Cricket_EDA

June 14, 2024

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
#importing the dataset
df = pd.read_excel('cricket_data.xlsx')
df.head()
```

[2]:

	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s	6s
0	DG Bradman (AUS)	1928-1948	52	80	10	6996	334	99.94	9800+	58.60	29	13	7	626	6
1	HC Brook (ENG)	2022-2023	12	20	1	1181	186	62.15	1287	91.76	4	7	1	141	23
2	AC Voges (AUS)	2015-2016	20	31	7	1485	269*	61.87	2667	55.68	5	4	2	186	5
3	RG Pollock (SA)	1963-1970	23	41	4	2256	274	60.97	1707+	54.48	7	11	1	246	11
4	GA Headley (WI)	1930-1954	22	40	4	2190	270*	60.83	416+	56.00	10	5	2	104	1

[3]:
```
#Renaming the column headers

df = df.rename(columns = {'Mat': 'Matches', 'NO':'Not Outs', 'HS': 'Highest␣
↪Inns Score', 'BF':'Balls Faced',
                  'SR': 'Batting Strike Rate'})
```

[4]:
```
df.head(5)
```

[4]:

	Player	Span	Matches	Inns	Not Outs	Runs	Highest Inns Score	Balls Faced	Batting Strike Rate	100	50	0	4s	6s
0	DG Bradman (AUS)	1928-1948	52	80	10	6996	334	9800+	58.60	29	13	7	626	6
1	HC Brook (ENG)	2022-2023	12	20	1	1181	186	1287	91.76	4	7	1	141	23
2	AC Voges (AUS)	2015-2016	20	31	7	1485	269*	2667	55.68	5	4	2	186	5
3	RG Pollock (SA)	1963-1970	23	41	4	2256	274	1707+	54.48	7	11	1	246	11
4	GA Headley (WI)	1930-1954	22	40	4	2190	270*	416+	56.00	10	5	2	104	1

```
4   GA Headley (WI)  1930-1954       22    40         4  2190

  Highest Inns Score    Ave Balls Faced  Batting Strike Rate  100  50  0   4s  \
0                334  99.94       9800+                58.60   29  13  7  626
1                186  62.15        1287                91.76    4   7  1  141
2               269*  61.87        2667                55.68    5   4  2  186
3                274  60.97       1707+                54.48    7  11  1  246
4               270*  60.83        416+                56.00   10   5  2  104

   6s
0   6
1  23
2   5
3  11
4   1
```

[5]:
```
df =df.dropna()
```

[6]:
```
#checking the null values
df.isnull().any()
```

[6]:
```
Player                False
Span                  False
Matches               False
Inns                  False
Not Outs              False
Runs                  False
Highest Inns Score    False
Ave                   False
Balls Faced           False
Batting Strike Rate   False
100                   False
50                    False
0                     False
4s                    False
6s                    False
dtype: bool
```

[7]:
```
df.describe()
```

[7]:
```
          Matches        Inns   Not Outs          Runs        Ave  \
count   62.000000   62.000000  62.000000     62.000000  62.000000
mean    80.306452  136.096774  14.145161   6437.016129  53.595806
std     50.839699   86.582796  11.154425   4019.583102   7.211395
min     12.000000   20.000000   1.000000    990.000000  48.000000
25%     30.250000   51.500000   5.250000   2379.000000  49.500000
50%     79.000000  134.500000  12.000000   6520.500000  51.580000
```

```
75%    117.750000  197.250000  19.000000   8844.000000  56.345000
max    200.000000  329.000000  49.000000  15921.000000  99.940000

       Batting Strike Rate         100         50          0           4s  \
count            62.000000  62.000000  62.000000  62.000000    62.000000
mean             47.485161  19.370968  28.661290   7.838710   654.403226
std              15.610293  12.378238  18.633606   5.549165   523.323035
min               0.000000   1.000000   3.000000   1.000000    37.000000
25%              42.442500   7.250000  11.500000   3.000000   150.500000
50%              49.645000  19.500000  28.500000   7.000000   567.000000
75%              55.297500  29.000000  42.500000  12.000000  1059.500000
max              91.760000  51.000000  68.000000  22.000000  2058.000000

              6s
count  62.000000
mean   27.403226
std    27.149531
min     0.000000
25%     5.250000
50%    20.500000
75%    42.750000
max    97.000000
```

[8]:
```
#drop duplicates

df.duplicated()
```

[8]:
```
0     False
1     False
2     False
3     False
4     False
      …
57    False
58    False
59    False
60    False
61    False
Length: 62, dtype: bool
```

[9]:
```
#split up span into and End Date

df['Span'].str.split(pat='-')
```

[9]:
```
0     [1928, 1948]
1     [2022, 2023]
2     [2015, 2016]
```

```
3     [1963, 1970]
4     [1930, 1954]
          …
57    [1965, 1981]
58    [2002, 2014]
59    [1924, 1934]
60    [1930, 1938]
61    [1928, 1934]
Name: Span, Length: 62, dtype: object
```

[10]:
```
df['Span'].str.split(pat = '-').str[1]
```

[10]:
```
0     1948
1     2023
2     2016
3     1970
4     1954
      …
57    1981
58    2014
59    1934
60    1938
61    1934
Name: Span, Length: 62, dtype: object
```

[11]:
```
df['Rookie Year'] = df['Span'].str.split(pat='-').str[0]
```

[12]:
```
df['Final Year'] = df['Span'].str.split(pat='-').str[1]
```

[13]:
```
df
```

[13]:
```
                Player       Span  Matches  Inns  Not Outs  Runs  \
0    DG Bradman (AUS)  1928-1948       52    80        10  6996
1      HC Brook (ENG)  2022-2023       12    20         1  1181
2      AC Voges (AUS)  2015-2016       20    31         7  1485
3     RG Pollock (SA)  1963-1970       23    41         4  2256
4     GA Headley (WI)  1930-1954       22    40         4  2190
..                 …        …      …   …         …   …
57   KD Walters (AUS)  1965-1981       74   125        14  5357
58  GC Smith (ICC/SA)  2002-2014      117   205        13  9265
59  WH Ponsford (AUS)  1924-1934       29    48         4  2122
60    SJ McCabe (AUS)  1930-1938       39    62         5  2748
61   DR Jardine (ENG)  1928-1934       22    33         6  1296

   Highest Inns Score    Ave Balls Faced  Batting Strike Rate  100  50   0  \
0                 334  99.94       9800+                58.60   29  13   7
1                 186  62.15        1287                91.76    4   7   1
```

```
2                269*  61.87        2667                55.68    5   4   2
3                 274  60.97       1707+                54.48    7  11   1
4                270*  60.83        416+                56.00   10   5   2
..                …     …          …                   …  …   ..  ..
57                250  48.26       8662+                49.16   15  33   4
58                277  48.25       15525                59.67   27  38  11
59                266  48.22       3118+                44.77    7   6   1
60                232  48.21       3217+                60.02    6  13   4
61                127  48.00       2110+                25.59    1  10   2

      4s  6s Rookie Year Final Year
0    626   6        1928       1948
1    141  23        2022       2023
2    186   5        2015       2016
3    246  11        1963       1970
4    104   1        1930       1954
..   …   ..          …          …
57   525  23        1965       1981
58  1165  24        2002       2014
59   119   0        1924       1934
60   241   5        1930       1938
61    53   0        1928       1934

[62 rows x 17 columns]
```

[14]:
```
#drop span columns

df = df.drop(columns=['Span'])
```

[15]:
```
df.head(5)
```

[15]:
```
             Player  Matches  Inns  Not Outs  Runs Highest Inns Score    Ave  \
0  DG Bradman (AUS)       52    80        10  6996                334  99.94
1    HC Brook (ENG)       12    20         1  1181                186  62.15
2    AC Voges (AUS)       20    31         7  1485               269*  61.87
3   RG Pollock (SA)       23    41         4  2256                274  60.97
4   GA Headley (WI)       22    40         4  2190               270*  60.83

  Balls Faced  Batting Strike Rate  100  50  0   4s  6s Rookie Year Final Year
0       9800+                58.60   29  13  7  626   6        1928       1948
1        1287                91.76    4   7  1  141  23        2022       2023
2        2667                55.68    5   4  2  186   5        2015       2016
3       1707+                54.48    7  11  1  246  11        1963       1970
4        416+                56.00   10   5  2  104   1        1930       1954
```

[16]:
```
#Spliting up the country from the year
```

```
df['Player'].str.split(pat = ',')
```

[16]:
```
0       [DG Bradman (AUS)]
1         [HC Brook (ENG)]
2         [AC Voges (AUS)]
3        [RG Pollock (SA)]
4        [GA Headley (WI)]
              ...
57      [KD Walters (AUS)]
58     [GC Smith (ICC/SA)]
59     [WH Ponsford (AUS)]
60       [SJ McCabe (AUS)]
61      [DR Jardine (ENG)]
Name: Player, Length: 62, dtype: object
```

[17]:
```
df['Country'] = df['Player'].str.split(pat='(').str[1]
df.head()
```

[17]:

	Player	Matches	Inns	Not Outs	Runs	Highest Inns Score	Ave	\
0	DG Bradman (AUS)	52	80	10	6996	334	99.94	
1	HC Brook (ENG)	12	20	1	1181	186	62.15	
2	AC Voges (AUS)	20	31	7	1485	269*	61.87	
3	RG Pollock (SA)	23	41	4	2256	274	60.97	
4	GA Headley (WI)	22	40	4	2190	270*	60.83	

	Balls Faced	Batting Strike Rate	100	50	0	4s	6s	Rookie Year	\
0	9800+	58.60	29	13	7	626	6	1928	
1	1287	91.76	4	7	1	141	23	2022	
2	2667	55.68	5	4	2	186	5	2015	
3	1707+	54.48	7	11	1	246	11	1963	
4	416+	56.00	10	5	2	104	1	1930	

	Final Year	Country
0	1948	AUS)
1	2023	ENG)
2	2016	AUS)
3	1970	SA)
4	1954	WI)

[18]:
```
# Assuming the DataFrame is named 'df' and the column is named 'Player'
df['Country'] = df['Country'].str.replace(')', '', regex=False)
```

[19]:
```
df.head()
```

[19]:

	Player	Matches	Inns	Not Outs	Runs	Highest Inns Score	Ave	\
0	DG Bradman (AUS)	52	80	10	6996	334	99.94	
1	HC Brook (ENG)	12	20	1	1181	186	62.15	

```
2    AC Voges (AUS)       20    31         7  1485               269*  61.87
3   RG Pollock (SA)       23    41         4  2256                274  60.97
4   GA Headley (WI)       22    40         4  2190               270*  60.83

  Balls Faced  Batting Strike Rate  100  50  0   4s  6s Rookie Year  \
0       9800+                58.60   29  13  7  626   6         1928
1        1287                91.76    4   7  1  141  23         2022
2        2667                55.68    5   4  2  186   5         2015
3       1707+                54.48    7  11  1  246  11         1963
4        416+                56.00   10   5  2  104   1         1930

  Final Year Country
0       1948     AUS
1       2023     ENG
2       2016     AUS
3       1970      SA
4       1954      WI
```

[20]:
```
df['Player'] = df['Player'].str.split(pat='(').str[0]
df.head()
```

[20]:
```
         Player  Matches  Inns  Not Outs  Runs Highest Inns Score    Ave  \
0  DG Bradman         52    80        10  6996                334  99.94
1    HC Brook         12    20         1  1181                186  62.15
2    AC Voges         20    31         7  1485               269*  61.87
3  RG Pollock         23    41         4  2256                274  60.97
4  GA Headley         22    40         4  2190               270*  60.83

  Balls Faced  Batting Strike Rate  100  50  0   4s  6s Rookie Year  \
0       9800+                58.60   29  13  7  626   6         1928
1        1287                91.76    4   7  1  141  23         2022
2        2667                55.68    5   4  2  186   5         2015
3       1707+                54.48    7  11  1  246  11         1963
4        416+                56.00   10   5  2  104   1         1930

  Final Year Country
0       1948     AUS
1       2023     ENG
2       2016     AUS
3       1970      SA
4       1954      WI
```

[21]:
```
#Checking the data types
df.dtypes
```

[21]:
```
Player                 object
Matches                 int64
```

```
Inns                     int64
Not Outs                 int64
Runs                     int64
Highest Inns Score      object
Ave                    float64
Balls Faced             object
Batting Strike Rate    float64
100                      int64
50                       int64
0                        int64
4s                       int64
6s                       int64
Rookie Year             object
Final Year              object
Country                 object
dtype: object
```

[22]:
```
# Removing the symbols from the 'Hghest Inns Score' column
df['Highest Inns Score'] = df['Highest Inns Score'].str.replace('*', '',␣
 ↪regex=False)
df.head()
```

[22]:

	Player	Matches	Inns	Not Outs	Runs	Highest Inns Score	Ave	\
0	DG Bradman	52	80	10	6996	334	99.94	
1	HC Brook	12	20	1	1181	186	62.15	
2	AC Voges	20	31	7	1485	269	61.87	
3	RG Pollock	23	41	4	2256	274	60.97	
4	GA Headley	22	40	4	2190	270	60.83	

	Balls Faced	Batting Strike Rate	100	50	0	4s	6s	Rookie Year	\
0	9800+	58.60	29	13	7	626	6	1928	
1	1287	91.76	4	7	1	141	23	2022	
2	2667	55.68	5	4	2	186	5	2015	
3	1707+	54.48	7	11	1	246	11	1963	
4	416+	56.00	10	5	2	104	1	1930	

	Final Year	Country
0	1948	AUS
1	2023	ENG
2	2016	AUS
3	1970	SA
4	1954	WI

[23]:
```
# Removing the symbols from the 'Balls Faced' column
df['Balls Faced'] = df['Balls Faced'].str.replace('+','', regex=False)
df.head()
```

[23]:
```
         Player  Matches  Inns  Not Outs  Runs Highest Inns Score    Ave  \
0  DG Bradman         52    80        10  6996                 334  99.94
1    HC Brook         12    20         1  1181                 186  62.15
2    AC Voges         20    31         7  1485                 269  61.87
3  RG Pollock         23    41         4  2256                 274  60.97
4  GA Headley         22    40         4  2190                 270  60.83

  Balls Faced  Batting Strike Rate  100  50  0   4s  6s Rookie Year  \
0        9800                58.60   29  13  7  626   6        1928
1        1287                91.76    4   7  1  141  23        2022
2        2667                55.68    5   4  2  186   5        2015
3        1707                54.48    7  11  1  246  11        1963
4         416                56.00   10   5  2  104   1        1930

  Final Year Country
0       1948     AUS
1       2023     ENG
2       2016     AUS
3       1970      SA
4       1954      WI
```

[24]:
```
df['Highest Inns Score'] = df['Highest Inns Score'].astype('int')
```

[25]:
```
df.dtypes
```

[25]:
```
Player                 object
Matches                 int64
Inns                    int64
Not Outs                int64
Runs                    int64
Highest Inns Score      int32
Ave                   float64
Balls Faced            object
Batting Strike Rate   float64
100                     int64
50                      int64
0                       int64
4s                      int64
6s                      int64
Rookie Year            object
Final Year             object
Country                object
dtype: object
```

[26]:
```
#Converting Rookie Year from object to an int
df['Rookie Year'] = df['Rookie Year'].astype('int')
```

[27]:
```
df.dtypes
```

[27]:
```
Player                  object
Matches                  int64
Inns                     int64
Not Outs                 int64
Runs                     int64
Highest Inns Score       int32
Ave                    float64
Balls Faced             object
Batting Strike Rate    float64
100                      int64
50                       int64
0                        int64
4s                       int64
6s                       int64
Rookie Year              int32
Final Year              object
Country                 object
dtype: object
```

[28]:
```
#Converting Final Year from object to an int

df['Final Year'] = df['Final Year'].astype('int')
```

[29]:
```
df.dtypes
```

[29]:
```
Player                  object
Matches                  int64
Inns                     int64
Not Outs                 int64
Runs                     int64
Highest Inns Score       int32
Ave                    float64
Balls Faced             object
Batting Strike Rate    float64
100                      int64
50                       int64
0                        int64
4s                       int64
6s                       int64
Rookie Year              int32
Final Year               int32
Country                 object
dtype: object
```

[30]:
```
df.head()
```

[30]:
```
         Player  Matches  Inns  Not Outs  Runs  Highest Inns Score    Ave  \
0  DG Bradman         52    80        10  6996                 334  99.94
1    HC Brook         12    20         1  1181                 186  62.15
2    AC Voges         20    31         7  1485                 269  61.87
3  RG Pollock         23    41         4  2256                 274  60.97
4  GA Headley         22    40         4  2190                 270  60.83

  Balls Faced  Batting Strike Rate  100  50  0   4s  6s  Rookie Year  \
0        9800                58.60   29  13  7  626   6         1928
1        1287                91.76    4   7  1  141  23         2022
2        2667                55.68    5   4  2  186   5         2015
3        1707                54.48    7  11  1  246  11         1963
4         416                56.00   10   5  2  104   1         1930

   Final Year Country
0        1948     AUS
1        2023     ENG
2        2016     AUS
3        1970      SA
4        1954      WI
```

[31]:
```
#Dropping rows with missing values 'Hon.FS Jackson'
df = df.drop(54, axis = 0)
```

[32]:
```
df.head()
```

[32]:
```
         Player  Matches  Inns  Not Outs  Runs  Highest Inns Score    Ave  \
0  DG Bradman         52    80        10  6996                 334  99.94
1    HC Brook         12    20         1  1181                 186  62.15
2    AC Voges         20    31         7  1485                 269  61.87
3  RG Pollock         23    41         4  2256                 274  60.97
4  GA Headley         22    40         4  2190                 270  60.83

  Balls Faced  Batting Strike Rate  100  50  0   4s  6s  Rookie Year  \
0        9800                58.60   29  13  7  626   6         1928
1        1287                91.76    4   7  1  141  23         2022
2        2667                55.68    5   4  2  186   5         2015
3        1707                54.48    7  11  1  246  11         1963
4         416                56.00   10   5  2  104   1         1930

   Final Year Country
0        1948     AUS
1        2023     ENG
2        2016     AUS
3        1970      SA
4        1954      WI
```

[33]:
```
df.dtypes
```

[33]:
```
Player                 object
Matches                 int64
Inns                    int64
Not Outs                int64
Runs                    int64
Highest Inns Score      int32
Ave                   float64
Balls Faced            object
Batting Strike Rate   float64
100                     int64
50                      int64
0                       int64
4s                      int64
6s                      int64
Rookie Year             int32
Final Year              int32
Country                object
dtype: object
```

[34]:
```
#Replace all '-' values with '0' and then convert the column to 'int' data type
df['Balls Faced'] = df['Balls Faced'].str.replace('-', '0').astype('int')
```

[35]:
```
df.dtypes
```

[35]:
```
Player                 object
Matches                 int64
Inns                    int64
Not Outs                int64
Runs                    int64
Highest Inns Score      int32
Ave                   float64
Balls Faced             int32
Batting Strike Rate   float64
100                     int64
50                      int64
0                       int64
4s                      int64
6s                      int64
Rookie Year             int32
Final Year              int32
Country                object
dtype: object
```

[36]:
```
#Carrer lenght of the players
```

```
df['Career Length'] = (df['Final Year'].astype(int) - df['Rookie Year'].
↪astype(int))
```

[37]:
```
df
```

[37]:

	Player	Matches	Inns	Not Outs	Runs	Highest Inns Score	Ave	\
0	DG Bradman	52	80	10	6996	334	99.94	
1	HC Brook	12	20	1	1181	186	62.15	
2	AC Voges	20	31	7	1485	269	61.87	
3	RG Pollock	23	41	4	2256	274	60.97	
4	GA Headley	22	40	4	2190	270	60.83	
..	…	…	…	…	…	…	…	
57	KD Walters	74	125	14	5357	250	48.26	
58	GC Smith	117	205	13	9265	277	48.25	
59	WH Ponsford	29	48	4	2122	266	48.22	
60	SJ McCabe	39	62	5	2748	232	48.21	
61	DR Jardine	22	33	6	1296	127	48.00	

	Balls Faced	Batting Strike Rate	100	50	0	4s	6s	Rookie Year	\
0	9800	58.60	29	13	7	626	6	1928	
1	1287	91.76	4	7	1	141	23	2022	
2	2667	55.68	5	4	2	186	5	2015	
3	1707	54.48	7	11	1	246	11	1963	
4	416	56.00	10	5	2	104	1	1930	
..	…	…	…	..	..	…	..	…	
57	8662	49.16	15	33	4	525	23	1965	
58	15525	59.67	27	38	11	1165	24	2002	
59	3118	44.77	7	6	1	119	0	1924	
60	3217	60.02	6	13	4	241	5	1930	
61	2110	25.59	1	10	2	53	0	1928	

	Final Year	Country	Career Length
0	1948	AUS	20
1	2023	ENG	1
2	2016	AUS	1
3	1970	SA	7
4	1954	WI	24
..	…	…	…
57	1981	AUS	16
58	2014	ICC/SA	12
59	1934	AUS	10
60	1938	AUS	8
61	1934	ENG	6

[61 rows x 18 columns]

[38]:
```
#Average carrer length of the players

df['Career Length'].mean()
```

[38]:
```
13.21311475409836
```

[39]:
```
#Average career length of Batting Strike Rate for players who played over 10 years

df[ df['Career Length']>10] ['Batting Strike Rate'].mean()
```

[39]:
```
49.061860465116276
```

[40]:
```
#Players that played before 1980

df[df['Rookie Year']<1980]
```

[40]:

	Player	Matches	Inns	Not Outs	Runs	Highest Inns Score	Ave	\
0	DG Bradman	52	80	10	6996	334	99.94	
3	RG Pollock	23	41	4	2256	274	60.97	
4	GA Headley	22	40	4	2190	270	60.83	
5	H Sutcliffe	54	84	9	4555	194	60.73	
6	E Paynter	20	31	5	1540	243	59.23	
7	KF Barrington	82	131	15	6806	256	58.67	
8	ED Weekes	48	81	5	4455	207	58.61	
9	WR Hammond	85	140	16	7249	336	58.45	
10	GS Sobers	93	160	21	8032	365	57.78	
13	JB Hobbs	61	102	7	5410	211	56.94	
14	CL Walcott	44	74	7	3798	220	56.68	
15	L Hutton	79	138	15	6971	364	56.67	
17	GE Tyldesley	14	20	2	990	122	55.00	
19	CA Davis	15	29	5	1301	183	54.20	
21	GS Chappell	87	151	19	7110	247	53.86	
22	AD Nourse	34	62	7	2960	231	53.81	
25	Javed Miandad	124	189	21	8832	280	52.57	
30	J Ryder	20	32	5	1394	201	51.62	
34	SM Gavaskar	125	214	16	10122	236	51.12	
38	AR Border	156	265	44	11174	205	50.56	
40	IVA Richards	121	182	12	8540	291	50.23	
41	DCS Compton	78	131	15	5807	278	50.06	
46	FMM Worrell	51	87	9	3860	261	49.48	
47	CP Mead	17	26	2	1185	182	49.37	
52	KC Bland	21	39	5	1669	144	49.08	
53	B Mitchell	42	80	9	3471	189	48.88	
56	RN Harvey	79	137	10	6149	205	48.41	
57	KD Walters	74	125	14	5357	250	48.26	
59	WH Ponsford	29	48	4	2122	266	48.22	

60	SJ McCabe	39	62	5	2748	232	48.21
61	DR Jardine	22	33	6	1296	127	48.00

	Balls Faced	Batting Strike Rate	100	50	0	4s	6s	Rookie Year	\
0	9800	58.60	29	13	7	626	6	1928	
3	1707	54.48	7	11	1	246	11	1963	
4	416	56.00	10	5	2	104	1	1930	
5	6558	34.59	16	23	2	202	6	1924	
6	1288	45.88	4	7	3	125	4	1931	
7	4957	42.42	20	35	5	591	27	1955	
8	0	0.00	15	19	6	258	2	1948	
9	7491	38.07	22	24	4	419	27	1927	
10	4063	53.58	26	30	12	593	32	1954	
13	5363	46.22	15	28	4	276	8	1908	
14	0	0.00	15	14	1	107	11	1948	
15	2844	39.34	19	33	5	358	7	1937	
17	178	29.21	3	6	2	37	2	1921	
19	665	35.48	4	4	1	65	1	1968	
21	13079	51.53	24	31	12	755	16	1970	
22	108	63.88	9	14	3	115	6	1935	
25	15164	45.99	23	43	6	788	48	1976	
30	2035	44.86	3	9	1	102	3	1920	
34	14184	43.35	34	45	12	908	26	1971	
38	27002	41.09	27	63	11	1161	28	1978	
40	9613	69.77	24	45	10	952	84	1974	
41	2731	38.81	17	28	10	342	3	1937	
46	488	40.16	9	22	11	240	11	1948	
47	619	30.04	4	3	3	38	0	1911	
52	885	37.74	3	9	2	123	20	1961	
53	158	29.11	8	21	3	41	1	1929	
56	1799	43.74	21	24	7	427	1	1948	
57	8662	49.16	15	33	4	525	23	1965	
59	3118	44.77	7	6	1	119	0	1924	
60	3217	60.02	6	13	4	241	5	1930	
61	2110	25.59	1	10	2	53	0	1928	

	Final Year	Country	Career Length
0	1948	AUS	20
3	1970	SA	7
4	1954	WI	24
5	1935	ENG	11
6	1939	ENG	8
7	1968	ENG	13
8	1958	WI	10
9	1947	ENG	20
10	1974	WI	20
13	1930	ENG	22

```
14          1960       WI              12
15          1955      ENG              18
17          1929      ENG               8
19          1973       WI               5
21          1984      AUS              14
22          1951       SA              16
25          1993      PAK              17
30          1929      AUS               9
34          1987      IND              16
38          1994      AUS              16
40          1991       WI              17
41          1957      ENG              20
46          1963       WI              15
47          1928      ENG              17
52          1966       SA               5
53          1949       SA              20
56          1963      AUS              15
57          1981      AUS              16
59          1934      AUS              10
60          1938      AUS               8
61          1934      ENG               6
```

[41]:
```
#Count of Players that played before 1980

df[df['Rookie Year']<1980]['Player'].count()
```

[41]:
```
31
```

[42]:
```
#Maximum Highest Inns Scored by a country and sorting in ascending order
df.groupby('Country')['Highest Inns Score'].max().sort_values(ascending=False)
```

[42]:
```
Country
ICC/WI     400
AUS        380
SL         374
WI         365
ENG        364
ICC/PAK    329
ICC/IND    319
PAK        313
SA         278
ICC/SA     277
IND        254
NZ         251
ZIM        232
Name: Highest Inns Score, dtype: int32
```

[43]:
```
#Maximum Highest Inns Scored by a country and sorting in ascending order and␣
 ↪adding a data frame to it.

df.groupby('Country')['Highest Inns Score'].max()\
  .to_frame('Highest by Country')\
  .reset_index()\
  .sort_values('Highest by Country', ascending=False)
```

[43]:

	Country	Highest by Country
5	ICC/WI	400
0	AUS	380
10	SL	374
11	WI	365
1	ENG	364
3	ICC/PAK	329
2	ICC/IND	319
8	PAK	313
9	SA	278
4	ICC/SA	277
6	IND	254
7	NZ	251
12	ZIM	232

[44]:
```
#Find the average of Hundreds, fifties, ducks, (0) by country

df.groupby('Country',)[['100', '50', '0']].mean()
```

[44]:

Country	100	50	0
AUS	20.625000	28.375000	8.562500
ENG	13.000000	22.083333	4.416667
ICC/IND	29.500000	47.500000	12.000000
ICC/PAK	25.000000	46.000000	15.000000
ICC/SA	36.000000	48.000000	13.500000
ICC/WI	34.000000	48.000000	17.000000
IND	29.500000	36.500000	10.750000
NZ	18.500000	22.000000	6.000000
PAK	21.250000	28.000000	9.500000
SA	9.800000	20.200000	3.400000
SL	28.666667	44.000000	12.333333
WI	16.625000	25.625000	7.250000
ZIM	12.000000	27.000000	5.000000

[45]:
```
#Find the average scored of Matches, Inns, Not Outs, Runs, Highest Inns Score,␣
 ↪Balls Faced by Player
```

```
df.groupby('Player',)[['Matches', 'Inns', 'Not Outs', 'Runs', 'Highest Inns␣
 ↪Score', 'Balls Faced']].mean()
```

[45]:

```
                 Matches   Inns  Not Outs     Runs  Highest Inns Score  \
Player
A Flower            63.0  112.0      19.0   4794.0               232.0
AB de Villiers     114.0  191.0      18.0   8765.0               278.0
AC Voges            20.0   31.0       7.0   1485.0               269.0
AD Nourse           34.0   62.0       7.0   2960.0               231.0
AR Border          156.0  265.0      44.0  11174.0               205.0
…                    …     …         …       …                    …
V Sehwag           104.0  180.0       6.0   8586.0               319.0
VG Kambli           17.0   21.0       1.0   1084.0               227.0
WH Ponsford         29.0   48.0       4.0   2122.0               266.0
WR Hammond          85.0  140.0      16.0   7249.0               336.0
Younis Khan        118.0  213.0      19.0  10099.0               313.0

                 Balls Faced
Player
A Flower             10636.0
AB de Villiers       16077.0
AC Voges              2667.0
AD Nourse              108.0
AR Border            27002.0
…                      …
V Sehwag             10441.0
VG Kambli             1823.0
WH Ponsford           3118.0
WR Hammond            7491.0
Younis Khan          19375.0

[61 rows x 6 columns]
```

[ ]: