**【数据处理实习生】笔试题**

|  |
| --- |
| 以下为笔试题目，需要个人独立完成，将代码粘贴到指定文档中 |

**题目1**

1. 要求：编写1个函数，输入是字符串文本（字符串，变量名固定为 text），实现统计文本中每个字符的出现次数，并按照出现次数从高到低排序输出（列表，变量名固定为 result）。区分大小写
2. 示例：

|  |
| --- |
| Python from collections import Counter  def question\_1(text):  counter = Counter(text)  result = sorted(counter.items(), key=lambda x: x[1], reverse=True)  return result  #text = "Hello, World!"  text = "sdasaa"  print(question\_1(text)) 示例： 输入： text = "sdasaa"  输出： result = [('a', 3), ('s', 2), ('d', 1)] |

**题目2**

1. 要求：编写一个函数，输入1段话（字符串，变量名固定为 text），和1个词汇（字符串，变量名固定为 word），输出该词汇在文本中出现的次数（整数，变量名固定为 count），和每次出现时，词汇第1个字符的下标index（列表，变量名固定为 index\_list）。区分大小写
2. 示例：

|  |
| --- |
| Python  def question\_2(text, word):  count = 0  index\_list = []  index = text.find(word)  while index != -1:  count += 1  index\_list.append(index)  index = text.find(word, index + 1)  return count, index\_list  text = "我买了华为的手机，因为华为手机的芯片是自研的"  word = "手机"  count, index\_list = question\_2(text, word)  print("count:", count)  print("index list:", index\_list)   输入： text = "我买了华为的手机，因为华为手机的芯片是自研的" word = "手机"  输出： count = 2 index\_list = [6, 13] |

**题目3**

1. 要求：编写1个函数，输入是 pandas 的 DataFrame（变量名固定为 dataset），筛选 Excel 中 语文大于90，数学小于120，英语在70-110之间的学生，函数返回 学生列表（列表，变量名固定为 students）
2. 示例：object-1-1.xlsx

|  |
| --- |
| Python  import pandas as pd  def question\_3(dataset):  condition = (dataset['语文'] > 90) & (dataset['数学'] < 120) & (dataset['英语'] >= 70) & (dataset['英语'] <= 110)  students = dataset[condition]['学生姓名']  return students  dataset = pd.read\_excel('object-1-1.xlsx')  students = question\_3(dataset)  print(students.tolist())  输入： dataset：用 pandas 读取 object-1-1.xlsx  输出： students = ["小明", "小特"] |

**题目4**

1. 要求：编写1个函数，输入是2个 pandas 的 DataFrame（变量名固定为 dataset\_1 和 dataset\_2），合并两个DataFrame 为1个新的 DataFrame，使得每一行都可以查看1位同学的4科成绩。输出新的 DataFrame（变量名固定为 new\_set）
2. 示例：object-1-2-1.xlsx object-1-2-2.xlsx

|  |
| --- |
| Python import pandas as pd  def question\_4(df1, df2):  new\_set = pd.merge(df1, df2, on='学生姓名')  return new\_set  df1 = pd.read\_excel("object-1-2-1.xlsx")  df2 = pd.read\_excel("object-1-2-2.xlsx")  merged\_df = question\_4(df1, df2)  print(merged\_df) 输入： dataset\_1：用 pandas 读取 object-1-2-1.xlsx dataset\_2：用 pandas 读取 object-1-2-2.xlsx  输出： new\_set：DataFrame，数据见下表 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **学生姓名** | **语文** | **数学** | **英语** | **物理** |
| 小明 | 92 | 119 | 120 | 92 |
| 小红 | 80 | 108 | 93 | 123 |
| 小张 | 108 | 146 | 92 | 146 |
| 小文 | 125 | 86 | 152 | 87 |
| 小新 | 88 | 92 | 88 | 90 |
| 小葵 | 90 | 120 | 89 | 27 |
| 小特 | 99 | 67 | 90 | 102 |

**题目5**

1. 要求：输入年（整数，变量名固定为 year）、月（整数，变量名固定为 month）、日（整数，变量名固定为 day），计算这一日期是这年的第几天，输出天数（整数，变量名固定为 day\_count）
2. 示例：

|  |
| --- |
| Python from datetime import datetime  def question\_5(year, month, day):  date = datetime(year, month, day)  day\_count = date.timetuple().tm\_yday  return day\_count  year = 2024  month = 2  day = 2  day\_count = question\_5(year, month, day)  print("day\_count:", day\_count) 输入： year = 2024 month = 2 day = 2  输出： day\_count = 33 |

**题目6**

1. 要求：编写1个函数，输入是字符串文本（字符串，变量名固定为 text），找出文本中**连续**出现最多次的字符和其最长连续出现次数，输出字符（str，变量名固定为 char）和出现次数（int，变量名固定为 count）
2. 示例：

|  |
| --- |
| Python  def question\_6(text):  max\_count = 0  max\_char = ''  prev\_char = ''  count = 0  for char in text:  if char == prev\_char:  count += 1  else:  count = 1  if count > max\_count:  max\_count = count  max\_char = char  prev\_char = char  return max\_char, max\_count  text = "165896514101ddnidndngooooooeur15541"  char, count = question\_6(text)  print("char:", char)  print("count:", count) 输入： text = "165896514101ddnidndngooooooeur15541"  输出： char = "o" count = 6 |