PANDAS

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

In [6]:
```
series = pd.Series([1,2,3,4,5])
```

In [7]:
```
series
```

Out[7]:
```
0    1
1    2
2    3
3    4
4    5
dtype: int64
```

Inserting array in series

In [13]:
```
import numpy as np
```

In [46]:
```
array = np.array([1,2,3,4,5,6,7,8,9])
```

In [47]:
```
array
```

Out[47]:
```
array([1, 2, 3, 4, 5, 6, 7, 8, 9])
```

In [48]:
```
pd.Series(array)
```

Out[48]:
```
0    1
1    2
2    3
3    4
4    5
5    6
6    7
7    8
8    9
dtype: int32
```

inserting index into series

In [49]:
```
ind = pd.Series(array, index = ("one","two","three","four","five","six","seven","eight
```

In [50]:
```
ind
```

Out[50]:
```
one      1
two      2
three    3
four     4
five     5
six      6
seven    7
eight    8
nine     9
dtype: int32
```

In [51]:
```
ind['one']
```

Out[51]:
```
1
```

In [55]:
```
exp = pd.Series([1,2,3,4,5],index = ("one","two","three","four","five"))
```

In [56]:
```
exp
```

Out[56]:
```
one      1
two      2
three    3
four     4
five     5
dtype: int64
```

now we are having two series...exp and ind

simple maths of addition on both series

In [57]:
```
exp+ind
```

Out[57]:
```
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
```

In [58]:
```
#nan number generated because index was not matching out
```

In [83]:
```
pd.concat([exp,ind],axis=1)
```

Out[83]:

	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

In [65]:
```
pd.concat([exp,ind])
```

Out[65]:
```
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
```

DATAFRAME CREATION

In [77]:
```
df1 = pd.DataFrame(np.random.rand(5,5))#here we created dataframe using numpy random r
```

In [78]:
```
df1
```

Out[78]:

	0	1	2	3	4
0	0.182469	0.320200	0.371576	0.840374	0.554736
1	0.818659	0.217008	0.753545	0.159630	0.123517
2	0.058443	0.951232	0.852316	0.706507	0.641665
3	0.293392	0.445356	0.460223	0.269861	0.262624
4	0.194180	0.781211	0.673131	0.161167	0.648876

indexing and giving column names

In [88]:
```
df1 = pd.DataFrame(np.random.rand(5,5),index=["one","two","three","four","five"],
                                       columns=["first","second","third","fourth","fi
```

In [89]:
```
df1
```

Out[89]:

	first	second	third	fourth	fifth
one	0.053503	0.106235	0.915407	0.880849	0.246622
two	0.190750	0.828302	0.217411	0.468420	0.515375
three	0.993074	0.518722	0.626595	0.239213	0.869956
four	0.380941	0.427077	0.891345	0.781259	0.957811
five	0.282786	0.775472	0.969148	0.548966	0.203651

Access

In [107…
```
df1['first']
```

Out[107]:
```
one      0.053503
two      0.190750
three    0.993074
four     0.380941
five     0.282786
Name: first, dtype: float64
```

In [108…
```
df1[['first','second','third']]
```

Out[108]:

	first	second	third
one	0.053503	0.106235	0.915407
two	0.190750	0.828302	0.217411
three	0.993074	0.518722	0.626595
four	0.380941	0.427077	0.891345
five	0.282786	0.775472	0.969148

In [113…
```
df1.loc['one']
```

Out[113]:
```
first     0.053503
second    0.106235
third     0.915407
fourth    0.880849
fifth     0.246622
Name: one, dtype: float64
```

In [114…
```
df1.iloc[1]
```

Out[114]:
```
first     0.190750
second    0.828302
third     0.217411
fourth    0.468420
fifth     0.515375
Name: two, dtype: float64
```

drop column

In [117…
```
df1.drop('fifth',axis=1)
```

Out[117]:

	first	second	third	fourth
one	0.053503	0.106235	0.915407	0.880849
two	0.190750	0.828302	0.217411	0.468420
three	0.993074	0.518722	0.626595	0.239213
four	0.380941	0.427077	0.891345	0.781259
five	0.282786	0.775472	0.969148	0.548966

In [120…
```
df1.reset_index()
```

Out[120]:

	index	first	second	third	fourth	fifth
0	one	0.053503	0.106235	0.915407	0.880849	0.246622
1	two	0.190750	0.828302	0.217411	0.468420	0.515375
2	three	0.993074	0.518722	0.626595	0.239213	0.869956
3	four	0.380941	0.427077	0.891345	0.781259	0.957811
4	five	0.282786	0.775472	0.969148	0.548966	0.203651

In [121…
```
df1['fifth']=[10,20,30,40,50]
```

In [122…
```
df1
```

Out[122]:

	first	second	third	fourth	fifth
one	0.053503	0.106235	0.915407	0.880849	10
two	0.190750	0.828302	0.217411	0.468420	20
three	0.993074	0.518722	0.626595	0.239213	30
four	0.380941	0.427077	0.891345	0.781259	40
five	0.282786	0.775472	0.969148	0.548966	50

Creating new dataframe

In [177…
```
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])
```

In [178…
```
new_df
```

Out[178]:

	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

In [179…
```
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])
```

In [180…
```
new_df1
```

Out[180]:

	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

In [181…
```
new_df1+new_df
```

Out[181]:

	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

In [182…
```
pd.concat([new_df,new_df1],axis=0,sort=True)
```

Out[182]:

	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

In [188…
```
pd.merge(new_df,new_df1,how='right',on='age')
```

Out[188]:

	name_x	age	last name_x	company_x	level	name_y	last name_y	company_y	experience
0	NaN	12	NaN	NaN	NaN	shivesh	mishra	QUESS	5
1	NaN	13	NaN	NaN	NaN	mansi	kapoor	NISG	4
2	NaN	14	NaN	NaN	NaN	baljeet	kaur	QUESS	5
3	NaN	15	NaN	NaN	NaN	sana	naseem	NA	2
4	NaN	16	NaN	NaN	NaN	nupur	chawla	QUESS	5

In [191…
```
final = pd.merge(new_df,new_df1,how='outer',on='age')
```

In [192…
```
final
```

Out[192]:

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

In [195…
```
final.to_csv('final.csv',index=True)
```

In [196…
```
pd.read_csv('final.csv')
```

Out[196]:

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

In [199…
```
pd.isnull(final).sum().sum()
```

Out[199]:
```
40
```