****

**课 程 设 计 报 告**

**课程名称： 疫情大数据分析**

**学 院： 信息工程学院**

**专业班级： 大数据1801班**

**学 号： 180808011022**

**姓 名： 刘锐**

**起止日期: 7月3日-- 7月 12日**

**成 绩:**

**指导老师：**

**2020 年 07 月 12 日**

**西京学院教务处制**

目录

[一、序言 1](#_Toc45027332)

[1.1课程背景 1](#_Toc45027333)

[1.2课程意义 1](#_Toc45027334)

[1.3预期效果 1](#_Toc45027335)

[二、实训内容 2](#_Toc45027336)

[2.1实训内容一 2](#_Toc45027337)

[2.2实训内容二 3](#_Toc45027338)

[2.3实训内容三 4](#_Toc45027339)

[2.4实训内容四 5](#_Toc45027340)

[2.5实训内容五 6](#_Toc45027341)

[2.6实训内容六 7](#_Toc45027342)

[2.7实训内容七 8](#_Toc45027343)

[2.8实训内容八 9](#_Toc45027344)

[2.9实训内容九 10](#_Toc45027345)

[2.10实训内容十 11](#_Toc45027346)

[课程总结与展望 12](#_Toc45027347)

# 一、序言

## 1.1课程背景

100多年前的1918年，西班牙流感席卷全球，造成人类至今难以忘却的伤痛。100多年后的今天，新冠肺炎成为人类生存的又一次挑战。不同的是，大数据与人工智能（AI）等技术手段正在成为人类战胜疫情的新型“武器”。

疫情之下，大数据与AI技术能够为疫情的预防控制做些什么？4月18日，在爱思唯尔云论坛上，多位来自学术界和工业界的专家，分享了此次新冠疫情中大数据与AI的相关研究与应用。

作为大学生的我们新冠肺炎疫情席卷全球之际，我们旨在使用大数据技术和人工智能理念来解析疫情的发展动态，为科学防疫提供决策支持.

## 1.2课程意义

1、通过此次实习实训，我们要将一个疫情分析的案例所用到的工具模块衔接起来。

2、让学生了解到我们之前学的爬虫，数据分析，可视化技术等再实际应用中的表现。

## 1.3预期效果

1、学员对疫情数据的采集和爬虫的概念及过程形成清晰的认识；

2、掌握获取数据后，对数据的分析和清晰等流程；

3、掌握数据可视化的技能技巧。

# 二、实训内容

## 2.1实训内容一

内容：今天是实训的第一天，老师带着我们大概了解了一下爬虫的知识，带着我们使用Python爬取了一些网易实时疫情的数据。

使用Python库：

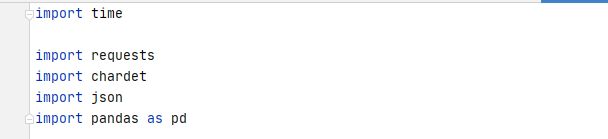


图1-1 导入第三方库

使用函数：

**1.请求目标url的信息，然后设置编码**

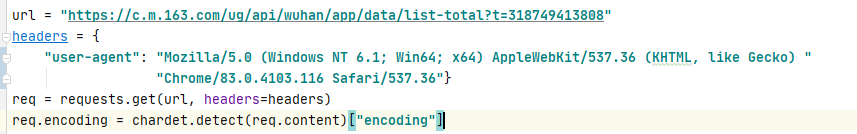


图1-2 编写url和header信息

**2.将收到的信息转换为json格式**



图1-3 格式转换

**3.pandas 合并表格**

****

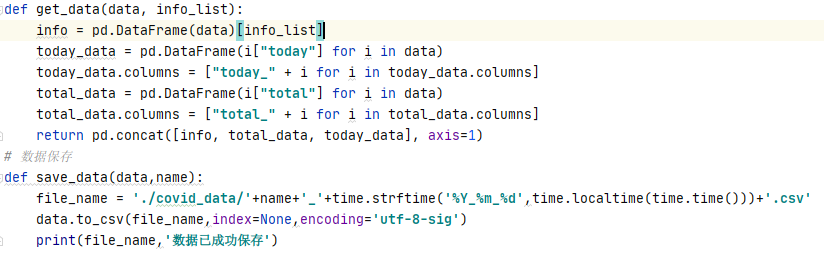
图1-4 表格合并

**4.pandas保存数据为csv文件**



图1-5文件保存为csv格式的文件文件

**5.函数封装**（将今天所写得爬虫代码封装好，以便以后使用）

图1-6函数封装代码

## 2.2实训内容二

内容：今天是实训的第二天，今天我们进行了全国实时数据和历史数据的爬取、各省历史数据爬取。

1.全国实时数据爬取

①在网易实时疫情网站里面查找对应的url

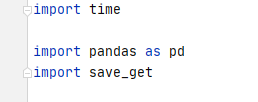


图2-1 导入第三方库

②设置url

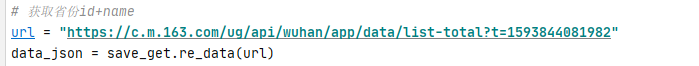


图2-2 设置url爬取数据

③用第一天封装好的函数获取返回的json文件

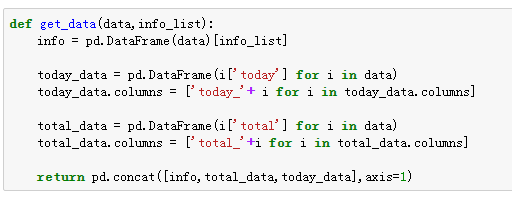
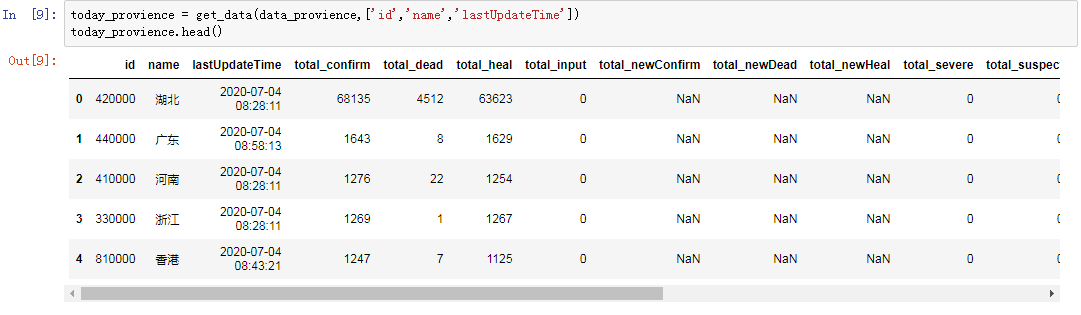


图2-3、2-4 封装的爬虫函数和全国实时数据

④保存文件

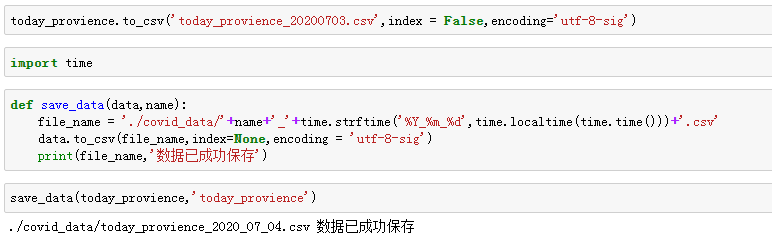


图2-5 文件保存

2.全国历史数据爬取

①导入数据库

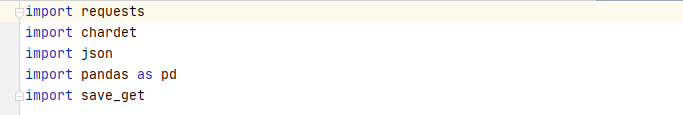


图2-6 导入第三方库

②设置url,配置反反爬

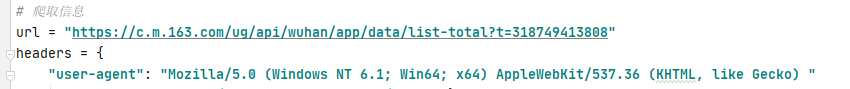


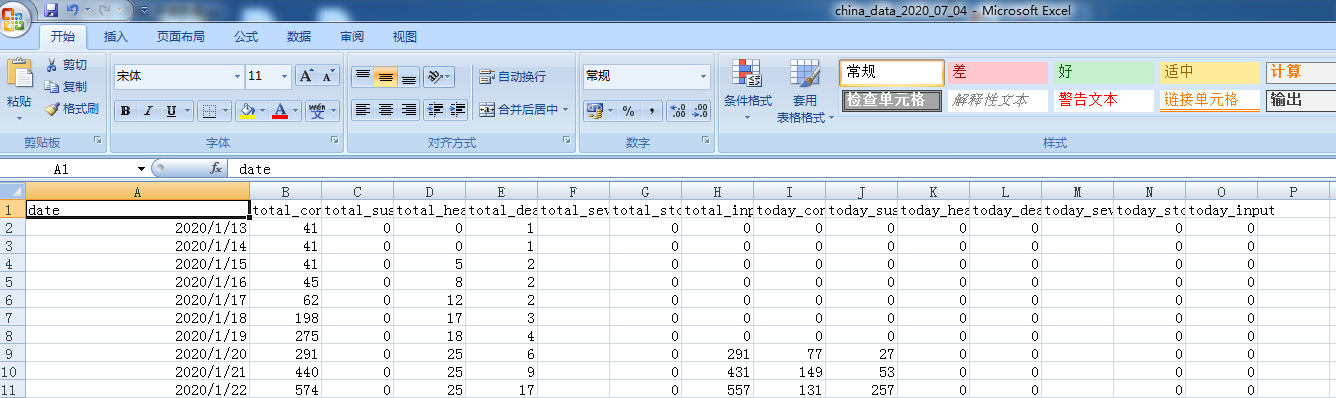
图2-7 设置url,配置反反爬

③数据爬取处理



图2-8 数据爬取

④插看爬到的数据

图2-9 全国历史数据

2.各省历史数据爬取

与前两个相似，数据如下：

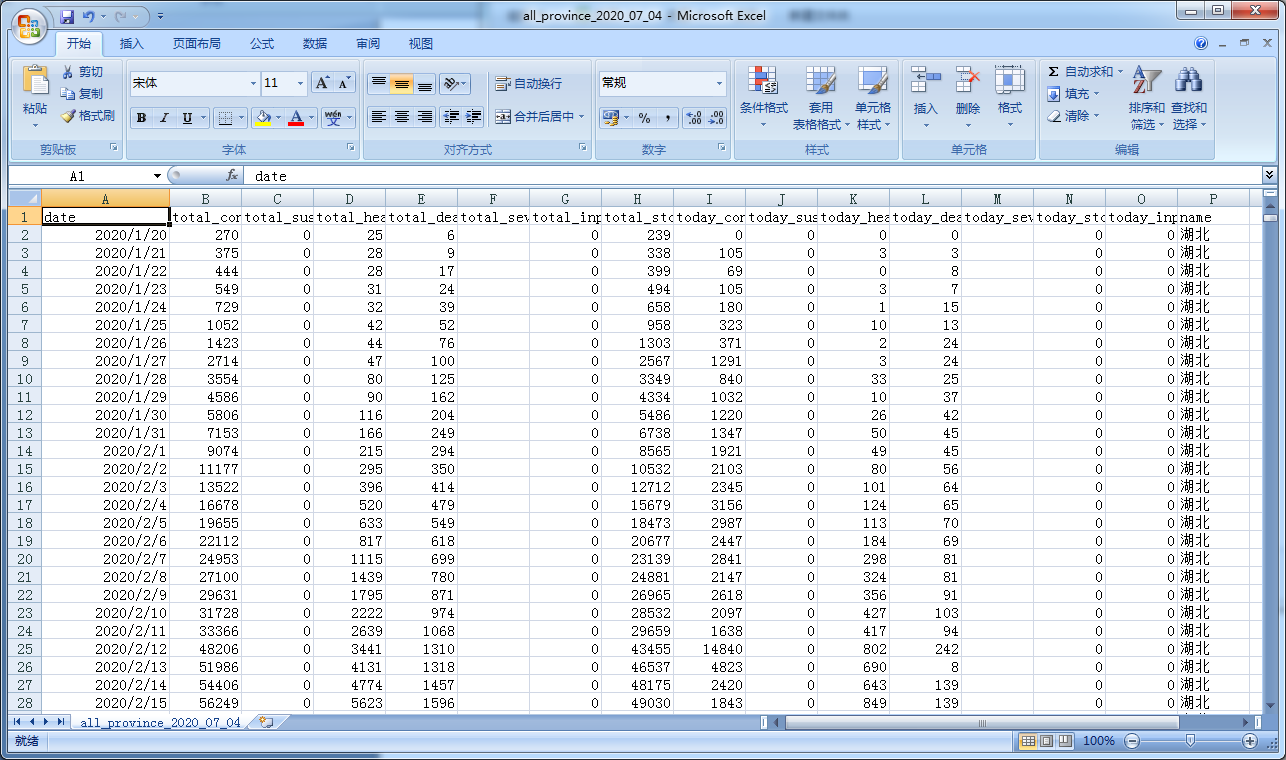


图2-10 各省历史数据

## 2.3实训内容三

内容：今天是实训的第三天，今天学习了用Pandas对疫情探索性分析性分析。

使用库：

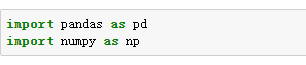


图3-1 导入第三方库

使用函数：

1.读取csv文件

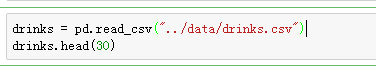


图 3-2 导入csv文件

2. 查看数据集中有多少列



图3-3 查看数据集

3.查看数据集的列名

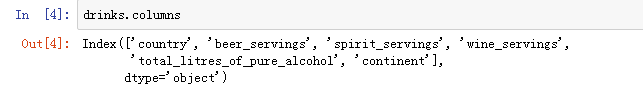


图3-4 查看数据集列名

4.查看索引



图3-5 查看索引

5.分组聚合

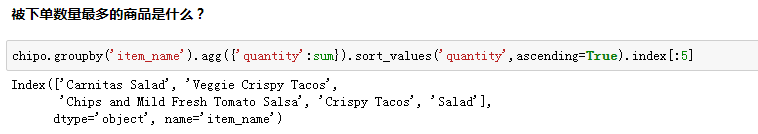


图3-6 分组聚合

6.去重

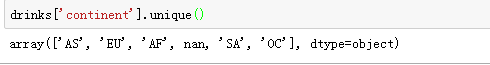


图3-7 去重查看

7.数据类型转换

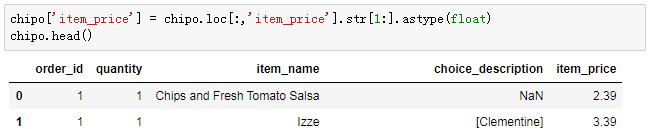


图3-8 类型转换

## 2.4实训内容四

内容：今天是实训的第四天，今天对之前学习的知识进行了复习，除此之外，仅仅是分析还是不够的，我们还简单学习了用matplotlib绘图，在绘图前，需要对数据进行一定的处理

1.数据处理

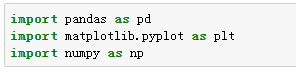


图4-1导入第三方库

读取之前爬取的数据



图4-2 导入数据

对读取的表格进行简单的查看

①使用head(5)查看前5行数据

②使用pd.info()查看列数据类型

③写一个列名字典，更改列名(pd.rename(columns=name\_dict,inplace=True))

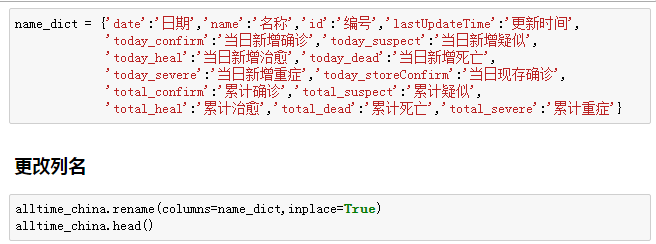


图4-3 更改列名

④对表格中的数据进行计算，不需要的列删除



图4-4 删除列

⑤更改列数据类型

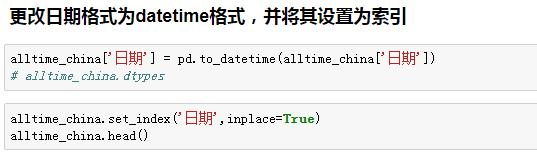


图4-5 类型转换

2.绘图

①导入需要的库以及设置字体

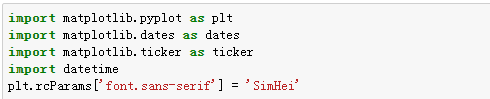


图4-6 导入第三方库

②设置画板和子图



图4-7 设置画板和字图

③绘图



图4-8 使用数据绘图

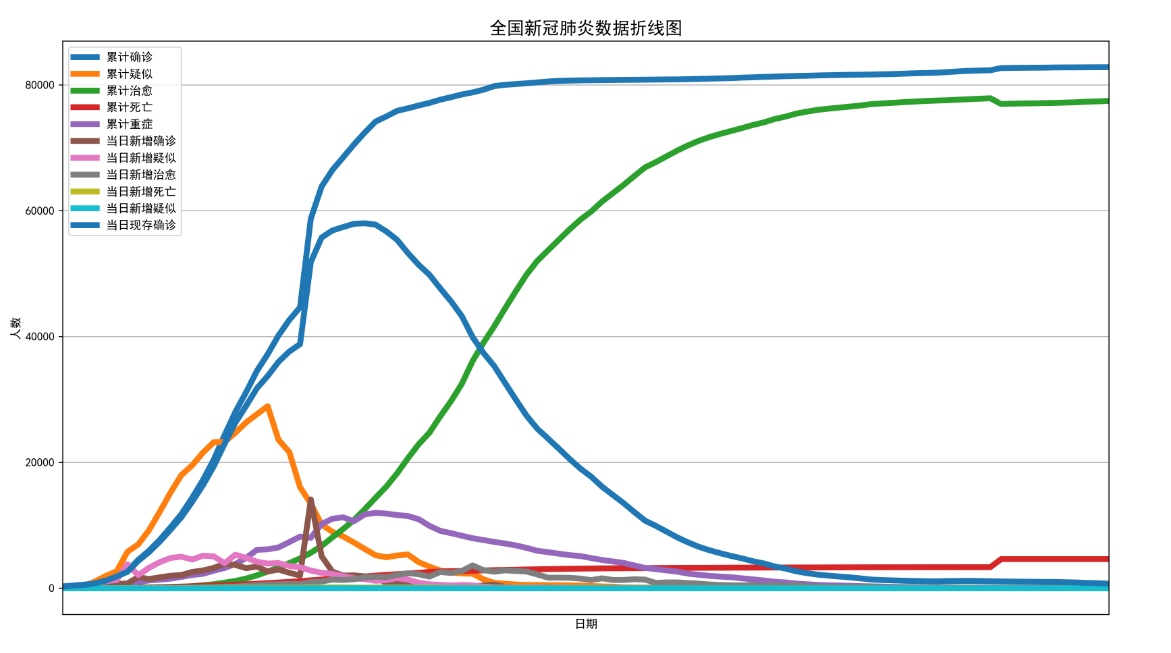


图4-9 全国新冠肺炎数据折线图

## 2.5实训内容五

内容：今天是实训的第五天，今天我们要用之前所学得只是爬取世界历史数据然后用这些数据绘制图。

1.导入库

save\_get:之前封装的函数



图5-1 导入库

2.爬取数据



图5-2 设置url,爬取数据



图5-3 数据处理

3.查看爬取的数据

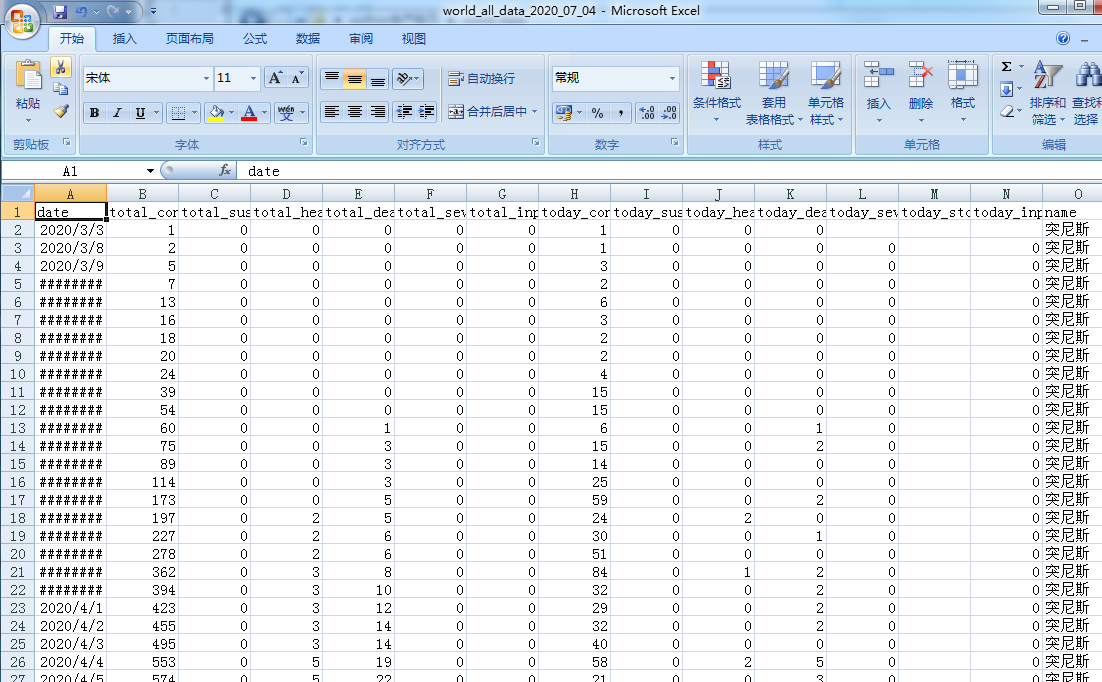


图5-4 查看数据

4.绘制图形

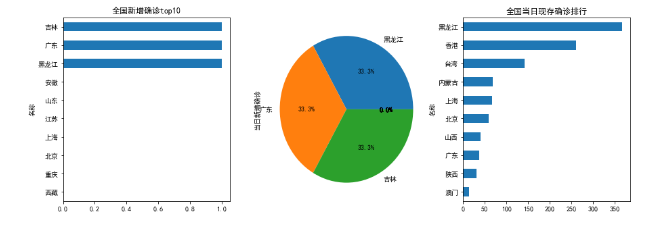


图5-5简单的绘图

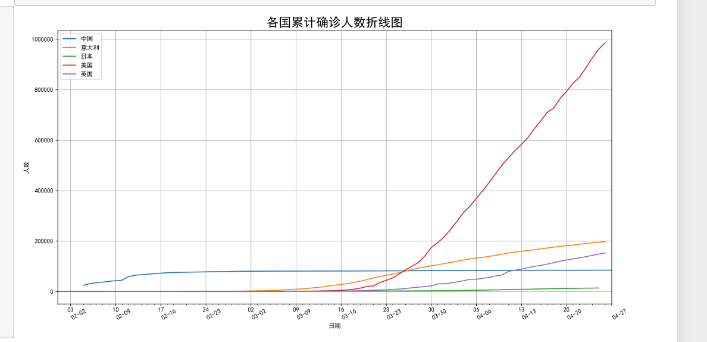


图5-6 各国累计确诊人数折线图

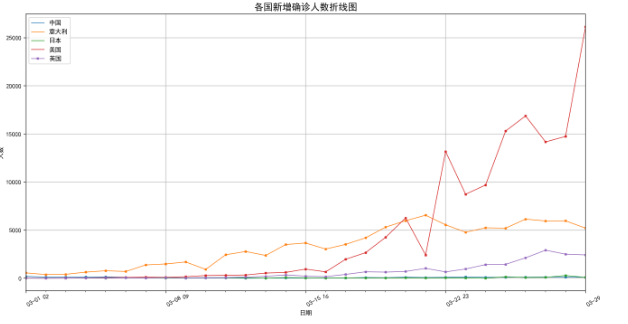


图5-7 各国新增确诊人数折线图

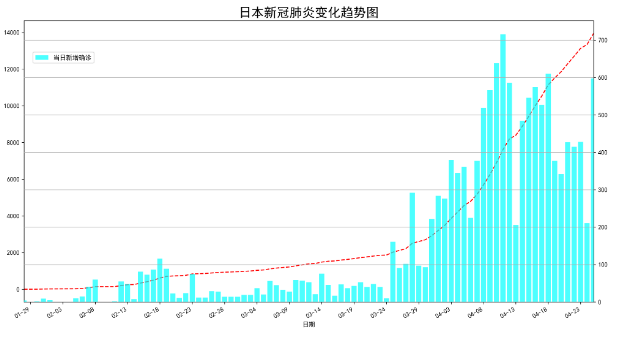


图5-8 日本新冠肺炎变化趋势图

## 2.6实训内容六

内容：今天是实训的第六天，今天学习了使用Python的pyechart库来绘图，这个库有很多好看的主题，就是各个版本之间不兼容，对萌新不太友好。

1.导入库

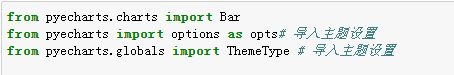


图6-1 导入库

## 2.7实训内容七

目标：

内容：

## 2.8实训内容八

目标：

内容：

## 2.9实训内容九

目标：

内容：

## 2.10实训内容十

目标：

内容：

# 课程总结与展望

|  |
| --- |
| **总结与展望：**  **学生签名：**  **日期： 年 月 日** |