

Pandas-DataFrame 1

December 23, 2022

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

```
[2]: # create data frame using list
student_data = [
    [100,70,7],
    [90,69,6],
    [80,78,8]
]
students=pd.DataFrame(student_data,columns=['iq','marks','package'])
```

```
[3]: # create dataframe using dict
student_dict = {
    'iq':[100,90,80,70],
    'marks':[80,70,85,79],
    'package':[10,7,8,9]
}
pd.DataFrame(student_dict)
```

```
[3]:    iq  marks  package
0  100     80         10
1   90     70          7
2   80     85          8
3   70     79          9
```

```
[4]: #using read csv
pd.read_csv('insurance.csv')
```

```
[4]:    age  sex  bmi  children  smoker  region  expenses
0    19 female  27.9         0     yes southwest  16884.92
1    18  male  33.8         1     no  southeast   1725.55
2    28  male  33.0         3     no  southeast   4449.46
3    33  male  22.7         0     no northwest  21984.47
4    32  male  28.9         0     no northwest   3866.86
...  ...  ...  ...  ...  ...  ...  ...
1333  50  male  31.0         3     no northwest  10600.55
1334  18 female  31.9         0     no northeast   2205.98
```

1335	18	female	36.9	0	no	southeast	1629.83
1336	21	female	25.8	0	no	southwest	2007.95
1337	61	female	29.1	0	yes	northwest	29141.36

[1338 rows x 7 columns]

```
[5]: data=pd.read_csv('insurance.csv')
data.head()
```

```
[5]:
```

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
1	18	male	33.8	1	no	southeast	1725.55
2	28	male	33.0	3	no	southeast	4449.46
3	33	male	22.7	0	no	northwest	21984.47
4	32	male	28.9	0	no	northwest	3866.86

```
[6]: pd.read_csv('ipl-matches.csv')
```

```
[6]:
```

	ID	City	Date	Season	MatchNumber	\
0	1312200	Ahmedabad	2022-05-29	2022	Final	
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	
2	1312198	Kolkata	2022-05-25	2022	Eliminator	
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	
4	1304116	Mumbai	2022-05-22	2022	70	
..	
945	335986	Kolkata	2008-04-20	2007/08	4	
946	335985	Mumbai	2008-04-20	2007/08	5	
947	335984	Delhi	2008-04-19	2007/08	3	
948	335983	Chandigarh	2008-04-19	2007/08	2	
949	335982	Bangalore	2008-04-18	2007/08	1	

	Team1	Team2	\
0	Rajasthan Royals	Gujarat Titans	
1	Royal Challengers Bangalore	Rajasthan Royals	
2	Royal Challengers Bangalore	Lucknow Super Giants	
3	Rajasthan Royals	Gujarat Titans	
4	Sunrisers Hyderabad	Punjab Kings	
..	
945	Kolkata Knight Riders	Deccan Chargers	
946	Mumbai Indians	Royal Challengers Bangalore	
947	Delhi Daredevils	Rajasthan Royals	
948	Kings XI Punjab	Chennai Super Kings	
949	Royal Challengers Bangalore	Kolkata Knight Riders	

	Venue	TossWinner	\
0	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	
1	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	

2	Eden Gardens, Kolkata	Lucknow Super Giants
3	Eden Gardens, Kolkata	Gujarat Titans
4	Wankhede Stadium, Mumbai	Sunrisers Hyderabad
..
945	Eden Gardens	Deccan Chargers
946	Wankhede Stadium	Mumbai Indians
947	Feroz Shah Kotla	Rajasthan Royals
948	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
949	M Chinnaswamy Stadium	Royal Challengers Bangalore

	TossDecision	SuperOver	WinningTeam	WonBy	Margin	\
0	bat	N	Gujarat Titans	Wickets	7.0	
1	field	N	Rajasthan Royals	Wickets	7.0	
2	field	N	Royal Challengers Bangalore	Runs	14.0	
3	field	N	Gujarat Titans	Wickets	7.0	
4	bat	N	Punjab Kings	Wickets	5.0	
..	
945	bat	N	Kolkata Knight Riders	Wickets	5.0	
946	bat	N	Royal Challengers Bangalore	Wickets	5.0	
947	bat	N	Delhi Daredevils	Wickets	9.0	
948	bat	N	Chennai Super Kings	Runs	33.0	
949	field	N	Kolkata Knight Riders	Runs	140.0	

	method	Player_of_Match	Team1Players	\
0	NaN	HH Pandya	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	
1	NaN	JC Buttler	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...	
2	NaN	RM Patidar	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...	
3	NaN	DA Miller	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	
4	NaN	Harpreet Brar	['PK Garg', 'Abhishek Sharma', 'RA Tripathi', ...	
..	
945	NaN	DJ Hussey	['WP Saha', 'BB McCullum', 'RT Ponting', 'SC G...	
946	NaN	MV Boucher	['L Ronchi', 'ST Jayasuriya', 'DJ Thorneley', '...	
947	NaN	MF Maharooof	['G Gambhir', 'V Sehwag', 'S Dhawan', 'MK Tiwa...	
948	NaN	MEK Hussey	['K Goel', 'JR Hopes', 'KC Sangakkara', 'Yuvra...	
949	NaN	BB McCullum	['R Dravid', 'W Jaffer', 'V Kohli', 'JH Kallis...	

	Team2Players	Umpire1	\
0	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	CB Gaffaney	
1	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	CB Gaffaney	
2	['Q de Kock', 'KL Rahul', 'M Vohra', 'DJ Hooda...	J Madanagopal	
3	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	BNJ Oxenford	
4	['JM Bairstow', 'S Dhawan', 'M Shahrukh Khan', ...	AK Chaudhary	
..	
945	['AC Gilchrist', 'Y Venugopal Rao', 'VVS Laxma...	BF Bowden	
946	['S Chanderpaul', 'R Dravid', 'LRPL Taylor', '...	SJ Davis	
947	['T Kohli', 'YK Pathan', 'SR Watson', 'M Kaif'...	Aleem Dar	
948	['PA Patel', 'ML Hayden', 'MEK Hussey', 'MS Dh...	MR Benson	

