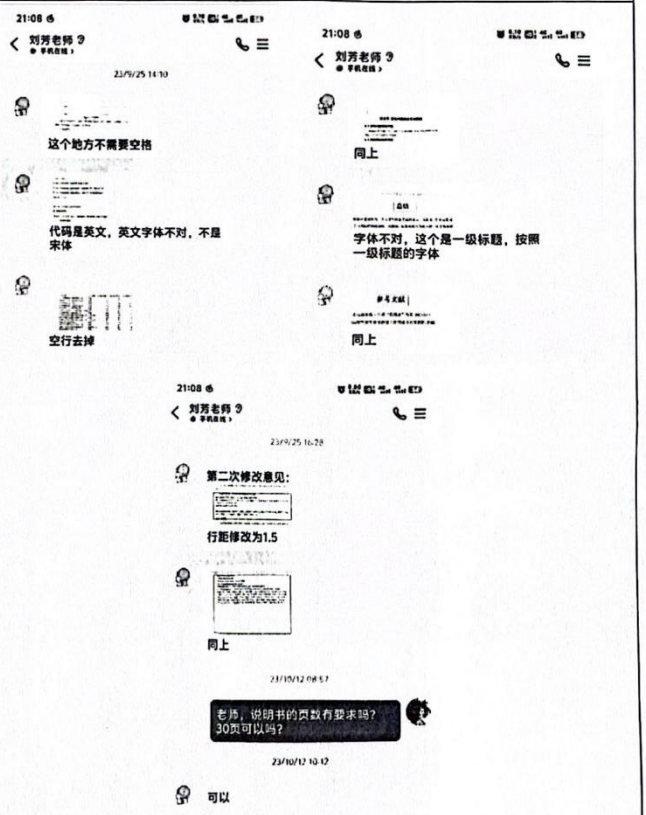
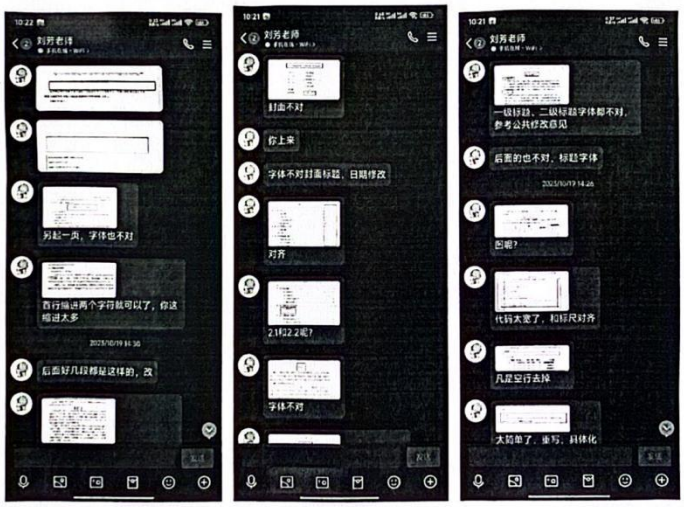
湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	饶蕾	学号	202165110021	指导教师	刘芳
二级学院	信息工程	专业	大数据技术	班级	大数据技术二班
毕业设计题目	基于智联卓聘网的招聘数据分析和可视化报告				
过程性指导记录内容					

指导教师（签字）：刘芳

2024 年 5 月 21 日

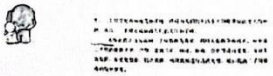
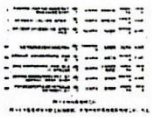
湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	粟杨林	学号	202165110022	指导教师	刘芳
二级学院	信息工程学院	专业	大数据技术	班级	大数据技术二班
毕业设计题目	基于中国地震台网的历史数据分析与可视化报告				
过程性指导记录内容	说明每一次指导情况及提供原始支撑记录（如：QQ、微信、网络空间等互动截图） 				

指导教师（签字）：刘芳

2024 年 5 月 21 日

湖南网络工程职业学院毕业设计教师指导记录表

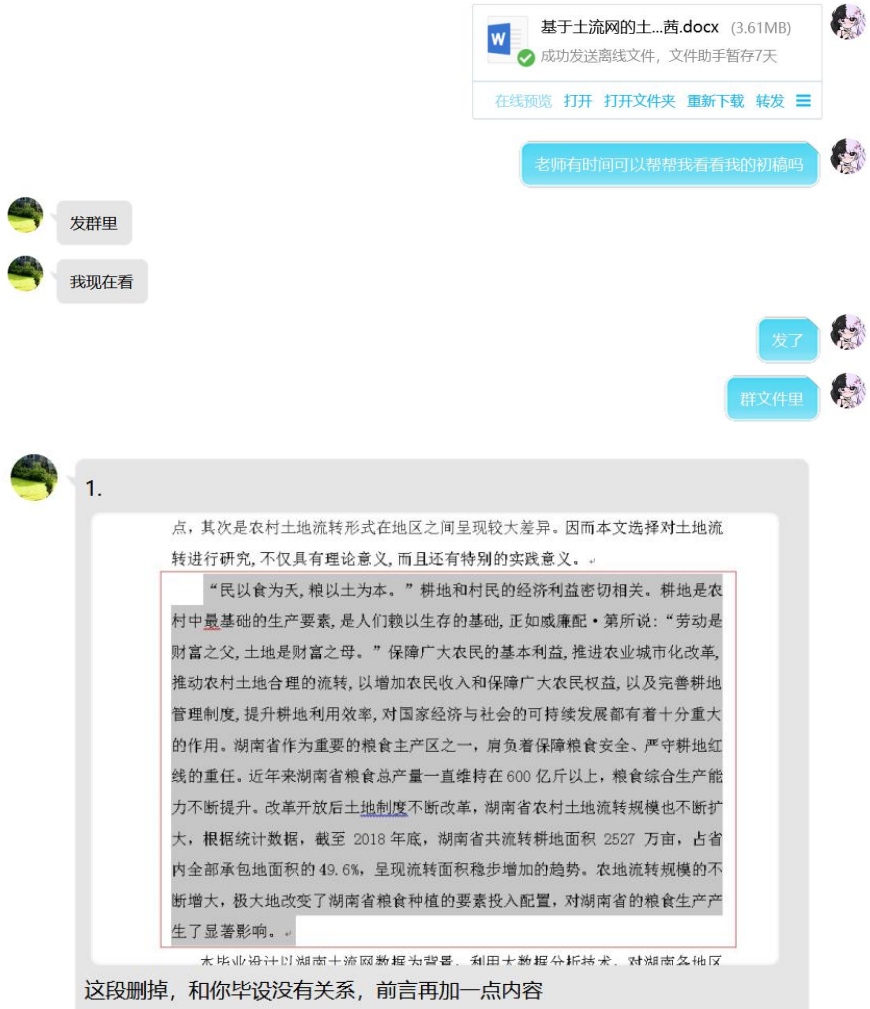
学生姓名	罗奥	学号	202165110019	指导教师	刘芳
二级学院	信息工程学院	专业	大数据技术	班级	大数据技术二班
毕业设计题目	基于智房网二手房数据分析与可视化报告				
过程性指导记录内容	说明每一次指导情况及提供原始支撑记录（如：QQ、微信、网络空间等互动截图） 毕业设计系统问题：1、 题目不一致 1、内容少，重写，2，没有对齐 23/9/25 15:53 4.7小结重写 参考文献至少十条  缩进两个字符！  图片模糊，建议替换				

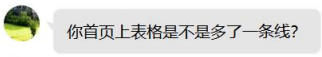
指导教师（签字）：刘芳

2024 年 5 月 21 日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	李欣茜	学号	202165110064	指导教师	曹清清

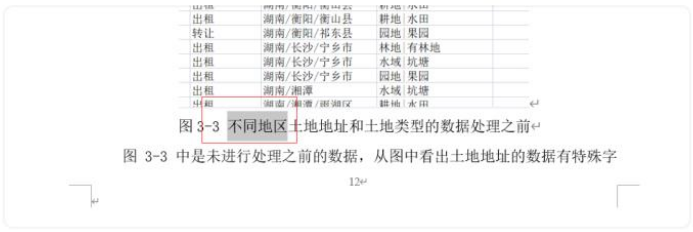
二级学院	信息工程学院	专业	大数据技术	班级	21 级大数据一班
毕业设计题目	基于土流网的土地流转数据分析与可视化报告				
过程性指导记录内容	说明每一次指导情况及提供原始支撑记录（如：QQ、微信、网络空间等互动截图） 2023 年 12 月 10 日 1、向指导老师发送毕业设计说明书初稿，请老师帮助检查并给出建议。老师查看后对初稿给出建议。 				



2023 年 12 月 15 日

2、初稿修改后，再次发送二稿请老师检查给予建议。老师查看后

再次给出建议。

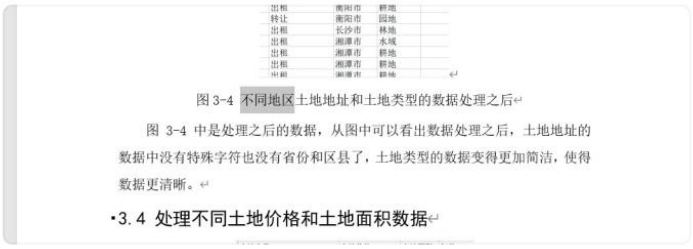


出租	湖南/衡阳/衡山县	耕地	水田
转让	湖南/衡阳/祁东县	园地	果园
出租	湖南/长沙/宁乡市	林地	有林地
出租	湖南/长沙/宁乡市	水域	坑塘
出租	湖南/长沙/宁乡市	园地	果园
出租	湖南/湘潭	水域	坑塘

图 3-3 不同地区土地地址和土地类型的数据处理之前

图 3-3 中是未进行处理之前的数据，从图中看出土地地址的数据有特殊字

12



出租	衡阳市	耕地	
转让	衡阳市	园地	
出租	长沙市	林地	
出租	湘潭市	水域	
出租	湘潭市	耕地	
出租	湘潭市	耕地	

图 3-4 不同地区土地地址和土地类型的数据处理之后

图 3-4 中是处理之后的数据，从图中可以看出数据处理之后，土地地址的数据中没有特殊字符也没有省份和区县了，土地类型的数据变得更加简洁，使得数据更清晰。

·3.4 处理不同土地价格和土地面积数据

安全。 获取正版 Office 了解详细信息

·3.4 处理不同土地价格和土地面积数据

土地名称	土地价格	土地面积	年份
其它林地出租	40元/亩/年	3180亩	40年
水田出租	320元/亩/年	3100亩	5年
畜牧养殖用地出租	40元/亩/年	3133亩	30年
水浇地转让	340元/亩/年	2000亩	5年
水田出租	420元/亩/年	5000亩	5年
旱地出租	430元/亩/年	4000亩	5年
天然牧草地经营权出租	600元/亩/年	200亩	5年
旱地经营权出租	550元/亩/年	2000亩	5年
水田经营权出租	500元/亩/年	3000亩	5年
旱地经营权出租	500元/亩/年	1000亩	5年
旱地经营权出租	500元/亩/年	1000亩	5年
人工草地经营权出租	680元/亩/年	300亩	8年
空闲用地出租	300元/亩/年	1800亩	20年
水田经营权出租	500元/亩/年	4000亩	5年
水田经营权出租	450元/亩/年	5000亩	5年
水田经营权出租	450元/亩/年	6000亩	5年
果园经营权转让	290元/亩/年	63.14亩	12年
有林地经营权出租	138元/亩/年	600亩	20年
坑塘经营权出租	6.8万元	110亩	10年
果园经营权出租	8万元	240亩	20年
坑塘经营权出租	200元/亩/年	10亩	10年

图 3-5 土地价格和土地面积数据处理之前

13

湖南网络工程职业学院毕业设计说明书

·3.8 按各类转让方式统计土地面积

使用权出租	32705.2
使用权转让	19874.71
出租	1039825.27
转包	7721.48
转让	18611.46

图 3-10 各类销售方式所占面积

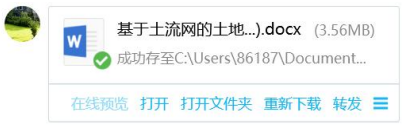
参考文献

[1] 周安娜. 农村土地流转的问题与策略探析，2022.08

[2] 董黎霞. 土地流转对农村经济发展的影响，2022.10

2023 年 12 月 21 日

3、老师帮我调整细节之后发我自己保存。

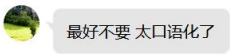


2024 年 1 月 2 日

4、字段不对，修改字段，再次发送请老师检查给予建议，老师检查完并给出建议。

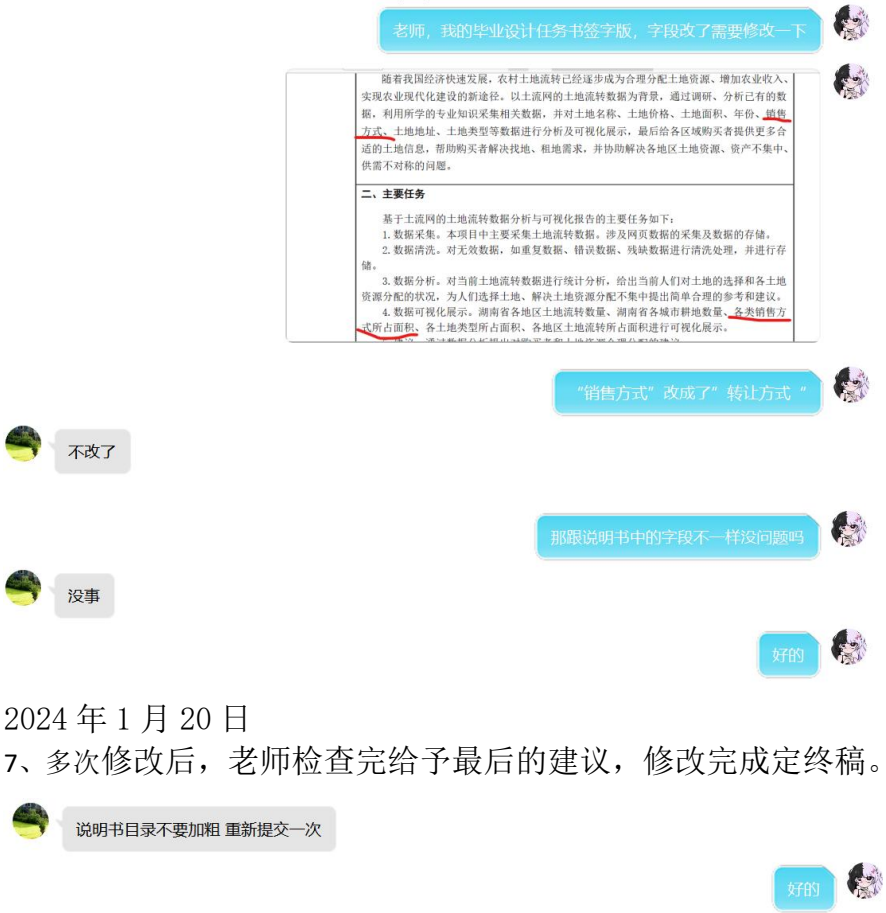
2024 年 1 月 7 日

5、询问毕业设计说明书的一些要求问题，成功得到了解决，并将修改完的毕业设计说明书再次发给老师检查。



2024 年 1 月 13 日

6、询问毕业设计任务签字书的字段问题，并成功得到了解决。



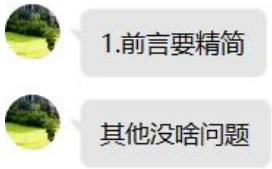
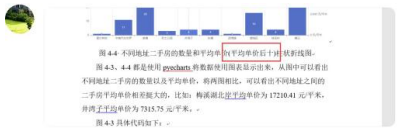
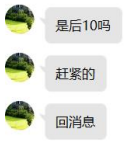
2024 年 1 月 20 日

7、多次修改后，老师检查完给予最后的建议，修改完成定终稿。

指导教师（签字）：曹倩倩

2024 年 5 月 10 日

湖南网络工程职业学院毕业设计教师指导记录表

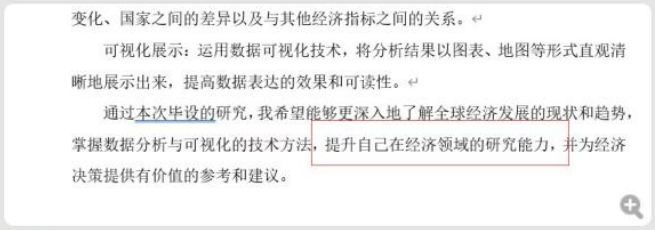
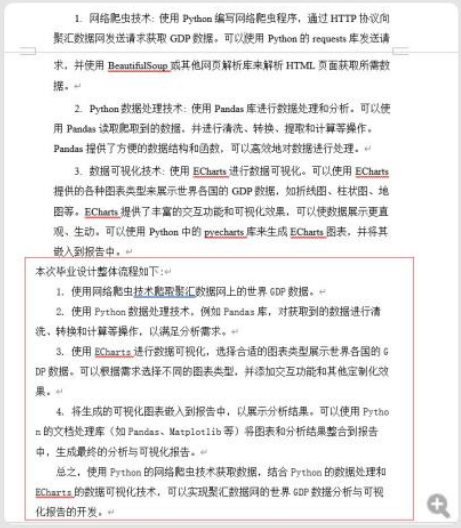
学生姓名	李晶	学号	202165110063	指导教师	曹清清
二级学院	信息工程学院	专业	大数据技术	班级	21 大数据技术 1 班
毕业设计题目	基于链家的长沙二手房数据分析与可视化报告				
过程性指导记录内容	1. 初稿修改 2. 第二稿修改  3. 第三稿修改   					

	4. 确定终稿

指导教师（签字）：曹清清

2024 年 5 月 10 日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	管智琛	学号	20216511059	指导教师	曹清清
二级学院	信息学院	专业	大数据技术	班级	21 级大数据一班
毕业设计题目	基于聚汇数据网的世界 GDP 数据分析与可视化报告				
过程性指导记录内容	1.  变化、国家之间的差异以及与其他经济指标之间的关系。 可视化展示：运用数据可视化技术，将分析结果以图表、地图等形式直观清晰地展示出来，提高数据表达的效果和可读性。 通过本次毕设的研究，我希望能够更深入地了解全球经济发展的现状和趋势，掌握数据分析与可视化的技术方法，提升自己在经济领域的研究能力，并为经济决策提供有价值的参考和建议。 能提升吗? 2.  1. 网络爬虫技术：使用 Python 编写网络爬虫程序，通过 HTTP 协议向聚汇数据网发送请求获取 GDP 数据。可以使用 Python 的 requests 库发送请求，并使用 BeautifulSoup 或其他网页解析库来解析 HTML 页面获取所需数据。 2. Python 数据处理技术：使用 Pandas 库进行数据处理和分析。可以使用 Pandas 读取爬取到的数据，并进行清洗、转换、提取和计算等操作。Pandas 提供了方便的数据结构和函数，可以高效地对数据进行处理。 3. 数据可视化技术：使用 ECharts 进行数据可视化。可以使用 ECharts 提供的各种图表类型来展示世界各国的 GDP 数据，如折线图、柱状图、地图等。ECharts 提供了丰富的交互功能和可视化效果，可以使数据展示更直观、生动。可以使用 Python 中的 pyecharts 库来生成 ECharts 图表，并将其嵌入到报告中。 本次毕业设计整体流程如下: 1. 使用网络爬虫技术爬取聚汇数据网上的世界 GDP 数据。 2. 使用 Python 数据处理技术，例如 Pandas 库，对获取到的数据进行清洗、转换和计算等操作，以满足分析需求。 3. 使用 ECharts 进行数据可视化，选择合适的图表类型展示世界各国的 GDP 数据。可以根据需求选择不同的图表类型，并添加交互功能和其他定制化效果。 4. 将生成的可视化图表嵌入到报告中，以展示分析结果。可以使用 Python 的文档处理库（如 Pandas、Matplotlib 等）将图表和分析结果整合到报告中，生成最终的分析与可视化报告。 总之，使用 Python 的网络爬虫技术获取数据，结合 Python 的数据处理和 ECharts 的数据可视化技术，可以实现聚汇数据网的世界 GDP 数据分析与可视化报告的开发。 这个流程是不是可以删除 与前言是不是重复的? 3. 大致浏览网站内容，确定需要爬取的字段有：排名 国家/地区 年份					

适合目标和用户群的表现方式。同样的数据可以呈现出多种不同的可视化形式，每种形式都有着独特的视觉效果。

•4. 2 根据每个国家的颜色来显示其 GDP 值

图 4-1 根据每个国家的颜色来显示其 GDP 值

在图 4-1 中使用颜色来表示每个国家的 GDP 值，可以直观地反映出各个国家之间的经济发展水平的差异和分布情况。较深的颜色通常表示该国家的

图片的无效区域去掉，让有效区域尽可能的大

11.

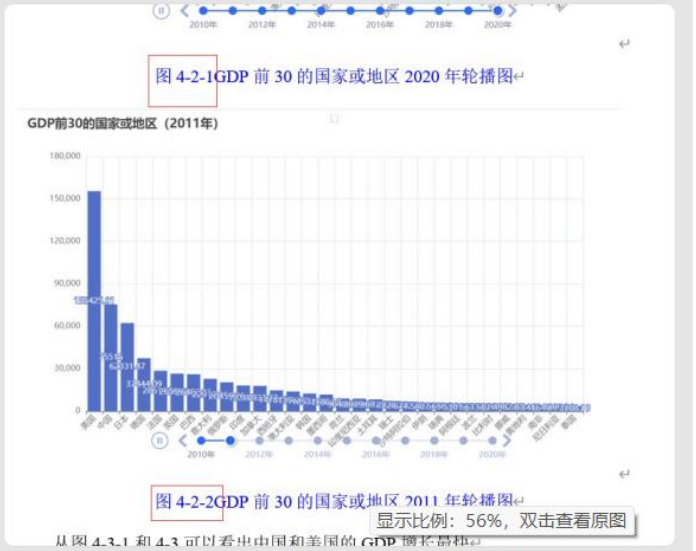


图 4-2-1GDP 前 30 的国家或地区 2020 年轮播图

图 4-2-2GDP 前 30 的国家或地区 2011 年轮播图

显示比例：56%，双击查看原图

从图 4-3-1 和 4-3 可以看出中国和美国的 GDP 增长情况

图片编号格式为 4-1 4-2 4-3 不要出现4-2-1，另外要改为黑色

12.参考文献格式要调整

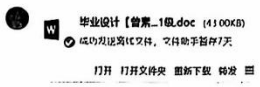
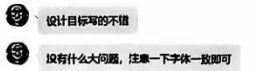
2023/10/12 8:58:03

	1. 变化、国家之间的差异以及与其他经济指标之间的关系。 可视化展示：运用数据可视化技术，将分析结果以图表、地图等形式直观清晰地展示出来，提高数据表达的效果和可读性。 通过本次毕设的研究，我希望能够更深入地了解全球经济发展的现状和趋势，掌握数据分析与可视化的技术方法，提升自己在经济领域的研究能力，并为经济决策提供有价值的参考和建议。 能提升吗? 2. 1. 网络爬虫技术：使用 Python 编写网络爬虫程序，通过 HTTP 协议向聚汇数据网发送请求获取 GDP 数据。可以使用 Python 的 requests 库发送请求，并使用 BeautifulSoup 或其他网页解析库来解析 HTML 页面获取所需数据。 2. Python 数据处理技术：使用 Pandas 库进行数据处理和分析。可以使用 Pandas 读取爬取到的数据，并进行清洗、转换、提取和计算等操作。Pandas 提供了方便的数据结构和函数，可以高效地对数据进行处理。 3. 数据可视化技术：使用 ECharts 进行数据可视化。可以使用 ECharts 提供的各种图表类型来展示世界各国的 GDP 数据，如折线图、柱状图、地图等。ECharts 提供了丰富的交互功能和可视化效果，可以使数据展示更直观、生动。可以使用 Python 中的 pyecharts 库来生成 ECharts 图表，并将其嵌入到报告中。 本次毕业设计整体流程如下: 1. 使用网络爬虫技术爬取聚汇数据网上的世界 GDP 数据。 2. 使用 Python 数据处理技术，例如 Pandas 库，对获取到的数据进行清洗、转换和计算等操作，以满足分析需求。 3. 使用 ECharts 进行数据可视化，选择合适的图表类型展示世界各国的 GDP 数据。可以根据需求选择不同的图表类型，并添加交互功能和其他定制化效果。 4. 将生成的可视化图表嵌入到报告中，以展示分析结果。可以使用 Python 的文档处理库（如 Pandas、Matplotlib 等）将图表和分析结果整合到报告中，生成最终的分析与可视化报告。 总之，使用 Python 的网络爬虫技术获取数据，结合 Python 的数据处理和 ECharts 的数据可视化技术，可以实现聚汇数据网的世界 GDP 数据分析与可视化报告的开发。 这个流程是不是可以删除 与前言是不是重复的?

指导教师（签字）：曾清雪

2024 年 5 月 10 日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	曾素琴	学号	202165110051	指导教师	马萌
二级学院	信息工程学院	专业	大数据技术	班级	一班
毕业设计题目	基于前程无忧网的招聘数据分析与可视化报告				
	1、11 月 6 日，指导毕业设计任务书   				

3、11 月 10 日毕业设计说明书初稿修改建议

1.

所属学院：↵ 信息工程学院↵
专　　业：↵ 大数据技术↵
班　　级：↵ 21 级大数据一班↵
姓　　名：↵ 曾素琴↵
学　　号：↵ 202165110061↵
指导老师：↵ 马萌↵
提交时间：↵ 2024-4-15↵

2.

▪2.1 项目数据获取思路介绍↵

根据项目所需要的数据和用于数据可视化处理的数据来寻找合适的网站进行爬取。大致浏览该网站的内容来分析所需要爬取以及进行可视化的数据，并且构思这些数据该如何进行分析。在网页中点击翻页查看不同页面之间的变化规律以便在设计代码时可以进行对网页的自动翻页。↵

这一段落字间距有问题

3.

7,5-7年,本科,合资 , 5000-10000人
,中技/中专,外资（非欧美） , 1000-500(

图 2-3↵

每一张图要有名称

4.

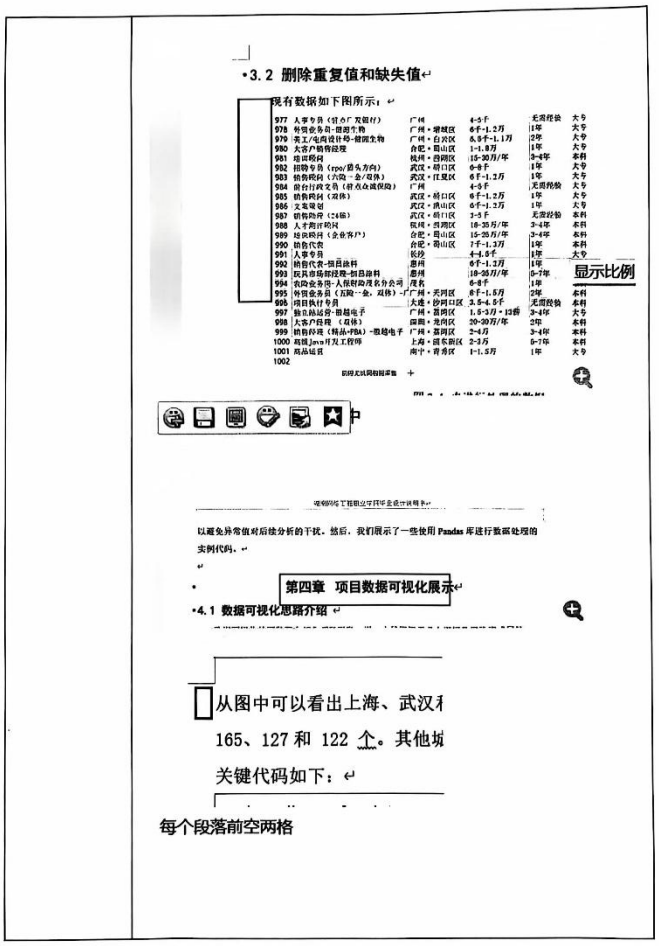
↵
↵
↵
↵

▪ 第三章 项目数据处理↵

▪3.1 数据处理思路介绍↵

在进行数据处理之前，首先需要进行数据清洗。之前在进行数据爬取和保存

每一章节的开头都要另起一页



•3.2 删除重复值和缺失值

现有数据如下图所示：

977	人事专员（社会广发招行）	广州		4-5千	无需经验	大专
978	外贸业务员-德国生物	广州·增城区		6千-1.2万	1年	大专
979	美工/电商设计师-德国生物	广州·白云区		6.5千-1.1万	2年	大专
980	大客户销售经理	合肥·蜀山区		1-1.8万	1年	大专
981	培训顾问	杭州·西湖区		15-30万/年	3-4年	本科
982	招聘专员（cpo/猎头方向）	武汉·硚口区		6-8千	1年	大专
983	销售顾问（六险一金/双休）	武汉·江夏区		6千-1.2万	1年	大专
984	前台行政文员（研点点滴保险）	广州		4-5千	无需经验	大专
985	销售顾问（双休）	武汉·硚口区		6千-1.2万	1年	大专
986	文案策划	武汉·洪山区		6千-1.2万	1年	大专
987	销售助理（24级）	武汉·硚口区		3-5千	无需经验	本科
988	人才测评顾问	杭州·西湖区		10-35万/年	3-4年	本科
989	培训顾问（企业客户）	合肥·蜀山区		15-25万/年	3-4年	本科
990	销售代表	合肥·蜀山区		7千-1.3万	1年	本科
991	人事专员	长沙		4-4.5千	1年	大专
992	销售代表-恒昌涂料	惠州		6千-1.2万	1年	
993	区负市场部经理-恒昌涂料	惠州		10-35万/年	5-7年	本科
994	农险业务员-人保财险茂名分公司	茂名		6-8千	1年	
995	外贸业务员（五险一金，双休）	广州·天河区		8千-1.5万	2年	本科
996	项目执行专员	大连·沙河口区		3.5-4.5千	无需经验	本科
997	独立站运营-跨越电子	广州·荔湾区		1.5-3万·13薪	3-4年	大专
998	大客户经理 （双休）	深圳·龙岗区		20-30万/年	3年	本科
999	销售经理（精品+BA）-跨越电子	广州·荔湾区		2-4万	3-4年	本科
1000	高级Java开发工程师	上海·闵东新区		2-3万	5-7年	本科
1001	商品运营	南宁·青秀区		1-1.5万	1年	大专
1002						

显示比例

以避免异常值对后续分析的干扰。然后，我们展示了一些使用 Pandas 库进行数据处理的实例代码。

第四章 项目数据可视化展示

•4.1 数据可视化思路介绍

从图中可以看出上海、武汉和
165、127 和 122 个。其他城
关键代码如下：

每个段落前空两格

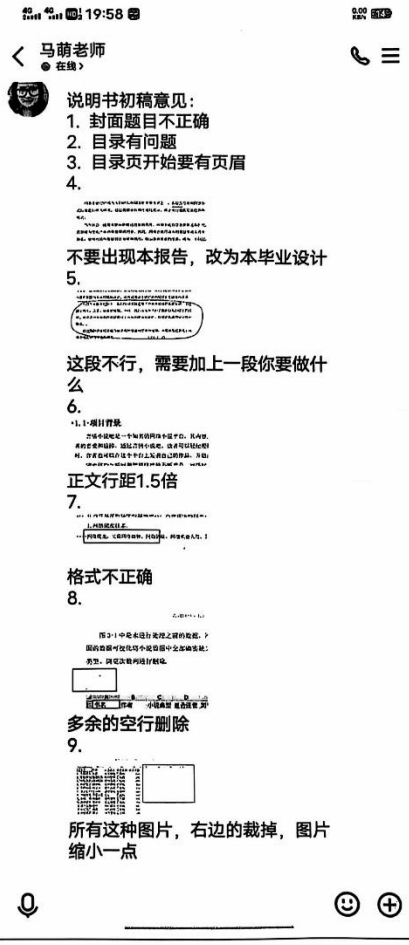
指导教师（签字）：

2020年 05月 10日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	龙晓宇	学号	202165110051	指导教师	马萌
二级学院	信息工程学院	专业	大数据技术	班级	21 级大数据一班
毕业设计题目	基于言情小说吧的数据分析与可视化报告				
过程性指导记录内容	1、9 月 4 日，指导毕业设计任务书 2、11 月 7 日指导毕业设计成果报告 				

3、11 月 10 日毕业设计说明书初稿修改建议



参考文献

[1]张理 案例教学法在 Python 语言程序设计教学中的应用. 计算 2021

[2]樊潮，秦婧 基于最新 Python 的 Excel 应用浅析. 计算机时代

[3]陈丛，周力恩 基于 Python 爬虫技术的虚假数据溯源与过滤. 真，2021

[4]张宝刚. 基于 Python 的网络爬虫与反爬虫技术的研究. 电子世

[5]李彦. 基于 Python 的网络爬虫技术的研究. 电子世界，2021

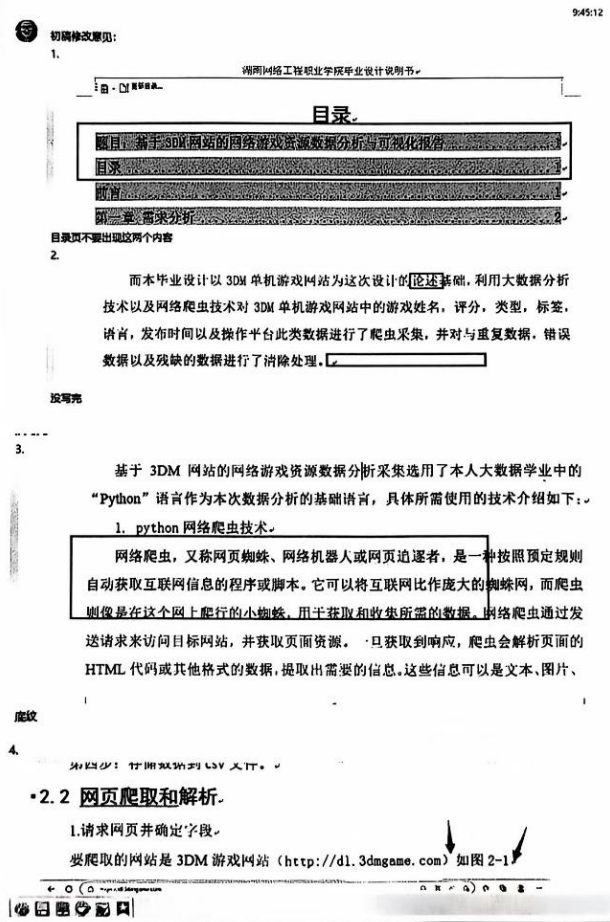
[6]吴文哲. 基于 Python 的服务器运维可视化研究与应用. 科技与

[7]我国游戏产业发展现状及对策研究_雷杨,2021

指导教师（签字）：

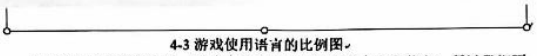
2024年 05月 09日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	黄靖	学号	202165110043	指导教师	马萌
二级学院	信息工程学院	专业	大数据技术	班级	21 级大数据一班
毕业设计题目	基于 3DM 网站的网络游戏资源数据分析与可视化报告				
过程性指导记录内容	说明每一次指导情况及提供原始支撑记录（如：QQ、微信、网络空间等互动截图）  9:45:12 初稿修改意见： 1. 湖南网络工程职业学院毕业设计说明书 目录 题目：基于3DM网站的网络游戏资源数据分析与可视化报告 2 目录 2 前言 2 第一章 需求分析 2 目录页不要出现这两个内容 2. 而本毕业设计以 3DM 单机游戏网站为这次设计的论述基础，利用大数据分析技术以及网络爬虫技术对 3DM 单机游戏网站中的游戏姓名，评分，类型，标签，语言，发布时间以及操作平台此类数据进行了爬虫采集，并对与重复数据，错误数据以及残缺的数据进行了消除处理。 没写完 3. 基于 3DM 网站的网络游戏资源数据分析采集选用了本人大数据学业中的“Python”语言作为本次数据分析的基础语言，具体所需使用的技术介绍如下： 1. python 网络爬虫技术 网络爬虫，又称网页蜘蛛、网络机器人或网页追逐者，是一种按照预定规则自动获取互联网信息的程序或脚本。它可以将互联网比作庞大的蜘蛛网，而爬虫则像是在这个网上爬行的小蜘蛛，用于获取和收集所需的数据。网络爬虫通过发送请求来访问目标网站，并获取页面资源。一旦获取到响应，爬虫会解析页面的HTML 代码或其他格式的数据，提取出需要的信息。这些信息可以是文本、图片、 底纹 4. 第四步：存储数据到 CSV 文件。 2.2 网页爬取和解析 1.请求网页并确定字段 要爬取的网站是 3DM 游戏网站（http://dl.3dmgame.com）如图 2-1				

缺标点符号

5.



4-3 游戏使用语言的比例图

从图中展示的数据来看，游戏中普遍的使用语言是中文和英文，所以我们不难分析出中国在游戏市场内的发展群体是庞大的。

操作代码如下：

格式问题

4-7 小结

本次章节进行的是数据可视化进程，是在基于前一章数据处理后的数据前提下，对数据来进行画图操作。数据视图化可以让复杂的数据变得简洁明了，让人更好的去理解和分析，也更容易让人理清数据之间的关系。

格式问题

4.4 基于游戏
4-5 基于各年
4-6 基于热度
4-7 小结

量还是与之有一定的差距。

5.从图 5-4 中可以看出，在热度前十的游

领先的。其他游戏的热度相比者说尽管不及，

的。

3.从游戏数据几年来的游戏数据中可以看出，现如今的游戏发行数量变的比起前几年的数量少了很多，但是可以花更多的时间去对游戏的品质精益求精，创造出更多富有价值的游戏。

4 从发布游戏的热度来看，很多评分高的，品质优秀的游戏反而热度比可以通过发布前的预热，广告的推销来积攒出自己的热度。让游戏的知名度起来。

总结

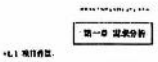
通过这次的毕业设计，我发觉到我有将自己曾经学习过的知识联系并运用起来，让我充分认识到大数据这门专业的实用性，也让我对自己学习过的知识得到了巩固和熟悉。在本次过程中，经过自己从错误的代码到正确运行和在错误中的摸爬滚打，我知晓了自己学习的还只是太片面了，很多知识还是得靠我自己去慢慢得探索，当然，有过本次的经历，我得到远远多于我付出了的，我看到了自己的不足，注意到了自己在做事情时的许多问题。

在这次过程中，我过的十分的充实，让我知道了劳动才能有所得，在一次次的翻阅资料和与同学的沟通中，我发现的所学过的知识又再一次得到了巩固，也让我学习到了很多我所没有学会或者没有学过的知识。

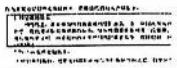
也是通过这次的毕业设计，我对自己的专业知识有了更深的了解和掌握，因此，我很感激能在这次毕业设计中的到马萌老师的指导与支持，在此特别感谢我的指导老师们。

正文宋体小四1.5倍行距

说明书中所有的我字修改一下，尽量不要出现

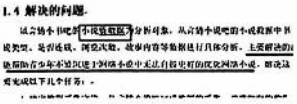


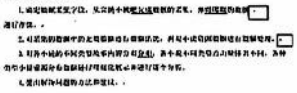
所有一级标题是黑体小二加粗　二级标题是黑体小三加粗



格式不对

正文是宋体小四1.5倍行距

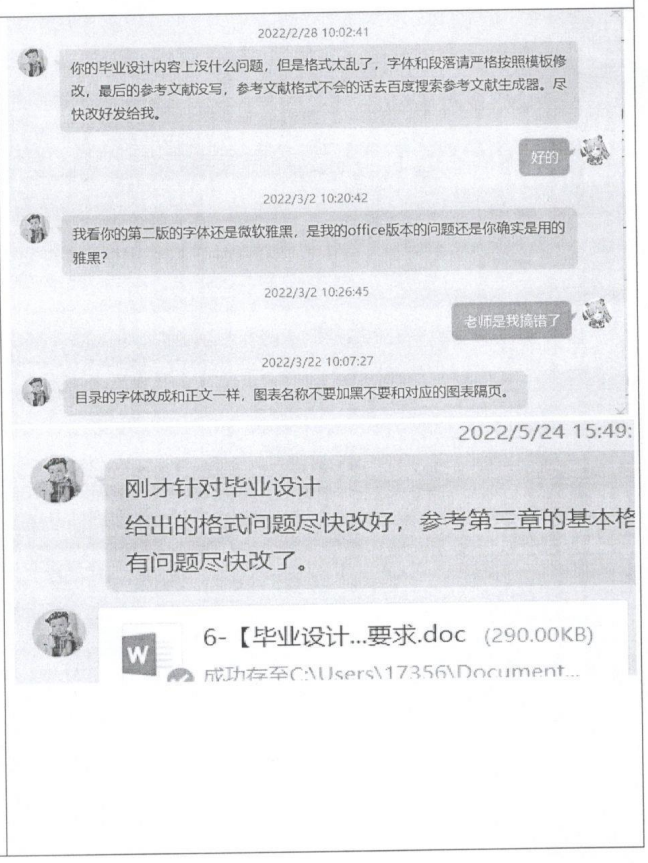




指导教师（签字）：

2020 年 05 月 09 日

湖南网络工程职业学院毕业设计教师指导记录表

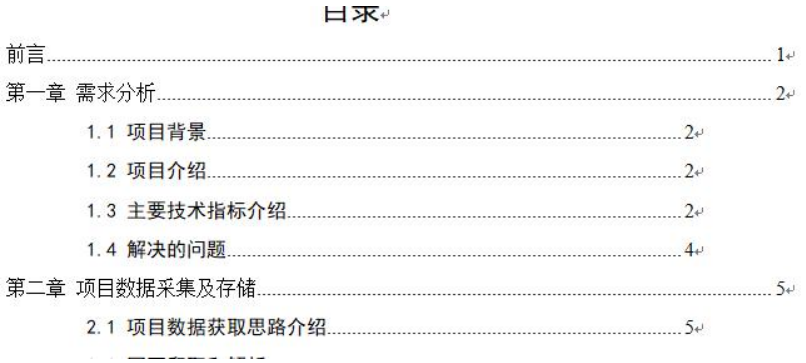
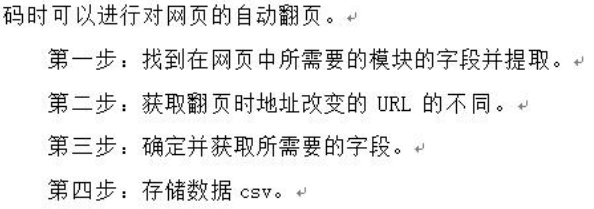
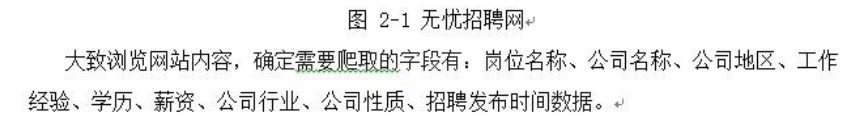
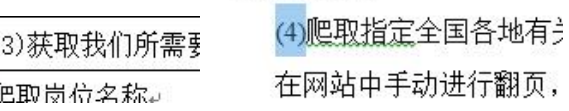
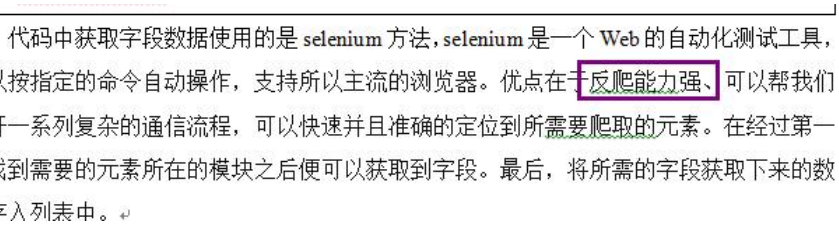
学生姓名	曹景琦	学号	202165110011	指导教师	宁可
二级学院	信息学院	专业	大数据技术与应用	班级	21 大数据二班
毕业设计题目	《基于上海黄金交易所的数据分析与可视化报告》				
过程性指导记录内容					

指导教师（签字）：

2023 年 12 月 17 日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	谢伟枝	**学号**	202165110031	**指导教师**	曹守富

二级学院	信息工程学院	专业	大数据技术	班级	21 大数据 2 班
毕业设计题目	基于无忧招聘网的计算机类招聘数据分析与可视化报告				
过程性指导记录内容	第一次主要指导内容 1、目录内容没有对齐。  2、这部分应结合自己实际项目，写的详细些。现在这种写法，放到哪个项目都可以，和项目基本没什么融合了，要结合自己的项目写。  3、各字段数据内容如何定位，要详细描述下。比如：使用 xpath，可以通过//@a，获取岗位名称。也就是要有对代码特征分析的过程描述，根据代码特征，如果来定位到相应字段的。  4、全文这种标号字体格式要一致。  5、项目本身在用爬虫技术，还反爬？应该是反反爬。  第二次主要指导内容。 1、语句不通顺，并且数据清洗内容及代码要放到第三章，这里只写存储。				

为了便于后期画图读取数据，将获取模块中我们所需要的字段并将对无
复数据、错误数据、残缺数据进行清洗处理，并直接存入到 csv 中。

全国各地有关互联网岗位数据处理并存储。

2、这个数据差距有点悬殊啊，河南岗位比广东还多么？好好检查下数据是否是合理的。

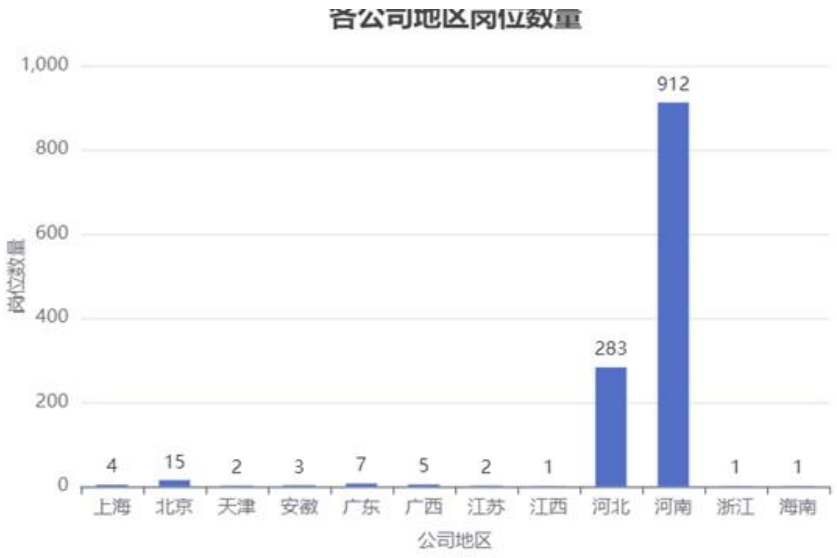


图 4-1 统计各公司地区岗位数量数据

3、语句不通顺。从……中，干……事，你这是从……的，干……事。

(1)确定数据采集字段，从无忧招聘网的招聘资源完成数据的采集，
取的数据进行存储。

(2)在获取到的数据中清除无效的数据，再对无忧招聘网的招聘资源
数据进行数据处理。

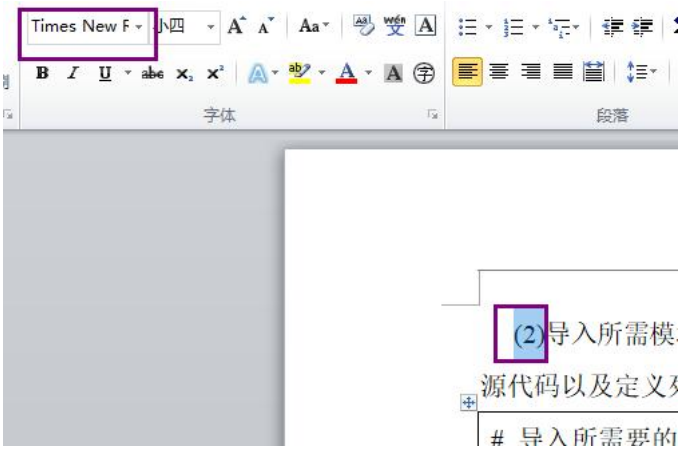
4、提出解决什么问题的方法和建议？要写详细点。

有关互联网岗位公司行业分布进行可视化展示并进行逐个分析。

(4)提出解决问题的方法和建议。

第三次主要指导内容

1、这种编号格式要一样，全文检查。要格式统一、规范。



2、语句不够通顺，一个内容说了两遍。

3. / 小结

本节主要是通过 Pandas 对爬取到的大量数据进行数据处理。由于大量数据较为混乱，而需对其进行数据处理。本节介绍了项目的数据处理思路，

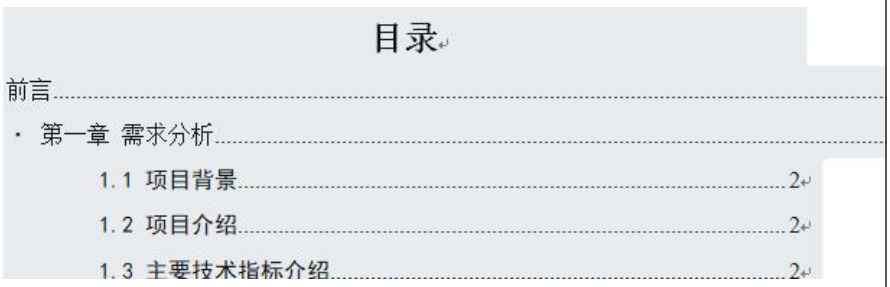
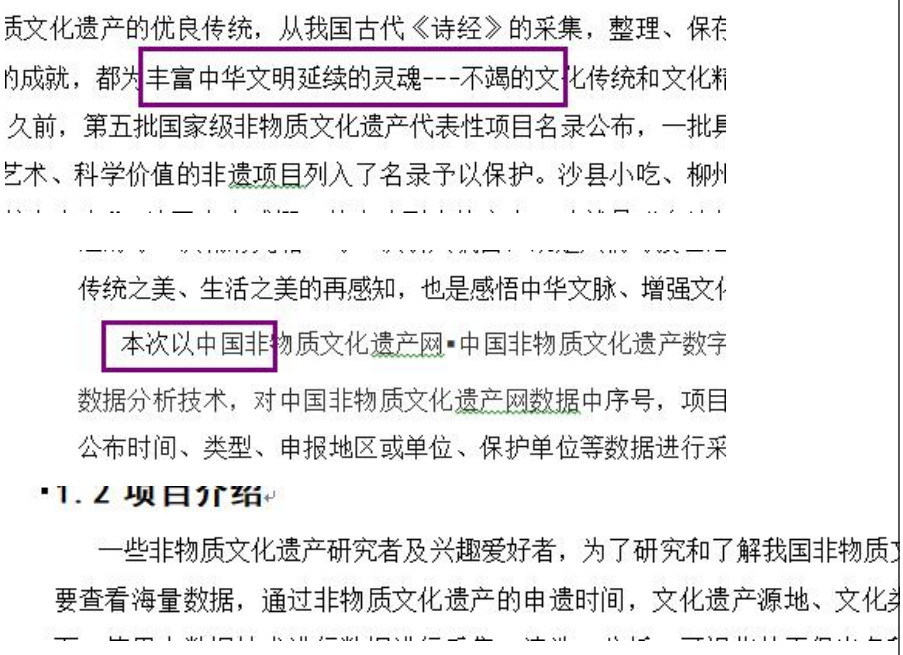
3、这里可以截个图，举个例子，重复的或者缺失的。

图 3-1 中是未进行处理之前的数据，刚获取到的各字段的数据有些是重复的、有些是缺失的，为了方便数据可视化，将数据中重复值和缺失值进行删除。

指导教师（签字）：曹宇富

2024 年 5 月 22 日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	杨炯健	学号	202165110033	指导教师	曹守富
二级学院	信息工程学院	专业	大数据技术	班级	21 大数据 2 班
毕业设计题目	基于中国非物质文化遗产网▪中国非物质文化遗产数字博物馆的数据分析与可视化报告				
过程性指导记录内容	第一次指导主要内容 1、目录内容没有对齐。  目录 前言 第一章 需求分析 1.1 项目背景 2 1.2 项目介绍 2 1.3 主要技术指标介绍 2 2、语句不通顺。  质文化遗产的优良传统，从我国古代《诗经》的采集，整理、保存 的成就，都为丰富中华文明延续的灵魂---不竭的文化传统和文化精 久前，第五批国家级非物质文化遗产代表性项目名录公布，一批具 艺术、科学价值的非遗项目列入了名录予以保护。沙县小吃、柳州 传统之美、生活之美的再感知，也是感悟中华文脉、增强文化 本次以中国非物质文化遗产网▪中国非物质文化遗产数字 数据分析技术，对中国非物质文化遗产网数据中序号，项目 公布时间、类型、申报地区或单位、保护单位等数据进行采 1.2 项目介绍 一些非物质文化遗产研究者及兴趣爱好者，为了研究和了解我国非物质文 要查看海量数据，通过非物质文化遗产的申遗时间，文化遗产源地、文化类 3、这种标号都去掉，后面还有，仔细检查。				

特别是民俗学建设的成就，都为丰富中
做出了贡献。[1]不久前，第五批国家级
重大历史、文学、艺术、科学价值的非
蛳粉的制作技艺“榜上有名”，让不少
遗”。截至目前，国家级非遗代表性项目
国家、省、市、县四级的名录体系，共
性项目名录体系建设的成果，也是中华
习近平总书记强调：“文化是一个国
族强。”[2]非遗从来不是孤芳自赏的“
个五年规划和 2035 年远景目标纲要》也提出，
与传承体系，[3]保护和利用好我国非物质文化
的全面、协调、可持续发展具有重要意义。加

4、海量去掉。

现以中国非物质文化遗产网为背景，通过调
识，通过分析统计海量数据并作可视化展示将
程度上更利于用户节省时间和高效阅读数据，

5、发起请求这块开始，和前面的内容没有一个过渡语句衔接，突然就出来一个“发起请求”了。这部分又自己语言重新写。

网络爬虫也称为“蜘蛛”，它可以在海量的互联网信息爬取需要的信息。简单地说它是模拟人类请求网站的行为，即自动请求网页、抓取数据，然后从中提取有价值的数据

[5]从功能上来讲，爬虫一般分为数据采集，处理，储存三个部分。发起请求:通过 HTTP 协议向目标站点发送请求（request），然后等待目标站点服务器的响应。获取响应内容:如果服务器能正常响应，会得到一个 Response。Response 的内容便是所要获取的页面内容，响应的内容可能有 HTML，Json 串，二进制数据（如图片视频）等等。解析内容:得到的内容可能是 HTML，可以用正则表达式、网页解析库进行解析;可能是 Json，可以直接转为 Json 对象解析;可能是二进制数据，可以做保存或者进一步的处理。保存数据:数据解析完成后，将保存下来。既可以存为文本文档、可以存到数据库中。

2.Python 数据处理技术

6、这部分应结合自己实际项目，写的详细些。现在这种写法，放到哪个项目都可以，和项目基本没什么关联了，要结合自己的项目写。

码时可以进行对网页的自动翻页。

第一步：找到在网页中所需要的模块的字段并提取。

第二步：获取翻页时地址改变的 URL 的不同。

第三步：确定并获取所需要的字段。

第四步：存储数据 csv。

7、格式有问题，跑到图下面了。

图 2-1。

第二次指导主要内容

1、各字段数据内容如何定位，要详细描述下。比如：使用 xpath，可以通过//@a，获取岗位名称。也就是要有对代码特征分析的过程描述，根据代码特征，如果来定位到相应字段的。

大致浏览网站内容，确定需要爬取的字段有：序号，项目序号、编号、名称、公布时间、类型、申报地区或单位、保护单位数据。

2、xpath 不叫爬虫模块。

2.4 小结

本项目使用的是爬虫模块 XPATH，Xpath 称为 XML 路径语言 Language。简单来说，Xpath 使用路径表达式在文档中进行导航以

3、格式错误，文字居中了，全文检查格式问题。

图 3-2 进行处理之后的数据图

从图中可以看出，与进行数据处理之前的数据相比，现在的数据中，列名为‘序号列已不存在。

·3.3 将数据中列名为申报地区或单位修改为申报省份或地区

4、分两列显示，一列太长了。

	  图 3-4 对各地区的申报非遗数量进行统计 5、小结内容太粗略，请围绕项目实际描述详细。 ·3.9 小结 本节使用到 pandas 处理数据，介绍了项目数据处理的思路，讲述了数据 数据处理的内容，并展示了相关的部分代码和图片。 6、你后面可视化代码用的是 pyechart，不是 js 的 echart，这里描述有问题。这段要好好修改。弄清楚什么是 echart、pyechart。 本次毕业设计用的是 ECharts，ECharts 是一个纯 Javascript 的图 直观，生动，可交互，可高度个性化定制的数据可视化图表。其创新的 据视图、值域漫游等特性大大增强了用户体验，赋予了用户对数据进行 7、这句话感觉是没说完，对齐也有问题，检查全文格式。 4-4 中国非物质文化遗产地区数量对比图 由图可知我国各地区的非物质文化遗产数量 关键代码如下： from pyecharts import options as opts from pyecharts.charts import Bar 8、都没有句号，内容需要进一步描述的细致一些。

根据可视化的结果分析如下：

1、从图 4-1 中可以从图中可以看出我国非物质文化遗产的空间分布，主要居多的地区为新疆，云南，湖北，山东，浙江

2、从 4-2 图中可以看出我国不同年份申请非物质文化遗产的时间变化，2008 年为申遗数量最多的年份

3、从图 4-3 可以得出我国申请非物质文化遗产的年份，2008 年占比最多，最少的是 2021 年

4、从图 4-4 可以得出由图可知我国各地区的非物质文化遗产数量

5、从图 4-5 可得出我国非物质文化遗产分类，主要是以传统音乐类别占比多。

第三次指导主要内容

1、提交时间 2024 年。

指导老师：曹守富

提交时间：2023-4-21

2、目录内容没有对齐。

目录

前言

· 第一章 需求分析

1.1 项目背景 2

1.2 项目介绍 2

1.3 主要技术指标介绍 2

3、这个图字符都看不清。

图 2-3 控制台具体字段图

4、这和前面出现的内容有重复的啊。

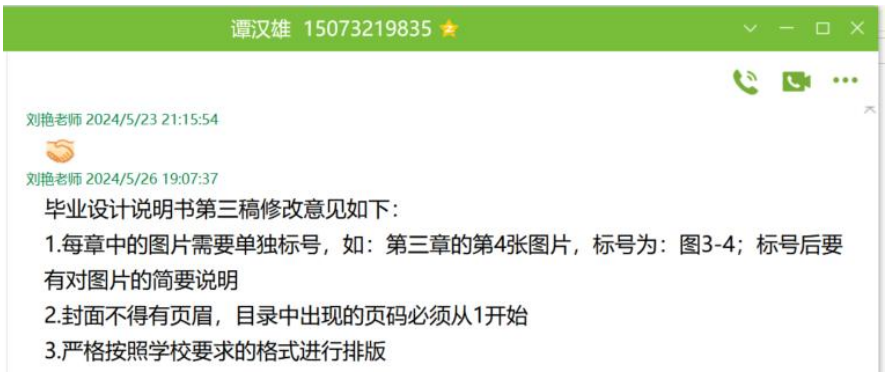
	·2.4 小结 本项目是通过 Xpath 去爬取所需字段数据，Xpath 称为 XML 路径语言，它的 XML Path Language。简单来说，Xpath 使用路径表达式在文档中进行导航以获取档中的节点或者节点集。路径表达式和我们在电脑文件系统中看到的路径表达式似。XPath 包含一个标准函数库使得它具有非常强大的选择功能，该库含有 100 个数用于元素或节点的比较和处理。 通过 HTTP 库向目标站点发起请求，即发送一个 Request，请求可以包含 headers 等信息，等待服务器响应。得到的内容可能是 HTML，可以用正则表达式、析库进行解析。可以直接转为 Json 对象解析，爬取中国非物质文化遗产网序号，号、编号、名称、类别、公布时间、类型、申报地区或单位、保护单位等字段数存储到 csv 中，爬取的过程中要注意找到数据存放的接口，注意在爬各日期多页对 url 的处理。 5、语句不通顺。 ·3.9 小结 本节使用到 pandas 处理数据，Pandas 的数据类型主要有以下几种，它们分别是：Series，DataFrame，Panel，Panel4D，PanelND。其中 Series 和 DataFrame 应用的最为广泛，几乎占据了使用频率 90% 以上。并介绍了项目数据处理的思路，讲述了数据的预处理和数据处理的内容，先对数据进行清洗，将数据中的无效数据，如错误数据、残缺数据、重复数据进行清洗处理。然后再根据要做的可视化展示对数据进行分组、统计、排序等操作便于下一步的可视化展示操作。在对数据进行处理的同时，也对之前的知识 6、文字格式对齐方式错误。 4-1 中国非物质文化遗产地区分布地图 从图中可以看出我国非物质文化遗产的空间分布，主要居多的地区为新疆，湖北、山东、浙江关键代码如下： from pyecharts import options as opts

指导教师（签字）：曹守富

2024 年 5 月 22 日

湖南网络工程职业学院毕业设计教师指导记录表

<table>
<tr><td>学生姓名</td><td>谭汉雄</td><td>学号</td><td>202165110001</td><td>指导教师</td><td>刘艳</td></tr>
<tr><td>二级学院</td><td>信息工程学院</td><td>专业</td><td>大数据技术</td><td>班级</td><td>21 大数据二班</td></tr>
<tr><td>毕业设计题目</td><td colspan="5">基于飞卢小说网的小说数据分析与可视化报告</td></tr>
<tr><td>过程性指导记录内容</td><td colspan="5">说明每一次指导情况及提供原始支撑记录（如：QQ、微信、网络空间等互动截图）
</td></tr>
</table>



指导教师（签字）：刘艳

2024 年 5 月 30 日

湖南网络工程职业学院毕业设计教师指导记录表

学生姓名	朱振阳	**学号**	2021651100010	**指导教师**	刘艳
二级学院	信息工程学院	**专业**	大数据技术	**班级**	21 级大数据二班
毕业设计题目	基于苏州看房网的二手房数据分析与可视化报告				
过程性指导记录内容	说明每一次指导情况及提供原始支撑记录（如：QQ、微信、网络空间等互动截图） 				

指导教师（签字）：刘艳

2024 年 5 月 30 日