**《Python玩转大数据》实验报告（第 次实验）**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **学 号** |  | | | | | | | **姓名** | | | **NaN** | | | | | | | |
| **专业班级** |  | | | | | | | **日期** | | | **NaN** | | | | | | | |
| **邮箱** |  | | | | | | | | | | | | | | | | | |
| **实验名称： 第一次编程** | | | | | | | | | | | | | | | | | | |
| **实验目的：** | | | | | | | | | | | | | | | | | | |
| **实验内容：见《第一次作业》文档** | | | | | | | | | | | | | | | | | | |
| **实验分析：**   1. **实现绘制各国碳排放量和GDP之间的关系图: 分别进行如下操作:去除空数据:dropna(),填补空白:fillna(),删除’Country Code’为NaN的行. 之后按如下方式获取数据:li = data[data[‘Country Cod’].isin(Country Array)].li = li.values.tolist().这种方式可以在一定程度上降低程序的耦合性,每次获取指定国家只需要在数组之间写入简称即可.之后,将GDP数据调整为以百万美元为单位,最后调用matplotlib绘制图像** 2. **主要使用bs4和requeests两个库.学校官网相对来说比较好爬,先是设置好请求头,之后传入指定的URL,由于需要100条数据,这里使用了格式化字符串来创建多个新闻页面的URL,之后将response.content转化为BeautifulSoup对象,搜索class\_为’news\_title’的span元素,在之后使用正则表达式获取标题.简简单单.** 3. **正则表达式分别匹配身份证号和手机号.简简单单.** | | | | | | | | | | | | | | | | | | |
| **实验结果（结果截屏，只需要运行结果窗口）** | | | | | | | | | | | | | | | | | | |
| **第1题结果** | |  | | | | | | | | | | | | | | | | |
| **第2题结果** | |  | | | | | | | | | | | | | | | | |
| **第3题结果** | |  | | | | | | | | | | | | | | | | |
| **第4题结果** | |  | | | | | | | | | | | | | | | | |
| **第5题结果** | |  | | | | | | | | | | | | | | | | |
| **第6题结果** | |  | | | | | | | | | | | | | | | | |
| **第7题结果** | |  | | | | | | | | | | | | | | | | |
| **第8题结果** | |  | | | | | | | | | | | | | | | | |
| **第9题结果** | |  | | | | | | | | | | | | | | | | |
| **第10题结果** | |  | | | | | | | | | | | | | | | | |
| **第11题结果** | |  | | | | | | | | | | | | | | | | |
| **第12题结果** | |  | | | | | | | | | | | | | | | | |
| **第13题结果** | |  | | | | | | | | | | | | | | | | |
| **第14题结果** | |  | | | | | | | | | | | | | | | | |
| **第15题结果** | |  | | | | | | | | | | | | | | | | |
| **以下单元格请勿删、增、修改！** | | | | | | | | | | | | | | | | | | |
|  | |  | | | | | | | | | | | | | | | | |
|  | |  | | | | | | | | | | | | | | | | |
|  | |  | | | | | | | | | | | | | | | | |
| **题 号** | | **1** | **2** | **3** | **4** | **5** | **6** | | **7** | **8** | | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
|  | |  |  |  |  |  |  | |  |  | |  |  |  |  |  |  |  |
|  | |  |  |  |  |  |  | |  |  | |  |  |  |  |  |  |  |
|  | |  | | | | | | | | | | | | | | | | |
|  | |  | | | | | | | | | | | | | | | | |

**由于程序自动判定，请仔细阅读以下说明！**

说明：

⒈请勿修改表格结构，即不得增加或删除任何行、列及单元格。每个单元格尺寸可以改变。请正确填写电子邮箱地址。

⒉实验报告命名为”学号实验序号.docx”，如202221623501.docx。2022216235为学号，01两位实验序号（即第1次实验），请勿在学号和序号之间添加任何字符。

⒊表格必须保存为.docx格式。

⒋不符合以上任意一条规则将影响数据采集，实验报告得分为0。