949 ['SC Ganguly', 'BB McCullum', 'RT Ponting', 'D... Asad Rauf

Umpire2
 0 Nitin Menon
 1 Nitin Menon
 2 MA Gough
 3 VK Sharma
 4 NA Patwardhan
 ..
 945 K Hariharan
 946 DJ Harper
 947 GA Pratapkumar
 948 SL Shastri
 949 RE Koertzen

[950 rows x 20 columns]

```
[7]: ipl = pd.read_csv('ipl-matches.csv')
      ipl.head()
```

```
[7]:      ID      City      Date Season  MatchNumber \
0  1312200  Ahmedabad  2022-05-29   2022         Final
1  1312199  Ahmedabad  2022-05-27   2022  Qualifier 2
2  1312198   Kolkata  2022-05-25   2022  Eliminator
3  1312197   Kolkata  2022-05-24   2022  Qualifier 1
4  1304116   Mumbai  2022-05-22   2022          70
```

```
      Team1      Team2 \
0  Rajasthan Royals  Gujarat Titans
1  Royal Challengers Bangalore  Rajasthan Royals
2  Royal Challengers Bangalore  Lucknow Super Giants
3  Rajasthan Royals  Gujarat Titans
4  Sunrisers Hyderabad  Punjab Kings
```

```
      Venue      TossWinner TossDecision \
0  Narendra Modi Stadium, Ahmedabad  Rajasthan Royals  bat
1  Narendra Modi Stadium, Ahmedabad  Rajasthan Royals  field
2  Eden Gardens, Kolkata  Lucknow Super Giants  field
3  Eden Gardens, Kolkata  Gujarat Titans  field
4  Wankhede Stadium, Mumbai  Sunrisers Hyderabad  bat
```

```
      SuperOver      WinningTeam  WonBy  Margin method \
0  N  Gujarat Titans  Wickets  7.0  NaN
1  N  Rajasthan Royals  Wickets  7.0  NaN
2  N  Royal Challengers Bangalore  Runs  14.0  NaN
3  N  Gujarat Titans  Wickets  7.0  NaN
4  N  Punjab Kings  Wickets  5.0  NaN
```

	Player_of_Match	Team1Players \	Team2Players	Umpire1 \	Umpire2
0	HH Pandya	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...		CB Gaffaney	Nitin Menon
1	JC Buttler	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...		CB Gaffaney	Nitin Menon
2	RM Patidar	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...		J Madanagopal	MA Gough
3	DA Miller	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...		BNJ Oxenford	VK Sharma
4	Harpreet Brar	['PK Garg', 'Abhishek Sharma', 'RA Tripathi', ...		AK Chaudhary	NA Patwardhan

DataFrame Attributes and Methods

```
[8]: #shape
data.shape
```

```
[8]: (1338, 7)
```

```
[9]: ipl.shape
```

```
[9]: (950, 20)
```

```
[10]: #dtypes gives series and its types of every
data.dtypes
```

```
[10]: age          int64
sex            object
bmi           float64
children       int64
smoker         object
region         object
expenses       float64
dtype: object
```

```
[11]: ipl.dtypes
```

```
[11]: ID                int64
      City              object
      Date              object
      Season            object
      MatchNumber       object
      Team1             object
      Team2             object
      Venue             object
      TossWinner        object
      TossDecision       object
      SuperOver         object
      WinningTeam       object
      WonBy             object
      Margin            float64
      method            object
      Player_of_Match   object
      Team1Players      object
      Team2Players      object
      Umpire1           object
      Umpire2           object
      dtype: object
```

```
[12]: #index
      data.index
```

```
[12]: RangeIndex(start=0, stop=1338, step=1)
```

```
[13]: ipl.index
```

```
[13]: RangeIndex(start=0, stop=950, step=1)
```

```
[14]: # columns
      data.columns
```

```
[14]: Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'expenses'],
      dtype='object')
```

```
[15]: ipl.columns
```

```
[15]: Index(['ID', 'City', 'Date', 'Season', 'MatchNumber', 'Team1', 'Team2',
      'Venue', 'TossWinner', 'TossDecision', 'SuperOver', 'WinningTeam',
      'WonBy', 'Margin', 'method', 'Player_of_Match', 'Team1Players',
      'Team2Players', 'Umpire1', 'Umpire2'],
      dtype='object')
```

```
[16]: #values
      ipl.values
```

```
[16]: array([[1312200, 'Ahmedabad', '2022-05-29', ...,
          '['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pandya', 'DA Miller', 'R
          Tewatia', 'Rashid Khan', 'R Sai Kishore', 'LH Ferguson', 'Yash Dayal', 'Mohammed
          Shami']"',
          'CB Gaffaney', 'Nitin Menon'],
          [1312199, 'Ahmedabad', '2022-05-27', ...,
          '['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D Padikkal', 'SO Hetmyer',
          'R Parag', 'R Ashwin', 'TA Boult', 'YS Chahal', 'M Prasidh Krishna', 'OC
          McCoy']"',
          'CB Gaffaney', 'Nitin Menon'],
          [1312198, 'Kolkata', '2022-05-25', ...,
          '['Q de Kock', 'KL Rahul', 'M Vohra', 'DJ Hooda', 'MP Stoinis', 'E
          Lewis', 'KH Pandya', 'PVD Chameera', 'Mohsin Khan', 'Avesh Khan', 'Ravi
          Bishnoi']"',
          'J Madanagopal', 'MA Gough'],
          ...,
          [335984, 'Delhi', '2008-04-19', ...,
          '['T Kohli', 'YK Pathan', 'SR Watson', 'M Kaif', 'DS Lehmann', 'RA
          Jadeja', 'M Rawat', 'D Salunkhe', 'SK Warne', 'SK Trivedi', 'MM Patel']"',
          'Aleem Dar', 'GA Pratapkumar'],
          [335983, 'Chandigarh', '2008-04-19', ...,
          '['PA Patel', 'ML Hayden', 'MEK Hussey', 'MS Dhoni', 'SK Raina', 'JDP
          Oram', 'S Badrinath', 'Joginder Sharma', 'P Amarnath', 'MS Gony', 'M
          Muralitharan']"',
          'MR Benson', 'SL Shastri'],
          [335982, 'Bangalore', '2008-04-18', ...,
          '['SC Ganguly', 'BB McCullum', 'RT Ponting', 'DJ Hussey', 'Mohammad
          Hafeez', 'LR Shukla', 'WP Saha', 'AB Agarkar', 'AB Dinda', 'M Kartik', 'I
          Sharma']"',
          'Asad Rauf', 'RE Koertzen']]), dtype=object)
```

```
[17]: #head and tail
data.head()
```

```
[17]:   age    sex    bmi  children  smoker    region  expenses
0   19  female  27.9         0     yes  southwest  16884.92
1   18   male  33.8         1     no   southeast  1725.55
2   28   male  33.0         3     no   southeast  4449.46
3   33   male  22.7         0     no  northwest  21984.47
4   32   male  28.9         0     no  northwest  3866.86
```

```
[18]: ipl.head()
```

```
[18]:   ID      City      Date Season  MatchNumber  \
0  1312200  Ahmedabad  2022-05-29   2022         Final
1  1312199  Ahmedabad  2022-05-27   2022  Qualifier 2
2  1312198   Kolkata  2022-05-25   2022  Eliminator
```

3	1312197	Kolkata	2022-05-24	2022	Qualifier 1
4	1304116	Mumbai	2022-05-22	2022	70

	Team1	Team2 \
0	Rajasthan Royals	Gujarat Titans
1	Royal Challengers Bangalore	Rajasthan Royals
2	Royal Challengers Bangalore	Lucknow Super Giants
3	Rajasthan Royals	Gujarat Titans
4	Sunrisers Hyderabad	Punjab Kings

	Venue	TossWinner	TossDecision \
0	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	bat
1	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	field
2	Eden Gardens, Kolkata	Lucknow Super Giants	field
3	Eden Gardens, Kolkata	Gujarat Titans	field
4	Wankhede Stadium, Mumbai	Sunrisers Hyderabad	bat

	SuperOver	WinningTeam	WonBy	Margin	method \
0	N	Gujarat Titans	Wickets	7.0	NaN
1	N	Rajasthan Royals	Wickets	7.0	NaN
2	N	Royal Challengers Bangalore	Runs	14.0	NaN
3	N	Gujarat Titans	Wickets	7.0	NaN
4	N	Punjab Kings	Wickets	5.0	NaN

	Player_of_Match	Team1Players \
0	HH Pandya	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...
1	JC Buttler	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...
2	RM Patidar	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...
3	DA Miller	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...
4	Harpreet Brar	['PK Garg', 'Abhishek Sharma', 'RA Tripathi', ...

	Team2Players	Umpire1 \
0	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	CB Gaffaney
1	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	CB Gaffaney
2	['Q de Kock', 'KL Rahul', 'M Vohra', 'DJ Hooda...	J Madanagopal
3	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	BNJ Oxenford
4	['JM Bairstow', 'S Dhawan', 'M Shahrukh Khan', ...	AK Chaudhary

	Umpire2
0	Nitin Menon
1	Nitin Menon
2	MA Gough
3	VK Sharma
4	NA Patwardhan

```
[19]: #sample
data.sample(5)
```



```
[19]:      age      sex    bmi  children smoker      region  expenses
      1214    27  female  31.3         1     no  northwest   3956.07
      575    58  female  27.2         0     no  northwest  12222.90
      837    56  female  28.3         0     no  northeast  11657.72
      527    51  female  25.8         1     no  southwest   9861.03
      142    34   male   25.3         2    yes  southeast  18972.50
```

```
[20]: ipl.sample(5)
```

```
[20]:      ID      City      Date      Season  MatchNumber  \
134  1237181      NaN  2020-11-10  2020/21          Final
136  1237178  Abu Dhabi  2020-11-06  2020/21  Eliminator
135  1237180  Abu Dhabi  2020-11-08  2020/21  Qualifier 2
484   829719  Bangalore  2015-04-13    2015           8
330  1082634  Hyderabad  2017-05-06    2017          44

      Team1      Team2  \
134      Delhi Capitals      Mumbai Indians
136  Royal Challengers Bangalore      Sunrisers Hyderabad
135      Delhi Capitals      Sunrisers Hyderabad
484  Royal Challengers Bangalore      Sunrisers Hyderabad
330      Sunrisers Hyderabad  Rising Pune Supergiant

      Venue      TossWinner  \
134  Dubai International Cricket Stadium      Delhi Capitals
136      Sheikh Zayed Stadium  Sunrisers Hyderabad
135      Sheikh Zayed Stadium      Delhi Capitals
484      M Chinnaswamy Stadium  Sunrisers Hyderabad
330  Rajiv Gandhi International Stadium, Uppal  Sunrisers Hyderabad

      TossDecision SuperOver      WinningTeam  WonBy  Margin method  \
134      bat      N      Mumbai Indians  Wickets      5.0      NaN
136      field      N      Sunrisers Hyderabad  Wickets      6.0      NaN
135      bat      N      Delhi Capitals      Runs      17.0      NaN
484      field      N      Sunrisers Hyderabad  Wickets      8.0      NaN
330      field      N  Rising Pune Supergiant      Runs      12.0      NaN

      Player_of_Match      Team1Players  \
134      TA Boult  ['MP Stoinis', 'S Dhawan', 'AM Rahane', 'SS Iy...
136      KS Williamson  ['V Kohli', 'D Padikkal', 'AJ Finch', 'AB de V...
135      MP Stoinis  ['MP Stoinis', 'S Dhawan', 'SS Iyer', 'SO Hetm...
484      DA Warner  ['CH Gayle', 'V Kohli', 'KD Karthik', 'AB de V...
330      JD Unadkat  ['DA Warner', 'S Dhawan', 'KS Williamson', 'Yu...

      Team2Players      Umpire1  \
134  ['RG Sharma', 'Q de Kock', 'SA Yadav', 'Ishan ...      CB Gaffaney
136  ['DA Warner', 'SP Goswami', 'MK Pandey', 'KS W...
```

```

135 ['PK Garg', 'DA Warner', 'MK Pandey', 'KS Will... PR Reiffel
484 ['DA Warner', 'S Dhawan', 'KS Williamson', 'KL... RM Deshpande
330 ['AM Rahane', 'RA Tripathi', 'SPD Smith', 'BA ... KN Ananthapadmanabhan

        Umpire2
134     Nitin Menon
136         S Ravi
135         S Ravi
484 RK Illingworth
330     AK Chaudhary

```

```
[21]: #info - gives high level information about the dataset
data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   age          1338 non-null   int64
1   sex          1338 non-null   object
2   bmi          1338 non-null   float64
3   children     1338 non-null   int64
4   smoker       1338 non-null   object
5   region       1338 non-null   object
6   expenses     1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB

```

```
[22]: ipl.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 20 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ID          950 non-null    int64
1   City        899 non-null    object
2   Date        950 non-null    object
3   Season      950 non-null    object
4   MatchNumber 950 non-null    object
5   Team1       950 non-null    object
6   Team2       950 non-null    object
7   Venue       950 non-null    object
8   TossWinner  950 non-null    object
9   TossDecision 950 non-null    object
10  SuperOver   946 non-null    object
11  WinningTeam 946 non-null    object

```

```

12 WonBy          950 non-null    object
13 Margin         932 non-null    float64
14 method         19 non-null     object
15 Player_of_Match 946 non-null    object
16 Team1Players    950 non-null    object
17 Team2Players    950 non-null    object
18 Umpire1         950 non-null    object
19 Umpire2         950 non-null    object
dtypes: float64(1), int64(1), object(18)
memory usage: 148.6+ KB

```

```
[23]: #describe - it will give the mathematical summary of a of a numerical columns
data.describe()
```

```
[23]:
```

	age	bmi	children	expenses
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.665471	1.094918	13270.422414
std	14.049960	6.098382	1.205493	12110.011240
min	18.000000	16.000000	0.000000	1121.870000
25%	27.000000	26.300000	0.000000	4740.287500
50%	39.000000	30.400000	1.000000	9382.030000
75%	51.000000	34.700000	2.000000	16639.915000
max	64.000000	53.100000	5.000000	63770.430000

```
[24]: ipl.describe()
```

```
[24]:
```

	ID	Margin
count	9.500000e+02	932.000000
mean	8.304852e+05	17.056867
std	3.375678e+05	21.633109
min	3.359820e+05	1.000000
25%	5.012612e+05	6.000000
50%	8.297380e+05	8.000000
75%	1.175372e+06	19.000000
max	1.312200e+06	146.000000

```
[25]: #isnull
data.isnull()
```

```
[25]:
```

	age	sex	bmi	children	smoker	region	expenses
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...
1333	False	False	False	False	False	False	False

1334	False	False	False	False	False	False	False
1335	False	False	False	False	False	False	False
1336	False	False	False	False	False	False	False
1337	False	False	False	False	False	False	False

[1338 rows x 7 columns]

```
[26]: data.isnull().sum()
```

```
[26]: age          0
sex          0
bmi          0
children     0
smoker       0
region       0
expenses     0
dtype: int64
```

```
[27]: ipl.isnull().sum()
```

```
[27]: ID          0
City          51
Date          0
Season        0
MatchNumber   0
Team1         0
Team2         0
Venue         0
TossWinner    0
TossDecision  0
SuperOver     4
WinningTeam   4
WonBy         0
Margin        18
method        931
Player_of_Match 4
Team1Players  0
Team2Players  0
Umpire1       0
Umpire2       0
dtype: int64
```

```
[28]: #duplicated - tell about dpllicate value in dataframe
data.duplicated()
```

```
[28]: 0      False
      1      False
```

```

2      False
3      False
4      False
...
1333   False
1334   False
1335   False
1336   False
1337   False
Length: 1338, dtype: bool

```

```
[29]: data.duplicated().sum()
```

```
[29]: 1
```

```
[30]: data.duplicated().sum()
```

```
[30]: 1
```

```
[31]: #rename - rename the dataframe column name
students
```

```
[31]:
   iq  marks  package
0  100    70        7
1   90    69        6
2   80    78        8

```

```
[32]: students.rename(columns={'marks': 'percentage', 'package': 'lpa'})
```

```
[32]:
   iq  percentage  lpa
0  100         70    7
1   90         69    6
2   80         78    8

```

```
[33]: # students.rename(columns={'marks': 'percentage', 'package': 'lpa'}, inplace=True)
```

Mathematical function

```
[34]: #sum() - it will apply sum on every column in the dataframe
data.sum()
```

```
[34]: age                                52459
sex      femalemalemalemalefemalefemalefemalemalefe...
bmi                                41030.4
children                                1465
smoker    yesnonononononononononoyesnonoyesnonononoyesno...
region    southwestsoutheastseasteastnorthwestnorthwests...
```

```
expenses
dtype: object
```

17755825.19

```
[35]: students
```

```
[35]:
```

	iq	marks	package
0	100	70	7
1	90	69	6
2	80	78	8

```
[36]: students.sum()
```

```
[36]:
```

iq	270
marks	217
package	21

```
dtype: int64
```

```
[37]: #mean() - column wise it is.
students.mean()
```

```
[37]:
```

iq	90.000000
marks	72.333333
package	7.000000

```
dtype: float64
```

```
[38]: # for row wise mean
students.mean(axis=1)
```

```
[38]:
```

0	59.000000
1	55.000000
2	55.333333

```
dtype: float64
```

```
[39]: #min() - it is column wise
students.min()
```

```
[39]:
```

iq	80
marks	69
package	6

```
dtype: int64
```

```
[40]: #max()
students.max()
```

```
[40]:
```

iq	100
marks	78
package	8

dtype: int64

selecting columns from dataframe

```
[41]: #single columns
      print(data['bmi'])
      data['bmi']
```

```
0      27.9
1      33.8
2      33.0
3      22.7
4      28.9
...
1333   31.0
1334   31.9
1335   36.9
1336   25.8
1337   29.1
Name: bmi, Length: 1338, dtype: float64
```

```
[41]: 0      27.9
      1      33.8
      2      33.0
      3      22.7
      4      28.9
      ...
      1333   31.0
      1334   31.9
      1335   36.9
      1336   25.8
      1337   29.1
      Name: bmi, Length: 1338, dtype: float64
```

```
[42]: data.columns
```

```
[42]: Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'expenses'],
      dtype='object')
```

```
[43]: data['age']
```

```
[43]: 0      19
      1      18
      2      28
      3      33
      4      32
      ..
```

```

1333    50
1334    18
1335    18
1336    21
1337    61
Name: age, Length: 1338, dtype: int64

