

# 安全生产事故统计分析系统

## 摘要

本作品是针对区域性、行业性及生产企业的安全生产事故数据进行深度分析和可视化系统开发。作品的展呈依托的数据是国内 10 余年来的安全生产事故数据，以《安全与环境学报》刊登的系列关于国内安全生产事故统计分析的论文为对象，收集其中 2005-2019 年（即“十一五”至“十三五”期间）的事故数据，构建完整的国内安全生产事故案例数据库。基于此数据库，利用 Python 编程语言进行数据提取和处理，并对事故发生特征开展了深度挖掘、分析及结果可视化，实现对事故发生时间序列分析、事故类型占比分析、事故增长率分析、各省份事故数量及占比分析、各省份事故死亡人数与 GDP 平均增速关联分析、事故发生工作日与节假日的关联度的分析、及各省份事故数量的历史变化动态分析。此外，基于完备的事故统计基础，本系统可用来分析：（1）某区域内（如省、市、区等）事故特征；（2）某行业内事故特征；（3）某生产企业事故特征。尤其是针对于特定的企业，由于具体企业的统计事故的指标（统计事故内容）是既定的，那么该系统就可以直接与企业事故案例数据库进行对接，嵌入到企业安全生产信息平台，实现企业事故数据的深度分析及可视化。

**关键词：** Python，安全生产事故，统计分析，软件

## 目录

安全生产事故统计分析系统 .....	1
第一章 设计背景 .....	4
第二章 特色与创新 .....	6
2.1 安全生产事故统计现状.....	6
2.2 本项目的特色与创新.....	6
2.2.1 项目特色 .....	6
2.2.2 项目创新 .....	6
第三章 功能设计 .....	8
3.1 实际需求.....	8
3.2 系统功能.....	8
3.2.1 系统登陆.....	8
3.2.2 数据加载、查看、筛选功能.....	9
3.2.3 事故分析功能.....	10
3.2.4 数据绘图功能.....	16
3.2.5 总体框图.....	17
第四章 系统实现方案 .....	19
4.1 系统方案架构.....	19
4.2 系统技术方案.....	19
4.2.1 编程语言选择 .....	19
4.2.2 数据获取层面 .....	20
4.2.3 数据处理层面 .....	21
4.2.4 数据显示层面 .....	24
第五章 项目效果与演示实验 .....	26

<b>5.1 企业事故统计分析系统初始界面介绍 .....</b>	<b>26</b>
5.1.1 登录界面及注册界面 .....	26
5.1.2 数据分析系统界面 .....	27
<b>5.2 系统运行效果演示实验.....</b>	<b>28</b>
5.2.1 系统功能演示 .....	28
5.2.2 详细设置介绍及演示 .....	32
<b>第六章 项目小结 .....</b>	<b>45</b>
<b>6.1 系统完成功能.....</b>	<b>45</b>
<b>6.2 系统待优化功能.....</b>	<b>45</b>
<b>6.3 展望 .....</b>	<b>46</b>
<b>参考文献.....</b>	<b>47</b>

## 第一章 设计背景

随着现代科学技术和我国国民经济的蓬勃发展，企业生产规模不断扩大。但是，安全管理却跟不上生产规模的扩大进度，再加上安全生产基础管理建设发展不平衡，致使生产安全事故不断。<sup>[1]</sup>当前，我国的安全生产事故虽呈现出总量逐年下降趋势，但仍处于易发多发阶段，尤其特别重大安全生产事故，一旦发生则会严重影响影响了各地方政府的社会形象及经济发展，对于事故家庭及个人来说，也将是不可磨灭的灾难性悲剧。比如最近的 2019 年 3 月 21 日，位于江苏省盐城市响水县生态化工园区的天嘉宜化工有限公司、发生特别重大爆炸事故，造成 78 人死亡、76 人重伤，640 人住院治疗，直接经济损失高达 198635.07 万元；同年 9 月 28 日，江苏省宜兴市 G25 长深高速路段，大客车冲破道路中央隔离带与半挂货车相撞，造成大客车及货车上共 36 人死亡、36 人受伤。

安全生产事故的发生不仅造成了重大的人员伤亡和财产损失，而且强烈冲击了人民群众的安全感，严重影响了政府执政的公信力，同时对国家治理能力现代化建设提出了严峻挑战。因此，寻找有效预防和减少事故发生的方法迫在眉睫。而安全生产事故统计分析基于真实事故数据，可以使领导机构及时、准确、全面地掌握安全生产状况，发现问题并作出正确的决策，并针对安全生产工作中的薄弱环节，有的放矢地采取避免事故发生的对策，最终为制定有关安全规章制度提供科学的依据，从而有效预防和减少事故发生。

但是，庞大的事故数量、繁杂的事故类型、多种的事故影响因素都给企业的事故统计分析工作带来不便。

2015 年底，以规范性文件形式印发的《生产安全事故统计管理办法(暂行)》(安监总厅统计[2015]111 号)，揭开了生产安全事故统计体系改革的大幕。<sup>[2]</sup>由于生产安全事故统计分析改革时间较短，当今无论是全国性的、区域性的还是针对于行业企业性的事故，都缺乏系统性、全面性的统计，也没有深度的数据分析和直观形象的分析结果展呈。于是，本安全生产事故统计分析系统应运而生。

为提高事故信息统计的及时性、准确性和全面性，增强统计分析的针对性、预测性和预警性，本系统顺应大数据、“互联网+”等先进技术发展大势，选用 Python 编程语言来作为系统载体。Python 拥有自己独特的优点，在科学计算和数据

分析领域应用十分广泛，受到很多热爱编程的人的喜爱。<sup>[3]</sup>它不仅支持各种数学运算，还可以绘制高质量的 2D 和 3D 图像，与科学领域最流行的软件 MATLAB 相比，Python 处理的文件和数据类型更多，可操作性更强，应用范围更广。<sup>[4]</sup>

本系统基于国内安全生产事故案例数据库，对事故的指标数据进行分析及可视化，旨在了解我国各行业事故发展趋势及总体安全生产形势，分析的结果可供安全生产管理者参考，以便提出科学可行的安全对策及措施建议，有助于进一步搞好安全生产管理。进一步的，由于具体企业的统计事故的指标（统计事故内容）是既定的，那么该系统就可以直接与企业的事故案例数据库进行对接，实现企业事故数据的分析及可视化。此外，本系统可扩展应用于安全专业的研究领域或师生的学习实践中，为安全学科建设提供技术支持。

## 第二章 特色与创新

### 2.1 安全生产事故统计现状

(1) 安全生产事故屡见不鲜，但是国内关于安全生产事故统计分析的较为完整的系统却十分少见，不利于企业总结事故的经验教训。

(2) 现有的安全生产事故统计不够完善，多为单行业的事故统计，不利于预测总体安全生产形势。

(3) 安全生产事故的分析范围较为狭窄，分析因素不够全面，数据研究不够深入，企业无法全面且透彻地掌握安全生产情况、学习总结事故分析的结果，导致其总结分析事故的角度过于单一，进而影响事故防范措施的效果。

### 2.2 本项目的特色与创新

#### 2.2.1 项目特色

(1) 实现国内的安全生产各行业事故汇总统计。收集《安全与环境学报》所公布的详细的事故统计数据，利用 Python 编程创建各个时期的数据文件夹，存放数据并进行汇总，让事故统计系统的覆盖的数据范围更广，事故统计系统的数据更详尽，使其更加全面化，精准化。

(2) 实现各类安全生产事故数据的全方位分析。在基本安全生产事故分析的基础上，再将收集的数据与地区、工作日及非工作日、省份 GDP 等因素建立联系，让企业从不同角度了解事故数据，掌握生产情况。

(3) 系统实用性强，可信度高，结果显示直观、生动。通过 Python 实现数据导入、数据分析、结果显示等工作，并绘制相应的折线图、饼状图、条形图、动态图，利于企业及时、准确、全面了解安全生产状况，制定具有针对性的正确决策，高效完成安全管理工作。

#### 2.2.2 项目创新

(1) 利用 Python 编程设计完成了一个健全完善的安全生产事故统计分析系统。

国内完整的安全生产事故统计分析系统较为少见，因此，本系统顺应信息时代潮流，满足安全专业实践的需求，建立了完整的事故案例库，填补了安全生产事故统计分析系统领域的空缺。

（2）相比于单行业的、基本的事故统计分析，各行业事故汇总分析覆盖的范围更广，分析数据的角度更多样，能够给相应的安全管理工作提供更多更好的方法与途径。

（3）系统登陆及使用需输入正确的密码，防止无关人员查看企业的机密文件，保证了企业的安全运营。在系统退出时，会弹出提示框，避免文档数据的丢失或工作人员的误操作。

## 第三章 功能设计

### 3.1 实际需求

- (1) 系统具有保密性，无关人员无权使用系统。
- (2) 能够存放企业本地数据，加载国内事故案例库，精准筛查所需的事故数据。
- (3) 从不同角度对事故进行分析对比，直观地显示分析结果。

### 3.2 系统功能

#### 3.2.1 系统登陆

通过Python编写的一个具有保障数据安全作用的身份验证页面，即用户登陆界面。初始账号与密码均为admin，新用户可以在身份注册后登陆。注册成功后，需要输入正确账号及密码方可对系统进行下一步操作，若输入错误，则无登陆，界面显示输入密码错误（如图1, 图2，图3）。



图1 身份注册





图2 身份验证



图3 账号信息错误

### 3.2.2数据加载、查看、筛选功能

用户可以选择导入本地文件，也可以选择加载事故案例库中的数据，数据加载完成后即在数据查看框中显示，为避免事故重复，在每一个事故前都标明序号。

在筛选框中输入筛选条件，可精准筛选出每一行每一列的相符数据。如：在索引行标题为省份、事故类型、发生时间的数据中，输入筛选条件“上海”，系统则会对每行条目进行搜索，筛选出省份为“上海”的事故数据；输入筛选条件为“事故类型”，系统则会在“事故类型”一列中进行筛查并导出事故类型数据；随后数据筛选框会罗列出所需数据条目，选择保存数据，即可将筛选出的数据以EXCEL形式存放至本地（如图4）。

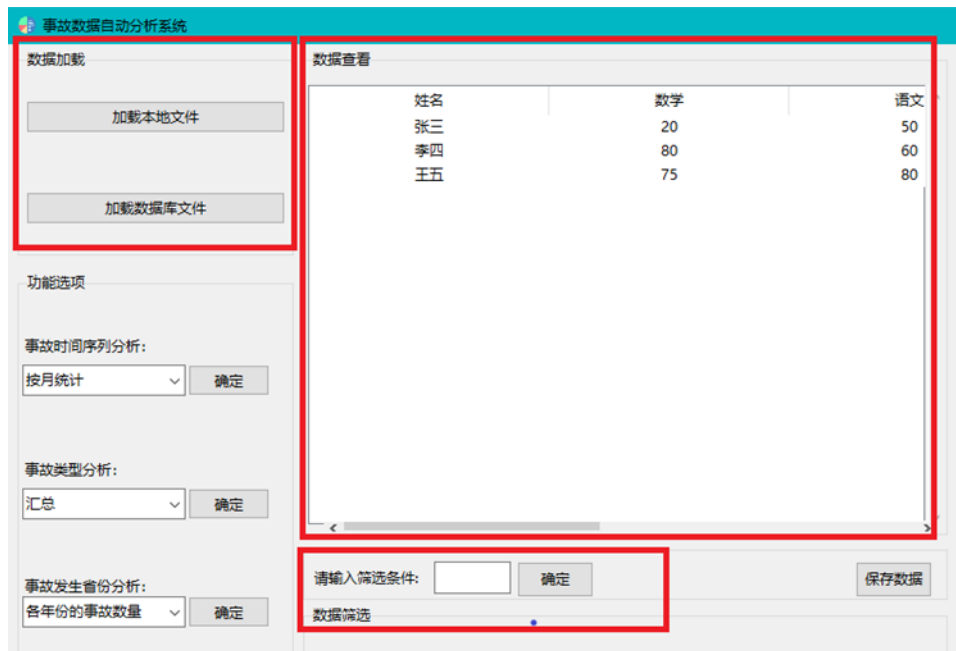


图4 数据加载、查看、筛选功能示意图

### 3.2.3 事故分析功能

事故分析功能是基于Python编程，可以对所选数据进行不同角度的数据分析，满足用户的多种需求的一项重要功能。

下面详细介绍事故的不同类型的分析功能。

#### (1) 事故时间序列分析功能

事故时间序列分析功能是基于Python编程，对导入的数据进行事故数量与时间的分析，其中时间分为年、季度、月。

以年为时间单位进行事故数量分析，有上半年年事故总量、下半年年事故总量、每年事故总量、年平均事故数量。（如图5）

以月为时间单位进行事故数量分析，即每年每个月的事数量。（如图6）

以季度为时间单位进行事故数量分析，有第一季度、第二季度、第三季度、第四季度的事故数量。

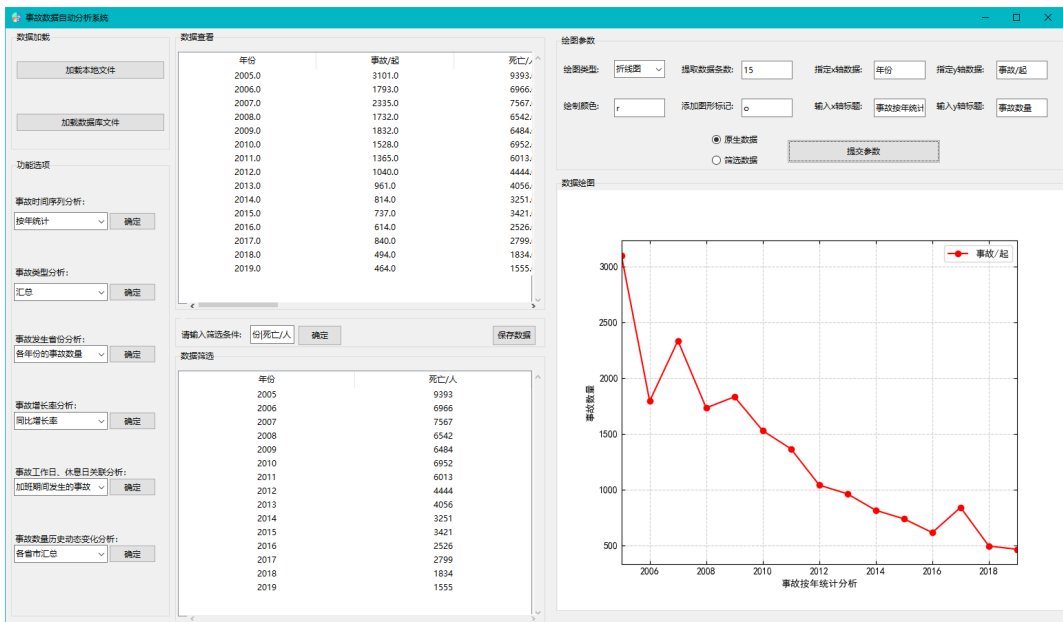


图5

按年统计的事故时间序列分析

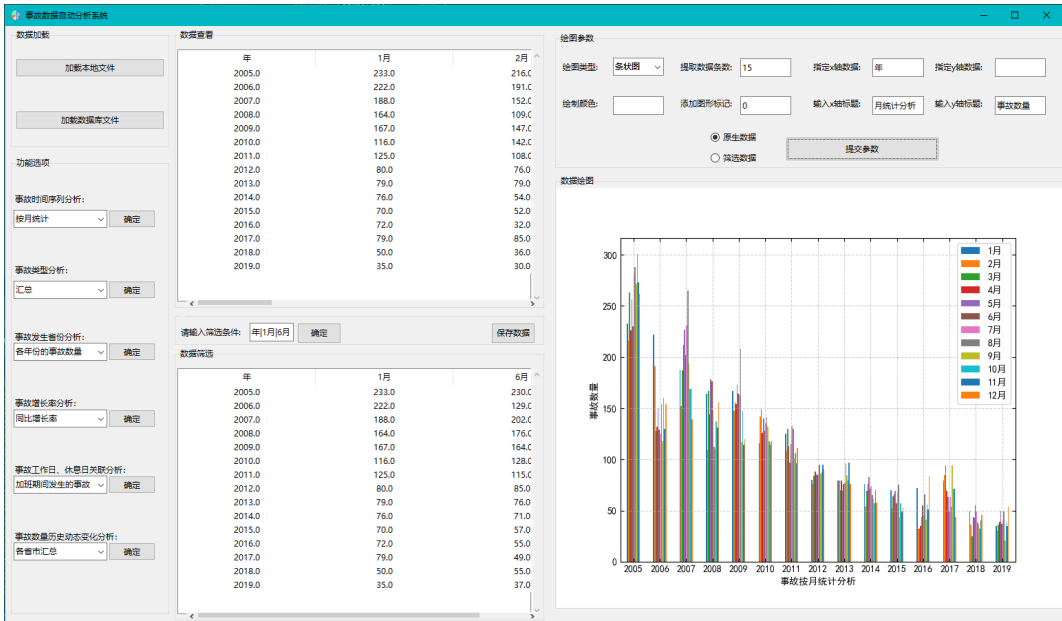


图6 按月统计的事故时间序列分析

(2) 事故类型占比分析功能

事故类型占比分析有以每月、每季度、每年为时间单位的事故类型占比分析。

(如图7，图8)

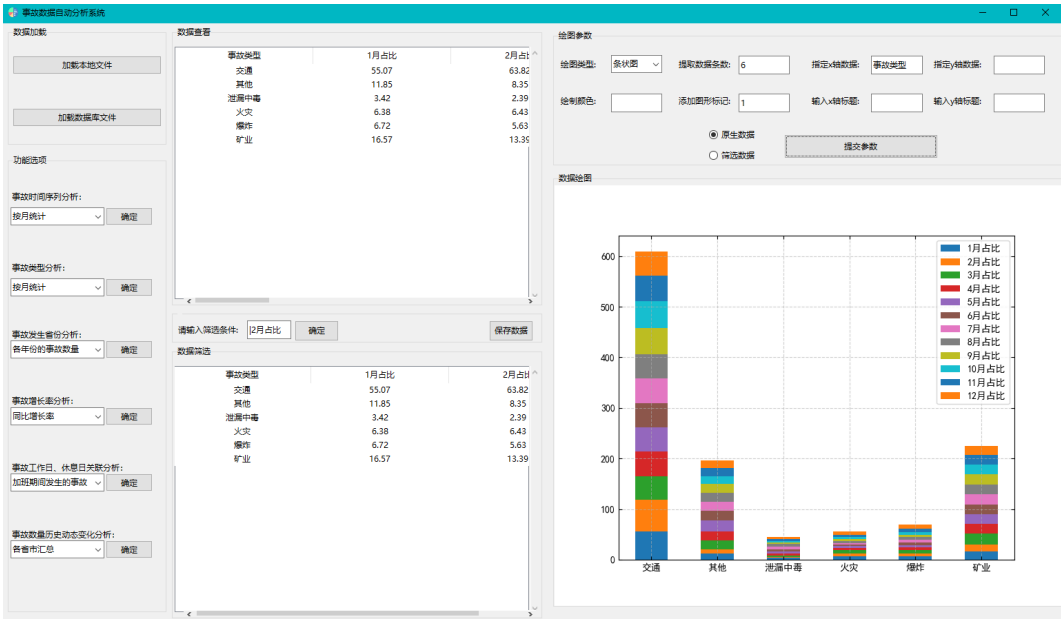


图7 按月统计的事故类型占比分析

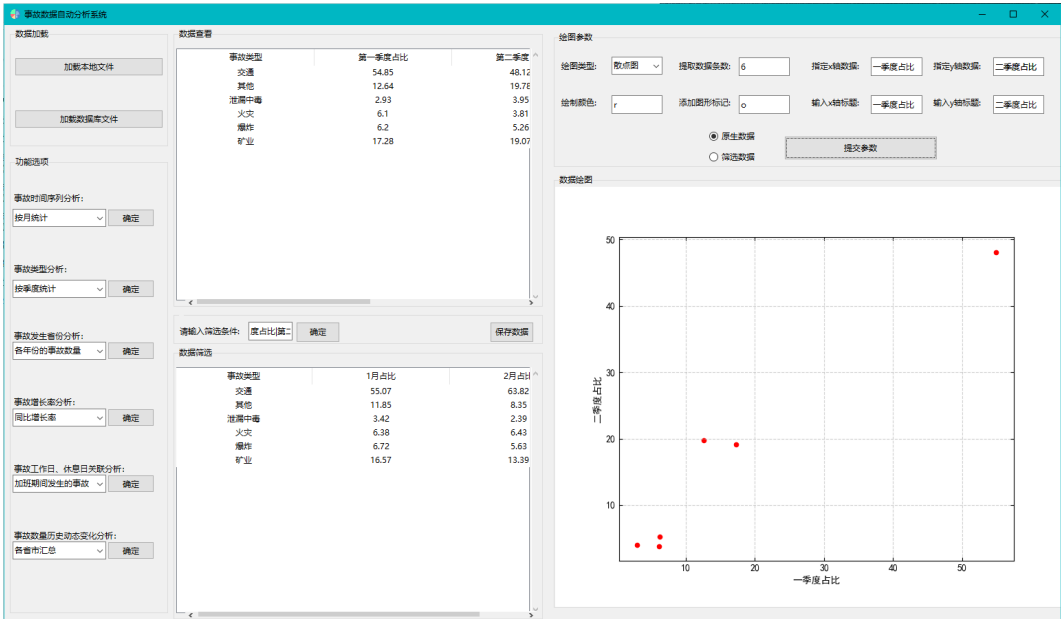


图8 按季度统计的事故类型占比分析（散点图说明每个季度的事故类型占比的变化不大）

(3) 事故发生省份分析功能

事故发生省份分析功能，即以发生过事故的省份为对象，统计其每年事故总量以及历史总数量并绘制表格，其中每格的背景颜色表示数量的多少，颜色越深，事故数量越多。此外，以中国地图为底图加气泡以显示省份与GDP关系，气泡颜色越深，则表示其省份发生的事故越多，气泡面积越大，则表示其省份GDP增长速度越快。（如图9，图10）

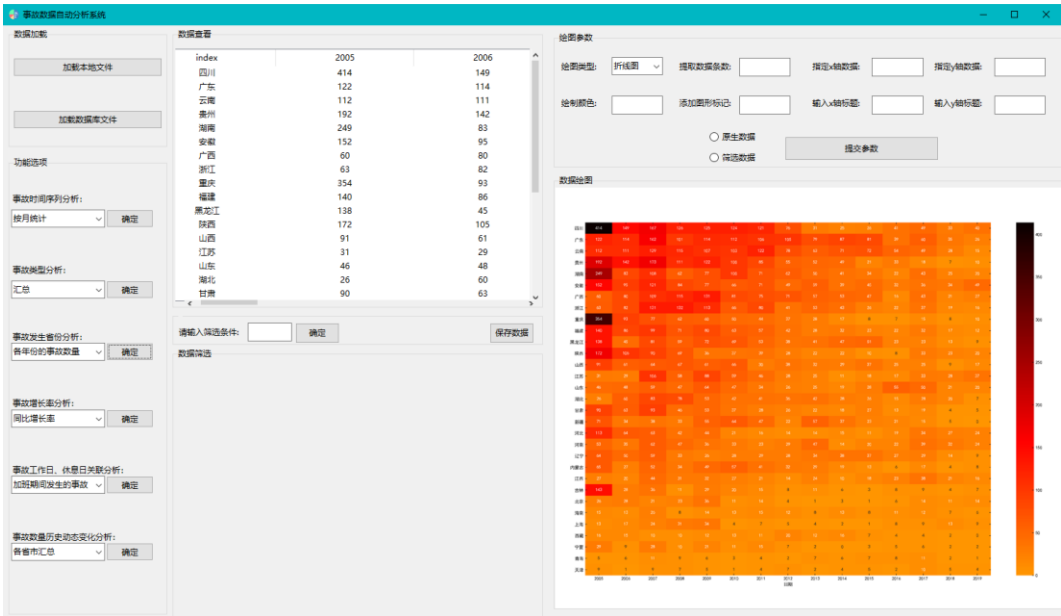


图9 省份与事故数量关系图

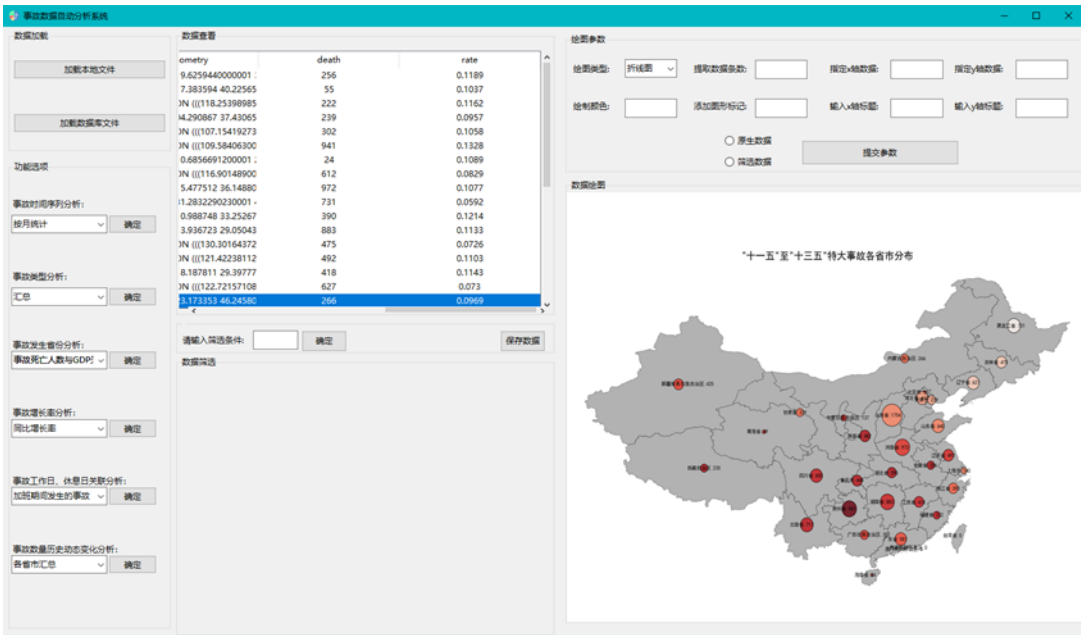


图10 事故数量与省份GDP关系图

(4) 事故增长率分析功能

事故增长率分析是计算事故的环比和同比增长率。计算环比增长率时，与《安全与环境学报》保持一致，采用两个月的数据进行环比。

环比增长率=（本期数-上期数）/上期数×100%，反映本期比上期增长了多少。比如1-2月，3-4月。

同比增长率=（本期数-同期数）/|同期数|×100%，反应同一统计时段不同间隔下的情况，比如2010年1-2月，2011年1-2月（如图7，图8）。

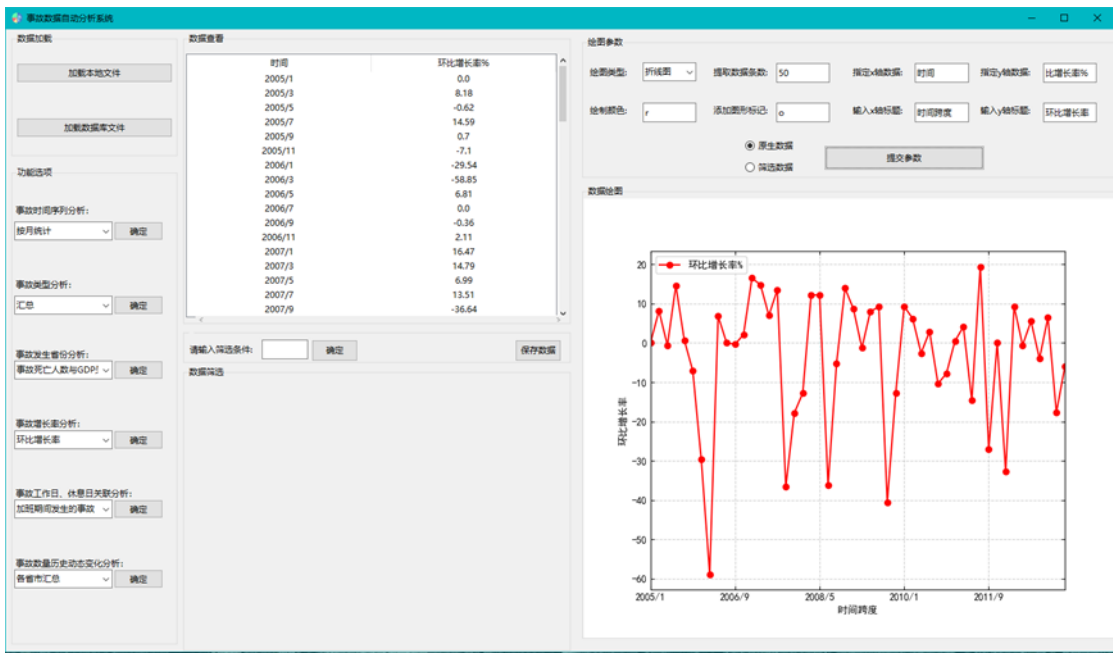


图11 环比增长率

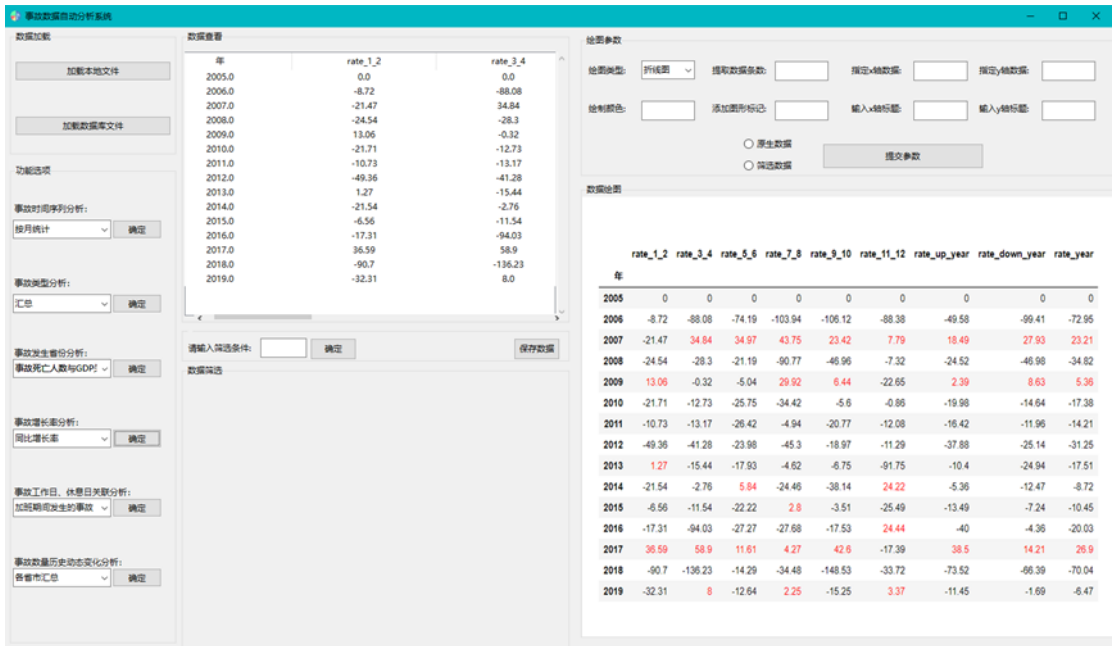


图12 同比增长率

(5) 事故与工作日、休息日关联度分析功能

事故与工作日、休息日关联度分析是利用Python编程，加载工作日、节假日、加班数据，与“十一五”至“十三五”的特大事故发生时间进行对应的筛查，分类出工作日事故与休息日事故并进行分析计算事故率和在各个时段的平均事故率。（如图13，图14）

结果

①各时段事故数量及占比

- 节假日发生事故起数： 64 事故占比为： 7.50%
- 正常周末休息发生的事故起数为： 224 事故占比为： 26.26%
- 因节假日调休周末上班发生的事故起数为： 15 事故占比为： 1.76%
- 正常上班发生的事故起数为： 853 - 64 - 224 - 15 = 550 事故占比为： 64.48%

②工作日与非工作日事故占比

- 工作日特重大事故占比：  $(550 + 15) / 853 = 66.24\%$
- 非工作日特重大事故占比：  $(224 + 64) / 853 = 33.76\%$

③事故率计算

按照一年是365天，我国实际节假日是虽然从05年到18年有所变动（实际的有效假日变化不大），有效工作提250天，剩下的就是休息日115天 2007年往后将五一长假缩短，添加了端午和中秋（实际有效假日几乎没有发生变化），然后15年增加一个胜利日（这一年是增加一天假）

- 平均事故率：  $853 / (365 \times 15) = 15.56\%$
- 工作日平均事故率  $(550 + 15) / (250 \times 15) = 15.07\%$
- 非工作日平均事故率  $(224 + 64) / (115 \times 15) = 16.70\%$

图13 “十一五”至“十三五”的事故与工作日、休息日关联度分析结果

(6) 事故数量历史动态图变化分析功能

事故数量历史动态图变化分析是以各省份为单位，记录事故数量并以动态条状图的形式展出，其中统计方式有双月数量累计统计，年数量累计统计（如图14，图15）。

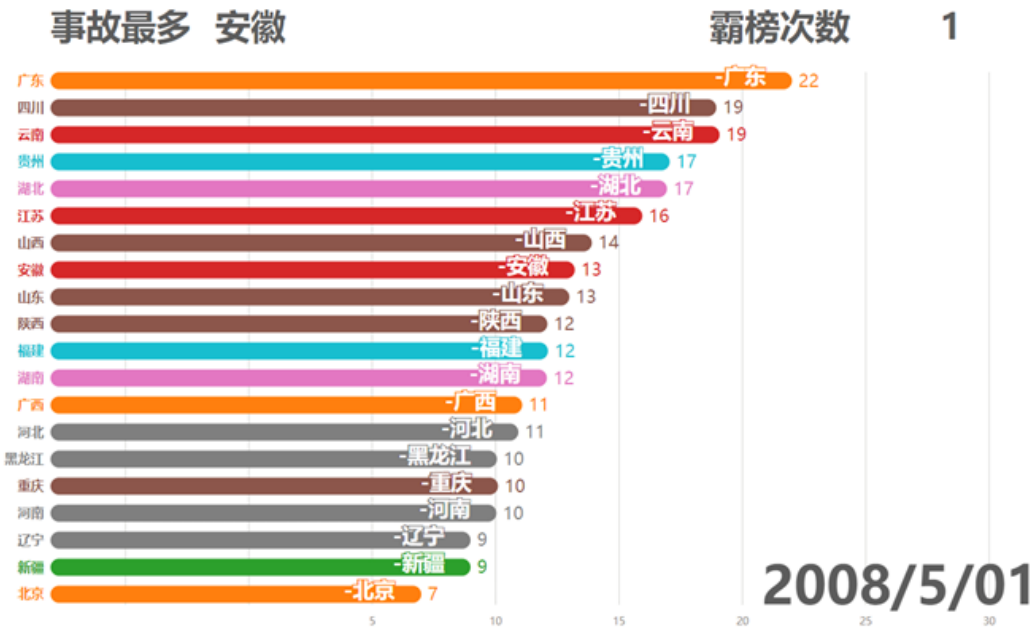


图14 按双月统计的事故数量历史动态图变化分析

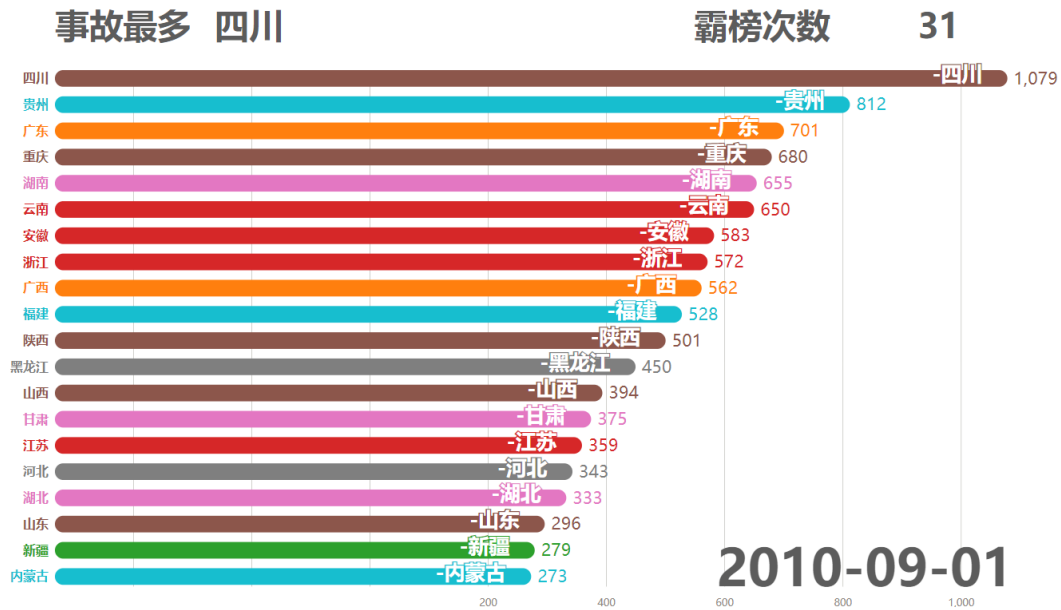


图15 按年统计的事故数量历史动态图变化分析

### 3.2.4数据绘图功能

用户可根据自己的需求，可对原生数据或筛选数据实行绘图功能，并自动显示于绘图框。

依照数据类型与数量，设置绘图参数：选择绘图类型（折线图、条状图、横向条状图、饼图），设置提取数据的条数，指定x，y轴的数据及标题，设置绘制图形的颜色（填入颜色的英文首字母）。对于多种类型数据绘图，需要区别颜色可使用

“|” 隔开颜色首字母，需要区别折线图中线条可以使用“添加图形标记”添加“○或×” 记号。在条状图中，“添加图形标记”可利用0、1来设置图形分开或堆叠。设置完成后，点击提交参数即可生成对应的图形（如图16，图17）。





绘图参数

绘图类型: 折线图 提取数据条数: 指定x轴数据: 指定y轴数据:

绘制颜色: 添加图形标记: 输入x轴标题: 输入y轴标题:

☐ 原生数据 ☐ 筛选数据

提交参数

图16 绘图参数设置

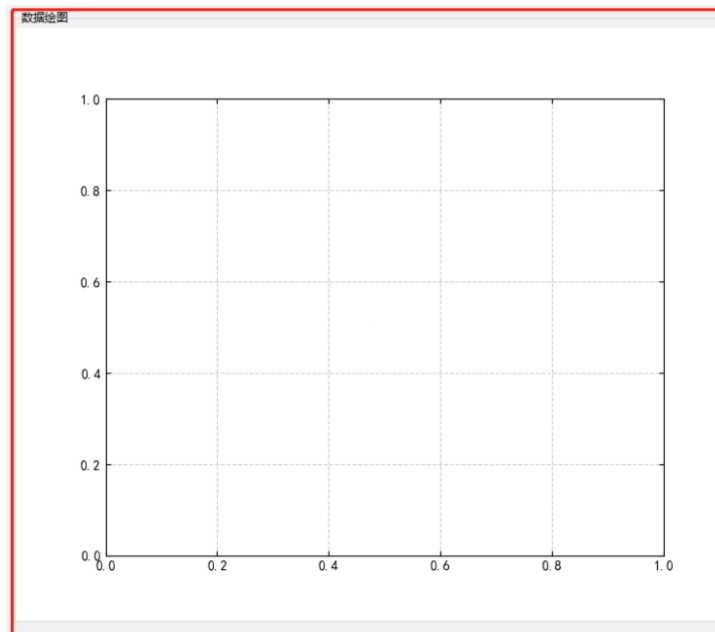


图17 绘图框

### 3.2.5 总体框图

系统基于Python编程，开发了身份验证、数据导入、数据筛查、事故分析、数据绘图等一系列功能。对导入的本地数据或加载的案例库数据进行符合要求的筛查，选定数据后，施行事故分析功能，包括事故时间序列分析（按年、月统计）、事故类型分析（年、月统计）、事故发生省份分析（各年、月事故数量）、事故增长率分析（同比增长率、环比增长率）、事故工作日和休息日关联分析、事故数量历史动态变化分析（各省市双月、各省市汇总），随后生成相对应的条形图、折线图、饼图等图表，实现数据的可视化（如图18）。

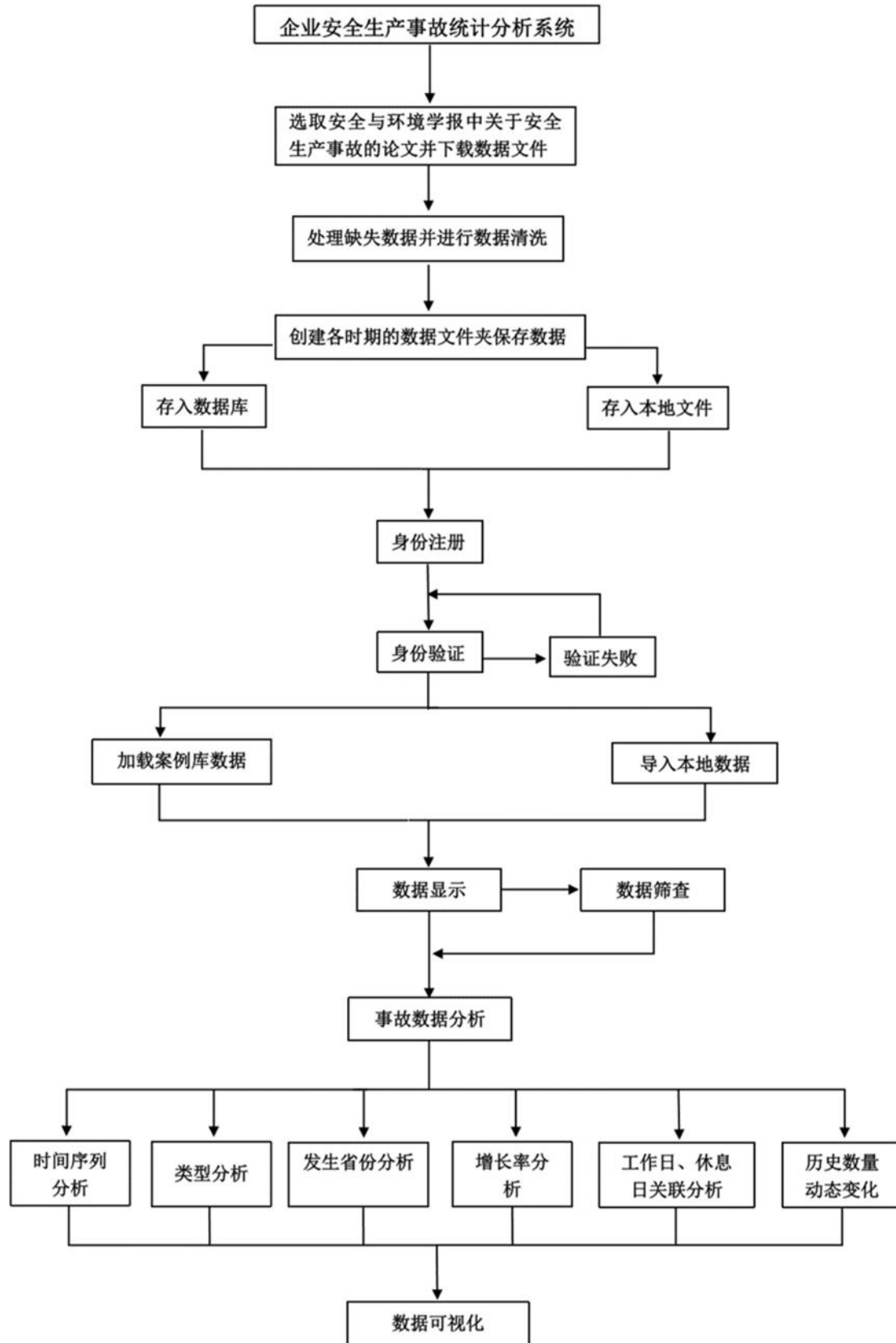


图18 系统的总体功能框图

## 第四章 系统实现方案

### 4.1系统方案架构

本系统基于Python编程语言技术，利用Tkinter制作系统实现界面可视化。通过Python获取《安全与环境学报》中的相关事故数据并对数据进行预处理，根据企业和安全专业研究领域的实际需求设计并实现六大分析功能选项，选取Tkinter可视化制作系统实现功能操作的界面显示。一方面，系统可服务于企业，实现企业自有事故数据的全面分析，促进企业搞好安全生产管理，合理制定安全事故管理策略。另一方面，系统可用于安全专业领域的研究中，各种企业事故数据的汇总统计利于安全研究员了解现在的安全生产的总体形势，根据系统的各种分析结果预测未来的安全生产事故的发生概率等，更好地把握安全生产规律，防范化解重大安全风险（如图19）。

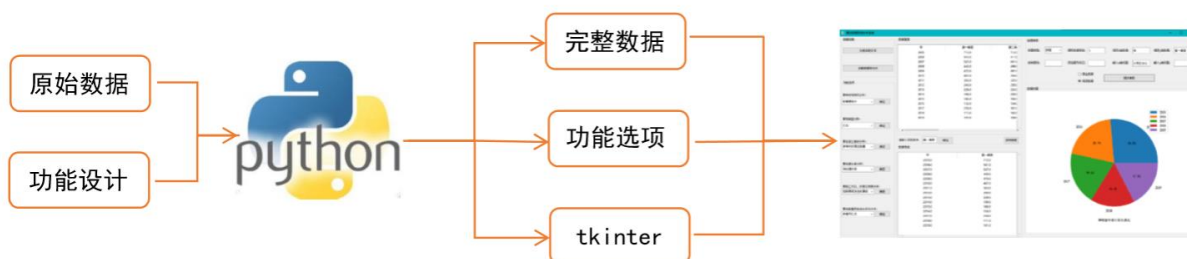


图19 系统方案架构流程图

### 4.2系统技术方案

#### 4.2.1编程语言选择

在系统初期构思设计时，本项目小组就系统编程语言、界面可视化框架的选用，结合当今时代的信息潮流与编程语言使用的限制条件进行了深层讨论。被普遍应用的编程语言有Python和Matlab两种。但是由于美国下达Matlab禁令，Matlab编程语言的使用受到限制，本项目小组便选择Python编程语言作为本系统的语言技术支持。Python是一种跨平台的计算机程序设计语言，随着其版本的不断更新和语言新功能的添加，被广泛应用于独立、大型项目的开发。此外，Python gui中的框架tkinter性能稳定，界面美观，易于学习和操作。综上，本小组确定系统编程语言选用

Python，界面可视化制作选用Tkinter。

## 4.2.2数据获取层面

数据获取的技术支持主要是Python编程语言，数据来源为中国知网——《安全与环境学报》中关于安全生产事故统计分析的论文。获取知网论文数据有两种方式，一是在线阅读，二是下载文件，下载格式为PDF或CAJ。其中在线阅读的获取数据方式又分为两种，一种为下载图表，另一种为直接导出Excel数据。本项目的首选获取方式为在线阅读，即直接导出Excel数据，不能直接导出的下载PDF文件，利用PDF编辑器获取文件中的事故数据。数据获取工作结束后，批量创建2004-2019年各时期的安全生产事故数据文件夹，再将原始数据存放至相应的文件夹中（如图20，图21，图22，图23）。



图20 在线阅读与文件下载

表1 2020年3—4月国内重大事故  
Table 1 Some major industrial accidents happened in China during period from March to April in 2020

时间	地点	事故	死亡/人
03-07	福建省泉州市	欣佳酒店坍塌	28
04-15	吉林省乾安县	3车相撞	12
总计	2起		40

图21 下载图表

表1 2004年11-12月国内特大事故 (死亡10人以上)

Table 1 Some major accidents happened in China during November to December in 2004 (more than ten people dead)

时间	地点	事故	死亡	受伤
11-04	山东日照市境内日东高速公路	追尾	20	
11-05	山西朔州市平鲁区石崖湾煤矿	瓦斯爆炸	16	
11-08	上海奉贤县境内A30高速公路近沿钱公路	撞车	10	16
11-10	上海奉贤区A30高速公路	撞车	10	15
11-11	四川成都市彭州市白鹿镇宏盛煤矿	瓦斯爆炸	12	7
11-11	河南平顶山市鲁山县新生煤矿南店报废井区	瓦斯爆炸	33	6

图22 导出EXCEL数据

```
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2003_11-2004_2.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2004_11-12.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2004_3-6.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2004_7-10.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2005_1-2.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2005_11-12.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2005_3-4.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2005_5-6.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2005_7-8.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2005_9-10.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2006_1-2.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2006_11-12.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2006_3-4.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2006_5-6.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2006_7-8.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2006_9-10.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2007_1-2.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2007_11-12.xls
C:\Users\86177\Desktop\新建文件夹\中国安全生产事故数据\2007_3-4.xls
```

图23 批量创建各时期文件夹

### 4.2.3数据处理层面

本系统的数据处理同样是依靠Python编程语言实现的，主要包括三大模块，即原始数据的清洗、数据的筛查和事故数据的分析。

在原始数据清洗前，对数据进行略览，查找异常数据与缺失数据。利用Python编写功能函数，使其可以计算数据量、输出数据条目、输出索引列、计算缺失值数量，程序运行后会统计每一列的数据缺失值。发现自2015年5-6月份起就没有统计事故受伤人数，所以剔除受伤人数一栏。剔除整理后对各时期文件夹数据进行逐一排查，发现2010年1-2月文件夹里缺失196条数据，2015年9-10月文件夹里缺失19条数据，除此之外其余文件夹内数据均完整无误。针对两部分的缺失数据，利用函数编写提取缺失数据前后各三年的对应月份的事故数据，采用插值法来计算缺失数值（如图24，图25，图26）。

```
def Data_glimpse():
    import glob
    path_dir = glob.glob("D:\\product_accident\\zone_accident_data\\*")
    path_files = [path_dir][0]
    n = 1
    for path in path_files:
        data = pd.read_excel(path, index_col = 0)
        data_counts = len(data) # 计算数据量
        columns = data.columns.tolist() # 输出数据columns
        #indexs = data.index.tolist() # 输出索引列
        nan_counts = len(data[data.isnull().values == True]) # 计算缺失值数量
        time = path.split("\\")[-1]
        #nan_counts = data.isna().sum() # 这种方式会把数据在每一列的数据缺失值给统计出来
        print("第{}个文件{} \n数据量为: {} \n列表数据字段: {} \n其中缺失数据量: {}".format(n, time, data_counts, columns,
        print('-----')
        n += 1
```

图24 数据略览

```
def data_filter():
    import glob
    path_dir = glob.glob("D:\\product_accident\\zone_accident_data\\*")
    path_files = [path_dir][0]

    ls_accident_num = []
    ls_death_num = []
    ls_injured_num = []
    for path in path_files[18:54:6]:
        data_complete = pd.read_excel(path, index_col = 0)

        ls_accident_num.append(data_complete["事故/起"].iloc[2:-6].values.tolist())
        ls_death_num.append(data_complete["死亡/人"].iloc[2:-6].values.tolist())
        ls_injured_num.append(data_complete["受伤/人"].iloc[2:-6].values.tolist())
    return(ls_accident_num, ls_death_num, ls_injured_num)
#该部分的函数功能是提取前后各三年1-2月份的数据
```

图25 提取前后三年1-2月份的数据

```
def data_interpolate(ls):
    num = len(ls[0])
    ls_mix = [[] for i in range(num)] #创建收纳的列表
    for item in ls:
        for i in range(num):
            ls_mix[i].append(item[i])

    ls_final = []
    for item in ls_mix:
        ts = pd.Series(item)
        u = round(ts.mean(), 0)
        ts.fillna(u, inplace=True)
        ls_final.append(int(ts.iloc[3]))
    return(ls_final)
#该部分函数的功能是进行缺失数据的插值（采用均值进行插值）

ls_accident_num, ls_death_num, ls_injured_num = data_filter()

data_accident = data_interpolate(ls_accident_num)
data_death = data_interpolate(ls_death_num)
data_injured = data_interpolate(ls_injured_num)
print(data_accident, data_death, data_injured)
```

[9, 2, 16, 20, 16, 2, 1, 3, 9, 7, 13, 2, 6, 3, 1, 6, 3, 0, 8, 5, 4, 0, 23, 0] [37, 6, 73, 69, 68, 1, 3, 30, 35, 40, 54, 8, 37, 8, 3, 27, 20, 0, 41, 20, 20, 0, 106, 0] [26, 5, 30, 25, 62, 13, 0, 44, 14, 51, 48, 34, 42, 3, 30, 16, 37, 0, 10, 5, 9, 0, 68, 0]

图26插值法处理缺失数据

事故数据分析依照功能设计共有六个分析模块：

### (1) 事故时间序列分析

通过功能函数编写剔除全国事故汇总数据中的省份，填补 2015 年 5-6 月份后的缺失的受伤人数数据，填补数值为 0。对填补后的数据按照时间字段划分，分为年、上半年和下半年、月、季度并按其时间单位提取相对应的事故数量、死亡人数、受伤人数。

### (2) 事故类型分析

与事故时间序列分析相同，填补缺失数据，对填补后的数据按照时间字段划分，分为年、月、季度并按其时间单位提取相对应的事故类型。

### (3) 事故省份分析

事故省份分析包括各省份随年份的事故数量关系、各省份死亡人数与 GDP 之间的关联分析。前者是以年为时间单位对事故发生省份的累积统计，固定省份，提取每年发生的事故数量并计算总和和占比。后者的实现需要先读取各省市的“十一五”

至“十三五”期间的 GDP 数据，利用功能函数计算各省市每年 GDP 的增速平均值，添加中国地图、气泡图、省份信息，设置网格线，合并最终数据。

#### (4) 事故增长率分析

事故增长率分析分为环比增长率和同比增长率，均通过 Python 编程语言实现处理计算。在实现环比增长率分析时，与《安全与环境学报》一致，采用两个月的数据进行环比。在计算同比增长率时，需要制定参考对象，这里分别以两个月、上半年、年为参考对象进行函数编写及封装。

#### (5) 事故工作日、休息日关联度分析

实现此分析功能的主要思路为筛选出在节假日、工作日、加班和正常休的事故数量，进行事故占比和各个时段的平均事故率分析，时段分为十一五、十二五、十三五三个时段。

首先加载“十一五”至“十三五”期间特大事故数据，编写获取事故发生的年月日函数并运行，再加载 2005 年-2019 年节假日与加班数据，将其转变为可用的数据格式便于后续工作。对提取的特大事故发生年月日进行判断其是否时休息日，是休息日即 holiday (true)，否即 holiday (false)，判断后对发生在节假日的事故条目添加节假日数据。由于节假日存在调休情况，特此添加因调休周末上班发生的事故数据，最后加载正常假期内周末的事故情况，用总事故数减去节假日事故数、周末休息日事故数、因调休发生的事故数得出正常工作日事故数量，综上即可得出工作日与非工作日之间的事故比例。

由于事故率的计算牵涉到天数，特此说明：本系统时间天数按照一年 365 天，虽然我国实际节假日在 2005 年-2018 年间有所变动，但实际有效假日变化不大，有效工作日为 250 天，剩下的为休息日 115 天。2007 年以后将五一假期缩短，但增加了端午节和中秋节假日，实际有效假日几乎没有变动，此外 2015 年增加一个胜利日，即这一年增加一天假，实际计算以下表和所示公式为准。（如表 1）

类别		事故数量（起）	事故占比
非工作日 (33.76%)	节假日发生特大事故	64	7.50%
	正常周末休息发生的特大事故	224	26.26%
工作日 (66.24%)	正常上班发生的特大事故	550	64.48%
	因节假日调休周末上班发生的特大事故	15	1.76%

表 1 工作日与非工作日特大事故起数

结合公式①②③，分别计算日平均特大事故率、工作日平均特大事故率和非工作日平均特大事故率（以下分别称为“日平均”、“工作日平均”、“非工作日平均”）。

$$\text{日 平均特大事故率} = \frac{\text{统计时长内发生特大事故数量}}{\text{统计时长}} = \frac{832}{365 \times 13} * 100\% = 16.481\% \quad \text{①}$$

$$\text{工作日平均特大事故率} = \frac{\text{统计工作日时长内发生特大事故数量}}{\text{统计工作日时长}} = \frac{539+14}{250 \times 13} * 100\% = 17.02\% \quad \text{②}$$

$$\text{非工作日平均特大事故率} = \frac{\text{统计非工作日时长内发生特大事故数量}}{\text{统计非工作日时长}} = \frac{62+217}{115 \times 13} * 100\% = 18.66\% \quad \text{③}$$

## (6) 事故数量历史动态变化分析

按照事故统计的两月和事故汇总进行动态分析，利用 Python 编写功能函数，将用于分析的数据生成 history\_rank\_month.csv 或 history\_rank\_acc.csv 文件保存至本地。

### 4.2.4 数据显示层面

数据显示即为数据可视化的表现，本系统的可视化制作依赖于Python gui框架中的Tkinter，实现可视化的具体编码文件包括：创建登陆界面、程序主窗体和设置数据筛选、绘图的project.py主文件，负责设置事故时间序列分析、事故类型分析、事故发生省份分析、事故增长率分析、事故工作日和休息日关联分析、事故数量历史动态变化分析这六个功能选项的编码文件分别是time\_series\_analysis.py文件、accident\_type\_analysis.py文件、accident\_zone\_analysis.py文件、accident\_rate\_analysis.py文件、work\_day\_and\_rest\_day\_analysis.py文件、history\_rank.py文件。

事故分析的前五个模块都可利用tkinter在软件内部进行绘图展示，事故数量历史动态变化分析是动态变化的图形，无法直接内嵌于软件之中，因此，动态分析可视化采用网页模式，进行数据的加载和可视化展示。（如图27）



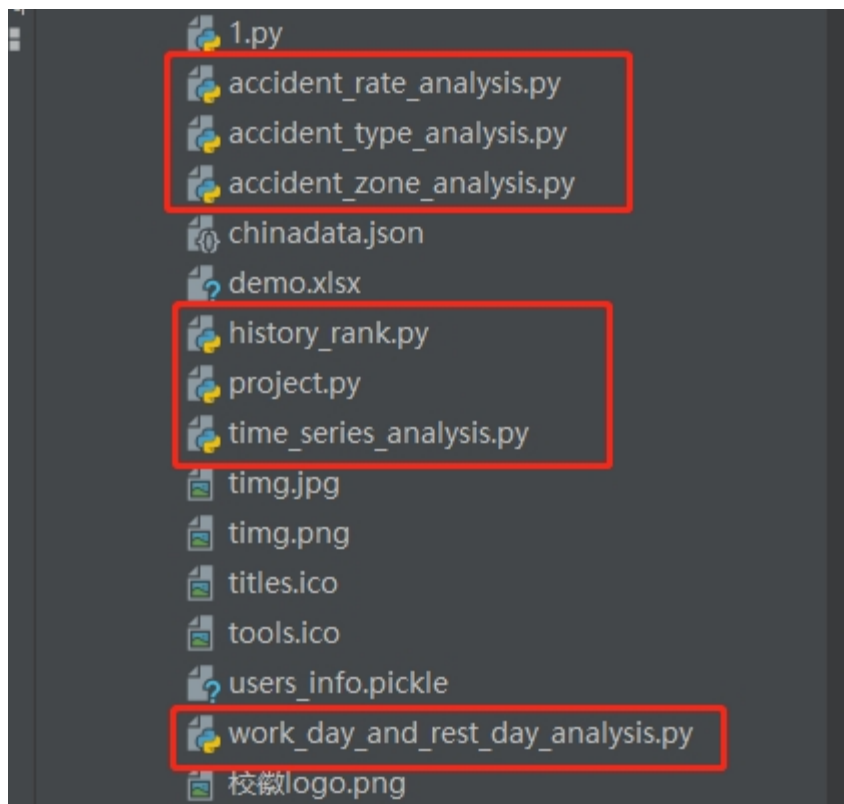


图27 Tkinter框架文件

## 第五章 项目效果与演示实验

### 5.1 企业事故统计分析系统初始界面介绍

#### 5.1.1 登录界面及注册界面

企业事故统计分析系统初始界面是一个基础的登录界面（如图 28），只要输入账号及密码即可登录系统，系统设置的初始账号密码都是 admin。



图 28 登录界面

如果输入的账号不在系统的储存名单内，会检测到账号未注册，弹出注册的提醒（如图 29），这时候点击“是”选项，或者点击“注册”选项都会跳转到注册的页面（如图 30）用户也可以自己申请一个账号进行使用。



图 29 提醒情况下的登录界面



图 30 注册界面

注册完毕，点击登录后，账号密码系统核实无误会提醒登录成功，欢迎使用的提示语（如图 31）。



图 31 登陆成功界面

5.1.2数据分析系统界面

在登陆之后用户就会看到本系统最重要的数据分析系统界面，该界面主要分为六大部分：①数据加载；②功能选项设置；③数据查看；④数据筛选；⑤绘图参数设置；⑥数据绘图（如图 32）

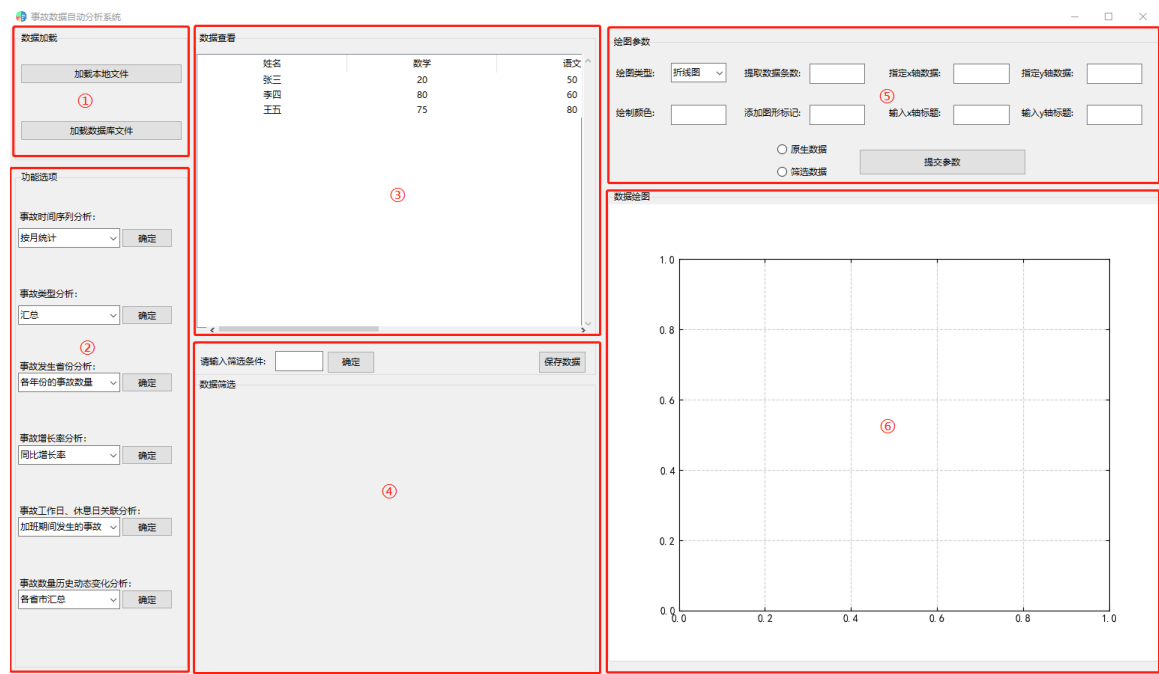


图 32 数据分析系统界面

数据加载部分，就是选择数据导入系统中；功能选择部分则是选择用户要使用的功能；数据查看部分就是查看用户所导入的数据；数据筛选部分就是按用户要求筛选出需要的数据；绘图参数设置部分就是设置用户要绘制的图表的具体要求；数据绘图部分就是绘制出用户设置好的图表。

## 5.2 系统运行效果演示实验

### 5.2.1 系统功能演示

在打开数据分析系统界面后，会自动加载 demo 文件（存放在本地计算机内）中的数据，然后将数据放置在数据查看的板块中。整个系统的工作流程如下：

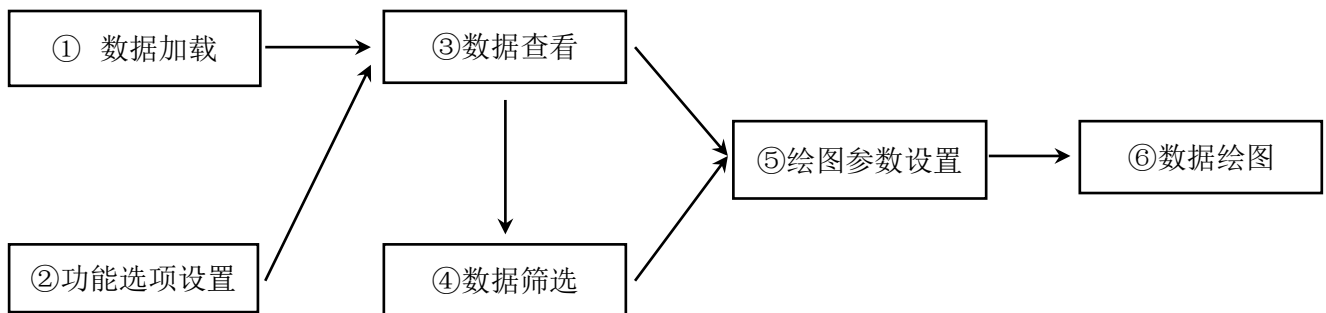


图 33 系统的工作流程图

本系统共有四个工作流，接下来会分别演示：

（1）第一种工作流：①③⑤⑥

首先，加载默认的 demo 文件数据，实现本地数据文件导入到软件系统的数据查看板块，然后进行绘图参数的设置，假定选择默认的折线图。其中选择参数这里，要多选的时候，使用|分离标题字段，直接有③到⑤，选择原生数据进行绘图。（如图 34）

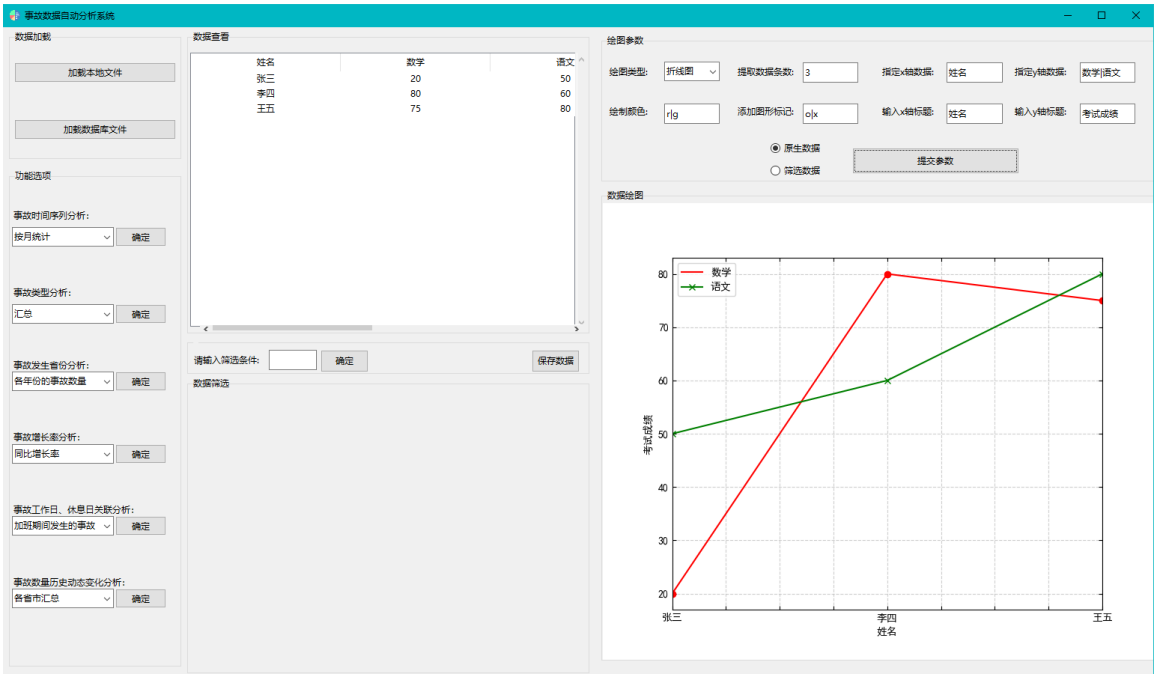


图 34 ①③⑤⑥ workflow 结果演示

(2) 第二种 workflow：①③④⑤⑥

这个 workflow 就是将数据查看中的内容进行筛选，然后进行数据绘图，要注意筛选功能的设置。这里会提到三种情况：

**第一种情况：**如果没有输入任何内容点击确认筛选，系统会提示筛选条件有误，请重新输入（如图 35）。

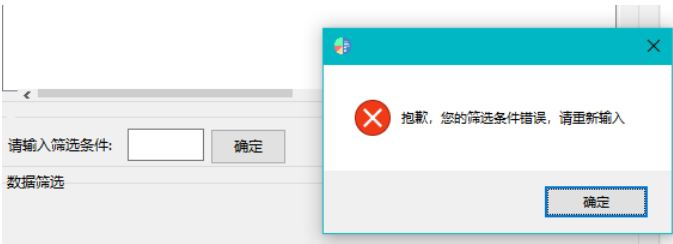


图 35 未输入任何筛选条件情况

**第二种情况：**如果输入的筛选值是在数据的标题上面，系统会返回该标题所在的那一列数据，多选的时候就返回多列（如图 36）。



图 36 筛选值在数据的标题上



图 37 筛选值在数据的标题下

第三种情况：如果选择的是标题下面的内容，系统就会返回筛选内容该行的数据，并包含标题（如图 37）。

接着就是选择筛选中的数据进行绘图，这里以条状图为例（如图 38,图 39）：

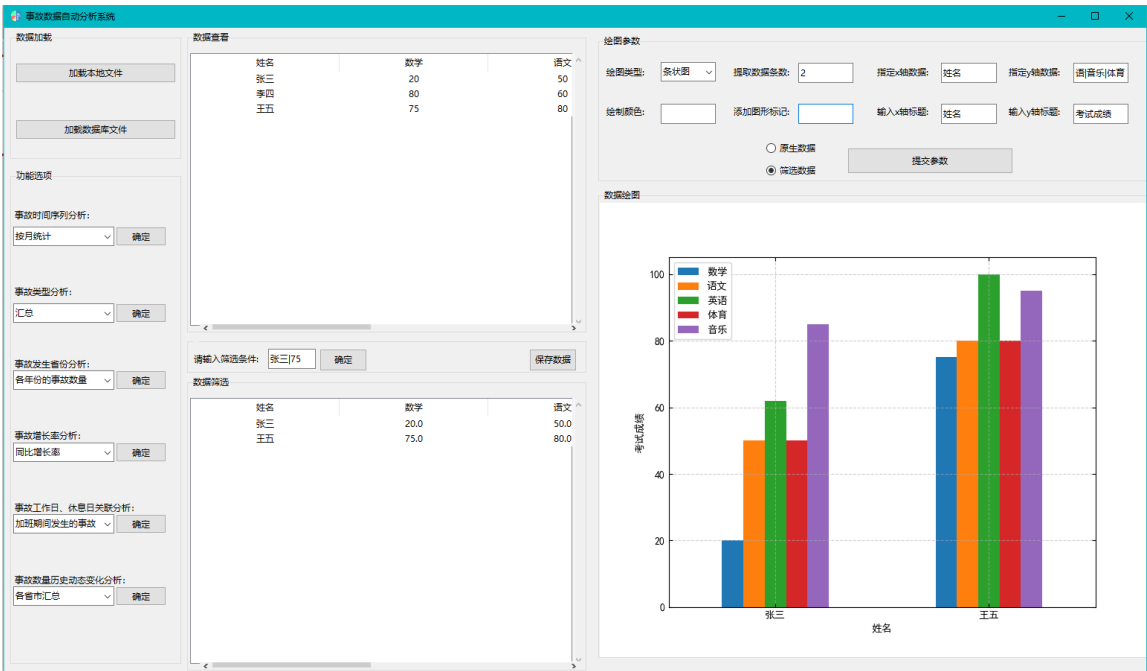


图 38 不堆叠

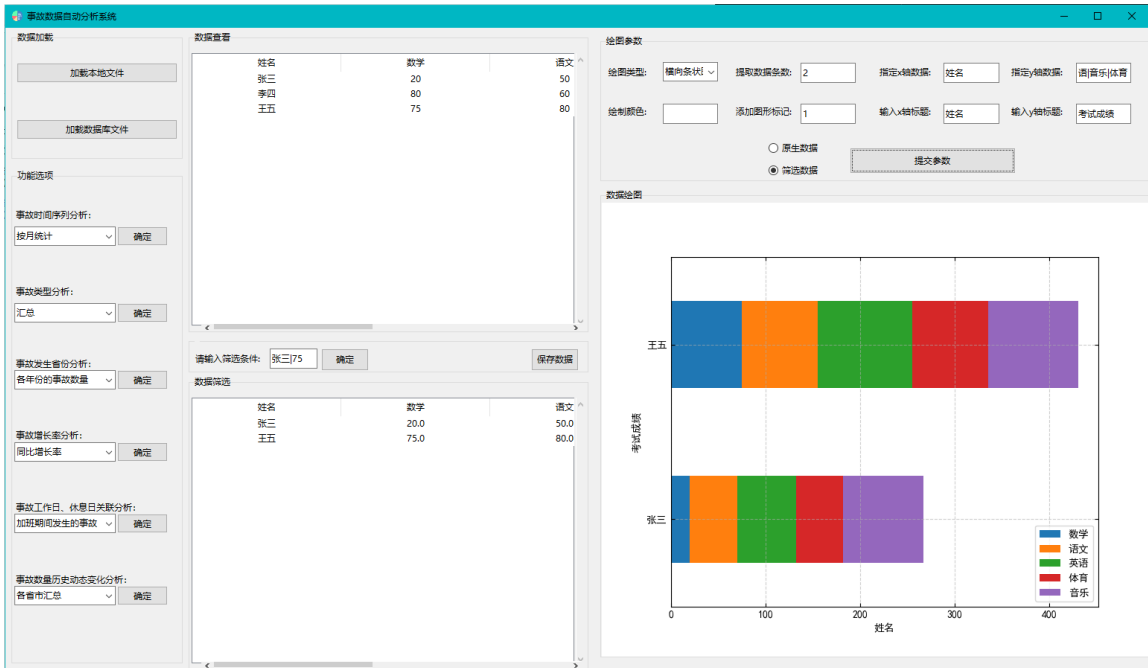


图 39 堆叠

(3) 第三/四种 workflow：②③⑤⑥/②③④⑤⑥

这里选择事故类型汇总，一般是要出饼图占比，此处选择在功能选项中的事故类型分析，然后默认即为汇总，点击确认即可（如图 40）。若使用筛选数据绘制（即使选择了绘制 6 条数据，但是只用 4 条，所以也只能绘制 4 条数据）（如图 41）。

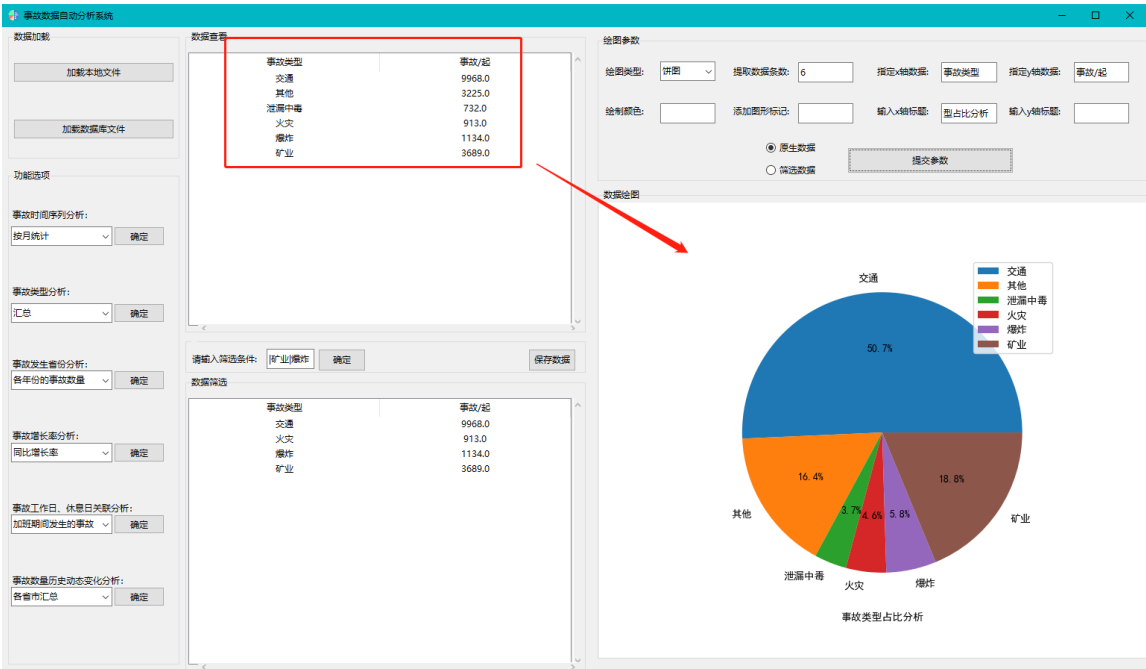


图 40 第三/四种 workflow 结果演示

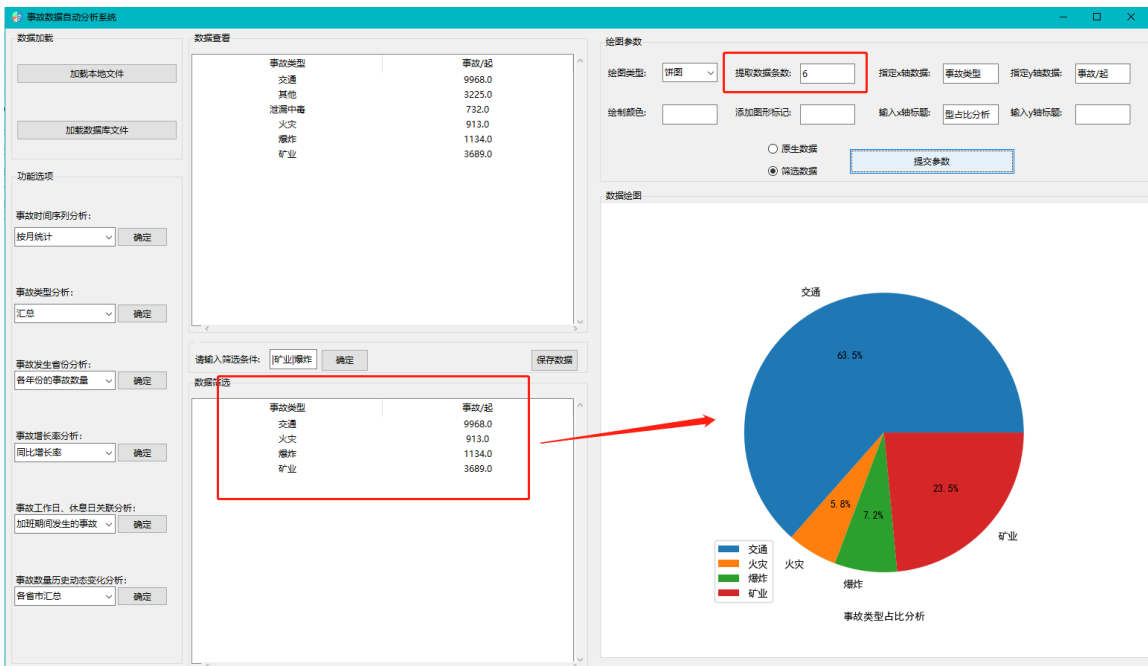


图 41 筛选数据后结果演示

5.2.2详细设置介绍及演示

(1) 数据加载

在此版块中，有两个选项，可以根据需要，选择对应的数据加载，如果点击加载本地数据，系统会打开文件对话框，提示选择要加载的数据文件，最后点击确定即可（如图 42）。



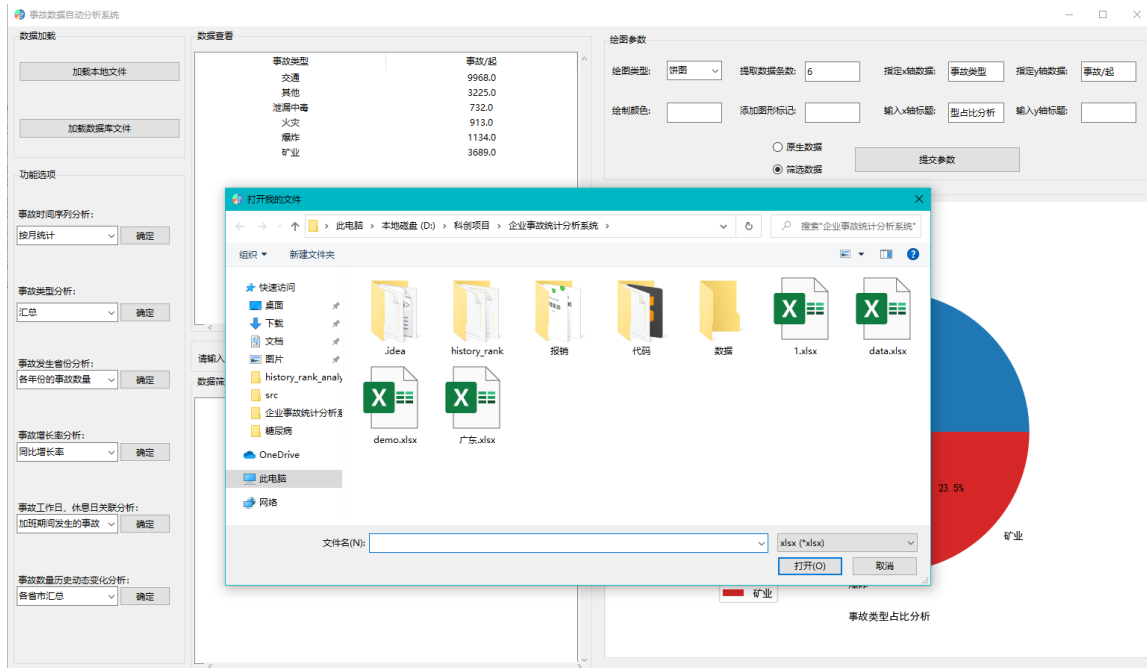


图 42 数据加载操作

举个例子，比如选择 data.xlsx 文件，里面记录了我国安全生产事故的数据，最后加载到系统中如下（会覆盖掉之前的 demo.xlsx 中的数据）（如图 43）。

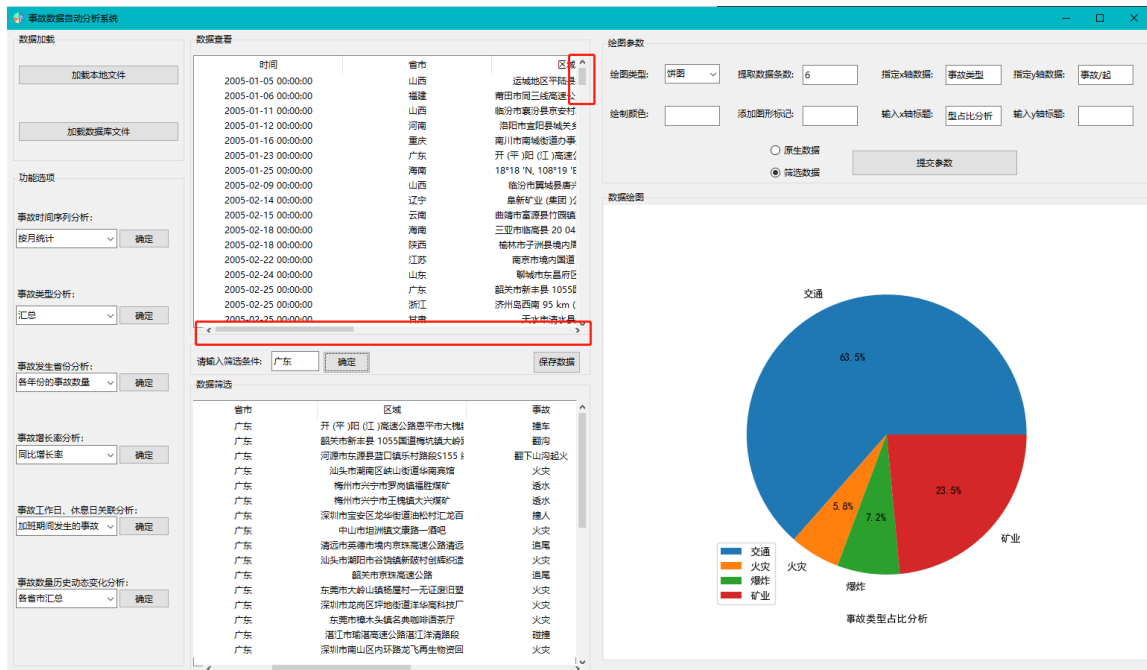


图 43 选择 data.xlsx 文件的实例演示

如果选择的是加载数据库文件，系统后台已经连接到 mysql 数据库，只要一点击，系统就自动将数据库中 accident\_data 表单中的数据全部加载进来（这里的数据和本地的数据一致，只不过多了一列主键 id 列）（如图 44）。

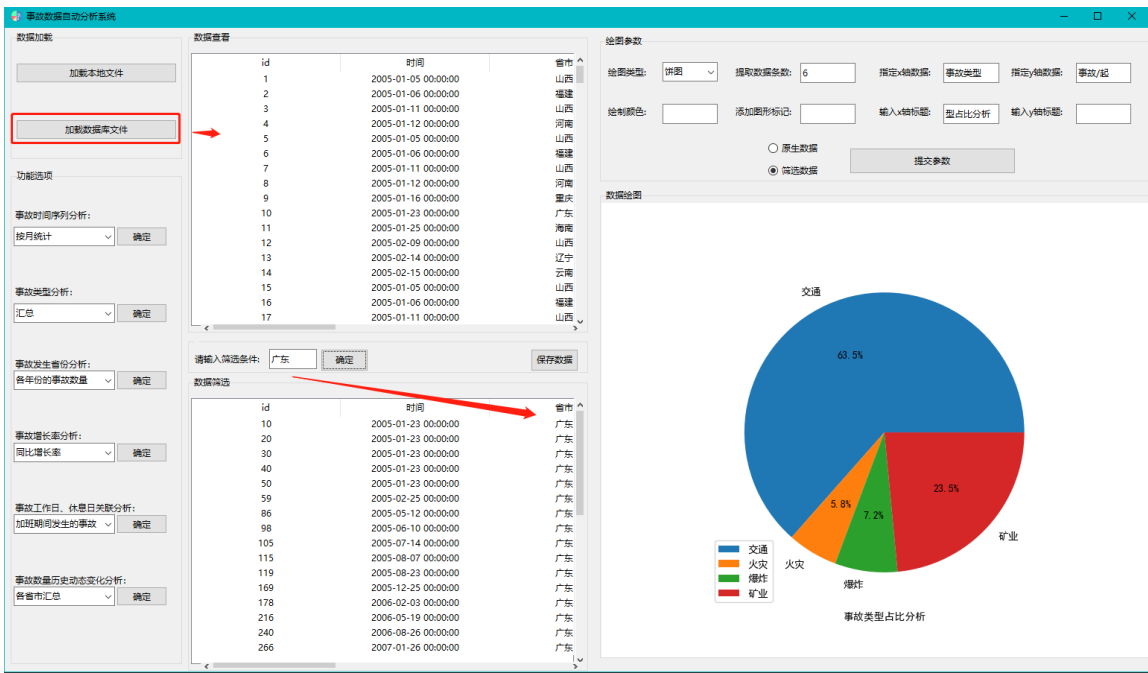


图 44 选择加载数据库文件的实例演示

(2) 功能具体演示

下面分为六个模块逐一说明，分别为事故发生时间序列分析、事故类型分析、事故发生省份分析、事故增长率分析、事故发生同工作日及节假日关联度分析、事故数量历史动态变化分析。

**事故发生时间序列分析：**在时间序列分析模式中可以选
 择按年、按月、按季度分析，接下来会分别演示到（如图 45，图 46，图 47，图 48，图 49）（注：原生数据指的是直接导入未经筛选的数据；筛选数据就是经过筛选的原生数据）：

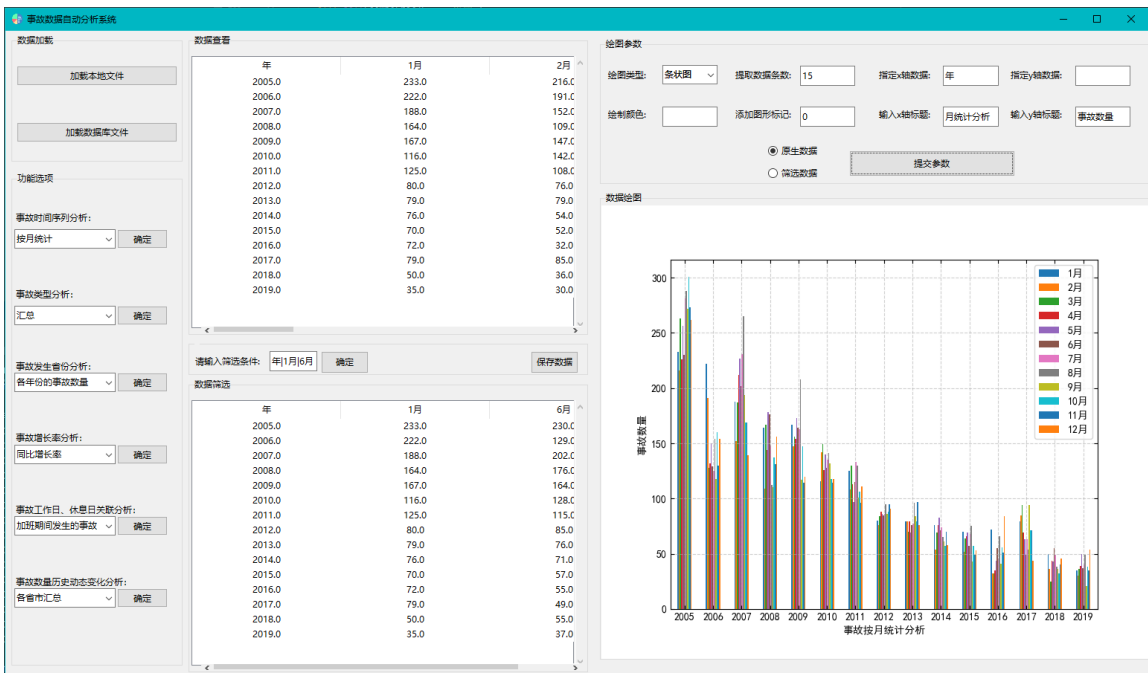


图 45 按月统计（选择原生数据）

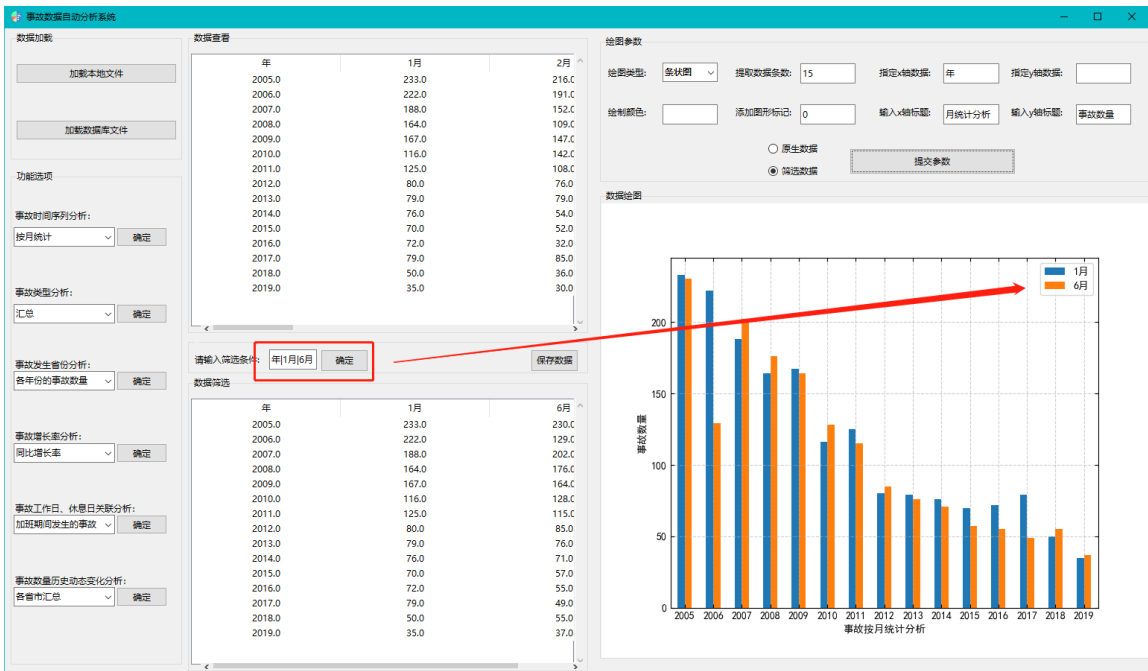


图 46 按月统计（选择筛选数据）

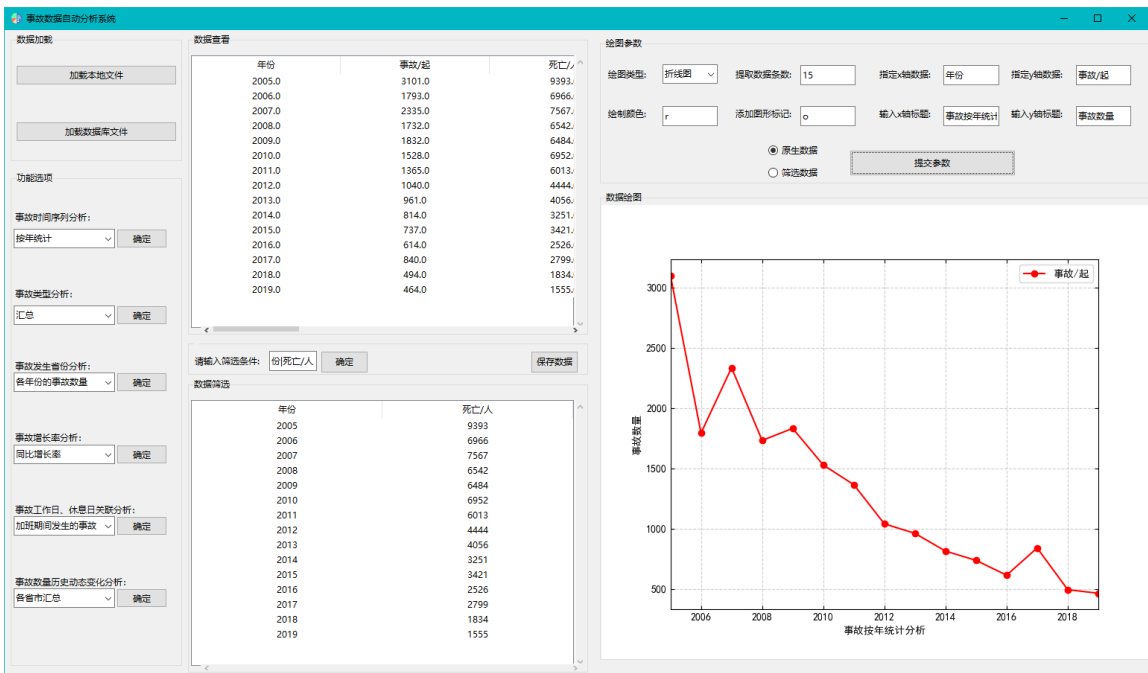


图 47 按年统计（此处拿原生数据绘制每年的事故数量的折线图）

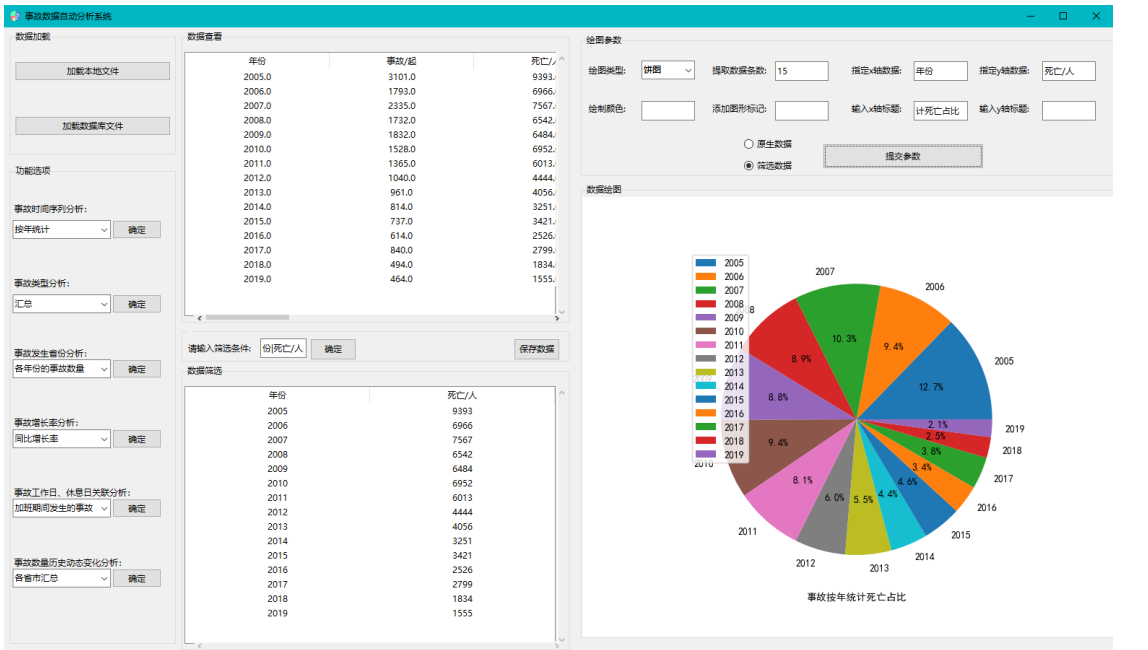


图 48 按年统计（此处用筛选数据绘制按年统计的死亡人数占比图）

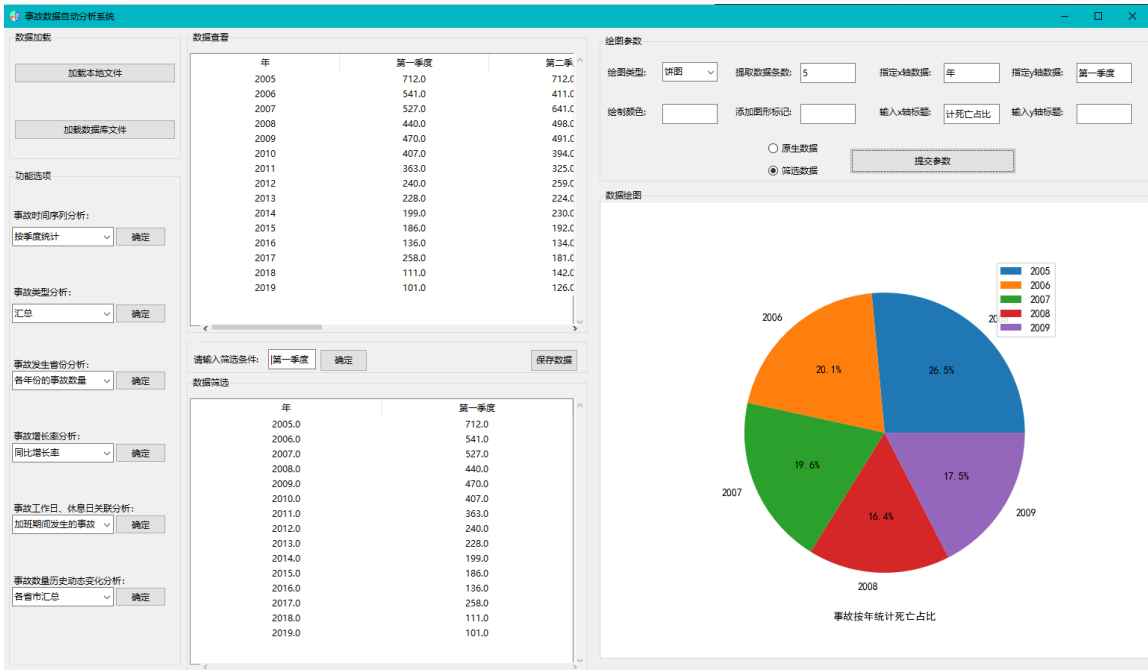


图 49 按季度统计（此处演示绘制前五年的事故数据占比）

**事故类型分析：**该模式默认情况下是进行各类事故的数据的统计，这个在之前的功能测试 demo 中已经进行分析，接下来就是具体演示（如图 50，图 51，图 52）：

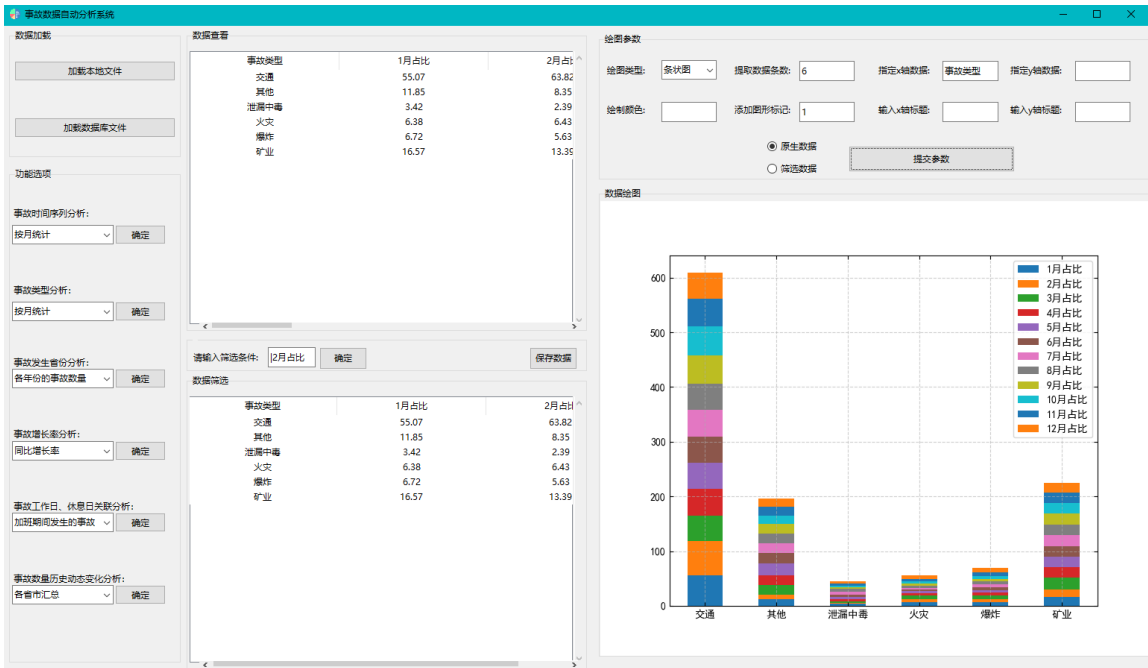


图 50 按月统计（原生数据）

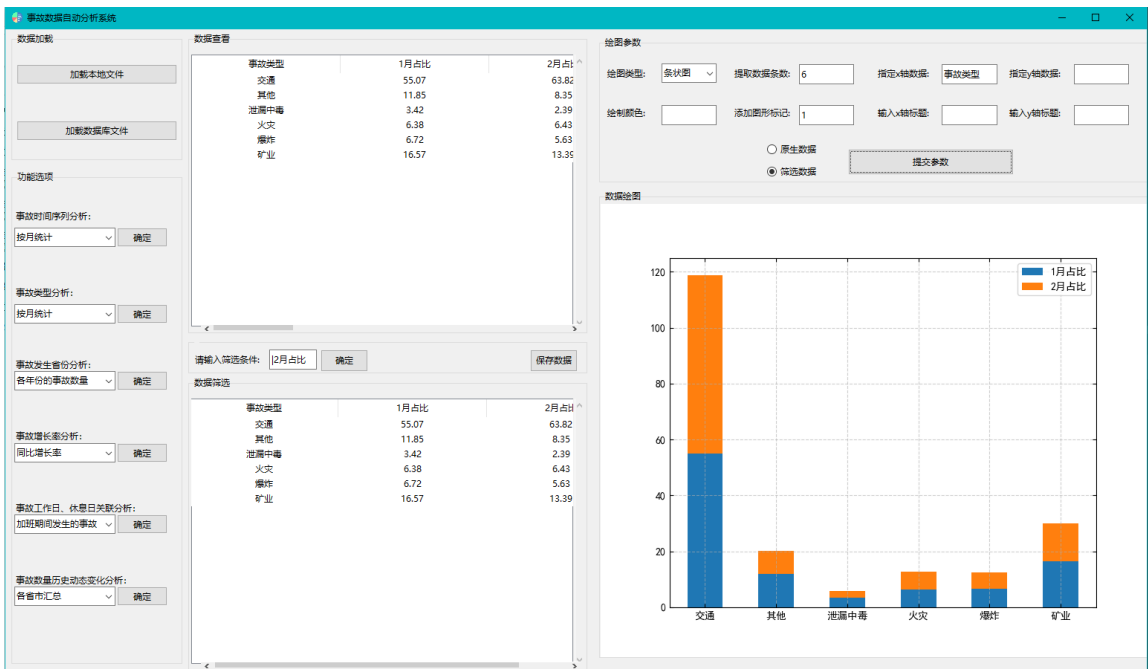


图 51 按月统计（筛选数据）

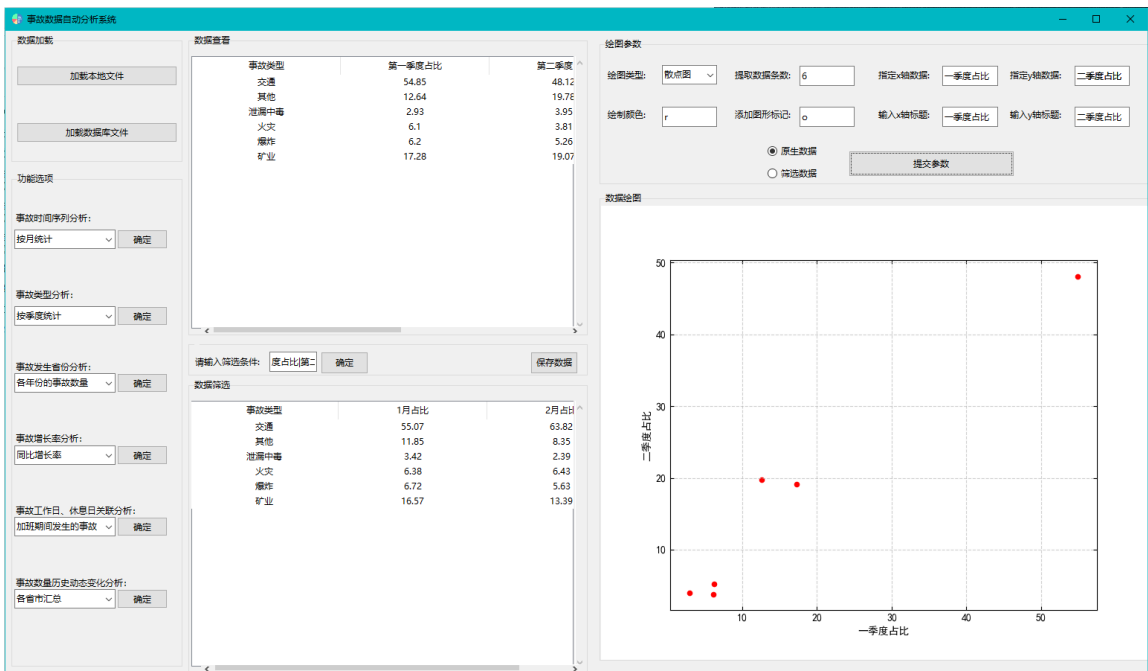


图 52 按季度分析事故类型占比（散点图，点分散在斜对角线，说明每个季度的事故类型占比的变化不大）

**事故发生省份分析：**事故省份分析可以分析中国各省份随年份的事故数量关系以及各省份死亡人数与 GDP 之间的关联分析（如图 53，图 54）。

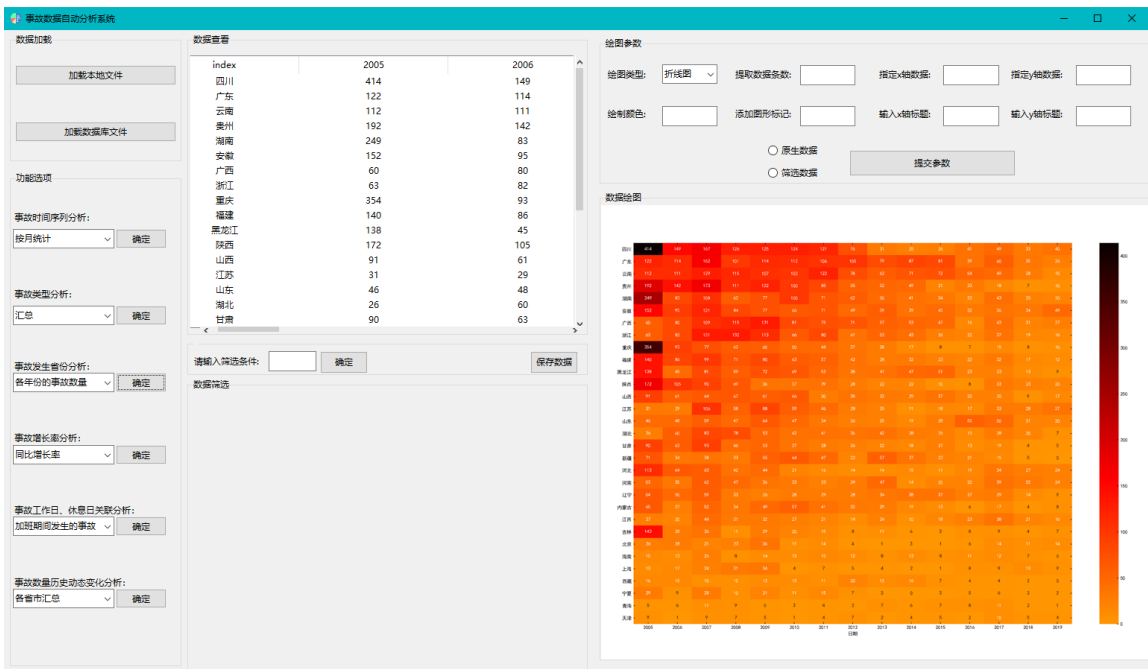


图 53 各省份随年份的事故数量关系

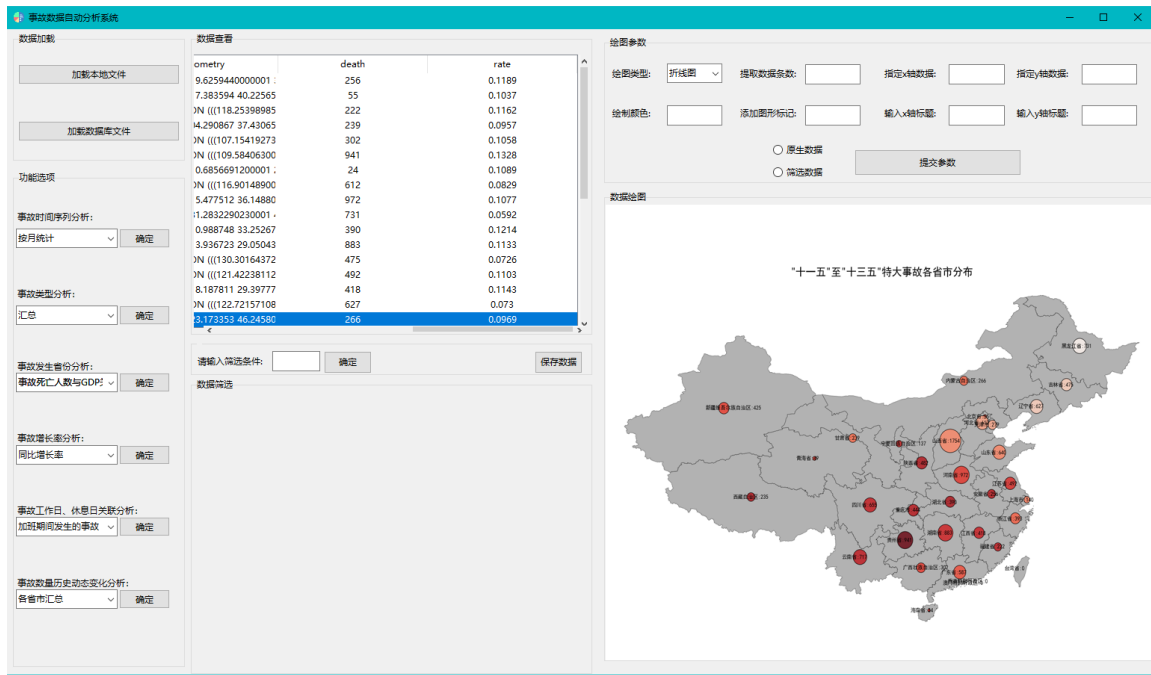


图 54 各省份死亡人数与 GDP 之间的关联分析

**事故增长率分析：**增长率分析能直观地反映出每年事故发生的同比以及环比增长率，以此来判断事故发生次数的增减，来推断安全措施的实施是否有效果（如图 55，图 56）。

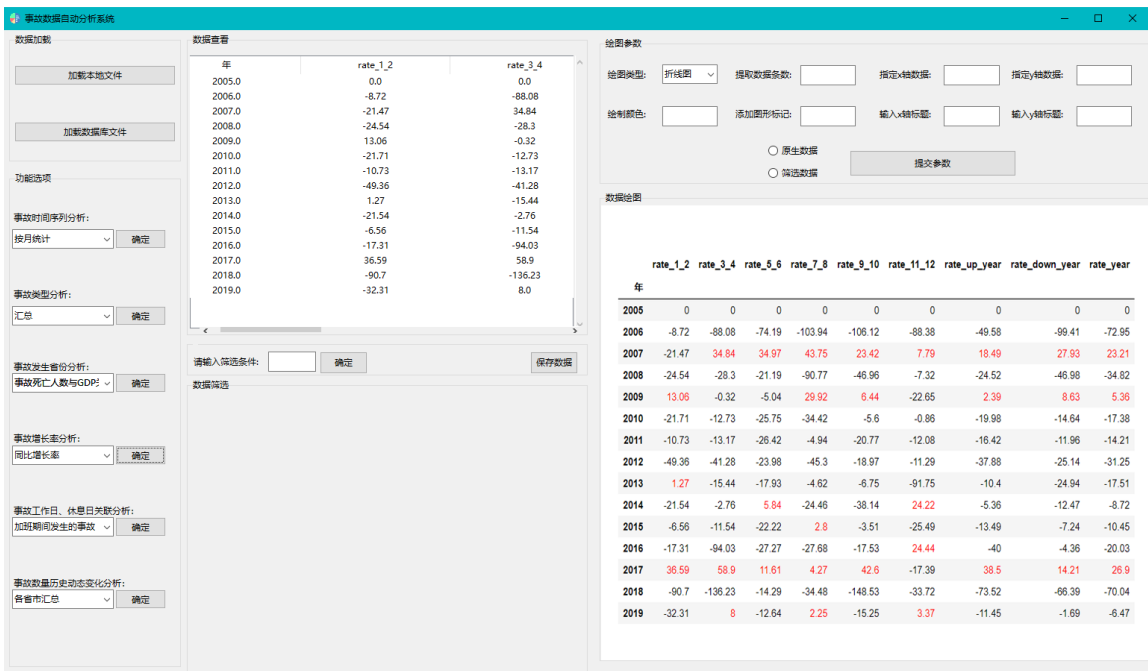


图 55 同比增长率分析

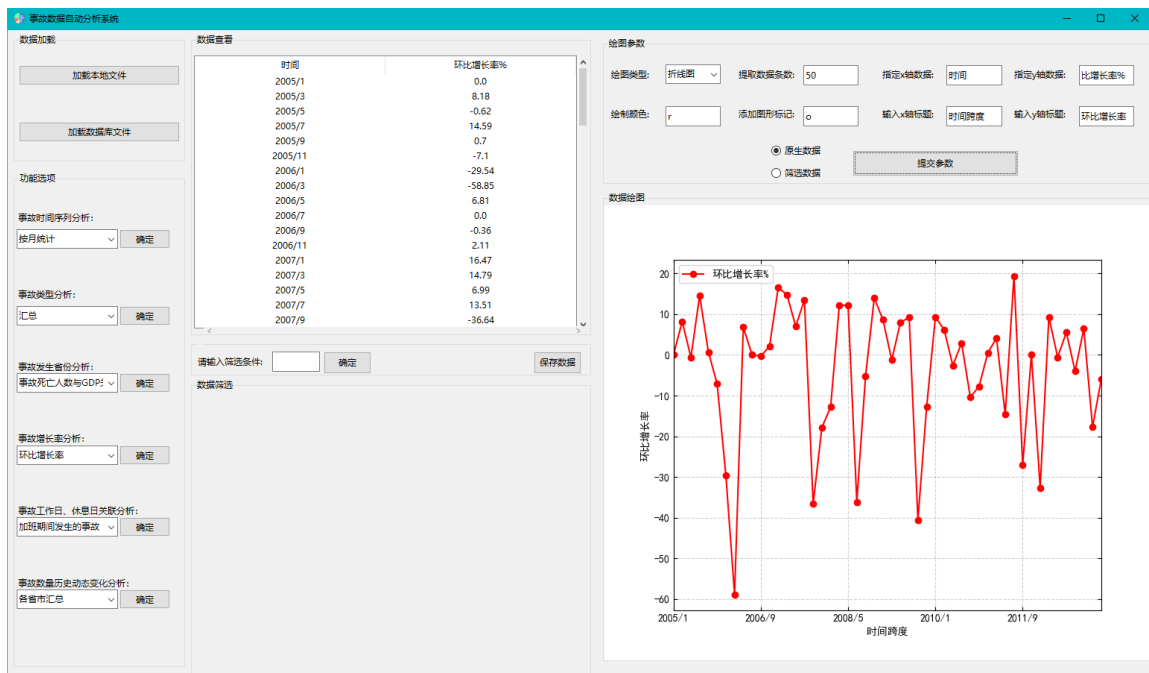


图 56 环比增长率分析

**事故同工作日及节假日关联度分析：**这一部分主要筛选的是安全生产事故发生在节假日、工作日、加班、和正常休息的事故数量，进行分析事故次数（绝对指标）和平均特大事故率（相对指标）。

由于这部分计算牵涉到时间天数，特此说明：2007 年以前，法定假日为 10 天；2007 年以后，国家取消了“五一”长假，增添了“清明”、“端午”、“中秋”假日，



节假日增加 1 天，2015 年为了纪念抗日战争胜利该年增加了一天。为了合理安排人员节假日旅游、交通运输、生产经营等有关工作，国务院办公厅常常会把法定节假日与相邻的周末休息时间连接在一起。因此，工作日包含正常上班的时间加上因节假日调整的周末时间，非公工作日为正常的周末休息时间加上法定节假日与其相邻的周末时间。按照一年 365 天进行计算，法定节假日为 11 天进行统计，每年正常的周末时间为 104 天，故每年工作日时长为 250 天，非工作日时长为 115 天。此处以“十一五”至“十三五”的事故情况（这里针对特大事故）来举例（如表 2、图 57，图 58，图 59，图 60）

类别		事故数量（起）	事故占比
非工作日 (33.76%)	节假日发生特大事故	64	7.50%
	正常周末休息发生的特大事故	224	26.26%
工作日 (66.24%)	正常上班发生的特大事故	550	64.48%
	因节假日调休周末上班发生的特大事故	15	1.76%

表 2 工作日与非工作日特大事故起数

结合公式①②③，分别计算日平均特大事故率、工作日平均特大事故率和非工作日平均特大事故率（以下分别称为“日平均”、“工作日平均”、“非工作日平均”）。

$$\text{日 平均特大事故率} = \frac{\text{统计时长内发生特大事故数量}}{\text{统计时长}} = \frac{832}{365 \times 13} * 100\% = 16.481\% \quad \text{①}$$

$$\text{工作日平均特大事故率} = \frac{\text{统计工作日时长内发生特大事故数量}}{\text{统计工作日时长}} = \frac{539+14}{250 \times 13} * 100\% = 17.02\% \quad \text{②}$$

$$\text{非工作日平均特大事故率} = \frac{\text{统计非工作日时长内发生特大事故数量}}{\text{统计非工作日时长}} = \frac{62+217}{115 \times 13} * 100\% = 18.66\% \quad \text{③}$$

结果

①各时段事故数量及占比

- 节假日发生事故起数： 64 事故占比为： 7.50%
- 正常周末休息发生的事故起数为： 224 事故占比为： 26.26%
- 因节假日调休周末上班发生的事故起数为： 15 事故占比为： 1.76%
- 正常上班发生的事故起数为：  $853 - 64 - 224 - 15 = 550$  事故占比为： 64.48%

②工作日与非工作日事故占比

- 工作日特重大事故占比：  $(550 + 15) / 853 = 66.24\%$
- 非工作日特重大事故占比：  $(224 + 64) / 853 = 33.76\%$

③ 事故率计算

按照一年是365天，我国实际节假日是虽然从05年到18年有所变动（实际的有效假日变化不大），有效工作提250天，剩下的就是休息日115天 2007年往后将五一长假缩短，添加了端午和中秋（实际有效假日几乎没有发生变化），然后15年增加一个胜利日（这一年是增加一天假）

- 平均事故率：  $853 / (365 \times 15) = 15.56\%$
- 工作日平均事故率  $(550 + 15) / (250 \times 15) = 15.07\%$
- 非工作日平均事故率  $(224 + 64) / (115 \times 15) = 16.70\%$

图 57 “十一五”至“十三五”的事故情况（这里针对特重大事故）

“十一五”安全生产特重大事故情况

①各时段事故数量及占比

- 节假日发生事故起数： 37 事故占比为： 7.82%
- 正常周末休息发生的事故起数为： 118 事故占比为： 24.95%
- 因节假日调休周末上班发生的事故起数为： 8 事故占比为： 1.69%
- 正常上班发生的事故起数为：  $473 - 118 - 37 - 8 = 310$  事故占比为： 65.54%

②工作日与非工作日事故占比

- 工作日特重大事故占比：  $(310 + 8) / 473 = 67.23\%$
- 非工作日特重大事故占比：  $(118 + 37) / 473 = 32.77\%$

③ 事故率计算

按照一年是365天，我国实际节假日是虽然从05年到18年有所变动（实际的有效假日变化不大），有效工作提250天，剩下的就是休息日115天 2007年往后将五一长假缩短，添加了端午和中秋（实际有效假日几乎没有发生变化），然后15年增加一个胜利日（这一年是增加一天假）

- 平均事故率：  $473 / (365 \times 5) = 25.92\%$
- 工作日平均事故率  $(310 + 8) / (250 \times 5) = 25.44\%$
- 非工作日平均事故率  $(118 + 37) / (115 \times 5) = 26.96\%$

图 58 “十一五”的事故情况（这里针对特重大事故）

①各时段事故数量及占比

- 节假日发生事故起数： 21 事故占比为： 7.98%
- 正常周末休息发生的事故起数为： 75 事故占比为： 28.75%
- 因节假日调休周末上班发生的事故起数为： 5 事故占比为： 1.90%
- 正常上班发生的事故起数为：  $263 - 75 - 21 - 5 - 8 = 162$  事故占比为： 61.60%

②工作日与非工作日事故占比

- 工作日特重大事故占比：  $(162 + 5) / 263 = 63.50\%$
- 非工作日特重大事故占比：  $(75 + 21) / 263 = 36.50\%$

③ 事故率计算

按照一年是365天，我国实际节假日是虽然从05年到18年有所变动（实际的有效假日变化不大），有效工作提250天，剩下的就是休息日115天 2007年往后将五一长假缩短，添加了端午和中秋（实际有效假日几乎没有发生变化），然后15年增加一个胜利日（这一年是增加一天假）

- 平均事故率：  $263 / (365 \times 5) = 14.41\%$
- 工作日平均事故率  $(162 + 5) / (250 \times 5) = 13.36\%$
- 非工作日平均事故率  $(75 + 21) / (115 \times 5) = 16.70\%$

图 59 “十二五”的事故情况（这里针对特重大事故）

①各时段事故数量及占比

- 节假日发生事故起数： 5 事故占比为： 4.27%
- 正常周末休息发生的事故起数为： 31 事故占比为： 26.50%
- 因节假日调休周末上班发生的事故起数为： 2 事故占比为： 1.71%
- 正常上班发生的事故起数为：  $117 - 31 - 6 - 2 = 78$  事故占比为： 66.67%

② 工作日与非工作日事故占比

- 工作日特重大事故占比：  $(78 + 2) / 117 = 68.38\%$
- 非工作日特重大事故占比：  $(31 + 6) / 117 = 31.62\%$

③ 事故率计算

按照一年是365天，我国实际节假日是虽然从05年到19年有所变动（实际的有效假日变化不大），有效工作提250天，剩下的就是休息日115天 2007年往后将五一长假缩短，添加了端午和中秋（实际有效假日几乎没有发生变化），然后15年增加一个胜利日（这一年是增加一天假）

- 平均事故率：  $116 / (365 \times 5) = 6.41\%$
- 工作日平均事故率  $(78 + 2) / (250 \times 5) = 6.40\%$
- 非工作日平均事故率  $(31 + 6) / (115 \times 5) = 6.43\%$

图 60 “十三五”的事故情况（这里针对特重大事故）

**事故数量历史动态变化分析：**这一部分数据的体现是动态的变化的图形，没有办法直接内嵌到软件之中，因此采用的方式通过网页直接加载数据，进行可视化展示。此处将按照事故统计的两月进行动态分析，执行该功能选项后，系统会自动生成 history\_rank\_month.csv 文件保存本地，并将数据加载到软件当中（如图 61）。

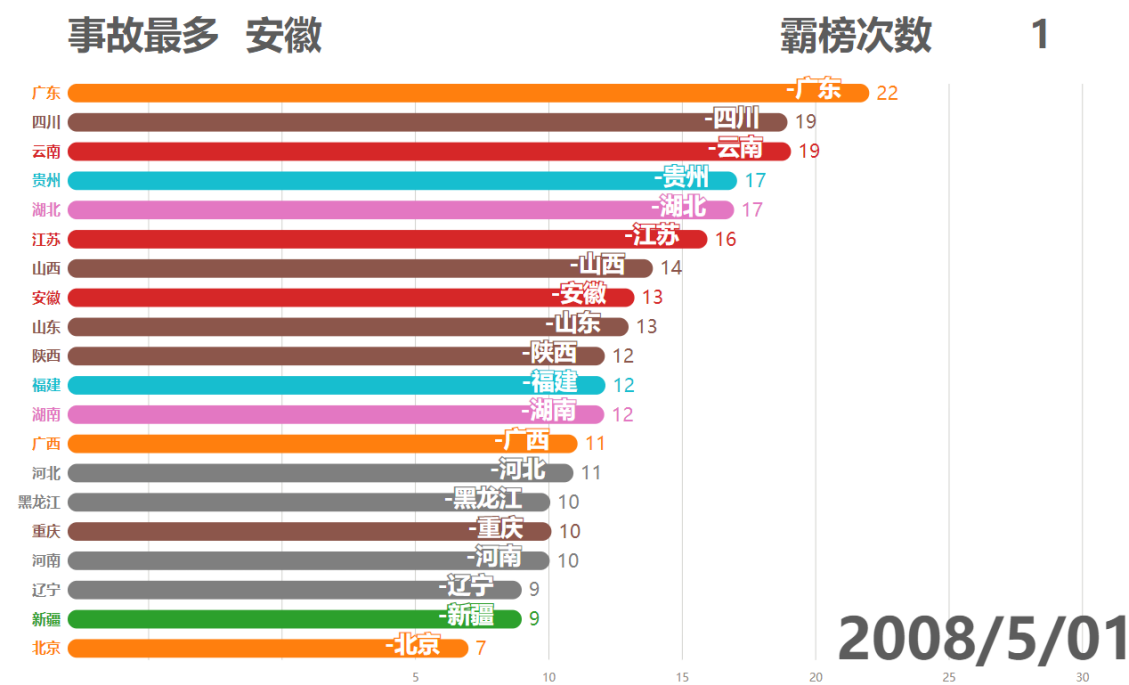


图 61 按照事故统计的结果演示

若按照事故汇总来进行动态分析，执行后，系统会自动生成 history\_rank\_acc.csv 文件保存本地，并将数据加载到软件当中（如图 62）。

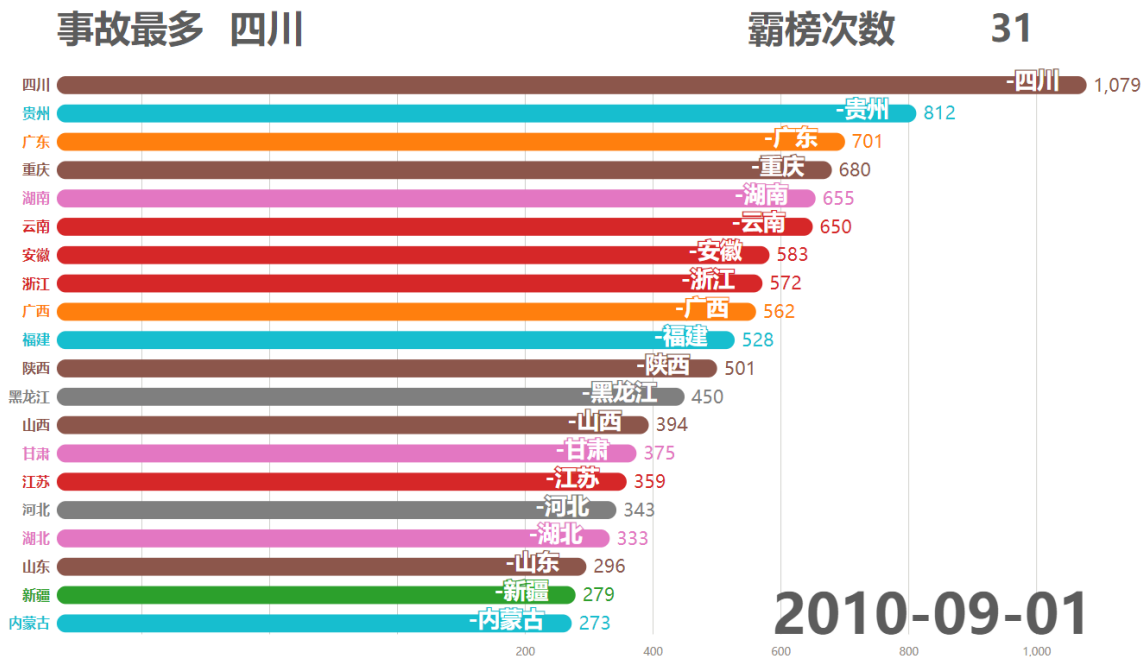


图 62 按照事故汇总的结果演示

## 第六章 项目小结

### 6.1 系统完成功能

#### (1) 登录

系统登录初始账号密码皆为admin，用户可根据需求注册账号。

#### (2) 数据加载

若用户需要对本地已有的Excel文件进行数据分析，可选用“加载本地数据”选项，若用户需要连接系统后台MySQL数据库，可直接点击“加载数据库文件”。

#### (3) 功能选项

该模块包含事故时间序列分析、事故类型分析、事故省份分析、增长率分析、事故与工作日或休息日关联度分析、事故数量历史动态变化分析。其中，每个选项都可进行不同类型的统计与分析。

#### (4) 数据查看、数据筛选

用户加载文件后，系统界面的“数据查看”模块会立即显示文件中的所有数据。若用户直接分析文件数据，可以跳过数据筛选，直接进行绘图参数设置。若用户需要对文件中的数据进行二次加工，可在“请输入筛选条件”中输入筛选要求，点击“确定”后，会直接显示已筛选的内容，并且可加以保存至本地。

#### (5) 绘图参数设置

该模块包含绘图类型、提取数据条数、绘制颜色、图形标记、指定x或y轴数据、输入x或y轴标题和原生数据与筛选数据的选择。

#### (6) 数据绘图

用户根据绘图参数的具体设置，结合功能选项，最后所呈现的数据图各不相同。其中，事故数量历史动态变化分析的数据图暂时不可在系统内界面展示，因此会在网页中显示。

### 6.2 系统待优化功能

#### (1) 绘图参数设置

目前绘图参数的设置需要用户手动输入，无法根据不同的功能选项自动生成默

认数据图，给用户造成一定程度上的不便。

#### (2) 数据可视化

本系统界面目前只能显示静态图，部分动态展示数据变化的图，无法直接内嵌到软件中，需要采用网页直接加载数据的方式，进行可视化展示。

#### (3) 系统界面

本系统目前处于初级开发阶段，在系统界面、功能布局上较为简单，可结合大众审美设计更美观的系统界面。

#### (4) 保密性

鉴于各企业数据需要保密，后续可以考虑绑定用户身份或添加其他密保辅助系统以增加系统的安全性和保密性。

### 6.3 展望

本系统开发的初心是为企业的安全生产保驾护航，紧跟安全生产专项整治三年行动路线，减少生产事故发生、员工伤亡和经济损失。

对企业安全管理部门来说，本系统可以有效直观地展示安全生产事故数据的全方位分析，推动企业的安全管理工作。希望企业的安全管理部门可以借用本系统，对已发生的事故进行总结归纳，针对事故原因或安全隐患采取积极有效的措施，减少事故发生的概率，提高安全生产的可靠性。

对安全专业的师生来说，本系统可供师生在实验实践中对安全生产事故的统计分析理论操作有更加深入的理解和感悟。希望该系统可以运用在事故分析相关的课程学习中，让学生能够探索总结真实发生的事故，而不是局限于仿真模拟的实验数据中。

## 参考文献

- [1] 叶伟林. 事故统计体系改革进行时[J]. 劳动保护, 2016(09):42-44.
- [2] 武耀东, 袁凤英. 企业安全事故分析及其对策[J]. 科技情报开发与经济, 2007(11):254-255.
- [3] 康计良. Python 语言的可视化编程环境的设计与实现[D]. 西安电子科技大学, 2012.
- [4] 张楠. Python 语言及其应用领域研究[J]. 科技创新导报, 2019, 16(17):122-123.