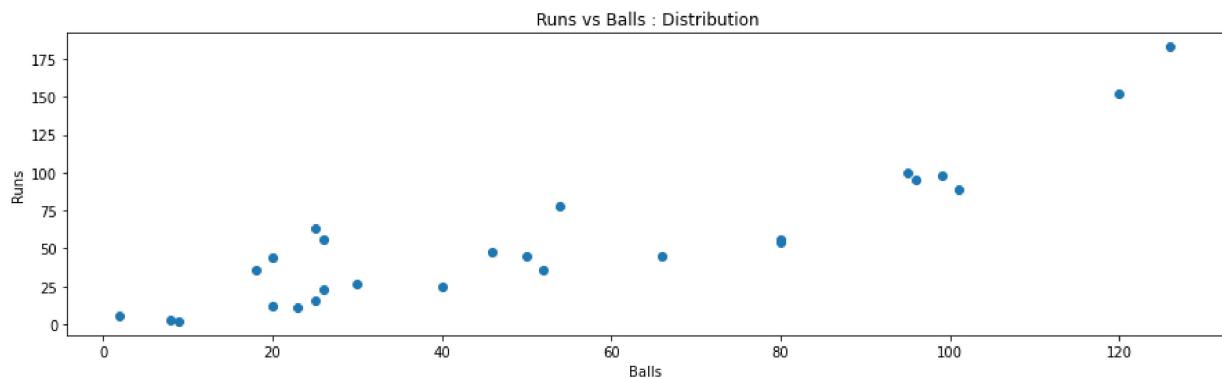


```
In [43]: plt.figure(figsize = (15,4))
plt.scatter(df.Balls, df.Runs)
plt.title(' Runs vs Balls : Distribution')
plt.xlabel("Balls")
plt.ylabel("Runs")
plt.show()
```



```
In [44]: # we have different plots like scatter, line, box-plot, bar charts, histograms
```

Sachin vs Virat : Analytical Debate

```
In [10]: import warnings
warnings.filterwarnings('ignore')

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

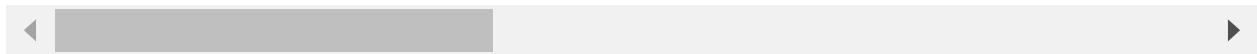
# to display all rows columns
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.expand_frame_repr', False)
pd.set_option('max_colwidth', -1)
```

```
In [11]: df= pd.read_csv('C:\\\\Users\\\\Smital Bhalerao\\\\Desktop\\\\ODI_data.csv')
```

In [12]: `df.head(2)`

Out[12]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate	In Nu
0	E Lewis	65	65	128	1.0	0.0	80	8	1	81.25	
1	N Pooran	42	42	69	1.0	0.0	52	4	1	80.76	



In [13]: `df.columns`

Out[13]:

```
Index(['Innings Player', 'Innings Runs Scored', 'Innings Runs Scored Num',
       'Innings Minutes Batted', 'Innings Batted Flag', 'Innings Not Out Flag',
       'Innings Balls Faced', 'Innings Boundary Fours',
       'Innings Boundary Sixes', 'Innings Batting Strike Rate',
       'Innings Number', 'Opposition', 'Ground', 'Innings Date', 'Country',
       '50\'s', '100\'s', 'Innings Runs Scored Buckets', 'Innings Overs Bowled',
       'Innings Bowled Flag', 'Innings Maidens Bowled',
       'Innings Runs Conceded', 'Innings Wickets Taken', '4 Wickets',
       '5 Wickets', '10 Wickets', 'Innings Wickets Taken Buckets',
       'Innings Economy Rate'],
      dtype='object')
```

In [14]: `len(df.columns)`

Out[14]: 28

In [15]: `len(df)`

Out[15]: 171968

In [16]:

```
# Runs per innings
# SR
# 100's
# 50's
# Team contribution
```

In [17]: `df['Innings Runs Scored Num'].unique()`

```
Out[17]: array(['65', '42', '18', '17', '13', '11', '5', '0', '120', '71', '20',
   '16', '3', '2', '1', '1', nan, '40', '6', '4', '87', '54', '46',
   '30', '12', '69', '39', '14', '10', '9', '8', '7', '82', '52',
   '41', '15', '98', '43', '19', '111', '48', '36', '25', '67', '60',
   '84', '59', '55', '47', '85', '49', '45', '34', '29', '22', '74',
   '28', '77', '50', '32', '23', '35', '122', '100', '95', '103',
   '113', '53', '96', '27', '64', '58', '33', '31', '73', '56', '86',
   '62', '106', '24', '57', '104', '26', '66', '51', '118', '105',
   '101', '21', '79', '44', '102', '88', '80', '72', '97', '68', '89',
   '38', '83', '63', '148', '166', '90', '76', '37', '70', '124',
   '94', '140', '153', '107', '117', '121', '92', '78', '75', '114',
   '115', '130', '128', '151', '110', '138', '135', '109', '61',
   '179', '170', '112', '116', '91', '143', '93', '123', '145', '81',
   '150', '162', '108', '131', '133', '137', '146', '139', '125',
   '129', '157', '152', '144', '99', '127', '210', '147', '126',
   '181', '160', '180', '208', '176', '168', '141', '132', '119',
   '154', '185', '134', '156', '164', '173', '178', '171', '149',
   '237', '159', '161', '215', '264', '136', '169', '209', '174',
   '189', '183', '163', '219', '158', '175', '177', '200', '194',
   '142', '172', '186', '188', '167'], dtype=object)
```

In [18]: `df = df[df['Innings Runs Scored Num'] != '-']`

In [19]: `df['Innings Runs Scored Num'].unique()`

```
Out[19]: array(['65', '42', '18', '17', '13', '11', '5', '0', '120', '71', '20',
   '16', '3', '2', '1', nan, '40', '6', '4', '87', '54', '46', '30',
   '12', '69', '39', '14', '10', '9', '8', '7', '82', '52', '41',
   '15', '98', '43', '19', '111', '48', '36', '25', '67', '60', '84',
   '59', '55', '47', '85', '49', '45', '34', '29', '22', '74', '28',
   '77', '50', '32', '23', '35', '122', '100', '95', '103', '113',
   '53', '96', '27', '64', '58', '33', '31', '73', '56', '86', '62',
   '106', '24', '57', '104', '26', '66', '51', '118', '105', '101',
   '21', '79', '44', '102', '88', '80', '72', '97', '68', '89', '38',
   '83', '63', '148', '166', '90', '76', '37', '70', '124', '94',
   '140', '153', '107', '117', '121', '92', '78', '75', '114', '115',
   '130', '128', '151', '110', '138', '135', '109', '61', '179',
   '170', '112', '116', '91', '143', '93', '123', '145', '81', '150',
   '162', '108', '131', '133', '137', '146', '139', '125', '129',
   '157', '152', '144', '99', '127', '210', '147', '126', '181',
   '160', '180', '208', '176', '168', '141', '132', '119', '154',
   '185', '134', '156', '164', '173', '178', '171', '149', '237',
   '159', '161', '215', '264', '136', '169', '209', '174', '189',
   '183', '163', '219', '158', '175', '177', '200', '194', '142',
   '172', '186', '188', '167'], dtype=object)
```

In [20]: `df = df.dropna(subset=['Innings Runs Scored Num'])`

In [21]: `df['Innings Runs Scored Num'].unique()`

Out[21]: `array(['65', '42', '18', '17', '13', '11', '5', '0', '120', '71', '20', '16', '3', '2', '1', '40', '6', '4', '87', '54', '46', '30', '12', '69', '39', '14', '10', '9', '8', '7', '82', '52', '41', '15', '98', '43', '19', '111', '48', '36', '25', '67', '60', '84', '59', '55', '47', '85', '49', '45', '34', '29', '22', '74', '28', '77', '50', '32', '23', '35', '122', '100', '95', '103', '113', '53', '96', '27', '64', '58', '33', '31', '73', '56', '86', '62', '106', '24', '57', '104', '26', '66', '51', '118', '105', '101', '21', '79', '44', '102', '88', '80', '72', '97', '68', '89', '38', '83', '63', '148', '166', '90', '76', '37', '70', '124', '94', '140', '153', '107', '117', '121', '92', '78', '75', '114', '115', '130', '128', '151', '110', '138', '135', '109', '61', '179', '170', '112', '116', '91', '143', '93', '123', '145', '81', '150', '162', '108', '131', '133', '137', '146', '139', '125', '129', '157', '152', '144', '99', '127', '210', '147', '126', '181', '160', '180', '208', '176', '168', '141', '132', '119', '154', '185', '134', '156', '164', '173', '178', '171', '149', '237', '159', '161', '215', '264', '136', '169', '209', '174', '189', '183', '163', '219', '158', '175', '177', '200', '194', '142', '172', '186', '188', '167'], dtype=object)`

In [22]: `# convert to datetime
df['Innings Date']= pd.to_datetime(df['Innings Date'])`

In [23]: `df['year']= df['Innings Date'].dt.year`

In [24]: `df.head(1)`

Out[24]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate	In Nu
0	E Lewis	65	65	128	1.0	0.0	80	8	1	81.25	

◀ ▶

In [25]: `df.tail(1)`

Out[25]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Inning Battin Strik Ral
171941	RW Marsh	10*	10	24	1.0	1.0	18	2	0	55.5

◀ ▶

```
In [26]: # sachin 1994 - 2004
# virat 2009 - 2019
```

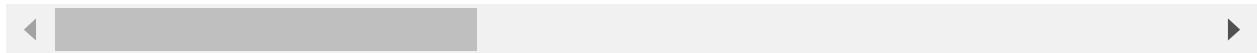
```
In [27]: sachin_df = df[(df.year >= 1994) & (df.year <= 2004)]
```

```
In [28]: virat_df = df[(df.year >= 2009) & (df.year <= 2019)]
```

```
In [29]: sachin_df.head(2)
```

Out[29]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate
77610	V Sehwag	70	70	85	1.0	0.0	52	9	2	134.61
77611	Yuvraj Singh	69	69	34	1.0	0.0	32	8	3	215.62



```
In [30]: virat_df.head(2)
```

Out[30]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate	In Nu
0	E Lewis	65	65	128	1.0	0.0	80	8	1	81.25	
1	N Pooran	42	42	69	1.0	0.0	52	4	1	80.76	



```
In [31]: # Runs per innings = Total Runs/Total Innings
```

```
# SR = 100*(Total Runs/Total Balls)
```

```
# 100's = sum(100's)
```

```
# 50's = sum(50's)
```

```
# Team contribution = Player Runs/Total Runs (ex. Virat scored 50/ Team IND 150 =
```

In [32]: df.dtypes

```
Out[32]: Innings Player          object
Innings Runs Scored          object
Innings Runs Scored Num       object
Innings Minutes Batted       object
Innings Batted Flag          float64
Innings Not Out Flag         float64
Innings Balls Faced          object
Innings Boundary Fours       object
Innings Boundary Sixes        object
Innings Batting Strike Rate   object
Innings Number                object
Opposition                     object
Ground                         object
Innings Date                  datetime64[ns]
Country                        object
50's                           float64
100's                          float64
Innings Runs Scored Buckets   object
Innings Overs Bowled          object
Innings Bowled Flag           float64
Innings Maidens Bowled        object
Innings Runs Conceded          object
Innings Wickets Taken          object
4 Wickets                      float64
5 Wickets                      float64
10 Wickets                     float64
Innings Wickets Taken Buckets  object
Innings Economy Rate           object
year                           int64
dtype: object
```

In [33]: # converting or change the datatypes

In [34]: df['Innings Runs Scored Num'] = df['Innings Runs Scored Num'].astype('int')

In [35]: df['Innings Balls Faced'] = df['Innings Balls Faced'].astype('int')

In [36]: df['Innings Not Out Flag'] = df['Innings Not Out Flag'].astype('int')

In [37]: # sachin 1994 - 2004
virat 2009 - 2019

In [38]: sachin_df = df[(df.year >= 1994) & (df.year <= 2004)]

In [39]: virat_df = df[(df.year >= 2009) & (df.year <= 2019)]

In [40]: `sachin_df.head(2)`

Out[40]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate
77610	V Sehwag	70	70	85	1.0	0	52	9	2	134.61
77611	Yuvraj Singh	69	69	34	1.0	0	32	8	3	215.62



In [41]: `virat_df.head(2)`

Out[41]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate	In Nu
0	E Lewis	65	65	128	1.0	0	80	8	1	81.25	
1	N Pooran	42	42	69	1.0	0	52	4	1	80.76	



In [42]: df.dtypes

```
Out[42]: Innings Player          object
Innings Runs Scored          object
Innings Runs Scored Num      int32
Innings Minutes Batted       object
Innings Batted Flag          float64
Innings Not Out Flag         int32
Innings Balls Faced          int32
Innings Boundary Fours       object
Innings Boundary Sixes        object
Innings Batting Strike Rate  object
Innings Number                object
Opposition                   object
Ground                       object
Innings Date                 datetime64[ns]
Country                      object
50's                         float64
100's                        float64
Innings Runs Scored Buckets  object
Innings Overs Bowled         object
Innings Bowled Flag          float64
Innings Maidens Bowled       object
Innings Runs Conceded        object
Innings Wickets Taken        object
4 Wickets                     float64
5 Wickets                     float64
10 Wickets                    float64
Innings Wickets Taken Buckets object
Innings Economy Rate         object
year                          int64
dtype: object
```

In [43]: df.head(2)

Out[43]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate	In Nu
0	E Lewis	65	65	128	1.0	0	80	8	1	81.25	
1	N Pooran	42	42	69	1.0	0	52	4	1	80.76	

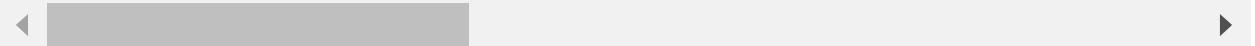
In [44]: # sachin_df.to_csv('sachin_data.csv')

In [45]: # what is the total runs scored by the sachin in these time frames?

In [46]: `sachin_df.head(2)`

Out[46]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate
77610	V Sehwag	70	70	85	1.0	0	52	9	2	134.61
77611	Yuvraj Singh	69	69	34	1.0	0	32	8	3	215.62



In [47]: `sachin_df.head(20)`

Out[47]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Inni Bat St F
77610	V Sehwag	70	70	85	1.0	0	52	9	2	13
77611	Yuvraj Singh	69	69	34	1.0	0	32	8	3	21
77612	R Dravid	60	60	87	1.0	0	68	5	0	8
77613	SC Ganguly	55	55	98	1.0	0	80	2	2	6
77614	SR Tendulkar	47	47	60	1.0	0	42	9	0	1
77615	M Kaif	29*	29	52	1.0	1	24	2	0	12
77616	MS Dhoni	7*	7	2	1.0	1	2	0	1	
77621	Rajin Saleh	82	82	149	1.0	0	114	8	1	7
77622	Mashrafe Mortaza	39	39	12	1.0	0	20	3	3	
77623	Mohammad Ashraful	32	32	52	1.0	0	34	4	0	9
77624	Mushfiqur Rahman	27*	27	42	1.0	1	44	3	0	6
77625	Khaled Mahmud	14	14	32	1.0	0	31	0	0	4
77626	Nafees Iqbal	10	10	33	1.0	0	20	2	0	
77627	Khaled Mashud	10	10	25	1.0	0	16	0	0	1
77628	Aftab Ahmed	9	9	24	1.0	0	13	1	0	6
77629	Habibul Bashar	2	2	13	1.0	0	7	0	0	2
77630	Hasibul Hossain	1*	1	2	1.0	1	1	0	0	
77631	Mohammad Rafique	0	0	3	1.0	0	4	0	0	
77654	SP Fleming	77*	77	136	1.0	1	92	8	2	8
77655	MS Sinclair	31	31	65	1.0	0	53	4	0	5

In [48]: `# SR Tendulkar
sdf = sachin_df[sachin_df['Innings Player'] == 'SR Tendulkar']`

In [49]: `sdf.head()`

Out[49]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Inning Battir Stril Ra
77614	SR Tendulkar	47	47	60	1.0	0	42	9	0	111
77747	SR Tendulkar	19	19	46	1.0	0	32	3	0	59.
78054	SR Tendulkar	16	16	30	1.0	0	17	2	0	94.
79590	SR Tendulkar	74	74	170	1.0	0	100	7	1	;
79681	SR Tendulkar	18	18	30	1.0	0	21	3	0	85.

In [50]: `sum(sdf['Innings Runs Scored Num'])`

Out[50]: 11818

In [51]: `# to look for player name
virat_df['Innings Player'].unique()`

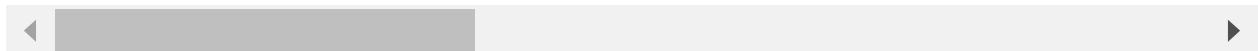
In [52]: `# virat_df.head(30)`

In [53]: `# similarly for virat
V Kohli
vdf = virat_df[virat_df['Innings Player'] == 'V Kohli']`

In [54]: `vdf.head()`

Out[54]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate
11	V Kohli	120	120	179	1.0	0	125	14	1	96
327	V Kohli	1	1	8	1.0	0	6	0	0	16.66
420	V Kohli	34*	34	61	1.0	1	41	3	0	82.92
664	V Kohli	26	26	45	1.0	0	27	3	0	96.29
804	V Kohli	66	66	103	1.0	0	76	7	0	86.84



In [55]: `sum(vdf['Innings Runs Scored Num'])`

Out[55]: 11247

In [56]: `len(vdf), len(sdf)`

Out[56]: (224, 271)

In [77]: *# RPI - for both sachin and virat*

```
sum(vdf['Innings Runs Scored Num'])/len(vdf), sum(sdf['Innings Runs Scored Num'])
```

Out[77]: (50.20982142857143, 43.608856088560884)

In [80]: *# SR for both*

```
100*sum(vdf['Innings Runs Scored Num'])/sum(vdf['Innings Balls Faced']), 100*sum(sdf['Innings Runs Scored Num'])/sum(sdf['Innings Balls Faced'])
```



Out[80]: (93.56126778138258, 88.21377920429947)

```
In [81]: # 100's for both
```

```
sum(vdf["100's"]), sum(sdf["100's"])
```

```
Out[81]: (42.0, 37.0)
```

```
In [82]: # 50's for both
```

```
sum(vdf["50's"]), sum(sdf["50's"])
```

```
Out[82]: (53.0, 57.0)
```

```
In [83]: # team contribution - Runs scored by each player, Runs by team
```

```
sum(vdf['Innings Runs Scored Num']), sum(sdf['Innings Runs Scored Num'])
```

```
Out[83]: (11247, 11818)
```

```
In [84]: sum(virat_df[virat_df.Country== 'India']['Innings Runs Scored Num'])
```

```
Out[84]: 63867
```

```
In [85]: sum(sachin_df[sachin_df.Country== 'India']['Innings Runs Scored Num'])
```

```
Out[85]: 69715
```

```
In [86]: 100*sum(vdf['Innings Runs Scored Num'])/sum(virat_df[virat_df.Country== 'India'][
```

```
Out[86]: 17.610033350556627
```

```
In [87]: 100*sum(sdf['Innings Runs Scored Num'])/sum(sachin_df[sachin_df.Country== 'India'][
```

```
Out[87]: 16.951875493078965
```

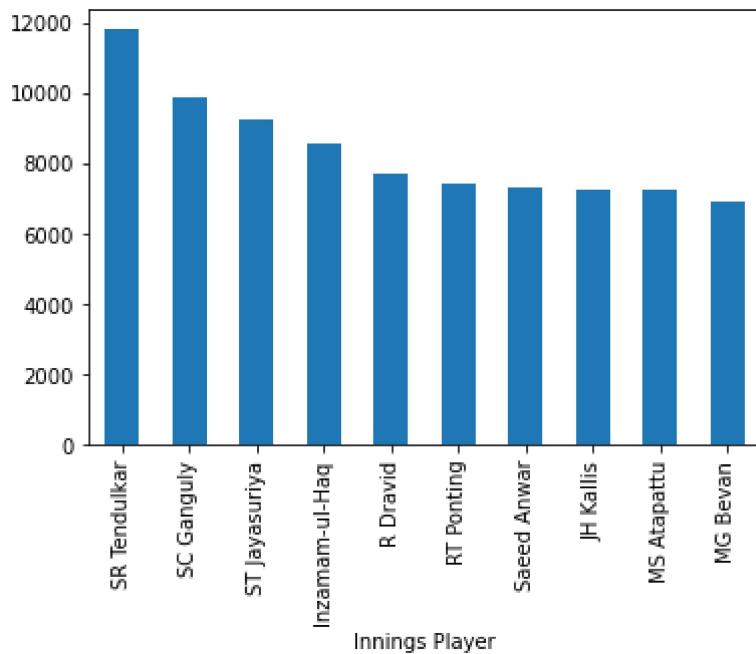
Visualizations:

```
In [65]: sachin_df.groupby(['Innings Player'])['Innings Runs Scored Num'].sum().sort_values
```

```
Out[65]: Innings Player
SR Tendulkar      11818
SC Ganguly        9911
ST Jayasuriya    9297
Inzamam-ul-Haq   8561
R Dravid          7751
RT Ponting        7422
Saeed Anwar       7320
JH Kallis         7267
MS Atapattu       7253
MG Bevan          6912
Name: Innings Runs Scored Num, dtype: int32
```

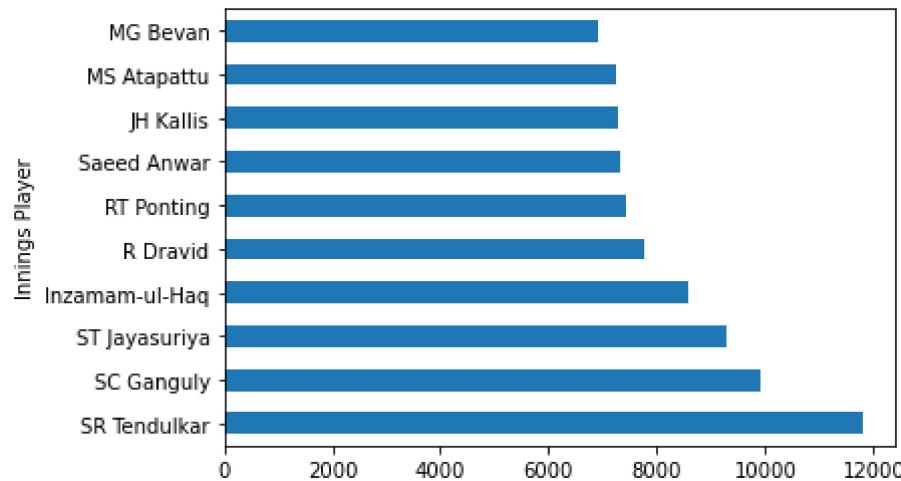
```
In [63]: # for vertical
```

```
sachin_df.groupby(['Innings Player'])['Innings Runs Scored Num'].sum().sort_values
plt.show()
```



In [62]: # for horizontal

```
sachin_df.groupby(['Innings Player'])['Innings Runs Scored Num'].sum().sort_values(ascending=False).head(10)
```



In [64]: virat_df.groupby(['Innings Player'])['Innings Runs Scored Num'].sum().sort_values(ascending=False).head(10)

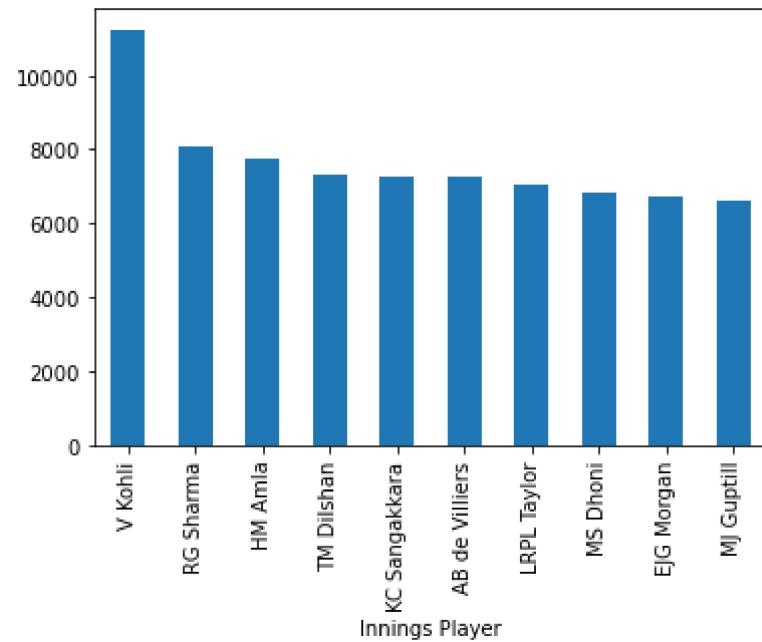
Out[64]: Innings Player

Innings Player	Innings Runs Scored Num
V Kohli	11247
RG Sharma	8083
HM Amla	7745
TM Dilshan	7296
KC Sangakkara	7275
AB de Villiers	7247
LRPL Taylor	7059
MS Dhoni	6838
EJG Morgan	6748
MJ Guptill	6626

Name: Innings Runs Scored Num, dtype: int32

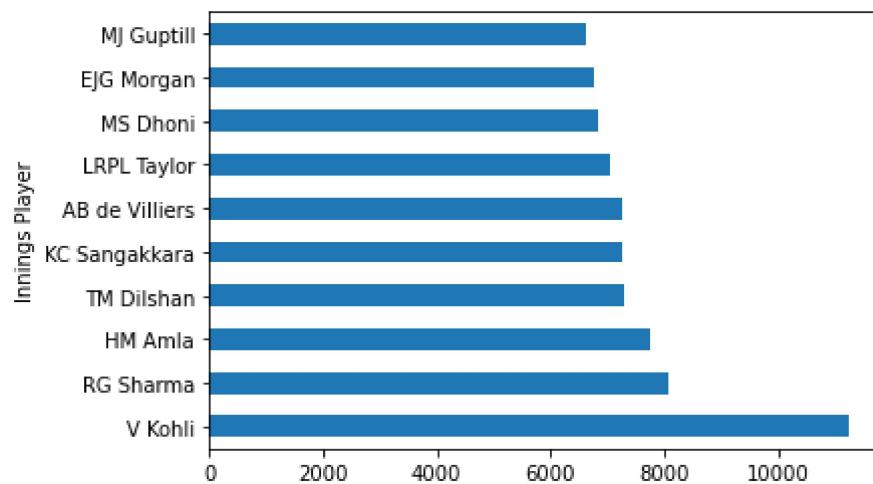
In [66]: #for vertical

```
virat_df.groupby(['Innings Player'])['Innings Runs Scored Num'].sum().sort_values  
plt.show()
```



In [69]: # for horizontal

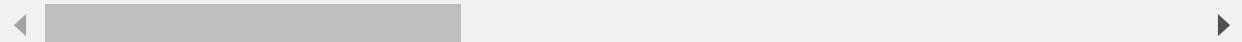
```
virat_df.groupby(['Innings Player'])['Innings Runs Scored Num'].sum().sort_values  
plt.show()
```



In [70]: `sdf.head(2)`

Out[70]:

	Innings Player	Innings Runs Scored	Innings Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Battin Strl Ra
77614	SR Tendulkar	47	47	60	1.0	0	42	9	0	111
77747	SR Tendulkar	19	19	46	1.0	0	32	3	0	59.

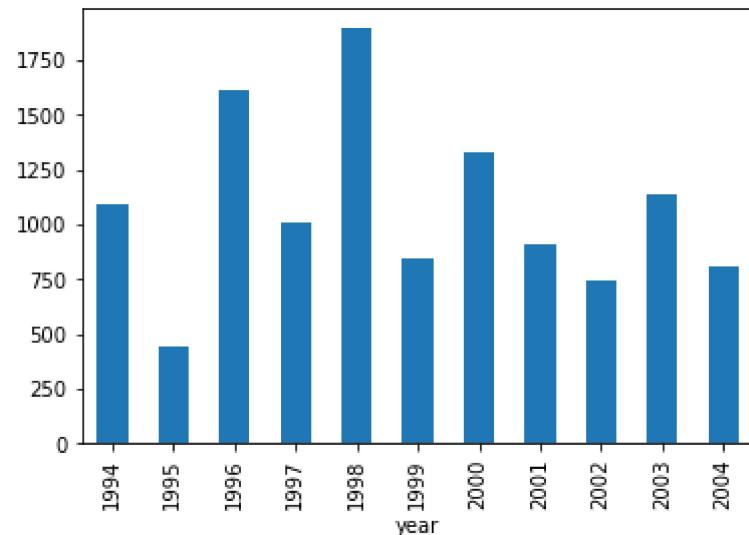


In [76]: `sdf.groupby(['year'])['Innings Runs Scored Num'].sum()`

Out[76]:

```
year
1994    1089
1995    444
1996   1611
1997   1011
1998   1894
1999    843
2000   1328
2001    904
2002    741
2003   1141
2004    812
Name: Innings Runs Scored Num, dtype: int32
```

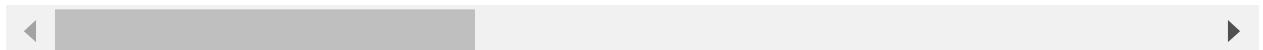
In [77]: `sdf.groupby(['year'])['Innings Runs Scored Num'].sum().plot(kind='bar')`
`plt.show()`



In [72]: `vdf.head(2)`

Out[72]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate
11	V Kohli	120	120	179	1.0	0	125	14	1	96
327	V Kohli	1	1	8	1.0	0	6	0	0	16.66



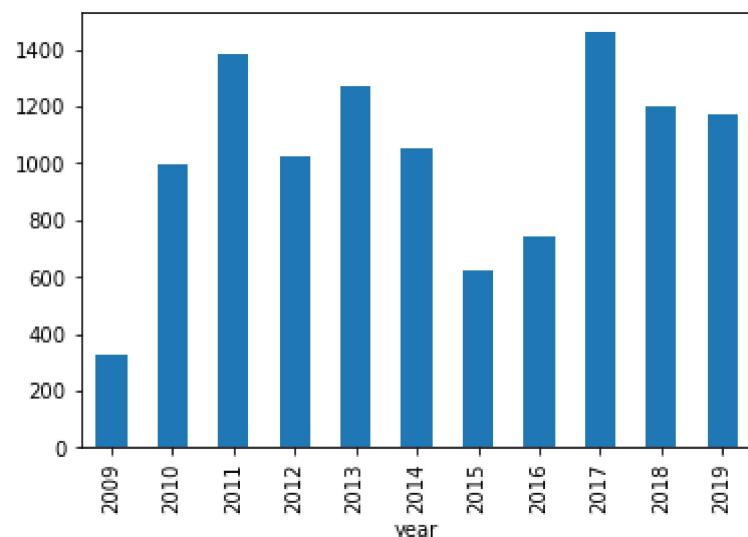
In [74]: `vdf.groupby(['year'])['Innings Runs Scored Num'].sum()`

Out[74]:

```
year
2009    325
2010    995
2011   1381
2012   1026
2013   1268
2014   1054
2015    623
2016    739
2017   1460
2018   1202
2019   1174
```

Name: Innings Runs Scored Num, dtype: int32

In [79]: `vdf.groupby(['year'])['Innings Runs Scored Num'].sum().plot(kind='bar')`
`plt.show()`



Normalization:

```
In [80]: # RPI - for both sachin and virat

sum(vdf['Innings Runs Scored Num'])/len(vdf), sum(sdf['Innings Runs Scored Num'])
```

Out[80]: (50.20982142857143, 43.608856088560884)

```
In [81]: # virat_df = player runs with virat
# player runs excluding virat => not_virat = virat_df[virat_df.player_name != 'V
```

```
In [82]: # RPI - for both sachin and virat
```

```
sum(virat_df['Innings Runs Scored Num'])/len(virat_df)
```

Out[82]: 24.99673202614379

```
In [83]: virat_df.head(2)
```

Out[83]:

	Innings Player	Innings Runs Scored	Innings Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate	In Nu
0	E Lewis	65	65	128	1.0	0	80	8	1	81.25	
1	N Pooran	42	42	69	1.0	0	52	4	1	80.76	



```
In [84]: non_virat_df = virat_df[virat_df['Innings Player'] != "V Kohli"]
```

```
In [87]: # average 25 runs
# where as virat has 50
```

```
In [85]: sum(non_virat_df['Innings Runs Scored Num'])/len(non_virat_df)
```

Out[85]: 24.737756786500366

```
In [86]: (sum(vdf['Innings Runs Scored Num'])/len(vdf))/(sum(non_virat_df['Innings Runs Sc
```



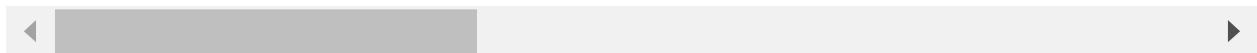
Out[86]: 2.029683688052565

```
In [88]: # similarly for sachin
```

In [89]: `sachin_df.head(2)`

Out[89]:

	Innings Player	Innings Runs Scored	Innings Runs Scored Num	Innings Minutes Batted	Innings Batted Flag	Innings Not Out Flag	Innings Balls Faced	Innings Boundary Fours	Innings Boundary Sixes	Innings Batting Strike Rate
77610	V Sehwag	70	70	85	1.0	0	52	9	2	134.61
77611	Yuvraj Singh	69	69	34	1.0	0	32	8	3	215.62



In [90]: `non_sachin_df = sachin_df[sachin_df['Innings Player'] != "SR Tendulkar"]`

In [91]: `sum(non_sachin_df['Innings Runs Scored Num'])/len(non_sachin_df)`

Out[91]: 22.43512946754194

In [92]: `(sum(sdf['Innings Runs Scored Num'])/len(sdf))/(sum(non_sachin_df['Innings Runs Scored Num'])/len(non_sachin_df))`



Out[92]: 1.9437755485945407

In [93]: `# virat => other
SR = V= 93, Others=80, V/Other, S/Other
100's - Number of matches to score a 100
50's - Number of matches to score a 50
Team Contribution - V_cont/O_cont`

In [94]: `#ex.
for V (i.e after every five match he will score century)
200/40`

Out[94]: 5.0

In [95]: `#for S (i.e after every 5.4 match he will score century)
200/37`

Out[95]: 5.405405405405405

In []: