Data Visualization on Heart Disease Dataset

**Prerequisites for this project:**Pandas, Seaborn, Matplotlib.

Data visualization is a powerful tool for understanding and communicating insights from a dataset. It can help you to identify patterns, trends, and relationships in your data, and can be a useful way to communicate your findings to others.

There are many different types of visualizations that you can use, and the best choice for a particular dataset will depend on the nature of the data and the insights you are trying to convey. Some common types of visualizations for exploring a heart disease dataset include scatter plots, line plots, bar plots, and histograms.

Scatter plots are useful for visualizing the relationship between two numerical variables. For example, you could use a scatter plot to visualize the relationship between age and cholesterol levels in a heart disease dataset.

Line plots are similar to scatter plots, but they show the trends in the data over time. For example, you could use a line plot to visualize changes in blood pressure over time.

Bar plots are useful for comparing the values of a categorical variable between different groups. For example, you could use a bar plot to compare the prevalence of heart disease among different age groups.

Histograms are useful for visualizing the distribution of a numerical variable. For example, you could use a histogram to visualize the distribution of cholesterol levels in a heart disease dataset.

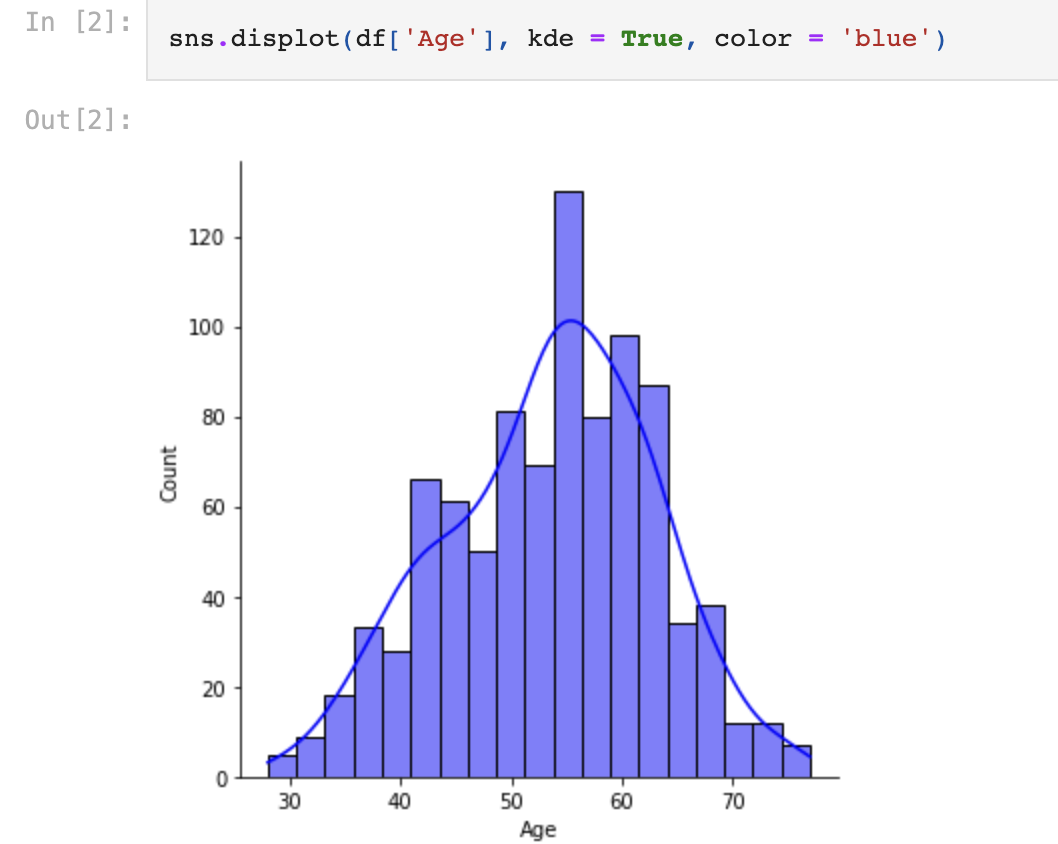
Overall, the key to effective data visualization is to choose the right type of plot for the data you are working with, and to use clear and informative labels and titles to help communicate your insights to others.

We are going to divide the project into 6 parts:

* Heart Disease EDA - Age (DistPlot)
* Heart Disease EDA - Categorical Columns (Pie Charts)
* Heart Disease EDA - ViolinPlot
* Heart Disease EDA - Correlation (HeatMap)
* Heart Disease EDA - Corrlation (PairPlot)
* Heart Disease EDA - Correlation - (JointPlot)

**Heart Disease EDA - Age (DistPlot)**

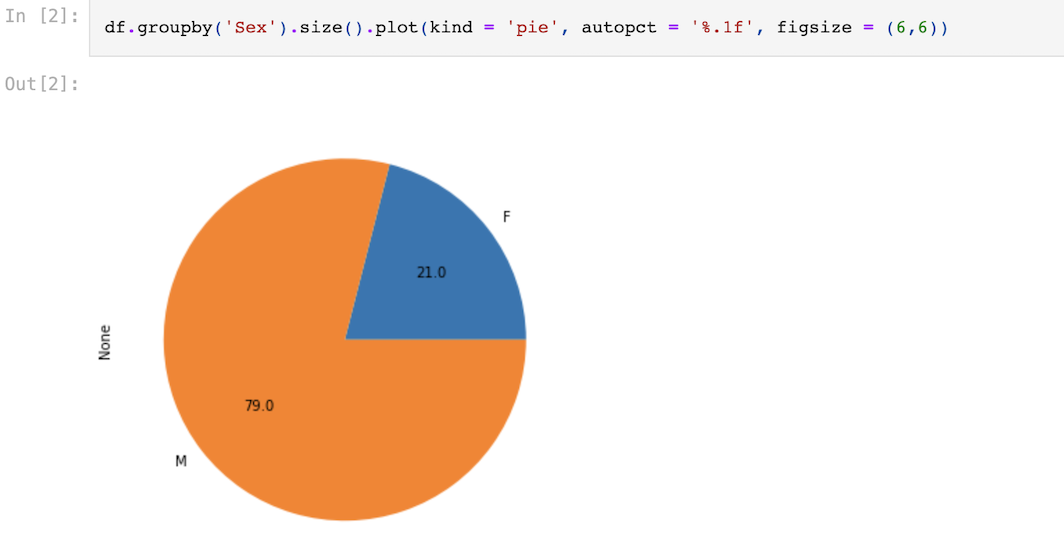
In seaborn, a distplot is a function that plots a histogram, kernel density estimate (KDE) plot, and/or rug plot on a single figure. It is used to visualize the distribution of a single continuous variable.



So after doing the visualization we can infer that minimum age is around 30 and maximum age is around 80. Maximum of the data falls within the range of 40 to 70.

**Heart Disease EDA - Categorical Columns (Pie Charts)**

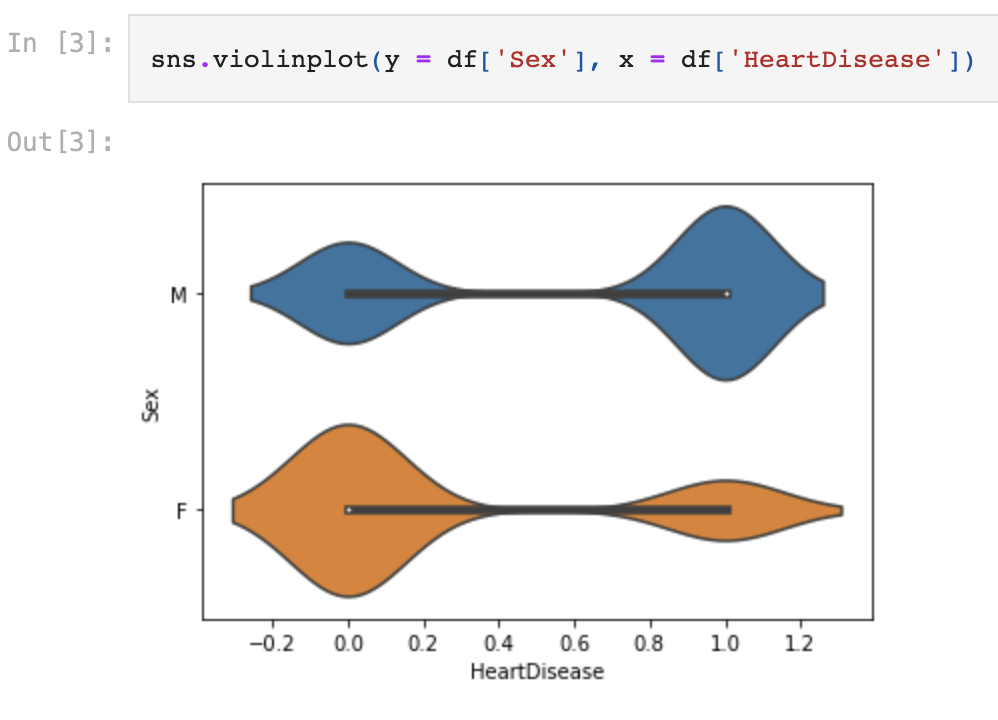
A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice is proportional to the quantity it represents. Pie charts are generally used to show the distribution of a categorical variable.



With the help of pie chart we can see that 79% of the people who are having heart disease are males and 21% are females in our dataset.

**Heart Disease EDA - ViolinPlot**

In seaborn, a violinplot is a graphical representation of a continuous distribution, showing the probability density of the data at different values. It is a combination of a box plot and a kernel density plot, with a rotated kernel density plot on each side.

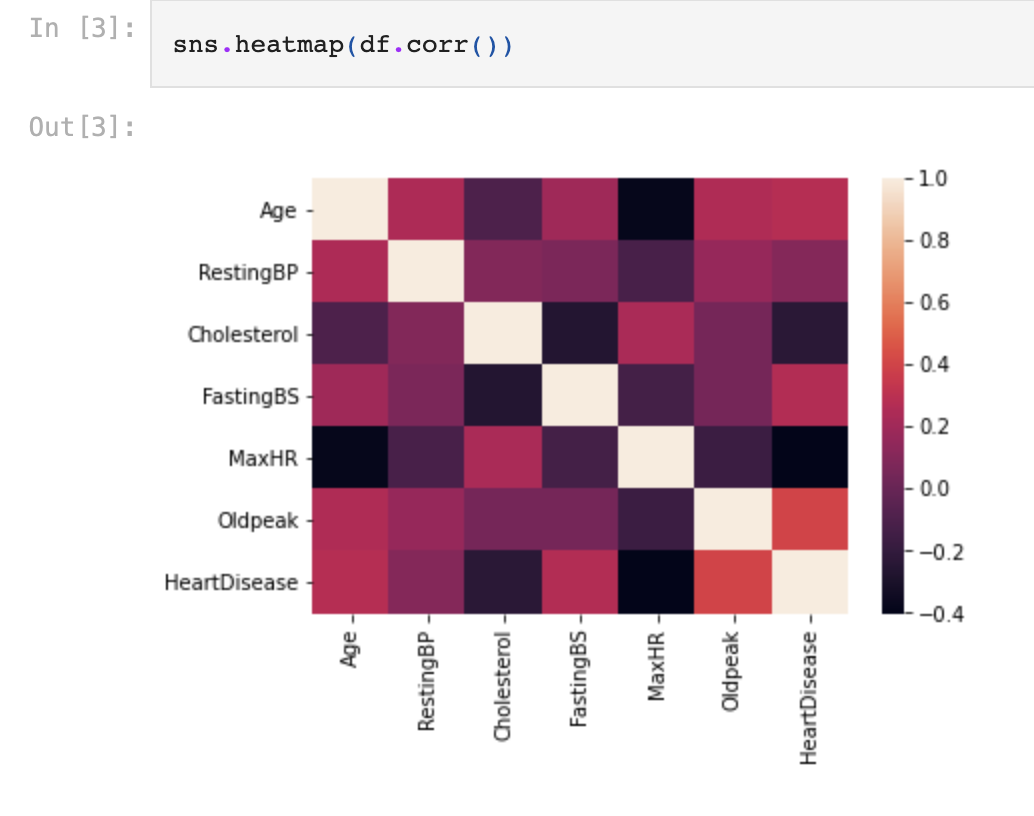


So from this we can see there are more males who are having the chances of having a heart disease compared to females.

**Heart Disease EDA - Correlation (HeatMap)**

A correlation heatmap is a graphical representation of the correlation matrix of a dataset, which shows the correlation coefficients between the different variables in the dataset. The values in the matrix are represented as colors, with darker colors indicating a stronger positive or negative correlation.

The values in the matrix are the Pearson correlation coefficients between the different variables, with 1 indicating a strong positive correlation, -1 indicating a strong negative correlation, and 0 indicating no correlation. The colors in the heatmap represent the strength of the correlation, with darker colors indicating a stronger correlation.



**Heart Disease EDA - Corrlation (PairPlot)**

A correlation pairplot is a graphical representation of the pairwise relationships between variables in a dataset. It is a matrix of scatter plots, where each scatter plot shows the relationship between two variables. The diagonal of the matrix is a histogram or kernel density plot showing the distribution of one of the variables.

The scatter plots on the upper and lower triangles of the matrix show the relationships between pairs of variables, and the histograms or kernel density plots on the diagonal show the distribution of each variable.

You can customize the appearance of the pairplot by setting various options, such as the color map, marker type, and histogram bin width. You can also plot only a subset of the variables by specifying a list of column names in the **vars** argument.

Correlation pairplots are useful for visualizing the relationships between variables in a dataset and identifying potential correlations that may be of interest. They can also be used to identify variables that are highly correlated, which may be indicative of multicollinearity in a statistical model.

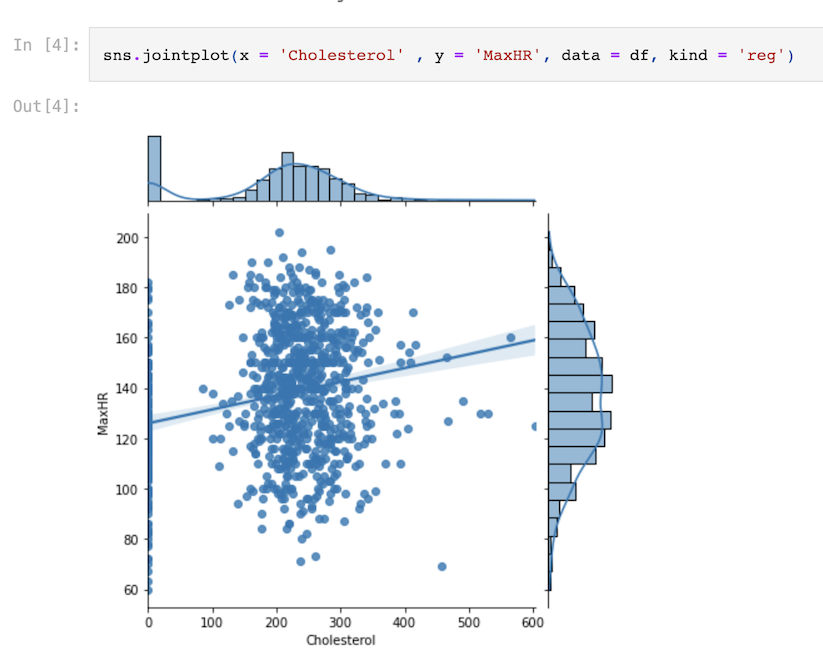
**Heart Disease EDA - Correlation - (JointPlot)**

A correlation jointplot is a graphical representation of the joint distribution of two variables, along with their individual distributions. In seaborn, you can create a jointplot using the **jointplot** function.

The plot includes a scatterplot of the data, with a regression line showing the trend, as well as histograms or kernel density plots of the individual variables on the x- and y-axes.

You can customize the appearance of the jointplot by setting various options, such as the marker type, color map, and bin width. You can also specify the type of plot to use for the individual variables, such as a histogram or kernel density plot.

Jointplots are useful for visualizing the relationship between two variables and the individual distributions of the variables. They can also be used to identify patterns in the data and to estimate the strength of the relationship between the variables.



So from this we can see based on Cholesterol how the MaxHR is changing.