2025 年春季学期

《Python 数据分析》课程教学大纲

课程名称

中文名称：Python 数据分析

英文名称：Data Analytics with Python

注意：《Python 数据分析》和《Python 数据分析上机》必须在系统中同时选课。

学分

2 学分，总计约 64 学时（含上机实践约 32 学时）

教学团队

授课教师：步一（buyi@pku.edu.cn）、孟凡（mengfan@pku.edu.cn）

助教： 23 硕 陈洪侃（chenhongkan@pku.edu.cn）、21 本周凯乐（2100016604@stu.pku.edu.cn）、21 本田逸凡（tianyifan@stu.pku.edu.cn）

上课时间

每周一 15:10-17:00 和 18:40-20:30

上课地点

理教 309

先修课程

1. 《计算概论》（Introduction to Computation，原《文科计算机基础》）及以上或其他相关课程；
2. 使用 Python 语言的基本能力。

中文简介

本课程讲授数据分析的方法，并采用 Python 作为实践工具。本课程要求学生在开课前具备使用 Python 的基本能力。本课程共分为三个模块：第一个模块（基础篇）首先简要介绍和回顾 Python 程序设计的语法，并介绍使用 Python 进行数据分析的基础，如使用 Numpy、pandas、matplotlib 和 seaborn 等；第二个模块（应用篇）展示多个使用 Python 进行数据分析的应用场景，如探索式

数据分析、时间序列分析、机器学习、社会网络分析、图像分析、文本挖掘等；第三个模块（总结篇）将由学生在课堂内展示小组作业，并进行一次期末考试。本课程是“大数据管理与应用”专业必修课，也是信息管理系其他专业和政府管理学院数字治理方向的专业限选课，同时也欢迎其他相关专业同学选修。

英文简介

This course teaches data analytics methods and adopts Python as the practical tool. Note that this is not a Python programming course; thus, students enrolling this course should have some basic capacity of Python. There are three modules in this course. In the first module, we introduce (and/or recap) the basic syntax of Python and talk about some basics of how to use Python to do data analytics, such as the Numpy, pandas, matplotlib, and seaborn packages. In the second module, we showcase many real data-analytics scenarios, such as exploratory data analysis, time series data analytics, machine learning, social network analysis, image data analysis, and text mining. The last module extends the previous modules and summarizes the course. This is a required course for the “Big Data Management and Application” major and is an elective course for other undergraduate-level majors in the Department of Information Management (e.g., Library Science and Information Management and Information Systems) and for “Digital Governance” major in the School of Government. We also welcome students from other schools/departments to enroll in this course.

学习目标

1. 熟练掌握 Python 的语法，能正确而熟练地使用 Python 进行程序的设计，并识读和编写较复杂程度的程序；
2. 熟练掌握 Python 数据分析的几个包，包括 Numpy、pandas、matplotlib 等，并能使用这些包完成一定的数据分析操作；
3. 能够使用 Python 解决实际应用问题，培养计算思维能力、创新能力和发现问题、分析问题和解决问题的能力。

建议授课学期

本科一年级第二学期

授课语言

中英双语

教材

1. Wes McKinney 著，徐敬一 译. 利用 Python 进行数据分析[M]. 北京：机械工业出版社，2018.
2. Daniel Y. Chen 著，武传海 译. Python 数据分析：活用 Pandas 库[M]. 北京：人民邮电出版社，2019.

参考书

1. Prabhanjan Tattar, Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta 著，刘旭华、李晗、闫晗 译. 数据科学实战手册（第 2 版）[M]. 北京：人民邮电出版社，2019.
2. David Natingga 著，封强、赵运枫、范东来 译. 精通数据科学算法[M]. 北京：人民邮电出版社，2018.
3. Mark Summerfield 著，王弘博、孙传庆 译. Python 3 程序开发指南（第 2 版） [M]. 北京：人民邮电出版社，2015.
4. 埃里克·马瑟斯 著，袁国忠 译. Python 编程：从入门到实践（第 2 版） [M]. 北京：人民邮电出版社，2020.
5. Ivan Idris 著，张驭宇 译. Python 数据分析基础教程：Numpy 学习指南（第 2 版） [M]. 北京：人民邮电出版社，2014.
6. Robert Johansson 著，黄强 译. Python 科学计算和数据科学应用（第 2 版） [M]. 北京：清华大学出版社，2020.
7. 李宁 编著. Python 爬虫技术 [M]. 北京：清华大学出版社，2019.

课程要求

1. 课程个人作业必须独立完成，严禁合作、讨论、抄袭、套作。
2. 课程个人作业、小组作业不得照搬或抄袭他人观点文字，需列出全部参考资料，必须遵照学术规范与诚信。
3. 期末考试须遵守学校关于考试的有关要求。
4. 所有作业必须在规定上课日期的课前提交（如上课时间为某天下午 15:10，则必须在当天 15:09 前提交到指定位置）。除遇不可抗力（不包括时间管理不善、课程冲突、数据或文档丢失等问题），如作业迟交在 24 小时以内，总分扣除 20%；迟交在 48 小时以内，总分扣除 40%；迟交在 72 小时内，总分扣除 60%；迟交在 96 小时内，总分扣除 80%；迟交 96 小时以上，该次作业不计入总分。

教学大纲

注：以下内容可能会根据实际情况进行调整，请以教学网或课程微信群的通知为准。

周	日期	授课教师	模块	授课内容	课堂练习与讲评	作业
0	寒假	-	基础篇	Python 语法基础（自学）		寒假期间自选完成
1	2/17	步一		课程简介；数据分析基础 I（numpy）	Python 语法练习；numpy 练习	-
2	2/24	孟凡		数据分析基础 II（pandas）	pandas 练习 C、A	-
3	3/3	孟凡		数据分析基础 II（pandas）；数据分析基础 III（数据读写）	pandas 练习 B	-

4	3/10	孟凡		数据分析基础 III（数据清洗；数据聚合与分组）	pandas 练习 D、E	布置个人作业 1（数据分析基础）
5	3/17	孟凡		pandas 补充练习；数据可视化 I（matplotlib）	pandas 补充练习；matplotlib 练习 A	-
6	3/24	步一		数据可视化 I（matplotlib）	matplotlib 练习 B	个人作业 1 截止；布置个人作业 2（数据可视化）
7	3/31	步一		数据可视化 II（seaborn）	seaborn 练习；个人作业 1 讲评	-
8	4/7	步一、孟凡		课程讲座；探索式数据分析	期中练习；因子分析练习；探索式数据分析练习；个人作业 2 讲评	个人作业 2 截止；布置期中练习；提示开始组队
9	4/14	步一	应用篇	社会网络分析	社会网络分析练习	再次提醒组队事宜
10	4/21	步一		时序数据分析；使用生成式人工智能工具辅助数据分析工作	时序数据分析练习；期中练习讲评	期中练习截止；提交小组成员名单；布置个人作业 3（时序数据分析+社会网络分析）
11	4/28	孟凡		机器学习基础	机器学习基础练习	-
12	5/5	-		劳动节及校庆假期		
13	5/12	孟凡		图像数据分析	图像数据分析练习	个人作业 3 截止；提交小组作业选题；布置个人作业 4（机器学习基础+图像数据分析+文本数据分析）
14	5/19	步一		文本数据分析	文本数据分析练习；个人作业 3 讲评	布置期末考试样题，介绍期末考试情况
15	5/26	步一、孟凡		期末大作业课堂展示；课程总结		-
16	6/2	-	总结篇	端午节假期		-
17	6/9	-		（考试周）无面授		个人作业 4 和小组作业截止
18	6/16	-		期末考试（14:00 开始）		-

成绩构成

本课程满分 100 分。成绩 60 分及以上者为合格，85 分及以上者为优秀。成绩构成包括：

1. 考勤及课堂参与：不单独占分，但无故缺勤会在总评中扣分。如果因故不能出席需要提前向教师以邮件形式请假，并说明原因。
2. 平时小作业：共 4 次作业，总计 40 分；其中个人作业 1 和个人作业 2 每次 10 分，个人作业 3 共 7 分，个人作业 4 共 13 分。
3. 期中练习：总计 10 分。
4. 期末大作业及展示：组队完成（每组不得少于 3 人，不得多于 5 人，并且要在报告中明确注明每人的贡献；2024 级学生不得与高年级学生组队），总计 15 分，包括课堂汇报和期末报告两部分内容。展示的要求（如时长、形式等）另行通知。
5. 期末考试：总计 35 分。
6. 其他额外加分：不超过 5 分，总成绩不超过 100 分。