

Assignment - 1

EDA



Data Set – Please [Click here](#) to access the Dataset

About – Dataset:

The dataset provided for this project is focused on heart failure, a critical medical condition characterized by the heart's inability to pump blood effectively, leading to inadequate circulation throughout the body. Here's an overview of the dataset:

Title: Heart Failure Clinical Records Dataset

Source: The dataset was collected at the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad, Punjab, Pakistan, during the period of April to December in 2015. It was later made available on Kaggle.

Description: The dataset comprises medical records of 299 patients diagnosed with heart failure. These records include various clinical, body, and lifestyle information of the patients. The dataset encompasses 13 features, providing insights into the patients' health status and potential risk factors for heart failure.

Features:

Age: The age of the patient (numeric).

Anaemia: Indicates the presence of anemia (binary: 0 for no, 1 for yes).

Creatinine Phosphokinase (CPK): Level of creatinine phosphokinase enzyme in the blood (numeric).

Diabetes: Indicates whether the patient has diabetes (binary: 0 for no, 1 for yes).

Ejection Fraction: Percentage of blood leaving the heart at each contraction (numeric).

High Blood Pressure: Indicates whether the patient has high blood pressure (binary: 0 for no, 1 for yes).

Platelets: Platelet count in the blood (numeric).

Serum Creatinine: Level of creatinine in the blood (numeric).

Serum Sodium: Level of sodium in the blood (numeric).

Sex: Gender of the patient (binary: 0 for female, 1 for male).

Smoking: Indicates whether the patient smokes (binary: 0 for no, 1 for yes).

Time: Follow-up period (in days) for the patient's condition (numeric).

Death Event:

Indicates whether the patient died during the follow-up period (binary: 0 for no, 1 for yes).

1. What is the distribution of age among heart failure patients in the dataset?
2. How does the death rate vary with age?
3. What is the percentage of male and female patients in the dataset?

4. How does the platelet count vary among different age groups?
5. Is there a correlation between creatinine and sodium levels in the blood?
6. How does the prevalence of high blood pressure differ between male and female patients?
7. What is the relationship between smoking habits and the occurrence of heart failure?
8. Are there any noticeable patterns in the distribution of death events across different age groups?
9. Is there any significant difference in ejection fraction between patients with and without diabetes?
10. How does the serum creatinine level vary between patients who survived and those who did not?