

潍坊环境工程职业学院 2023-2024 学年第 1 学期期末考试

2022 级 3 年制《python 采集与分析》试题

座号：_____ 姓名：_____ 班级：_____ 成绩：_____

考试时间 90 分钟，满分 100 分。考试结束后，将本试卷交回。

一、配置 pycharm 开发环境（15 分）

1、打开《完美者博客》www.wmzos.com，搜索 pycharm，找到 pycharm 配置教程，按照教程完成开发环境的配置，配置完成后，全屏截图，需要看到 pycharm 开发界面，并看到电脑右下角任务栏上的当前时间。

二、打开《完美者博客》www.wmzos.com，搜索“大数据 2023 期末考试 B 卷数据集”，下载 用户购买记录.rar。数据集包括以下字段：用户 ID、购买日期、购买金额、商品类别、地理位置。代码用到的模块有：pandas, matplotlib.pyplot

1、读取数据集，并检查数据的结构，是否存在缺失值。（15 分）

要求在代码中给出注释，注释中把题目标题写入后再开始写代码，如下：

1、读取数据集，并检查数据的结构，是否存在缺失值。

这里开始写代码

2、对数据进行清洗，处理缺失值，购买日期缺失的数据，用前一条数据中的日期填充，其他字段数据的缺失，删除处理。（20 分）

要求在代码中给出注释，注释中把题目标题写入后再开始写代码，如下：

2、对数据进行清洗，处理缺失值。

这里开始写代码

3、计算每个用户的购买总额，并列前 10 名购买总额最高的用户。（15 分）

要求在代码中给出注释，注释中把题目标题写入后再开始写代码，如下：

3、计算每个用户的购买总额，并列前 10 名购买总额最高的用户

这里开始写代码

4、分析不同商品类别的销售情况，找出销售额最高的商品类别。（15 分）

要求在代码中给出注释，注释中把题目标题写入后再开始写代码，如下：

4、找出销售额最高的商品类别

这里开始写代码

5、可视化分析结果，绘制销售额最高的商品类别的销售趋势图。（20 分）

要求在代码中给出注释，注释中把题目标题写入后再开始写代码，如下：

5、可视化分析结果，绘制销售额最高的商品类别的销售趋势图

这里开始写代码

6、将写好的代码文件改成自己的名字上传到教室机。

如：22.3.1 李子洁.py