《大数据计算机基础》2022 秋季学期作业 4

本次作业数据链接如下：

https://disk.pku.edu.cn:443/link/A4BF4410A41CA90BF78E710CEDFADFFC

有效期限：2022-11-03 23:59

本次作业要求如下：

1、 数据为货车 GPS 数据。GPS 数据可能出现遮蔽、飘移等多种异常，请定义算法整理数据，可考虑滑动窗口，即时间轴上，用一定大小的窗口移动处理；

2、 数据格式为：以英文逗号间隔的文本文件。第一列数据为脱敏车号、第二列数据为日期时间，第三、四列数据为经纬度；

3、 数据处理目标 1：判断一辆货车在指定时间段内的是否位移，位移时长、位移距离。直接两点距离和或多个点累积距离之和，都常与真实距离相差甚大；

4、 数据处理目标 2：寻找数据密级区域。一般要求是找出 1 平方公里范围数据最多的区域，并利用该数据图形化显示该数据。在此基础上，形成路网连线。附加要求（完成可加分，不完成不减分）：寻找 100 平方米范围内的数据最密集区域，并根据数据特征自动识别数据范围，即识别密集区与非密集区边界（如同城市建成区和郊区计算是利用手机信令数据），本要求要考虑停车区排除，即要识别停车区并将其排除在外；

5、 数据量较大，可以编写一个程序比如：topN.py，读取最前面 N 行（比如：10 万行），利于调试程序。调试完毕后，跑全部数据。

6、 上述数据禁止外传。如外传，需承担其相应后果。