第4章 pandas统计分析基础

教案

**课程名称：**Python数据分析与应用

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**64学时（其中理论36学时，实验28学时）

**总学分：**4.0学分

**本章学时**：7学时

# 材料清单

* 1. 《Python数据分析与应用》教材。
  2. 配套PPT。
  3. 数据。
  4. 代码。
  5. 引导性提问。
  6. 探究性问题。
  7. 拓展性问题。

# 教学目标与基本要求

### 教学目标

以餐饮数据为例，介绍数据库数据，csv数据，Excel数据三种常用的数据读取与写入方式。阐述DataFrame的常用属性，方法与描述性统计相关内容。介绍时间数据的转换，信息提取与算术运算。剖析分组聚合方法groupby的原理，用法和三种聚合方法。展现透视表与交叉表的制作方法。通过本章的学习，能够对pandas库有一个整体了解并能够利用pandas库进行基础的统计。

### 基本要求

1. 掌握常见的数据读取方式。
2. 掌握DataFrame常用属性与方法。
3. 掌握基础时间数据处理方法。
4. 掌握分组聚合的原理与方法。
5. 掌握透视表与交叉表的制作。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

1. 数据分析中常见的结构话数据读取方式有那些？
2. DataFrame和数组有什么相似之处？
3. Excel透视表如何制作？

### 探究性问题

探究性问题需要教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问的基础上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. 读取数据库数据为什么需要别的库？
2. 时间数据中存在哪些信息？
3. 以百万级的数据为例，Excel制作透视表和Python制作透视表哪个速度更快？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 能否读取word数据，该如何做？
2. 为什么索引的时候有loc和iloc，设计者的意何在？

# 主要知识点、重点与难点

### 主要知识点

1. 常见的数据读取方式。
2. DataFrame常用属性与方法。
3. 基础时间数据处理方法。
4. 分组聚合的原理与方法。
5. 透视表与交叉表的制作。

### 重点

1. 读写数据。
2. DataFrame数据的常用属性与方法。
3. 基础时间数据处理方法。
4. 分组聚合的原理与方法。
5. 透视表与交叉表的制作。

### 难点

1. 读写数据。
2. 分组聚合的原理与方法。
3. 透视表与交叉表的制作。

# 教学过程设计

### 理论教学过程

1. 读写数据库数据。
2. 读写文本文件。
3. 读写Excel文件。
4. 查看DataFrame的常用属性。
5. 查改增删DataFrame数据。
6. 描述分析DataFrame数据。
7. 转换字符串时间为标准时间。
8. 提取时间序列数据信息。
9. 加减时间数据。
10. 使用groupby方法拆分数据。
11. 使用agg方法聚合数据。
12. 使用apply方法聚合数据。
13. 使用transform方法聚合数据。
14. 使用povit\_table函数创建透视表。
15. 使用crosstab函数创建交叉表实验教学过程。

### 实验教学过程

1. 读写数据库数据。
2. 读写文本文件。
3. 读写Excel文件。
4. 查看DataFrame的常用属性。
5. 查改增删DataFrame数据。
6. 描述分析DataFrame数据。
7. 转换字符串时间为标准时间。
8. 提取时间序列数据信息。
9. 加减时间数据。
10. 使用groupby方法拆分数据。
11. 使用agg，apply，transform方法聚合数据。
12. 制作透视表。
13. 制作交叉表。

# 教材与参考资料

### 教材

黄红梅，张良均．Python数据分析与应用[M]．北京：人民邮电出版社．2018．

### 参考资料

[1] 张良均．Python数据分析与挖掘实战[M]．北京：机械工业出版社．2015．

[2] 张良均．Python与数据挖掘[M]．北京：机械工业出版社．2016．