**人工智能训练师（三级）操作技能考核**

**试题单**

准考证号：

试题代码：

试题名称：Pandas数据清洗培训大纲编写

考核时间：10min

**1.**场地设备要求

（1）人工智能训练师主机；   
（2）Office软件。

**2.**工作任务

A企业是一家康复训练机构，需要利用人工智能技术为康复患者提供个性化的康复训练计划。为了提高人工智能技术的应用效果，其采集了大量的康复数据，现计划对新进技术人员进行数据清洗工具的使用培训。因为Pandas是Python中最为广泛使用的数据分析和操作库之一，因此这次培训将使新进技术人员掌握Pandas的主要数据清洗功能使用，能对大量的康复数据进行清洗，达到人工智能训练师四级/中级工的技能水平。  
请你根据要求补全素材4.1.4.docx中的培训大纲。

**3.**技能要求

（1）能编写初级培训讲义；   
（2）能对初级工、中级工开展知识和技术培训。

**4.**质量指标

（1）培训讲义编写逻辑合理；   
（2）培训资料正确有效；   
（3）语句组织通顺。

**人工智能训练师（三级）操作技能考核**

**试题评分表**

准考证号：

试题代码：

试题名称：Pandas数据清洗培训大纲编写

考核时间：10min

测量分评分表

| 细则编号 | 配分 | 评分细则描述 | 规定或  标称值 | 结果或  实际值 | 得分 |
| --- | --- | --- | --- | --- | --- |
| M1 | 5 | 回答内容与答案类似得1分；最多5分； | 根据数据 |  |  |
| 合计配分 | 5 | 合计得分 | |  |  |

**参考答案**（尽量将细则内容写在上面的表格内，写不下可另写，但要具体可评判）

**在下划线上填写答案。**

# 根据学习大纲补充学习目标

**数据清洗基础理论**

学习目标：掌握数据清洗的基本概念和流程。

内容：

什么是数据清洗？

数据清洗的常见任务：数据去重、缺失值处理、数据格式转换等。

数据清洗在数据分析和模型训练中的重要性。

**Pandas简介与安装**

学习目标**：能够熟练使用Pandas完成数据导入与导出，掌握多格式数据的读写方法**

内容：

读取CSV、Excel、SQL等格式的数据。

保存数据到CSV、Excel等格式。

实践操作：导入和导出康复数据集。

**数据筛选与过滤**

学习目标**：** **能够高效筛选、去重和修复数据，解决康复数据中的常见质量问题**

内容：

条件筛选：基于单列或多列的条件筛选数据。

数据去重：去除重复记录。

缺失值处理：识别、删除和填充缺失值。

实践操作：筛选和过滤康复数据。

**数据转换**

学习目标**：** **掌握复杂数据类型的转换方法，优化数据结构的可分析性**。

内容：

数据类型转换：转换数据类型（如字符串到日期）。

时间序列处理：处理和分析时间序列数据。

数据分列和合并：分割和合并数据列。

实践操作：转换和处理康复数据。

**数据合并与重塑**

学习目标**：** **能够整合多源数据并重塑结构，满足后续分析与建模需求**。

内容：

数据合并：concat和merge方法。

数据重塑：pivot和melt方法。

实践操作：合并和重塑康复数据。

**数据分组与聚合**

学习目标**：** 通过分组与聚合技术，挖掘康复数据的群体特征与统计规律。

内容：

数据分组：groupby方法。

聚合函数：sum、mean、count等。

实践操作：分组和聚合康复数据。

**数据可视化**

学习目标：综合运用所学知识进行数据清洗项目。

内容：

项目介绍：清洗一个多来源的康复数据集。

数据采集：导入多种格式的数据。

数据清洗：使用Pandas进行数据清洗和转换。

项目评审：展示清洗成果，讲解实现思路和遇到的问题。