

**Author : Sanjoy Biswas**

**Topic : Pandas Tutorial: DataFrames in Python**

**Email : sanjoy.eee32@gmail.com**

In [3]:

```
import pandas as pd
import numpy as np
```

## missing data

In [5]:

```
d = {
    'A': [1,2, np.nan],
    'B': [5, np.nan, np.nan],
    'C': [1,2,3]
}
```

In [6]:

```
df = pd.DataFrame(d)
```

In [7]:

```
df
```

Out[7]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |
| 1 | 2.0 | NaN | 2 |
| 2 | NaN | NaN | 3 |

## drop nan method

In [10]:

```
df.dropna()
```

Out[10]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |

In [11]:

```
df.dropna(axis = 0)
```

Out[11]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |

In [13]:

```
df.dropna??
```

In [14]:

```
df.dropna(axis = 1)
```

Out[14]:

|   | C |
|---|---|
| 0 | 1 |
| 1 | 2 |
| 2 | 3 |

## filling value

In [15]:

```
df
```

Out[15]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |
| 1 | 2.0 | NaN | 2 |
| 2 | NaN | NaN | 3 |

In [16]:

```
df.fillna(value = 'Filling Value')
```

Out[16]:

|   | A             | B             | C |
|---|---------------|---------------|---|
| 0 | 1             | 5             | 1 |
| 1 | 2             | Filling Value | 2 |
| 2 | Filling Value | Filling Value | 3 |

In [19]:

```
df = pd.DataFrame(d)
```

In [22]:

```
a = df['A'].fillna(value = df['A'].mean())
```

In [23]:

```
df
```

Out[23]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |
| 1 | 2.0 | NaN | 2 |
| 2 | NaN | NaN | 3 |

```

A    B    C

```

In [24]:

```
a
```

Out[24]:

```
0    1.0
1    2.0
2    1.5
Name: A, dtype: float64
```

In [25]:

```
df
```

Out[25]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |
| 1 | 2.0 | NaN | 2 |
| 2 | NaN | NaN | 3 |

In [26]:

```
df['A'].fillna(value = df['A'].mean())
```

Out[26]:

```
0    1.0
1    2.0
2    1.5
Name: A, dtype: float64
```

In [27]:

```
df
```

Out[27]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |
| 1 | 2.0 | NaN | 2 |
| 2 | NaN | NaN | 3 |

In [28]:

```
df.fillna??
```

In [29]:

```
df['A'].fillna(value = df['A'].mean(), inplace = True)
```

In [30]:

```
df
```

Out[30]:

|   | A   | B   | C |
|---|-----|-----|---|
| 0 | 1.0 | 5.0 | 1 |

|   |     |     |   |
|---|-----|-----|---|
| 1 | 2A  | NaB | Q |
| 2 | 1.5 | NaN | 3 |

## Group By

In [31]:

```
data = {  
    'Company': ['Google', 'Google', 'MSFT', 'FB', 'FB', 'IBM'],  
    'Person': ['Sam', 'Nihad', 'Any', 'Van', 'Rakib', 'Ovi'],  
    'Sales': [200, 120, 340, 124, 243, 350]  
}
```

In [32]:

```
df = pd.DataFrame(data)
```

In [33]:

```
df
```

Out[33]:

|   | Company | Person | Sales |
|---|---------|--------|-------|
| 0 | Google  | Sam    | 200   |
| 1 | Google  | Nihad  | 120   |
| 2 | MSFT    | Any    | 340   |
| 3 | FB      | Van    | 124   |
| 4 | FB      | Rakib  | 243   |
| 5 | IBM     | Ovi    | 350   |

In [34]:

```
byComp = df.groupby('Company')
```

In [35]:

```
byComp.mean()
```

Out[35]:

|         | Sales |
|---------|-------|
| Company |       |
| FB      | 183.5 |
| Google  | 160.0 |
| IBM     | 350.0 |
| MSFT    | 340.0 |

In [36]:

```
byComp.sum()
```

Out[36]:

|         | Sales |
|---------|-------|
| Company |       |
| FB      | 367   |

|         |     |
|---------|-----|
| Google  | 350 |
| Company | 350 |
| MSFT    | 340 |

In [37]:

```
df
```

Out[37]:

|   | Company | Person | Sales |
|---|---------|--------|-------|
| 0 | Google  | Sam    | 200   |
| 1 | Google  | Nihad  | 120   |
| 2 | MSFT    | Any    | 340   |
| 3 | FB      | Van    | 124   |
| 4 | FB      | Rakib  | 243   |
| 5 | IBM     | Ovi    | 350   |

In [38]:

```
byComp.std()
```

Out[38]:

|         | Sales     |
|---------|-----------|
| Company |           |
| FB      | 84.145707 |
| Google  | 56.568542 |
| IBM     | NaN       |
| MSFT    | NaN       |

In [39]:

```
df.groupby('Company').count()
```

Out[39]:

|         | Person | Sales |
|---------|--------|-------|
| Company |        |       |
| FB      | 2      | 2     |
| Google  | 2      | 2     |
| IBM     | 1      | 1     |
| MSFT    | 1      | 1     |

In [40]:

```
df
```

Out[40]:

|   | Company | Person | Sales |
|---|---------|--------|-------|
| 0 | Google  | Sam    | 200   |
| 1 | Google  | Nihad  | 120   |
| 2 | MSFT    | Any    | 340   |
| 3 | FB      | Van    | 124   |

| 4 | Company | Periskib | Sales |
|---|---------|----------|-------|
| 5 | IBM     | Ovi      | 350   |

In [41]:

```
df.groupby('Company').describe()
```

Out[41]:

|         | Sales |       |           |       |        |       |        |       |  |
|---------|-------|-------|-----------|-------|--------|-------|--------|-------|--|
|         | count | mean  | std       | min   | 25%    | 50%   | 75%    | max   |  |
| Company |       |       |           |       |        |       |        |       |  |
| FB      | 2.0   | 183.5 | 84.145707 | 124.0 | 153.75 | 183.5 | 213.25 | 243.0 |  |
| Google  | 2.0   | 160.0 | 56.568542 | 120.0 | 140.00 | 160.0 | 180.00 | 200.0 |  |
| IBM     | 1.0   | 350.0 | NaN       | 350.0 | 350.00 | 350.0 | 350.00 | 350.0 |  |
| MSFT    | 1.0   | 340.0 | NaN       | 340.0 | 340.00 | 340.0 | 340.00 | 340.0 |  |

## Merging, Joining and Concatenating

### Concatenating

In [42]:

```
df1 = pd.DataFrame(np.random.randint(1, 20, size = (4, 3)), [1,2,3,4], ['W', 'X', 'Y'])
```

In [51]:

```
df2 = pd.DataFrame(np.random.randint(20,40, size = (4, 3)), [5,6,7,8], ['W', 'X', 'Y'])
```

In [52]:

```
df3 = pd.DataFrame(np.random.randint(40,60, size = (4, 3)), [9,10,11,12], ['W', 'X', 'Y'])
```

In [53]:

```
df1
```

Out[53]:

|   | W  | X  | Y  |
|---|----|----|----|
| 1 | 9  | 10 | 19 |
| 2 | 14 | 15 | 8  |
| 3 | 6  | 13 | 6  |
| 4 | 4  | 11 | 14 |

In [54]:

```
df2
```

Out[54]:

|   | W  | X  | Y  |
|---|----|----|----|
| 5 | 32 | 26 | 24 |
| 6 | 25 | 35 | 20 |
| 7 | 29 | 36 | 25 |

8 23 27 29  
~~W~~ ~~X~~ ~~Y~~

In [55]:

```
df3
```

Out[55]:

|    | W  | X  | Y  |
|----|----|----|----|
| 9  | 52 | 50 | 47 |
| 10 | 54 | 53 | 40 |
| 11 | 57 | 53 | 51 |
| 12 | 43 | 58 | 47 |

In [56]:

```
farme = [df1, df2, df3]
```

In [57]:

```
pd.concat(farme)
```

Out[57]:

|    | W  | X  | Y  |
|----|----|----|----|
| 1  | 9  | 10 | 19 |
| 2  | 14 | 15 | 8  |
| 3  | 6  | 13 | 6  |
| 4  | 4  | 11 | 14 |
| 5  | 32 | 26 | 24 |
| 6  | 25 | 35 | 20 |
| 7  | 29 | 36 | 25 |
| 8  | 23 | 27 | 29 |
| 9  | 52 | 50 | 47 |
| 10 | 54 | 53 | 40 |
| 11 | 57 | 53 | 51 |
| 12 | 43 | 58 | 47 |

