附件3

**佛山市中小学人工智能展示活动**

**数据耦合项目要求**

1. 项目说明。

　　从一个大数据集中发现数据之间的内在规律，用于现状分析、关系分析、预测分析。参加代表队根据“数据耦合任务”设计方案，通过设计程序完成数据预处理、分析数据、数据可视化。同时，撰写涵括设计理念、原理、技术、实现、成果的报告。

1. 组别设置。

分别设置小学、初中、高（职）中三个组别。每个展示队应由2名学生及1—2名指导教师组成。

1. 名额分配

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 区域 | 小学组 | 初中组 | 高（职）中组 | 总数 |
| 禅城区（含市直） | 2 | 2 | 2 | 6 |
| 南海区 | 3 | 3 | 3 | 9 |
| 顺德区 | 3 | 3 | 3 | 9 |
| 高明区 | 1 | 1 | 1 | 3 |
| 三水区 | 1 | 1 | 1 | 3 |

1. 展示规则。

　　活动分为程序设计、撰写报告和答辩三个环节。参加队伍根据题目进行程序设计，经过专家评审后，遴选优秀队伍进入撰写PPT报告和答辩环节。

1. 程序设计。组委会现场发布题目，选手在规定时间内设计、编写和调试程序，时间结束前运行和提交结果。使用的电脑由组委会统一提供。预装编程环境为anaconda3 （附带python3、idle、jupyter notebook等工具）、sklearn 0.22.1。预装图形化编程工具为Mixpy、海龟编辑器。程序设计过程中电脑不能连接外网，但可在展示前提出所需安装的软件和库，经组委会审核后统一部署。
2. 撰写报告。根据程序运行结果撰写分析报告，对程序运行结果进行描述。
3. 答辩环节。上交程序文件（.py或.ipynb）及分析报告（.doc或.docx），由专家对项目进行评核，遴选优秀成果进入答辩。答辩环节每组5分钟，评委根据选手现场表现进行评分。
4. 展示内容。

　　1、数据准备：导入和导出；

　　2、数据预处理：合并、计算、分组、清洗降噪；

　　3、数据分析：分析与建模；

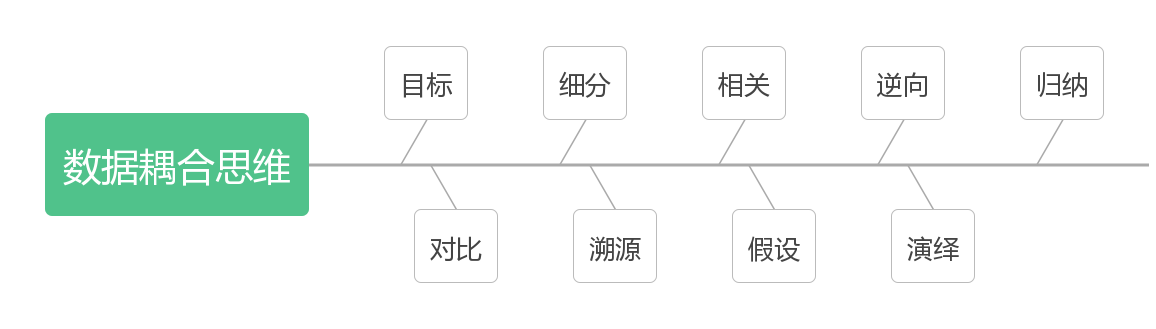
　　4、数据可视化：组织数据、展示数据，得出结果与推论。

1. 成果展示要求

|  |  |  |  |
| --- | --- | --- | --- |
| 评价条目 | 说 明 | 评分 | 评判导向 |
| 项目理解 | 从项目开展和分析报告中展现出对项目目标、内容与任务的正确理解。 | 10 | 能选用适当的数据分析工具，呈现数据。 |
| 小组协作 | 小组分工合理，协作紧密，合作有成效。 | 10 | 都参与了合作过程、同心协力地解决问题、做出了贡献。 |
| 数据预处理 | 获取数据后，对数据整理成适合数据分析的样式。参与程度、格式符合、清洗技术、输出种类。 | 15 | 能利用适当的工具对数据进行采集和分类，做数学运算，数据预处理后是否符合数据分析的要求与规范。 |
| 数据分析 | 选择合适的分析方法及工具，对预处理的数据进行分析，提取有价值的信息，形成有效结论。 | 15 | 能够了解常用的数据分析方法，有一定的数理逻辑能力。 |
| 数据可视化 | 能够有效、直观地传达分析人员表达的观点，把原始数据简单、直观化，能引导用户分析和推理出有效信息。多样准确。报告形式、技术难度、发布方式。 | 15 | 选用适当的数据分析工具，分析、呈现是否反应分析结果，选择的图表类型是否合适，  智慧发现并解析数据。 |
| 分析报告 | 有一个好的分析框架，层次明晰、图文并茂，有明确的结论、建议或解决方案。 | 20 |
| 代码规范 | 所编写代码规范简洁，可读性强 | 15 | 代码变量及方法设置合理，逻辑清晰，有适当的注释 |
| 展示交流 | 仪态大方，语言整洁明了，能够清晰阐述自己的观点。 | 20 | 一等奖后答辩，附加分  PPT制作精美，内容清晰、逻辑性强，包含所有要求内容，并在规定时间内有条理地介绍研究成果。 |

1. 评审标准

活动以培养数据耦合思维为中心，将按照目标思维、对比思维、细分思维、溯源思维、相关思维、假设思维、逆向思维、演绎思维、归纳思维共九大思维要求开展活动评审。



1. 目标思维：根据题目要求，有明确且合理的数据分析目标。
2. 对比思维：在分析过程中，适当地采用多维度数据对比方法。
3. 细分思维：围绕数据分析目标，能发现异常数据，并从不同的维度进行细分分析，得出有支撑性的结论。
4. 溯源思维：具有追根溯源精神，有一定的数据敏感度，从数据中找到蕴含的逻辑关系，并提出解决方案。
5. 相关思维：主动寻找数据变量之间的相互关系，甚至得出定量的关联程度。
6. 假设思维：能大胆地提出合理假设，并采用科学方法得到引证。
7. 逆向思维：适当采取反向思维寻求解决问题的方法，如利用反证法分析数据规律。
8. 演绎思维：能充分利用客观世界的一般规律，推导出数据中呈现的特定规律。
9. 归纳 思维：能从数据反映的特定规律中，归纳总结出客观世界的一般规律。

学生成绩分析报告示例

组委会将会提供几万条以上的学生成绩数据，如下：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学号 | 班级 | 姓名 | 性别 | 英语 | 体育 | 军训 | 数分 | 高代 | 解几 |
| 2308024241 | 23080242 | 成龙 | 男 | 76 | 78 | 77 | **40** | **23** | 60 |
| 2308024244 | 23080242 | 周怡 | 女 | 66 | 91 | 75 | **47** | **47** | **44** |
| 2308024251 | 23080242 | 张波 | 男 | 85 | 81 | 75 | **45** | **45** | 60 |
| 2308024249 | 23080242 | 朱浩 | 男 | 65 | **50** | 80 | 72 | 62 | 71 |
| 2308024219 | 23080242 | 封印 | 女 | 73 | 88 | 92 | 61 | **47** | **46** |
| 2308024201 | 23080242 | 迟培 | 男 | 60 | **50** | 89 | 71 | 76 | 71 |
| 2308024347 | 23080243 | 李华 | 女 | 67 | 61 | 84 | 61 | 65 | 78 |
| 2308024307 | 23080243 | 陈田 | 男 | 76 | 79 | 86 | 69 | **40** | 69 |
| 2308024326 | 23080243 | 余皓 | 男 | 66 | 67 | 85 | 65 | 61 | 71 |
| 2308024320 | 23080243 | 李嘉 | 女 | 62 | 作弊 | 90 | 60 | 67 | 77 |
| 2308024342 | 23080243 | 李上初 | 男 | 76 | 90 | 84 | 60 | 66 | 60 |
| 2308024310 | 23080243 | 郭窦 | 女 | 79 | 67 | 84 | 64 | 64 | 79 |
| 2308024435 | 23080244 | 姜毅涛 | 男 | 77 | 71 | 缺考 | 61 | 73 | 76 |
| 2308024432 | 23080244 | 赵宇 | 男 | 74 | 74 | 88 | 68 | 70 | 71 |
| 2308024446 | 23080244 | 周路 | 女 | 76 | 80 |  | 61 | 74 | 80 |
| 2308024421 | 23080244 | 林建祥 | 男 | 72 | 72 | 81 | 63 | 90 | 75 |
| 2308024433 | 23080244 | 李大强 | 男 | 79 | 76 | 77 | 78 | 70 | 70 |
| 2308024428 | 23080244 | 李侧通 | 男 | 64 | 96 | 91 | 69 | 60 | 77 |
| 2308024402 | 23080244 | 王慧 | 女 | 73 | 74 | 93 | 70 | 71 | 75 |
| 2308024422 | 23080244 | 李晓亮 | 男 | 85 | 60 | 85 | 72 | 72 | 83 |
| 2308024201 | 23080242 | 迟培 | 男 | 60 | **50** | 89 | 71 | 76 | 71 |

