第3章 数据探索

教案

**课程名称：**Python数据分析与挖掘实战

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**112学时（其中理论49学时，实验63学时）

**总学分：**7.0学分

**本章学时**：6学时

# 材料清单

* 1. 《Python数据分析与挖掘实战》教材。
  2. 配套PPT。
  3. 引导性提问。
  4. 探究性问题。
  5. 拓展性问题。

# 教学目标与基本要求

### 教学目标

本章从应用的角度出发，从数据质量分析和数据特征分析两个方面对数据进行探索分析，最后介绍了Python常用的数据探索函数及用例。数据质量分析要求拿到数据后要先检测是否存在缺失值和异常值；数据特征分析要求在数据挖掘建模前，通过频率分布分析、对比分析、帕累托分析、周期性分析、相关性分析等方法，对采集的样本数据的特征规律进行分析，以了解数据的规律和趋势，为数据挖掘的后续环节提供支持。

### 基本要求

1. 掌握Python中数据质量分析的方法。
2. 掌握Python中数据特征分析的方法。
3. 熟悉Python主要数据探索函数。

# 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. 为什么要进行数据探索？
  2. Python中的数据探索函数有哪些？

### 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. 数据质量分析的主要任务是什么？
2. 有什么手段可以进行数据特征分析？
3. 周期性分析有什么意义？

### 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. 如何处理数据质量分析中发现的缺失值和异常值数据？
2. 处理不一致数据的方法有什么?

# 主要知识点、重点与难点

### 主要知识点

1. 数据质量分析。
2. 数据特征分析。
3. Python主要数据探索函数。

### 重点

1. 数据质量分析。
2. 数据特征分析。

### 难点

1. 数据质量分析。
2. 数据特征分析。

# 教学过程设计

### 理论教学过程

1. 缺失值分析的方法。
2. 异常值分析的方法。
3. 一致性分析的方法。
4. 分布分析的方法。
5. 对比分析的方法。
6. 统计量分析的方法。
7. 周期性分析的方法。
8. 贡献度分析的方法。
9. 相关性分析的方法。
10. 基本的统计特征函数。
11. 统计绘图函数。

### 实验教学过程

1. 使用Python进行数据质量分析。
2. 使用Python进行数据特征分析。
3. 使用Python进行特征统计。
4. 使用Python进行绘图。

# 教材

### 教材

张良均，谭立云，刘名军，江建明．《Python数据分析与挖掘实战（第2版）》[M]．北京：机械工业出版社．2019．