

# 《Python 数据分析与应用:从数据获取到可视化》 课程教学大纲

#### (课程英文名称)

课程编号: 201800522073

学 分: 5 学分

学 时: 46 学时 (其中: 讲课学时 33 上机学时: 13)

先修课程: 《Python 快速编程入门》

后续课程: Python 数据挖掘

适用专业: 信息技术及其计算机相关专业

开课部门: 计算机系

#### 一、课程的性质与目标

《Python 数据分析与应用:从数据获取到可视化》是高校 xx 专业(大数据、人工智能)的一门校定必修专业基础课,涉及科学计算库 NumPy、数据分析工具 Pandas、数据可视化、时间序列分析和文本数据分析。

通过对本课程的学习,学生能够熟悉数据分析的流程和思想,可以利用数据分析技术解决特定业务领域的问题。

## 二、教学条件要求

操作系统: Windows7 版本

Python 环境: Anconda 4.5.9

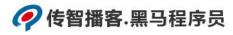
开发工具: Jupyter Notebook

#### 三、课程的主要内容及基本要求

#### 第1章 数据分析概述

章名	数据分析概述	学时	2
----	--------	----	---

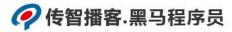
网址: yx.boxuegu.com 教学交流QQ/微信号: 2011168841



	1.	了解数据分	析的背景及应用	月场景				
	2.	. 掌握什么是数据分析以及数据分析的流程						
学习目标	3.	3. 会创建 Pyhton 环境,使用 Anaconda 管理 Python 包						
	4.	会简单使用	Jupyter Notek	ook				
	5.	认识常见的	数据分析工具					
知识点		了解	熟悉	掌握	重点			
数据分析的背景		<b>√</b>						
什么是数据分析		√						
数据分析的应用场景		√						
数据分析的流程					√			
Python 做数据分析的优势		√						
Anconda 的介绍			√					
Anconda 的安装及使用					√			
启用 Jupyter Notebook					√			
Jupyter Notebook 界面介绍及		√						
使用								
常见的数据分析工具			√					

# 第2章 科学计算库 NumPy

章名	ź	科学计算库 NumPy 学		时		5		
	1. 认识 NumPy 数组对象, 会创建 NumPy 数组							
	2. 熟悉 ndarray 对象的数据类型,并会转换数据类型							
W = 1 L L-	3. 掌握数组运算方式							
学习目标	4.	掌握数组的索	引和切	片				
	5.	会使用数组进	行数据	处理				
	6.	熟悉线性代数	模块和	随机数	枚模块的	り使用		
知识点		了解	熟	悉	掌	握	重点	
认识 NumPy 数组对象					<b>√</b>	′		
创建 NumPy 数组							$\checkmark$	
ndarray 对象的数据类型							$\checkmark$	
数组运算			^	/				
整数索引和切片的基本使用							√	
花式(数组)索引的基本使用							√	
布尔型索引的基本使用							$\checkmark$	
数组的转置和轴对称					<b>√</b>	1		
NumPy 通用函数			^	/				
将条件逻辑转为数组运算					<b>√</b>	′		
数组统计运算			1	/				
数组排序			1	/				
检索数组元素					<b>√</b>	′		



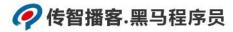
线性代数模块	√			
随机数模块			√	
案例一酒鬼漫步		<b>√</b>		

# 第3章 数据分析工具 Pandas

章名	2	数据分析工具 Pandas		学时	7		
	1. 掌握 Pandas 的两种数据结构						
	2.	掌握 Pandas 雾	<b>尽引的相关</b> 护	操作			
学习目标	3.	掌握 Pandas 的	的常见操作,	包括算术运算	草、排序、统		
		计计算					
	4.	掌握 Pandas 诱	<b>英写数据的</b> フ	方式			
知识点		了解	熟悉	掌握	重点		
Series					√		
DataFrame					<b>√</b>		
索引对象			√				
重置索引				√			
索引操作				√			
算术运算与数据对齐			√				
数据排序				√			
统计计算与描述		√					
认识层次化索引			√				
层次化索引的操作					√		
读写文本文件					<b>√</b>		
读写 Excel 文件					√		
读写 HTML 表格数据					√		
读写数据库					√		
案例一北京高考分数线统计分析			√				

### 第4章 数据预处理

章名	数据预处理			学时		6
	1.	掌握数据清洗常	用操作,	会检查和外	<b>心理各</b>	类有问题的
		数据				
	2.	掌握数据合并的	常用方法	,会使用艺	不同的	方式合并数
<b>光</b>		据				
学习目标	3.	掌握数据重塑的	常见操作	,会重塑	Panda	s 对象的结
		构				
	4.	掌握数据转换的	常见操作	,可以实现	见离散	(化和哑变量
		处理				
知识点		了解	熟悉	掌	握	重点



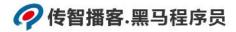
空值和缺失值的处理			√	
重复值的处理			√	
异常值的处理			√	
更改数据类型			√	
轴向堆叠数据				√
主键合并数据				√
根据行索引合并数据				√
合并重塑数据				√
重塑层次化索引				√
轴向旋转			√	
重命名轴索引			√	
离散化连续数据		√		
哑变量处理类别型数据	<b>√</b>			
案例一预处理部分地区信息		√		

# 第5章 数据聚合与分组运算

章名	数扫	数据聚合与分组运		学	时		4	
		算						
	1.	1. 理解分组与聚合的原理						
学习目标	2.	2. 掌握 groupby()方法,可以按照不同的规则进						
		行分组						
	3.	掌握聚合	操作,	会使用	统计划	方法和	聚合方法聚	
		合数据						
		4. 掌握其他分组级运算方法的使用						
知识点		了解	熟	悉	掌	握	重点	
分组与聚合的原理		<b>√</b>						
通过 groupby()方法将数据拆分成组							<b>√</b>	
使用内置统计方法聚合数据					√	′		
面向列的聚合方法	√						<b>✓</b>	
数据转换	√						√	
数据应用	√						<b>√</b>	
案例一运动员信息的分组与聚合			^	/				

# 第6章 数据可视化

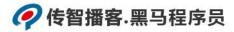
章名	数据可视化		学时	6		
	1. 了解什么是数据可视化					
学习目标	2.	掌握 Matplotli	b绘制图表的	常用方法		
	3.	了解 seaborn 的基本使用				



	4. 了解 bokeh 的基本使用						
知识点	了解	熟悉	掌握	重点			
什么是数据可视化	√						
常见的图表类型		<b>√</b>					
数据可视化的工具	√						
通过 figure()函数创建画布		<b>√</b>					
通过 subplot()函数创建单个子图			√				
通过 subplots()函数创建多个子图			√				
通过 add_subplot()方法添加和选中			<b>√</b>				
子图							
添加各类标签			√				
绘制常见图表			√				
本地保存图形			<b>√</b>				
可视化数据的分布	√						
用分类数据绘图		√					
认识 Bokeh 库	√						
通过 Plotting 绘制图形		√					
案例—画图分析某年旅游景点数据		√	_				

第7章 时间序列数据分析

章名	H	时间序列数据分析		学	:时		6
	1. 了解什么是时间序列,会创建时间序列对象						对象
	2. 会使用时间戳索引和切片选取子集						
	3.	学会创建固定	频率的	时间序	列,能	够调	整时间序列
学习目标		的频率					
	4.	掌握 Period(	时期)	,转	<b>奂时期的</b>	り频率	
	5.	掌握重采样,	实现降	采样利	1升采样	Ė	
	6.	熟悉滑动窗口	的使用				
知识点		了解	熟	悉	掌握	屋	重点
创建时间序列					√		
通过时间戳索引选取子集			^	/			
创建固定频率的时间序列					√		
时间序列的频率、偏移量					√		
时间序列的移动					√		
创建时期对象					√		
时期的频率转换					√		
重采样方法 (resample)		<u> </u>			√		
降采样					√		
升采样					√		
数据统计—滑动窗口					√		



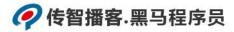
时序模型—ARIMA	√		
案例—股票收盘价分析		√	

# 第8章 文本数据分析

章名	文本	数据分析	沂	学	时		6
	1.	1. 了解文本分析的工具 NLTK 与 jieba,				ba,会安装	
		和使用这	这些工具	具			
	2. 掌握文本预处理的流程						
	3. 掌握文本情感分析,可以用 NLTK 分				分析情感倾		
学习目标		向					
	4. 掌握文本相似度,可以结合 NLTK 与家 度实现相似度分析				与余弦相似		
	5. 掌握文本分类,可以结合 NLTK 与算法对文本						
		进行分类	É				
知识点	-	了解	熟	悉	掌	握	重点
NLTK 与 jieba 概述		√					
安装 NLTK 和下载语料库			^	/			
jieba 库的安装			^	/			
预处理的流程					√	1	
分词					√	′	
词性标注					<b>√</b>	′	
词形归一化					<b>√</b>	′	
删除停用词					<b>√</b>	′	
文本情感分析			~	/			
文本相似度			^	/			
文本分类			^	/			
案例—商品评价分析			^	/			

# 第9章 实战一北京租房数据统计分析

章名	实战一北	京租 学	:时	4		
	房数据统计	计分				
	析					
	1. 掌握数据预处理的方法					
₩ <b>코</b> □1=	2. 掌握使用 Matplotlib 绘制图像的方法					
学习目标	3. 了解百	度地图 API	的使用			
	4. 会基于	数据进行独	立分析			
知识点	了解	熟悉	掌握	重点		
数据读取			√			



重复值和空值处理		√	
数据转换类型		√	
房源数量、位置分布分析		√	
户型数量分析		√	
平均租金分析		√	
面积区间分析		√	

### 四、学时分配

章目	讲课	上机	合计
第1章 数据分析概述	2 学时	0 学时	2 学时
第 2 章 科学计算库 NumPy	4 学时	1 学时	5 学时
第3章 数据分析工具 Pandas	5 学时	2 学时	7 学时
第4章 数据预处理	4 学时	2 学时	6 学时
第5章 数据聚合与分组运算	3 学时	1 学时	4 学时
第6章 数据可视化	4 学时	2 学时	6学时
第7章 时间序列数据分析	4 学时	2 学时	6 学时
第8章 文本数据分析	4 学时	2 学时	6 学时
第9章 实战一北京租房数据统计分析	3 学时	1 学时	4 学时
合计	33 学时	13 学时	46 学时

# 五、考核模式与成绩评定办法

本课程为考试课程,期末考试采用百分制的闭卷考试模式。学生的考试成绩由平时成绩(30%)和期末考试(70%)组成,其中,平时成绩包括出勤(5%)、作业(5%)、上机成绩(20%)。

### 六、选用教材和主要参考书

本大纲是根据教材《Python 快速编程入门》所设计的。

参考书籍:

传智播客.《Python 快速编程入门》人民邮电出版社. 201709

### 七、大纲说明

本课程的授课模式为:课堂授课+上机,其中,课堂主要采用多媒体的方式进行授课,

网址: yx.boxuegu.com 教学交流QQ/微信号: 2011168841



并且会通过测试题阶段测试学生的掌握程度;上机主要是编写程序,要求学生动手完成指定的程序设计或验证。

撰写人:

审定人:

批准人:

执行时间: