EXPLORATORY DATA ANALYSIS

Since the data was already cleaned, we went ahead with refining it for analysis and presented out findings with graphs and tables.

EDA:

1. Relation between people with high BP and heart disease by gender-

To analyze this relation between people of certain gender which have both heart disease and high blood pressure, we considered three columns- HighBP, HeartDiseaseorAttack, and Sex. At first,we modified column values from '0' or '1' to 'Yes' or 'No' for better visualization and totaled them by heart attack occurrence and gender. So we have total 4 categories on X-axis- male with high BP, female with high BP, male with no BP issues and female with no BP issues. On Y-axis, we plotted total counts of heart disease cases in that category. We can see the graph as below-
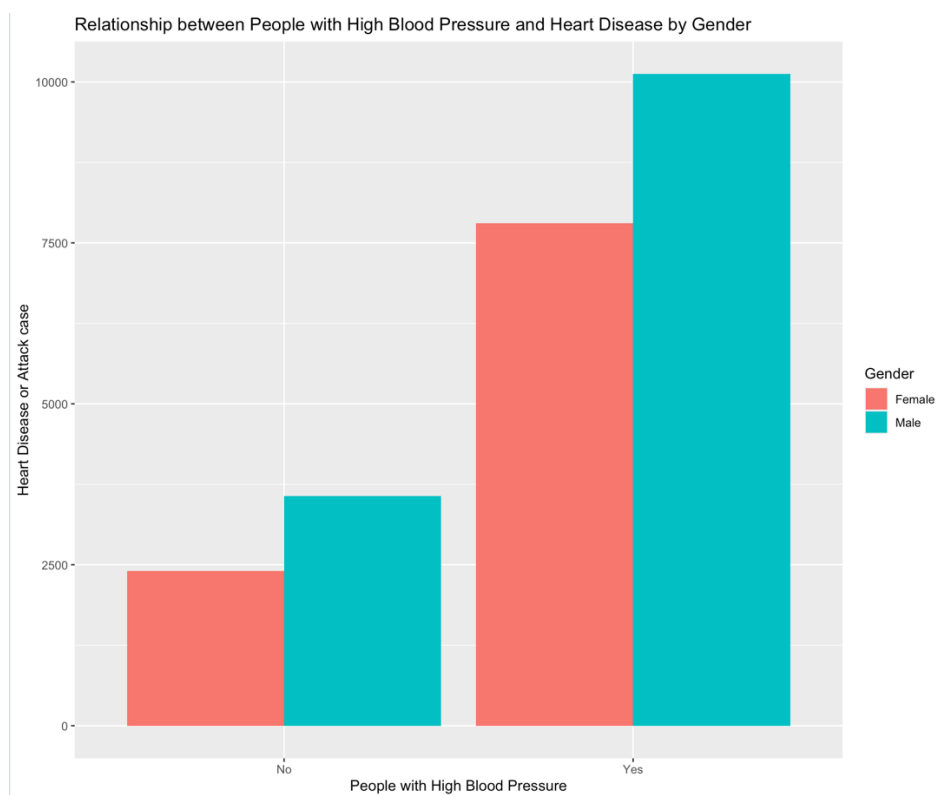
Graph-

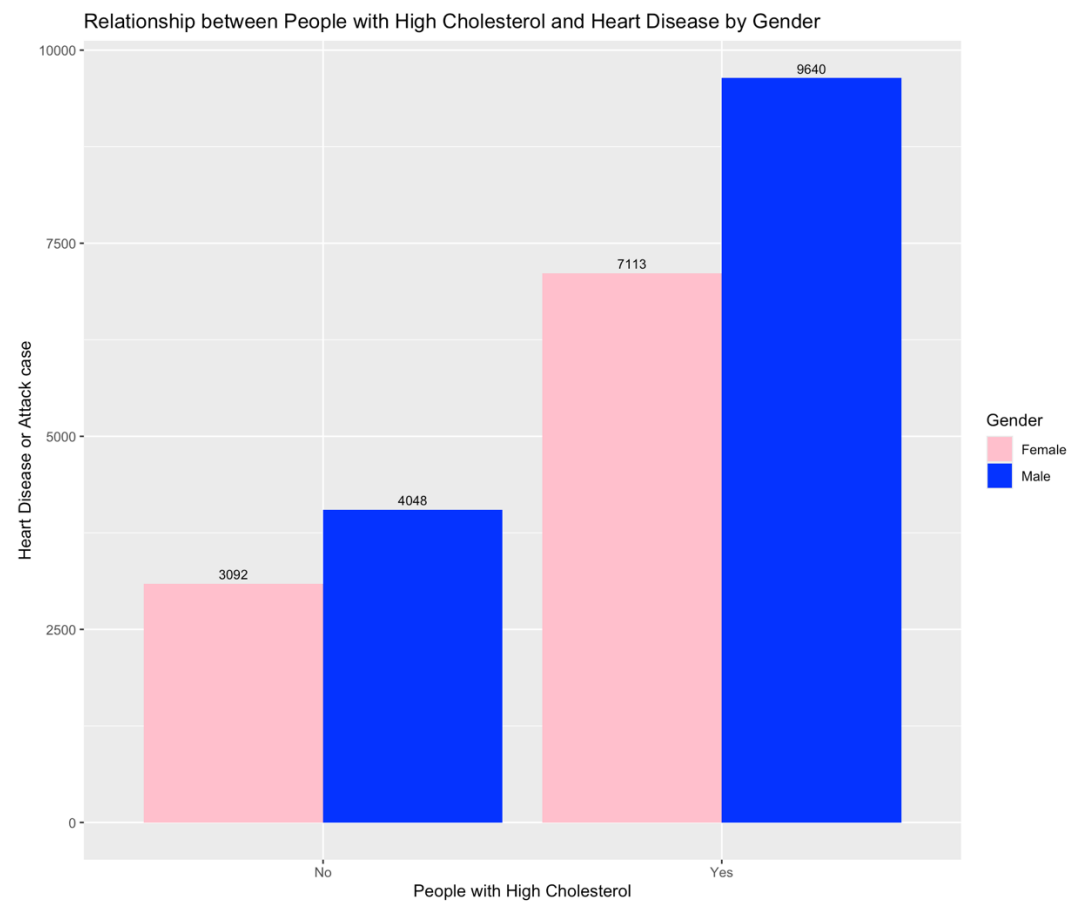Relationship between People with High Blood Pressure and Heart Disease by Gender

Finding - From the graph it is clear that occurrence of heart disease is more likely to be present in male compared to female irrespective of the High BP issues. Furthermore, having high BP adds to the increased risk of developing heart diseases significantly as can be seen by bars in 'Yes' columns in above graph.

2. Relation between people with high cholesterol issues and heart disease occurrence by gender-

To analyze this relation between men and women with high cholesterol and heart disease occurence, we considered three columns- HighChol, HeartDiseaseorAttack, and Sex. First, we modified 'HighChol' column values from '0' or '1' to 'No or 'Yes' values in a new column. Then categorized those further based on 'Sex' column values in 2 categories– 'Male' and 'Female'. We then totaled them by heart attack occurrence and plotted both in a graph. So we have total 4 categories on X-axis- male with high cholesterol concern, female with high cholesterol concern, male with no cholesterol concern and female with no cholesterol concerns. We can see the graph as below-

Graph-

Relationship between People with High Cholesterol and Heart Disease by Gender

Finding - From the graph it is clear that occurrence of heart disease is more likely to be present in male compared to female irrespective of the High cholesterol issues. Furthermore, having high cholesterol adds to the increased risk of developing heart diseases significantly as can be seen by bars in 'Yes' columns in

above graph.

3. Relation between people heart disease occurrence with age –

For this analysis, we first divided the 'Age' column values in 13 age ranges, each of 7 year span and stored the total count for each range in a temp column of ageRange. We used this column for plotting on Y-axis, with total count of heart disease occurrence on X-asix. Graph can be seen below-

Graph-
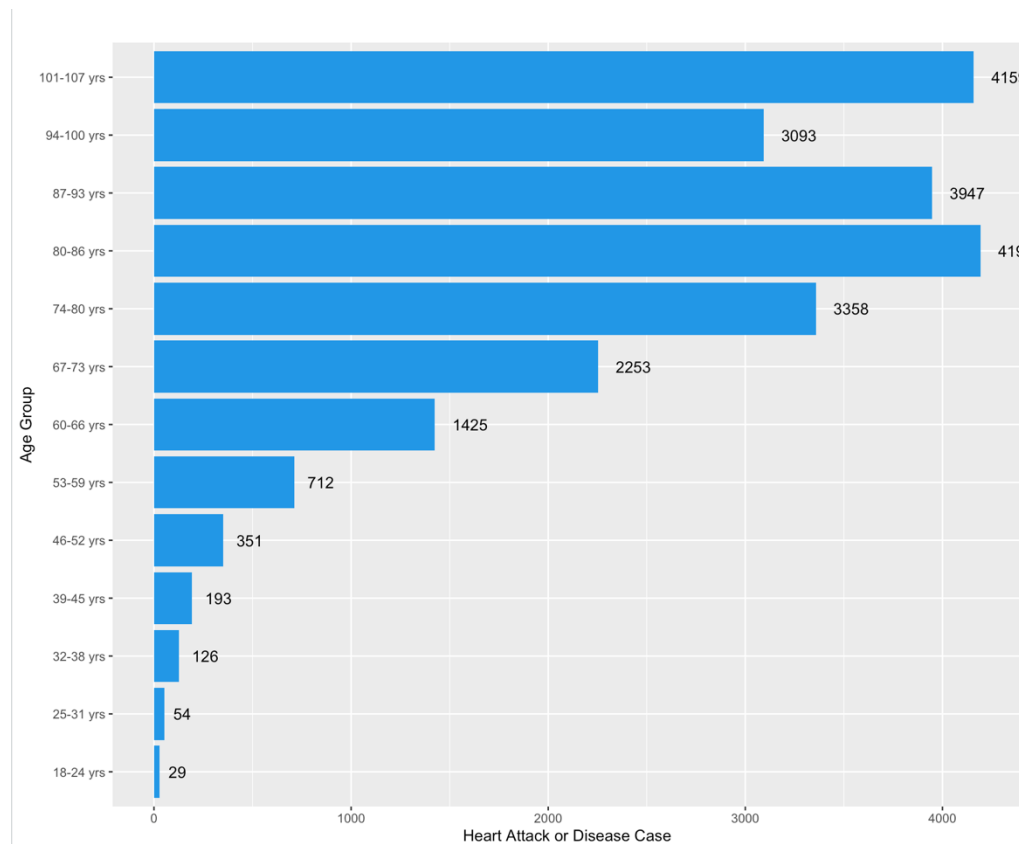
Finding - In the graph, we can see the possibility of developing a heart disease increases exponentially as one ages. Although the data might suggest that the risk peaks at 80-86 years and then drops ahead till 100 years, it is just because we've lesser data points in that range (fewer people older than 80 years of age).

4. Correlation plot for all variables–
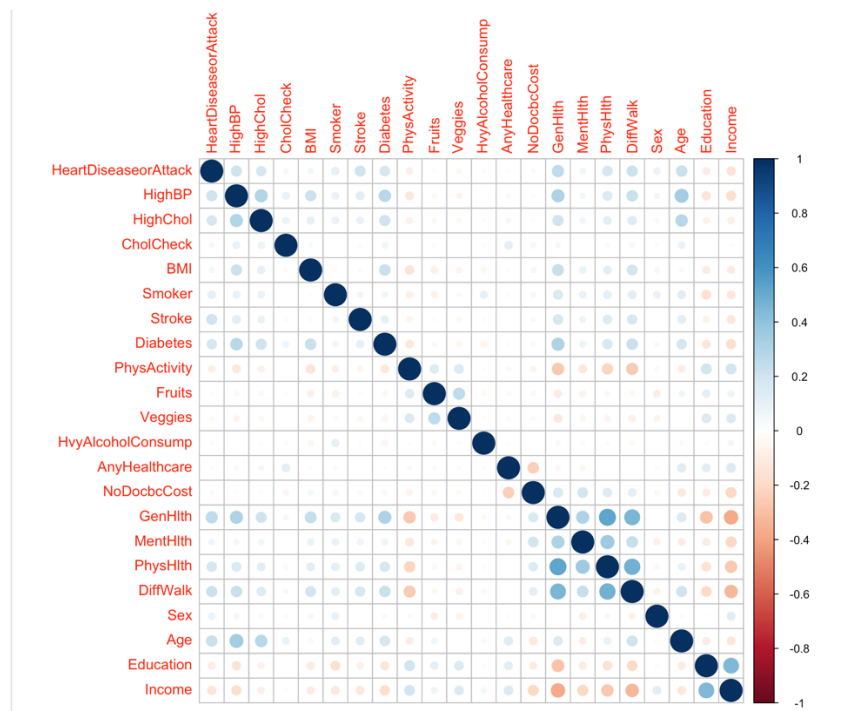
5. Heat Map for all the 22 Variables present in our dataset: