Assignment - 1 (Intermediate Level)

Dataset:

Dataset is in the same zipped folder by name Heart.csv.

Heart.csv :

The dataset contains 14 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 cholesterol 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 fluoroscopy
- thal : 3 = normal; 6 = fixed defect; 7 = reversible defect
- target : 1 or 0

Questions to be solved:

1. How many are suffering from heart disease? Also plot the stats.
2. How many males and females have heart disease out of total?
3. Visualize frequency distribution of the thalach variable and find what's the heart rate and heart disease relation?
4. Find correlation matrix for all the variables with target.
5. Find Mean,Min & Max of age and plot its distribution.
6. Age and its relation to heart disease. Are young people more prone to heart disease?
7. Plot chest pain type pie chart.
8. What is the max heart rate achieved in non heart disease patients?

Reference work on the same data.

https://www.kaggle.com/ronitf/heart-disease-uci/kernels