信息工程系

教

案

课程名称： 专业技能实训 III

教　　师： 黄梅佳、邱煜佳

总 学 时 ： 54

理论学时： 0

实训学时： 54

上课班级： 计算机应用技术 241、计算机应用技术 241（3+）

授课学期： 2025-2026 第一学期

课题名称	第 1 章 数据分析概述	计划课时	2 课时
教学引入	随着大数据时代的到来，数据得到了前所未有的爆发性增长，我们每天都生活在庞大的数据群体中，能够从数据中挖掘有价值的信息变得愈发重要，数据分析技术应运而生。 数据分析可以运用计算机工具和数学知识处理数据，并从海量数据中发现规律性的信息，帮助企业规避自身问题以及预测未来趋势。由此可见，数据分析在大数据时代扮演着不可估量的角色。接下来，我们就正式进入数据分析的学习吧！		
教学目标	● 使学生了解数据分析产生的背景，能够简述数据分析产生的背景 ● 使学生了解数据分析的概念及层次，能够说出数据分析的概念以及数据分析的 4 个层次 ● 使学生了解数据分析的应用领域，能够列举至少 3 个数据分析的应用领域 ● 使学生熟悉数据分析的流程，能够归纳数据分析的基本流程 ● 使学生了解 Python 做数据分析的优势，能够说出 Python 在数据分析方面有哪些优势 ● 使学生了解 Anaconda 工具，能够说出 Anaconda 工具的特点 ● 使学生掌握 Anaconda 的安装与使用，能够独立在计算机中安装 Anaconda 工具，并通过 Anaconda 工具安装、更新、卸载包 ● 使学生掌握 Jupyter Notebook 的启用方式，能够通过 Anaconda 或命令的方式启用 Jupyter Notebook 工具		
教学重点	● Anaconda 的安装 ● Jupyter Notebook 的基本使用		
教学难点	● 通过 Anaconda 管理包 ● Jupyter Notebook 的基本使用		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	1. **家国情怀与时代责任** 结合 “大数据时代背景”，介绍国家《“十四五” 数字经济发展规划》中 “培育数据要素市场” 的战略部署，引导学生认识数据分析是服务国家数字经济的重要工具，树立 “用数据助力国家发展” 的意识。 2. **科学精神与严谨态度** 在讲解 “数据分析流程” 时，强调 “数据预处理不可篡改数据”“结果需反复验证”，培养学生求真务实的科学态度；在 Anaconda 安装实操中，要求学生耐心排查环境变量问题，渗透 “精益求精” 的工匠精神。 3. **数据伦理与社会责任** 以 “校园 APP 数据采集” 为例，讨论 “如何保护学生隐私（如兼职信息、消费数据）”，引导学生树立数据伦理意识，明确 “技术服务于		

	人，不可滥用数据” 的底线。
教学过程	**第一课时** **（数据分析产生的背景、什么是数据分析、数据分析的应用场景、数据分析的流程、为什么选择 Python 做数据分析、Anaconda 概述、Anaconda 的安装）** 一、创设情景，导入新课 教师通过给学生列举一些数据分析的例子，比如智能推荐等，引导学生了解到数据分析的重要性，从而实现导入新课的目的。 二、新课讲解 **知识点 1-数据分析产生的背景** 教师通过 PPT 讲解数据分析产生的背景。 （1）大数据时代带来的问题。 （2）数据分析技术应运而生。 （3）数据分析的好处。 **知识点 2-什么是数据分析** 教师通过 PPT 讲解什么是数据分析。 （1）数据分析的概念 （2）数据分析的四个层次 ● 描述性分析 ● 诊断性分析 ● 预测性分析 ● 规范性分析 **知识点 3-数据分析的应用场景** 教师通过 PPT 讲解数据分析的应用场景。 （1）营销方面的应用 （2）医疗方面的应用 （3）零售方面的应用 （4）网络安全方面的应用 （5）交通物流方面的应用 **知识点 4-数据分析的流程** 教师通过 PPT 讲解数据分析的流程。 （1）明确目的和思路 （2）数据收集 （3）数据处理 （4）数据分析 （5）数据展现 **知识点 5-为什么选择 Python 做数据分析** 教师通过 PPT 讲解为什么选择 Python 做数据分析。 （1）语法简单精炼，适合初学者入门 （2）拥有一个巨大且活跃的科学计算社区 （3）拥有强大的通用编程能力 （4）人工智能时代的通用语言 （5）方便对接其它语言

知识点 6-Anaconda 概述

教师通过 PPT 讲解 Anaconda 概述。

（1）Anaconda 工具是什么。

（2）Anaconda 工具的特点。

知识点 7-Anaconda 的安装

教师通过 PPT 结合实操的形式讲解 Anaconda 的安装。

（1）Anaconda 工具的下载与安装

① 打开 Anaconda 官方网站的首页。

② 把鼠标放到 Products 菜单上方自动弹出下拉列表。

③ 单击“Anaconda Distribution”选项切换到下载页面。

④ 在“Download”按钮上方右击，选择“在新窗口中打开链接”打开一个新窗口，并开始下载相应的安装包。

⑤ 以管理员身份运行安装程序，打开欢迎界面。

⑥ 单击“Next”按钮进入用户许可协议界面。

⑦ 单击“I Agree”按钮，进入选择安装类型的界面。

⑧ 选择 Just Me 选项，单击“Next”按钮进入选择安装类型的界面。

⑨ 保持默认配置，单击“Next”按钮进入高级安装选项界面。

⑩ 勾选两个复选框，单击“Install”按钮进入正在安装界面。

⑪ 等待片刻后自动进入安装完成界面。

⑫ 单击“Next”按钮进入 Anaconda3 2022.10(64-bit)界面。

⑬ 单击“Next”按钮进入完成安装的界面。

（2）验证 Anaconda 工具安装成功

打开开始菜单，找到 Anaconda3(64-bit)文件夹，单击“Anaconda Navigator”图标，若能够正常启动 Anaconda Navigator，则说明 Anaconda 工具安装成功。

（3）通过计算机给学生演示如何安装 Anaconda 工具。

三、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第二课时

（通过 Anaconda 管理包、启动 Anaconda 自带的 Jupyter Notebook、Jupyter Notebook 界面详解、Jupyter Notebook 的基本使用、常见的数据分析库）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要认识了数据分析，包括背景、概念、应用场景、基本流程、开发工具等，本节课将学习开发工具的使用，包括通过 Anaconda 管理包、启用 Jupyter Notebook。

三、新课讲解

知识点 1-通过 Anaconda 管理包

教师通过 PPT 结合实操的形式讲解通过 Anaconda 管理包。

（1）常用的 Conda 命令。

- 查看当前版本

	● 查看当前环境下的包信息 ● 查找包 ● 安装包 ● 卸载包 ● 更新包 （2）通过 Anaconda Prompt 工具演示常用的 Conda 命令。 **知识点 2-启动 Anaconda 自带的 Jupyter Notebook** 教师通过实操的形式讲解启动 Anaconda 自带的 Jupyter Notebook。 （1）Jupyter Notebook 是什么 （2）启动 Jupyter Notebook 的方式 ● 通过 Anaconda Navigator 启动 Jupyter Notebook ● 通过命令启动 Jupyter Notebook （3）通过 Anaconda Navigator 和命令演示如何启动 Jupyter Notebook **知识点 3-Jupyter Notebook 界面详解** 教师通过 PPT 讲解 Jupyter Notebook 界面详解。 （1）新建 Python 脚本文件 （2）Python 脚本文件窗口的组成部分 ● 标题栏 ● 菜单栏 ● 快捷键区域 ● 编辑区域 **知识点 4-Jupyter Notebook 的基本使用** 教师通过 PPT 讲解 Jupyter Notebook 的基本使用。 （1）Jupyter Notebook 工具的基本操作 ● 编辑和运行代码。 ● 设置标题。 ● 导出文件。 （2）通过 Jupyter Notebook 演示如何编辑运行代码、设置标题和导出文件 **知识点 5-常见的数据分析库** 教师通过 PPT 讲解常见的数据分析库。 （1）NumPy （2）pandas （3）Matplotlib （4）Seaborn （5）Pyecharts （6）NLTK （7）scikit-learn 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 五、布置作业

教学后记	要让学生意识到 Anaconda 管理包作为工具箱的方便性，但也可以先从安装所需的各类依赖库开始做起，后续再感知该工具箱的作用。

课题名称	第 2 章　科学计算库 NumPy	计划课时	6 课时
教学引入	NumPy 作为高性能科学计算和数据分析的基础包，它是本书数据分析相关库的基础，掌握 NumPy 的功能及其用法，将有助于后续其他数据分析相关库的学习。接下来，本章将带领大家学习 NumPy 的基本用法。		
教学目标	● 使学生了解 NumPy 数组的相关概念，能够说出什么是 NumPy 数组、维度、轴和秩 ● 使学生熟悉 NumPy 数组的属性，能够归纳 ndim 和 shape 属性的作用 ● 使学生掌握数据的创建方式，能够灵活创建一维数组和二维数组 ● 使学生掌握数组的数据类型，能够查看与转换数组的数据类型 ● 使学生掌握数组的索引和切片操作，能够灵活地通过不同形式的索引获取数组元素 ● 使学生掌握数组的算术运算，能够实现数组与数组或数组与标量的算术运算 ● 使学生掌握数组的通用函数，能够熟练地使用一元通用函数和二元通用函数进行数学运算 ● 使学生掌握数组的重塑操作，能够通过 reshape()方法实现数组的重塑操作 ● 使学生掌握数组的转置操作，能够通过 T 属性或 transpose()方法实现数组的转置操作 ● 使学生掌握数组的其他操作，能够实现数组的条件逻辑、统计运算、排序操作 ● 使学生熟悉线性代数模块，能够通过 linalg 模块的功能完成矩阵操作 ● 使学生掌握随机数模块，能够通过 random 模块的功能生成包含随机数的数组		
教学重点	● 创建数组 ● 整数索引和切片 ● 花式索引 ● 布尔索引 ● 形状相同的数组间的算术运算 ● 形状不同的数组间的算术运算		
教学难点	● 形状不同的数组间的算术运算 ● 数组的转置 ● 线性代数模块		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	• **基础能力与科技自立** 强调 NumPy 是数据分析的 “底层基石”，如同 “盖房子的地基”，引导学生认识 “扎实掌握基础技术是参与科技自立自强的前提”，避免 “好高骛远、轻视基础” 的心态。 • **数学思维与严谨性** 在讲解 “数组广播机制”“线性代数运算” 时，要求学生验证每一步计算结果（如矩阵相乘的维度匹配），渗透 “数学是科技的语言，严谨是技术的生命” 的理念，培养 “不主观臆断、不省略步骤” 的习惯。		

教学过程	**第一课时** **（NumPy 数组的相关概念、NumPy 数组的属性、创建数组、查看数据类型）** 一、创设情景，导入新课 教师通过给学生提问问题，例如问题是：用什么数据结构存储棋盘上的棋子，并根据学生的问题进行总结，引出像这种形式数据可以使用数组存储，从而实现导入新课的目的。 二、新课讲解 **知识点 1-NumPy 数组的相关概念** 教师通过 PPT 讲解 NumPy 数组的相关概念。 （1）数组 ● 数组在创建时具有固定的大小，不会动态地增长。 ● 数组中所有元素必须具有相同的类型。 ● 数组适用于大量数据的高级数学操作，执行效率更高、代码量更少。 （2）维度 ● 零维是一个无限小的点，没有长度。 ● 一维是一条无限长的直线，只有长度。 ● 二维是一个平面，由长度和宽度组成。 ● 三维是一个立方体，由长度、宽度和高度组成。 （3）轴 ● 一维数组只有一个轴，轴编号为 0。 ● 二维数组有沿行方向和列方向的两个轴，轴编号分别为 0、1。 ● 三维数组有沿着列、行以及由行列组成平面的三个轴，这三个轴的编号分别为 0、1、2。 （4）秩 秩是轴的个数。 **知识点 2-NumPy 数组的属性** 教师通过 PPT 讲解 NumPy 数组的属性。 （1）ndarray 对象 （2）ndarray 对象的常用属性 **知识点 3-创建数组** 教师通过 PPT 结合实操的形式讲解创建数组。 （1）创建数组的方式 ● array()函数：直接传入列表或元组。 ● zeros()函数：创建元素值都是 0 的数组。 ● ones()函数：创建元素值都为 1 的数组。 ● empty()函数：创建一个新的数组，该数组只分配了内存空间，它里面填充的元素都是随机的。 ● arange()函数：创建一个等差数组，需要指定步长。 ● linspace()函数：创建一个等差数组，需要指定元素的数量。 （2）指定数组元素的类型 （3）通过代码演示使用上述几种方式创建数组，并指定类型 **知识点 4-查看数据类型**

	教师通过 PPT 结合实操的形式讲解查看数据类型。 （1）查看数据类型的方式 先通过数组访问 dtype 属性得到 numpy.dtype 类型的对象，再通过该对象访问 name 属性进行获取。 （2）通过代码演示查看数组的数据类型 （3）数组常用的数据类型 三、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第二课时** **（转换数据类型、数组的索引方式、整数索引和切片、花式索引、布尔索引）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了 NumPy 数组的相关概念、NumPy 数组的属性、创建数组、查看数据类型，本节课将学习转换数据类型以及索引和切片，包括整数索引、花式索引、布尔索引和切片。 三、新课讲解 **知识点 1-转换数据类型** 教师通过 PPT 结合实操的形式讲解转换数据类型。 （1）转换数据类型的方式 astype()方法可以将数组中元素的数据类型转换其他的数据类型。 （2）转换数据类型的示例 ● 整数类型转换为浮点数类型 ● 浮点数类型转换为整数类型 ● 字符串类型转换为数值类型 （3）通过代码演示上述几种形式的转换。 **知识点 2-数组的索引方式** 教师通过 PPT 讲解数组的索引方式。 （1）一维数组的索引方式 一维数组与 Python 中的序列类型的结构类似，它们的索引方式相同。 （2）二维数组的索引方式 在二维数组中，每个元素对应行索引和列索引，其中行索引和列索引可以是正向索引或反向索引。 **知识点 3-整数索引和切片** 教师通过 PPT 结合实操的形式讲解整数索引和切片。 （1）获取二维数组的一行元素 格式为“数组[行索引]”。 （2）获取二维数组的单个元素 格式为“数组[行索引，列索引]”。 （3）获取二维数组的多行元素 格式为“数组[行索引的切片]”。

（4）获取二维数组的部分元素

- 使用“数组[行索引的切片，列索引的切片]”
- 混合使用切片与整数索引

（5）通过代码演示如何使用整数索引和切片获取二维数组的元素

知识点 4-花式索引

教师通过 PPT 结合实操的形式讲解花式索引。

（1）什么是花式索引

花式索引是指用整数数组或整数列表作为索引。

（2）花式索引的基本用法

- 若花式索引操作的目标对象是一维数组，则会把花式索引的每个整数作为索引，通过索引获取相应位置的元素。
- 若花式索引操作的目标对象是二维数组，则通过花式索引获取的结果是一行或多行元素。

（3）通过代码演示如何使用花式索引操作一维数组和二维数组

知识点 5-布尔索引

教师通过 PPT 结合实操的形式讲解布尔索引。

（1）什么是布尔索引

布尔索引指的是将一个布尔数组或布尔列表作为数组索引。

（2）布尔索引的基本用法

当使用布尔索引访问一维数组时，会将一维数组中与布尔数组或布尔列表位置相同的元素进行匹配，并返回布尔数组或布尔列表中与 True 位置对应的元素。

（3）通过代码演示如何使用布尔索引操作二维数组

（4）通过代码演示如何混合使用布尔索引和切片操作二维数组

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

五、布置作业

第三课时

（形状相同的数组间的算术运算、形状不同的数组间的算术运算、数组与标量的算术运算、通用函数、数组的重塑）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了数组的索引和切片操作，本节课将学习数组的算术运算、通用函数以及数组的重塑操作。

三、新课讲解

知识点 1-形状相同的数组间的算术运算

教师通过 PPT 结合实操的形式讲解形状相同的数组间的算术运算。

（1）运算规则

形状相同的数组之间进行任何算术运算时，会将位置相同的元素进行算术运

	算，所得的运算结果组成一个新的数组。 （2）通过代码演示形状相同的数组进行加减乘除运算。 **知识点 2-形状不同的数组间的算术运算** 教师通过 PPT 讲解形状不同的数组间的算术运算。 （1）运算规则 形状不同的数组在执行算术计算时可能会触发广播机制，该机制会对参与运算的数组进行扩展，使扩展后的数组具有相同的形状，这样就可以对数组进行算术运算了。 （2）形状兼容的判定规则 以两个数组为例，这两个数组的形状右对齐，之后沿着从右向左的顺序逐个比较同一纬度是否满足以下任意一种情况： ①维度相等； ②有一方维度为1。 如果数组的形状的每个维度都满足上述任意一种情况，说明两个数组的形状兼容。 （3）形状兼容的示例 （4）广播机制扩展数组的过程 （5）通过代码演示形状不同的数组进行加法运算。 **知识点 3-数组与标量的算术运算** 教师通过 PPT 结合实操的形式讲解数组与标量的算术运算。 （1）运算规则 ● 数组与标量执行算术运算时会将标量应用到各元素。 ● 数组与标量执行算术运算后会产生一个新数组，每个元素的值分别是原数组中每个元素与标量算术运算后得到的结果。 （2）通过代码演示二维数组与标量的算术运算 **知识点 4-通用函数** 教师通过 PPT 结合实操的形式讲解通用函数。 （1）通用函数的分类 ● 一元通用函数 ● 二元通用函数 （2）常见的一元通用函数 （3）常见的二元通用函数 （4）通过代码演示一元通用函数的用法 （5）通过代码演示二元通用函数的用法 **知识点 5-数组的重塑** 教师通过 PPT 结合实操的形式讲解数组的重塑。 （1）什么是数组的重塑 数组的重塑是指重新将数组的形状变成指定的形状，重塑前后元素的总数量不变。 （2）通过代码演示使用 reshape()方法将一维数组重塑为二维数组 （3）通过代码演示使用 reshape()方法将二维数组重塑为二维数组 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第四课时

（数组的转置、条件逻辑、统计运算、数组元素排序、检索数组元素是否满足条件、查找数组的唯一元素、判断元素是否在其他数组中）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了数组的算术运算、通用函数以及数组的重塑操作，本节课将继续学习数组的一些操作，包括数组的转置、条件逻辑、统计运算、数组元素排序、检索数组元素是否满足条件、查找数组的唯一元素、判断元素是否在其他数组中。

三、新课讲解

知识点 1-数组的转置

教师通过 PPT 结合实操的形式讲解数组的转置。

（1）什么是数组的转置

数组的转置指的是将数组中的每个元素按照一定的规则进行位置变换。

（2）数组的转置方式

- T 属性
- transpose()方法

（3）通过代码演示如何使用 T 属性实现数组转置操作

（4）通过代码演示如何使用 transpose()方法实现数组转置操作

知识点 2-条件逻辑

教师通过 PPT 结合实操的形式讲解条件逻辑。

（1）条件逻辑函数 where()的功能

（2）通过代码演示如何使用 where()函数实现条件逻辑操作

知识点 3-统计运算

教师通过 PPT 结合实操的形式讲解统计运算。

（1）统计运算的方法

（2）通过代码演示部分统计运算方法的用法

知识点 4-数组元素排序

教师通过 PPT 结合实操的形式讲解数组元素排序。

（1）排序方法 sort()

（2）通过代码演示如何使用 sort()方法对每行元素排序

（3）通过代码演示如何使用 sort()方法对每列元素排序

知识点 5-检索数组元素是否满足条件

教师通过 PPT 结合实操的形式讲解检索数组元素是否满足条件。

（1）检索元素的函数

- all()函数：判断整个数组中的元素的值是否全部满足条件。
- any()函数：判断整个数组中的元素至少有一个满足条件。

（2）通过代码演示使用 all()和 any()函数检索数组的元素

知识点 6-查找数组的唯一元素

教师通过 PPT 结合实操的形式讲解查找数组的唯一元素。

	（1）查找唯一元素的函数 unique() unique()函数用于找出数组中的唯一值，并返回一个升序排列的数组。 （2）通过代码演示如何使用布 unique()函数查找数组的唯一元素 **知识点 7-判断元素是否在其他数组中** 教师通过 PPT 结合实操的形式讲解判断元素是否在其他数组中。 （1）in1d()函数的功能 （2）in1d()函数的语法格式 （3）通过代码演示使用 in1d()函数判断一个数组元素是否在其他数组中 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第五课时** **（线性代数模块、随机数模块、案例：计算股票收益率和波动率）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了数组的一些操作，本节课将学习两个 NumPy 模块，分别是线性代数模块和随机数模块。 三、新课讲解 **知识点 1-线性代数模块** 教师通过 PPT 结合实操的形式讲解线性代数模块。 （1）线性代数模块 linalg （2）通过代码演示如何使用 dot()方法实现矩阵相乘操作。 （3）矩阵相乘图解 （4）linalg 模块的其他常见函数 **知识点 2-随机数模块** 教师通过 PPT 结合实操的形式讲解随机数模块。 （1）通过代码演示生成包含随机数的二维数组 （2）通过代码演示生成包含随机数的三维数组 （3）random 模块的常见函数 （4）seed()函数的语法格式 （5）通过代码演示如何使用 seed()函数生成相同的一组随机数 **知识点 3-案例：计算股票收益率和波动率** 教师通过 PPT 结合实操的形式讲解案例。 （1）通过 PPT 介绍收益率和波动率 （2）通过 PPT 介绍案例的要求 （3）通过代码演示案例的实现步骤 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 五、布置作业

第六课时（上机练习）

上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。

上机一：（考察知识点为创建数组）

形式：单独完成

题目：

请按照要求完成操作，具体要求如下：

（1）根据列表创建一维数组、二维数组和三维数组。

（2）创建一个 3 行 3 列的数组，元素都是 0，指定类型是 int32。

（2）创建一个 3 行 3 列的数组，元素都是 1。

（4）创建一个等差数组，起始值是 1，终止值是 20，步长为 5。

（5）创建一个等差数组，起始值是 1，终止值是 20，数量为 5。

上机二：（考察知识点为整数索引和切片、花式索引、布尔索引）

形式：单独完成

题目：

请按照要求完成操作，具体要求如下：

（1）根据列表创建一维数组、二维数组。

（2）使用整数索引和切片访问数组。

（3）使用花式索引访问数组。

（4）使用布尔索引访问数组。

上机三：（考察知识点为通用函数）

形式：单独完成

题目：

请按照要求完成操作，具体要求如下：

（1）练习 2.6 节一元通用函数的示例代码。

（2）练习 2.6 节二元通用函数的示例代码。

上机四：（考察知识点为条件逻辑、统计运算、数组元素排序、检索数组元素是否满足条件、查找数组的唯一元素、判断元素是否在其他数组中、随机数模块）

形式：单独完成

题目：

请按照要求完成操作，具体要求如下：

（1）练习 2.8 节全部的示例代码。

（2）练习 2.10 节全部的示例代码。

上机五：（考察知识点为案例：计算股票收益率和波动率）

形式：单独完成

题目：

请按照 2.11 节案例的要求，编写代码，计算一组股票书籍的简单收益率、

	对数收益率、年波动率和月波动率。
教学后记	后续可增加 “用 NumPy 分析国家 GDP 增长数据” 的拓展任务，将数学运算与家国情怀结合；针对广播机制，制作 “维度兼容判断流程图”，降低理解难度。

课题名称	第 3 章 数据分析库 pandas 基础	计划课时	6 课时
教学引入	pandas 是一个以 NumPy 为基础，专门为数据分析而设计的库，该库中不仅提供了一些标准的数据模型，而且提供了高效操作大量数据的数据结构，被广泛地应用到诸如金融、统计等一些领域中。pandas 库是本书的重点内容，本章只介绍一些 pandas 的基础知识，更多知识会在后面的章节进行介绍。		
教学目标	● 使学生掌握 Series 的结构和创建方式，能够通过 Series 类的构造方法创建 Series 类的对象 ● 使学生掌握 DataFrame 的结构和创建方式，能够通过 DataFrame 类的构造方法创建 DataFrame 类的对象 ● 使学生熟悉索引对象的类型和特点，能够归纳索引对象的类型和特点 ● 使学生掌握重置索引的方式，能够通过 reindex()方法重置 Series 或 DataFrame 类对象的索引 ● 使学生掌握索引与切片的基本用法，能够通过索引或切片获取 Series 或 DataFrame 的数据 ● 使学生掌握 loc 和 iloc 属性的基本用法，能够使用 loc 和 iloc 属性获取 Series 或 DataFrame 的数据 ● 使学生掌握读写数据的方式，能够熟练地读取网页表格的数据，以及读写 CSV 文件、TXT 文件、Excel 文件、数据库 ● 使学生掌握数据的排序方式，能够使用索引或值对 Series 或 DataFrame 进行排序 ● 使学生掌握 pandas 的算术运算，能够通过算术运算符或算术方法对 DataFrame 中的数据进行算术运算 ● 使学生掌握 pandas 的统计计算，能够通过统计方法对 DataFrame 中的数据进行统计计算 ● 使学生掌握 pandas 的统计描述，能够通过 describe()方法描述 DataFrame 类的对象的统计指标 ● 使学生掌握分层索引的相关操作，能够熟练地创建有分层索引的 DataFrame，并使用分层索引获取数据		
教学重点	● Series ● DataFrame ● 通过索引和切片获取数据 ● 通过 loc 和 iloc 属性获取数据 ● 读写 CSV 和 TXT 文件的数据 ● 读写 Excel 文件的数据		
教学难点	● 读写数据库 ● 使用分层索引获取数据 ● 交换索引层级的顺序 ● 分层索引排序		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	• **数据诚信与职业道德** 讲解 “读写文件”“数据筛选” 时，强调 “原始数据不可篡改”“缺失值需		

	标注原因”，以 “学术造假案例（如篡改实验数据）” 警示学生，培养 “做数据分析先做‘诚信人’” 的职业道德。 • **服务意识与社会责任** 以 “陕西高考分数线统计分析” 案例为基础，拓展 “用 pandas 分析校园奖学金数据，优化评选规则”，引导学生思考 “数据技术如何服务同学、公平分配资源”，树立 “技术为民” 的服务意识。
教学过程	**第一课时** **（Series、DataFrame、索引对象、重置索引、通过索引和切片获取数据）** 一、创设情景，导入新课 教师通过给学生提问问题，例如问题是：大家觉得 NumPy 库相比列表有什么优势，并根据学生的问题进行总结，引出基于 NumPy 库的基础上设计了 pandas 库，这个库不仅提供了操作大量数据的数据结构，还提供了数据处理和可视化的功能，从而实现导入新课的目的。 二、新课讲解 **知识点 1-Series** 教师通过 PPT 结合实操的形式讲解 Series。 （1）Series 的特点 ● 由数据和索引两部分组成。 ● 数据可以是任意类型的。 （2）Series 的结构图 （3）Series 类构造方法的语法格式 （4）通过代码演示如何根据列表创建 Series 类的对象 （5）通过代码演示如何创建 Series 类对象的同时指定标签索引 （6）通过代码演示如何根据字典创建 Series 类的对象 （7）Series 类的属性 ● index：获取 Series 类对象的索引。 ● values：获取 Series 类对象的数据。 （8）通过代码演示如何获取 Series 类对象的索引和数据 **知识点 2-DataFrame** 教师通过 PPT 结合实操的形式讲解 DataFrame。 （1）DataFrame 的特点 ● 由数据和索引两部分组成，既有行索引又有列索引。 ● 每列数据可以是不同的数据类型。 （2）DataFrame 的结构图 （3）DataFrame 类构造方法的语法格式 （4）通过代码演示如何根据二维数组创建 DataFrame 类的对象 （5）通过代码演示如何创建 DataFrame 类对象的同时指定列索引 （6）通过代码演示如何使用列标签索引获取一列数据 （7）通过代码演示如何使用 info()方法查看摘要信息。 **知识点 3-索引对象** 教师通过 PPT 结合实操的形式讲解索引对象。 （1）常见的 Index 子类 ● RangeIndex：位置索引。

- Int64Index：整数索引。
- Float64Index：浮点数索引。
- DatetimeIndex：时间戳索引。
- PeriodIndex：时间间隔索引。
- MultiIndex：分层索引。

（2）索引对象的特性

- 不可变性
- 可重复性

（3）通过代码验证索引对象的两个特性

知识点 4-重置索引

教师通过 PPT 结合实操的形式讲解重置索引。

（1）什么是重置索引

重置索引是指重新为对象设定索引，以构建一个符合新索引的对象。

（2）reindex()方法的作用

（3）reindex()方法的语法格式

（4）通过代码演示如何使用 reindex()方法重置行索引

（5）通过代码演示如何使用同一个值填充空缺位置

（6）通过代码演示如何使用不同的值填充空缺位置

知识点 5-通过索引和切片获取数据

教师通过 PPT 结合实操的形式讲解通过索引和切片获取数据。

（1）索引的用法

（2）通过代码演示如何使用索引获取 Series 类对象的单个数据

（4）通过代码演示如何使用索引获取 Series 类对象的多个数据

（5）布尔索引的用法

将 Series 类的对象中每个数据进行逻辑运算，只要运算结果为 True，就返回 Series 类对象中位置为 True 对应的数据

（6）通过代码演示使用布尔索引获取 Series 类对象中符合条件的数据

（7）切片的用法

- 如果切片使用的索引是位置索引，则切片结果包含起始位置但不包含结束位置对应的数据。
- 如果切片使用的索引是标签索引，则切片结果既包含起始位置又包含结束位置对应的数据。

（8）通过代码演示使用切片获取 Series 类对象的部分数据

三、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第二课时

（通过 loc 和 iloc 属性获取数据、读写 CSV 和 TXT 文件的数据、读写 Excel 文件的数据、读取网页表格的数据）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了 pandas 的两种数据结构，以及索引和切片的基本用法，本节课将学习 loc 和 iloc 属性，读写 CSV、TXT、Excel 文件的数据，以及读取网页表格的数据。

三、新课讲解

知识点 1-通过 loc 和 iloc 属性获取数据

教师通过 PPT 结合实操的形式讲解通过 loc 和 iloc 属性获取数据。

（1）loc 和 iloc 属性

- loc 是基于标签索引的索引器
- iloc 是基于位置索引的索引器

（2）loc 属性的使用格式

（3）通过代码演示如何使用 loc 属性获取 Series 对象的数据

（4）通过代码演示如何使用 loc 属性获取 DataFrame 对象的数据

（5）iloc 属性的使用格式

（6）通过代码演示如何使用 iloc 属性获取 DataFrame 对象的数据

知识点 2-读写 CSV 和 TXT 文件的数据

教师通过 PPT 讲解读写 CSV 和 TXT 文件的数据。

（1）CSV 和 TXT 文件的特点

- 只能保存文本的内容，不能保存文本的样式。
- CSV 文件通常以逗号或制表符为分隔符。

（2）to_csv()方法的作用

- 向指定路径下的 CSV 或 TXT 文件中写入部分或全部数据。
- 如果指定路径下文件不存在，则会新建一个文件。
- 如果指定路径下文件已经存在，则会覆盖文件中的内容。

（3）to_csv()方法的语法格式

（4）通过代码演示如何使用 to_csv()方法向文件写入数据

（5）read_csv()函数的作用

read_csv()函数会从指定路径下的 CSV 或 TXT 文件中读取数据，读取成功后会根据数据形式转换成一个 Series 或 DataFrame 类的对象。

（6）read_csv()函数的语法格式

（7）通过代码演示如何使用 read_csv()函数从文件中读取数据

（8）read_table()函数的作用

- 用于从 TXT 文件中读取数据。
- TXT 文件使用的分隔符是制表符。

（9）通过代码演示如何使用 read_table()函数从文件中读取数据

知识点 3-读写 Excel 文件的数据

教师通过 PPT 结合实操的形式讲解读写 Excel 文件的数据。

（1）Excel 文件的特点

- 可以添加若干个工作表。
- 每个工作表都是以表格的形式显示数据。

（2）to_excel()方法的作用

- 用于将 Series 或 DataFrame 类的对象写入到 Excel 文件中。
- 如果 Excel 文件不存在，则会新建一个文件。
- 如果 Excel 文件存在，则会覆盖原文件中的内容。

（3）to_excel()方法的语法格式

（4）通过代码演示如何使用 to_excel()方法向文件写入数据

（5）read_excel()函数的作用

read_excel()函数用于读取 Excel 文件中的数据，并根据数据的形式转换成 Series 或 DataFrame 类的对象。

（6）read_excel()函数的语法格式

（7）通过代码演示如何使用 read_excel()函数从 Excel 文件中读取数据

知识点 4-读取网页表格的数据

教师通过 PPT 结合实操的形式讲解读取网页表格的数据。

（1）read_html()函数的语法格式

（2）通过代码演示如何使用 read_html()函数读取网页上表格的数据

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第三课时

（读写数据库、按索引排序、按值排序、算术运算与数据对齐、统计计算、统计描述）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了 loc 和 iloc 属性，读写 CSV、TXT、Excel 文件的数据，以及读取网页表格的数据，本节课将学习读写数据库、排序操作、算术运算、统计计算、统计描述。

三、新课讲解

知识点 1-读写数据库

教师通过 PPT 结合实操的形式讲解读写数据库。

（1）读写数据库的函数或方法

（2）to_sql()方法的语法格式

（3）通过代码演示如何使用 to_sql()方法向数据库写入数据

（4）read_sql()函数的语法格式

（5）通过代码演示如何使用 read_sql()函数读取数据库的数据

知识点 2-按索引排序

教师通过 PPT 结合实操的形式讲解按索引排序。

（1）sort_index()方法的语法格式

（2）通过代码演示使用 sort_index()方法按照索引排序

知识点 3-按值排序

教师通过 PPT 结合实操的形式讲解按值排序。

（1）sort_values()方法的语法格式

（2）通过代码演示使用 sort_values()方法按照数据排序

知识点 4-算术运算与数据对齐

教师通过 PPT 结合实操的形式讲解算术运算与数据对齐。

（1）算术运算的规则

Series 类或 DataFrame 类的对象进行算术运算时，会先将对象中索引相同的数据按位置对齐，对齐后再进行相应的运算，没有对齐的位置会用 NaN 补齐。

（2）通过代码演示两个 Series 类的对象的加法运算

（3）处理 NAN 值的方式

调用 add()方法时给 fill_value 参数传值

（4）通过代码演示在执行加法运算时如何处理 NAN 值

知识点 5-统计计算

教师通过 PPT 结合实操的形式讲解统计计算。

（1）常见的统计计算方法

（2）通过代码演示部分统计计算方法的基本使用

知识点 6-统计描述

教师通过 PPT 结合实操的形式讲解统计描述。

（1）describe()方法的语法格式

（2）通过代码演示 describe()方法的基本使用

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

五、布置作业

第四课时

（创建分层索引、创建有分层索引的对象、使用分层索引获取数据、交换索引层级的顺序、分层索引排序、案例：陕西高考分数线统计分析）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了读写数据库、排序操作、算术运算、统计计算、统计描述，本节课将继续学习分层索引的操作，以及围绕所学的知识完成一个案例。

三、新课讲解

知识点 1-数组的转置创建分层索引

教师通过 PPT 结合实操的形式讲解创建分层索引。

（1）什么是分层索引

分层索引可以理解为单层索引的延伸，即在一个轴方向上具有两层或两层以上的索引。

（2）分层索引的示意图

（3）创建分层索引的方法

（4）通过代码演示如何使用 from_tuples()方法创建分层索引

（5）通过代码演示如何使用 from_arrays()方法创建分层索引

（6）通过代码演示如何使用 from_product()方法创建分层索引

知识点 2-创建有分层索引的对象

教师通过 PPT 结合实操的形式讲解创建有分层索引的对象。

	（1）创建有分层索引对象的基本方式 在 Series 类和 DataFrame 类构造方法的 index 参数中传入一个嵌套列表。 （2）通过代码演示使用上述方式创建有分层索引的对象 （3）创建有分层索引对象的其他方式 在 Series 类和 DataFrame 类构造方法的 index 参数中传入一个 MultiIndex 类的对象。 （4）通过代码演示使用上述方式创建有分层索引的对象 **知识点 3-使用分层索引获取数据** 教师通过 PPT 结合实操的形式讲解使用分层索引获取数据。 （1）分层索引的用法 ● 对象[外层索引]：访问外层索引嵌套的索引及其数据。 ● 对象[外层索引，内层索引]：访问索引对应的数据。 （2）通过代码演示如何使用分层索引获取数据 **知识点 4-交换索引层级的顺序** 教师通过 PPT 结合实操的形式讲解交换索引层级的顺序。 （1）什么是交换分层顺序 交换分层顺序是指交换外层索引和内层索引的位置。 （2）通过代码演示如何使用 swaplevel()方法交换分层顺序 **知识点 5-分层索引排序** 教师通过 PPT 结合实操的形式讲解分层索引排序。 （1）sort_index()方法的作用 使用 sort_index()方法进行排序时，会优先按外层索引排序，然后再按照内层索引排序。 （2）通过代码演示使用 sort_index()方法对有分层索引的对象排序 **知识点 6-案例：陕西高考分数线统计分析** 教师通过 PPT 结合实操的形式讲解案例。 （1）通过 PPT 介绍案例的需求 （2）通过 PPT 介绍准备的数据 （3）通过代码演示案例的实现步骤 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第五、六课时（上机练习）** 上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。 **上机一：（考察知识点为 Series、DataFrame、重置索引、通过索引和切片获取数据、通过 loc 和 iloc 属性获取数据）** **形式：单独完成** **题目：** 请按照要求完成操作，具体要求如下： （1）练习 3.1 节全部的示例代码。

	（2）练习 3.2.2 到 3.2.4 小节全部的示例代码。 **上机二：（考察知识点为读写 CSV 和 TXT 文件的数据、读写 Excel 文件的数据、读取网页表格的数据、读写数据库）** **形式：单独完成** **题目：** 练习 3.3 节全部的示例代码 **上机三：（考察知识点为按索引排序、按值排序、算术运算与数据对齐、统计计算、统计描述）** **形式：单独完成** **题目：** 请按照要求完成操作，具体要求如下： （1）练习 3.4 节全部的示例代码。 （2）练习 3.5 节全部的示例代码。 （3）练习 3.6 节全部的示例代码。 **上机四：（考察知识点为创建分层索引、创建有分层索引的对象、使用分层索引获取数据、交换索引层级的顺序、分层索引排序）** **形式：单独完成** **题目：** 练习 3.7 节全部的示例代码。 **上机五：（考察知识点为案例：陕西高考分数线统计分析）** **形式：单独完成** **题目：** 请按照 3.8 节案例的要求，编写代码，从 scores.xlsx 文件中读取数据，并按照设定的目标操作数据。
教学后记	“分层索引” 难度较大，需通过 “类比文件夹层级” 强化理解。

课题名称	第 4 章 数据预处理	计划课时	8 课时
教学引入	在数据分析工作前期收集的数据或多或少会存在着一些瑕疵或不足，比如数据缺失、重复、格式不统一等，因此我们在分析数据之前需要先对数据进行预处理，包括数据清洗、数据合并、数据重塑和数据转换。为了处理这些问题数据，pandas 提供了很多用于数据预处理的函数与方法。接下来，本章将针对 pandas 中数据预处理的内容进行详细地讲解。		
教学目标	● 使学生掌握缺失值的检测方式，能够通过 isnull()和 notnull()函数检测数据中是否存在缺失值 ● 使学生掌握缺失值的处理方式，能够通过 dropna()或 fillna()方法删除缺失值或填充缺失值 ● 使学生掌握重复值的检测方式，能够通过 duplicated()方法检测数据中是否存在重复值 ● 使学生掌握重复值的处理方式，能够通过 drop_duplicates()方法删除重复值 ● 使学生熟悉异常值的检测方式，能够通过 3σ 原则和箱形图检测数据中是否存在异常值 ● 使学生掌握异常值的处理方式，能够通过 replace()方法替换数据中的异常值 ● 使学生熟悉数据类型的转换方式，能够通过 astype()方法或 to_numberic()函数转换数据类型 ● 使学生掌握数据合并的相关操作，能够根据需求选择适合的方案实现数据合并的操作 ● 使学生掌握数据重塑的相关操作，能够根据需求选择适合的方案实现数据重塑的相关操作 ● 使学生掌握数据转换的相关操作，能够根据需求选择适合的方案实现数据转换的相关操作		
教学重点	● 缺失值的检测 ● 缺失值的处理 ● 重复值的检测 ● 重复值的处理 ● 堆叠合并 ● 主键合并		
教学难点	● 异常值的检测 ● 主键合并 ● 面元划分 ● 哑变量处理		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	• **求真务实与客观态度** 讲解 “缺失值处理” 时，对比 “删除缺失值” 与 “填充缺失值” 的适用场景，强调 “根据数据实际情况选择方法，不主观决定”，培养学生 “尊重客观数据，不回避问题” 的求真态度。 • **批判性思维与风险意识**		

	在 “异常值检测” 环节，以 “校园消费数据中的‘高额消费异常值’” 为例，引导学生思考 “是真实消费还是数据错误”，培养 “不盲目相信数据，主动排查异常” 的批判性思维，渗透 “风险识别” 的意识。
教学过程	**第一课时** **（缺失值的检测、缺失值的处理、重复值的检测、重复值的处理）** 一、创设情景，导入新课 教师提前准备两份数据，一份不包含缺失值、重复值、异常值的数据，一份包含缺失值、重复值、异常值的数据，给学生提问问题，例如问题是：如果要计算平均值，大家觉得哪份数据得到的结果相对是比较准确的，并根据学生的问题进行总结，引出数据清洗的好处，也就是提高数据的质量，从而实现导入新课的目的。 二、新课讲解 **知识点 1-缺失值的检测** 教师通过 PPT 结合实操的形式讲解缺失值的检测。 （1）什么是缺失值 ● 缺失值是指数据集中某个或某些属性的值是不完整的。 ● 缺失值一般使用 None 或 np.nan 表示，统一标记为 NaN。 （2）检测缺失值的方式 ● isnull()：在检测到缺失值的位置标记 True，其他位置标记为 False。 ● notnull()：在检测到缺失值的位置标记 False，其他位置标记为 True。 （3）通过代码演示如何使用 isnull()函数检测缺失值 （4）通过代码演示如何使用 notnull()函数检测缺失值 （5）通过代码演示如何自定义函数来了解缺失值的占比情况 **知识点 2-缺失值的处理** 教师通过 PPT 结合实操的形式讲解缺失值的处理。 （1）处理缺失值的方式 ● 删除缺失值 ● 填充缺失值 （2）dropna()方法的语法格式 （3）通过代码演示如何使用 dropna()方法删除缺失值 （4）fillna()方法的语法格式 （5）通过代码演示如何使用 fillna()方法填充缺失值 （6）填充不同的值 在调用 fillna()方法填充缺失值时传入一个字典给 value 参数，其中字典的键为列索引，字典的值为待替换的值。 （7）通过代码演示如何使用 fillna()方法填充不同的值。 （8）填充缺失值相邻的前面的有效值 在调用 fillna()方法时给 method 参数传入值 ffill，指定填充方式为前向填充。 （9）通过代码演示如何使用 fillna()方法实现前向填充的效果。 **知识点 3-重复值的检测** 教师通过 PPT 结合实操的形式讲解重复值的检测。 （1）什么是重复值 重复值是指数据集中某个或某些记录是完全相同的。

（2）检测重复值的方式

duplicated()方法默认会对所有数据进行检测，检测的标准为：只要一行数据与其他行数据的所有值是完全相同的，就会将这一行数据判定为重复值，并标记为 True，非重复值标记为 False。

（3）duplicated()方法的语法格式

（4）通过代码演示如何使用 duplicated()方法检测缺失值

知识点 4-重复值的处理

教师通过 PPT 结合实操的形式讲解重复值的处理。

（1）重复值的处理方式

重复值会影响分析结果的准确性，一般情况下需要进行删除。

（2）drop_duplicates()方法的语法格式

（3）通过代码演示如何使用 drop_duplicates()方法删除重复值

三、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第二课时

（异常值的检测、异常值的处理、转换数据类型、堆叠合并）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了缺失值的检测与处理、重复值的检测与处理，本节课将继续学习异常值的检测与处理、转换数据类型和堆叠合并。

三、新课讲解

知识点 1-异常值的检测

教师通过 PPT 结合实操的形式讲解异常值的检测。

（1）什么是异常值

异常值是指数据集中的个别值明显偏离它所属数据集的其余值，这些数值是不合理的或错误的。

（2）异常值的检测方式

- 3σ 原则：适用于符合或近似正态分布的数据集。
- 箱形图：可以检测任意的数据集。

（3）基于 3σ 原则检测的原理

凡是误差超过（μ-3σ, μ+3σ)区间的数值就认为是异常值。

（4）基于 3σ 原则检测的函数

（5）通过代码演示如何基于 3σ 原则检测异常值

（6）基于箱形图检测的原理

- 箱形图可以展示异常值。
- 异常值的范围一般是小于 Q1 – 1.5IQR 或大于 Q3 + 1.5IQR。

（7）通过代码演示如何绘制绘制箱形图

（8）通过代码演示如何确定异常值的位置

知识点 2-异常值的处理

	教师通过 PPT 结合实操的形式讲解异常值的处理。 （1）异常值的处理方式 ● 异常值被检测出来之后，需要进一步确认是否为真正的异常值。 ● 通常情况下会使用指定的值或根据一些算法计算的值替换异常值。 （2）replace()方法的语法格式 （3）通过代码演示如何使用 replace()方法替换一个异常值 （4）通过代码演示如何使用 replace()方法替换多个异常值 **知识点 3-转换数据类型** 教师通过 PPT 结合实操的形式讲解转换数据类型。 （1）转换数据类型的使用场景 （2）转换数据类型的方式 ● 通过 astype()方法转换数据的类型。 ● 通过 to_numeric()函数转换数据类型。 （3）astype()方法的语法格式 （4）通过代码演示如何使用 astype()方法转换数据的类型 （5）to_numeric()方法的作用 to_numeric()函数用于将字符串、混合类型等一些复杂类型的数据转换为数值类型的数据，并能够按照不同的参数配置灵活地处理这些复杂类型的数据。 （6）to_numeric()方法的语法格式 （7）通过代码演示如何使用 to_numeric()方法转换数据的类型 **知识点 4-堆叠合并** 教师通过 PPT 结合实操的形式讲解堆叠合并。 （1）什么是堆叠合并 堆叠合并指的是沿着某个轴的方向将两个或两个以上的对象按照一定的逻辑关系进行合并。 （2）concat()函数的语法格式 （3）横向堆叠与外连接 （4）通过代码演示如何实现横向堆叠与外连接的效果 （5）纵向堆叠与内连接 （6）通过代码演示如何实现纵向堆叠与内连接的效果 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 五、布置作业 教师通过学习通布置本节课作业以及下节课的预习作业。 **第三课时** **（主键合并、根据索引合并、合并重叠数据、重塑分层索引）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了异常值的检测、异常值的处理、转换数据类型和堆叠

	合并，本节课将学习其他几种合并数据的方式，包括主键合并、根据索引合并、合并重叠数据，以及重塑分层索引。 三、新课讲解 **知识点 1-主键合并** 教师通过 PPT 结合实操的形式讲解主键合并。 （1）什么是主键合并 主键合并类似于关系型数据库的主键查询操作，它指的是根据一个或多个键将两个对象进行合并，大多数情况下会将这两个对象中共有的列作为合并的键。 （2）merge()函数的语法格式 （3）通过代码演示一个键合并的效果 （4）通过代码演示两个键合并的效果 （5）通过代码演示全外连接合并的效果 （6）通过代码演示左连接合并的效果 **知识点 2-根据索引合并** 教师通过 PPT 结合实操的形式讲解根据索引合并。 （1）什么是根据索引合并 根据索引合并指的是根据行索引或列索引将多个对象合并成一个对象。 （2）join()方法的语法格式 （3）通过代码演示如何使用 join()方法实现没有重叠列合并的效果 （4）通过代码演示如何使用 join()方法实现有重叠列合并的效果 **知识点 3-合并重叠数据** 教师通过 PPT 结合实操的形式讲解合并重叠数据。 （1）combine_first()方法的语法格式 （2）通过代码演示如何使用 combine_first()方法实现合并重叠数据的效果 **知识点 4-重塑分层索引** 教师通过 PPT 结合实操的形式讲解重塑分层索引。 （1）重塑分层索引的方法 ● stack()方法用于将数据的列“旋转”为行。 ● unstack()方法用于将数据的行“旋转”为列。 （2）通过代码演示如何使用 stack()方法实现重塑索引的操作 （3）通过代码演示如何使用 unstack()方法实现重塑索引的操作 （4）通过代码演示如何使用 stack()方法实现重塑分层索引的效果 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第四、五、六课时** **（轴向旋转、面元划分、哑变量处理、案例：预处理二手房数据）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了主键合并、根据索引合并、合并重叠数据、重塑分层

索引，本节课将继续学习轴向旋转、面元划分、哑变量处理，以及围绕所学的知识完成一个案例。

三、新课讲解

知识点 1-轴向旋转

教师通过 PPT 结合实操的形式讲解轴向旋转。

（1）轴向旋转的举例

（2）pivot()方法的语法格式

（3）通过代码演示如何使用 pivot()方法实现轴向旋转的效果

知识点 2-面元划分

教师通过 PPT 结合实操的形式讲解面元划分。

（1）什么是面元划分

面元划分是指连续数据被离散化处理，按一定的映射关系划分为相应的面元，这里的面元可以理解为区间。

（2）面元划分的举例

（3）cut()函数的语法格式

（4）通过代码演示如何使用 cut()函数实现面元划分操作

知识点 3-哑变量处理

教师通过 PPT 结合实操的形式讲解哑变量处理。

（1）什么是哑变量

哑变量又称虚拟变量、名义变量等，它是人为虚设的变量，用来反映某个变量的不同类别，常用的取值为 0 和 1。

（2）get_dummies()函数的语法格式

（3）通过代码演示如何使用 get_dummies()函数实现哑变量处理的效果

知识点 4-案例：预处理二手房数据

教师通过 PPT 结合实操的形式讲解案例。

（1）通过 PPT 介绍案例的需求

（2）通过代码演示如何读取数据和合并数据

（3）通过代码演示案例的实现步骤

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

五、布置作业

第七、八课时（上机练习）

上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。

上机一：（考察知识点为缺失值的检测、缺失值的处理、重复值的检测、重复值的处理、异常值的检测、异常值的处理）

形式：单独完成

题目：

练习 4.1.1 到 4.1.6 小节的示例代码。

	上机二：（考察知识点为堆叠合并、主键合并、根据索引合并、合并重叠数据） **形式：单独完成** **题目：** 练习 4.2 节全部的示例代码 **上机三：（考察知识点为重塑分层索引、轴向旋转、面元划分、哑变量处理）** **形式：单独完成** **题目：** 练习 4.3 到 4.4 节的示例代码。 **上机四：（考察知识点为案例：预处理二手房数据）** **形式：单独完成** **题目：** 请按照 4.5 节案例的要求，编写代码，分别从 secondhandhouse_one.xlsx 和 secondhandhouse_two.xlsx 文中读取数据，合并数据，并按照设定的目标操作数据。
教学后记	异常值处理可引入更多校园真实场景，避免机械套用公式。

<table>
<tr><td>课题名称</td><td>第 5 章 数据聚合和分组运算</td><td>计划课时</td><td>4 课时</td></tr>
<tr><td>教学引入</td><td colspan="3">在进行数据分析工作时，我们可能会遇到这样的场景：现在要求从日志数据中找出每天访问次数最多的 IP，这时需要先把所有的日志数据按天拆分成每天的日志数据，再对每天的日志数据进行统计运算，最后把所有的统计结果放到一起，这样便完成了最初设定的要求，这个过程中用到的思想就是分组与聚合——数据重组后再合并。pandas 中提供了一些用于分组与聚合的方法，另外还提供一些其他的分组级运算，本章将针对这些内容进行详细地讲解。</td></tr>
<tr><td>教学目标</td><td colspan="3">● 使学生了解分组与聚合的原理，能够说出分组与聚合的原理
● 使学生掌握分组方法的使用，能够通过 groupby()方法按照不同的拆分标准对数据进行分组
● 使学生掌握分组信息的查看方式，能够通过多种方式查看分组的信息
● 使学生熟悉内置统计方法的使用，能够通过统计方法聚合数据
● 使学生掌握 agg()方法的使用，能够通过 agg()方法聚合数据
● 使学生掌握 transform()方法的使用，能够通过 transform()方法转换数据
● 使学生掌握 apply()方法的使用，能够通过 apply()方法聚合数据</td></tr>
<tr><td>教学重点</td><td colspan="3">● 通过 groupby()对数据进行分组
● 通过 agg()聚合数据
● 数据转换</td></tr>
<tr><td>教学难点</td><td colspan="3">● 通过 agg()聚合数据
● 数据转换</td></tr>
<tr><td>教学方式</td><td colspan="3">课堂教学以 PPT 讲授为主，并结合多媒体进行教学</td></tr>
<tr><td>思政元素</td><td colspan="3">• 分类思维与高效解决问题
讲解 groupby () 分组时，类比 “学校按‘学院’分组统计人数”，引导学生理解 “分类是高效处理复杂问题的核心思维”，培养 “化繁为简、有序处理” 的能力。
• 数据价值与社会服务
以 “篮球运动员信息分析” 案例为基础，拓展 “用分组运算分析校园各专业学生的‘图书馆打卡时长’”，引导学生思考 “如何用数据为学校‘优化图书馆开放时间’‘增设专业书籍’提供依据”，体现数据服务社会的价值。</td></tr>
<tr><td>教学过程</td><td colspan="3">第一课时
（分组与聚合的原理、通过 groupby()对数据进行分组、查看分组信息、通过统计方法聚合数据、通过 agg()聚合数据）
一、创设情景，导入新课
教师提前准备一份日志数据，这份数据是乱序的，通过给学生提问问题，例如问题是：我们要找出每天访问次数最多的 IP，需要怎么实现，并根据学生的问题进行总结，引出解决这个问题的过程中用到的思想就是分组与聚合，从而实现导入新课的目的。
二、新课讲解
知识点 1-分组与聚合的原理</td></tr>
</table>

	教师通过 PPT 讲解分组与聚合的原理。 （1）什么是分组与聚合 分组与聚合是数据分析工作中比较常见的操作，它主要根据一定的拆分标准将原数据拆分成若干个分组，然后对每个分组应用统计运算，并把运算后的结果合并到一起。 （2）分组与聚合的基本过程 ① 拆分 ② 应用 ③ 合并 **知识点 2-通过 groupby()对数据进行分组** 教师通过 PPT 结合实操的形式讲解通过 groupby()对数据进行分组。 （1）groupby()的语法格式 （2）通过代码演示如何按照列标签对 DataFrame 类的对象进行分组 （3）通过代码演示如何按照 Series 对 DataFrame 类的对象进行分组 （4）通过代码演示如何按照字典对 DataFrame 类的对象进行分组 （5）通过代码演示如何按照函数对 DataFrame 类的对象进行分组 **知识点 3-查看分组信息** 教师通过 PPT 结合实操的形式讲解查看分组信息。 （1）GroupBy 对象 无论是 SeriesGroupBy 对象和 DataFrameGroupBy 对象，它们其实都属于 GroupBy 对象。 （2）查看分组信息的方式 ● for 语句 ● groups 属性 ● get_group()方法 （3）通过代码演示如何使用 groups 属性查看分组的信息 （4）通过代码演示如何使用 get_group()方法查看分组的信息 **知识点 4-通过统计方法聚合数据** 教师通过实操的形式讲解通过统计方法聚合数据。 通过代码演示如何使用统计方法聚合数据。 **知识点 5-通过 agg()聚合数据** 教师通过 PPT 结合实操的形式讲解通过 agg()聚合数据。 （1）agg()方法的语法格式 （2）通过代码演示聚合数据时所有列应用一个函数 （3）通过代码演示聚合数据时所有列应用多个函数 （4）通过代码演示聚合数据时不同列应用不同函数 三、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第二课时** **（数据转换、数据应用、案例：篮球运动员信息分析）**

	一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们学习了分组与聚合的原理、分组操作、聚合操作，本节课将继续学习其他的分组级操作，包括数据转换、数据应用，以及围绕所学的知识完成一个案例。 三、新课讲解 **知识点 1-数据转换** 教师通过 PPT 结合实操的形式讲解数据转换。 （1）什么是数据转换 数据转换是 pandas 中强大的功能之一，它可以对分组执行一些汇总操作，且不改变分组之前的对象形状，使转换后对象的形状与分组前对象的形状保持一致。 （2）transform()方法的语法格式 （3）通过代码演示如何使用 transform()方法实现数据转换的功能 **知识点 2-数据应用** 教师通过 PPT 结合实操的形式讲解数据应用。 （1）apply()方法的语法格式 （2）通过代码演示如何使用 apply()方法替换多个异常值 **知识点 3-案例：篮球运动员信息分析** 教师通过 PPT 结合实操的形式讲解案例。 （1）通过 PPT 介绍案例的需求 （2）通过代码演示如何读取数据、合并数据，以及查看摘要信息 （3）通过代码演示案例的实现步骤 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 五、布置作业 **第三、四课时（上机练习）** 上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。 **上机一：（考察知识点为通过 groupby()对数据进行分组、查看分组信息）** **形式：单独完成** **题目：** 练习 5.2 节全部的示例代码。 **上机二：（考察知识点为通过统计方法聚合数据、通过 agg()聚合数据）** **形式：单独完成** **题目：** 练习 5.3 节全部的示例代码

	上机三：（考察知识点为数据转换、数据应用） **形式：单独完成** **题目：** 练习 5.4 节全部的示例代码。 **上机四：（考察知识点为案例：篮球运动员信息分析）** **形式：单独完成** **题目：** 请按照 5.5 节案例的要求，编写代码，分别从“运动员信息采集 01.csv”和“运动员信息采集 02.xlsx”文件中读取数据，合并数据，并按照设定的目标操作数据。
教学后记	下次教学可增加类似 “用分组运算分析国家各省份 GDP 增长数据” 的拓展任务，强化家国情怀；针对 transform ()，可设计 “分组后计算学生绩点排名”的简单案例，降低理解难度。

课题名称	第 6 章 数据可视化	计划课时	10 课时
教学引入	大多数情况下，我们获取的数据是以文字或数字的形式进行呈现的，这种密密麻麻的文字或数字不仅会降低数据信息的可读性，而且无法很好地展示数据之间的关系和规律。为了解决这些问题，数据可视化应运而生，它可以使数据变得更直观，更容易被人们理解与接受。Python 提供了许多优秀的用于实现数据可视化功能的库，比如 Matplotlib、Seaborn、Pyecharts 等，本章将围绕着这些库的基本使用进行详细地讲解。		
教学目标	● 使学生了解数据可视化，能够说出可视化的概念以及意义 ● 使学生熟悉常见的图表类型，能够说出图表的特点以及适用场景 ● 使学生熟悉图表的基本组成元素，能够说出每个组成元素的用途 ● 使学生掌握 Matplotlib 库的基本使用，能够使用 Matplotlib 库绘制常见的图表 ● 使学生掌握 Seaborn 库的基本使用，能够使用 Seaborn 库绘制常见的图表 ● 使学生掌握 Pyecharts 库的基本使用，能够使用 Pyecharts 库绘制常见的图表		
教学重点	● 使用 Matplotlib 绘制折线图 ● 使用 Matplotlib 绘制柱形图 ● 使用 Matplotlib 绘制直方图 ● 使用 Matplotlib 绘制散点图 ● 使用 Pyecharts 绘制柱形图 ● 使用 Pyecharts 绘制词云图		
教学难点	● 用分类数据绘图 ● 使用 Pyecharts 绘制柱形图 ● 使用 Pyecharts 绘制词云图		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	• **真实表达与社会责任** 强调 “数据可视化不可误导”，如 “不刻意拉伸 Y 轴制造数据差异”“图表需标注单位和来源”，以 “虚假数据图表误导公众” 案例警示学生，培养 “用图表真实传递信息” 的社会责任。 • **家国情怀与文化自信** 设计 “用 Pyecharts 绘制‘国家近 10 年科技投入增长图’”“用 Matplotlib 展示‘校园少数民族学生比例图’” 任务，引导学生通过图表感受国家发展成就、尊重文化多样性，增强家国情怀和文化自信。		
教学过程	**第一、二课时** **（什么是数据可视化、常见的图表类型、图表的辅助元素、使用 Matplotlib 绘制折线图、使用 Matplotlib 绘制柱形图）** 一、创设情景，导入新课 教师提前准备两份数据，一份表格形式的，一份图表形式的，引出数据可视化的好处，从而实现导入新课的目的。 二、新课讲解		

	知识点 1-什么是数据可视化 教师通过 PPT 讲解什么是数据可视化。 （1）数据可视化的概念 数据可视化是指将大型数据集中的数据以图形、图像的形式表示，并利用数据分析和开发工具发现其中未知信息的处理过程。 （2）数据可视化的过程 **知识点 2-常见的图表类型** 教师通过 PPT 讲解常见的图表类型。 （1）直方图 （2）折线图 （3）柱形图 （4）饼图 （5）散点图 （6）箱形图 **知识点 3-图表的辅助元素** 教师通过 PPT 讲解图表的辅助元素。 （1）什么是辅助元素 辅助元素是指除了根据数据绘制的图像之外的内容，用于对图形进行补充说明。 （2）常用的辅助元素 ● 坐标轴 ● 标题 ● 图例 ● 网格 ● 参考线 ● 参考区域 ● 注释文本 **知识点 4-使用 Matplotlib 绘制折线图** 教师通过 PPT 结合实操的形式讲解使用 Matplotlib 绘制折线图。 （1）绘制折线图的基本思路 ① 导入 pyplot 模块。 ② 使用 plot()函数绘制线条。 ③ 完善图表，添加辅助元素。 ④ 使用 show()函数展示图表。 （2）通过代码演示绘制包含一条线的折线图 （3）绘制包含多条线的折线图 ● 多次调用 plot()函数绘制多条线 ● 调用 plot()函数时一次传入多组数据 （4）通过代码演示绘制包含两条线的折线图 （5）给线条添加数据标记的方式 在调用 plot()函数绘制线条时将标记取值传递给 marker 参数，另外还可以传入 markersize 或 ms 参数，用于设置标记的大小。 （6）通过代码演示如何给折线图的线条添加数据标记 （7）通过代码演示如何给折线图添加标题、设置坐标轴的标签

知识点 5-使用 Matplotlib 绘制柱形图

教师通过 PPT 结合实操的形式讲解使用 Matplotlib 绘制柱形图。

（1）绘制折线图的基本思路

① 导入 pyplot 模块。

② 使用 bar()函数绘制柱形。

③ 完善图表，添加辅助元素。

④ 使用 show()函数展示图表。

（2）通过代码演示绘制包含一组柱形的柱形图

（3）绘制包含多组柱形的柱形图

如果希望在绘制区域上再绘制另一组柱形，也就是说绘制包含两组柱形的柱形图，则需要再次调用 bar()函数，并在该函数中通过第一个参数控制另一组柱形显示的位置。

（4）通过代码演示绘制包含两组柱形的柱形图

（5）通过代码演示如何给柱形图添加坐标轴的标签、标题和注释文本、图例

三、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第三、四课时

（使用 Matplotlib 绘制直方图、使用 Matplotlib 绘制散点图、可视化数据的分布）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了什么是数据可视化、常见的图表类型、图表的辅助元素，以及如何使用 Matplotlib 绘制折线图和柱形图，本节课将继续学习 Matplotlib 的知识，包括使用 Matplotlib 绘制直方图、使用 Matplotlib 绘制散点图，另外再介绍一些 Seaborn 库的知识。

三、新课讲解

知识点 1-使用 Matplotlib 绘制直方图

教师通过 PPT 结合实操的形式讲解使用 Matplotlib 绘制直方图。

（1）绘制折线图的基本思路

① 导入 pyplot 模块。

② 使用 hist()函数绘制图像。

③ 完善图表，添加辅助元素。

④ 使用 show()函数展示图表。

（2）通过代码演示如何基于上述思路绘制直方图

（3）解决刻度标签与图形不匹配的问题

- 获取每个矩形对应的数值区间。
- 重新设置 x 轴上的刻度标签。

（4）通过代码演示如何优化直方图

知识点 2-使用 Matplotlib 绘制散点图

教师通过 PPT 结合实操的形式讲解使用 Matplotlib 绘制散点图。

（1）绘制折线图的基本思路

① 导入 pyplot 模块。

② 使用 scatter()函数绘制点。。

③ 完善图表，添加辅助元素。

④ 使用 show()函数展示图表。

（2）通过代码演示绘制一个散点图

（3）通过代码演示如何给散点图添加网格

知识点 3-可视化数据的分布

教师通过 PPT 结合实操的形式讲解可视化数据的分布。

（1）查看数据分布情况的方式

- 单变量的数据采用直方图或核密度曲线。
- 双变量的数据可以采用散点图、二维直方图、核密度估计图形。

（2）通过代码演示如何使用 displot()函数绘制直方图

（3）什么是核密度估计曲线

核密度估计曲线是一种统计图表，它通过在每个数据点周围放置一个核函数，并通过调整核函数的宽度得到一个平滑的曲线。

（4）通过代码演示如何使用 displot()函数绘制核密度估计曲线

（5）通过代码演示绘制核密度估计曲线时如何显示密度观察条

（6）通过代码演示如何使用 jointplot()函数绘制散点图

（7）什么是六边形二维直方图

六边形二维直方图的效果类似于蜂巢的形状，它主要用于显示落在六边形区域内的观察值的计数，适用于数量较大的数据集。

（8）通过代码演示如何使用 jointplot()函数绘制六边形二维直方图

（9）通过代码演示如何使用 jointplot()函数绘制二维密度图

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

五、布置作业

第五、六课时

（用分类数据绘图、Pyecharts 简介、使用 Pyecharts 绘制柱形图）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了使用 Matplotlib 绘制直方图、使用 Matplotlib 绘制散点图、可视化数据的分布，本节课将学习用分类数据绘图、Pyecharts 简介、使用 Pyecharts 绘制柱形图。

三、新课讲解

知识点 1-用分类数据绘图

教师通过 PPT 结合实操的形式讲解用分类数据绘图。

（1）什么是分类数据

分类数据是按照现象的某种属性对其进行分类或分组而得到的反映事物类型的数据。

（2）通过代码演示如何使用 stripplot()函数绘制一个散点图

（3）通过代码演示如何使用 boxplot()函数绘制箱形图

（4）什么是小提琴图

- 小提琴图是箱式图与核密度图的结合。
- 小提琴图不仅能显示一组数据的中位数、上四分位数、下四分位数等信息，还能显示数据在不同数值下的概率密度。

（5）通过代码演示如何使用 violinplot()函数绘制小提琴图

（6）通过代码演示如何使用 barplot()函数绘制柱形图

（7）通过代码演示如何使用 pointplot()函数绘制点图

知识点 2-Pyecharts 简介

教师通过 PPT 结合实操的形式讲解 Pyecharts 简介。

（1）使用 Pyecharts 绘制图表的思路

①创建图表类的对象。

②添加图表用到的数据。

③添加图表配置项。

④渲染图表。

（2）创建图表类的对象

- 常用的图表类
- 创建图表类对象的方式

（3）添加图表用到的数据

- add()方法
- add_xx()方法，直角坐标系图表类一般使用 add_yaxis()或 add_xaxis()方法

（4）添加图表配置项

- 全局配置项
- 全局配置项的设置方式
- 系列配置项
- 系列配置项的设置方式

（5）渲染图表

- render()
- render_notebook()

知识点 3-使用 Pyecharts 绘制柱形图

教师通过 PPT 结合实操的形式讲解使用 Pyecharts 绘制柱形图。

（1）绘制柱形图的基本思路

① 创建 Bar 类的对象。

② 使用 add_xaxis()和 add_yaxis()方法添加数据。

③ 添加图表配置项。

④ 渲染图表。

（2）通过代码演示如何基于上述思路绘制柱形图

四、归纳总结

	教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第七、八课时** **（使用 Pyecharts 绘制词云图、使用 Pyecharts 绘制气泡图、使用 Pyecharts 绘制圆环图、案例：电影数据分析）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了用分类数据绘图、Pyecharts 简介、使用 Pyecharts 绘制柱形图，本节课将继续学习使用 Pyecharts 绘制词云图、使用 Pyecharts 绘制气泡图、使用 Pyecharts 绘制圆环图，以及围绕所学的知识完成一个案例。 三、新课讲解 **知识点 1-使用 Pyecharts 绘制词云图** 教师通过 PPT 结合实操的形式讲解使用 Pyecharts 绘制词云图。 （1）什么是词云图 ● 词云图会将文本中出现频率较高的关键词予以视觉上的突出。 ● 每个词出现的频率代表的是词的重要性。 ● 字号越大，说明这个关键词越重要。 （2）绘制词云图的基本思路 ① 创建 WordCloud 类的对象。 ② 使用 add ()方法添加数据。 ③ 添加图表配置项。 ④ 渲染图表。 （3）通过代码演示如何基于上述思路绘制词云图 **知识点 2-使用 Pyecharts 绘制气泡图** 教师通过 PPT 结合实操的形式讲解使用 Pyecharts 绘制气泡图。 （1）什么是气泡图 （2）绘制气泡图的基本思路 ① 创建 Scatter 类的对象。 ② 使用 add ()方法添加数据。 ③ 添加图表配置项。 ④ 渲染图表。 （3）通过代码演示如何基于上述思路绘制气泡图 **知识点 3-使用 Pyecharts 绘制圆环图** 教师通过 PPT 结合实操的形式讲解使用 Pyecharts 绘制圆环图。 （1）什么是圆环图 （2）绘制圆环图的基本思路 ① 创建 Pie 类的对象。 ② 使用 add ()方法添加数据。 ③ 添加图表配置项。 ④ 渲染图表。 （3）通过代码演示如何基于上述思路绘制圆环图

	知识点 4-案例：电影数据分析 教师通过 PPT 结合实操的形式讲解案例。 （1）通过 PPT 介绍案例的需求 （2）通过代码演示如何读取数据、查看摘要信息 （3）通过代码演示案例的实现步骤 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第九、十课时（上机练习）** 上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。 **上机一：（考察知识点为使用 Matplotlib 绘制折线图、使用 Matplotlib 绘制柱形图、使用 Matplotlib 绘制直方图、使用 Matplotlib 绘制散点图）** **形式：单独完成** **题目：** 练习 6.2 节全部的示例代码。 **上机二：（考察知识点为可视化数据的分布、用分类数据绘图）** **形式：单独完成** **题目：** 练习 6.3 节全部的示例代码。 **上机三：（考察知识点为使用 Pyecharts 绘制柱形图、使用 Pyecharts 绘制词云图、使用 Pyecharts 绘制气泡图、使用 Pyecharts 绘制圆环图）** **形式：单独完成** **题目：** 练习 6.4.2 到 6.4.5 小节全部的示例代码 **上机四：（考察知识点为案例：电影数据分析）** **形式：单独完成** **题目：** 请按照 6.5 节案例的要求，编写代码，从 IMDB-Movie-Data.csv 文件中读取数据，查看摘要信息，并按照设定的目标操作数据。
教学后记	后续可增加 “用可视化展示校园‘绿色低碳’成果（如垃圾分类准确率变化）”任务，强化环保意识；针对图表细节，可提供 “图表检查清单”（标题、标签、单位、来源），规范学生操作。

课题名称	第 7 章 文本数据分析	计划课时	12 课时
教学引入	自然语言处理（NLP）是人工智能领域一个重要方向，在这一方向上文本数据占据着很大的市场，由于文本中可能包含中文、英文等一些语言的内容，所以 Python 针对不同语言的文本提供了相应的库进行处理，常见的有用于处理英文文本的 NLTK 库，用于处理中文文本的 jieba 库。接下来，本章主要围绕着 NLTK 和 jieba 库介绍文本预处理的基本流程，以及文本数据分析的经典应用，包括文本情感分析、文本相似度和文本分类。		
教学目标	● 使学生了解 NLTK 与 jieba 库，能够说明 NLTK 与 jieba 库的用途 ● 使学生掌握 NLTK 语料库的安装，能够在计算机中成功安装 NLTK 语料库 ● 使学生熟悉文本预处理的流程，能够归纳出文本预处理的基本流程 ● 使学生掌握分词的方式，能够通过 NLTK 与 jieba 库对文本进行分词 ● 使学生掌握词性标注的方式，能够通过 pos_tag()函数对英文文本进行分词标注 ● 使学生掌握词形归一化操作，能够通过 nltk.stem 模块实现词形归一化的操作 ● 使学生掌握删除停用词操作，能够通过 stopwords 模块实现删除停用词的操作 ● 使学生熟悉文本情感分析，能够通过多种方式实现简单的文本情感分析 ● 使学生熟悉文本相似度，可以结合 NLTK 与余弦相似度实现简单的文本相似度分析 ● 使学生熟悉文本分类，可以结合 NLTK 与朴素贝叶斯算法实现简单的文本分类分析		
教学重点	● 分词 ● 词性标注 ● 词形归一化 ● 删除停用词		
教学难点	● 文本情感分析 ● 文本相似度 ● 文本分类		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	• **舆情责任与网络素养** 以 “商品评论情感分析” 为基础，拓展 “用文本分析识别校园论坛中的‘负面舆情’（如恶意吐槽食堂）”，引导学生思考 “如何理性对待网络言论”“用技术维护健康网络环境”，培养 “不传播谣言、不煽动情绪” 的网络素养。 • **隐私保护与伦理意识** 强调 “文本数据分析不可侵犯个人隐私”，如 “分析学生评论时需脱敏（隐藏姓名、学号）”，以 “非法抓取用户评论用于商业营销” 案例警示学生，树立 “技术服务于人，不损害他人权益” 的伦理意识。		
教学	**第一、二课时** **（认识 NLTK 与 jieba、安装 jieba 和 NLTK 语料库、文本预处理基本流程、**		

过 程	**分词、词性标注）** 一、创设情景，导入新课 教师通过给学生展示一些应用文本分析的场景，例如场景是：锤子新发布的功能“BigBang”分词功能、智能客服、网络舆情监控等，并根据这些场景，引出这些场景都是应用的文本数据分析，从而实现导入新课的目的。 二、新课讲解 **知识点 1-认识 NLTK 与 jieba** 教师通过 PPT 讲解认识 NLTK 与 jieba。 （1）NLTK 是什么 NLTK 是一套基于 Python 的自然语言处理工具包，可以方便地完成自然语言处理的任务，包括分词、词性标注、命名实体识别（NER）及句法分析等。 （2）NLTK 的常用模块 （3）jieba 库的特点 ● 支持三种分词模式。 ● 支持繁体分词。 ● 支持自定义词典。 ● MIT 授权协议。 **知识点 2-安装 jieba 和 NLTK 语料库** 教师通过 PPT 结合实操的形式讲解安装 jieba 和 NLTK 语料库。 （1）安装 jieba 库的方式 （2）通过 Anaconda 命令行工具演示如何安装 jieba 库 （3）下载 NLTK 语料库 ● 打开 NLTK 下载器 ● 安装所有选项 ● 单独安装某个词料库或模型 （4）通过代码演示如何打开下载器和安装部分语料库 （5）通过代码验证语料库是否安装成功 **知识点 3-文本预处理基本流程** 教师通过 PPT 讲解文本预处理基本流程。 （1）文本预处理的基本流程图 （2）分词 （3）词形归一化 （4）删除停用词 **知识点 4-分词** 教师通过 PPT 结合实操的形式讲解分词。 （1）什么是分词 分词是指将由连续词或字组成的语句，按照一定的规则划分成独立词语的过程。 （2）英文文本分词的方式 NLTK 库的 word_tokenize()函数用于以空格或标点符号为分隔符对英文文本进行分词，并返回分词后的单词列表。 （3）通过代码演示如何使用 word_tokenize()函数实现英文文本分词效果 （4）中文文本分词的方式 jieba 模块的 cut() 函数用于实现中文文本分词的效果

	（5）通过代码演示如何使用 cut()函数实现中文文本分词的效果 **知识点 5-词性标注** 教师通过 PPT 结合实操的形式讲解词性标注。 （1）词性的分类 （2）什么是词性标注 词性标注是指为分词结果中的每个单词标注一个正确的词性。 （3）NLTK 库的词性标注集 （4）词性标注的方式 先下载 averaged_perceptron_tagger 模块，再通过该模块的 pos_tag()函数进行标注。 （5）通过代码演示对一段英文文本进行词性标注 三、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第三、四课时** **（词形归一化、删除停用词、文本情感分析）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了安装 jieba 和 NLTK 语料库、文本预处理基本流程、分词、词性标注，本节课将学习词形归一化、删除停用词、文本情感分析。 三、新课讲解 **知识点 1-词形归一化** 教师通过 PPT 结合实操的形式讲解词形归一化。 （1）为什么要进行词形归一化 有些单词有多个变种，这会影响语料库学习的准确度，为了解决这个问题，我们需要对一个词的不同形态进行规范化，以提高文本处理的效率。 （2）词形规一化的方式 ● 词干提取：删除不影响词性的词缀，得到单词词干。 ● 词形还原：捕捉基于词根的规范单词形式。 （3）实现词干提取的方式 ● nltk.stem 模块的 PorterStemmer 类：波特词干提取器。 ● nltk.stem 模块的 PorterStemmer 类：兰卡斯特词干提取器。 ● nltk.stem 模块的 SnowballStemmer 类：其它词干提取器。 （4）通过代码演示如何使用上述三种方式实现词干提取的操作 （5）实现词形还原的方式 WordNetLemmatizer 类的 lemmatize()方法会比对 wordnet 语料库，并采用递归技术删除词缀，直至在词汇网络中找到匹配项，最终返回输入词的基本形式。如果没有找到匹配项，则直接返回输入词，不做任何变化。 （6）通过代码演示如何使用上述方式实现词形还原的操作 **知识点 2-删除停用词**

	教师通过 PPT 结合实操的形式讲解删除停用词。 （1）什么是停用词 停用词是指在信息检索中，为节省存储空间和提高搜索效率，在处理自然语言文本之前或之后会自动过滤掉某些没有具体意义的字或词。 （2）删除停用词的实现方式 （3）通过代码演示如何使用 NLTK 库实现删除停用词的功能 **知识点 3-文本情感分析** 教师通过 PPT 结合实操的形式讲解文本情感分析。 （1）什么是文本情感分析 （2）情感分析的分类 ● 情感极性分析 ● 情感程度分析 ● 主客观分析 （3）情感极性分析的方法 ● 基于情感词典的情感分析 ● 基于机器学习的情感分析 （4）基于情感词典的情感分析的基本思路 ①找出正向情感词、负向情感词、否定词以及程度副词。 ②如果情感词前面存在否定词，则将情感词的情感权值乘以-1；如果情感词前面有程度副词，就情感词的情感权值乘以程度副词的程度值。 ③将所有组得分相加，结果大于 0 归正向，小于 0 归负向。 （5）基于情感词典情感分析的举例 （6）朴素贝叶斯算法的概念 朴素贝叶斯是经典的机有监督的机器学习算法之一，它的思想是对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个概率最大就认为此待分类项属于哪个类别。 （7）基于朴素贝叶斯算法实现情感分析的方式 NaiveBayesClassifier 类封装了朴素贝叶斯分类算法的功能，该类中有一个类方法 train()，用于根据训练集来训练模型。 （8）通过代码演示如何使用 NaiveBayesClassifier 类实现情感分析 四、归纳总结 教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。 **第五、六课时** **（文本相似度、文本分类、案例：商品评论分析）** 一、复习巩固 教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。 二、通过直接引入的方式导入新课 上节课我们主要学习了词形归一化、删除停用词、文本情感分析，本节课将继续学习文本相似度、文本分类，以及围绕所学的知识完成一个案例。 三、新课讲解

知识点 1-文本相似度

教师通过 PPT 结合实操的形式讲解文本相似度。

（1）文本相似度的应用场景

（2）文本相似度的方法

- 基于关键字匹配的传统方法。
- 将文本映射到向量空间，再利用余弦相似度等方法进行计算。
- 基于深度学习的方法。

（3）文本映射到向量空间的实现思路

①找出两篇文章的关键词。

②从每篇文章中各取出若干个关键词，把这些关键词合并成一个集合，然后计算每篇文章中各个词对于这个集合中的关键词的词频。

③生成两篇文章中各自的词频向量。

④计算两个向量的余弦相似度，值越大则表示越相似。

（4）通过代码演示如何根据上述步骤实现文本相似度的操作

知识点 2-文本分类

教师通过 PPT 结合实操的形式讲解文本分类。

（1）什么是文本分类

（2）文本分类的实现思路

①数据集准备

②特征抽取。

③模型训练。

④分类结果评价。

（3）通过代码演示根据上述思路使用 NLTK 库实现文本分类的操作

知识点 3-案例：商品评论分析

教师通过 PPT 结合实操的形式讲解案例。

（1）通过 PPT 介绍案例的需求

（2）通过代码演示如何读取数据

（3）通过代码演示案例的实现步骤

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

第七、八学时（Python 爬虫基础）

一、什么是爬虫

爬虫：一段自动抓取互联网信息的程序，从互联网上抓取对于我们有价值的信息。

二、Python 爬虫架构

Python 爬虫架构主要由五个部分组成，分别是调度器、URL 管理器、网页下载器、网页解析器、应用程序（爬取的有价值数据）。

调度器：相当于一台电脑的 CPU，主要负责调度 URL 管理器、下载器、解析器之间的协调工作。

URL 管理器：包括待爬取的 URL 地址和已爬取的 URL 地址，防止重复抓取 URL 和循环抓取 URL，实现 URL 管理器主要用三种方式，通过内存、数据库、缓存数据库来实现。

网页下载器：通过传入一个 URL 地址来下载网页，将网页转换成一个字符串，网页下载器有 urllib2（Python 官方基础模块）包括需要登录、代理、和 cookie，requests(第三方包)

网页解析器：将一个网页字符串进行解析，可以按照我们的要求来提取出我们有用的信息，也可以根据 DOM 树的解析方式来解析。网页解析器有正则表达式（直观，将网页转成字符串通过模糊匹配的方式来提取有价值的信息，当文档比较复杂的时候，该方法提取数据的时候就会非常的困难）、html.parser（Python 自带的）、beautifulsoup（第三方插件，可以使用 Python 自带的 html.parser 进行解析，也可以使用 lxml 进行解析，相对于其他几种来说要强大一些）、lxml（第三方插件，可以解析 xml 和 HTML），html.parser 和 beautifulsoup 以及 lxml 都是以 DOM 树的方式进行解析的。

应用程序：就是从网页中提取的有用数据组成的一个应用。

下面用一个图来解释一下调度器是如何协调工作的:

三、urllib2 实现下载网页的三种方式

```
#!/usr/bin/python
# -*- coding: UTF-8 -*-

import cookielib
import urllib2

url = "http://www.baidu.com"
response1 = urllib2.urlopen(url)
print "第一种方法"
#获取状态码，200 表示成功
print response1.getcode()
#获取网页内容的长度
print len(response1.read())

print "第二种方法"
request = urllib2.Request(url)
#模拟 Mozilla 浏览器进行爬虫
request.add_header("user-agent","Mozilla/5.0")
response2 = urllib2.urlopen(request)
print response2.getcode()
print len(response2.read())

print "第三种方法"
cookie = cookielib.CookieJar()
#加入 urllib2 处理 cookie 的能力
opener = urllib2.build_opener(urllib2.HTTPCookieProcessor(cookie))
urllib2.install_opener(opener)
response3 = urllib2.urlopen(url)
```

```
print response3.getcode()
print len(response3.read())
print cookie
```

四、第三方库 Beautiful Soup 的安装

Beautiful Soup: Python 的第三方插件用来提取 xml 和 HTML 中的数据，官网地址 https://www.crummy.com/software/BeautifulSoup/

1、安装 Beautiful Soup

打开 cmd(命令提示符),进入到 Python(Python2.7版本)安装目录中的 scripts 下,输入 dir 查看是否有 pip.exe, 如果用就可以使用 Python 自带的 pip 命令进行安装，输入以下命令进行安装即可：

```
pip install beautifulsoup4
```

2、测试是否安装成功

编写一个 Python 文件，输入:

```
import bs4
print bs4
```

运行该文件，如果能够正常输出则安装成功。

五、使用 Beautiful Soup 解析 html 文件

```
#!/usr/bin/python
# -*- coding: UTF-8 -*-

import re

from bs4 import BeautifulSoup
html_doc = """
<html><head><title>The Dormouse's story</title></head>
<body>
<p class="title"><b>The Dormouse's story</b></p>
<p class="story">Once upon a time there were three little sisters; and
their names were
<a           href="http://example.com/elsie"           class="sister"
id="link1">Elsie</a>,
<a href="http://example.com/lacie" class="sister" id="link2">Lacie</a>
and
<a           href="http://example.com/tillie"           class="sister"
id="link3">Tillie</a>;
and they lived at the bottom of a well.</p>
```

```
<p class="story">...</p>
"""
#创建一个 BeautifulSoup 解析对象
soup = BeautifulSoup(html_doc,"html.parser",from_encoding="utf-8")
#获取所有的链接
links = soup.find_all('a')
print "所有的链接"
for link in links:
    print link.name,link['href'],link.get_text()

print "获取特定的 URL 地址"
link_node = soup.find('a',href="http://example.com/elsie")
print
link_node.name,link_node['href'],link_node['class'],link_node.get_tex
t()

print "正则表达式匹配"
link_node = soup.find('a',href=re.compile(r"ti"))
print
link_node.name,link_node['href'],link_node['class'],link_node.get_tex
t()

print "获取 P 段落的文字"
p_node = soup.find('p',class_='story')
print p_node.name,p_node['class'],p_node.get_text()
```

第九、十课时

1.1.1八爪鱼采集器 官网地址

https://www.bazhuayu.com/?campaign=baidu&terminal=pc&plan=%E5%93%81%E7%89%8C&unit=%E5%85%AB%E7%88%AA%E9%B1%BC&keyword=%E5%85%AB%E7%88%AA%E9%B1%BC&bd_vid=9187269027533183325

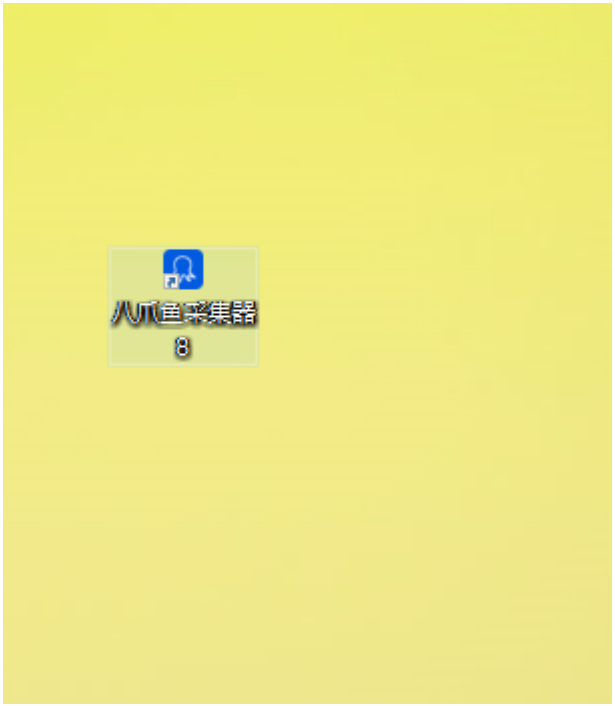
1.1.2安装篇

1、登录官方网站下载软件进行安装

https://www.bazhuayu.com/?campaign=baidu&terminal=pc&plan=%E5%93%81%E7%89%8C&unit=%E5%85%AB%E7%88%AA%E9%B1%BC&keyword=%E5%85%AB%E7%88%AA%E9%B1%BC&bd_vid=9187269027533183325

2、选择下载软件，并进行安装

3、安装完成后，打开软件



1.1.3采集示例：

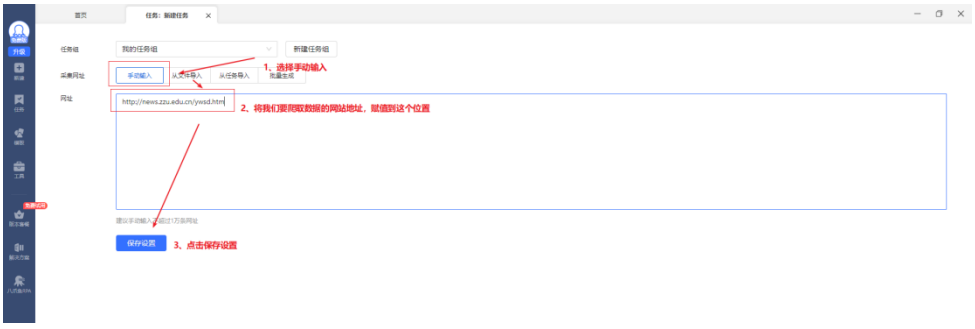
1、打开要采集的网站地址，赋值里面的 url 连接

2、保存对应的 url 连接地址

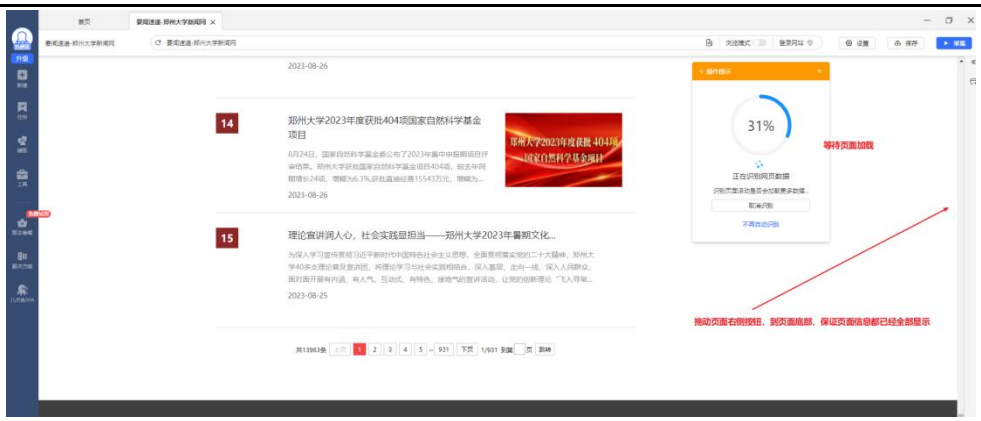
http://news.zzu.edu.cn/ywsd.htm

3、打开八爪鱼采集器

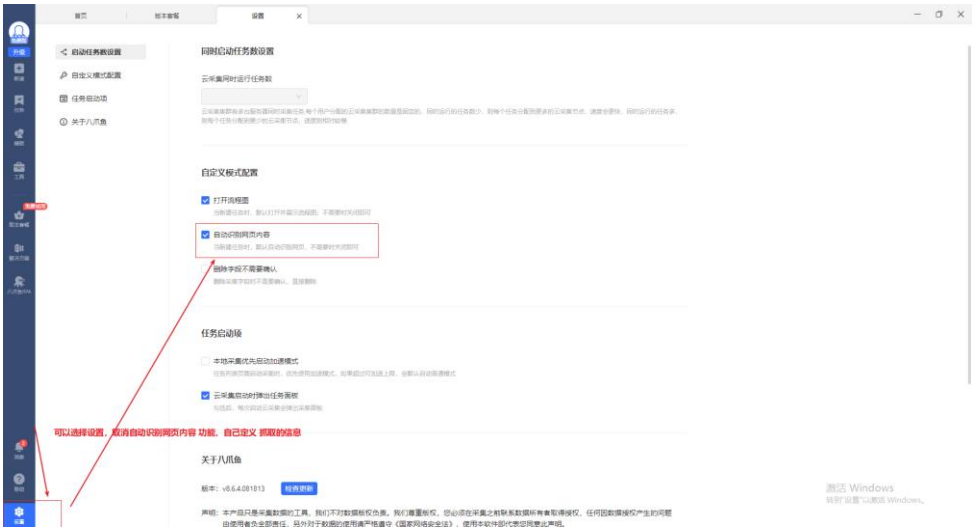
4、手动添加目标网站地址



5、将页面 拖动到最底部 查看信息是否加载成功， 这里可以取消自动识别

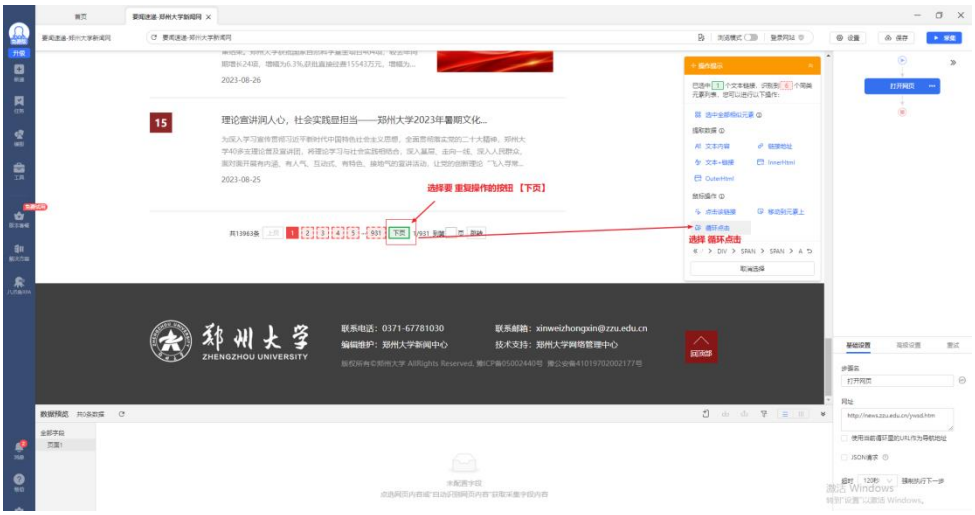


6、也可以通过设置进行自动识别的取消



7、页面拖动到最下方

8、循环点击翻页按钮设置



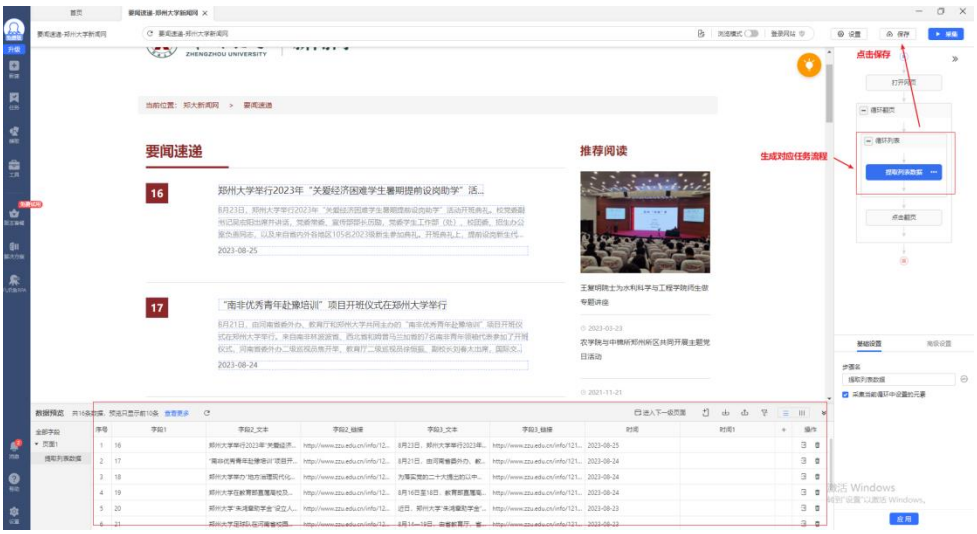
9、确认任务创建完成

10、示例：选择要爬取的页面 div 块内容

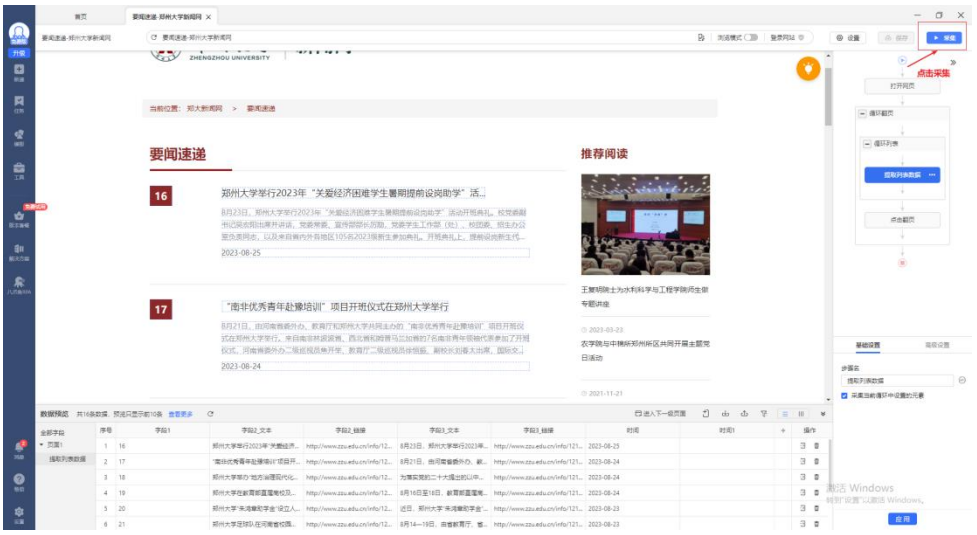
11、选择匹配所有相似内容模块

12、选择 获取当前元素中的所有内容（这里自动获取，也可以根绝需求 手动设置 帮助文档）

13、保存当前定义的执行任务



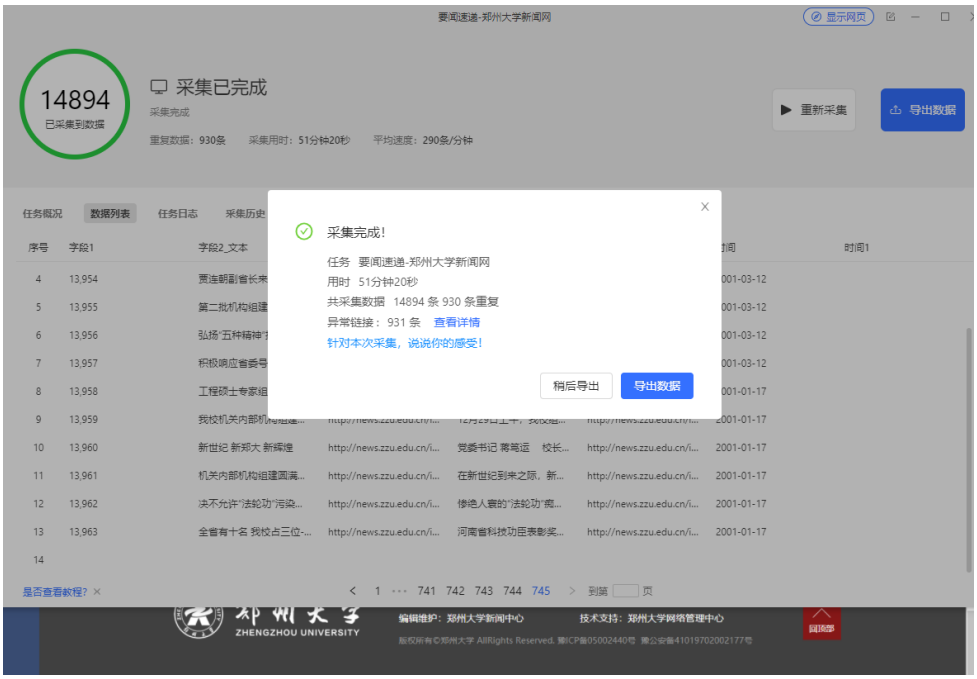
14、点击采集按钮，运行当前执行任务进行数据采集



15、选择普通采集，开始进行数据采集

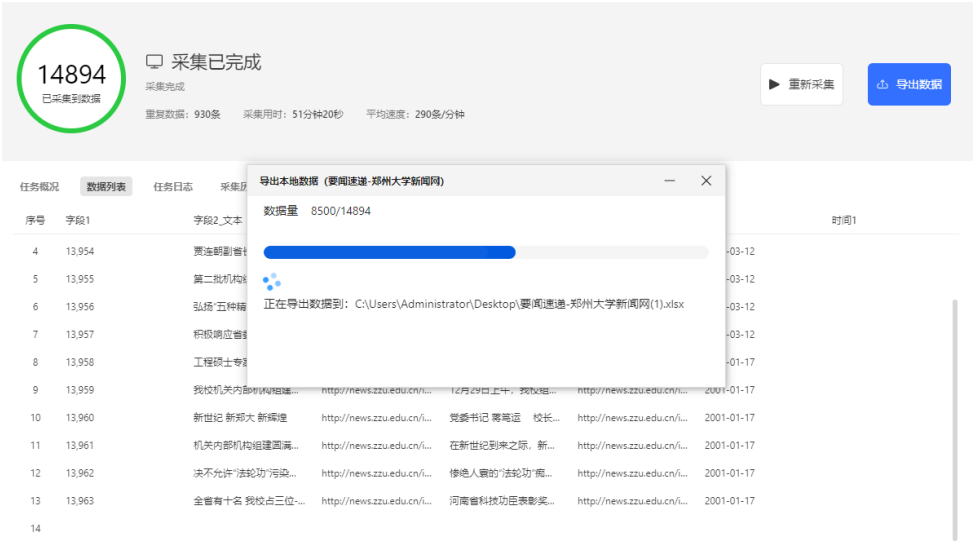
16、页面采集中，内容介绍

17、 数据采集完毕，14894 条数据，用时 51 分钟 20 秒

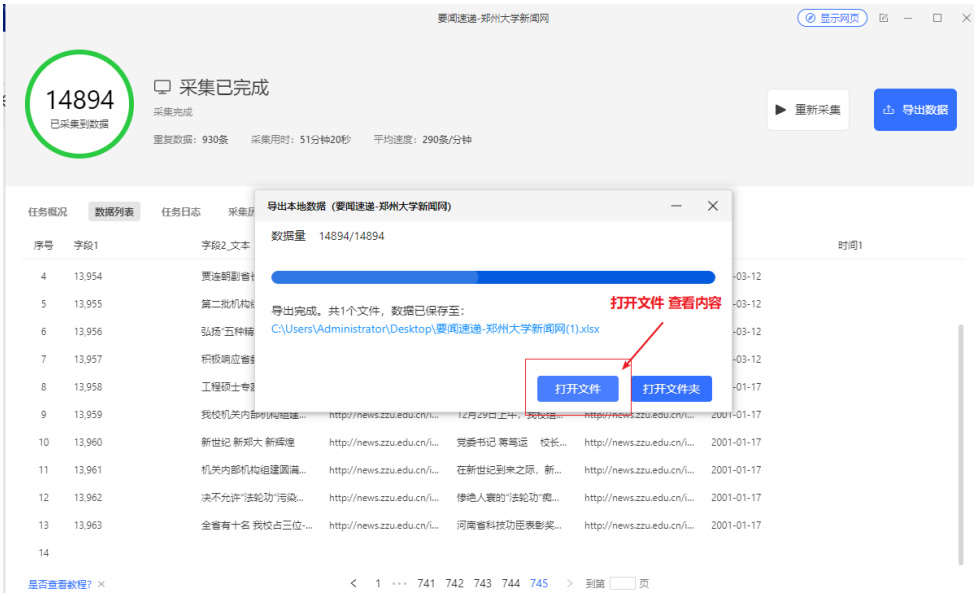


18、采集完毕后进行数据的导出，目前官方解释没有倒数数据的数量限制

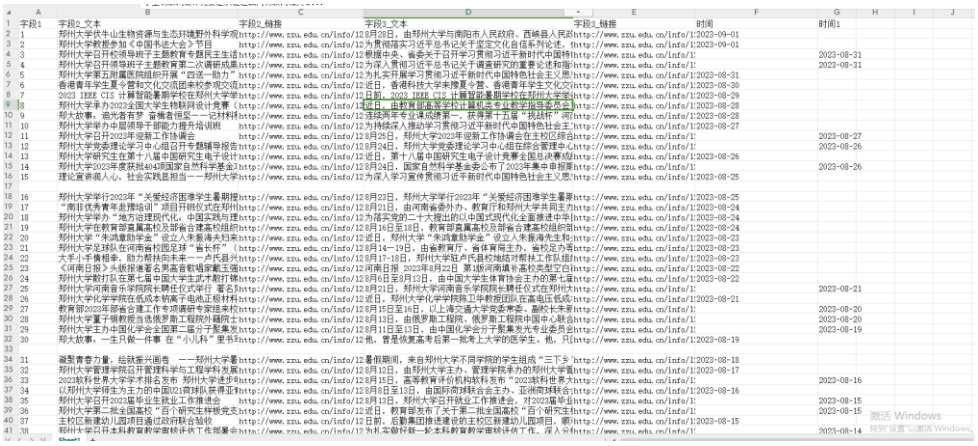
19、本地数据进行保存



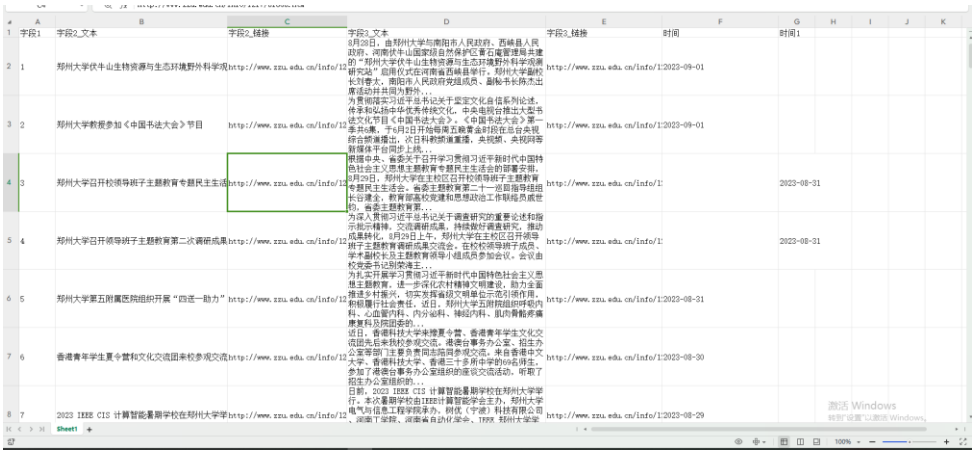
20、数据采集完成后，打开本地生成的 excle 文件，查看采集的数据信息



21、查看采集数据内容，行数 ，数据内容准确性，是否有数据丢失



22、整理数据文本信息，内容居中展示，文本内容自动换行



	1.1.4官网帮助视频文档 *https://www.bazhuayu.com/tutorial8/81spwbxh* 学生上机，选择某一网站进行信息爬取操作，强调不能设置高频次，避免对造成测试的网站产生压力。 **第十一、十二课时（上机练习）** 上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。 **上机一：（考察知识点为安装 jieba 和 NLTK 语料库、分词、词性标注、词形归一化、删除停用词）** **形式：单独完成** **题目：** 练习 8.1 和 8.2 节全部的示例代码。 **上机二：（考察知识点为文本情感分析、文本相似度、文本分类）** **形式：单独完成** **题目：** 练习 8.3 到 8.5 节全部的示例代码。 **上机三：（考察知识点为案例：商品评论分析）** **形式：单独完成** **题目：** 请按照 8.6 节案例的要求，编写代码，从“商品评价信息.csv”文件中读取数据，并按照设定的目标操作数据。
教学后记	下次教学可增加 “用文本分析解读‘二十大报告’关键词” 的任务，强化家国情怀；针对情感分析的中性评论问题，可提供 “扩展情感词典”，帮助学生更精准判断。

课题名称	第 8 章 机器学习入门	计划课时	4 课时
教学引入	在这个数据为王的时代，海量数据已经远远超出了直接计算的可能性，我们若想要从海量的数据中高效地提取有价值的信息，需要用到专门的学习算法，这就是机器学习的作用所在。本章为大家介绍机器学习的相关内容，相信你们会惊叹机器学习的神奇并为之着迷的。		
教学目标	● 使学生了解机器学习，能够说出什么是机器学习 ● 使学生熟悉机器学习的基本概念，能够归纳机器学习涉及的基本概念 ● 使学生了解机器学习算法的分类，能够区分监督学习、无监督学习和强化学习 ● 使学生了解机器学习解决问题的流程，能够说出机器学习解决问题的流程 ● 使学生熟悉机器学习库 scikit-learn，能够列举至少 3 个 scikit-learn 的模块和数据集 ● 使学生掌握 KNN 算法，能够使用 scikit-learn 的 API 实现 KNN 算法		
教学重点	● KNN 算法的思想 ● 使用 sklearn 实现 KNN 算法 ● 超参数 ● 使用 sklearn 实现归一化		
教学难点	● 超参数 ● 网格搜索与交叉验证 ● 归一化		
教学方式	课堂教学以 PPT 讲授为主，并结合多媒体进行教学		
思政元素	1. **算法伦理与公平正义** 讲解 KNN 算法时，讨论 “用机器学习预测‘学生奖学金资格’” 的公平性，强调 “算法不可因‘家庭背景’‘性别’等无关特征歧视学生”，引导学生树立 “算法公平、不偏袒” 的伦理意识，避免技术滥用。 2. **科技向善与社会责任** 以 “预测签到位置” 案例为基础，拓展 “用机器学习预测‘校园共享单车停放热点’，帮助学校规划停车区”，引导学生思考 “如何用机器学习解决校园实际问题”，培养 “科技向善、服务社会” 的责任。 3. **家国情怀与科技自强** 介绍 scikit-learn 等库时，提及 “我国在机器学习领域的突破（如华为 MindSpore 框架）”，引导学生认识 “核心技术自主可控的重要性”，树立 “学好技术，为我国人工智能发展贡献力量” 的家国情怀。		
教学过程	**第一课时** **（什么是机器学习、机器学习的基本概念、机器学习算法的分类、机器学习解决问题的流程、认识机器学习库 scikit-learn）** 一、创设情景，导入新课		

	教师播放机器人与机器人对话和机器人与人对话的视频，给学生提问问题，例如问题是：机器人为什么可以跟人类交流，并对学生的回答进行总结，引出机器学习的概念，从而实现导入新课的目的。 二、新课讲解 **知识点 1-什么是机器学习** 教师通过 PPT 讲解什么是机器学习。 （1）什么是机器学习 机器学习，从字面意思来看，就是让机器具备和人类一样的学习能力，包括决策、推理、认知、识别等智能行为。 （2）学习的三个概念 ● 任务 T ● 训练经验 E ● 性能目标 P **知识点 2-机器学习的基本概念** 教师通过 PPT 讲解机器学习的基本概念。 （1）鸢尾花数据集 ● 记录了变色鸢尾花、山鸢尾花和维吉尼亚鸢尾花这三类花。 ● 每一类鸢尾花收集了 50 条记录，整个鸢尾花数据集共 150 条记录。 ● 每条记录代表的是一朵花的数据。 ● 每一条记录包括萼片长度、萼片宽度、花瓣长度和花瓣宽度共 4 个属性。 （2）数据相关概念 ● 数据集 ● 样本 ● 特征 ● 标签 ● 特征空间 ● 特征向量 **知识点 3-机器学习算法的分类** 教师通过 PPT 讲解机器学习算法的分类。 （1）监督学习 （2）无监督学习 （3）强化学习 **知识点 4-机器学习解决问题的流程** 教师通过 PPT 讲解机器学习解决问题的流程。 （1）搜集业务相关的带标签的原始数据 （2）对数据进行预处理 （3）建立特征工程 （4）训练机器学习模型 （5）在验证数据集上面测试模型，评估模型性能的好坏 **知识点 5-认识机器学习库 scikit-learn** 教师通过 PPT 讲解认识机器学习库 scikit-learn。 （1）什么是 scikit-learn ● 机器学习领域中热门的 Python 语言开源库。 ● 依赖 NumPy、pandas、sciPy、matplotlib 等一些扩展库。

- 涵盖机器学习的经典样例数据集。
- 拥有很多用于回归、分类、聚类等问题的高效算法。
- 提供了预处理、模型拟合、模型评估等一些功能。

（2）scikit-learn 库的常用模块

（3）scikit-learn 库的常用数据集

三、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

四、布置作业

第二课时

（KNN算法的思想、使用sklearn实现KNN算法、超参数、网格搜索与交叉验证）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了什么是机器学习、机器学习的基本概念、机器学习算法的分类、机器学习解决问题的流程、认识机器学习库 scikit-learn，本节课将学习 KNN 算法的思想、使用 sklearn 实现 KNN 算法、超参数、网格搜索与交叉验证。

三、新课讲解

知识点 1-KNN 算法的思想

教师通过 PPT 讲解 KNN 算法的思想。

（1）什么是 KNN 算法

KNN 算法又称为 K 近邻算法，它思想简单、应用数学知识少，通常被应用在分类或回归的场景中。

（2）KNN 算法的思想

如果一个样本在特征空间中的 K 个最相邻的样本中，大多数属于某个类别，则该样本属于该类别。

（3）诊断肿瘤的场景举例

知识点 2-使用 sklearn 实现 KNN 算法

教师通过 PPT 结合实操的形式讲解使用 sklearn 实现 KNN 算法。

（1）KNeighborsClassifier 类的构造方法

（2）通过代码演示如何使用 KNeighborsClassifier 类实现 KNN 算法

知识点 3-超参数

教师通过 PPT 结合实操的形式讲解超参数。

（1）什么是超参数

超参数指的是在模型训练之前要设置的参数，用来控制算法行为。

（2）KNN 算法的超参数

- K 值
- 距离权重 weight
- 闵可夫斯基距离计算公式中的 p

（3）超参数 K

（4）通过代码演示如何寻找合适的超参数 K

（5）超参数 weights

（6）通过代码演示如何寻找合适的超参数 weights

（7）超参数 p

（8）通过代码演示如何寻找合适的超参数 p

知识点 4-网格搜索与交叉验证

教师通过 PPT 结合实操的形式讲解网格搜索与交叉验证。

（1）什么是网格搜索

网格搜索是一种穷举搜索方法，它通过遍历给定的参数组合来优化模型，其原理类似在数组里面找最大值的过程。

（2）网格搜索工具类 GridSearchCV

网格搜索工具类 GridSearchCV 可以自动调参，只要把所有的参数可能性输入，就会得到一个合适的调参器以及合适参数。

（3）通过代码演示如何使用网格搜索的方式寻找合适的超参数

（4）什么是交叉验证

- 交叉验证是机器学习建立模型和验证模型参数的一种方法。
- 常见的是 K 折交叉验证。
- K 折交叉验证的做法。

```
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
data = load_breast_cancer()
X = data["data"]
y = data["target"]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.2)
from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier()
param_grid = [
    {
        "weights": ['uniform'],
        "n_neighbors": [i for i in range(1, 11)]
    },
    {
        "weights": ['distance'],
        "n_neighbors": [i for i in range(1, 11)],
        'p':[i for i in range(1, 6)]
    }
]

from sklearn.model_selection import GridSearchCV
grid_search = GridSearchCV(model, param_grid)
grid_search.fit(X_train, y_train)
result = grid_search.best_estimator_
```

print(result)
print(grid_search.best_score_)

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

五、布置作业

第三课时

（归一化、使用 sklearn 实现归一化、案例：预测签到位置）

一、复习巩固

教师通过上节课作业的完成情况，对学生吸收不好的知识点进行再次巩固讲解。

二、通过直接引入的方式导入新课

上节课我们主要学习了 KNN 算法的思想、使用 sklearn 实现 KNN 算法、超参数、网格搜索与交叉验证，本节课将继续学习归一化、使用 sklearn 实现归一化，以及围绕所学的知识完成一个案例。

三、新课讲解

知识点 1-归一化

教师通过 PPT 结合实操的形式讲解归一化。

（1）归一化的使用场景

（2）直线方程与单变量线性回归的对应关系

（3）常用的数据归一化

- 最值归一化
- 均值方差归一化

（4）什么是最值归一化

最值归一化用于将数据映射到 0~1 之间，适用于数据有明显边界的情况。

（5）通过代码演示最值归一化的计算方式

（6）什么是均值方差归一化

均值方差归一化可以将数据映射到均值为 0，方差为 1 的分布，适用于数据分部没有明显边界，有可能存在极端值的情况。

（7）通过代码演示均值方差归一化

知识点 2-使用 sklearn 实现归一化

教师通过 PPT 结合实操的形式讲解使用 sklearn 实现归一化。

（1）StandardScaler 类的方法

- fit()：计算训练集的均值和标准差。
- transform()：通过居中和缩放执行数据归一化。
- fit_transform()：等同于 fit()和 transform()方法结合。

（2）通过代码演示如何使用 StandardScaler 类实现归一化

知识点 3-案例：预测签到位置

教师通过 PPT 结合实操的形式讲解案例。

（1）通过 PPT 介绍案例的需求

（2）通过 PPT 介绍案例用到的数据

（3）通过代码演示案例的实现步骤

```
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
data = load_breast_cancer()
X = data["data"]
y = data["target"]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
print(X_train)
from sklearn.preprocessing import StandardScaler
ssc = StandardScaler()
ssc.fit(X_train)
X_train_ssc = ssc.transform(X_train)
print(X_train_ssc)
X_test_ssc = ssc.fit_transform(X_test)
print(X_test_ssc)
from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier()
from sklearn.model_selection import GridSearchCV
param_grid = [ {"weights": ['uniform', 'distance'],
                "n_neighbors": [i for i inrange(1, 11)]},
                ]
grid_search = GridSearchCV(model, param_grid)
grid_search.fit(X_train_ssc, y_train)
grid_search.best_score_
```

四、归纳总结

教师回顾本节课所讲的内容，并通过测试题的方式引导学生解答问题并给予指导。

五、布置作业

第四课时（上机练习，可选）

上机练习主要针对本章中需要重点掌握的知识点，以及在程序中容易出错的内容进行练习，通过上机练习可以考察同学对知识点的掌握情况，对代码的熟练程度。

上机一：（考察知识点为使用 sklearn 实现 KNN 算法、超参数）

形式：单独完成

题目：

练习 9.2.2 和 9.2.3 小节全部的示例代码。

上机二：（考察知识点为网格搜索与交叉验证、使用 sklearn 实现归一化）

形式：单独完成

题目：

练习 9.2.4 和 9.2.6 小节全部的示例代码。

	上机三：（考察知识点为案例：预测签到位置） **形式：单独完成** **题目：** 请按照 9.3 节案例的要求，编写代码，按照设定的目标操作数据。
教学后记	后续可根据课时情况适当增加其它机器学习算法的实例