**2022学年第一学期期末考试试卷**

**高一信息技术**

**综合题一：（30分）**

1. 中国城镇化数据分析研究

城镇化是人类社会发展的客观趋势，是国家现代化的重要标志，城镇化水平又称为城镇化率，是衡量城镇化程度的数量指标，也是衡量一个国家和一个地区社会经济发展水平的重要标志。中国正在加快推进新型城镇化建设，计划在2020年实施1亿左右农业转移人口和其它常住人口在城镇落户，常住人口城镇化率达到60%，户籍城镇化率达到45%。

（1）在研究某国城镇化率之前，需要先了解当前的人口数据。伴随着科学技术的发展，数据采集也变得越来越方便快捷，1953年我国进行第一次人口普查时，普查登记的实际时间历时半年多，而如今伴随着互联网等技术的应用，人口普查耗时越来越短，但是数据采集的基本流程是不变的，正确的数据采集流程为\_\_\_\_\_\_\_→\_\_\_\_\_\_\_→\_\_\_\_\_\_\_→\_\_\_\_\_\_\_（注：按顺序填写正确的步骤）。

①实施数据采集 ②明确数据要求 ③选择采集方法 ④确定数据来源

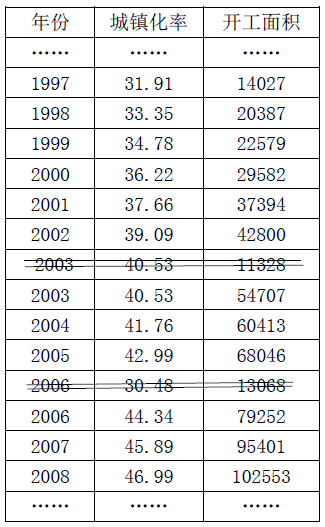
（2）2021年5月的数据显示，我国人口年龄构成方面，0至14岁人口为25338万人，占17. 95%，15至59岁人口为89438万人，占63. 35%，60岁及以上人口为26402万人，占18. 70%。为了更好地研究数据，可以运用以下哪种图表形象地呈现我国人口年龄结构？\_\_\_\_\_\_\_\_

A 拆线图 B. 柱状图 C. 饼图 D. 散点图

（3）在研究中，嘉嘉发现，城镇化率与国民经济GDP和商品住宅新开工、竣工、销售面积等都有一定关联，于是决定通过互联网获取相关数据，互联网数据采集的三个步骤分别是：获取网页、解析网页、保存数据。在获取网页的过程中，可以使用\_\_\_\_\_\_\_\_库。

A. Pandas B. BeautifulSoup C. math D. Request

（4）在采集历年中国城镇化率以及商品住宅新开工面积的相关数据之后，将其存储在”urbanization. csv”中，初步浏览数据，发现数据存在一些问题，



对图中划线数据所在记录进行删除处理的代码如下，请将划线处代码补充完整

I、第一空应该填入\_\_\_\_\_\_\_\_（单选题）

A. drop\_duplicates B. dropna C. del D. delete

II、第二空和第三空分别应填入

import pandas as pd

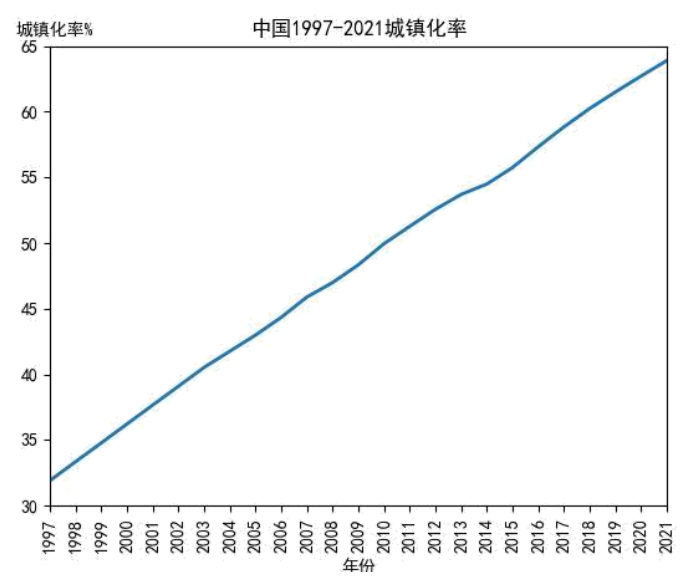
df=pd. read\_csv('urbanization. csv', encoding="ANSI")#读取数据文件

mydf=df. \_\_\_\_\_\_\_\_\_(subset=\_\_\_\_\_\_\_\_\_, keep='\_\_\_\_\_\_\_', inplace=False)

#去除年份和城镇化率重复的数据，重复数据保留最后一条，不在原表上修改

mydf. to\_csv('urbanizationnew. csv', encoding="ANSI")

（5）将整理好的数据保存在urbanizationnew. csv文件并将其可视化，如下图所示，具体代码如下，请将划线处代码补充完整：（填空）



import pandas as pd

import matplotlib. pyplot as plt

plt. rcParams['font. sans-serif']=['SimHei']

mydf=pd. read\_csv('urbanization. csv', encoding='ANSI')#读取数据文件

plt. xlim((1999, 2021))

plt. ylim((30, 65))

plt. title('中国1997-2021城镇化率')

plt. xlabel('年份')

plt. ylabel('\_\_\_\_\_\_\_\_\_', labelpad=-20, y=1. 02, rotation=0)#设置y轴标签（参照前面可视化图形设置）

kedustr=[1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021]

plt. \_\_\_\_\_\_\_\_\_\_（ ）, mydf['城镇化率'], linewidth=2)#绘制图形

plt. xticks（ ）#设置x轴刻度

plt. xticks(rotation=90)#设置x轴刻度方向

plt. savefig('中国1997-2021城镇化率. jpeg')

plt. show

（6）数据分析报告根据类型的不同，其作用也不相同，合理选择确定数据分析报告的类型对后续的编写工作至关重要，我们所要编写的分析历年城镇化进程及其影响因素的分析报告属于\_\_\_\_\_\_\_\_。

A. 综合分析报告 B. 专题分析报告 C. 日常数据通报 D. 个人数据报告

【答案】 ①. ② ②. ④ ③. ③ ④. ① ⑤. C ⑥. D ⑦. A ⑧. [‘年份’, ‘城镇化率’] ⑨. last ⑩. 城镇化率 ⑪. plot ⑫. mydf['年份'] ⑬. kedustr ⑭. B

【解析】

【详解】本题主要考查Python程序的综合应用。（1）正确的数据采集流程为：明确数据需求→确定数据来源→选择采集方法→实施数据采集。（2）饼图适用于部分与整体的占比情况，柱形图适用于数据间的比较，折线图适用于描述趋势。为了更好地研究数据，可以运用饼图形象地呈现我国人口年龄结构。（3）Pandas库用于数据分析；Beautiful Soup是python的一个库，最主要的功能是从网页抓取数据并进行解析；math库提供大量数学函数；**requests 库**就是用来发送各种请求的，用于获取网页，故本题选D选项。（4）drop\_duplicates函数主要用于删除重复记录，dropna函数用于删除缺失记录，故本题选A选项。去除年份和城镇化率重复的数据，重复数据保留最后一条，不在原表上修改，因此第二空填[‘年份’, ‘城镇化率’]，第三空填last。（5）由图可知，纵坐标标签是城镇化率，故第一空填城镇化率。绘制点和线的函数是plot，第二空填plot。横坐标是年份，故第三空填mydf['年份']。第四空设置x轴刻度，刻度保存再kedustr列表中，故第四空填kedustr。（6）专题分析报告专题分析报告是对社会经济现象的某一方面或某一个问题进行专门研究的一种数据分析报告，它的主要作用是为决策者制定某项政策、解决某个问题提供决策参考和依据。我们所要编写的分析历年城镇化进程及其影响因素的分析报告属于专题分析报告，故本题选B选项。

