DSE 315: Data Science in Practice

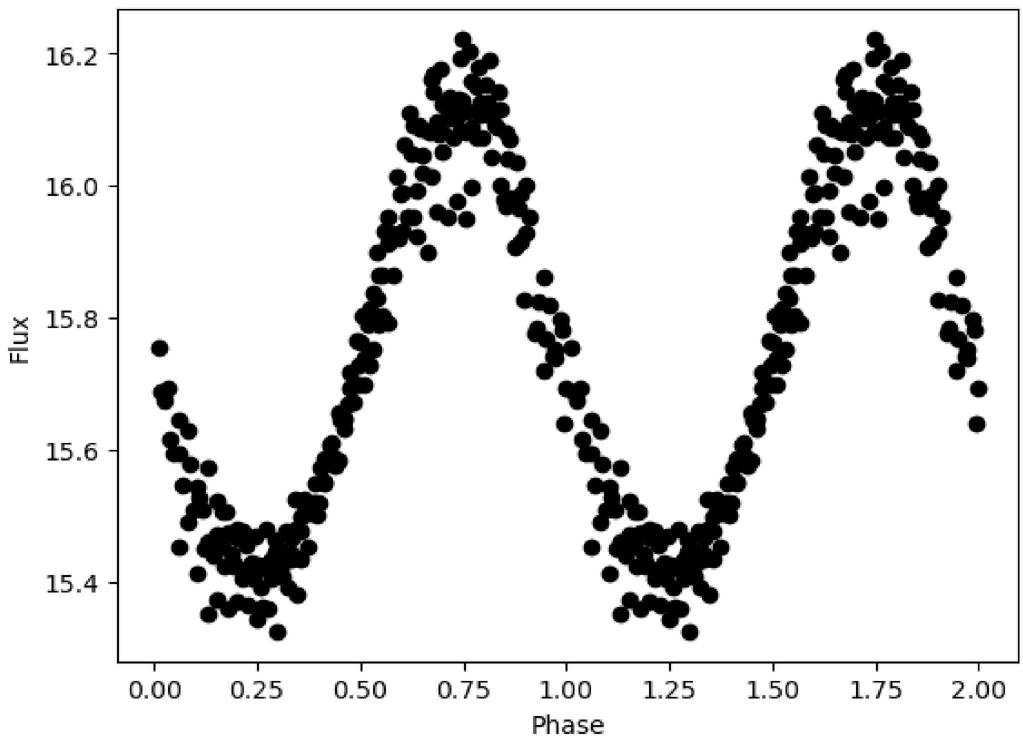
Final project

LIGHT CURVE ANALYSIS OF VARIABLE STARS FROM GAIA Data

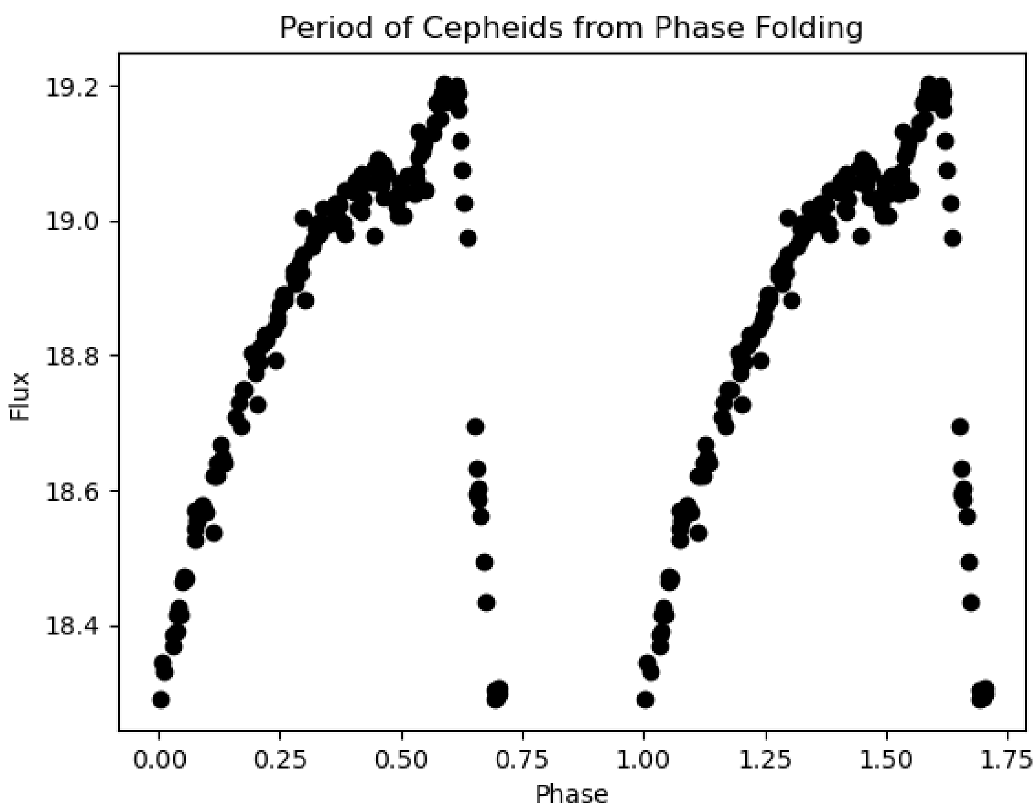
Divyansh Tripathi
Department of Physics (20108)

Stars, as we know, are one of the essential things in astronomy. Life is possible on earth also due to a star, **Sun**. Due to our everyday experience and prior knowledge, we would expect a star to be an astronomical body whose brightness does not change appreciably. This constant brightness is not valid in the case of **Variable Stars,** which, as the name suggests, are stars whose brightness varies in a detectable amount in a matter of days or months.
In this project, I analysed the **Light Curve Data** of the **Cepheid Variable** and Stars like **RR Layre**. I also analysed data on various astrophysical parameters of these stars and tried to mimic the results of previous research works.
The project uses data from the **ESA** (European Space Agency) and space observation mission **GAIA** (Global Astrometric Interferometer for Astrophysics).

The data used is part of 3 data releases of the mission, with the latest data released DR3 in 2022. The observation is the most extensive collection of astronomical data of the Milkyway to this day, containing about 1 billion objects.
The data was obtained from this archival data [GAIA Data] using two methods. First, the data was directly downloaded from the website in CSV files using a built-in SQL query. Second, data was queried from inside the Jupyter Notebook using the astroquery package of GAIA and then stored in CSV files.
The approach for this project was an instructive one, and I first tested my method on the synthetic, clean and processed data and then went on to do the more rigorous analysis of the actual dataset.
First, to do the Light Curve Analysis, I synthetically used a Sin signal of period one and then found its period using two methods. First, using the Phase-Fold Method, all the data was compressed to one phase observation of the star, and then the period was found by repetition of maxima in phase. This method is crude and needs a very educated guess of the period in the data first to start the phase fold and derive the pattern.
The second method is the Lomb Scargle Periodogram, a relatively new light curve analysis method designed to overcome the limitation of the phase fold method. This method produces quite accurate results of the Period of the pulsation of the Cepheid Variables.
After that, the periodogram method is used on the dataset and for the RR-Layre Data, a pulsation period of 1.02 days is very close to the universally accepted value.

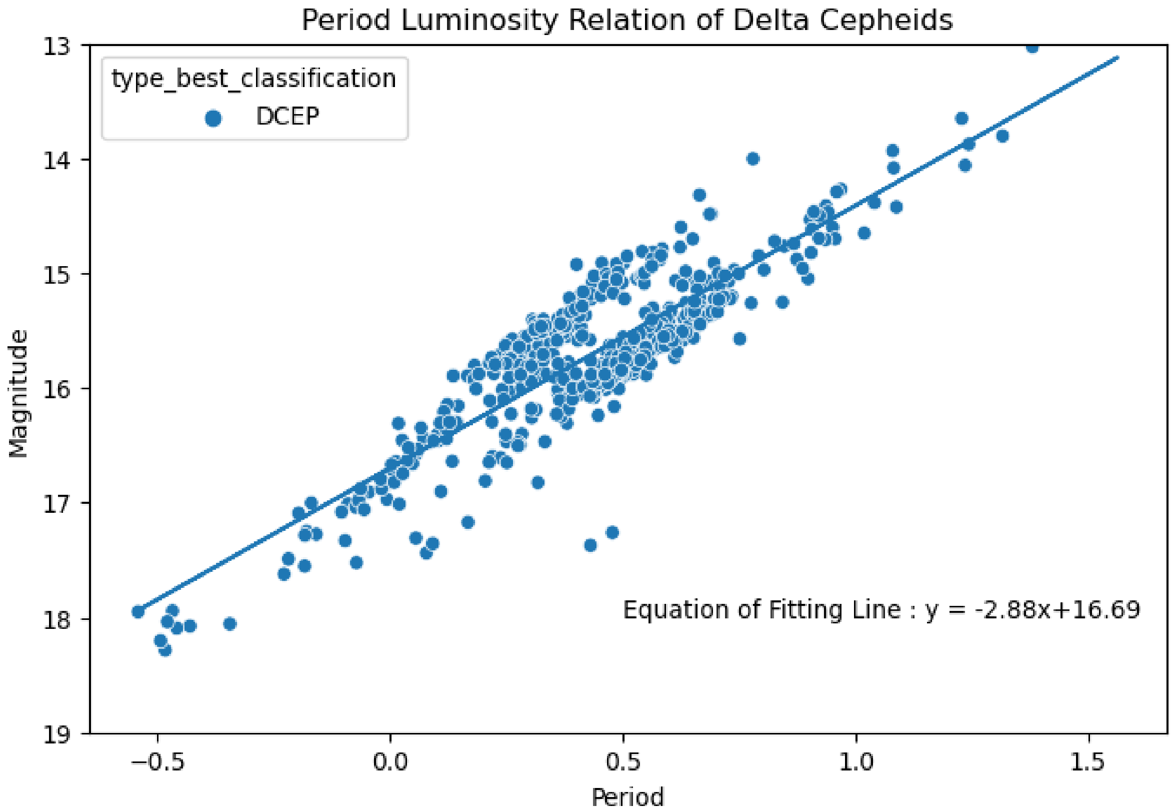


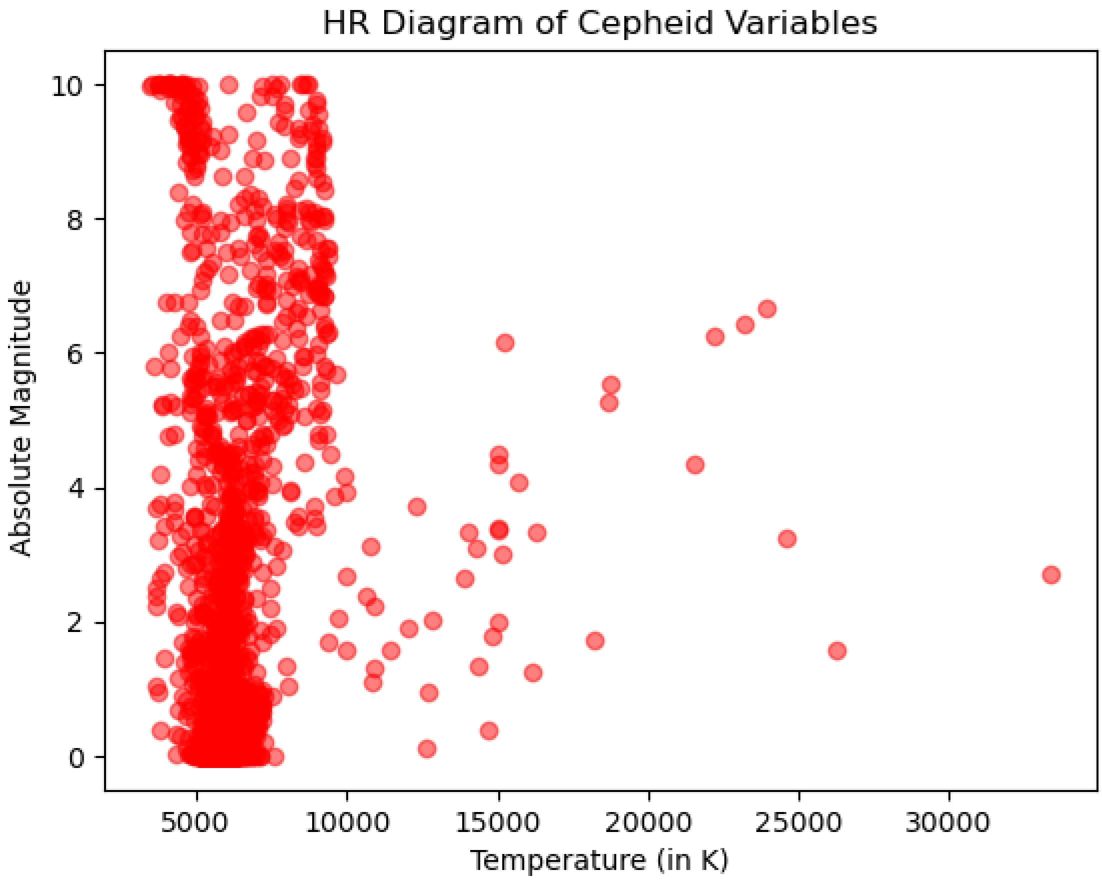
Flux
16.2
16.0
15.8
15.6
15.4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Phase



Period of Cepheids from Phase Folding
Flux
19.2
19.0
18.8
18.6
18.4
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Phase

Then I go on the data analysis of the astrophysical parameters and find some exciting patterns in the dataset.
Some of the plots I obtained are :





Results :

The final Results obtained from this project are:

1. The variable stars lie in HR-Diagram in a narrow region known as the Instability Strip.

2. The temperature in the instability strip is relatively constant between 5000-8000 K, with some stars reaching 10,000 K

3. The Period Luminosity Relation obtained by this data is **Magnitude = -2.02*log(Period)+16.706**, when the whole dataset is considered

4. When only taking Delta Cepheids, The Period Luminosity Relation obtained is **Magnitude = -2.28*log(Period)+16.697**

Some variations from the accepted value are there in the obtained value, and further analysis is required for the following, and there is a scope for further work in this.