湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">全国二手房房价数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>胡晓峰</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550109</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：通过什么去重？
利用 Pandas 库清洗数据。

问题二：数据来源？数据量是多少？
贝壳二手房网；
数据量 17 万。

问题三：：数据清理后有多少？一共有多少个字段？
清洗完后数据量 15 万清洗了空字段和不符合格式的字段；
一共有 9 个字段分别是 ID、楼层高度、层数、年建、户型、占地面积、朝向、总价（万）、价格（平方米）。

问题四：一共画了几种图，从什么方面分析？
一共画了 5 个图分别是柱状图、散点图、折线图、饼图、南丁格尔玫瑰图；
分别从房价高低、建房年份、平均价、户型分析的。

会议主持人：唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：80　　答辩小组（组长）签名：　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">Python 图书管理系统</td></tr>
<tr><td>姓　　名</td><td>李萬豪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550114</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题一：用户模块的查询功能是模糊查询还是精确查询
答：精确查询 。
问题二：用户模块有什么功能？
答：用户模块有登陆、注册、查询、借书和还书功能。
问题三：管理用户功能具体怎么管理？
答：管理员到管理用户接界面进行管理，可对用户进行查询，可查看用户的具体信息，以及对用户进行注销。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

成绩： 65　　答辩小组（组长）签名：　　2021 年 11 月 4 日日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">贪吃蛇游戏的设计与实现</td></tr>
<tr><td>姓　　名</td><td>彭丞霄</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大　数　据 1901</td><td>学号</td><td>201927550118</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="5">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：（1）代码为多少？
答：100 多行
（2）主要用了什么？
答：主要用了 payame 第三方库 hile 循环和多个 if 条件语句实现，图形界面设计来辅助完成。
（3）是否使用数据库。是否有存档功能？
答：没有使用数据库，代码直接运行。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11　月　4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

成绩：　65　　答辩小组（组长）签名：　　2021 年 11　月　4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">app 软件数据分析可视化</td></tr>
<tr><td>姓　　名</td><td>秦汉文</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大　数　据 1901</td><td>学号</td><td>19</td></tr>
</table>

指导教师	罗伟	所在部门	网络空间安全学院	职　务（职　称）	讲师
答辩小组成员	姓名	职务（职称）			
	蒋亮	讲师			
	罗伟	讲师			
	唐紫珺	讲师			

答辩记录：
问题一
本次毕业设计使用到了哪些软件？
本次实验主要开发工具使用了 pycharm 程序测试工具使用了 selenium 数据清洗组件 kettle
问题二
本次实验在哪里爬的数据？
历趣网
问题三
本次实验一共爬了多数数据？清洗之后还剩多少数据？
本次实验一共爬取了 12000 条数据，清洗之后还剩下 8000 条数据

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11　月　4 日

答辩小组意见（评语）：

成绩：70　　答辩小组（组长）签名：　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

课题名称	图书管理系统设计与实现						
姓　　名	粟俊威	院（部）	网络空间安全学院	专业/班级	大　数　据 1901	学号	20
指导教师	罗伟	所在部门	网络空间安全学院	职　务（职　称）	讲师		
答	姓名	职务（职称）					

<table>
<tr><td rowspan="3">辩
小
组
成
员</td><td>蒋亮</td><td>讲师</td></tr>
<tr><td>罗伟</td><td>讲师</td></tr>
<tr><td>唐紫珺</td><td>讲师</td></tr>
<tr><td colspan="3">答辩记录：
1.怎么实现对数据的增删改查。
答：普通用户对图书只有查找和借书、还书功能，只有管理员才有对图书的增删改查的权限，管理员对 Gui 界面进行操作。

2.注册界面中的电话号码是否能为空
答：根据每个人唯一号码确定信息不重复，所以电话号码不能为空

3.登录界面是否区分用户与管理员
答：是的，在主界面选择登录模式后才能进行登录

会议主持人： 唐紫珺 记 录 人：蒋亮 2021 年 11 月 4 日</td></tr>
<tr><td colspan="3">答辩小组意见（评语）：

成绩：65 答辩小组（组长）签名： 2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">炸弹人游戏的设计与实现</td></tr>
<tr><td>姓　　名</td><td>王金河</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550123</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
1 问有么有设置关卡？设置了几关？
答：设置了关卡，总共设置了两关
2..这两个关卡有什么区别？区别在哪里？
答：难度增加了，地图也发生了变化，关卡 2 地图更大，难度也更大，而且随机生产的回血物品也减少了。
3.你用了那些模块，这些模块的作用是什么？
用了 pygame、random、sys，
pygame 是用来写游戏的模块集合
random 这是一个随机数发生器，在游戏中主要用来随机生成补给。
sys 主要用来从程序外部获取参数的桥梁。
会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：70　　答辩小组（组长）签名：　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的书籍数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>王鹏杰</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>24</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题一：数据采集过程中怎么获取的数据？
答：用 requests 库的 get 方法获取网页，然后再用复制 xpath 获取数据。

问题二：可视化做了多少个图？分别展示了什么？
答：可视化做了 6 个图，展示了前十书籍类型的人气占比图、前四出版社的人气占比图、是否完结的人气的横向柱状图、人气前五作者的柱状图、前四各个出版社的书籍数量折线图等。

答案三：kettle 做了什么操作？缺失数据是怎么处理的？
答：处理无用列、处理重复数据、处理缺失数据、字符串类型转换等的操作，缺失数据是用的过滤检查是否为空值，如果为空就丢弃改条数据。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

成绩：80　　答辩小组（组长）签名：　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">五子棋人机对战游戏</td></tr>
<tr><td>姓　　名</td><td>向阳</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>26</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
用到了那些库？
Sys，pygame,randon
代码有多少行？
500 多
有那些功能？
简单的人机对战，可以在对战中重新开始

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

成绩：60　　答辩小组（组长）签名：　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">疫情防控数据的可视化</td></tr>
<tr><td>姓　　名</td><td>颜文和</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>28</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
用到了什么库？
使用了 lxml、pymsql 等库
爬了什么数据？
爬了百度的疫情事实大报告
画了几个图？
折线图、排名图、横向柱状图等。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

成绩：60　　答辩小组（组长）签名：　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">坦克大战游戏设计与实现</td></tr>
<tr><td>姓　　名</td><td>阳志祥</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550129</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题 1：游戏能否进行存档？
答：游戏目前没有数据库来进行存档。
问题 2：游戏是否有关卡？多少关？
答：游戏有关卡设计，但目前只有两关。
问题 3：游戏采用了什么库？
答：游戏用 python 语言编写，采用了 pygame 库。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：60	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的植物大战僵尸小游戏设计与实现</td></tr>
<tr><td>姓　　名</td><td>叶频</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901 班</td><td>学号</td><td>201927550131</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 游戏是如何实现的，设计了哪些细节？

包括植物的攻击力，防御力、向日葵产生的阳光速率和防御力、僵尸的攻击力和防御力等一系列基础设定，游戏背景和游戏输赢的设计。

2. 为什么要选择 python 来实现？

因为首先 Python 是面向对象的程序语言，可以比较方便的封装游戏中的各个角色属性以及动作，其次 python 开发迭代速度快。

3. 请详细描述如何用 python 实现？

（1）设定游戏背景：利用 Python 定义函数，设定地图的长、宽大小，进行初始化地图。

（2）设定植物属性：利用 python 定义植物类，进行植物属性的设定，包括样式、攻击力、防御力等一系列属性。

（3）设定僵尸属性：利用 python 定义僵尸类，进行僵尸属性的设定，包括样式、防御力、攻击力等一系列属性。

（4）设定计算函数：利用 python 定义计算类，进行计算植物的死亡和对僵尸的攻击产生到的伤害已经僵尸的死亡和对植物的伤害，并且控制游戏输赢等一系列计算过程。

（5）设定主函数：利用 python 定义开始游戏函数，进行调用地图、植物类、僵尸类、计算函数等进行整个项目的运行，保证运行能够正常运行。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：60	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">有声读物网页的数据采集与分析</td></tr>
<tr><td>姓　　名</td><td>俞强</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901 班</td><td>学号</td><td>33</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：采集到的数据量是多少？
答：从喜马拉雅网页上采集到一万三千条数据。
问题二：采集的数据类型是什么？
答：作品名称、作者 id、作品收听量、作者粉丝数、作者作品数、作者作品条数。
问题三：工程代码总量是多少？
答：代码工程总量大致为两百条。
问题四：数据清洗工具是什么？
答：使用的是 kettle 中的 spoon 组件。
问题五：清洗过程中进行了那些步骤？
答：连接 MySQL，设置表输入、连接过滤记录、连接处理字符串、执行 SQL 语句、进行表输出结束。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：70	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的 B 站视频弹幕的数据分析与实现</td></tr>
<tr><td>姓　　名</td><td>曾圣轩</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550135</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：爬取了多少部电影和弹幕数据
答：目前只有 1 部电影，一万条弹幕数据，后续可增加。
问题二：爬取的弹幕有哪些字段
答:，弹幕出现时间、弹幕类型、弹幕颜色、发送时间戳、弹幕 ID、发送者 ID、
问题三：所用的 jieba，snownlp 有什么作用
答：jieba 由于中文分词，snownlp 用于情感分析。

会议主持人： 唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：65　　答辩小组（组长）签名： 罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于苏宁的手机数据的分析与可视化系统</td></tr>
<tr><td>姓　　名</td><td>曾宪源</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550136</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 做这个选题的原因？

答：提供一个可供参考的大数据岗位数据的可视化系统

2. 用了哪些库？

答：requests、pandas、numpy、matpolib

3. 数据采集了那些字段

答：手机品牌、手机处理器、手机内存、运行内存

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：65	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 51job 的大数据岗位数据的分析与可视化系统</td></tr>
<tr><td>姓　　名</td><td>朱红伟</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550139</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

4. 做这个选题的原因？

答：提供一个可供参考的大数据岗位数据的可视化系统

5. 用了哪些库？

答：requests、pandas、numpy、matpolib

6. 数据采集了那些字段

答：职位名称、公司名称、薪资待遇、企业类型、行业、企业资金类、公司人数

会议主持人：唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：60　　答辩小组（组长）签名：罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于大数据技术的波士顿民宿数据分析系统</td></tr>
<tr><td>姓　　名</td><td>段丽洁子</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550140</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 爬取的是哪里的数据

美国 Data ju 网站获取的数据

2. 做了那些

爬取数据、存储数据、分析数据、绘图

3. 绘了那些图

词云图、柱状图、饼图、折线图

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：60　　答辩小组（组长）签名： 罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Java 的俄罗斯方块游戏开发</td></tr>
<tr><td>姓　　名</td><td>刘莉</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550142</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

7. 做这个选题的原因？

答：因为从小就玩，现在长大了，觉得实现这个游戏不难，就想着做出来。

8. 为什么选 Java 语言来做?

答：Java 语言应用范围广，我学的没有 Python 好，就像加深一下 Java 语言的学习。

3．你的做的这个游戏有没有什么特点？

答：游戏简单，益智，可操作强，是一个很好的休闲小游戏，而且我也有设置游戏难度，增加游戏方块样式，提高了游戏的趣味性。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：70	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的全国旅游数据的系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>屈怡洁</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1901</td><td>学号</td><td>201927550144</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 爬取的是哪里的数据

去哪儿网、国家统计局

2. 做了哪些

爬取数据、存储数据、分析数据、绘图

3. 画了那些图

横向柱状图、动态柱状图、中国地图、折线图、双柱状图

会议主持人：唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：60	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">安卓平台 2048 手游开发</td></tr>
<tr><td>姓　　名</td><td>盛丹</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550145</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题一：你为什么要做这个游戏？
答：喜欢玩游戏，想尝试一下自己能不能做出一款小游戏。
问题二：这个游戏有多少行代码？
答：有三百多行。
问题三：有没有认真测试游戏的可行性？
答：有找同学室友测试。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：</td></tr>
</table>

成绩：60	答辩小组（组长）签名：罗伟	2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">对《长津湖》电影热点影评的数据爬取分析与可视化的实现</td></tr>
<tr><td>姓　　名</td><td>汤雅青</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901 班</td><td>学号</td><td>201927550146</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

4. 在什么平台上面爬取的数据，一共爬取了多少条，清洗之后还剩多少条数据。
在猫眼电影官网爬取的数据，一共爬取了 7000 多条，清洗之后还剩 3000 多条数据。

5. 数据清洗步骤你主要是利用什么工具清洗的？一共绘制了多少种图像？
主要是利用 kettle 工具清洗数据的，一共绘制了八中图像。

6. 为什么选择这个课题？
在国庆期间，这部以爱国主义为主题的电影，在两周的时间内就突破了 45 亿票房，引起了很大的反响，通过研究这部电影的影评，了解本部电影的意义，让我们意识到了爱国教育非常重要，之后我们要拍更多具有正能量的意义的电影，也应该更加具有爱国意识。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：70	答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的笔记本电脑的数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>姚雨杰</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550148</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：python 爬虫遇到什么问题？有没有遇到反爬？怎么解决的？

答：爬取京东数据的时候，不知道怎么找 URL 和 header，爬取的时候爬取次数多了，就被反爬了，换了一个 IP 地址之后就可以继续爬取数据了。

问题二：爬取了多少数据？清洗之后还有多少数据？用了什么字段？

答：爬取了 1 万 6 千条数据，清洗完成之后还有 1 万 1 条数据，使用了 Varchar 字段。

问题三：爬取的什么网站数据？

答：爬取的京东网站的笔记本电脑数据。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：75　　答辩小组（组长）签名：　罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">Bilibili 数据获取与可视化</td></tr>
<tr><td>姓　　名</td><td>曹炜</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据技术与应用1901班</td><td>学号</td><td>201927550101</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>大数据与云计算教研室</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 爬取了什么类型的数据？

答：爬取了 Bilibili 排行榜排名和各视频数据的弹幕，排行榜属性有：作者、热度、播放量、评论数和视频链接，视频弹幕数据的属性有：弹幕发送时间、弹幕在视频中的时间戳、弹幕所属者、弹幕文本信息等。

2. 用了什么库？

答：爬虫使用了 urllib 和 requests，数据存储使用了 PyMySQL 连接数据库，数据可视化使用了 PyEcharts 进行可视化。

3. 分析了什么属性？

答：弹幕文本情绪、弹幕高频词汇、各时间段弹幕发送数量、弹幕在视频各时间中数量。

会议主持人：唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：73　　答辩小组（组长）签名：罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的显卡和 CPU 处理器的各类数据分析系统与实现</td></tr>
<tr><td>姓　　名</td><td>陈波</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550102</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：爬取什么网址的数据，什么类型的数据？

答：京东商场，价格、销量、店铺，店铺的 ID 类型等等
问题二：数据的量是多少？，清洗完后的数据有多少？
答：数据量在 5w 条左，清洗完后可用数据量在 2w 左右。
问题三：可视化的成果类型有几类图？
答：有五类，分别为饼图，柱状图，词云图，折线图，双柱状图。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：　68　　答辩小组（组长）签名：　罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 爬虫的相机销售数据分析</td></tr>
<tr><td>姓　　名</td><td>陈文杰</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550104</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：你用了数据仓库吗？
答：没有使用数据库
问题二：你爬取了多少条数据？
答：6 万左右
问题三：你爬取的数据是什么网址的，爬取了什么内容。
答：爬取的是京东电商平台上的相机商品，爬取了它的价格，名字，销售量，和详细网址。
问题四：怎样清洗数据的？
答：先使用 pandas 将商品分类分类，再使用 kettle 把同一类商品，销售量为空的给清理出去，然后使用 python 中的函数保存，使用继承在调用想要的数据。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：　67　　答辩小组（组长）签名：　罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Requests 的网上笔记本销售数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>段俊如</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550106</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：爬取什么网站数据，数据信息是什么？
答：爬取京东网，信息为笔记本信息。

问题二：利用什么工具进行爬取数据？
答：利用 Python 的 Requests 库爬取。

问题三：爬取数据量有多少？清洗后有多少？
答：爬取数据量 3 万，清洗数据后为 1 万。

问题四：可视化图的类型？
答：柱状图、双柱状图、词云图。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：65　　答辩小组（组长）签名： 罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">北上广租房信息的数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>关书富</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550108</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：哪里来的数据？
答：链家进行采集的数据

问：爬了多少数据？用什么方法提取数据？
答：爬了 10 万+的租房信息数据，通过 selenium 与 requests 进行数据的采集，使用 xpath 进行页面数据的提取，通过 pymysql 进行数据库数据的提取

问：几个类型的图？
答：用使用了四个类型的图，柱形、折线、饼图、柱形折线交互图

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：　63　　答辩小组（组长）签名：　罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">京东商品历史价格离线分析与可视化</td></tr>
<tr><td>姓　　名</td><td>黄广发</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550110</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题一，字段处理处理了那些字段，用了什么方法？
答：主要处理了商品名称中的特殊符号，使用的正则表达式 。
问题二，爬取了几个字段？
答：两个。
问题三，爬取数据时碰见什么问题了？
答：数据爬取速度不够快，在使用多线程时碰见了错误，数据字段获取时有些页面数据缺失
问题四，画了那些类型的图？
答：折线图，饼图等。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

答辩合格

成绩：72　　答辩小组（组长）签名： 罗伟　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 爬虫的直播市场可视化分析</td></tr>
<tr><td>姓　　名</td><td>鞠志成</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据技术与应用</td><td>学号</td><td>201927550111</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：爬取什么网址的数据？什么类型的数据？
答：爬取了斗鱼跟虎牙的数据。类型有：主播名称，时间，游戏名称，活跃观众，弹幕数量，弹幕人数，礼物总量，峰值热度，礼物人数。
问题二：爬取了多少数据？
答：总共爬取了大概 13 万条数据。
问题三：用了那些第三方库？
答：pymysql 和 requests。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：65　　答辩小组（组长）签名：　罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">电商交易系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>李家鼎</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550112</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题一：系统是用什么语言编写的？使用的数据库是什么？
答：系统的后台是使用 Java 编写的，前端是使用了 jsp 技术。数据库使用的是 MySQL
问题二：这个电商系统的支持的最大并发数多少？
答：在笔记本上部署，经过测试，并发大概在 200 左右。
问题三;电商系统的支付系统退款功能有没有做事务性限制？
答:、做了事务性的限制的，保证支付资金的安全。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11　月　4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

答辩合格

成绩：　61　　答辩小组（组长）签名：　罗伟　　2021 年 11　月　4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">前程无忧招聘数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>李琳</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>13</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题一：数据清洗主要过程？数据的量是多少？
答：Kettle 连接数据库，清洗空值，写入数据库，总共爬取了二十万条数据。
问题二：提取解析数据用的是什么？
答：Json。
问题三;数据的存储是怎样的？
答：数据存储在 mysql 里面。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

答辩合格

成绩：67　　答辩小组（组长）签名： 罗伟　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于贝壳的全国房价数据的可视化分析系统</td></tr>
<tr><td>姓　　名</td><td>李希奇</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550115</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3"></td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：爬取了什么网站的数据，爬了什么数据？

答：爬取了贝壳网新房数据;主要爬取了该网站新房房价，楼盘面积，地址和总价格的数据。

问题二：数据量有多少？清洗后还有多少？

答:主要是用了 Scrapy 框架爬取了 23000 多条数据。清洗后还剩 18000 多条数据。

问题三：有几个可视化的图形？

答：1 地图(全国各地房价展示)2 柱状图 3 条形图 4 折线图 5 饼图

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：66	答辩小组（组长）签名：　罗伟	2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于爬虫的鞋类商品可视化报表分析</td></tr>
<tr><td>姓　　名</td><td>刘小瑞</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550117</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td></td><td>职 务（职 称）</td><td colspan="3"></td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题 1：爬取了哪些字段？数据量是多少？
答：字段：商品名称，商品价格，销售地，购买人数，评论人数，店名。
数据量：八万余条

问题 2：数据清洗的主要过程？
答：对商品购买人数的字段进行删减，然后和评论人数一起转变成整数类型。

问题 3：爬虫使用的第三方库？
答：re 库，requests 库

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：67　　答辩小组（组长）签名： 罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的 IT 行业职位需求分析与可视化</td></tr>
<tr><td>姓　　名</td><td>王方鸣</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>22</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td></td><td>职 务（职 称）</td><td colspan="3"></td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
1. 爬了多少数据，有哪些字段？
答：四万多条、职位名称、工资、公司、服务行业、学历、经验等。
2. 有没有用到数据库？
答：有，使用了 mysql。
3. 数据清洗做了什么清洗？
答：空值处理、特殊字符替换操作。

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

答辩合格

成绩：72　　答辩小组（组长）签名： 罗伟　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的书籍管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>邬自强</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550125</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td></td><td>职 务（职 称）</td><td colspan="3"></td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1，　　还有什么特点？

答：除了这些，没有别的特点了，以后会增加其他的东西。

2，　　用的什么图形界面，用了数据库吗？

答：用的是 python 的 GUI 界面，用的 MySQL 数据库

3，　　如果借书人没有还书，有没有什么机制？

答：没有，这个我还没有想到，是我考虑的不全面，不过后续我会慢慢改进。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：67　　答辩小组（组长）签名：　罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的学生成绩可视化管理系统</td></tr>
<tr><td>姓　　名</td><td>杨浩</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>30</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td></td><td>职　务（职 称）</td><td colspan="3"></td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 你的系统有几个用户

答：有两个用户，学生和老师。

2. 你的系统有没有连接数据库，老师删除学生，数据库里是不是也进行了删除。

答：系统连接了数据库，老师删除学生信息后，数据库里也进行了删除。

3. 老师在求及格率，会不会出现图形。

答：老师求及格率会出现两个图，分别是两个科目的及格率和不及格率饼图。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：64　　答辩小组（组长）签名：　罗伟　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于大数据技术的 MLB 数据分析系统</td></tr>
<tr><td>姓　　名</td><td>余锜</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550132</td></tr>
<tr><td>指导教师</td><td>蒋亮</td><td>所在部门</td><td>网络空间完全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：数据的量是多少？，清洗完后的数据有多少？

答：数据量在 10w 条左，清洗完后数据量在 2w 左右

问题二：有没有用到数据库？

答：有，使用了关系型数据库 MySQL。

问答三：数据清洗做了什么清洗？

答：空值处理、特殊字符替换操作。

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：62　　答辩小组（组长）签名：蒋亮　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于大数据的股票数据的分析与可视化</td></tr>
<tr><td>姓　　名</td><td>喻建文</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550134</td></tr>
<tr><td>指导教师</td><td>陈书奇</td><td>所在部门</td><td>网络空间完全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>陈书奇</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：爬取了什么网站的数据，爬了什么数据？

答：通过 tushare 财经接口爬取沪深近 10 年的股票涨幅，开收盘，总收益等数据

问题二：数据量有多少？清洗后还有多少？

答:爬取拿到了 75000 多条数据。清洗后还剩 18000 多条数据。

问答三：是否使用到数据库

答：刚开始是将数据保存再 CSV 文件，后面将清洗的数据放到了数据库中

会议主持人：唐紫珺　　记　录　人：陈书奇　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：66　　答辩小组（组长）签名：陈书奇　　2021 年 11 月 4 日

37 湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Web 的学生信息可视化管理系统</td></tr>
<tr><td>姓　　名</td><td>郑琪琪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201816414016</td></tr>
<tr><td>指导教师</td><td>陈书奇</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>陈书奇</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：系统安全怎么做的，制作的怎么样了？

答：目前还没有做系统安全，后续会补上。

问题二：权限怎么控制的

答：权限是管理员全局控制，然后控制学生与教师，教师又控制学生

问题三：老师选课管理可以管理什么？

答：添加学生的选课、退选与搜索课程

会议主持人：唐紫珺　　记　录　人：陈书奇　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：70　　答辩小组（组长）签名：陈书奇　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于微信小程序的服装管理公平台的开发与实现</td></tr>
<tr><td>姓　　名</td><td>周攀</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901 班</td><td>学号</td><td>201927550138</td></tr>
<tr><td>指导教师</td><td>陈书奇</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>陈书奇</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 系统的安全性如何保证？

依靠微信云开发数据库，数据安全由微信安全部门进行保障。

2. 有什么几个用户权限级别？

有三种用户级别，分别为普通员工用户、公司管理员和初始管理员

3. 管理员用户有那些功能？

有提交订单数据、查看订单数据、删除订单数据、查看员工数据等一系列功能

会议主持人：唐紫珺　　记　录　人：陈书奇　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩：74　　答辩小组（组长）签名：陈书奇　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">汽车销售数据分析系统与实现</td></tr>
<tr><td>姓　　名</td><td>刘爱娟</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550141</td></tr>
<tr><td>指导教师</td><td>陈书奇</td><td>所在部门</td><td></td><td>职　务（职 称）</td><td colspan="3"></td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>陈书奇</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 绘图几个以及大概的意义

总共五个图

（1）2021 年每月汽车总销量条形图：对于每个月的汽车销量数据进行分析与对比情况

（2）2021 年汽车类型总销量南丁格尔玫瑰图：三种汽车类型总销量的占比。轿车总销量占比 48.71%，SUV 总销量占比 47.22%，MPV 总销量占比 4.07%，从占比可得出轿车与 SUV 比 MPV 畅销。

（3）2021 年各车型总销量折线图：对于各车型总销量的分析，汽车各车型类型每个月的总销量趋势的对比分析情况

（4）2021 年 9 月畅销车型销量 TOP10：对 9 月畅销车型销量 TOP10 的分析情况

（5）2021 年最畅销销量柱状图：对于最畅销三种车型销量的分析。2021 年最畅销轿车-日产轩逸每月销量,2021 年最畅销 SUV-哈弗 H6 每月销量,2021 年最畅销 MPV-别克 GL8 每月销量情况

2. 数据清洗分析用的是 pandas。大致做了删除空值，重复值清洗

3. 多少数据，包含的字段

Python 爬虫爬取一万多数据，包含的字段分别为：cartype,date,name,sale_rank,price,sale.

会议主持人：唐紫珺　　记　录　人：陈书奇　　2021 年 11 月 4 日

答辩小组意见（评语）： 答辩合格 成绩：____67_____ 答辩小组（组长）签名： 陈书奇 2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的打砖块游戏的设计与实现</td></tr>
<tr><td>姓　　名</td><td>罗剑琦</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901 班</td><td>学号</td><td>201927550143</td></tr>
<tr><td>指导教师</td><td>陈书奇</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>陈书奇</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问题:一：在过程中遇到了什么问题，测试时有没有什么 bug？
答：在创建碰撞检测类时遇到了困难，不知道怎么通过代码检测小球、砖块、挡板之间发生的碰撞，通过查阅资料。解决了这个问题。测试时很顺利，没有 bug。
问题二 ：游戏可以多人玩吗？
答：是一款单人游戏，不可以多人玩。
问题三：游戏有几个关卡？
答：游戏只有一个关卡。但是挡板每接中 5 次小球，小球速度将会增加 1 倍。

会议主持人： 唐紫珺　　记　录　人：陈书奇　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

答辩合格

成绩：66　　答辩小组（组长）签名： 陈书奇　　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">北京二手房房源数据的分析与可视化</td></tr>
<tr><td>姓　　名</td><td>杨婧</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550147</td></tr>
<tr><td>指导教师</td><td>陈书奇</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>陈书奇</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：画这么多图得出了什么结论？

得出了北京世纪城房价最高，房源面积集中在 50-200 平方，其最高价格为 14590 万元。装修中精装位居第一，说明大部分二手房房主都属家居生活享受者。2 室 1 厅户型在二手房销售中最受欢迎。房屋朝向中南北方占据最多。

问题二：多少数据，清洗之后还有多少，都有什么字段？

爬取了 84473 条，清洗后为 70906 条，分别有 varchar 和 int 类型。

问题三：论文遇到什么问题，如何解决的？

编写代码时遇到不会的知识点上网查找解决问题。在 pycharm 下载三方库时出现了 Non-zero exit code (2)，首先上网查找尝试所有方法未能解决，找到指导老师，直接重装，安装配置 Anaconda 就可以用了。

会议主持人： 唐紫珺　　记　录　人：陈书奇　　2021 年 11 月 4 日

答辩小组意见（评语）：

答辩合格

成绩： 64　　答辩小组（组长）签名： 陈书奇　　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的象棋游戏开发</td></tr>
<tr><td>姓　　名</td><td>陈嘉和</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550103</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：代码是否可以运行
答：代码可以运行
问题二：运行后有几个界面
答：只有一个界面
问题三：开发的时常是多少
答：开发的时间相对较长，
问题四：代码的长度
答：代码长度有一两千行
问题五：使用的库有哪些
答：主要运用到 pygame 游戏库

会议主持人： 唐紫珺　　记 录 人：蒋亮　　2021 年 11 月 4 日

答辩小组意见（评语）：

成绩：70　　答辩小组（组长）签名： 罗伟　2021 年 11 月 4 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">国内疫情可视化</td></tr>
<tr><td>姓　　名</td><td>付豪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1901</td><td>学号</td><td>201927550107</td></tr>
<tr><td>指导教师</td><td>罗伟</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>蒋亮</td><td colspan="6">讲师</td></tr>
<tr><td>罗伟</td><td colspan="6">讲师</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
1. Django 的作用是什么？
答：用于 web 服务器的开发尽量让模块 功能单一化，方便使用和调用

2. 做的是系统还是爬虫？做了几种类型的图？数据从哪里爬的？
答：系统，折线，饼图，柱状还有全国整个疫情点的显示图，阿里云ＡＰＩ

3. 做的系统里面的数据能不能实时更新？
答：做的系统　采用前后端分离数据可以实时更新

会议主持人：　唐紫珺　　记　录　人：蒋亮　　2021 年 11 月 4 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

成绩：65　　答辩小组（组长）签名：　罗伟　2021 年 11 月 4 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的美团商家数据分析可视化</td></tr>
<tr><td>姓　　名</td><td>王定鑫</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>22</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1，　　数据去空行前和去空行后还有多少？

答：去空前显示五万多行，去空后显示二万五千多条。

2.，数据采集使用了哪些技术？

答：使用了 Python 代码进行了网页数据爬取，在将数据存储在数据库。

3．可视化分析中，所在商圈店铺数量前十排行榜分析的意义？

答：通过此可视化分析图，以便于分析不同地区的外卖发展行业以及侧面看出地区的今后及水平。

4，对这次结果发现了什么不足？

答：这次设计虽然基本满足最开始的需求，但对于网页的反爬机制处理不够到位，爬取时出现多次被封 IP，导致数据采集时间过长。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：65　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的二手房数据分析可视化</td></tr>
<tr><td>姓　　名</td><td>吴家豪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1902</td><td>学号</td><td>23</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：你的数据量爬取了多少？
答：从链家网爬取了大概 30000 条数据
问：是否对数据进行了清洗，用的什么工具清洗的，主要清洗了什么？
答：对数据在 kettle 行了清洗，主要是将数据的重复值以及空值清洗了出来，并将需要的字段分出来
问：对数据的可视化的内容是什么？
答：对二手房市场的价格以及房屋的类型等方面进行了可视化。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：62　　答辩小组（组长）签名：　彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的搜狗电影数据分析系统与可视化</td></tr>
<tr><td>姓　　名</td><td>吴亚楠</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1902</td><td>学号</td><td>24</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：数据清洗做了什么步骤？
答：对获取的数据进行去空格，去标点符号，对一些值为空时设置一个默认值。

问：数据量够吗，有多少？
答：数据清洗保存到数据库后有 26000 多行。

问：系统实现的流程是什么？
答：首先通过 python 代码获取原始数据，之后对数据进行处理并保存到数据库再在 flask 中获取数据库的数据，通过 jinja2 将数据返回到前端进行可视化。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：65　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">网上汽车租赁系统的设计与实践</td></tr>
<tr><td>姓　　名</td><td>向彩民</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>25</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
彭成辉：为什么这个课题？
回答：因为人们的生活越来越离不开汽车这种交通工具，而相对于汽车购买的昂贵价格，汽车租赁有时更加实惠和便利，所以我想做一个汽车租赁系统，使汽车在租赁过程中更有效率。
唐紫珺：为什么选择 PHP 语言来做这个系统？
回答：因为 PHP 语言运行于 lamp 框架，整体框架较为完善，语言也结合了 c+、Java、Python 的一些优势，相对于初学者来说相对容易，且成本较低。
王红军：运行可行性？
回答：系统采用图形化 web 界面，管理员或用户只需要使用任意操作系统上的浏览器就可以操作整个系统，易于管理；PHP 代码是运行在服务端的，PHP 代码在服务器端被解释转变成普通的 HTML 页面内容，送给浏览器端，客户端无法得知其背后的代码是如何运作的。这样做保障了运行的安全性。

会议主持人：唐紫珺　　记　录　人：向彩全　　2021 年 11 月 1 日

答辩小组意见（评语）：

不合格

成绩：50　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 ___2022___ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的淘宝月饼销量销售数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>肖志琳</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550226</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.数据清洗的作用是什么？

　　数据清洗很重要，这个对于我们后续做可视化展示，极其重要。因此我们需要根据后面要做的图形，然后进行对应的进行数据清洗。

2.数据可视化分为那几个模块？

分 5 个模块：月饼销量 Top10 的柱形图；2.店铺月饼销量 Top10 的柱形图；全国月饼销量的地域分布地图；不同价格区间的月饼销量圆环图；月饼销售关键字的词云图。

提问:问什么标识符不建议使用下划线开头？

答:因为在 Python 中以下划线开头的变量为私有变量，如果你不想让变量私有，就不要使用下划线开头

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：60　　答辩小组（组长） 签名：　唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的已加入联合国的公司的分析与可视化</td></tr>
<tr><td>姓　　名</td><td>杨胤</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550228</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.问：该项目一共爬取了多少条数据？

答：一共爬取了三万多条数据。

2.问：该项目是怎么去实现数据的可视化的？

答：通过 pycharm 中 pandas，matplotlib 等库来实现的。

3.问：为什么使用 python 来进行数据的分析与可视化？

答：Python 语言简单易用,现成的爬虫框架和工具包降低了使用门槛,具体使用时配合正则表达式的运用,使得数据抓取工作变得生动有趣。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：62　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 ___2022___ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的电影网站数据分析统计</td></tr>
<tr><td>姓　　名</td><td>银洋</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1902</td><td>学号</td><td>201927550229</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.为什么选择这个课题研究、写作它有什么现实意义？

电影随着时代的进步技术也越来越好，研究电影网站的电影数据，可以侧面反映出什么样的电影更受大众的喜欢，什么样的题材更受大众喜欢。

2.这个课题由几个步骤组成？

此课题从三个步骤出发，分别是数据爬取、数据清洗、数据可视化。

3.课题可视化部分运用到了哪些库，简述一下？

运用到很多库，比如：pandas，numpy，worldcould，jieba，matplotlib

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：__65_____　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Java+oracle 学生信息管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>于海博</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550230</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.该系统有哪些功能？

该系统包含如下的基本模块:1.账号登录模块。2.添加信息模块。3.改动信息模块。4.查看信息模块。

2.该系统针对用户在使用方面有哪些好处？

系统运作后，就用户来讲，因为用户在运用时并不关注系统的构造和完成方式，即对用户全透明，因而系统定坐落于页面友善、实际操作便捷、功能完善的标准。

3.系统能否运行？运行是否稳定？

当前系统还不是很完善，不能运行，功能还有很多没有实现，测试不稳定，后续需要继续开发。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：60　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Web 网上购物系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>于淇民</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550231</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一： 为什么会选择做这个题目？

答：为所有大学在校生的大学生和在校教师们免费提供网上便捷的商品购物咨询服务和消费体验,满足所有大学在校生和大学生对于学校商品的基本购买要求。

问题二：java 的作用是什么？

答：Java 是一门通用编程语言，你可以用它处理商业订单，处理表格和数据库、控制数字机顶盒、控制网站的用户请求、控制计算机视频游戏交互等等，总的来说是一门全能的但有时有些低效的工具。

问题三：什么是 E-R 图?

答：E-R 图也称实体-联系图(Entity Relationship Diagram)，提供了表示实体类型、属性和联系的方法，用来描述现实世界的概念模型。它是描述现实世界关系概念模型的有效方法。是表示概念关系模型的一种方式。用“矩形框”表示实体型，矩形框内写明实体名称；用“椭圆图框”或圆角矩形表示实体的属性，并用“实心线段”将其与相应关系的“实体型”连接起来；用”菱形框”表示实体型之间的联系成因，在菱形框内写明联系名，并用”实心线段“分别与有关实体型连接起来，同时在”实心线段“旁标上联系的类型（1:1, 1:n 或 m:n)。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

不合格

成绩：51　　答辩小组（组长）签名： 彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的 Java 工程师职位数据可视化</td></tr>
<tr><td>姓　　名</td><td>湛鹏宇</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>2019275502 32</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.此项目的设计的思路是什么？

利用 Python 软件进行招聘网站目标数据进行采集，对爬取得数据进行可视化并将数据进行清洗，将数据以图表的形式表达出来，让数据更加的直观。

2.此项目解决了什么问题？

本项目通过在往上爬取具有权威性的职位信息，帮助求职者获取更多更有价值的数据，从而让求职人更加容易寻找到自己心仪的工作岗位。

3.该系统在性能方面要解决哪些需求？

本系统是一个典型的数据分析系统，主要爬取数据进行数据分析。作为分析系统，各功能模块都应在较短的时间内执行完毕，能提高系统的执行效率。本系统除数据采集模块之外的各功能模块应在 2S 内执行完毕，其中执行效率较慢的是数据采集及数据可视化操作。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：62	答辩小组（组长） 签名： 唐紫珺	2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的汽车数据的分析与可视化</td></tr>
<tr><td>姓　　名</td><td>陈聪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>33</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题 1：绘制流程图的步骤？

圆角矩形表示“开始”与“结束”；矩形表示行动方案、普通工作环节用；菱形表示问题判断或判定（审核/审批/评审）环节；用平行四边形表示输入输出；箭头代表工作流方向。

问题 2：E-R 图是什么？

E-R 图也称实体-联系图(Entity Relationship Diagram)，提供了表示实体类型、属性和联系的方法，用来描述现实世界的概念模型。它是描述现实世界关系概念模型的有效方法。是表示概念关系模型的一种方式。用“矩形框”表示实体型，矩形框内写明实体名称；用“椭圆图框”或圆角矩形表示实体的属性，并用“实心线段”将其与相应关系的“实体型”连接起来；

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：__61__　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 jsp 的医院挂号系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>付缓缓</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>2019275502 34</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.该系统实现了哪些功能？

该系统功能细分为三大块功能，分别是：用户对自己的个人信息以及挂号信息的管理功能、新用户的注册与已有账户的登录模块、用户对医院门诊的线上挂号并查询详细信息功能。

2.此系统包含哪些实体，它们之间的关系是怎样的？

本系统共有 7 个 E-R 实体属性图，分别为系统公告属性图、医生信息属性图、挂号信息属性图、医院信息属性图、科室信息属性图、用户信息属性图、管理员信息属性图，分别表明了实体间的关系。

3.该系统有哪些不足？

系统最终没有达到预期所要展示的个性化和智能化。在页面以及设计上只完成了各项功能的实现和模块显示，界面设计不够完善，还需设计完备。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：65	答辩小组（组长） 签名： 唐紫珺		2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 JSP 人力资源管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>郭海潮</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550235</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问：自己为什么选择这个课题？

答：一方面是社会的进步使得越来越多的小企业变成大企业、也越来越多的人开始创业；第二个方面是自己将步入社会会接触到这个系统，所以面对越来越多的大型企业，应该有新的人力资源管理系统去适应当下的企业状况，所以我选择了这个课题。

问：全文的基本框架是什么？

答：整个系统需要前端、后端和用户端相互联系，所以然我联想到了 SSM 框架（spring+springMVC+Mybatis），使用 SSM 框架贯穿全系统，设计思想为 MVC 模式。

问：研究这个课题的意义和目的是什么？

答：这个课题的实现可以很大程度的方便公司管理人员对内部员工的信息进行管理变动，有利于提高公司的工作效率和打好公司内部基础。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：__65__　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的大数据职位信息的分析与可视化</td></tr>
<tr><td>姓　　名</td><td>胡海燕</td><td>院(部)</td><td>网络空间安全学院</td><td>专业/班级</td><td>大　数　据 1902</td><td>学号</td><td>201927550236</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、　　你为什么会选择这样一个课题名称？

答：对于我们来说，部分人以后就业想要往大数据方向进行发展。另外，我的选题也有一定的意义，原因主要如下所示：一、帮助在校生更好地了解该专业就业时招聘者更看重哪些能力，以及该行业的薪资、学历要求等等；二、为应届毕业生更好地了解该专业的发展前景，给予她们一个宝贵的参考意见；三、为求职者或者有意向转行的人员提供一些有用的数据，帮助他们更好地规划未来职业方向。

2、　　你是爬取哪个招聘网站的，爬取了多少条数据？

答：我爬取的网站是前程无忧，采用 pycharm 结合多线程技术爬取了上万条招聘数据。

3、　　用简洁的语言概括，你主要采用了什么技术？

答：本次研究中主要用到了四大技术，其一是爬虫技术，即模拟头发送请求、获取响应并利用 json 字符串提取数据；其二是数据预处理，将多余、重复、不合理的数据进行清洗；其三是利用数据库 Navicat 将清洗完的数据进行存储；其四是利用可视化技术实现数据向图表的转换，将有用的数据利用不同可视化图表将其呈现出来，挖掘数据间隐藏的关联性，最大程度地展现数据间的特征。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：78　　答辩小组（组长） 签名：　唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的网上化妆品数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>江唯</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1902</td><td>学号</td><td>201927550237</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职　称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问：你的化妆品数据库表的字段有哪些？
答：化妆品数据库表里面包含了 id、商品名称、店铺名称、链接、价格等字段。
问：论文当中的数据是来源于哪里呢？
答：化妆品数据是来源于淘宝网站里面的。
问：你选择这个论文题目的目的是什么？
答：通过对数据的分析，发现销售数据信息的内在规律，从而就可以知道用户一般偏爱于使用哪种品牌的化妆品，同时可以将自己学习的综合专业知识融合到系统当中去。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：
合格

成绩：__62__　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日</td></tr>
</table>

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的网上手机销售数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>姜雪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550238</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.在爬取网站数据时，采取怎样反爬措施进行数据爬取？

面对爬虫时代，各个网站基本上都设置了相应的反爬虫机制，当我们遇到拒绝访问错误提示 404 时，可通过获取 User-Agent 来将自己的爬虫程序伪装成由人亲自来完成的信息的获取，而非一个程序进而来实现网页内容的获取。

2.数据去重的方法有哪些？

python 列表去重的 5 种常见方，使用 for 循环实现列表去重、使用列表推导式去重、使用集合转换函数 set()实现列表去重、使用新建字典方式实现列表去重、删除列表中存在重复的数据。

3.系统数据可视化有哪些模块？

可视化模块有各品牌手机销量占比、各品牌手机均价、各种品牌手机最高价、各个手机最低价、各种品牌手机销量等模块。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：70　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 JSP 学生选课系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>李静仪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550239</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问：为什么选这个课题？

答：可以方便学生更方便的选课，实现全校学生选课在线查询，还可以对新来的学生信息进行信息录入。由于学生上课情况复杂，可以实现信息修改，及时调整学生的基本情况。

问：JSP 系统中，你是如何实现图片滚动的？

答：通过 JS 的 setlnerval 实现图片滚动。

问：javaweb 中你是如何连接数据库的？

答：1.创建测试工程（java web 功能一样）；

2.创建数据库连接类，最主要的是创建数据库连接类，封装了 Connection,Statem...

3.编写数据库连接类方法：根据数据库驱动包，获取数据连接，以便程序使用；

4.封装执行方法，此方法执行 SQL，返回受影响的行数

5.数据库查询。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：62　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 jsp 的社团管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>李珍</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550240</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.选择这个课题的原因是什么？

步入大学参加社团后，觉得大学社团很多，而且每年还会有不同的社团诞生，且每个社团的社团活动也非常多，隔三差五就举办，对于社团成员来说，选择参加什么社团和什么社团活动就会非常的麻烦和复杂，社长对于管理统筹社团和社团活动的工作量也非常大，所以我就想设计一个社团管理系统来解决这方面的问题，提高社长工作效率，使学生参加社团和社团活动更加方便。

2.这个课题的目的和意义是什么？

信息化社团管理，使日益复杂的社团管理工作简单化，减少学生选择社团活动的时间和过程，大大提高了工作效率，让社团管理工作顺应时代发展。

3.毕业设计项目有什么不足？

项目模块不够丰富，社长功能尚未开发完全，页面还不够美观，对于项目的验证可以进行进一步的完善。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：74　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的家具数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>刘燕</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1902</td><td>学号</td><td>41</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 封面在课题名称和姓名之间有回车符等。
2. 系统详细设计缺少可视化流程图。
3. 数据处理未连接数据库。
4. 可视化图表单一。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

不合格

成绩：__50__　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的网上招聘岗位数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>潘希芮</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902 班</td><td>学号</td><td>2019275502 42</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.系统包含哪些模块？

该项目的模块有数据采集模块、数据清洗模块、数据分析模块，数据可视化模块。

2.数据从哪里采集，共爬取了多少数据？

在前程无忧网中搜索相关职位数据。翻页查看招聘职位信息，可以找到 URL 翻页的规律。检查网页源代码以查找要提取的数据。爬取了 600 页的招聘数据，共 30498 条招聘信息。

3.数据可视化包含哪些？

数据可视化包含招聘岗位数最多的城市 Top10，计算岗位薪酬情况，学历的要求的情况，工作经验要求情况，招聘公司所属行业的分布情况等。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：75　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于大数据技术的新闻网数据的分析可视化系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>唐亚丽</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902 班</td><td>学号</td><td>201927550243</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、请你描述一下整个系统的大致功能模块？

答：本系统设计主要有五个功能，分别是大数据场景模拟、日志采集服务、离线统计服务、数据可视化服务。

2、请问整个系统的可靠型是怎么样保证的呢？

答：本系统基于 Hadoop 分布式文件系统，由 Zookeeper 组件实时监控分布式集群的运行动态，能有效确保整个系统的稳定性和可靠性。

3、请问你设计的系统中离线统计服务是有哪些功能？

答：本系统的离线统计服务主要在大数据组件 Hive 上实现，本系统将数据存储在 hbase 数据库中，将 Hive 与 hbase 数据库做好集成后，能够在 Hive 读取数据，也能实现 Hive 的离线数据分析。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：__77__　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的湖南天气的数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>田丰良</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1902</td><td>学号</td><td>201927550244</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.本系统采取了哪些工具和技术实现？

基于 Python 工具，实现对湖南天气数据信息的网络爬取，运用 Pycharm 工具，实现数据分析系统的运行以及数据的可视化，并运用到 Requests、Pandas 等库。

2.数据是怎样采集的？

数据采集是依据数据源，选用湖南天气数据，应用 Requests 库发动 get 请求,猎取页面信息，通过 re 库从源代码中精准的匹配到需求信息，最后以 csv 文件存储数据。

3.数据怎样进行存储的？

该项目是采用.csv 文件的形式对数据进行数据库的存储，其中包含的字段有：city、date、maxWendu、minWendu、weather、wind direction、wind power

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：62　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的斗鱼直播数据的数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>吴艳娇</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据技术与应用</td><td>学号</td><td>201927550245</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、请你描述一下你是怎么爬取数据的？

答：网络爬虫通过向 start_URL 属性中定义的 URL 发出请求开始，并调用默认回调方法 parse，将响应对象作为参数传递。在解析回调中，使用 CSS 选择器循环引用元素，生成包含提取的引用文本，然后将所爬取的数据写入文件，以存储文件格式存储在系统。

2、请你简单描述一下技术架构图？

答：先从网站上爬取数据，然后再用 kettle 将无意义数据进行处理，最后用 python 代码将数据以图表形式可视化显示出来。

3、请描述一下数据处理的过程？

答：因获取到的数据有脏数据，先用字段选择将无用字段进行删除，然后用过滤记录将无意义数据进行处理，最后运用字符串处理将“工会名称”字段中的数据进项处理，以便后面更好地进行数据可视化，最后用文本文件输出。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：63　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 ___2022___ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 jsp 的网上书城系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>伍健</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1920 班</td><td>学号</td><td>2019275502 46</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.此系统的设计目标是什么？

通过本软件系统，能帮助顾客利用网络快速便捷地进行购物，而书店管理员则可以更加方便的管理顾客信息、以及书籍的入库与整理，并对顾客订单进行处理，使购书流程更加地安全、方便、快捷。

2.此系实体间的关系有哪些？

本系统 E-R 属性图共有 5 个，分别为商店书籍属性图，管理员、用户信息属性图，购物车订单属性图，销售榜单属性图，公告发布属性图等。

3.此系统有哪些功能？

此书城管理系统分用户注册功能、用户登陆功能、用户查看功能、用户支付功能、管理员登陆功能、管理员管理功能。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：74　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的全国景区数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>肖晚霞</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大　数　据 1902</td><td>学号</td><td>20192755 0247</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.请问你是如何获取想要的数据的？

先确定好要找什么数据，比如说我是需要全国景区数据，再找到去哪儿旅行的官网，然后再利用爬虫把它保存到本地文件

2.请介绍一下你此次毕业设计的目的。

本课题要完成的主要任务是实现基于 Python 的全国景区数据分析与可视化，将通过全国景区数据的查找、获取、清洗以及对数据的可视化等方法，来对数据信息进行全面分析，找出全国的景区数据，得到景区分布情况，发现热门景点，了解各个景区的价格和月销量，掌握旅游景区的类型，分析行业现状和发展趋势

3.你设计的数据库中的数据表有几个字段

七个，分别是 province（地区）、name（名称）、level（级别）、hot（热度）、count（相关景点个数）、sales（月销量）、price（门票）

会议主持人：唐紫珺　　记　录　人：　唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：　63　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的租房信息数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>曹伯龙</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550201</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1：pandas 是什么？

pandas 是基于 NumPy 的一种工具，该工具是为了解决数据分析任务而创建的。

2：kettle 的处理流程是怎样的？

先把数据放到 csv’文件内容里面，然后过滤掉不需要的数据在将错误重复的字段去掉，最后将数据导入数据库中。

会议主持人：唐紫珺　　记　录　人：　唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：61　　答辩小组（组长）签名：　彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的网络小说网站统计数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>董涛</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550202</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.使用爬虫抓取了哪些网站？

为了对比分析数据，我抓取了飞卢小说网，纵横小说网，晋江小说城，起点小说网的数据。

2.抓取的数据存储在哪里？

为了避免原始数据丢失，我将其存放在 csv 文件中，经过清洗后，我将数据存入了数据库。

3.数据如何清洗？

为了使数据更干净，我对每个数据的类型进行了判断，并去除了空格特殊符号，最后还比较了是否是正确的数据。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩。

成绩：__70__　　答辩小组（组长）签名：唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的招聘信息数据处理与可视化</td></tr>
<tr><td>姓　　名</td><td>龚诗瑭</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550203</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

（1）　　爬虫的流程是什么？

确定自己所爬取的网站，获取 URL、解析页面，获取 JS 中所需数据、将所获取的数据存储到 CSV 文件中。

（2）　　数据库的 er 表中包含那些内容？

数据库招聘信息（intro_job）ER 表中包含了招聘职业（job_array）、招聘公司（company_array）、招聘城市（place_array_city）、薪资情况（salary_array_unite）、招聘信息发布时间（date_array）。

（3）　　毕业设计的目的是什么？

通过爬虫与可视化对招聘信息进行爬取与可视化，能够让我们这些刚才出学校的学生更加直观的了解自己所在城市的工作岗位与薪资情况，让我们能够更好地判断自己的实习薪资。解决广大实习生的就业问题！

会议主持人：唐紫珺　　记录人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：62　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的失踪人口数据预处理与可视化</td></tr>
<tr><td>姓　　名</td><td>龚赞强</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550204</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

（1）　　爬虫的流程是什么？

确定目标网站、获取 URL、爬取网页源代码、定位筛选需要爬取的数据、爬取需要数据、存储数据。

（2）　　毕业设计的目的是什么？

通过爬虫与可视化技术对我国失踪人口的数据进行爬取和可视化！更加直观地了解失散人员户籍地的数据与失散原因的对比，以及失散日期的统计，更加了解失散人员这个群体从而预防类似事情的发生。

（3）　　可视化了一些什么功能？

失散人员的户籍地可视化、失散人员的失散原因可视化、失散日期的可视化、失散人员的性别比例可视化。

会议主持人：唐紫珺　　记录人：彭成辉　　2022 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：72　　答辩小组（组长）签名：唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 ___2022___ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的平板电脑销售数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>何泽阳</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550205</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

（1）　　你的爬虫是流程是什么？

URL 链接、数据抓取、数据清洗、数据存储、数据可视化。

（2）　　数据清洗清洗哪些东西？

有些商品地区重复可以利用数据清洗筛选掉，去除空格等特殊字符，去除错误的数据

（3）　　选这个题目的意义是什么？毕业设计的目的是什么？

意义在于分析平板网站其潜藏的重要数据，通过数据看出顾客的购买倾向，了解不同顾客的需求。也能在过程中提高自己的动手能力和对知识的理解能力。毕业设计也是大学期间对学生的一个阶段性的知识总结,将学习中的各个知识点进行融合贯通，提升编码能力和开发效率。

会议主持人：唐紫珺　　记　录　人：　唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

成绩：__62_____　　答辩小组（组长）签名：　彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的去哪儿旅行的数据分析和可视化</td></tr>
<tr><td>姓　　名</td><td>黄少将</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550207</td></tr>
<tr><td>指导教师</td><td>李逸</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

（1）　　数据清洗的流程是什么？

读取数据内容，检查数据，处理缺失数据，逻辑错误清洗，非需求数据清洗。

（2）　　毕业设计的目的是什么

通过去哪儿网各省门票的销售情况，简单分析一下全国比较热门的景点分布和国民出游情况，看看哪些景点比较受欢迎。

（3）　　爬虫的流程是什么？

选择爬取网站，获取 URL，爬取网页源代码，选择要爬取的数据，爬取数据，存储数据。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

成绩：合格　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的长沙市租房信息数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>蒋博宇</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550208</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

（4）　　数据存储的位置是哪？

数据首先通过爬虫从网站中获取后导入 csv 文件中，在通过数据清洗后存入数据库中。

（5）　　数据表如何制作？

根据 csv 文件的结构创建 shuju 表，并创建 name，new、diqu、huxin、zuj、mianji、dic、louc 等八个字段。

（6）　　毕业设计的目的是什么？

通过爬虫与可视化对招聘信息进行爬取与可视化，能够让我们这些将要进入社会工作的学生了解长沙市的租房信息，并可以作为租房时的参考，让我们能够更加理性的选择要租的房子。

会议主持人：唐紫珺　　记录人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：74　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的空调商品可视化与分析</td></tr>
<tr><td>姓　　名</td><td>李余胜</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1902</td><td>学号</td><td>201927550209</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
1. 图 2-1 技术架构图有问题，应进行更改
2. 功能描述中表 3-1 数据采集时间要进行更改
3. 图 3-1 功能需求图不对，要进行更改
4. 数据可视化类型单一，要丰富可视化类型
5. 系统数据库设计中，要对图和表进行更加详细的描述
6. 图 5-3 数据处理流程，应改成数据处理过程
7. 清洗和处理数据中，应对所进行的操作，添加相应的描述
8. 参考文献 3 和 4 引用不对

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

成绩：__62__　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的淘宝显卡价格分析与可视化</td></tr>
<tr><td>姓　　名</td><td>梁康祺</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>2019275502 10</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.你的数据采集模块和数据分析模块是什么？

答：采集模块主要由爬虫程序和代理池组成，负责显卡数据采集工作，分析模块负责对数据库中数据进行分析、可视化处理、训练并得出有用信息。

2.Scrapy 框架的爬取过程是什么？

答：爬取过程是 Scrapy 引擎发送请求，之后调度器把初始 URL 交给下载器，然后下载器向服务器发送服务请求，得到响应后将下载的网页内容交与蜘蛛来处理，尔后蜘蛛会对网页进行详细的解析。

3.淘宝显卡价格的分析和可视化包含几个模块？

答：分为三个模块：数据采集，数据清洗，数据可视化。

会议主持人：唐紫珺　　记录人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：78　答辩小组（组长）　　签名：　唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 对求职信息的数据分析系统与实现</td></tr>
<tr><td>姓　　名</td><td>刘飚</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902 班</td><td>学号</td><td>201927550211</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1,请你描述一下数据可视化系统的大致功能？

答：利用计算机图形和数据处理技术，把数据转变成数字信息变成为直观的，以图形图像信息表示的，随时间和空间变化的物理现象呈现处理。

2，请你描述一下 Hadoop 的三大组件？

答：HDFS，具有可靠，高吞吐量的分布式文件系统　　MaReduce 分布式的离线框架　Yarn 分布式资源管理框架。

3，请问这次你分析的数据有哪些好处？

答：本项目主要完成的是基于 Python 对求职信息的数据分析系统与实现，简单的来说，就是确定分析目标，使企业领导者能够更快地理解和处理他们的信息，使利益最大化，对市场变化更快的调整和对新机会的快速识别每个行业的竞争优势，通过对数据信息进行分析求职信息的数据分析主要在于产品搜索维度、趋势、用户、市场、交易等，其中包含对求职各方面的数据分析，及对外市场大趋势分析。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：61　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的手机商品数据可视化与分析</td></tr>
<tr><td>姓　　名</td><td>刘强</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550212</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.数据来源于哪里？

本数据来源于淘宝，是其随机选择约 100 万用户在 2021 年 6 月至 9 月之间发生的行为记录，具有包括点击、购买、加购物车和收藏商品的行为。

2.数据处理包含哪些内容？

处理数据，主要包括对异常值、重复值和缺失值的处理。

3.手机参数信息怎么提取？

手机参数信息以字典形式保存，创建一个函数，将每个键值对提取出来，以列的形式呈现。

会议主持人：唐紫珺　　记录人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：70　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的 TK 电影网站数据的分析系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>刘威驿</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1902</td><td>学号</td><td>13</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

提问 1：为什么选择电影网站数据进行分析与处理？

学生回答：首先，我觉得在现在激烈的国际文化交融的情况下，人们的文娱活动的次数增加，分析电影网站的数据有助于让人们了解到不同人们的喜好，从而在看电影的过程中了解到不同国家的文化，有助于丰富人们的视野，同时这对中国的电影厂商来说也是很重要的，他们能够了解到人们的喜好，从而推动本国的产业发展。

提问 2：通过可视化数据如何才能助于本国电影产业的发展？

学生回答：首先通过对数据的分析与处理我们从图表的情况能够明显的看出什么电影类型人们比较爱看，什么不爱看，我们应该多致力于人们爱看的电影进行发展，同时注重电影的质量。多学习其他国家电影产业的发展，从而运用到本国电影产业的发展。

提问 3：说说中国比较有代表性的电影并说明其有什么意义？

学生回答：《战狼 2》作为一部以英雄主义主旋律为基调的作品，是让人倍感惊喜的，吴京不但在这部作品中塑造了一个有血有肉，能去打动人心的英雄人物，也更让人见证了中国军人那永不褪色的军魂本色。同时也在电影中一些细节中体现出中国的力量。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 1 月 1 日

答辩小组意见（评语）：

合格

成绩：75　　答辩小组（组长）签名：彭成辉　　2021 年 1 月 1 日

湖南信息职业技术学院 __2022__ 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于大数据的离线电商数仓系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>罗胜</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902 班</td><td>学号</td><td>14</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、请问数据采集流程是如何实现的呢？

答：数据采集分为两种形式，MySQL 业务数据直接通过 Sqoop 传输到 HDFS 上，埋点数据则通过 Flume-Kafka-Flume 的方式进行传输。

2、请问整个数仓系统大致分为几层呢？

答：本离线数仓系统分为五层，分别为 ODS(原始数据层)、DWD(明细数据层)、DWS(服务数据层)、DWT(数据主题层)、ADS(数据应用层)。

3、请问你是如何实现各个任务的调度？

答：本系统通过 Azkaban 框架实现工作流程的调度，使复制工作流简单化，大程度的节省了人力成本。

会议主持人：唐紫珺　　记录人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：75　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的长沙二手房数据的可视化与分析</td></tr>
<tr><td>姓　　名</td><td>马海</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550215</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：

1、系统设计部分缺少总体 ER 图。
2、图片缺少解释，不能只放单一图片。
3、可视化绘图种类太少。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年　11 月　1 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

不合格

成绩：50　　答辩小组（组长）签名：彭成辉　　2021 年　11 月　1 日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的全国疫情数据可视化</td></tr>
<tr><td>姓　　名</td><td>欧阳海波</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1902</td><td>学号</td><td>201927550216</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.系统由几部分构成？

中国疫情数据分析系统由五大部分构成，分别是数据源、数据采集、数据清洗、数据分析和数据可视化。

2.数据那个哪里获取？

数据采集的数据一般来源于网页，主要从中国政府网采取中国疫情数据。

3.数据数量不够，可视化图片不完美，怎么解决？

加大数据采集量，增加清洗过滤规则，美化可视化效果。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

通过答辩！

成绩：70　　答辩小组（组长） 签名： 唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的网易云用户信息分析与可视化</td></tr>
<tr><td>姓　　名</td><td>欧阳鹏</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1920</td><td>学号</td><td>17</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：爬虫代码怎么实现的？总共爬取了多少数据？
答：使用 Python 代码编写，并且引入 request 库获取网页请求，在通过 xpath 库解析网页，对数据进行爬取。总共爬取了 30000 条数据。

问：数据清洗时，用了什么工具，主要清洗了什么？
答：清洗数据用的 kettle 软件，主要清除了空值，错值，无意义值，并将字段剪切成利于可视化的长度。

问：可视化使用了什么库？
答：使用了 pyecharts 库来可视化。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：61　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的二手车市场数据分析处理与可视化</td></tr>
<tr><td>姓　　名</td><td>石磊</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1902</td><td>学号</td><td>201927550219</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：程序爬取的数据量是多少？可视化图做了几个？数据处理怎么处理的？
答：数据爬取量是 32000 多条，数据可视化图做了 6 个。
问：爬虫代码是怎样实现的？
答：在 Python 上写代码，用 requests 加 parsel 组合去实现，先采用 requests 获取目标网页内容，因为 Parsel 库支持标准的 css 选择器和 xpath 选择器，所以可以通过 get 方法或 getall 方法获取文本或属性值。然后处理得出数据存入数据库中。
问：数据处理怎么处理的？
答：使用 kettle 工具对数据进行排序，去重和过滤处理，然后就能得到处理后的数据了。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

合格

成绩：62　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的豆瓣书籍数据分析可视化</td></tr>
<tr><td>姓　　名</td><td>谭智明</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据技术与应用 1902 班</td><td>学号</td><td>201927550220</td></tr>
<tr><td>指导教师</td><td>王红军</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：毕业设计的目标是什么？
答：利用 Python 网络爬取以及可视化对豆瓣网上的书籍进行爬取分析并对其进行可视化。让书籍的类型更加直观的呈现在读者面前，使读者能更加方便的选择适合自己以及自己喜欢的书籍，更方便的分析自己的阅读意向。

问：此毕业设计的技术架构由哪几个部分组成？
答：由数据源，数据采集，数据清洗，数据分析和数据可视化五部分组成。

问：数据分析的目的是什么？
答：数据分析是指通过分析方法和分类与算法等发现数据中有价值的信息，为后续可视化而做准备工作。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 11 月 1 日

答辩小组意见（评语）：

答辩通过！

成绩：　60　　答辩小组（组长）签名：唐紫珺　　2021 年 11 月 1 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的疫情数据地图的数据采集及实现</td></tr>
<tr><td>姓　　名</td><td>陶汉宇</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大 数 据 1902</td><td>学号</td><td>201927550221</td></tr>
<tr><td>指导教师</td><td>彭成辉</td><td>所在部门</td><td>网络空间安全学院</td><td>职 务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>王红军</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、　　请你描述一下这个系统的功能模块？

答：本系统主要爬取可视化有三大功能模块，分别是国内的疫情数据信息地图、所有国家疫情信息地图、全球累计疫情以及当前疫情地图。

2、　　请问这些功能模块的信息主要通过哪些工具实现爬取的呢？

答：爬取基于 Python 工具实现运用到 Requests，BeautifulSoup4，Echarts 等库，用 BeautifulSoup4，对收集的全球及国内疫情的数据信息进行抽取、转换、加载，形成分析型数据。

3、　　疫情数据信息的可视化地图是是通过哪些工具实现可视化的呢？

答：数据分析系统与可视化运用到了 Pycharm 工具，通过采用可视化工具 Echarts 、前端 HTML 进行可视化，绘出疫情数据地图。

会议主持人：唐紫珺　　记　录　人：唐紫珺　　2021 年 11 月 1 日

答辩小组意见（评语）：

不合格

成绩：50　　答辩小组（组长）签名：彭成辉　　2021 年 11 月 1 日

<table>
<tr><td>课题名称</td><td colspan="7">基于 jsp 办公管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>易虎</td><td>院（部）</td><td>网络空间安全学院</td><td>大数据1903</td><td>大数据技术</td><td>学号</td><td>201927550329</td></tr>
<tr><td>指导教师</td><td>胡钰芊</td><td>所在部门</td><td>网络空间安全学院</td><td>职务（职称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>胡柳</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
<tr><td>胡钰芊</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题 1：　解释一下 CSS 是什么？
答：　CSS（Cascading Style Sheet）是一组格式设置规则，用于控制 Web 页面的外观。

问题 2：　什么是 C/S 结构？什么是 B/S 结构？两者又有哪些不同？
答：　C/S 结构（Client/Server）是软件系统体系结构的一种。简单地讲就是基于企业内部网络的应用系统。
B/S 结构（Browser/Server）即浏览器和服务器结构。
1.硬件环境不同。　2.对安全要求不同。　3.处理问题不同。

问题 3：MYSQL 数据库中 innoDB 和 myisam 的区别？
答：1.innoDB 支持事务，myisam 不支持。
2. innoDB 支持行级锁，myisam 支持表级锁。
3. innoDB 支持外键，myisam 不支。
4. innoDB 不支持全文索引，myisam 支持。

会议主持人：　刘彪　　记　录　人：　胡柳　　2021 年 11 月 3 日

答辩小组意见（评语）：

通过答辩！

成绩：　68　　答辩小组（组长）签名：　刘彪　　2021 年 11 月 3 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的运动鞋消费数据分析与可视化成果书</td></tr>
<tr><td>姓　　名</td><td>唐自华</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>19</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：爬取了多少条的数据？
回答：一共爬取了 22000 多条数据，清洗完之后还有 15000 多条。
问题二：爬取数据出现报错怎么办？
回答：出现报错后，刷新网页，复制新 cookie 后，重新开始爬取。
问题三：做了哪些清洗？
回答：首先由于我是分品牌爬取的，首先整合成一份数据，再进行去重，还有去除多余数据，最后分割出省份。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的腾讯视频数据可视化</td></tr>
<tr><td>姓　　名</td><td>涂特朗</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>20</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：1.清洗过程：流程代码不够详细
2.　　流程图：流程图格式不对，要重新画
3.　　可视化画图：需要从数据库中提取数据，不能读取本地文件

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，个别图表需要进行修改

成绩：　65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的 IMDb 电影数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>王胡涛</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>21</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：爬虫用什么解析的页面？
答：使用 xpath 解析 html 页面
问：数据怎么清洗的？
答：使用 mapreduce 清洗，去除空值、重复值，删除部分不完整数据
问：数据可视化画了几张图？展示一下
答：6 张，各类电影数量柱状图、电影数量环形图、1902 年至 2021 年每年发布的电影数量散点图、电影时长分析折线图等等

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的淘宝衣服数据可视化</td></tr>
<tr><td>姓　　名</td><td>王念铖</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>22</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
唐紫珺：这次做毕业设计一共爬取了多少条数据，有没有用到反扒。爬取过程是否顺利，用了哪些解决办法？
王念铖：这次做毕业设计一共爬取了两万多将近三万条数据，因为是爬取淘宝网的数据使用没有用到反扒的技术，爬取过程中经常会被封网页，只能等一段时间重新刷新验证，进入页面后改一个“cookie”部分继续爬取数据。
彭成辉：做可视化用了哪些包？做了什么图？
王念铖：做可视化的时候用了 pyecharts、pandas 这两个包，做了柱状图，饼图，还有一个中国地图。
刘彪：数据都爬取了哪些字段，有没有做数据分析？
王念铖：这次爬取了网站上的付款人数，地区，商品名称，店铺名称，商品价格这些字段，有用到 pandas 做数据分析。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日</td></tr>
</table>

答辩小组意见（评语）：
毕业设计内容符合要求，文字和格式还需进一步调整
成绩：65　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

课题名称	基于 python 的书包销售数据分析与·可视化						
姓　　名	王增龙	院（部）	网络空间安全学院	专业/班级	大数据1904	学号	24
指导教师	熊旺	所在部门	网络空间安全学院	职　务（职 称）	讲师		
答辩小组成员	姓名	职务（职称）					
	唐紫珺	讲师					
	彭成辉	讲师					
	刘彪	讲师					

答辩记录：
1.爬了哪个网站
答爬的是淘宝网页版，具体操作：登陆淘宝官网，输入书包，用浏览器的开发者工具，手动获取 cooki。
2.有没有用反扒
答：没有，网页被封的解决方案是刷新网页或者休息一下获取 URL。
3.爬了多少数据
答：一共爬了二万多条，数据清洗完之后剩余一万五千多条。
4.用了什么包
答：使用的是正则表达式。
5.清洗怎么清洗的
答：先对数据进行合并，然后去除，检查并去除缺失值，再剪切字段，最后排序。
6.可视化做了那些图
答;五个图；一个环形图，一个柱状图，一个折线图，一个中国地理图，一个词云图。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 7 月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：67　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的淘宝网电脑销售数据可视化分析</td></tr>
<tr><td>姓　　名</td><td>吴江</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>25</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1，爬了多少数据？

一共爬了 2 万多条。

2，有哪些字段？

字段有：产品名称，单个商品的的价格，地区，购买人数。

3，清洗用什么清洗，清洗了些什么？

使用了 kettle 清洗，清洗了空值重复值。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

答辩合格

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Vue 的电商后台管理系统</td></tr>
<tr><td>姓　　名</td><td>吴平</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550426</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：前端用什么写的？
Css 部分是用 scss 框架用的 VUE.JS 前端框架，使用 vue-clii 搭建项目整体框架。
问题二：后端用什么写的？
用 Node.js 写的，通过 node.js 编写接口跟前端页面进行数据对接
问题三：开发过程中遇到了什么困难，怎么解决的。
遇到过一个缓存问题
症状：代码没有问题，配置也没有问题，整理半天也没有找到问题。
解决办法：重启项目，重启浏览器，清理缓存，重新编译，不行再重启 idea，不行再删除项目重拉取项目，不行再重启电脑。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，流程图需要规范一下

成绩：65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">网上门诊预约挂号系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>易威</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>27</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.预约挂号的流程没用介绍清楚。

答：需要挂号的用户现在网站上进行注册，注册好了在进行登录，登录成功了，在预约栏目进行选择，医生和科室，点击预约就 ok 了。

2.SSH 框架是什么？

答：S.S.H 框架是 J2EE 应用中 struts2+spring+hibernate 三大免费开源框架的结合使用，它可以看成工具，也是中间件。他是用来提高我们的开发效率，提高我们软件产品的可维护性、可扩展性乃至敏捷性的。他们里面有很多优秀的设计理念及模式应用

3、普通用户和管理员用户和医生怎么区别怎么实现？

答：在语句中做一个判断，使之产生区别

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 7 月 28 日

答辩小组意见（评语）：

答辩合格

成绩：67　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 springboot 的书籍论坛系统毕业设计</td></tr>
<tr><td>姓　　名</td><td>易宇航</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550428</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.为什么选择这个主题作为毕业设计？

答：开发书籍论坛系统的目的是提供一个供阅读学习交流的平台，为优秀的书籍提供交流经验、探讨问题的社区。

2.你的系统有哪些用户？

答：分为普通用户和管理员用户

3.什么是 spring boot？

答：Spring Boot 是所有基于 Spring 开发的项目的起点。SpringBoot 默认配置了很多框架的使用方式，就像 maven 整合了所有的 jar 包，spring boot 整合了所有的框架。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">个人博客系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>禹乐田</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>29</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1 网站搭建用的是什么环境?
答：apache,mysql 数据库,和 phpstudy 集成环境。

2 登录是如何实现的？
登录页面中一个 form　当用户输入账号密码时，点击 提交按钮时，调用 form 表单的 login.php 文件，运行时，通过 get 获取 form 表单中标签内的 value 值。php 变量获取到账号后，调用链接 MySQL 数据库函数，将变量与数据库中的用户账号密码匹配，符合要求则传输数据，跳转其他页面。

3. 搜索功能是如何实现的？
搜索文章的原理在用户输入数据时点击搜索按钮，调用搜索函数，ajax 调用相关 PHP 文件，例如　http://sou.php？key=”语文”，php 运行时，$key 变量通过 get 获取值 语文 调用链接数据库函数，将变量与数据库文章表中的名称比较，是否包含，包含的直接一天 json 格式输出。ajax 接受后将数据保存在本地 localStorage 或者调用函数，以 innhtml 形式插入。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：
答辩合格

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于音乐用户画像的可视化系统</td></tr>
<tr><td>姓　　名</td><td>曾露</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550430</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：什么是用户画像？
答：给用户打标签

问：数据生成脚本的数据怎样生成？
答：依赖爬虫的数据使用随机数生成数据

问：可视化是怎样实现的
答：使用 sql 语法获取到符合画图的数据，使用 supperset 画图

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字还需润色，格式还需进一步调整

成绩：　69　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">人力资源管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>张扬</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550431</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、ssm 框架是什么？

SSM 全称是 Spring+SpringMVC+MyBatis。

SSM 框架集由 Spring、MyBatis 两个开源框架整合而成（SpringMVC 是 Spring 中的部分内容）。常作为数据源较简单的 web 项目的框架。

2、BS 模式是什么意思？

BS 开发模式是指以浏览器和服务器结构开发软件的模式。B/S 是 Browser/Server,即浏览器/服务器的意思

3、普通用户和管理员用户区别怎么实现？

在语句中做一个判断，使之产生区别

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

基本合格，还需进一步改进

成绩：60　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 ssm 的图书馆管理系统</td></tr>
<tr><td>姓　　名</td><td>周森亮</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550432</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.唐紫珺问：图书馆管理系统如何区分管理员和读者之间的登录？

答：a.通过账号区分。管理员和读者同页面登录，管理员账号唯一，已写死数据库，部可更改，读者账号以 10000 开始。b.通过登录背景图区分。账号登录成功，进入页面不同，管理员和读者登录成功后，都会进入自己特定的背景图，点击功能按钮后才会进入图书管理系统。

2.刘彪问：什么是 ssm 框架？

答：ssm 是 Spring MVC ＋ Spring ＋ Mybatis 的简称，是标准的 MVC 设计模式，将整个系统划分为 Bean 层，Controller 层，Service 层，DAO 层四层，Spring MVC 框架是有一个 MVC 框架，通过实现 Model-View-Controller 模式来很好地将数据、业务与展现进行分离
Spring 是一个开源框架，Spring 是一个轻量级的控制反转（IoC）和面向切面（AOP）的容器框架，还能更好的让其他框架整合，MyBatis 是一个基于 Java 的持久层框架，Spring MVC 负责请求的转发和视图管理，Spring 实现业务对象管理，Mybatis 作为数据对象的持久化引擎。

3.彭成辉问：图书管理系统读者如何实现借阅功能？

答：通过 jdbc 连接创建的 library 数据库，在 tomcat 运行呈现的网页上，读者点击 jsp 编译的全部图书页面的借阅按钮进行借阅，此时，个人借阅功能中会新增借阅日志，借阅的图书出于为归还状态，借阅按钮也会变成归还按钮，管理员的借阅日志中也会新增读者借阅的图书。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了图书管理系统的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：69　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">信院闲置物品交易平台</td></tr>
<tr><td>姓　　名</td><td>程珍春</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>34</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
提问：管理员和用户怎么区别实现的？
答：使用过滤器。首先把用户登录时的用户名保存到 session 范围中，然后在过滤器中根据用户名判断过滤哪个 jsp 页面，这样就可以达到用户权限的管理了。

提问：该系统由什么功能？
答：其主要功能有用户注册、登录、浏览闲置、发布闲置、购买闲置、留言、收藏、个人中心、退出等功能。管理员模块主要是用来对系统进行一个粗略管理，其主要功能为登录、闲置管理、订单管理、用户管理、退出等操作。

提问：系统用到了哪些开发工具？
答：系统后端用了 IDEA 实现、前端运用了 WebStrom，数据库使用的时 MySQL，基于 SSM 框架开发而成，设计过程中页面通过 element-Ui 框架完成。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 7 月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了闲置物品交易平台的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：65　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的天气数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>黄思琪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>35</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：数据有多少，有哪些字段。
答：有两个数据，一个有六万多条，有省级行政区、城市、小时、温度、风力方向、风级、降水量、相对湿度和空气质量 9 个字段。另一个有三万多条，有省级行政区、城市、日期、天气、最低气温、最高气温、风向 1、风向 2 和风级 9 个字段。
问：数据是如何进行清洗的。
答：用 python 进行缺失值查询，如何对缺失值补 0，最后对数据进行重复值查询。
问：可视化汇图用了什么包。
答：用了 matplotlib.pyplot、pyecharts 和 pyecharts.charts。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了天气数据分析与可视化的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的全国各地房价数据的数据爬取与可视化分析</td></tr>
<tr><td>姓　　名</td><td>凌慧</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550436</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题 1：爬取数据的数据量多大？

答：使用 Python 爬虫技术，从链家网上爬取数据，通过模拟用户浏览的过程来获取数据，爬虫通过分析和过滤进行资源的获取，根据 HTML 的相关知识分析爬取网站的结构，调用函数库中的函数，通过 URL 来确定对链家网站全国房价进行资源请求，获取页面信息，通过正则表达式来解析页面信息，爬取大约五万条数据，最后通过 CSV 文件存储。

问题 2：使用哪个软件做的清洗？做了哪些清洗？

答：使用 kettle 软件进行数据清洗。用 kettle 进行数据去除缺失值和重复值，选择 kettle 流程类别中的过滤记录步骤组件过滤掉缺失值，用以 Excel 文件输出。再选择 kettle 的转换类别中的排序记录步骤进行排序，然后选择 kettle 的转换类别中的去除重复记录步骤，最后以 Excel 文件输出。

问题 3：数据爬取时间多长？

答：由于使用的 python 爬虫技术爬取数据，数据量较大，且需要网络，因此爬取数据的时间较长，大约需要 1 小时。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了全国各地房价数据的数据爬取与可视化分析的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：67	答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

课题名称	基于 Python 的当当网书籍数据的数据分析与可视化						
姓　　名	杨明明	院（部）	网络空间安全学院	专业/班级	大数据1904	学号	201927550438
指导教师	熊旺	所在部门	网络空间安全学院	职　务（职 称）	讲师		
答辩小组成员	姓名	职务（职称）					
	唐紫珺	讲师					
	彭成辉	讲师					
	刘彪	讲师					

答辩记录：
唐紫珺：爬了多少数据？
杨明明：清洗前爬了五万多条，清洗后爬了一万五多条数据。
唐紫珺：清洗了那些数据？
杨明明：去掉多余的符号，比如价格前的“¥”，出版社名称前的“/”，或者评论数里的多余的数字单位“条评论”。还有删除了名称一样的书籍，以及有空值数据的书籍等等。
刘彪：过程中是否有反爬？
杨明明：没有。
唐紫珺：可以考虑将一些冷门书籍中的 0 数据做一个统计绘图，不然数据清洗掉的数据太广了。
杨明明：好的。
彭成辉：爬取数据是用到了什么库？
杨明明：在爬取数据时用到了 requests 库进行请求，另外在数据分析时用到了 lxml 库用来提取图书 HTML 网页信息，在数据可视化时用到了 Matplotlib 包等等。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 7 月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了当当网书籍数据的数据分析与可视化的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：65　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

、

湖南信息职业技术学院 2019 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的人口数据爬虫与可视化分析</td></tr>
<tr><td>姓　　名</td><td>杨斯宇</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>39</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问：数据采集用到了哪个库？采集到的数据是如何处理的？
答：数据采集直接用了 requests 库，采集到的数据直接保存为 excel 文件。
问：数据可视化部分用的是什么包？
答：数据可视化用的是 pyecharts。
问：数据可视化绘了哪些图？
答：数据可视化绘制了饼图、折线图、柱状图。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了人口数据爬虫与可视化分析的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">网易云音乐歌单数据的采集与可视化</td></tr>
<tr><td>姓　　名</td><td>郑津</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>40</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题 1：如何爬取数据，爬取了多少数据？
回答：获取网站地址，获取 User-Agent，找到歌单风格网络链接，通过请求 url，爬取了 3 万多的数据，后通过清洗数据，还剩 1 万多数据。
问题 2：有没有反爬？
回答：因为是每个风格的歌单单独爬取，所以没有被反爬。
问题 3：爬了多久？
回答：每一个风格爬了 15 分钟左右，一共爬了 4 个半小时左右。
问题 4：一般哪些地方存在空值？
回答：一般是歌单收藏量、转发量、评论量和标签存在空值。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了 Boss 直聘就业职位数据分析与可视化的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：65	答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的瓜子二手车数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>周诗妍</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>41</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.爬虫使用的库？
答：使用的 requests 库。利用 requests 库发起 get 请求，从而获取页面信息，再通过 re 库从源代码中批量精确匹配到想要相关的信息，最后再以 csv 文件类型保存数据。
2.爬虫总共爬取了多少条数据，爬取了多久？
答：爬取了 16000 多条，爬取了一晚上。
3.数据进行了哪些清洗步骤？
答：爬取的数据并不十分混乱，只需进行补全缺失值和删除重复值这两个步骤

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了瓜子二手车数据分析与可视化的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的 Boss 直聘就业职位数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>祝晓湘</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>2019275504 42</td></tr>
<tr><td>指导教师</td><td>熊旺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、这个你爬了多少条数据？

答：在 boss 直聘网站的一共爬取了 3 万条左右关于就业职位的数据，然后再进行清洗，分析，可视化等。

2、经过清洗后有多少？

答：还有 2 万多，2 万 5 千条的样子，包括的字段有最低工资，最高工作，公司名称，平均工资，学习等。

3、你的数据清洗是用什么做的？

答：pandas，pandas 是基于 NumPy 的一种工具，提供了高效地操作大型数据集所需的工具。pandas 提供了大量能使我们快速便捷地处理数据的函数和方法。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了 Boss 直聘就业职位数据分析与可视化的实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的安居客长沙地区房租可视化</td></tr>
<tr><td>姓　　名</td><td>王天豪</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>23</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
一、爬取安居客数据的时候是否遇到过反爬虫技术？你是怎么解决的。
答：遇到了，我是通过更换采用改变 header 头的方式防止请求过于频繁导致 IP 被封，虽然繁琐但是效果显著。
二、你爬取这些数据一共花费了多少时间？
答：因为安居客一个词条页数搜索最多 50 页，所以我需要不断更换 url 和 cookie 来达到爬取想要所有数据的目的，算上更换 url 的时间差不多 17 分钟
三、数据清洗前后有什么区别？以及数据差距
答：因为爬取的数据需要组合，我使用的是 Windows 自带的 cmd 复制命令导致有部分的重复数据和乱码，以及有部分非长沙地区的数据，数据清洗前 30000 条左右数据清洗后 10000 多条

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：
答辩合格

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 ssm 的网上订餐系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>朱清文</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550433</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.ssm 框架分别指的是什么？

答:ssm 全称 Spring+SpringMvc+MyBatis。Spring 是一个轻量级的控制反转和面向切面的容器框架，springMVC 使用了 MVC 设计模式，对 web 层进行解耦，开发更简洁，MyBatis 是一个基于 java 的持久层框架。

2.如何实现管理员和普通用户登陆页面的验证？

答:使用表单制作 jsp 登录页面，发送请求通过 controller 调用 service 层与数据库进行交互，获取数据进行验证登录。

3.B/S 模式是什么？

答:B/S 开发模式是以浏览器和服务器结构开发软件的模式，这种模式统一了客户端，将系统功能集中到服务器上，简化了系统的开发、维护和使用。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

答辩合格

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的汽车品牌销量数据分析可视化</td></tr>
<tr><td>姓　　名</td><td>邓奕雄</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>01</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

问题一：爬虫爬了多少条数据？

答案：爬虫有三个部分，第一个部分爬取 10500 左右条数据，第二部分爬取 7800 左右条数据，第三部分爬取 31750 条数据。

问题二：你清洗数据的数据是怎么清晰的？

答案：我在爬取数据的过程中就将爬取的数据结合成了 dataframe 形式存入 csv 文件，因为是循环存入后面追加的 dataframe 数据的列索引也变成了数据，所以我清洗的数据就是将后边 dataframe 的列索引删除。

问题三：数据存入了哪些数据库？

答案：全国经销商数据存入了全国经销商数据库中，车型信息和汽车品牌销量存入汽车品牌数据库中。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　69　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 实现八哥电影可视化分析</td></tr>
<tr><td>姓　　名</td><td>邓在平</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据 1904</td><td>学号</td><td>02</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
问：爬虫用什么解析的页面？
答：使用 xpath 解析 html 页面
问：数据怎么清洗的？
答：使用 mapreduce 清洗，去除空值、重复值，删除部分不完整数据
问：数据可视化画了几张图？展示一下
答：3 张，各类电影类型柱状图、电影语言柱状图、使用 pyEcharts 模块绘制饼图等

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：答辩合格

成绩：　65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">学生信息成绩管理系统</td></tr>
<tr><td>姓　　名</td><td>冯涛</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>03</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：系统有几种用户？
答案：一共有 3 种用户，分别是管理员，教师和学生这三种用户

问题二：你系统一共构建的多少数据库表？
答案：我这个系统一共设计了 9 个数据库表，有管理员表、教师表、学生表、学期表、专业表、班级表和成绩表等等。

问题三：学生用户主要有哪些功能？
答案：学生用户主要有学习信息管理、学生选课、和成绩管理等功能

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，但是除了文字和格式还需进一步调整，整个系统还需要不断完善，基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　60　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 实现中国 2000-2015 各项发展数据可视化分析</td></tr>
<tr><td>姓　　名</td><td>贺兴林</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>04</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

你爬了多少数据清，洗完之后还剩多少条数据？
10 万左右，1 万多。
使用什么清洗的？
首先使用 pandas 的 iloc 通过行列索引得到想要的数据再将这数据分别存入数据库
用什么包画的图？画的什么图？
使用 matplotlib 包绘图画的图包括折线图，饼图，柱状图，横向柱状图

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：
答辩合格

成绩：　68　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 世纪佳缘网站的男女婚姻数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>黄惠</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550405</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、这个你爬了多少条数据？

答：先导入 requests 包之后，使用 get 函数，采取 post 的方式，对“性别”，“年龄”，“学历”等 10 个字段进行爬取，并通过性别字段进行分类，并放入定义好的列表当中，大约 10 万左右。

2、jinja2 是用来做什么的？

答：jinja2 是 Flask 作者开发的一个模板系统，起初是仿 django 模板的一个模板引擎，为 Flask 提供模板支持，由于其灵活，快速和安全等优点被广泛使用。该模板在 Python 的 web 开发中广泛使用，它能够有效的将业务逻辑和页面逻辑分开，使代码可读性增强、并且更加容易理解和维护。

3、你用什么做的清洗？

答：首先将数据爬取之后，导入 pandas 模块然后

1，使用 drop_duplicates 函数删除重复值。

2，使用 drop 函数删除掉不必要的字段，例如“helloUrl”，“randAttr”，等没有用的字段。3，使用 dropna 删除列表当中的缺失值或者空值。

4，使用 describe 方法看看是否有异常值。

5，使用 to_excel 函数保存到本地，并最终完成清洗。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：69　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 爬虫疫情数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>雷圆有</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>06</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

（1）刘彪老师：总共爬取多少条数据？

雷圆有：总共爬取数据十万条，总共有三种数据　全球疫情数据，各市疫情数据以及各省疫情数据。

（2）唐紫珺老师：爬取的字段有哪些？

雷圆有：新冠新增人数　治愈人数，　死亡人数，确诊人数，日期，累计死亡人数。新增死亡，国家。

（3）唐紫珺老师：可视化用的什么包？

雷圆有：主要是用到 matplotlib　这个是绘图的的，pandas 数据分析包，pyecharts，数据图表这三个。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，个别图表需要修改，基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">财务系统管理系统的设计与实现</td></tr>
<tr><td>姓　　名</td><td>李志高</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>08</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 管理员和普通员工之间的权限怎么解决？

做一个判断，登陆界面选择管理员或员工，选择不同的登陆给予不同的权限。

2. 程序中连接数据库怎么实现的？

下载 MySQL 驱动，部署到环境中，连接数据库，创建 statement 对象，关闭连接。

3. 部门信息写入怎么实现的？

输入数据都通过 form 表单中定义的方法 onsubmit="return checkForm()"来检查，checkForm()函数中是各种的校验输入数据的方式。

会议主持人：唐紫珺　　记　录　人：彭成辉　　2021 年 7 月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：67　　答辩小组（组长）签名：刘 彪　　2021 年 7 月 29 日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的前程无忧网招聘信息的爬取与可视化</td></tr>
<tr><td>姓　　名</td><td>廖佳德</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>09</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、爬虫使用的库？

本项目爬虫使用的使 requests 库，使用 get 方法对网页发起访问请求，还使用 XPath 解析 HTML 页面数据，以此来获取所需 URL。

2、爬取的数据量？

答：使用爬虫程序共爬取了 39839 条数据，在使用 kettle 对数据进行去重和去除脏数据和无效数据之后后还剩下 36601 条数据。

3、kettle 清洗做了哪些操作？

答：对工资字段进行拆分、拆分为最小工资和最大工资、并且对其进行清洗、将工资单位统一单位、去除工资为空的行、对数据进行去重、根据参照表新增省份字段、把数据中一些空值替换为不详。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整,基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 java 小区管理系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>廖昭嗣</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>10</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.你使用了什么框架

答：我使用了两个框架，分别是 Springboot、Mybatispuls

2.你对于 ssm 的理解。

答：ssm 全称 Spring+SpringMVC+MyBatis，是目前比较主流的 Java　EE 企业级框架，适用于搭建各种大型的企业级应用系统。ssm 还可指“软系统方法论”，是一项运用系统思考解决非系统问题的定性研究技术。

3.你这个系统实现的功能。

答：主要包括业务管理部门、安保部门和服务部门，业务部门有分为资产管理、业主管理、费用管理、智慧停车管理、报修管理、组织管理、系统管理、巡查管理。

4.数据库用什么方式链接的

答：数据库链接我运用了 jdbc 加 dao

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

该同学在毕业设计答辩过程中，阐述了小区管理系统实现过程，回答问题基本正确。基本完成了毕业设计任务，经答辩小组集体评议后认定该毕业设计成绩为合格。

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的人人影视的数据可视化</td></tr>
<tr><td>姓　　名</td><td>刘秉科</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>11</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 用什么来解析数据的？
   先编写好爬虫的相应部位，在利用 python 里面的一个 etree 里面的 xpath 语法来解析网页数据。
2. 用什么进行的数据清洗？
   利用纯 python 的基础用法来清洗数据，去除不需要的列，去重复值，去除缺失值。
3. 数据用什么存储的？
   利用 python 编写代码，连接上数据库，在把本地的 csv 文件数据写入到 mysql 里面相应的数据库里。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 裤子数据采集与分析爬虫可视化</td></tr>
<tr><td>姓　　名</td><td>刘励</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550412</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：爬取了哪些字段
本课题共爬取了 6 个字段，分别是名称、价格、地区、销量、评论数、店名
问题二：爬了多少条数据，清洗做了什么
利用 requests 库进行爬取，利用 xpath 解析网页，然后爬取网站，开始爬了 2 万多条，得出有些数据有空值，利用 kettle 清洗一下，把空值删除，最后筛选出 1 万多条的清洗的数据。
问题三：可视化运用了那些包
本课题运用了 Pandas 包，numpy 包，bar 包 ,pie 包 ,map 包 ,pie 包 ，pyecharts 包，scatter 包

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 实现手机销售评论数据采集与分析可视化</td></tr>
<tr><td>姓　　名</td><td>刘梁有</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550413</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：用什么清洗，清洗做了什么
使用 pycharms 对已爬取的数据进行去重，删除空值的清洗处理
问题二：用了什么防止反爬
使用 fake-useragent 进行伪装生成 header 请求访问，这样可以频繁的更换 ua 防止爬虫中断
问题三：用什么解析网页
使用 beautifulsoup 从 HTML 中获取数据，解析网页获取标签信息。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：
答辩合格

成绩：　69　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的长沙各地区薪资水平的分析与可视化</td></tr>
<tr><td>姓　　名</td><td>罗彬</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>14</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：
问题一：爬虫爬了多少条数据？
答案：爬了 50 同城 7 万条数据跟 boss 直聘 3 万条

问题二：你清洗数据的数据是怎么清晰的？
答案：我在爬取数据的过程中就将爬取的数据结合成了 dataframe 形式存入 csv 文件，因为是循环存入后面追加的 dataframe 数据的列索引也变成了数据，所以我清洗的数据就是将后边 dataframe 的列索引删除。

问题三：数据存入了哪些数据库？
答案：存入了 MySQL

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，个别图表需要修改

成绩：　65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Java 的学生管理系统设计与实现</td></tr>
<tr><td>姓　　名</td><td>罗雄超</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>15</td></tr>
<tr><td>指导教师</td><td>刘彪</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1.系统面向的用户？实现了哪些功能？

答：系统面向的用户是教师。主要包括系统管理模块，教师管理模块和成绩管理模块这三大模块。本系统对学生成绩进行统计、分析，方便教师等快速把握学生、班级、学校的学习状况。

2.数据库创建了哪些表？

答：本系统数据库由 7 张表组成，分别为 tb_Department（所属院系表）、tb_Major（专业表）、tb_Classe（班级表）、tb_Student（学生表）、tb_User（用户表）、tb_Course（课程表）、tb_Score（成绩表）。

3.连接数据库和构建图形界面运用的是哪些类库？

答：本系统使用 JDBC 类库连接数据库，构建图形界面运用的类库则是 swing。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

答辩合格

成绩：　66　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 python 的品牌耳机数据分析与可视化</td></tr>
<tr><td>姓　　名</td><td>莫敏</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>201927550416</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1. 数据清洗用的是什么工具 ？

答：数据清洗，我用的是 Kettle 工具，主要是清洗数据中的空值、空值替换、重复值及剪切字符串。空值清洗是在商品名称上做空值查询和删除，重复值查询也是在商品名称上查找和删除，空值替换是把评论数字段是空值的替换成 0，而剪切字符串是在地区字段上剪切的是我所需要的省份。

2.做可视化用的什么包？

答：数据做可视化用到的是 pyecharts 包和 pandas 包，pyecharts 是一个用于生成 Echarts 图表的类库，用 它生成的图可视化效果非常棒，pyecharts 还可以可以生成独立的网页，而 pandas 包，就用来读取我们需要的数据文件，为了解决数据分析任务。

3.爬虫有用到反爬虫技术吗？

答：没有用的，我们在爬取网页信息数据的过程中，不能打开文件，只有在爬完后才能打开文件查看数据，如果文件中没有数据，就是淘宝网上的淘宝号被暂时封了或者 cookie 更新了爬不到数据，要是淘宝网封了就多刷新几次，再换一个 cookie 就可以继续爬取数据了。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于大数据技术实现股票数据的传输与存储可视化</td></tr>
<tr><td>姓　　名</td><td>欧阳灿林</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>17</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
<tr><td colspan="8">答辩记录：
唐紫珺问：股票数据怎么来的?
股票数据为离线数据，spark 课程提供的数据集。
彭成辉问：图不够规范
通过 pyecharts 库完成标准绘图。
刘彪问：清洗用的什么？
用 spark core 对股票数据按照需求过滤不符要求的数据。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日</td></tr>
<tr><td colspan="8">答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整，个别图表需要进行修改

成绩：65　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日</td></tr>
</table>

湖南信息职业技术学院 2022 届学生毕业设计
评阅、答辩及成绩评定表

<table>
<tr><td>课题名称</td><td colspan="7">基于 Python 的全国旅游数据分析可视化</td></tr>
<tr><td>姓　　名</td><td>唐一涛</td><td>院（部）</td><td>网络空间安全学院</td><td>专业/班级</td><td>大数据1904</td><td>学号</td><td>18</td></tr>
<tr><td>指导教师</td><td>唐紫珺</td><td>所在部门</td><td>网络空间安全学院</td><td>职　务（职 称）</td><td colspan="3">讲师</td></tr>
<tr><td rowspan="4">答辩小组成员</td><td>姓名</td><td colspan="6">职务（职称）</td></tr>
<tr><td>唐紫珺</td><td colspan="6">讲师</td></tr>
<tr><td>彭成辉</td><td colspan="6">讲师</td></tr>
<tr><td>刘彪</td><td colspan="6">讲师</td></tr>
</table>

答辩记录：

1、　　数据清洗用了什么工具，怎么清洗的

答：数据清洗用的是 kettle，先把多个文档合并成一个文档，再选择需要的字段，最后输出文档。

2、　　数据可视化用的是什么库，做了几张图

答：数据可视化是用 Request 库，一共画了 5 张图，二个柱状图，一个折线图，二个中国地理图。

3、用了什么包

答：使用的是正则表达式。

会议主持人：唐紫珺　　记　录　人：　彭成辉　　2021　年　7　月 28 日

答辩小组意见（评语）：

毕业设计内容符合要求，文字和格式还需进一步调整

成绩：　67　　答辩小组（组长）签名：刘 彪　　2021　年　7　月 29　日