**Introduction**:

This dataset contains information of different patients who are susceptible to heart failure.

It is predicted by considering various factors like blood sugar, age, cholesterol, type of chest pain and various other factors.

**Descriptive analysis of numerical columns (explaining columns with significant observation):**

1. The mean age is 53.51 years. Age ranges from 28 to 77 years. The interquartile range suggests that people between 47 to 60 years old are majorly affected, hence heart disease is predominant in middle aged individuals.
2. Resting blood pressure ranges from 0 to 200mmhg. 0 indicates possible error values since blood pressure cannot be 0 in anyone in reality. The Interquartile range suggests that there are many individuals who fall in High BP category.
3. Mean value of cholesterol is 198.8 which falls in the higher limit of the normal cholesterol values(<200mg/dl). It is ranging from 0 to 603, there might be error values since cholesterol cannot be 0 in human body. cholesterol levels are varying broadly (can be observed from interquartile range) indicating it can be a risk factor for heart diseases.

**Interpretation of cholesterol, age and Resting blood pressure using KDE:**

1. Age: Distributed over a range of values.
2. Cholesterol: No uniform distribution of data, outliers/extreme values can be observed.
3. RestingBP: No uniform distribution of data, some erroneous values might be present.
4. Oldpeak, maxHR-values can fluctuate since they pertain to medical variables.
5. HeatDisease, Fastingbs-binary values.

**Interpretation of box-plot:**

1. Box-plot confirms that the data is distributed over a range of values.
2. Median is dividing dataset into two equal halves.
3. The points outside the Box-plot confirms the presence of outliers in both the parameters.

**Skewness and kurtosis:**

1. Cholesterol-Negatively skewed with less number of higher values on the left side.
2. Also means that more number of values falls under the mean indicating the presence of lower error values.
3. Kurtosis explains that data is not normally distributed since it shows some flatter peaks.
4. RestingBP-Positively skewed with higher density of lower values on right.
5. Kurtosis elucidates that data is not normally distributed since it shows taller peaks with outliers.

**Class Imbalance:**

1. This is to check if the target column (heart disease) is imbalanced. Imbalanced dataset can result in models that favor majority class hence resulting in biased decision.
2. Imbalance ratio is: 1.2365853658536585, which indicates that class is fairly balanced
3. Might not need rigorous measures to correct imbalance. Although there is slight

imbalance suggesting majority class is more than minority class.

**Conclusion**:

From this analysis, we can conclude that older males who have high cholesterol and high blood pressure tend to be mostly asymptomatic and are prone to heart disease rather than females. Additionally, analysis highlights the importance of factors like St\_slope, oldpeak in predicting heat attack.