## 心脏病个人指标数据集处理

### 数据预处理

将数据集（heart\_2020\_cleaned.csv）处理成可以使用的状态，如所有类别列（HeartDisease列等）处理成数值列（如1代表Yes，0代表No）等，或将数据中你认为不合适的数值所在的行或列删除掉。

### 数据可视化

使用合适的库进行数据的可视化，展示你认为有意义的数据，包括但不限于（折线图、柱状图、饼图、词云图等）

* 1. 绘制BMI指数的箱线图，查找BMI指数的标准，绘制出受访者肥胖人数和正常BMI人数和偏瘦人数的饼图。
  2. 绘制每24小数睡眠时间的词云图。
  3. 绘制出受访者各个年龄段的人数柱状图，如果人数过多或者不平均，尝试使用占比进行柱状图绘制。
  4. 其他你认为有必要展示的图形。

### 数据分析

* 1. 使用合适的方法，给出心脏病和哪些因素关系较大，相关程度如何，分析哪种生活习惯对心脏病得病影响最大。
  2. （选做）使用sklearn进行分类算法模型（SVM、随机森林等）的设计和训练，研究不同列对心脏病的影响；
  3. （选做）使用test.csv中的数据测试训练好的模型，并给出模型的评价指标数值。
  4. （选做）在未得心脏病的受访者中，使用上述模型，设计模型求解出得心脏病的可能性（提示：使用predict\_proba方法）。