

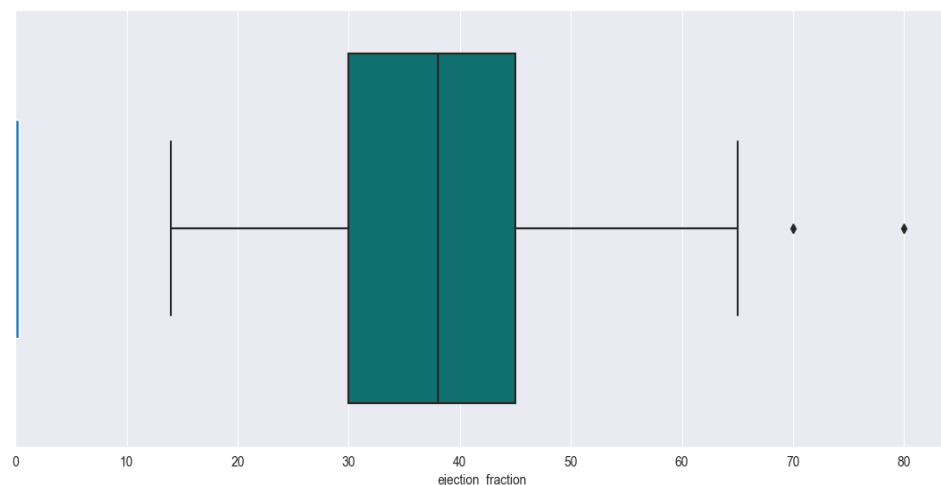
Overview of Dataset: Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Heart failure is a common event caused by CVDs and this dataset contains 12 features that can be used to predict mortality by heart failure.

Most cardiovascular diseases can be prevented by addressing behavioural risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies.

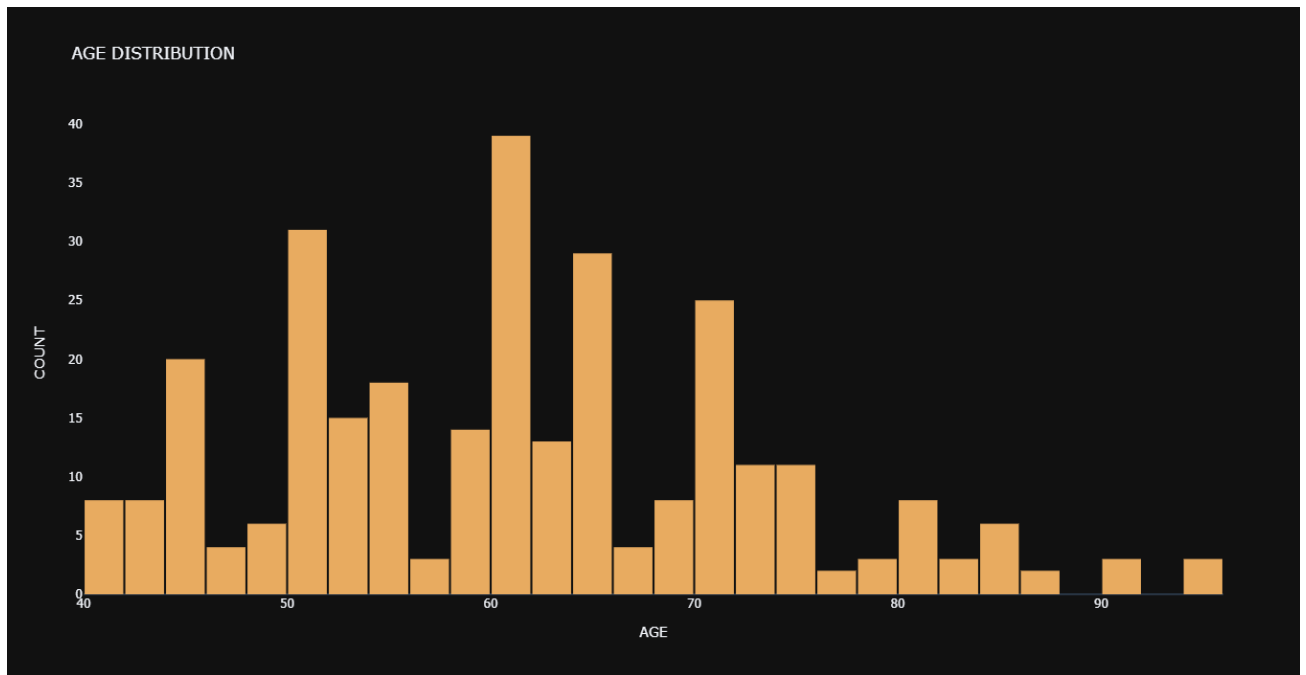
People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.

Data Visualisations for heart Failure Dataset.

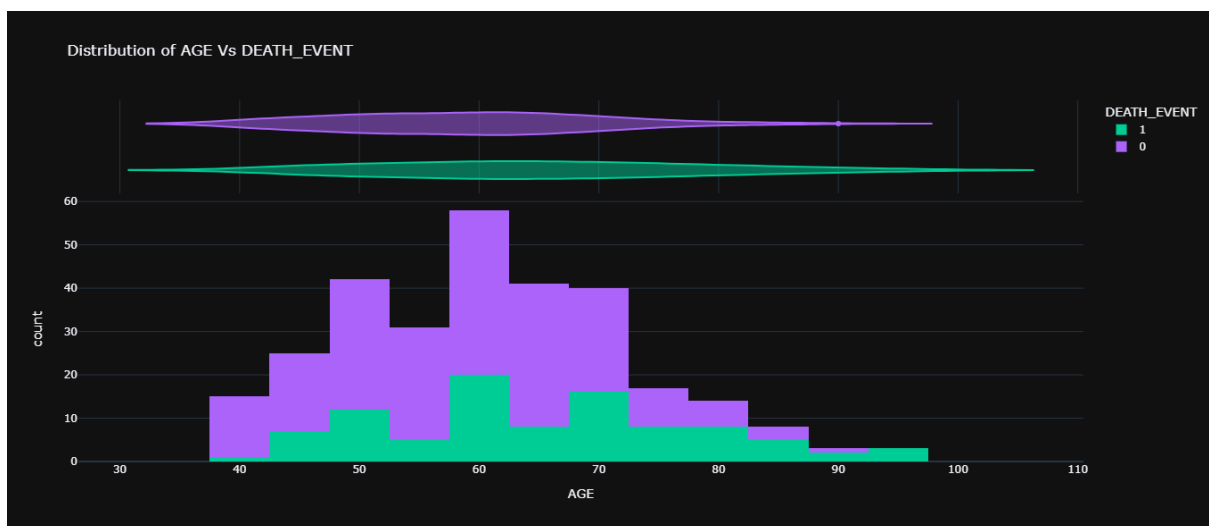
1) Boxplots for 3 most important features



As we can see above there are only 2 outliers in the dataset for ejection fraction. For data analysis purposes these two will be removed.

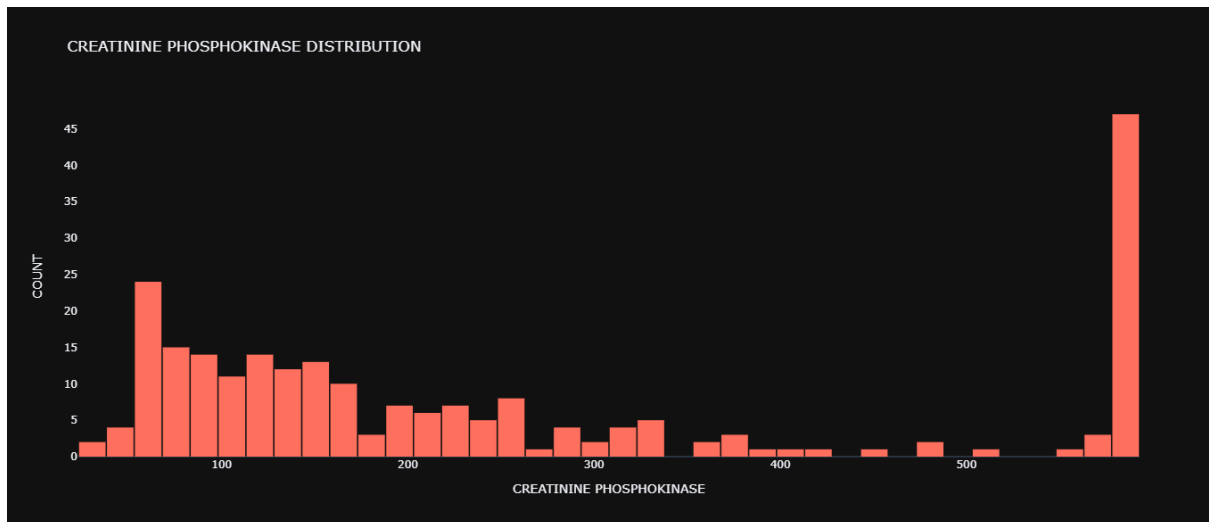


Now we will categorize the above histogram by death event.

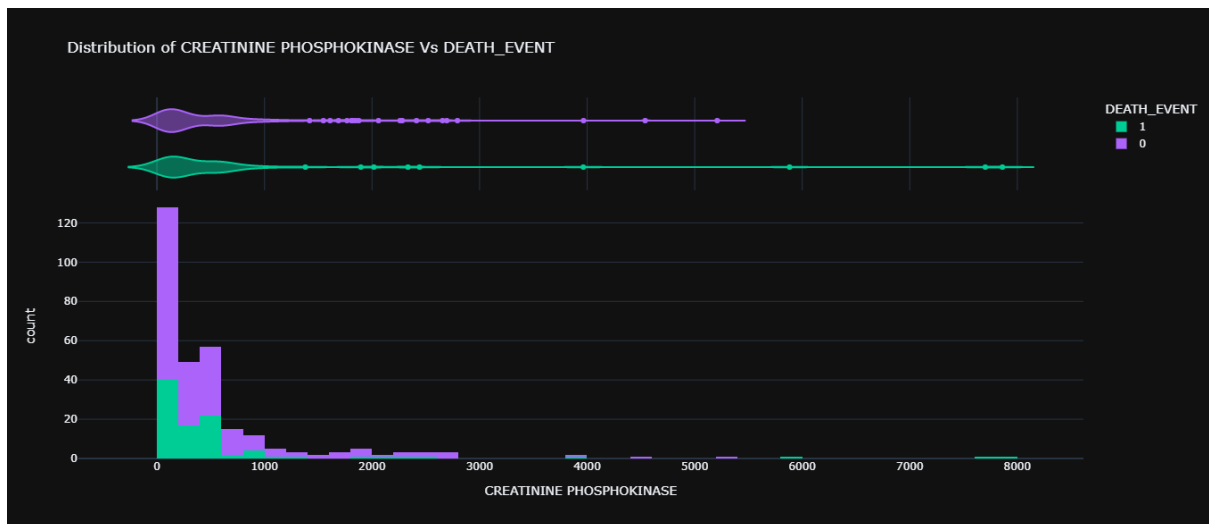


Wider sections of the violin plot represent a higher probability of observations taking a given value, the thinner sections correspond to a lower probability and the value of probability is given by kde value for given x

Distribution of Creatinine Phosphokinase

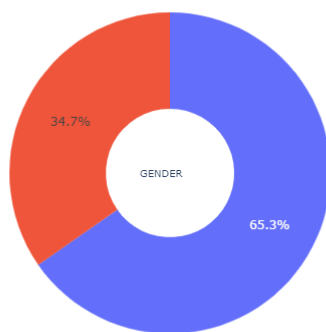


Now we will categorize the above histogram by death event

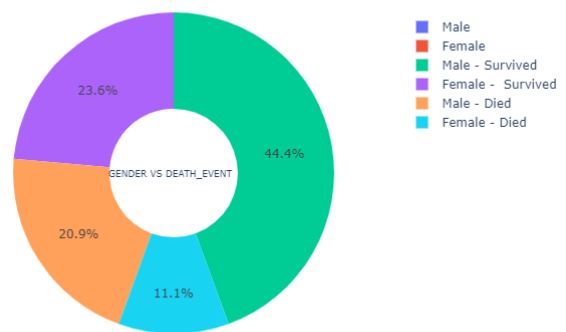


Gender distribution vs death event in dataset

GENDER DISTRIBUTION IN THE data



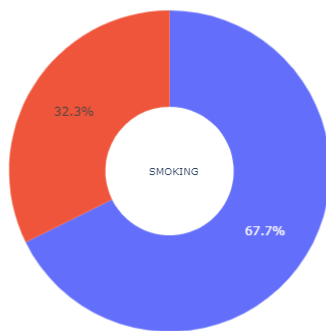
GENDER VS DEATH_EVENT



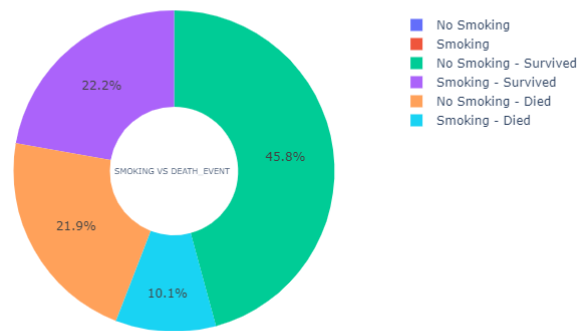
Insight: From the above subplot we can conclude that in our dataset 65.3% are MALE (out of which 44.4% survived and 20.9% died) and 34.7% are FEMALE (out of which 23.6% survived and 11.1% died).

Smoking distribution vs death event in dataset

SMOKING DISTRIBUTION IN THE data



SMOKING VS DEATH_EVENT



Insight: From the above subplot we can conclude that in our dataset 67.7% do not SMOKE (out of which 45.8% survived and 21.9% died) and 32.3% do SMOKE (out of which 22.2% survived and 10.1% died).