Homework10数据处理说明文档

1. 写了一个爬虫程序将所有数据爬取到本地。然后投放到六个列表里。
2. 考虑每个数据文件是否有缺失值，如有缺失值，则将其去掉。
3. 将采集时间过长或者过短的文件过滤掉，由于采集频率是5HZ，所以只需要用数据条数除以5即可得到时间，然后我的标准是保留10分钟到30分钟的文件。
4. 将前四步得到的文件列表进行分类，按文件名可分为6类（一共三个大类，每个大类分为健康和焦虑）