学生成绩分析报告

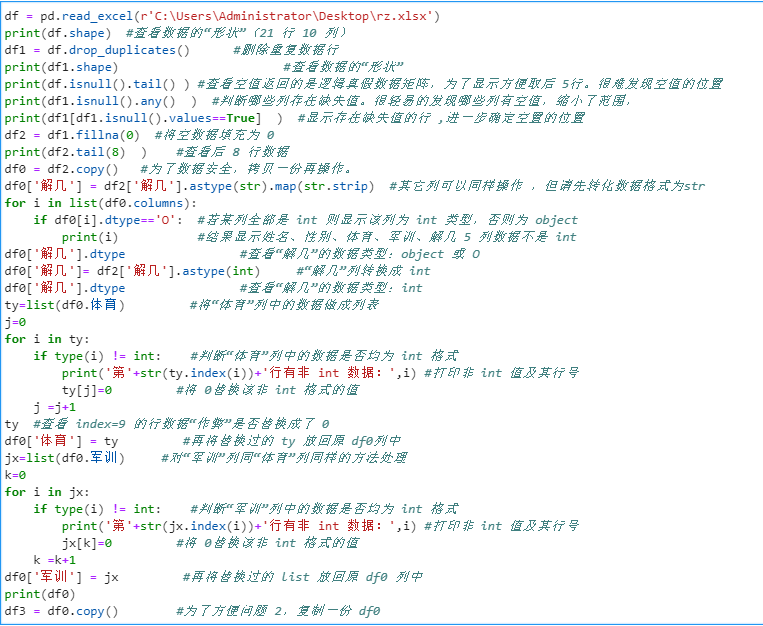
探究背景：对学生成绩数据进行统计、分析，能够为学校、老师提供翔实的统计和分析数据，使之了解学生的各科成绩情况，及时采取科学的干预措施。通过学科相关性探究，学生科目之间可能存在一定的内部联系，作品根据所有学生的各个科目的考试成绩，可以做出学生科目成绩之间的分布关系，探究学科之间的关联性。

