

20200305 作业

1、从以下链接下载数据：

北大网盘：<https://disk.pku.edu.cn:443/link/79F94A579E75FDC15FD8133A4B69B761>

百度云盘：<https://pan.baidu.com/s/11OhcikgUaAUo4QWMcwsQFw> 提取码：fclb

2、数据是北京地铁刷卡数据。第一行是数据标题，含义数据：

6 a) carType: SUB 是唯一取值，表明数据是地铁；

1 b) iclD: 交通卡编号，每个编号对应一张卡；

2 c) cardType: 卡类型

3 d) tradeType: 交易类型

4 e) UpLine: 上车线路编号

5 f) UpTime: 上车时间

6 g) UpStation: 上车站编号

7 h) DownLine: 下车线路编号

8 i) DownTime: 下车时间

9 j) DownStation: 下车站编号

10 k) State: 弃用

3、为便于程序调试，建议编写名为 TopN.py 程序，可以根据输入行数，获得前 N 行数据，调试过程中，先用少量数据，待语法、逻辑正确后用全部数据运行。本程序不用提交。

4、计算乘车耗时人数。以分钟单位，计算出每个时间段内的人数。如 1 分钟：10 人，10 分钟 500 人，30 分钟 5000 人。0-59 秒按 1 分钟计算，60-119 秒按 2 分钟计算，以此类推，有关时间的计算，如无特别申明，本文档均以此为准。以形成的数据为基准，以最多人数的时间段为中心，向两侧逐分钟拓展(不能有空隙，某个时间段内人数可以为 0，假如 30 分钟人数最多，则分别计算出 29、31 分钟乘车人数；28、32 分钟乘车人数)，将累积人数覆盖到 98%的时间段及人数数据存储到 PeopleInCarTime.txt 文件中。

5、计算在外时长，同天最后一次下车减去同天最早一次上车为工作在外时长，只计算间隔时间大于 180 分钟的数据。以 5 分钟为时间间隔，计算出每个时段人数。将人数从高到低排序，将人数覆盖到 98% (统计范围的人数) 时间段数据存储到 PeopleOutHomeTime.txt 文件中。

6、计算站对流量排名。计算两个站之间进出人数排名前 200 的组合。如进北京大学东门站和出北京南站以及进北京南站和出北京大学东门站之间的人数。将上述数据存储到 PeopleBtwStation.txt 文件中；

7、4、5、6 分别对应一个程序，命名为学号_题号.py，题号取值为 4、5、6。将上述数据转存到 Excel 之中，合并成一个文件，学号_Data.xlsx，并形成对应的柱状图。

8、提交内容包括：三个源程序代码和 Excel 文件，压缩为学号_Work03.zip。

9、作业提交截止时间：2020-03-11 18:00 前。