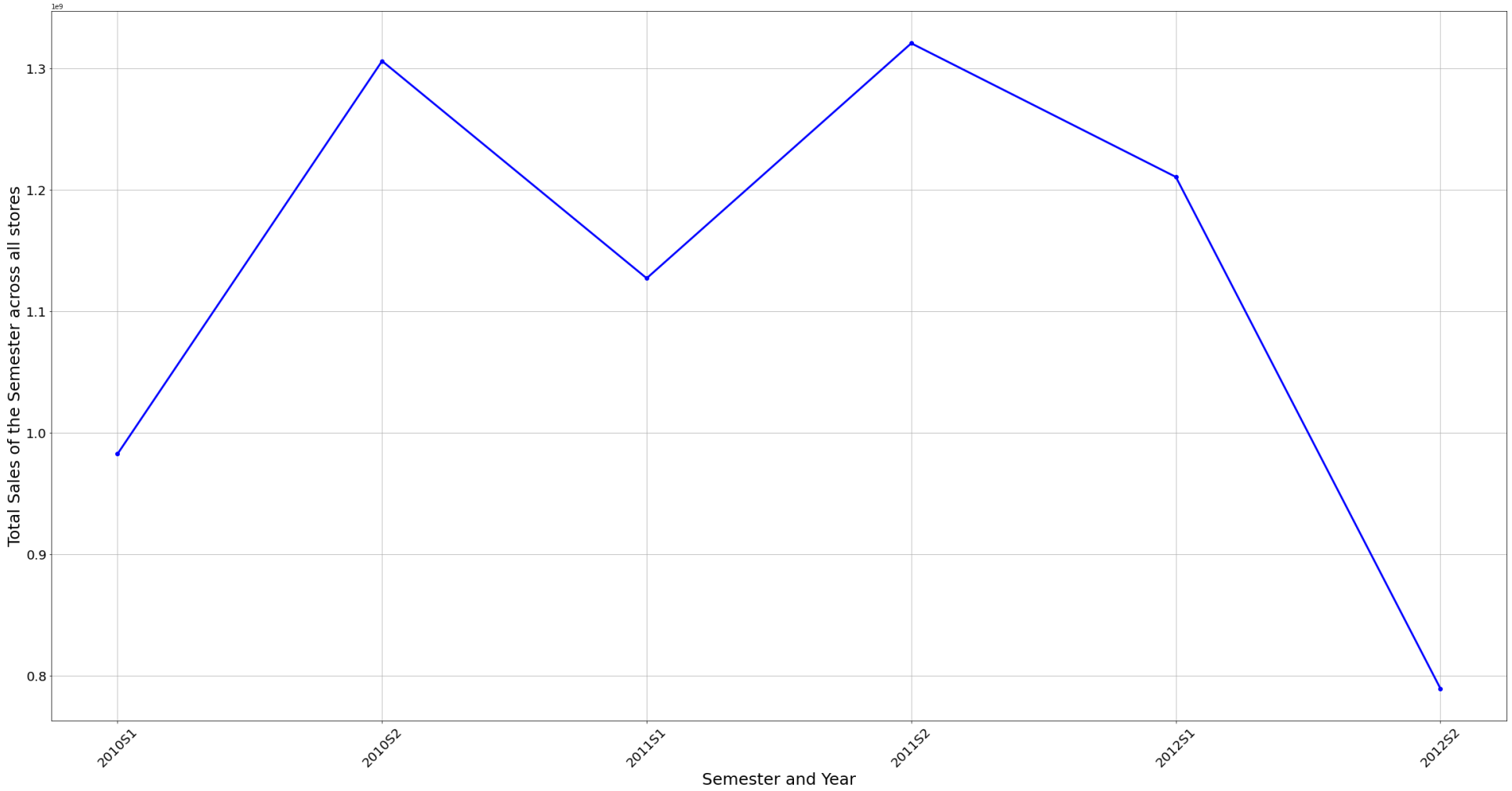
Out[241…

| Semester | sum |
|---|---|
| **2010S1** | 9.826223e+08 |
| **2010S2** | 1.306264e+09 |
| **2011S1** | 1.127340e+09 |
| **2011S2** | 1.320860e+09 |
| **2012S1** | 1.210765e+09 |
| **2012S2** | 7.893674e+08 |

# The following plot shows us the semester view of sales in units across all stores

In [242…

```
plt.figure(figsize = (40,20))
plt.xlabel("Semester and Year", size = 25)
plt.ylabel("Total Sales of the Semester across all stores", size = 25)
plt.plot(df_semester.index,df_semester['sum'].values, color = 'b', marker = 'o', linestyle = '-', linewidth = 3)
plt.grid()
_ = plt.xticks(rotation=45,size=20)
_ = plt.yticks(size = 20)
```




In [ ]:

# Statistical Modelling

In [243...

```
df
```

Out[243...

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 20 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 20 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 20 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 20 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 | 20 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 | 20 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 | 20 |
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 | 20 |
| | | 2012- | | | | | | | | | | | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6434** | 45 | 10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 | 20 |

6435 rows × 14 columns

In [244...
```
df_store1 = df.copy().loc[df['Store'] == 1]
```

In [245...
```
df_store1
```

Out[245...

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 2010 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 2010 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 2010 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 2010 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 2010 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **138** | 1 | 2012-09-28 | 1437059.26 | 0 | 76.08 | 3.666 | 222.981658 | 6.908 | 2012Q3 | 9 | 28 | 2012 | 2012 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **139** | 1 | 2012-10-05 | 1670785.97 | 0 | 68.55 | 3.617 | 223.181477 | 6.573 | 2012Q4 | 10 | 5 | 2012 | 2012 |
| **140** | 1 | 2012-10-12 | 1573072.81 | 0 | 62.99 | 3.601 | 223.381296 | 6.573 | 2012Q4 | 10 | 12 | 2012 | 2012 |
| **141** | 1 | 2012-10-19 | 1508068.77 | 0 | 67.97 | 3.594 | 223.425723 | 6.573 | 2012Q4 | 10 | 19 | 2012 | 2012 |
| **142** | 1 | 2012-10-26 | 1493659.74 | 0 | 69.16 | 3.506 | 223.444251 | 6.573 | 2012Q4 | 10 | 26 | 2012 | 2012 |

143 rows × 14 columns

In [246…
```
X = df_store1[['CPI', 'Unemployment', 'Fuel_Price']]
y = df_store1['Weekly_Sales']
```

In [247…
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
```

In [248…
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train,y_train)
```

Out[248… LinearRegression()

In [249…
```
coeff_df = pd.DataFrame(regressor.coef_, X.columns, columns=['Coefficient'])
coeff_df
```

Out[249…

| | Coefficient |
|---|---|
| **CPI** | 24250.648186 |
| **Unemployment** | 176080.650986 |
| **Fuel_Price** | -72367.009121 |

In [250…

```
y_pred = regressor.predict(X_test)
```

In [251…

```
temp_df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
temp_df
```

Out[251…

| | Actual | Predicted |
|---|---|---|
| **45** | 1891034.93 | 1.499042e+06 |
| **118** | 1611096.05 | 1.567672e+06 |
| **16** | 1432069.95 | 1.489387e+06 |
| **56** | 1636263.41 | 1.510455e+06 |
| **22** | 1546074.18 | 1.499289e+06 |
| **7** | 1404429.92 | 1.546754e+06 |
| **108** | 1688420.76 | 1.586683e+06 |
| **134** | 1582083.40 | 1.543973e+06 |
| **130** | 1631135.79 | 1.551342e+06 |
| **101** | 1459601.17 | 1.590215e+06 |
| **94** | 2033320.66 | 1.648679e+06 |
| **127** | 1527014.04 | 1.562369e+06 |

| | | |
|---|---|---|
| **8** | 1594968.28 | 1.490431e+06 |
| **96** | 1799682.38 | 1.666309e+06 |
| **140** | 1573072.81 | 1.513752e+06 |
| **33** | 1351791.03 | 1.512445e+06 |
| **84** | 1514259.78 | 1.585750e+06 |
| **120** | 1555444.55 | 1.577330e+06 |
| **119** | 1595901.87 | 1.572282e+06 |
| **24** | 1385065.20 | 1.504823e+06 |
| **63** | 1564819.81 | 1.502019e+06 |
| **86** | 1394561.83 | 1.614362e+06 |
| **60** | 1495064.75 | 1.507397e+06 |
| **26** | 1605491.78 | 1.509984e+06 |
| **62** | 1559889.00 | 1.502575e+06 |
| **18** | 1542561.09 | 1.509536e+06 |
| **113** | 1899676.88 | 1.545948e+06 |
| **106** | 1819870.00 | 1.585127e+06 |
| **44** | 1682614.26 | 1.502394e+06 |
| **126** | 1769854.16 | 1.563489e+06 |
| **2** | 1611968.17 | 1.569104e+06 |
| **27** | 1508237.76 | 1.508548e+06 |
| **125** | 1540421.49 | 1.599619e+06 |
| **10** | 1466058.28 | 1.475945e+06 |
| **43** | 1548033.78 | 1.515588e+06 |

| | | |
|---|---|---|
| **139** | 1670785.97 | 1.507749e+06 |
| **54** | 1686842.78 | 1.514085e+06 |
| **95** | 1584083.95 | 1.659303e+06 |
| **51** | 1316899.31 | 1.491134e+06 |
| **85** | 1380020.27 | 1.598138e+06 |
| **110** | 1677472.78 | 1.587979e+06 |
| **93** | 1539483.70 | 1.637476e+06 |
| **90** | 1445249.09 | 1.619656e+06 |

