```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

teams=pd.read_csv("teams.csv")

teams

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

      prev_medals  prev_3_medals  
0             0.0            0.0  
1             0.0            0.0  
2             0.0            0.0  
3             0.0            0.0  
4             0.0            0.0  
...           ...            ...  
2139          0.0            0.0  
2140          0.0            0.0  
2141          3.0            1.0  
2142          4.0            2.3  
2143          0.0            2.3  

[2144 rows x 11 columns]

teams=teams[["team",
"country","year","athletes","age","prev_medals","medals"]]
```

```
teams
```

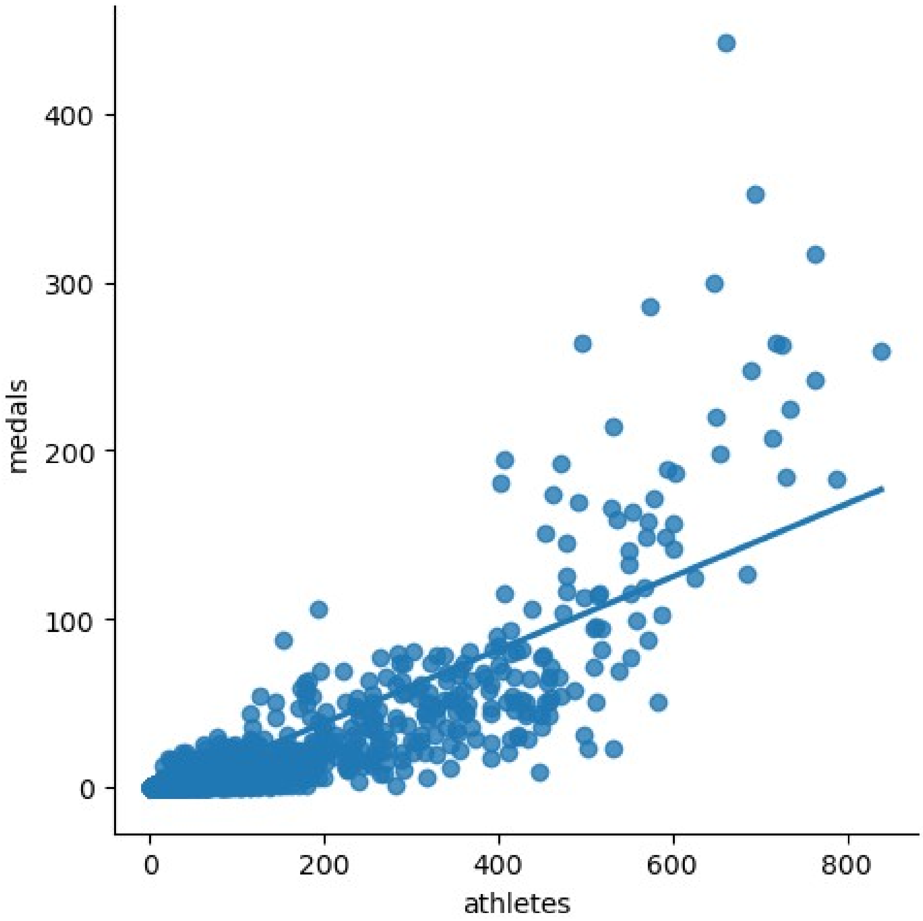
```
     team      country  year  athletes   age  prev_medals  medals
0     AFG  Afghanistan  1964         8  22.0          0.0       0
1     AFG  Afghanistan  1968         5  23.2          0.0       0
2     AFG  Afghanistan  1972         8  29.0          0.0       0
3     AFG  Afghanistan  1980        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]
```

```
import seaborn as sns
```