**综合题二：**

2. 随着社会的高速发展，每天所产生的垃圾也越来越多，为了减轻环境压力，小申所住的小区里新安装了一台具有“用户识别”“自动称重”“满溢提醒”等功能的智能垃圾回收机，如图所示。居民投递可回收物时，能够根据自动称重的重量获得相应金额。这些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）智能垃圾回收系统的“满溢提醒”功能是借助超声波传感器来获取相关数据。这种数据获取的方式属于（ ）。（单选题）

A 自动采集 B. 人工输入 C. 智能识别 D. 自动控制

（2）垃圾分类机器人在算法模型的设计过程中，通过对大量已标记样本的学习，能够快速地对可回收物进行识别。这属于机器学习中的\_\_\_\_\_\_\_\_\_\_（选填：监督学习/非监督学习）。

（3）在机器学习的过程中，可以使用欧式距离来判断回收物测试数据的类型，其中，二维平面上点P1(x1,y1)与P2(x2,y2)之间的欧式距离为，设置变量dist存放测试数据与某一类回收物均值之间的距离，则将以上欧式距离公式转化为python语句为dist=\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_（注：两点坐标为P1(x1, y1)与P2(x2, y2)，python中计算算术平方根用math库的sqrt()函数，写法为math. sqrt()）

（4）下图所示是一张可回收物图片及其属性信息。其占用的存储空间为\_\_\_\_\_\_\_\_\_\_\_\_MB。（只需填入计算结果，3分）



（5）因为需要用到大量的训练图像，为了减小存储量，可对这些图像进行压缩，对此以下说法正确的是\_\_\_\_\_\_。（单选题）

A. 将此文件拓展名改为tif属于有损压缩

B. 将此文件用WinRAR软件进行压缩属于无损压缩

C. 将此文件重命名为玻璃瓶. zip属于有损压缩

D. 将此文件另存为jpg格式属于无损压缩

（6）下图的可回收物图片采用RGB颜色模型来描述颜色，其中某像素点的颜色用十六进制数表示是E63A46，则此像素在计算内部存储的编码是\_\_\_\_\_\_。



A11011010、00110101、01011010

B.11100110、00111010、01000110

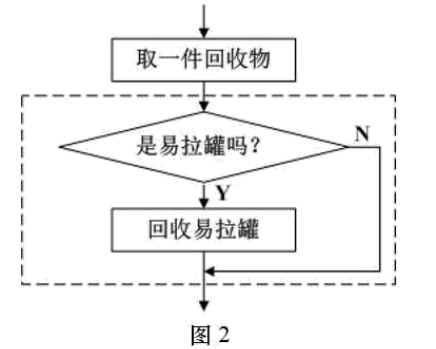
C.11001011、01010101、01011011

D.11011101、01011100、01001101

（7）智能垃圾回收机在完成回收后会播放一段音频，此音频采用wav格式，双声道立体声，采样频率为48kHz，量化位数为16bit，文件大小为4800KB，这段音频的时长是\_\_\_\_\_\_\_\_\_\_秒。（只需填入计算结果）

（8）在机器学习的过程中，有2000张各种各样的可回收物图片。某轮测试将其中80%作为训练集，20%作为测试集。测试数据中共有380张图片被正确分类，则该轮机器学习的识别正确率为\_\_\_\_\_\_\_\_\_\_。）

（9）垃圾分拣机器人判断某回收物是否为易拉罐的简要过程如图2所示，虚线框中算法的基本控制结构是\_\_\_\_\_\_\_\_\_\_。

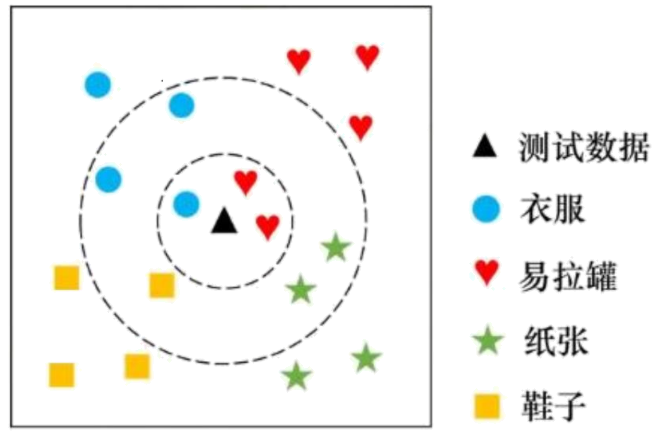


（10）为了测试识别模型的识别率，小申做了多轮实验，实验数据如表所示。

|  |  |  |  |
| --- | --- | --- | --- |
| 轮次 | 训练数据量 | 测试数据量 | 识别正确率 |
| 第1轮 | 80 | 20 | 50% |
| 第2轮 | 160 | 20 | 60% |
| …… | …… | …… | …… |
| 第N-1轮 | 1600 | 400 | 95% |
| 第N轮 | 2000 | 400 | 975% |

结合已学知识并分析以上实验数据，你可以得到的结论：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_。

（11）若用K-近邻算法对可回收物数据集的测试数据进行分类，某个测试数据▲到训练数据的距离如下图所示，当K值分别为3和8时，这个测试数据的分类结果分别为\_\_\_\_\_\_。



A. 衣服、易拉罐 B. 易拉罐、纸张 C. 衣服、纸张 D. 易拉罐、衣服

（12）目前人工智能技术已应用到生活中的很多领域，以下能够体现人工智能应用的是\_\_\_\_\_\_。

①将书本扫描成图像保存在计算机中

②用智能扫拖地机器人打扫房间

③某系统通过设置“记住密码”自动登录

④通过语音命令天猫精灵播放歌曲

⑤用某软件对身处各地的学生进行在线直播授课

⑥某软件识别用户拍摄的花卉照片并返回花卉品种等信息

A ②④⑥ B. ②③④⑤ C. ①②④⑥ D. ②③④⑤⑥

（13）关于人工智能对社会发展的影响，以下表述正确的是\_\_\_\_\_\_。（多选题）

A. 人工智能将减少人类追求知识的热情，阻碍社会发展

B. 人工智能正在改变着人类的学习和生活方式

C. 人工智能的发展应尊重隐私保护，防止数据滥用厂

D. 人工智能的应用将人类从部分脑力劳动中解放出来

【答案】 ①. A  
 ②. 监督学习 ③. math.sqrt((x1-x2)\*\*2+(y1-y2)\*\*2) ④. 2.25 ⑤. B ⑥. B ⑦. 25.6 ⑧. 95% ⑨. 分支结构 ⑩. 测试数据量相同的情况下，训练数据量越大，识别正确率越高 ⑪. D ⑫. A ⑬. BCD

【解析】

【详解】本题主要考查人工智能技术及机器学习。（1）智能垃圾回收系统的“满溢提醒”功能是借助超声波传感器来获取相关数据。这种数据获取的方式属于自动采集，故选A选项。（2）监督学习和无监督学习区分：是否有监督（supervised），就看输入数据是否有标签（label），输入数据有标签，则为有监督学习，没标签则为无监督学习。通过对大量已标记样本的学习，能够快速地对可回收物进行识别。这属于机器学习中的监督学习。（3）python中计算算术平方根用math库的sqrt()函数，写法为math. sqrt()。幂运算符是“\*\*”，因此该表达式写成Python表达式是math.sqrt((x1-x2)\*\*2+(y1-y2)\*\*2)。（4）未经压缩图像文件容量=分辨率\*颜色深度/8=1024\*768\*24/8/1024/1024=2.25MB。（5）将此文件用WinRAR软件进行压缩属于无损压缩，其余选项表述均错误，故本题选B选项。（6）十六进制数转换为二进制数：一个十六进制数分成四个二进制数，用四位二进制按权相加，最后得到二进制。E6、3A、46转换为二进制数分别是11100110、00111010、01000110，故选B选项。（7）音频文件容量=采样频率\*量化位数\*声道数\*时长/8，因此这段音频的时长是4800\*8\*1024/48000/16/2=25.6秒。（8）有2000张各种各样的可回收物图片。某轮测试将其中80%作为训练集，20%作为测试集。测试数据中共有380张图片被正确分类，则该轮机器学习的识别正确率为380/(2000\*20%)=95%。（9）由图可知，虚线框中算法的基本控制结构是分支结构。（10）由表可知，得到的结论是：测试数据量相同的情况下，训练数据量越大，识别正确率越高。（11）当k=3时，易拉罐出现次数最多，对应的分类结果是易拉罐。当k=8时，衣服出现次数最多，因此对应的分类结果是衣服，故本题选D选项。（12）人工智能研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。②用智能扫拖地机器人打扫房间④通过语音命令天猫精灵播放歌曲⑥某软件识别用户拍摄的花卉照片并返回花卉品种等信息均体现了人工智能技术，故本题选A选项。（13）人工智能有利有弊，但总体弊大于利，会推动社会的发展；人工智能正在改变着人类的学习和生活方式；人工智能的发展应尊重隐私保护，防止数据滥用厂；人工智能的应用将人类从部分脑力劳动中解放出来，故本题选BCD选项。

**综合题三：**

3. 疫情的传播和现代信息技术的迅速发展，在一定程度上改变着人们的学习方式与学习习惯，移动学习（Mobile Learning）是一种在移动设备帮助下的能够在任何时间、任何地点发生的学习。随着智能手机的普及使用，越来越多的人都习惯在手机上安装学习软件，利用碎片化的时间背单词、听英语、上网课。某学习APP为了激励用户每天登录平台进行学习，推出了“打卡送积分，积分兑好书”的活动。积分赠送规则如下：

若用户每天不间断登录平台，打卡参加学习，则第1天可获得1点积分，之后2天（第2天和第3天）里，每天可获得2点积分，之后3天（第4、5、6天）里，每天可获得3点积分…这种模式会一直持续下去，当连续n天每天获得n点积分后，用户在接下来的n+1天里，每天可获得n+1点积分（n为正整数）。如果中间某天没有打卡，则积分规则重置，从下一个连续打卡的日期开始重新累计计算。



小言从第1天开始参加活动，他想知道经过连续x天的打卡之后，一共可以获得多少点积分，请帮助小言设计算法并编写程序解决问题。

一、抽象与建模

（1）若将每段积分点数相同的日期视为一个周期，记第1天为第1个周期，第2、3天为第2个周期，第4、5、6天为第3个周期，以此类推，则第k周期有\_\_\_\_\_\_\_\_\_\_天，每天可获得\_\_\_\_\_\_\_\_\_\_点积分。

（2）设变量x表示连续打卡的天数，变量k表示某天恰好处于第k个周期内，total表示累计获得的积分总数，则该算法的输入为\_\_\_\_\_\_\_\_\_\_，输出为\_\_\_\_\_\_\_\_\_\_。（填写变量名，可参考下一页流程图进行分析）

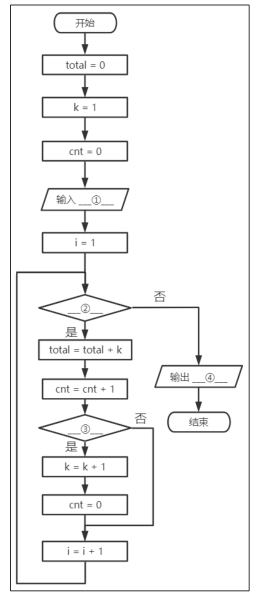
（3）为了计算total，需要知道每天可获得的积分点数，而积分点数与当前所处的周期k的值相同，且会持续k天，因此小言使用变量cnt来记录每个周期内累计积分的天数。每个周期开始时将cnt设置为0，每当累加了一天的积分，令cnt的值增加1，则在第k个周期内，当cnt=\_\_\_\_\_\_\_\_\_\_时，表示该周期内的所有天数的积分全部累加完毕，下一天需进入下一个周期，下一个周期每天可获得\_\_\_\_\_\_\_\_\_\_点积分。（可参考下一页流程图进行分析）

（4）小言设计算法从第1天开始依次遍历到第x天，计算每天可获得的积分点数，并将该积分点数累加到total中求和，即可得出答案。算法中从第1天遍历至第x天，可以使用\_\_\_\_\_\_\_\_\_\_结构快速实现。

A. 顺序 B. 分支 C. 判断 D. 循环

二、设计算法

（5）小言使用流程图描述他所设计的算法，根据算法设计的思路将如下流程图补充完整。



①\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_；②\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_。③\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_；④\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_。

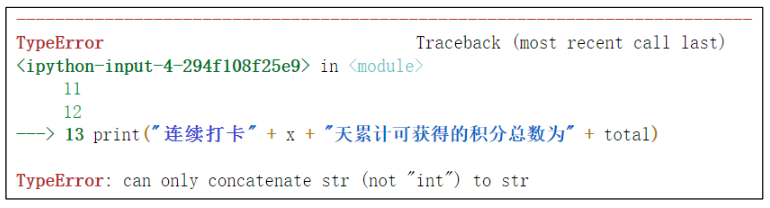
三、编写程序

（6）小言使用Python语言编写程序，请在以下横线上程序段补充完整。

调试运行

|  |
| --- |
| total = 0  k = 1  cnt = 0  x = int(input("请输入连续打卡的天数 x：")  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ # 使用 for 循环遍历第 1 至第 x 天  total = total + k  cnt = cnt + 1  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ # 判断是否进入下一个周期，并更新变量值  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  print("连续打卡" + x + "天累计可获得的积分总数为" + total) |

（7）小言编写程序运行时发现程序报错，根据下图所示的信息分析报错原因，并阐述如何将程序修改正确\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_。



【答案】 ①. k ②. k ③. x ④. total ⑤. k ⑥. k+1 ⑦. D ⑧. 连续打卡的天数x ⑨. i<=x? ⑩. cnt==k? ⑪. 连续打卡x天累计可获得的积分总数为total ⑫. for i in range(1,x+1): ⑬. if cnt==k: ⑭. k=k+1 ⑮. cnt=0 ⑯. 报错原因：整型数据不能直接和字符型数据连接 修改：print("连续打卡"+str(x)+"天累计可获得的积分总数为"+str(total))

【解析】

【详解】本题主要考查Python程序的综合应用。（1）若将每段积分点数相同的日期视为一个周期，记第1天为第1个周期，第2、3天为第2个周期，第4、5、6天为第3个周期，以此类推，则第k周期有k天，每天可获得k点积分。（2）由流程图可知，该算法的输入为x，输出为total。（3）每个周期开始时将cnt设置为0，每当累加了一天的积分，令cnt的值增加1，则在第k个周期内，当cnt=k时，表示该周期内的所有天数的积分全部累加完毕，下一天需进入下一个周期，下一个周期每天可获得k+1点积分。（4）算法中从第1天遍历至第x天，可以使用循环结构快速实现，故选D选项。（5）第一空是输入连续打卡的天数x。第二空是总的循环条件，即i<=x时进入循环，计算积分。第三空处理当cnt=k时，表示该周期内的所有天数的积分全部累加完毕，下一天需进入下一个周期，下一个周期每天可获得k+1点积分，故填cnt==k？。第四空是输出连续打卡x天累计可获得的积分总数为total。（6）第一空使用 for 循环遍历第 1 至第 x 天，循环变量是i。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)。因此填for i in range(1,x+1):。第二空、第三空、第四空判断是否进入下一个周期，并更新变量值，对照流程图可知，依次填if cnt==k:、k=k+1、cnt=0。（7）分析报错原因，可知报错原因：整型数据不能直接和字符型数据连接，应修改：print("连续打卡"+str(x)+"天累计可获得的积分总数为"+str(total))。