In [252…

```
from sklearn import metrics
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

```
Mean Absolute Error: 117170.84944018991
Mean Squared Error: 22693168190.011814
Root Mean Squared Error: 150642.51786933135
```

In [253…

```
metrics.r2_score(y_test, y_pred)
```

Out[253…

```
-0.010728484446370645
```

## Since the R2 score is negative, this tells us that the model poorly fits the data in consideration. Therefore, we shall consider a different model for Weekly Sales forecasting

# Time Series Modelling

In [254...]
```
df
```

Out[254...]

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment | Quarter | Month | Day | Year | month_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2010-02-05 | 1643690.90 | 0 | 42.31 | 2.572 | 211.096358 | 8.106 | 2010Q1 | 2 | 5 | 2010 | 20 |
| **1** | 1 | 2010-02-12 | 1641957.44 | 1 | 38.51 | 2.548 | 211.242170 | 8.106 | 2010Q1 | 2 | 12 | 2010 | 20 |
| **2** | 1 | 2010-02-19 | 1611968.17 | 0 | 39.93 | 2.514 | 211.289143 | 8.106 | 2010Q1 | 2 | 19 | 2010 | 20 |
| **3** | 1 | 2010-02-26 | 1409727.59 | 0 | 46.63 | 2.561 | 211.319643 | 8.106 | 2010Q1 | 2 | 26 | 2010 | 20 |
| **4** | 1 | 2010-03-05 | 1554806.68 | 0 | 46.50 | 2.625 | 211.350143 | 8.106 | 2010Q1 | 3 | 5 | 2010 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **6430** | 45 | 2012-09-28 | 713173.95 | 0 | 64.88 | 3.997 | 192.013558 | 8.684 | 2012Q3 | 9 | 28 | 2012 | 20 |
| **6431** | 45 | 2012-10-05 | 733455.07 | 0 | 64.89 | 3.985 | 192.170412 | 8.667 | 2012Q4 | 10 | 5 | 2012 | 20 |
| **6432** | 45 | 2012-10-12 | 734464.36 | 0 | 54.47 | 4.000 | 192.327265 | 8.667 | 2012Q4 | 10 | 12 | 2012 | 20 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6433** | 45 | 2012-10-19 | 718125.53 | 0 | 56.47 | 3.969 | 192.330854 | 8.667 | 2012Q4 | 10 | 19 | 2012 | 20 |
| **6434** | 45 | 2012-10-26 | 760281.43 | 0 | 58.85 | 3.882 | 192.308899 | 8.667 | 2012Q4 | 10 | 26 | 2012 | 20 |

6435 rows × 14 columns

In [255…
```
df_time_series = df.copy().loc[df['Store'] == 1][['Date','Weekly_Sales']]
```

In [256…
```
df_time_series
```

Out[256...

| | Date | Weekly_Sales |
|---|---|---|
| **0** | 2010-02-05 | 1643690.90 |
| **1** | 2010-02-12 | 1641957.44 |
| **2** | 2010-02-19 | 1611968.17 |
| **3** | 2010-02-26 | 1409727.59 |
| **4** | 2010-03-05 | 1554806.68 |
| **...** | ... | ... |
| **138** | 2012-09-28 | 1437059.26 |
| **139** | 2012-10-05 | 1670785.97 |
| **140** | 2012-10-12 | 1573072.81 |
| **141** | 2012-10-19 | 1508068.77 |
| **142** | 2012-10-26 | 1493659.74 |

143 rows × 2 columns

In [257...

```
df_time_series = df_time_series.set_index("Date")
```

In [258...

```
df_time_series
```

Out[258...

| Date | Weekly_Sales |
|---|---|
| **2010-02-05** | 1643690.90 |
| **2010-02-12** | 1641957.44 |
| **2010-02-19** | 1611968.17 |
| **2010-02-26** | 1409727.59 |
| **2010-03-05** | 1554806.68 |
| **...** | ... |
| **2012-09-28** | 1437059.26 |
| **2012-10-05** | 1670785.97 |
| **2012-10-12** | 1573072.81 |
| **2012-10-19** | 1508068.77 |
| **2012-10-26** | 1493659.74 |

143 rows × 1 columns

In [259...

```
df_time_series.index.freq = "W-FRI"
```

In [260...

```
df_time_series
```

Out[260…

| Date | Weekly_Sales |
|---|---|
| **2010-02-05** | 1643690.90 |
| **2010-02-12** | 1641957.44 |
| **2010-02-19** | 1611968.17 |
| **2010-02-26** | 1409727.59 |
| **2010-03-05** | 1554806.68 |
| **…** | … |
| **2012-09-28** | 1437059.26 |
| **2012-10-05** | 1670785.97 |
| **2012-10-12** | 1573072.81 |
| **2012-10-19** | 1508068.77 |
| **2012-10-26** | 1493659.74 |

143 rows × 1 columns

In [261…

```
df_time_series.head()
```

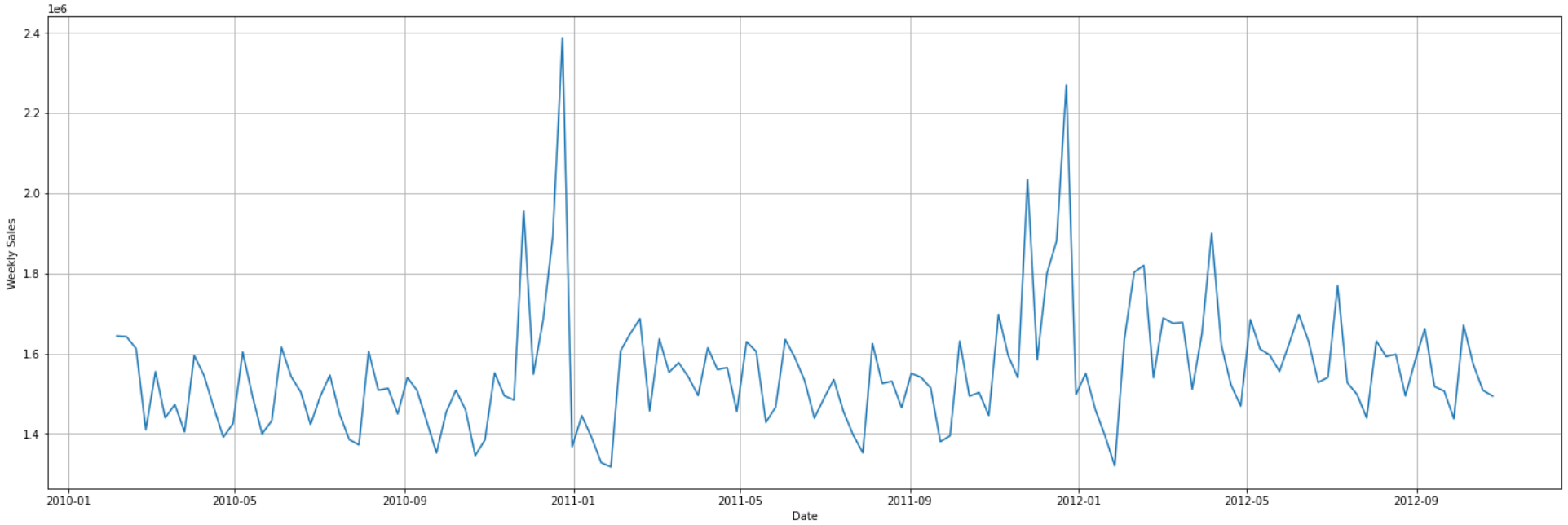
Out[261...

| Date | Weekly_Sales |
|---|---|
| **2010-02-05** | 1643690.90 |
| **2010-02-12** | 1641957.44 |
| **2010-02-19** | 1611968.17 |
| **2010-02-26** | 1409727.59 |
| **2010-03-05** | 1554806.68 |

In [262...

```
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = [25,8]
plt.grid()
plt.plot(df_time_series)
plt.xlabel("Date")
plt.ylabel("Weekly Sales")
```

Out[262…
```
Text(0, 0.5, 'Weekly Sales')
```




In [263…
```
df_time_series.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 143 entries, 2010-02-05 to 2012-10-26
Freq: W-FRI
Data columns (total 1 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Weekly_Sales  143 non-null    float64
dtypes: float64(1)
memory usage: 2.2 KB
```

In [264…
```
df_time_series.describe()
```

Out[264...

| | Weekly_Sales |
|---|---|
| **count** | 1.430000e+02 |
| **mean** | 1.555264e+06 |
| **std** | 1.559808e+05 |
| **min** | 1.316899e+06 |
| **25%** | 1.458105e+06 |
| **50%** | 1.534850e+06 |
| **75%** | 1.614892e+06 |
| **max** | 2.387950e+06 |

# Check for Stationarity : Dickey Fuller Test

## $H0$ : The Data is not Stationary

## $Ha$ : The Data is Stationary

### If p-value of the test is less than 0.05 (5% significance) then reject the null hypothesis and conclude that data is stationary

In [265...

