**Pre-processing presentation script**

# Slide 07

---

**Part 1**

---

In the pre-processing stage -

We had to perform extensive data cleaning to prepare the original data for testing and analysis. During the inspection, we discovered numerous missing data points and realised that some of the column data needed to be converted into float data type.

After converting the necessary columns and dropping rows with null values, we were left with a reduced dataset that accounted for 3.2% of the original data. Although this represented a loss, it was deemed acceptable.

Finally, we reset the index as some pre-processing steps had altered it.

---

**Part 2**

---

We needed to calculate the absolute magnitude (amag) in a set of stars. To do this, we researched and sourced an appropriate equation.

We then used two pieces of data that were readily available to us: the visual magnitude (vmag) and the parallax (plx) of each star. Using the sourced equation, we created a new column in our dataset that contained the calculated absolute magnitude (amag) values for each star. This column was based on the relationship between vmag and plx that the equation describes.

With the new absolute magnitude (amag) column added to our dataset, we were able to conduct further analysis and gain a deeper understanding of the characteristics of the stars we were studying.

---

**Part 3**

---

We needed to analyse a large dataset of stars that contained spectral classifications. However, these classifications were not in a format that was easy to work with, as they contained a mix of information on both the spectral type and the target type.

To make our analysis more manageable, we decided to split the spectral classifications into two categories: the B-V colour index and the absolute magnitude (amag). Using the B-V colour index, we identified eight distinct target types within the main category (0-6), while any stars that fell outside of these categories were placed in the miscellaneous category (7). We used a for loop to run through the data and assign each star a numerical value between 0 and 7 based on its spectral target class.

We then created a classification model based on the absolute magnitude (amag) of each star, which allowed us to categorise stars by their size (dwarf or giant) within their target class. To do this, we used Roman numerals to assign each star into either the dwarf or giant category based on its spectral type with an intuitive for loop. This approach provided us with a more accurate and detailed understanding of the characteristics of the stars in our dataset.

# Slide 08

The 7 main types of stellar classification, known as the Harvard spectral classification system, categorize stars based on their temperature and spectral lines. The classes, from hottest to coolest, are O, B, A, F, G, K, and M. These classifications help to study the evolution and properties of stars.

---

O: These are the hottest, with surface temperatures above 30,000 K. They emit most of their light in the ultraviolet part of the spectrum.

B: These stars are also very hot, with surface temperatures between 10,000 K and 30,000 K. They emit most of their light in the blue part of the spectrum.

A: These stars are hot, with surface temperatures between 7,500 K and 10,000 K. They emit most of their light in the blue part of the spectrum.

F: These stars are moderately hot, with surface temperatures between 6,000 K and 7,500 K. They emit most of their light in the blue and green parts of the spectrum.

G: These stars include our Sun and are moderately cool, with surface temperatures between 5,200 K and 6,000 K. They emit most of their light in the yellow part of the spectrum.

K: These stars are cool, with surface temperatures between 3,700 K and 5,200 K. They emit most of their light in the orange part of the spectrum.

M: These stars are the coolest, with surface temperatures below 3,700 K. They emit most of their light in the red part of the spectrum.

---

The "K" in the temperature scale used for stellar classification stands for Kelvin, which is the unit of measurement used to express temperature in the International System of Units. The Kelvin scale is based on absolute zero and is commonly used in astronomy and astrophysics to compare the temperatures of celestial objects.