❤️Helthcare Domain❤️

\*\*Heart disease\*\* or \*\*Cardiovascular disease (CVD)\*\* is a class of diseases that involve the heart or blood vessels. Cardiovascular diseases are the leading cause of death globally. This is true in all areas of the world except Africa. Together CVD resulted in 17.9 million deaths (32.1%) in 2015. Deaths, at a given age, from CVD are more common and have been increasing in much of the developing world, while rates have declined in most of the developed world since the 1970s.

So, in this kernel, I have conducted \*\*Exploratory Data Analysis\*\* or \*\*EDA\*\* of the heart disease dataset. \*\*Exploratory Data Analysis\*\* or \*\*EDA\*\* is a critical first step in analyzing a new dataset. The primary objective of EDA is to analyze the data for distribution, outliers and anomalies in the dataset. It enable us to direct specific testing of the hypothesis. It includes analysing the data to find the distribution of data, its main characteristics, identifying patterns and visualizations. It also provides tools for hypothesis generation by visualizing and understanding the data through graphical representation.

To visualize the data import matplotlib,Seaborn,pandas and Streamlit for creating frontend.

The dataset contains several columns which are as follows -

- age : age in years

- sex : (1 = male; 0 = female)

- cp : chest pain type

- trestbps : resting blood pressure (in mm Hg on admission to the hospital)

- chol : serum cholestoral in mg/dl

- fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

- restecg : resting electrocardiographic results

- thalach : maximum heart rate achieved

- exang : exercise induced angina (1 = yes; 0 = no)

- oldpeak : ST depression induced by exercise relative to rest

- slope : the slope of the peak exercise ST segment

- ca : number of major vessels (0-3) colored by flourosopy

- thal : 3 = normal; 6 = fixed defect; 7 = reversable defect

- target : 1 or 0

Based on the Target from the data set performed Univariate Analysis.

compared other data with Target using Boxplot,countplot.

Interpretation of correlation coefficient

- The correlation coefficient ranges from -1 to +1.

- When it is close to +1, this signifies that there is a strong positive correlation. So, we can see that there is no variable which has strong positive correlation with `target` variable.

- When it is clsoe to -1, it means that there is a strong negative correlation. So, we can see that there is no variable which has strong negative correlation with `target` variable.

- When it is close to 0, it means that there is no correlation. So, there is no correlation between `target` and `fbs`.

- We can see that the `cp` and `thalach` variables are mildly positively correlated with `target` variable. So, I will analyze the interaction between these features and `target` variable.

Using this Correlation plot Heatmap and Pair plot.

checked for Outliers.

In this kernel, we have explored the heart disease dataset. The feature variable of interest is `target` variable. We have analyzed it alone and check its interaction with other variables.

We have also discussed how to detect missing data and outliers.