

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

In [2]: sehwaḡ = pd.read_csv("sehwaḡ.csv")
dravid = pd.read_csv("dravid.csv")

In [3]: sehwaḡ.shape

Out[3]: (245, 14)

In [4]: dravid.shape

Out[4]: (318, 14)

In [6]: sehwaḡ

Out[6]:
```

	Runs	Mins	BF	4s	6s	SR	Pos	Dismissal	Inns	Unnamed: 9	Opposition	Ground	Start Date	Unnamed: 13	
0	1	1	5	2	0	0	50.00	7	lbw	1	NaN	v Pakistan	Mohali	1 Apr 1999	ODI # 1427
1	19	18	24	0	1	79.16	6	caught	1	NaN	v Zimbabwe	Rajkot	14 Dec 2000	ODI # 1660	
2	58	62	54	8	0	107.40	6	bowled	1	NaN	v Australia	Bengaluru	25 Mar 2001	ODI # 1696	
3	2	7	7	0	0	28.57	6	caught	2	NaN	v Zimbabwe	Bulawayo	27 Jun 2001	ODI # 1730	
4	11	19	16	1	0	68.75	6	not out	2	NaN	v West Indies	Bulawayo	30 Jun 2001	ODI # 1731	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
240	15	21	15	2	0	100.00	2	caught	1	NaN	v Sri Lanka	Hambantota	24 Jul 2012	ODI # 3292	
241	3	6	6	0	0	50.00	2	caught	2	NaN	v Sri Lanka	Colombo (RPS)	28 Jul 2012	ODI # 3293	
242	34	46	29	6	0	117.24	2	caught	2	NaN	v Sri Lanka	Colombo (RPS)	31 Jul 2012	ODI # 3294	
243	4	20	11	1	0	36.36	2	bowled	1	NaN	v Pakistan	Chennai	30 Dec 2012	ODI # 3314	
244	31	70	43	3	0	72.09	2	lbw	2	NaN	v Pakistan	Kolkata	3 Jan 2013	ODI # 3315	

245 rows × 14 columns

```
In [7]: sehwaḡ["Runs"].describe()

Out[7]: count    245.000000
mean      33.767347
std       34.809419
min        0.000000
25%        8.000000
50%       23.000000
75%       46.000000
max       219.000000
Name: Runs, dtype: float64

In [9]: s_25=np.percentile(sehwaḡ["Runs"],25) # percentile of Q1
s_25

Out[9]: 8.0

In [10]: s_50=np.percentile(sehwaḡ["Runs"],50) # percentile of Q2
s_50

Out[10]: 23.0

In [11]: s_75=np.percentile(sehwaḡ["Runs"],75) # percentile of Q3
s_75

Out[11]: 46.0

In [12]: np.percentile(sehwaḡ["Runs"],25)

Out[12]: 8.0

In [13]: sehwaḡ["Runs"].quantile(.25)

Out[13]: 8.0

In [14]: sehwaḡ["Runs"].quantile(.5)

Out[14]: 23.0

In [15]: sehwaḡ["Runs"].quantile(.75)

Out[15]: 46.0

In [17]: s_range= sehwaḡ["Runs"].max()-sehwaḡ["Runs"].min()
s_range

Out[17]: 219

In [18]: s_iqr=s_75-s_25
s_iqr

Out[18]: 38.0

In [20]: s_lower_whisker= max(s_25-(1.5*s_iqr),sehwaḡ["Runs"].min())
s_lower_whisker

Out[20]: 0

In [22]: s_upper_whisker= min(s_75+(1.5*s_iqr),sehwaḡ["Runs"].max())
s_upper_whisker

Out[22]: 193.0

In [25]: s_outlier=sehwaḡ.loc[sehwaḡ["Runs"]>s_upper_whisker]
s_outlier

Out[25]:
```

	Runs	Mins	BF	4s	6s	SR	Pos	Dismissal	Inns	Unnamed: 9	Opposition	Ground	Start Date	Unnamed: 13
37	126	136	104	21	1	121.15	1	caught	2	NaN	v England	Colombo (RPS)	22 Sep 2002	ODI # 1884
43	114	-	82	17	2	139.02	2	not out	2	NaN	v West Indies	Rajkot	12 Nov 2002	ODI # 1895
49	108	160	119	9	2	90.75	2	run out	2	NaN	v New Zealand	Napier	29 Dec 2002	ODI # 1927
53	112	175	139	11	3	80.57	2	caught	2	NaN	v New Zealand	Auckland	11 Jan 2003	ODI # 1933
75	130	193	134	15	2	97.01	1	caught	1	NaN	v New Zealand	Hyderabad (Deccan)	15 Nov 2003	ODI # 2062
104	108	156	95	9	3	113.68	1	bowled	1	NaN	v Pakistan	Kochi	2 Apr 2005	ODI # 2235
163	114	127	87	17	3	131.03	3	caught	1	NaN	v Bermuda	Port of Spain	19 Mar 2007	ODI # 2542
182	119	150	95	12	5	125.26	2	caught	2	NaN	v Pakistan	Karachi	26 Jun 2008	ODI # 2717
192	116	165	90	17	0	128.88	1	run out	1	NaN	v Sri Lanka	Colombo (RPS)	3 Feb 2009	ODI # 2813
198	125	106	74	14	6	168.91	2	not out	2	NaN	v New Zealand	Hamilton	11 Mar 2009	ODI # 2824
206	146	166	102	17	6	143.13	1	caught	1	NaN	v Sri Lanka	Rajkot	15 Dec 2009	ODI # 2932
220	110	136	93	16	1	118.27	2	caught	1	NaN	v New Zealand	Dambulla	25 Aug 2010	ODI # 3039
222	175	198	140	14	5	125.00	1	bowled	1	NaN	v Bangladesh	Mirpur	19 Feb 2011	ODI # 3100
233	219	208	149	25	7	146.97	2	caught	1	NaN	v West Indies	Indore	8 Dec 2011	ODI # 3223

```
In [28]: s_outlier.shape[0]

Out[28]: 14

In [29]: sehwaḡ.shape[0]

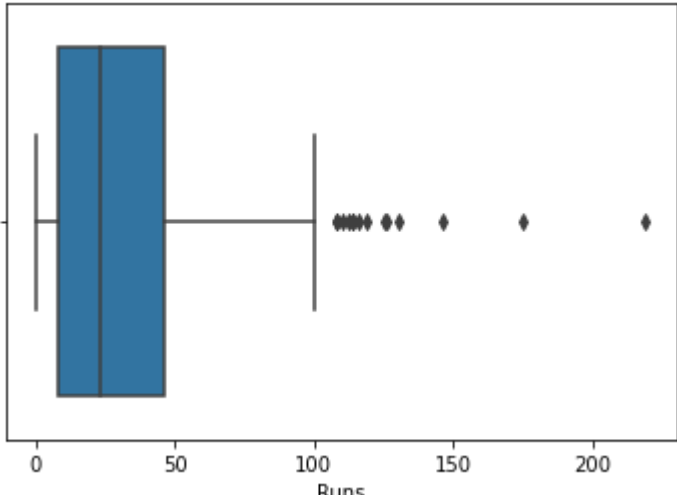
Out[29]: 245

In [30]: s_outlier.shape[0]/sehwaḡ.shape[0]

Out[30]: 0.05714285714285714

In [32]: sns.boxplot(x=sehwaḡ["Runs"])

Out[32]: <AxesSubplot: xlabel='Runs'>
```



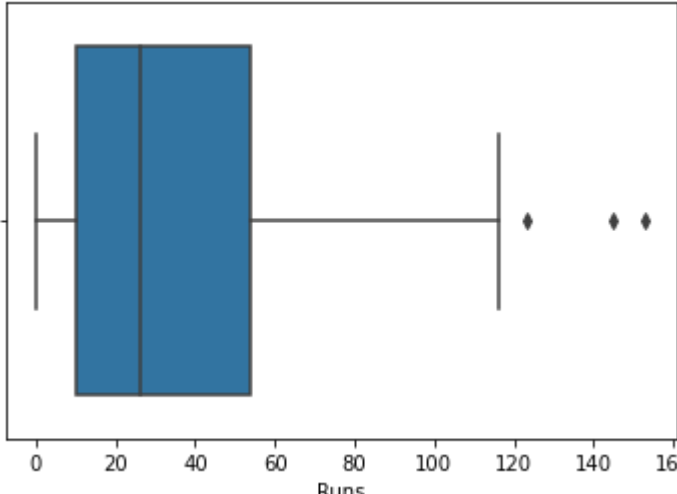
```
In [33]: sns.boxplot?

In [34]: dravid["Runs"].describe()

Out[34]: count    318.000000
mean      34.242138
std       29.681822
min        0.000000
25%       10.000000
50%       26.000000
75%       54.000000
max       153.000000
Name: Runs, dtype: float64

In [35]: sns.boxplot(x=dravid["Runs"])

Out[35]: <AxesSubplot: xlabel='Runs'>
```



```
In [40]: d_25=np.percentile(dravid["Runs"],25) # percentile of Q1
d_50=np.percentile(dravid["Runs"],50) # percentile of Q2
d_75=np.percentile(dravid["Runs"],75) # percentile of Q3
d_iqr=d_75-d_25

In [41]: d_lower_whisker= max(d_25-(1.5*d_iqr),dravid["Runs"].min())
d_lower_whisker

Out[41]: 0

In [42]: d_upper_whisker= min(d_75+(1.5*d_iqr),dravid["Runs"].max())
d_upper_whisker

Out[42]: 120.0

In [44]: d_outlier=dravid.loc[dravid["Runs"]>d_upper_whisker]
d_outlier

Out[44]:
```

	Runs	Mins	BF	4s	6s	SR	Pos	Dismissal	Inns	Unnamed: 9	Opposition	Ground	Start Date	Unnamed: 13
58	123	201	123	10	1	100.0	3	not out	1	NaN	v New Zealand	Taupo	9 Jan 1999	ODI # 1378
76	145	179	129	17	1	112.4	3	run out	1	NaN	v Sri Lanka	Taunton	26 May 1999	ODI # 1463
97	153	196	153	15	2	100.0	3	caught	1	NaN	v New Zealand	Hyderabad (Deccan)	8 Nov 1999	ODI # 1523

```
In [45]: d_outlier.shape[0]/dravid.shape[0]

Out[45]: 0.009433962264150943

In [46]: df=pd.read_html("https://www.espnccricinfo.com/records/most-runs-in-career-83548")
df

Out[46]:
```

		Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s	6s
0		SR Tendulkar (IND)	1989-2012	463	452	41	18426	200*	44.83	21368	86.23	49	96	20	2016	195
1		KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25	93	15	1385	88
2		RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30	82	20	1231	162
3		ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.20	28	68	34	1500	270
4		V Kohli (IND)	2008-2023	274	265	40	12898	183	57.32	13776	93.62	46	65	15	1211	138
...		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
88		JP Duminy (SA)	2004-2019	199	179	40	5117	150*	36.81	6052	84.55	4	27	7	347	75
89		PD Collingwood (ENG)	2001-2011	197	181	37	5092	120*	35.36	6614	76.98	5	26	7	365	74
90		Babar Azam (PAK)	2015-2023	100	98	12	5089	158	59.17	5702	89.24	18	26	3	458	51
91		A Symonds (AUS)	1998-2009	198	161	33	5088	156	39.75	5504	92.44	6	30	15	449	103
92		Abdul Razzaq (Asia/PAK)	1996-2011	265	228	57	5080	112	29.70	6252	81.25	3	23	14	382	124

[93 rows x 15 columns]]

```
In [47]: type(df)

Out[47]: list

In [49]: df1=df[0]
df1

Out[49]:
```

		Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s	6s
0		SR Tendulkar (IND)	1989-2012	463	452	41	18426	200*	44.83	21368	86.23	49	96	20	2016	195
1		KC Sangakkara (Asia/ICC/SL)	2000-2015	404	380	41	14234	169	41.98	18048	78.86	25	93	15	1385	88
2		RT Ponting (AUS/ICC)	1995-2012	375	365	39	13704	164	42.03	17046	80.39	30	82	20	1231	162
3		ST Jayasuriya (Asia/SL)	1989-2011	445	433	18	13430	189	32.36	14725	91.20	28	68	34	1500	270
4		V Kohli (IND)	2008-2023	274	265	40	12898	183	57.32	13776	93.62	46	65	15	1211	138
...		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
88		JP Duminy (SA)	2004-2019	199	179	40	5117	150*	36.81	6052	84.55	4	27	7	347	75
89		PD Collingwood (ENG)	2001-2011	197	181	37	5092	120*	35.36	6614	76.98	5	26	7	365	74
90		Babar Azam (PAK)	2015-2023	100	98	12	5089	158	59.17	5702	89.24	18	26	3	458	51
91		A Symonds (AUS)	1998-2009	198	161	33	5088	156	39.75	5504	92.44	6	30	15	449	103
92		Abdul Razzaq (Asia/PAK)	1996-2011	265	228	57	5080	112	29.70	6252	81.25	3	23	14	382	124

93 rows x 15 columns

```
In [ ]: s_outlier= data.loc[(data["col"]> s_upper_whisker)| (data["col"]< s_lower_whisker)]
```