```
from statsmodels.tsa.stattools import adfuller
adfuller(df_time_series)
```

Out[265…
```
(-5.102186145192291,
 1.38777883307592e-05,
 4,
 138,
 {'1%': -3.47864788917503,
  '5%': -2.882721765644168,
  '10%': -2.578065326612056},
 3412.7325502876756)
```

In [266…
```
adfuller(df_time_series, autolag="AIC")
```

Out[266…
```
(-5.102186145192291,
 1.38777883307592e-05,
 4,
 138,
 {'1%': -3.47864788917503,
  '5%': -2.882721765644168,
  '10%': -2.578065326612056},
 3412.7325502876756)
```

In [267…
```
X = df_time_series.values
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
        print('\t%s: %.3f' % (key, value))
```

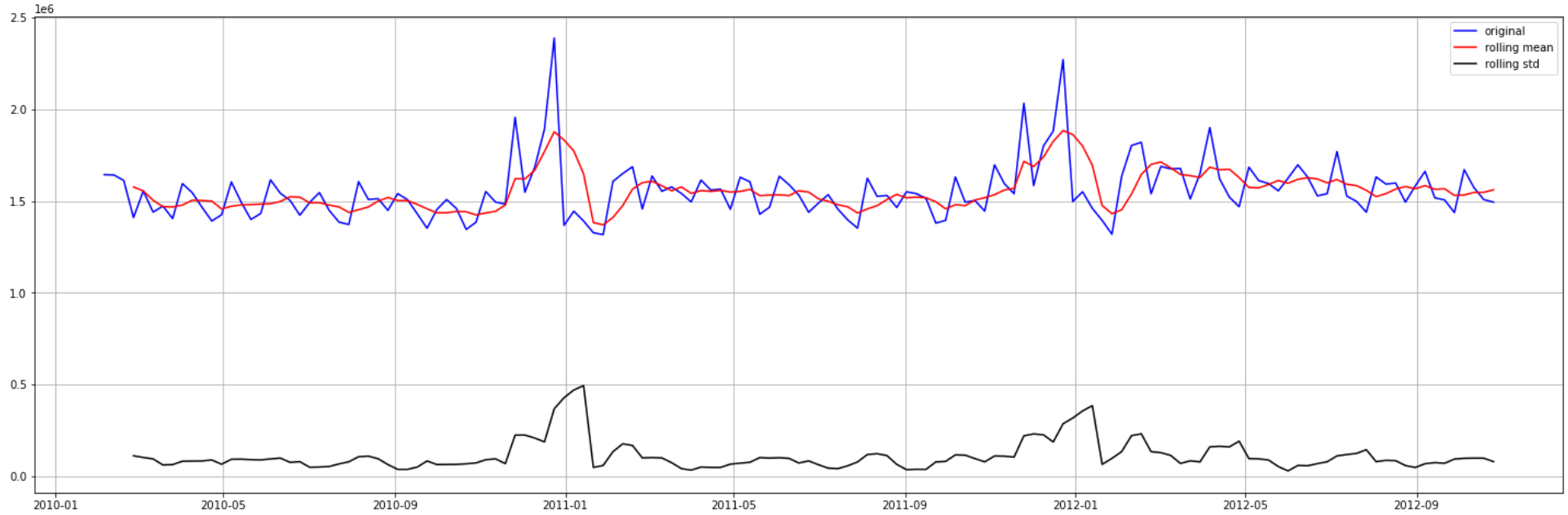
```
ADF Statistic: -5.102186
p-value: 0.000014
Critical Values:
        1%: -3.479
        5%: -2.883
        10%: -2.578
```

## Since the p-value is less than 0.05(significance level (alpha)), we can reject the null hypothesis and conclude that the data is stationary

## Since there are 4 weeks in a month and our data is weekly data, we have chosen a window size of 4

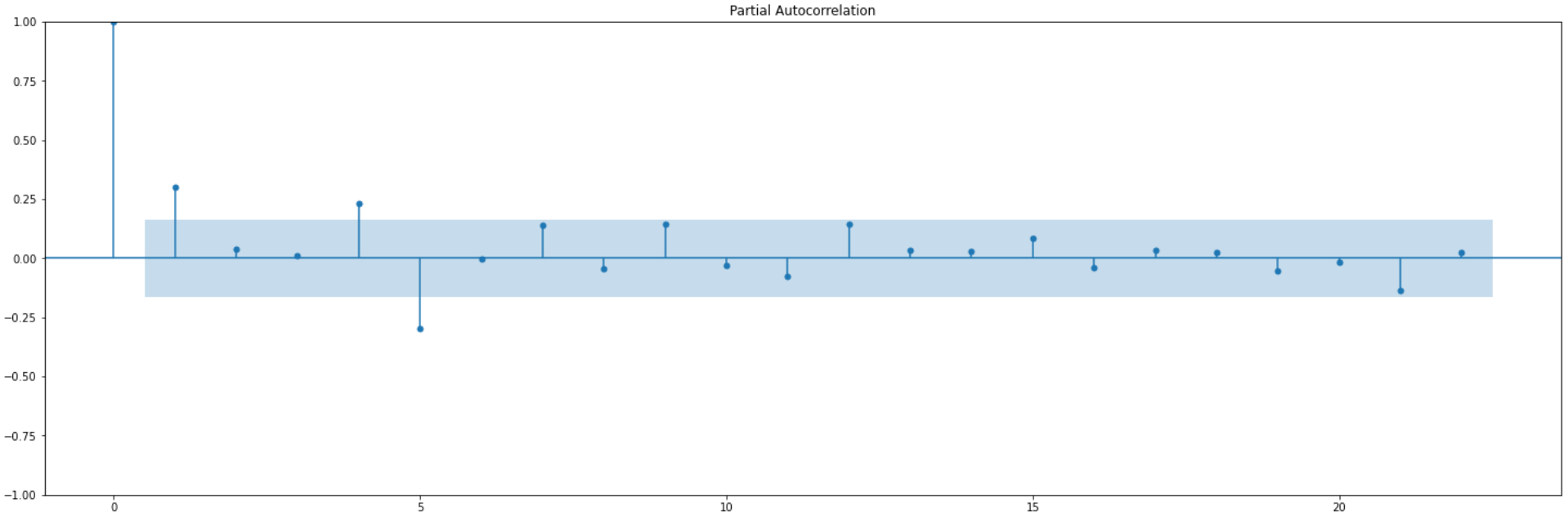
In [268...

```
rollmean = df_time_series.rolling(window = 4).mean()
rollstd = df_time_series.rolling(window = 4).std()
plt.plot(df_time_series, color='blue',label='original')
plt.plot(rollmean, color='red', label='rolling mean')
plt.plot(rollstd, color='black', label = 'rolling std');
plt.grid()
plt.legend();
```



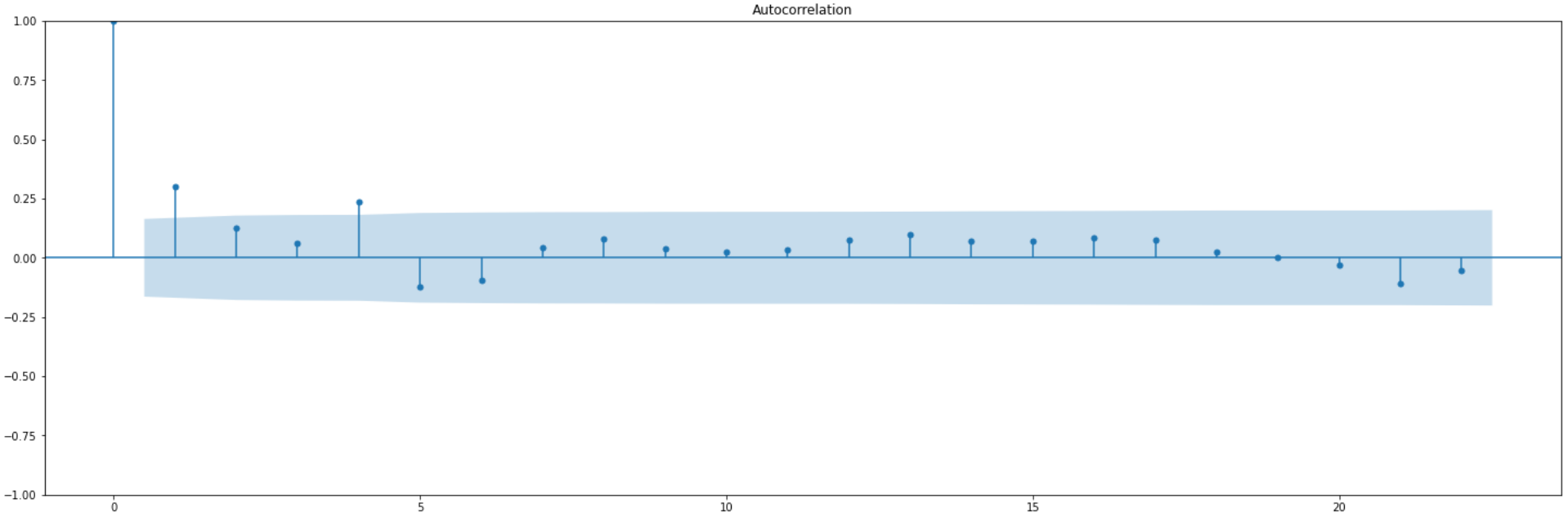
In [269…

```
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
plot_pacf(df_time_series.dropna(), method = 'ywm'); # p = 1
```




In [270…

```
plot_acf(df_time_series.dropna()); # q = 1
```




In [271...
```
train = df_time_series.iloc[:95]
test = df_time_series.iloc[95:]
```

In [272...
```
len(df_time_series)
```

Out[272... 143

# Predicting the forecast for the last 48 weeks (1 year)

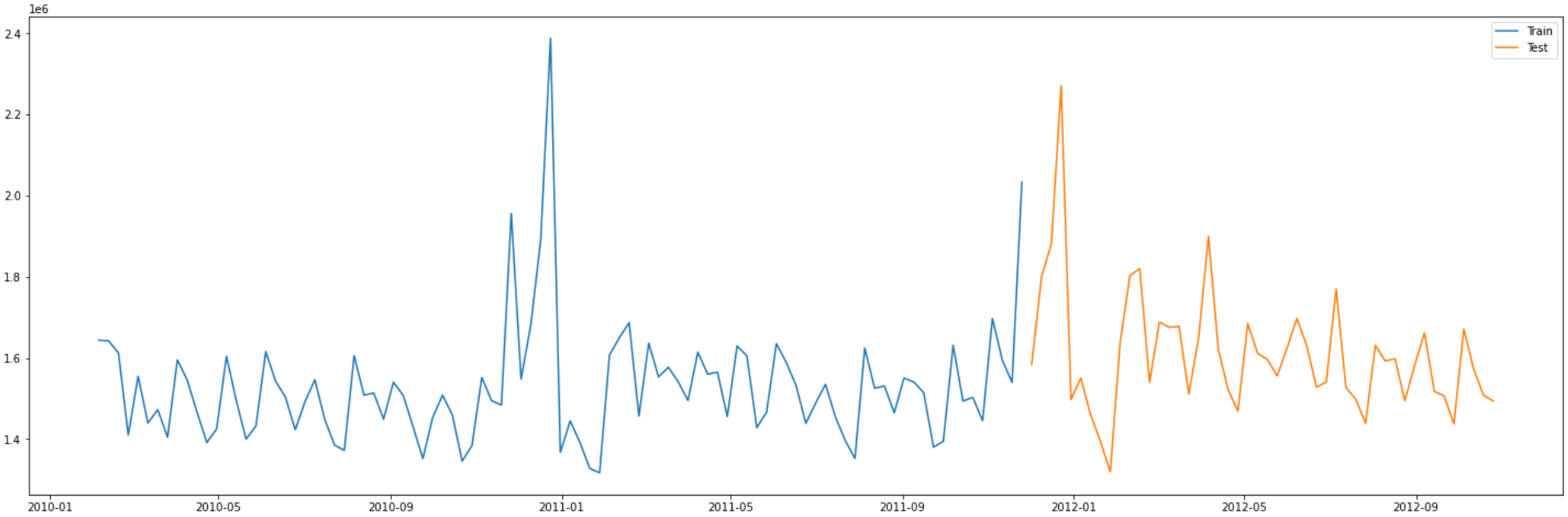
In [273...
```
len(test)
```

Out[273... 48

In [274...
```
len(train)
```

Out[274...
```
95
```

In [275...
```
plt.plot(train, label = 'Train')
plt.plot(test, label = 'Test')
plt.legend();
```



In [276…

```
from statsmodels.tsa.arima.model import ARIMA
# train an  ARIMA (1,0,1) model & get summary

model = ARIMA(train, order = (1,0,1), freq = 'W-FRI')
model_fit = model.fit()
model_fit.summary()
```

Out [276...

SARIMAX Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | Weekly_Sales | **No. Observations:** | 95 |
| **Model:** | ARIMA(1, 0, 1) | **Log Likelihood** | -1263.625 |
| **Date:** | Sat, 09 Apr 2022 | **AIC** | 2535.250 |
| **Time:** | 11:34:43 | **BIC** | 2545.466 |
| **Sample:** | 02-05-2010 | **HQIC** | 2539.378 |
| | - 11-25-2011 | | |
| **Covariance Type:** | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 1.528e+06 | 2.33e+04 | 65.542 | 0.000 | 1.48e+06 | 1.57e+06 |
| **ar.L1** | 0.2435 | 0.736 | 0.331 | 0.741 | -1.200 | 1.687 |
| **ma.L1** | 0.0086 | 0.819 | 0.011 | 0.992 | -1.596 | 1.614 |
| **sigma2** | 2.137e+10 | 0.208 | 1.03e+11 | 0.000 | 2.14e+10 | 2.14e+10 |

| | | | |
|---|---|---|---|
| **Ljung-Box (L1) (Q):** | 0.00 | **Jarque-Bera (JB):** | 378.98 |
| **Prob(Q):** | 0.99 | **Prob(JB):** | 0.00 |
| **Heteroskedasticity (H):** | 2.22 | **Skew:** | 2.17 |
| **Prob(H) (two-sided):** | 0.03 | **Kurtosis:** | 11.77 |

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

[2] Covariance matrix is singular or near-singular, with condition number 1.31e+26. Standard errors may be unstable.

In [277…

```
# forecast for the periods of the test datasets
fcast = model_fit.get_forecast(steps = 100)
fcast # this is a forecast object which cant be output directly
```

Out[277… <statsmodels.tsa.statespace.mlemodel.PredictionResultsWrapper at 0x1483ddcd0>

In [278…

