**TITLE: Visualizing Data using Python**

**Introduction**

The project focuses on visualizing quantitative data using certain libraries in python.

**Objective**

Data related to heart diseases was collected from Kaggle and various graphs were plotted to understand the relation between various factors.

**Methodology**

Used python libraries like scipy, seaborn, matplotlib to plot graphs and understand the data. Various graphs like scatterplot, histogram etc. were plotted.

About the data considered

1. The data has been taken from Kaggle
2. It contains numerical data
3. There is total 14 columns having various

Data contains

1.Age of people in years

2. Gender: 1 for male and 0 for female

3. CP: Type of chest pain Chest Pain

4. trestbps: Resting Blood Pressure (in mm Hg)

5. chol: serum cholesterol in mg/dl

6. fbs: (fasting sugar>120mg/dl)

(We will consider 0 as False and 1 as True)

7. restecg: Resting electrocardiographic results

8. thalach: maximum heart rate achieved

9. exang: Exercise induced angina (1: Yes 2: No)

10. oldpeak: ST depression induced by exercise relative to test

11. slope: The slope of the peak exercise relative to rest

12. ca: Number of major vessels (0-3) colored by fluoroscopy

(fluoroscopy: similar to radiography and X-ray computed tomography (X-ray CT) in that it generates images using X-rays.)

13. thal: 3: normal 6: fixed defect 7: reversable defect

14. target: have disease or not (1: yes 0: no)

Understanding our data:

Certain medical expressions are considered in the data file. Few important variables are defined here for reference.

1. **restecg**: Resting ECG is a simple, quick test. It can detect certain heart conditions such as hypertrophy id heart, ischemia, myocardial infraction, etc. Normal range up to 120ms.
2. **fbs**: Fasting blood sugar: Blood sample taken after overnight fast. Less than 100 mg/dL = **normal**. Between **110**–125 mg/dL = impaired **fasting glucose** (i.e., prediabetes) Greater than 126 mg/dL on two or more samples = diabetes.
3. **Exang**: Exercise induced angina: When you increase the demand for oxygen, such as when you **exercise**, **angina** can result. Stable **angina**. Stable **angina** is usually triggered by **physical activity**.
4. **Thal**: Thallium Stress Result
5. **chol**: serum cholesterol: normal – **less than 200mg/dL**

Understanding graphs plotted

1. **Graph of target:** Target variables represents if the person had any previous disease or not. A graph is plotted to understand how much of the data set had previous history of diseases.
2. **Gender:** To understand the ratios of male female patients we have in our data.
3. **Age vs Target**: The plot helps to understand the relation and pattern between age of a person and appearance of a disease.
4. **Age vs Talach**: Helps in interfering the relation between age and maximum heart rate achieved by a person
5. **Fbs vs Target**: Helps in understanding how having disease can fluctuate fasting sugar levels.

**Result**

A pattern and many strong relations between the variables considered were observed.