

附件

湖南省职业院校技能竞赛
高职高专组电子信息专业类
“Python 程序开发”赛项竞赛规程

一、竞赛内容

Python 程序开发赛项以企业真实项目为基础,采用市场主流软件开发架构和实际操作形式进行现场编程设计。竞赛基于 Django 框架搭建,采用“网络爬虫”、“数据清洗”、“数据分析与可视化”、“机器学习”4 个模块。主要涉及知识和技能如下:

模块	模块名称	工作任务
一	网络爬虫	理解业务需求,编写代码实现功能。
		通过 requests、urllib、selenium、scrapy、scrapy-redis 等爬虫库爬取数据。
		通过 Xpath、BeautifulSoup 等库或使用正则表达式等方式从响应内容中解析数据。
		通过 Python 文件操作、Pandas 库以及 Django ORM 等方式将爬取的数据存储到文件或数据库中。
二	数据清洗	理解业务需求,编写代码实现功能。
		通过 Django ORM 或 Pandas 完成数据读取、数据存储等操作。
		对数据进行全方位探查分析并梳理清洗思路。
		使用 NumPy、Pandas 等库对数据进行清洗,

		如格式统一，数据类型转换，缺失值、异常值、重复值处理等。
		使用 Navicat 工具管理数据库。
三	数据分析与可视化	对业务需求及功能的理解，规划业务流程，并编码实现功能。
		使用 NumPy、Pandas 等数据分析工具对数据从各维度进行分析。
		使用 Matplotlib、PyEcharts、seaborn 等可视化工具将分析结果转化为合适的图表，如柱状图、散点图、折线图、饼图、关系图、雷达图、热力图等。
		将可视化结果保存到本地或通过 Django 框架渲染出来。
		使用 Navicat 工具管理数据库。
四	机器学习	根据业务需求编写代码实现功能。
		通过 pandas、Django ORM 读取数据，或使用机器学习库自带数据集并进行特征提取。
		采取合适的手段对数据进行预处理。
		基于 sklearn 库提供的接口选取算法模型、训练模型并对模型进行调优。
		使用模型进行预测，将预测结果进行可视化展示，或通过 Django 框架渲染到页面上。
		使用 Navicat 工具管理数据库。
职业素养		操作规范、代码清晰易读、遵守考场纪律、收纳整理干净整洁、文明竞赛。

二、竞赛方式

个人赛。

三、竞赛时量

总时量：240 分钟。

四、名次确定办法

以竞赛总成绩从高到低排序确定名次，不设并列名次。

成绩相同的，按照竞赛项目题型的顺序对选手得分的高低进行排序，如果第一个题型能得到排序则按此确定名次，否则对后续题型得分依次进行排序，只要能区分名次就终止后续题型的判断。如果以上都相同，按照代码行数从少到多排序（以行数少为优，空行不计入行数）。

五、评分标准与评分细则

（一）评分标准

竞赛总分为 100 分，各模块分数分配如下表所示。其中各模块均是技术技能操作考核共 95 分，职业素养 5 分，评分细则表详细列写了模块中每个具体工作任务、达成目标和对应分值。规定每个具体任务达成目标的获得对应分值，未达成目标的不得分，评分标准如下：

模块	模块名称	分数权重
一	网络爬虫	20
二	数据清洗	30
三	数据分析与可视化	30
四	机器学习	15
职业素养		5

合计	100
----	-----

(二) 评分细则

考核环节	工作任务	分值	评分细则
网络爬虫 (20 分)	理解业务需求,编写代码实现功能。	2	结果评分 (1) 运行结果完全达标(业务逻辑、发送请求、解析数据、数据存储): 100% (2) 运行结果部分达标: 按实现结果占总要求的百分比给分。 (3) 未实现: 0%
	通过 Requests 或 urllib、selenium、scrapy、scrapy-redis 等爬虫库爬取数据。	8	
	通过 Xpath 或 Beautiful Soup 等库或正则表达式等方式从响应内容中解析数据。	7	
	通过 Python 文件操作或 Pandas 库以及 Django ORM 等方式将爬取的数据存储到文件或数据库中。	3	
数据清洗 (30 分)	理解业务需求,编写代码实现功能。	5	结果评分 (1) 运行结果完全达标(业务逻辑、数据读取、数据清洗和处理、数据存储): 100% (2) 运行结果部分达标: 按实现结果占总要求的百分比给分。 (3) 未实现: 0%
	通过 Django ORM 或 Pandas 完成数据读取、数据存储等操作。	10	
	对数据进行全方位探查分析并梳理清洗思路。	5	
	使用 NumPy、Pandas 等库对数据进行清洗,如格式统一,数据类型转换,缺失值、异常值、重复值处理等。	5	
	使用 Navicat 工具管理数据库。	5	
数据分析与	对业务需求及功能的理解,规划	3	结果评分

可视化 (30 分)	业务流程，并编码实现功能。		(1) 运行结果完全达标(业务逻辑，数据统计分析、统计图绘制、Django 视图、Django 模板、Django ORM、Django 路由)：100% (2) 运行结果部分达标：按实现结果占总要求的百分比给分。 (3) 未实现：0%
	使用 NumPy、Pandas 等数据分析工具对数据从各维度进行分析。	10	
	使用 Matplotlib、PyEcharts、seaborn 等可视化工具将分析结果转化为合适的图表,如柱状图、散点图、折线图、饼图、关系图、雷达图、热力图等。	13	
	将可视化结果保存到本地或通过 Django 框架渲染出来。	2	
	使用 Navicat 工具管理数据库。	2	
机器学习 (15 分)	通过 pandas、Django ORM 读取数据,或使用机器学习库自带数据集，并进行特征提取。	2	结果评分 (1) 运行结果完全达标(业务逻辑，数据读取、特征提取、数据预处理、模型选择、模型训练、模型调优，Django 视图、Django 模板、Django ORM、Django 路由)：100% (2) 运行结果部分达标：按实现结果占总要求的百分比给分。 (3) 未实现：0%
	采取合适的手段对数据进行预处理。	3	
	基于 sklearn 库提供的接口选取算法模型、训练模型并对模型进行调优。	8	
	使用模型进行预测,将预测结果进行可视化展示或通过 Django 框架渲染到页面上。	2	
职业素养 (5 分)		5	操作规范、代码清晰易读、遵守考场纪律、收纳整理干净整洁、文明竞赛。

六、赛点提供的设施设备仪器清单

（一）竞赛设备

选手以个人为单位参赛，每人需要 1 台 PC 机（用于运行服务端与客户端开发）。

裁判区域：供裁判休息及工作场地。配电脑，A4 激光打印机 1 台，桌椅，饮水机，纸杯，文具用品等。

（二）硬件环境及配置

类别	部件	参数
选手工位客户端	CPU	Intel 10 代 i5 及以上
	内存	8G 及以上
	硬盘	固态硬盘 256GB 及以上
	显示器	19 寸及以上
选手工位网络	\	200Mbps 及以上
U 盘或移动硬盘	\	64GB 及以上

（三）软件环境

类别	名称	版本号
竞赛平台	中慧云启 Python 程序开发平台	V2.0
选手工位	PyCharm Community Edition	Version 2021 or upper
	Python	Version 3.7.x
	Chrome	Version 90.x or upper
	Navicat	Version 16 or upper
	MySQL	Version 5.7 or upper
	WPS	Version 2019 or upper
	django	Version 3.2.x or upper
	PyMySQL	Version 1.0.x or upper

	BeautifulSoup	Version 4.11.x or upper
	requests	Version 2.26.x or upper
	lxml	Version 4.6.x or upper
	pyecharts	Version 1.9.x or upper
	Matplotlib	Version 3.4.x or upper
	Numpy	Version 1.19.x or upper
	Pandas	Version 1.3.x or upper
	Redis 数据库	Version 5.0.x or upper
	redis	Version 3.5.x or upper
	django_redis	Version 5.1.x or upper
	sklearn	Version 1.0.x or upper
	openpyxl	Version 3.0.x or upper
	Scrapy	Version 1.8.0 or upper
	Scrapy-redis	Version 0.7.1 or upper
	Selenium	Version 4.1.3 or upper
	seaborn	Version 0.11.2 or upper

（四）云端环境

类别	部件	参数
服务器	CPU	2 颗 Intel Xeon 银牌 4214R 以上
	内存	服务器内存 256GB 以上
	硬盘	480GB 以上 SSD*2 （Raid1）
	网卡	2*1GbE 及以上
网络	\	1000Mbps 及以上

备注：具体设备由赛点提供。

七、选手须知

（一）选手自带工（量）具及材料清单

无需选手自带工具。

（二）主要技术规范及要求

该赛项主要涉及以下国家标准，参赛选手在实施竞赛项目中要求遵循如下规范：

序号	标准号	中文标准名称
1	DB21/T 2347.3-2014	信息技术行业职业技能第3部分：软件开发
2	GB/T 32423-2015	系统与软件工程 验证与确认
3	GB/T 32424-2015	系统与软件工程用户文档的设计者要求
4	GB 8566-1988	计算机软件开发规范
5	SJ/T 10367-1993	计算机过程控制软件开发规程
6	GB/T 36475-2018	软件产品分类
7	GB/T 36964-2018	软件工程 软件开发成本度量规范
8	GB/T 37691-2019	可编程逻辑器件软件安全性设计指南
9	GB/T 25000.2-2018	系统与软件工程系统与软件质量要求和评价
10	GB/T 28174.1-2011	统一建模语言(UML) 第1部分：基础结构
11	GB/T 11457-1995	软件工程术语
12	GB/T 16260.1-2006	软件工程 产品质量 第1部分：质量模型
13	GB/T 32421-2015	软件工程 软件评审与审核
14	GB/T 32423-2015	系统与软件工程 验证与确认
15	GB/T 32904-2016	软件质量量化评价规范
16	GB/T 30998-2014	信息技术 软件安全保障规范

（三）选手注意事项

1. 参赛选手在比赛前应认真阅读赛项规程，严格按照赛项规程参加比赛，避免不必要失误。
2. 各参赛选手应在竞赛开始前一天按照规定的时间段进入赛场熟悉环境。
3. 各参赛选手不得统一着装，并不得穿有身份标识的服装。
4. 参赛选手应按照规定时间抵达赛场，凭身份证、学生证，以

及统一发放的参赛证，完成入场检录、抽签确定竞赛工位号，不得迟到早退。并按工位号入座，检查比赛所需竞赛设备齐全后选手签字方可开始参赛。选手在比赛中应注意随时存盘或存档。竞赛期间不准出场，竞赛结束后方开离场。

5. 参赛选手不得私自携带任何竞赛软硬件工具（各种便携式电脑、各种移动存储设备等）、设计资源、通信工具进入考场。

6. 参赛选手要严格遵守竞赛现场规则，如发现有冒名顶替等舞弊行为者，均取消竞赛资格。如遇到电脑或其他比赛用设备故障，可向裁判提出，获得及时解决。

7. 竞赛过程中，各参赛选手不得与其他人员讨论问题，也不得向裁判、巡视员和其他必须进入考场的工作人员询问竞赛项目的操作流程和操作方法，如有竞赛题目文字不清、软硬件环境故障问题时，可向裁判员询问。

8. 竞赛过程中除裁判和其他必须进入考场的工作人员外，任何其他非竞赛选手不得进入竞赛场地。

9. 竞赛结束后，参赛选手要确认成功提交竞赛要求的文件，裁判员与参赛选手一起签字确认，参赛选手在确认后不得再进行任何操作。

10. 其它未尽事宜，将在赛前向各领队做详细说明。

（四）竞赛直播

1. 赛点提供全程无盲点录像。

2. 可在赛点指定区域通过网络监控观摩比赛。

八、样题（竞赛任务书）

2024 年度“楚怡杯”湖南省职业院校技能竞赛

高职高专组电子信息类

Python 程序开发赛项

（样题）

【时量：240 分钟】

竞
赛
试
题

工位号： ____

一、注意事项

1. 请根据大赛所提供的竞赛环境，检查所列的硬件设备、软件清单、材料清单是否齐全，计算机设备是否能正常。

2. 竞赛结束前，在竞赛平台提供的虚拟机中，根据赛题将各项目代码进行完善整合，并运行；根据竞赛平台左侧的答题区进行答题，根据题目对运行代码及结果进行截图。

3. 竞赛结束时，请将答题区的答题报告进行保存提交操作，答题报告在竞赛结束前可重复提交。

二、竞赛环境

1. PC 机：系统已安装 Python 相关环境、MySQL 数据库、用户名密码分别为：root/123456。

2. 根据任务书说明，从竞赛平台虚拟机桌面获取程序开发项目工程项目包。桌面的工程项目包可以直接使用虚拟机中的 Pycharm 导入、编译、运行和发布。

3. 竞赛平台虚拟机内为断网环境。

三、软件组件

1. Python 编程语言及相关开发环境（Python、PyCharm）
2. Web 框架（Django）
3. Python 爬虫组件（Requests、lxml、BeautifulSoup、Scrapy、Selenium）
4. Python 数据分析组件（NumPy、Pandas）
5. Python 可视化组件（Matplotlib、Pyecharts、seaborn）
6. Python 机器学习组件（Scikit_Learn）

四、赛题

模块一：网络爬虫（20 分）

任务一：获取网站数据

【任务说明】

数据是很多企业的生命，没有数据就没有一切。企业首先要解决的问题就是数据问题，获取数据的手段有很多种，其中爬虫就是性价比最高的一种。现有一个视频网站，网站上有大量用户对各种视频的播放、评论、点赞等数据，请根据具体要求，编写爬虫实现数据抓取。

【任务要求】

以网站首页为入口，从该页面获取各大视频分类，每个分类下面均有大量视频及相应的点击、播放等数据。现需要通过爬虫抓取相应数据，具体要求如下：

- 1.使用 requests 库向 url 发送请求；
- 2.使 BeautifulSoup 或 Xpath 从响应内容中解析数据；
- 3.从首页中获取视频分类名和各类别链接；
- 4.向各类别链接发送请求，从响应内容中获取视频具体的播放、评论、点赞等数据；
- 5.将抓取的数据存入 MySQL 数据库中；
- 6.绕过网站的反爬虫检测；

【操作说明】

1.从虚拟机桌面“赛题/01_网络爬虫/01_获取网站数据”文件夹中获取相关项目工程代码。

2.完成任务后，根据平台“答题区”对应内容要求，完成答题报

告并点击“保存”按钮实时保存。

模块二：数据清洗（30分）

任务一：数据清洗

【任务说明】

数据清洗是数据分析过程中很重要的一个环节，没有高质量的数据清洗就没有高质量的数据分析。在不准确的数据基础上做出的分析，结论将变得毫无价值和意义。

现有一份某医疗机构的药品销售数据，请根据任务要求完成数据清洗功能。

【任务要求】

数据集中有购药时间、社保卡号、商品编码、商品名称、销售数量、应收金额、实收金额几个字段，请你使用 NumPy 和 Pandas 按如下要求对数据进行清洗：

- 1.购药时间就是销售时间，为了后续分析更好的理解字段，将"购药时间"改为"销售时间"；
- 2.任何一条数据中只要有一个缺失值就删除该条数据；
- 3.将销售数量、应收金额、实收金额三列的数据类型转换为 float64；
- 4.销售时间中包含日期和星期几，要求数据最终只保留日期，并把销售日期从字符串类型转换为日期数据类型；
- 5.转换日期过程中不符合日期格式的数值会被转换为空值，需要删除列（销售时间，社保卡号）中为空行；
- 6.按销售日期进行升序排列，排序后的索引已被打乱，需要修改

成从 0 到 N 按顺序的索引值；

7.将有异常值（如销售数量、应付金额、实付金额为负数）的数据进行删除。

8.将清洗后的数据保存为 CSV 文件。

【操作说明】

1.从虚拟机桌面“赛题/02_数据清洗/01_数据清洗”文件夹中获取相关项目工程代码。

2.完成任务后，根据平台“答题区”对应内容要求，完成答题报告并点击“保存”按钮实时保存。

模块三：数据分析及可视化（30 分）

任务一：药品消费趋势分析

【任务说明】

现有某医疗机构 10 年的药品销售数据，请根据任务要求完成数据分析。

【任务要求】

读取所需数据集后对数据进行必要的清洗，分析中成药和西药在这 10 年中销售占比以及销售额的变化情况，绘制出堆叠面积图并用 Django 框架渲染到前端页面，绘图要求如下：

- 1.使用 PyEcharts 库绘制堆叠面积图；
- 2.使用 Django 框架在前端页面中渲染展示堆叠面积图；
- 3.示意图如下：

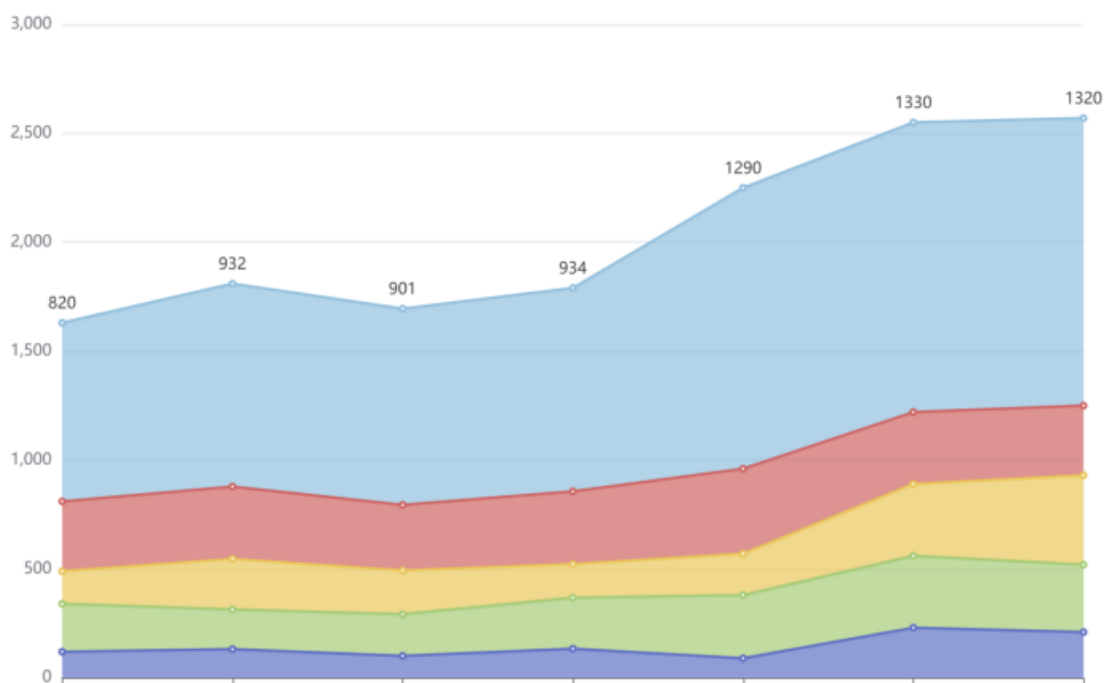


图 1：堆叠面积示意图

【操作说明】

- 1.从虚拟机桌面“赛题/03_数据分析及可视化/01_药品消费趋势分析”文件夹中获取相关项目工程代码。
- 2.完成任务后，根据平台“答题区”对应内容要求，完成答题报告并点击“保存”按钮实时保存。

任务二：感冒高发期分析

【任务说明】

感冒是一种常见的急性上呼吸道病毒性感染性疾病，多由鼻病毒、副流感病毒、呼吸道合胞病毒、埃可病毒、柯萨奇病毒、冠状病毒、腺病毒等引起。临床表现为鼻塞、喷嚏、流涕、发热、咳嗽、头痛等，多呈自限性。

现有某一医疗机构 10 年的患者诊断数据，请你根据诊断结果进行感冒高发期分析。

【任务要求】

读取所需数据集后，对数据进行必要的清洗，绘制出感冒高发期热力图，分析感冒高发期在每一年中的什么时候，并出具分析报告。
绘图要求如下：

- 1.使用 PyEcharts 库绘制热力图；
- 2.热力图横轴为年份，纵轴为月份；
- 3.使用 Django 框架在前端页面中渲染展示热力图；
- 4.示意图如下：

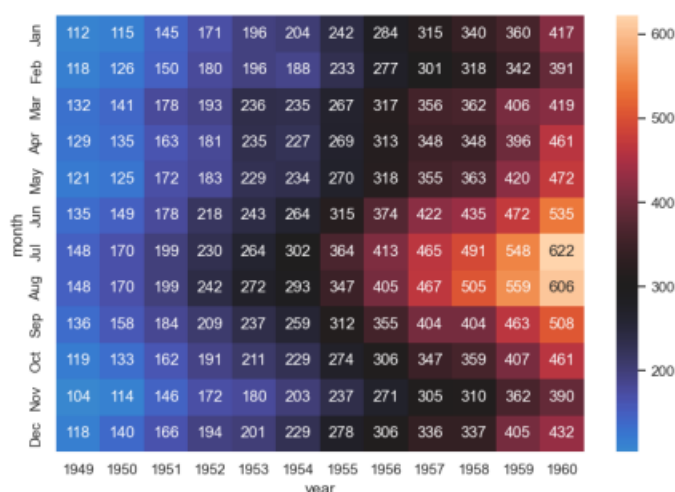


图 2：热力示意图

【操作说明】

- 1.从虚拟机桌面“赛题/03_数据分析及可视化/02_感冒高发期分析”文件夹中获取相关项目工程代码。
- 2.完成任务后，根据平台“答题区”对应内容要求，完成答题报告并点击“保存”按钮实时保存。

模块四：机器学习（15 分）

任务一：药品销量预测

【任务说明】

现有一医疗机构多年来的药品销售数据，基于该数据，选择合适的算法训练模型，并对该机构的药品销量进行预测。

【任务要求】

- 1.读取药品销售数据；
- 2.提取合适的特征；
- 3.对数据进行编码等预处理；
- 4.划分训练集和测试集；
- 5.构建机器学习模型；
- 6.编写模型训练相关代码，完成模型训练；
- 7.将训练好的模型进行保存；
- 8.使用模型进行预测。

【操作说明】

1.从虚拟机桌面“赛题/04_机器学习/01_药品销量预测”文件夹中获取项目工程文档。

2.完成任务后，根据平台“答题区”对应内容要求，完成答题报告并点击“保存”按钮实时保存。

任务二：疾病辅助诊断

【任务说明】

现有一医疗机构十余年来的患者病例数据，数据中包含每个患者的症状，比如咳嗽、发烧、流鼻涕等，以及最后的诊断结果，现要基

于该数据集构建机器学习模型，用于疾病的辅助诊断。

【任务要求】

- 1.读取患者病例数据；
- 2.对数据进行清洗以及预处理；
- 3.特征提取以及特征工程；
- 4.划分训练集及测试集；
- 5.选取合适的机器学习模型；
- 6.使用数据训练模型并进行模型调优；
- 7.将患者症状数据输入模型，预测患者所患疾病；
- 8.将训练好的模型进行保存。

【操作说明】

1.从虚拟机桌面“赛题/04_机器学习/02_疾病辅助诊断”文件夹中获取项目工程文档。

2.完成任务后，根据平台“答题区”对应内容要求，完成答题报告并点击“保存”按钮实时保存。

职业素养（5分）

【任务要求】

参赛选手操作规范、代码清晰易读、遵守考场纪律、收纳整理干净整洁、文明竞赛。