

# Welcome to Pandas Tutorials!!!

Pandas library in python is used for data manipulation and data analysis Pandas two main Data Structure : 1.DataFrame 2.Series

## Fetch data from csv in jupyter notebook :

```
import pandas as pd
df = pd.read_csv('retail_product.csv')
```

df

	Invoice	StockCode	Description
Quantity \			
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS
12			
1	489434	79323P	PINK CHERRY LIGHTS
12			
2	489434	79323W	WHITE CHERRY LIGHTS
12			
3	489434	22041	RECORD FRAME 7" SINGLE SIZE
48			
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX
24			
...	...	...	...
..			
1067366	581587	22899	CHILDREN'S APRON DOLLY GIRL
6			
1067367	581587	23254	CHILDRENS CUTLERY DOLLY GIRL
4			
1067368	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE
4			
1067369	581587	22138	BAKING SET 9 PIECE RETROSPOT
3			
1067370	581587	POST	POSTAGE
1			
	InvoiceDate	Price	Customer ID
0	2009-12-01 07:45:00	6.95	13085.0
1	2009-12-01 07:45:00	6.75	13085.0
2	2009-12-01 07:45:00	6.75	13085.0
3	2009-12-01 07:45:00	2.10	13085.0
4	2009-12-01 07:45:00	1.25	13085.0
...	...	...	...
1067366	2011-12-09 12:50:00	2.10	12680.0
1067367	2011-12-09 12:50:00	4.15	12680.0

1067368	2011-12-09	12:50:00	4.15	12680.0	France
1067369	2011-12-09	12:50:00	4.95	12680.0	France
1067370	2011-12-09	12:50:00	18.00	12680.0	France

[1067371 rows x 8 columns]

## Basic Attributes and Functions useful in every Analysis :

`df.head()` *#head() gives top 5 rows of the dataframe*

	Invoice	StockCode	Description	Quantity	\
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	
1	489434	79323P	PINK CHERRY LIGHTS	12	
2	489434	79323W	WHITE CHERRY LIGHTS	12	
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	

	InvoiceDate	Price	Customer ID	Country
0	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

`df.head(2)` *#custom values in head*

	Invoice	StockCode	Description	Quantity	\
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	
1	489434	79323P	PINK CHERRY LIGHTS	12	

	InvoiceDate	Price	Customer ID	Country
0	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	2009-12-01 07:45:00	6.75	13085.0	United Kingdom

`df.tail()` *#tail() gives last 5 rows*

	Invoice	StockCode	Description	Quantity	\
1067366	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	
1067367	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	
1067368	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	
1067369	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	
1067370	581587	POST	POSTAGE	1	

	InvoiceDate	Price	Customer ID	Country
1067366	2011-12-09 12:50:00	2.10	12680.0	France
1067367	2011-12-09 12:50:00	4.15	12680.0	France
1067368	2011-12-09 12:50:00	4.15	12680.0	France
1067369	2011-12-09 12:50:00	4.95	12680.0	France
1067370	2011-12-09 12:50:00	18.00	12680.0	France

```
df.tail(2)
```

	Invoice	StockCode	Description	Quantity
1067369	581587	22138	BAKING SET 9 PIECE RETROSPOT	3
1067370	581587	POST	POSTAGE	1

	InvoiceDate	Price	Customer ID	Country
1067369	2011-12-09 12:50:00	4.95	12680.0	France
1067370	2011-12-09 12:50:00	18.00	12680.0	France

## To find the Shape of the data ie. How big is the data

```
df.shape # gives a tuple (x,y) x- rows y- cols
```

```
(1067371, 8)
```

```
df.shape[0]
```

```
1067371
```

```
df.shape[1]
```

```
8
```

## To get details of the dataset use .info()

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1067371 entries, 0 to 1067370
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Invoice                1067371 non-null  object
1   StockCode             1067371 non-null  object
2   Description           1062989 non-null  object
3   Quantity              1067371 non-null  int64
```

```

4 InvoiceDate 1067371 non-null object
5 Price      1067371 non-null float64
6 Customer ID 824364 non-null float64
7 Country    1067371 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 65.1+ MB

```

## df.describe() for all the statistical info

```
df.describe() #works only on numerical columns
```

	Quantity	Price	Customer ID
count	1.067371e+06	1.067371e+06	824364.000000
mean	9.938898e+00	4.649388e+00	15324.638504
std	1.727058e+02	1.235531e+02	1697.464450
min	-8.099500e+04	-5.359436e+04	12346.000000
25%	1.000000e+00	1.250000e+00	13975.000000
50%	3.000000e+00	2.100000e+00	15255.000000
75%	1.000000e+01	4.150000e+00	16797.000000
max	8.099500e+04	3.897000e+04	18287.000000

## Fetch cols from the dataset

```

df['Price']

0      6.95
1      6.75
2      6.75
3      2.10
4      1.25
...
1067366  2.10
1067367  4.15
1067368  4.15
1067369  4.95
1067370 18.00
Name: Price, Length: 1067371, dtype: float64

type(df['Price']) #This is a series datatype
pandas.core.series.Series

df['Price'].shape
(1067371,)

```

```
df2 = df[['Description', 'Quantity', 'Price']]
df2
```

	Description	Quantity	Price
0	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	6.95
1	PINK CHERRY LIGHTS	12	6.75
2	WHITE CHERRY LIGHTS	12	6.75
3	RECORD FRAME 7" SINGLE SIZE	48	2.10
4	STRAWBERRY CERAMIC TRINKET BOX	24	1.25
...	...	...	...
1067366	CHILDREN'S APRON DOLLY GIRL	6	2.10
1067367	CHILDRENS CUTLERY DOLLY GIRL	4	4.15
1067368	CHILDRENS CUTLERY CIRCUS PARADE	4	4.15
1067369	BAKING SET 9 PIECE RETROSPOT	3	4.95
1067370	POSTAGE	1	18.00

```
[1067371 rows x 3 columns]
```

```
df2.shape
```

```
(1067371, 3)
```

## For fetching multiple rows use .iloc[]

```
df.iloc[0] #just like array indexing
```

Invoice	489434
StockCode	85048
Description	15CM CHRISTMAS GLASS BALL 20 LIGHTS
Quantity	12
InvoiceDate	2009-12-01 07:45:00
Price	6.95
Customer ID	13085.0
Country	United Kingdom
Name: 0, dtype: object	

```
df.iloc[1:5] #we can do slicing also
```

	Invoice	StockCode	Description	Quantity	\
1	489434	79323P	PINK CHERRY LIGHTS	12	
2	489434	79323W	WHITE CHERRY LIGHTS	12	
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	

	InvoiceDate	Price	Customer ID	Country
1	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

```
df.iloc[1:21:3]
```

	Invoice	StockCode	Description	Quantity	\
1	489434	79323P	PINK CHERRY LIGHTS	12	
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	
7	489434	21523	FANCY FONT HOME SWEET HOME DOORMAT	10	
10	489435	22195	HEART MEASURING SPOONS LARGE	24	
13	489436	21755	LOVE BUILDING BLOCK WORD	18	
16	489436	22119	PEACE WOODEN BLOCK LETTERS	3	
19	489436	22295	HEART FILIGREE DOVE LARGE	12	

	InvoiceDate	Price	Customer ID	Country
1	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
7	2009-12-01 07:45:00	5.95	13085.0	United Kingdom
10	2009-12-01 07:46:00	1.65	13085.0	United Kingdom
13	2009-12-01 09:06:00	5.45	13078.0	United Kingdom
16	2009-12-01 09:06:00	6.95	13078.0	United Kingdom
19	2009-12-01 09:06:00	1.65	13078.0	United Kingdom

```
df.iloc[[1,4,21,7700]]
```

	Invoice	StockCode	Description	Quantity	\
1	489434	79323P	PINK CHERRY LIGHTS	12	
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	
21	489436	22107	PIZZA PLATE IN BOX	4	
7700	490064	21507	ELEPHANT, BIRTHDAY CARD,	12	

	InvoiceDate	Price	Customer ID	Country
1	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
21	2009-12-01 09:06:00	3.75	13078.0	United Kingdom
7700	2009-12-03 13:51:00	0.42	16139.0	United Kingdom

```
df.iloc[:, [1,3,5]] # here the first half is for selection all the rows and the second half is from selecting the indexes of columns
```

	StockCode	Quantity	Price
0	85048	12	6.95
1	79323P	12	6.75
2	79323W	12	6.75
3	22041	48	2.10
4	21232	24	1.25
...	...	...	...
1067366	22899	6	2.10
1067367	23254	4	4.15
1067368	23255	4	4.15
1067369	22138	3	4.95
1067370	POST	1	18.00

```
[1067371 rows x 3 columns]
```

What if you want only specific part of the data ,  
like only the rows from country = France

```
df['Country'] == 'France' #this will return a boolean Series
```

```
0      False
1      False
2      False
3      False
4      False
```

```
...
1067366    True
1067367    True
1067368    True
1067369    True
1067370    True
```

```
Name: Country, Length: 1067371, dtype: bool
```

```
# store the above boolean Series in a mask variable
```

```
mask = df['Country'] == 'France'
```

```
df[mask] #here we are masking the boolean series where only true values will be taken from the data
```

```
# it is boolean Indexing
```

	Invoice	StockCode	Description	Quantity
\				
71	489439	22065	CHRISTMAS PUDDING TRINKET POT	12
72	489439	22138	BAKING SET 9 PIECE RETROSPOT	9
73	489439	22139	RETRO SPOT TEA SET CERAMIC 11 PC	9
74	489439	22352	LUNCHBOX WITH CUTLERY RETROSPOT	12
75	489439	85014A	BLACK/BLUE DOTS RUFFLED UMBRELLA	3
...	...	...	...	...
1067366	581587	22899	CHILDREN'S APRON DOLLY GIRL	6
1067367	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4
1067368	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4
1067369	581587	22138	BAKING SET 9 PIECE RETROSPOT	3
1067370	581587	POST	POSTAGE	1

	InvoiceDate	Price	Customer ID	Country
71	2009-12-01 09:28:00	1.45	12682.0	France
72	2009-12-01 09:28:00	4.95	12682.0	France
73	2009-12-01 09:28:00	4.95	12682.0	France
74	2009-12-01 09:28:00	2.55	12682.0	France
75	2009-12-01 09:28:00	5.95	12682.0	France
...	...	...	...	...
1067366	2011-12-09 12:50:00	2.10	12680.0	France
1067367	2011-12-09 12:50:00	4.15	12680.0	France
1067368	2011-12-09 12:50:00	4.15	12680.0	France
1067369	2011-12-09 12:50:00	4.95	12680.0	France
1067370	2011-12-09 12:50:00	18.00	12680.0	France

[14330 rows x 8 columns]

```
mask1 = df['Country']=='France'
mask2 = df['InvoiceDate'] > '2010-12-06 00:00:00'
```

*#df[mask1 and mask2] this 'and' will give error as the comparison is between boolean series and Not with the boolean values*  
*df[mask1 & mask2] # this will give all the rows with the country = France and the InvoiceDate after the 2010-12-06 00:00:00 ie.AFTER 12th June 2010 12AM*

Quantity \	Invoice	StockCode	Description
517064	537463	22961	JAM MAKING SET PRINTED
12			
517065	537463	21224	SET/4 SKULL BADGES
10			
517066	537463	22326	ROUND SNACK BOXES SET OF4 WOODLAND
18			
517067	537463	21124	SET/10 BLUE POLKADOT PARTY CANDLES
24			
517068	537463	21121	SET/10 RED POLKADOT PARTY CANDLES
24			
...	...	...	...
..			
1067366	581587	22899	CHILDREN'S APRON DOLLY GIRL
6			
1067367	581587	23254	CHILDRENS CUTLERY DOLLY GIRL
4			
1067368	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE
4			
1067369	581587	22138	BAKING SET 9 PIECE RETROSPOT
3			
1067370	581587	POST	POSTAGE
1			

InvoiceDate	Price	Customer ID	Country
-------------	-------	-------------	---------



517064	2010-12-07	10:08:00	1.45	12681.0	France
517065	2010-12-07	10:08:00	1.25	12681.0	France
517066	2010-12-07	10:08:00	2.95	12681.0	France
517067	2010-12-07	10:08:00	1.25	12681.0	France
517068	2010-12-07	10:08:00	1.25	12681.0	France
...					
1067366	2011-12-09	12:50:00	2.10	12680.0	France
1067367	2011-12-09	12:50:00	4.15	12680.0	France
1067368	2011-12-09	12:50:00	4.15	12680.0	France
1067369	2011-12-09	12:50:00	4.95	12680.0	France
1067370	2011-12-09	12:50:00	18.00	12680.0	France

[8617 rows x 8 columns]

To find the frequencies for any column ie. the series, we can use `value_counts()`

```
df3 = df['Description'].value_counts()
df3.shape

(5698,)
```

```
df3.head(10)
```

Description	
WHITE HANGING HEART T-LIGHT HOLDER	5918
REGENCY CAKESTAND 3 TIER	4412
JUMBO BAG RED RETROSPOT	3469
ASSORTED COLOUR BIRD ORNAMENT	2958
PARTY BUNTING	2765
STRAWBERRY CERAMIC TRINKET BOX	2613
LUNCH BAG BLACK SKULL.	2529
JUMBO STORAGE BAG SUKI	2434
HEART OF WICKER SMALL	2319
JUMBO SHOPPER VINTAGE RED PAISLEY	2297

Name: count, dtype: int64

```
df3.tail(10)
```

Description	
VINTAGE WOOD ORGANISER	1
for online retail orders	1
add stock to allocate online orders	1
ORANGE CANDLE	1
HAND OPEN SHAPE CHROME	1
IVORY PENDANT SHELL NECKLACE	1
allocate stock for dotcom orders ta	1
PAINTED HEART WREATH WITH BELL	1

```

FLOWER FAIRY MUG+COASTER,ROSE      1
Zebra invcing error                  1
Name: count, dtype: int64

df3.iloc[1:500:20]

Description
REGENCY CAKESTAND 3 TIER              4412
HEART OF WICKER LARGE                 2112
SWEETHEART CERAMIC TRINKET BOX       1639
ROUND SNACK BOXES SET OF4 WOODLAND   1472
CREAM HEART CARD HOLDER              1350
ROUND SNACK BOXES SET OF 4 FRUITS    1237
SPOTTY BUNTING                      1172
SPACEBOY BIRTHDAY CARD               1060
BEWARE OF THE CAT METAL SIGN         1017
DOORMAT RED RETROSPOT                 950
4 TRADITIONAL SPINNING TOPS          902
COLOUR GLASS. STAR T-LIGHT HOLDER    869
WHITE SPOT RED CERAMIC DRAWER KNOB   848
RED SPOTTY CHARLOTTE BAG              808
TRADITIONAL CHRISTMAS RIBBONS         781
SET OF 6 SOLDIER SKITTLES            755
KITCHEN METAL SIGN                   732
COLOURING PENCILS BROWN TUBE         699
ALARM CLOCK BAKELIKE ORANGE          673
PINK VINTAGE PAISLEY PICNIC BAG       655
10 COLOUR SPACEBOY PEN               626
15CM CHRISTMAS GLASS BALL 20 LIGHTS  604
36 PENCILS TUBE SKULLS               586
SET OF 12 FAIRY CAKE BAKING CASES    567
DOORMAT ENGLISH ROSE                 550
Name: count, dtype: int64

```

## Lets see the Plot function in Pandas

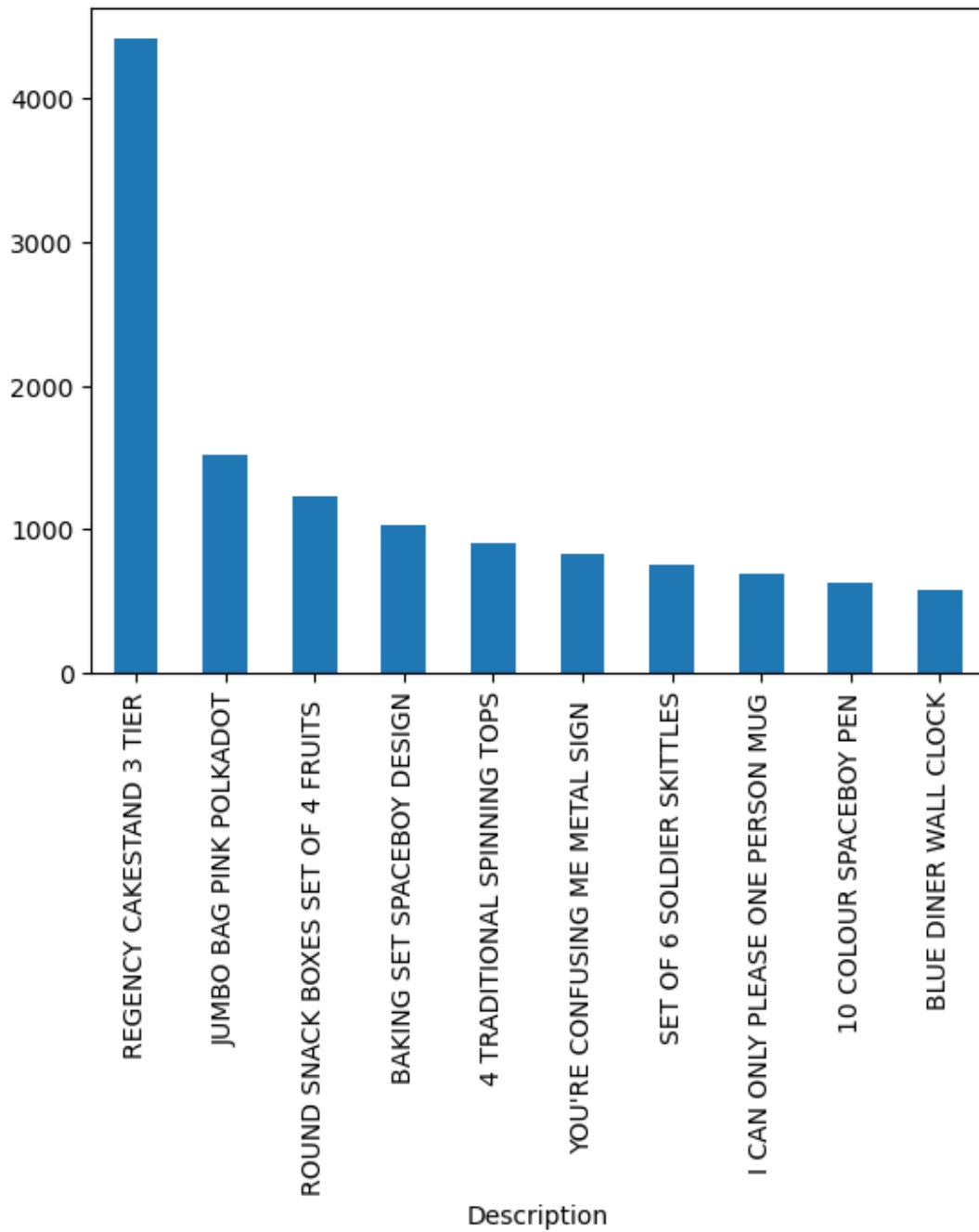
```

import matplotlib.pyplot as plt

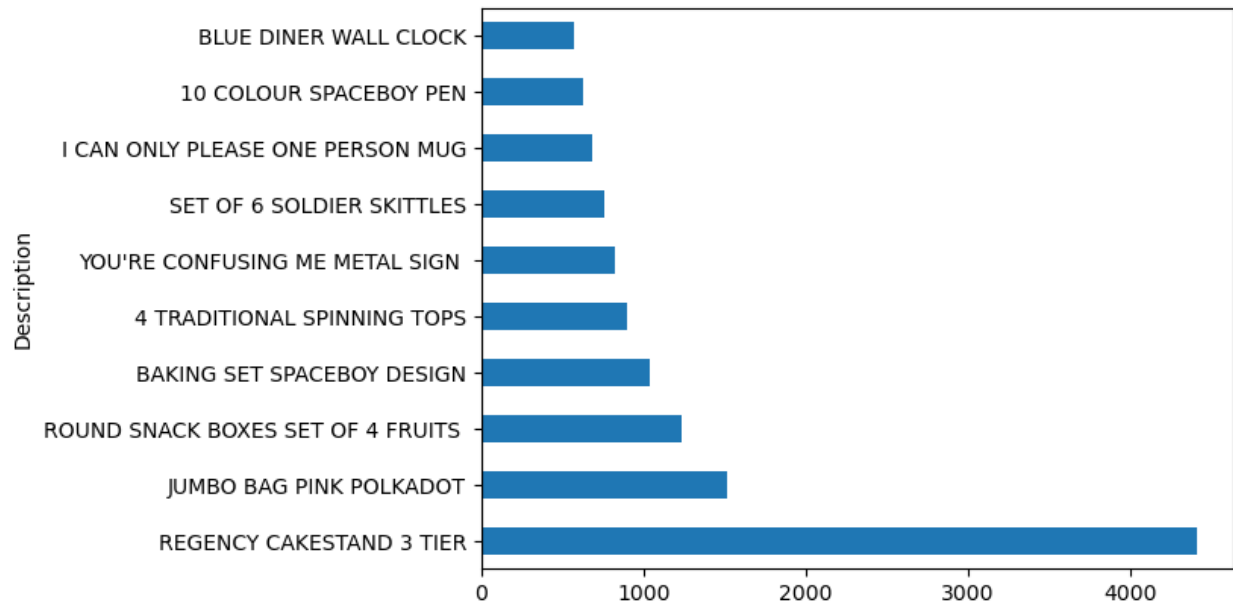
df3.iloc[1:500:50].plot(kind='bar')

<Axes: xlabel='Description'>

```



```
# we can draw horizontal bar : using 'barh'  
df3.iloc[1:500:50].plot(kind='barh')  
  
<Axes: ylabel='Description'>
```

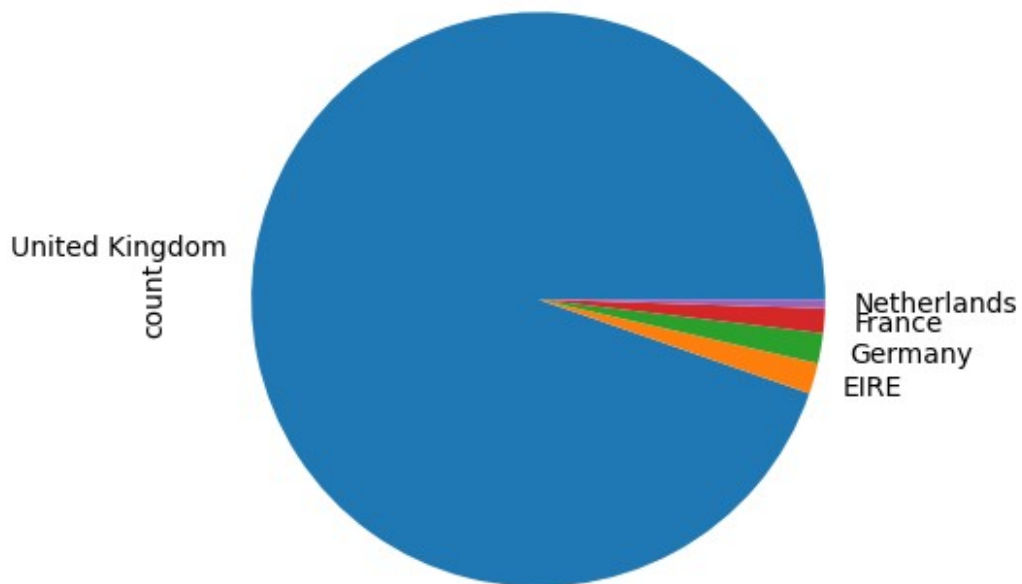


```
#we can make a pie chart also : using 'pie'  
df4 = df['Country'].value_counts().head()  
df4
```

```
Country  
United Kingdom    981330  
EIRE              17866  
Germany           17624  
France            14330  
Netherlands        5140  
Name: count, dtype: int64
```

```
df4.plot(kind='pie')
```

```
<Axes: ylabel='count'>
```



```
df['Price'].value_counts()
```

Price

1.25 104428

1.65 73808

0.85 69218

2.95 65862

0.42 45438

...

122.15 1

90.94 1

40.04 1

51.87 1

224.69 1

Name: count, Length: 2807, dtype: int64

*#you can also make a Histogram plot*

```
dff = pd.read_csv('IPL_data.csv')
```

```
dff.head()
```

	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				

```

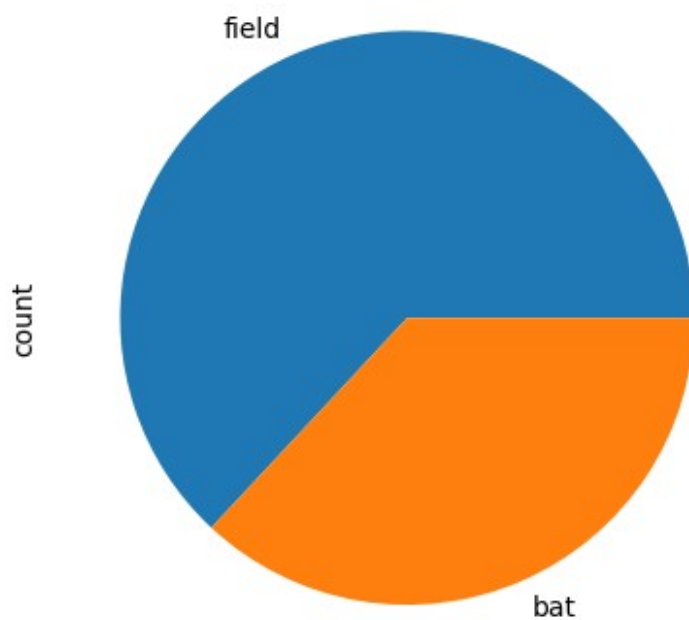
3     VK Sharma
4     NA Patwardhan

dff2 = dff['TossDecision'].value_counts()
dff2

TossDecision
field      599
bat        351
Name: count, dtype: int64

dff2.plot(kind='pie')
<Axes: ylabel='count'>

```



```

dff2 = dff['Margin'].value_counts()
dff2

Margin
6.0      114
7.0      113
5.0       98
8.0       75
4.0       62
...
118.0     1
146.0     1

```

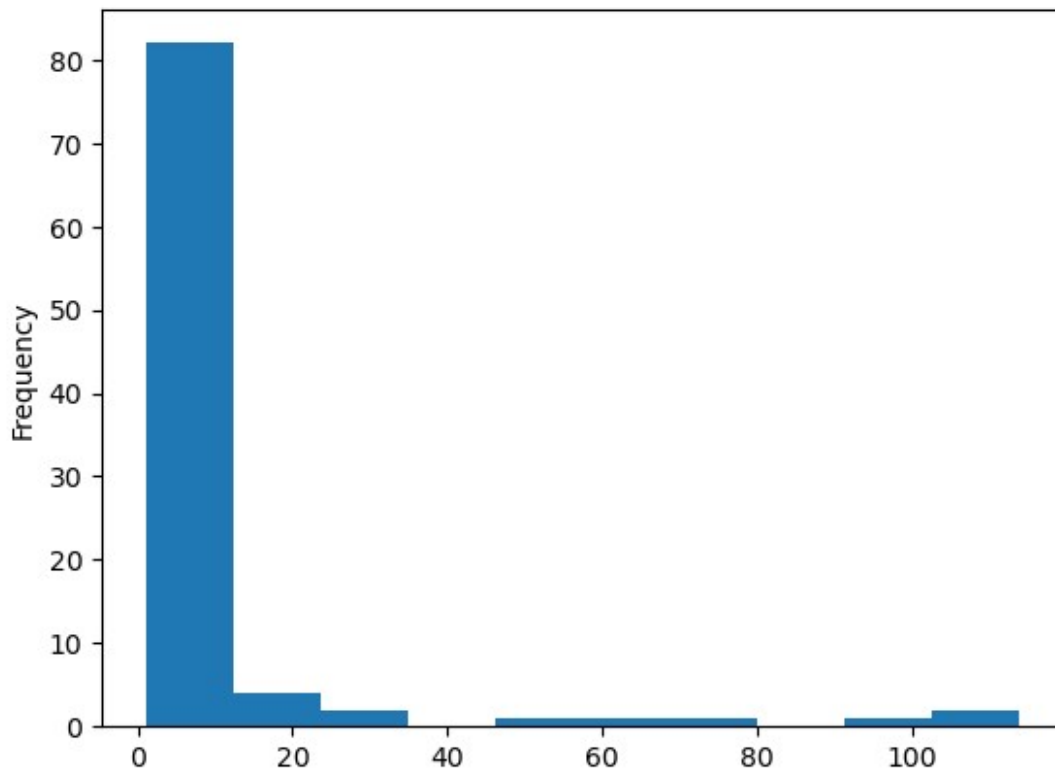
```

91.0      1
52.0      1
140.0     1
Name: count, Length: 94, dtype: int64

dff2.plot(kind= 'hist')

<Axes: ylabel='Frequency'>

```



## Now Talking about SERIES

```

mySeries = dff['WinningTeam'].value_counts()
mySeries  # this returns as series as we already know
# The left side is the index and it is always unique and the right
# side can be anything and my repeat for 2 or more indices values

```

```

WinningTeam
Mumbai Indians      131
Chennai Super Kings 121
Kolkata Knight Riders 114
Royal Challengers Bangalore 109
Rajasthan Royals     96
Kings XI Punjab      88
Sunrisers Hyderabad  75

```



Delhi Daredevils	67
Delhi Capitals	36
Deccan Chargers	29
Gujarat Lions	13
Punjab Kings	13
Pune Warriors	12
Gujarat Titans	12
Rising Pune Supergiant	10
Lucknow Super Giants	9
Kochi Tuskers Kerala	6
Rising Pune Supergiants	5

Name: count, dtype: int64

mySeries.index

```
Index(['Mumbai Indians', 'Chennai Super Kings', 'Kolkata Knight
Riders',
      'Royal Challengers Bangalore', 'Rajasthan Royals', 'Kings XI
Punjab',
      'Sunrisers Hyderabad', 'Delhi Daredevils', 'Delhi Capitals',
      'Deccan Chargers', 'Gujarat Lions', 'Punjab Kings', 'Pune
Warriors',
      'Gujarat Titans', 'Rising Pune Supergiant', 'Lucknow Super
Giants',
      'Kochi Tuskers Kerala', 'Rising Pune Supergiants'],
      dtype='object', name='WinningTeam')
```

mySeries['Delhi Daredevils']

67

dff['Team1'].value\_counts()

Team1	
Royal Challengers Bangalore	126
Mumbai Indians	112
Chennai Super Kings	111
Kolkata Knight Riders	105
Kings XI Punjab	92
Rajasthan Royals	89
Delhi Daredevils	85
Sunrisers Hyderabad	69
Deccan Chargers	39
Delhi Capitals	29
Pune Warriors	23
Punjab Kings	18
Gujarat Lions	16
Lucknow Super Giants	8
Gujarat Titans	7
Rising Pune Supergiant	7
Rising Pune Supergiants	7

```
Kochi Tuskers Kerala      7
Name: count, dtype: int64
```

```
dff['Team2'].value_counts() + dff['Team1'].value_counts()
#YOU CAN ALSO PERFORM PYTHON COMPARATOR OPERATIONS ON SERIES
```

```
Chennai Super Kings      208
Deccan Chargers          75
Delhi Capitals            63
Delhi Daredevils         161
Gujarat Lions            30
Gujarat Titans           16
Kings XI Punjab          190
Kochi Tuskers Kerala     14
Kolkata Knight Riders    223
Lucknow Super Giants     15
Mumbai Indians           231
Pune Warriors            46
Punjab Kings             28
Rajasthan Royals         192
Rising Pune Supergiant   16
Rising Pune Supergiants  14
Royal Challengers Bangalore 226
Sunrisers Hyderabad     152
Name: count, dtype: int64
```

```
(dff['Team2'].value_counts() + dff['Team1'].value_counts()).head()
#we got top 5 teams but they are not sorted so they are not top 5
technically
```

```
Chennai Super Kings      208
Deccan Chargers          75
Delhi Capitals            63
Delhi Daredevils         161
Gujarat Lions            30
Name: count, dtype: int64
```

```
#to sort this we have sort_values() function
(dff['Team2'].value_counts() +
dff['Team1'].value_counts()).sort_values(ascending= False).head()
```

```
#hence we got the top 5 teams in IPL
```

```
Mumbai Indians           231
Royal Challengers Bangalore 226
Kolkata Knight Riders    223
Chennai Super Kings      208
Rajasthan Royals         192
Name: count, dtype: int64
```

```
dff.head()
```

	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

```

```

# suppose we want to sort dataframe according to city names
dff.sort_values('City')

```

	ID	City	Date	Season	MatchNumber \
135	1237180	Abu Dhabi	2020-11-08	2020/21	Qualifier 2
95	1254098	Abu Dhabi	2021-09-26	2021	38
533	729315	Abu Dhabi	2014-04-29	2014	19
181	1216503	Abu Dhabi	2020-10-01	2020/21	13
92	1254099	Abu Dhabi	2021-09-28	2021	42
..	...	...	...	...	...
543	729295	NaN	2014-04-22	2014	9
545	729291	NaN	2014-04-20	2014	7
546	729289	NaN	2014-04-19	2014	6
547	729287	NaN	2014-04-19	2014	5
550	729281	NaN	2014-04-17	2014	2

	Team1	Team2 \
135	Delhi Capitals	Sunrisers Hyderabad
95	Kolkata Knight Riders	Chennai Super Kings
533	Kolkata Knight Riders	Rajasthan Royals
181	Mumbai Indians	Kings XI Punjab
92	Punjab Kings	Mumbai Indians
..	...	...
543	Kings XI Punjab	Sunrisers Hyderabad
545	Rajasthan Royals	Kings XI Punjab
546	Kolkata Knight Riders	Delhi Daredevils
547	Royal Challengers Bangalore	Mumbai Indians
550	Delhi Daredevils	Royal Challengers Bangalore

	Venue	TossWinner
135	Sheikh Zayed Stadium	Delhi Capitals
95	Zayed Cricket Stadium, Abu Dhabi	Kolkata Knight Riders
533	Sheikh Zayed Stadium	Rajasthan Royals
181	Sheikh Zayed Stadium	Kings XI Punjab
92	Zayed Cricket Stadium, Abu Dhabi	Mumbai Indians

```

..          ...
543          Sharjah Cricket Stadium          Sunrisers Hyderabad
545          Sharjah Cricket Stadium          Kings XI Punjab
546  Dubai International Cricket Stadium          Kolkata Knight Riders
547  Dubai International Cricket Stadium  Royal Challengers Bangalore
550          Sharjah Cricket Stadium  Royal Challengers Bangalore

```

```

      TossDecision SuperOver          WinningTeam      WonBy
Margin \
135          bat          N          Delhi Capitals      Runs
17.0
95          bat          N          Chennai Super Kings      Wickets
2.0
533          bat          Y          Rajasthan Royals  SuperOver
NaN
181          field          N          Mumbai Indians      Runs
48.0
92          field          N          Mumbai Indians      Wickets
6.0

```

```

..          ...          ...          ...          ...
...
543          field          N          Kings XI Punjab      Runs
72.0
545          field          N          Kings XI Punjab      Wickets
7.0
546          bat          N          Delhi Daredevils      Wickets
4.0
547          field          N  Royal Challengers Bangalore      Wickets
7.0
550          field          N  Royal Challengers Bangalore      Wickets
8.0

```

```

      method Player_of_Match
Team1Players \
135      NaN      MP Stoinis  ['MP Stoinis', 'S Dhawan', 'SS Iyer', 'SO
Hetm...
95      NaN      RA Jadeja  ['Shubman Gill', 'VR Iyer', 'RA Tripathi',
'EJ...
533      NaN      JP Faulkner  ['G Gambhir', 'MS Bisla', 'JH Kallis', 'MK
Pan...
181      NaN      KA Pollard  ['Q de Kock', 'RG Sharma', 'SA Yadav',
'Ishan ...
92      NaN      KA Pollard  ['KL Rahul', 'Mandeep Singh', 'CH Gayle',

```

```

'AK ...
..    ...    ...
...
543   NaN      GJ Maxwell  ['CA Pujara', 'V Sehwag', 'GJ Maxwell',
'DA Mi...
545   NaN      GJ Maxwell  ['AM Nayar', 'AM Rahane', 'SV Samson', 'SR
Wat...
546   NaN      JP Duminy  ['JH Kallis', 'G Gambhir', 'MK Pandey',
'RV Ut...
547   NaN      PA Patel   ['NJ Maddinson', 'PA Patel', 'V Kohli',
'Yuvra...
550   NaN      YS Chahal  ['MA Agarwal', 'M Vijay', 'KD Karthik',
'MK Ti...

```

	Team2Players	Umpire1 \
135	['PK Garg', 'DA Warner', 'MK Pandey', 'KS Will...	PR Reiffel
95	['RD Gaikwad', 'F du Plessis', 'MM Ali', 'AT R...	CB Gaffaney
533	['AM Rahane', 'KK Nair', 'SV Samson', 'SR Wats...	Aleem Dar
181	['KL Rahul', 'MA Agarwal', 'KK Nair', 'N Poora...	VK Sharma
92	['RG Sharma', 'Q de Kock', 'SA Yadav', 'SS Tiw...	S Ravi
..	...	...
543	['AJ Finch', 'S Dhawan', 'DA Warner', 'KL Rahu...	M Erasmus
545	['CA Pujara', 'V Sehwag', 'WP Saha', 'GJ Maxwe...	BF Bowden
546	['M Vijay', 'MA Agarwal', 'KD Karthik', 'LRPL ...	Aleem Dar
547	['MEK Hussey', 'AP Tare', 'AT Rayudu', 'RG Sha...	Aleem Dar
550	['PA Patel', 'NJ Maddinson', 'V Kohli', 'Yuvra...	Aleem Dar

	Umpire2
135	S Ravi
95	Tapan Sharma
533	AK Chaudhary
181	S Ravi
92	VK Sharma
..	...
543	S Ravi
545	M Erasmus
546	VA Kulkarni
547	AK Chaudhary
550	S Ravi

[950 rows x 20 columns]

All that functions we used on the dataframe does not make the real dataset to be modified, inorder to modify the real dataframe , we can use INPLACE=TRUE

```
#suppose we want to make the above changes to the dff permanent
dff.sort_values('City',inplace=True)
dff
```

	ID	City	Date	Season	MatchNumber	\
135	1237180	Abu Dhabi	2020-11-08	2020/21	Qualifier 2	
95	1254098	Abu Dhabi	2021-09-26	2021	38	
533	729315	Abu Dhabi	2014-04-29	2014	19	
181	1216503	Abu Dhabi	2020-10-01	2020/21	13	
92	1254099	Abu Dhabi	2021-09-28	2021	42	
..	...	...	...	...	...	
543	729295	NaN	2014-04-22	2014	9	
545	729291	NaN	2014-04-20	2014	7	
546	729289	NaN	2014-04-19	2014	6	
547	729287	NaN	2014-04-19	2014	5	
550	729281	NaN	2014-04-17	2014	2	
	Team1		Team2		\	
135	Delhi Capitals		Sunrisers Hyderabad			
95	Kolkata Knight Riders		Chennai Super Kings			
533	Kolkata Knight Riders		Rajasthan Royals			
181	Mumbai Indians		Kings XI Punjab			
92	Punjab Kings		Mumbai Indians			
..	...		...			
543	Kings XI Punjab		Sunrisers Hyderabad			
545	Rajasthan Royals		Kings XI Punjab			
546	Kolkata Knight Riders		Delhi Daredevils			
547	Royal Challengers Bangalore		Mumbai Indians			
550	Delhi Daredevils		Royal Challengers Bangalore			
	Venue		TossWinner		\	
135	Sheikh Zayed Stadium		Delhi Capitals			
95	Zayed Cricket Stadium, Abu Dhabi		Kolkata Knight Riders			
533	Sheikh Zayed Stadium		Rajasthan Royals			
181	Sheikh Zayed Stadium		Kings XI Punjab			
92	Zayed Cricket Stadium, Abu Dhabi		Mumbai Indians			

```

..                                     ...
543          Sharjah Cricket Stadium          Sunrisers Hyderabad
545          Sharjah Cricket Stadium          Kings XI Punjab
546  Dubai International Cricket Stadium          Kolkata Knight Riders
547  Dubai International Cricket Stadium  Royal Challengers Bangalore
550          Sharjah Cricket Stadium  Royal Challengers Bangalore

```

```

      TossDecision SuperOver          WinningTeam      WonBy
Margin \
135          bat          N          Delhi Capitals      Runs
17.0
95          bat          N          Chennai Super Kings      Wickets
2.0
533          bat          Y          Rajasthan Royals  SuperOver
NaN
181          field          N          Mumbai Indians      Runs
48.0
92          field          N          Mumbai Indians      Wickets
6.0
..          ...          ...          ...          ...
...
543          field          N          Kings XI Punjab      Runs
72.0
545          field          N          Kings XI Punjab      Wickets
7.0
546          bat          N          Delhi Daredevils      Wickets
4.0
547          field          N  Royal Challengers Bangalore      Wickets
7.0
550          field          N  Royal Challengers Bangalore      Wickets
8.0

```

```

      method Player_of_Match
Team1Players \
135      NaN      MP Stoinis  ['MP Stoinis', 'S Dhawan', 'SS Iyer', 'SO
Hetm...
95      NaN      RA Jadeja  ['Shubman Gill', 'VR Iyer', 'RA Tripathi',
'EJ...
533      NaN      JP Faulkner  ['G Gambhir', 'MS Bisla', 'JH Kallis', 'MK
Pan...
181      NaN      KA Pollard  ['Q de Kock', 'RG Sharma', 'SA Yadav',
'Ishan ...
92      NaN      KA Pollard  ['KL Rahul', 'Mandeep Singh', 'CH Gayle',
'AK ...

```



```

..      ...      ...
...
543      NaN      GJ Maxwell  ['CA Pujara', 'V Sehwag', 'GJ Maxwell',
'DA Mi...
545      NaN      GJ Maxwell  ['AM Nayar', 'AM Rahane', 'SV Samson', 'SR
Wat...
546      NaN      JP Duminy  ['JH Kallis', 'G Gambhir', 'MK Pandey',
'RV Ut...
547      NaN      PA Patel   ['NJ Maddinson', 'PA Patel', 'V Kohli',
'Yuvra...
550      NaN      YS Chahal  ['MA Agarwal', 'M Vijay', 'KD Karthik',
'MK Ti...

```

```

                                Team2Players      Umpire1 \
135  ['PK Garg', 'DA Warner', 'MK Pandey', 'KS Will...  PR Reiffel
95   ['RD Gaikwad', 'F du Plessis', 'MM Ali', 'AT R...  CB Gaffaney
533  ['AM Rahane', 'KK Nair', 'SV Samson', 'SR Wats...  Aleem Dar
181  ['KL Rahul', 'MA Agarwal', 'KK Nair', 'N Poora...  VK Sharma
92   ['RG Sharma', 'Q de Kock', 'SA Yadav', 'SS Tiw...  S Ravi
..      ...      ...      ...
543  ['AJ Finch', 'S Dhawan', 'DA Warner', 'KL Rahu...  M Erasmus
545  ['CA Pujara', 'V Sehwag', 'WP Saha', 'GJ Maxwe...  BF Bowden
546  ['M Vijay', 'MA Agarwal', 'KD Karthik', 'LRPL ...  Aleem Dar
547  ['MEK Hussey', 'AP Tare', 'AT Rayudu', 'RG Sha...  Aleem Dar
550  ['PA Patel', 'NJ Maddinson', 'V Kohli', 'Yuvra...  Aleem Dar

```

```

      Umpire2
135      S Ravi
95   Tapan Sharma
533  AK Chaudhary
181      S Ravi
92   VK Sharma
..      ...
543      S Ravi
545      M Erasmus
546  VA Kulkarni
547  AK Chaudhary
550      S Ravi

```

```
[950 rows x 20 columns]
```

```

# to sort according to multiple columns
dff.sort_values(['City', 'Date'])

```

```

      ID      City      Date      Season      MatchNumber \
551  729279  Abu Dhabi  2014-04-16      2014              1
549  729283  Abu Dhabi  2014-04-18      2014              3
548  729285  Abu Dhabi  2014-04-18      2014              4
544  729293  Abu Dhabi  2014-04-21      2014              8
538  729305  Abu Dhabi  2014-04-26      2014             14

```

143	1216502	NaN	2020-10-31	2020/21	52
140	1216530	NaN	2020-11-01	2020/21	54
138	1216495	NaN	2020-11-03	2020/21	56
137	1237177	NaN	2020-11-05	2020/21	Qualifier 1
134	1237181	NaN	2020-11-10	2020/21	Final

	Team1	Team2
551	Mumbai Indians	Kolkata Knight Riders
549	Chennai Super Kings	Kings XI Punjab
548	Sunrisers Hyderabad	Rajasthan Royals
544	Chennai Super Kings	Delhi Daredevils
538	Rajasthan Royals	Royal Challengers Bangalore
143	Royal Challengers Bangalore	Sunrisers Hyderabad
140	Kolkata Knight Riders	Rajasthan Royals
138	Mumbai Indians	Sunrisers Hyderabad
137	Mumbai Indians	Delhi Capitals
134	Delhi Capitals	Mumbai Indians

	Venue	TossWinner
551	Sheikh Zayed Stadium	Kolkata Knight Riders
bat		
549	Sheikh Zayed Stadium	Chennai Super Kings
bat		
548	Sheikh Zayed Stadium	Rajasthan Royals
field		
544	Sheikh Zayed Stadium	Chennai Super Kings
bat		
538	Sheikh Zayed Stadium	Rajasthan Royals
field		
...	...	...
143	Sharjah Cricket Stadium	Sunrisers Hyderabad
field		
140	Dubai International Cricket Stadium	Rajasthan Royals
field		
138	Sharjah Cricket Stadium	Sunrisers Hyderabad
field		
137	Dubai International Cricket Stadium	Delhi Capitals
field		
134	Dubai International Cricket Stadium	Delhi Capitals
bat		

SuperOver	WinningTeam	WonBy	Margin	method
Player_of_Match				
551	N Kolkata Knight Riders	Runs	41.0	NaN JH
Kallis				
549	N Kings XI Punjab	Wickets	6.0	NaN GJ

Maxwell						
548	N	Rajasthan Royals	Wickets	4.0	NaN	AM
Rahane						
544	N	Chennai Super Kings	Runs	93.0	NaN	SK
Raina						
538	N	Rajasthan Royals	Wickets	6.0	NaN	PV
Tambe						
..	...	...	...	...	...	
...						
143	N	Sunrisers Hyderabad	Wickets	5.0	NaN	Sandeep
Sharma						
140	N	Kolkata Knight Riders	Runs	60.0	NaN	PJ
Cummins						
138	N	Sunrisers Hyderabad	Wickets	10.0	NaN	S
Nadeem						
137	N	Mumbai Indians	Runs	57.0	NaN	JJ
Bumrah						
134	N	Mumbai Indians	Wickets	5.0	NaN	TA
Boult						

	Team1Players \
551	['MEK Hussey', 'AP Tare', 'AT Rayudu', 'RG Sha...
549	['DR Smith', 'BB McCullum', 'SK Raina', 'MS Dh...
548	['AJ Finch', 'S Dhawan', 'DA Warner', 'KL Rahu...
544	['DR Smith', 'BB McCullum', 'SK Raina', 'F du ...
538	['AM Rahane', 'KK Nair', 'SV Samson', 'AM Naya...
..	...
143	['JR Philippe', 'D Padikkal', 'V Kohli', 'AB d...
140	['Shubman Gill', 'N Rana', 'RA Tripathi', 'SP ...
138	['RG Sharma', 'Q de Kock', 'SA Yadav', 'Ishan ...
137	['Q de Kock', 'RG Sharma', 'SA Yadav', 'Ishan ...
134	['MP Stoinis', 'S Dhawan', 'AM Rahane', 'SS Iy...

	Team2Players
Umpire1 \	
551	['G Gambhir', 'JH Kallis', 'MK Pandey', 'RV Ut... M
Erasmus	
549	['CA Pujara', 'V Sehwag', 'GJ Maxwell', 'AR Pa... RK
Illingworth	
548	['AM Nayar', 'AM Rahane', 'SV Samson', 'SR Wat... BF
Bowden	
544	['MA Agarwal', 'M Vijay', 'KD Karthik', 'MK Ti... RK
Illingworth	
538	['PA Patel', 'YV Takawale', 'V Kohli', 'Yuvraj... HDPK
Dharmasena	
..	...
...	
143	['DA Warner', 'WP Saha', 'MK Pandey', 'KS Will... KN
Ananthapadmanabhan	

```

140 ['RV Uthappa', 'BA Stokes', 'SPD Smith', 'SV S...
Nitin Menon
138 ['DA Warner', 'WP Saha', 'MK Pandey', 'KS Will...      C
Shamshuddin
137 ['PP Shaw', 'S Dhawan', 'AM Rahane', 'SS Iyer'...      CB
Gaffaney
134 ['RG Sharma', 'Q de Kock', 'SA Yadav', 'Ishan ...      CB
Gaffaney

```

```

                Umpire2
551 RK Illingworth
549 C Shamshuddin
548 RK Illingworth
544 C Shamshuddin
538 C Shamshuddin
..      ...
143 K Srinivasan
140 PR Reiffel
138 RK Illingworth
137 Nitin Menon
134 Nitin Menon

```

```
[950 rows x 20 columns]
```

```

# we can choose the ascending orders for both the columns
dff.sort_values(['City', 'Date'], ascending=[True, False])

```

	ID	City	Date	Season	MatchNumber	\
79	1254088	Abu Dhabi	2021-10-08	2021	55	
82	1254095	Abu Dhabi	2021-10-06	2021	52	
88	1254089	Abu Dhabi	2021-10-02	2021	47	
92	1254099	Abu Dhabi	2021-09-28	2021	42	
95	1254098	Abu Dhabi	2021-09-26	2021	38	
..	...	...	...	...	...	...
543	729295	NaN	2014-04-22	2014	9	
545	729291	NaN	2014-04-20	2014	7	
546	729289	NaN	2014-04-19	2014	6	
547	729287	NaN	2014-04-19	2014	5	
550	729281	NaN	2014-04-17	2014	2	

	Team1	Team2	\
79	Mumbai Indians	Sunrisers Hyderabad	
82	Sunrisers Hyderabad	Royal Challengers Bangalore	
88	Chennai Super Kings	Rajasthan Royals	
92	Punjab Kings	Mumbai Indians	
95	Kolkata Knight Riders	Chennai Super Kings	
..	...	...	...
543	Kings XI Punjab	Sunrisers Hyderabad	
545	Rajasthan Royals	Kings XI Punjab	
546	Kolkata Knight Riders	Delhi Daredevils	

547	Royal Challengers Bangalore		Mumbai Indians
550	Delhi Daredevils	Royal Challengers Bangalore	
		Venue	TossWinner
\			
79	Zayed Cricket Stadium, Abu Dhabi		Mumbai Indians
82	Zayed Cricket Stadium, Abu Dhabi	Royal Challengers Bangalore	
88	Zayed Cricket Stadium, Abu Dhabi		Rajasthan Royals
92	Zayed Cricket Stadium, Abu Dhabi		Mumbai Indians
95	Zayed Cricket Stadium, Abu Dhabi		Kolkata Knight Riders
..	...		...
543	Sharjah Cricket Stadium		Sunrisers Hyderabad
545	Sharjah Cricket Stadium		Kings XI Punjab
546	Dubai International Cricket Stadium		Kolkata Knight Riders
547	Dubai International Cricket Stadium	Royal Challengers Bangalore	
550	Sharjah Cricket Stadium	Royal Challengers Bangalore	
		WinningTeam	WonBy
TossDecision	SuperOver		
Margin \			
79	bat	N	Mumbai Indians
42.0			Runs
82	field	N	Sunrisers Hyderabad
4.0			Runs
88	field	N	Rajasthan Royals
7.0			Wickets
92	field	N	Mumbai Indians
6.0			Wickets
95	bat	N	Chennai Super Kings
2.0			Wickets
..	...	...	...
..			...
543	field	N	Kings XI Punjab
72.0			Runs
545	field	N	Kings XI Punjab
7.0			Wickets
546	bat	N	Delhi Daredevils
4.0			Wickets
547	field	N	Royal Challengers Bangalore
7.0			Wickets

550 field N Royal Challengers Bangalore Wickets  
8.0

method Player\_of\_Match

Team1Players \

79 NaN Ishan Kishan ['RG Sharma', 'Ishan Kishan', 'HH Pandya', 'KA...]

82 NaN KS Williamson ['JJ Roy', 'Abhishek Sharma', 'KS Williamson', ...]

88 NaN RD Gaikwad ['RD Gaikwad', 'F du Plessis', 'SK Raina', 'MM...]

92 NaN KA Pollard ['KL Rahul', 'Mandeep Singh', 'CH Gayle', 'AK ...]

95 NaN RA Jadeja ['Shubman Gill', 'VR Iyer', 'RA Tripathi', 'EJ...]

.. ...

