PANDAS

May 19, 2023

[1]:
```
import pandas as pd
```

[2]:
```
series = pd.Series([1,2,3,4,5])
series
```

[2]:
```
0    1
1    2
2    3
3    4
4    5
dtype: int64
```

1 Inserting Arrays In Series

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

[4]:
```
array = np.array([1,2,3,4,5,6,7,8,9])
array
```

[4]:
```
array([1, 2, 3, 4, 5, 6, 7, 8, 9])
```

[6]:
```
pd.Series(array)
```

[6]:
```
0    1
1    2
2    3
3    4
4    5
5    6
6    7
7    8
8    9
dtype: int32
```

2 Inserting Index into Series

[7]:
```
Sr = pd.Series(array, index =␣
↪("one","two","three","four","five","six","seven","eight","nine"))
Sr
```

[7]:
```
one      1
two      2
three    3
four     4
five     5
six      6
seven    7
eight    8
nine     9
dtype: int32
```

[8]:
```
Sr['one']
```

[8]:
```
1
```

[9]:
```
exp = pd.Series([1,2,3,4,5],index = ("one","two","three","four","five"))
exp
```

[9]:
```
one      1
two      2
three    3
four     4
five     5
dtype: int64
```

3 Now We are having Two Series...Add Both exp+Sr

[10]:
```
exp+Sr
```

[10]:
```
eight     NaN
five     10.0
four      8.0
nine      NaN
one       2.0
seven     NaN
six       NaN
three     6.0
two       4.0
dtype: float64
```

[11]:
```
#nan number generated because index was not matching out
```

[12]:
```
pd.concat([exp,Sr],axis=1)
```

[12]:

	0	1
one	1.0	1
two	2.0	2
three	3.0	3
four	4.0	4
five	5.0	5
six	NaN	6
seven	NaN	7
eight	NaN	8
nine	NaN	9

[13]:
```
pd.concat([exp,Sr])
```

[13]:
```
one      1
two      2
three    3
four     4
five     5
one      1
two      2
three    3
four     4
five     5
six      6
seven    7
eight    8
nine     9
dtype: int64
```

4 DATAFRAME CREATION

[14]:
```
df1 = pd.DataFrame(np.random.rand(5,5))#here we created dataframe using numpy␣
 ↪random numbers
df1
```

[14]:

	0	1	2	3	4
0	0.000246	0.216858	0.530353	0.390946	0.276777
1	0.539195	0.918449	0.753328	0.786820	0.124434
2	0.958271	0.091281	0.844990	0.623345	0.585116
3	0.736231	0.483943	0.909906	0.783644	0.838569
4	0.590180	0.977407	0.837250	0.383525	0.248370

5 indexing and giving column names

[15]:
```
df1 = pd.DataFrame(np.random.
 ↪rand(5,5),index=["one","two","three","four","five"],
                                          ␣
 ↪columns=["first","second","third","fourth","fifth"])
df1
```

[15]:

	first	second	third	fourth	fifth
one	0.417353	0.010517	0.473127	0.099021	0.670368
two	0.529190	0.297085	0.818766	0.325834	0.083521
three	0.429977	0.849677	0.415422	0.730958	0.893193
four	0.166678	0.032881	0.305859	0.918278	0.923379
five	0.585812	0.009920	0.588297	0.111847	0.034634

6 Access

[16]:
```
df1['first']
```

[16]:
```
one      0.417353
two      0.529190
three    0.429977
four     0.166678
five     0.585812
Name: first, dtype: float64
```

[17]:
```
df1[['first','second','third']]
```

[17]:

	first	second	third
one	0.417353	0.010517	0.473127
two	0.529190	0.297085	0.818766
three	0.429977	0.849677	0.415422
four	0.166678	0.032881	0.305859
five	0.585812	0.009920	0.588297

[18]:
```
df1.loc['one']
```

[18]:
```
first     0.417353
second    0.010517
third     0.473127
fourth    0.099021
fifth     0.670368
Name: one, dtype: float64
```

[19]:
```
df1.iloc[1]
```

[19]:
```
first     0.529190
second    0.297085
third     0.818766
fourth    0.325834
fifth     0.083521
Name: two, dtype: float64
```

7 Drop Column

[20]:
```
df1.drop('fifth',axis=1)
```

[20]:

	first	second	third	fourth
one	0.417353	0.010517	0.473127	0.099021
two	0.529190	0.297085	0.818766	0.325834
three	0.429977	0.849677	0.415422	0.730958
four	0.166678	0.032881	0.305859	0.918278
five	0.585812	0.009920	0.588297	0.111847

[21]:
```
df1.reset_index()
```

[21]:

	index	first	second	third	fourth	fifth
0	one	0.417353	0.010517	0.473127	0.099021	0.670368
1	two	0.529190	0.297085	0.818766	0.325834	0.083521
2	three	0.429977	0.849677	0.415422	0.730958	0.893193
3	four	0.166678	0.032881	0.305859	0.918278	0.923379
4	five	0.585812	0.009920	0.588297	0.111847	0.034634

[22]:
```
df1['fifth']=[10,20,30,40,50]
df1
```

[22]:

	first	second	third	fourth	fifth
one	0.417353	0.010517	0.473127	0.099021	10
two	0.529190	0.297085	0.818766	0.325834	20
three	0.429977	0.849677	0.415422	0.730958	30
four	0.166678	0.032881	0.305859	0.918278	40
five	0.585812	0.009920	0.588297	0.111847	50

8 Creating New DataFrame

[23]:
```
new_df = pd.DataFrame({'name':['shivesh','mansi','baljeet','sana','nupur'],
                       'age':[23,22,27,19,25],
                       'last name':['mishra','kapoor','kaur','naseem','chawla'],
                       'company':['QUESS','NISG','QUESS','NA','QUESS'],
                       'level':[5,4,5,2,5]},index=[1,2,3,4,5])
new_df
```

[23]:
```
      name  age last name company  level
1  shivesh   23    mishra   QUESS      5
2    mansi   22    kapoor    NISG      4
3  baljeet   27      kaur   QUESS      5
4     sana   19    naseem      NA      2
5    nupur   25    chawla   QUESS      5
```

[24]:
```
new_df1 = pd.DataFrame({'name':['shivesh','mansi','baljeet','sana','nupur'],
                        'age':[12,13,14,15,16],
                        'last name':['mishra','kapoor','kaur','naseem','chawla'],
                        'company':['QUESS','NISG','QUESS','NA','QUESS'],
                        'experience':[5,4,5,2,5]},index=[1,2,3,4,5])
new_df1
```

[24]:
```
      name  age last name company  experience
1  shivesh   12    mishra   QUESS           5
2    mansi   13    kapoor    NISG           4
3  baljeet   14      kaur   QUESS           5
4     sana   15    naseem      NA           2
5    nupur   16    chawla   QUESS           5
```

[25]:
```
new_df1+new_df
```

[25]:
```
   age     company  experience     last name  level            name
1   35  QUESSQUESS         NaN  mishramishra    NaN  shiveshshivesh
2   35    NISGNISG         NaN  kapoorkapoor    NaN      mansimansi
3   41  QUESSQUESS         NaN      kaurkaur    NaN  baljeetbaljeet
4   34        NANA         NaN  naseemnaseem    NaN        sanasana
5   41  QUESSQUESS         NaN  chawlachawla    NaN      nupurnupur
```

[26]:
```
pd.concat([new_df,new_df1],axis=0,sort=True)
```

[26]:
```
   age company  experience last name  level     name
1   23   QUESS         NaN    mishra    5.0  shivesh
2   22    NISG         NaN    kapoor    4.0    mansi
3   27   QUESS         NaN      kaur    5.0  baljeet
4   19      NA         NaN    naseem    2.0     sana
5   25   QUESS         NaN    chawla    5.0    nupur
1   12   QUESS         5.0    mishra    NaN  shivesh
2   13    NISG         4.0    kapoor    NaN    mansi
3   14   QUESS         5.0      kaur    NaN  baljeet
4   15      NA         2.0    naseem    NaN     sana
5   16   QUESS         5.0    chawla    NaN    nupur
```

[27]:
```
pd.merge(new_df,new_df1,how='right',on='age')
```

[27]:
```
  name_x  age last name_x company_x  level   name_y last name_y company_y  \
0    NaN   12         NaN       NaN    NaN  shivesh      mishra     QUESS
```

```
1    NaN   13         NaN       NaN    NaN    mansi      kapoor      NISG
2    NaN   14         NaN       NaN    NaN  baljeet        kaur     QUESS
3    NaN   15         NaN       NaN    NaN     sana      naseem        NA
4    NaN   16         NaN       NaN    NaN    nupur      chawla     QUESS

   experience
0           5
1           4
2           5
3           2
4           5
```

[28]:
```
final = pd.merge(new_df,new_df1,how='outer',on='age')
final
```

[28]:
```
     name_x  age last name_x company_x  level   name_y last name_y company_y  \
0  shivesh   23      mishra     QUESS    5.0      NaN         NaN        NaN
1    mansi   22      kapoor      NISG    4.0      NaN         NaN        NaN
2  baljeet   27        kaur     QUESS    5.0      NaN         NaN        NaN
3     sana   19      naseem        NA    2.0      NaN         NaN        NaN
4    nupur   25      chawla     QUESS    5.0      NaN         NaN        NaN
5      NaN   12         NaN       NaN    NaN  shivesh      mishra     QUESS
6      NaN   13         NaN       NaN    NaN    mansi      kapoor      NISG
7      NaN   14         NaN       NaN    NaN  baljeet        kaur     QUESS
8      NaN   15         NaN       NaN    NaN     sana      naseem        NA
9      NaN   16         NaN       NaN    NaN    nupur      chawla     QUESS

   experience
0         NaN
1         NaN
2         NaN
3         NaN
4         NaN
5         5.0
6         4.0
7         5.0
8         2.0
9         5.0
```

[29]:
```
final.to_csv('final.csv',index=True)
pd.read_csv('final.csv')
```

[29]:
```
   Unnamed: 0   name_x  age last name_x company_x  level   name_y last name_y  \
0           0  shivesh   23      mishra     QUESS    5.0      NaN         NaN
1           1    mansi   22      kapoor      NISG    4.0      NaN         NaN
2           2  baljeet   27        kaur     QUESS    5.0      NaN         NaN
3           3     sana   19      naseem       NaN    2.0      NaN         NaN
```

4	4	nupur	25	chawla	QUESS	5.0	NaN	NaN
5	5	NaN	12	NaN	NaN	NaN	shivesh	mishra
6	6	NaN	13	NaN	NaN	NaN	mansi	kapoor
7	7	NaN	14	NaN	NaN	NaN	baljeet	kaur
8	8	NaN	15	NaN	NaN	NaN	sana	naseem
9	9	NaN	16	NaN	NaN	NaN	nupur	chawla

	company_y	experience
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	QUESS	5.0
6	NISG	4.0
7	QUESS	5.0
8	NaN	2.0
9	QUESS	5.0

[30]:
```
pd.isnull(final).sum().sum()
40
```

[30]: 40

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: