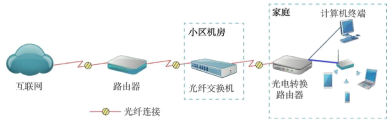
**建平中学2023学年第二学期期末教学质量检测**

**高一信息科技**

**试卷说明：**

**本试卷总分100分，考试时间为60分钟，答题过程中不可使用计算器、手机、电子词典等辅助计算工具。**

1. 小明学校的智能体育系统采用了物联网技术，用来检测同学健康及体质达标情况。室内及室外场馆均接入了网络，可实现智能管理。



学校体育场馆 小明家

（1）该系统利用智能手环实时采集学生心率，这属于物联网基本架构中的（ ）

A.感知层 B.网络层 C.应用层 D.逻辑层

（2）场馆内各类计算机终端通过设备A与服务器相连，设备A应该是（ ）

A.交换机 B.路由器 C.无线AP D.基站

（3）若要使体育场馆内的计算机连入因特网，除了配备通信线路和连接设备之外，还需安装的网络协议是（ ）协议。

A.SMTP B.HTTP C.FTP D.TCP/IP

（4）要想使体育馆内的网络接入因特网，并访问因特网所提供的各类服务与丰富的信息资源，除了需要通过公用或私有的线路将设备或局域网连接到因特网，还需要利用（ ）提供的接入服务。

A.通信设备生产商 B.网络安全管理中心 C.因特网服务提供商ISP D.学校信息办公室

（5）小明将笔记本电脑接入图书馆的网络，获取的IP地址是10.68.5.203，他想知道与中心服务器（IP地址为10.68.0.168）是否顺利通信，可以使用的命令是（ ）

A.ipconfig 10.68.0.168 B.ping 10.68.5.203 C.ping 10.68.0.168 D.cmd 10.68.5.203

（6）小明家宽带采用光纤接入方式，带宽为300Mbps。小明的手机和家中无线路由都支持500Mbps的最大传输速率，小明用手机通过WLAN下载一部2GB的视频，理论上需要的下载时间最少是（ ）秒

A.2\*1024\*8/500 B.2\*1000\*8/500 C.2\*1024\*8/300 D.2\*1000\*8/300

（7）小明发现，使用同一移动终端访问学校场馆的中心服务器，在家里比在学校图书馆的速度更快，原page number 0

因可能有\_\_\_\_。（至少2条）

（8）同学们使用学生平板上的智慧校园APP就可以查询到个人信息，这个APP属于信息系统组成要素中的（ ）

A.数据 B.软件 C.硬件 D.人

（9）小明通过智慧校园APP下载了学校的校歌，该WAV音频采样频率为44.1KHz，采用16位立体声录制，时长为3分钟，则该音频文件的数据存储量为（ ）

A.44.1\*16\*2\*3/8 B B.44.1\*4\*2\*3\*60/8 B C.44.1\*1000\*4\*3\*60/8/1024 KB D.44.1\*1000\*4\*3\*60/1024 KB

（10）小明为了完成智慧校园研究报告，下载了一些学习资料和软件工具等，以下不能进行有损压缩的是（ ）

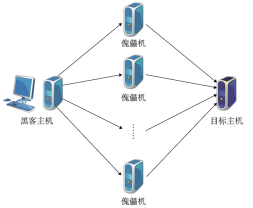
A.智慧校园.exe B.智慧校园宣传片.AVI C.智慧校园海报.bmp D.智慧校园项目调查问卷.docx

（11）信息系统为同学们带来便利的同时，也存在一定的安全风险，比如病毒破坏、数据丢失、隐私泄露等。针对不同的风险，以下应对措施合理的是（ ）

A.安装防火墙并定期用杀毒软件查杀 B.定期对系统的数据进行备份

C建立网络安全制度，规范网络管理流程 D.设置访问控制，只允许管理员访问系统

（12）为进一步了解信息系统安全，小明通过查阅资料了解到，如下图所示，有一种黑客非法行为，是通过控制许多傀儡机向目标主机发送大量数据包，使目标主机的所有可用资源都被消耗完，导致合法用户的请求无法回应，这种系统安全风险属于（ ）



A.管理安全风险 B.网络安全风险 C.软件安全风险 D.物理安全风险

【答案】 ①. A ②. A ③. D ④. C ⑤. C ⑥. C ⑦. 学校图书馆的网络带宽被共享；学校图书馆的网络设备性能较差；移动终端连接学校无线AP信号受干扰等 ⑧. B ⑨. D ⑩. AD ⑪. ABC ⑫. B

【解析】page number 1

【详解】 本题考查物联网。

（1）智能手环实时采集学生心率，属于获取数据的感知设备，这是物联网基本架构中的感知层。感知层负责收集和获取信息。故答案为：A。

（2）场馆内各类计算机终端通过设备与服务器相连，交换机常用于在局域网内连接多个终端设备和服务器，实现数据交换。故答案为：A。

（3）TCP/IP是因特网中广泛使用的网络协议，要使计算机连入因特网，需要安装该协议。故答案为：D。

（4）因特网服务提供商（ISP）提供接入因特网的服务，通过他们的网络设施和服务，用户能够连接到因特网。故答案为：C。

（5）ping命令用于测试网络连接，ping 10.68.0.168可以检查与中心服务器是否能通信。故答案为：C。

（6）带宽为300Mbps，换算为字节是300/8Mbps，文件大小2GB换算为字节是2\*1024\*8M字节，下载时间=文件大小/下载速度，即2\*1024\*8/(300/8)秒。故答案为：C。

（7）学校图书馆的网络带宽可能被多个用户共享，导致每个用户可用带宽减少；学校图书馆的网络设备可能性能较差，影响数据传输速度；移动终端连接学校无线AP时，信号可能受到干扰或衰减，影响通信质量。

（8）智慧校园APP属于软件，是信息系统中的软件要素。故答案为：B。

（9）音频存储量=采样频率×量化位数×声道数×时长/8，44.1KHz换算为44100Hz，时长3分钟换算为3×60秒，计算为44100×16×2×3×60/8字节，再换算为KB为44100×16×2×3×60/8/1024KB。故答案为：D。

（10）.exe是可执行文件，.docx是文档格式，通常这两种格式不进行有损压缩。故答案为：AD。

（11）安装防火墙和定期杀毒可防范病毒破坏；定期备份数据可防止数据丢失；建立网络安全制度和规范流程有助于整体的网络安全管理。只允许管理员访问系统会极大限制正常使用，不是合理的应对措施。故答案为：ABC。

（12）黑客通过控制傀儡机发送大量数据包攻击目标主机，属于网络攻击行为，是网络安全风险。故答案为：B。

2. 为了更好地保护环境，落实垃圾回收与分类，小明尝试开发基于人工智能技术的可回收垃圾识别系统。

（1）用于训练人工智能模型的可回收垃圾的图片对应的类别已提前做好了人工标记，该机器学习方法属于（ ）

A.非监督学习 B.半监督学习 C.监督学习 D.强化学习

（2）下图是一张用于训练的图片，其分辨率为300\*400，位深度为16bit，文件类型为bmp。理论上，该图像文件的数据存储量是（ ）B。page number 2



A.300\*400\*16/8/1024 B.300\*400\*4/8 C.300\*400\*16/8 D.300\*400\*16/1024

（3）该图像色彩采用RGB格式，其中某像素点的蓝色值用十六进制表示是BE，转换为二进制数是（ ）

A.10101111 B.10111111 C.11011110 D.10111110

（4）小明先实现系统对可回收垃圾中易拉罐和玻璃瓶的识别，训练样本特征值如下表所示。现有一个未知分类样本，特征值一为8，特征值二为7。根据欧式距离法，可判定该样本属于\_\_\_\_类；若采用k-近邻算法，k取5时，该样本应判定属于\_\_\_\_类。

A.易拉罐 B.玻璃瓶

|  |  |  |
| --- | --- | --- |
| 类别 | 特征值一 | 特征值二 |
| 易拉罐 | 10 | 4 |
| 易拉罐 | 8 | 3 |
| 易拉罐 | 9 | 3 |
| 玻璃瓶 | 6 | 9 |
| 玻璃瓶 | 6 | 12 |
| 玻璃瓶 | 5 | 10 |

（5）小明通过计算准确率，来客观地评估该系统的分类效果。已知训练数据有1000条，测试数据有200条，识别错误的数据有20条，则该系统的准确率是\_\_\_\_ %。

（6）人工智能技术如今已得到广泛应用，例如进行人脸识别、人脸检测。人脸检测最常用的方式是调用某些平台SDK来实现。将以下步骤按正确顺序排列为：\_\_\_\_。

①调用SDK，获取人脸检测（或识别）任务的返回信息 ②从返回信息中提取人脸的定位信息

③根据人脸定位信息绘制矩形框 ④导入智能工具图像处理SDK

A.④①②③ B.④③②① C.④②①③ D.④①③②

（7）小明积极学习人脸检测技术，导入了一张照片facel.jpg进行测试，使用Pillow库，将照片中的人脸位置标记出来，即以（110，90）为起点，划宽、高分别为120、100的蓝色矩形。补全如下代码正确的选项是（ ）。page number 3

|  |
| --- |
| from PIL import Image, ImageDraw  im = Image.open('facel.jpg')  draw=ImageDraw.Draw(im)  draw.rectangle((110,90, , ), outline=( )) #颜色按RGB方式  im.show() |

A.120 100 255,0,0 B.230 190 255,0,0 C.120 100 0,0,255 D.230 190 0,0,255

【答案】 ①. C ②. C ③. D ④. A ⑤. A ⑥. 90 ⑦. A ⑧. D

【解析】

【详解】本题考查人工智能。

（1）监督学习是一种机器学习方法，其中训练数据集中的每个样本都被标记正确的输出。在这种学习模式下，模型从带标签的数据中学习输入和输出之间的关系，并使用这个学习到的关系来对新的、未见过的数据进行预测。可回收垃圾的图片对应的类别已提前做好了人工标记，符号监督学习的特征。故正确答案为：选项C。

（2）为了计算一张分辨率为300\*400、位深度为16bit、文件类型为bmp的图像文件的数据存储量，我们可以使用以下公式：图像文件数据存储量 = 水平像素 × 垂直像素 × 每个像素所需位数 / 8（字节）=300\*400\*16/8。故正确答案为：选项C。

（3）要将十六进制数BE转换为二进制数，我们需要了解十六进制和二进制之间的转换规则。十六进制数BE由两个十六进制位组成：B和E。十六进制中的B等于二进制的1011（因为十六进制中的B是10进制的11，而11的二进制是1011）。十六进制中的E等于二进制的1110（因为十六进制中的E是10进制的14，而14的二进制是1110）。故二进制数为：10111110。故正确答案为：选项D。

（4）首先，我们计算未知分类样本（特征值一为8，特征值二为7）与每个已知分类样本之间的欧式距离。

欧式距离的计算公式为：d=sqrt((x2​−x1​)\*\*2+(y2​−y1​)\*\*2)，其中(x1​,y1​) 和 (x2​,y2​) 是两个样本的特征值。对于未知样本（8,7）与各个已知样本的距离：

与第一个易拉罐样本（10,4）的距离：d=sqrt((10−8)\*\*2+(4−7)\*\*2​)=sqrt(13​)。

与第二个易拉罐样本（8,3）的距离：d=sqrt((8−8)\*\*2+(3−7)\*\*2​)=sqrt(16)

与第三个易拉罐样本（9,3）的距离：d=sqrt((9−8)\*\*2+(3−7)\*\*2)​​=sqrt(17​)

与第一个玻璃瓶样本（6,9）的距离：d=sqrt((6−8)\*\*2+(9−7)\*\*2)​=sqrt(8)

与第二个玻璃瓶样本（6,12）的距离：d=sqrt((6−8)\*\*2+(12−7)\*\*2​)=sqrt(29)​page number 4

与第三个玻璃瓶样本（5,10）的距离：d=sqrt((5−8)\*\*2+(10−7)2\*\*2)=sqrt(18)​

按照欧式距离法，选择距离最小的类别作为未知样本的类别。这里，未知样本与三个易拉罐样本的距离相对最短，所以根据欧式距离法，该样本应判定属于A.易拉罐类。接下来，考虑k-近邻算法（k-NN），k取5时：对未知样本（8,7）来说，最近的5个样本中，有3个是易拉罐（第二个、第一个、第三个易拉罐样本），2个是玻璃瓶（第一个、第三个玻璃瓶样本）。因此，根据k-近邻算法（多数投票原则），该样本应判定属于A.易拉罐类。综上，无论是欧式距离法还是k-近邻算法（k=5），该样本都应判定属于A.易拉罐类。

（5）要计算该系统的准确率，我们需要使用以下公式：准确率=正确识别的样本数/总测试样本数​×100%。根据题目，测试数据有200条，识别错误的数据有20条。因此，正确识别的样本数是：正确识别的样本数=总测试样本数−识别错误的数据数=200−20=180。将上述值代入准确率公式中，我们得到：准确率=180/200​×100%=90%。所以，该系统的准确率是90%。

（6）对于人脸检测任务中使用SDK的步骤排序，我们需要按照逻辑和执行流程来确定正确的顺序。下面是每个步骤的解释和排序理由：

① 调用SDK，获取人脸检测（或识别）任务的返回信息：这个步骤是在SDK被正确加载并准备好之后执行的，用于启动人脸检测或识别任务，并从SDK接收处理结果。

② 从返回信息中提取人脸的定位信息：在获取到SDK的返回信息后，需要解析这些信息以获取人脸在图像中的具体位置（如坐标和大小）。

③ 根据人脸定位信息绘制矩形框：有了人脸的定位信息后，就可以在图像上绘制矩形框来标记人脸的位置。

④ 导入智能工具图像处理SDK：这是整个流程第一步，需要首先导入或加载用于人脸检测的SDK，以便后续调用其提供的功能。

根据上述解释，正确的执行顺序应该是先导入SDK（④），然后调用SDK的功能（①），从返回的结果中提取人脸定位信息（②），最后根据这些信息在图像上绘制矩形框（③）。因此，正确的步骤顺序是④①②③。故正确答案为：选项A。

（7）在这个问题中，我们需要补全draw.rectangle()方法的参数来绘制一个蓝色的矩形，该矩形以(110, 90)为起点，宽为120，高为100。首先，draw.rectangle()方法的参数需要两个点来定义一个矩形，这两个点分别是矩形的左上角和右下角。既然我们已经知道左上角的坐标是(110, 90)，并且矩形的宽是120，高是100，那么右下角的坐标可以通过左上角坐标加上宽和高来得到，即(110+120, 90+100) = (230, 190)。接下来，我们需要设置矩形的轮廓颜色。在Pillow库中，颜色通常按照RGB（红、绿、蓝）的顺序以元组的形式给出，其中每个颜色的取值范围是0到255。蓝色在RGB中的表示是(0, 0, 255)。故正确答案为：选项D。page number 5

3. 2024年9月，建平中学将迎来80周年校庆，为了方便广大校友进行返校，小申想利用所学，搭建一个建平校友返校预约系统。

（1）小申设计了一项功能，校友返校预约后，系统会自动为其分配活动室。学校内有两个教学楼教室可用作活动室，其中一个楼有5层，每层有14个门；另一个楼有6层，每层有20个门。在进行代码设计时，对门牌号进行二进制编码，至少需要 位二进制数？若编码需要体现楼栋、楼层和门牌号信息，则至少需要 位二进制数（ ）

A.7 8 B.8 9 C.7 9 D.9 10

（2）用户需要通过浏览器和URL地址访问并获取该系统提供的预约服务，则该系统的软件结构是\_\_\_\_结构

A.HTML B.P2P C.B/S D.C/S

（3）小申设计了搭建系统所需的网页模板文件，应该默认存放到项目包中的\_\_\_\_文件夹中

A.instance B.static C.templates D.venv

（4）小申为“添加返校预约”功能设计的用户界面如下图所示，采集校友的“毕业年份”信息时，最好是采用以下哪种方式（ ）



A.使用键盘手动输入 B.下拉框中选择年份 C.获取系统当前年份 D.查询获取毕业年份

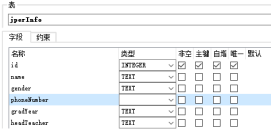
（5）校友使用该系统添加返校预约时，填写并提交的预约信息数据的流向是（ ）

A.业务逻辑层→数据访问层→数据库→用户界面层 B.数据访问层→业务逻辑层→数据库→用户界面层

C.业务逻辑层→用户界面层→数据访问层→数据库 D.用户界面层→业务逻辑层→数据访问层→数据库

（6）小申设计了系统所需的数据库，并用数据表jperInfo存放校友返校预约信息，如下图所示，则手机号phoneNumber字段应该设置为\_\_\_\_类型

A.integer B.text C.numeric D.realpage number 6



（7）小申为了提供对数据库的数据访问，编写了自定义函数于程序db.py中，该程序文件对应系统的软件架构中的\_\_\_\_层

A.用户界面 B.数据访问 C.业务逻辑 D.网络传输

（8）程序db.py中，将校友返校预约信息写入数据库的函数add\_one()部分代码如下，请帮小申补全空白处的代码。

|  |
| --- |
| def add\_one(xm, xb, phone, year, bzr):  conn=sqlite3.connect("instance/test.db")  cur=conn.cursor()  cur.execute("①\_\_\_\_ (name, gender, phoneNumber, gradYear, headTeacher)  values(?,?,?,?,?)",(xm, xb,②\_\_\_\_, year, bzr))  cur.close()  conn.commit()  conn.close() |

（9）小申使用Flask框架搭建网站服务端，主程序run.py中的关键代码的顺序应该是（ ）

① app= Flask(\_\_name\_\_)

② from flask import Flask, render\_template, request, redirect

③ app.run()

④ @app.route(’/’)

A.②①④③ B.②③④① C.②④①③ D.②①③④

（10）小申编程实现校友“添加返校预约”功能，并通过在浏览器中输入URL（http://127.0.0.1:8000/add）进行本地测试，请帮其补全下方关键代码。

|  |
| --- |
| @app.route("①\_\_\_\_ ", methods=["GET","POST"])  def add\_info():  if request.method=="②\_\_\_\_ ":  xm= request.form["name"] |

page number 7

|  |
| --- |
| xb=request.form["gender"]  phone=request.form["phoneNumber"]  year=request.form["grad Year"]  bzr= request.form["headTeacher"]  db.add\_one(xm, xb, phone, date, year, bzr)  return redirect('/')  return render\_template("add.html")  app.run(debug=True, host=’0.0.0.0’, port=③\_\_\_\_) |

（11）小申设计的系统除了“添加返校预约”外，还有“查看所有预约”和“查询个人预约”等功能。为了保护用户隐私，查看预约信息时，对手机号信息的显示做了一定处理，例如“198\*\*\*\*1234”，这是进行了\_\_\_\_。

A.数据加密 B.数据脱敏 C.数据整理 D.数据修改

（12）在系统运行与测试阶段，小梅提出——所有用户都能进入“查看所有预约”功能页面不合理。为解决该问题，进一步完善该系统，可以采用的手段有\_\_\_\_。

A.身份认证 B.访问控制 C.修补漏洞 D.定期查毒

【答案】 ①. B ②. C ③. C ④. B ⑤. D ⑥. B ⑦. B ⑧. insert into jperInfo ⑨. phone ⑩. A ⑪. /add ⑫. POST（须大写） ⑬. 8000 ⑭. B ⑮. AB

【解析】

【详解】本题考查信息系统。

（1）首先，对于门牌号的二进制编码，其中一个楼有5层，每层有14个门；另一个楼有6层，每层有20个门。共计190个门。故需要8位二进制数表示。接下来，若编码需要体现楼栋、楼层和门牌号信息，我们可以这样分配。楼栋：2个选项（两栋楼），需要1位二进制数。楼层：最高6层，需要3位二进制数。门牌号：20个门，需要5位二进制数。所以总共需要 1(楼栋)+3(楼层)+5(门牌号)=9 位二进制数。故正确答案为：选项B。

（2）用户通过浏览器和URL地址访问系统，这是典型的浏览器/服务器（B/S）结构。故正确答案为：选项C。

（3）在Web项目中，网页模板文件通常存放在名为templates的文件夹中。故正确答案为：选项C。

（4）对于“毕业年份”这样的信息，最好是让用户从下拉框中选择，这样可以减少输入错误，并快速选择。故正确答案为：选项B。

（5）用户填写并提交预约信息时，数据的流向是从用户界面层开始，经过业务逻辑层处理，然后到数据page number 8

访问层与数据库交互。故正确答案为：选项D。

（6）手机号通常包含数字、加号等字符，因此应该使用文本类型来存储。故正确答案为：选项B。

（7）自定义函数用于数据访问，因此它对应的是数据访问层。故正确答案为：选项B。

（8）使用SQL语言中的insert into插入数据。表名为 jperInfo，自定义函数的形参为phone。故正确答案为：①insert into jperInfo；②phone。

（9）Flask程序的关键代码顺序应该是先导入Flask和其他需要的模块，然后创建Flask应用实例，接着定义路由和视图函数，最后运行应用。故正确答案为：选项A。

（10）①根据浏览器中输入URL（http://127.0.0.1:8000/add）可知路由函数的路径为/add。故正确答案为：/add。②检查请求方法，是否是POST方法。故正确答案为：POST（须大写）。③根据浏览器输入的地址，可知端口号为：8000。故正确答案为：8000。

（11）对于小申设计的系统中，手机号信息的显示方式“198\*\*\*\*1234”是进行了数据脱敏处理。这种处理方式旨在保护用户隐私，通过隐藏或替换数据中的敏感信息（如手机号码的中间几位），从而在不泄露敏感数据的前提下，满足数据处理和展示的需求。因此，正确答案是：选项B。

（12）在系统运行与测试阶段，针对小梅提出的问题——所有用户都能进入“查看所有预约”功能页面不合理，为了解决该问题，需要采取的手段是确保只有授权用户能够访问特定的资源或功能。这通常通过身份认证和访问控制来实现。身份认证用于确认用户身份，而访问控制则用于基于用户身份和权限来决定用户是否能够访问某个资源或执行某个操作。因此，正确答案为：选项AB。

4. “建平校友返校预约”系统运行了一段时间后，小达想对该系统收集到的校友返校预约数据进行一定分析。

（1）小达想要利用网络爬虫技术，采集“查看所有预约”功能页面上的全部预约信息。第一步是获取网页源代码，小达编写的部分程序如下：

|  |
| --- |
| h =① .get(url)  h.encoding=‘utf-8’ |

则①处应该使用的模块是（ ）

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

（2）第二步是解析获取到的网页内容，导入解析网页所需模块的语句是（ ）

A.import requests B.import BeautifulSoup C.import bs4.BeautifulSoup D.from bs4 import BeautifulSoup

（3）小达获取到的部分网页源代码如下：page number 9



小达编写了下方的程序，将解析后的网页保存到对象sp中，然后从sp对象中提取数据单元格中的数据，则代码空缺处应该是（ ）

|  |
| --- |
| ……  sp =BeautifulSoup(h.text, "html.parser")  data=① ("② ")  for i in data:  print(i.text) |

A.sp.find tr B.sp.find\_all tr C.sp.find td D.sp.find\_all td

（4）小达将获取到的数据保存到文件“校友信息.csv”中，部分数据如下所示：



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 编号 | 姓名 | 性别 | 手机号 | 毕业年份 | 班主任 |
| 1 | 小平 | 男 | 198\*\*\*\*1234 | 1999 | 金老师 |
| 2 | 小果 | 女 | 133\*\*\*\*0009 | 2012 | 郑老师 |
| 3 | 张三 | 男 | 198\*\*\*\*5678 | 2021 | 郑老师 |
| 4 | 李四 | 女 | 133\*\*\*\*3456 | 2013 | 王老师 |
| 5 | 刘二 | 男 | 133\*\*\*\*2222 |  |  |

page number 10

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 6 | 王五 | 男 | 198\*\*\*\*1234 | 2012 | 王老师 |
| 7 | 小平 | 男 | 198\*\*\*\*1234 | 1999 | 金老师 |

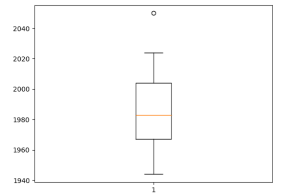
小达发现数据中存在重复记录和缺失值，想利用Python程序对数据进行整理，编写了如下代码，请根据小达的注释，补全其中空白处的代码。

|  |
| --- |
| import pandas as pd  #读取数据  df=pd.read\_csv("校友信息.csv", encoding='①\_\_\_\_ ')  #去除存在任意缺失值的记录，替换原表  df.dropna(how="②\_\_\_\_ ", inplace=True)  #去除重复记录，保留第一条，替换原表  df.drop\_duplicates(subset=["姓名","③\_\_\_\_ "], keep="first", inplace=True)  #保存数据  df.to\_csv("校友信息整理.csv") |

（5）小达想要检测毕业年份数据中是否存在异常值，可以通过绘制\_\_\_\_图和\_\_\_\_图。

（6）小达绘制的数据分析图如下，下方代码空缺处应该是 ？毕业年份数据中 异常值。（ ）

|  |
| --- |
| import matplotlib.pyplot as plt  plt. (df["毕业年份"], sym='o', whis=1.5)  plt.show() |



A.plot 存在 B.plot 不存在 C.boxplot 存在 D.boxplot 不存在

（7）对数据进行整理后，小达想要统计返校校友中毕业年份在1990年到1999年的人数，下列Python语句中可以实现该需求的是（ ）page number 11

A.len(df[(1990<=df["毕业年份"]<=1999)])

B.len(df [(df ["毕业年份"]>=1990) and (df ["毕业年份"]<=1999)])

C.len(df [(df ["毕业年份"]>=1990) & (df ["毕业年份"]<=1999)])

D.len(df[(df ["毕业年份"]>=1990) | (df["毕业年份"]<=1999)])

（8）小达想要分析不同年代毕业校友的返校情况，并将1940s、1950s、……2010s、2020s相应返校校友人数保存到列表sn中，准备按上题思路进行人数统计。小明却给了小达一个更好的建议一一找到毕业年份、毕业年代和列表sn索引三者间关系。

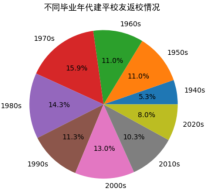
小达根据小明的建议，编写的代码如下，代码空白处应该为（ ）

A.i/10 B.i/10-194 C.i//10-194 D.int(i/10)-194

|  |
| --- |
| import pandas as pd  df=pd.read\_csv("校友信息整理.csv")  sn=[0,0,0,0,0,0,0,0,0,0] #不同毕业年代返校校友人数  for i in df[‘毕业年份’]: #遍历毕业年份数据  sn[ ]+=1 #对应毕业年代人数加1 |

（9）为了直观显示不同年代毕业校友的返校占比情况，小达编写程序绘制了下方的饼图，请补全小达代码中的空白处。

|  |
| --- |
| #代码续上题  plt.rcParams['font.sans-serif']=['SimHei']  plt.title('不同毕业年代建平校友返校情况')  nd=['1940s','1950s','1960s','1970s','1980s','1990s','2000s','2010s','2020s']  plt.pie(sn, labels=\_\_\_\_, autopet='%1.1f%%')  plt.show() |

page number 12

从上图中，你可以得出哪些结论或信息：预约返校校友中，\_\_\_\_。（至少两点）占比最高；1940年代的占比最低等

（10）小达想要利用枚举法，找出这次返校的毕业生姓名中，含有多少个“建”或“平”字，枚举法的控制结构是\_\_\_\_

A.循环结构 B.循环嵌套分支结构 C.分支嵌套分支结构 D.循环嵌套循环结构

（11）小达先将所有人的姓名拼接为一个字符串，接着统计其中含有多少个“建”或“平”字，请补全下方程序中的空白。

|  |
| --- |
| import pandas as pd  df=pd.read\_csv("校友信息整理.csv")  names =""  for i in df["姓名"]:  names +=i  c=0  #补全下方程序  \_\_\_\_  print(c) |

【答案】 ①. B ②. D ③. D ④. ANSI或ansi ⑤. any ⑥. 手机号 ⑦. 散点 ⑧. 箱形 ⑨. C ⑩. C ⑪. CD ⑫. nd ⑬. 1970年代的占比最高；1940年代的占比最低等 ⑭. B ⑮. for i in names:

if i == '建' or i =='平':

c=c+1 #c+=1

合理即可，说明：

1、i是循环变量名，写成a、b、c都可以

2、names是字符串序列，可以直接for遍历

3、'='是赋值符，'=='是关系比较符'等于'

【解析】

【详解】本题考查Python程序设计。

（1）对于第一个问题，小达想要使用网络爬虫技术来获取网页源代码。在Python中，用于发送HTTP请求并获取网页内容常用的库是requests。因此，在①处应该填写的模块是requests。故正确答案为：选项page number 13

B。

（2）第二步是解析获取到的网页内容。在Python中，BeautifulSoup是一个非常流行的HTML和XML解析库，用于从网页中提取数据。要导入这个库，应该使用from bs4 import BeautifulSoup。故正确答案为：选项D。

（3）对于第三个问题，小达想要从解析后的网页中提取数据单元格中的数据。由于BeautifulSoup的find\_all方法用于查找文档中所有匹配的标签，而find方法仅返回第一个匹配的标签，所以在这里应该使用find\_all方法来获取所有匹配的标签。另外，从给出的网页源代码片段中，数据在td标签内，故需要查找td标签。故正确答案为：选项D。

（4）①在处理中文文件时，通常需要指定正确的编码。常用的编码是ANSI。故正确答案为：ANSI或ansi。②dropna()函数用于删除包含缺失值的行或列。参数how可以是'any'（如果行/列中有任何NA值就删除）或'all'（只有行/列中的所有值都是NA才删除）。由于小达想要去除存在任意缺失值的记录，所以应该使用'any'。故正确答案为：any。③drop\_duplicates()函数用于删除重复的行。参数subset指定了哪些列的组合被用来识别重复项。由于小达想要基于姓名和手机号来识别重复项所以应该包含这两列。故正确答案为：手机号。

（5）箱形图：箱线图是一种用作显示一组数据分散情况资料的统计图。它能显示出一组数据的最大值、最小值、中位数、下四分位数及上四分位数。通过箱线图，可以很容易地识别出数据中的异常值（即那些远离其他观测值的观测值），因为异常值通常位于箱线图的上下须线之外。散点图也是检测数据中异常值的一种有效方式，特别是在涉及两个或多个变量的关系中。虽然散点图通常不直接用于单一变量（如毕业年份）来检测异常值，但如果小达想要将毕业年份与其他变量（如年龄、入学年份等）相关联，并观察是否存在异常模式或点，那么散点图将非常有用。故正确答案为：散点、箱形。

（6）代码中缺少部分为调用函数，画出箱线图。画出箱线图的函数为boxplot。观察图像可知，存在一个孤立点，为异常数据。故正确答案为：选项C。

（7）A. len(df[(1990<=df["毕业年份"]<=1999)])这个选项的语法是错误的。在Python中，1990<=df["毕业年份"]<=1999 这样的链式比较并不直接适用于Pandas的DataFrame或Series对象。B. len(df [(df ["毕业年份"]>=1990) and (df ["毕业年份"]<=1999)])这个选项使用了and而不是&，并且and在Pandas的DataFrame或Series的条件筛选中是不适用的。它会引发一个错误，因为and不能用于数组或序列之间的元素级比较。C. len(df [(df ["毕业年份"]>=1990) & (df ["毕业年份"]<=1999)])这个选项是正确的。它使用了&来组合条件，并且每个条件都被正确地应用在了df["毕业年份"]上，同时被括号包围以确保逻辑运算的优先级。D. len(df[(df ["毕业年份"]>=1990) | (df["毕业年份"]<=1999)])这个选项使用了逻辑或运算符（|），这会导致选择所有毕业年份大于或等于1990年或小于或等于1999年的记录，即实际上选择了所有年份的记录，因为任何年份都会满足这两个条件之一。这不是小达想要的结果。故正确答案为：选项C。page number 14

（8）A. i/10：这会给出毕业年份的十年份的浮点数表示，但没有转换为所需的索引。B. i/10-194：同样，这会给出浮点数，并且没有正确地转换为索引。故正确答案为：选项CD。

（9）为了补全小达代码中的空白处，使得饼图能够正确显示不同年代毕业校友的返校占比情况，你需要将nd列表作为标签（labels）参数的值传递给plt.pie()函数。故正确答案为：nd。根据饼图显示的数据可知，1970年代的占比最高；1940年代的占比最低等。

（10）使用循环控制需要遍历总次数，在循环内使用分支结构判断出姓名中有“建”或“平”字并统计。故正确答案为：选项B。

（11）首先需要取到姓名列，并遍历姓名，将所有的姓名累加到变量names中。其次使用for循环遍历字符串，使用分支结构判断每个字是否是“建”或“平”字。最后将符合判断条件次数累加到变量c中。

学科网(www.zxxk.com)--教育资源门户，提供试卷、教案、课件、论文、素材以及各类教学资源下载，还有大量而丰富的教学相关资讯！ **Python帮助小贴士**

·pandas库

|  |  |  |
| --- | --- | --- |
| 函数 | 简单示例 | 功能 |
| pandas.read\_csv() | pandas.read\_csv('文件名.扩展名', encoding=' utf-8') | 读取当前目录的csv文件，encoding表示文件的编码格式 |
| DataFrame.dropna() | DataFrame.dropna  (how='any', inplace=False) | 去除缺失记录。how表示记录中存在任意(any)/全部(all)缺失时进行去缺；inplace表示是否在原表上进行操作(True/False)。 |
| DataFrame.drop\_duplicates() | DataFrame.drop\_duplicates (subset=['x'], keep='first', inplace= False) | 去除重复记录。subset表示进行去重的列，None表示全部列；keep表示保留第几个数据(first/last/False)；inplace表示是否在原表上进行操作(True/False)。 |

·matplotlib库

|  |  |  |
| --- | --- | --- |
| 函数 | 简单示例 | 功能 |
| matplotlib.pyplot.plot() | matplotlib.pyplot.plot(x,y) | 绘制折线图 |
| matplotlib.pyplot.boxplot() | matplotlib.pyplot.boxplot(x, sym='o', whis=1.5) | 绘制箱形图 |
| matplotlib.pyplot.pie() | matplotlib.pyplot.pie(x, labels=y) | 绘制饼图 |

page number 15

|  |  |  |
| --- | --- | --- |
| matplotlib.pyplot.show() | matplotlib.pyplot.show() | 在本机显示图形 |

·操作数据库的命令SQL语句

|  |  |
| --- | --- |
| 操作 | 简单示例 |
| 查询 | select字段名1,字段名2,字段名3 from 数据表名 where 字段=值  例: select userid, bjxh, ys from xuanke where userid="21001" |
| 添加 | insert into 数据表名(字段名1,字段名2,字段名3) values(值1,值2,值3)  例: insert into xuanke (userid, bjxh) values ("21001","211012") |
| 修改 | update数据表名 set字段名1=值1,字段名2=值2 where 字段名=值  例: update xuanke set bjxh="210308" where userid="21001" |
| 删除 | delete from 数据表名 where 字段名=值  例: delete from xuanke where userid="21001" |

·Pyth on内置函数

|  |  |  |
| --- | --- | --- |
| 函数 | 简单示例 | 功能 |
| len() | len(x) | 返回序列x的长度或个数 |
| range() | range(x,y, step) | 返回一个等差整数序列 |
| round() | round(x,a) | 返回x四舍五入保留的a位小数 |

page number 16