

BASKETBALL

```
In [ ]: # ANAKHA R MENON  
# CH.EN.U4CSE20103  
# CSE- B
```

```
In [1]: import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt  
sns.set(color_codes=True)  
%matplotlib inline  
import numpy as np
```

```
In [2]: basketballdata=pd.read_csv('DS--+Part2--+Basketball.csv')  
basketballdata
```

```
Out[2]:
```

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion	Run
0	Team 1	86	4385	2762	1647	552	563	5947	3140	33	
1	Team 2	86	4262	2762	1581	573	608	5900	3114	25	
2	Team 3	80	3442	2614	1241	598	775	4534	3309	10	
3	Team 4	82	3386	2664	1187	616	861	4398	3469	6	
4	Team 5	86	3368	2762	1209	633	920	4631	3700	8	
...	
56	Team 57	1	34	38	8	10	20	38	66	-	
57	Team 58	1	22	30	7	8	15	37	57	-	
58	Team 59	1	19	30	7	5	18	51	85	-	
59	Team 60	1	14	30	5	4	21	34	65	-	
60	Team 61	1	-	-	-	-	-	-	-	-	

61 rows × 13 columns

```
In [3]: basketballdata.info #clean the dataset
```

```

Out[3]: <bound method DataFrame.info of
0 Team 1 86 4385 2762 1647 552 563
1 Team 2 86 4262 2762 1581 573 608
2 Team 3 80 3442 2614 1241 598 775
3 Team 4 82 3386 2664 1187 616 861
4 Team 5 86 3368 2762 1209 633 920
... ..
56 Team 57 1 34 38 8 10 20
57 Team 58 1 22 30 7 8 15
58 Team 59 1 19 30 7 5 18
59 Team 60 1 14 30 5 4 21
60 Team 61 1 - - - - -

BasketScored BasketGiven TournamentChampion Runner-up TeamLaunch \
0 5947 3140 33 23 1929
1 5900 3114 25 25 1929
2 4534 3309 10 8 1929
3 4398 3469 6 6 1931to32
4 4631 3700 8 7 1929
... ..
56 38 66 - - 2009-10
57 37 57 - - 1956-57
58 51 85 - - 1951-52
59 34 65 - - 1955-56
60 - - - - 2017~18

HighestPositionHeld
0 1
1 1
2 1
3 1
4 1
... ..
56 20
57 16
58 16
59 15
60 9

[61 rows x 13 columns]>

```

```
In [4]: basketballdata.info() #clean the dataset brackets difference here
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61 entries, 0 to 60
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Team                   61 non-null    object
1   Tournament              61 non-null    int64
2   Score                   61 non-null    object
3   PlayedGames             61 non-null    object
4   WonGames                61 non-null    object
5   DrawnGames              61 non-null    object
6   LostGames               61 non-null    object
7   BasketScored            61 non-null    object
8   BasketGiven             61 non-null    object
9   TournamentChampion      61 non-null    object
10  Runner-up               61 non-null    object
11  TeamLaunch              61 non-null    object
12  HighestPositionHeld     61 non-null    int64
dtypes: int64(2), object(11)
memory usage: 6.3+ KB

```

```
In [5]: basketballdata.isnull().sum()
```

```

Out[5]: Team                0
Tournament                0
Score                    0
PlayedGames               0
WonGames                  0
DrawnGames                0
LostGames                 0
BasketScored              0
BasketGiven               0
TournamentChampion        0
Runner-up                 0
TeamLaunch                0
HighestPositionHeld       0
dtype: int64

```

```
In [6]: basketballdata.isna().sum()
```

Out[6]: Team 0
Tournament 0
Score 0
PlayedGames 0
WonGames 0
DrawnGames 0
LostGames 0
BasketScored 0
BasketGiven 0
TournamentChampion 0
Runner-up 0
TeamLaunch 0
HighestPositionHeld 0
dtype: int64

In [7]: basketballdata.head()

Out[7]:

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion	Runn
0	Team 1	86	4385	2762	1647	552	563	5947	3140		33
1	Team 2	86	4262	2762	1581	573	608	5900	3114		25
2	Team 3	80	3442	2614	1241	598	775	4534	3309		10
3	Team 4	82	3386	2664	1187	616	861	4398	3469		6
4	Team 5	86	3368	2762	1209	633	920	4631	3700		8

In [8]: basketballdata.tail()

Out[8]:

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion	Runn
56	Team 57	1	34	38	8	10	20	38	66		-
57	Team 58	1	22	30	7	8	15	37	57		-
58	Team 59	1	19	30	7	5	18	51	85		-
59	Team 60	1	14	30	5	4	21	34	65		-
60	Team 61	1	-	-	-	-	-	-	-		-

