

```
In [44]: import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt

In [6]: ds=pd.read_csv("teams.csv")

In [7]: ds

Out[7]:
   team  country  year  events  athletes  age  height  weight  medals  prev_medals  prev_3_medals
0  AFG  Afghanistan  1964      8         8  22.0  161.0   64.2      0          0.0          0.0
1  AFG  Afghanistan  1968      5         5  23.2  170.2   70.0      0          0.0          0.0
2  AFG  Afghanistan  1972      8         8  29.0  168.3   63.8      0          0.0          0.0
3  AFG  Afghanistan  1980     11        11  23.6  168.4   63.2      0          0.0          0.0
4  AFG  Afghanistan  2004      5         5  18.6  170.8   64.8      0          0.0          0.0
...   ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
2139 ZIM  Zimbabwe  2000     19        26  25.0  179.0   71.1      0          0.0          0.0
2140 ZIM  Zimbabwe  2004     11        14  25.1  177.8   70.5      3          0.0          0.0
2141 ZIM  Zimbabwe  2008     15        16  26.1  171.9   63.7      4          3.0          1.0
2142 ZIM  Zimbabwe  2012      8         9  27.3  174.4   65.2      0          4.0          2.3
2143 ZIM  Zimbabwe  2016     13        31  27.5  167.8   62.2      0          0.0          2.3

2144 rows x 11 columns

In [8]: ds.head()

Out[8]:
   team  country  year  events  athletes  age  height  weight  medals  prev_medals  prev_3_medals
0  AFG  Afghanistan  1964      8         8  22.0  161.0   64.2      0          0.0          0.0
1  AFG  Afghanistan  1968      5         5  23.2  170.2   70.0      0          0.0          0.0
2  AFG  Afghanistan  1972      8         8  29.0  168.3   63.8      0          0.0          0.0
3  AFG  Afghanistan  1980     11        11  23.6  168.4   63.2      0          0.0          0.0
4  AFG  Afghanistan  2004      5         5  18.6  170.8   64.8      0          0.0          0.0

In [9]: ds.tail()

Out[9]:
   team  country  year  events  athletes  age  height  weight  medals  prev_medals  prev_3_medals
2139 ZIM  Zimbabwe  2000     19        26  25.0  179.0   71.1      0          0.0          0.0
2140 ZIM  Zimbabwe  2004     11        14  25.1  177.8   70.5      3          0.0          0.0
2141 ZIM  Zimbabwe  2008     15        16  26.1  171.9   63.7      4          3.0          1.0
2142 ZIM  Zimbabwe  2012      8         9  27.3  174.4   65.2      0          4.0          2.3
2143 ZIM  Zimbabwe  2016     13        31  27.5  167.8   62.2      0          0.0          2.3

In [10]: ds.describe()

Out[10]:
   count  2144.000000  2144.000000  2144.000000  2144.000000  2144.000000  2144.000000  2144.000000  2014.000000  2014.000000
   mean  1994.507463   35.724813   73.693097   24.778591   173.900233   69.271595   10.556437   10.248759   9.449901
   std    15.384205   49.490270  127.907577   2.808559     5.357367    7.606507   33.028143   31.951920   28.232227
   min   1964.000000   1.000000   1.000000   17.000000  151.000000   43.000000   0.000000   0.000000   0.000000
   25%   1984.000000   6.000000   7.000000  23.275000   170.500000   64.500000   0.000000   0.000000   0.000000
   50%   1996.000000  13.000000  19.000000  24.700000   174.400000   69.400000   0.000000   0.000000   0.300000
   75%   2008.000000  44.000000  70.250000  26.100000   177.300000   73.400000   4.000000   4.000000   4.650000
   max   2016.000000  270.000000  839.000000  66.000000  193.000000  148.000000  442.000000  442.000000  314.000000

In [11]: teams=ds[['team','country','year','athletes','age','prev_medals','medals']]

In [12]: teams

Out[12]:
   team  country  year  athletes  age  prev_medals  medals
0  AFG  Afghanistan  1964      8  22.0          0.0      0
1  AFG  Afghanistan  1968      5  23.2          0.0      0
2  AFG  Afghanistan  1972      8  29.0          0.0      0
3  AFG  Afghanistan  1980     11  23.6          0.0      0
4  AFG  Afghanistan  2004      5  18.6          0.0      0
...   ...      ...      ...      ...      ...      ...
2139 ZIM  Zimbabwe  2000     26  25.0          0.0      0
2140 ZIM  Zimbabwe  2004     14  25.1          0.0      3
2141 ZIM  Zimbabwe  2008     16  26.1          3.0      4
2142 ZIM  Zimbabwe  2012      9  27.3          4.0      0
2143 ZIM  Zimbabwe  2016     31  27.5          0.0      0

2144 rows x 7 columns

In [13]: teams.corr()

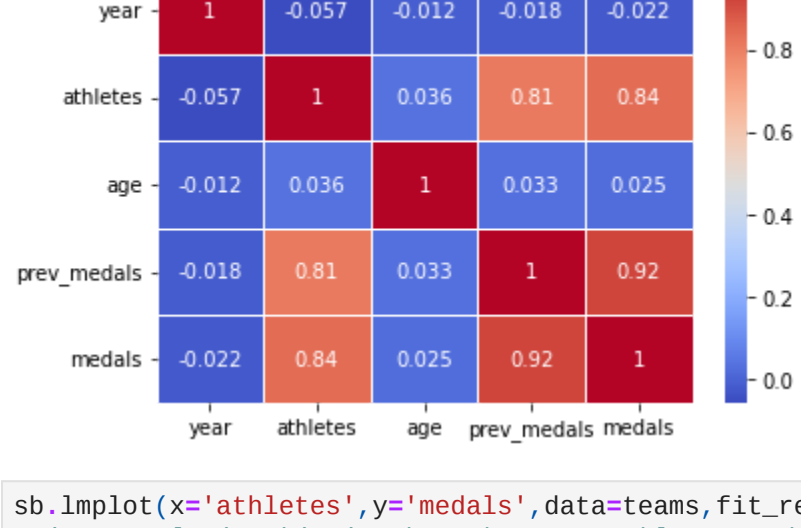
Out[13]:
   year  athletes  age  prev_medals  medals
year    1.000000  -0.056957 -0.012179 -0.017917 -0.021603
athletes -0.056957  1.000000  0.036358  0.810576  0.840817
age      -0.012179  0.036358  1.000000  0.032596  0.025096
prev_medals -0.017917  0.810576  0.032596  1.000000  0.920048
medals   -0.021603  0.840817  0.025096  0.920048  1.000000

In [14]: teams.corr()[['medals']]

Out[14]:
year      -0.021603
athletes  0.840817
age        0.025096
prev_medals 0.920048
medals     1.000000
Name: medals, dtype: float64

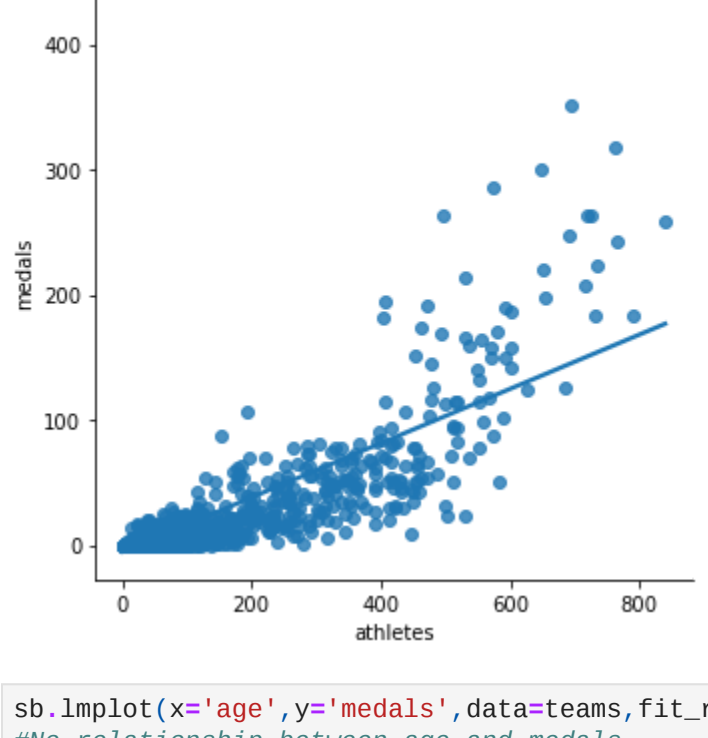
In [15]: sb.heatmap(teams.corr(), annot=True, cmap='coolwarm', linewidths=0.5)

Out[15]:
<AxesSubplot:~>



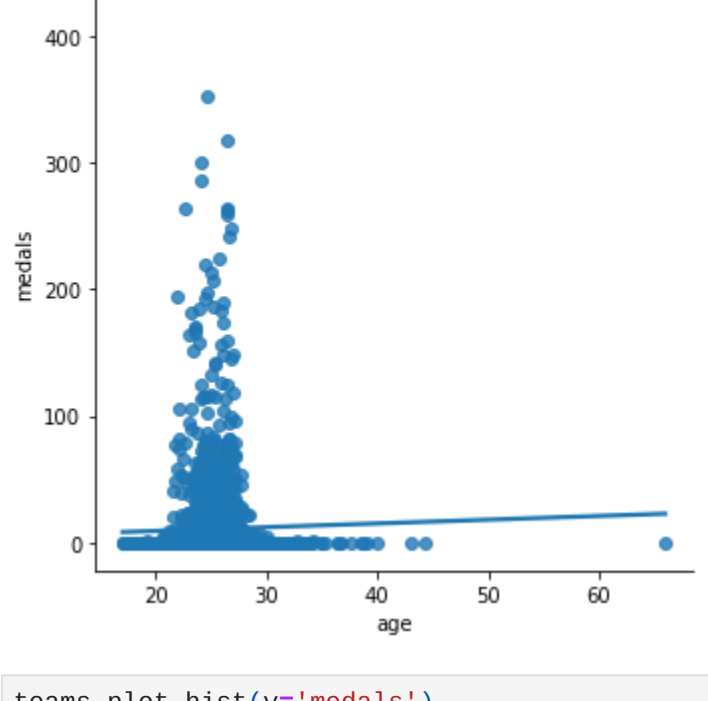
In [16]: sb.lmplot(x='athletes',y='medals',data=teams,fit_reg=True,ci=None)
#Linear Relationship is there between athletes and medals

Out[16]:
<seaborn.axisgrid.FacetGrid at 0x26466e8dc0>



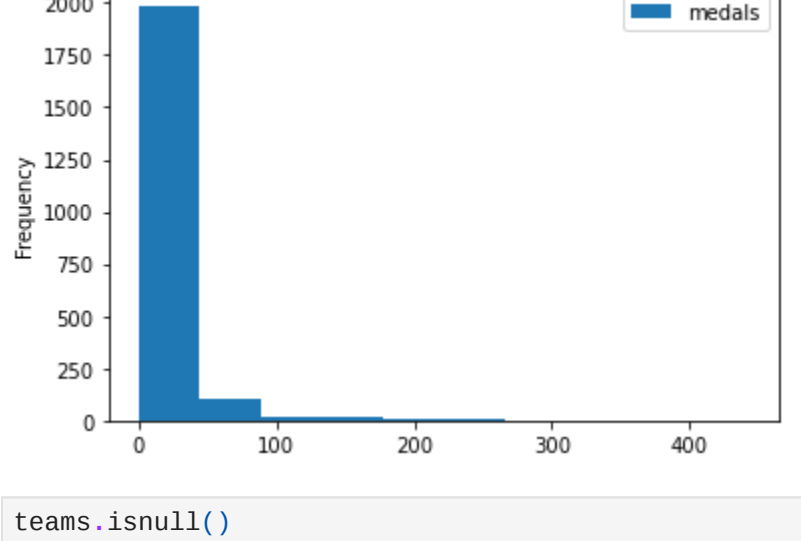
In [17]: sb.lmplot(x='age',y='medals',data=teams,fit_reg=True,ci=None)
#No Relationship between age and medals

Out[17]:
<seaborn.axisgrid.FacetGrid at 0x26466f69700>



In [18]: teams.plot.hist(y='medals')
#yaha bataya hai 0 to 100 ke bich medals bohot mile , 200 to 300 bohot kam mile hai

Out[18]:
<AxesSubplot:ylabel='Frequency'~>



In [19]: teams.isnull()

Out[19]:
   team  country  year  athletes  age  prev_medals  medals
0  False  False  False  False  False  False  False
1  False  False  False  False  False  False  False
2  False  False  False  False  False  False  False
3  False  False  False  False  False  False  False
4  False  False  False  False  False  False  False
...   ...      ...      ...      ...      ...      ...
2139 False  False  False  False  False  False  False
2140 False  False  False  False  False  False  False
2141 False  False  False  False  False  False  False
2142 False  False  False  False  False  False  False
2143 False  False  False  False  False  False  False

2144 rows x 7 columns

In [20]: teams.isnull().sum()

Out[20]:
team          0
country       0
year          0
athletes      0
age           0
prev_medals   130
medals        0
dtype: int64

In [21]: teams[teams.isnull().any(axis=1)]

Out[21]:
   team  country  year  athletes  age  prev_medals  medals
19  ALB  Albania  1992      9  25.3          NaN      0
26  ALG  Algeria  1964      7  26.0          NaN      0
39  AND  Andorra  1976      3  28.3          NaN      0
50  ANG  Angola   1980     17  17.4          NaN      0
59  ANT  Antigua and Barbuda  1976     17  23.2          NaN      0
...   ...      ...      ...      ...      ...      ...
2092 VIN  Saint Vincent and the Grenadines  1988      6  20.5          NaN      0
2103 YAR  North Yemen  1984      3  27.7          NaN      0
2105 YEM  Yemen        1992      8  19.6          NaN      0
2112 YMD  South Yemen  1988      5  23.6          NaN      0
2120 ZAM  Zambia      1964     15  21.7          NaN      0

130 rows x 7 columns

In [22]: teams=teams.dropna()

In [23]: teams

Out[23]:
   team  country  year  athletes  age  prev_medals  medals
0  AFG  Afghanistan  1964      8  22.0          0.0      0
1  AFG  Afghanistan  1968      5  23.2          0.0      0
2  AFG  Afghanistan  1972      8  29.0          0.0      0
3  AFG  Afghanistan  1980     11  23.6          0.0      0
4  AFG  Afghanistan  2004      5  18.6          0.0      0
...   ...      ...      ...      ...      ...      ...
2139 ZIM  Zimbabwe  2000     26  25.0          0.0      0
2140 ZIM  Zimbabwe  2004     14  25.1          0.0      3
2141 ZIM  Zimbabwe  2008     16  26.1          3.0      4
2142 ZIM  Zimbabwe  2012      9  27.3          4.0      0
2143 ZIM  Zimbabwe  2016     31  27.5          0.0      0

2014 rows x 7 columns

In [24]: teams.isnull().sum()
#Now there are no null values

Out[24]:
team          0
country       0
year          0
athletes      0
age           0
prev_medals   0
medals        0
dtype: int64

In [25]: # train=teams[teams["year"]<2012].copy()
# test=teams[teams["year"]>=2012].copy()

In [35]: # test.shape

Out[35]:
(405, 7)

In [27]: from sklearn.linear_model import LinearRegression #importing linearRegree algo
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error,r2_score

In [28]: model=LinearRegression() #Initialization of model

In [29]: # predictors=["athletes","prev_medals"]
# target="medals"

x=teams[['athletes','prev_medals']]
y=teams['medals']

In [30]: X_train,X_test,y_train,y_test =train_test_split(X,y, test_size=0.3 ,random_state=42)

In [ ]: model=LinearRegression()

In [32]: model.fit(X_train,y_train)

Out[32]:
LinearRegression()

In [34]: y_pred=model.predict(X_test)

In [35]: mse=mean_squared_error(y_test,y_pred)
r2=r2_score(y_test,y_pred)

In [40]: print(f"Mean Squared Error:{mse}")
print(f"R2 Score: {r2}")

Mean Squared Error:124.7719117695567
R2 Score: 0.8392615290167755

In [63]: future_years = pd.DataFrame({ # <- Correct way to create a DataFrame
'athletes': [117], # Future years you want to predict for
'prev_medals': [7] # Estimated number of athletes for those years
})

In [67]: future_prediction = model.predict(future_years)

In [69]: print(f"Predicted Medals for 2024: {future_prediction[0]:.2f}")

Predicted Medals for 2024: 11.32
```