

# 题目

数据大学第一届毕业生就要诞生了，但是由于管教务系统服务器的小哥哥昨晚喝大了，然后服务器被不知名的神秘力量攻击到宕机了，现在还不能恢复服务，预计恢复服务起码是一周后的事情了。但学校教务处必须马上对成绩进行汇总和处理，以便通知不能毕业的学生要补修课程以及提交学生办公室评出优秀毕业生。教务处的邓主任找到你，看你天资聪颖，骨骼精奇，希望你能帮他们处理这件棘手的事情，并说如果你解决不了，那现在你在学的这门课程就需要重修了，还要给你留校察看处分。

为了继续醉生梦死的生活，你不得不接下这个棘手的任务，并从邓主任手上拿到了这些资料。

## 一、三份 Excel 文件

第一份是学生花名册，字段名如下

{班级, 班号, 座号, 姓名, 入学时间, 性别, 生源地} 第二份是学生成绩汇总

{学号, 姓名, 课程, 课程编号, 平时成绩, 卷面成绩, 学分, 录入时间, 教师编号} 第三份是课程表

{课程编号, 学分, 年级, 课程名称, 教师, 教师编号}

## 二、学号

已知学生的学号构成方式如下：入学年份后两位+班号+座号，如座号是个位数，则前面补零组成 2 位数。

如：A 同学 2010 年入学，班号是 1，座号是 2，那么 A 同学的学号是：101002

### 三、绩点

绩点计算公式：绩点=成绩/10-5 计算结果取整平均绩点（GPA）的计算

公式：GPA=(课程 1 学分 x 课程 1 绩点+课程 2 学分 x 课程 2 绩点+.....)/(课程 1 学分+课程 2 学分+.....) 计算结果保留 2 位小数（末尾四舍五入）

如：A 同学今年参加两门课程，成绩如下

课程	学分	成绩
数据分析	4	93
数据挖掘	5	81

那么 A 同学每门课的绩点是

数据分析课程绩点=93/10-5=4

数据挖掘课程绩点=81/10-5=3

GPA=(4x4+5x3)/(4+5)=3.44

### 四、绩点评级

GPA 计算出来后，需要按照绩点进行评级，并以评级后的绩点作为最终的 GPA。绩点评级表如下

成绩区间	绩点	评级
100	5.0	S
[90,100)	4.0	A
[80,90)	3.0	B
[70,80)	2.0	C

[60,70)	1.0	D
[0,60)	0	E

## 五、总成绩计算公式

为了体现学校的人性化，学生的成绩由两部分组成：平时成绩、卷面成绩。

卷面成绩就是期末考试的试卷得分；

平时成绩就是老师看着给，如果你跟老师的关系比较好的话，也许可以给你比较高的分数；

在计算绩点时使用的成绩为：平时成绩  $\times 0.3$  + 卷面成绩  $\times 0.7$

## 六、数据格式化

教务处的老师们只会用简单的 excel 操作，所以为了方便他们对数据进行分类，请按照以下格式

输出文件：

- 1、excel 文件不能带上序号（即 index）；
- 2、年级统一为：一年级、二年级、三年级、四年级；
- 3、班级统一为：一班、二班、三班；

那么问题来了：

- 1、对于本次考试（录入时间大于 2017 年 9 月）最终成绩（用第五提到的算法计算得出的成绩）低于 60 分（ $< 60$  分）的同学本次课程是需要补考的，请将需要补考的同学名单按以下格式

式保存到文件：failed.xlsx

{学号, 姓名, 年级, 班级, 性别, 课程, 课程编号, 平时成绩, 卷面成绩, 最终成绩}

ps: 如果有一位同学挂了多门课程, 记得把他们排列到一起, 方便教务处进行通知。

2、学校规定, 绩点评级为 E 的学生, 按规定是不能毕业的, 请帮忙找出这部分学生的名单,

按以下格式保存到文件: unfortunately.xlsx

{学号, 姓名, 年级, 班级, 性别}

3、学校规定, 大四的同学(2014 年入学)中, GPA 排名占全年级前 10% 的同学可以参加优秀毕业生的评选, 学校希望评选的时候基于公平公正的原则来进行。经过各个部门研究决定: 要不还是做个随机抽样程序来选好了。如果待评选的学生不足 5( $< 5$ ) 人, 那么就都评为优秀; 如果待评选的学生人数超过 5 人( $\geq 5$ ), 那么随机取出 20% 的人作为优秀学生。

并将

优秀毕业生的名单按以下格式存储到文件: perfect.xlsx

{学号, 姓名, GPA, 年级, 班级, 性别} 4、每年辅导员总说: 你们是我带过的最差的一届学生。请你用数据来印证一下这个观点

5、你的发小小明同学由于得罪了高中班主任王老师, 现在还在高中单着, 听说你拿到了学校所有学生的名单, 小明想让你计算一下你们学校

的男女比例大概是什么样子的，各个年级的男女比例又是怎样的。当然还是给个统计图比较容易看啦。

6、胶水儿老师想让你帮忙统计一下她所教的课程各个分数段的分布大概是怎样的，请你用直方图帮她做一下统计。

7、学校在课程设置上把 Python 从入门到精通、计算机网络 这两门课程作为 Python 网络

爬虫的前置课程，但由于学时限制，计算机网络只能跟 Python 网络爬虫同时进行，试探索一下这两门前置课程对 Python 网络爬虫是否存在必要前置的关系，方便教务处调整下一学年的教学计划。

8、学校以往都是把同个地区出来的学生尽量安排住在同个宿舍，一来是方便各个宿舍的学生快速熟络，组队吃鸡比较方便，二来是彼此敦促好好学习。请你通过数据分析印证一下这个方案的实施效果。