```

```
[44]: #multiple columns fetch
data[['age','smoker','region']]
```

```
[44]:
```

	age	smoker	region
0	19	yes	southwest
1	18	no	southeast
2	28	no	southeast
3	33	no	northwest
4	32	no	northwest
...
1333	50	no	northwest
1334	18	no	northeast
1335	18	no	southeast
1336	21	no	southwest
1337	61	yes	northwest

```
[1338 rows x 3 columns]
```

Selecting rows from DataFrame

- `iloc` = searches using index positions
- `loc` = searches using index labels

```
[45]: student_dict = {
        'name': ['A', 'B', 'C', 'D'],
        'iq': [100, 90, 80, 70],
        'marks': [80, 70, 85, 79],
        'package': [10, 7, 8, 9]
    }
s_data = pd.DataFrame(student_dict)
```

```
[46]: s_data
```

```
[46]:
```

	name	iq	marks	package
0	A	100	80	10
1	B	90	70	7
2	C	80	85	8
3	D	70	79	9

```
[47]: # set_index - To set a column as an index
```



```
[48]: s_data.set_index('name')
```

```
[48]:      iq  marks  package
name
A     100    80      10
B      90    70       7
C      80    85       8
D      70    79       9
```

```
[49]: s_data.set_index('name',inplace=True)
```

```
[50]: s_data
```

```
[50]:      iq  marks  package
name
A     100    80      10
B      90    70       7
C      80    85       8
D      70    79       9
```

```
[51]: #single row
data.iloc[0]
```

```
[51]: age          19
sex          female
bmi          27.9
children         0
smoker         yes
region    southwest
expenses    16884.92
Name: 0, dtype: object
```

```
[52]: #multiple rows
data.iloc[0:5] #first five moves
```

```
[52]:   age  sex  bmi  children  smoker  region  expenses
0   19  female  27.9         0    yes  southwest  16884.92
1   18   male  33.8         1    no   southeast   1725.55
2   28   male  33.0         3    no   southeast   4449.46
3   33   male  22.7         0    no  northwest  21984.47
4   32   male  28.9         0    no  northwest   3866.86
```

```
[53]: data.iloc[10:17]
```

```
[53]:   age  sex  bmi  children  smoker  region  expenses
10  25  male  26.2         0    no  northeast   2721.32
11  62  female  26.3         0   yes  southeast  27808.73
```

12	23	male	34.4	0	no	southwest	1826.84
13	56	female	39.8	0	no	southeast	11090.72
14	27	male	42.1	0	yes	southeast	39611.76
15	19	male	24.6	1	no	southwest	1837.24
16	52	female	30.8	1	no	northeast	10797.34

```
[54]: #want alternate moves between in range of 10 to 20
data.iloc[10:20:2]
```

```
[54]:
```

	age	sex	bmi	children	smoker	region	expenses
10	25	male	26.2	0	no	northeast	2721.32
12	23	male	34.4	0	no	southwest	1826.84
14	27	male	42.1	0	yes	southeast	39611.76
16	52	female	30.8	1	no	northeast	10797.34
18	56	male	40.3	0	no	southwest	10602.39

```
[55]: #fancy indexing - for particular index
data.iloc[[0,5,2,1]]
```

```
[55]:
```

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
5	31	female	25.7	0	no	southeast	3756.62
2	28	male	33.0	3	no	southeast	4449.46
1	18	male	33.8	1	no	southeast	1725.55

```
[56]: #loc
#for single items
s_data.loc['A']
```

```
[56]: iq          100
marks         80
package        10
Name: A, dtype: int64
```

```
[57]: # loc for multiple use it in below way
s_data.loc['A':'C']
```

```
[57]:
```

	iq	marks	package
name			
A	100	80	10
B	90	70	7
C	80	85	8

selecting both rows and columns can use loc and iloc

```
[58]: data.iloc[0:2,0:3]
```

```
[58]:   age      sex    bmi
      0   19  female  27.9
      1   18   male  33.8
```

Filtering the dataframe

```
[ ]:
```

```
[59]: ipl.head(2)
```

```
[59]:      ID      City      Date Season  MatchNumber  \
0  1312200  Ahmedabad  2022-05-29   2022         Final
1  1312199  Ahmedabad  2022-05-27   2022  Qualifier 2

      Team1      Team2  \
0      Rajasthan Royals  Gujarat Titans
1  Royal Challengers Bangalore  Rajasthan Royals

      Venue      TossWinner TossDecision SuperOver  \
0  Narendra Modi Stadium, Ahmedabad  Rajasthan Royals      bat      N
1  Narendra Modi Stadium, Ahmedabad  Rajasthan Royals      field      N

      WinningTeam  WonBy  Margin method Player_of_Match  \
0      Gujarat Titans  Wickets      7.0      NaN      HH Pandya
1  Rajasthan Royals  Wickets      7.0      NaN      JC Buttler

      Team1Players  \
0  ['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...
1  ['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...

      Team2Players      Umpire1      Umpire2
0  ['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...  CB Gaffaney  Nitin Menon
1  ['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...  CB Gaffaney  Nitin Menon
```

```
[62]: #find all the final winners
mask = ipl['MatchNumber'] == 'Final'
```

```
[64]: new_df = ipl[mask]
```

```
[65]: new_df[['Season', 'WinningTeam']]
```

```
[65]:      Season      WinningTeam
0      2022      Gujarat Titans
74     2021  Chennai Super Kings
134  2020/21      Mumbai Indians
194     2019      Mumbai Indians
254     2018  Chennai Super Kings
```

314	2017	Mumbai Indians
373	2016	Sunrisers Hyderabad
433	2015	Mumbai Indians
492	2014	Kolkata Knight Riders
552	2013	Mumbai Indians
628	2012	Kolkata Knight Riders
702	2011	Chennai Super Kings
775	2009/10	Chennai Super Kings
835	2009	Deccan Chargers
892	2007/08	Rajasthan Royals

```
[68]: # or one liner
ipl[ipl['MatchNumber'] == 'Final'][['Season', 'WinningTeam']]
```

```
[68]:      Season      WinningTeam
0      2022      Gujarat Titans
74     2021     Chennai Super Kings
134  2020/21      Mumbai Indians
194     2019      Mumbai Indians
254     2018     Chennai Super Kings
314     2017      Mumbai Indians
373     2016     Sunrisers Hyderabad
433     2015      Mumbai Indians
492     2014     Kolkata Knight Riders
552     2013      Mumbai Indians
628     2012     Kolkata Knight Riders
702     2011     Chennai Super Kings
775  2009/10     Chennai Super Kings
835     2009      Deccan Chargers
892  2007/08     Rajasthan Royals
```

```
[69]: # How many super over finishes have occurred
```

```
[70]: ipl.columns
```

```
[70]: Index(['ID', 'City', 'Date', 'Season', 'MatchNumber', 'Team1', 'Team2',
        'Venue', 'TossWinner', 'TossDecision', 'SuperOver', 'WinningTeam',
        'WonBy', 'Margin', 'method', 'Player_of_Match', 'Team1Players',
        'Team2Players', 'Umpire1', 'Umpire2'],
        dtype='object')
```

```
[71]: ipl['SuperOver'].head()
```

```
[71]: 0    N
      1    N
      2    N
      3    N
```

```
4      N
Name: SuperOver, dtype: object
```

```
[72]: ipl['SuperOver'].value_counts()
```

```
[72]: N      932
      Y      14
      Name: SuperOver, dtype: int64
```

```
[76]: ipl[ipl['SuperOver'] == 'Y'].shape
```

```
[76]: (14, 20)
```

```
[77]: # or
      ipl[ipl['SuperOver'] == 'Y'].shape[0]
```

```
[77]: 14
```

```
[86]: #How many matches csk won in in kolkata
      ipl[(ipl['City'] == 'Kolkata') & (ipl['WinningTeam'] == 'Chennai Super Kings')]
```

```
[86]:
```

	ID	City	Date	Season	MatchNumber	Team1	\
224	1178404	Kolkata	2019-04-14	2019	29	Kolkata Knight Riders	
602	598022	Kolkata	2013-04-20	2013	26	Kolkata Knight Riders	
641	548368	Kolkata	2012-05-14	2012	63	Kolkata Knight Riders	
827	419113	Kolkata	2010-03-16	2009/10	8	Kolkata Knight Riders	
908	336025	Kolkata	2008-05-18	2007/08	41	Kolkata Knight Riders	

		Team2	Venue	TossWinner	TossDecision	\
224	Chennai Super Kings	Eden Gardens	Chennai Super Kings	field		
602	Chennai Super Kings	Eden Gardens	Kolkata Knight Riders	bat		
641	Chennai Super Kings	Eden Gardens	Chennai Super Kings	field		
827	Chennai Super Kings	Eden Gardens	Chennai Super Kings	bat		
908	Chennai Super Kings	Eden Gardens	Kolkata Knight Riders	bat		

	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	\
224	N	Chennai Super Kings	Wickets	5.0	NaN	Imran Tahir	
602	N	Chennai Super Kings	Wickets	4.0	NaN	RA Jadeja	
641	N	Chennai Super Kings	Wickets	5.0	NaN	MEK Hussey	
827	N	Chennai Super Kings	Runs	55.0	NaN	MS Dhoni	
908	N	Chennai Super Kings	Runs	3.0	D/L	M Ntini	

	Team1Players	\
224	['CA Lynn', 'SP Narine', 'N Rana', 'RV Uthappa...]	
602	['G Gambhir', 'YK Pathan', 'JH Kallis', 'EJG M...]	
641	['G Gambhir', 'BB McCullum', 'JH Kallis', 'MK ...]	
827	['BJ Hodge', 'MK Tiwary', 'SC Ganguly', 'WP Sa...]	

```
908 ['Salman Butt', 'Mohammad Hafeez', 'SC Ganguly...
```

```

                                Team2Players      Umpire1 \
224 ['SR Watson', 'F du Plessis', 'SK Raina', 'AT ...  CK Nandan
602 ['MEK Hussey', 'R Ashwin', 'M Vijay', 'SK Rain...  Asad Rauf
641 ['MEK Hussey', 'M Vijay', 'SK Raina', 'F du Pl...  JD Cloete
827 ['M Vijay', 'ML Hayden', 'SK Raina', 'S Badrin...  HDPK Dharmasena
908 ['PA Patel', 'SP Fleming', 'S Vidyut', 'SK Rai...  Asad Rauf

                                Umpire2
224      RJ Tucker
602      AK Chaudhary
641      SJA Taufel
827      AM Saheba
908      K Hariharan
```

```
[87]: ipl[(ipl['City'] == 'Kolkata') & (ipl['WinningTeam'] == 'Chennai Super Kings')].
      ↪shape[0]
```

```
[87]: 5
```

```
[88]: #toss winner is match winner in percentage
```

```
[90]: ipl[ipl['TossWinner'] == ipl['WinningTeam']]
```

```
[90]:
```

	ID	City	Date	Season	MatchNumber	\
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	
5	1304115	Mumbai	2022-05-21	2022	69	
8	1304112	Navi Mumbai	2022-05-18	2022	66	
11	1304109	Mumbai	2022-05-15	2022	63	
..	
935	335996	Bangalore	2008-04-28	2007/08	15	
937	335994	Mumbai	2008-04-27	2007/08	14	
939	335992	Bangalore	2008-04-26	2007/08	12	
941	335990	Hyderabad	2008-04-24	2007/08	9	
948	335983	Chandigarh	2008-04-19	2007/08	2	

```

                                Team1      Team2 \
1      Royal Challengers Bangalore      Rajasthan Royals
3              Rajasthan Royals      Gujarat Titans
5              Delhi Capitals      Mumbai Indians
8              Lucknow Super Giants      Kolkata Knight Riders
11             Rajasthan Royals      Lucknow Super Giants
..              ...
935      Royal Challengers Bangalore      Chennai Super Kings
937             Mumbai Indians      Deccan Chargers
```

939	Royal Challengers Bangalore	Rajasthan Royals
941	Deccan Chargers	Rajasthan Royals
948	Kings XI Punjab	Chennai Super Kings

	Venue	TossWinner \
1	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals
3	Eden Gardens, Kolkata	Gujarat Titans
5	Wankhede Stadium, Mumbai	Mumbai Indians
8	Dr DY Patil Sports Academy, Mumbai	Lucknow Super Giants
11	Brabourne Stadium, Mumbai	Rajasthan Royals
..
935	M Chinnaswamy Stadium	Chennai Super Kings
937	Dr DY Patil Sports Academy	Deccan Chargers
939	M Chinnaswamy Stadium	Rajasthan Royals
941	Rajiv Gandhi International Stadium, Uppal	Rajasthan Royals
948	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings

	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method \
1	field	N	Rajasthan Royals	Wickets	7.0	NaN
3	field	N	Gujarat Titans	Wickets	7.0	NaN
5	field	N	Mumbai Indians	Wickets	5.0	NaN
8	bat	N	Lucknow Super Giants	Runs	2.0	NaN
11	bat	N	Rajasthan Royals	Runs	24.0	NaN
..
935	bat	N	Chennai Super Kings	Runs	13.0	NaN
937	field	N	Deccan Chargers	Wickets	10.0	NaN
939	field	N	Rajasthan Royals	Wickets	7.0	NaN
941	field	N	Rajasthan Royals	Wickets	3.0	NaN
948	bat	N	Chennai Super Kings	Runs	33.0	NaN

	Player_of_Match	Team1Players \
1	JC Buttler	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...
3	DA Miller	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...
5	JJ Bumrah	['PP Shaw', 'DA Warner', 'MR Marsh', 'RR Pant'...
8	Q de Kock	['Q de Kock', 'KL Rahul', 'E Lewis', 'DJ Hooda...
11	TA Boult	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...
..
935	MS Dhoni	['B Chipli', 'W Jaffer', 'LRPL Taylor', 'JH Ka...
937	AC Gilchrist	['L Ronchi', 'ST Jayasuriya', 'AM Rahane', 'RV...
939	SR Watson	['S Chanderpaul', 'R Dravid', 'LRPL Taylor', '...
941	YK Pathan	['AC Gilchrist', 'VVS Laxman', 'Shahid Afridi'...
948	MEK Hussey	['K Goel', 'JR Hopes', 'KC Sangakkara', 'Yuvra...

	Team2Players	Umpire1 \
1	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	CB Gaffaney
3	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	BNJ Oxenford
5	['Ishan Kishan', 'RG Sharma', 'D Brevis', 'Til...	Nitin Menon

```

8      ['VR Iyer', 'A Tomar', 'N Rana', 'SS Iyer', 'S...      R Pandit
11     ['Q de Kock', 'KL Rahul', 'A Badoni', 'DJ Hood...      PG Pathak
..
935    ['PA Patel', 'ML Hayden', 'MEK Hussey', 'SK Ra...      BR Doctrove
937    ['AC Gilchrist', 'VVS Laxman', 'A Symonds', 'R...      Asad Rauf
939    ['GC Smith', 'M Rawat', 'YK Pathan', 'SR Watso...      MR Benson
941    ['Kamran Akmal', 'GC Smith', 'YK Pathan', 'SR ...      Asad Rauf
948    ['PA Patel', 'ML Hayden', 'MEK Hussey', 'MS Dh...      MR Benson

```

```

      Umpire2
1      Nitin Menon
3      VK Sharma
5      Tapan Sharma
8      YC Barde
11     Tapan Sharma
..
935     RB Tiffin
937     SL Shastri
939     IL Howell
941     MR Benson
948     SL Shastri

```

[489 rows x 20 columns]

```
[96]: ipl[ipl['TossWinner'] == ipl['WinningTeam']].shape[0]/ipl.shape[0]
```

```
[96]: 0.5147368421052632
```

Add columns in DataFrame

```
[97]: ipl.head()
```

```
[97]:
```

	ID	City	Date	Season	MatchNumber	\
0	1312200	Ahmedabad	2022-05-29	2022	Final	
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	
2	1312198	Kolkata	2022-05-25	2022	Eliminator	
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	
4	1304116	Mumbai	2022-05-22	2022	70	

	Team1	Team2	\
0	Rajasthan Royals	Gujarat Titans	
1	Royal Challengers Bangalore	Rajasthan Royals	
2	Royal Challengers Bangalore	Lucknow Super Giants	
3	Rajasthan Royals	Gujarat Titans	
4	Sunrisers Hyderabad	Punjab Kings	

	Venue	TossWinner	TossDecision	\
--	-------	------------	--------------	---

0	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	bat
1	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	field
2	Eden Gardens, Kolkata	Lucknow Super Giants	field
3	Eden Gardens, Kolkata	Gujarat Titans	field
4	Wankhede Stadium, Mumbai	Sunrisers Hyderabad	bat

	SuperOver	WinningTeam	WonBy	Margin	method	\
0	N	Gujarat Titans	Wickets	7.0	NaN	
1	N	Rajasthan Royals	Wickets	7.0	NaN	
2	N	Royal Challengers Bangalore	Runs	14.0	NaN	
3	N	Gujarat Titans	Wickets	7.0	NaN	
4	N	Punjab Kings	Wickets	5.0	NaN	

	Player_of_Match	Team1Players	\
0	HH Pandya	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	
1	JC Buttler	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...	
2	RM Patidar	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...	
3	DA Miller	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	
4	Harpreet Brar	['PK Garg', 'Abhishek Sharma', 'RA Tripathi', ...	

	Team2Players	Umpire1	\
0	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	CB Gaffaney	
1	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	CB Gaffaney	
2	['Q de Kock', 'KL Rahul', 'M Vohra', 'DJ Hooda...	J Madanagopal	
3	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	BNJ Oxenford	
4	['JM Bairstow', 'S Dhawan', 'M Shahrukh Khan', ...	AK Chaudhary	

	Umpire2
0	Nitin Menon
1	Nitin Menon
2	MA Gough
3	VK Sharma
4	NA Patwardhan

```
[98]: ipl['Country'] = 'India'
```

```
[99]: ipl.head()
```

```
[99]:
```

	ID	City	Date	Season	MatchNumber	\
0	1312200	Ahmedabad	2022-05-29	2022	Final	
1	1312199	Ahmedabad	2022-05-27	2022	Qualifier 2	
2	1312198	Kolkata	2022-05-25	2022	Eliminator	
3	1312197	Kolkata	2022-05-24	2022	Qualifier 1	
4	1304116	Mumbai	2022-05-22	2022	70	

	Team1	Team2	\
0	Rajasthan Royals	Gujarat Titans	

1	Royal Challengers Bangalore	Rajasthan Royals		
2	Royal Challengers Bangalore	Lucknow Super Giants		
3	Rajasthan Royals	Gujarat Titans		
4	Sunrisers Hyderabad	Punjab Kings		

	Venue	TossWinner	TossDecision	...	\
0	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	bat	...	
1	Narendra Modi Stadium, Ahmedabad	Rajasthan Royals	field	...	
2	Eden Gardens, Kolkata	Lucknow Super Giants	field	...	
3	Eden Gardens, Kolkata	Gujarat Titans	field	...	
4	Wankhede Stadium, Mumbai	Sunrisers Hyderabad	bat	...	

	WinningTeam	WonBy	Margin	method	Player_of_Match	\
0	Gujarat Titans	Wickets	7.0	NaN	HH Pandya	
1	Rajasthan Royals	Wickets	7.0	NaN	JC Buttler	
2	Royal Challengers Bangalore	Runs	14.0	NaN	RM Patidar	
3	Gujarat Titans	Wickets	7.0	NaN	DA Miller	
4	Punjab Kings	Wickets	5.0	NaN	Harpreet Brar	

	Team1Players	\
0	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	
1	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...	
2	['V Kohli', 'F du Plessis', 'RM Patidar', 'GJ ...	
3	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	
4	['PK Garg', 'Abhishek Sharma', 'RA Tripathi', ...	

	Team2Players	Umpire1	\
0	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	CB Gaffaney	
1	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...	CB Gaffaney	
2	['Q de Kock', 'KL Rahul', 'M Vohra', 'DJ Hooda...	J Madanagopal	
3	['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan...	BNJ Oxenford	
4	['JM Bairstow', 'S Dhawan', 'M Shahrukh Khan', ...	AK Chaudhary	

	Umpire2	Country
0	Nitin Menon	India
1	Nitin Menon	India
2	MA Gough	India
3	VK Sharma	India
4	NA Patwardhan	India

[5 rows x 21 columns]

```
[100]: # create a column from a existing column
```

important DataFrame function

```
[101]: #astype -it changes the data type of the column
ipl.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ID                    950 non-null    int64
1   City                  899 non-null    object
2   Date                  950 non-null    object
3   Season                950 non-null    object
4   MatchNumber           950 non-null    object
5   Team1                 950 non-null    object
6   Team2                 950 non-null    object
7   Venue                 950 non-null    object
8   TossWinner            950 non-null    object
9   TossDecision          950 non-null    object
10  SuperOver             946 non-null    object
11  WinningTeam           946 non-null    object
12  WonBy                 950 non-null    object
13  Margin                932 non-null    float64
14  method                19 non-null     object
15  Player_of_Match       946 non-null    object
16  Team1Players           950 non-null    object
17  Team2Players           950 non-null    object
18  Umpire1               950 non-null    object
19  Umpire2               950 non-null    object
20  Country               950 non-null    object
dtypes: float64(1), int64(1), object(19)
memory usage: 156.0+ KB
```

```
[103]: ipl['ID'] = ipl['ID'].astype('int32')
```

```
[104]: ipl.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ID                    950 non-null    int32
1   City                  899 non-null    object
2   Date                  950 non-null    object
3   Season                950 non-null    object
4   MatchNumber           950 non-null    object
5   Team1                 950 non-null    object
6   Team2                 950 non-null    object
```

```

7   Venue          950 non-null    object
8   TossWinner     950 non-null    object
9   TossDecision   950 non-null    object
10  SuperOver      946 non-null    object
11  WinningTeam    946 non-null    object
12  WonBy          950 non-null    object
13  Margin         932 non-null    float64
14  method         19 non-null     object
15  Player_of_Match 946 non-null    object
16  Team1Players   950 non-null    object
17  Team2Players   950 non-null    object
18  Umpire1        950 non-null    object
19  Umpire2        950 non-null    object
20  Country        950 non-null    object
dtypes: float64(1), int32(1), object(19)
memory usage: 152.3+ KB

```

```
[105]: ipl['Season'] = ipl['Season'].astype('category')
```

```
[106]: ipl.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 21 columns):
#   Column          Non-Null Count  Dtype
---  -
0   ID              950 non-null   int32
1   City            899 non-null   object
2   Date            950 non-null   object
3   Season          950 non-null   category
4   MatchNumber     950 non-null   object
5   Team1           950 non-null   object
6   Team2           950 non-null   object
7   Venue           950 non-null   object
8   TossWinner      950 non-null   object
9   TossDecision    950 non-null   object
10  SuperOver       946 non-null   object
11  WinningTeam     946 non-null   object
12  WonBy           950 non-null   object
13  Margin          932 non-null   float64
14  method          19 non-null    object
15  Player_of_Match 946 non-null   object
16  Team1Players    950 non-null   object
17  Team2Players    950 non-null   object
18  Umpire1         950 non-null   object
19  Umpire2         950 non-null   object
20  Country         950 non-null   object
dtypes: category(1), float64(1), int32(1), object(18)

```

memory usage: 146.4+ KB

```
[107]: ipl['Team1'] = ipl['Team1'].astype('category')
ipl['Team2'] = ipl['Team2'].astype('category')
```

```
[108]: ipl.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   ID                    950 non-null    int32
1   City                  899 non-null    object
2   Date                  950 non-null    object
3   Season                950 non-null    category
4   MatchNumber           950 non-null    object
5   Team1                 950 non-null    category
6   Team2                 950 non-null    category
7   Venue                 950 non-null    object
8   TossWinner            950 non-null    object
9   TossDecision          950 non-null    object
10  SuperOver             946 non-null    object
11  WinningTeam           946 non-null    object
12  WonBy                 950 non-null    object
13  Margin                932 non-null    float64
14  method                19 non-null     object
15  Player_of_Match       946 non-null    object
16  Team1Players           950 non-null    object
17  Team2Players           950 non-null    object
18  Umpire1               950 non-null    object
19  Umpire2               950 non-null    object
20  Country               950 non-null    object
dtypes: category(3), float64(1), int32(1), object(16)
memory usage: 134.8+ KB
```

```
[ ]:
```

```
[ ]:
```