543 NaN GJ Maxwell ['CA Pujara', 'V Sehwag', 'GJ Maxwell', 'DA Mi...]

545 NaN GJ Maxwell ['AM Nayar', 'AM Rahane', 'SV Samson', 'SR Wat...]

546 NaN JP Duminy ['JH Kallis', 'G Gambhir', 'MK Pandey', 'RV Ut...]

547 NaN PA Patel ['NJ Maddinson', 'PA Patel', 'V Kohli', 'Yuvra...]

550 NaN YS Chahal ['MA Agarwal', 'M Vijay', 'KD Karthik', 'MK Ti...]

Team2Players

Umpire1 \

79 ['JJ Roy', 'Abhishek Sharma', 'MK Pandey', 'Mo... Tapan Sharma

82 ['V Kohli', 'D Padikkal', 'DT Christian', 'KS ... S Ravi

88 ['E Lewis', 'YBK Jaiswal', 'SV Samson', 'S Dub... CB Gaffaney

92 ['RG Sharma', 'Q de Kock', 'SA Yadav', 'SS Tiw... S Ravi

95 ['RD Gaikwad', 'F du Plessis', 'MM Ali', 'AT R... CB Gaffaney

.. ...

543 ['AJ Finch', 'S Dhawan', 'DA Warner', 'KL Rahu... M Erasmus

545 ['CA Pujara', 'V Sehwag', 'WP Saha', 'GJ Maxwe... BF Bowden

546 ['M Vijay', 'MA Agarwal', 'KD Karthik', 'LRPL ... Aleem Dar

547 ['MEK Hussey', 'AP Tare', 'AT Rayudu', 'RG Sha... Aleem Dar

```

550  ['PA Patel', 'NJ Maddinson', 'V Kohli', 'Yuvra...      Aleem Dar
      Umpire2
79    VK Sharma
82    UV Gandhe
88    VK Sharma
92    VK Sharma
95    Tapan Sharma
..
543    S Ravi
545    M Erasmus
546    VA Kulkarni
547    AK Chaudhary
550    S Ravi

[950 rows x 20 columns]

```

When we have duplicates in the data or We want to find total unique objects we can use `drop_duplicates()` in pandas

```

dff.drop_duplicates(subset=['City'])
dff.drop_duplicates('City')

```

	ID	City	Date	Season	MatchNumber \
135	1237180	Abu Dhabi	2020-11-08	2020/21	Qualifier 2
112	1254079	Ahmedabad	2021-04-27	2021	22
729	501244	Bangalore	2011-05-06	2011	47
200	1178429	Bengaluru	2019-05-04	2019	54
848	392226	Bloemfontein	2009-05-15	2009	46
891	392181	Cape Town	2009-04-18	2009	1
839	392235	Centurion	2009-05-21	2009	55
766	501206	Chandigarh	2011-04-13	2011	9
623	598002	Chennai	2013-04-06	2013	5
524	733987	Cuttack	2014-05-07	2014	29
624	598001	Delhi	2013-04-06	2013	4
781	419159	Dharamsala	2010-04-18	2009/10	54
91	1254103	Dubai	2021-09-29	2021	43
871	392204	Durban	2009-05-01	2009	24
859	392215	East London	2009-05-08	2009	35
649	548329	Hyderabad	2012-05-10	2012	55
719	501254	Indore	2011-05-13	2011	57
285	1136588	Jaipur	2018-04-29	2018	28
842	392232	Johannesburg	2009-05-19	2009	52
379	981007	Kanpur	2016-05-21	2016	54

854	392220	Kimberley	2009-05-11	2009	40
772	501200	Kochi	2011-04-09	2011	3
646	548363	Kolkata	2012-05-12	2012	58
892	336040	Mumbai	2008-06-01	2007/08	Final
789	419151	Nagpur	2010-04-12	2009/10	46
39	1304081	Navi Mumbai	2022-04-23	2022	35
877	392197	Port Elizabeth	2009-04-27	2009	17
371	1082592	Pune	2017-04-06	2017	2
590	598035	Raipur	2013-04-28	2013	39
347	1082616	Rajkot	2017-04-23	2017	26
503	734029	Ranchi	2014-05-22	2014	50
90	1254091	Sharjah	2021-09-30	2021	44
396	980973	Visakhapatnam	2016-05-08	2016	37
134	1237181	NaN	2020-11-10	2020/21	Final

	Team1	Team2	\
135	Delhi Capitals	Sunrisers Hyderabad	
112	Royal Challengers Bangalore	Delhi Capitals	
729	Royal Challengers Bangalore	Kings XI Punjab	
200	Sunrisers Hyderabad	Royal Challengers Bangalore	
848	Delhi Daredevils	Kings XI Punjab	
891	Chennai Super Kings	Mumbai Indians	
839	Delhi Daredevils	Mumbai Indians	
766	Kings XI Punjab	Chennai Super Kings	
623	Chennai Super Kings	Mumbai Indians	
524	Kings XI Punjab	Chennai Super Kings	
624	Delhi Daredevils	Rajasthan Royals	
781	Kings XI Punjab	Chennai Super Kings	
91	Rajasthan Royals	Royal Challengers Bangalore	
871	Royal Challengers Bangalore	Kings XI Punjab	
859	Delhi Daredevils	Mumbai Indians	
649	Deccan Chargers	Delhi Daredevils	
719	Kochi Tuskers Kerala	Kings XI Punjab	
285	Sunrisers Hyderabad	Rajasthan Royals	
842	Royal Challengers Bangalore	Delhi Daredevils	
379	Gujarat Lions	Mumbai Indians	
854	Deccan Chargers	Rajasthan Royals	
772	Kochi Tuskers Kerala	Royal Challengers Bangalore	
646	Kolkata Knight Riders	Mumbai Indians	
892	Chennai Super Kings	Rajasthan Royals	
789	Deccan Chargers	Royal Challengers Bangalore	
39	Gujarat Titans	Kolkata Knight Riders	
877	Kolkata Knight Riders	Mumbai Indians	
371	Rising Pune Supergiant	Mumbai Indians	
590	Delhi Daredevils	Pune Warriors	
347	Gujarat Lions	Kings XI Punjab	
503	Chennai Super Kings	Sunrisers Hyderabad	
90	Sunrisers Hyderabad	Chennai Super Kings	
396	Mumbai Indians	Sunrisers Hyderabad	



134	Delhi Capitals	Mumbai Indians	
		Venue \	
135		Sheikh Zayed Stadium	
112	Narendra Modi Stadium, Ahmedabad		
729	M Chinnaswamy Stadium		
200	M.Chinnaswamy Stadium		
848	OUTsurance Oval		
891	Newlands		
839	SuperSport Park		
766	Punjab Cricket Association Stadium, Mohali		
623	MA Chidambaram Stadium, Chepauk		
524	Barabati Stadium		
624	Feroz Shah Kotla		
781	Himachal Pradesh Cricket Association Stadium		
91	Dubai International Cricket Stadium		
871	Kingsmead		
859	Buffalo Park		
649	Rajiv Gandhi International Stadium, Uppal		
719	Holkar Cricket Stadium		
285	Sawai Mansingh Stadium		
842	New Wanderers Stadium		
379	Green Park		
854	De Beers Diamond Oval		
772	Nehru Stadium		
646	Eden Gardens		
892	Dr DY Patil Sports Academy		
789	Vidarbha Cricket Association Stadium, Jamtha		
39	Dr DY Patil Sports Academy, Mumbai		
877	St George's Park		
371	Maharashtra Cricket Association Stadium		
590	Shaheed Veer Narayan Singh International Stadium		
347	Saurashtra Cricket Association Stadium		
503	JSCA International Stadium Complex		
90	Sharjah Cricket Stadium		
396	Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket St...		
134	Dubai International Cricket Stadium		
	TossWinner	TossDecision	SuperOver \
135	Delhi Capitals	bat	N
112	Delhi Capitals	field	N
729	Kings XI Punjab	field	N
200	Royal Challengers Bangalore	field	N
848	Kings XI Punjab	field	N
891	Chennai Super Kings	field	N
839	Delhi Daredevils	field	N
766	Kings XI Punjab	field	N
623	Mumbai Indians	bat	N
524	Chennai Super Kings	field	N

624	Rajasthan Royals	bat	N
781	Chennai Super Kings	field	N
91	Royal Challengers Bangalore	field	N
871	Royal Challengers Bangalore	bat	N
859	Mumbai Indians	bat	N
649	Deccan Chargers	bat	N
719	Kings XI Punjab	field	N
285	Sunrisers Hyderabad	bat	N
842	Delhi Daredevils	bat	N
379	Gujarat Lions	field	N
854	Deccan Chargers	bat	N
772	Kochi Tuskers Kerala	bat	N
646	Mumbai Indians	bat	N
892	Rajasthan Royals	field	N
789	Royal Challengers Bangalore	field	N
39	Gujarat Titans	bat	N
877	Mumbai Indians	bat	N
371	Rising Pune Supergiant	field	N
590	Pune Warriors	field	N
347	Gujarat Lions	field	N
503	Sunrisers Hyderabad	field	N
90	Chennai Super Kings	field	N
396	Mumbai Indians	field	N
134	Delhi Capitals	bat	N
WinningTeam WonBy Margin method			
Player_of_Match \			
135	Delhi Capitals	Runs	17.0 NaN MP
Stoinis			
112	Royal Challengers Bangalore	Runs	1.0 NaN AB de
Villiers			
729	Royal Challengers Bangalore	Runs	85.0 NaN CH
Gayle			
200	Royal Challengers Bangalore	Wickets	4.0 NaN SO
Hetmyer			
848	Kings XI Punjab	Wickets	6.0 NaN B
Lee			
891	Mumbai Indians	Runs	19.0 NaN SR
Tendulkar			
839	Delhi Daredevils	Wickets	4.0 NaN V
Sehwag			
766	Kings XI Punjab	Wickets	6.0 NaN PC
Valthaty			
623	Mumbai Indians	Runs	9.0 NaN KA
Pollard			
524	Kings XI Punjab	Runs	44.0 NaN GJ
Maxwell			
624	Rajasthan Royals	Runs	5.0 NaN R
Dravid			

781	Chennai Super Kings	Wickets	6.0	NaN	MS
Dhoni					
91	Royal Challengers Bangalore	Wickets	7.0	NaN	YS
Chahal					
871	Royal Challengers Bangalore	Runs	8.0	NaN	Yuvraj
Singh					
859	Delhi Daredevils	Wickets	7.0	NaN	A
Nehra					
649	Delhi Daredevils	Wickets	9.0	NaN	DA
Warner					
719	Kings XI Punjab	Wickets	6.0	NaN	KD
Karthik					
285	Sunrisers Hyderabad	Runs	11.0	NaN	KS
Williamson					
842	Royal Challengers Bangalore	Wickets	7.0	NaN	JH
Kallis					
379	Gujarat Lions	Wickets	6.0	NaN	SK
Raina					
854	Deccan Chargers	Runs	53.0	NaN	DR
Smith					
772	Royal Challengers Bangalore	Wickets	6.0	NaN	AB de
Villiers					
646	Mumbai Indians	Runs	27.0	NaN	RG
Sharma					
892	Rajasthan Royals	Wickets	3.0	NaN	YK
Pathan					
789	Deccan Chargers	Runs	13.0	NaN	Harmeet
Singh					
39	Gujarat Titans	Runs	8.0	NaN	Rashid
Khan					
877	Mumbai Indians	Runs	92.0	NaN	SR
Tendulkar					
371	Rising Pune Supergiant	Wickets	7.0	NaN	SPD
Smith					
590	Delhi Daredevils	Runs	15.0	NaN	DA
Warner					
347	Kings XI Punjab	Runs	26.0	NaN	HM
Amla					
503	Sunrisers Hyderabad	Wickets	6.0	NaN	DA
Warner					
90	Chennai Super Kings	Wickets	6.0	NaN	JR
Hazlewood					
396	Sunrisers Hyderabad	Runs	85.0	NaN	A
Nehra					
134	Mumbai Indians	Wickets	5.0	NaN	TA
Boult					
Team1Players \					
135	['MP Stoinis', 'S Dhawan', 'SS Iyer', 'SO Hetm...				

112 ['V Kohli', 'D Padikkal', 'RM Patidar', 'GJ Ma...  
 729 ['CH Gayle', 'TM Dilshan', 'V Kohli', 'AB de V...  
 200 ['WP Saha', 'MJ Guptill', 'MK Pandey', 'KS Wil...  
 848 ['G Gambhir', 'V Sehwag', 'TM Dilshan', 'AB de...  
 891 ['PA Patel', 'ML Hayden', 'SK Raina', 'A Flint...  
 839 ['G Gambhir', 'DA Warner', 'V Sehwag', 'TM Dil...  
 766 ['PC Valthaty', 'AC Gilchrist', 'SE Marsh', 'S...  
 623 ['MEK Hussey', 'M Vijay', 'SK Raina', 'DJ Brav...  
 524 ['V Sehwag', 'Mandeep Singh', 'GJ Maxwell', 'D...  
 624 ['DA Warner', 'UBT Chand', 'DPMD Jayawardene',...  
 781 ['DPMD Jayawardene', 'SE Marsh', 'KC Sangakkara...  
 91 ['E Lewis', 'YBK Jaiswal', 'SV Samson', 'MK Lo...  
 871 ['JD Ryder', 'SP Goswami', 'R Bishnoi', 'JH Ka...  
 859 ['G Gambhir', 'DA Warner', 'TM Dilshan', 'AB d...  
 649 ['DJ Harris', 'S Dhawan', 'KC Sangakkara', 'CL...  
 719 ['BB McCullum', 'DPMD Jayawardene', 'RA Jadeja...  
 285 ['AD Hales', 'S Dhawan', 'KS Williamson', 'MK ...  
 842 ['RV Uthappa', 'JH Kallis', 'R Dravid', 'LRPL ...  
 379 ['AJ Finch', 'BB McCullum', 'SK Raina', 'KD Ka...  
 854 ['TL Suman', 'AC Gilchrist', 'RG Sharma', 'A S...  
 772 ['BB McCullum', 'VVS Laxman', 'DPMD Jayawarden...  
 646 ['MS Bisla', 'G Gambhir', 'JH Kallis', 'MK Tiw...  
 892 ['PA Patel', 'S Vidyut', 'SK Raina', 'JA Morke...  
 789 ['HH Gibbs', 'AC Gilchrist', 'TL Suman', 'RG S...  
 39 ['WP Saha', 'Shubman Gill', 'HH Pandya', 'DA M...  
 877 ['BB McCullum', 'CH Gayle', 'BJ Hodge', 'SC Ga...  
 371 ['AM Rahane', 'MA Agarwal', 'SPD Smith', 'BA S...  
 590 ['DPMD Jayawardene', 'V Sehwag', 'UBT Chand', ...  
 347 ['BB McCullum', 'AJ Finch', 'SK Raina', 'KD Ka...  
 503 ['DR Smith', 'F du Plessis', 'SK Raina', 'DJ H...  
 90 ['JJ Roy', 'WP Saha', 'KS Williamson', 'PK Gar...  
 396 ['RG Sharma', 'PA Patel', 'AT Rayudu', 'KH Pan...  
 134 ['MP Stoinis', 'S Dhawan', 'AM Rahane', 'SS Iy...

## Team2Players

Umpire1 \

135 ['PK Garg', 'DA Warner', 'MK Pandey', 'KS Will... PR

Reiffel

112 ['PP Shaw', 'S Dhawan', 'SPD Smith', 'RR Pant'...

S Ravi

729 ['AC Gilchrist', 'PC Valthaty', 'AM Nayar', 'S...

Aleem Dar

200 ['PA Patel', 'V Kohli', 'AB de Villiers', 'SO ... AK

Chaudhary

848 ['S Sohal', 'SM Katich', 'LA Pomersbach', 'Yuv... HDPK

Dharmasena

891 ['ST Jayasuriya', 'SR Tendulkar', 'S Dhawan', ... BR

Doctrove

839 ['ST Jayasuriya', 'SR Tendulkar', 'JP Duminy',... IL

Howell  
766 ['S Anirudha', 'M Vijay', 'SK Raina', 'S Badri...  
Asad Rauf  
623 ['RT Ponting', 'SR Tendulkar', 'KD Karthik', '... M  
Erasmus  
524 ['DR Smith', 'BB McCullum', 'SK Raina', 'RA Ja... HDPK  
Dharmasena  
624 ['MDKJ Perera', 'AM Rahane', 'R Dravid', 'STR ...  
S Das  
781 ['M Vijay', 'ML Hayden', 'SK Raina', 'S Badrin... BF  
Bowden  
91 ['V Kohli', 'D Padikkal', 'KS Bharat', 'GJ Max... AY  
Dandekar  
871 ['Yuvraj Singh', 'K Goel', 'SM Katich', 'KC Sa... HDPK  
Dharmasena  
859 ['L Ronchi', 'JP Duminy', 'PR Shah', 'SR Tendu... M  
Erasmus  
649 ['V Sehwag', 'DA Warner', 'NV Ojha', 'DPMD Jay... JD  
Cloete  
719 ['AC Gilchrist', 'PC Valthaty', 'SE Marsh', 'K... S  
Asnani  
285 ['AM Rahane', 'RA Tripathi', 'SV Samson', 'BA ... BNJ  
Oxenford  
842 ['G Gambhir', 'V Sehwag', 'M Manhas', 'AB de V... IL  
Howell  
379 ['RG Sharma', 'MJ Guptill', 'N Rana', 'KH Pand... AK  
Chaudhary  
854 ['GC Smith', 'SA Asnodkar', 'LA Carseldine', '... GAV  
Baxter  
772 ['MA Agarwal', 'TM Dilshan', 'V Kohli', 'AB de... HDPK  
Dharmasena  
646 ['HH Gibbs', 'SR Tendulkar', 'RG Sharma', 'KD ...  
S Ravi  
892 ['NK Patel', 'SA Asnodkar', 'Kamran Akmal', 'S... BF  
Bowden  
789 ['MK Pandey', 'JH Kallis', 'R Dravid', 'RV Uth... RE  
Koertzen  
39 ['SW Billings', 'SP Narine', 'SS Iyer', 'N Ran... KN  
Ananthapadmanabhan  
877 ['ST Jayasuriya', 'SR Tendulkar', 'Harbhajan S... BG  
Jerling  
371 ['PA Patel', 'JC Buttler', 'RG Sharma', 'N Ran... A Nand  
Kishore  
590 ['RV Uthappa', 'AJ Finch', 'Yuvraj Singh', 'LJ... CK  
Nandan  
347 ['M Vohra', 'HM Amla', 'SE Marsh', 'GJ Maxwell... AK  
Chaudhary  
503 ['DA Warner', 'S Dhawan', 'NV Ojha', 'AJ Finch... BNJ  
Oxenford

```

90    ['RD Gaikwad', 'F du Plessis', 'MM Ali', 'SK R...
Nitin Menon
396    ['DA Warner', 'S Dhawan', 'KS Williamson', 'Yu...
S Ravi
134    ['RG Sharma', 'Q de Kock', 'SA Yadav', 'Ishan ...      CB
Gaffaney

```

```

                                Umpire2
135                                S Ravi
112                                VK Sharma
729                                RB Tiffin
200                                NJ Llong
848                                IL Howell
891                                K Hariharan
839                                S Ravi
766                                SL Shastri
623                                VA Kulkarni
524                                PG Pathak
624                                C Shamshuddin
781                                AM Saheba
91    KN Ananthapadmanabhan
871                                S Ravi
859                                SK Tarapore
649                                SJA Taufel
719                                RJ Tucker
285                                A Nand Kishore
842                                RB Tiffin
379                                CK Nandan
854                                HDPK Dharmasena
772                                K Hariharan
646                                SJA Taufel
892                                RE Koertzen
789                                RB Tiffin
39                                UV Gandhe
877                                RB Tiffin
371                                S Ravi
590                                S Ravi
347                                M Erasmus
503                                C Shamshuddin
90                                YC Barde
396                                C Shamshuddin
134                                Nitin Menon

```

```
dff.drop_duplicates(subset=['City'],keep='last').shape
```

```
(34, 20)
```

```
dff.drop_duplicates(subset=['City','Season'])
```

	ID	City	Date	Season	MatchNumber	\
135	1237180	Abu Dhabi	2020-11-08	2020/21	Qualifier 2	
95	1254098	Abu Dhabi	2021-09-26	2021	38	
533	729315	Abu Dhabi	2014-04-29	2014	19	
112	1254079	Ahmedabad	2021-04-27	2021	22	
522	733989	Ahmedabad	2014-05-08	2014	30	
..	...	...	...	...	...	
478	829731	Visakhapatnam	2015-04-18	2015	13	
693	548314	Visakhapatnam	2012-04-09	2012	9	
196	1181766	Visakhapatnam	2019-05-08	2019	Eliminator	
134	1237181	NaN	2020-11-10	2020/21	Final	
532	729317	NaN	2014-04-30	2014	20	
		Team1		Team2	\	
135		Delhi Capitals		Sunrisers Hyderabad		
95		Kolkata Knight Riders		Chennai Super Kings		
533		Kolkata Knight Riders		Rajasthan Royals		
112	Royal	Challengers Bangalore		Delhi Capitals		
522		Rajasthan Royals		Sunrisers Hyderabad		
..		...		...		
478		Sunrisers Hyderabad		Delhi Daredevils		
693		Deccan Chargers		Mumbai Indians		
196		Sunrisers Hyderabad		Delhi Capitals		
134		Delhi Capitals		Mumbai Indians		
532		Mumbai Indians		Sunrisers Hyderabad		
					Venue	
TossWinner	\					
135					Sheikh Zayed Stadium	Delhi
Capitals						
95					Zayed Cricket Stadium, Abu Dhabi	Kolkata Knight
Riders						
533					Sheikh Zayed Stadium	Rajasthan
Royals						
112					Narendra Modi Stadium, Ahmedabad	Delhi
Capitals						
522					Sardar Patel Stadium, Motera	Rajasthan
Royals						
..					...	
...						
478	Dr. Y.S. Rajasekhara Reddy	ACA-VDCA Cricket St...				Delhi
Daredevils						
693	Dr. Y.S. Rajasekhara Reddy	ACA-VDCA Cricket St...				Deccan
Chargers						
196	Dr. Y.S. Rajasekhara Reddy	ACA-VDCA Cricket St...				Delhi
Capitals						
134					Dubai International Cricket Stadium	Delhi
Capitals						
532					Dubai International Cricket Stadium	Mumbai
Indians						

Margin \	TossDecision	SuperOver	WinningTeam	WonBy
135	bat	N	Delhi Capitals	Runs
17.0				
95	bat	N	Chennai Super Kings	Wickets
2.0				
533	bat	Y	Rajasthan Royals	SuperOver
NaN				
112	field	N	Royal Challengers Bangalore	Runs
1.0				
522	field	N	Sunrisers Hyderabad	Runs
32.0				
..	...	...	...	...
...				
478	bat	N	Delhi Daredevils	Runs
4.0				
693	bat	N	Mumbai Indians	Wickets
5.0				
196	field	N	Delhi Capitals	Wickets
2.0				
134	bat	N	Mumbai Indians	Wickets
5.0				
532	field	N	Sunrisers Hyderabad	Runs
15.0				

method	Player_of_Match
Team1Players \	
135 NaN MP Stoinis	['MP Stoinis', 'S Dhawan', 'SS Iyer', 'SO Hetm...']
95 NaN RA Jadeja	['Shubman Gill', 'VR Iyer', 'RA Tripathi', 'EJ...']
533 NaN JP Faulkner	['G Gambhir', 'MS Bisla', 'JH Kallis', 'MK Pan...']
112 NaN AB de Villiers	['V Kohli', 'D Padikkal', 'RM Patidar', 'GJ Ma...']
522 NaN B Kumar	['AM Rahane', 'KK Nair', 'SV Samson', 'SR Wats...']
.. ...	...
...	
478 NaN JP Duminy	['DA Warner', 'S Dhawan', 'RS Bopara', 'KL Rah...']
693 NaN RG Sharma	['PA Patel', 'S Dhawan', 'B Chipli', 'DT Chris...']
196 NaN RR Pant	['WP Saha', 'MJ Guptill', 'MK Pandey', 'KS Wil...']
134 NaN TA Boult	['MP Stoinis', 'S Dhawan', 'AM Rahane', 'SS Iy...']
532 NaN B Kumar	['BR Dunk', 'RG Sharma', 'CJ Anderson', 'AT Ra...']



```

Team2Players
Umpire1 \
135 ['PK Garg', 'DA Warner', 'MK Pandey', 'KS Will... PR
Reiffel
95 ['RD Gaikwad', 'F du Plessis', 'MM Ali', 'AT R... CB
Gaffaney
533 ['AM Rahane', 'KK Nair', 'SV Samson', 'SR Wats... Aleem
Dar
112 ['PP Shaw', 'S Dhawan', 'SPD Smith', 'RR Pant'... S
Ravi
522 ['S Dhawan', 'AJ Finch', 'KL Rahul', 'NV Ojha'... AK
Chaudhary
.. ... ..
.
478 ['MA Agarwal', 'SS Iyer', 'JP Duminy', 'Yuvraj... PG
Pathak
693 ['TL Suman', 'RE Levi', 'RG Sharma', 'AT Rayud... AK
Chaudhary
196 ['PP Shaw', 'S Dhawan', 'SS Iyer', 'RR Pant', ... BNJ
Oxenford
134 ['RG Sharma', 'Q de Kock', 'SA Yadav', 'Ishan ... CB
Gaffaney
532 ['AJ Finch', 'S Dhawan', 'KL Rahul', 'DA Warne... HDPK
Dharmasena

Umpire2
135 S Ravi
95 Tapan Sharma
533 AK Chaudhary
112 VK Sharma
522 NJ Llong
.. ...
478 S Ravi
693 JD Cloete
196 S Ravi
134 Nitin Menon
532 M Erasmus

[140 rows x 20 columns]

dff.drop_duplicates(subset=['City', 'Season']).shape

(140, 20)

# to find winners of every season in IPL , we can use drop_duplicates
smartly here
# as every season's last match is the final match we can drop
duplicate seasons and keep the last one
dff2 = dff.sort_values(['Season', 'Date'])

```

```
dff2
dff2.drop_duplicates('Season',keep='last')
```

	ID	City	Date	Season	MatchNumber	\
892	336040	Mumbai	2008-06-01	2007/08	Final	
835	392239	Johannesburg	2009-05-24	2009	Final	
775	419165	Mumbai	2010-04-25	2009/10	Final	
702	501271	Chennai	2011-05-28	2011	Final	
628	548381	Chennai	2012-05-27	2012	Final	
552	598073	Kolkata	2013-05-26	2013	Final	
492	734049	Bangalore	2014-06-01	2014	Final	
433	829823	Kolkata	2015-05-24	2015	Final	
373	981019	Bangalore	2016-05-29	2016	Final	
314	1082650	Hyderabad	2017-05-21	2017	Final	
254	1136620	Mumbai	2018-05-27	2018	Final	
194	1181768	Hyderabad	2019-05-12	2019	Final	
134	1237181	NaN	2020-11-10	2020/21	Final	
74	1254117	Dubai	2021-10-15	2021	Final	
0	1312200	Ahmedabad	2022-05-29	2022	Final	

	Team1	Team2	\
892	Chennai Super Kings	Rajasthan Royals	
835	Royal Challengers Bangalore	Deccan Chargers	
775	Chennai Super Kings	Mumbai Indians	
702	Chennai Super Kings	Royal Challengers Bangalore	
628	Kolkata Knight Riders	Chennai Super Kings	
552	Chennai Super Kings	Mumbai Indians	
492	Kolkata Knight Riders	Kings XI Punjab	
433	Mumbai Indians	Chennai Super Kings	
373	Royal Challengers Bangalore	Sunrisers Hyderabad	
314	Mumbai Indians	Rising Pune Supergiant	
254	Sunrisers Hyderabad	Chennai Super Kings	
194	Mumbai Indians	Chennai Super Kings	
134	Delhi Capitals	Mumbai Indians	
74	Chennai Super Kings	Kolkata Knight Riders	
0	Rajasthan Royals	Gujarat Titans	

	Venue	
TossWinner	\	
892	Dr DY Patil Sports Academy	Rajasthan Royals
835	New Wanderers Stadium	Royal Challengers Bangalore
775	Dr DY Patil Sports Academy	Chennai Super Kings
702	MA Chidambaram Stadium, Chepauk	Chennai Super Kings
628	MA Chidambaram Stadium, Chepauk	Chennai Super Kings
552	Eden Gardens	Mumbai

Indians		
492	M Chinnaswamy Stadium	Kolkata Knight
Riders		
433	Eden Gardens	Chennai Super
Kings		
373	M Chinnaswamy Stadium	Sunrisers
Hyderabad		
314	Rajiv Gandhi International Stadium, Uppal	Mumbai
Indians		
254	Wankhede Stadium	Chennai Super
Kings		
194	Rajiv Gandhi International Stadium	Mumbai
Indians		
134	Dubai International Cricket Stadium	Delhi
Capitals		
74	Dubai International Cricket Stadium	Kolkata Knight
Riders		
0	Narendra Modi Stadium, Ahmedabad	Rajasthan
Royals		

	TossDecision	SuperOver	WinningTeam	WonBy	Margin
method \					
892	field	N	Rajasthan Royals	Wickets	3.0
NaN					
835	field	N	Deccan Chargers	Runs	6.0
NaN					
775	bat	N	Chennai Super Kings	Runs	22.0
NaN					
702	bat	N	Chennai Super Kings	Runs	58.0
NaN					
628	bat	N	Kolkata Knight Riders	Wickets	5.0
NaN					
552	bat	N	Mumbai Indians	Runs	23.0
NaN					
492	field	N	Kolkata Knight Riders	Wickets	3.0
NaN					
433	field	N	Mumbai Indians	Runs	41.0
NaN					
373	bat	N	Sunrisers Hyderabad	Runs	8.0
NaN					
314	bat	N	Mumbai Indians	Runs	1.0
NaN					
254	field	N	Chennai Super Kings	Wickets	8.0
NaN					
194	bat	N	Mumbai Indians	Runs	1.0
NaN					
134	bat	N	Mumbai Indians	Wickets	5.0
NaN					
74	field	N	Chennai Super Kings	Runs	27.0

NaN					
0	bat	N	Gujarat Titans	Wickets	7.0
NaN					
	Player_of_Match			Team1Players	
\					
892	YK Pathan	['PA Patel', 'S Vidyut', 'SK Raina', 'JA Morke...			
835	A Kumble	['JH Kallis', 'MK Pandey', 'RE van der Merwe',...			
775	SK Raina	['M Vijay', 'ML Hayden', 'SK Raina', 'S Badrin...			
702	M Vijay	['MEK Hussey', 'M Vijay', 'MS Dhoni', 'SK Rain...			
628	MS Bisla	['MS Bisla', 'G Gambhir', 'JH Kallis', 'LR Shu...			
552	KA Pollard	['MEK Hussey', 'M Vijay', 'SK Raina', 'S Badri...			
492	MK Pandey	['RV Uthappa', 'G Gambhir', 'MK Pandey', 'YK P...			
433	RG Sharma	['LMP Simmons', 'PA Patel', 'RG Sharma', 'KA P...			
373	BCJ Cutting	['CH Gayle', 'V Kohli', 'AB de Villiers', 'KL ...			
314	KH Pandya	['LMP Simmons', 'PA Patel', 'AT Rayudu', 'RG S...			
254	SR Watson	['SP Goswami', 'S Dhawan', 'KS Williamson', 'S...			
194	JJ Bumrah	['Q de Kock', 'RG Sharma', 'SA Yadav', 'Ishan ...			
134	TA Boult	['MP Stoinis', 'S Dhawan', 'AM Rahane', 'SS Iy...			
74	F du Plessis	['RD Gaikwad', 'F du Plessis', 'RV Uthappa', '...			
0	HH Pandya	['YBK Jaiswal', 'JC Buttler', 'SV Samson', 'D ...			
			Team2Players		
Umpire1	\				
892	['NK Patel', 'SA Asnodkar', 'Kamran Akmal', 'S...				BF
Bowden					
835	['AC Gilchrist', 'HH Gibbs', 'TL Suman', 'A Sy...				RE
Koertzen					
775	['S Dhawan', 'SR Tendulkar', 'AM Nayar', 'Harb...				RE
Koertzen					
702	['MA Agarwal', 'CH Gayle', 'V Kohli', 'AB de V...				Asad
Rauf					
628	['MEK Hussey', 'M Vijay', 'SK Raina', 'MS Dhon...				BF
Bowden					
552	['DR Smith', 'AP Tare', 'KD Karthik', 'RG Shar...				HDPK
Dharmasena					

```

492 ['V Sehwag', 'M Vohra', 'GJ Bailey', 'WP Saha'... HDPK
Dharmasena
433 ['DR Smith', 'MEK Hussey', 'SK Raina', 'MS Dho... HDPK
Dharmasena
373 ['DA Warner', 'S Dhawan', 'MC Henriques', 'Yuv... HDPK
Dharmasena
314 ['AM Rahane', 'RA Tripathi', 'SPD Smith', 'MS ... NJ
Llong
254 ['SR Watson', 'F du Plessis', 'SK Raina', 'AT ... M
Erasmus
194 ['F du Plessis', 'SR Watson', 'SK Raina', 'AT ... IJ
Gould
134 ['RG Sharma', 'Q de Kock', 'SA Yadav', 'Ishan ... CB
Gaffaney
74 ['Shubman Gill', 'VR Iyer', 'N Rana', 'SP Nari... Nitin
Menon
0 ['WP Saha', 'Shubman Gill', 'MS Wade', 'HH Pan... CB
Gaffaney

```

```

        Umpire2
892 RE Koertzen
835 SJA Taufel
775 SJA Taufel
702 SJA Taufel
628 SJA Taufel
552 SJA Taufel
492 BNJ Oxenford
433 RK Illingworth
373 BNJ Oxenford
314 S Ravi
254 S Ravi
194 Nitin Menon
134 Nitin Menon
74 RK Illingworth
0 Nitin Menon

```

```
dff2.drop_duplicates('Season',keep='last')[['Season','WinningTeam']]
```

	Season	WinningTeam
892	2007/08	Rajasthan Royals
835	2009	Deccan Chargers
775	2009/10	Chennai Super Kings
702	2011	Chennai Super Kings
628	2012	Kolkata Knight Riders
552	2013	Mumbai Indians
492	2014	Kolkata Knight Riders
433	2015	Mumbai Indians
373	2016	Sunrisers Hyderabad
314	2017	Mumbai Indians
254	2018	Chennai Super Kings

194	2019	Mumbai Indians
134	2020/21	Mumbai Indians
74	2021	Chennai Super Kings
0	2022	Gujarat Titans

## drop\_duplicates is a Tricky function

## groupby function

```
company = pd.read_csv('fortune500.csv')
company
```

	company	rank	revenue	profit	num. of employees \
0	Walmart	1	572754.0	13673.0	2300000.0
1	Amazon	2	469822.0	33364.0	1608000.0
2	Apple	3	365817.0	94680.0	154000.0
3	CVS Health	4	292111.0	7910.0	258000.0
4	UnitedHealth Group	5	287597.0	17285.0	350000.0
..	...	...	...	...	...
995	Vizio Holding	996	2124.0	-39.4	800.0
996	1-800-Flowers.com	997	2122.2	118.7	4800.0
997	Cowen	998	2112.8	295.6	1534.0
998	Ashland	999	2111.0	220.0	4100.0
999	DocuSign	1000	2107.2	-70.0	7461.0

	sector	city	state	profitable
0	Retailing	Bentonville	AR	yes
1	Retailing	Seattle	WA	yes
2	Technology	Cupertino	CA	yes
3	Health Care	Woonsocket	RI	yes
4	Health Care	Minnetonka	MN	yes
..	...	...	...	...
995	Industrials	Irvine	CA	no
996	Retailing	Jericho	NY	yes
997	Financials	New York	NY	yes
998	Chemicals	Wilmington	DE	yes
999	Technology	San Francisco	CA	no

```
[1000 rows x 9 columns]
```

```
company.groupby('sector')
```

*#this will give unexpected output but don't panic , it looks like that only, but we can do soo many things with this*

```
sectors = company.groupby('sector')
```

```
sectors
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x00000282C1253410>
```

```
len(sectors) # gives number of sectors present in the dataset
```

```
21
```

```
sectors.size() # gives each sector's size separately
```

```
sector
Aerospace & Defense    17
Apparel                16
Business Services     52
Chemicals              29
Energy                100
Engineering & Construction 32
Financials            166
Food & Drug Stores      9
Food, Beverages & Tobacco 34
Health Care           77
Hotels, Restaurants & Leisure 28
Household Products    23
Industrials           50
Materials             46
Media                 28
Motor Vehicles & Parts   20
Retailing             77
Technology            121
Telecommunications     9
Transportation        35
Wholesalers           31
dtype: int64
```

```
sectors.size().sort_values(ascending = False)
```

```
# this gives us all the sorted list of sectors in terms on numbers of companies belonging to them
```

```
sector
Financials            166
Technology            121
Energy               100
Health Care           77
Retailing            77
Business Services     52
Industrials           50
Materials             46
Transportation        35
Food, Beverages & Tobacco 34
Engineering & Construction 32
Wholesalers           31
Chemicals             29
```

Media	28
Hotels, Restaurants & Leisure	28
Household Products	23
Motor Vehicles & Parts	20
Aerospace & Defense	17
Apparel	16
Telecommunications	9
Food & Drug Stores	9

dtype: int64

```
sectors.first()
```

*#this will give name of company occuring at top in each sector with rank*

*#we can get the top companies in each sector through this*

revenue \ sector	company	rank
Aerospace & Defense 67044.0	Lockheed Martin	55
Apparel 44538.0	Nike	83
Business Services 25371.0	PayPal	143
Chemicals 54968.0	Dow	64
Energy 285640.0	ExxonMobil	6
Engineering & Construction 27774.2	D.R. Horton	124
Financials 276094.0	Berkshire Hathaway	7
Food & Drug Stores 148579.0	Walgreens Boots Alliance	18
Food, Beverages & Tobacco 85249.0	Archer Daniels Midland	38
Health Care 292111.0	CVS Health	4
Hotels, Restaurants & Leisure 29060.6	Starbucks	120
Household Products 76118.0	Procter & Gamble	47
Industrials 74196.0	General Electric	48
Materials 36483.9	Nucor	98
Media 67418.0	Walt Disney	53
Motor Vehicles & Parts	Ford Motor	22



136341.0			
Retailing		Walmart	1
572754.0			
Technology		Apple	3
365817.0			
Telecommunications		AT&T	13
168864.0			
Transportation		UPS	34
97287.0			
Wholesalers		Sysco	70
51297.8			
	profit	num. of employees	city
state \			
sector			
Aerospace & Defense	6315.0	114000.0	Bethesda
MD			
Apparel	5727.0	73300.0	Beaverton
OR			
Business Services	4169.0	30900.0	San Jose
CA			
Chemicals	6311.0	35700.0	Midland
MI			
Energy	23040.0	63000.0	Irving
TX			
Engineering & Construction	4175.8	11788.0	Arlington
TX			
Financials	89795.0	372000.0	Omaha
NE			
Food & Drug Stores	2542.0	258500.0	Deerfield
IL			
Food, Beverages & Tobacco	2709.0	39979.0	Chicago
IL			
Health Care	7910.0	258000.0	Woonsocket
RI			
Hotels, Restaurants & Leisure	4199.3	383000.0	Seattle
WA			
Household Products	14306.0	101000.0	Cincinnati
OH			
Industrials	-6520.0	168000.0	Boston
MA			
Materials	6827.5	28800.0	Charlotte
NC			
Media	1995.0	171000.0	Burbank
CA			
Motor Vehicles & Parts	17937.0	183000.0	Dearborn
MI			
Retailing	13673.0	2300000.0	Bentonville

AR			
Technology	94680.0	154000.0	Cupertino
CA			
Telecommunications	20081.0	202600.0	Dallas
TX			
Transportation	12890.0	400945.0	Atlanta
GA			
Wholesalers	524.2	57710.0	Houston
TX			

profitable

sector	
Aerospace & Defense	yes
Apparel	yes
Business Services	yes
Chemicals	yes
Energy	yes
Engineering & Construction	yes
Financials	yes
Food & Drug Stores	yes
Food, Beverages & Tobacco	yes
Health Care	yes
Hotels, Restaurants & Leisure	yes
Household Products	yes
Industrials	no
Materials	yes
Media	yes
Motor Vehicles & Parts	yes
Retailing	yes
Technology	yes
Telecommunications	yes
Transportation	yes
Wholesalers	yes

sectors.last() *#gives the worst ranked company of each sector*

	company	rank	revenue
\			
sector			
Aerospace & Defense	BWX Technologies	995	2124.1
Apparel	Crocs	954	2313.4
Business Services	Harsco	991	2147.0
Chemicals	Ashland	999	2111.0
Energy	Genesis Energy	994	2125.5

Engineering & Construction	Beazer Homes USA	992	2140.3
Financials	Cowen	998	2112.8
Food & Drug Stores	Grocery Outlet Holding	816	3079.6
Food, Beverages & Tobacco	Brown-Forman	759	3461.0
Health Care	Chemed	993	2139.3
Hotels, Restaurants & Leisure	Acushnet Holdings	990	2147.9
Household Products	HNI	983	2184.4
Industrials	Vizio Holding	996	2124.0
Materials	Century Aluminum	979	2212.5
Media	E.W. Scripps	962	2283.5
Motor Vehicles & Parts	Cooper-Standard Holdings	952	2330.2
Retailing	1-800-Flowers.com	997	2122.2
Technology	DocuSign	1000	2107.2
Telecommunications	Telephone & Data Systems	573	5329.0
Transportation	Kirby	971	2246.7
Wholesalers	Kar Auction Services	969	2251.6

city state \ sector	profit	num. of employees	
Aerospace & Defense Lynchburg VA	305.9	6600.0	
Apparel Broomfield CO	725.7	5770.0	
Business Services Hill PA	-3.2	12000.0	Camp
Chemicals Wilmington DE	220.0	4100.0	
Energy Houston TX	-165.1	1898.0	
Engineering & Construction Atlanta GA	122.0	1052.0	
Financials York NY	295.6	1534.0	New

Food & Drug Stores	62.3	875.0	
Emeryville CA			
Food, Beverages & Tobacco	903.0	4700.0	
Louisville KY			
Health Care	268.6	14137.0	
Cincinnati OH			
Hotels, Restaurants & Leisure	178.9	6500.0	
Fairhaven MA			
Household Products	59.8	8000.0	
Muscatine IA			
Industrials	-39.4	800.0	
Irvine CA			
Materials	-167.1	2512.0	
Chicago IL			
Media	122.7	5600.0	
Cincinnati OH			
Motor Vehicles & Parts	-322.8	21300.0	
Northville MI			
Retailing	118.7	4800.0	
Jericho NY			
Technology	-70.0	7461.0	San
Francisco CA			
Telecommunications	156.0	8800.0	
Chicago IL			
Transportation	-247.0	5125.0	
Houston TX			
Wholesalers	66.5	8928.0	
Carmel IN			

#### profitable

sector	
Aerospace & Defense	yes
Apparel	yes
Business Services	no
Chemicals	yes
Energy	no
Engineering & Construction	yes
Financials	yes
Food & Drug Stores	yes
Food, Beverages & Tobacco	yes
Health Care	yes
Hotels, Restaurants & Leisure	yes
Household Products	yes
Industrials	no
Materials	no
Media	yes
Motor Vehicles & Parts	no
Retailing	yes
Technology	no

Telecommunications	yes
Transportation	no
Wholesalers	yes

sectors.groups

```
{'Aerospace & Defense': [54, 57, 59, 93, 100, 205, 301, 369, 611, 623,
639, 697, 732, 861, 971, 980, 994], 'Apparel': [82, 379, 384, 457,
505, 539, 540, 545, 657, 753, 812, 884, 917, 922, 933, 953], 'Business
Services': [142, 146, 166, 189, 202, 207, 226, 241, 254, 267, 313,
328, 352, 362, 406, 416, 469, 493, 509, 511, 573, 598, 607, 610, 616,
617, 648, 651, 662, 670, 674, 676, 677, 685, 714, 739, 803, 808, 828,
830, 852, 867, 881, 901, 915, 929, 965, 972, 974, 984, 986, 990],
'Chemicals': [63, 101, 174, 212, 217, 292, 304, 319, 321, 348, 349,
392, 405, 409, 491, 501, 517, 602, 615, 621, 765, 779, 786, 836, 940,
946, 948, 949, 998], 'Energy': [5, 15, 18, 28, 29, 53, 76, 87, 88, 98,
110, 129, 132, 134, 144, 152, 156, 167, 169, 192, 196, 214, 215, 218,
222, 223, 232, 238, 240, 242, 247, 256, 275, 277, 287, 307, 308, 314,
320, 332, 342, 343, 345, 346, 357, 361, 399, 413, 415, 430, 446, 450,
454, 468, 478, 484, 485, 499, 534, 543, 546, 557, 565, 568, 580, 618,
633, 638, 643, 701, 713, 715, 727, 730, 734, 750, 761, 767, 775, 780,
788, 806, 818, 820, 845, 848, 857, 858, 862, 869, 900, 904, 923, 941,
957, 964, 973, 979, 988, 993], 'Engineering & Construction': [123,
130, 258, 259, 261, 266, 284, 312, 356, 388, 396, 428, 452, 459, 542,
579, 587, 637, 672, 694, 700, 721, 748, 751, 754, 798, 811, 819, 822,
878, 921, 991], 'Financials': [6, 23, 32, 35, 40, 41, 43, 49, 50, 55,
56, 60, 65, 66, 71, 77, 78, 79, 84, 86, 89, 95, 96, 99, 102, 107, 125,
137, 149, 154, 158, 159, 160, 176, 177, 183, 184, 186, 187, 221, 228,
235, 237, 245, 246, 250, 255, 263, 265, 276, 280, 281, 288, 302, 315,
316, 317, 323, 324, 327, 335, 338, 344, 350, 355, 366, 367, 370, 371,
374, 375, 382, 383, 385, 386, 394, 397, 410, 414, 417, 421, 426, 433,
436, 441, 442, 448, 453, 463, 467, 472, 486, 488, 492, 502, 518, 520,
523, 531, 536, ...], 'Food & Drug Stores': [17, 20, 51, 75, 147, 519,
608, 671, 815], 'Food, Beverages & Tobacco': [37, 44, 80, 92, 94, 109,
120, 138, 164, 200, 231, 236, 257, 294, 295, 326, 330, 351, 381, 387,
402, 408, 425, 475, 504, 510, 559, 563, 622, 630, 656, 663, 729, 759],
'Health Care': [3, 4, 8, 9, 11, 14, 19, 25, 36, 39, 42, 61, 62, 70,
81, 85, 117, 121, 124, 128, 139, 172, 180, 194, 203, 213, 229, 230,
268, 291, 296, 298, 303, 318, 322, 337, 341, 360, 437, 439, 447, 512,
528, 538, 544, 550, 581, 583, 591, 599, 627, 642, 667, 687, 691, 693,
698, 744, 762, 797, 801, 814, 826, 834, 850, 851, 868, 872, 890, 903,
925, 928, 943, 975, 976, 977, 992], 'Hotels, Restaurants & Leisure':
[119, 151, 269, 334, 358, 363, 449, 464, 489, 529, 537, 566, 661, 680,
688, 704, 718, 756, 770, 778, 781, 809, 810, 823, 847, 870, 950, 989],
'Household Products': [46, 181, 210, 211, 227, 329, 347, 412, 443,
458, 495, 553, 569, 585, 597, 614, 784, 825, 829, 896, 908, 926, 982],
'Industrials': [47, 72, 83, 104, 148, 161, 168, 198, 251, 252, 253,
262, 333, 427, 432, 438, 471, 549, 552, 578, 584, 606, 634, 640, 675,
678, 696, 705, 707, 712, 717, 722, 743, 745, 758, 766, 787, 792, 817,
837, 846, 859, 860, 885, 886, 909, 920, 938, 945, 995], 'Materials':
```

```
[97, 163, 170, 171, 175, 191, 195, 260, 270, 271, 285, 311, 353, 400,
407, 411, 440, 451, 466, 483, 500, 533, 547, 548, 555, 558, 560, 561,
564, 567, 603, 625, 644, 679, 684, 731, 749, 795, 805, 874, 889, 905,
918, 936, 954, 978], 'Media': [52, 114, 115, 286, 309, 325, 373, 395,
507, 516, 551, 575, 595, 612, 636, 741, 769, 800, 816, 831, 841, 865,
873, 888, 911, 919, 934, 961], 'Motor Vehicles & Parts': [21, 24, 64,
150, 185, 201, 208, 243, 305, 390, 420, 586, 652, 709, 735, 736, 883,
939, 942, 951], 'Retailing': [0, 1, 10, 16, 31, 34, 67, 74, 105, 136,
140, 141, 143, 157, 173, 182, 188, 219, 220, 239, 244, 248, 264, 272,
274, 278, 289, 293, 299, 306, 336, 359, 376, 377, 378, 380, 389, 393,
401, 403, 419, 434, 444, 460, 461, 474, 479, 480, 487, 497, 508, 514,
525, 535, 589, 604, 641, 645, 658, 706, 719, 724, 738, 763, 773, 802,
804, 863, 880, 882, 897, 910, 930, 932, 981, 985, 996], 'Technology':
[2, 7, 13, 26, 30, 45, 48, 58, 73, 90, 91, 106, 118, 122, 126, 127,
133, 135, 155, 165, 193, 197, 206, 209, 216, 225, 234, 249, 273, 300,
339, 340, 365, 372, 404, 422, 423, 424, 435, 455, 456, 462, 465, 470,
473, 481, 482, 494, 496, 503, 515, 521, 527, 530, 541, 556, 562, 570,
588, 594, 596, 600, 613, 620, 631, 632, 650, 653, 660, 665, 666, 673,
682, 683, 690, 703, 708, 711, 720, 723, 725, 728, 733, 737, 742, 757,
771, 777, 790, 796, 799, 807, 813, 827, 832, 833, 835, 838, 842,
843, ...], 'Telecommunications': [12, 22, 27, 68, 178, 204, 354, 498,
572], 'Transportation': [33, 38, 112, 113, 145, 153, 162, 224, 233,
290, 297, 310, 331, 364, 418, 429, 490, 513, 522, 526, 532, 554, 571,
577, 590, 647, 668, 686, 695, 702, 794, 891, 893, 960, 970],
'Wholesalers': [69, 103, 108, 111, 116, 131, 179, 190, 199, 279, 282,
283, 368, 391, 398, 431, 445, 476, 477, 506, 524, 576, 605, 664, 755,
772, 785, 789, 791, 854, 968]}
```

*#fetching a company by rank*

```
company.iloc[23, : ]
```

```
company      JPMorgan Chase
rank          24
revenue      127202.0
profit       48334.0
num. of employees  271025.0
sector       Financials
city         New York
state        NY
profitable    yes
Name: 23, dtype: object
```

```
sectors.get_group('Financials')
```

	company	rank	revenue	profit	num. of employees
6	Berkshire Hathaway	7	276094.0	89795.0	372000.0
23	JPMorgan Chase	24	127202.0	48334.0	271025.0

32	Fannie Mae	33	101543.0	22176.0	7400.0
35	Bank of America	36	93851.0	31978.0	208248.0
40	Wells Fargo	41	82407.0	21548.0	247848.0
..	...	...	...	...	...
955	Signature Bank	956	2311.3	918.4	1854.0
956	SBA Communications	957	2308.8	237.6	1596.0
959	AvalonBay Communities	960	2294.9	1004.3	2883.0
963	Sun Communities	964	2272.6	380.2	5455.0
997	Cowen	998	2112.8	295.6	1534.0

	sector	city	state	profitable
6	Financials	Omaha	NE	yes
23	Financials	New York	NY	yes
32	Financials	Washington	DC	yes
35	Financials	Charlotte	NC	yes
40	Financials	San Francisco	CA	yes
..	...	...	...	...
955	Financials	New York	NY	yes
956	Financials	Boca Raton	FL	yes
959	Financials	Arlington	VA	yes
963	Financials	Southfield	MI	yes
997	Financials	New York	NY	yes

[166 rows x 9 columns]

using sum() and mean() function on the sectors which we made using groupby function on company (dataframe)

```
sectors.sum(numeric_only=True)
#to get only numeric columns use numeric_only
```

	rank	revenue	profit	num. of
employees				
sector				
Aerospace & Defense	8363	338982.0	21097.0	
857200.0				

Apparel 313245.0	10278	117313.1	11799.3
Business Services 1521413.0	31152	383672.9	54894.2
Chemicals 519363.0	15229	297776.6	39577.5
Energy 876073.0	47231	1884551.1	139828.6
Engineering & Construction 468877.0	17829	253801.0	20963.0
Financials 3695473.0	78191	2910251.7	556635.2
Food & Drug Stores 1334835.0	2932	446986.5	10033.5
Food, Beverages & Tobacco 1122281.0	12104	617540.7	62982.9
Health Care 3384082.0	34412	2688237.8	209442.4
Hotels, Restaurants & Leisure 2122533.0	17290	188472.8	18927.8
Household Products 599696.0	12450	243560.9	30236.1
Industrials 1456299.0	29193	507474.6	43100.4
Materials 756134.0	24277	398430.4	38176.3
Media 461661.0	17138	256759.8	19439.1
Motor Vehicles & Parts 1046698.0	9364	470407.9	39308.0
Retailing 7868015.0	34051	2359968.2	126835.7
Technology 4353295.0	66945	2184129.1	459503.3
Telecommunications 687640.0	1944	529942.9	66548.9
Transportation 1841312.0	16468	503219.3	39568.5
Wholesalers 466755.0	13656	405322.1	11499.2

sectors.mean(numeric\_only=True)

	rank	revenue	
profit \ sector			
Aerospace & Defense	491.941176	19940.117647	1241.000000
Apparel	642.375000	7332.068750	737.456250



Business Services	599.076923	7378.325000	1055.657692
Chemicals	525.137931	10268.158621	1364.741379
Energy	472.310000	18845.511000	1412.410101
Engineering & Construction	557.156250	7931.281250	655.093750
Financials	471.030120	17531.636747	3353.224096
Food & Drug Stores	325.777778	49665.166667	1114.833333
Food, Beverages & Tobacco	356.000000	18162.961765	1852.438235
Health Care	446.909091	34912.179221	2720.031169
Hotels, Restaurants & Leisure	617.500000	6731.171429	675.992857
Household Products	541.304348	10589.604348	1314.613043
Industrials	583.860000	10149.492000	862.008000
Materials	527.760870	8661.530435	829.919565
Media	612.071429	9169.992857	694.253571
Motor Vehicles & Parts	468.200000	23520.395000	2068.842105
Retailing	442.220779	30648.937662	1647.216883
Technology	553.264463	18050.653719	3797.547934
Telecommunications	216.000000	58882.544444	8318.612500
Transportation	470.514286	14377.694286	1130.528571
Wholesalers	440.516129	13074.906452	370.941935
num. of employees			
sector			
Aerospace & Defense	50423.529412		
Apparel	19577.812500		
Business Services	29257.942308		
Chemicals	17909.068966		
Energy	8760.730000		
Engineering & Construction	14652.406250		
Financials	22261.885542		
Food & Drug Stores	148315.000000		
Food, Beverages & Tobacco	33008.264706		

Health Care	43949.116883
Hotels, Restaurants & Leisure	75804.750000
Household Products	26073.739130
Industrials	29125.980000
Materials	16437.695652
Media	17098.555556
Motor Vehicles & Parts	52334.900000
Retailing	102182.012987
Technology	35977.644628
Telecommunications	76404.444444
Transportation	52608.914286
Wholesalers	15056.612903

```
sectors['revenue'].mean().sort_values(ascending=False)
```

sector	
Telecommunications	58882.544444
Food & Drug Stores	49665.166667
Health Care	34912.179221
Retailing	30648.937662
Motor Vehicles & Parts	23520.395000
Aerospace & Defense	19940.117647
Energy	18845.511000
Food, Beverages & Tobacco	18162.961765
Technology	18050.653719
Financials	17531.636747
Transportation	14377.694286
Wholesalers	13074.906452
Household Products	10589.604348
Chemicals	10268.158621
Industrials	10149.492000
Media	9169.992857
Materials	8661.530435
Engineering & Construction	7931.281250
Business Services	7378.325000
Apparel	7332.068750
Hotels, Restaurants & Leisure	6731.171429
Name: revenue, dtype: float64	

```
sectors['profit'].mean().sort_values(ascending=False)
```

sector	
Telecommunications	8318.612500
Technology	3797.547934
Financials	3353.224096
Health Care	2720.031169
Motor Vehicles & Parts	2068.842105
Food, Beverages & Tobacco	1852.438235
Retailing	1647.216883
Energy	1412.410101

Chemicals	1364.741379
Household Products	1314.613043
Aerospace & Defense	1241.000000
Transportation	1130.528571
Food & Drug Stores	1114.833333
Business Services	1055.657692
Industrials	862.008000
Materials	829.919565
Apparel	737.456250
Media	694.253571
Hotels, Restaurants & Leisure	675.992857
Engineering & Construction	655.093750
Wholesalers	370.941935

Name: profit, dtype: float64

This is a new dataset which tell about every ball played in Cricket match.

```
deliveries = pd.read_csv('deliveries.csv')
```

```
deliveries.shape
```

```
(150460, 21)
```

```
deliveries.head()
```

	match_id	inning	batting_team		bowling_team		
over \							
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore			
1							
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore			
1							
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore			
1							
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore			
1							
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore			
1							

	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs
\							
0	1	DA Warner	S Dhawan	TS Mills	0	...	0
1	2	DA Warner	S Dhawan	TS Mills	0	...	0
2	3	DA Warner	S Dhawan	TS Mills	0	...	0
3	4	DA Warner	S Dhawan	TS Mills	0	...	0

4	5	DA Warner	S Dhawan	TS Mills	0	...	0
---	---	-----------	----------	----------	---	-----	---

	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	\
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	4	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	2

	total_runs	player_dismissed	dismissal_kind	fielder
0	0	NaN	NaN	NaN
1	0	NaN	NaN	NaN
2	4	NaN	NaN	NaN
3	0	NaN	NaN	NaN
4	2	NaN	NaN	NaN

[5 rows x 21 columns]

```
batsman = deliveries.groupby('batsman')
batsman
```

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x00000282C345A9D0>

```
batsman.sum('batsman').sort_values('batsman_runs',ascending= False)
```

	match_id	inning	over	ball	is_super_over	wide_runs
\						
batsman						
SK Raina	1079285	4748	32746	12256	3	124
V Kohli	1304291	5112	33426	12284	2	110
RG Sharma	1081920	4822	36662	11601	0	66
G Gambhir	1072760	5510	23650	12313	0	151
DA Warner	1085379	4258	21153	10188	1	108
...	...	...	...	...	...	...
Abdur Razzak	202	4	38	11	0	0
S Kaushik	620	2	19	4	0	0
S Ladda	1836	20	155	47	0	1
U Kaul	91	2	20	6	0	0

RR Bhatkal	640	2	40	9	0	0
------------	-----	---	----	---	---	---

	bye_runs	legbye_runs	noball_runs	penalty_runs
batsman_runs \				
batsman				

SK Raina	0	57	16	0
4548				

V Kohli	5	39	11	0
4423				

RG Sharma	2	48	11	0
4207				

G Gambhir	10	85	10	0
4132				

DA Warner	23	52	16	0
4014				

...	...	...	...	...
...				

Abdur Razzak	0	0	0	0
0				

S Kaushik	0	0	0	0
0				

S Ladda	0	0	0	0
0				

U Kaul	0	0	0	0
0				

RR Bhatkal	0	0	0	0
0				

	extra_runs	total_runs
batsman		

SK Raina	197	4745
----------	-----	------

V Kohli	165	4588
---------	-----	------

RG Sharma	127	4334
-----------	-----	------

G Gambhir	256	4388
-----------	-----	------

DA Warner	199	4213
-----------	-----	------

...	...	...
-----	-----	-----

Abdur Razzak	0	0
--------------	---	---

S Kaushik	0	0
-----------	---	---

S Ladda	1	1
---------	---	---

U Kaul	0	0
--------	---	---

RR Bhatkal	0	0
------------	---	---

[461 rows x 13 columns]

```
batsman['batsman_runs'].sum().sort_values(ascending = False).head(10)
# this is the answer for the top batsmen
```

```

batsman
SK Raina      4548
V Kohli       4423
RG Sharma     4207
G Gambhir     4132
DA Warner     4014
RV Uthappa    3778
CH Gayle      3651
S Dhawan      3561
MS Dhoni      3560
AB de Villiers 3486
Name: batsman_runs, dtype: int64

```

```

mask = deliveries['batsman_runs']==4
new_deliveries = deliveries[mask]
new_deliveries

```

	match_id	inning		batting_team	\
2	1	1		Sunrisers Hyderabad	
8	1	1		Sunrisers Hyderabad	
13	1	1		Sunrisers Hyderabad	
30	1	1		Sunrisers Hyderabad	
32	1	1		Sunrisers Hyderabad	
...	...	...		...	
150403	636	2	Royal Challengers	Bangalore	
150407	636	2	Royal Challengers	Bangalore	
150418	636	2	Royal Challengers	Bangalore	
150447	636	2	Royal Challengers	Bangalore	
150459	636	2	Royal Challengers	Bangalore	

		bowling_team	over	ball	batsman	\
2	Royal Challengers	Bangalore	1	3	DA Warner	
8	Royal Challengers	Bangalore	2	2	DA Warner	
13	Royal Challengers	Bangalore	2	7	MC Henriques	
30	Royal Challengers	Bangalore	5	5	MC Henriques	
32	Royal Challengers	Bangalore	6	1	S Dhawan	
...	...	...	...	...	...	
150403	Sunrisers	Hyderabad	11	4	V Kohli	
150407	Sunrisers	Hyderabad	12	2	V Kohli	
150418	Sunrisers	Hyderabad	14	1	KL Rahul	
150447	Sunrisers	Hyderabad	18	6	Sachin Baby	
150459	Sunrisers	Hyderabad	20	6	Iqbal Abdulla	

	non_striker	bowler	is_super_over	...	bye_runs	\
2	S Dhawan	TS Mills	0	...	0	
8	S Dhawan	A Choudhary	0	...	0	
13	S Dhawan	A Choudhary	0	...	0	
30	S Dhawan	S Aravind	0	...	0	
32	MC Henriques	SR Watson	0	...	0	
...	...	...	...	...	...	

150403	AB de Villiers	BCJ Cutting	0	...	0
150407	AB de Villiers	Bipul Sharma	0	...	0
150418	AB de Villiers	Bipul Sharma	0	...	0
150447	STR Binny	B Kumar	0	...	0
150459	Sachin Baby	B Kumar	0	...	0

	legbye_runs	noball_runs	penalty_runs	batsman_runs
extra_runs \				
2	0	0	0	4
0				
8	0	0	0	4
0				
13	0	0	0	4
0				
30	0	0	0	4
0				
32	0	0	0	4
0				
...	...	...	...	...
..				
150403	0	0	0	4
0				
150407	0	0	0	4
0				
150418	0	0	0	4
0				
150447	0	0	0	4
0				
150459	0	0	0	4
0				

	total_runs	player_dismissed	dismissal_kind	fielder
2	4	NaN	NaN	NaN
8	4	NaN	NaN	NaN
13	4	NaN	NaN	NaN
30	4	NaN	NaN	NaN
32	4	NaN	NaN	NaN
...	...	...	...	...
150403	4	NaN	NaN	NaN
150407	4	NaN	NaN	NaN
150418	4	NaN	NaN	NaN
150447	4	NaN	NaN	NaN
150459	4	NaN	NaN	NaN

[17033 rows x 21 columns]

```
highest_boundries = new_deliveries.groupby('batsman')
```

```
highest_boundries['batsman_runs'].count().sort_values(ascending=False).head(5)
```

```

batsman
G Gambhir      484
SK Raina       402
DA Warner       401
S Dhawan       401
V Kohli        384
Name: batsman_runs, dtype: int64

#we can do for 3 runs also
mask = deliveries['batsman_runs']==3
new_deliveries = deliveries[mask]
new_deliveries.groupby('batsman')
['batsman_runs'].count().sort_values(ascending=False).head(5)

batsman
M Vijay        17
S Dhawan       16
G Gambhir      15
AM Rahane      14
DA Warner      13
Name: batsman_runs, dtype: int64

vk = deliveries[deliveries['batsman']=='V Kohli']

vk.groupby('bowling_team')
['batsman_runs'].sum().sort_values(ascending=False).head(3)

bowling_team
Chennai Super Kings    706
Delhi Daredevils       661
Kings XI Punjab        483
Name: batsman_runs, dtype: int64

```

we can create a function that returns name of team against which a particular batsman scored the most

```

def highest_runs_scored(batsman_name):
    vk = deliveries[deliveries['batsman']==batsman_name]
    return vk.groupby('bowling_team')
    ['batsman_runs'].sum().sort_values(ascending=False).head(1).index[0]

highest_runs_scored('DA Warner')

'Kolkata Knight Riders'

```



```
highest_runs_scored('G Gambhir')
'Kings XI Punjab'
highest_runs_scored('MS Dhoni')
'Royal Challengers Bangalore'
```

## isin() Function

*#getting the death overs*

```
death_over = deliveries[deliveries['over']>15]
death_over.head()
```

	match_id	inning	batting_team		bowling_team	
over \						
93	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore		
16						
94	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore		
16						
95	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore		
16						
96	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore		
16						
97	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore		
16						

	ball	batsman	non_striker	bowler	is_super_over	...
bye_runs \						
93	1	MC Henriques	Yuvraj Singh	YS Chahal	0	...
0						
94	2	MC Henriques	Yuvraj Singh	YS Chahal	0	...
0						
95	3	Yuvraj Singh	DJ Hooda	YS Chahal	0	...
0						
96	4	DJ Hooda	Yuvraj Singh	YS Chahal	0	...
0						
97	5	Yuvraj Singh	DJ Hooda	YS Chahal	0	...
0						

	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	
93	0	0	0	0	0	
94	0	0	0	0	0	
95	0	0	0	1	0	

96	0	0	0	1	0
97	0	0	0	1	0

	total_runs	player_dismissed	dismissal_kind	fielder
93	0	NaN	NaN	NaN
94	0	MC Henriques	caught	Sachin Baby
95	1	NaN	NaN	NaN
96	1	NaN	NaN	NaN
97	1	NaN	NaN	NaN

[5 rows x 21 columns]

*# finding the number of balls played by each batsman who are also present in the deathover*

```
req_batsman = death_over.groupby('batsman')['batsman_runs'].count()
req_batsman
```

*# now we do masking for batsman playing 200+ balls*

```
mask = req_batsman > 200
req_batsman[mask]
```

```
batsman
A Mishra      225
AB de Villiers 570
AD Mathews    289
AM Rahane     268
AR Patel      229
AT Rayudu     425
BJ Hodge      385
DA Miller     360
DA Warner     228
DJ Bravo      409
DJ Hussey     234
DPMD Jayawardene 246
Harbhajan Singh 418
IK Pathan     465
JA Morkel     425
JH Kallis     231
JP Duminy     518
JP Faulkner   294
KA Pollard    838
KD Karthik    463
KM Jadhav     338
LRPL Taylor   204
MK Pandey     224
MK Tiwary     423
MS Dhoni     1224
NV Ojha       304
P Kumar       268
```

PP Chawla	311
R Vinay Kumar	235
RA Jadeja	576
RG Sharma	748
RV Uthappa	275
S Badrinath	283
S Dhawan	243
SK Raina	458
SPD Smith	316
SS Tiwary	300
STR Binny	218
V Kohli	546
WP Saha	295
Y Venugopal Rao	260
YK Pathan	584
Yuvraj Singh	516

Name: batsman\_runs, dtype: int64

*# now we extract all batsman names*

req\_batsman[mask].index *#this gives us a list*

```
Index(['A Mishra', 'AB de Villiers', 'AD Mathews', 'AM Rahane', 'AR
Patel',
      'AT Rayudu', 'BJ Hodge', 'DA Miller', 'DA Warner', 'DJ Bravo',
      'DJ Hussey', 'DPMD Jayawardene', 'Harbhajan Singh', 'IK
Pathan',
      'JA Morkel', 'JH Kallis', 'JP Duminy', 'JP Faulkner', 'KA
Pollard',
      'KD Karthik', 'KM Jadhav', 'LRPL Taylor', 'MK Pandey', 'MK
Tiwary',
      'MS Dhoni', 'NV Ojha', 'P Kumar', 'PP Chawla', 'R Vinay Kumar',
      'RA Jadeja', 'RG Sharma', 'RV Uthappa', 'S Badrinath', 'S
Dhawan',
      'SK Raina', 'SPD Smith', 'SS Tiwary', 'STR Binny', 'V Kohli',
      'WP Saha',
      'Y Venugopal Rao', 'YK Pathan', 'Yuvraj Singh'],
      dtype='object', name='batsman')
```

*#we can make a list in proper format using tolist()*

```
req_list=req_batsman[mask].index.tolist()
req_list
```

```
['A Mishra',
 'AB de Villiers',
 'AD Mathews',
 'AM Rahane',
 'AR Patel',
 'AT Rayudu',
 'BJ Hodge',
 'DA Miller',
```

```

'DA Warner',
'DJ Bravo',
'DJ Hussey',
'DPMD Jayawardene',
'Harbhajan Singh',
'IK Pathan',
'JA Morkel',
'JH Kallis',
'JP Duminy',
'JP Faulkner',
'KA Pollard',
'KD Karthik',
'KM Jadhav',
'LRPL Taylor',
'MK Pandey',
'MK Tiwary',
'MS Dhoni',
'NV Ojha',
'P Kumar',
'PP Chawla',
'R Vinay Kumar',
'RA Jadeja',
'RG Sharma',
'RV Uthappa',
'S Badrinath',
'S Dhawan',
'SK Raina',
'SPD Smith',
'SS Tiwary',
'STR Binny',
'V Kohli',
'WP Saha',
'Y Venugopal Rao',
'YK Pathan',
'Yuvraj Singh']

```

```

# now for the strike rate we req runs scored by these batsmen and
# number of balls played by them
# we can use deliveries data to find runs scored but it also has other
# batsmen data
# here we can use isin() function

```

```

final = deliveries[deliveries['batsman'].isin(req_list)]
final

```

	match_id	inning	batting_team \
0	1	1	Sunrisers Hyderabad
1	1	1	Sunrisers Hyderabad
2	1	1	Sunrisers Hyderabad

3	1	1	Sunrisers	Hyderabad
4	1	1	Sunrisers	Hyderabad
...	...	...		...
150442	636	2	Royal Challengers	Bangalore
150443	636	2	Royal Challengers	Bangalore
150445	636	2	Royal Challengers	Bangalore
150446	636	2	Royal Challengers	Bangalore
150448	636	2	Royal Challengers	Bangalore
bowling_team over ball batsman				
non_striker \				
0	Royal Challengers	Bangalore	1	1 DA Warner S
Dhawan				
1	Royal Challengers	Bangalore	1	2 DA Warner S
Dhawan				
2	Royal Challengers	Bangalore	1	3 DA Warner S
Dhawan				
3	Royal Challengers	Bangalore	1	4 DA Warner S
Dhawan				
4	Royal Challengers	Bangalore	1	5 DA Warner S
Dhawan				
...		...	...	...
.				
150442	Sunrisers	Hyderabad	18	1 STR Binny Sachin
Baby				
150443	Sunrisers	Hyderabad	18	2 STR Binny Sachin
Baby				
150445	Sunrisers	Hyderabad	18	4 STR Binny Sachin
Baby				
150446	Sunrisers	Hyderabad	18	5 STR Binny Sachin
Baby				
150448	Sunrisers	Hyderabad	19	1 STR Binny Sachin
Baby				
bowler is_super_over ... bye_runs				
legbye_runs \				
0	TS Mills		0	... 0 0
1	TS Mills		0	... 0 0
2	TS Mills		0	... 0 0
3	TS Mills		0	... 0 0
4	TS Mills		0	... 0 0
...	...		...	... 0 0
150442	B Kumar		0	... 0 0

150443	B Kumar	0	...	0	0
150445	B Kumar	0	...	0	0
150446	B Kumar	0	...	0	1
150448	Mustafizur Rahman	0	...	0	0

	noball_runs	penalty_runs	batsman_runs	extra_runs
total_runs \				
0	0	0	0	0
0				
1	0	0	0	0
0				
2	0	0	4	0
4				
3	0	0	0	0
0				
4	0	0	0	2
2				
...	...	...	...	...
.				
150442	0	0	0	0
0				
150443	0	0	1	0
1				
150445	0	0	0	0
0				
150446	0	0	0	1
1				
150448	0	0	1	0
1				

	player_dismissed	dismissal_kind	fielder
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
...	...	...	...
150442	NaN	NaN	NaN
150443	NaN	NaN	NaN
150445	NaN	NaN	NaN
150446	NaN	NaN	NaN
150448	STR Binny	run out	DJ Hooda

[66006 rows x 21 columns]

```
runs = final.groupby('batsman')['batsman_runs'].sum()
```

```
balls = final.groupby('batsman')['batsman_runs'].count()
strike_rate = (runs/balls)*100
```

hence we have the deadliest batsmen in IPL

```
strike_rate.sort_values(ascending = False).head()
```

```
batsman
AB de Villiers    145.129059
YK Pathan         140.751445
KA Pollard        140.621266
DA Warner         138.318401
DA Miller         137.709251
Name: batsman_runs, dtype: float64
```

## merge() opeartions

Merge in pandas combines two datasets based on a common column, aligning rows with matching values and creating a new DataFrame. The merge function allows you to perform SQL-like joins, such as inner, outer, left, and right joins, providing flexibility in combining data from different sources.

```
-----
-----
NameError                                Traceback (most recent call
last)
Cell In[133], line 1
----> 1 match

NameError: name 'match' is not defined

match = pd.read_csv("IPL_data.csv")
match.head()
```

	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



## deliveries

	match_id	inning	batting_team \	
0	1	1	Sunrisers	Hyderabad
1	1	1	Sunrisers	Hyderabad
2	1	1	Sunrisers	Hyderabad
3	1	1	Sunrisers	Hyderabad
4	1	1	Sunrisers	Hyderabad
...	...	...	...	...
150455	636	2	Royal Challengers	Bangalore
150456	636	2	Royal Challengers	Bangalore
150457	636	2	Royal Challengers	Bangalore
150458	636	2	Royal Challengers	Bangalore
150459	636	2	Royal Challengers	Bangalore

	bowling_team		over	ball	batsman	
non_striker \						
0	Royal Challengers	Bangalore	1	1	DA Warner	S
Dhawan						
1	Royal Challengers	Bangalore	1	2	DA Warner	S
Dhawan						
2	Royal Challengers	Bangalore	1	3	DA Warner	S
Dhawan						
3	Royal Challengers	Bangalore	1	4	DA Warner	S
Dhawan						
4	Royal Challengers	Bangalore	1	5	DA Warner	S
Dhawan						
...	...	...	...	...	...	...
...						
150455	Sunrisers	Hyderabad	20	2	Sachin Baby	CJ
Jordan						
150456	Sunrisers	Hyderabad	20	3	Sachin Baby	CJ
Jordan						
150457	Sunrisers	Hyderabad	20	4	Iqbal Abdulla	
Sachin Baby						
150458	Sunrisers	Hyderabad	20	5	Sachin Baby	Iqbal
Abdulla						
150459	Sunrisers	Hyderabad	20	6	Iqbal Abdulla	
Sachin Baby						

	bowler	is_super_over	...	bye_runs	legbye_runs
noball_runs \					
0	TS Mills	0	...	0	0
0					
1	TS Mills	0	...	0	0
0					
2	TS Mills	0	...	0	0
0					
3	TS Mills	0	...	0	0
0					

4	TS Mills	0	...	0	0	
0						
...	...	...	...	...	...	.
..						
150455	B Kumar	0	...	0	0	
0						
150456	B Kumar	0	...	0	0	
0						
150457	B Kumar	0	...	0	1	
0						
150458	B Kumar	0	...	0	0	
0						
150459	B Kumar	0	...	0	0	
0						
	penalty_runs	batsman_runs	extra_runs	total_runs		
player_dismissed \						
0	0	0	0	0		
NaN						
1	0	0	0	0		
NaN						
2	0	4	0	4		
NaN						
3	0	0	0	0		
NaN						
4	0	0	2	2		
NaN						
...	...	...	...	...		
...						
150455	0	2	0	2		
NaN						
150456	0	0	0	0	CJ	
Jordan						
150457	0	0	1	1		
NaN						
150458	0	1	0	1		
NaN						
150459	0	4	0	4		
NaN						
	dismissal_kind	fielder				
0	NaN	NaN				
1	NaN	NaN				
2	NaN	NaN				
3	NaN	NaN				
4	NaN	NaN				
...	...	...				
150455	NaN	NaN				
150456	run out	NV Ojha				

```

150457      NaN      NaN
150458      NaN      NaN
150459      NaN      NaN

```

```
[150460 rows x 21 columns]
```

```
deliveries.corr(numeric_only = True)
```

	match_id	inning	over	ball	
is_super_over	\				
match_id	1.000000	0.001033	0.004365	-0.002464	-0.001220
inning	0.001033	1.000000	-0.051496	-0.003618	0.091904
over	0.004365	-0.051496	1.000000	-0.010953	-0.037394
ball	-0.002464	-0.003618	-0.010953	1.000000	-0.001258
is_super_over	-0.001220	0.091904	-0.037394	-0.001258	1.000000
wide_runs	-0.004850	0.001055	-0.016226	-0.005931	-0.001155
bye_runs	-0.004662	-0.000395	0.013669	0.006755	0.001516
legbye_runs	-0.004324	-0.001385	-0.005907	-0.004449	0.001717
noball_runs	-0.005849	0.000389	0.015717	0.002350	0.014389
penalty_runs	-0.000972	0.003763	-0.001055	0.000774	-0.000085
batsman_runs	0.006273	-0.005668	0.089945	0.008944	0.011497
extra_runs	-0.008833	0.000132	-0.007649	-0.004176	0.003613
total_runs	0.004366	-0.005680	0.088896	0.008086	0.012377

	wide_runs	bye_runs	legbye_runs	noball_runs	
penalty_runs	\				
match_id	-0.004850	-0.004662	-0.004324	-0.005849	-
0.000972					
inning	0.001055	-0.000395	-0.001385	0.000389	
0.003763					
over	-0.016226	0.013669	-0.005907	0.015717	-
0.001055					
ball	-0.005931	0.006755	-0.004449	0.002350	
0.000774					
is_super_over	-0.001155	0.001516	0.001717	0.014389	-
0.000085					
wide_runs	1.000000	-0.006230	-0.016186	-0.008703	
0.013633					

bye_runs	-0.006230	1.000000	-0.004751	-0.002555	-
0.000156					
legbye_runs	-0.016186	-0.004751	1.000000	-0.006637	-
0.000405					
noball_runs	-0.008703	-0.002555	-0.006637	1.000000	-
0.000218					
penalty_runs	0.013633	-0.000156	-0.000405	-0.000218	
1.000000					
batsman_runs	-0.111690	-0.032785	-0.085177	0.000272	-
0.002795					
extra_runs	0.723726	0.318850	0.557403	0.196724	
0.061842					
total_runs	0.047354	0.037402	0.037322	0.043722	
0.010843					

	batsman_runs	extra_runs	total_runs
match_id	0.006273	-0.008833	0.004366
inning	-0.005668	0.000132	-0.005680
over	0.089945	-0.007649	0.088896
ball	0.008944	-0.004176	0.008086
is_super_over	0.011497	0.003613	0.012377
wide_runs	-0.111690	0.723726	0.047354
bye_runs	-0.032785	0.318850	0.037402
legbye_runs	-0.085177	0.557403	0.037322
noball_runs	0.000272	0.196724	0.043722
penalty_runs	-0.002795	0.061842	0.010843
batsman_runs	1.000000	-0.141761	0.975809
extra_runs	-0.141761	1.000000	0.078085
total_runs	0.975809	0.078085	1.000000

## dropping missing values using dropna()

```
data = pd.read_csv('titanic.csv')
```

```
data.isnull
```

```
<bound method DataFrame.isnull of      PassengerId  Survived  \
Pclass  \
0          1         0       3
1          2         1       1
2          3         1       3
3          4         1       1
4          5         0       3
..        ...        ...    ...
886        887         0       2
887        888         1       1
888        889         0       3
889        890         1       1
```

```

890      891      0      3
SibSp  \
0      Braund, Mr. Owen Harris    male  22.0
1
1      Cumings, Mrs. John Bradley (Florence Briggs Th... female  38.0
1
2      Heikkinen, Miss. Laina    female  26.0
0
3      Futrelle, Mrs. Jacques Heath (Lily May Peel)    female  35.0
1
4      Allen, Mr. William Henry    male  35.0
0
..      ...      ...      ...
...
886      Montvila, Rev. Juozas    male  27.0
0
887      Graham, Miss. Margaret Edith    female  19.0
0
888      Johnston, Miss. Catherine Helen "Carrie"    female   NaN
1
889      Behr, Mr. Karl Howell    male  26.0
0
890      Dooley, Mr. Patrick    male  32.0
0

```

```

Parch      Ticket      Fare Cabin Embarked
0      0      A/5 21171    7.2500   NaN      S
1      0      PC 17599   71.2833   C85      C
2      0  STON/O2. 3101282    7.9250   NaN      S
3      0      113803   53.1000  C123      S
4      0      373450    8.0500   NaN      S
..      ...      ...      ...      ...
886      0      211536   13.0000   NaN      S
887      0      112053   30.0000   B42      S
888      2      W./C. 6607   23.4500   NaN      S
889      0      111369   30.0000  C148      C
890      0      370376    7.7500   NaN      Q

```

```
[891 rows x 12 columns]>
```

```
data.shape
```

```
(891, 12)
```

```
data.dropna().shape
```

```
(183, 12)
```