Report on Correlation of the Characteristics of Stars

Name: Edward Leen
UHID: 23090780
GitHub: https://github.com/edwardleen95/Applied-Data-Science-Assignment-1

1. Introduction

This report consists of exploratory data analysis (EDA) on a dataset containing information about stars. Through various visualizations such as histograms, scatter plots and a correlation heatmap, the aim is to uncover patterns, relationships and distributions in the data.

2. Analysis and Results

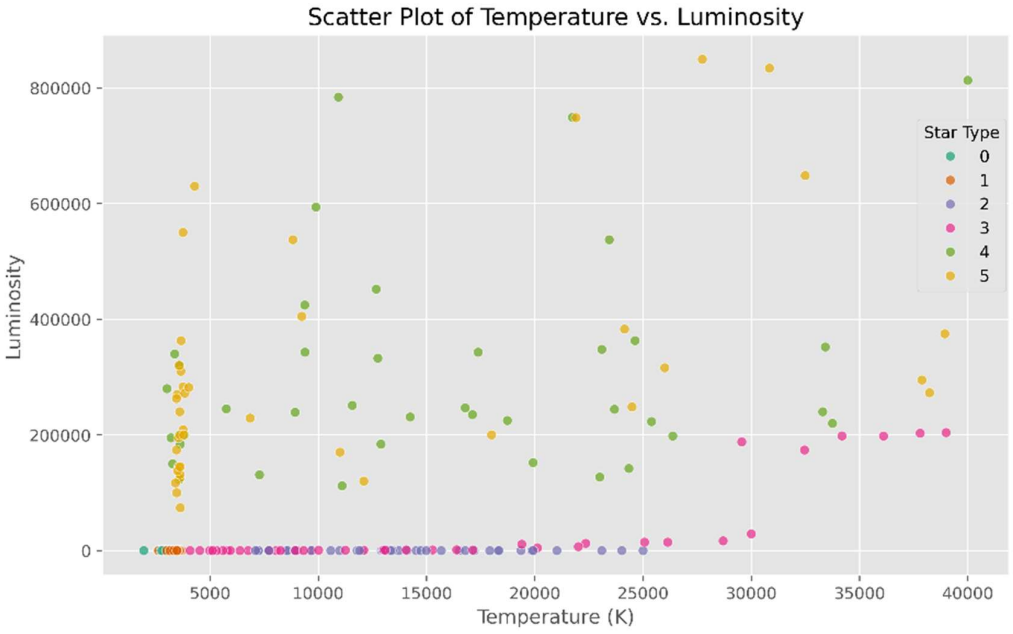
Statistical Depth Analysis

Statistical depth provides a summary of the central tendency, dispersion, and overall range of numeric features. This summary helps to quickly understand the general properties of the data and identify any significant outliers or anomalies.

	Temperature	Luminosity	Radius	Absolute Magnitude	Type
mean	10497.46	107188.36	237.16	4.38	2.50
std	9552.43	179432.24	517.16	10.53	1.71
min	1939.00	0.00	0.01	-11.92	0.00
max	40000.00	849420.00	1948.50	20.06	5.00

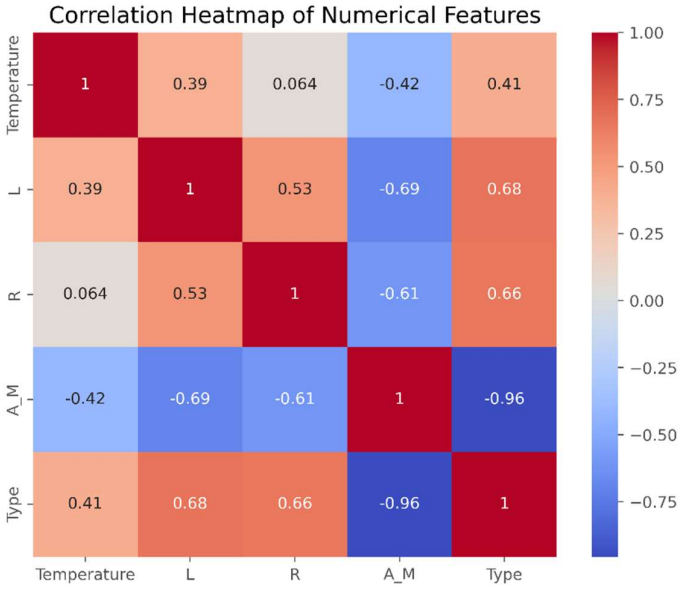
Scatter Plot of Relative Luminosity and Absolute Magnitude:

This scatter plot shows the relationship between temperature and luminosity, with colours representing different star types. By examining this relationship, we can understand how star temperature influences brightness, a critical factor in stellar classification.



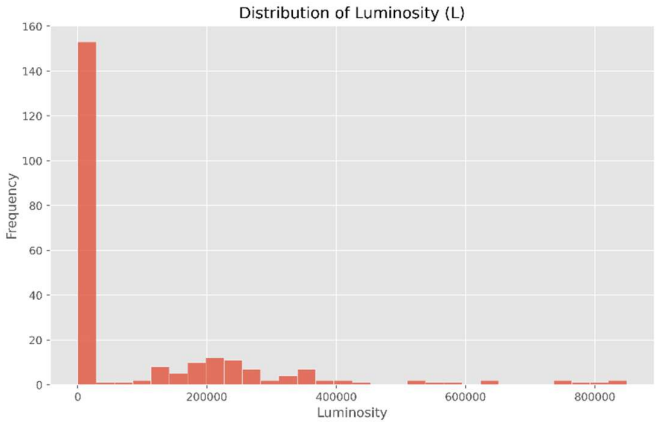
Correlation Heatmap:

The correlation heatmap reveals relationships between numerical features, helping us identify pairs of variables that change together. Strong correlations between temperature and luminosity, might suggest that as one variable increases, the other tends to follow. The heatmap indicates a strong positive correlation between temperature and luminosity, as well as temperature and radius. This suggests that hotter stars are generally both larger and more luminous.



Distribution of Luminosity

The luminosity distribution provides insight into the brightness range across the stars in the dataset. Stars with higher luminosity are often hotter and larger, so this visualization helps in categorizing stars by brightness levels. The luminosity distribution is highly skewed, indicating a small number of stars with extremely high luminosity. Most stars have low luminosity, suggesting that dimmer stars are more prevalent. A skew of 2.06 indicates a positive skewness to the right. A kurtosis of 4.46 in the luminosity indicates a moderate departure from the normality.



Conclusion

In conclusion, the scatter plot demonstrates how a star's temperature influences its luminosity across different types of stars. Generally, hotter stars tend to be brighter. This pattern is further supported by the correlation heatmap, which reveals strong positive correlations between temperature, luminosity, and radius, indicating that hotter stars are typically larger and more luminous. Additionally, the histogram of absolute magnitude shows a varied distribution of stellar brightness, with a smaller number of extremely bright stars.