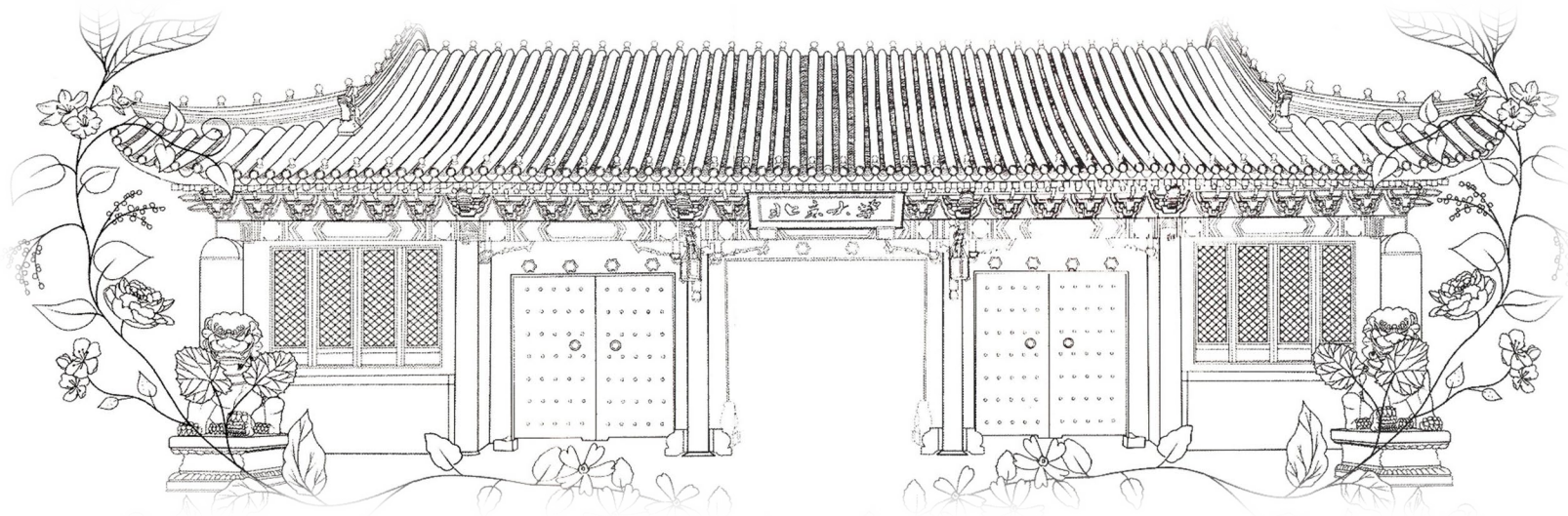


《Python数据分析》

课程介绍

北京大学信息管理系

2021/3/7





课程简介

- 课程性质
 - “大数据管理与应用”专业：专业必修课
 - “图书馆学”和“信息管理与信息系统”专业：专业选修课（2020级：“信息技术”模块）
- 建议先修课：
 - 《数据科学导论》或其他相关课程
 - 《计算概论（C）》（原《文科计算机基础》）及以上或其他相关课程
 - 使用Python语言的基本能力
- 修课学期：
 - 大一下学期





课程简介

- 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的基本语法能力。
- **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。
- **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。



课程简介

- **本课程不是语法课！**
 - 课程重点：数据分析的理论和实践
 - 使用工具：Python



教学目标

- 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。
- 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。
- 能够使用Python解决实际问题，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。



教材与参考书

• 教材

- 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.
- 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

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 李宁 编著. Python爬虫技术 [M]. 北京：清华大学出版社，2019.





课程内容：基础篇

- 寒假：Python语法基础
 - 你完成了吗 😊
- 第1周：课程简介；数据分析基础（I）：Numpy
- 第2周：数据分析基础（II）：pandas
- 第3周：数据分析基础（III）：数据读写；数据清洗；数据聚合与分组
- 第4周：数据可视化（I）：matplotlib
- 第5周：清明节假期
- 第6周：数据可视化（II）：seaborn



课程内容：应用篇

- 第7周：探索式数据分析（I）
- 第8周：探索式数据分析（II）
- 第9周：机器学习（I）
- 第10周：机器学习（II）；时间序列分析
- 第11周：社会网络分析
- 第12周：文本数据分析
- 第13周：图像数据分析
- 第14周：期末大作业汇报
- 第15周：端午节假期
- 第16-17周：期末考试





成绩评定

- 考勤及课堂参与：5分。
- 个人作业：每次作业10分，共5次作业，总计50分。
- 期中练习：15分。
- 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个人的贡献），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。
- 期末考试：总计15分。
- 其他额外加分：不超过5分，总成绩不超过100分。





注意事项

- 2学分，4课时/周
 - 1班：二教301（周一7-8节）、二教306（周一10-11节）
 - 2班：二教420（周一7-8节）、二教304（周一10-11节）
- 本课程和上机课必须同时选
- 上课需要携带自己的笔记本电脑（和充电器）
- 学号单号选1班，双号选2班
- 两个平行班每周的课堂进度几乎相同





注意事项

- 期末大作业允许跨班组队，第14周两个班在同一个教室上课
 - 2020级学生不得与高年级学生组队
- 两个平行班除选课、上课分开外，均在一起评定成绩



课程要求

- 个人作业必须独立完成，严禁合作、讨论、抄袭、套作。
- 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。
- 期末考试须遵守学校关于考试的有关要求。
- 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）。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。



教学团队

- 1班：
 - 教师：步一 (buyi@pku.edu.cn)
 - 助教：高嘉骐 20硕、王越千 17本 (wangyueqian@pku.edu.cn)
 - 教师Office Hour：周一13-15点，方李邦琴楼513
- 2班：
 - 教师：孟凡 (mengfan@pku.edu.cn)
 - 助教：刘千慧 19硕、刘奕 17本 (imliuyi@pku.edu.cn)
 - 教师Office Hour：周一13-15点，方李邦琴楼504



课程微信群



Python 数据分析-21春季



该二维码 7 天内 (3月13日前) 有效, 重新进入将更新





谢谢！

