PANDAS Tutorial

Install Libararies

In [5]:
```
# pip install pandas
# pip install numpy
```

Import Libararies

In [6]:
```
import pandas as pd
import numpy as np
```

In [14]:
```
# object creation
s=pd.Series([1,2,3,np.nan,5,6])
s
```

Out[14]:
```
0    1.0
1    2.0
2    3.0
3    NaN
4    5.0
5    6.0
dtype: float64
```

In [16]:
```
dates= pd.date_range("20220120",periods=16)
dates
```

Out[16]:
```
DatetimeIndex(['2022-01-20', '2022-01-21', '2022-01-22', '2022-01-23',
               '2022-01-24', '2022-01-25', '2022-01-26', '2022-01-27',
               '2022-01-28', '2022-01-29', '2022-01-30', '2022-01-31',
               '2022-02-01', '2022-02-02', '2022-02-03', '2022-02-04'],
              dtype='datetime64[ns]', freq='D')
```

In [18]:
```
df= pd.DataFrame(np.random.randn(16,4),index=dates,columns=list("BABA"))
df
```

Out[18]:

	B	A	B	A
2022-01-20	2.137841	-0.769085	0.605918	-0.081160
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352
2022-01-23	-0.873729	1.477522	-0.487166	0.726665
2022-01-24	-1.484645	0.051141	1.470947	1.742746
2022-01-25	1.403945	0.883918	-0.892503	-0.863513
2022-01-26	-1.623025	-2.708877	-1.051823	1.202084
2022-01-27	-1.122181	0.313951	1.233480	1.349252
2022-01-28	-0.445214	-1.932293	-0.241790	0.769331

	B	A	B	A
2022-01-29	1.587859	0.047537	-1.225914	-0.052216
2022-01-30	0.765112	-0.488646	0.934667	-1.780728
2022-01-31	0.461617	-0.618401	-0.543748	0.197366
2022-02-01	-0.010508	-0.166281	-0.600458	-0.521830
2022-02-02	0.545949	-0.026106	0.242712	-1.146708
2022-02-03	-2.264509	-1.219256	-0.882976	0.086152
2022-02-04	-0.883709	-0.176698	-1.145405	0.461645

In [30]:
```
df2= pd.DataFrame(
    {
        "A":np.zeros(4),
        "B":pd.date_range("20220201",periods=4),
        "C":pd.Series(1,index=list(range(4)),dtype="float32"),
        "D":np.array([6]*4,dtype="int32"),
        "E":pd.Categorical(["black","blue","pink","gold"]),
        "F":"colors" }
)
df2
```

Out[30]:

	A	B	C	D	E	F
0	0.0	2022-02-01	1.0	6	black	colors
1	0.0	2022-02-02	1.0	6	blue	colors
2	0.0	2022-02-03	1.0	6	pink	colors
3	0.0	2022-02-04	1.0	6	gold	colors

In [27]:
```
df2.dtypes
```

Out[27]:
```
A           float64
B    datetime64[ns]
C           float32
D             int32
E          category
F            object
dtype: object
```

In [33]:
```
df.head()
```

Out[33]:

	B	A	B	A
2022-01-20	2.137841	-0.769085	0.605918	-0.081160
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352
2022-01-23	-0.873729	1.477522	-0.487166	0.726665
2022-01-24	-1.484645	0.051141	1.470947	1.742746

In [34]:

In [ ]:
```
df.head(2)
```

Out[34]:

	B	A	B	A
2022-01-20	2.137841	-0.769085	0.605918	-0.081160
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606

In [35]:
```
df.tail(2)
```

Out[35]:

	B	A	B	A
2022-02-03	-2.264509	-1.219256	-0.882976	0.086152
2022-02-04	-0.883709	-0.176698	-1.145405	0.461645

In [36]:
```
df.index
```

Out[36]:
```
DatetimeIndex(['2022-01-20', '2022-01-21', '2022-01-22', '2022-01-23',
               '2022-01-24', '2022-01-25', '2022-01-26', '2022-01-27',
               '2022-01-28', '2022-01-29', '2022-01-30', '2022-01-31',
               '2022-02-01', '2022-02-02', '2022-02-03', '2022-02-04'],
              dtype='datetime64[ns]', freq='D')
```

In [37]:
```
df2.index
```

Out[37]:
```
Int64Index([0, 1, 2, 3], dtype='int64')
```

In [38]:
```
df.to_numpy
```

Out[38]:
```
<bound method DataFrame.to_numpy of                    B         A         B
A
2022-01-20  2.137841 -0.769085  0.605918 -0.081160
2022-01-21 -0.203228 -1.249527 -0.847458  1.251606
2022-01-22 -0.599437 -0.227804  1.297838 -0.848352
2022-01-23 -0.873729  1.477522 -0.487166  0.726665
2022-01-24 -1.484645  0.051141  1.470947  1.742746
2022-01-25  1.403945  0.883918 -0.892503 -0.863513
2022-01-26 -1.623025 -2.708877 -1.051823  1.202084
2022-01-27 -1.122181  0.313951  1.233480  1.349252
2022-01-28 -0.445214 -1.932293 -0.241790  0.769331
2022-01-29  1.587859  0.047537 -1.225914 -0.052216
2022-01-30  0.765112 -0.488646  0.934667 -1.780728
2022-01-31  0.461617 -0.618401 -0.543748  0.197366
2022-02-01 -0.010508 -0.166281 -0.600458 -0.521830
2022-02-02  0.545949 -0.026106  0.242712 -1.146708
2022-02-03 -2.264509 -1.219256 -0.882976  0.086152
2022-02-04 -0.883709 -0.176698 -1.145405  0.461645>
```

In [39]:
```
df2.to_numpy
```

Out[39]:
```
<bound method DataFrame.to_numpy of      A          B    C  D      E       F
0  0.0 2022-02-01  1.0  6  black  colors
1  0.0 2022-02-02  1.0  6   blue  colors
2  0.0 2022-02-03  1.0  6   pink  colors
3  0.0 2022-02-04  1.0  6   gold  colors>
```

In [40]:

```
df.describe()
```

Out[40]:

	B	A	B	A
count	16.000000	16.000000	16.000000	16.000000
mean	-0.162991	-0.425557	-0.133355	0.155771
std	1.241053	1.020263	0.950516	1.003103
min	-2.264509	-2.708877	-1.225914	-1.780728
25%	-0.943327	-0.881628	-0.885358	-0.603461
50%	-0.324221	-0.202251	-0.515457	0.141759
75%	0.600740	0.048438	0.688105	0.877519
max	2.137841	1.477522	1.470947	1.742746

In [41]:

```
#transpose Data,rows into column,column into rows
df2.T
```

Out[41]:

	0	1	2	3
A	0.0	0.0	0.0	0.0
B	2022-02-01 00:00:00	2022-02-02 00:00:00	2022-02-03 00:00:00	2022-02-04 00:00:00
C	1.0	1.0	1.0	1.0
D	6	6	6	6
E	black	blue	pink	gold
F	colors	colors	colors	colors

In [42]:

```
df2.sort_index(axis=0,ascending=False)
```

Out[42]:

	A	B	C	D	E	F
3	0.0	2022-02-04	1.0	6	gold	colors
2	0.0	2022-02-03	1.0	6	pink	colors
1	0.0	2022-02-02	1.0	6	blue	colors
0	0.0	2022-02-01	1.0	6	black	colors

In [45]:

```
df.sort_index(axis=0,ascending=False)
```

Out[45]:

	B	A	B	A
2022-02-04	-0.883709	-0.176698	-1.145405	0.461645
2022-02-03	-2.264509	-1.219256	-0.882976	0.086152
2022-02-02	0.545949	-0.026106	0.242712	-1.146708
2022-02-01	-0.010508	-0.166281	-0.600458	-0.521830
2022-01-31	0.461617	-0.618401	-0.543748	0.197366

	B	A	B	A
2022-01-30	0.765112	-0.488646	0.934667	-1.780728
2022-01-29	1.587859	0.047537	-1.225914	-0.052216
2022-01-28	-0.445214	-1.932293	-0.241790	0.769331
2022-01-27	-1.122181	0.313951	1.233480	1.349252
2022-01-26	-1.623025	-2.708877	-1.051823	1.202084
2022-01-25	1.403945	0.883918	-0.892503	-0.863513
2022-01-24	-1.484645	0.051141	1.470947	1.742746
2022-01-23	-0.873729	1.477522	-0.487166	0.726665
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606
2022-01-20	2.137841	-0.769085	0.605918	-0.081160

In [47]:
```
df2.sort_values(by="B")
```

Out[47]:

	A	B	C	D	E	F
0	0.0	2022-02-01	1.0	6	black	colors
1	0.0	2022-02-02	1.0	6	blue	colors
2	0.0	2022-02-03	1.0	6	pink	colors
3	0.0	2022-02-04	1.0	6	gold	colors

In [49]:
```
df2.sort_values(by="B",ascending=False)
```

Out[49]:

	A	B	C	D	E	F
3	0.0	2022-02-04	1.0	6	gold	colors
2	0.0	2022-02-03	1.0	6	pink	colors
1	0.0	2022-02-02	1.0	6	blue	colors
0	0.0	2022-02-01	1.0	6	black	colors

In [50]:
```
df["B"]
```

Out[50]:

	B	B
2022-01-20	2.137841	0.605918
2022-01-21	-0.203228	-0.847458
2022-01-22	-0.599437	1.297838
2022-01-23	-0.873729	-0.487166
2022-01-24	-1.484645	1.470947
2022-01-25	1.403945	-0.892503

	B	B
2022-01-26	-1.623025	-1.051823
2022-01-27	-1.122181	1.233480
2022-01-28	-0.445214	-0.241790
2022-01-29	1.587859	-1.225914
2022-01-30	0.765112	0.934667
2022-01-31	0.461617	-0.543748
2022-02-01	-0.010508	-0.600458
2022-02-02	0.545949	0.242712
2022-02-03	-2.264509	-0.882976
2022-02-04	-0.883709	-1.145405

In [51]:
```
df2["E"]
```

Out[51]:
```
0    black
1     blue
2     pink
3     gold
Name: E, dtype: category
Categories (4, object): ['black', 'blue', 'gold', 'pink']
```

In [52]:
```
df["B"]
```

Out[52]:

	B	B
2022-01-20	2.137841	0.605918
2022-01-21	-0.203228	-0.847458
2022-01-22	-0.599437	1.297838
2022-01-23	-0.873729	-0.487166
2022-01-24	-1.484645	1.470947
2022-01-25	1.403945	-0.892503
2022-01-26	-1.623025	-1.051823
2022-01-27	-1.122181	1.233480
2022-01-28	-0.445214	-0.241790
2022-01-29	1.587859	-1.225914
2022-01-30	0.765112	0.934667
2022-01-31	0.461617	-0.543748
2022-02-01	-0.010508	-0.600458
2022-02-02	0.545949	0.242712
2022-02-03	-2.264509	-0.882976
2022-02-04	-0.883709	-1.145405

In [55]:
```
#row wise selection
df[0:]
df[0:3]
df[1:3]
```

Out[55]:

	B	A	B	A
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352

In [59]:
```
df.loc[dates[3]]
```

Out[59]:
```
B   -0.873729
A    1.477522
B   -0.487166
A    0.726665
Name: 2022-01-23 00:00:00, dtype: float64
```

In [62]:
```
df.loc[dates[1:3]]
```

Out[62]:

	B	A	B	A
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352

In [69]:
```
df.loc[dates[1:3],["A"]]
```

Out[69]:

	A	A
2022-01-21	-1.249527	1.251606
2022-01-22	-0.227804	-0.848352

In [71]:
```
df.loc[:,["A"]]
```

Out[71]:

	A	A
2022-01-20	-0.769085	-0.081160
2022-01-21	-1.249527	1.251606
2022-01-22	-0.227804	-0.848352
2022-01-23	1.477522	0.726665
2022-01-24	0.051141	1.742746
2022-01-25	0.883918	-0.863513
2022-01-26	-2.708877	1.202084
2022-01-27	0.313951	1.349252
2022-01-28	-1.932293	0.769331
2022-01-29	0.047537	-0.052216
2022-01-30	-0.488646	-1.780728

	A	A
2022-01-31	-0.618401	0.197366
2022-02-01	-0.166281	-0.521830
2022-02-02	-0.026106	-1.146708
2022-02-03	-1.219256	0.086152
2022-02-04	-0.176698	0.461645

In [73]:
```
df.loc["20220120":"20220124",["A"]]
```

Out[73]:

	A	A
2022-01-20	-0.769085	-0.081160
2022-01-21	-1.249527	1.251606
2022-01-22	-0.227804	-0.848352
2022-01-23	1.477522	0.726665
2022-01-24	0.051141	1.742746

In [75]:
```
df.loc[["20220120","20220124"],["A"]]
```

Out[75]:

	A	A
2022-01-20	-0.769085	-0.081160
2022-01-24	0.051141	1.742746

In [76]:
```
df.loc["20220130",["A","B"]]
```

Out[76]:
```
A   -0.488646
A   -1.780728
B    0.765112
B    0.934667
Name: 2022-01-30 00:00:00, dtype: float64
```

In [83]:
```
df.at[dates[14],"A"]
```

Out[83]:
```
A   -1.219256
A    0.086152
Name: 2022-02-03 00:00:00, dtype: float64
```

In [84]:
```
df.at[dates[2],"A"]
```

Out[84]:
```
A   -0.227804
A   -0.848352
Name: 2022-01-22 00:00:00, dtype: float64
```

In [85]:
```
df.iloc[3]
```

Out[85]:
```
B   -0.873729
A    1.477522
B   -0.487166
```

```
A    0.726665
Name: 2022-01-23 00:00:00, dtype: float64
```

In [86]:
```
df.iloc[3:10]
```

Out[86]:

	B	A	B	A
2022-01-23	-0.873729	1.477522	-0.487166	0.726665
2022-01-24	-1.484645	0.051141	1.470947	1.742746
2022-01-25	1.403945	0.883918	-0.892503	-0.863513
2022-01-26	-1.623025	-2.708877	-1.051823	1.202084
2022-01-27	-1.122181	0.313951	1.233480	1.349252
2022-01-28	-0.445214	-1.932293	-0.241790	0.769331
2022-01-29	1.587859	0.047537	-1.225914	-0.052216

In [87]:
```
df.iloc[2:6,0:3]
```

Out[87]:

	B	A	B
2022-01-22	-0.599437	-0.227804	1.297838
2022-01-23	-0.873729	1.477522	-0.487166
2022-01-24	-1.484645	0.051141	1.470947
2022-01-25	1.403945	0.883918	-0.892503

In [88]:
```
df.iloc[:, 0:1]
```

Out[88]:

	B
2022-01-20	2.137841
2022-01-21	-0.203228
2022-01-22	-0.599437
2022-01-23	-0.873729
2022-01-24	-1.484645
2022-01-25	1.403945
2022-01-26	-1.623025
2022-01-27	-1.122181
2022-01-28	-0.445214
2022-01-29	1.587859
2022-01-30	0.765112
2022-01-31	0.461617
2022-02-01	-0.010508
2022-02-02	0.545949

	B
2022-02-03	-2.264509
2022-02-04	-0.883709

In [ ]:
```
#indexing not working in my kernal
df[df["B"]>0]
df[df["B","A"]>1.2]
df[df["B"]>0]
```

In [96]:
```
df2=df.copy()
df2["G"]=["A","A","M","M","A","R","A","A","M","M","A","R","A","A","M","M"]
df2
```

Out[96]:

	B	A	B	A	G
2022-01-20	2.137841	-0.769085	0.605918	-0.081160	A
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606	A
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352	M
2022-01-23	-0.873729	1.477522	-0.487166	0.726665	M
2022-01-24	-1.484645	0.051141	1.470947	1.742746	A
2022-01-25	1.403945	0.883918	-0.892503	-0.863513	R
2022-01-26	-1.623025	-2.708877	-1.051823	1.202084	A
2022-01-27	-1.122181	0.313951	1.233480	1.349252	A
2022-01-28	-0.445214	-1.932293	-0.241790	0.769331	M
2022-01-29	1.587859	0.047537	-1.225914	-0.052216	M
2022-01-30	0.765112	-0.488646	0.934667	-1.780728	A
2022-01-31	0.461617	-0.618401	-0.543748	0.197366	R
2022-02-01	-0.010508	-0.166281	-0.600458	-0.521830	A
2022-02-02	0.545949	-0.026106	0.242712	-1.146708	A
2022-02-03	-2.264509	-1.219256	-0.882976	0.086152	M
2022-02-04	-0.883709	-0.176698	-1.145405	0.461645	M

In [99]:
```
df2["yoo"]=[44,55,77,22,7,5,444,66,77,22,11,77,22,7,5,444]
df2
```

Out[99]:

	B	A	B	A	G	yoo
2022-01-20	2.137841	-0.769085	0.605918	-0.081160	A	44
2022-01-21	-0.203228	-1.249527	-0.847458	1.251606	A	55
2022-01-22	-0.599437	-0.227804	1.297838	-0.848352	M	77
2022-01-23	-0.873729	1.477522	-0.487166	0.726665	M	22
2022-01-24	-1.484645	0.051141	1.470947	1.742746	A	7

	B	A	B	A	G	yoo
2022-01-25	1.403945	0.883918	-0.892503	-0.863513	R	5
2022-01-26	-1.623025	-2.708877	-1.051823	1.202084	A	444
2022-01-27	-1.122181	0.313951	1.233480	1.349252	A	66
2022-01-28	-0.445214	-1.932293	-0.241790	0.769331	M	77
2022-01-29	1.587859	0.047537	-1.225914	-0.052216	M	22
2022-01-30	0.765112	-0.488646	0.934667	-1.780728	A	11
2022-01-31	0.461617	-0.618401	-0.543748	0.197366	R	77
2022-02-01	-0.010508	-0.166281	-0.600458	-0.521830	A	22
2022-02-02	0.545949	-0.026106	0.242712	-1.146708	A	7
2022-02-03	-2.264509	-1.219256	-0.882976	0.086152	M	5
2022-02-04	-0.883709	-0.176698	-1.145405	0.461645	M	444

In [100…

```
df2= df2.iloc[:,2:4]
df2
```

Out[100…

	B	A
2022-01-20	0.605918	-0.081160
2022-01-21	-0.847458	1.251606
2022-01-22	1.297838	-0.848352
2022-01-23	-0.487166	0.726665
2022-01-24	1.470947	1.742746
2022-01-25	-0.892503	-0.863513
2022-01-26	-1.051823	1.202084
2022-01-27	1.233480	1.349252
2022-01-28	-0.241790	0.769331
2022-01-29	-1.225914	-0.052216
2022-01-30	0.934667	-1.780728
2022-01-31	-0.543748	0.197366
2022-02-01	-0.600458	-0.521830
2022-02-02	0.242712	-1.146708
2022-02-03	-0.882976	0.086152
2022-02-04	-1.145405	0.461645