In [9]: basketballdata.describe()

Out[9]:

	Tournament	HighestPositionHeld
count	61.000000	61.000000
mean	24.000000	7.081967
std	26.827225	5.276663
min	1.000000	1.000000
25%	4.000000	3.000000
50%	12.000000	6.000000
75%	38.000000	10.000000
max	86.000000	20.000000

In [10]: basketballdata.duplicated()

Out[10]: 0 False
1 False
2 False
3 False
4 False
...
56 False
57 False
58 False
59 False
60 False
Length: 61, dtype: bool

In [11]: basketballdata['Score'].unique()

```
Out[11]: array(['4385', '4262', '3442', '3386', '3368', '2819', '2792', '2573',
                '2109', '1884', '1814', '1789', '1471', '1416', '1389', '1351',
                '1314', '1174', '1148', '1020', '970', '667', '662', '606', '553',
                '538', '510', '445', '421', '416', '375', '353', '343', '293',
                '285', '277', '242', '230', '190', '188', '168', '150', '148',
                '132', '107', '96', '91', '83', '81', '76', '71', '56', '52', '42',
                '40', '35', '34', '22', '19', '14', '-'], dtype=object)
```

```
In [12]: basketballdata
```

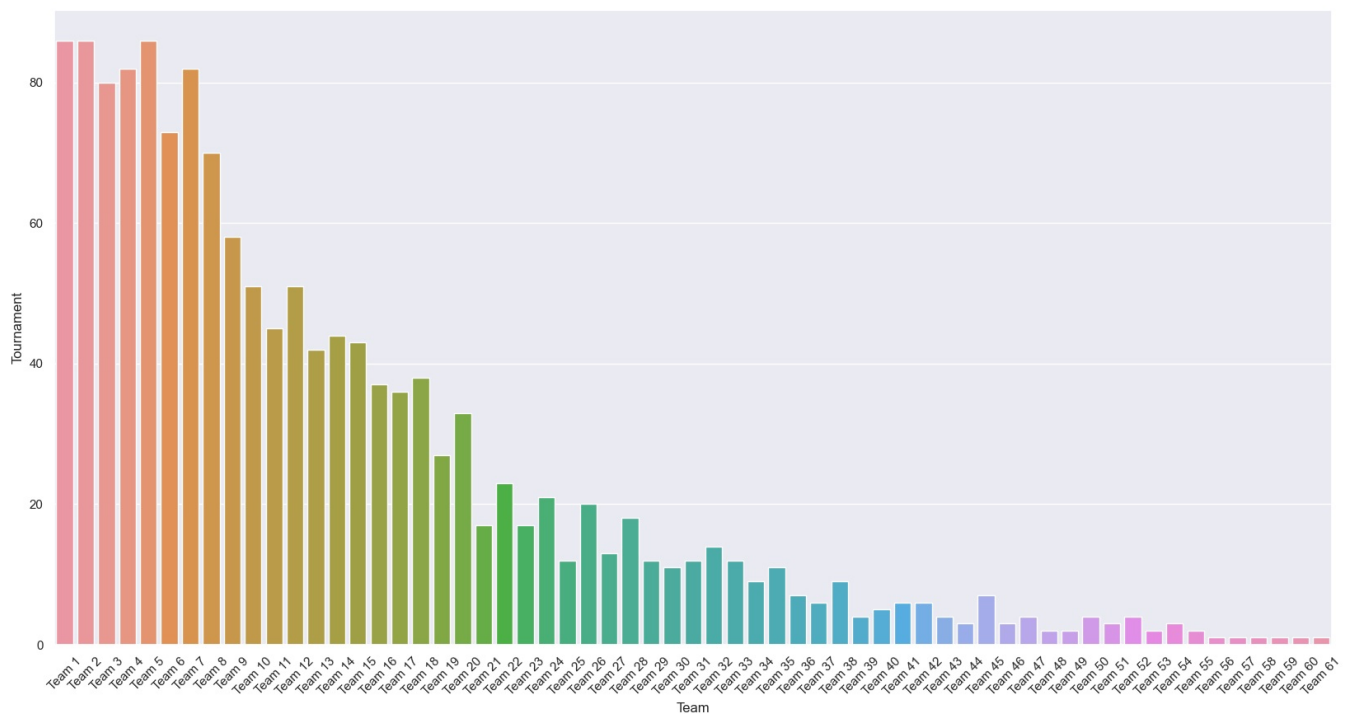
Out[12]:

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion	Run
0	Team 1	86	4385	2762	1647	552	563	5947	3140	33	
1	Team 2	86	4262	2762	1581	573	608	5900	3114	25	
2	Team 3	80	3442	2614	1241	598	775	4534	3309	10	
3	Team 4	82	3386	2664	1187	616	861	4398	3469	6	
4	Team 5	86	3368	2762	1209	633	920	4631	3700	8	
...	
56	Team 57	1	34	38	8	10	20	38	66	-	
57	Team 58	1	22	30	7	8	15	37	57	-	
58	Team 59	1	19	30	7	5	18	51	85	-	
59	Team 60	1	14	30	5	4	21	34	65	-	
60	Team 61	1	-	-	-	-	-	-	-	-	

61 rows × 13 columns

```
In [14]: # Q2 No of tournaments played by teams
plt.figure(figsize=(20,10))
sns.barplot(x = basketballdata['Team'], y = basketballdata['Tournament'])
plt.xticks(rotation=45)
```

```
Out[14]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
                17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
                34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
                51, 52, 53, 54, 55, 56, 57, 58, 59, 60]),
         [Text(0, 0, 'Team 1'),
          Text(1, 0, 'Team 2'),
          Text(2, 0, 'Team 3'),
          Text(3, 0, 'Team 4'),
          Text(4, 0, 'Team 5'),
          Text(5, 0, 'Team 6'),
          Text(6, 0, 'Team 7'),
          Text(7, 0, 'Team 8'),
          Text(8, 0, 'Team 9'),
          Text(9, 0, 'Team 10'),
          Text(10, 0, 'Team 11'),
          Text(11, 0, 'Team 12'),
          Text(12, 0, 'Team 13'),
          Text(13, 0, 'Team 14'),
          Text(14, 0, 'Team 15'),
          Text(15, 0, 'Team 16'),
          Text(16, 0, 'Team 17'),
          Text(17, 0, 'Team 18'),
          Text(18, 0, 'Team 19'),
          Text(19, 0, 'Team 20'),
          Text(20, 0, 'Team 21'),
          Text(21, 0, 'Team 22'),
          Text(22, 0, 'Team 23'),
          Text(23, 0, 'Team 24'),
          Text(24, 0, 'Team 25'),
          Text(25, 0, 'Team 26'),
          Text(26, 0, 'Team 27'),
          Text(27, 0, 'Team 28'),
          Text(28, 0, 'Team 29'),
          Text(29, 0, 'Team 30'),
          Text(30, 0, 'Team 31'),
          Text(31, 0, 'Team 32'),
          Text(32, 0, 'Team 33'),
          Text(33, 0, 'Team 34'),
          Text(34, 0, 'Team 35'),
          Text(35, 0, 'Team 36'),
          Text(36, 0, 'Team 37'),
          Text(37, 0, 'Team 38'),
          Text(38, 0, 'Team 39'),
          Text(39, 0, 'Team 40'),
          Text(40, 0, 'Team 41'),
          Text(41, 0, 'Team 42'),
          Text(42, 0, 'Team 43'),
          Text(43, 0, 'Team 44'),
          Text(44, 0, 'Team 45'),
          Text(45, 0, 'Team 46'),
          Text(46, 0, 'Team 47'),
          Text(47, 0, 'Team 48'),
          Text(48, 0, 'Team 49'),
          Text(49, 0, 'Team 50'),
          Text(50, 0, 'Team 51'),
          Text(51, 0, 'Team 52'),
          Text(52, 0, 'Team 53'),
          Text(53, 0, 'Team 54'),
          Text(54, 0, 'Team 55'),
          Text(55, 0, 'Team 56'),
          Text(56, 0, 'Team 57'),
          Text(57, 0, 'Team 58'),
          Text(58, 0, 'Team 59'),
          Text(59, 0, 'Team 60'),
          Text(60, 0, 'Team 61')])
```



```
In [15]: basketballdata.replace(to_replace='-', value=0, inplace=True)

basketballdata['TeamLaunch'] = basketballdata['TeamLaunch'].apply(lambda year: str(year)[:4])
```

```
In [16]: basketballdata
```

```
Out[16]:
```

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion	Run
0	Team 1	86	4385	2762	1647	552	563	5947	3140		33
1	Team 2	86	4262	2762	1581	573	608	5900	3114		25
2	Team 3	80	3442	2614	1241	598	775	4534	3309		10
3	Team 4	82	3386	2664	1187	616	861	4398	3469		6
4	Team 5	86	3368	2762	1209	633	920	4631	3700		8
...
56	Team 57	1	34	38	8	10	20	38	66		0
57	Team 58	1	22	30	7	8	15	37	57		0
58	Team 59	1	19	30	7	5	18	51	85		0
59	Team 60	1	14	30	5	4	21	34	65		0
60	Team 61	1	0	0	0	0	0	0	0		0

61 rows × 13 columns

```
In [17]: columns = basketballdata.columns
basketballdata[columns[1:]] = basketballdata[columns[1:]].astype('int')
basketballdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61 entries, 0 to 60
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Team                   61 non-null    object
1   Tournament             61 non-null    int32
2   Score                  61 non-null    int32
3   PlayedGames            61 non-null    int32
4   WonGames               61 non-null    int32
5   DrawnGames             61 non-null    int32
6   LostGames              61 non-null    int32
7   BasketScored           61 non-null    int32
8   BasketGiven            61 non-null    int32
9   TournamentChampion     61 non-null    int32
10  Runner-up              61 non-null    int32
11  TeamLaunch              61 non-null    int32
12  HighestPositionHeld    61 non-null    int32
dtypes: int32(12), object(1)
memory usage: 3.5+ KB
```

```
In [18]: basketballdata.dtypes
```

```
Out[18]: Team                object
Tournament              int32
Score                   int32
PlayedGames             int32
WonGames                int32
DrawnGames              int32
LostGames               int32
BasketScored            int32
BasketGiven             int32
TournamentChampion      int32
Runner-up               int32
TeamLaunch               int32
HighestPositionHeld     int32
dtype: object
```

```
In [19]: basketballdata.describe()
```

Out[19]:

	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion
count	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000
mean	24.000000	901.426230	796.819672	303.967213	188.934426	303.754098	1140.344262	1140.229508	1.426230
std	26.827225	1134.899121	876.282765	406.991030	201.799477	294.708594	1506.740211	1163.710766	5.472535
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.000000	96.000000	114.000000	34.000000	24.000000	62.000000	153.000000	221.000000	0.000000
50%	12.000000	375.000000	423.000000	123.000000	95.000000	197.000000	430.000000	632.000000	0.000000
75%	38.000000	1351.000000	1318.000000	426.000000	330.000000	563.000000	1642.000000	1951.000000	0.000000
max	86.000000	4385.000000	2762.000000	1647.000000	633.000000	1070.000000	5947.000000	3889.000000	33.000000

```
In [20]: basketballdata
```

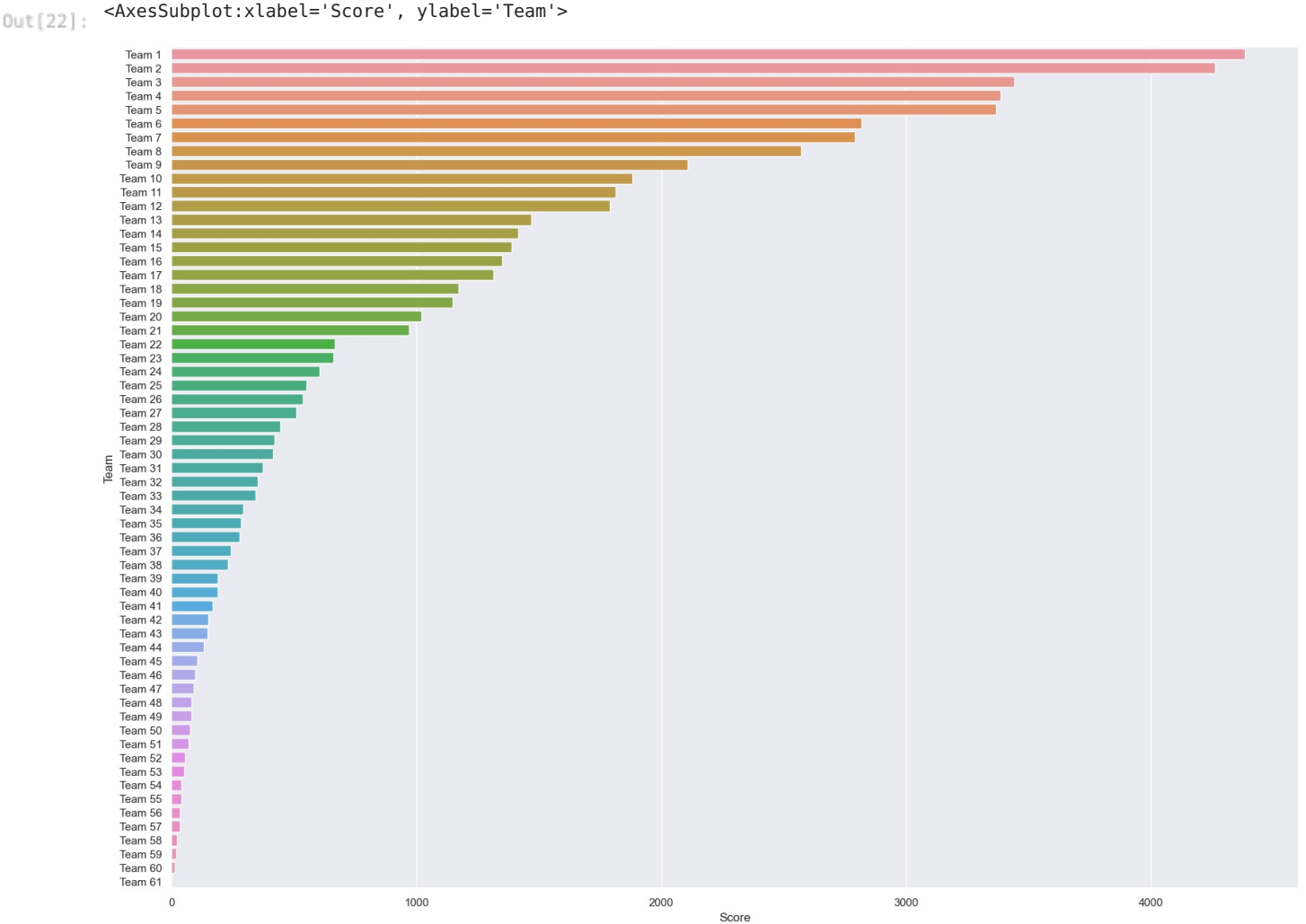
Out[20]:

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	BasketScored	BasketGiven	TournamentChampion	Run
0	Team 1		86	4385	2762	1647	552	563	5947	3140	33
1	Team 2		86	4262	2762	1581	573	608	5900	3114	25
2	Team 3		80	3442	2614	1241	598	775	4534	3309	10
3	Team 4		82	3386	2664	1187	616	861	4398	3469	6
4	Team 5		86	3368	2762	1209	633	920	4631	3700	8
...
56	Team 57		1	34	38	8	10	20	38	66	0
57	Team 58		1	22	30	7	8	15	37	57	0
58	Team 59		1	19	30	7	5	18	51	85	0
59	Team 60		1	14	30	5	4	21	34	65	0
60	Team 61		1	0	0	0	0	0	0	0	0

61 rows × 13 columns

In [22]:

```
#Q3 Scores of the teams
plt.figure(figsize=(20,15))
sns.barplot(x = basketballdata['Score'], y = basketballdata['Team'])
```



In [23]:

```
diff = basketballdata['PlayedGames'] - (basketballdata['DrawnGames'] + basketballdata['WonGames'] + basketballdata['LostGames'])
basketballdata.insert(loc=7, column='Balance', value=diff)
print(basketballdata)
```


	TeamLaunch	HighestPositionHeld
0	1929	1
1	1929	1
2	1929	1
3	1931	1
4	1929	1
..
56	2009	20
57	1956	16
58	1951	16
59	1955	15
60	2017	9

```
[61 rows x 14 columns]
```

```
In [24]: indexs_to_drop = basketballdata[basketballdata['Balance']>0].index
basketballdata.drop(indexs_to_drop, inplace=True)
basketballdata
```

Out[24]:

[illegible]

16	17	36	1314	1255	390	330	535	0	1421	1763
17	Team 18	38	1174	1192	408	292	492	0	1642	1951
18	Team 19	27	1148	988	333	256	399	0	1182	1371
19	Team 20	33	1020	1096	367	242	487	0	1347	1746
20	Team 21	17	970	646	266	172	208	0	892	789
21	Team 22	23	667	742	218	175	349	0	819	1157
23	Team 24	21	606	678	203	180	295	0	750	1022
24	Team 25	12	553	456	147	112	197	0	520	633
25	Team 26	20	538	628	184	149	295	0	716	1050
26	Team 27	13	510	494	155	128	211	0	619	744
27	Team 28	18	445	586	145	143	298	0	607	992
28	Team 29	12	421	380	125	81	174	0	458	623
29	Team 30	11	416	402	113	95	194	0	430	632
30	Team 31	12	375	423	123	102	198	0	422	581
31	Team 32	14	353	426	129	95	202	0	492	720
32	Team 33	12	343	448	104	127	217	0	393	662
33	Team 34	9	293	346	96	92	158	0	291	489
34	Team 35	11	285	334	103	79	152	0	419	588
35	Team 36	7	277	270	76	76	118	0	320	410
36	Team 37	6	242	228	62	56	110	0	244	366
37	Team 38	9	230	282	82	63	137	0	285	430
38	Team 39	4	190	160	52	45	63	0	199	241
39	Team 40	5	188	186	50	46	90	0	202	296
40	Team 41	6	168	204	59	50	95	0	216	310
41	Team 42	6	150	180	53	44	83	0	165	221
42	Team 43	4	148	152	37	37	78	0	155	253
43	Team 44	3	132	114	35	27	52	0	139	167
44	Team 45	7	107	130	43	21	66	0	227	308
45	Team 46	3	96	114	26	44	44	0	101	139
46	Team 47	4	91	116	34	16	66	0	181	295
47	Team 48	2	83	80	20	23	37	0	62	117
48	Team 49	2	81	80	19	24	37	0	70	115
49	Team 50	4	76	108	30	16	62	0	145	252
50	Team 51	3	71	90	29	13	48	0	121	183
51	Team 52	4	56	72	21	14	37	0	153	184
52	Team 53	2	52	68	17	18	33	0	71	116

53	Team 54	3	42	54	18	6	30	0	97	131
54	Team 55	2	40	68	13	14	41	0	70	182
55	Team 56	1	35	38	8	11	19	0	36	55
56	Team 57	1	34	38	8	10	20	0	38	66
57	Team 58	1	22	30	7	8	15	0	37	57
58	Team 59	1	19	30	7	5	18	0	51	85
59	Team 60	1	14	30	5	4	21	0	34	65
60	Team 61	1	0	0	0	0	0	0	0	0

```
In [25]: basketballdata.groupby(['Team'])['Tournament'].sum().sort_values(ascending=False).head(5)
```

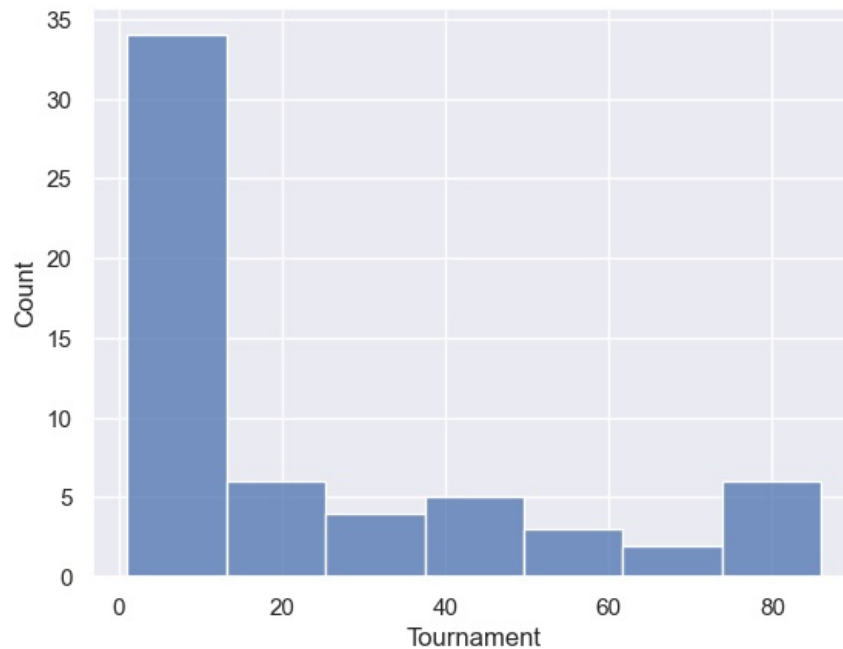
```
Out[25]: Team
Team 1    86
Team 2    86
Team 5    86
Team 7    82
Team 4    82
Name: Tournament, dtype: int32
```

```
In [26]: basketballdata.groupby(['Team'])['Tournament'].sum().sort_values(ascending=True).head(5)
```

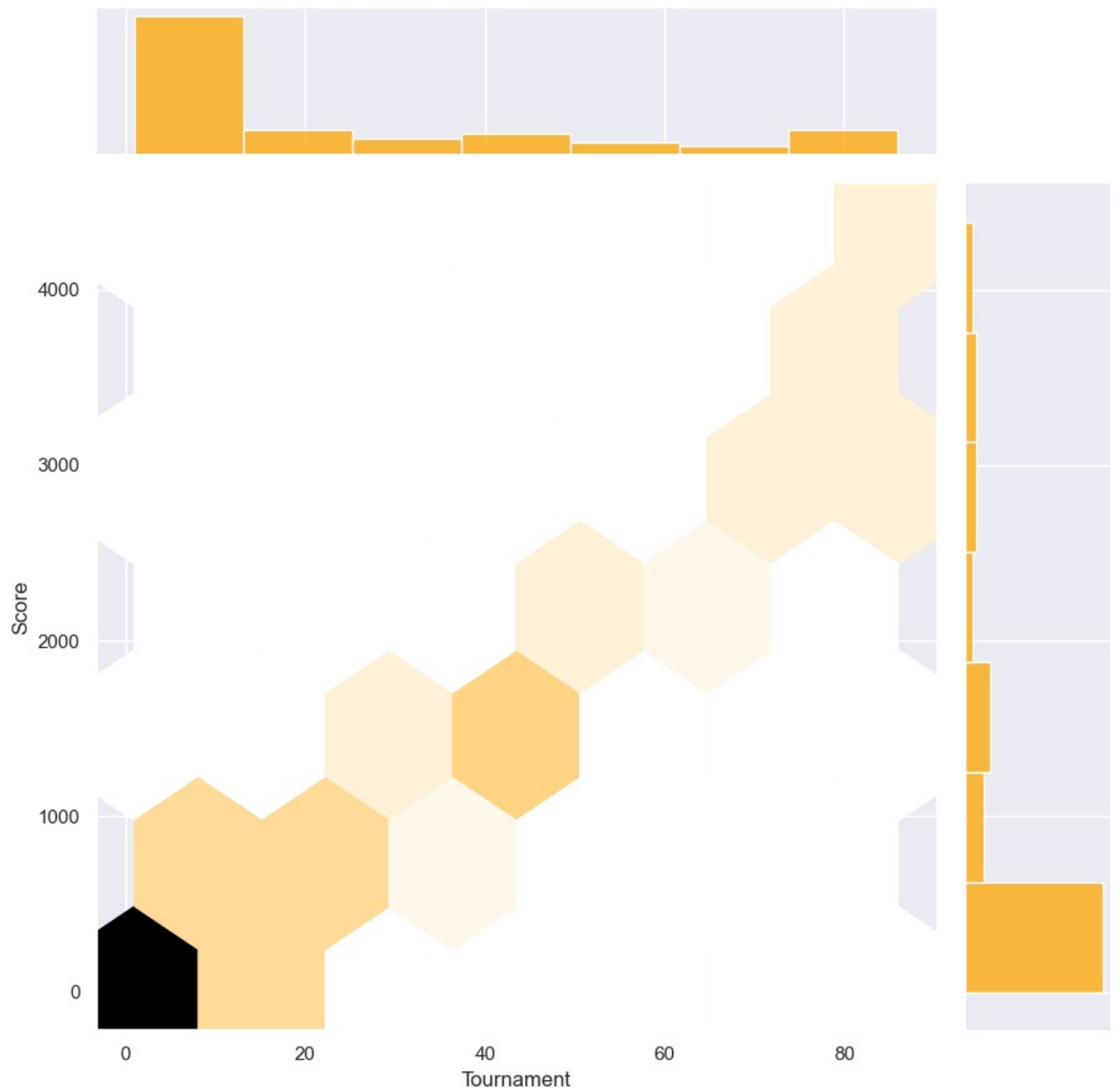
```
Out[26]: Team
Team 56    1
Team 59    1
Team 58    1
Team 61    1
Team 57    1
Name: Tournament, dtype: int32
```

```
In [27]: sns.histplot(basketballdata['Tournament'])
```

```
Out[27]: <AxesSubplot:xlabel='Tournament', ylabel='Count'>
```



```
In [29]: sns.jointplot(x = basketballdata['Tournament'], y = basketballdata['Score'], kind='hex', height=9, color='orange');
```

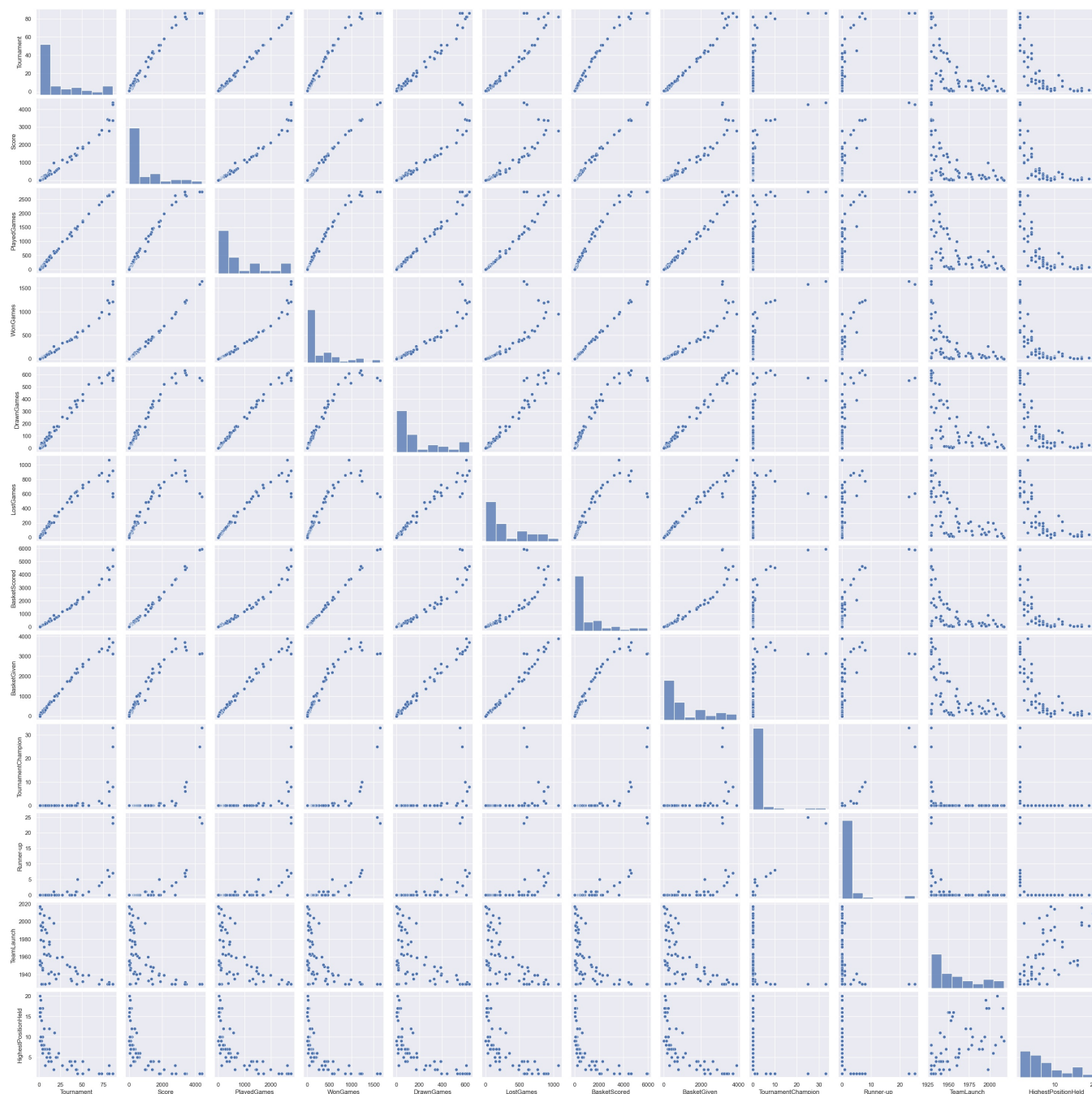


In [30]: `basketballdata.dtypes`

```
Out[30]: Team                object
Tournament              int32
Score                  int32
PlayedGames            int32
WonGames               int32
DrawnGames             int32
LostGames              int32
Balance                int32
BasketScored           int32
BasketGiven            int32
TournamentChampion     int32
Runner-up              int32
TeamLaunch             int32
HighestPositionHeld    int32
dtype: object
```

In [35]: `sns.pairplot(basketballdata[['Tournament', 'Score', 'PlayedGames', 'WonGames', 'DrawnGames', 'LostGames', 'BasketScored']])`

Out[35]: <seaborn.axisgrid.PairGrid at 0x288977f94e0>



```
In [36]: basketballdata['Champions'] = basketballdata['TournamentChampion']/basketballdata['Tournament']
basketballdata['Champions'].dtype
TopChampions = basketballdata['Champions'].sort_values(ascending =False).head(10)
print('The top 5 teams which became Champions are:')
basketballdata.groupby(['Team'])['Champions'].sum().sort_values(ascending =False).head(5)
```

The top 5 teams which became Champions are:

```
Out[36]:
Team
Team 1    0.383721
Team 2    0.290698
Team 3    0.125000
Team 5    0.093023
Team 4    0.073171
Name: Champions, dtype: float64
```

```
In [37]: #Team Age
basketballdata['TeamAge'] = 2020-basketballdata['TeamLaunch']
basketballdata[basketballdata['TeamAge'] == basketballdata['TeamAge'].max()]
```

Out[37]:

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	Balance	BasketScored	BasketGiven	TournamentChampion
0	Team 1		86	4385	2762	1647	552	563	0	5947	3140
1	Team 2		86	4262	2762	1581	573	608	0	5900	3114
2	Team 3		80	3442	2614	1241	598	775	0	4534	3309
4	Team 5		86	3368	2762	1209	633	920	0	4631	3700
6	Team 7		82	2792	2626	948	608	1070	0	3609	3889
7	Team 8		70	2573	2302	864	577	861	0	3228	3230
13	Team 14		44	1416	1428	453	336	639	0	1843	2368
44	Team 45		7	107	130	43	21	66	0	227	308
51	Team 52		4	56	72	21	14	37	0	153	184
53	Team 54		3	42	54	18	6	30	0	97	131

In [38]:

basketballdata[basketballdata['TeamAge'] == basketballdata['TeamAge'].min()]

Out[38]:

	Team	Tournament	Score	PlayedGames	WonGames	DrawnGames	LostGames	Balance	BasketScored	BasketGiven	TournamentChampion
60	Team 61		1	0	0	0	0	0	0	0	

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js