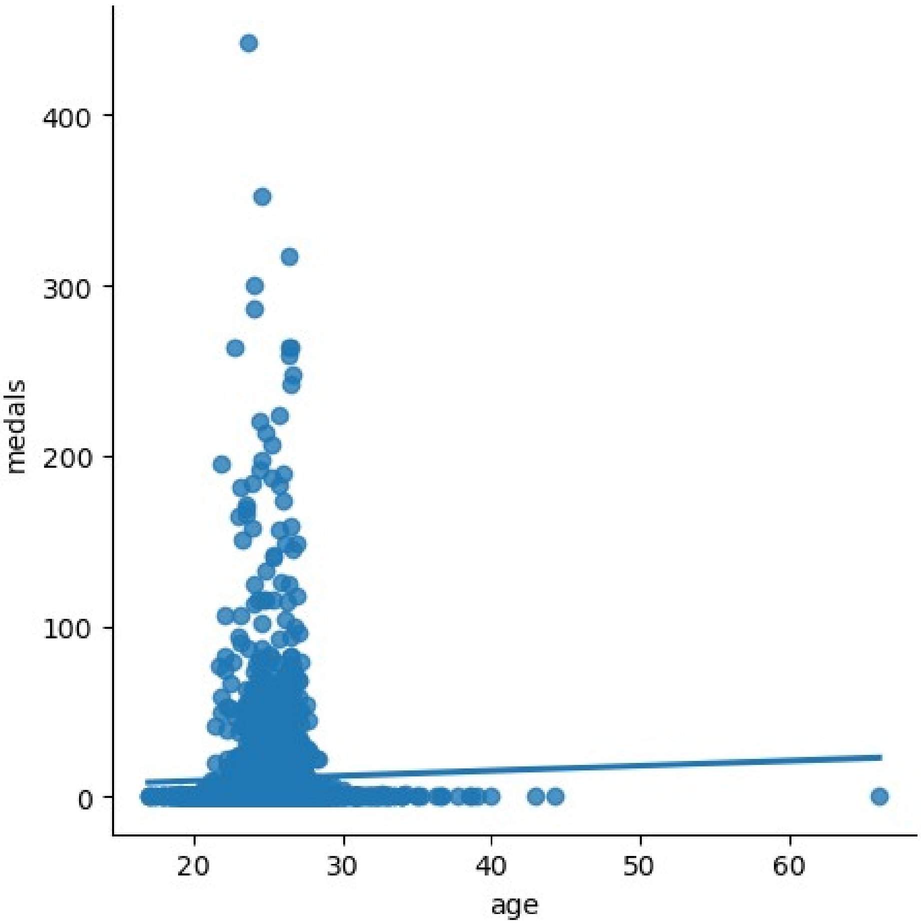
```
sns.lmplot(x="athletes",y="medals",data=teams,fit_reg=True,ci=None)
```

```
<seaborn.axisgrid.FacetGrid at 0x197103ebb10>
```



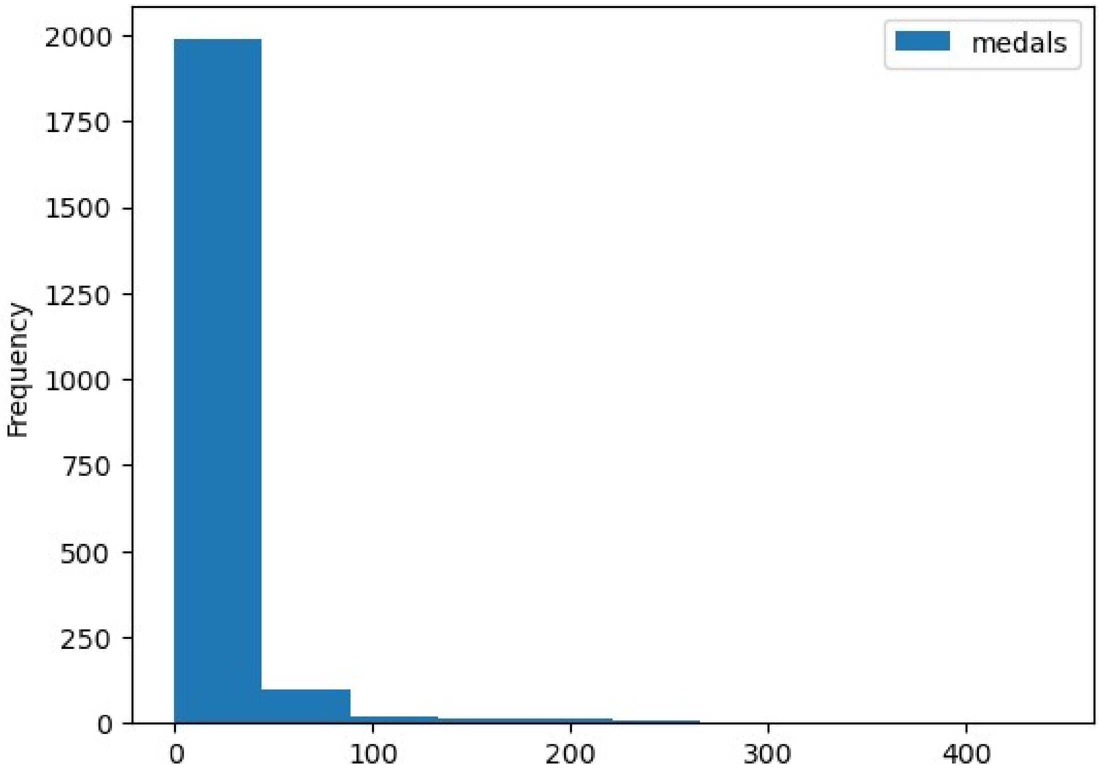
```
sns.lmplot(x="age",y="medals",data=teams,fit_reg=True,ci=None)
```

```
<seaborn.axisgrid.FacetGrid at 0x197104360d0>
```



```
 teams.plot.hist(y="medals")
```

```
<Axes: ylabel='Frequency'>
```



```
teams[teams.isnull().any(axis=1)]
```

