# CMSE 830 midterm project

Shuangyu Zhao

1. Goal

This topic can help to increase the accuracy of primary diagnosis and to prevent the worsening by early detection and treatment. In this project, I explore the relationship between heart disease and the other appearances, and predict the heart disease by other attributes.

2. What do you learn?

In the app, you can know where I got this dataset, and the introduction of the dataset. You can know how I preprocess the data and the process of EDA, which concludes this dataset is good for classification. Because, the scatter plots shows targets can be distinguished by the attributes. And I also fit a KNN prediction model, which can help user to diagnosis whether has heart disease preliminarily.

3. Visualization

I use heatmap to represent missing data and the correlations between attributes(seaborn). The box plots show the outliers(seaborn). I add color on the tables(dataframe.style), which can help me tell my story better. Scatter plot is used to check whether this dataset is good for classification. I use altair to draw my 2d scatter plot and plotly.express to draw my 3d scatter plot. When you put your mouse on the scatter plots, the information of this point will be shown and you can zoom in and zoom out. Stacked bar chart shows the proportion of target=0 and 1 in every element of the attributes with discrete values, which also can help to check whether this is good for classification, and show some relations between heart disease and attributes. I use streamlit.bar_chart() to draw this. When mouse is put on the bar, the information is shown either.

4. Preprocess

First, I use heatmap to visualize the missing data, which shows no missing data. And them, I use box plot to visualize the outliers. All attributes with continuous values have outliers. So I delete all the outliers.

5. App design

The siderbar contain the navigator/ catalogue of my web app. I use container to make my app more better constructed. In “data analysis” chapter, you can choose the option in the selectbox and see the stacked bar chart you like. In “prediction” chapter, slider and radio are used to input the user’s information, and you can use them to predict whether you have heart disease..