**报告**

**学号: 姓名: 班级:**

1. **目标**

针对实际获取的某大楼光伏电站的数据，判断数据中存在的异常。

1. **数据**

数据文件pv.csv记录了来源于某大楼的光伏电站的数据，其中数据时间段为2020/5/7 15:00:00---2023/10/25 16:15:00，采样频率为15min，每个时刻获取了两个物理量：功率和发电量，共121542个数据时刻。

1. **任务**
2. 分别画出两个物理量的趋势折线图以及散点图、盒图，并分析数据中可能存在的异常；
3. 直接利用某时刻的信息来判断异常。如现有121542个数据时刻形成同样大小的数据样本，每个样本由两个属性：功率和发电量组成，用两种异常检测的方法**如DBSCAN或一类SVM判断异常样本（也可采用较先进的方法）**，并进行对比分析。**（注意：数据中存在遗失值的情况）**
4. 利用功率这一物理量的数据趋势信息来判断异常。一般针对功率这一时间序列数据采用滑动窗口+步长的方式形成多属性样本，采用两种异常检测的方法如DBSCAN和一类SVM判断异常样本，并进行相应的对比分析。

注：利用时间序列的数据采用滑动窗口+步长的方式生成样本的代码见dataset\_power.py，修改窗口、步长的大小，统计样本数量，并分析异常检测的结果

4）探索同时利用功率和发电量两个物理量的数据变化趋势信息来判断异常样本。**（该任务选做，有一定程度的加分）**