



## What is Pandas?

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

## Why Use Pandas?

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

Relevant data is very important in data science.

More information about Pandas [here](#)

## Create a DataFrame

	Amit	Ramen
Ice Cream	I liked it.	Pretty good.
Burger	It was awful.	Bland.

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

df = pd.DataFrame({'Amit': ['I liked it.', 'It was awful.'],
                  'Ramen': ['Pretty good.', 'Bland.']],
                  index=['Ice Cream', 'Burger'])

df
```

```
Out[6]:
```

	Amit	Ramen
Ice Cream	I liked it.	Pretty good.
Burger	It was awful.	Bland.

```
In [7]: import numpy as np

data = pd.DataFrame({'Name': ['Arunesh', 'Abhinav', 'Adela', 'Amit', 'Arghadyuti', 'Prabhat', 'Ramen', 'Reena', 'Shrav'],
                    'Favorite Number': np.random.randint(123, 234, size=(11,)),
                    'Random Number': np.random.randint(0, 100, size=(11,)),
                    'Color': ['red', 'green', 'blue', 'black', 'yellow', 'magenta', 'pink', 'magenta', 'orange', 'viol']})

data.to_csv('virginia.csv', index=False) #way to save as csv(comma seperated values)
```

```
In [10]: data_read = pd.read_csv('virginia.csv')
```

```
#data_read = data_read.drop(columns=['Unnamed: 0'])  
data_read
```

Out[10]:

	Name	Favorite Number	Random Number	Color
0	Arunesh	172	47	red
1	Abhinav	138	91	green
2	Adela	210	27	blue
3	Amit	150	35	black
4	Arghadyuti	176	76	yellow
5	Prabhat	145	98	magenta
6	Ramen	157	3	pink
7	Reena	155	93	magenta
8	Shravan	208	10	orange
9	Soumya	210	22	violet
10	Subhashish	201	53	cyan

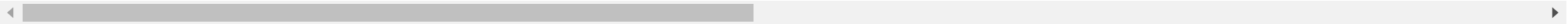
In [11]:

```
fifa_22 = pd.read_csv('/home/suman/Downloads/FIFA22_official_data.csv')  
#fifa_22.sort_values(['Overall'], ascending=False)  
fifa_22
```

Out[11]:

	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club
0	212198	Bruno Fernandes	26	https://cdn.sofifa.com/players/212/198/22_60.png	Portugal	https://cdn.sofifa.com/flags/pt.png	88	89	Manchester United
1	209658	L. Goretzka	26	https://cdn.sofifa.com/players/209/658/22_60.png	Germany	https://cdn.sofifa.com/flags/de.png	87	88	FC Bayern München
2	176580	L. Suárez	34	https://cdn.sofifa.com/players/176/580/22_60.png	Uruguay	https://cdn.sofifa.com/flags/uy.png	88	88	Atlético de Madrid
3	192985	K. De Bruyne	30	https://cdn.sofifa.com/players/192/985/22_60.png	Belgium	https://cdn.sofifa.com/flags/be.png	91	91	Manchester City
4	224334	M. Acuña	29	https://cdn.sofifa.com/players/224/334/22_60.png	Argentina	https://cdn.sofifa.com/flags/ar.png	84	84	Sevilla FC
...	...	...	...	...	...	...	...	...	...
16705	240558	18 L. Clayton	17	https://cdn.sofifa.com/players/240/558/18_60.png	England	https://cdn.sofifa.com/flags/gb-eng.png	53	70	Cheltenham Town
16706	262846	🇷🇴. Dobre	20	https://cdn.sofifa.com/players/262/846/22_60.png	Romania	https://cdn.sofifa.com/flags/ro.png	53	63	FC Academica Clinceni
16707	241317	21 Xue Qinghao	19	https://cdn.sofifa.com/players/241/317/21_60.png	China PR	https://cdn.sofifa.com/flags/cn.png	47	60	Shanghai Shenhua FC
16708	259646	A. Shaikh	18	https://cdn.sofifa.com/players/259/646/22_60.png	India	https://cdn.sofifa.com/flags/in.png	47	67	ATK Mohun Bagan FC
16709	178453	07 A. Censori	17	https://cdn.sofifa.com/players/178/453/07_60.png	Italy	https://cdn.sofifa.com/flags/it.png	28	38	Arezzo

16710 rows × 65 columns



# Read values from DataFrame

read rows

In [12]:

```
data_read
```

Out[12]:

	Name	Favorite Number	Random Number	Color
0	Arunesh	172	47	red
1	Abhinav	138	91	green
2	Adela	210	27	blue
3	Amit	150	35	black
4	Arghadyuti	176	76	yellow
5	Prabhat	145	98	magenta
6	Ramen	157	3	pink
7	Reena	155	93	magenta
8	Shravan	208	10	orange
9	Soumya	210	22	violet
10	Subhashish	201	53	cyan

In [39]:

```
# using the dataframe itself  
data_read[0:5]
```

Out[39]:

	Name	Favorite Number	Random Number	Color	Country
0	Arunesh	172	47	red	UK
1	Abhinav	138	91	green	Australia
2	Adela	210	27	blue	UAE
3	Amit	150	35	black	Australia
4	Arghadyuti	176	76	yellow	UAE

In [40]:

```
data_read.iloc[0]
```

```
Out[40]: Name          Arunesh
Favorite Number      172
Random Number        47
Color                red
Country              UK
Name: 0, dtype: object
```

```
In [41]: #using loc

data_read.loc[data_read['Name']=='Arghadyuti']
```

```
Out[41]:
```

	Name	Favorite Number	Random Number	Color	Country
4	Arghadyuti	176	76	yellow	UAE

```
In [17]: import numpy as np
data_read.loc[np.logical_or(data_read['Favorite Number']==206, data_read['Color']=='red')]
```

```
Out[17]:
```

	Name	Favorite Number	Random Number	Color
0	Arunesh	172	47	red

## Question

find the players who play for both Argentina(Country) and Paris Saint-Germain F.C. (Club)

```
In [42]: fifa_22.loc[np.logical_and(fifa_22['Nationality']=='Argentina', fifa_22['Club']=='Paris Saint-Germain')]
```

Out[42]:

	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club
29	158023	L. Messi	34	https://cdn.sofifa.com/players/158/023/22_60.png	Argentina	https://cdn.sofifa.com/flags/ar.png	93	93	Paris Saint-Germain
76	183898	A. Di María	33	https://cdn.sofifa.com/players/183/898/22_60.png	Argentina	https://cdn.sofifa.com/flags/ar.png	87	87	Paris Saint-Germain
81	207439	L. Paredes	27	https://cdn.sofifa.com/players/207/439/22_60.png	Argentina	https://cdn.sofifa.com/flags/ar.png	81	83	Paris Saint-Germain
1959	201399	M. Icardi	28	https://cdn.sofifa.com/players/201/399/22_60.png	Argentina	https://cdn.sofifa.com/flags/ar.png	83	83	Paris Saint-Germain

4 rows × 65 columns



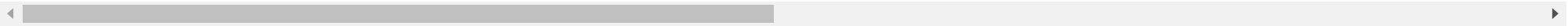
In [43]:

```
fifa_22.loc[np.logical_and(fifa_22['Nationality']=='Brazil', fifa_22['Club']=='Paris Saint-Germain')]
```

Out[43]:

	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club
40	146530	19 Dani Alves	35	https://cdn.sofifa.com/players/146/530/19_60.png	Brazil	https://cdn.sofifa.com/flags/br.png	82	82	Paris Saint-Germain
64	190871	Neymar Jr	29	https://cdn.sofifa.com/players/190/871/22_60.png	Brazil	https://cdn.sofifa.com/flags/br.png	91	91	Paris Saint-Germain
185	201400	Rafinha	28	https://cdn.sofifa.com/players/201/400/22_60.png	Brazil	https://cdn.sofifa.com/flags/br.png	80	80	Paris Saint-Germain
212	53405	17 Maxwell	34	https://cdn.sofifa.com/players/053/405/17_60.png	Brazil	https://cdn.sofifa.com/flags/br.png	79	79	Paris Saint-Germain
371	207865	Marquinhos	27	https://cdn.sofifa.com/players/207/865/22_60.png	Brazil	https://cdn.sofifa.com/flags/br.png	87	90	Paris Saint-Germain

5 rows × 65 columns



# Read Columns

gather data about column

In [22]:

```
data_read['Name']
```



```
Out[22]: 0      Arunesh
1      Abhinav
2      Adela
3      Amit
4      Arghadyuti
5      Prabhat
6      Ramen
7      Reena
8      Shravan
9      Soumya
10     Subhashish
Name: Name, dtype: object
```

```
In [23]: fifa_22['Club'][0:5]
```

```
Out[23]: 0      Manchester United
1      FC Bayern München
2      Atlético de Madrid
3      Manchester City
4      Sevilla FC
Name: Club, dtype: object
```

## Add a New Column

```
In [24]: country = ['UK', 'Australia', 'UAE', 'Australia', 'UAE', 'USA', 'USA', 'UK', 'UK', 'UK', 'UK' ]

data_read['Country'] = country

d = data_read.sort_values(['Favorite Number'])  #sort_values
```

```
In [25]: d
```

Out[25]:

	Name	Favorite Number	Random Number	Color	Country
1	Abhinav	138	91	green	Australia
5	Prabhat	145	98	magenta	USA
3	Amit	150	35	black	Australia
7	Reena	155	93	magenta	UK
6	Ramen	157	3	pink	USA
0	Arunesh	172	47	red	UK
4	Arghadyuti	176	76	yellow	UAE
10	Subhashish	201	53	cyan	UK
8	Shravan	208	10	orange	UK
2	Adela	210	27	blue	UAE
9	Soumya	210	22	violet	UK

## Question

we want he sort the FIFA 22 dataset based on overall score

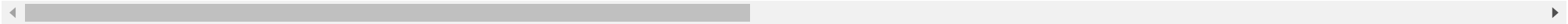
In [26]:

```
fifa_22.sort_values(['Overall'], ascending=False)
```

Out[26]:

	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Cl
29	158023	L. Messi	34	https://cdn.sofifa.com/players/158/023/22_60.png	Argentina	https://cdn.sofifa.com/flags/ar.png	93	93	Paris Sai Germ
33	188545	R. Lewandowski	32	https://cdn.sofifa.com/players/188/545/22_60.png	Poland	https://cdn.sofifa.com/flags/pl.png	92	92	FC Bay Münch
14244	200389	J. Oblak	28	https://cdn.sofifa.com/players/200/389/22_60.png	Slovenia	https://cdn.sofifa.com/flags/si.png	91	93	Atlético Mac
3	192985	K. De Bruyne	30	https://cdn.sofifa.com/players/192/985/22_60.png	Belgium	https://cdn.sofifa.com/flags/be.png	91	91	Manches C
64	190871	Neymar Jr	29	https://cdn.sofifa.com/players/190/871/22_60.png	Brazil	https://cdn.sofifa.com/flags/br.png	91	91	Paris Sai Germ
...	...	...	...	...	...	...	...	...	
15593	235352	18 T. Käßermodel	28	https://cdn.sofifa.com/players/235/352/18_60.png	Germany	https://cdn.sofifa.com/flags/de.png	46	46	Erzgebir A
15685	219735	15 T. Fletcher	19	https://cdn.sofifa.com/players/219/735/15_60.png	England	https://cdn.sofifa.com/flags/gb-eng.png	46	52	Wycom Wander
16572	19334	10 I. Baraclough	38	https://cdn.sofifa.com/players/019/334/10_60.png	England	https://cdn.sofifa.com/flags/gb-eng.png	44	65	N
15999	220806	16 E. Redman	18	https://cdn.sofifa.com/players/220/806/16_60.png	Wales	https://cdn.sofifa.com/flags/gb-wls.png	44	57	Newp Cou
16709	178453	07 A. Censori	17	https://cdn.sofifa.com/players/178/453/07_60.png	Italy	https://cdn.sofifa.com/flags/it.png	28	38	Arez

16710 rows × 65 columns



In [27]:

```
new_data = pd.DataFrame({'Name': ['Suman'], 'Favorite Number': [np.nan], 'Random Number': [''], 'Color': 'Olive', 'Country': 'Japan'})
new_data
```

Out[27]:

	Name	Favorite Number	Random Number	Color	Country
0	Suman	NaN		Olive	Japan

```
In [28]: data_new = pd.concat([data_read, new_data], ignore_index=True)
data_new
```

```
Out[28]:
```

	Name	Favorite Number	Random Number	Color	Country
0	Arunesh	172.0	47	red	UK
1	Abhinav	138.0	91	green	Australia
2	Adela	210.0	27	blue	UAE
3	Amit	150.0	35	black	Australia
4	Arghadyuti	176.0	76	yellow	UAE
5	Prabhat	145.0	98	magenta	USA
6	Ramen	157.0	3	pink	USA
7	Reena	155.0	93	magenta	UK
8	Shravan	208.0	10	orange	UK
9	Soumya	210.0	22	violet	UK
10	Subhashish	201.0	53	cyan	UK
11	Suman	NaN		Olive	Japan

```
In [29]: data_new.drop(0)
```

Out[29]:

	Name	Favorite Number	Random Number	Color	Country
1	Abhinav	138.0	91	green	Australia
2	Adela	210.0	27	blue	UAE
3	Amit	150.0	35	black	Australia
4	Arghadyuti	176.0	76	yellow	UAE
5	Prabhat	145.0	98	magenta	USA
6	Ramen	157.0	3	pink	USA
7	Reena	155.0	93	magenta	UK
8	Shravan	208.0	10	orange	UK
9	Soumya	210.0	22	violet	UK
10	Subhashish	201.0	53	cyan	UK
11	Suman	NaN		Olive	Japan

## Clean Up data

### Remove empty cells, NaN values

In [30]:

```
cleaned_data = data_new.dropna()  
  
cleaned_data
```

Out[30]:

	Name	Favorite Number	Random Number	Color	Country
0	Arunesh	172.0	47	red	UK
1	Abhinav	138.0	91	green	Australia
2	Adela	210.0	27	blue	UAE
3	Amit	150.0	35	black	Australia
4	Arghadyuti	176.0	76	yellow	UAE
5	Prabhat	145.0	98	magenta	USA
6	Ramen	157.0	3	pink	USA
7	Reena	155.0	93	magenta	UK
8	Shravan	208.0	10	orange	UK
9	Soumya	210.0	22	violet	UK
10	Subhashish	201.0	53	cyan	UK

## drop unnecessary columns

In [31]:

```
cleaned_data2 = cleaned_data.drop(columns=['Random Number'])
cleaned_data2[0:5]
```

Out[31]:

	Name	Favorite Number	Color	Country
0	Arunesh	172.0	red	UK
1	Abhinav	138.0	green	Australia
2	Adela	210.0	blue	UAE
3	Amit	150.0	black	Australia
4	Arghadyuti	176.0	yellow	UAE

In [32]:

```
fifa_23 = fifa_22.drop(columns=['GKReflexes', 'GKPositioning', 'GKKicking', 'GKHandling', 'GKDividing', 'SlidingTackle',
                                ])
fifa_23_top = fifa_23.sort_values(['Overall'], ascending=False)[0:10]
```

fifa\_23\_top

Out[32]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Jersey Number	Penalties	Best Position	Best Overall Rating	Release Clause	DefensiveAwareness
29	L. Messi	34	Argentina	93	93	Paris Saint-Germain	€78M	€320K	30.0	75.0	RW	93.0	€144.3M	
33	R. Lewandowski	32	Poland	92	92	FC Bayern München	€119.5M	€270K	9.0	90.0	ST	92.0	€197.2M	
14244	J. Oblak	28	Slovenia	91	93	Atlético de Madrid	€112M	€130K	13.0	11.0	GK	91.0	€238M	
3	K. De Bruyne	30	Belgium	91	91	Manchester City	€125.5M	€350K	17.0	83.0	CM	91.0	€232.2M	
64	Neymar Jr	29	Brazil	91	91	Paris Saint-Germain	€129M	€270K	10.0	93.0	LW	91.0	€238.7M	
82	K. Mbappé	22	France	91	95	Paris Saint-Germain	€194M	€230K	7.0	79.0	ST	92.0	€373.5M	
36	Cristiano Ronaldo	36	Portugal	91	91	Manchester United	€45M	€270K	7.0	88.0	ST	91.0	€83.3M	
39	H. Kane	27	England	90	90	Tottenham Hotspur	€129.5M	€240K	10.0	91.0	ST	90.0	€246.1M	
13890	M. ter Stegen	29	Germany	90	92	FC Barcelona	€99M	€250K	1.0	25.0	GK	90.0	€210.4M	
71	N. Kanté	30	France	90	90	Chelsea	€100M	€230K	7.0	54.0	CDM	90.0	€185M	

In [33]: `fifa_23_top['Club'].unique()`

Out[33]: `array(['Paris Saint-Germain', 'FC Bayern München', 'Atlético de Madrid', 'Manchester City', 'Manchester United', 'Tottenham Hotspur', 'FC Barcelona', 'Chelsea'], dtype=object)`

In [36]: `fifa_23.groupby(['Nationality']).mean()`

Out[36]:

	Age	Overall	Potential	Jersey Number	Penalties	Best Overall Rating	DefensiveAwareness
Nationality							
Afghanistan	25.500000	65.000000	67.500000	9.000000	56.000000	65.500000	47.500000
Albania	25.770833	67.583333	71.812500	20.583333	52.687500	68.500000	51.444444
Algeria	27.120690	71.517241	74.034483	22.051724	57.948276	72.258621	49.452830
Andorra	31.000000	64.000000	64.000000	3.000000	45.000000	64.000000	56.000000
Angola	25.100000	69.050000	74.550000	20.150000	55.200000	69.800000	46.950000
...	...	...	...	...	...	...	...
Venezuela	24.292135	66.707865	73.898876	19.449438	50.382022	67.719101	46.367816
Vietnam	25.500000	63.000000	66.500000	8.000000	36.500000	62.500000	9.000000
Wales	25.297468	63.930380	69.493671	18.411392	48.962025	64.987342	49.186207
Zambia	23.700000	67.800000	73.900000	23.000000	59.700000	69.000000	38.666667
Zimbabwe	26.166667	67.500000	71.083333	20.166667	52.833333	68.666667	49.363636

168 rows × 7 columns

In [37]:

```
fifa_23.groupby(['Nationality']).mean().sort_values('Overall', ascending=False)[0:10]
```



Out[37]:

	Age	Overall	Potential	Jersey Number	Penalties	Best Overall Rating	DefensiveAwareness
Nationality							
Tanzania	28.000000	74.000000	74.000000	70.000000	69.000000	74.000000	32.000000
Mozambique	30.750000	72.750000	73.000000	15.000000	42.250000	72.750000	72.000000
Brazil	26.927961	72.411477	74.910867	21.150551	57.239316	73.045177	49.959270
Libya	27.250000	72.250000	73.250000	12.000000	53.250000	73.250000	50.500000
Fiji	33.000000	72.000000	72.000000	21.000000	74.000000	73.000000	41.000000
Namibia	23.000000	72.000000	80.000000	2.000000	36.000000	72.000000	70.000000
Egypt	27.071429	71.678571	73.642857	16.962963	55.000000	72.392857	51.259259
Czech Republic	26.524752	71.653465	74.900990	19.584158	49.534653	72.425743	48.697917
Ukraine	25.913043	71.550725	76.492754	25.318841	47.913043	72.405797	48.476190
Algeria	27.120690	71.517241	74.034483	22.051724	57.948276	72.258621	49.452830