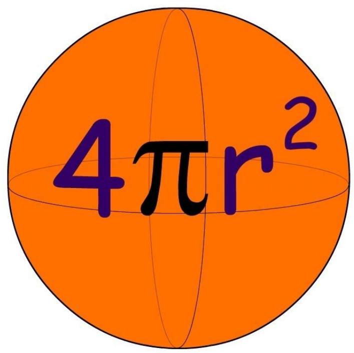
STAR DATASET TO PREDICT STAR TYPES:

1. The column **‘Luminosity(L/Lo)’** in our data frame contains values of the ratio of the **Luminosity of the Star:Luminosity of the Sun.**
   And similarly column, ‘**Radius(R/Ro)’** contains the values of the ratio of the **Radius of the Star:Radius of the Sun.**
   Hence to find the Luminosity and Radius, let us multiply the values under the columns with **‘Luminosity of the Sun’** and **‘Radius of Sun’** respectively**.** We can create new columns containing the values of the Luminosities and Radii of the stars.

2. When we try to carry out calculations related to a star, we assume the star to be a perfectly spherical body, to make our calculations easily computable.
   The Surface Area of any spherical shaped object is:



The formula to find the Luminosity of a star on a Main Sequence is:

$$L = 4\pi R^2 \sigma T^4$$

L= Luminosity
R= Stellar Radius
T= Surface Temperature

σ = Stefan-Boltzmann Constant

As we are already given the **‘Luminosity’** and the **‘Temperature’** in the data frame, and have already computed the **‘Surface Area’**, we can use this data to calculate the **‘Stefan-Boltzmann Constant’**.
The value of the Stefan–Boltzmann constant is given in SI units by:

$$\sigma = 5.6703x10^{-8}\, watt / m^2 K^4$$

As per calculations, the median value of the Stefan-Boltzmann constant is:

5.59*e-08

That is close to the value of the constant.

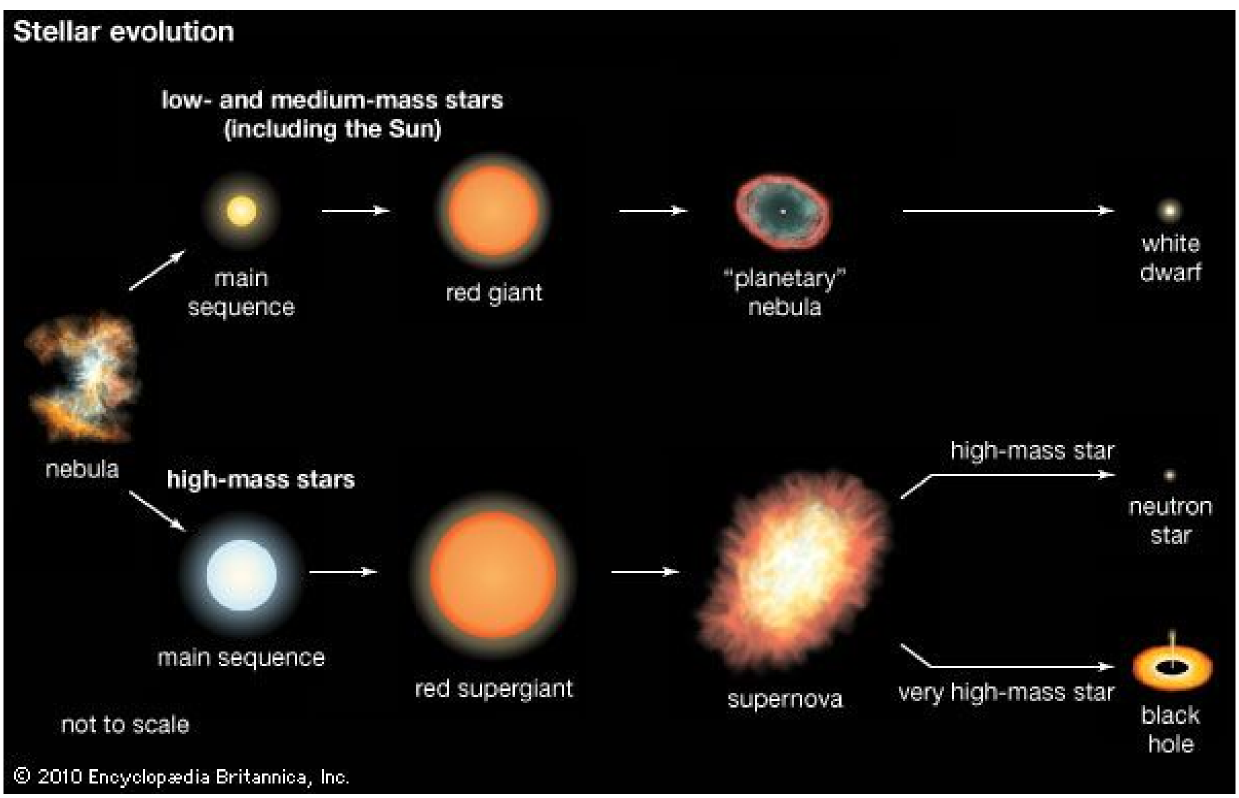
3. Stars can be classified based on their Luminosity and Temperature. We can group our data based on their type. They are usually classified into 6 categories:

- Brown Dwarf -> Star Type = 0

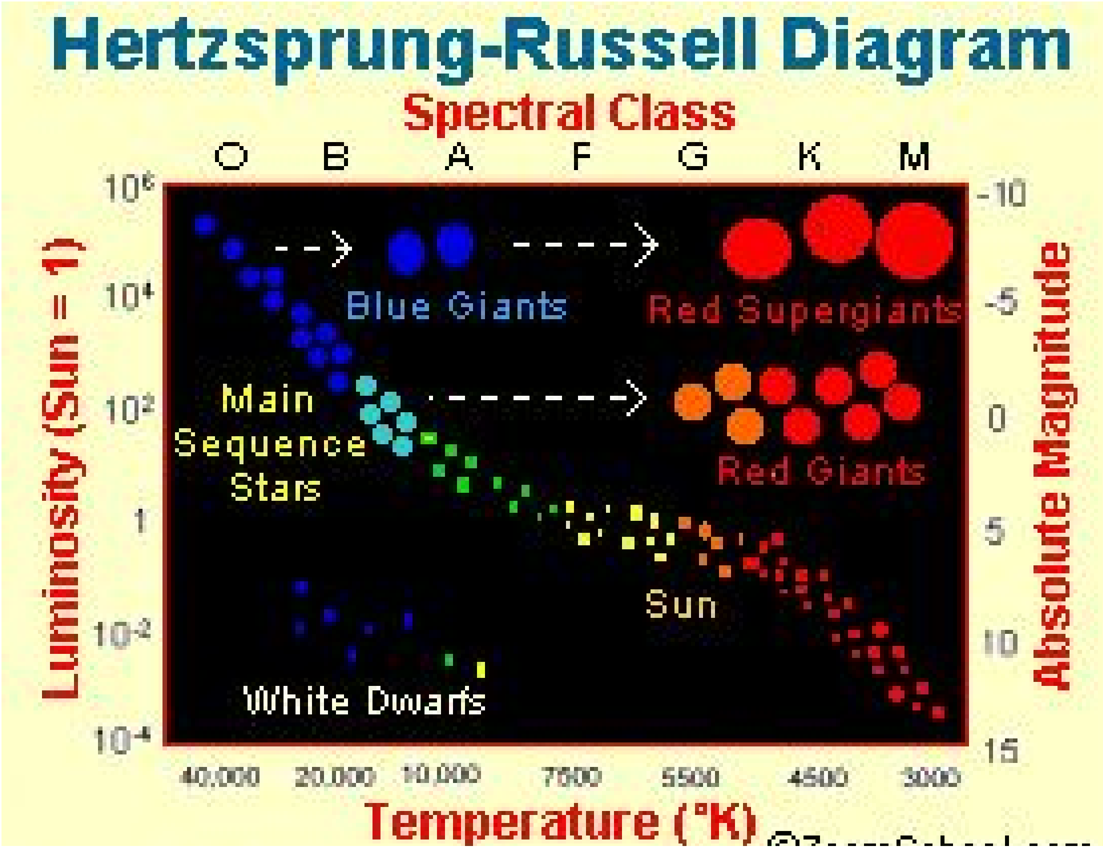
- Red Dwarf -> Star Type = 1
- White Dwarf-> Star Type = 2
- Main Sequence -> Star Type = 3
- Supergiant -> Star Type = 4
- Hypergiant -> Star Type = 5

Now, we can find the average Luminosity and Radius of the stars in each group. As we can see in the data frame 'df_g', Brown dwarfs, White and Red dwarfs have the smallest radii (as the name suggests). And Supergiants and Hypergiants have the greatest average radii.

- **Brown Dwarfs** are stars that have an extremely small radius and surface area, and can not reach the stage of nuclear fusion in their core. They do not ignite.
- **Red Dwarfs** are the smallest type of star on the main sequence. They are extremely common in the Milky Way, especially around the Sun.
- **White Dwarfs** are small in size (surface area) and also have a comparatively low luminosity. But they are very large in mass, and hence they are among the densest type of star.
- 90% of the stars in our universe are in the **Main Sequence**.Our Sun is a Main Sequence star. Main sequence stars fuse hydrogen atoms to form helium atoms in their cores.
- **Supergiants** are the most massively sized stars. There are two types of Supergiants, namely 'Red' and 'Blue'. These stars have short life spans (going as low as a few hundred thousand years).
- **Hypergiant** is a very rare kind of star. They usually shows extremely high luminosities and very high rates of mass loss.



4. The Hertzsprung - Russell (H-R) Diagram is a graph that plots **Absolute Magnitude** against the **Temperature** of the surface of the star. It is used by astronomers to classify stars according to their luminosity, spectral type, color, temperature and evolutionary stage. The diagram usually looks like this:



5. We can filter our data set to see the stars in the Main Sequence.

In [17]:

```
import pandas as pd
df= pd.read_csv('6 class csv.csv')
df.sample(n=25)
```

Out[17]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)	Star type	Star color	Spectral Class
78	2621	0.000600	0.09800	12.81	1	Red	M
217	19400	10920.000000	6.03000	-3.08	3	Blue-white	B
2	2600	0.000300	0.10200	18.70	0	Red	M
210	22350	12450.000000	6.36000	-3.67	3	Blue-white	B
85	9675	0.000450	0.01090	13.98	2	Blue White	A
184	3453	0.000621	0.07730	17.08	0	Red	M
115	3553	145000.000000	1324.00000	-11.03	5	Red	M
17	3692	0.003670	0.47000	10.80	1	Red	M
238	9235	404940.000000	1112.00000	-11.23	5	White	A
126	2935	0.000870	0.09320	16.88	0	Red	M
77	3342	0.001500	0.30700	11.87	1	Red	M
43	3200	195000.000000	17.00000	-7.22	4	Red	M
76	2890	0.003400	0.24000	13.46	1	Red	M
108	24345	142000.000000	57.00000	-6.24	4	Blue	O
99	36108	198000.000000	10.20000	-4.40	3	Blue	O
154	25070	14500.000000	5.92000	-3.98	3	Blue-white	B
53	3749	550000.000000	1648.00000	-8.05	5	Orange	M
1	3042	0.000500	0.15420	16.60	0	Red	M
133	2989	0.008700	0.34000	13.12	1	Red	M
9	2700	0.000180	0.13000	16.05	0	Red	M
15	3340	0.003800	0.24000	13.07	1	Red	M
204	18340	0.001340	0.01240	11.22	2	Blue	B
209	19360	0.001250	0.00998	11.62	2	Blue	B
13	3628	0.005500	0.39300	10.48	1	Red	M
134	3542	0.000900	0.62000	14.23	1	Red	M

In [18]:

```
#1
Sun_luminosity= 3.828 * (10**26) #Watts
df['Luminosity']=  df['Luminosity(L/Lo)']*Sun_luminosity

Sun_radius= 6.95700 * (10**8) #metres
df['Radius']= df['Radius(R/Ro)']*Sun_radius

df
```

Out[18]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)	Star type	Star color	Spectral Class	Lum
0	3068	0.002400	0.1700	16.12	0	Red	M	9.1872
1	3042	0.000500	0.1542	16.60	0	Red	M	1.9140
2	2600	0.000300	0.1020	18.70	0	Red	M	1.1484
3	2800	0.000200	0.1600	16.65	0	Red	M	7.6560
4	1939	0.000138	0.1030	20.06	0	Red	M	5.2826
...	...	...	...	...	...	...	...	
235	38940	374830.000000	1356.0000	-9.93	5	Blue	O	1.4348
236	30839	834042.000000	1194.0000	-10.63	5	Blue	O	3.1927
237	8829	537493.000000	1423.0000	-10.73	5	White	A	2.0575
238	9235	404940.000000	1112.0000	-11.23	5	White	A	1.5501
239	37882	294903.000000	1783.0000	-7.80	5	Blue	O	1.1288

240 rows × 9 columns

In [19]:

```python
#2
import math
df['Surface Area']= 4*math.pi*( (df['Radius'])**2 )
df
```

Out[19]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)	Star type	Star color	Spectral Class	Lum
0	3068	0.002400	0.1700	16.12	0	Red	M	9.1872
1	3042	0.000500	0.1542	16.60	0	Red	M	1.9140
2	2600	0.000300	0.1020	18.70	0	Red	M	1.1484
3	2800	0.000200	0.1600	16.65	0	Red	M	7.6560
4	1939	0.000138	0.1030	20.06	0	Red	M	5.2826
...	...	...	...	...	...	...	...	
235	38940	374830.000000	1356.0000	-9.93	5	Blue	O	1.4348
236	30839	834042.000000	1194.0000	-10.63	5	Blue	O	3.1927
237	8829	537493.000000	1423.0000	-10.73	5	White	A	2.0575
238	9235	404940.000000	1112.0000	-11.23	5	White	A	1.5501
239	37882	294903.000000	1783.0000	-7.80	5	Blue	O	1.1288

240 rows × 10 columns

In [20]:

```python
#3
group= df.groupby('Star type')
df_g= group.agg("mean").reset_index()
df_g
```

Out[20]:

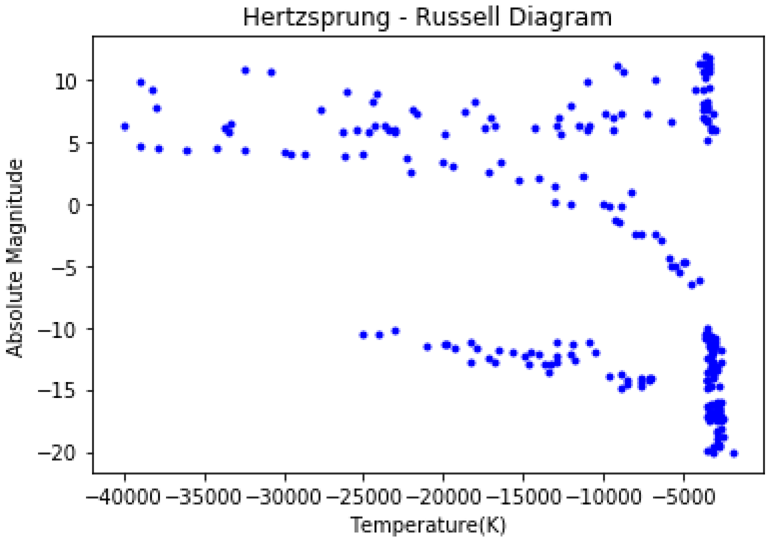
	Star type	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)	Luminosity	Radi
0	0	2997.950	0.000693	0.110015	17.563500	2.653857e+23	7.653744e+
1	1	3283.825	0.005406	0.348145	12.539975	2.069321e+24	2.422045e+
2	2	13931.450	0.002434	0.010728	12.582500	9.315917e+23	7.463644e+
3	3	16018.000	32067.386275	4.430300	-0.367425	1.227540e+31	3.082160e+
4	4	15347.850	301816.250000	51.150000	-6.369925	1.155353e+32	3.558506e+
5	5	11405.700	309246.525000	1366.897500	-9.654250	1.183796e+32	9.509506e+

In [21]:

```
#4
import matplotlib.pyplot as plt
plt.plot(-(df['Temperature (K)']),-(df['Absolute magnitude(Mv)']),'b.')
plt.title('Hertzsprung - Russell Diagram')
plt.xlabel('Temperature(K)')
plt.ylabel('Absolute Magnitude')
```

Out[21]:

```
Text(0, 0.5, 'Absolute Magnitude')
```



In [22]:

```
#5
#Let's see just the stars in the main sequence

plt.plot(-(df_mainseq['Temperature (K)']),-(df_mainseq['Absolute magnitude(Mv)']),'ro')
plt.title('Hertzsprung - Russell Diagram')
plt.xlabel('Temperature(K)')
plt.ylabel('Absolute Magnitude')
```

Out[22]:

```
Text(0, 0.5, 'Absolute Magnitude')
```