

《Python 数据分析》课程大纲

课程基本信息 (Course Information)					
课程编号 (Course ID)	1163020011	*学时 (Credit Hours)	32	*学分 (Credits)	2
*课程名称 (Course Name)	Python 数据分析				
	Data Analysis with Python				
先修课程 (Prerequisite Courses)					
*课程简介 (Description)	<p>在信息时代，以几何级数增加的数据形成了海量数据集，如何从这些数据中寻找有价值的信息就变得十分重要，大数据及数据分析技术已经成为信息技术中十分重要的一个环节。随着互联网+的不断发展，在所有行业中，数据分析技术都能发挥一定作用。中国特色社会主义进入新时代，实现中华民族伟大复兴的中国梦开启新征程。党中央决定实施国家大数据战略，吹响了加快发展数字经济、建设数字中国的号角。习近平总书记在十九届中共中央政治局第二次集体学习时的重要讲话中指出：“大数据是信息化发展的新阶段”，并做出了“推动大数据技术产业创新发展、构建以数据为关键要素的数字经济、运用大数据提升国家治理现代化水平、运用大数据促进保障和改善民生、切实保障国家数据安全”的战略部署，为我国构筑大数据时代国家综合竞争新优势指明了方向。</p> <p>近年来，Python 已经成为学习数据分析的首选编程语言。首先，Python 语言简单易用，学习门槛不高，对于非计算机专业人员十分友好；其次，Python 有很多功能强大的第三方库，可以覆盖数据分析中从数据获取、数据清洗、数据可视化、数据挖掘、机器学习等各个环节，是名副其实的数据分析全栈语言。本课程的主要内容就是如何使用 Python 进行数据分析，包括了 NumPy、pandas、matplotlib、pycharts、scikit-learn 等最常用的数据分析库的使用，以及如何使用这些库完成数据分析的整个流程。课程中也会包含一些实际的案例，通过这些案例能够更好的掌握数据分析相关技术。课程中大部分内容都以实践操作为主，涉及纯理论的环节并不太多。课程的重点在于各种数据分析库的基本用法，综合运行这些库进行数据分析是课程的难点。本课程采用课堂讲授与练习相结合的教学方法，课程成绩由考勤、课后作业完成情况以及期末成绩构成。另外，课程不包含 Python 基础语法的介绍，选择该课程之前必须修过 Python 基础语法相关课程或者自学过相关内容。</p>				
*课程简介 (Description)	<p>In the information age, massive data sets are formed by geometrically increasing data. How to find valuable information from these data becomes very important. Big data and data analysis technology have become a very important part of information technology. With the continuous development of Internet+, data analysis technology can play a certain role in all industries. Socialism with Chinese characteristics has entered a new era, and a new journey has begun to realize the Chinese dream of the great rejuvenation of the Chinese nation. The Party Central Committee decided to implement the national big data strategy, sounding the clarion call to accelerate the development of the digital economy and build a digital China. In his important speech at the second collective study of the Political Bureau of the 19th CPC Central Committee, General Secretary Xi Jinping pointed out that "big data is a new stage of informatization development", and made "promoting the innovation and development of the big data technology industry,</p>				

	<p>building a The strategic deployment of the digital economy with data as the key element, the use of big data to improve the level of national governance modernization, the use of big data to promote the protection and improvement of people's livelihood, and the effective protection of national data security has pointed out the direction for my country to build a new national comprehensive competitive advantage in the era of big data. . In recent years, Python has become the programming language of choice for learning data analysis. First of all, the Python language is simple and easy to use, with a low learning threshold, and is very friendly to non-computer professionals; secondly, Python has many powerful third-party libraries, which can cover data acquisition, data cleaning, data visualization, and data mining in data analysis. , machine learning and other links, is a veritable full-stack language for data analysis. The main content of this course is how to use Python for data analysis, including the use of the most used data analysis libraries such as NumPy, pandas, matplotlib, pycharts, and scikit-learn, and how to use these libraries to complete the entire process of data analysis. The course will also include some practical cases, through which you can better master data analysis related techniques. Most of the content in the course is mainly based on practical operation, and there are not too many links involving pure theory. The focus of the course is on the basic usage of various data analysis libraries. The difficulty of the course is to comprehensively run these libraries for data analysis. This course adopts a teaching method that combines classroom lectures and exercises. The course grades are composed of attendance, homework completion and final grades. In addition, the course does not include an introduction to the basic Python grammar. Before choosing this course, you must have taken a Python basic grammar related course or self-taught related content.</p>		
<p>*教材 (Textbooks)</p>	<p>《Python 数据科学手册》，作者:[美]杰克·万托布拉斯 (Jake VanderPlas) 出版社:人民邮电出版社 出版时间:2018 年 02 月，版次 1，国际标准书号 ISBN: 9787115475893。</p>		
<p>参考资料 (Other References)</p>			
<p>*课程类别 (Course Category)</p>	<p><input type="checkbox"/>公共基础课/全校公共必修课 <input checked="" type="checkbox"/>通识教育课 <input type="checkbox"/>专业基础课 <input type="checkbox"/>专业核心课/专业必修课 <input type="checkbox"/>专业拓展课/专业选修课 <input type="checkbox"/>其他_____</p>		
<p>*授课对象 (Target Students)</p>	<p>全校本科生</p>	<p>*授课模式 (Mode of Instruction)</p>	<p><input type="checkbox"/>线上，教学平台_____</p> <p><input checked="" type="checkbox"/>线下 <input type="checkbox"/>混合式 <input type="checkbox"/>其他</p> <p><input type="checkbox"/>实践类 (70%以上学时深入基层) _</p>
<p>*开课院系 (School)</p>	<p>计算机教研部</p>	<p>*授课语言 (Language of Instruction)</p>	<p><input checked="" type="checkbox"/>中文 <input type="checkbox"/>全外语_____</p> <p><input type="checkbox"/>双语: 中文+ _____ (外语讲授不低于 50%)</p>
<p>*授课教师信息 (Teacher Information)</p>	<p>课程负责人 姓名及简介</p>	<p>徐卫克，1980 年生，男，中国社会科学院大学计算机教研部教师，计算社会科学研究中心成员，主要研究方向为数据分析、人工智能。教授课程有《大学计算机》、《Python 数据分析》、《Python 深度学习》、《Python 编程导论》、《数据获取与网络爬虫》等。</p>	

		团队成员 姓名及简介					
学习目标 (Learning Outcomes)		1. 学会搭建数据分析开发环境; 2. 了解 NumPy 库的基本用法; 3. 熟练掌握 Pandas 库的基本操作; 4. 了解和掌握数据处理常用操作; 5. 掌握 matplotlib、seaborn 等可视化库的用法; 6. 了解时间序列数据分析的基本方法; 7. 掌握使用 Python 进行简单统计分析; 8. 了解使用 scikit-learn 库进行机器学习。					
*考核方式 (Grading)		平时成绩 30%，期末成绩 70%					
*课程教学计划 (Teaching Plan)							
周次	周学时	其中					教学内容摘要 (必含章节名称、讲述的内容提要、实验的名称、教学方法、课堂讨论的题目、阅读文献参考书目及作业等)
		讲授	实验课	习题课	课程讨论	其他环节	
第一周	3	3					教学方法：讲授和案例教学。 第一章 Python 数据分析简介 1.1 数据分析流程简介 1.2 本机环境配置 1.3 线上环境配置 1.4 Jupyter Notebook 启动及使用 1.5 Python123 环境介绍 案例：数据分析入门案例—泰坦尼克
第二周	3	3					教学方法：讲授和案例教学。 第二章 Numpy 2.1 多维数组 2.2 索引与切片 2.3 数组的运算 2.4 数组的存取 案例：图片像素修改 求解线性方程组
第三周	3	3					教学方法：讲授和案例教学。 第三章 pandas 3.1 基本数据结构 3.2 基本操作 3.3 运算 3.4 层次化索引 3.5 可视化 案例：探索 Pokemon 数据

						探索 Chipotle 快餐数据 数据分析流程 california 房价数据展示
第四周	3	3				教学方法：讲授和案例教学。 第四章 数据读取与存储 4.1 CSV 文件 4.2 txt 文件 4.3 JSON 文件 4.4 Excel 文件 案例：出生人口姓名性别统计 随机漫步数据分析 探索 2012 欧洲杯数据
第五周	3	3				教学方法：讲授和案例教学。 第五章 数据清洗与整理 5.1 数据清洗 5.2 数据合并与重排 案例：鸢尾花数据清洗 探索 Iris 鸢尾花数据 探索虚拟姓名数据
第六周	3	3				教学方法：讲授和案例教学。 第六章 数据分组与聚合 6.1 分组 6.2 聚合 6.3 分组运算 6.4 数据透视表 案例：公务员工资数据分析 探索酒类消费数据 探索风速数据
第七周	3	3				教学方法：讲授和案例教学。 第七章 matplotlib 可视化 7.1 线形图 7.2 柱状图 7.3 其它基本图表 7.4 自定义设置 案例：星巴克店铺数据分析 宝可梦图表展示 泰坦尼克数据探索
第八周	3	3				教学方法：讲授和案例教学。 第八章 seaborn 可视化 8.1 样式与分布图 8.2 分类图 8.3 回归图与网格 第九章 pyecharts 可视化 9.1 基础图表

						<p>9.2 其他图表</p> <p>案例：泰坦尼克号 seaborn 绘图</p> <p>数据特征分析</p> <p>航空公司客户价值分析</p> <p>糗事百科数据分析</p> <p>荔枝数据分析</p>
第九周	3	3				<p>教学方法：讲授和案例教学。</p> <p>第十章 时间序列</p> <p>10.1 datetime 模块</p> <p>10.2 时间序列基础</p> <p>10.3 日期</p> <p>10.4 时期</p> <p>10.5 频率转换与重采样</p> <p>案例：自行车租赁数据分析</p> <p>探索美国犯罪数据</p> <p>探索 Apple 公司股价数据</p> <p>Tushare API 介绍</p>
第十周	3	3				<p>教学方法：讲授和案例教学。</p> <p>第十一章 统计分析基础</p> <p>11.1 统计量</p> <p>11.2 统计分布</p> <p>11.3 统计推断</p> <p>11.4 相关性分析</p> <p>案例：</p> <p>回归分析</p> <p>数据特征分析</p>
第十一周	2	2				<p>教学方法：讲授和案例教学。</p> <p>第十二章 机器学习</p> <p>12.1 简介</p> <p>12.2 分类</p> <p>12.3 回归</p> <p>12.4 聚类</p> <p>案例：</p> <p>共享单车需求分析</p> <p>估算 boston 房屋价格</p> <p>手写数字识别</p>
总计	32	32				
备注 (Notes)						