In [1]:
```python
import pandas as pd
```

In [2]:
```python
teams= pd.read_csv("teams.csv")
```

In [3]:
```python
teams
```

Out[3]:

	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 × 11 columns

In [4]:
```python
teams= teams[["team", "country", "year", "athletes", "age", "prev_medals", "medals"]]
```

In [5]:
```python
teams
```

Out[5]:

	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 × 7 columns

In [9]:
```python
correlation = teams["medals"].corr(teams["year"])
print(correlation)
```

```
-0.021603233614383134
```

In [11]:
```python
correlation = teams["medals"].corr(teams["athletes"])
print(correlation)
```

```
0.8408174672847653
```

In [12]:
```python
correlation = teams["medals"].corr(teams["age"])
print(correlation)
```

```
0.02509580254295867
```

In [13]:
```python
correlation = teams["medals"].corr(teams["prev_medals"])
print(correlation)
```

```
0.9200483125630238
```

In [14]:
```python
correlation = teams["medals"].corr(teams["medals"])
print(correlation)
```

```
1.0
```

In [16]:
```python
pip install seaborn
```

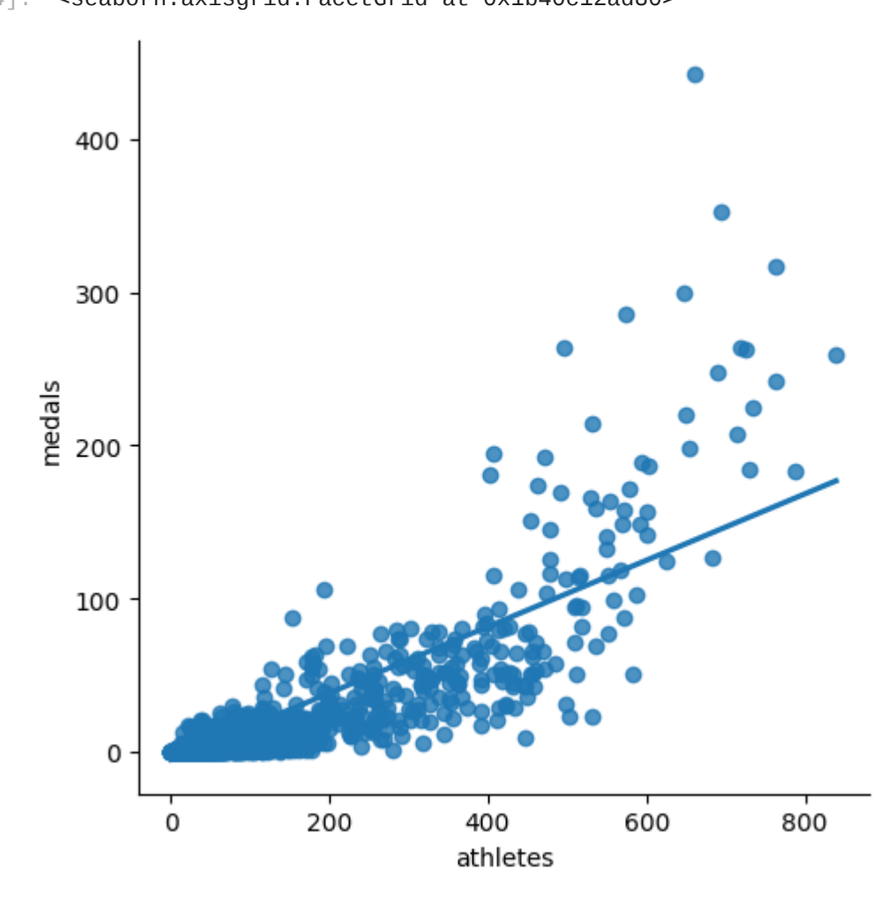
```
Collecting seabornNote: you may need to restart the kernel to use updated packages.

  Obtaining dependency information for seaborn from https://files.pythonhosted.org/packages/7b/e5/83fcd7e9db036c179e0352bfcd20f81d728197a16f883e7b90307a88e65e/seaborn-0.13.0-py3-none-any.whl.metadata
  Downloading seaborn-0.13.0-py3-none-any.whl.metadata (5.3 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (1.26.2)
Requirement already satisfied: pandas>=1.2 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (2.1.3)
Collecting matplotlib!=3.6.1,>=3.3 (from seaborn)
  Obtaining dependency information for matplotlib!=3.6.1,>=3.3 from https://files.pythonhosted.org/packages/2e/51/c77a14869b7eb9d6fb440e811b754fc3950d6868c38ace57d0632b674415/matplotlib-3.8.2-cp312-cp312-win_amd64.whl.metadata
  Downloading matplotlib-3.8.2-cp312-cp312-win_amd64.whl.metadata (5.9 kB)
Collecting contourpy>=1.0.1 (from matplotlib!=3.6.1,>=3.3->seaborn)
  Obtaining dependency information for contourpy>=1.0.1 from https://files.pythonhosted.org/packages/8e/ae/a6353db548bff1a592b85ae6bb80275f0a51dc25a0410d059e5b33183e36/contourpy-1.2.0-cp312-cp312-win_amd64.whl.metadata
  Downloading contourpy-1.2.0-cp312-cp312-win_amd64.whl.metadata (5.8 kB)
Collecting cycler>=0.10 (from matplotlib!=3.6.1,>=3.3->seaborn)
  Obtaining dependency information for cycler>=0.10 from https://files.pythonhosted.org/packages/e7/05/c19819d5e3d95294a6f5947fb9b9629efb316b96de511b418c53d245aae6/cycler-0.12.1-py3-none-any.whl.metadata
  Downloading cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting fonttools>=4.22.0 (from matplotlib!=3.6.1,>=3.3->seaborn)
  Obtaining dependency information for fonttools>=4.22.0 from https://files.pythonhosted.org/packages/71/00/46562cbdf644aeaa20666d920e2d6b71a0096ff987ef878ec510e4c1d886/fonttools-4.45.1-cp312-cp312-win_amd64.whl.metadata
  Downloading fonttools-4.45.1-cp312-cp312-win_amd64.whl.metadata (158 kB)
     ---------------------------------------- 0.0/158.4 kB ? eta -:--:--
     ---------- ----------------------------- 41.0/158.4 kB 991.0 kB/s eta 0:00:01
     ---------------------------------------- 158.4/158.4 kB 2.4 MB/s eta 0:00:00
Collecting kiwisolver>=1.3.1 (from matplotlib!=3.6.1,>=3.3->seaborn)
  Obtaining dependency information for kiwisolver>=1.3.1 from https://files.pythonhosted.org/packages/63/50/2746566bdf4a6a842d117367d05c90cfb87ac04e9e2845aa1fa21f071362/kiwisolver-1.4.5-cp312-cp312-win_amd64.whl.metadata
  Downloading kiwisolver-1.4.5-cp312-cp312-win_amd64.whl.metadata (6.5 kB)
Requirement already satisfied: packaging>=20.0 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.3->seaborn) (23.2)
Collecting pillow>=8 (from matplotlib!=3.6.1,>=3.3->seaborn)
  Obtaining dependency information for pillow>=8 from https://files.pythonhosted.org/packages/32/e4/978865107d097dd9cb650331676d8dc29ed9fcd0aaab46486e9d6e5123f0/Pillow-10.1.0-cp312-cp312-win_amd64.whl.metadata
  Downloading Pillow-10.1.0-cp312-cp312-win_amd64.whl.metadata (9.6 kB)
Collecting pyparsing>=2.3.1 (from matplotlib!=3.6.1,>=3.3->seaborn)
  Obtaining dependency information for pyparsing>=2.3.1 from https://files.pythonhosted.org/packages/39/92/8486ede85fcc088f1b3dba4ce92dd29d126fd96b0008ea213167940a2475/pyparsing-3.1.1-py3-none-any.whl.metadata
  Downloading pyparsing-3.1.1-py3-none-any.whl.metadata (5.1 kB)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.3->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\genet\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.3->seaborn) (1.16.0)
Downloading seaborn-0.13.0-py3-none-any.whl (294 kB)
   ---------------------------------------- 0.0/294.6 kB ? eta -:--:--
   ----------------------------------- ---- 256.0/294.6 kB 5.3 MB/s eta 0:00:01
   ------------------------------------- -- 286.7/294.6 kB 4.5 MB/s eta 0:00:01
   ---------------------------------------- 294.6/294.6 kB 3.1 MB/s eta 0:00:00
Downloading matplotlib-3.8.2-cp312-cp312-win_amd64.whl (7.6 MB)
   ---------------------------------------- 0.0/7.6 MB ? eta -:--:--
   -- ------------------------------------- 0.4/7.6 MB 8.1 MB/s eta 0:00:01
   ----- ---------------------------------- 1.1/7.6 MB 11.3 MB/s eta 0:00:01
   ------------ --------------------------- 2.4/7.6 MB 17.1 MB/s eta 0:00:01
   ------------------- -------------------- 3.6/7.6 MB 19.3 MB/s eta 0:00:01
   ---------------------------- ----------- 5.4/7.6 MB 23.1 MB/s eta 0:00:01
   -------------------------------------- - 7.3/7.6 MB 25.9 MB/s eta 0:00:01
   ---------------------------------------- 7.6/7.6 MB 24.5 MB/s eta 0:00:00
Downloading contourpy-1.2.0-cp312-cp312-win_amd64.whl (187 kB)
   ---------------------------------------- 0.0/187.7 kB ? eta -:--:--
   ---------------------------------------- 187.7/187.7 kB 11.1 MB/s eta 0:00:00
Downloading cycler-0.12.1-py3-none-any.whl (8.3 kB)
Downloading fonttools-4.45.1-cp312-cp312-win_amd64.whl (2.1 MB)
   ---------------------------------------- 0.0/2.1 MB ? eta -:--:--
   ------------------------------ --------- 1.6/2.1 MB 52.5 MB/s eta 0:00:01
   ---------------------------------------  2.1/2.1 MB 33.8 MB/s eta 0:00:01
   ---------------------------------------- 2.1/2.1 MB 27.4 MB/s eta 0:00:00
Downloading kiwisolver-1.4.5-cp312-cp312-win_amd64.whl (56 kB)
   ---------------------------------------- 0.0/56.0 kB ? eta -:--:--
   ---------------------------------------- 56.0/56.0 kB 2.9 MB/s eta 0:00:00
Downloading Pillow-10.1.0-cp312-cp312-win_amd64.whl (2.6 MB)
   ---------------------------------------- 0.0/2.6 MB ? eta -:--:--
   ----------------------------- ---------- 1.9/2.6 MB 39.9 MB/s eta 0:00:01
   ---------------------------------------- 2.6/2.6 MB 33.3 MB/s eta 0:00:00
Downloading pyparsing-3.1.1-py3-none-any.whl (103 kB)
   ---------------------------------------- 0.0/103.1 kB ? eta -:--:--
   ---------------------------------------- 103.1/103.1 kB ? eta 0:00:00
Installing collected packages: pyparsing, pillow, kiwisolver, fonttools, cycler, contourpy, matplotlib, seaborn
Successfully installed contourpy-1.2.0 cycler-0.12.1 fonttools-4.45.1 kiwisolver-1.4.5 matplotlib-3.8.2 pillow-10.1.0 pyparsing-3.1.1 seaborn-0.13.0
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [17]:
```python
import seaborn as sns
```

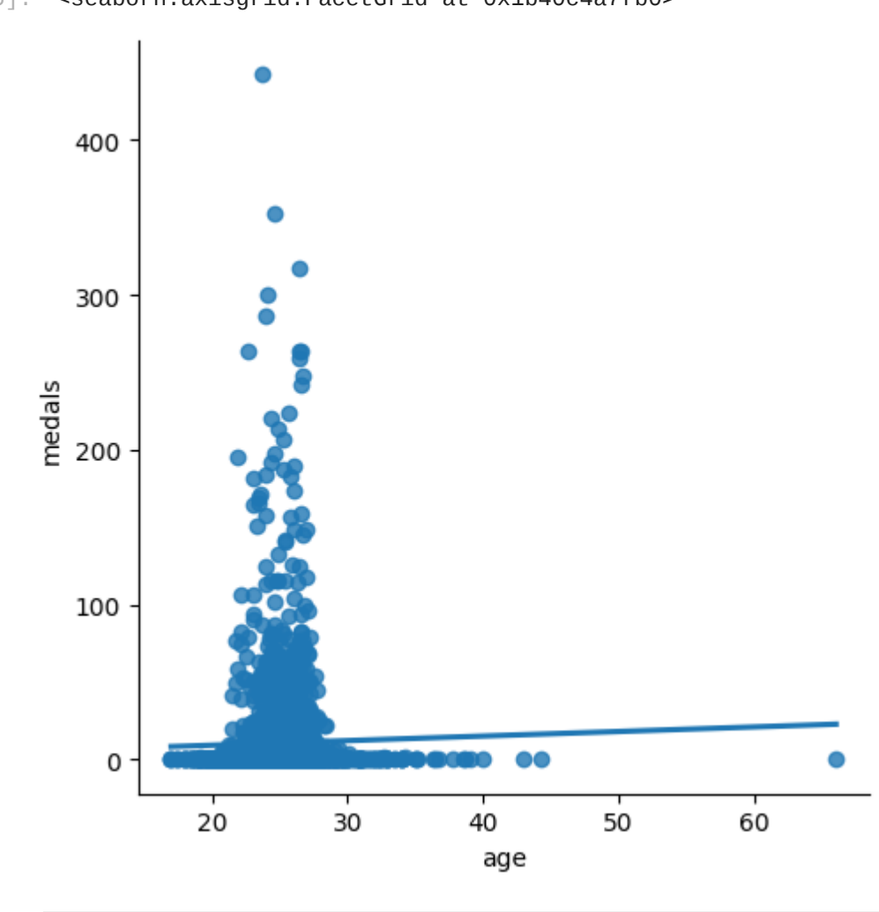
In [24]:
```python
sns.lmplot(x="athletes", y="medals", data=teams, fit_reg=True, ci=None)
```

Out[24]: <seaborn.axisgrid.FacetGrid at 0x1b40c12ad80>



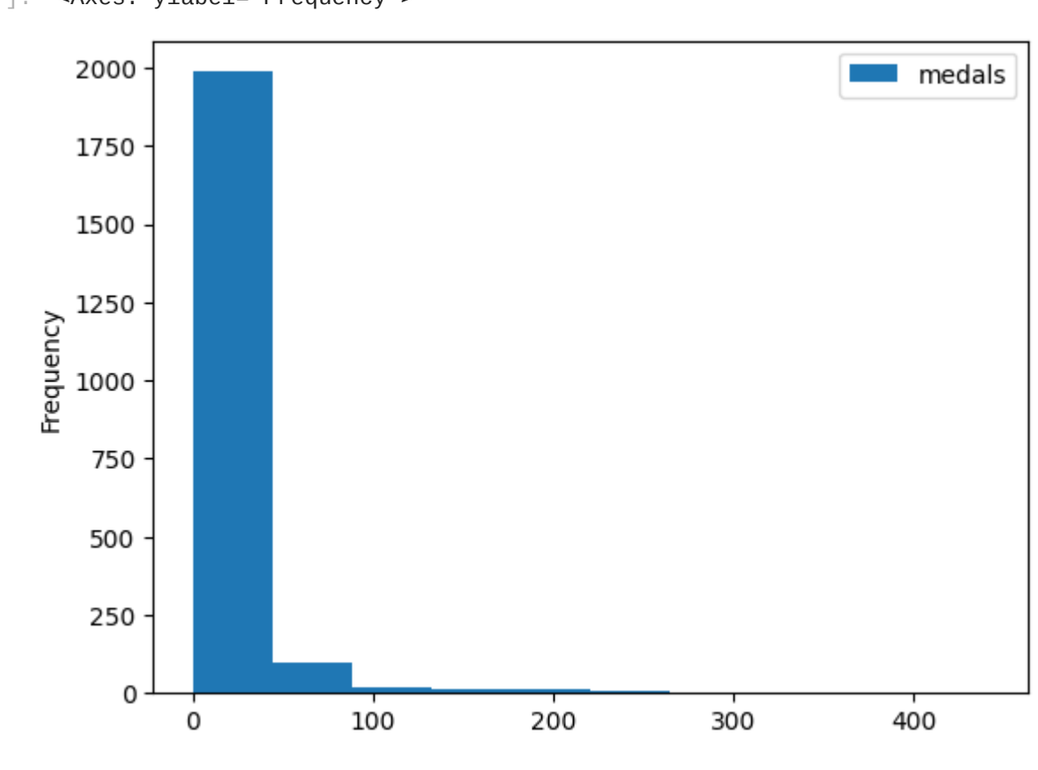
In [25]:
```python
sns.lmplot(x="age", y="medals", data=teams, fit_reg=True, ci=None)
```

Out[25]: <seaborn.axisgrid.FacetGrid at 0x1b40c4a7fb0>



In [27]:
```python
teams.plot.hist(y="medals")
```

Out[27]: <Axes: ylabel='Frequency'>



In [28]:
```python
teams[teams.isnull().any(axis=1)]
```

Out[28]:

	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 × 7 columns

In [29]:
```python
teams=teams.dropna()
```

In [30]:
```python
teams
```

Out[30]:

	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 × 7 columns

In [66]:
```python
 train=teams[teams["year"]<2012].copy()
 newtest=teams[teams["year"]>=2012].copy()
```

In [67]:
```python
train.shape
```

Out[67]: (1609, 7)

In [70]:
```python
newtest.shape
```

Out[70]: (405, 7)

In [64]:
```python
train
```

Out[64]:

	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
...	...	...	...	...	...	...	...
2137	ZIM	Zimbabwe	1992	28	21.2	0.0	0
2138	ZIM	Zimbabwe	1996	21	23.8	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

1609 rows × 7 columns

In [71]:
```python
newtest
```

Out[71]:

	team	country	year	athletes	age	prev_medals	medals
6	AFG	Afghanistan	2012	6	24.8	1.0	1
7	AFG	Afghanistan	2016	3	24.7	1.0	0
24	ALB	Albania	2012	10	25.7	0.0	0
25	ALB	Albania	2016	6	23.7	0.0	0
37	ALG	Algeria	2012	39	24.8	2.0	1
...	...	...	...	...	...	...	...
2111	YEM	Yemen	2016	3	19.3	0.0	0
2131	ZAM	Zambia	2012	7	22.6	0.0	0
2132	ZAM	Zambia	2016	7	24.1	0.0	0
2142	ZIM	Zimbabwe	2012	9	27.3	4.0	0
2143	ZIM	Zimbabwe	2016	31	27.5	0.0	0

405 rows × 7 columns

In [73]:
```python
from sklearn.linear_model import LinearRegression
reg = LinearRegression()
```

In [84]:
```python
predictors = ["athletes", "prev_medals" ]
target = "medals"
```

In [85]:
```python
reg.fit(train[predictors], train["medals"])
```

Out[85]:

▾ LinearRegression

LinearRegression()

In [91]:
```python
predictions = reg.predict(newtest[predictors])
```

In [92]:
```python
newtest["predictions"] = predictions
```

In [93]:
```python
newtest
```

Out[93]:

	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 × 8 columns

In [104…
```python
newtest.loc[newtest["predictions"] < 0, "predictions"] = 0
```

In [105…
```python
newtest["predictions"] = newtest["predictions"].round()
```

In [106…
```python
newtest
```

Out[106…

	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 × 8 columns

In [107…
```python
from sklearn.metrics import mean_absolute_error

error = mean_absolute_error(newtest["medals"], newtest["predictions"])
```

In [108…
```python
error
```

Out[108… 3.2987654320987656

In [109…
```python
teams.describe () ["medals"]
```

Out[109…
```
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
```

In [111…
```python
newtest[newtest["team"] =="USA"]
```

Out[111…

	team	country	year	athletes	age	prev_medals	medals	predictions
2053	USA	United States	2012	689	26.7	317.0	248	285.0
2054	USA	United States	2016	719	26.4	248.0	264	236.0

In [112…
```python
newtest[newtest["team"] =="IND"]
```

Out[112…

	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

In [115…
```python
errors= (newtest["medals"] - newtest["predictions"]).abs ()
```

In [116…
```python
errors
```

Out[116…
```
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
```

In [117…
```python
error_by_team = errors.groupby(newtest["team"]).mean()
```

In [118…
```python
error_by_team
```

Out[118…
```
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
```

In [121…
```python
medals_by_team = newtest["medals"].groupby(newtest["team"]).mean()
```

```python
medals_by_team
```

In [123…
```python
error_ratio=error_by_team/medals_by_team
```

In [124…
```python
error_ratio
```

Out[124…
```
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
```

In [125…
```python
error_ratio[~pd.isnull(error_ratio)]
```

Out[125…
```
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
```

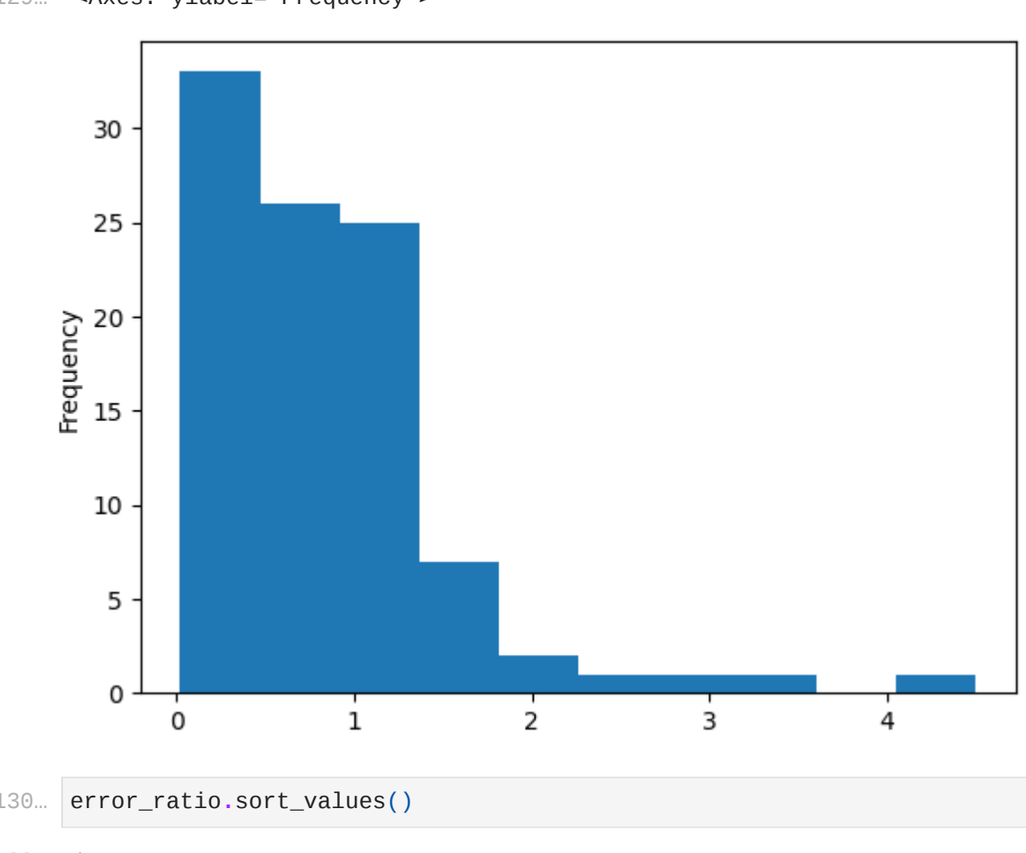
In [127…
```python
import numpy as np
error_ratio = error_ratio[np.isfinite(error_ratio)]
```

In [128…
```python
error_ratio
```

Out[128…
```
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
```

In [129…
```python
error_ratio.plot.hist()
```

Out[129… <Axes: ylabel='Frequency'>



In [130…
```python
error_ratio.sort_values()
```

Out[130…
```
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
```

In [ ]: