# 

# Python数据分析

# 详细设计文档

版本3.0

IMG_256

江苏润和软件股份有限公司

南京润云科技有限公司

2019年7月

履历

|  |  |  |  |
| --- | --- | --- | --- |
| 版本 | 更新日期 | 更新描述 | 负责人 |
| 1.0 | 2019-07-09 | 初稿 | 郑能锦、蒋安全、陈国华  晁圣博、赵韩兵、黄优 |
| 2.0 | 2019-7-10 | 修改了内容等问题，详情见第一次评审报告 | 郑能锦、蒋安全、陈国华  晁圣博、赵韩兵、黄优 |
| 3.0 | 2019-7-15 | 修正了部分格式 | 郑能锦、蒋安全、陈国华  晁圣博、赵韩兵、黄优 |
| 4.0 | 2019-7-22 | 删除了数据库模块内容，重新绘制说明图，修正了目录内容，添加了图片表格标号 | 郑能锦、蒋安全、陈国华  晁圣博、赵韩兵、黄优 |

目 录

[1.引言 1](#_Toc4502)

[1.1编写目的 1](#_Toc32148)

[1.2开发背景 1](#_Toc13614)

[1.3项目概述 1](#_Toc28240)

[2.总体设计 1](#_Toc27399)

[2.1需求概述 1](#_Toc25790)

[2.2运行环境 2](#_Toc12151)

[2.3基本设计概念和处理流程 2](#_Toc20832)

[2.4结构设计 3](#_Toc10330)

[2.5功能概述 3](#_Toc25727)

[3.接口设计 4](#_Toc12623)

[4.模块设计 4](#_Toc26036)

[4.1用例图 4](#_Toc13960)

[4.2功能模块设计 5](#_Toc12851)

[5.系统数据结构设计 5](#_Toc9594)

[6.UI设计 6](#_Toc12778)

[6.1用户登陆界面设计 6](#_Toc15318)

[6.2主要界面设计 6](#_Toc18671)

[附录 7](#_Toc17085)

[参考资料 7](#_Toc19700)

## 

## **1.引言**

### 1.1编写目的

本项目是对部门的人员信息进行收集、分析处理以及图形化展示的一个网站。在企业中，通过分析企业内的大量人员数据信息，形成一个可供参考的判断标准，可以使管理层更容易的按照标准的管理方法进行决策。而其中数据是必不可少的，大量的数据能让决策更加规范。因此我们做数据分析，而更多的是为了客户根据我们分析出来的结果对他们的需求做出判断与预测。

### 1.2开发背景

大数据开启了一次重大的时代转型，它的核心就是挖掘出庞大的数据库独有的价值。随着数据分析处理技术的发展，数据的获取越来越容易，但统计学作为从数据中读取信息的科学，应该永远和获取数据的学问相伴相生。大数据时代的数据量巨大，我们需要引入统计学思想，更好地进行数据挖掘和数据分析任务。

大数据的科学价值和社会价值正体现在这里。数据分析可以帮助决策者相处更加合理的决策并制定可以影响商业绩效的策略和计划。在大数据时代，需要挖掘数据的隐藏价值，对数据加强分析。

### 1.3项目概述

本项目通过批量提取规格化的员工数据文件，使用Python的Pandas库方法，对数据进行预处理分析，通过Echarts工具将分析的结果以图表的方式动态展示在前端网页。同时在前端提供了获取存储相同类型数据的文件接口。

## **2.总体设计**

### 2.1需求概述

* 业务需求:对企业内部员工信息进行分析处理，方便对人员分布、构成、比例等进行管理。
* 输入形式:信息数据表，Excel表格
* 输出形式:图形化动态展示，以图表的形式（柱状图、饼图、折线图等）
* 主要功能:实现了对员工信息按分布、构成、比例等方面的统计

### 2.2运行环境

* 操作系统: Windows 7/10
* 开发语言: Python / Css / JavaScript
* 开发工具: PyCharm
* 文档编写工具: WPS / NotePad++
* 版本管理工具: SVN / Git

### 2.3基本设计概念和处理流程

数据分析设计六部曲：明确分析目的和思路---数据收集---数据处理---数据分析---数据展现---报告撰写。

1. 明确分析目的和确定思路：明确分析目的首先要思考为什么要开展数据分析？通过这次数据分析要解决什么问题？分析目的确定之后梳理分析思路，搭建分析框架。把分析目的分解成若干个不同的分析要点，即如何具体开展数据分析、需要从哪几个角度进行分析，采用哪些分析指标。
2. 数据收集：按照确定的数据分析框架，收集相关数据的过程。这儿的数据分为第一手数据为可直接获取的数据，第二手数据为加工整理后得到的数据。数据来源包括数据库、公开出版物，互联网和市场调查。
3. 数据处理：对收集到的数据进行加工处理，形成适合数据分析的样式。基本目的是从大量的，杂乱无章，难以理解的数据中，提取并推导出对解决问题有价值、有意义的数据。包括数据清洗，数据转换、数据提取和数据计算。
4. 数据分析：使用适当的数据分析方法和工具，对处理过的数据进行分析，提取有价值的信息，形成有效结论的过程。数据分析与数据处理的区别：数据处理与数据分析的基础，如果数据本身存在错误，那么即使采用最先进的数据分析方法，得到的结果也是错误的，不具备任何参考价值，甚至还会误导决策。数据分析与数据挖掘的关系：数据挖掘是一种高级的数据分析方法。数据挖掘侧重解决四类数据分析问题：分类、聚类、关联和预测，重点在寻找模式和规律。
5. 数据展现：数据通过表格和图形的方式来呈现。常见的数据图表包括饼图、柱形图、条形图、折线图、散点图和雷达图等，也可对这些图表进一步加工整理，变为我们需要的图形，金字塔图，矩阵图、漏斗图，帕累托图等。
6. 报告撰写：数据分析报告是对整个数据分析过程的一个总结和呈现。通过报告把数据分析的起因、经过、结果和建议完整的表现出来，供决策者参考。

### 2.4结构设计

本项目是一个Python数据分析项目，是利用Python作为编程语言，通过使用一些主流的数据分析库，对客户提供的数据文件进行批量导入并读取信息，然后进行数据清洗，即对符合客户需求的信息进行整理，对需求的内容进行数据数据提取，以表格的形式整理，并在前端Web页面通过图表的形式对不同的信息进行展示，其中包括员工构成、工作环境分布、学历、年龄、性别等，如图2-1所示

python数据分析

员工构成

已婚比例

性别比例

年龄分布

学历分布

司龄分布

图2-1

2.5功能概述

### **2.5.1快速导入数据**

可以通过导入本地文件接口将Excel信息表中的信息快速的导入信息库，并且支持同时导入多个Excel表或者陆续导入多个Excel信息表；系统会将多个信息表中的数据进行对比分析，去除重复数据，合并不同数据。

### **2.5.2 可视化**

系统对导入进来的人员信息分类统计处理后，用Pandas进行数据分析处理，然后会将处理结果通过Matplotlib绘制成柱状图、饼状图等图表，用Echarts显示在Web前端界面。

同时，当用户信息数据发生变化时，图表信息也将随之变化。

## **接口设计**

前端通过web上传Excel文件到后端，后端分析处理后返回json数据给前端展示

## **4.模块设计**

### 4.1用例图

本项目为数据分析项目，客户可以通过批量导入员工资料信息数据，经过后台对数据的处理，生成按需分类的图表，在网页端进行可视化展示，并且客户可以通过对信息的增删，而实现图表的实时更新，同时。客户还可以查看用户信息列表及详情。用例图如图4-1所示。

管理员

登录

数据统计结果查看

图4-1

### 4.2功能模块设计

### **4.2.1快速导入数据**

可以通过导入本地文件接口将Excel信息表中的信息快速的导入信息库，并且支持同时导入多个Excel表或者陆续导入多个Excel信息表；系统会将多个信息表中的数据进行对比分析，去除重复数据，合并不同数据。

### **4.2.2 数据分析可视化处理**

系统对人员信息进行分析、分类统计处理后，会将处理结果绘制成柱状图、饼状图等图表，然后显示出来。

## **5.系统数据结构设计**

用户打开首页，页面初始化的时候发送ajax请求到服务器，服务器从数据库查询数据并对数据进行分析，把结果以图表的形式返回首页，具体过程如图5-1所示。

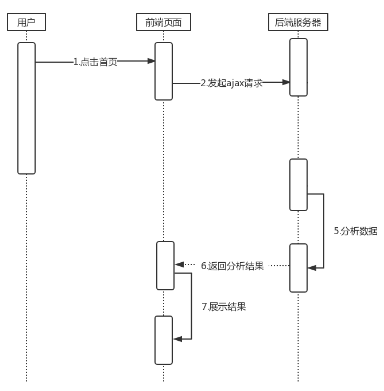


图5-1

## **7.UI设计**

### 7.1用户登陆界面设计



图7-1

### 7.2主要界面设计

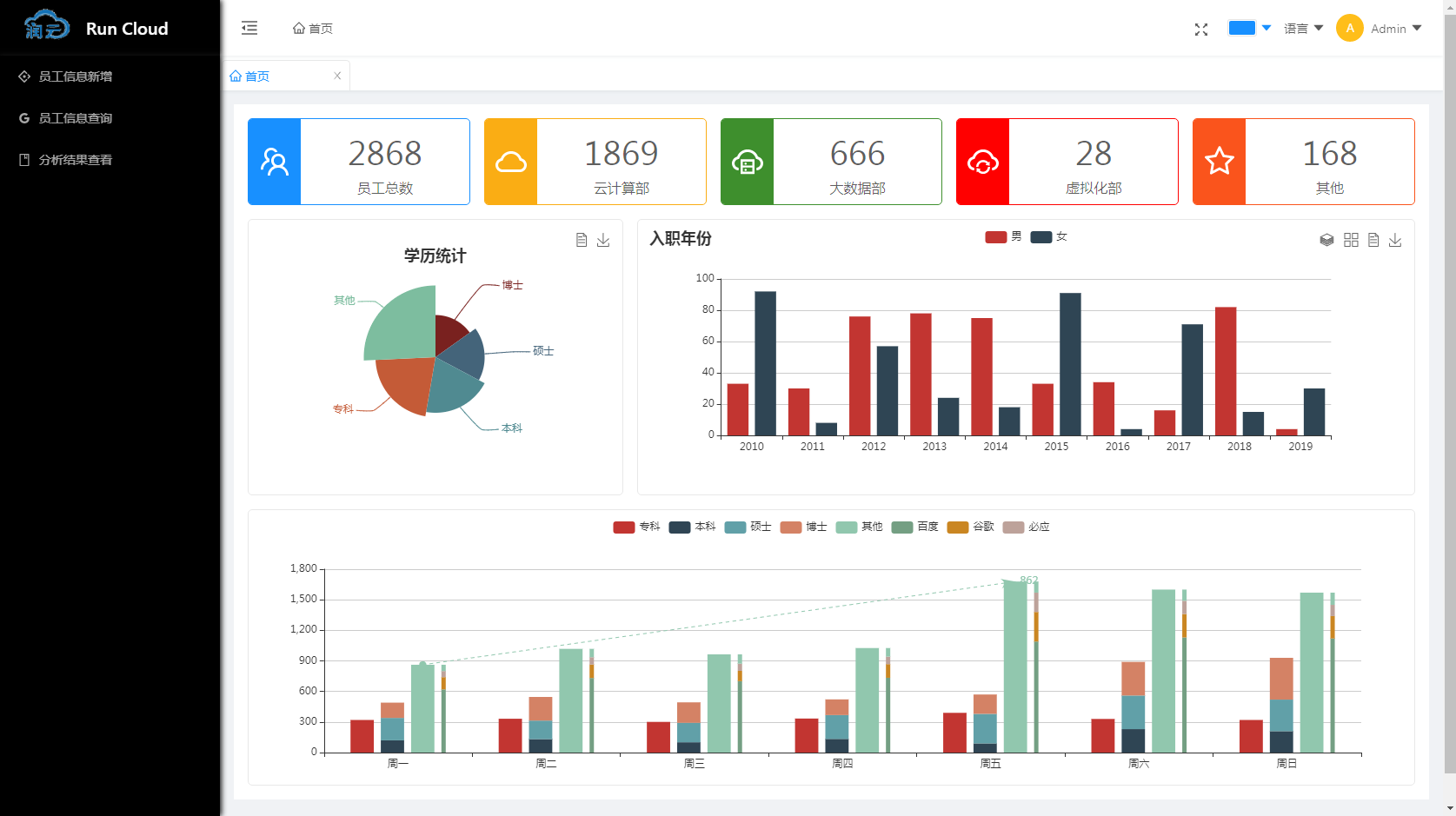


图7-2

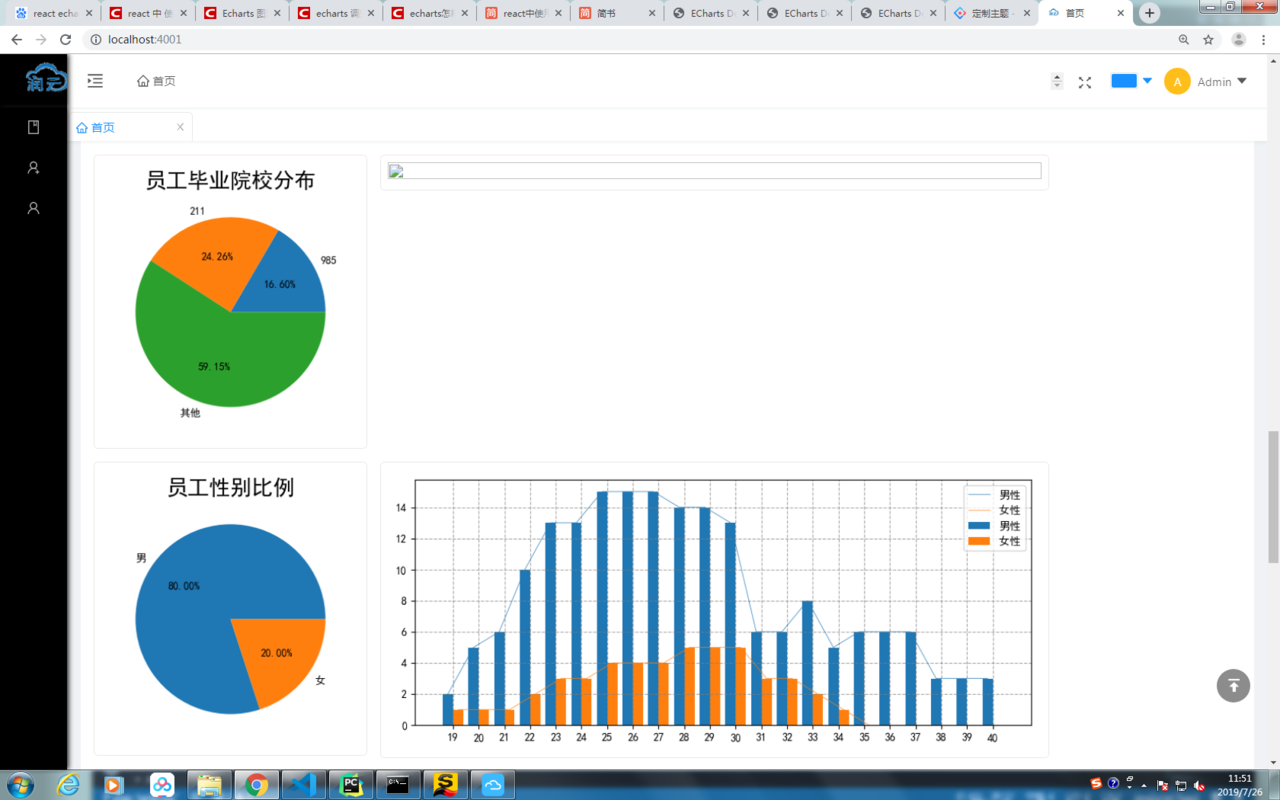


图7-3

## **附录**

### 参考资料

《Python编程从入门到实践》----Eric Matthes 著

《利用Python进行数据分析》----Wes McKinney 著