```
# get detailed forecast using summary_fram()
detailed_forecast = fcast.summary_frame()
detailed_forecast
```

Out[278…

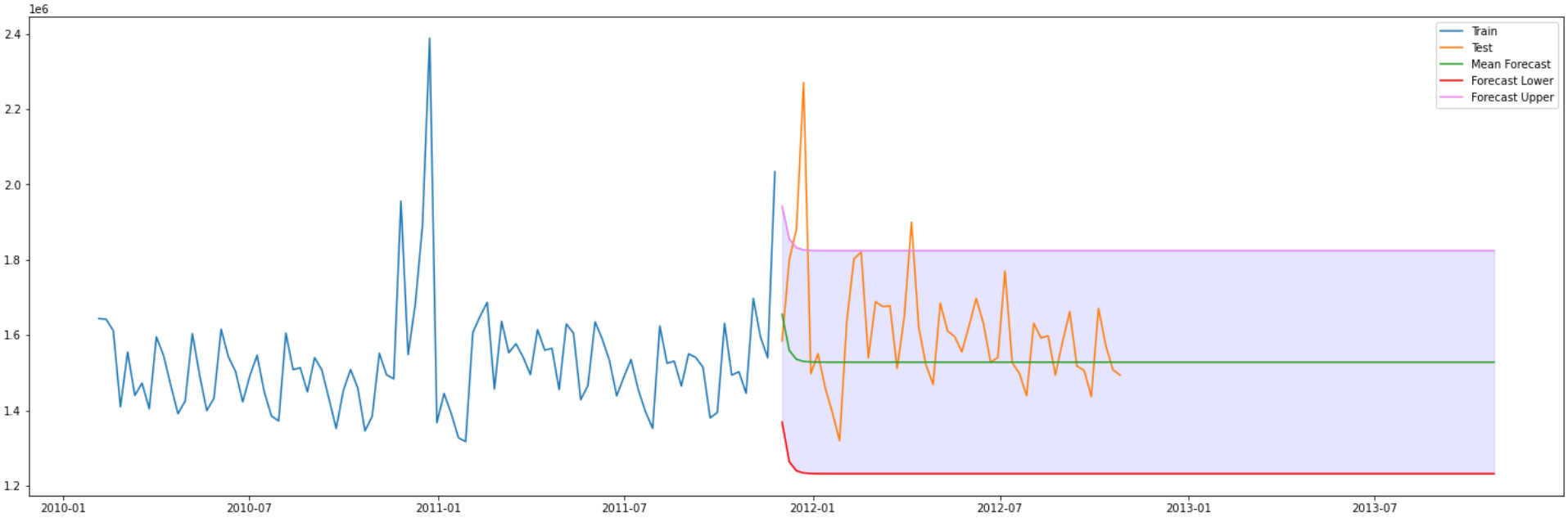
| Weekly_Sales | mean | mean_se | mean_ci_lower | mean_ci_upper |
|---|---|---|---|---|
| **2011-12-02** | 1.655432e+06 | 146200.460667 | 1.368885e+06 | 1.941980e+06 |
| **2011-12-09** | 1.559091e+06 | 150774.760667 | 1.263578e+06 | 1.854605e+06 |
| **2011-12-16** | 1.535635e+06 | 151041.574558 | 1.239599e+06 | 1.831671e+06 |
| **2011-12-23** | 1.529924e+06 | 151057.376420 | 1.233857e+06 | 1.825991e+06 |
| **2011-12-30** | 1.528533e+06 | 151058.313099 | 1.232465e+06 | 1.824602e+06 |
| **…** | … | … | … | … |
| **2013-09-27** | 1.528086e+06 | 151058.372124 | 1.232017e+06 | 1.824155e+06 |
| **2013-10-04** | 1.528086e+06 | 151058.372124 | 1.232017e+06 | 1.824155e+06 |
| **2013-10-11** | 1.528086e+06 | 151058.372124 | 1.232017e+06 | 1.824155e+06 |
| **2013-10-18** | 1.528086e+06 | 151058.372124 | 1.232017e+06 | 1.824155e+06 |
| **2013-10-25** | 1.528086e+06 | 151058.372124 | 1.232017e+06 | 1.824155e+06 |

100 rows × 4 columns

In [279…
```
x = np.array(detailed_forecast.index)
```

In [280…
```
y1 = np.array(detailed_forecast['mean_ci_lower'])
y2 = np.array(detailed_forecast['mean_ci_upper'])
```

In [281…
```
plt.plot(train, label = 'Train')
plt.plot(test, label = 'Test')
plt.plot(detailed_forecast['mean'],label = 'Mean Forecast')
plt.plot(detailed_forecast['mean_ci_lower'],color = 'red',label = 'Forecast Lower')
plt.plot(detailed_forecast['mean_ci_upper'],color = 'violet',label = 'Forecast Upper')
plt.fill_between(x,y1,y2,color = 'b',alpha = 0.1)
plt.legend();
```



## The above plot is able to show a forecast of the data and predicts the mean, upper bound and lower bound prediction with a 95% confidence interval

In [ ]: