Pandas

October 14, 2023

1 Pandas

```
[2]: import pandas as pd
     print(pd.__version__)
```

```
1.5.3
```

2 DataFrame

```
[15]: #using List
      import pandas as pd
      data = [1,2,3,4,5]
      df = pd.DataFrame(data,columns=["numbers"])
      print(df)
```

```
   numbers
0        1
1        2
2        3
3        4
4        5
```

```
[7]: #using List
     data=[['aaa',68,'M'],['bbb',87,'F'],['ggg',98,'M'],['ppp',76,'F']]
     a=pd.DataFrame(data,columns=['name','marks','gender'])
     print(a)
```

```
  name  marks gender
0  aaa     68      M
1  bbb     87      F
2  ggg     98      M
3  ppp     76      F
```

```
[2]: #using dict
     data={'name':['aa','bb','cc'],'age':[25,26,28],'city':
      ↪['madurai','chennai','tuticorin']}
```

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

```
   name  age       city
0    aa   25    madurai
1    bb   26    chennai
2    cc   28  tuticorin
```

3 Empty DataFrame

```
[2]: df=pd.DataFrame()
     print(df)
```

```
Empty DataFrame
Columns: []
Index: []
```

4 DataFrame Manipulation

4.1 Adding a new column to a DataFrame

```
[7]: data={'name':['aa','bb','cc'],'blood':['B+','A1B+','O+'],'weight':[50,40,67]}
     df=pd.DataFrame(data)
     print(df)
     df['address']=['India','United States','California']
     print(df)
```

```
   name blood  weight
0    aa    B+      50
1    bb  A1B+      40
2    cc    O+      67
   name blood  weight        address
0    aa    B+      50          India
1    bb  A1B+      40  United States
2    cc    O+      67     California
```

4.2 Adding a new row to a DataFrame

```
[8]: data={'name':['aa','bb','cc'],'blood':['B+','A1B+','O+'],'weight':[50,40,67]}
     df=pd.DataFrame(data)
     print(df)
     df.loc[len(df.index)]=['dd','A+',77]
     print(df)
```

```
   name blood  weight
0    aa    B+      50
1    bb  A1B+      40
2    cc    O+      67
   name blood  weight
0    aa    B+      50
1    bb  A1B+      40
2    cc    O+      67
3    dd    A+      77
```

4.3 Remove Rows from a Pandas DataFrame

```python
[11]: data={'name':['aa','bb','cc','dd'],'blood':['B+','A1B+','O+','A+'],'weight':
      ↪[50,40,67,78]}
      df=pd.DataFrame(data)
      print(df)
      print()
      df.drop(index=1, inplace=True)
      print(df)
```

```
   name blood  weight
0    aa    B+      50
1    bb  A1B+      40
2    cc    O+      67
3    dd    A+      78

   name blood  weight
0    aa    B+      50
2    cc    O+      67
3    dd    A+      78
```

4.4 Remove Columns from a Pandas DataFrame

```python
[14]: data={'name':['aa','bb','cc','dd'],'blood':['B+','A1B+','O+','A+'],'weight':
      ↪[50,40,67,78]}
      df=pd.DataFrame(data)
      print(df)
      print()
      df.drop(columns='weight', inplace=True)
      print(df)
```

```
   name blood  weight
0    aa    B+      50
1    bb  A1B+      40
2    cc    O+      67
3    dd    A+      78
```

```
     name blood
0     aa    B+
1     bb  A1B+
2     cc    O+
3     dd    A+
```

4.5   Read csv File

```
[10]: a=pd.read_csv('D:\wholesale.csv')
      print(a)
```

```
       Series_reference    Period   Data_value   Suppressed STATUS    UNITS  \
0           WTSQ.SFA1CA    1995.03      2368.69          NaN      F  Dollars
1           WTSQ.SFA1CA    1995.06      2100.44          NaN      F  Dollars
2           WTSQ.SFA1CA    1995.09      2070.21          NaN      F  Dollars
3           WTSQ.SFA1CA    1995.12      2284.77          NaN      F  Dollars
4           WTSQ.SFA1CA    1996.03      2134.76          NaN      F  Dollars
...                 ...        ...          ...          ...    ...      ...
5229        WTSQ.SFZ9CT    2022.06          NaN          NaN      C  Dollars
5230        WTSQ.SFZ9CT    2022.09          NaN          NaN      C  Dollars
5231        WTSQ.SFZ9CT    2022.12          NaN          NaN      C  Dollars
5232        WTSQ.SFZ9CT    2023.03          NaN          NaN      C  Dollars
5233        WTSQ.SFZ9CT    2023.06          NaN          NaN      C  Dollars

       Magnitude                           Subject  \
0              6   Wholesale Trade Survey - WTS
1              6   Wholesale Trade Survey - WTS
2              6   Wholesale Trade Survey - WTS
3              6   Wholesale Trade Survey - WTS
4              6   Wholesale Trade Survey - WTS
...          ...                               ...
5229           6   Wholesale Trade Survey - WTS
5230           6   Wholesale Trade Survey - WTS
5231           6   Wholesale Trade Survey - WTS
5232           6   Wholesale Trade Survey - WTS
5233           6   Wholesale Trade Survey - WTS

                                            Group  \
0       Industry by variable - (ANZSIC06) - Subannual ...
1       Industry by variable - (ANZSIC06) - Subannual ...
2       Industry by variable - (ANZSIC06) - Subannual ...
3       Industry by variable - (ANZSIC06) - Subannual ...
4       Industry by variable - (ANZSIC06) - Subannual ...
...                                            ...
5229    Industry by variable - (ANZSIC06) - Subannual ...
5230    Industry by variable - (ANZSIC06) - Subannual ...
```

```
5231  Industry by variable - (ANZSIC06) - Subannual ...
5232  Industry by variable - (ANZSIC06) - Subannual ...
5233  Industry by variable - (ANZSIC06) - Subannual ...

                      Series_title_1           Series_title_2  Series_title_3  \
0      Basic material wholesaling  Sales (operating income)  Current prices
1      Basic material wholesaling  Sales (operating income)  Current prices
2      Basic material wholesaling  Sales (operating income)  Current prices
3      Basic material wholesaling  Sales (operating income)  Current prices
4      Basic material wholesaling  Sales (operating income)  Current prices
...                           ...                       ...             ...
5229             Total wholesaling            Total stocks  Current prices
5230             Total wholesaling            Total stocks  Current prices
5231             Total wholesaling            Total stocks  Current prices
5232             Total wholesaling            Total stocks  Current prices
5233             Total wholesaling            Total stocks  Current prices

        Series_title_4 Series_title_5
0          Unadjusted            NaN
1          Unadjusted            NaN
2          Unadjusted            NaN
3          Unadjusted            NaN
4          Unadjusted            NaN
...               ...            ...
5229            Trend            NaN
5230            Trend            NaN
5231            Trend            NaN
5232            Trend            NaN
5233            Trend            NaN

[5234 rows x 14 columns]
```

5   DataFrame Analysis

5.0.1   head()

```python
a=pd.read_csv('D:\wholesale.csv')
print(a.head(3))
```

```
  Series_reference    Period  Data_value  Suppressed STATUS    UNITS  \
0     WTSQ.SFA1CA   1995.03     2368.69         NaN      F  Dollars
1     WTSQ.SFA1CA   1995.06     2100.44         NaN      F  Dollars
2     WTSQ.SFA1CA   1995.09     2070.21         NaN      F  Dollars


   Magnitude                        Subject  \
0          6  Wholesale Trade Survey - WTS
```

```
1            6  Wholesale Trade Survey - WTS
2            6  Wholesale Trade Survey - WTS

                                             Group  \
0  Industry by variable - (ANZSIC06) - Subannual ...
1  Industry by variable - (ANZSIC06) - Subannual ...
2  Industry by variable - (ANZSIC06) - Subannual ...

              Series_title_1            Series_title_2 Series_title_3  \
0  Basic material wholesaling  Sales (operating income)  Current prices
1  Basic material wholesaling  Sales (operating income)  Current prices
2  Basic material wholesaling  Sales (operating income)  Current prices

  Series_title_4 Series_title_5
0     Unadjusted            NaN
1     Unadjusted            NaN
2     Unadjusted            NaN
```

5.0.2 tail()

```
[13]: a=pd.read_csv('D:\wholesale.csv')
      print(a.tail(3))
```

```
      Series_reference   Period  Data_value  Suppressed STATUS    UNITS  \
5231       WTSQ.SFZ9CT  2022.12         NaN         NaN      C  Dollars
5232       WTSQ.SFZ9CT  2023.03         NaN         NaN      C  Dollars
5233       WTSQ.SFZ9CT  2023.06         NaN         NaN      C  Dollars

      Magnitude                     Subject  \
5231          6  Wholesale Trade Survey - WTS
5232          6  Wholesale Trade Survey - WTS
5233          6  Wholesale Trade Survey - WTS

                                             Group    Series_title_1  \
5231  Industry by variable - (ANZSIC06) - Subannual ...  Total wholesaling
5232  Industry by variable - (ANZSIC06) - Subannual ...  Total wholesaling
5233  Industry by variable - (ANZSIC06) - Subannual ...  Total wholesaling

     Series_title_2  Series_title_3 Series_title_4 Series_title_5
5231   Total stocks  Current prices          Trend            NaN
5232   Total stocks  Current prices          Trend            NaN
5233   Total stocks  Current prices          Trend            NaN
```

5.0.3 Get DataFrame Information

```
[14]: a=pd.read_csv('D:\wholesale.csv')
      df=pd.DataFrame(a)
      df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5234 entries, 0 to 5233
Data columns (total 14 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Series_reference  5234 non-null   object
 1   Period            5234 non-null   float64
 2   Data_value        5038 non-null   float64
 3   Suppressed        0 non-null      float64
 4   STATUS            5234 non-null   object
 5   UNITS             5234 non-null   object
 6   Magnitude         5234 non-null   int64
 7   Subject           5234 non-null   object
 8   Group             5234 non-null   object
 9   Series_title_1    5234 non-null   object
 10  Series_title_2    5234 non-null   object
 11  Series_title_3    5234 non-null   object
 12  Series_title_4    5234 non-null   object
 13  Series_title_5    446 non-null    object
dtypes: float64(3), int64(1), object(10)
memory usage: 572.6+ KB
```

6 Series

```
[18]: #using List
      a=[10,20,30,50]
      b=pd.Series(a)
      print(b)
      print(b[3])
```

```
0    10
1    20
2    30
3    50
dtype: int64
50
```

6.1 Labels

```
[16]: a=[1,3,'f',7,9]
      b=pd.Series(a,index=['a','b','c','d','e'])
      print(b)
      print(b['b'])
```

```
a    1
b    3
c    f
d    7
e    9
dtype: object
3
```

6.2 Create Series from a Dict

```
[10]: per={"sem1":79,"sem2":70,"sem3":77,"sem4":90}
      s=pd.Series(per)
      print(s)
```

```
sem1    79
sem2    70
sem3    77
sem4    90
dtype: int64
```

```
[8]: per={"sem1":79,"sem2":80,"sem3":77,"sem4":90}
     s=pd.Series(per,index=['sem1','sem2'])
     print(s)
```

```
sem1    79
sem2    80
dtype: int64
```

```
[29]: per={"IBCA":60,"IIBCA":47,"IIIBCA":53}
      s=pd.Series(per)
      print(s)
      print(s['IBCA'])
```

```
IBCA      60
IIBCA     47
IIIBCA    53
dtype: int64
60
```

7 Setting Index

```
[7]: data={'id':[100,121,321,343],'name':['ajith','darun','anto','velan'],'branch':
     →['sappani Kovil','villapuram',

                                                                                    ␣
     →'valluvar illam','annanagar']}
     df.set_index('id', inplace=True)
     print(df)
```

```
        name          branch
id
100  ajith   sappani Kovil
121  darun      villapuram
321   anto  valluvar illam
343  velan        annanagar
```

```
[5]: data={'id':[100,121,321,343],'name':['ajith','darun','anto','velan'],'branch':
     →['sappani Kovil','villapuram',

                                                                                    ␣
     →'valluvar illam','annanagar']}
     df=pd.DataFrame(data, index=pd.RangeIndex(5,9,name='index'))
     print(df)
```

```
        id    name          branch
index
5      100  ajith   sappani Kovil
6      121  darun      villapuram
7      321   anto  valluvar illam
8      343  velan        annanagar
```