```
In [1]:   # import the required libraries
          import pandas as pd

In [28]:  # reading the csv file
          df= pd.read_csv("D:\\1577429980_walmart_store_sales unzip\\Walmart_Store_sales.csv")

In [15]:  # checking the df
          df
```

Out[15]:

	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 [19]:  # checking the initial headings of the dataset
          df.head(5)
```

Out[19]:

	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

```
In [21]:  # checking the last 5 elements of the dataset
          df.tail()
```

Out[21]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
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

```
In [23]:  # checking the shape of the dataset
          df.shape
```

Out[23]: (6435, 8)

```
In [24]:  # checking the information present in the dataset
          df.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
```

```
In [25]:  # checking the numerical contents of the dataset
```

```python
df.describe()
```

Out[25]:

	Store	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
count	6435.000000	6.435000e+03	6435.000000	6435.000000	6435.000000	6435.000000	6435.000000
mean	23.000000	1.046965e+06	0.069930	60.663782	3.358607	171.578394	7.999151
std	12.988182	5.643666e+05	0.255049	18.444933	0.459020	39.356712	1.875885
min	1.000000	2.099862e+05	0.000000	-2.060000	2.472000	126.064000	3.879000
25%	12.000000	5.533501e+05	0.000000	47.460000	2.933000	131.735000	6.891000
50%	23.000000	9.607460e+05	0.000000	62.670000	3.445000	182.616521	7.874000
75%	34.000000	1.420159e+06	0.000000	74.940000	3.735000	212.743293	8.622000
max	45.000000	3.818686e+06	1.000000	100.140000	4.468000	227.232807	14.313000

```python
In [40]: #  checking the unique values present in the column temperature
df ['Temperature'].nunique()
```

Out[40]: 3528

```python
In [44]: # checking the nunique values that means all the unique values of the elements
df.nunique()
```

Out[44]:
```
Store             45
Date             143
Weekly_Sales    6435
Holiday_Flag       2
Temperature     3528
Fuel_Price       892
CPI             2145
Unemployment     349
dtype: int64
```

```python
In [47]: # checking the value counts of the element called store
df['Store'].value_counts()
```

Out[47]:
```
1     143
24    143
26    143
27    143
28    143
29    143
30    143
31    143
32    143
33    143
34    143
35    143
36    143
37    143
38    143
39    143
40    143
41    143
42    143
43    143
44    143
25    143
23    143
2     143
22    143
3     143
4     143
5     143
6     143
7     143
8     143
9     143
10    143
11    143
12    143
13    143
14    143
15    143
16    143
17    143
18    143
19    143
20    143
21    143
45    143
Name: Store, dtype: int64
```

```python
In [48]: # checking the value counts of the dataset
df.value_counts()
```

```
Out[48]:    Store  Date        Weekly_Sales  Holiday_Flag  Temperature  Fuel_Price  CPI         Unemployment
            1      01-04-2011  1495064.75    0             59.17        3.524       214.837166  7.682        1
            30     30-09-2011  387001.13     0             78.91        3.355       216.362033  7.852        1
            31     02-07-2010  1311704.92    0             82.29        2.669       210.880373  8.099        1
                   02-04-2010  1357600.68    0             64.12        2.719       210.479887  8.200        1
                   02-03-2012  1427881.22    0             59.30        3.630       220.486689  7.057        1
                                                                                                            ..
            15     30-12-2011  603460.79     1             31.44        3.566       136.643258  7.866        1
                   30-09-2011  521297.31     0             64.87        3.858       136.419500  7.806        1
                   30-07-2010  619224.06     0             72.04        2.932       132.598387  8.099        1
                   30-04-2010  570791.11     0             49.09        3.042       132.064433  8.185        1
            45     31-12-2010  679156.20     1             29.67        3.179       182.571448  8.724        1
            Length: 6435, dtype: int64
```

In [49]: `# checking the sum of the null values present in the dataset`
`df.isna().sum()`

```
Out[49]:  Store           0
          Date            0
          Weekly_Sales    0
          Holiday_Flag    0
          Temperature     0
          Fuel_Price      0
          CPI             0
          Unemployment    0
          dtype: int64
```

In [50]: `# checking the nunique values of the dataset`
`df.nunique()`

```
Out[50]:  Store             45
          Date             143
          Weekly_Sales    6435
          Holiday_Flag       2
          Temperature     3528
          Fuel_Price       892
          CPI             2145
          Unemployment     349
          dtype: int64
```

In [52]: `# creating the list for the element called store and date`
`df[['Store','Date']]`

Out[52]:

	Store	Date
0	1	05-02-2010
1	1	12-02-2010
2	1	19-02-2010
3	1	26-02-2010
4	1	05-03-2010
...	...	...
6430	45	28-09-2012
6431	45	05-10-2012
6432	45	12-10-2012
6433	45	19-10-2012
6434	45	26-10-2012

6435 rows × 2 columns

In [53]: `# creating the list for the element called store and date`
`# store it in the new variable called new data`
`new_data=df[['Store','Date']]`

In [54]: `# checking the type of new data`
`type(new_data)`

Out[54]: `pandas.core.frame.DataFrame`

In [55]: `# printing the new data`
`new_data`

	Store	Date
0	1	05-02-2010
1	1	12-02-2010
2	1	19-02-2010
3	1	26-02-2010
4	1	05-03-2010
...	...	...
6430	45	28-09-2012
6431	45	05-10-2012
6432	45	12-10-2012
6433	45	19-10-2012
6434	45	26-10-2012

6435 rows × 2 columns

In [60]:
```python
# checking for the columns present in the dataset
df .columns
```

Out[60]:
```
Index(['Store', 'Date', 'Weekly_Sales', 'Holiday_Flag', 'Temperature',
       'Fuel_Price', 'CPI', 'Unemployment'],
      dtype='object')
```

In [65]:
```python
# checking for the elements fro 0 to 7 in the dataset
df.loc[0:7]
```

Out[65]:

	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

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js