PANDAS

We shorten the imported name to np for better readability of code using NumPy. This is a widely adopted convention that makes your code more readable for everyone working on it. We recommend to always use import numpy as np . Same goes with pd.

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df=pd.read_csv("C:\\Users\\Asus\\Downloads\\IPL.csv") #to read a (comma seperated file)
#which is stord in a given location
```

In [3]:
```
df #to see the values of data frame that we named as df.
```

Out[3]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10
3	4	Indore	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	0	6
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0
5	6	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrisers Hyderabad	0	9
6	7	Mumbai	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
7	8	Indore	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	Kings XI Punjab	0	8
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0
9	10	Mumbai	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4

In [4]:
```
df.shape #to see the no. of rows & columns of data frame that we named as df.
```

Out[4]:
```
(10, 9)
```

In [5]:
```
len(df) #to see the no. of rows of data frame that we named as df.
```

Out[5]:
```
10
```

In [6]:
```
df.describe() #to see the statistical values of colums having numeric values of data frame that we named
```

Out[6]:

	id	win_by_runs	win_by_wickets
count	10.00000	10.000000	10.000000
mean	5.50000	14.700000	4.800000
std	3.02765	31.087868	3.823901
min	1.00000	0.000000	0.000000
25%	3.25000	0.000000	1.000000
50%	5.50000	0.000000	5.000000
75%	7.75000	11.250000	7.750000
max	10.00000	97.000000	10.000000

In [7]:

```
df.index
```

Out[7]:

```
RangeIndex(start=0, stop=10, step=1)
```

In [8]:

```
df.columns
```

Out[8]:

```
Index(['id', 'city', 'team1', 'team2', 'toss_winner', 'toss_decision',
       'winner', 'win_by_runs', 'win_by_wickets'],
      dtype='object')
```

In [9]:

```
df.T #Transposed. Rows and columns will be interchanged.
```

Out[9]:

	0	1	2	3	4	5	6	7	8	9
id	1	2	3	4	5	6	7	8	9	10
city	Hyderabad	Pune	Rajkot	Indore	Bangalore	Hyderabad	Mumbai	Indore	Pune	Mumbai
team1	Sunrisers Hyderabad	Mumbai Indians	Gujarat Lions	Rising Pune Supergiant	Royal Challengers Bangalore	Gujarat Lions	Kolkata Knight Riders	Royal Challengers Bangalore	Delhi Daredevils	Sunrisers Hyderabad
team2	Royal Challengers Bangalore	Rising Pune Supergiant	Kolkata Knight Riders	Kings XI Punjab	Delhi Daredevils	Sunrisers Hyderabad	Mumbai Indians	Kings XI Punjab	Rising Pune Supergiant	Mumbai Indians
toss_winner	Royal Challengers Bangalore	Rising Pune Supergiant	Kolkata Knight Riders	Kings XI Punjab	Royal Challengers Bangalore	Sunrisers Hyderabad	Mumbai Indians	Royal Challengers Bangalore	Rising Pune Supergiant	Mumbai Indians
toss_decision	field	field	field	field	bat	field	field	bat	field	field
winner	Sunrisers Hyderabad	Rising Pune Supergiant	Kolkata Knight Riders	Kings XI Punjab	Royal Challengers Bangalore	Sunrisers Hyderabad	Mumbai Indians	Kings XI Punjab	Delhi Daredevils	Mumbai Indians
win_by_runs	35	0	0	0	15	0	0	0	97	0
win_by_wickets	0	7	10	6	0	9	4	8	0	4

In [10]:

```
df.sort_values('city')
```

Out[10]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0
5	6	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrisers Hyderabad	0	9
3	4	Indore	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	0	6
7	8	Indore	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	Kings XI Punjab	0	8
6	7	Mumbai	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
9	10	Mumbai	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10

In [11]:

```
df.sort_values('win_by_runs')
```

Out[11]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10
3	4	Indore	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	0	6
5	6	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrisers Hyderabad	0	9
6	7	Mumbai	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
7	8	Indore	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	Kings XI Punjab	0	8
9	10	Mumbai	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0

In [12]:

```
df #original file never got changed
```

Out[12]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10
3	4	Indore	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	0	6
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0
5	6	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrisers Hyderabad	0	9
6	7	Mumbai	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
7	8	Indore	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	Kings XI Punjab	0	8
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0
9	10	Mumbai	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4

In [13]:

```
df1=df.sort_values('city') #new file with permanent changes is created.
```

In [14]:

```
df1
```

Out[14]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0
5	6	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrisers Hyderabad	0	9
3	4	Indore	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	0	6
7	8	Indore	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	Kings XI Punjab	0	8
6	7	Mumbai	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
9	10	Mumbai	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10

In [15]:

```
df['team1']
#df.team1   ----but this will not work if column name has space in it otherwise it will work.
```

Out[15]:

```
0             Sunrisers Hyderabad
1                  Mumbai Indians
2                   Gujarat Lions
3          Rising Pune Supergiant
4     Royal Challengers Bangalore
5                   Gujarat Lions
6           Kolkata Knight Riders
7     Royal Challengers Bangalore
8                Delhi Daredevils
9             Sunrisers Hyderabad
Name: team1, dtype: object
```

In [16]:

```
df['win_by_runs']+10
```

Out[16]:

```
0     45
1     10
2     10
3     10
4     25
5     10
6     10
7     10
8    107
9     10
Name: win_by_runs, dtype: int64
```

In [17]:

```
df['new_win_by_runs']=df['win_by_runs']+10 #new column created.
```

In [18]:

```
df.head(3)
```

Out[18]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0	45
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7	10
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10	10

In [19]:

```
df['new_column']='POOJA VAJPAYEE'
```

In [20]:

```
df.head(3)
```

Out[20]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0	45
1	2	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	Rising Pune Supergiant	0	7	10
2	3	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	Kolkata Knight Riders	0	10	10

In [21]:

```
df[5:9] #will show specified rows [start from:til end-1]
```

Out[21]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
5	6	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	Sunrisers Hyderabad	0	9	10
6	7	Mumbai	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	Mumbai Indians	0	4	10
7	8	Indore	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	Kings XI Punjab	0	8	10
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0	107

In [22]:

```
df[df.win_by_runs>30] #conditional selection.
```

Out[22]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0	45
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0	107

In [23]:

```
df1=df[df.win_by_runs>10]
```

In [24]:

```
df1
```

Out[24]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0	45
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0	25
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0	107

In [25]:

```
len(df1)
```

Out[25]:

```
3
```

In [26]:

```
df['new_column'].fillna(10) #fill in place of NA/NULL only.in this sheet there is no NA/NULL values.
```

Out[26]:

```
0    POOJA VAJPAYEE
1    POOJA VAJPAYEE
2    POOJA VAJPAYEE
3    POOJA VAJPAYEE
4    POOJA VAJPAYEE
5    POOJA VAJPAYEE
6    POOJA VAJPAYEE
7    POOJA VAJPAYEE
8    POOJA VAJPAYEE
9    POOJA VAJPAYEE
Name: new_column, dtype: object
```

In [27]:

```
df1.pop('new_column') #to remove a column(method 1)
```

Out[27]:

```
0    POOJA VAJPAYEE
4    POOJA VAJPAYEE
8    POOJA VAJPAYEE
Name: new_column, dtype: object
```

In [28]:

```
df1
```

Out[28]:

	id	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
0	1	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0	45
4	5	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0	25
8	9	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0	107

In [29]:

```
df1.drop('id',axis=1) #to remove a column(method 2)
```

Out[29]:

	city	team1	team2	toss_winner	toss_decision	winner	win_by_runs	win_by_wickets	new_win_by_runs
0	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Sunrisers Hyderabad	35	0	45
4	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	Royal Challengers Bangalore	15	0	25
8	Pune	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	Delhi Daredevils	97	0	107

In [30]:

```
df['win_by_runs'].apply(np.sqrt)
```

Out[30]:

```
0    5.916080
1    0.000000
2    0.000000
3    0.000000
4    3.872983
5    0.000000
6    0.000000
7    0.000000
8    9.848858
9    0.000000
Name: win_by_runs, dtype: float64
```

In [31]:

```
df1 = pd.DataFrame({'A': ['A0', 'A1', 'A2', 'A3'],
                    'B': ['B0', 'B1', 'B2', 'B3'],
                    'C': ['C0', 'C1', 'C2', 'C3'],
                    'D': ['D0', 'D1', 'D2', 'D3']},
                  index=[0,1,2,3])
```

In [32]:

```
df2 = pd.DataFrame({'A': ['A4', 'A5', 'A6', 'A7'],
                    'B': ['B4', 'B5', 'B6', 'B7'],
                    'C': ['C0', 'C3', 'C6', 'C7'],
                    'D': ['D4', 'D5', 'D6', 'D7']},
                  index=[4,5,6,7])
```

In [33]:

```
df1
```

Out[33]:

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3

In [34]:

```
df2
```

Out[34]:

	A	B	C	D
4	A4	B4	C0	D4
5	A5	B5	C3	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7

In [35]:

```
result=pd.concat([df1,df2])
```

In [36]:

```
result
```

Out[36]:

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3
4	A4	B4	C0	D4
5	A5	B5	C3	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7

In [37]:

```
merge_df=pd.merge(df1,df2,on='C',how='inner')
```

In [38]:

```
merge_df
```

Out[38]:

	A_x	B_x	C	D_x	A_y	B_y	D_y
0	A0	B0	C0	D0	A4	B4	D4
1	A3	B3	C3	D3	A5	B5	D5

In [39]:

```
merge_df=pd.merge(df1,df2,on='C',how='left')
```

In [40]:

```
merge_df
```

Out[40]:

	A_x	B_x	C	D_x	A_y	B_y	D_y
0	A0	B0	C0	D0	A4	B4	D4
1	A1	B1	C1	D1	NaN	NaN	NaN
2	A2	B2	C2	D2	NaN	NaN	NaN
3	A3	B3	C3	D3	A5	B5	D5

In [41]:

```
merge_df=pd.merge(df1,df2,on='C',how='right')
```

In [42]:

```
merge_df
```

Out[42]:

	A_x	B_x	C	D_x	A_y	B_y	D_y
0	A0	B0	C0	D0	A4	B4	D4
1	A3	B3	C3	D3	A5	B5	D5
2	NaN	NaN	C6	NaN	A6	B6	D6
3	NaN	NaN	C7	NaN	A7	B7	D7

In [43]:

```
merge_df=pd.merge(df1,df2,on='C',how='outer')
```

In [44]:

```
merge_df
```

Out[44]:

	A_x	B_x	C	D_x	A_y	B_y	D_y
0	A0	B0	C0	D0	A4	B4	D4
1	A1	B1	C1	D1	NaN	NaN	NaN
2	A2	B2	C2	D2	NaN	NaN	NaN
3	A3	B3	C3	D3	A5	B5	D5
4	NaN	NaN	C6	NaN	A6	B6	D6
5	NaN	NaN	C7	NaN	A7	B7	D7

In [45]:

```
s = pd.Series(list("abcdef"), index=[49, 48, 47, 0, 1, 2])
```

In [46]:

```
s
```

Out[46]:

```
49    a
48    b
47    c
0     d
1     e
2     f
dtype: object
```

In [47]:

```
s.loc[0] #row/index 0
```

Out[47]:

```
'd'
```

In [48]:

```
s.iloc[0] #0th location
```

Out[48]:

```
'a'
```

In [49]:

```
#s.loc[3] ----error coz there is no index as 3
s.iloc[3]
```

Out[49]:

```
'd'
```

In [50]:

```
s.loc[0:1]
```

Out[50]:

```
0    d
1    e
dtype: object
```

In [51]:

```
s.iloc[0:1]
```

Out[51]:

```
49    a
dtype: object
```

In [52]:

```
s.iloc[0:2]
```

Out[52]:

```
49    a
48    b
dtype: object
```

In [53]:

```
s=s.sort_index() #sorting the series.
```

In [54]:

```
s
```

Out[54]:

```
0     d
1     e
2     f
47    c
48    b
49    a
dtype: object
```

In [55]:

```
s.loc[0] #same value after series got sorted
```

Out[55]:

```
'd'
```

In [56]:

```
s.iloc[0] #different value after series got sorted.
```

Out[56]:

```
'd'
```

In [57]:

```
s.iloc[3] #different value after series got sorted.
```

Out[57]:

```
'c'
```

In [58]:

```
s.loc[0:1] #same value after series got sorted
```

Out[58]:

```
0    d
1    e
dtype: object
```

In [59]:

```
s.iloc[0:1] #different value after series got sorted.
```

Out[59]:

```
0    d
dtype: object
```

In [60]:

```
s.iloc[0:2] #different value after series got sorted.
```

Out[60]:

```
0    d
1    e
dtype: object
```

In [ ]:

```

```