Ishita Sarkar

Python Assessment

1. **Read the data set and replace dashes with 0 to make sure you can perform arithmetic operations on the data. And check the distribution for the 'Best Position' and report the top position**

Answer 1

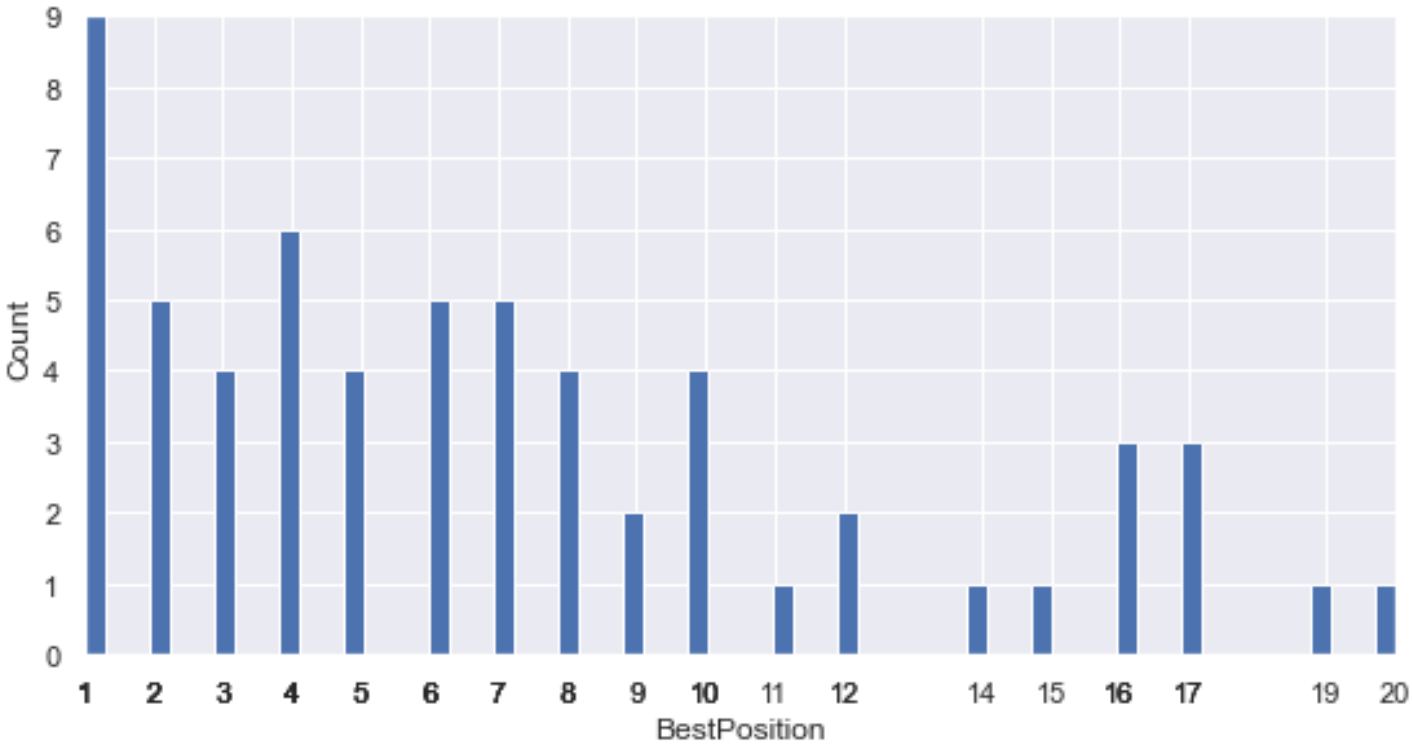
In [206]:
```
#1.*Read the data set and replace dashes with 0 to make sure you can perform arithmetic operations on the data.
#And check the distribution for the 'Best Position' and report the top position

sports = sports.replace('-',0)
sports
best=sports.sort_values(by=['BestPosition'],ascending = True).head()
best[['Team','BestPosition']]
```

Out[206]:

	Team	BestPosition
0	Real Madrid	1
1	Barcelona	1
2	Atletico Madrid	1
3	Valencia	1
4	Athletic Bilbao	1

To solve the issue of '-' which were present in the database '.replace' function was used to replace the hyphen characters with a 0 value.



After analysing the data and plotting a graph one can clearly observe that there are a total of 9 teams which are finishing up at the top of column 'BestPosition'

2. **Print all the teams which have started playing between 1930-1980 using "Debut" column (Include year 1930 only)**

Answer 2

In [247]:
```
#2. Print all the teams which have started playing between 1930-1980 using "Debut" column (Include year 1930 only)

sports['Debut']=sports['Debut'].astype(str)
sports_debut=sports[sports['Debut'].str[:4].between('1930','1980')]
sports_debut.shape
```

Out[247]: (37, 18)

So, to check for the teams who made their Debut in between the years 1930 and 1980 the ''.between" function is being used which accurately will choose all the teams based on the condition.

In [185]:
```
sports_debut[['Team','Debut']]
```

Out[185]:

	Team	Debut
3	Valencia	1931-32
5	Sevilla	1934-35
8	Zaragoza	1939-40
9	Real Betis	1932-33
10	Deportivo La Coruna	1941-42
11	Celta Vigo	1939-40
12	Valladolid	1948-49
14	Sporting Gijon	1944-45
15	Osasuna	1935-36
16	Malaga	1949-50
17	Oviedo	1933-34
18	Mallorca	1960-61
19	Las Palmas	1951-52
21	Granada	1941-42
22	Rayo Vallecano	1977-78

From the above output one can see all the teams which have Debuted between the years 1930 and 1980.