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

      prev_medals  medals
19            NaN       0
26            NaN       0
39            NaN       0
50            NaN       0
59            NaN       0
...           ...     ...
2092          NaN       0
2103          NaN       0
2105          NaN       0
```

```
2112          NaN       0
2120          NaN       0

[130 rows x 7 columns]
```

```python
teams=teams.dropna()
```

```python
teams
```

```
     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]
```

```python
train=teams[teams["year"]<2012].copy()
test=teams[teams["year"]>=2012].copy()
```

```python
train.shape
```

```
(1609, 7)
```

```python
test.shape
```

```
(405, 7)
```

```python
from sklearn.linear_model import LinearRegression
reg=LinearRegression()
```

```python
predictors=["athletes","prev_medals"]
target="medals"
```

```python
reg.fit(train[predictors],train["medals"])
```

```
LinearRegression()
```

```python
predictions=reg.predict(test[predictors])
```

```python
test["predictions"]=predictions
```

```python
test
```

```
     team      country  year  athletes   age  prev_medals  medals
predictions
```

```
6     AFG  Afghanistan  2012         6  24.8          1.0       1  -
0.961221
7     AFG  Afghanistan  2016         3  24.7          1.0       0  -
1.176333
24    ALB      Albania  2012        10  25.7          0.0       0  -
1.425032
25    ALB      Albania  2016         6  23.7          0.0       0  -
1.711847
37    ALG      Algeria  2012        39  24.8          2.0       1   
2.155629
...   ...          ...   ...       ...   ...          ...     ...   
...
2111  YEM        Yemen  2016         3  19.3          0.0       0  -
1.926958
2131  ZAM       Zambia  2012         7  22.6          0.0       0  -
1.640143
2132  ZAM       Zambia  2016         7  24.1          0.0       0  -
1.640143
2142  ZIM     Zimbabwe  2012         9  27.3          4.0       0   
1.505767
2143  ZIM     Zimbabwe  2016        31  27.5          0.0       0   
0.080748

[405 rows x 8 columns]
```

```
test.loc[test["predictions"]<0,"predictions"]=0
```

```
test["predictions"]=test["predictions"].round()
```

```
test
```

```
     team      country  year  athletes   age  prev_medals  medals  
predictions
6     AFG  Afghanistan  2012         6  24.8          1.0       1  
0.0
7     AFG  Afghanistan  2016         3  24.7          1.0       0  
0.0
24    ALB      Albania  2012        10  25.7          0.0       0  
0.0
25    ALB      Albania  2016         6  23.7          0.0       0  
0.0
37    ALG      Algeria  2012        39  24.8          2.0       1  
2.0
...   ...          ...   ...       ...   ...          ...     ...  
...
2111  YEM        Yemen  2016         3  19.3          0.0       0  
0.0
2131  ZAM       Zambia  2012         7  22.6          0.0       0  
0.0
2132  ZAM       Zambia  2016         7  24.1          0.0       0  
```

```
0.0
2142  ZIM     Zimbabwe  2012            9  27.3           4.0         0
2.0
2143  ZIM     Zimbabwe  2016           31  27.5           0.0         0
0.0

[405 rows x 8 columns]
```

```python
from sklearn.metrics import mean_absolute_error
error=mean_absolute_error(test["medals"],test["predictions"])
```

```python
error
```

```
3.2987654320987656
```

```python
teams.describe()["medals"]
```

```
count    2014.000000
mean       10.990070
std        33.627528
min         0.000000
25%         0.000000
50%         0.000000
75%         5.000000
max       442.000000
Name: medals, dtype: float64
```

```python
test[test["team"]=="USA"]
```

```
     team        country  year  athletes   age  prev_medals  medals  \
2053  USA  United States  2012       689  26.7        317.0     248
2054  USA  United States  2016       719  26.4        248.0     264

      predictions
2053        285.0
2054        236.0
```

```python
test[test["team"]=="IND"]
```

```
    team country  year  athletes   age  prev_medals  medals
predictions
907  IND   India  2012        95  26.0          3.0       6
7.0
908  IND   India  2016       130  26.1          6.0       2
12.0
```

```python
errors=(test["medals"]-test["predictions"]).abs()
```

```python
errors
```

```
6       1.0
7       0.0
```

```
24      0.0
25      0.0
37      1.0
       ...
2111    0.0
2131    0.0
2132    0.0
2142    2.0
2143    0.0
Length: 405, dtype: float64
```

```
error_by_team=errors.groupby(test["team"]).mean()
```

```
error_by_team
```

```
team
AFG    0.5
ALB    0.0
ALG    1.5
AND    0.0
ANG    0.0
      ...
VIE    1.0
VIN    0.0
YEM    0.0
ZAM    0.0
ZIM    1.0
Length: 204, dtype: float64
```

```
medals_by_team=test["medals"].groupby(test["team"]).mean()
```

```
error_ratio=error_by_team/medals_by_team
```

```
error_ratio
```

```
team
AFG    1.0
ALB    NaN
ALG    1.0
AND    NaN
ANG    NaN
      ...
VIE    1.0
VIN    NaN
YEM    NaN
ZAM    NaN
ZIM    inf
Length: 204, dtype: float64
```

```
error_ratio[~pd.isnull(error_ratio)]
```

```
team
AFG    1.000000
ALG    1.000000
ARG    0.853659
ARM    0.428571
AUS    0.367347
         ...
USA    0.126953
UZB    0.625000
VEN    1.750000
VIE    1.000000
ZIM         inf
Length: 102, dtype: float64
```

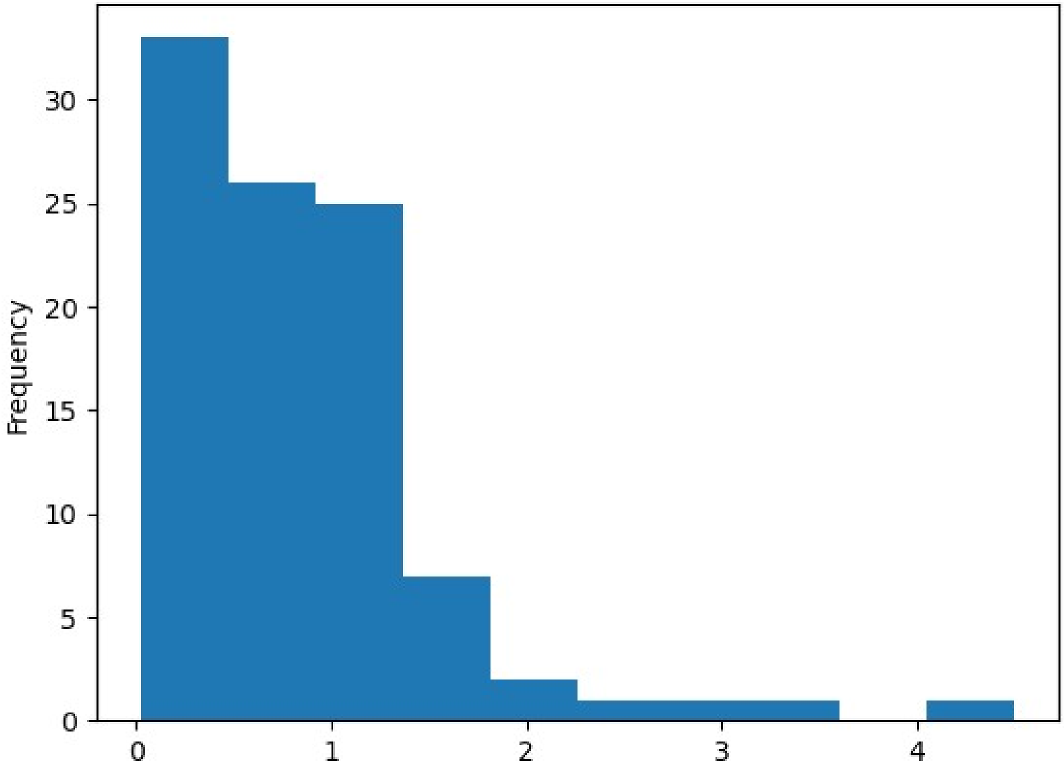
```
import numpy as np
error_ratio=error_ratio[np.isfinite(error_ratio)]
```

```
error_ratio
```

```
team
AFG    1.000000
ALG    1.000000
ARG    0.853659
ARM    0.428571
AUS    0.367347
         ...
UKR    0.951220
USA    0.126953
UZB    0.625000
VEN    1.750000
VIE    1.000000
Length: 97, dtype: float64
```

```
error_ratio.plot.hist()
```

```
<Axes: ylabel='Frequency'>
```



```
error_ratio.sort_values()

team
FRA    0.022472
CAN    0.048387
NZL    0.063492
RUS    0.082353
ITA    0.121429
         ...
MAR    2.000000
EGY    2.400000
HKG    3.000000
POR    3.333333
AUT    4.500000
Length: 97, dtype: float64
```