　数据整理

对于收集到的数据，我们需要对其进行整理，去除重复的项目、每条数据信息缺失较多的要去除，有逻辑错误或者有格式错误的要予以修正。

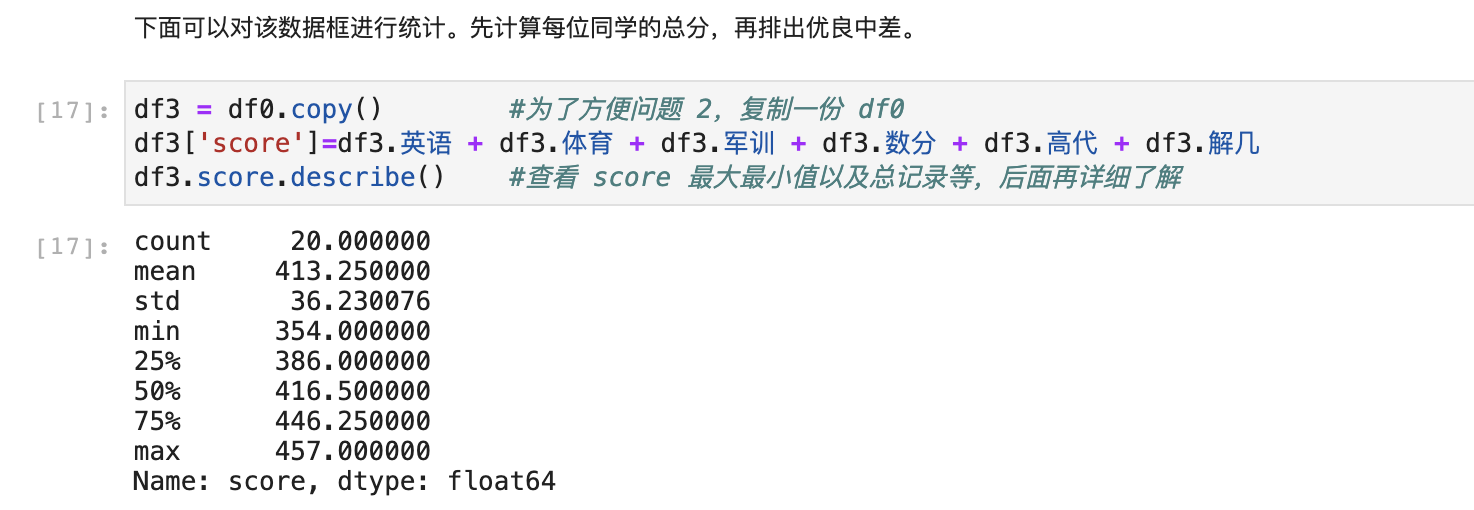
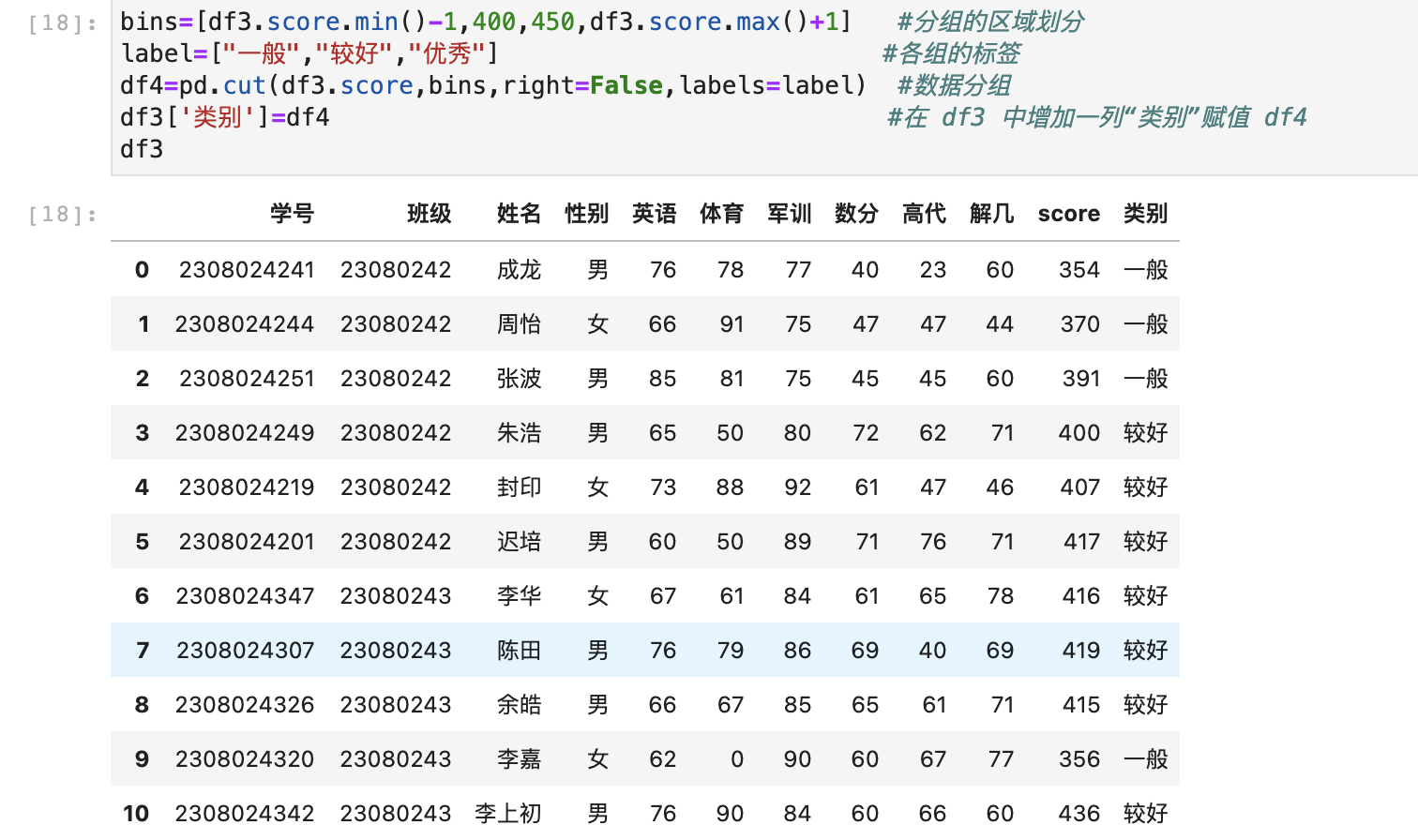


代码如下：

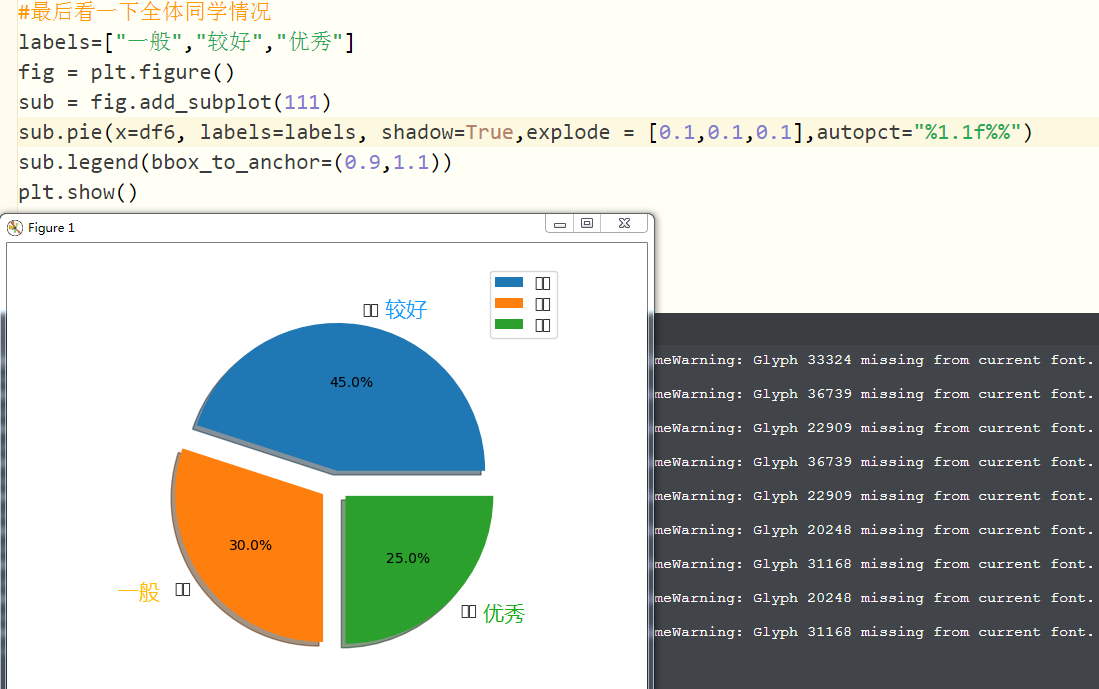


1. 数据分析与可视化（包括关联性因素分析） ：组织数据、展示数据，得出结果与推论。

分析（1) 将数据表添加两列：各科成绩总分(score)和每位同学的整体情况（类别），类别按 照[df.score.min()-1,400,450,df.score.max()+1]分为"一般","较好","优秀"三种情况。



绘制学生成绩占比图：

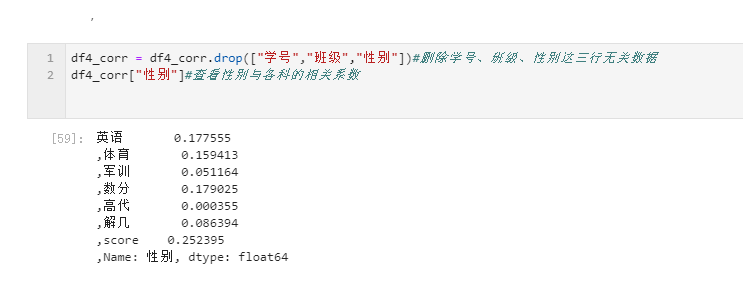


得出结论发现：成绩为“优秀”和“较好”的学生占大多数，总体成绩偏优，教学成果良好。

分析(2）人生而不同，许多研究表明男女思维方式有所不同，试找出性别与各个科目之间的相关性。

由于性别是字符串，而各科成绩是数字，类型不同无法直接进行分析，因此首先需要将性别数据标准化（转化为离散数字数据），再对性别和各科成绩进行相关性分析（如皮尔逊相关系数分析）





得出结论：