anthony_apwood_assignment2.ipynb

February 10, 2024

```python
[1]: import pandas as pd
     import matplotlib.pyplot as plt
```

```python
[2]: df = pd.read_csv("cleaned_star_data.csv")
     # I think there are no NA values, but dropping anyway.
     df.dropna(inplace=True)
     df
```

```
[2]:      Temperature (K)  Luminosity(L/Lo)  Radius(R/Ro)  Absolute magnitude(Mv)  \
     0               3068          0.002400        0.1700                   16.12
     1               3042          0.000500        0.1542                   16.60
     2               2600          0.000300        0.1020                   18.70
     3               2800          0.000200        0.1600                   16.65
     4               1939          0.000138        0.1030                   20.06
     ..               ...               ...           ...                     ...
     235            38940     374830.000000     1356.0000                   -9.93
     236            30839     834042.000000     1194.0000                  -10.63
     237             8829     537493.000000     1423.0000                  -10.73
     238             9235     404940.000000     1112.0000                  -11.23
     239            37882     294903.000000     1783.0000                   -7.80

          Star type Star color Spectral Class
     0            0        Red              M
     1            0        Red              M
     2            0        Red              M
     3            0        Red              M
     4            0        Red              M
     ..         ...        ...            ...
     235          5       Blue              O
     236          5       Blue              O
     237          5      White              A
     238          5      White              A
     239          5       Blue              O

     [240 rows x 7 columns]
```

```python
[3]: df[df.duplicated()]
```

[3]: Empty DataFrame
Columns: [Temperature (K), Luminosity(L/Lo), Radius(R/Ro), Absolute
magnitude(Mv), Star type, Star color, Spectral Class]
Index: []

No duplicated entries.

[4]: `df.describe()`

[4]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)
count	240.000000	240.000000	240.000000
mean	10497.462500	107188.361635	237.157781
std	9552.425037	179432.244940	517.155763
min	1939.000000	0.000080	0.008400
25%	3344.250000	0.000865	0.102750
50%	5776.000000	0.070500	0.762500
75%	15055.500000	198050.000000	42.750000
max	40000.000000	849420.000000	1948.500000

	Absolute magnitude(Mv)	Star type
count	240.000000	240.000000
mean	4.382396	2.500000
std	10.532512	1.711394
min	-11.920000	0.000000
25%	-6.232500	1.000000
50%	8.313000	2.500000
75%	13.697500	4.000000
max	20.060000	5.000000

[5]: `df.dtypes`

[5]:
```
Temperature (K)            int64
Luminosity(L/Lo)         float64
Radius(R/Ro)             float64
Absolute magnitude(Mv)   float64
Star type                  int64
Star color                object
Spectral Class            object
dtype: object
```

