



2024 一带一路暨金砖国家技能发展与技术创新大赛

【Python 程序开发赛项】

赛项编号: BRICS2024-ST-181

样题

金砖国家工商理事会技能发展、应用技术与创新（中方）工作组

一带一路暨金砖国家技能发展与技术创新大赛组委会

竞赛技术委员会专家组制定

2024 年 8 月

赛题说明

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，根据赛题将各试题代码进行完善整合，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，请将答题区的答卷进行提交操作，答卷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。
2. 根据考题说明，从竞赛平台虚拟机桌面获取程序开发项目工程代码包。桌面的工程代码可以直接使用虚拟机中的 Pycharm 导入、编译、运行和发布。

三、赛题说明

本套赛题实操内容包括三个模块：数据清洗、数据分析与可视化、机器学习，综合考察选手职业素养。竞赛时长 240 分钟。

四、样题

模块一：数据清洗（30分）

任务一：药品销售数据清洗

【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）中为空的行；

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/01_数据清洗/”路径下“01_医疗机构销售数据清洗”文件夹中获取相关资料，结果保存至桌面“提交文档/01 ”文件夹中。

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：数据分析及可视化（40 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：

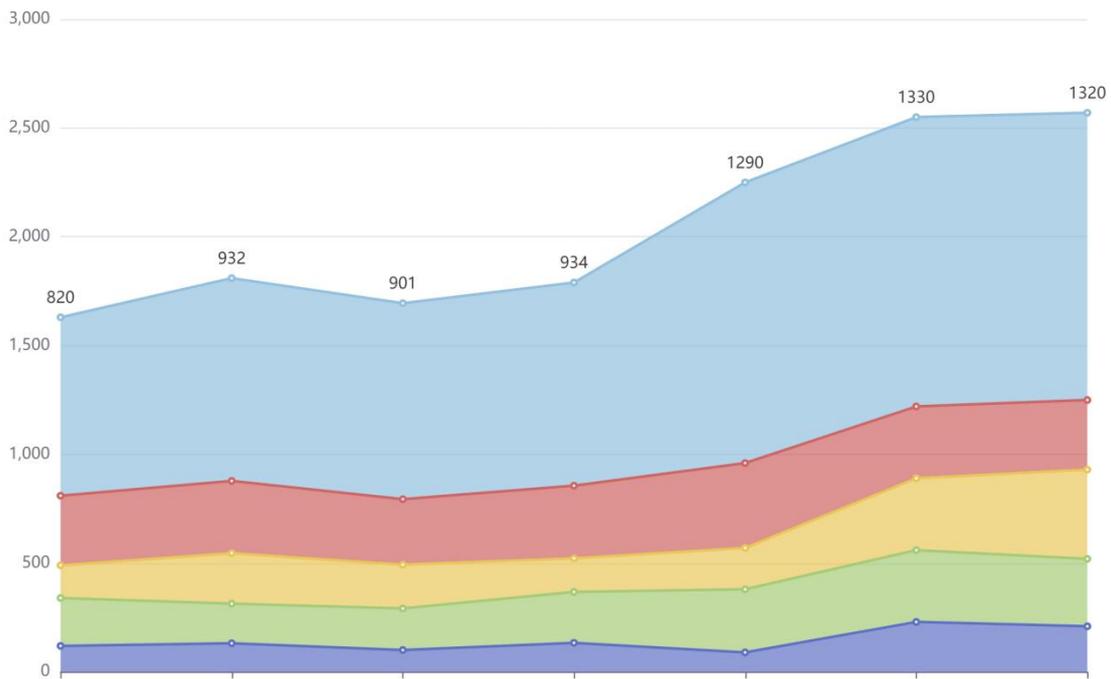


图 1：堆叠面积示意图

【操作说明】

1、从虚拟机桌面“赛题/02_数据分析及可视化/”路径下“02_药品消费趋势分析”文件夹中获取相关资料，结果保存至桌面“提交文档/02 ”文件夹中。

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

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：

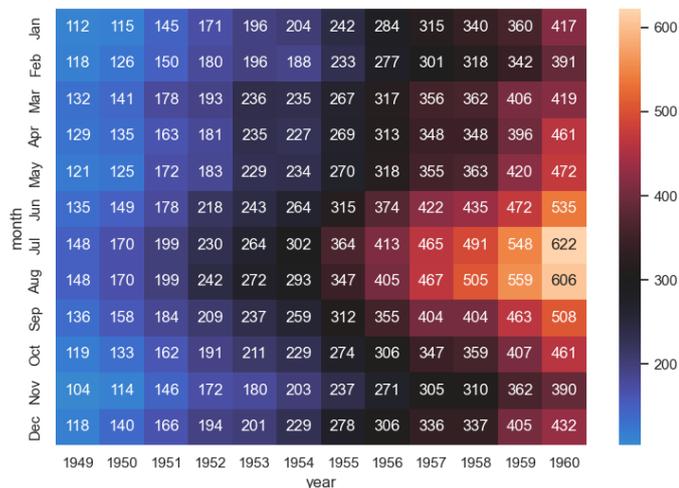


图 2：热力示意图

【操作说明】

- 1、从虚拟机桌面“赛题/02 数据分析及可视化/”路径下“03_

感冒高发期分析”文件夹中获取相关资料，结果保存至桌面“提交文档/02”文件夹中。

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：机器学习（25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/03_机器学习/”路径下“04_药品销量预测”文件夹中获取相关资料，结果保存至桌面“提交文档/03”

文件夹中。

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/03_机器学习/”路径下“05_疾病辅助诊断”文件夹中获取相关资料，结果保存至桌面“提交文档/03”文件夹中。

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、遵守考场纪律、收纳整理干净整洁、文明竞赛。