**Health Care Project**

**Steps performed:**

1. **Data Exploration:**

1. Performed descriptive analysis and recognised the statistical features of the data and also the null/missing values.

2. Visually explored the variables using histograms.

3. Treated the missing values by replacing the Zeros with the appropriate **Median values.**

4. There were integer and float data type variables in the dataset, so created a count (frequency) plot describing the data types and the count of variables. Only 2 variables had Float data type.

1. **Data Exploration:**

1. Checked the balance of the data by plotting the count of outcomes by their value. The records with the Outcome as **‘0’** were more than **‘1’**, but the difference is not unusual.

2. Created scatter charts between the pair of variables to understand the relationships. The correlation between variables like **‘BMI’ and ‘Skinthickness’ is Positive in nature.**

3. Performed correlation analysis and visually explored it using a heat map.

1. **Data Modelling:**

Firstly, the data is scaled by using the Standard Scaler. Then KNN Algorithm is applied on this scaled data.

Recall value in Classification report is given importance and the accuracy is improved by changing and trying different K-Neighbor values.

Then Random Forest Algorithm is also applied to check the performance. The Accuracy and Recall values are better with KNeighbors algorithm, with the prediction accuracy of 73%.

1. **Data Reporting:**

**In Tableau :**

1. Created a Pie chart to describe the diabetic or non-diabetic population.

2. Scatter charts prepared between relevant variables to analyze the relationships.

3. Created various Histogram or frequency charts to analyze the distribution of the data

4. Heatmap of correlation analysis created among the relevant variables.

5. Created bins of these age values: 20-25, 25-30, 30-35, etc. and analyzed different variables for these age brackets using a bubble chart.