[6]: `df.head()`

[6]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)
0	3068	0.002400	0.1700	16.12
1	3042	0.000500	0.1542	16.60
2	2600	0.000300	0.1020	18.70
3	2800	0.000200	0.1600	16.65
4	1939	0.000138	0.1030	20.06

```
     Star type Star color Spectral Class
0            0       Red              M
1            0       Red              M
2            0       Red              M
3            0       Red              M
4            0       Red              M
```
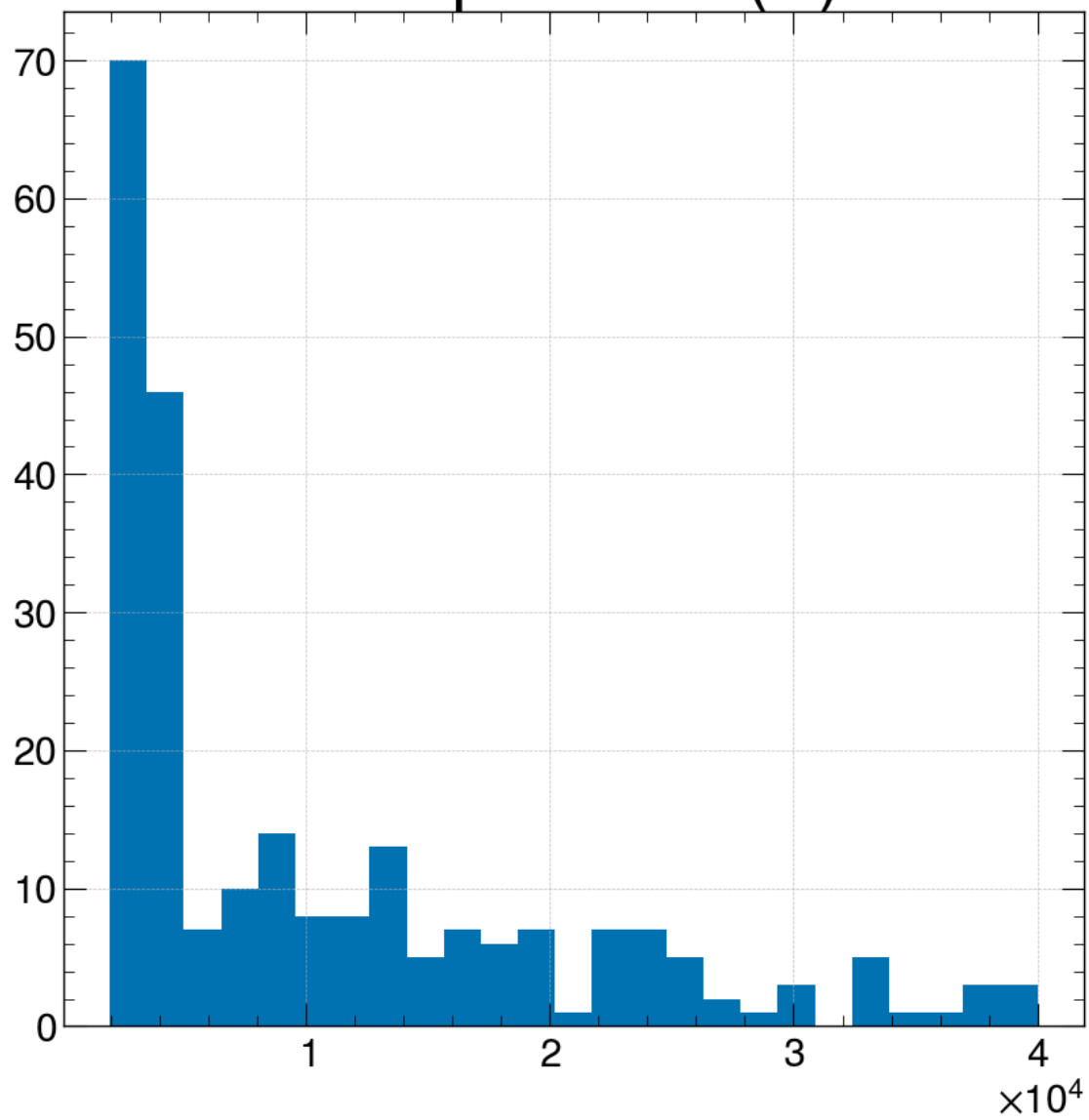
[7]: `df.tail()`

[7]:
```
     Temperature (K)  Luminosity(L/Lo)  Radius(R/Ro)  Absolute magnitude(Mv)  \
235            38940          374830.0        1356.0                   -9.93
236            30839          834042.0        1194.0                  -10.63
237             8829          537493.0        1423.0                  -10.73
238             9235          404940.0        1112.0                  -11.23
239            37882          294903.0        1783.0                   -7.80

     Star type Star color Spectral Class
235          5      Blue              O
236          5      Blue              O
237          5     White              A
238          5     White              A
239          5      Blue              O
```

[8]:
```python
df.hist("Temperature (K)", bins=25)

plt.show()
```

Temperature (K)

[ ]: