

2024 年广西职业院校技能大赛

中职组《大数据应用与服务》赛项

样题第 2 套

一、背景描述

在当今数字化飞速发展的时代，大数据技术在各行各业都扮演着至关重要的角色，而在学校数据管理领域，大数据的应用更是愈发凸显其重要性。学校作为一个庞大而复杂的组织体系，拥有众多的学生、教职工和各种资源，数据的产生和管理变得日益庞大而复杂。因此，大数据技术的引入成为学校数据治理的一项重要举措。

在中职教育领域，大数据比赛的举办对学生、学校和整个教育体系都具有重要的意义。首先，大数据比赛为中职学生提供了一个实践和展示自己技能的平台。通过参与比赛，学生可以运用所学的大数据技术知识，解决实际问题，提高实际操作能力。这不仅有助于加深学生对大数据领域的理解，还锻炼了他们的团队协作和问题解决能力。

其次，中职大数据比赛有助于激发学生对科学技术的兴趣。通过实际操作和创新性的比赛内容，学生更容易对大数据技术产生浓厚的兴趣，并在实践中感受到这一领域的魅力。这种积极的学习体验有助于激发学生对科技领域的兴趣，培养创新意识，为其未来的职业发展奠定基础。

另外，中职大数据比赛也为学校提供了一个展示自身教学水平和培养优秀学生的机会。优异的比赛成绩不仅可以提升学校的声誉，还有助于吸引更多学生报考相关专业。同时，学校通过组织和支持学生参与大数据比赛，也能够加深与企业、行业的合作关系，促进校企合作的深度发展。

总体而言，中职大数据比赛在推动学生综合素质提升、促进科技创新、推动学校发展等方面发挥着重要的作用。通过这样的比赛活动，中职教育能够更好地适应社会需求，培养更符合产业发展需求的人才，为学生未来的职业生涯打下坚实基础。

二、模块一：平台搭建与运维

（一）任务一：大数据平台搭建

1. 子任务一：Hadoop 完全分布式安装配置

本任务需要使用 root 用户完成相关配置。打开竞赛平台的赛题资源的容器资源，利用智联助手的 SSH 集成窗口打开容器，基于竞赛平台进行完全分布式模式 Hadoop 的搭建和管理。相关安装文件在容器 “/opt” 目录下，请选择对应的安装包进行安装，用不到的可忽略。主机名为 master、slave1、slave2 的三台节点都需要安装 JDK 和 Hadoop。

(1) 在 master 节点执行命令，创建 jdk 安装目录

“/data/jdk” 和 Hadoop 安装目录 “/data/hadoop”，执行 ls 命令查看创建的目录。提交创建目录的命令和查看的结果截图；

(2) 在 master 节点执行 tar 命令（请确保在解压缩文件时，不要保留原始文件路径中的第一个目录层级），将节点中 “/opt” 目录下 jdk 安装文件解压到节点中的 “/data/jdk” 目录，将 Hadoop 安装文件解压到节点中的 “/data/hadoop” 目录，执行 ls 命令分别查看解压后的文件。提交解压命令和查看的结果截图；

(3) 在三台节点分别修改 “/etc/hosts” 配置，添加三台节点 ip 和主机名映射；配置 ssh 免密，实现 master、slave1、slave2 三台节点间相互免密登录，在 master 节点执行 “ssh slave1” 命令和在 slave2 节点执行 “ssh master” 命令验证。提交 “/etc/hosts” 配置截图、在 master 执行的免密命令和执行结果截图、验证命令的结果截图；

(4) 在 master 节点使用 scp 命令并使用绝对路径将 jdk 目录拷贝到 slave1 和 slave2（若路径不存在，则需新建），并在三台节点的 “/etc/profile” 文件中配置 jdk 环境变量并使其生效，配置完毕后，在节点中执行 “java -version” 命令，检测 jdk 是否安装成功。提交 scp 命令、环境变量截图和使环境变量生效命令、检测命令结果截图；

(5) 依次配置 `hadoop-env.sh`、`core-site.xml`、`workers` 配置文件,其中 NameNode 的地址为 master 节点,端口为 9000, master、slave1、slave2 节点均作为 DataNode,配置好相关环境。在 master 节点使用 `scp` 命令将配置完的 Hadoop 安装目录直接拷贝至 slave1 和 slave2 (若路径不存在,则需新建),在三台节点的“`/etc/profile`”文件中配置 Hadoop 环境变量,配置完毕后,在容器中执行“`hadoop version`”命令,检测 Hadoop 是否安装成功。提交修改的配置内容截图、`scp` 命令、环境变量截图、检测命令结果截图;

(6) 在 master 节点初始化 Hadoop 环境 namenode。提交初始化命令和结果截图;

(7) 在 master 节点依次启动 HDFS、YARN 集群,在三个节点分别执行 `jps` 命令,查看节点中的进程。提交 `jps` 查看结果截图;

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

2. 子任务二: Flume 安装配置

本任务需要使用 root 用户完成相关配置,已安装 Hadoop 及需要配置前置环境,具体要求如下:

(1) 在节点中执行命令，创建 Flume 安装目录

“/data/flume”，执行 ls 命令查看创建的目录。提交创建目录命令、查看命令和结果截图；

(2) 在节点中执行 tar 命令（请确保在解压缩文件时，不要保留原始文件路径中的第一个目录层级），将 “/opt” 目录下 Flume 安装文件解压到节点中的 “/data/flume” 目录，在 “/etc/profile” 文件中配置 Flume 环境变量 FLUME_HOME 和 PATH 的值并使其生效，执行命令 “flume-ng version” 检测 Flume 是否安装成功。提交解压命令、环境变量截图、检测命令结果截图；

(3) 进入 \$FLUME_HOME/conf 目录下，使用 cp 命令将 flume-env.sh.template 文件复制一份，并重命名为 flume-env.sh；在 \$FLUME_HOME 目录下新建一份 test.txt 文件，在文件中输入内容 “这是一份 flume 的测试文档”，在 \$FLUME_HOME/conf 目录下新建 file-flume-hdfs.conf，配置 flume 将 test.txt 传输到 HDFS 的 “/tmp/flume” 目录下，运行 Flume 并查看 HDFS 中 “/tmp/flume” 目录下生成的内容。提交 cp 命令、file-flume-hdfs.conf 配置内容截图、运行 Flume 的命令、查看 HDFS 目录中文件的内容的命令和结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

（二）任务二：数据库配置维护

1. 子任务一：数据库配置

打开竞赛平台的赛题资源的数据库资源，利用智联助手的SSH集成窗口打开master节点，基于竞赛平台进行MySQL的搭建和管理。具体要求如下：

（1）在赛题资源的数据库资源对应的ip的节点中执行命令，创建MySQL安装目录“/data/mysql”；执行tar命令，将容器中“/opt”目录下Mysql安装文件解压到“/data/mysql”目录，执行ls命令查看解压后的文件。提交tar命令和ls结果截图；

（2）安装好MySQL后，对数据库进行初始化，将\$MYSQL/bin目录下的mysql脚本创建软链到/usr/bin目录，在/etc/systemd/system目录中配置mysql.service，配置开机自启动并启动数据库。提交初始化命令、创建软链命令、mysql.service配置内容截图、配置自启动命令、启动数据库结果截图；

（3）使用root用户和密码登录MySQL，然后将root用户的密码修改为赛题资源的数据库资源的密码，配置授予root用户在任何主机（'%'）上连接时对所有数据库（*.*）的全部权限，并允许该用户授予权限给其他用户，配置完刷新权限。提交修改密码sql语句和结果截图、配置授予权限sql语句和

结果截图;

(4) 连接数据库, 先用 sql 创建一个 “exam” 数据库(字符集设置 utf8mb4), 再创建一个 “test” 新用户, 密码与 root 密码一致, 将 “exam” 数据库权限授予 “test” 用户, 刷新权限后用 “test” 账号登录数据库, 执行 “show databases;” 命令查看数据库。提交创库 sql 语句、建用户语句、授权语句、“show databases;” 的结果截图;

(5) 将 “test” 用户删除。提交删除语句;

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

2. 子任务二：数据库表维护

打开竞赛平台的赛题资源的数据库资源, 利用智联助手的数据库工具打开 “exam” 数据库, 基于竞赛平台进行 MySQL 的数据库表维护。

(1) 在 MySQL 数据库的 “exam” 数据库中创建 “athlete_events” 数据表。数据表的表字段格式如下:

表 1 “athlete_events” 的表字段结构

| 字段 | 类型 | 注释 |
|------|---------|----|
| Name | varchar | 名称 |
| Sex | varchar | 性别 |
| Age | int | 年龄 |

| | | |
|--------|---------|------|
| Height | int | 身高 |
| Weight | int | 体重 |
| Games | varchar | 赛事 |
| City | varchar | 城市 |
| Event | varchar | 比赛项目 |
| Medal | varchar | 奖牌 |

提交建表 sql 语句和结果截图；

(2) 为 “athlete-events” 表添加数据记录。记录如下：

| Name | Sex | Age | Height | Weight | Games | City | Event | Medal |
|---------------|-----|-----|--------|--------|-------------|----------------|--|--------|
| Cao Yuan | M | 21 | 160 | 42 | 2016 Summer | Rio de Janeiro | Diving Men's Springboard | Gold |
| Chen Aisen | M | 20 | 168 | 60 | 2016 Summer | Rio de Janeiro | Diving Men's Platform | Gold |
| Chen Long | M | 27 | 188 | 81 | 2016 Summer | Rio de Janeiro | Badminton Men's Singles | Gold |
| Chen Peina | F | 27 | 172 | 63 | 2016 Summer | Rio de Janeiro | Sailing Women's Windsurfer | Silver |
| Chen Dequan | M | 18 | 176 | 66 | 2014 Winter | Sochi | Short Track Speed Skating Men's 5,000 metres Relay | Bronze |
| Jia Zongyangu | M | 22 | 175 | 70 | 2014 Winter | Sochi | Freestyle Skiing Men's Aerials | Bronze |
| Cai Yun | M | 32 | 181 | 68 | 2012 Summer | London | Badminton Men's Doubles | Gold |
| Zou Kai | M | 24 | 158 | 55 | 2012 Summer | London | Gymnastics Men's Team All-Around | Gold |
| Zhou Lulu | F | 24 | 175 | 133 | 2012 Summer | London | Weightlifting Women's Super-Heavyweight | Gold |
| Zhang Jike | M | 24 | 178 | 70 | 2012 Summer | London | Table Tennis Men's Singles | Gold |
| Guo Xinxin | F | 26 | 157 | 60 | 2010 Winter | Vancouver | Freestyle Skiing Women's Aerials | Bronze |
| Li Nina | F | 27 | 160 | 52 | 2010 Winter | Vancouver | Freestyle Skiing Women's Aerials | Silver |

| | | | | | | | | |
|---------------|---|----|-----|----|-------------|----------------|--|--------|
| Bao Yingying | F | 24 | 172 | 67 | 2008 Summer | Beijing | Fencing Women's Sabre, Team | Silver |
| Feng Kun | F | 29 | 183 | 75 | 2008 Summer | Beijing | Volleyball Women's Volleyball | Bronze |
| Han Xiaoping | M | 22 | 173 | 75 | 2006 Winter | Torino | Freestyle Skiing Men's Aerials | Gold |
| Chen Jing | F | 28 | 182 | 75 | 2004 Summer | Athina | Volleyball Women's Volleyball | Gold |
| Chen Zhong | F | 21 | 183 | 73 | 2004 Summer | Athina | Taekwondo Women's Heavyweight | Gold |
| Zhao Hongbo | M | 28 | 177 | 75 | 2002 Winter | Salt Lake City | Figure Skating Mixed Pairs | Bronze |
| Cai Yalin | M | 23 | 174 | 60 | 2000 Summer | Sydney | Shooting Men's Air Rifle, 10 metres | Gold |
| Chen Xiaomin | F | 23 | 158 | 63 | 2000 Summer | Sydney | Weightlifting Women's Middleweight | Gold |
| Chen Lu | F | 21 | 162 | 52 | 1998 Winter | Nagano | Figure Skating Women's Singles | Bronze |
| He Qi | F | 22 | 178 | 68 | 1996 Summer | Atlanta | Volleyball Women's Volleyball | Silver |
| Lu Lin | M | 27 | 174 | 60 | 1996 Summer | Atlanta | Table Tennis Men's Doubles | Silver |
| Li Xiaoshuang | M | 22 | 157 | 52 | 1996 Summer | Atlanta | Gymnastics Men's Individual All-Around | Gold |
| Ye Qiaobo | F | 29 | 168 | 70 | 1994 Winter | Lillehammer | Speed Skating Women's 1,000 metres | Bronze |

提交第一条数据的插入 sql 语句及所有数据添加结果截图;

(3) 新增 “Year” 字段(int), 注释为 “年份”, 将 “Games” 字段中的年份数字提取到 “Year” 字段, 并查询 “Year” 在 2000 年到 2008 年之间的数据。提交新增字段 sql 和结果截图、提取年份 sql、查询的 sql 和结果截图;

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

3.子任务三：数据表查询

(1)筛选出比赛项目为“Volleyball Women's Volleyball”且身高大于“180”的数据，并按照年份“Year”进行升序排序。提交 sql 和结果截图；

(2)查询 2016 年参赛人员年龄、身高和体重各自的最大值、最小值和平均值。提交 sql 和结果截图；

(3)按年份分组查询每年参赛人员的男女人数各多少。提交 sql 和结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

三、模块二：数据获取与处理

(一) 任务一：数据获取与清洗

1. 子任务一：数据获取

打开竞赛平台赛题资源的文件资源的 P2-M2-T1 文件分类，下载 parse-athlete-events.py 文件。赛题资源中静态资源 athlete-events.html 是奥运会运动员数据和奖牌成绩列表内容，点击进入可打开网页。parse-athlete-events.py 为 Python 脚本文件，程序读取 athlete-events.html，使用 lxml 对网页进行解析，提取相应的列表数据，并将结果输出。

(1) 用竞赛平台智联助手的 python 脚本工具打开

parse-athlete-events.py 文件，补全文件中【1】代码，配置公共资源地址 url。提交 url 地址代码；

（2）使用浏览器打开赛题资源中静态资源文件 athlete-events.html 网页文件，通过“审查”工具进行网页结构分析。补全 parse-athlete-events.py 中【2】代码，实现获取奥运会运动员数据和奖牌成绩 div 列表。提交补全代码；

（3）补全 parse-athlete-events.py 中【3】~【5】代码，实现“text”，“举办城市”和“代表队伍”文本内容提取。提交补全代码；

（4）运行 parse-athlete-events.py 脚本，完成奥运会运动员数据和奖牌成绩列表的解析并打印输出。提交运行结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

2. 子任务二：数据处理

打开竞赛平台赛题资源的文件资源的 P2-M2-T1 文件分类，下载 athlete-events.csv 文件。通过编写 python 代码完成对相关数据文件中数据的清洗和整理。请分析相关数据集，根据题目规定要求实现数据处理，具体要求如下：

(1) 查看 athlete_events.csv 中数据总数、标准差、均值、最小值、四分之一分位数、二分之一分位数、四分之三分位数和最大值。提交代码截图和结果截图;

(2) 对 athlete_events.csv 文件进行处理, 对于 Age 列数据, 采用 “Age” 填充缺失值, 并将年龄数字大于等于 90 的数据删除, 然后存入 athlete_events_c1.csv 中。提交代码截图;

(3) 对 athlete_events_c1.csv 文件进行处理, 对于 Height 列和 Weight 列数据, 将缺失值的数据删除, 然后存入 athlete_events_c2.csv 中。提交代码截图;

(4) 对 athlete_events_c2.csv 文件进行处理, 对于 Weight 列数据, 小数点后超过一位数的数据, 将小数四舍五入改为一位数, 然后存入 athlete_events_c3.csv 中。提交代码截图;

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

(二) 任务二：数据标注

对 athlete_events_c3.csv 进行标注, 判断运动员是否获得奖牌, 具体的标注规则如下:

(1) 如果 “Medal” 列数据不为空, 则数据标注为 “yes”;

(2) 如果 “Medal” 列数据为空, 则数据标注为 “no” ;

标注好的数据存储为列 “Have_Medal” 并和 athlete-events-c3.csv 数据合并存入 result.csv。提交代码截图和 csv 文件；

将上述任务的结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

（三）任务三：数据统计

1. 子任务一：处理异常数据

打开竞赛平台赛题资源的文件资源的 P2-M2-T3 文件分类，下载 TravelInsurancePrediction.csv 文件。该文件存储了旅游保险预测相关数据，数据中有以下内容：

Age: 年龄，int 类型

EmploymentType: 职业种类，varchar 类型

GraduateOrNot: 是否毕业生，varchar 类型

AnnualIncome: 年收入，int 类型

FamilyMembers: 家庭成员，int 类型

ChronicDiseases: 是否有慢性病，int 类型

FrequentFlyer: 是否经常坐飞机，varchar 类型

EverTravelledAbroad: 曾经出国旅游，varchar 类型

TravelInsurance: 是否购买旅游保险，int 类型

编写 MapReduce 程序，实现以下功能：清除曾经有出国旅游的人员数据，在控制台按顺序打印输出前 20 条数据，输出结果到 HDFS，使用命令查看结果。提交代码截图和结果截图；

将上述任务的程序和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

2. 子任务二：数据统计

使用竞赛平台赛题资源的文件资源的 P2-M2-T3 文件分类的 TravelInsurancePrediction.csv 文件，编写 MapReduce 程序，实现以下功能：根据 “GraduateOrNot” 字段统计毕业生与非毕业生人数，并在控制台输出，输出结果到 HDFS，使用命令查看结果。提交代码截图和结果截图；

将上述任务的程序和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

四、模块三：数据分析与可视化

（一）任务一：数据分析与可视化

使用竞赛平台可视化工具或者下载可视化操作文件至本地进行代码编写并以图表展示。

1. 子任务一：柱状图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具，点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源 P2-M3-T1-SUBT1 文件分类，下载 visualization.zip 文件至本地解压，编写补充代码，实现 Web 网页形式对 2008 年得奖数量前十的国家数据进行可视化展示：

【竞赛平台的数据分析与可视化工具】

(1) 编写补充 yAxis 对象，获取相关数据中 barData 数据，设置 y 轴显示类型为“类目轴”、设置坐标轴文字颜色值为：#999999，大小为：12、设置坐标轴在 grid 区域中的分隔线颜色为：#CAD3E0，线的类型为：虚线、设置 y 轴显示数据为"2008 年得奖数量前十的国家名称"。

(2) 编写补充 series 对象，获取相关数据中 barData 数据，设置图表显示类型为柱状图、设置系列名称为“奖牌数量”、设置柱条的宽度为 20，背景颜色为：rgba(180, 180, 180, 0.2)、将"2008 年得奖数量前十的国家", barData 对象中的数据设置为柱状图显示数据。

(3) 附上“2008 年得奖数量前十的国家柱状图”截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地，根据 visualization/data/data.js 文件中 barData 对象中的数据，

补充完整 visualization/js/chat.js 文件中 getBarChart() 函数的代码

(1) 编写补充 yAxis 对象, 获取 barData 数据, 设置 y 轴显示类型为“类目轴”、设置坐标轴文字颜色值为: #999999, 大小为: 12、设置坐标轴在 grid 区域中的分隔线颜色为: #CAD3E0, 线的类型为: 虚线、设置 y 轴显示数据为"2008 年得奖数量前十的国家名称"。

(2) 编写补充 series 对象, 获取 barData 数据, 设置图表显示类型为柱状图、设置系列名称为“奖牌数量”、设置柱条的宽度为 20, 背景颜色为: rgba(180, 180, 180, 0.2)、将"2008 年得奖数量前十的国家", barData 对象中的数据设置为柱状图显示数据。

(3) 运行网页, 附上“2008 年得奖数量前十的国家柱状图”截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

2. 子任务二：折线图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具, 点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源 P2-M3-T1-SUBT2 文件分类, 下载 visualization.zip 文件至本

地解压，编写补充代码，实现 Web 网页形式对中国在各届夏季奥运会上的奖牌数量变化数据进行可视化展示。任务点如下：

【竞赛平台的数据分析与可视化工具】

(1) 编写补充 tooltip 对象，设置提示框组件的触发类型为坐标轴触发、设置指示器类型为：直线指示器、设置提示框浮层的文字颜色：#666666，字体大小为：12。

(2) 编写补充 xAxis 对象，获取相关数据中 lineData 数据，设置 X 轴显示类型为“类目轴”、设置坐标文字显示为：#999999，文字大小设置为：12，文字间隔为 0，文字倾斜角度为 40 度、设置 X 轴显示坐标为“各界夏季奥运会”。

(3) 编写补充 series 对象，获取相关数据中 lineData 数据，设置图表显示类型为'line'、设置线条显示平滑，标记大小为 6 的三角形、设置折线图文字显示(将文字颜色设置为 #999999，文字大小设置为：12)、将"中国在各届夏季奥运会上的奖牌数量"对象中的数据设置为折线显示数据。

(4) 附上“中国在各届夏季奥运会上的奖牌数量变化折线图”截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地，根据 visualization/data/data.js 文件中 lineData 对象中的数据，

补充完整 visualization/js/chat.js 文件中 getLineChart() 函数的代码

(1) 编写补充 tooltip 对象，设置提示框组件的触发类型为坐标轴触发、设置指示器类型为：直线指示器、设置提示框浮层的文字颜色：#666666，字体大小为：12。

(2) 编写补充 xAxis 对象，获取 lineData 数据，设置 X 轴显示类型为“类目轴”、设置坐标文字显示为：#999999，文字大小设置为：12，文字间隔为 0，文字倾斜角度为 40 度、设置 X 轴显示坐标为“各界夏季奥运会”。

(3) 编写补充 series 对象，获取 lineData 数据，设置图表显示类型为'line'、设置线条显示平滑，标记大小为 6 的三角形、设置折线图文字显示(将文字颜色设置为#999999，文字大小设置为：12)、将"中国在各届夏季奥运会上的奖牌数量"对象中的数据设置为折线显示数据。

(4) 运行网页，附上“中国在各届夏季奥运会上的奖牌数量变化折线图”截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

3. 子任务三：饼图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具，点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源

P2-M3-T1-SUBT3 文件分类, 下载 visualization.zip 文件至本地解压, 编写补充代码, 实现 Web 网页形式对 2016 夏季奥运会运动员年龄分布数据进行可视化展示。任务点如下:

【竞赛平台的数据分析与可视化工具】

(1) 编写补充 legend 对象, 获取相关数据中 pieData 数据, 设置图例的朝向为: 垂直显示、设置图例在 X 轴方向上的位置为右、设置图例上显示的文字信息为: 年龄段, 图例项的图标都设置为圆形、设置图例文字颜色为: #999999, 大小为: 12。

(2) 编写补充 series 对象, 获取相关数据中 pieData 数据, 设置图表类型为 'pie'、设置系列名称为 '2016 夏季奥运会运动员年龄分布'、设置饼图半径为 ['30%', '70%']、设置饼图高亮状态, 标签文字颜色: #999999, 大小: 24, 居中显示、将 '2016 夏季奥运会运动员年龄分布' 对象中的数据设置为饼图显示数据。

(3) 附上 “2016 夏季奥运会运动员年龄分布饼图” 截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地, 根据 visualization/data/data.js 文件中 pieData 对象中的数据, 补充完整 visualization/js/chart.js 文件中 getPieChart() 函数的代码

(1) 编写补充 legend 对象, 获取 pieData 数据, 设置图例的朝向为: 垂直显示、设置图例在 X 轴方向上的位置为右、设置图例上显示的文字信息为: 年龄段, 图例项的图标都设置为圆形、设置图例文字颜色为: #999999, 大小为: 12。

(2) 编写补充 series 对象, 获取 pieData 数据, 设置图表类型为 'pie'、设置系列名称为 '2016 夏季奥运会运动员年龄分布'、设置饼图半径为 ['30%', '70%']、设置饼图高亮状态, 标签文字颜色: #999999, 大小: 24, 居中显示、将 '2016 夏季奥运会运动员年龄分布' 对象中的数据设置为饼图显示数据。

(3) 运行网页, 附上 “2016 夏季奥运会运动员年龄分布饼图” 截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

