**一、分析题目要求：**

总需求：数据分析预测与可视化。——挖掘重要信息，改善学生成绩，服务教务设计，优化学生服务。未来学生的学习行为画像、考试分数、发展潜力方向。

针对以下指标体系设计报表进行展示，形成可视化分析门户，支持按照条件查询工作实现。

详细需求：对三大域进行数据分析。

1. **学生域数据分析**
2. 对学生基本信息进行画像展示分析；包括性别、年级、班级、住址、教师等形成学生画像标签。
3. 对学生学业情况进行描述性统计分析。
4. 学生消费画像分析，通过对学生一卡通消费数据，分析学生消费情况。
5. 学生考勤画像分析。
6. **课程域数据分析**
7. 对课程域相关数据进行描述性性统计分析，汇总各个科目班级历史最高分趋势和历史最低分趋势。
8. 直观反应学生的选课分布，展示7选3全部组合分布情况：包括人数和占比；统计每种7选3组合下，选课重叠的人数分布情况。
9. **群体域数据分析**
10. 对班级域相关数据进行描述性统计分析，分析各个班级平均分排名，指标包括原始分、绝对分；支持总分和单科目；
11. 对班级学生排名分布情况进行统计，包括总分和单科目；
12. 对每个班级每个分数段学生排名人数进行统计;比对不同班级学生考试总分，在不同分数段的分布情况，比对指标包括：原始分、标准分、离均值；
13. 比对不同班级学生考试每个科目，在不同分数段的分布情况，比对指标包括：原始分、标准分、离均值。

**二、观测数据：**

**数据说明**

**1\_teacher.csv:包含了近五年各班各学科的教师信息**  
term:学期  
cla\_id:班级ID  
cla\_Name:班级名  
gra\_Name:年级名  
sub\_id:学科ID  
sub\_Name:学科名  
bas\_id:教师id  
bas\_Name:教师名

**2\_studentinfo.csv:包含了当前在校学生详细信息**  
bf\_StudentID,学生ID  
bf\_Name,学生姓名  
bf\_sex,性别  
bf\_nation,民族  
bf\_BornDate,出生日期（年）  
cla\_Name,班级名（与teacher.csv的cla\_name对应）  
bf\_NativePlace,家庭住址（省市或省）  
Bf\_ResidenceType,家庭类型  
bf\_policy,政治面貌  
cla\_id,班级ID  
cla\_term,班级学期  
bf\_zhusu,是否住校  
bf\_leaveSchool,是否退学  
bf\_qinshihao,宿舍号

**3\_kaoqin.csv**  
kaoqin\_id,考勤ID  
qj\_term,学期  
DataDateTime,时间和日期  
ControllerID,对应考勤类型表里的ControllerID  
controler\_name,考勤名称  
control\_task\_order\_id,对应考勤类型表里的control\_task\_order\_id  
bf\_studentID,学生ID，对应学生信息表  
bf\_Name,学生姓名  
cla\_Name,班级名  
bf\_classid班级ID

**4\_kaoqintype.csv：考勤类型**  
controler\_id,考勤类型id  
controler\_name,考勤类型名称  
control\_task\_order\_id,考勤事件id  
control\_task\_name：考勤事件名

**5\_chengji.csv:学生成绩**  
mes\_TestID,考试id  
exam\_number,考试编码  
exam\_numname,考试编码名称  
mes\_sub\_id,考试学科id  
mes\_sub\_name,考试学科名  
exam\_term,考试学期  
exam\_type,考试类型（对应考试类型表）  
exam\_sdate,考试开始时间  
mes\_StudentID,学生id  
mes\_Score,考试成绩(-1为作弊，-2为缺考，-3为免考)  
mes\_Z\_Score,换算成Z-score（Z-score、T-score、等第 是一种学生成绩评价方式，可以参考网络百科）  
mes\_T\_Score,换算成T-score  
mes\_dengdi：换算成等第  
\*参见：[数据补充说明—“等第”释义，21等第与5等第](https://tianchi.aliyun.com/forum/postDetail?postId=47777" \t "https://tianchi.aliyun.com/competition/entrance/231704/_blank)

**6\_exam\_type.csv:考试类型**  
EXAM\_KIND\_ID:考试类型id  
EXAM\_KIND\_NAME：考试类型名称

**7\_consumption.csv：本学年学生消费信息**  
DealTime,消费时间  
MonDeal,消费金额  
bf\_studentID:对应学生信息表studentid  
AccName,姓名  
PerSex,性别

**特别说明：**  
1.由于人为登记等不可避免原因，某些字段可能存在缺失或者异常值；  
2.从班级名可以看出，从2017年开始学校陆续启用了新校区，2018年新校区统一命名为型为“白-高二（01）”和“东-高二（01）”的班级名；  
3.考勤类型中的“校服[移动考勤]”指的是没穿校服；  
4.允许使用的其他数据包括但不限于任何其他外部公开数据或常识数据，如天气数据、日历节假日数据等，但使用的外部数据链接或者原始数据需要同步提交评审。

**存在问题 && 解决方案**

**1）有选手问'exam\_sdate'”为'2019-01-22'和'2019-01-26'的数据的考试科目id和名称都是缺失值**

缺失部分是人为因素，可能是因为没有及时录入，或者其他录入原因

**2）有的班级名称带IB的是啥意思？**

lb是ibdp的简称，说明这个班级是国际部

**3）关于 “7选3”有些选四门，有些选两门？**

7选3的课程数据前后口径有一定的问题。针对此，建议选手就当前高三学生的数据进行7选3的典型性分析。

**4）考勤类型表：默认信息是指什么？早上迟到和晚到学校有啥区别啊？**

由于校园考勤设备批次不同，来源供应商不同，其中某个品牌考勤机功能限制，导出全部是默认信息，其余来自于其他品牌如手持考勤机。迟到和晚到是一个意思，不同老师的录入风格不同。

**5）考勤信息里面是不是除了违规行为（迟到早退校服校徽）之外还有正常的进校离校考勤啊，哪些是违规的，那些是正常的，能不能让校方给一个考勤类型的详细说明啊。**

考勤类型说明在数据表里均已说明，均为字面意思。校服校徽即为没穿戴校服校徽。除操场考勤机、进校、离校为正常考勤外，其它项根据字面意思判断考勤违规行为即可。

**6）成绩和考勤，部分班级的数据全部没有，是不是可以请官方补足数据呢？**

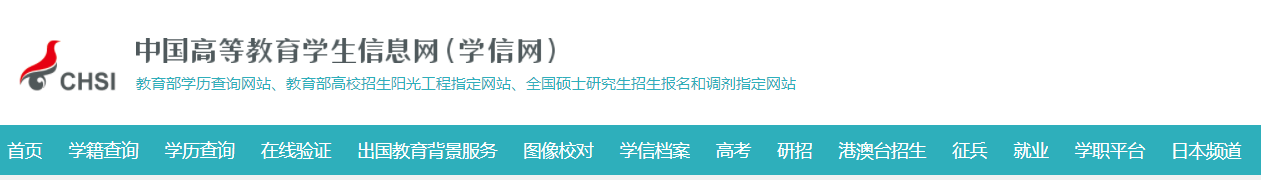
缺失的数据可以不必可视化，学校的情况比较复杂，比如国际部的成绩一般是不录入的，关于数据补充可能需要协调学校的老师和学生去统计录入，难度较大周期较长。

**7）为什么学生表（2\_student\_info.csv）里面学号在14454-16152范围，但是在考勤表（3\_kaoqin.csv）和成绩表（5\_chengji.csv），中有13539，13893等学号的学生，这些无法对应到学生信息中啊？**

我们在提取数据时是按照一个大致的时间为界限 取近几年的部分数据做为此次大赛数据 因此在每个项目的数据抽取的时间分割范围内 会有一些项目出现信息多余和缺少 这部分数据选手酌情处理即可

**三、收集智能教育相关领域软件的功能：**

1. 学信网



功能：1）学籍/学历查询。

1. 显示相关报考信息和报考功能。
2. 学习资料查询。
3. 汕大学分制系统



功能：1）查询空闲教室;

1. 教学反馈与课程评估;
2. 查看毕业审核情况;
3. 查看备选课程、教程订购;
4. 进行选课和申请课程;
5. 资格考试报名；
6. 查看个人信息，如考试安排、课表、成绩、学业导师信息。