In [58]:

```
pd.concat(farme, axis = 1)
```

Out[58]:

|    | W    | X    | Y    | W    | X    | Y    | W    | X    | Y    |
|----|------|------|------|------|------|------|------|------|------|
| 1  | 9.0  | 10.0 | 19.0 | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |
| 2  | 14.0 | 15.0 | 8.0  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |
| 3  | 6.0  | 13.0 | 6.0  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |
| 4  | 4.0  | 11.0 | 14.0 | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |
| 5  | NaN  | NaN  | NaN  | 32.0 | 26.0 | 24.0 | NaN  | NaN  | NaN  |
| 6  | NaN  | NaN  | NaN  | 25.0 | 35.0 | 20.0 | NaN  | NaN  | NaN  |
| 7  | NaN  | NaN  | NaN  | 29.0 | 36.0 | 25.0 | NaN  | NaN  | NaN  |
| 8  | NaN  | NaN  | NaN  | 23.0 | 27.0 | 29.0 | NaN  | NaN  | NaN  |
| 9  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | 52.0 | 50.0 | 47.0 |
| 10 | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | 54.0 | 53.0 | 40.0 |

|    |     |     |     |     |     |     |      |      |      |
|----|-----|-----|-----|-----|-----|-----|------|------|------|
| 11 | NaN | NaN | NaN | NaN | NaN | NaN | 57.0 | 53.0 | 51.0 |
| 12 | NaN | NaN | NaN | NaN | NaN | NaN | 43.0 | 58.0 | 47.0 |

## Merging

In [59]:

```
left = {
    'key': ['k0', 'k1', 'k2', 'k3'],
    'A': [10, 20, 30, 40],
    'B': [1, 2, 3, 4]
}
```

In [60]:

```
right = {
    'key': ['k0', 'k1', 'k2', 'k3'],
    'C': [100, 200, 300, 400],
    'D': [11, 12, 13, 14]
}
```

In [61]:

```
left = pd.DataFrame(left)
```

In [62]:

```
right = pd.DataFrame(right)
```

In [63]:

```
left
```

Out[63]:

|   | key | A  | B |
|---|-----|----|---|
| 0 | k0  | 10 | 1 |
| 1 | k1  | 20 | 2 |
| 2 | k2  | 30 | 3 |
| 3 | k3  | 40 | 4 |

In [64]:

```
right
```

Out[64]:

|   | key | C   | D  |
|---|-----|-----|----|
| 0 | k0  | 100 | 11 |
| 1 | k1  | 200 | 12 |
| 2 | k2  | 300 | 13 |
| 3 | k3  | 400 | 14 |

In [65]:

```
pd.merge(left, right, on='key')
```

Out[65]:



|   | key | A  | B | C   | D  |
|---|-----|----|---|-----|----|
| 0 | k0  | 10 | 1 | 100 | 11 |
| 1 | k1  | 20 | 2 | 200 | 12 |
| 2 | k2  | 30 | 3 | 300 | 13 |
| 3 | k3  | 40 | 4 | 400 | 14 |

In [66]:

```
pd.merge??
```

## Joining

In [67]:

```
left = {
    'A': ['A0', 'A1', 'A2'],
    'B': ['B0', 'B1', 'B2']
}
```

In [68]:

```
right = {
    'C': ['C0', 'C2', 'C3'],
    'D': ['D0', 'D2', 'D3']
}
```

In [69]:

```
left = pd.DataFrame(left, index = ['k0', 'k1', 'k2'])
```

In [70]:

```
right = pd.DataFrame(right, index = ['k0', 'k2', 'k3'])
```

In [71]:

```
left
```

Out[71]:

|    | A  | B  |
|----|----|----|
| k0 | A0 | B0 |
| k1 | A1 | B1 |
| k2 | A2 | B2 |

In [72]:

```
right
```

Out[72]:

|    | C  | D  |
|----|----|----|
| k0 | C0 | D0 |
| k2 | C2 | D2 |
| k3 | C3 | D3 |

In [76]:

```
left.join(right, how = 'inner')
```

Out[76]:

|    | A  | B  | C  | D  |
|----|----|----|----|----|
| k0 | A0 | B0 | C0 | D0 |
| k2 | A2 | B2 | C2 | D2 |

In [74]:

```
left.join??
```

## Operations

In [77]:

```
df = pd.DataFrame({
    'col1': [444, 555, 666, 444],
    'col2': [1, 2, 3, 4],
    'col3': ['abc', 'def', 'ghi', 'xyz']
})
```

In [78]:

```
df
```

Out[78]:

|   | col1 | col2 | col3 |
|---|------|------|------|
| 0 | 444  | 1    | abc  |
| 1 | 555  | 2    | def  |
| 2 | 666  | 3    | ghi  |
| 3 | 444  | 4    | xyz  |

In [79]:

```
df['col1'].unique()
```

Out[79]:

```
array([444, 555, 666], dtype=int64)
```

In [80]:

```
df['col1'].nunique()
```

Out[80]:

```
3
```

In [82]:

```
df['col1'].value_counts()
```

Out[82]:

```
444    2
555    1
666    1
Name: col1, dtype: int64
```

# Apply Method

In [84]:

```
def times2(x):  
    return 2*x
```

In [85]:

```
df['col2'].apply(times2)
```

Out[85]:

```
0    2  
1    4  
2    6  
3    8  
Name: col2, dtype: int64
```

In [87]:

```
df['col2'].apply(lambda x: x**2)
```

Out[87]:

```
0     1  
1     4  
2     9  
3    16  
Name: col2, dtype: int64
```

# Reading File

In [88]:

```
df = pd.read_excel('SampleData.xlsx', sheet_name = 'SalesOrders')
```

In [89]:

```
df
```

Out[89]:

|    | OrderDate  | Region  | Rep      | Item    | Units | Unit Cost | Total   |
|----|------------|---------|----------|---------|-------|-----------|---------|
| 0  | 2019-01-06 | East    | Jones    | Pencil  | 95    | 1.99      | 189.05  |
| 1  | 2019-01-23 | Central | Kivell   | Binder  | 50    | 19.99     | 999.50  |
| 2  | 2019-02-09 | Central | Jardine  | Pencil  | 36    | 4.99      | 179.64  |
| 3  | 2019-02-26 | Central | Gill     | Pen     | 27    | 19.99     | 539.73  |
| 4  | 2019-03-15 | West    | Sorvino  | Pencil  | 56    | 2.99      | 167.44  |
| 5  | 2019-04-01 | East    | Jones    | Binder  | 60    | 4.99      | 299.40  |
| 6  | 2019-04-18 | Central | Andrews  | Pencil  | 75    | 1.99      | 149.25  |
| 7  | 2019-05-05 | Central | Jardine  | Pencil  | 90    | 4.99      | 449.10  |
| 8  | 2019-05-22 | West    | Thompson | Pencil  | 32    | 1.99      | 63.68   |
| 9  | 2019-06-08 | East    | Jones    | Binder  | 60    | 8.99      | 539.40  |
| 10 | 2019-06-25 | Central | Morgan   | Pencil  | 90    | 4.99      | 449.10  |
| 11 | 2019-07-12 | East    | Howard   | Binder  | 29    | 1.99      | 57.71   |
| 12 | 2019-07-29 | East    | Parent   | Binder  | 81    | 19.99     | 1619.19 |
| 13 | 2019-08-15 | East    | Jones    | Pencil  | 35    | 4.99      | 174.65  |
| 14 | 2019-09-01 | Central | Smith    | Desk    | 2     | 125.00    | 250.00  |
| 15 | 2019-09-18 | East    | Jones    | Pen Set | 16    | 15.99     | 255.84  |

|    |            |         |          |         |    |        |         |
|----|------------|---------|----------|---------|----|--------|---------|
| 16 | 2019-10-05 | Central | Morgan   | Binder  | 28 | 8.99   | 251.72  |
| 17 | 2019-10-22 | East    | Jones    | Pen     | 64 | 8.99   | 575.36  |
| 18 | 2019-11-08 | East    | Parent   | Pen     | 15 | 19.99  | 299.85  |
| 19 | 2019-11-25 | Central | Kivell   | Pen Set | 96 | 4.99   | 479.04  |
| 20 | 2019-12-12 | Central | Smith    | Pencil  | 67 | 1.29   | 86.43   |
| 21 | 2019-12-29 | East    | Parent   | Pen Set | 74 | 15.99  | 1183.26 |
| 22 | 2020-01-15 | Central | Gill     | Binder  | 46 | 8.99   | 413.54  |
| 23 | 2020-02-01 | Central | Smith    | Binder  | 87 | 15.00  | 1305.00 |
| 24 | 2020-02-18 | East    | Jones    | Binder  | 4  | 4.99   | 19.96   |
| 25 | 2020-03-07 | West    | Sorvino  | Binder  | 7  | 19.99  | 139.93  |
| 26 | 2020-03-24 | Central | Jardine  | Pen Set | 50 | 4.99   | 249.50  |
| 27 | 2020-04-10 | Central | Andrews  | Pencil  | 66 | 1.99   | 131.34  |
| 28 | 2020-04-27 | East    | Howard   | Pen     | 96 | 4.99   | 479.04  |
| 29 | 2020-05-14 | Central | Gill     | Pencil  | 53 | 1.29   | 68.37   |
| 30 | 2020-05-31 | Central | Gill     | Binder  | 80 | 8.99   | 719.20  |
| 31 | 2020-06-17 | Central | Kivell   | Desk    | 5  | 125.00 | 625.00  |
| 32 | 2020-07-04 | East    | Jones    | Pen Set | 62 | 4.99   | 309.38  |
| 33 | 2020-07-21 | Central | Morgan   | Pen Set | 55 | 12.49  | 686.95  |
| 34 | 2020-08-07 | Central | Kivell   | Pen Set | 42 | 23.95  | 1005.90 |
| 35 | 2020-08-24 | West    | Sorvino  | Desk    | 3  | 275.00 | 825.00  |
| 36 | 2020-09-10 | Central | Gill     | Pencil  | 7  | 1.29   | 9.03    |
| 37 | 2020-09-27 | West    | Sorvino  | Pen     | 76 | 1.99   | 151.24  |
| 38 | 2020-10-14 | West    | Thompson | Binder  | 57 | 19.99  | 1139.43 |
| 39 | 2020-10-31 | Central | Andrews  | Pencil  | 14 | 1.29   | 18.06   |
| 40 | 2020-11-17 | Central | Jardine  | Binder  | 11 | 4.99   | 54.89   |
| 41 | 2020-12-04 | Central | Jardine  | Binder  | 94 | 19.99  | 1879.06 |
| 42 | 2020-12-21 | Central | Andrews  | Binder  | 28 | 4.99   | 139.72  |

In [91]:

```
df.head(10)
```

Out[91]:

|   | OrderDate  | Region  | Rep      | Item   | Units | Unit Cost | Total  |
|---|------------|---------|----------|--------|-------|-----------|--------|
| 0 | 2019-01-06 | East    | Jones    | Pencil | 95    | 1.99      | 189.05 |
| 1 | 2019-01-23 | Central | Kivell   | Binder | 50    | 19.99     | 999.50 |
| 2 | 2019-02-09 | Central | Jardine  | Pencil | 36    | 4.99      | 179.64 |
| 3 | 2019-02-26 | Central | Gill     | Pen    | 27    | 19.99     | 539.73 |
| 4 | 2019-03-15 | West    | Sorvino  | Pencil | 56    | 2.99      | 167.44 |
| 5 | 2019-04-01 | East    | Jones    | Binder | 60    | 4.99      | 299.40 |
| 6 | 2019-04-18 | Central | Andrews  | Pencil | 75    | 1.99      | 149.25 |
| 7 | 2019-05-05 | Central | Jardine  | Pencil | 90    | 4.99      | 449.10 |
| 8 | 2019-05-22 | West    | Thompson | Pencil | 32    | 1.99      | 63.68  |
| 9 | 2019-06-08 | East    | Jones    | Binder | 60    | 8.99      | 539.40 |

In [92]:

```
df.tail()
```

Out[92]:

|  | OrderDate | Region | Rep | Item | Units | Unit Cost | Total |
|--|-----------|--------|-----|------|-------|-----------|-------|
|--|-----------|--------|-----|------|-------|-----------|-------|

| 38 | OrderDate  | Region  | Rep     | Item   | Units | Unit Cost | Total   |
|----|------------|---------|---------|--------|-------|-----------|---------|
| 39 | 2020-10-31 | Central | Andrews | Pencil | 14    | 1.29      | 18.06   |
| 40 | 2020-11-17 | Central | Jardine | Binder | 11    | 4.99      | 54.89   |
| 41 | 2020-12-04 | Central | Jardine | Binder | 94    | 19.99     | 1879.06 |
| 42 | 2020-12-21 | Central | Andrews | Binder | 28    | 4.99      | 139.72  |

In [93]:

```
df.columns
```

Out[93]:

```
Index(['OrderDate', 'Region', 'Rep', 'Item', 'Units', 'Unit Cost', 'Total'], dtype='object')
```

In [94]:

```
df.columns.to_list()
```

Out[94]:

```
['OrderDate', 'Region', 'Rep', 'Item', 'Units', 'Unit Cost', 'Total']
```

In [95]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43 entries, 0 to 42
Data columns (total 7 columns):
OrderDate    43 non-null datetime64[ns]
Region       43 non-null object
Rep          43 non-null object
Item         43 non-null object
Units        43 non-null int64
Unit Cost    43 non-null float64
Total        43 non-null float64
dtypes: datetime64[ns](1), float64(2), int64(1), object(3)
memory usage: 2.5+ KB
```

In [97]:

```
df.nunique()
```

Out[97]:

```
OrderDate    43
Region        3
Rep          11
Item          5
Units        37
Unit Cost     12
Total        41
dtype: int64
```

In [98]:

```
df.head()
```

Out[98]:

|   | OrderDate  | Region  | Rep     | Item   | Units | Unit Cost | Total  |
|---|------------|---------|---------|--------|-------|-----------|--------|
| 0 | 2019-01-06 | East    | Jones   | Pencil | 95    | 1.99      | 189.05 |
| 1 | 2019-01-23 | Central | Kivell  | Binder | 50    | 19.99     | 999.50 |
| 2 | 2019-02-09 | Central | Jardine | Pencil | 36    | 4.99      | 179.64 |
| 3 | 2019-02-26 | Central | Gill    | Pen    | 27    | 19.99     | 539.73 |

|   | OrderDate  | Region | Rep     | Item   | Units | Unit Cost | Total  |
|---|------------|--------|---------|--------|-------|-----------|--------|
| 4 | 2019-03-15 | West   | Sorvino | Pencil | 56    | 2.99      | 167.44 |

In [99]:

```
df['20% Dis'] = df['Total'].apply(lambda x: x*(20/100))
```

In [100]:

```
df.head()
```

Out[100]:

|   | OrderDate  | Region  | Rep     | Item   | Units | Unit Cost | Total  | 20% Dis |
|---|------------|---------|---------|--------|-------|-----------|--------|---------|
| 0 | 2019-01-06 | East    | Jones   | Pencil | 95    | 1.99      | 189.05 | 37.810  |
| 1 | 2019-01-23 | Central | Kivell  | Binder | 50    | 19.99     | 999.50 | 199.900 |
| 2 | 2019-02-09 | Central | Jardine | Pencil | 36    | 4.99      | 179.64 | 35.928  |
| 3 | 2019-02-26 | Central | Gill    | Pen    | 27    | 19.99     | 539.73 | 107.946 |
| 4 | 2019-03-15 | West    | Sorvino | Pencil | 56    | 2.99      | 167.44 | 33.488  |

In [101]:

```
df['Payment'] = df['Total'] - df['20% Dis']
```

In [102]:

```
df.head()
```

Out[102]:

|   | OrderDate  | Region  | Rep     | Item   | Units | Unit Cost | Total  | 20% Dis | Payment |
|---|------------|---------|---------|--------|-------|-----------|--------|---------|---------|
| 0 | 2019-01-06 | East    | Jones   | Pencil | 95    | 1.99      | 189.05 | 37.810  | 151.240 |
| 1 | 2019-01-23 | Central | Kivell  | Binder | 50    | 19.99     | 999.50 | 199.900 | 799.600 |
| 2 | 2019-02-09 | Central | Jardine | Pencil | 36    | 4.99      | 179.64 | 35.928  | 143.712 |
| 3 | 2019-02-26 | Central | Gill    | Pen    | 27    | 19.99     | 539.73 | 107.946 | 431.784 |
| 4 | 2019-03-15 | West    | Sorvino | Pencil | 56    | 2.99      | 167.44 | 33.488  | 133.952 |

In [103]:

```
import matplotlib.pyplot as plt
```

In [105]:

```
plt.figure(figsize = (12,6))
plt.plot(df['OrderDate'], df['Total'])
plt.show()
```