4. 子任务四：雷达图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具, 点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源 P2-M3-T1-SUBT4 文件分类, 下载 visualization.zip 文件至本地解压, 编写补充代码, 实现 Web 网页形式对 2016 奥运会男篮各国年龄、身高、体重平均值数据进行可视化展示。任务点如下:

【竞赛平台的数据分析与可视化工具】

(1) 编写补充 radar 对象，设置雷达图的中心（圆心）坐标为百分比形式，第一项为相对于容器宽度 50%，第二项为相对于容器高度 70%、设置雷达图的外半径值为 180、设置雷达图的指示器，分别为：Age、Height、Weight，设置最大值分别为：39、218、137。

(2) 编写补充 series 对象，获取相关数据中 radarData 数据，设置图表显示类型为 'radar'、设置标记的图形为 'triangle'，标记的大小为 10、将 '2016 奥运会男篮各国年龄、身高、体重平均值' 对象中的数据设置为雷达图显示数据。

(3) 附上 “2016 奥运会男篮各国年龄、身高、体重平均值雷达图” 截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地，根据 visualization/data/data.js 文件中 radarData 对象中的数据，补充完整 visualization/js/chat.js 文件中 getRadarChart() 函数的代码

(1) 编写补充 radar 对象，设置雷达图的中心（圆心）坐标为百分比形式，第一项为相对于容器宽度 50%，第二项为相对于容器高度 70%、设置雷达图的外半径值为 180、设置雷达图的指示器，分别为：Age、Height、Weight，设置最大值分别为：39、218、137。

(2) 编写补充 series 对象, 获取 radarData 数据, 设置图表显示类型为 'radar'、设置标记的图形为 'triangle', 标记的大小为 10、将 '2016 奥运会男篮各国年龄、身高、体重平均值' 对象中的数据设置为雷达图显示数据。

(3) 运行网页, 附上 “2016 奥运会男篮各国年龄、身高、体重平均值雷达图” 截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

(二) 任务二：数据分析

使用电子表格软件进行查询统计并使用图表进行展示。

1. 子任务一：Excel 数据折线图分析

打开竞赛平台赛题资源的文件资源的 P2-M3-T2 文件分类, 下载 ANALYSE.xlsx 文件。使用电子表格软件打开 ANALYSE.xlsx 文件, 对表格内的数据进行操作分析, 并最终使用电子表格软件自带的图表功能, 展示 2016 年奥运会运动员年龄小于 30 岁人数数量折线图, 任务点如下:

(1) 统计每个 “Age” 的 “数量” ;

(2) 筛选 “Age” 小于 30 的 “数量”, 并按照年龄升序使用折线图进行显示;

(3) 根据 “Age” 与 “数量” 的变化, 写一段分析报告。

将上述任务的结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

2. 子任务二：Excel 数据饼图分析

打开竞赛平台赛题资源的文件资源的 P2-M3-T2 文件分类，下载 ANALYSE.xlsx 文件。使用电子表格软件打开 ANALYSE.xlsx 文件，对表格内的数据进行操作分析，并最终使用电子表格软件自带的图表功能，展示 2016 年奥运会运动员身高人数前十占比的饼图，任务点如下：

- （1）统计每个“Height”的“数量”；
- （2）取“Height”的“数量”前十，并计算每个“Height”的百分比，并使用饼图展示；
- （3）根据饼状图，写一段分析报告。

将上述任务的结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

五、模块四：职业素养

中职大数据比赛职业素养评分关注团队合作、创新、问题解决、沟通、项目管理、职业操守等方面。参赛者需展现学科知识运用、自我学习、文档撰写能力。全面素养将在比赛中为团队成功贡献关键因素。

任务要求：

(1) 团队合作：能力在团队中合作协调，有效沟通，共同完成项目任务。评价团队成员之间的合作默契、协同工作的能力。

(2) 创新能力：考察参赛者在解决问题时的创造性思维和创新能力，包括提出独特的解决方案、采用新方法和技术等。

(3) 问题解决能力：能够迅速识别问题，采用合适的方法解决问题，具备分析和解决实际问题的能力。

(4) 沟通能力：能够清晰、准确地表达自己的观点，有效地与团队成员和评审沟通，包括书面报告、口头演讲等。

(5) 文档撰写能力：能够撰写清晰、完整、规范的文档。