

2024 年陕西省职业院校技能大赛中职组 大数据应用与服务赛项竞赛规程

一、赛项名称

赛项名称：大数据应用与服务

赛项归属：电子信息

赛项编号：ZZ052

二、竞赛目标

本赛项旨在贯彻党的教育方针，落实立德树人根本任务，充分发挥海量数据和丰富应用场景优势，促进数字技术与实体经济深度融合，赋能传统产业转型升级，催生新产业、新业态、新模式，引领专业建设和教学改革，促进职普融通、产教融合、科教融汇。展现参赛选手团队合作、工匠精神等职业素养，大力推进大数据技术及相关专业的技术技能型人才培养，推动职业教育与产业深度互动，推动职业教育提档升级，服务学生全面发展、服务经济社会发展、服务国家发展战略。

三、竞赛内容

本赛项涉及大数据行业的典型工作场景，包括大数据平台搭建、数据库运行维护、数据清洗、数据标注、数据分析、数据可视化和业务分析等工作任务，考查的主要技术技能如下：

1) 大数据平台搭建：安装 Hadoop 全分布式平台，安装 Hadoop 平台相关的常用组件，包括但不限于 ZooKeeper、Flume、Kafka、Spark、Flink、Redis、HBase 等，验证 Hadoop 平台和相关组件的可用性。

2) 数据库配置维护：基于 MySQL 数据库进行建库建表，运用基本的 SQL 语句完成数据的增删改查等操作。

3) 数据获取与清洗：读取 CSV 数据源，对指定字段进行有效性检

查，正确处理无效值和异常值，对数据进行一致性检查，对数据进行清洗和转换。

4) 数据统计：使用 Hadoop 默认提供的 wordcount 示例来完成单词数统计任务，并将程序打包部署到 Hadoop 平台上运行。使用 HDFS 上传和下载文件。

5) 数据标注：使用 Python 程序对数据进行分类标注，将标注后的数据保存到指定位置。

6) 数据可视化：使用 Web 技术或 Python 可视化技术对数据进行呈现，包括但不限于柱状图、折线图、玫瑰图、气泡图、饼状图、条形图、雷达图、散点图等效果。

7) 业务分析：能够理解业务场景，对业务数据进行分析，编写分析报告。

表 1 赛项内容与分值

模块		主要内容	比赛时长	分值
模块一：平台搭建与运维	任务一：大数据平台搭建	Hadoop平台的安装部署和常用组件的安装部署。	4小时	10
	任务二：数据库配置维护	使用MySQL数据库建库建表，运用基本的SQL语言完成数据的增删改查等操作。		20
模块二：数据获取与处理	任务一：数据获取与清洗	对CSV数据文件进行加载、清洗和转换等操作，识别和处理无效值，检查数据的一致性，将清洗后的数据保存到指定位置。		10
	任务二：数据标注	使用Python语言对数据进行分类标注。		10
	任务三：数据统计	使用Hadoop默认提供的wordcount示例来完成单词数		15

		统计任务，并将程序打包部署到Hadoop平台上运行。使用HDFS上传和下载文件。		
模块三：业务分析与可视化	任务一：数据可视化	使用Web前端框架或者python可视化库对数据进行可视化展示。		20
	任务二：业务分析	报表分析，对大数据项目的业务场景和数据进行分析，撰写报告。		10
职业素养		团队分工明确合理、操作规范、文明竞赛		5

四、竞赛方式

本竞赛为线下比赛，组队方式为团体赛，具体要求如下：

（一）参赛选手须为中等职业学校全日制在籍学生，五年制高职一至三年级（含三年级）学生也可报名参赛。凡在往届全国职业院校技能大赛中获一等奖的选手，不能再参加同一项目同一组别的比赛；

（二）每支参赛队由 3 名选手组成，不得跨校组队，原则上每所院校报名参赛不超过 2 队。参赛队可配指导教师，指导教师须为本校专兼职教师，每队限报 2 名指导教师。参赛选手和指导教师报名获得确认后不得随意更换；

（三）本赛项为单一场次，所有参赛队在现场根据给定的任务说明，在 4 小时内相互配合，采用小组合作的形式完成任务，最后以提交的结果文档作为最终评分依据。

五、竞赛流程

（一）竞赛时间表(以最终公布时间为准)

表 2 竞赛时间表

日期	时间	内容
赛前	第一天 14:00-15:00	各参赛队报到
	第一天	领队会、一次加密：抽取抽签顺序号

	15:00-16:00	
	第一天 16:00-16:30	参赛队熟悉比赛场地
	第一天 16:30-17:00	现场裁判赛前检查，封闭赛场
赛中	第二天 07:30 之前	参赛队集合前往比赛现场
	第二天 07:30-07:50	赛场检录
	第二天 07:50-08:10	二次加密：参赛队抽取赛位号
	第二天 08:10-08:30	选手进入工位、任务发放、宣布竞赛注意事项、检查竞赛环境
	第二天 08:30-12:30	竞赛选手完成竞赛任务、提交竞赛结果
赛后	第二天 13:00-17:00	竞赛成绩评判
	第二天 17:00 以后	技术点评

(二) 竞赛流程图

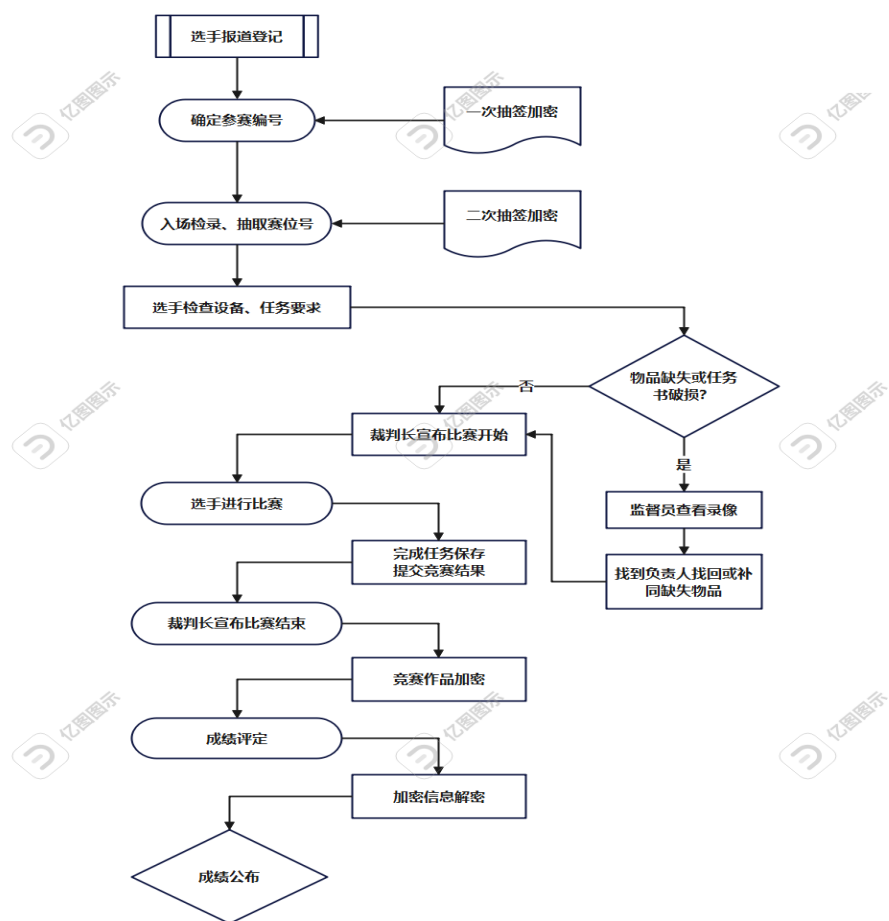


图 1 竞赛流程

六、竞赛规则

（一）选手报名：参赛选手须为中等职业学校全日制在籍学生，五年制高职一至三年级（含三年级）学生也可报名参赛。凡在往届全国职业院校技能大赛中获一等奖的选手，不能再参加同一项目同一组别的比赛。

（二）熟悉场地：竞赛前日安排各参赛队领队、指导教师、参赛选手熟悉赛场。

（三）入场规则：参赛选手按规定时间到达指定地点，必须携带参赛等证件，进行检录、一次加密、二次加密等流程，最终确定工位，选手迟到 10 分钟取消比赛资格。严禁参赛选手、赛项裁判、工作人员私自携带通讯、摄录设备进入比赛场地。参赛选手所需的硬件、软件和辅助工具统一提供，参赛队不得使用自带的任何有存储功能的设备，如手机、U 盘、移动硬盘等。

（四）赛场规则：在比赛过程中，参赛选手如有疑问，应举手示意，现场裁判应按要求及时予以答疑。如遇设备或软件等故障，参赛选手应举手示意，现场裁判、技术人员等应及时予以解决。确因计算机软件或硬件故障，致使操作无法继续，经裁判长确认，予以启用备用设备。参赛选手不得因各种原因提前结束比赛。如确因不可抗因素需要离开赛场的，须向现场裁判员举手示意，经裁判员许可并完成记录后，方可离开。凡在竞赛期间内提前离开的选手，不得返回赛场。

（五）离场规则：比赛时间结束，选手应全体起立，结束操作。参赛选手要确认已成功提交竞赛要求的文档，裁判员与参赛选手一起签字确认，经工作人员查收清点所有文档后方可离开赛场，离开赛场时不得带走任何资料。

（六）成绩评定

1.利用信息技术手段规避人为评分主观性、公平性等因素。

2.竞赛采取二次加密。第一次加密裁判组织参赛选手第一次抽签，抽取参赛编号，替代选手参赛证等个人信息;第二次加密裁判组织参赛选手进行第二次抽签，确定赛位号，替换选手参赛编号。每个环节结束后，数据立即封存，在评分结束后进行解密并统计成绩。

3.解密:裁判长正式提交评分结果并复核无误后，进行两层解密:工位号到参赛编号解密;参赛编号到参赛队名称解密。

4.抽检复核:为保障成绩评判的准确性，督察员对赛项总成绩排名前30%的参赛队伍的成绩进行复核;对其余成绩进行抽检复核，抽检覆盖率不得低于15%。督察员需将复检中发现的错误以书面方式及时告知裁判长，由裁判长更正成绩并签字确认。复核、抽检错误率超过5%的，则认定为非小概率事件，裁判组需对所有成绩进行复核。

(七)其它未尽事宜，将在赛前向各领队做详细说明。

七、技术规范

本赛项以专业技术标准、行业技能标准、软件开发标准为准则。

（一）技术规范

表3 专业技术标准

编号	标准号	标准名称
1	GB/T 11457-2006	信息技术、软件工程术语
2	GB8566-88	计算机软件开发规范
3	GB/T 12991-2008	信息技术数据库语言SQL第1部分: 框架
4	GB/T 21025-2007	XML使用指南
5	GB/T 28821-1012	关系数据管理系统技术要求

表4 大数据相关标准

编号	标准号	标准名称
1	GB/T 38672-2020	信息技术大数据接口基本要求
2	GB/T 38673-2020	信息技术大数据系统基本要求
3	GB/T 38676-2020	信息技术大数据存储与处理系统功能测试要求
4	GB/T 38643-2020	信息技术大数据分析系统功能测试要求
5	GB/T 38675-2020	信息技术大数据计算系统通用要求
6	GB/T 38633-2020	信息技术大数据系统运维和管理功能要求
7	GB/T 38672-2020	信息技术大数据接口基本要求
8	GB/T 38673-2020	信息技术大数据系统基本要求

表 5 软件开发标准

编号	标准号	标准名称
1	GB/T 8566-2001	信息技术软件生存周期过程
2	GB/T 15853-1995	软件支持环境
3	GB/T 14079-1993	软件维护指南
4	GB/T 17544-1998	信息技术软件包质量要求和测试

（二）设备使用与操作规范

1.计算机电源应保持良好的，插座不得松动，发现有漏电现象应立即切断电源。

2.开机前应检查有无异常情况。

3.开机前先接电源、开外设，最后开主机。

4.不能带电插拔外设及主机。

5.如发现计算机有不正常现象时应立即停止操作，请裁判员检查后方可用机。

6.做好数据资料的保密工作。

（三）操控人员应具备的专业知识

操控人员应具备数据采集与处理、数据分析与可视化、数据标注、大数据业务分析方法和方案架构、运行维护数据库系统等专业知识。

（四）操控人员应具备的技术技能

操控人员应具备分析系统数据来源、分析数据应用需求、设计数据资源整合解决方案、数据可视化、运行维护数据库系统、数据和信息处理等技术技能。

八、技术环境

（一）竞赛场地

赛场设选手检录区、选手休息区、竞赛区、裁判组工作区、技术支持区、服务保障区、观摩区等。

1.选手检录区：选手等待检录、一次加密、二次加密等职能区域；

2.选手休息区：选手检录前及竞赛离场休息、指导教师休息区；

3.竞赛区：每个竞赛工位设工位编号，竞赛工位相对独立，确保选手独立开展竞赛，不受外界影响。赛场内安装摄像头，确保每个竞赛工位无盲区监控。

（1）整个比赛场地应保持通畅和开放，并配备防火防爆及其他安全设施。

（2）赛场周边设有卫生间、医疗等公共服务区和紧急疏散通道，并在赛场周围设置隔离带。

（3）设立赛场开放区和安全通道。

（4）场地配套提供稳定的电、气源和供电应急设备，并有保安、消防、设备维修和电力抢险人员待命，以防突发事件。

4.裁判组工作区：供裁判工作及休息，相对封闭，配办公桌椅、电脑、打印机等；

5.服务保障服务区：提供医疗等服务保障，处理比赛中突发情况问题；

（二）技术平台

本赛项的硬件平台原则上采用通用技术实现。

表 6 硬件平台（承办校提供）

序号	设备名称	数量	说明
1	服务器	1	CPU：相当于或优于2颗 Intel Xeon Silver 4210性能； 内存：不少于128GB； 硬盘：不少于1TB SSD； 网卡：至少双千兆网口；
2	PC机	3	CPU：i5及以上 内存：不少于8GB 硬盘：不少于500GB 显示器：1920*1080及以上
3	交换机	1	8口及以上千兆交换机

本赛项的软件平台原则上采用通用开源技术实现。

表 7 软件平台（承办校提供）

序号	软件类别	软件名称和版本或要求说明	单位	数量
1	支撑系统	(1) 开源、免费； (2) 通用，与赛题无任何绑定； (3) 可开启和检测虚拟机操作、程序运行、程序代码；	套	1
2	大数据集群操作系统	CentOS Linux release 7	套	1
3	大数据平台组件	Hadoop 2.7.7 以上	套	1
4		Yarn 3.1.3	套	1
5		ZooKeeper 3.4.14 以上	套	1
6		Hive 2.3.4/3.1.2	套	1
7		JDK 1.8	套	1
8		Scala 2.11 以上	套	1

9		Spark 3.0.0/3.1.1	套	1
10		Flume 1.9.0	套	1
11		Kafka 2.1 以上	套	1
12		Sqoop 1.4.7 以上	套	1
13		Flink 1.14.0	套	1
14		Redis 6.2.6	套	1
15		Excel 2016	套	1
16		HBase 2.2.3	套	1
17	关系型数据库	MySQL 5.7 及以上	套	1
18	开发语言	JavaScript	套	1
19		html/html5+CSS	套	1
20		Python 3.6/3.7	套	1
21	开发库	numpy 1.18.5	套	1
22		pandas 1.3.4/2.1.3	套	1
23		matplotlib 3.5.0	套	1
24		Vue.js 3.2	套	1
25		ECharts 5.1 以上	套	1
26		pyecharts 2.0.4	套	1
27		snownlp 0.12.3	套	1
28		seaborn 0.11.2	套	1
29		openpyxl3.0.9	套	1
30		lxml 4.9.3	套	1
31	开发工具	IDEA 2022 (Community Edition)	套	1
32		PyCharm 2023 (Community Edition)	套	1
33		HBuilderX 3	套	1
34		Visual studio code 1.79 以上	套	1
35	数据库工具	Navicat	套	1
36	浏览器	Chrome	套	1
37	SSH 连接工具	Xshell 或 MobaXterm 或系统自带终端工具	套	1
38	文档编辑器	WPS	套	1
39	输入法	搜狗拼音输入法	套	1
40	PC 操作系统	Windows 10 64 位	套	1

九、竞赛样题或题库

项目背景

近年来随着社会经济的快速发展，百姓生活水平的不断提高，外出旅游成为很多人生活的热门选择，如何促进旅游业的发展成为各级政府高度重视的工作。为了更好地统筹管理城市的旅游资源，某省的旅游管理部门采集了本省若干城市的酒店经营数据和用户评论数据，其中酒店经营数据包括日期、城市、酒店名称、酒店星级、酒店当天预定房间数、酒店当天入住客户数、酒店当天最高房价和酒店当天最低房价等字段，这些数据保存到文件 `hotel.csv` 中。用户评论数据包括日期、城市、酒店名称、住客评分、评论内容等字段，这些数据保存到文件 `comments.csv` 中。

你作为技术人员，需要通过数据采集清洗、数据标注、数据分析、数据可视化、业务分析等步骤对酒店经营数据和用户评论数据进行处理，从而为政府制定旅游发展的政策提供决策依据。请按照下面的要求完成相关任务。

模块一：平台搭建与运维

任务一：大数据平台搭建

子任务 1Hadoop 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，安装 Hadoop 需要配置前置环境。命令中要求使用绝对路径，具体要求如下：

1) 从 Master 中的 `/opt/software` 目录下将文件 `hadoop-3.1.3.tar.gz`、`jdk-8u191-linux-x64.tar.gz` 安装包解压到 `/opt/module` 路径中(若路径不存在，则需新建)，将命令和结果复制粘贴至对应报告中；

2) 修改 Master 中 `/etc/profile` 文件，设置 JDK 环境变量并使其生效，

配置完毕后在 Master 节点分别执行 “java-version” 和 “javac” 命令，将命令和结果复制粘贴至对应报告中；

3)将三个节点分别命名为 master、slave1、slave2，并做免密登录，用 scp 命令并使用绝对路径从 Master 复制 JDK 解压后的安装文件到 slave1、slave2 节点（若路径不存在，则需新建），并配置 slave1、slave2 相关环境变量，将命令和结果复制粘贴至对应报告中；

4)在 Master 将 Hadoop 解压到/opt/module(若路径不存在，则需新建)目录下，并将解压包分发至 slave1、slave2 中，其中 master、slave1、slave2 节点均作为 datanode，配置好相关环境，初始化 Hadoop 环境 namenode，将命令和结果复制粘贴至对应报告中；

5)启动 Hadoop 集群（包括 hdfs 和 yarn），使用 jps 命令查看 Master 节点与 slave1 节点的 Java 进程，将命令和结果复制粘贴至对应报告中。

子任务 2Hive 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) 从 Master 中的 /opt/software 目录下将文件 apache-hive-3.1.2-bin.tar.gz、mysql-connector-java-5.1.37.jar 安装包解压到 /opt/module 目录下，将命令和结果复制粘贴至对应报告中。

2)设置 Hive 环境变量，并使环境变量生效，执行命令 hive--version 将命令和结果复制粘贴至对应报告中。

3)完成相关配置并添加所依赖包，将 MySQL 数据库作为 Hive 元数据库。初始化 Hive 元数据，并通过 schematool 相关命令执行初始化，将命令和结果复制粘贴至对应报告中。

任务二：数据库配置维护

子任务 1 在数据库中创建表

本任务在 MySQL 中创建表 t_comment 和表 t_hotel，并将用户评论数据 comments.csv 和酒店经营数据 hotel.csv 分别导入到表 t_comment 和表 t_hotel 中。具体要求如下：

1.创建用户评论表 t_comment，表 t_comment 的字段定义如下：

字段	类型	说明
comment_date	date	日期
city	varchar	城市
hotel_name	varchar	酒店名称
score	double	住客评分
content	varchar	评论内容

2.在 MySQL 中将 comments.csv 的数据导入表 t_comment。

3.创建酒店经营数据表 t_hotel,表 t_hotel 的字段定义如下：

字段	类型	说明
current_date	date	日期
city	varchar	城市
hotel_name	varchar	酒店名称
hotel_star	varchar	酒店星级
rooms_booked	int	酒店当天预定房间数
customers_checkedin	int	酒店当天入住客户数
highest_price	int	酒店当天最高房价
lowest_price	int	酒店当天最低房价

4.在 MySQL 中将 hotel.csv 的数据导入表 t_hotel。

5.将以上 SQL 语句和运行结果复制粘贴至对应报告中。

子任务 2 使用 SQL 查询数据

本任务具体要求如下：

- 1.查询指定酒店的评论数量。
- 2..查询指定酒店的住客评分的平均值。
- 3.查询每个城市的酒店数量。

- 4.查询指定酒店的最高房价和最低房价。
- 5.将以上 SQL 语句和运行结果复制粘贴至对应报告中。

模块二：数据获取与处理

任务一：数据获取与清洗

子任务 1 对空字段数据进行处理

1.使用 python 读取 comments.csv 文件，将字段“酒店名称”为空的数据删除，并打印输出删除条目数，将打印内容粘贴至对应报告中，打印内容格式如下：

```
=== “删除酒店名称为空的数据共***条” ===
```

2.将字段“酒店名称”非空的数据保存到 comments1.csv 文件。

3.将符合题目要求的代码答案和 comments1.csv 的前 10 条记录数据复制粘贴至对应报告中。

子任务 2 对异常字段数据进行处理

住客评分的取值范围为[0,5]，其中 5 表示评价最高，0 表示评价最低。如果住客评分超出此取值范围的，都视为异常数据。本任务

使用 python 读取 hotel.csv 文件的数据，将字段“住客评分”异常的数据删除，并打印输出删除条目数，将打印内容粘贴至对应报告中，打印内容格式如下：

```
=== “删除住客评分异常的数据共***条” ===
```

任务二：数据标注

本任务根据酒店的评论数据对酒店的类型打上标签，并将标签数据保存到指定位置。系统提前设定用户评价活跃阈值，如酒店的用户评价数量大于用户评价活跃阈值，则将该酒店的类型标注为“热门”，否则将该酒店的类型标注为“普通”，具体要求如下：

1.编写 python 程序读取 comments.csv 的数据，统计每个酒店的用户评价数量。

2.比较酒店的评价数量和用户评价活跃阈值，给该酒店的类型打上指定的标签（热门/普通），然后将打上标签的数据保存到 comments_tag.csv 中，comments_tag.csv 的字段定义如下：

酒店名称	评论数量	酒店类型
		热门/普通

任务三：数据统计

子任务一：HDFS 文件操作

本任务需要使用 Hadoop、HDFS 命令，已安装 Hadoop 及需要配置前置环境，具体要求如下：

1) 在 HDFS 目录下新建目录/file2，将新建目录的完整命令复制粘贴至对应报告中；

2) 修改权限，赋予目录/file2 最高 777 权限，将修改目录权限的完整命令复制粘贴至对应报告中；

子任务二：计算输入文件中的单词数

本任务需要使用 Hadoop 默认提供的 wordcount 示例来完成单词数统计任务，具体要求如下：

1) 在 HDFS 上创建/user/hadoop/input 目录；

2) 在 master 节点将/var/log/dmesg 文件上传到 HDFSuser/hadoop/input 目录下；

3) 使用 Hadoop 中提供的 wordcount 示例对 HDFS 上的 dmesg 文件进行单词统计，并将统计结果存储到 HDFS 的/user/hadoop/output 目录下；

4) 查看 HDFS 中的/user/hadoop/output 单词数统计结果并将结果前十行截图复制粘贴至对应报告中。

模块三：业务分析与可视化

任务一：数据可视化

子任务 1 使用堆叠图展示城市星级酒店的数量

本任务使用堆叠图展示每个城市星级酒店的数量，本任务具体要求如下：

1) 读取 hotel.csv，使用 pandas 分别统计每个城市的三星级酒店、四星级酒店和五星级酒店的数量。

2) 使用 matplotlib 绘制堆叠图，堆叠图的标题为“各城市星级酒店的数量”，堆叠图的横坐标为城市名称，纵坐标为星级酒店数量。将可视化结果复制粘贴至对应报告中。

子任务 2 使用散点图展示各城市酒店入住客户总人数

将每个城市的所有酒店的入住客户的数量进行累加，就获得了每个城市入住客户的总人数。使用散点图展示不同城市入住客户的总人数，可以直观地对比这些城市的旅游接待能力，本任务具体要求如下：

1) 读取 hotel.csv，使用 pandas 统计每个城市的所有酒店的入住客户总人数。

2) 使用 matplotlib 绘制散点图，散点图的标题为“各城市酒店入住客户总人数”，将可视化结果复制粘贴至对应报告中。

子任务 3 使用柱状图展示酒店的评分数据

本任务使用柱状图展示酒店的评分数据，具体要求如下：

1) 读取 hotel.csv，使用 pandas 统计分别统计三星级酒店、四星级酒店和五星级酒店的住客评分的平均值。

2) 使用 matplotlib 绘制柱状图，柱状图的标题为“不同星级酒店的住客评分数据”，柱状图的横坐标分别为三星级酒店、四星级酒店和五星级酒店，纵坐标为星级酒店对应的住客评分的平均值。柱状图为横向布局，将可视化结果复制粘贴至对应报告中。

任务二：业务分析

子任务 1 分析影响酒店入住客户数量的因素有哪些

结合模块三的任务一制作的可视化效果图，说明影响酒店入住客户数量的因素有哪些，并就如何提高酒店入住率给出相应的措施和建议。

子任务 2 分析影响酒店评分的因素有哪些

结合本模块三的任务一制作的可视化效果图，说明影响酒店评分的因素有哪些，并就如何提高酒店用户满意度和服务水平给出相应的措施和建议。

十、赛项安全

（一）比赛环境

- 1.赛场的布置，赛场内的器材、设备，应符合国家有关安全规定。
- 2.赛场周围设立警戒线，防止无关人员进入，发生意外事件。
- 3.承办院校提供保障应急预案实施的条件，明确制度和预案。
- 4.大赛期间，赛项承办院校须在赛场设置医疗医护工作站。
- 5.参赛选手、赛项裁判、工作人员严禁携带通讯、摄录设备和未经许可的记录用具进入比赛区域。

（二）生活环境

- 1.比赛期间，承办院校须尊重少数民族参赛人员的宗教信仰及文化习俗。
- 2.比赛期间安排的住宿场所应具有住宿、餐饮等经营许可资质。

（三）组队责任

1.各参赛队，须为参赛选手、指导老师、领队等购买大赛期间的人身意外伤害保险。

2.各学校代表队组成后，须制定相关管理制度，并对所有选手、指导教师进行安全教育。

3.各参赛队伍须加强对参与比赛人员的安全管理，实现与赛场安全管理的对接。

（四）应急处理

比赛期间发生意外事故，发现者应第一时间报告赛项执委会，同时采取措施避免事态扩大。

十一、成绩评定

（一）评分队伍组成

成绩评定实行裁判长负责制，裁判组独立完成成绩评定工作。由竞赛裁判经验丰富的人员组成，裁判和监督仲裁具体遴选要求参考《2023全国职业院校技能大赛专家和裁判工作管理办法》中裁判遴选条件。

（二）评分标准制定原则

大数据应用与服务赛项评分，始终贯彻落实公平、公正和公开原则。

1.参与竞赛成绩管理的组织机构包括裁判组、监督仲裁组等，裁判组实行“裁判长负责制”，监督仲裁组组成结构、人数和相关要求按上级规定执行。

2.裁判评分方法，根据评分标准，各项目评分裁判根据选手操作过程和操作结果进行评分，独立评分。

3.成绩产生方法。为保证公开、公平、公正、透明地进行成绩评定，在裁判员的评分中，取两个评分裁判平均分作为选手技能得分。

4.成绩审核方法为各裁判员首先审核自身对选手的原始打分成绩，并签名，裁判长对所有裁判员的打分成绩进行审核，并签名，由监督组按有关要求对竞赛成绩抽检复核。

（三）评分方法

选手在完成任务之后，将任务完成结果进行保存提交，由参赛选手队长签字确认（签工位号）。

评分采取分步得分、累计总分的计分方式。不计参赛选手的个人得分，只记录团体得分。

参赛队提交比赛任务结束请求或者在比赛时间终止后，不得再进行任何操作。否则，视为比赛作弊，给参赛队记警告一次。

在竞赛过程中，选手如有不服从裁判判决、扰乱赛场秩序、舞弊等不文明行为，由裁判长按照规定扣减相应分数并且给予警告，情节严重的取消竞赛资格。

（四）成绩复核与解密

监督仲裁组将对赛项总成绩排名前 30%的所有参赛队伍（选手）的成绩进行复核；对其余成绩进行抽检复核，抽检覆盖率不得低于 15%。如发现成绩错误以书面方式及时告知裁判长，由裁判长更正成绩并签字确认。复核、抽检错误率超过 5%的，裁判组将对所有成绩进行复核。如有以上异常情况，应在专家组组长主持下，由裁判长带领裁判员、监督仲裁员共同处理。

成绩复核、确认无误后进行成绩排名，得出排名结果后进行解密，不允许先解密后排序。如无以上异常情况，成绩单由裁判长、监督仲裁组长共同签字确认。

（五）成绩公布方法

竞赛成绩复核无误后，经裁判员、裁判长、仲裁人员审核签字后确定。若有异议，经过规定程序仲裁后，按照仲裁结果公布比赛成绩。

（六）评分标准

表 8 评分标准

模块	任务	主要知识与技能点	分值
模块一：平台搭建与运维	任务一：大数据平台搭建	Hadoop 完全分布式下的 JDK 的解压安装、JDK 环境变量配置、节点配置、Hadoop 配置文件修改、运行测试等。	10
	任务二：数据库配置维护	使用 SQL 语句建立数据库和表。 使用 SQL 语句对表进行增删改操作。 使用 SQL 语句对表进行统计查询操作。将 CSV 文件导入到数据库中。	20
	小计		30
模块二：数据获取与处理	任务一：数据获取与清洗	使用 Python 程序读取 CSV 文件。 使用 Python 程序处理空字段数据、异常字段数据。	10
	任务二：数据标注	使用 Python 对指定数据进行分类标注。 使用 Python 将标注后的数据保存到指定位置。	10
	任务三：数据统计	HDFS 上传 CVS 文件到指定目录下。 使用 Hadoop 默认提供的 wordcount 示例来完成单词数统计任务。 将计算结果保存到 HDFS 指定目录下。	15
	小计		35
模块三：业务数据分析与可视化	任务一：数据可视化	使用堆叠图展示相关统计数据 使用散点图展示相关统计数据 使用柱状图展示相关统计数据	20
	任务二：业务分析	理解业务场景，对数据报表进行分析研究，给出相应的建议和举措。	10
	小计		30
模块四：职业素养	考察职业素养	竞赛团队分工明确合理、操作规范、文明竞赛	5
	小计		5
总分			100

十二、奖项设置

本赛项奖项设团体奖。按照参赛队数量的 10%、20%、30%分别设立一等奖、二等奖、三等奖。

如出现参赛队总分相同情况，依序按照模块二、模块一、模块三得分高低进行排名，在前序模块得分相同的情况下，按照后续模块得分排名。如果所有任务分值相同，则查看文档撰写规范、职业素养的分值进行排序。

十三、申诉与仲裁

（一）各参赛队对不符合赛项规程规定的设备、工具、计算机软硬件、竞赛执裁、赛场管理及工作人员的不规范行为等，可向赛项仲裁组提出申诉，申诉主体为参赛队领队。

（二）申诉启动时，由参赛队领队向赛项仲裁工作组递交亲笔签字同意的书面申诉报告。申诉报告应对申诉事件的现象、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。非书面申诉不予受理。

（三）提出申诉应在赛项比赛结束后 2 小时内提出。超过 2 小时不予受理。

（四）赛项仲裁组在接到申诉报告后的 2 小时内组织复议，并及时将复议结果以书面形式告知申诉方。申诉方对复议结果仍有异议，可由领队向大赛仲裁工作组提出申诉。大赛仲裁工作组的仲裁结果为最终结果。

（五）申诉方不得以任何理由拒绝接收仲裁结果；不得以任何理由采取过激行为扰乱赛场秩序。仲裁结果由申诉人签收，不能代收；如在约定时间和地点申诉人离开，视为自行放弃申诉。

（七）申诉方可随时提出放弃申诉。

十四、竞赛观摩

竞赛期间采用线上观摩，观摩人员必须服从现场工作人员的指挥，保持安静，不得大声喧哗，不得干扰比赛。

十五、竞赛直播

条件允许时，本赛项进行网上直播。