

Distributions

Two numerical columns

```
In [3]: import pandas as pd
```

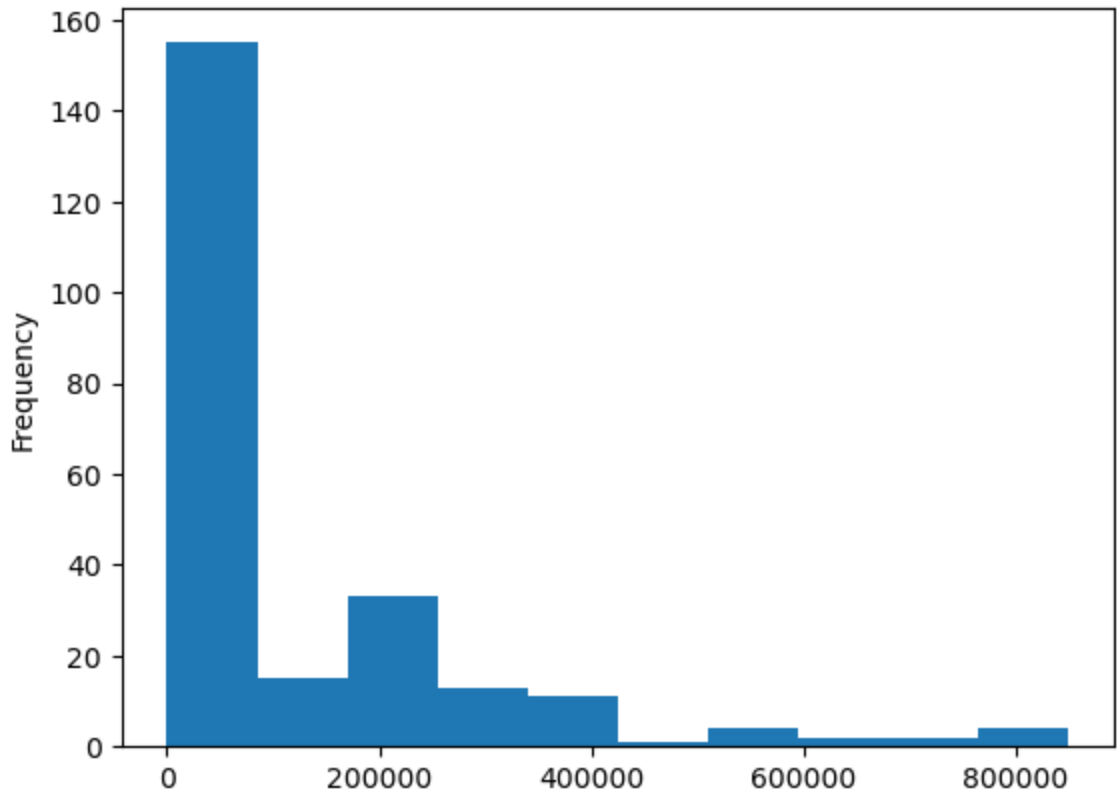
```
In [4]: data = pd.read_csv('star_dataset.csv')
data.head()
```

Out[4]:

	Temperature (K)	Luminosity(L/L _o)	Radius(R/R _o)	Absolute magnitude(M _v)	Star type	Star color	Spectral Class
0	3068	0.002400	0.1700	16.12	Red Dwarf	Red	M
1	3042	0.000500	0.1542	16.60	Red Dwarf	Red	M
2	2600	0.000300	0.1020	18.70	Red Dwarf	Red	M
3	2800	0.000200	0.1600	16.65	Red Dwarf	Red	M
4	1939	0.000138	0.1030	20.06	Red Dwarf	Red	M

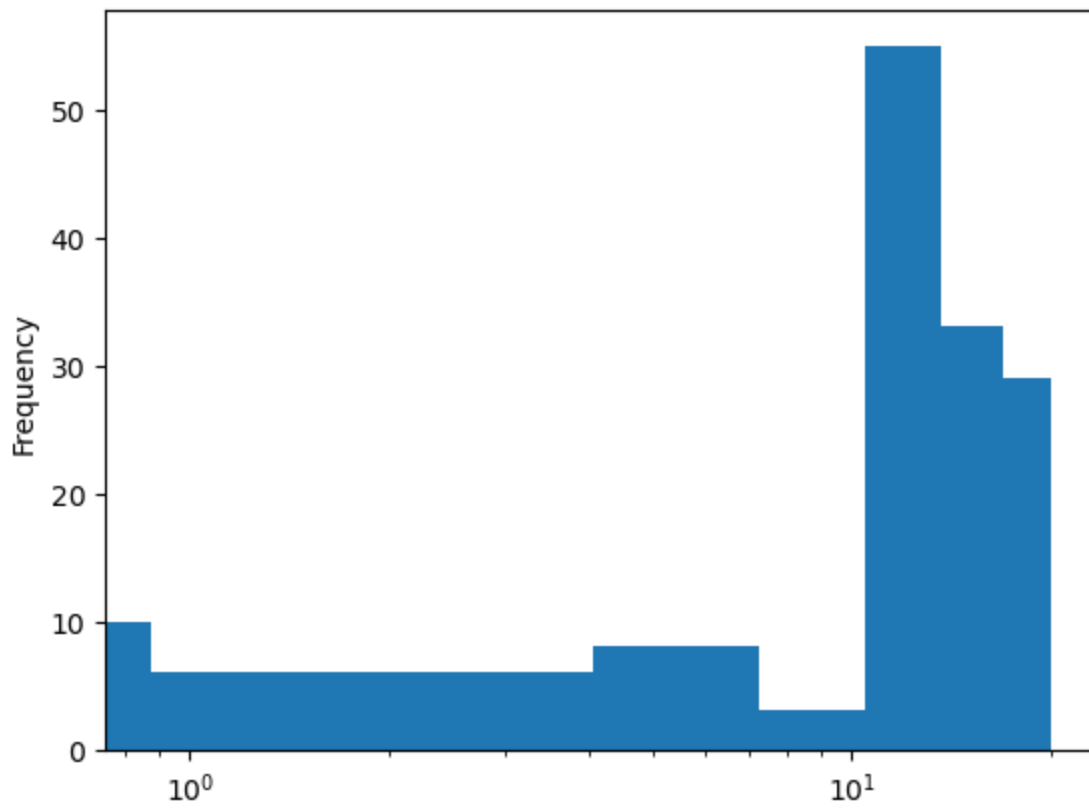
```
In [6]: data['Luminosity(L/Lo)'].plot(kind='hist', bins=10)
```

