**上海中学2022学年第一学期高一年级期末练习**

**信息技术**

**说明：**

**1.本次练习时间60分钟，总计100分。**

**2.可以使用计算器。**

**3.答题前，请先在答题纸和答题卡上正确填写班级、姓名、学号。**

**综合题一**

1. 让绿色生活成为新时尚，上海于2019年7月1日开始推动垃圾分类工作。为了提高垃圾回收的效率，在有些小区里安装了具有“用户识别”“自动称重”“满溢提醒”等功能的智能垃圾回收机，如下图所示。居民投递的可回收物送到分拣工厂后经机器人精细分类，最终进入相应的再生产工厂变废为宝。



（1）智能垃圾回收机放置在小区公共位置，屏幕上滚动显示可回收物类别等内容，以便广大居民了解相关信息。这主要体现信息特征是\_\_\_\_\_\_\_\_\_\_\_（选填：价值是相对的/具有时效性/具有共享性）。

（2）在小区没有安装智能垃圾回收机之前，我们在投放垃圾时需要自己分好类：废报纸属于可回收垃圾、坏的灯泡属于有害垃圾、过期的食物属于厨余垃圾……，这些描述的内容属于\_\_\_\_\_\_\_\_\_\_\_（数据/信息/知识）

（3）这款智能垃圾回收机有语音交互功能模块，方便用户进行功能选择、报修、意见收集等任务的实现。语音交互功能模块在工作时借助于内置的声2023.1线上音传感器，将采集到的\_\_\_\_\_\_\_\_\_\_\_信号转换为\_\_\_\_\_\_\_\_\_\_\_信号，即通过\_\_\_\_\_\_\_\_\_\_\_、\_\_\_\_\_\_\_\_\_\_\_、编码的过程，完成转换后才能让数字设备进行存储和处理（ ）。

A．数字 模拟 采样 量化 B．模拟 数字 量化 采样

C．模拟 数字 采样 量化 D．数字 模拟 量化 采样

（4）机器内置的SD卡用于存储收集的语音反馈和留言。若已知SD卡的剩余容量为64G，每个用户允许最多录制2分钟的语音，声音的格式如下图，则最多还能支持\_\_\_\_\_\_\_\_\_\_\_名用户的录制。（每人录满2分钟）



（5）为了鼓励更多人使用智能垃圾回收机进行垃圾分类和回收，该机器可以对每次合理的分类投放给予一定的积分奖励。具体做法是：投放前可扫脸或输入手机号及密码进行登录。从数据安全防护的角度来看，输入手机号及密码的方式才能登录，是采用了\_\_\_\_\_\_\_\_\_\_\_方法进行了用户数据的保护；登录成功后个人的界面（如下图）则采用了\_\_\_\_\_\_\_\_\_\_\_方法进行了用户的数据安全防护。



（6）下图所示的是一张可回收物图片及其属性信息，一张这样的图片需要占用\_\_\_\_\_\_\_\_\_\_\_MB空间；



（7）若对图片（上图）进行压缩，以下关于压缩的描述，正确的是（ ）。

A．将此文件用WinRAR软件进行压缩属于无损压缩

B．将此文件另存为jpg格式属于无损压缩

C．将此文件重命名为玻璃瓶.zip属于有损压缩

D．将此文件扩展名改为.bmp属于无损压缩

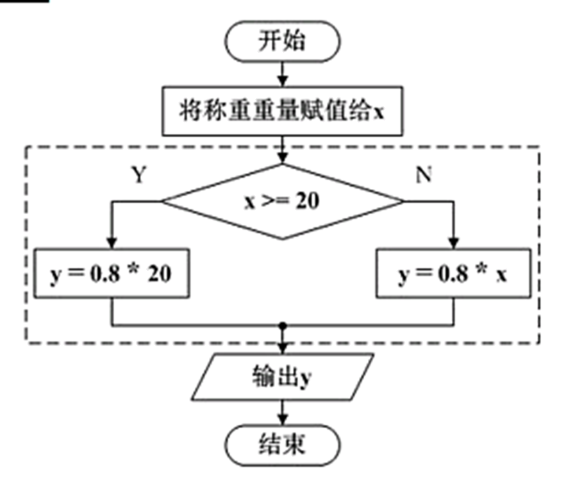
（8）如下图所示的易拉罐图片采用RGB颜色模型来描述颜色，其中某像素点的R、G、B原色用十进制数表示分别是230、58、70，将其转换为十六进制表示应为#\_\_\_\_\_\_\_\_\_\_\_。



（9）已知可回收物的当前回收价格为0.8元/公斤，单次投递超过20公斤最多按20公斤结算。若设计算法来实现某次投递所获金额的计算，算法描述如下：将称重重量赋值给x，如果x小于20，输出金额为0.8\*x元，否则输出金额为0.8\*20元，这种算法的描述方法属于（ ）。

A．程序设计语言 B．流程图 C．伪代码 D．自然语言

（10）小申用流程图设计了某次投递所获金额的算法（如下图），虚线框中算法的基本控制结构是\_\_\_\_\_\_\_\_\_\_\_，若可回收物的重量是20kg，表达式x>=20的值是\_\_\_\_\_\_\_\_\_\_\_。



【答案】 ①. 具有共享性 ②. 知识 ③. C ④. 6492 ⑤. 身份认证 ⑥. 隐蔽技术 ⑦. 0.56 ⑧. A ⑨. E63A46 ⑩. D ⑪. 分支结构 ⑫. True

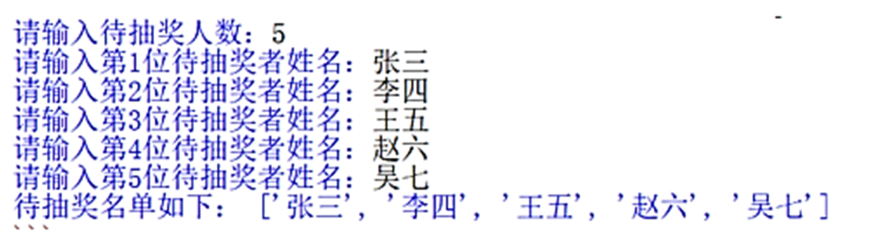
【解析】

【详解】本题主要考查信息技术的综合应用。（1）智能垃圾回收机放置在小区公共位置，屏幕上滚动显示可回收物类别等内容，以便广大居民了解相关信息。这主要体现的信息特征是具有共享性。（2）知识是人们运用大脑对所获取或积累的信息进行的提炼、研究和分析的结果，能够精确的反应事物的本质。废报纸属于可回收垃圾、坏的灯泡属于有害垃圾、过期的食物属于厨余垃圾……，这些描述的内容属于知识。（3）将采集到的模拟信号转换为数字信号，即通过采样、量化、编码的过程，完成转换后才能让数字设备进行存储和处理，故选C选项。（4）声音容量=采样频率\*量化位数\*声道数\*时长/8，根据题干，则最多还能支持用户数是64\*1024\*1024\*1024\*8/（22.050\*1000\*16\*2\*2\*60）=6492名。（5）从数据安全防护的角度来看，输入手机号及密码的方式才能登录，是采用了身份认证方法进行了用户数据的保护；登录成功后个人的界面（如下图）则采用了隐蔽技术方法进行了用户的数据安全防护，中间位用星号显示。（6）图像文件容量=分辨率\*颜色深度/8=512\*384\*24/8/1024/1024=0.56MB。（7）将此文件用WinRAR软件进行压缩属于无损压缩，其他说法均错误，故本题选A选项。（8）十进制数230、58、70分别通过“除权取余、逆序排列”的方法转换为十六进制数是E6、3A、46，故填E63A46。（9）分析题干可知，这种算法的描述方法属于自然语言，故本题选D选项。（10）由图可知，若可回收物的重量是20kg，则满足判断条件，因此表达式x>=20的值是True。

**综合题二**

2. 元旦将至，小申作为班级迎新活动的组织者，他想在活动过程中加入一个幸运抽奖环节。为了提高抽奖的效率与趣味性，抽奖功能是通过Python程序来实现的，请根据以上情境回答以下问题：

（1）抽奖前先要准备待抽奖名单，小申为了保持程序的灵活性，允许主持人输入待抽奖的同学人数，随后输入待抽奖者姓名，并用列表namelist保存待抽奖名单。若输入待抽奖人数5，准备抽奖名单程序预期运行效果如下图：

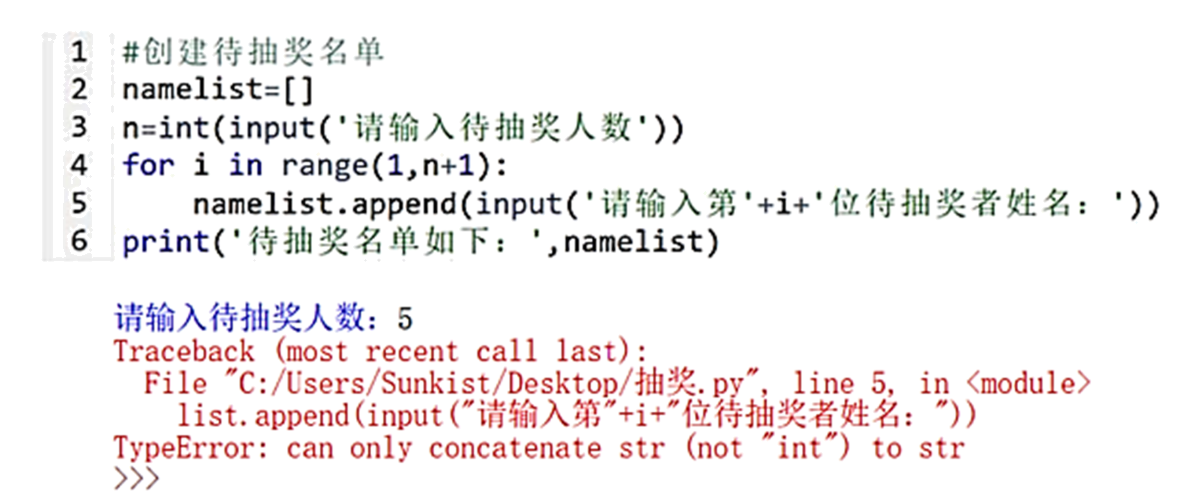


要向列表namelist中追加待抽奖者姓名，可以用的方法是（ ）。

A．namelist.insert（　　） B．namelist.input（　　）

C．namelist.add（　　） D．namelist.append（　　）

（2）编写好程序后，小申在调试时程序报错，出错信息如下：



出错语句是\_\_\_\_\_\_\_\_\_\_\_行，应修改为\_\_\_\_\_\_\_\_\_\_\_（\_\_\_\_\_\_）。

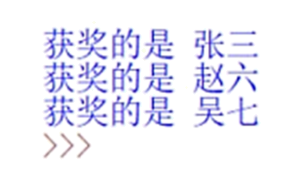
A．3 n=str（input（‘请输入待抽奖人数：’））

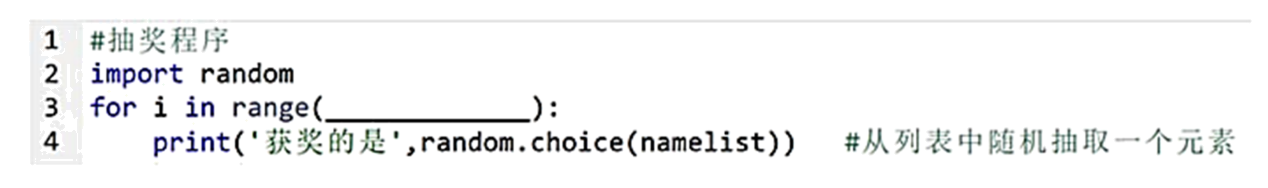
B．5 namelistappend（input（‘请输入第’+str（i）+‘位待抽奖者姓名：’））

C．6 print（‘待抽奖名单如下：’，str（namelist））

D．5 namelist.append（input（‘请输入第’+int（i）+‘位待抽奖者姓名：’））

（3）准备好了抽奖名单后小申使用random函数从待抽奖名单中抽出三名获奖者，预期效果如左下图所示，则代码中划线处应填写（ ）。

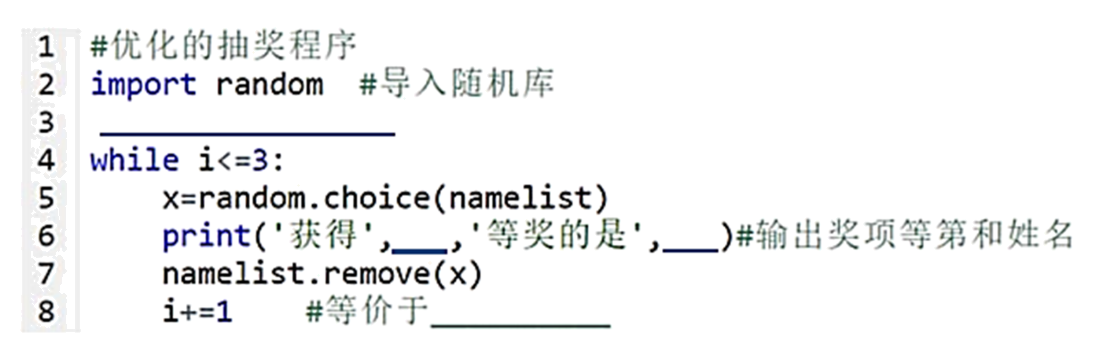




A．1，3 B．0，4 C．1，4 D．4

（4）上题中小申用到的random函数，是python中使用非常普遍的函数，它返回随机生成的一个实数。若小申想要随机生成一个[60，100）范围内的一个随机整数，表达式应写成\_\_\_\_\_\_\_\_\_\_\_

（5）小亚作为小申的助手，在小申设计的程序出问题时，他也尝试写了一个程序做备案；而且他在设计时还考虑到“中过奖的同学不能再抽”的情况，在程序上做了优化。思路是：将已经中奖的同学名字从待抽奖名单namelist中移除之后再进行下一次抽奖，预期运行效果如下图所示。



第③行空白划线处应该填写语句：\_\_\_\_\_\_\_\_\_\_\_

第⑥行的空白划线处应分别填写：\_\_\_\_\_\_\_\_\_\_\_，\_\_\_\_\_\_\_\_\_\_\_

第⑧行的注释划线处应填语句：\_\_\_\_\_\_\_\_\_\_\_

【答案】 ①. D ②. B ③. C ④. random.randint(60,99) ⑤. i=1 ⑥. str(i) ⑦. x ⑧. i=i+1

【解析】

【详解】本题主要考查Python程序的综合应用。（1）要向列表namelist中追加待抽奖者姓名，可以用的方法是append方法，即namelist.append（　　），故选D选项。（2）分析代码可知，出错的语句是第5行，原因是字符串不能和整型直接相连接，应修改为namelist.append（input（‘请输入第’+str（i）+‘位待抽奖者姓名：’）），故选B选项。（3）range(start, stop, [step])，start: 计数从 start 开始。默认是从 0 开始。例如range（5）等价于range（0， 5）；stop: 计数到 stop 结束，但不包括 stop。例如：range（0， 5） 是[0, 1, 2, 3, 4]没有5；step：步长，默认为1。例如：range（0， 5） 等价于 range(0, 5, 1)。一共循环了3次，结合选项可知，选C选项。（4）random.randint(a,b)用于生成一个指定范围内的整数。其中参数a是下限，参数b是上限，生成的随机数n: a <= n <= b。若小申想要随机生成一个[60，100）范围内的一个随机整数，表达式应写成random.randint(60,99)。（5）分析程序可知，变量i的初值是1，故第一空填i=1。第二空填str(i)表示获得第i等奖，获奖姓名是x。i+=1等价于i=i+1。

**综合题三**

3. 小申很喜欢编程，自学一段时间Python后，想体验它在数据采集和处理方面功能，在老师的建议下他首先尝试使用python到网络上“爬”数据。他试着编写了一个简单的程序：从某网站爬取了一些单车使用的数据，保存在bike.csv文件中，在用Excel打开此文件时出现了乱码，如下图所示。

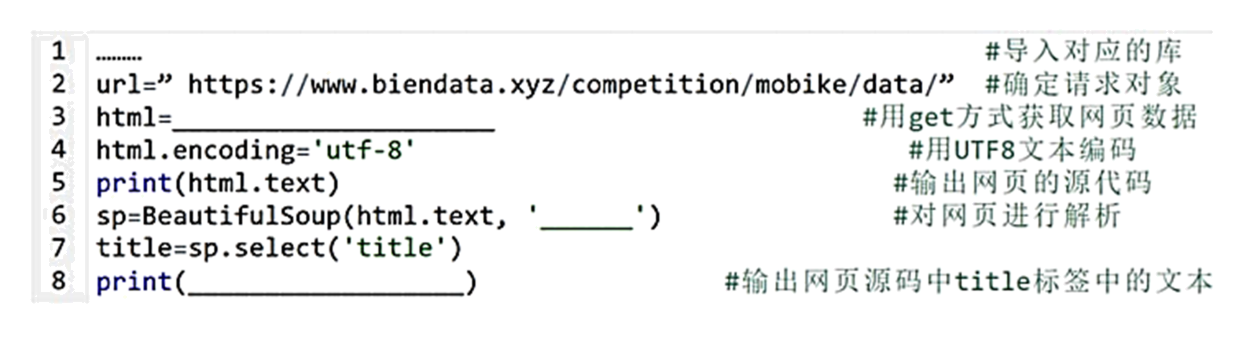


（1）小申从网络上“爬”数据，即通过互联网进行数据采集。其完整流程有三个步骤，分别是：\_\_\_\_\_\_\_\_\_\_\_、解析网页和保存数据。

（2）以下\_\_\_\_\_\_\_\_\_\_\_库与互联网数据采集有关的。（ ）

A．pandas B．numpy C．matplotlib D．requests

（3）网页请求方式有多种，GET是最常见的方式，一般用于获取或者查询资源信息，响应速度较快。若已经导入了方便发送HTTP请求的库，要用get方式获取网页数据的python代码（即下图第3行的横线上）需填写\_\_\_\_\_\_\_\_\_\_\_；获取了网页信息后需要使用BeautifulSoup库中的方法进行网页解析，则代码中对应横线（图第6行）处应填写\_\_\_\_\_\_\_\_\_\_\_



（4）接上题，要输出网页源码中title标签的文本则先要用select属性抓取title标题，然后再使用print命令输出，那么print后的横线处（上图第8行）应填写\_\_\_\_\_\_\_\_\_\_\_

A．title．Text B．sp．text C．html．title．text D．sp．title．text

（5）以下\_\_\_\_\_\_\_\_\_\_\_是爬虫技术可能存在风险。

A．大量占用爬取网站的资源

B．网站敏感信息的获取造成的不良后果

C．大量访问目标系统，带宽被强制占用

D．以上都是

（6）小申又观察了这份数据文件，发现所有的英文、数字均正常显示，这是因为西文字符集很小，只需要\_\_\_\_\_\_\_\_\_\_\_个字节就能表示所有西文字符。其中字节（Byte）是计算机中信息组织和存储的基本单位，生活中我们还用过KB、MB、TB等单位，若需要进行换算的化：1TB=\_\_\_\_\_\_\_\_\_\_\_Byte（用幂次表示）

（7）请简单说明bike.csv文件里中文字符出现乱码的原因。\_\_\_\_\_\_\_\_\_\_\_。

【答案】 ①. 获取网页 ②. D ③. requests.get(url) ④. html.parser ⑤. A ⑥. D ⑦. 1 ⑧. 240 ⑨. 中文出现乱码原因可能是 python 代码用的是 utf -8编码对获取的网页信息进行编码。而目标网站不是用 utf -8编码，可能是 gbk 等其他编码方式。也可能是解析时目标网站不是 html 代码

【解析】

【详解】本题主要考查Python程序的综合应用。（1）网络爬虫完整流程的三个步骤分别是获取网页、解析网页和保存数据。获取网页是解析网页的前提。（2）Pandas 和 numpy 和数据计算有关， matplotlib 和绘图有关， requests 和数据抓取有关。故选D选项。（3）要用 get 方式获取网页数据的 python 代码为 requests . get ( url )。其中 requests 为库名， get 为方法名。 url 是要抓取的网址。html . parser是Python内置的专门用来解析HTML的模块。利用html . parser我们可以分析出一段HTML里面的标签、数据等，是一种处理HTML的简便途径，因此第6行处填写 html . parser 。表示用 html 代码格式解析获得的信息。（4）title 表示获得的网页 title 部分。 title . Text 表示 title 标签中的内容。故本题选A选项。（5）抓取模板网站的信息会占用目标网站的服务器资源和宽带资源。同时抓取到的信息可能并不是目标网站想公开的内容。 ABC 三个选项说法都正确。故本题选 D选项 。（6）英文和数字一个字节就可以保存，汉字一般需要两个字节保存。从 Byte 到1TB需要经过 KB 、 MB 、 GB ，每次进制为2的十次方。故1TB等于240Byte。（7）中文出现乱码的原因可能是 python 代码用的是 utf -8编码对获取的网页信息进行编码。而目标网站不是用 utf -8编码，可能是 gbk 等其他编码方式。也可能是解析时目标网站不是 html 代码。

**综合题四**

4. 近几年全球的气候多变，夏季高温天增多。为了探究背后的原因，李老师给同学们布置了课后活动——选择一个城市进行相关的气候研究。小申他们组采集了某城市2022年7月的气象数据。数据中包含了每天的日平均气温（℃）、日平均相对湿度（%）、日降水量（mm）、日平均风速（km/h）、日照时数（h）。

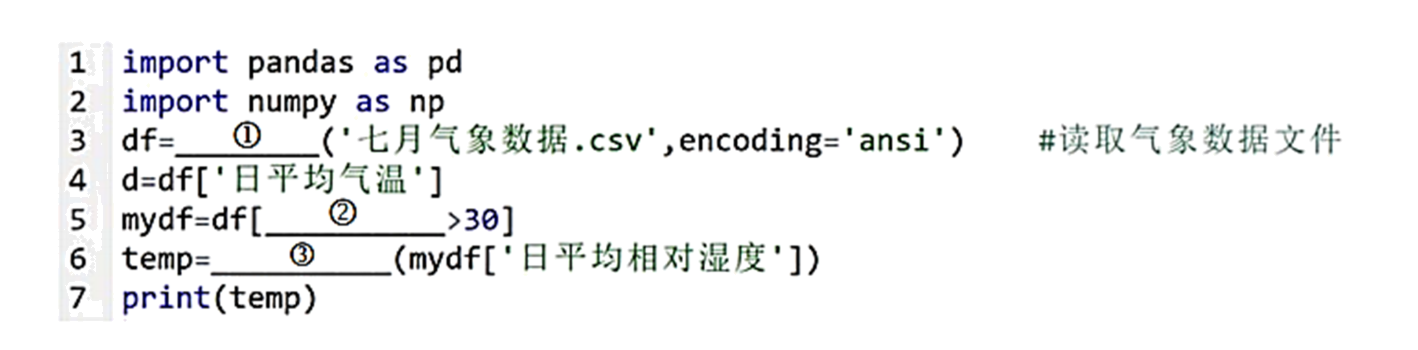
（1）小申采集的部分气象数据如下图所示，他使用Python第三方库pandas提供的方法进行数据整理，删除有缺失值的记录，若读取的数据存放在df中，以下方法使用正确的是\_\_\_\_\_\_\_\_\_\_\_。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 日期 | 日平均气温 | 日平均相对湿度 | 日降水量 | 日平均风速 | 日照时数 |
| 1日 | 29.3 | 8 | 0 | 1.9 | 6.6 |
| 2日 | 29.8 | 8.7 | 5.4 | 0.9 | 7.9 |
|  | 5.3 | 6.2 | 0 | 2 | 0 |
| …… |  |  |  |  |  |

A．df.isnull（　　） B．df.dropna（　　）

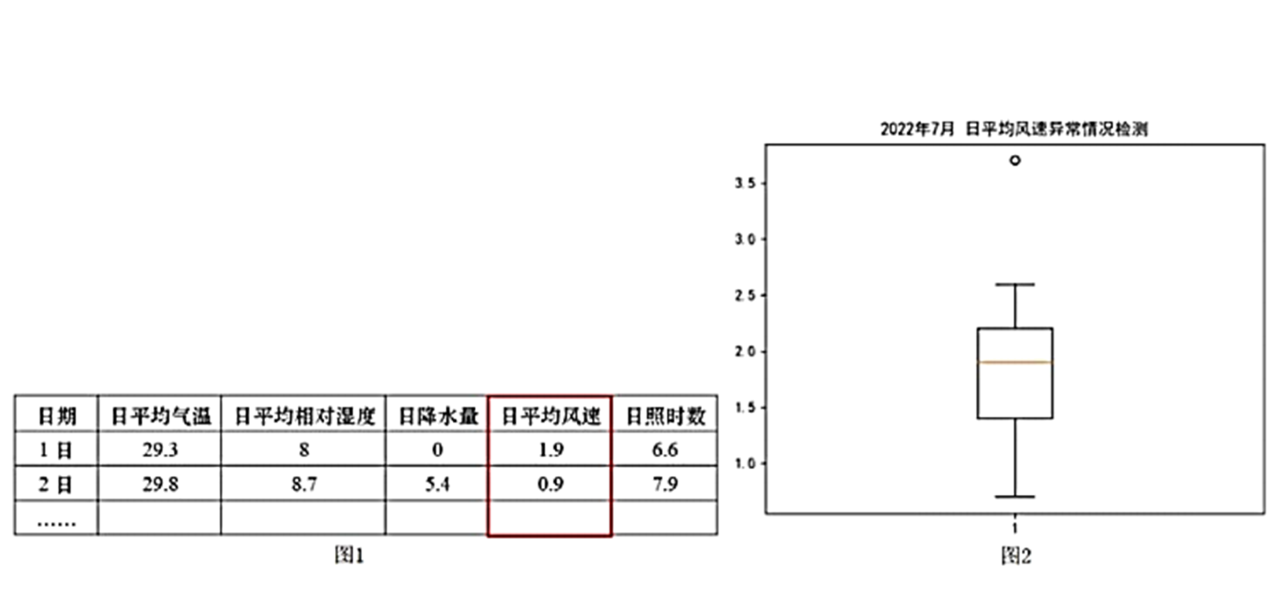
C．df.drop\_duplicates（　　） D．df.drop\_duplicates（subset=['日期']）

（2）小申把整理后的数据存储于“七月气象数据.csv”文件中。他编写程序对这些数据进行分析（如下图所示），在日平均气温大于30的记录中，找出日平均相对湿度的最大值（可使用numpy中的max函数），请将程序填写完整。



①\_\_\_\_\_\_\_\_\_\_\_②\_\_\_\_\_\_\_\_\_\_\_③\_\_\_\_\_\_\_\_\_\_\_

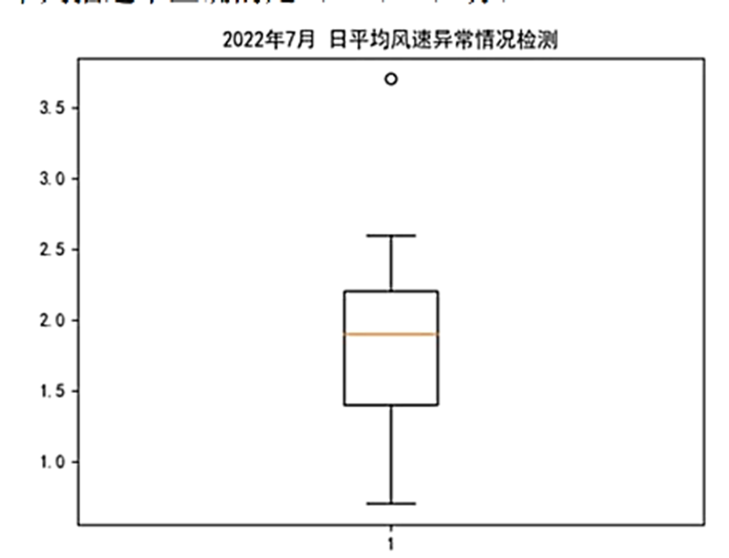
（3）小申根据如图1所示的七月份气象数据，编写程序绘制如图2所示的七月份日平均风速异常情况检测可视化图形。该图形叫\_\_\_\_\_\_\_\_\_\_\_，我们可以借助于\_\_\_\_\_\_\_\_\_\_\_函数完成它的绘制。\_\_\_\_\_\_



A．箱形图 box（　　） B．散点图 scatter（　　）

C．散点图 plot（　　） D．箱形图 boxplot（　　）

（4）小申根据下图所示的7月份日平均风速异常情况检测可视化图形进行了分析，下列描述中正确的是\_\_\_\_\_\_\_\_\_\_\_。



A．该月的日平均风速数据中存在异常值

B．必须立即删除用小圆圈表示的这个数据

C．该月的日平均风速数据中存在缺失值

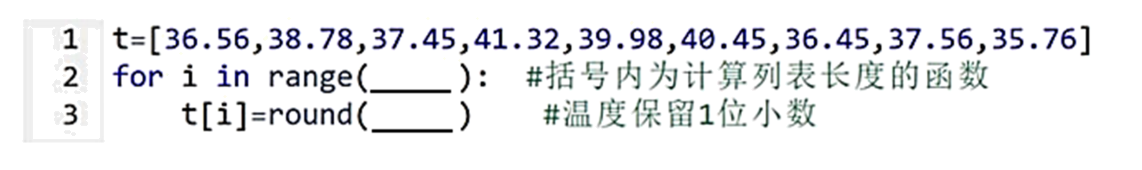
D．可以进一步分析用小圆圈表示的这个数据

（5）小申了解到某省气象大数据云平台可以存储地面、高空、海洋等气象数据，这些数据中有二维表格数据、文本数据、图片数据、地理位置数据等，这最能体现的大数据特征是\_\_\_\_\_\_\_\_\_\_\_。

A．数据规模大 B．处理速度快 C．数据类型多 D．价值密度低

（6）为了及时做好防暑降温工作，国家对不同的气温发布对应的预警信息。小申采集了部分城市气象部门发布的某日14时气温预报数据，并设计了一个算法，统计发布黄色预警的城市数量。

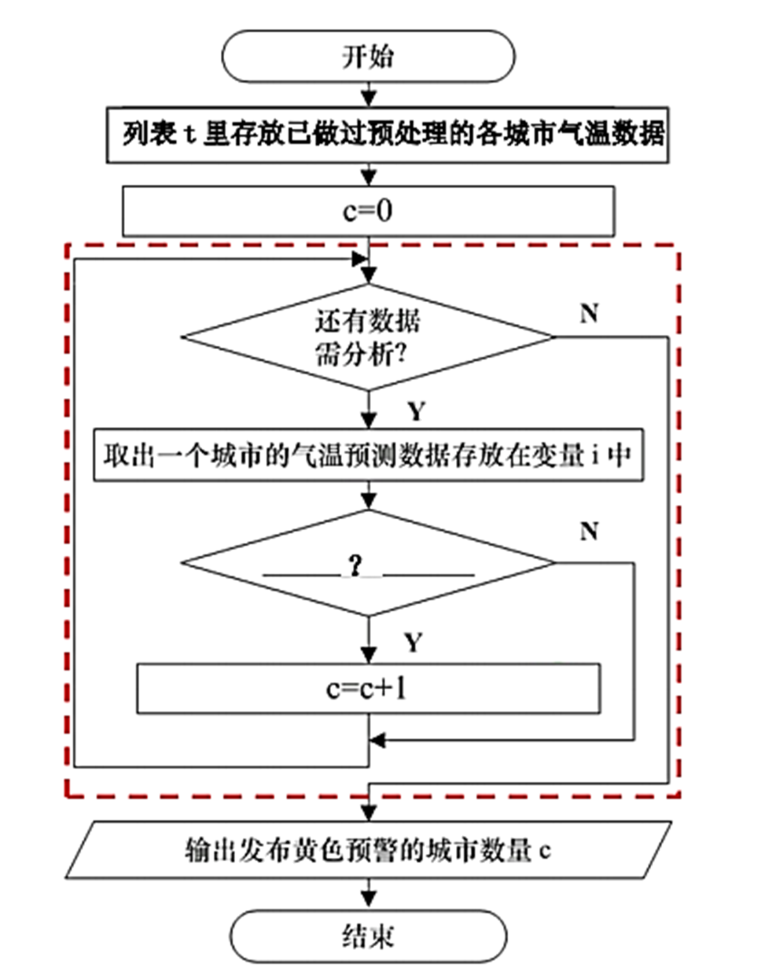
①采集的气温预报数据存在列表t中，在进行分析之前需要对温度作保留1位小数的预处理，小申写了下面的代码，根据注释补全其中的空白部分。



第2行空白划线处应填：\_\_\_\_\_\_\_\_\_\_\_第3行空白划线处应填：\_\_\_\_\_\_\_\_\_\_\_

②小申开始是用流程图设计的算法（下图），补全横线处的语句\_\_\_\_\_\_\_\_\_\_\_。

|  |  |
| --- | --- |
| 预警 | 气温范围 |
| 黄色预警 | 35℃≤气温<37℃ |
| 橙色预警 | 37℃≤气温<40℃ |
| 红色预警 | 气温≥40℃ |



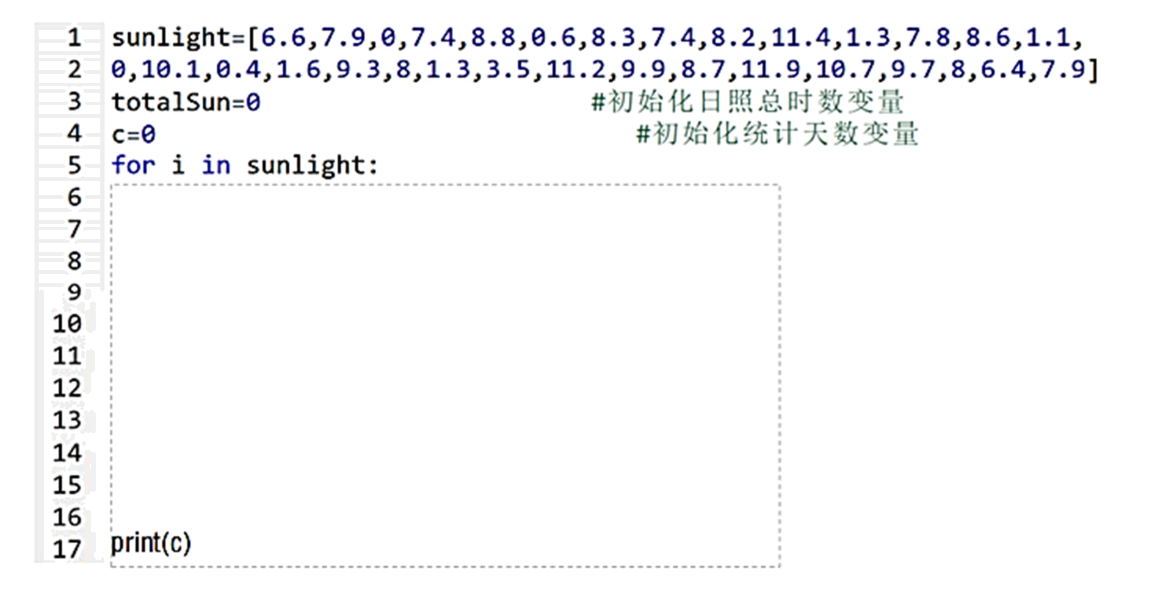
③以上流程图中，虚线框内是典型的\_\_\_\_\_\_\_\_\_\_\_算法

（7）为了统计出7月份低于平均日照时数的天数，小申打算编写Python程序解决。程序中sunlight用于记录“7月份每天的日照时数”的变量名，c是记录天数的变量名，i是循环控制的变量名，若要设置变量存放7月份日照总时数，从命名的规则来看，下面的变量名\_\_\_\_\_\_\_\_\_\_\_是不合法的

A．Import B．sum-days C．printJuly\_total D．Junetotal

（8）按照题4-7中的要求，统计出7月份（31天）低于平均日照时数的天数，请帮助小申完善程序。（在虚线框内补全）\_\_\_\_\_\_\_\_\_\_\_

【提示】：先计算7月份的日照总时数，即可求出平均日照时数；再将每日日照时数与之比较，小于的话则累计1。



【答案】 ①. B ②. pd.read\_csv ③. d ④. max ⑤. D ⑥. A ⑦. C ⑧. len(t) ⑨. t[i],1 ⑩. 35<=i <37? ⑪. 多分支结构 ⑫. A B ⑬. totalSun+=i  
for i in sunlight:  
 if i < totalSun/len(sunlight):  
 c=c+1

【解析】

【详解】本题主要考查Python程序的综合应用。（1） df . isnull ()判断是否为缺失值，df . dropna()为删除缺失数据， df . dropauplicates()为删除重复记录，故本题选B选项。（2）①处读取七月气象数据文件，所以填写为 pd .read .csv ，用变量 d 来表示日平均气温，所以如果日平均温度大于30，则赋值给 mydf ，最后 temp 求列表 mydf 中的最大值，所以②和③分别为 d 、 max 。（3）如图所示的图形为箱形图，箱线图( Box - plot ）又称为盒式图或箱型图，是一种用来显示一组数据分散情况的统计图，它能显示一组数据的上界、下界、中位数、上下四分位数以及异常值等。绘制箱型图的函数是boxplot，故本题选D选项。（4）箱型图中在最上方和最下方的圆圈分别表示样本数据中的极端值，所以如图所示该月的日平均风速数据中存在异常值，后期如果继续分析可以剔除用小圆圈表示的这个数据，故本题选 A选项。（5）这些数据中有二维表格数据、文本数据、图片数据、地理位置数据等，这最能体现的大数据特征是数据类型多，故本题选C选项。（6）range(start, stop, [step])，start: 计数从 start 开始。默认是从 0 开始。例如range（5）等价于range（0， 5）；stop: 计数到 stop 结束，但不包括 stop。例如：range（0， 5） 是[0, 1, 2, 3, 4]没有5；step：步长，默认为1。例如：range（0， 5） 等价于 range(0, 5, 1)。括号内为计算列表长度的函数，故填len(t)。温度保留1位小数，故此处填t[i],1。以上流程图中，虚线框内是典型的多分支结构。（7）变量的命名规则有：只能以字母或下划线开头，不能包含特殊符号，不能是关键字。Import是关键字，sum-days含有特殊符号，因此均不能作为变量名，故本题选AB选项。（8）根据题干，具体代码如下：

totalSun+=i  
for i in sunlight:  
 if i < totalSun/31:  
 c=c+1

