

MACHINE LEARNING STARS DATASET

<https://www.kaggle.com/brsdincer/star-type-classification>

Dataset zawierający specyfikacje gwiazd w tym ich typy.

Features:

Liczbowe:

- Temperatura (K)
- Jasność (L/Lo)
- Promień (R/Ro)
- Absolutna wielkość gwiazdowa (Mv)

Tekstowe:

- Kolor
- Klasa wydmowa

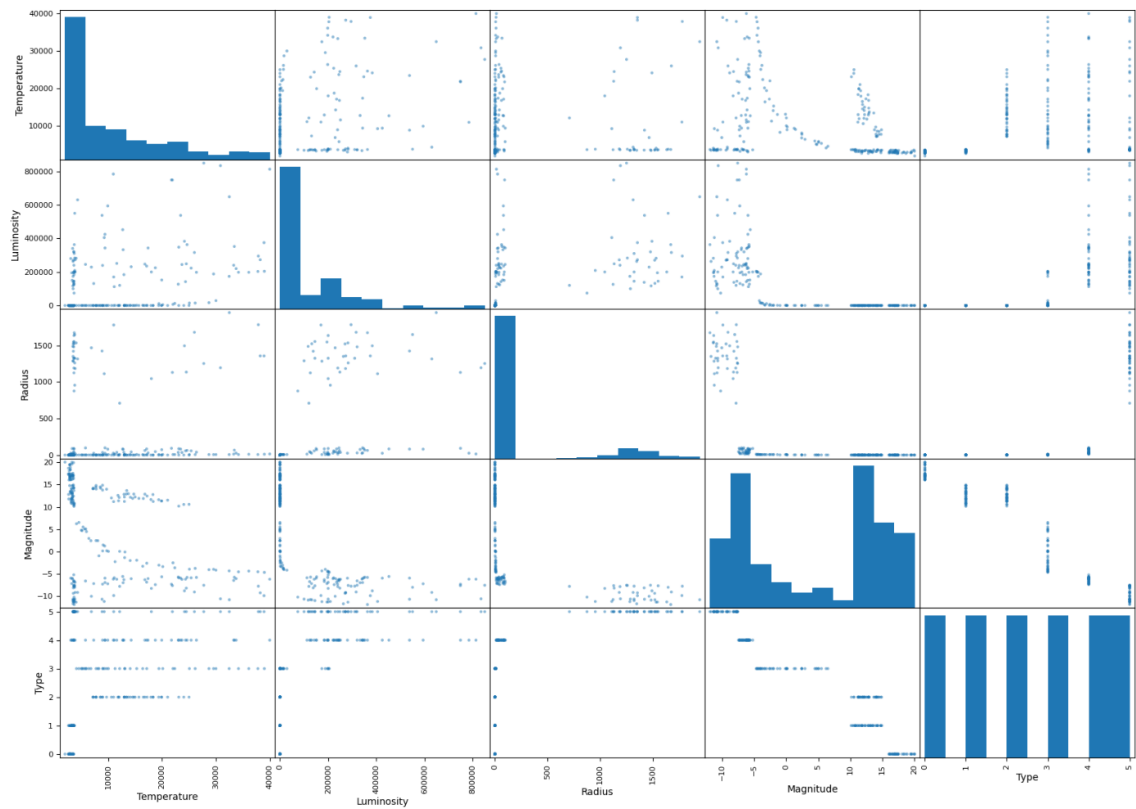
Target:

- Typ gwiazdy

Lo – jasność Słońca

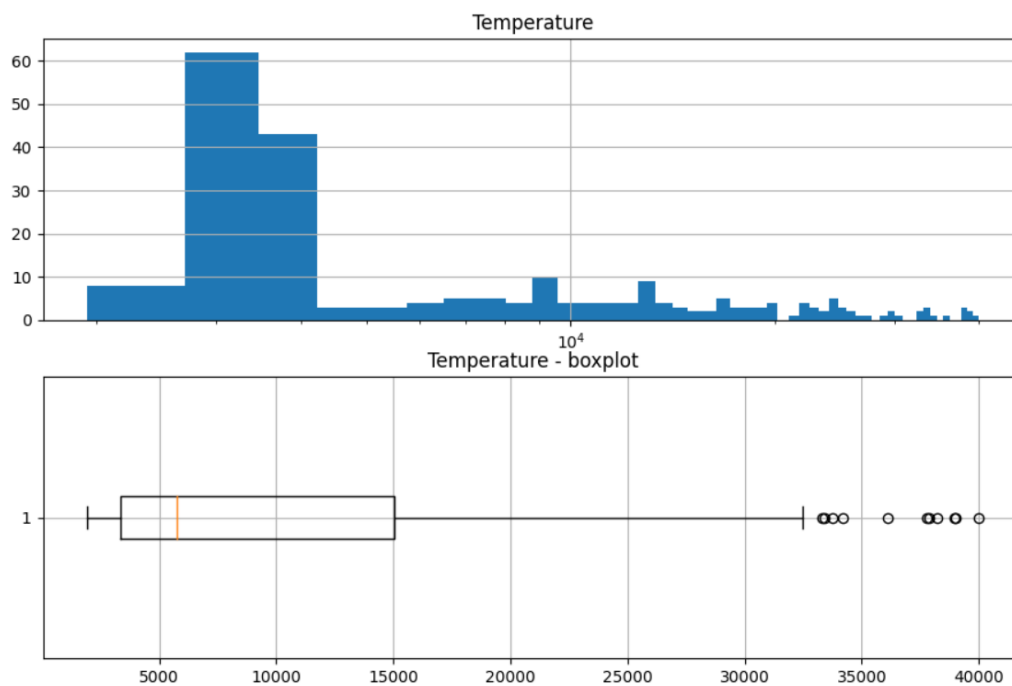
Ro – promień Słońca

Wizualizacja zbioru



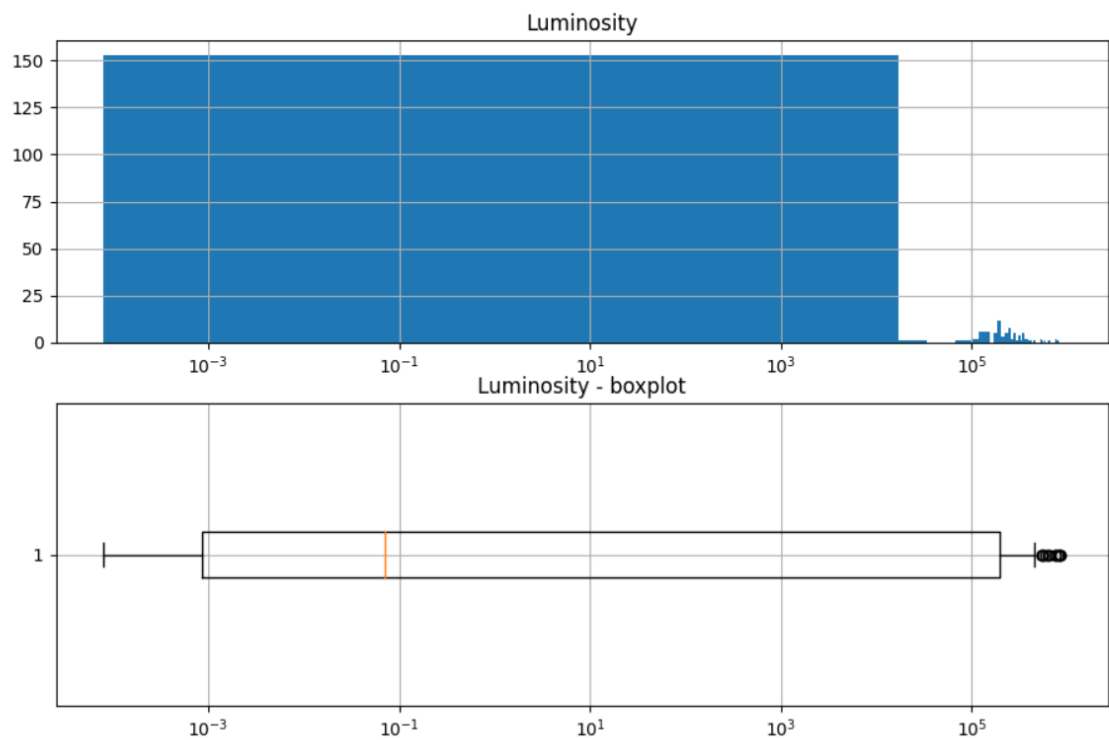
Matrix plot pokazujący jak zależą od siebie wartości

1. Temperatura



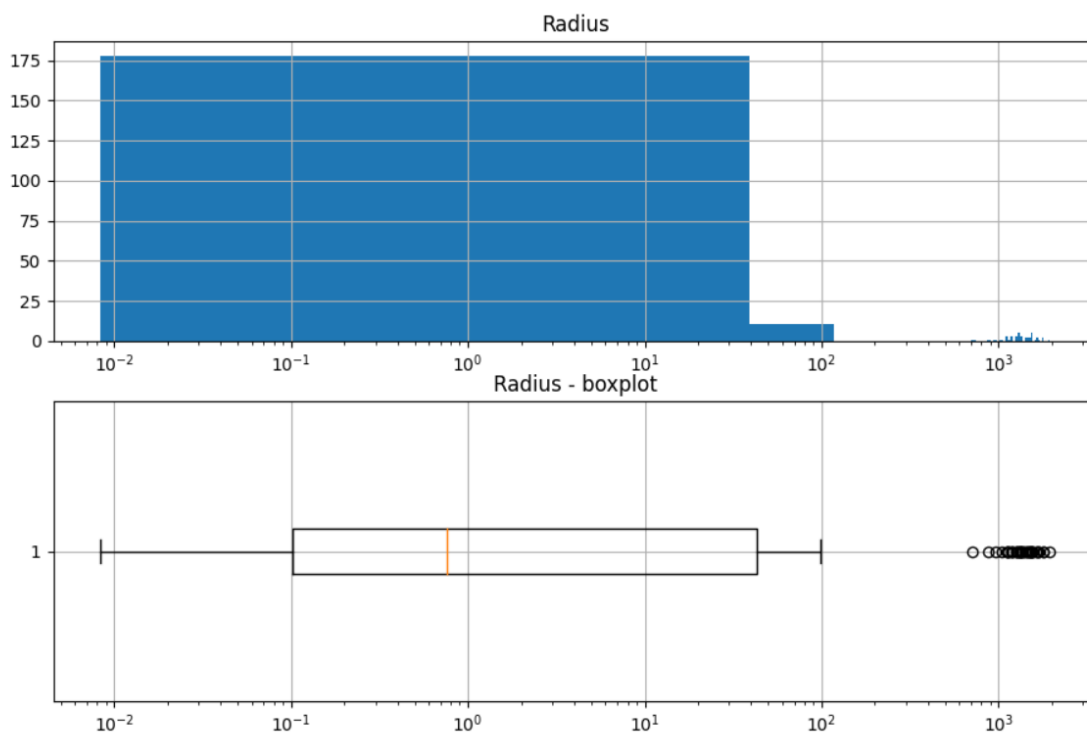
Rozkład temperatury (boxplot) pokazuje, że najwięcej gwiazd ma temperaturę z zakresu 4000K - 15000K. Mediana to ok 6000K. Niektóre gwiazdy mają jednak temperatury sięgające znacznie więcej.

2. Jasność



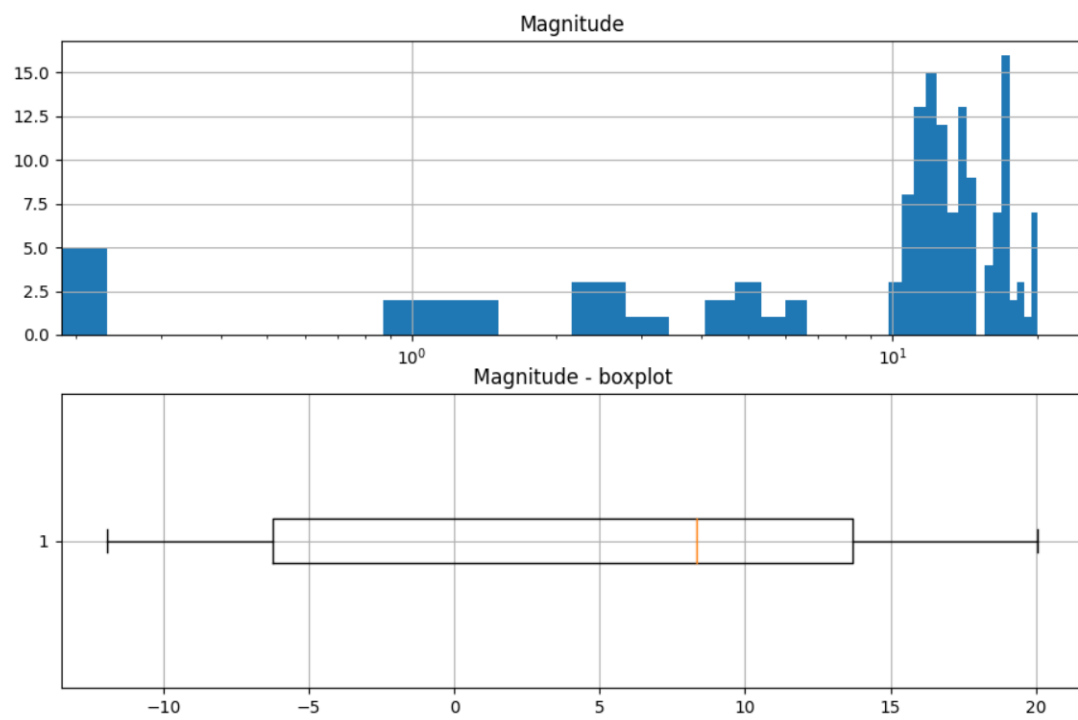
Światłość waha się bardziej niż temperatura, dlatego należy podawać ją logarytmicznie. Mediana jest mniejsza od wielkości naszego Słońca.

3. Promień

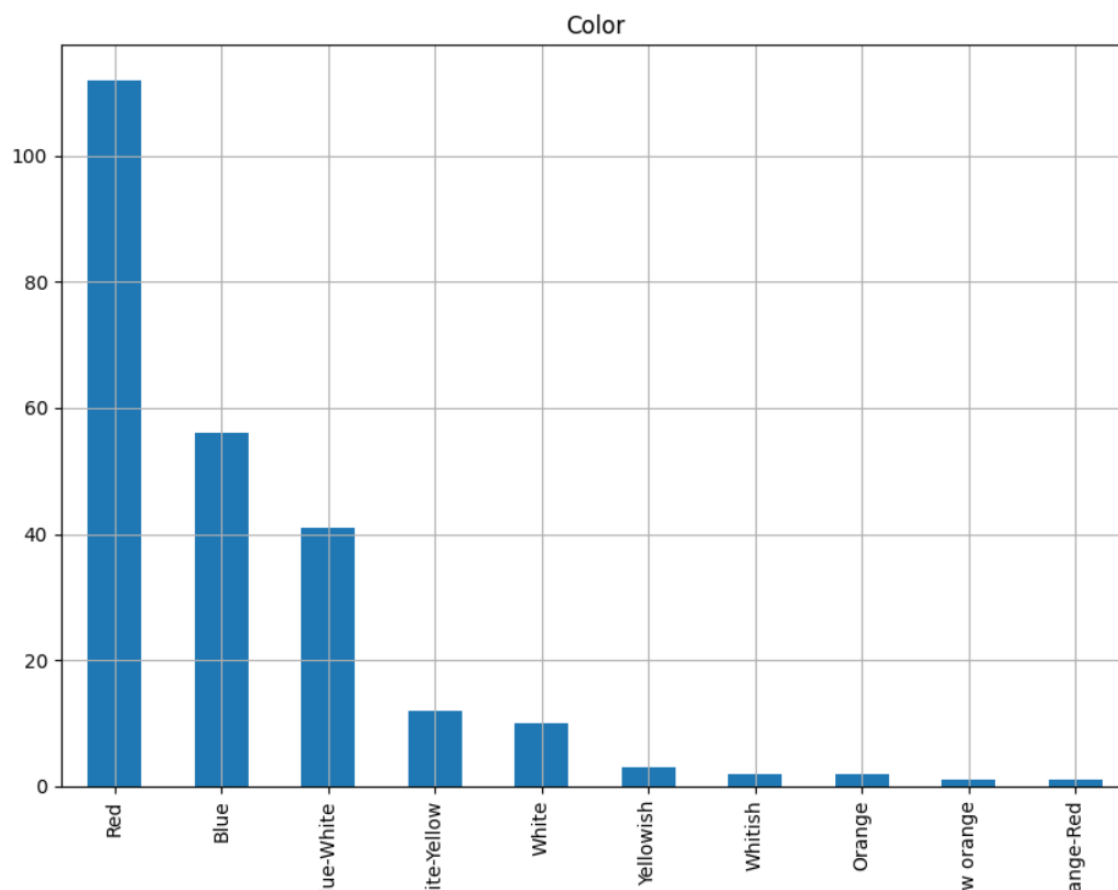


Promienie tak samo jak Jasność należy pokazywać logarytmicznie z uwagi na duży rozrzut. Średni promień gwiazdy jest w przybliżeniu równy Promieniowi naszego słońca. Większość gwiazd jest 100 razy większa lub mniejsza, lecz są też gwiazdy mocno oddalone od średniej na wykresie

4. Absolutna wielkość gwiazdowa

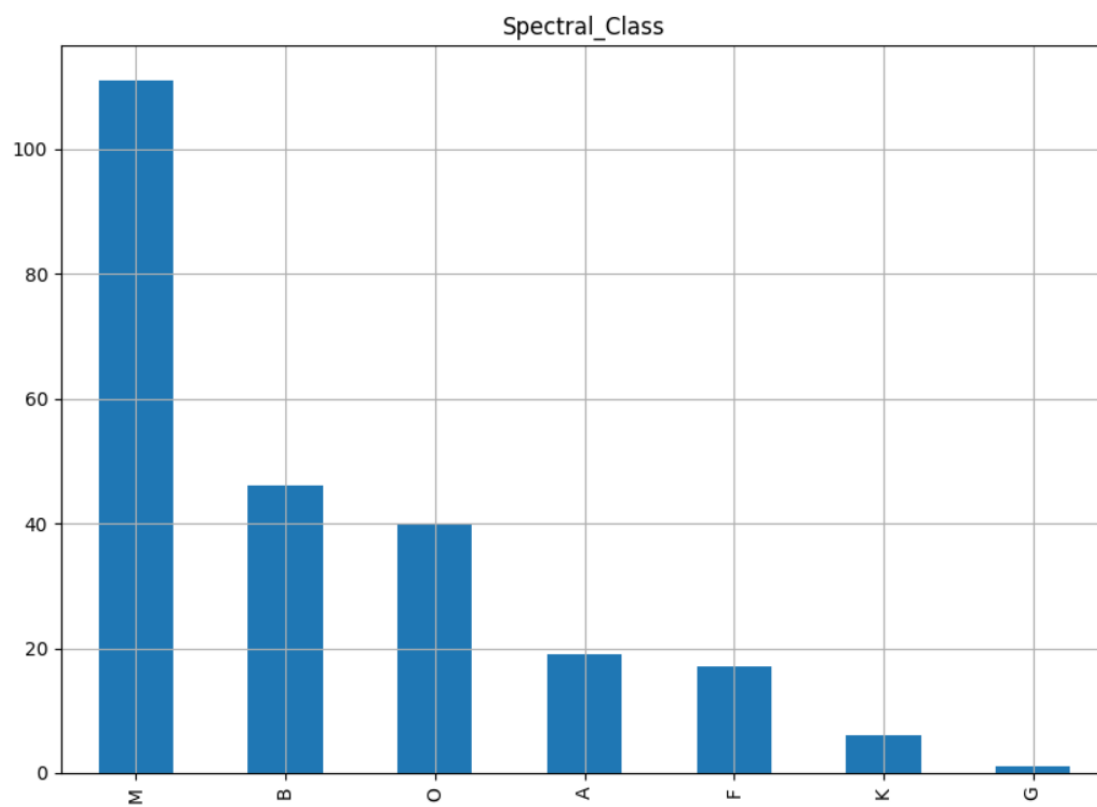


5. Kolor

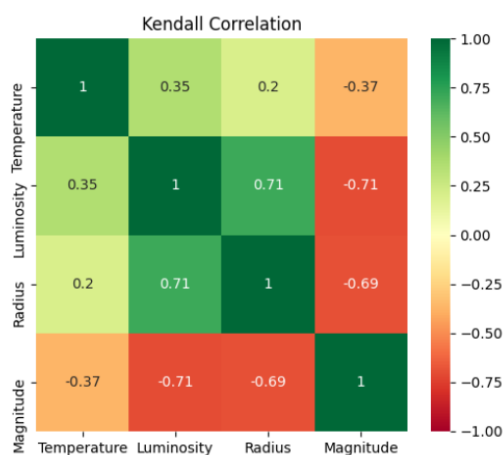
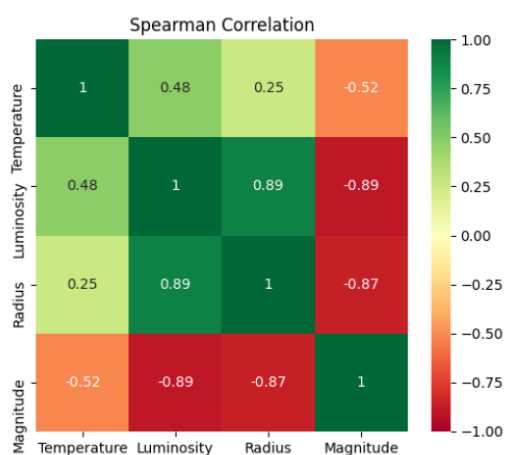
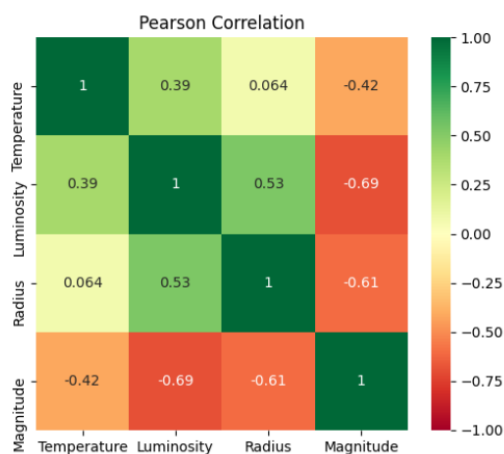


W kolorze przeważa czerwony (połowa wszystkich gwiazd ze zbioru jest tego koloru). Znaczący udział w kolorach biorą Czerwony, Niebieski oraz Biały (najczęściej gwiazda będzie miała jeden z tych lub ich mieszankę)

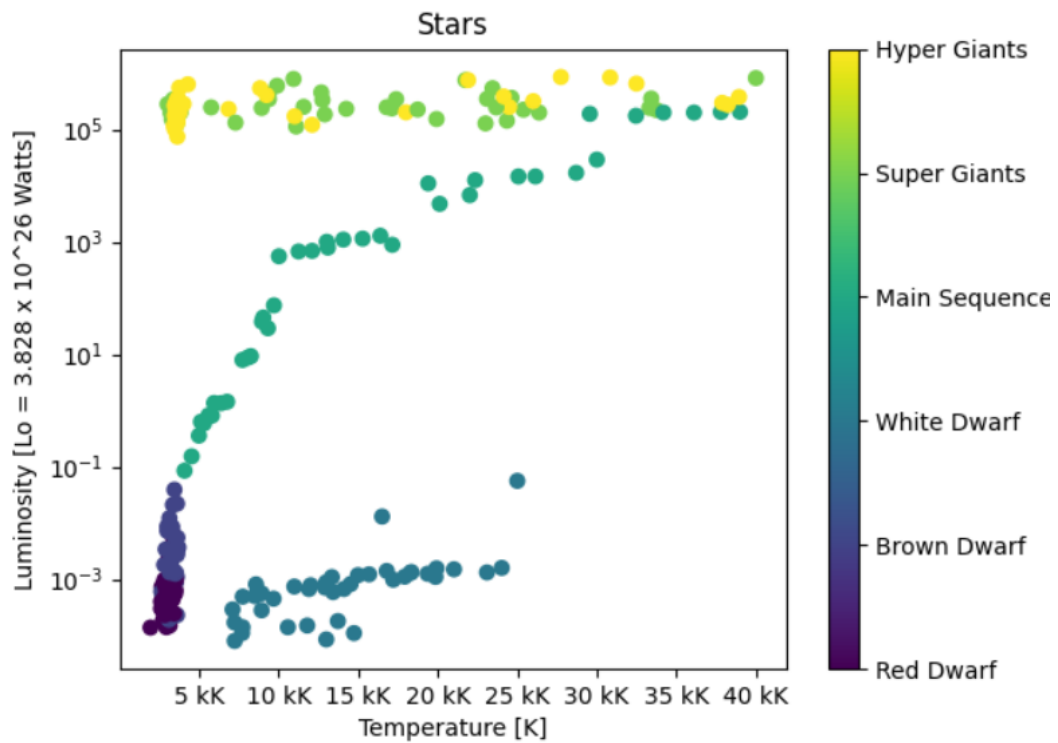
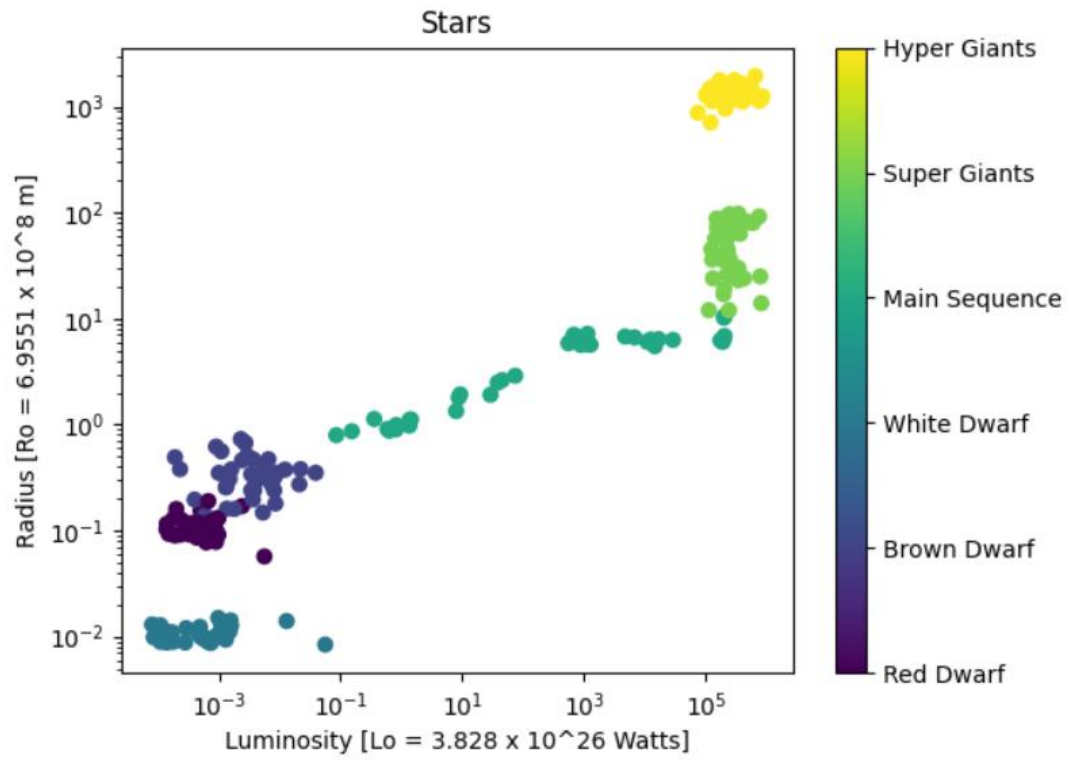
6. Klasa widmowa (https://en.wikipedia.org/wiki/Asteroid_spectral_types)



Zależności danych w zbiorze danych



Wykresy korelacji pokazują, że najbardziej zależne są Promień-Jasność oraz Temperatura-Jasność dlatego warto zobaczyć wykresy Typu gwiazdy od tych 2 zestawów.



Wykres Jasności od Temperatury jest właśnie dlatego dosyć popularny aby pokazać jak dzielą się gwiazdy ze względu na te dwa typy

