|  |  |
| --- | --- |
| 成绩 |  |



毕业设计

|  |  |
| --- | --- |
| 题目： | 基于python的公务员招录数据分析系统 |

|  |  |
| --- | --- |
| 学生姓名： | 代子龙 |
| 学 号： | 1908050205 |
| 所在学院： | 计算机学院 |
| 专业名称： | 数据科学与大数据技术 |
| 届 别： | 23届 |
| 指导教师： | 陈宝国 |
| 职称（学位）： |  |

淮南师范学院教务处制

**淮南师范学院本科毕业论文（设计）**

**诚信承诺书**

1.本人郑重承诺：所呈交的毕业论文（设计），题目《 基于Flink的社交媒体热点实时分析系统的设计与实现》是本人在指导教师指导下独立完成的，没有弄虚作假，没有抄袭、剽窃别人的内容；

2.毕业论文（设计）所使用的相关资料、数据、观点等均真实可靠，文中所有引用的他人观点、材料、数据、图表均已注释说明来源；

3. 毕业论文（设计）中无抄袭、剽窃或不正当引用他人学术观点、思想和学术成果，伪造、篡改数据的情况；

4.本人已被告知并清楚：学院对毕业论文（设计）中的抄袭、剽窃、弄虚作假等违反学术规范的行为将严肃处理，并可能导致毕业论文（设计）成绩不合格，无法正常毕业、取消学士学位资格或注销并追回已发放的毕业证书、学士学位证书等严重后果；

5.若在省教育厅、学院组织的毕业论文（设计）检查、评比中，被发现有抄袭、剽窃、弄虚作假等违反学术规范的行为，本人愿意接受学院按有关规定给予的处理，并承担相应责任。

学生（签名）：

日期： 年 月 日

目录

[第1章 绪论 3](#_Toc131726336)

[1.1研究背景及意义 3](#_Toc131726337)

[1.1.1研究背景 3](#_Toc131726338)

[1.1.1研究意义 4](#_Toc131726339)

[1.2 研究现状 5](#_Toc131726340)

[第2章 相关理论与技术研究 6](#_Toc131726341)

[2.1理论研究领域概述 6](#_Toc131726342)

[2.2相关技术 6](#_Toc131726343)

[第3章系统设计与实现 7](#_Toc131726344)

[3.1系统架构设计 7](#_Toc131726345)

[3.2数据采集模块 8](#_Toc131726346)

[3.3数据处理模块 9](#_Toc131726347)

[3.4数据可视化模块 12](#_Toc131726348)

[3.4.1 Echarts图表设计与实现 12](#_Toc131726349)

[3.4.2 Flask后端与前端交互 15](#_Toc131726350)

[3.5 MySQL数据库设计与实现 16](#_Toc131726351)

[4系统应用与分析 18](#_Toc131726352)

[4.1公务员招录数据的特点 18](#_Toc131726353)

[4.2公务员招录数据分析 18](#_Toc131726354)

[4.3系统运行效果 19](#_Toc131726355)

[第6章 总结和展望 28](#_Toc131726356)

[6.1 总结 28](#_Toc131726357)

[6.2 展望 29](#_Toc131726358)

[参考文献: 30](#_Toc131726359)

[致谢 31](#_Toc131726360)

基于python的公务员招录数据分析系统

学生：代子龙（指导老师：陈宝国）

（淮南师范学院计算机学院）

摘 要： 近年来，公务员考试已成为国内各大高校毕业生和社会人士备受关注的重要就业渠道之一，每年的考试报名人数逐年攀升，每年都有数以百万计的人参加考试。公务员招录信息庞杂、复杂，难以快速准确地获取和分析，因此构建一套高效的公务员招录数据分析系统，旨在通过爬虫技术获取公务员招录信息、通过数据处理分析技术对数据进行加工、并通过可视化技术呈现数据，为公务员招录提供数据支持。本文旨在利用Python编程语言，结合Echarts、Flask、Mysql和Scrapy等技术，构建一套高效的公务员招录数据分析系统，实现对公务员招录信息的高效获取、处理、分析和展示，对于各类考生和招录单位都具有重要意义。

本文的研究内容包括：系统需求分析、系统架构设计、数据爬取、数据处理与存储、数据分析与可视化、系统实现与测试。其中，系统需求分析主要针对公务员招录信息的获取、处理、分析和展示需求，对系统的功能、性能、安全性等进行详细的分析和规划；系统架构设计包括系统结构、模块划分、数据流程、数据库设计等方面；数据爬取模块采用Scrapy爬虫框架对招聘信息进行抓取和存储，数据处理与存储模块采用Mysql数据库对数据进行处理和存储；数据分析与可视化模块采用Echarts可视化框架，通过Python编程语言实现对数据的分析和可视化；系统实现与测试包括系统集成、系统测试和性能测试等方面。

本文旨在通过Python技术，构建一套基于Python的公务员招录数据分析系统，对公务员招录数据进行处理分析，将招录数据以直观的形式呈现。以便实现公务员招录数据的快速获取和分析，有针对性的为报考人员提供帮助，具有重要的理论意义和实际应用价值。

关键词: Python Echarts Flask Mysql数据分析 数据可视化 公务员招录 大数据

**Data analysis system for civil service recruitment based on Python**

Student: Dai Zilong (Faculty Adviser：Chen Baoguo)

(School of Computer Science, Huainan Normal University)

Abstract**:** In recent years, the Civil service examination has become one of the important employment channels that attract much attention from college graduates and social figures in China. The number of applicants for the exam is increasing year by year, and millions of people take the exam every year. The recruitment information of civil servants is complex and complex, making it difficult to quickly and accurately obtain and analyze. Therefore, an efficient data analysis system for civil servant recruitment is constructed, aiming to obtain civil servant recruitment information through crawler technology, process data through data processing and analysis technology, and present data through visualization technology to provide data support for civil servant recruitment. This article aims to use Python programming language, combined with technologies such as Echarts, Flask, MySQL, and Scrape, to build an efficient civil servant recruitment data analysis system, achieving efficient acquisition, processing, analysis, and display of civil servant recruitment information. It is of great significance for various candidates and recruitment units.

The research content of this article includes: system requirements analysis, system architecture design, data crawling, data processing and storage, data analysis and visualization, system implementation and testing. Among them, system requirement analysis mainly focuses on the acquisition, processing, analysis, and display requirements of civil servant recruitment information, and conducts detailed analysis and planning of the system's functions, performance, security, etc; The system architecture design includes system structure, module division, data flow, database design and other aspects; The data crawling module uses the Scrapy crawler framework to crawl and store recruitment information, while the data processing and storage module uses a MySQL database to process and store the data; The data analysis and visualization module adopts the Echarts visualization framework, and achieves data analysis and visualization through the Python programming language; System implementation and testing include system integration, system testing, and performance testing.

This article aims to construct a Python based civil service recruitment data analysis system using Python technology, process and analyze civil service recruitment data, and present the recruitment data in an intuitive form. In order to achieve rapid acquisition and analysis of civil servant recruitment data, and provide targeted assistance for applicants, it has important theoretical significance and practical application value.

Keywords**:** Python Echarts Flask Mysql Data analysis, data visualization, civil service recruitment, big data

# 绪论

## 1.1研究背景及意义

### 1.1.1研究背景

公务员招录是国家公务员制度的重要组成部分，也是政府重要的人才选拔渠道之一。近年来，随着社会的发展和人才竞争的激烈化，公务员考试的报考人数也不断攀升。自2009年国考报名人数首次突破100万，此后10余年，国考报名人数都在百万以上。2023年度国考计划招录3.71万人，相比2022年度扩招18.7%。截至11月3日17时30分，2023年度国考报名总人数达2504933人，与2022年同期相比增加近48万人，增幅23.8%，首次突破250万人。由此可见，公务员招录已成为了广大青年学子和社会人士追求稳定就业的热门选择之一。这庞大的数据规模和不断增长的数据量，使得人工处理数据的效率低下，不能满足现代社会对数据分析处理的快速需求。对于相关部门而言，如何快速、准确、高效地处理和分析招录数据成为了一个紧迫的问题。

数据分析在决策中的价值非常重要，可以帮助决策者更好地了解各种因素和变量，优化决策方案，预测未来的趋势和变化，推动组织的创新发展。通过对数据进行深入分析，可以发现很多潜在的问题和趋势，从而提供决策者更全面的信息，帮助其作出更准确的决策。基于数据分析的结果和结论，决策者可以对决策方案进行优化和调整，提高决策的效果和成功率，减少决策带来的风险和损失。

爬虫技术是一种自动化获取互联网数据的技术，常常应用在数据收集和数据分析领域。如网络舆情监测：通过爬虫技术，可以快速抓取各大媒体、论坛、社交媒体等平台上的相关信息，帮助企业、政府等对舆情进行监测和分析。产品竞品分析：通过爬虫技术，可以抓取竞品的价格、评论、销量等信息，帮助企业进行产品竞品分析，优化产品策略。数据挖掘：通过爬虫技术，可以抓取大量的数据，帮助企业或个人进行数据挖掘、分析和建模，从而得出有价值的结论。爬虫技术在数据收集中的应用非常广泛，可以帮助企业、政府、学者等快速获取大量的数据，从而支持各种决策和研究工作。但需要注意的是，爬虫技术的使用需要遵守相关法规，避免侵犯他人的权益。

在此背景下，基于Python的公务员招录数据分析系统应运而生。Python作为一种高级编程语言，具有易学易用、免费开源等特点，已成为数据科学领域最为常用的编程语言之一。此外，Python社区拥有丰富的第三方库和工具，可以实现数据爬取、处理、分析和可视化等一系列操作。因此，基于Python的公务员招录数据分析系统能够快速、高效地从招录网站爬取数据，实现数据处理、分析和可视化，为报考者提供决策支持。

随着互联网技术的不断发展，爬虫技术和数据处理分析和可视化等相关技术的研究也得到了广泛的关注。爬虫技术和数据处理分析和可视化技术的应用越来越广泛，为公务员招录数据的处理和分析提供了新的途径和工具。

公务员职位种类繁多，岗位竞争激烈，了解招录信息可以帮助报考者确定自己的职业方向和需求，选择适合自己的职位，提高报考成功率。也可以帮助报考者提前了解竞争者的情况，制定更好的备考策略，提高自己的竞争力，增加报考成功的机会。因此，本研究旨在结合爬虫技术和数据处理分析和可视化技术，建立一套完整的公务员招录数据分析系统，提供全面、准确、直观的公务员招录数据分析结果，为广大考生提供更好的参考和帮助，也为公务员招录的管理和决策提供依据。

### 1.1.1研究意义

随着社会的不断发展和经济的不断壮大，公务员已经成为越来越多人向往的职业。然而，公务员招录考试的报名人数也越来越多，报考者面临着激烈的竞争。因此，对公务员招录考试的数据进行分析和可视化处理具有很大的研究意义。

通过对历年公务员招录考试数据的分析，可以更好地了解公务员招录考试的趋势和规律，有助于报考者制定更为科学合理的备考计划和提高通过率。对历史招录数据进行分析，为招录决策提供支持和参考，提高决策的科学性和可靠性。通过对公务员招录考试数据的可视化处理，可以更加清晰直观地展示招录情况，为政府部门的决策提供参考。对于相关的招录机构来说，也可以通过数据分析和可视化处理更好地了解考生的报考情况和地域分布等信息，从而进行更加精准的招录计划。

而爬虫技术、数据处理分析和可视化等相关技术则为公务员招录数据分析提供了强有力的支持，能够实现对各类公务员招录信息的快速获取和分析处理。基于Python的公务员招录数据分析系统的研究和开发，不仅有利于提高公务员招录数据的获取和处理效率，也具有广泛的研究和应用价值，将为公务员招录提供更加高效、全面、精准的数据支持。

## 1.2 研究现状

随着数据科学与人工智能技术的发展爬虫技术、数据处理分析和可视化技术在许多领域得到了广泛应用。，越来越多的研究者开始关注如何应用这些技术来解决社会现实问题。在研究方面，越来越多的学者开始关注如何利用爬虫技术获取数据并进行分析和可视化。其中，以机器学习和深度学习为代表的数据处理分析技术在文本数据、图像数据、语音数据等方面均取得了重要进展。同时，基于这些技术的可视化方法，如热图、散点图、词云等也成为了研究数据的常用手段。以互联网舆情分析为例，爬虫技术被广泛应用于从社交媒体、新闻网站等获取大量文本数据，通过自然语言处理和文本挖掘技术进行情感分析、主题分析等，然后将结果可视化展示。这种分析方法对于政府、企业等进行品牌评估、舆情监控等具有重要的意义。

公务员是指在政府机关、事业单位、国有企业等行政机构中从事管理、执法、服务等工作的职员。由于公务员岗位待遇优厚、稳定性高、福利待遇好等优点，所以每年都有大量的考生参加公务员考试。一些研究者也关注公务员招录的市场化程度和与经济发展的关系。公务员职位的数量和结构是否与经济发展相适应；市场化程度高的地区公务员待遇是否更高，考试难度是否更大等。

近年来爬虫技术、数据处理分析和可视化技术在公务员招录领域得到了广泛应用，并且基于Python的公务员招录数据分析系统也得到了研究和实践。在这些相关研究的基础上，本文将提出一种基于Python的公务员招录数据分析系统，旨在为公务员招录相关人员提供数据分析和决策支持。

# 第2章 相关理论与技术研究

## 2.1理论研究领域概述

本文《基于Python的公务员招聘数据分析系统》是基于多个相关的理论和技术研究领域。理论研究领域包括数据收集、数据预处理、数据分析和可视化。在数据收集方面，该系统使用网络爬虫技术从公务员招聘的官方网站中提取信息。在公务员招录数据分析系统中，爬虫技术用于从招录网站上自动化地获取招录信息。其实现原理主要是通过模拟人类浏览器的行为，自动访问网页并解析其中的信息。在公务员招录数据分析系统中，数据处理分析用于对采集的公务员招录数据进行整合和处理，以获取各类统计数据和分析结果。数据预处理涉及到对原始数据进行清洗和过滤,确保其质量并准备好进一步分析使用。可视化则采用交互式动态图表等方式将结果更直观有效地传达给用户。

《基于Python的公务员招录数据分析系统》的理论研究领域主要包括数据挖掘和爬虫技术、数据处理和分析技术以及可视化技术。这些技术可以协同工作，构建出一个强大、高效的公务员招录数据分析系统，为报考人员提供有力的支持与帮助。

## 2.2相关技术

Python是一种通用型编程语言，可以帮助开发者完成各种数据处理、分析和可视化的任务，同时也可以使用Python作为爬虫技术的实现工具。在这篇论文中，Python是这篇论文的核心技术，用于实现数据爬取和处理、分析以及可视化等功能。在数据获取方面，Python的爬虫技术发挥了重要作用。爬虫程序可以根据设定的规则和策略，自动从网站中获取数据。在该论文中，作者使用Python语言编写爬虫程序，获取公务员招录数据，实现了数据的自动化获取。

在数据处理和分析方面，Python的Pandas库、Numpy库等成为了不可或缺的工具。Pandas库是Python中最受欢迎的数据处理库之一，提供了高效的数据操作、清洗和转换等功能。Numpy库是Python中用于科学计算的基础库之一，提供了高效的数组和矩阵运算等功能。在该论文中，作者使用了这些库进行数据处理和分析，以方便用户更好地理解和分析数据。

Echarts是一个非常优秀的可视化工具，可以快速生成各种图表。在该论文中，作者使用了Python中的pyecharts库，结合Echarts实现了数据可视化分析的功能，展示了招录人数、报考人数等各种招录数据的可视化效果。同时，Matplotlib、Seaborn等库也提供了丰富的数据可视化功能，可以满足不同需求的数据分析和可视化。

Flask框架则是一种轻量级的Web应用框架，可以方便快捷地开发小型Web应用。在该论文中，作者使用Flask框架实现了一个简单的Web应用，以便用户可以方便地通过浏览器访问数据分析结果。Flask框架通过路由配置实现URL映射、请求响应处理等功能，同时还可以与数据库、模板等组件集成使用。

MySQL作为一个开源的关系型数据库，也是这篇论文所使用的技术之一。MySQL可以方便地存储和管理数据，支持SQL语言进行数据查询和操作。在该论文中，作者使用MySQL数据库存储了爬取到的公务员招录数据，同时使用SQL语言进行数据查询和操作。这为用户提供了更加灵活的数据操作方式，可以方便地根据需要进行数据查询和筛选。

Anaconda是一个开源的Python发行版，包含了Python解释器、许多科学计算和数据处理相关的库以及一个方便的集成开发环境。Anaconda使得Python的科学计算和数据处理更加方便和快速，Anaconda包括了常用的科学计算和数据处理相关的库，比如Numpy、Pandas、Matplotlib、Scipy等，同时还支持虚拟环境的创建和管理，方便不同项目的库管理。Jupyter Notebook可以实现代码、文本和可视化结果的交互式展示，方便进行数据处理和分析的交互式编程。

# 第3章系统设计与实现

## 3.1系统架构设计

《基于Python的公务员招录数据分析系统》的系统架构设计主要分为数据获取、数据处理、数据分析和数据可视化四个模块。

数据获取模块通过Python编写的爬虫程序，从官方招考网站上获取公务员招录相关信息，包括岗位信息、报考人数、招录人数等，并将获取到的数据存储到MySQL数据库中。数据处理模块使用Python中的Pandas库和Numpy库，对获取到的数据进行清洗和整理，并且进行统计分析，得到各种指标的数值。

数据分析模块主要使用Python的数据分析库进行分析，包括描述性统计、假设检验、回归分析等，以及相关的数据可视化方法，如Matplotlib、Seaborn、Echarts等。同时，该模块还支持用户自定义分析方法和指标，以满足用户不同的需求。

数据可视化模块则是将分析结果以可视化的方式呈现给用户，提供直观的图表展示和数据可视化效果。这个模块使用了Python中的pyecharts库结合Echarts图表库，提供了各种类型的图表展示，如柱状图、折线图、饼图等。

为了方便用户访问和使用分析结果，该系统还实现了一个Web应用程序，使用了Flask框架实现了简单的Web界面。用户可以通过浏览器访问系统的Web界面，查看数据分析结果和可视化图表，，方便用户对公务员招录数据进行更直观的分析和理解。

整个系统的架构设计基于Python语言的丰富的生态系统和开源库，使用了MySQL数据库、Flask框架等技术，实现了高效、灵活、易用的公务员招录数据分析系统。该系统不仅可以满足用户对数据分析的需求，而且具有良好的可扩展性和可维护性，可以适应不同规模和复杂度的数据分析任务。

## 3.2数据采集模块

通过Python代码，对公考信息网站进行爬取数据，爬取的数据主要是2023年安徽各地公务员报考信息和招录信息，包括职位代码，地区，部门名称，职位，学历，专业，招考人数，报名人数和合格人数。针对公务员招录信息的网站，通过对该网站URL的拼接和GET请求，获取网站上的职位信息数据，利用BeautifulSoup对HTML页面进行解析，并遍历表格中的所有行，将其中的信息写入到本地的CSV文件中。具体实现步骤如下：

1：确定目标网站的URL链接，该爬虫代码针对的是“https://ah.huatu.com/zt/zwb/search/”这个网站，因此在代码中定义了一个变量base\_url，用于存储目标网站的URL链接。

2：在定义base\_url变量的同时，还定义了一个page\_params变量，该变量用于存储网站翻页的参数，即“?page={}”。

3：创建一个CSV文件并写入表头，该代码中使用了Python内置的csv模块进行文件操作，利用该模块创建了一个名为“job.csv”的CSV文件，并写入了表头，包括职位代码、地区、部门名称、职位、学历、专业、招考人数和报名人数。

4：通过循环遍历网站上的前179页数据（该网站共有178页数据），将每一页数据的URL链接通过base\_url和page\_params拼接出来，然后发送GET请求，获取返回的HTML页面。

5：利用BeautifulSoup解析HTML页面，找到所有职位信息所在的表格。

遍历表格中的所有行并写入CSV文件中，需要注意的是，有些表格中的行并不是职位信息，而是表格中的表头，因此需要判断一下当前行是否是表头，如果是，则跳过该行不处理，否则将该行的信息写入CSV文件中。

## 3.3数据处理模块

数据处理模块实现使用的是Jupyter Notebook, Jupyter Notebook是基于网页的用于交互计算的应用程序。其可被应用于全过程计算：开发、文档编写、运行代码和展示结果。还使用了SQL语句可以用于筛选和处理数据从关系型数据库中获取特定的数据，对数据进行筛选、排序、聚合等处理操作，并生成新的数据集合。具体实现步骤如下：

1: 读取datam目录下名为 '/job\_data.csv' 的CSV文件

2：查看缺失值情况，将position\_code列删除，，将number\_of\_recruits和number\_of\_applicants列的数据类型转换为int。然后，规范major列的数据。将major列中的NaN值替换为空字符串。

3：数据分组：使用 Pandas 的 groupby() 方法按部门和职位对数据进行分组。数据统计：对每个部门和职位的招聘人数和申请人数进行统计，使用 Pandas 的 sum() 方法对每个分组的数据进行求和。数据去重：对统计结果进行去重，使用 Pandas 的 drop\_duplicates() 方法删除重复的部门项和职位项。

4：SELECT position, number\_of\_recruits FROM job\_position ORDER BY CAST(number\_of\_recruits AS UNSIGNED) DESC LIMIT 10：选取安徽省2023年度考试录用公务员招录人数最多的十大职位

5：SELECT position, number\_of\_applicants FROM job\_position ORDER BY CAST(number\_of\_applicants AS UNSIGNED) DESC LIMIT 10：选取安徽省2023年度考试录用公务员报考人数最多的十大职位

6：SELECT area, COUNT(DISTINCT position) AS position\_count FROM job\_gk WHERE number\_of\_applicants = 0 GROUP BY area HAVING COUNT(\*) > 0：选取安徽各地区无人报考职位数量

7：SELECT position\_code, area, position, degree, major, number\_of\_recruits FROM job\_gk WHERE number\_of\_applicants=0：选取安徽各地区无人报考职位信息

8：SELECT department, number\_of\_recruits FROM job\_department ORDER BY CAST(number\_of\_recruits AS UNSIGNED) DESC LIMIT 10：选取安徽省2023年度考试录用公务员招录人数最多的十大部门

9：SELECT department, number\_of\_applicants FROM job\_department ORDER BY CAST(number\_of\_applicants AS UNSIGNED) DESC LIMIT 10：选取安徽省2023年度考试录用公务员报考人数最多的十大部门

10：SELECT area, COUNT(DISTINCT department) AS department\_count FROM job\_gk WHERE number\_of\_applicants = 0 GROUP BY area HAVING COUNT(\*) > 0：选取安徽各地区无人报考职位数量

11：SELECT degree, COUNT(\*) FROM job\_gk GROUP BY degree：选取招录学历分布

12：SELECT position\_code, area, department, degree, major, number\_of\_recruits FROM job\_gk WHERE number\_of\_applicants=0：选取安徽各地区无人报考部门信息表

13：SELECT degree, SUM(number\_of\_recruits) AS recruit\_count, COUNT(position) AS position\_count, COUNT(DISTINCT department) AS department\_count FROM job\_gk GROUP BY degree ORDER BY recruit\_count DESC：选取不同学历的部门数、职位数、招录人数

14：SELECT degree, SUM(number\_of\_applicants) AS applicants\_count FROM job\_gk GROUP BY degree ORDER BY applicants\_count DESC：选取不同学历对应的报考人数

15:SELECTposition,SUM(number\_of\_recruits)AStotal\_recruits, SUM(number\_of\_applicants) AS total\_applicants FROM job\_gk WHERE major LIKE '%不限%' GROUP BY position ORDER BY total\_applicants DESC LIMIT 10：选取十大不限专业职位招录人数和报考人数

16:SELECTdepartment,SUM(number\_of\_recruits)AStotal\_recruits,SUM(number\_of\_applicants) AS total\_applicants FROM job\_gk WHERE major LIKE '%不限%' GROUP BY department ORDER BY total\_applicants DESC LIMIT 10：选取十大不限专业部门招录人数和报考人数

17: SELECT TRIM(SUBSTRING\_INDEX(SUBSTRING\_INDEX(major, '、', n), '、', -1)) AS major, COUNT(\*) AS count

FROM (

SELECT major, 1 + LENGTH(major) - LENGTH(REPLACE(major, '、', '')) AS n

FROM job\_gk

WHERE major != '不限' AND major NOT LIKE '%门%'

) AS t

GROUP BY major

ORDER BY count DESC

LIMIT 10

：选取十大热门专业分布

18: SELECT TRIM(SUBSTRING\_INDEX(SUBSTRING\_INDEX(major, '、', n), '、', -1)) AS major,

SUM(number\_of\_recruits) AS total\_recruits,

SUM(number\_of\_applicants) AS total\_applicants

FROM (

SELECT major, 1 + LENGTH(major) - LENGTH(REPLACE(major, '、', '')) AS n, number\_of\_recruits, number\_of\_applicants

FROM job\_gk

WHERE major != '不限' AND major NOT LIKE '%门%'

) AS t

GROUP BY major

ORDER BY total\_recruits DESC

LIMIT 10;

:选取十大热门专业招录/报考人数

## 3.4数据可视化模块

### 3.4.1 Echarts图表设计与实现

基于 Python 的公务员招录数据分析系统需要对数据进行可视化展示，以便更加直观地观察数据趋势和分布。Echarts 是一个基于 JavaScript 的开源可视化库，可以用来制作各种类型的数据可视化图表，包括折线图、柱状图、饼图、散点图等。Echarts 图表设计的过程主要包含两个步骤：图表类型选择和图表样式设计。首先，需要根据数据分析需求选择合适的图表类型，例如饼图、柱状图、折线图等。其次，在确定图表类型的基础上，还需要逐步设计图表样式，包括颜色、字体、标签以及提示框等。在使用echarts绘图前我们需要为 ECharts 准备一个定义了高宽的 DOM 容器, 然后就可以通过 echarts.init 方法初始化一个 echarts 实例并通过 setOption 方法生成一个简单可视化图。

使用 Python 的 Echarts 库来绘制图表，主要包括以下几个步骤：定义图表对象：使用 Echarts 库提供的图表对象，例如 Line、Bar、Pie、Scatter 等，来定义需要绘制的图表类型。添加数据：使用 add\_xaxis 和 add\_yaxis 方法来添加需要展示的数据，其中 add\_xaxis 方法用于添加 x 轴数据，add\_yaxis 方法用于添加 y 轴数据。可以根据需要添加多组数据，并为每组数据设置不同的系列名称。配置图表：使用 set\_global\_opts 方法来配置图表的全局选项，例如标题、图例、提示框等，以及 x 轴和 y 轴的相关选项。还可以使用 set\_series\_opts 方法来配置每个系列的选项，例如线条样式、柱子样式、标记点样式等。渲染图表：使用 render 方法将图表渲染为 HTML 文件，并保存到指定的目录中。

以柱状图为例，我要绘制安徽省2023年度考试录用公务员职位数据汇总柱状图。

利用 echarts 库实现的动态数据可视化图表。主要分为三个部分：HTML 页面中的 div 元素、Echarts 配置项以及调用 Ajax 异步加载数据。具体实现如下：首先，使用一个 div 元素创建了一个用于呈现图表的容器，并设置其宽度和高度。

html

<div id = "chart" style="width:100%;height:600px">

这表示创建了一个 id 为 chart 的 div 元素，并设置其宽度为 100%，高度为 600px。

接下来，使用 Echarts 官方提供的配置选项来设置图表样式和数据等相关信息。

javascript

option = {

// 设置标题等信息

title: {...},

tooltip: {...},

legend: {...},

// 设置x轴和y轴等样式设置

grid: {...},

xAxis: {...},

yAxis: {...},

// 数据系列，例如柱状图、折线图等

series: [...]

};

其中，通过 title 设置标题信息；通过 tooltip 可以设置鼠标悬停时的提示信息；通过 legend 来设置图例信息；通过 grid 来设置画布布局等等。

最后，通过调用 Ajax 异步加载数据来更新图表。

javascript

// 异步加载数据

var url = '/chart\_data';

var xhr = new XMLHttpRequest();

xhr.open('GET', url, true);

xhr.onload = function() {

if (xhr.readyState === 4 && xhr.status === 200) {

var data = JSON.parse(xhr.responseText);

// 更新x轴和数据

myChart.setOption({

xAxis: {

data: data["area"]

},

series: [

{

name: '岗位数',

data: data["number\_of\_position"]

},

{

name: '招录人数',

data: data["number\_of\_recruits"]

},

{

name: '报考人数',

data: data["number\_of\_applicants"]

}

]

});

}

};

xhr.send();

这里创建了一个 XMLHTTPRequest 对象，并使用 GET 方法请求 '/chart\_data' 路径，获取服务器返回的 JSON 数据。然后通过解析 JSON 数据，将数据更新到图表中。

最后，使用 setOption 方法来将新的配置项和数据应用到图表上。

javascript

myChart.setOption(option);

这里调用了 Echarts 的 setOption 方法，将之前定义的 option 变量作为参数传入，来将图表的样式和数据更新为最新状态。以上就是利用 echarts 库实现数据可视化图表的主要步骤和内容，

### 3.4.2 Flask后端与前端交互

Flask 后端与前端交互可以通过 HTTP 请求和响应来实现。当用户在前端页面上针对某个功能进行操作时，例如点击某个按钮、填写表单等，前端会通过 Ajax 或者表单提交等方式向后端发送请求。后端收到请求后，根据请求的内容与类型，决定如何处理并返回相应的结果给前端展示。在构建公务员招录数据分析系统时，Flask可以用于实现后端API接口，提供数据查询、处理和分析功能。

使用Flask等web框架来搭建服务器端，实现与前端交互的接口，包括数据传递、响应请求等功能。在后端中对数据库中的数据进行查询和处理，然后将处理结果返回给前端。可以在应用程序中定义路由和视图函数，以及使用数据库连接池（如 PyMySQL）连接到 MySQL 数据库。

在Flask后端中，需要编写API接口处理HTTP请求并返回JSON格式的数据。这些API接口可以使用Flask提供的路由和视图函数实现。以绘制安徽省2023年度考试录用公务员招录人数最多的十大职位折线图为例，首先，创建了一个名为 app 的 Flask 实例，并设置了一个名为 secret\_key 的属性。secret\_key 用于为应用程序提供加密和安全性保护等功能。app = Flask(\_\_name\_\_) app.secret\_key = "session"。

接下来，定义了一个路由为 /data，表示访问该接口时需要在网址后面添加 /data。在 get\_data 函数中，连接了一个名为 fb 的 MySQL 数据库，并查询 job\_position 表中招录人数最多的前10个岗位名称和对应的招录人数信息。

conn = pymysql.connect(host='localhost', user='root', password='991109', database='fb', port=3306)

cursor = conn.cursor()

sql = "SELECT position, number\_of\_recruits FROM job\_position ORDER BY CAST(number\_of\_recruits AS UNSIGNED) DESC LIMIT 10"

cursor.execute(sql)

results = cursor.fetchall()

接着，将查询结果以字典形式封装为一个名为 data 的数据对象，其中键为 positions 和 recruits 分别表示岗位名称和对应的招录人数。data = {'positions': [x[0] for x in results], 'recruits': [x[1] for x in results]}。最后，使用 jsonify 方法将数据对象转换为 JSON 格式并返回给前端页面，return jsonify(data)。在主函数中，通过 app.run() 方法启动了 Flask 应用程序，并开启了 debug 模式以方便调试。if \_\_name\_\_ == "\_\_main\_\_":app.run(debug=True)。最后使用 jQuery 的 ajax 方法获取数据的代码块。这里，通过 $.get() 方法向之前定义的 /data 接口发起 GET 请求，获取 MySQL 数据库中查询得到的岗位名称和对应招录人数信息。$.get("/data", function(data, status){// 在这里处理获取到的数据});

总的来说，Flask后端与前端交互的过程可以分为以下几个步骤：

在Flask后端定义路由和视图函数，用于处理HTTP请求并返回JSON格式数据。在前端使用jQuery等JavaScript框架向Flask后端发送HTTP请求，获取JSON格式数据。在前端使用JSON数据更新ECharts等可视化图表的配置项和数据，实现动态更新图表数据的效果。

## 3.5 MySQL数据库设计与实现

MySQL是一种常用的关系型数据库管理系统，它能够提供高效、可靠的数据存储和管理服务。在基于Python的公务员招录数据分析系统中，使用MySQL作为数据存储和管理的工具，能够提高系统的数据处理效率和数据安全性。

首先需要进行需求分析，明确该数据库需要存储哪些信息以及数据之间的联系。根据公务员招录的特点，可以将数据库的信息分为以下几个方面：position\_code area department position degree major number\_of\_recruits number\_of\_applicants。其次，需要根据需求分析结果设计数据库结构，可以设计以下几张表：job\_position：存储职位名称、招录人数和报名人数信息。job\_department：存储部门名称、招录人数和报名人数信息。Job\_jk:存储职位代码 ,地区,部门名称,职位,学历,专业,招考人数.报名人数,合格人数信息。Result\_gk:存储各地区职位，部门，报考人数，招录人数汇总信息。最后是USER表，存储用户账号信息。最后，需要实现数据库表的创建、数据插入、更新和查询等功能，以实现公务员招录数据的有效管理和分析。

在MySQL数据库的实现中，可以使用Python的PyMySQL模块来连接数据库、执行SQL语句、查询数据等操作。例如，可以使用以下代码来连接MySQL数据库：

import pymysql

# 连接MySQL数据库

conn = pymysql.connect(host='localhost', user='root', password='991109', database='fb', port=3306)

# 执行SQL语句

cursor = conn.cursor()

cursor.execute('SELECT \* FROM mytable')

# 获取查询结果

result = cursor.fetchall()

# 关闭数据库连接

cursor.close()

conn.close()

在使用MySQL数据库时，需要注意数据库的安全性和数据的完整性。可以采取一些安全措施，如设置数据库密码、限制用户权限、备份数据等。同时，在数据库的设计和实现中，需要确保数据的正确性和完整性，可以采用一些数据校验和数据约束的方法来实现。

# 4系统应用与分析

## 4.1公务员招录数据的特点

公务员招录数据具有以下特点：规模较大。公务员招录是政府吸纳社会人才的重要渠道之一。每年国家和地方都会组织各种级别和类型的公务员考试，涉及到成千上万个岗位和数十万甚至上百万的报考者。政策性强。公务员招录通常受到相关的法律法规、政策文件等约束，包括招录标准、程序、分值、面试方式等各个方面。招录数据中的政策因素对于分析和预测公务员招录情况至关重要。数据质量高。公务员招录数据通常由政府部门或者官方机构发布，数据来源准确可靠，可以保证数据的质量和真实性。同时，还能够避免数据的失真和篡改等问题。数据多样性。公务员招录数据不仅包括招录岗位和人数等基础信息，还包括报名人员的性别、年龄、学历等基本信息，以及笔试、面试、体检等各个环节的考试成绩和录用结果等详细信息。数据时效性强。公务员招录是一个时间敏感性强的过程，各种招考信息发布、报名、笔试、面试等环节都有着明确的时间节点和周期。因此，公务员招录数据的时效性也很强，需要采用及时、准确的方法进行采集、整理和发布。这些特点决定了公务员招录数据不仅需要进行简单的描述统计，还需要进行深入的分析挖掘。

可以通过数据可视化等方法，对公务员招录数据进行简单的描述统计，例如招录人数、职位数、报名人数、竞争比例、专业分布等方面进行分析。同时，也可以针对不同层级、部门、地区等因素，对公务员招录数据进行比较和分析，以揭示其变化趋势和规律。

## 4.2公务员招录数据分析

2023安徽公务员共报名169905人，招录5974人，录取率仅为3.5%，共有职位数3549个。这意味着报考公务员的人员数量远远超过了招录人数。在这种情况下，竞争越来越激烈，需要拥有优秀的综合素质和能力，才能脱颖而出，获得录用的机会。招录人数最多的职位是人民警察，其次是机关综合。招录人数最多的部门是省白湖监狱管理分局，其次是桐城市乡镇机关。黄山市无人报考职位和部门数量都是安徽省各市中最多的。报考公务员人数最多的三个专业市法学类，中国语言文学，新闻传播学类。关于学历，本科的招录人数，职位数，部门数都是最多的，其次是专科和研究生。

黄山市无人报考职位和部门数量最多，这表明了当地较少的公务员工作机会、低薪水等原因可能使得应聘者不太感兴趣或是应聘者觉得该区域不适合自己生活发展目标，也反映了公务员职业吸引力不同区域之间的差异。

在专业方面，市法学类、中国语言文学、新闻传播学类是报考人数最多的三个专业。这说明这些专业容易涉及到与公务员工作相关的领域，在这些领域里具备一定的专业知识和技能对于公务员职业有助于提升竞争力。此外，本科的招录人数、职位数、部门数都是最多的，显示了本科学历对于公务员招录依然是具有优势的。但是，随着社会上对研究生教育重视的提高，未来研究生学历申请公务员的情况可能会逐渐增多。

综合分析，公务员职位的竞争越来越激烈，报考人数远高于实际招录人数，同时不同地区、职位及专业的公务员招录需求也存在很大差异。因此，应聘者需要根据自身情况，选择合适的职位和部门，并提早做好准备工作，比如加强学习、提升自身能力、提高综合素质和面试技巧等，提高自己的综合素质和能力，以获得更好的职业发展机会。以提高成功率。同时，公务员招录部门需要进一步了解市场需求和求职者的职业偏好，也需要加强对招录机制及政策的研究和制定，为应聘者提供更为公平、公正的招聘体系，促进公务员培养和管理水平的提高。

## 4.3系统运行效果

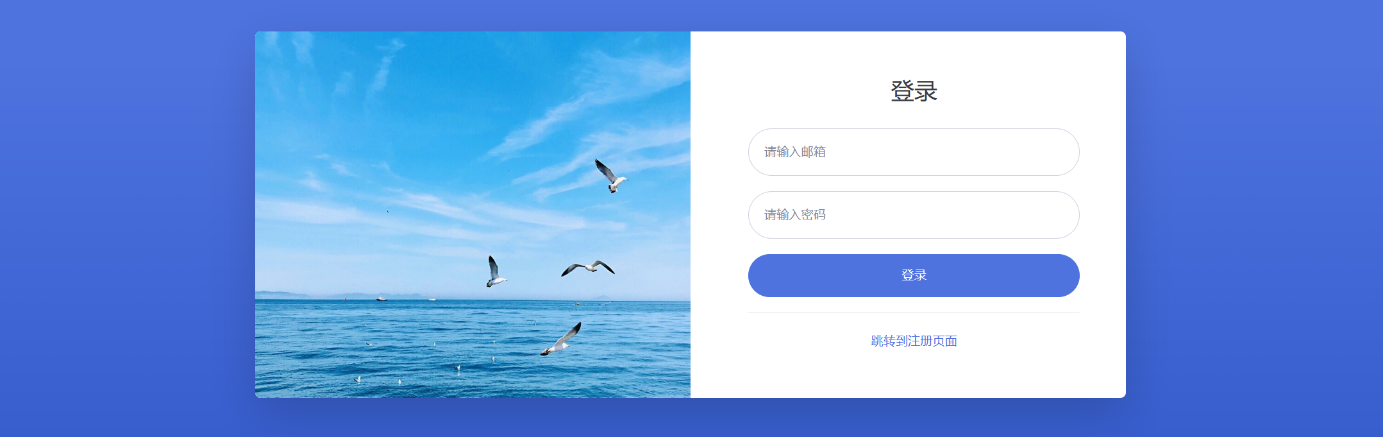


图 1 登录页面



图 2 注册页面



图 3 首页



图 4 招录人数最多的十大职位



图 5 报考人数最多的十大职位



图 6 无人报考职位数量



图 7 无人报考信息表



图 8 招录人数最多的十大职位



图 9 招录人数最多的十大部门



图 10 无人报考部门数量



图 11 无人报考部门信息

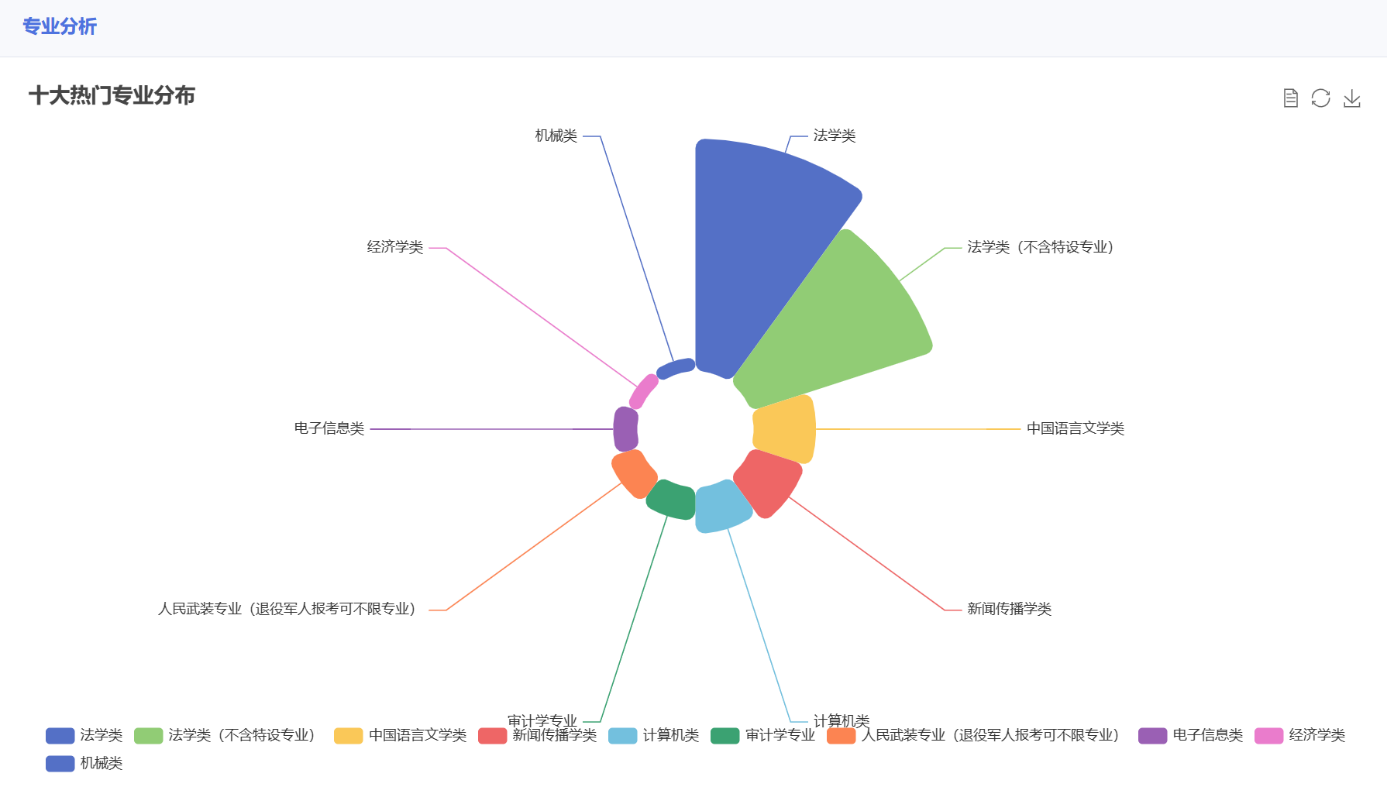


图 12 十大热门专业分布



图 13 十大热门专业招录/报考人数

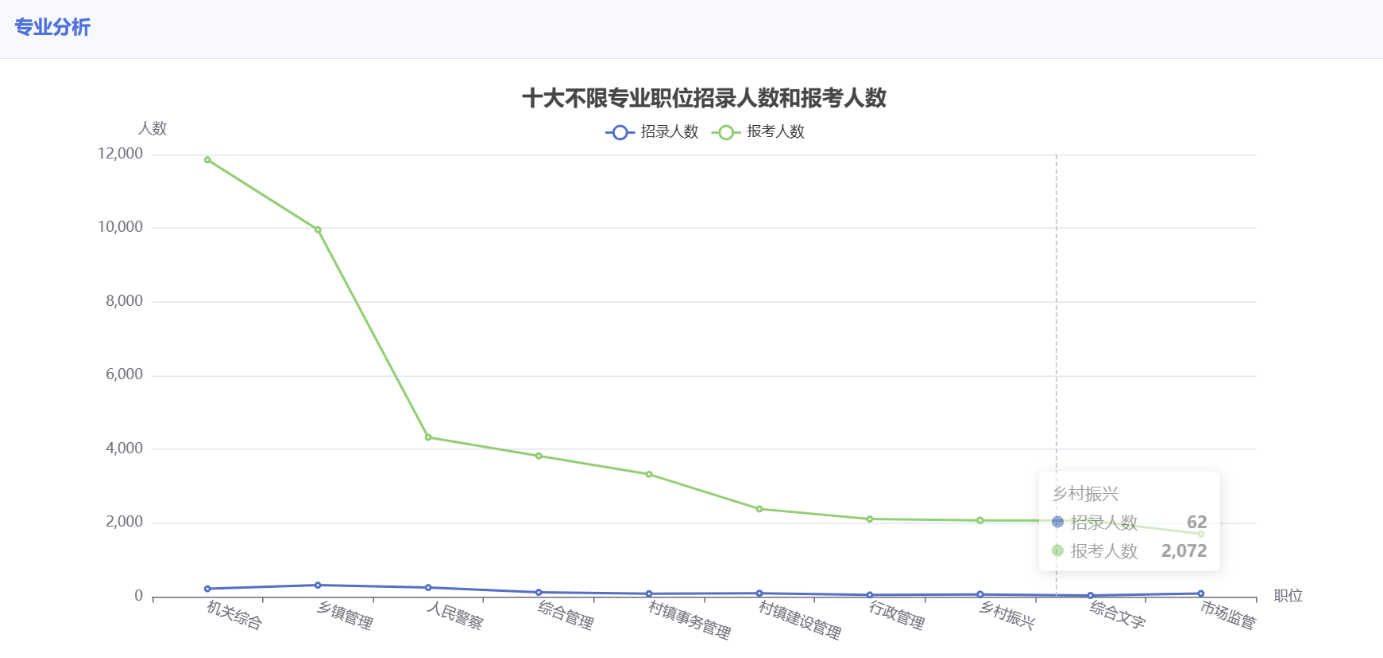


图 14 十大不限专业职位招录人数和报考人数



图 15 十大不限专业部门招录人数和报考人数

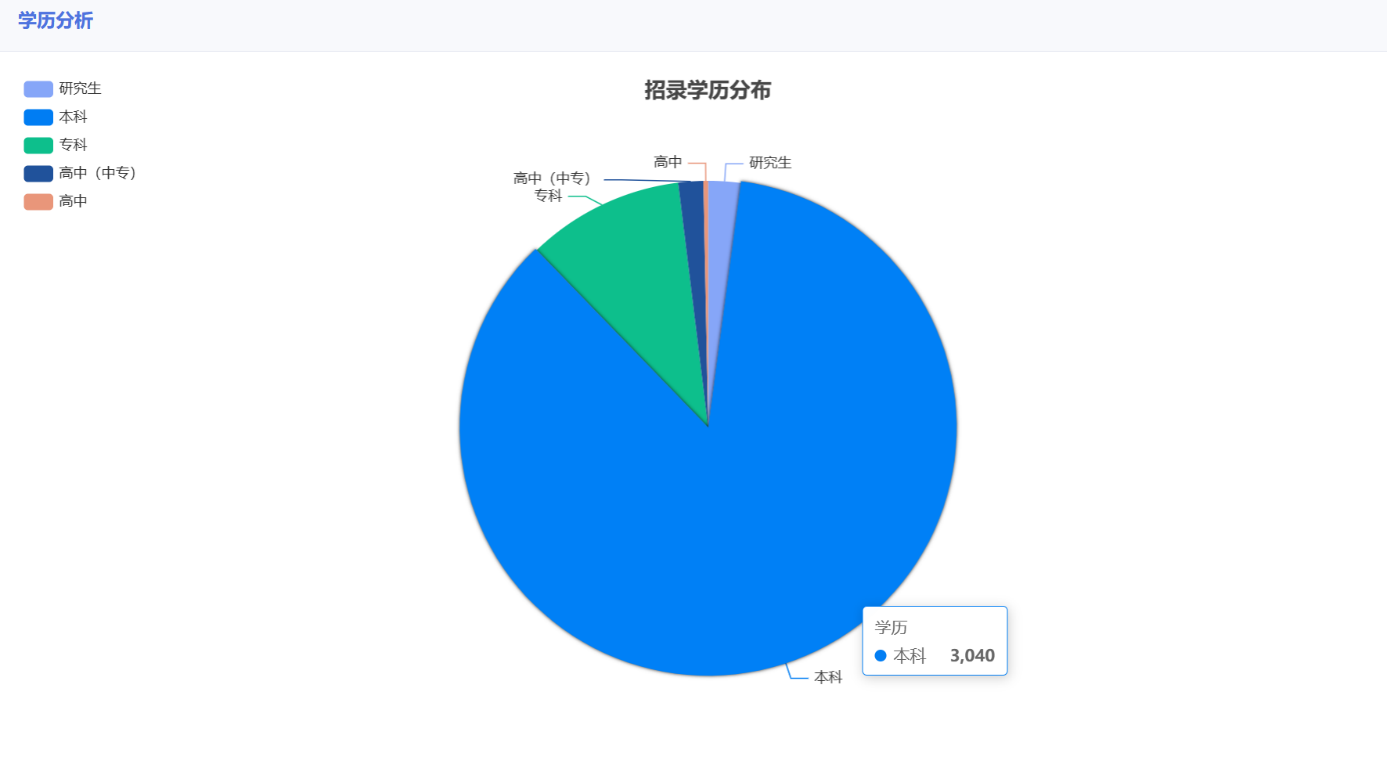


图 16 招录学历分布

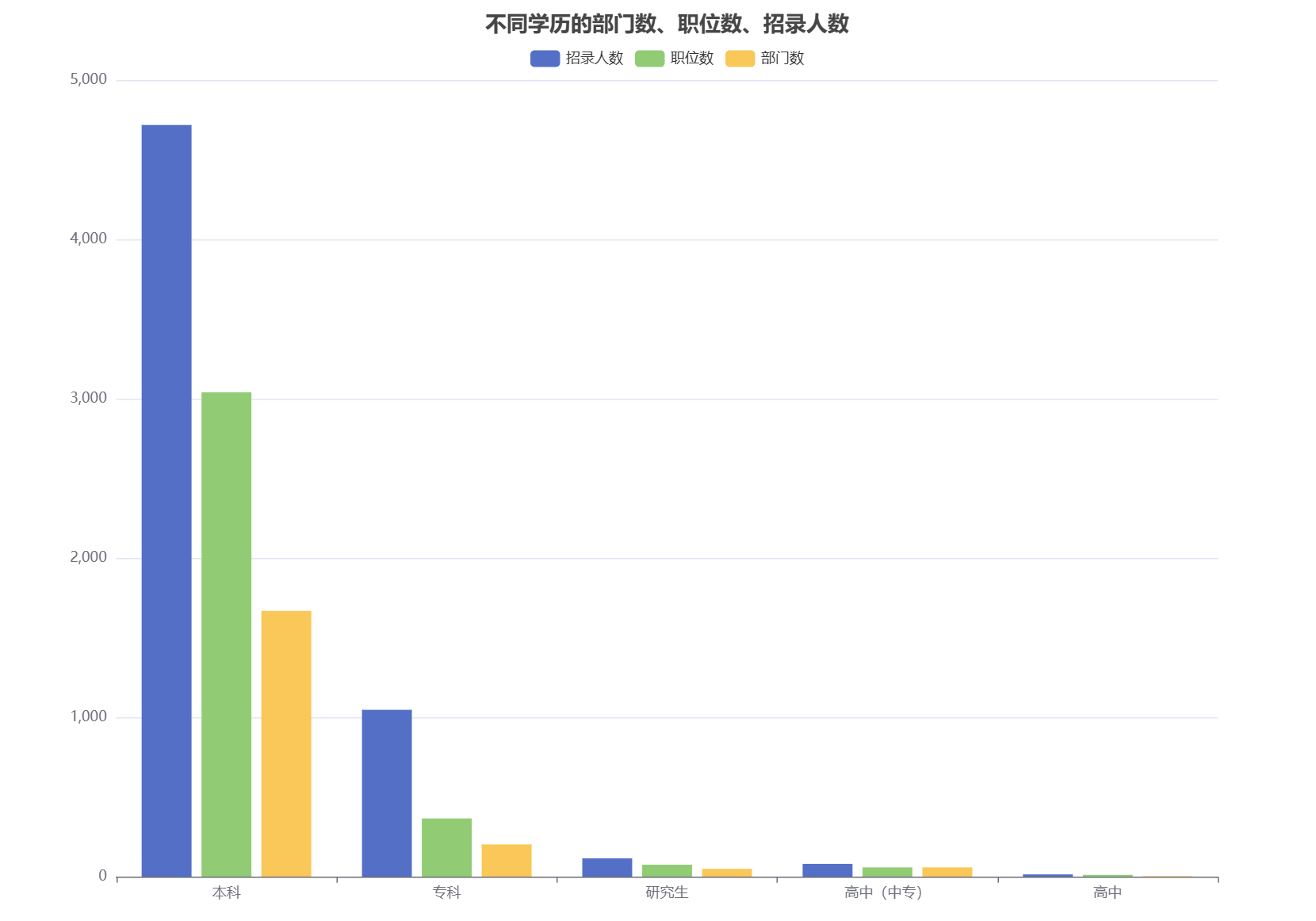


图 17 不同学历信息

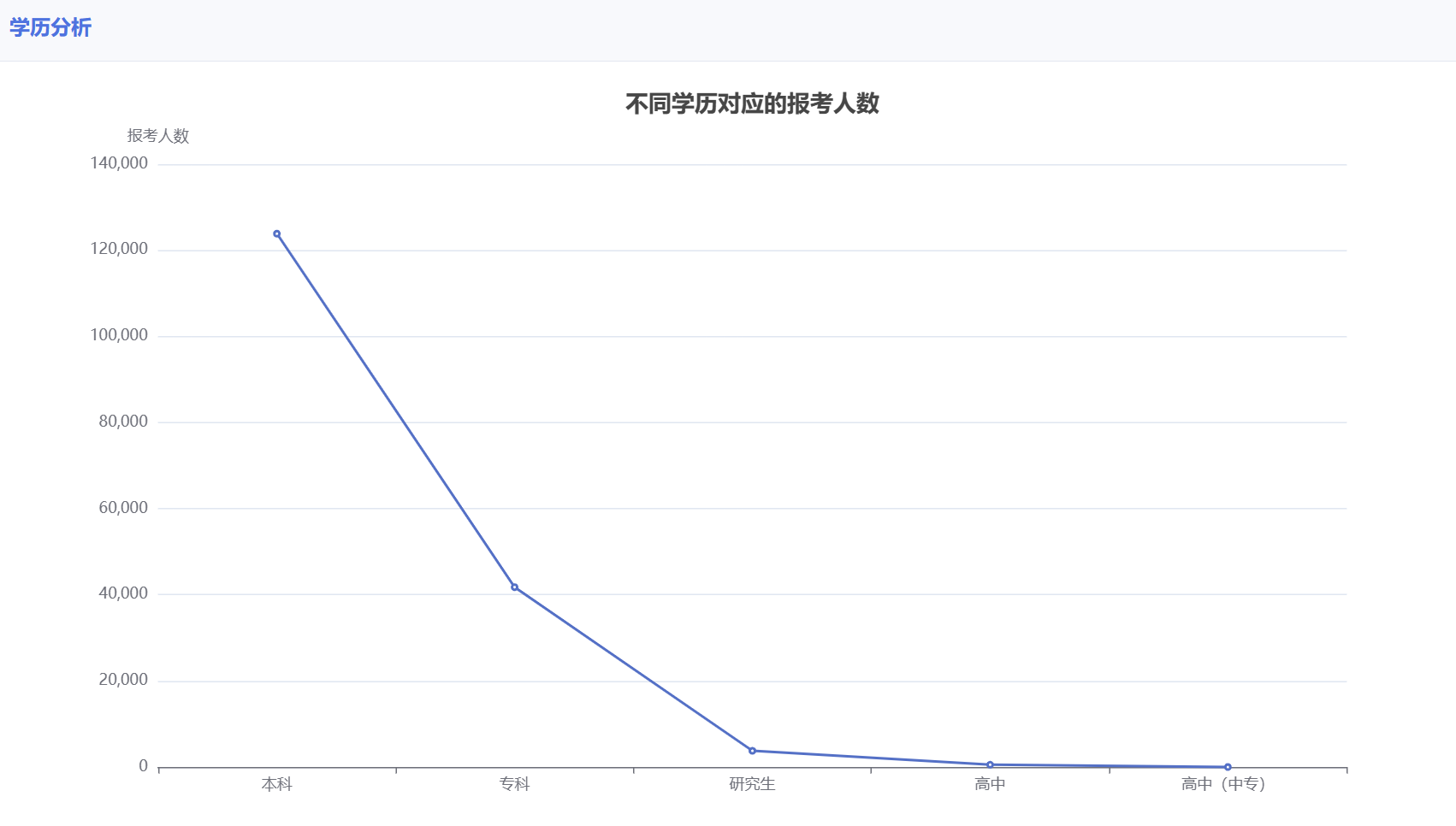


图 18 不同学历的报考人数

# 第6章 总结和展望

## 6.1 总结

近年来，公务员考试竞争非常激烈，公务员招录工作越来越重要对于政府机构的人才引进和管理至关重要，政府和社会对于公务员招录数据的需求也越来越高。因此，对公务员招录数据进行分析和研究，可以为政府机构提供更加精确的招录策略和决策，同时也可以帮助考生更好地了解招录情况，提高自身的竞争力。

本文主要介绍了一种基于Python的公务员招录数据分析系统该系统主要包括三个模块：数据获取模块、数据处理模块和数据可视化模块。数据获取模块使用Python爬虫技术对公考信息网站进行数据爬取，并将数据保存在本地CSV文件中。数据处理模块通过SQL语句对数据进行筛选、排序和统计等操作，得到相关的分析结果。数据可视化模块使用Python中的数据可视化库matplotlib对数据进行图表展示，更加直观地呈现出分析结果。

数据获取模块主要实现了对公考信息网站上的数据爬取，包括2023年安徽各地公务员报考信息和招录信息，包括职位代码、地区、部门名称、职位、学历、专业、招考人数、报名人数和合格人数等。通过对目标网站的URL进行拼接和GET请求，获取网站上的职位信息数据，利用BeautifulSoup对HTML页面进行解析，并遍历表格中的所有行，将其中的信息写入到本地的CSV文件中。随后，使用MySQL数据库进行数据存储，并通过Flask框架实现数据可视化。

数据分析模块主要实现了对公务员招录数据的分析，包括选取安徽省2023年度考试录用公务员招录人数最多的十大职位、选取安徽省2023年度考试录用公务员报考人数最多的十大职位和选取安徽各地区无人报考职位数量等。通过MySQL数据库进行管理和分析，使用SQL语句进行筛选和计算，得出相应的分析结果。

本文所介绍的基于Python的公务员招录数据分析系统具有一定的实用性和可行性。系统可以帮助政府机构更好地了解招录情况，提高招录效率和精度，同时也可以为考生提供更加详细和全面的招录信息，提高其竞争力和招录成功率。同时，该系统还可以进行更多的数据分析和挖掘，为公务员招录工作提供更加深入和精确的研究和分析。

在数据分析方面，系统运用了SQL语言对数据进行筛选、过滤、排序等操作，实现了选取招录人数最多的十大职位、报考人数最多的十大职位以及各地区无人报考职位数量等分析，为考生提供了重要参考信息。该系统的设计和实现充分体现了Python作为一种快速开发、易上手的编程语言的优势，同时也展示了Flask框架在数据可视化方面的便利性和灵活性。此外，通过对公务员报考招录信息的分析，也为考生提供了更加全面、准确的信息，有助于考生更好地了解和把握招录形势。

通过这个数据可以看出，虽然安徽省公务员招录比较稳定，但是仍然存在一些职位无人报考的情况，这需要相关部门对职位需求进行合理规划和调整。

未来，该系统可以进一步扩展其数据源，以覆盖更多的招录信息网站和更多的省份和地区。此外，可以考虑增加一些更加深入的数据分析方法，例如机器学习算法等，以提高数据分析的准确性和实用性。

## 6.2 展望

基于Python的公务员招录数据分析系统是一种基于数据分析和数据可视化技术的公务员招录信息处理工具。未来，这种系统将具有更加广泛的应用前景，将在数据分析和数据可视化技术的不断发展和创新下得到不断改进和升级，使得公务员招录信息处理更加方便、快捷、高效。

在技术方面，未来的系统将继续关注数据分析和可视化技术的发展趋势。针对现有系统的不足之处，未来将在数据分析方面加强数据挖掘技术的应用，以发掘更多的隐藏信息。例如，通过应用机器学习算法，可以在公务员招录信息中发现更多的关联规则，以帮助招聘机构更好地制定招聘计划和策略。另外，针对当前数据可视化的不足之处，未来的系统将进一步提升用户体验，优化数据呈现的方式和展示效果，例如增加动态图表、交互式可视化等，以更好地满足用户的需求。

在应用方面，公务员招录数据分析系统将在更多的领域得到应用。例如，未来的系统将面向更多的用户群体，除了招聘机构和考生之外，还可以应用于政府部门、媒体、学术机构等。在政府部门方面，公务员招录数据分析系统可以帮助政府制定更科学的招聘政策和计划，提高公务员招录的透明度和公正性。在媒体方面，公务员招录数据分析系统可以为媒体提供更加全面、准确的招录信息和分析报告，帮助媒体进行深度报道和分析。在学术机构方面，公务员招录数据分析系统可以为研究者提供更加详实的招录数据和分析报告，促进学术研究的深入和发展。

在商业方面，公务员招录数据分析系统将成为数据服务行业的一个重要组成部分。未来的系统将借助云计算、大数据、人工智能等新技术的发展，实现数据服务的智能化、个性化、高效化。例如，未来的系统可以通过数据挖掘技术发现新的商业机会，提供更加精准的数据服务，为客户提供更加优质的数据产品和服务。

基于Python的公务员招录数据分析系统是一个非常有前途和潜力的数据分析工具。通过不断地改进和发展，本系统将成为一个更加强大和智能的数据分析平台，为用户提供更优质和个性化的数据服务。

# 参考文献:

[1] 申兆慕.多源公众出行大数据综合可视化分析[D].北京建筑大学,2019.

[2] 姜枫，许桂秋.大数据可视化技术[M].人民邮电出版社,2019.

[3] 张小雨.大数据时代下软件工程技术的应用[J].张小雨.电脑知识与技术,2020(33)

[4] 林子雨.大数据技术原理与应用[M].人民邮电出版社,2017.

[5] 孙品一.信息可视化的应用研究[D].孙品一.湖北工业大学,2016

[7] 陈为.数据可视化的基本原理与方法[M].科学出版社,2013.

[9] 李希娟.大数据时代下的数据可视化研究[D].河北大学,2014.

[10] 崔蓬.ECharts在数据可视化中的应用[J].崔蓬.软件工程. 2019(06)

[11] 许文鹏.数据可视化系统架构的设计与实现[D].北京交通大学,2015.

[12] 展鹏.多维数据展现开发工具的设计与实现[D].山东大学,2013.

[13] 胡永凯.云平台下的大数据可视化实现.[D].电子科技大学，2022

[14] 邓智豪.基于大数据的自定义可视化分析引擎的设计与实现.[D].北京邮电大学，2021

[15] Data Visualization Model Based on the User Experience [J] . Wang Pengwen,Hu Yanan,Huo Jiaofei. Journal of Physics: Conference Series . 2021 (3)

# 致谢

时光荏苒，岁月如梭。转眼之间，我的四年大学生活即将步入尾声。回首这四年的校园时光，我收获了宝贵的知识，结识了众多良师益友。

在毕业设计和论文编写的过程中,我得到了陈保国老师的大力帮助和关心。老师渊博的知识和严谨的治学作风使我受益匪浅，对顺利完成本系统起到了极大的作用。借此我向导师表示我最诚挚的感谢和祝福。

此外还要感谢辛勤培育我的老师们，及朝夕相处的同学们，他们的关心使我充满动力。感谢同学们在专业知识上给予我的帮助，因为他们的协助，才得以顺利完成了设计。

感谢几年来传授我知识的老师们，感谢我的家人对我学业上的支持和鼓励，感谢所有关心帮助过我的人。

感谢学校为我提供了一个良好的学习平台，使我有机会在知识的海洋之中遨游。