Out[6]: <AxesSubplot:ylabel='Frequency'>



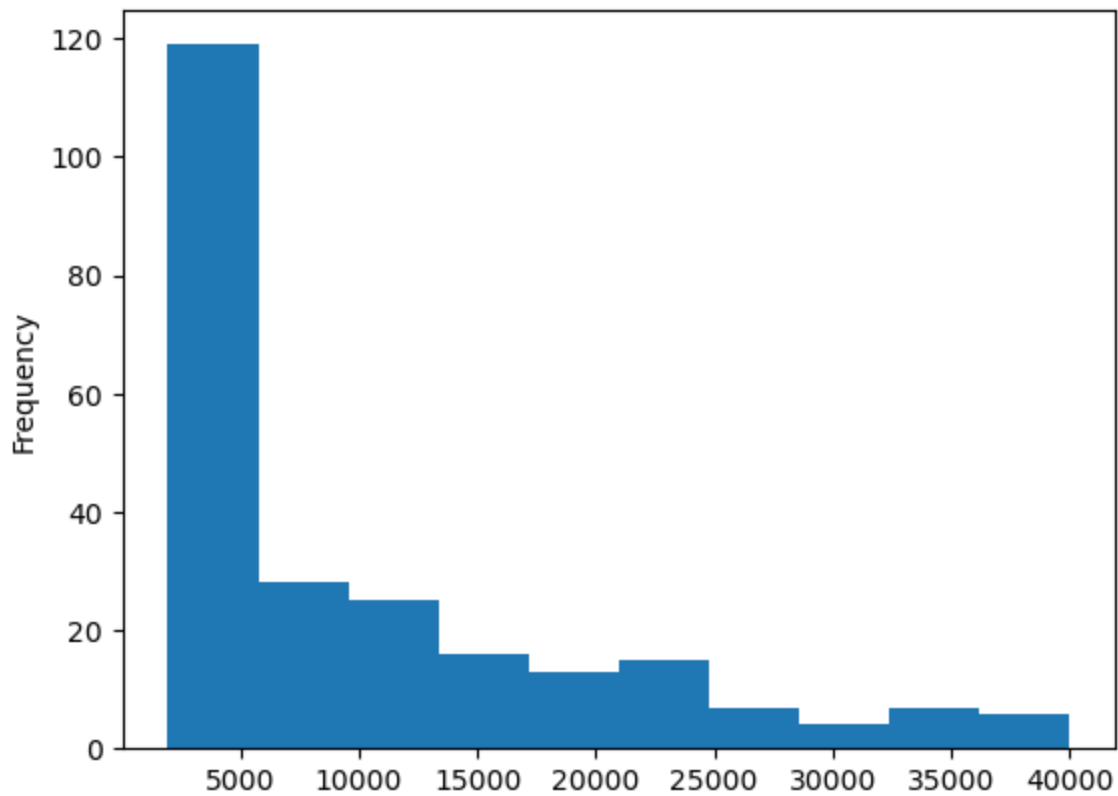
```
In [60]: data['Absolute magnitude(Mv)'].plot(kind='hist', bins=10, logx=True)
```

Out[60]: <AxesSubplot:ylabel='Frequency'>



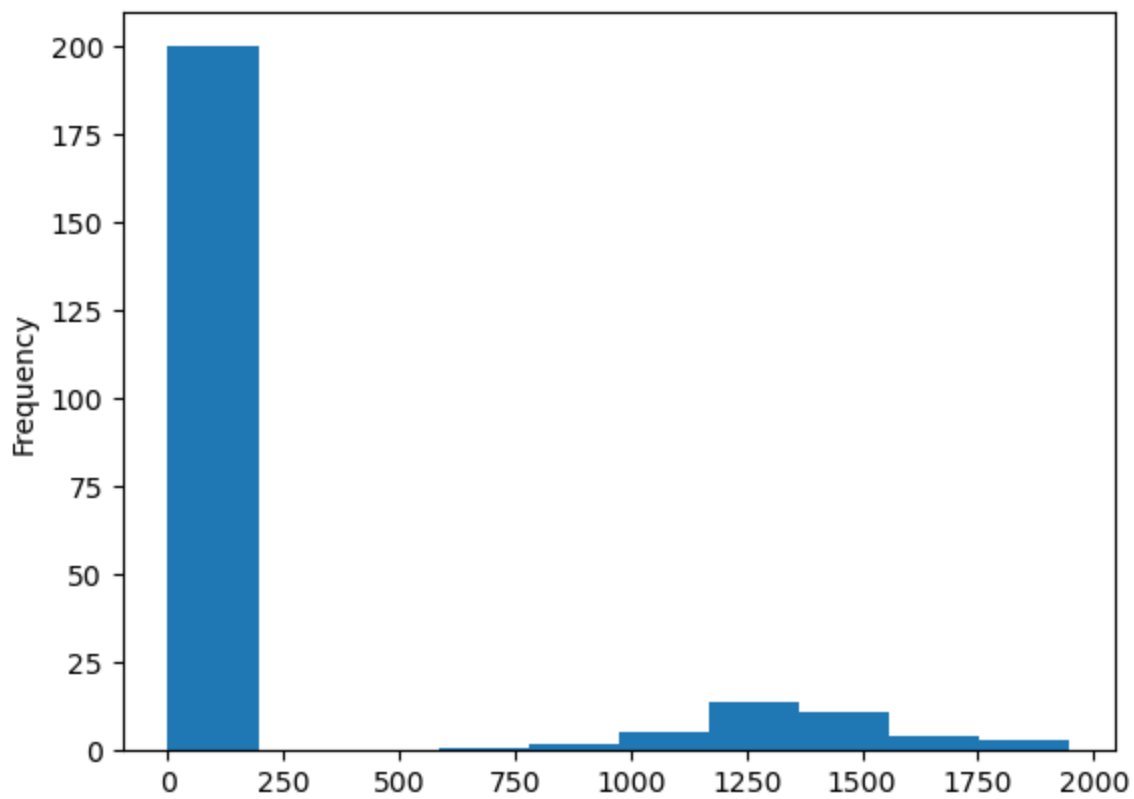
```
In [53]: data['Temperature (K)'].plot(kind='hist', bins=10)
```

```
Out[53]: <AxesSubplot:ylabel='Frequency'>
```



```
In [10]: data['Radius (R/Ro)'].plot(kind='hist', bins=10)
```

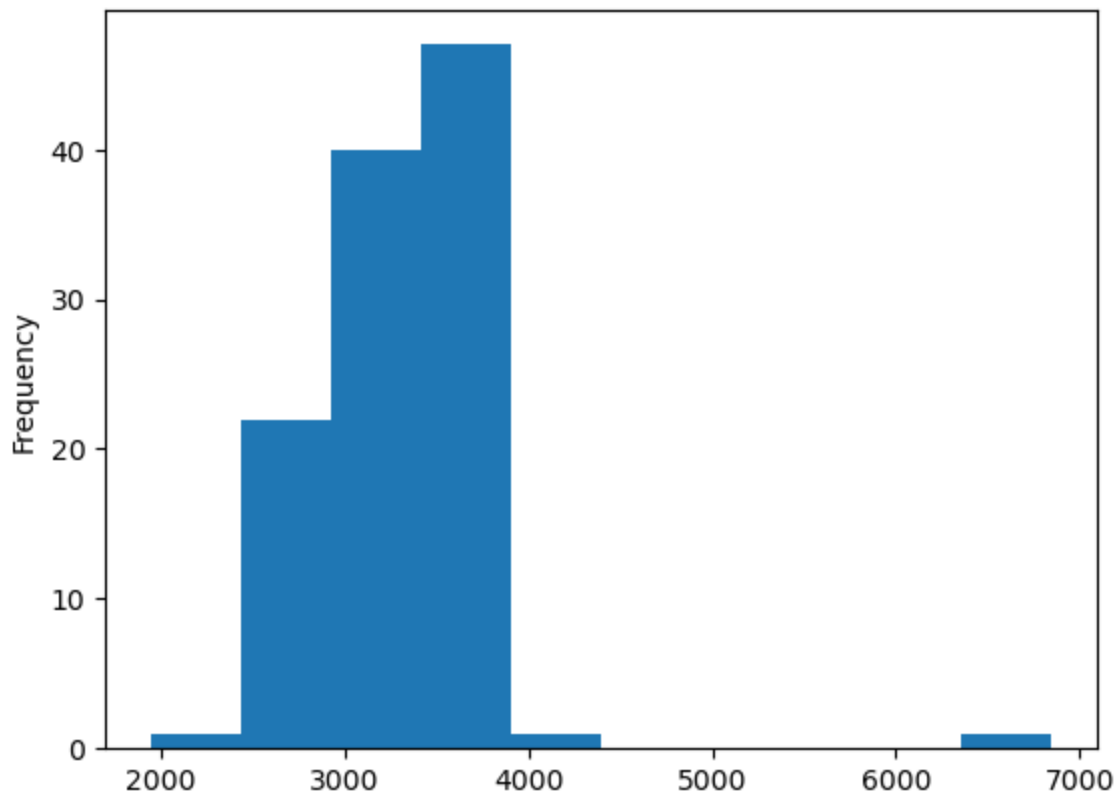
```
Out[10]: <AxesSubplot:ylabel='Frequency'>
```



Er is voor deze dataset gekozen om een gelijke verdeling aan star types te selecteren (dwerfen, main sequence en giants). Daarom is er een kolommen niet zomaar een normaalverdeling te vinden, er is een bias naar kleinere sterren (ook omdat de giants vele maten groter zijn). Je kan wel kijken naar rijen die gefilterd zijn op bijvoorbeeld kleur of type.

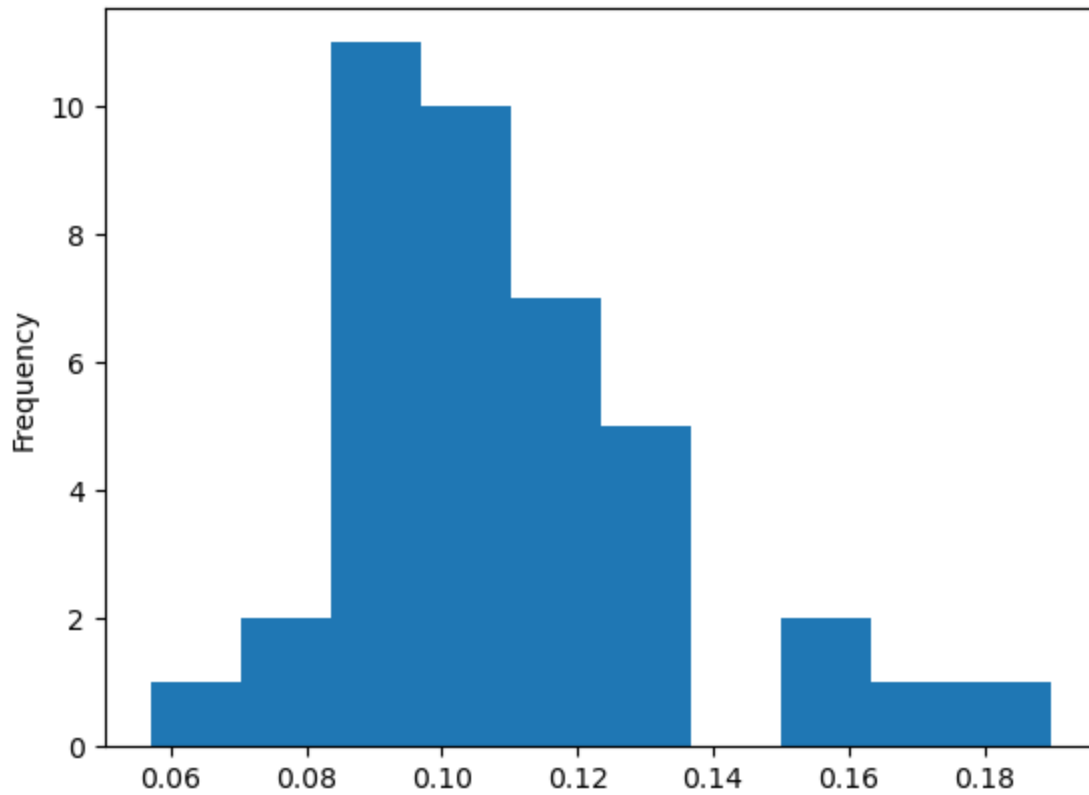
```
In [35]: color = 'Red'
data[(data['Star color'] == color)][['Temperature (K)']].plot(kind='hist', bins=10)
```

```
Out[35]: <AxesSubplot:ylabel='Frequency'>
```



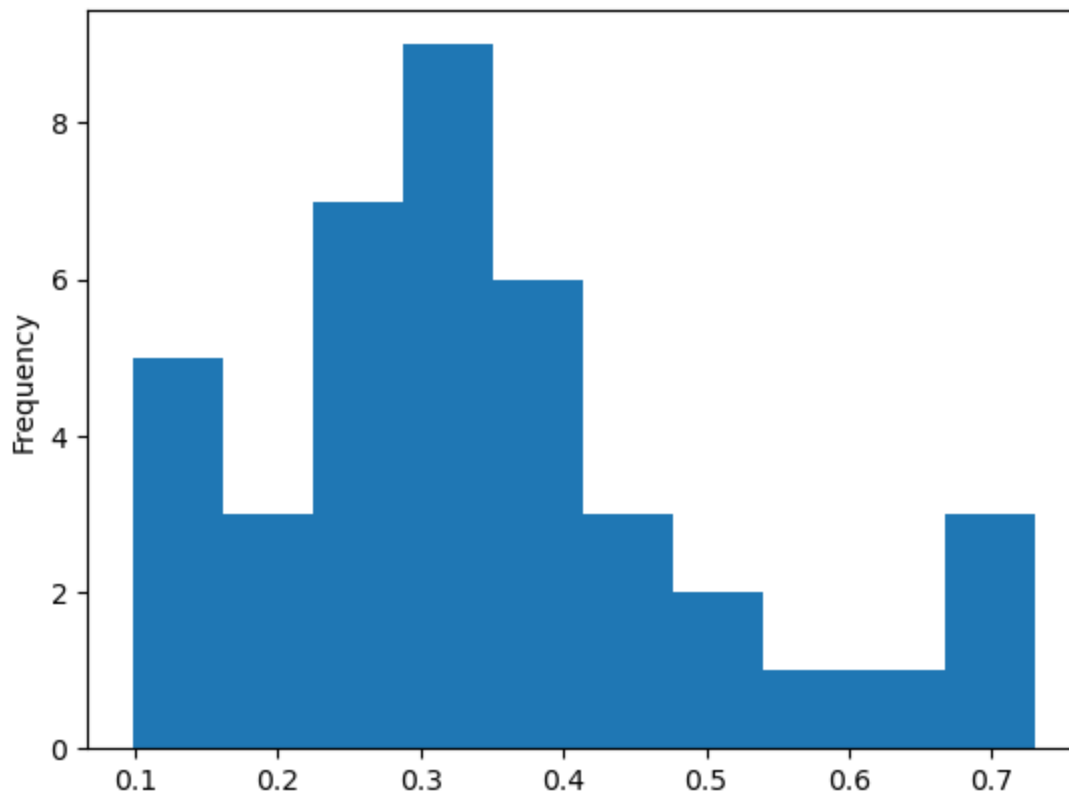
```
In [24]: data[(data['Star type'] == 'Red Dwarf')]['Radius (R/Ro)'].plot(kind='hist', bins=10)
```

```
Out[24]: <AxesSubplot:ylabel='Frequency'>
```



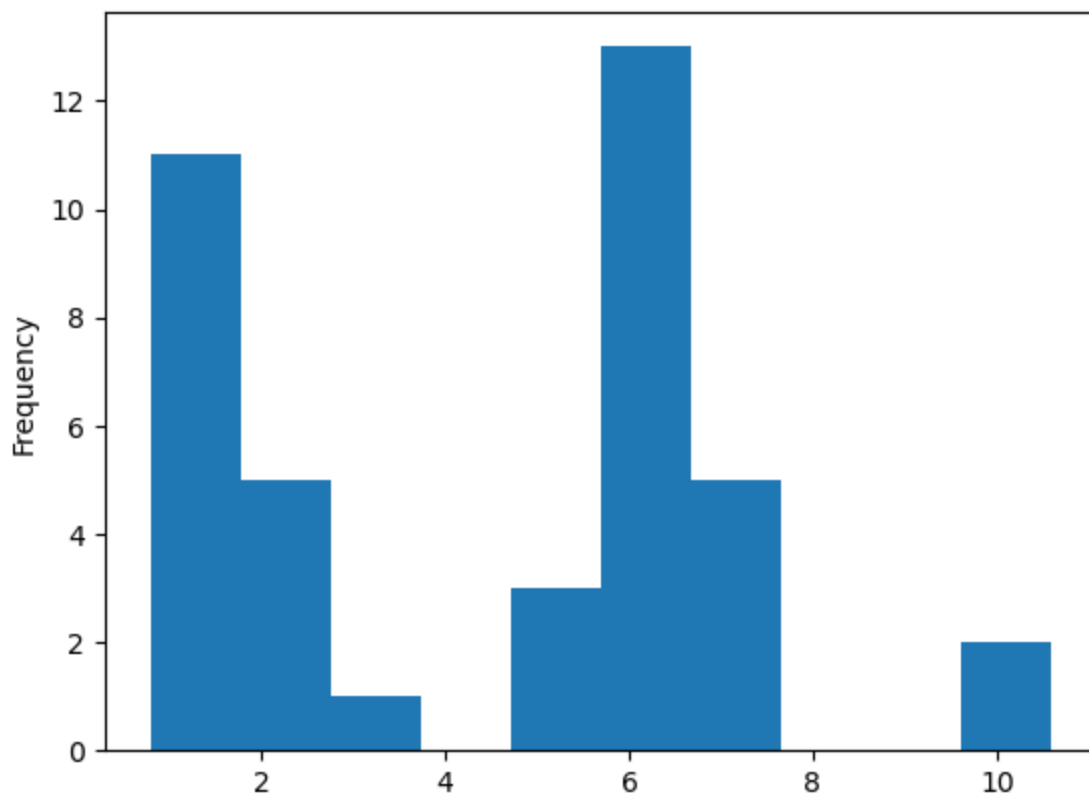
```
In [46]: data[(data['Star type'] == 'Brown Dwarf')]['Radius (R/Ro)'].plot(kind='hist', bins=10)
```

```
Out[46]: <AxesSubplot:ylabel='Frequency'>
```



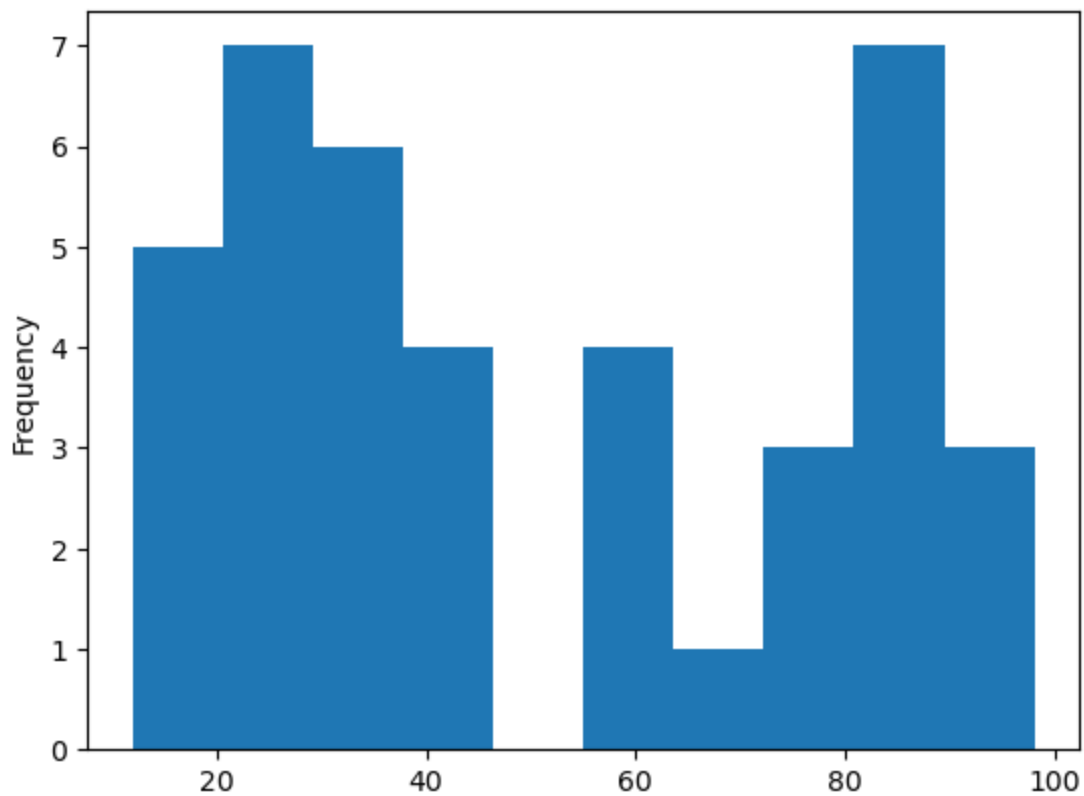
```
In [45]: data[(data['Star type'] == 'Main Sequence')]['Radius (R/Ro)'].plot(kind='hist', bins=10)
```

```
Out[45]: <AxesSubplot:ylabel='Frequency'>
```



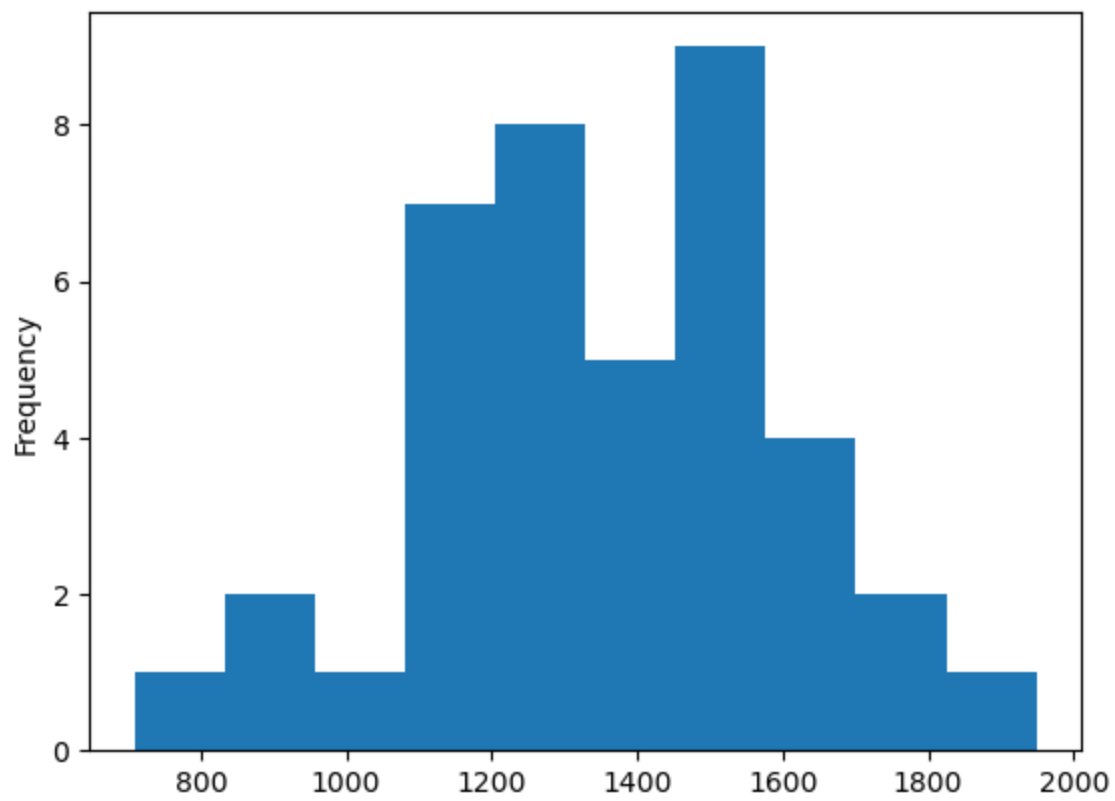
```
In [44]: data[(data['Star type'] == 'Super Giant')]['Radius (R/Ro)'].plot(kind='hist', bins=10)
```

```
Out[44]: <AxesSubplot:ylabel='Frequency'>
```



```
In [42]: data[(data['Star type'] == 'Hyper Giant')]['Radius (R/Ro)'].plot(kind='hist', bins=10)
```

```
Out[42]: <AxesSubplot:ylabel='Frequency'>
```



Bij deze slices van data lijken er wel normal distributions aanwezig te zijn.