

《Python 高级数据分析》教学大纲

(2020 年制定)

课程代码：42D2244

英文名称：Advanced Python Data Analysis

课程性质：专业限定选修课

前置课程：Python 大数据分析

后置课程：无

学 分：3 学分

课 时：51 课时（理论 34 课时+实践 17 课时）

课程负责人：李树青

主讲教师：李树青等

考核方式：闭卷考试

成绩构成：平时成绩（20%）+期中成绩（20%）+期末成绩（60%）

使用教材：张莉主编，Python 程序设计，北京：高等教育出版社，2019

课程概述：

在大数据时代，各种以大数据分析为基础的应用不断产生，并极大的影响了现有学科专业学习的知识结构。数据分析已经成为常见和必备的专业技能，广泛的应用于各个专业学习和应用实践中，也成为当代大学生应具备的基本信息素养之一。传统的数据分析方法存在着学习成本高、数据处理难度大等困难，而 Python 语言的出现，极大改变了这种局面，凭借着数据分析处理的先天设计优势，和较低的学习门槛，以及广泛的开源工具库包的支持，事实上成为现代大数据分析的主流工具软件。因此，本课程主要讲授 Python 语言基础、Python 数据计算基础、Python 数据查询分析、Python 时间序列分析、Python 可视化分析、Python 机器学习分析和 Python 大数据分析等内容，并结合数据分析案例和方法，以理论联系实际介绍其广泛用途。

教学目的：

在 Python 大数据分析能力层面，通过本课程以 Python 数据分析为主要内容的学习，重点强调常见数据分析方法的应用能力培养。学生在学习完以后，可以对各种专业学习的的海数据分析任务，能够结合常见 Python 库包，在数据方法选择、数据整理、数据分析和数据呈现等各个方面掌握基本的使用方法，能够运用 Python 解决数据分析的实际问题，培养学

生数据计算思维能力、数据科学创新能力，和发现数据问题、分析数据问题和解决数据问题的能力。同时，学生也能掌握简单的程序设计基本方法和面向对象的基础知识，为进一步的专业学习和自学提供基础。本课程也强调对学生工程理论教育目标的达成要求，通过讲解我国大数据分析应用在各个行业的应用案例，激发学生科技报国的家国情怀，培养學生的大国工匠精神。

教学方法：

本课程教学把 Python 大数据分析技能学习与科学精神培养有机结合，强调学生解决实际问题的方法掌握能力，尤其通过案例驱动等教学手段，注重科学思维方法训练，强调与当前社会热点数据分析应用结合。具体教学方法以课堂教学为主、上机实验为辅，并结合作业、测试、讨论、答疑等。课堂教学以多媒体教学手段为主，充分利用多媒体教学的直观性、动态性、可重复性、信息容量大、图文并茂、声像结合等特点来突破教学难点，多角度、多方面、多层次地调动学生的学习兴趣、注意力和学习热情。学生自主完成相关上机实验（结合教师辅导），在完成实验的过程中加深理解课堂教学中讲授的知识与方法。

各章教学要求及教学要点：

第一章 Python 数据查询分析

课时分配：12 课时（含实践教学 4 课时）

教学要求：

要求学生掌握 pandas 库包中常见数据查询分析方法，能结合专业数据资源进行数据案例分析。同时，能结合社会热点问题，提出面向实际问题处理的数据决策支持。

教学内容：

第一节 简单数据查询

- 一、数据建立
- 二、数据选择
- 三、数据显示
- 四、数据更新

第二节 高级数据查询

- 一、数据排序
- 二、数据分组聚合
- 三、数据连接

实践教学：

- 一、实践要求：结合 pandas 库包，以项目报告提交
- 二、选题方向：
 - 1) MovieLens 观影用户的兴趣分析
 - 2) 我国各省 GDP 发展数据分析
 - 3) 《习近平新时代中国特色社会主义思想三十讲》关键词分析

思考题：

- 1. 如何实现子查询？就是利用上次的查询结果来作为当前查询的条件。
- 2. 根据上证股市数据，分析上市公司特征分析。

第二章 Python 时间数据分析

课时分配：12 课时（含实践教学 4 课时）

教学要求：

要求学生了解时间数据的类型及其表达、转换和处理等常见方法，并掌握时间数据的常见分析策略，并了解时间数据分析在我国经济发展分析中的常见应用。

教学内容：

第一节 时间数据

- 一、时间数据类型
- 二、时间数据处理

第二节 时间数据分析

- 一、时间列的基本操作
- 二、时间索引
- 三、特殊时间序列的生成
- 四、时间信息的转换
- 五、时间采样

实践教学：

- 一、实践要求：结合 pandas 和 datetime 等库包，以项目报告提交
- 二、选题方向：
 - 1) 当前网络热点舆情趋势发展分析
 - 2) 我国 GDP 与世界各国 GDP 排名变化对比分析

思考题：

- 1. 基于亚马逊购买信息的用户购买兴趣分析。
- 2. 结合 tushare 库，进行股市趋势分析。

第三章 Python 可视化分析

课时分配：12 课时（含实践教学 4 课时）

教学要求：

要求学生了解各种常见的可视化图表类型及其特点，掌握利用图表多种特征完成多种数据的表达，并掌握可视化图表的绘制和数据分析方法，并了解在舆情管理、社交关系等典型领域的应用分析方法。

教学内容：

第一节 数据可视化基本方法

- 一、数据图表
- 二、Python 可视化工具包

第二节 可视化分析方法

一、线性图绘制

二、散点图绘制

三、饼状图绘制

四、柱状图绘制

五、立体图绘制

实践教学：

一、实践要求：结合 matplotlib、networkx 和 pyecharts 等库包，以项目报告提交

二、选题方向：

- 1) 我国改革开放以来省级人口流动的关系可视化
- 2) 基于网络图的社会主义核心价值观关系可视化分析

思考题：

1. 结合微博数据的社交人群可视化分析。
2. 古典文学著作中人物关系的可视化分析。
3. 如何设置可视化图表的样式以增强显示效果？

第四章 Python 大数据分析

课时分配：15 课时（含实践教学 5 课时）

教学要求：

要求学生了解机器学习的常见基本方法和 Python 机器学习库包的简单使用，并能够利用这些工具在专业数据集合上完成各种特定分析任务。同时，了解 Python 处理大数据的常见方法和策略。

教学内容：

第一节 机器学习

一、机器学习基本方法

二、分类分析

三、回归分析

四、决策数分析

第二节 大数据分析

一、大数据存储

二、大数据计算

实践教学：

一、实践要求：结合 sklearn 和 seaborn 等库包，以项目报告提交

二、选题方向：

- 1) 基于文献数据的学校专业影响力分析
- 2) 基于 CPI 的我国居民生活水平的变化分析
- 3) 基于网络信息抓取数据指标的我国与世界各国发展比较分析

思考题：

1. 通过我国近年来科研基金数据，分析了解我们科研整体特征。
2. 通过市场数据，了解我国产品品牌的类别分析及其发展预测。
3. 结合自己已有的数据集，尝试完成小规模的数据分析。

参考书目

1. 嵩天, 礼欣, 黄天羽. Python 语言程序设计基础 (第 2 版) [M]. 北京: 高等教育出版社, 2017.
2. [美] 韦斯·麦金尼. 利用 Python 进行数据分析 (第 2 版) [M]. 北京: 机械工业出版社, 2018.
3. 马克·卢茨. Python 学习手册 (第 5 版) [M]. 北京: 机械工业出版社, 2018.
4. 芒努斯·利·海特兰德. Python 基础教程 (第 3 版) [M]. 北京: 人民邮电出版社, 2016.
5. Wesley Chun. Python 核心编程 (第 3 版) [M]. 北京: 人民邮电出版社, 2016.
6. 李刚. 疯狂 Python 讲义[M]. 北京: 电子工业出版社, 2018.
7. 埃里克·马瑟斯. Python 编程从入门到实践[M]. 北京: 人民邮电出版社, 2016.
8. 弗朗索瓦·肖莱. Python 深度学习[M]. 北京: 人民邮电出版社, 2018.
9. 崔庆才. Python 3 网络爬虫开发实战[M]. 北京: 人民邮电出版社, 2018.
10. 王宇韬, 房宇亮, 肖金鑫. Python 金融大数据挖掘与分析全流程详解[M]. 北京: 机械工业出版社, 2019.
11. 余本国. 基于 Python 的大数据分析基础及实战[M]. 北京: 中国水利水电出版社, 2018.

执笔人：李树青

2020 年 6 月

审定人：曹杰

2020 年 6 月

学院负责人：曹杰

2020 年 6 月