Not only that but the total number of teams which made their Debut between the given time frame where 37.

3. **Print the list of teams which came Top 5 in terms of points**

Answer 3

In [248]:
```
#3.Print the list of teams which came Top 5 in terms of points

top5=sports[['Points','Team']]
```

In [187]:
```
top5['Points'] = top5['Points'].astype(int)
top5.sort_values(by='Points', ascending=False, inplace=True)
top5.head(5)
```

Now to find the top5 teams using the points first the Points column values were converted into integers using '.astype' function after which the data was simply sorted using the 'sort_values' function where the values were sorted in descending order on the basis of 'Points'.

	Points	Team
0	4385	Real Madrid
1	4262	Barcelona
2	3442	Atletico Madrid
3	3386	Valencia
4	3368	Athletic Bilbao

So, after sorting 'head()' function was used to identify the top 5 teams which were sorted using the 'Points' attribute.

4. **Write a function with the name “Goal_diff_count” which should return all the teams with their Goal Differences. Goal_diff_count = GoalsFor - GoalsAgainst**

Answer4

In [267]:
```
#4. Write a function with the name “Goal_diff_count” which should return all the teams with their Goal Differences.
# Goal_diff_count = GoalsFor - GoalsAgainst

#converting goal_For & goal_against into int
sports['GoalsFor']=sports['GoalsFor'].astype(int)
sports['GoalsAgainst']=sports['GoalsAgainst'].astype(int)


def Goal_diff_count():
    sports['Goal_diff_count'] = sports['GoalsFor']-sports['GoalsAgainst']
    return sports[['Team','Goal_diff_count']]

Goal=Goal_diff_count().sort_values(by=['Goal_diff_count'],ascending = False)
Goal
```

First the values are converted to int type using the ‘astype’ function so that we can perform arithmetic operations on the data, after which a user-defined function is used to calculate and store the data into the database.

	Team	Goal_diff_count
0	Real Madrid	2807
1	Barcelona	2786
2	Atletico Madrid	1225
4	Athletic Bilbao	931
3	Valencia	929
27	Murcia	-385
19	Las Palmas	-399
14	Sporting Gijon	-399
12	Valladolid	-413
13	Racing Santander	-525

Further the data is sorted using the same user-defined function and later printed.

5. **Using the same function, find the team which has a maximum and minimum goal difference.**

Answer5

In [270]:
```
# 5. Using the same function, find the team which has a maximum and minimum goal difference.


Goal.head(1)#max
```

Out[270]:

	Team	Goal_diff_count
0	Real Madrid	2807

In [271]:
```
Goal.tail(1)#min
```

Out[271]:

	Team	Goal_diff_count
13	Racing Santander	-525

To find the maximum and minimum goal difference the ‘head’ and ‘tail’ functions are used.

Where maximum goal difference value is; 2807 and,

Minimum goal difference value is -525

6. **Create a new column with the name “Winning Percent” and append it to the data set Percentage of Winning = (GamesWon / GamesPlayed)*100 If there are any numerical error, replace it with 0%**

Answer6

In [272]:
```
# 6. Create a new column with the name "Winning Percent" and append it to the data set
# Percentage of Winning = (GamesWon / GamesPlayed)*100 If there are any numerical error, replace it with 0%
sports['GamesWon']=sports['GamesWon'].astype(int)
sports['GamesPlayed']=sports['GamesPlayed'].astype(int)

sports['Winning Percentage'] = (sports['GamesWon']/sports['GamesPlayed']) *100

sports['Winning Percentage'].fillna(0,inplace=True)

sports[['Team','Winning Percentage']]
```

First the data/values are converted into int using the astype function, further the given formula is used to calculate the ‘Winning Percentage’, after which a check is performed where the system check for NaN/Null or wrong values and changes/replaces it with 0 using the ‘fillna’ function.

	Team	Winning Percentage
0	Real Madrid	59.630702
1	Barcelona	57.241130
2	Atletico Madrid	47.475134
3	Valencia	44.557057
4	Athletic Bilbao	43.772629
...	...	...
56	Xerez	21.052632
57	Condal	23.333333
58	Atletico Tetuan	23.333333
59	Cultural Leonesa	16.666667
60	Girona	0.000000

61 rows × 2 columns

7. Print the top 5 teams which have the highest Winning percentage

Answer7

In [283]:
```
# 7. Print the top 5 teams which have the highest Winning percentage

sports.sort_values('Winning Percentage', ascending=False).head(5)[['Team','Winning Percentage']]
```

	Team	Winning Percentage
0	Real Madrid	59.630702
1	Barcelona	57.241130
2	Atletico Madrid	47.475134
3	Valencia	44.557057
4	Athletic Bilbao	43.772629

8. Group teams based on their “Best position” and print the sum of their points for all positions

Answer8

In [287]:
```
# 8. Group teams based on their “Best position” and print the sum of their points for all positions

grp_BestPos = sports[['Team','Points','BestPosition']].groupby('BestPosition')
grp_BestPos
grp_BestPos.sum()
```

BestPosition	Points
1	27933
2	6904
3	5221
4	6563
5	1884
6	2113
7	1186
8	1134

BestPosition	Points
9	96
10	450
11	445
12	511
14	71
15	14
16	81
17	266
19	81
20	34