

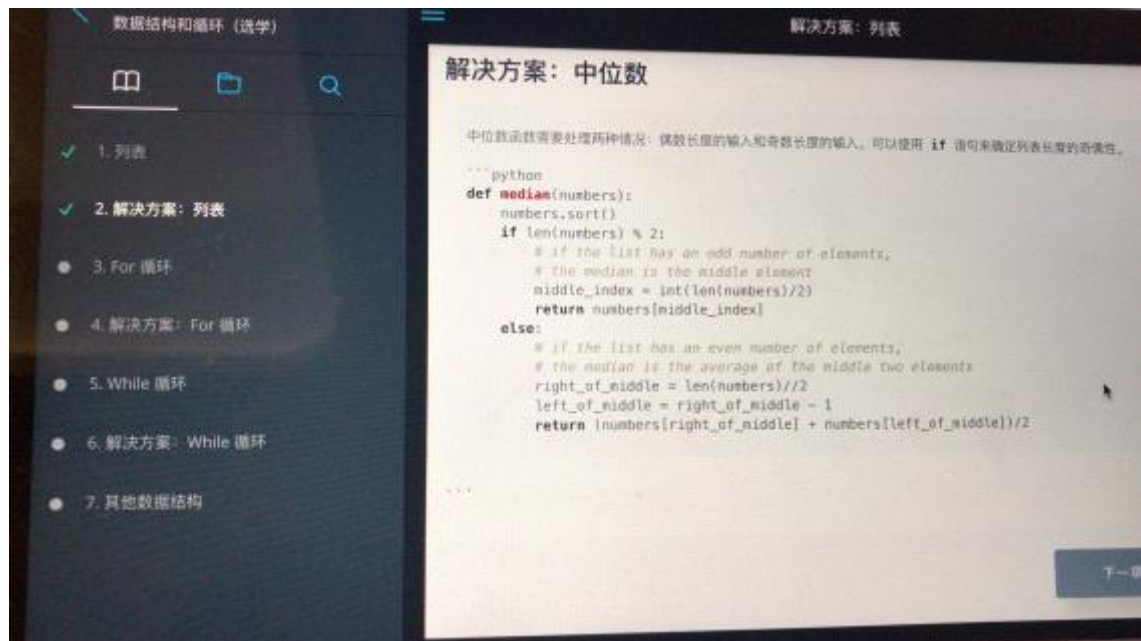
入门班级问题汇总

2018年5月25日星期五 13:59

注：黑色字体为学员提问，红色字体为助教回答。

学员问题1：

有关于中位数
习题



第一个if那里是不是应该有个!=0...

就是if len () %2!=0

!=0？

是想说不等于零的意思么？

啊对

嗯嗯这个函数好像没有问题的

越夫的意思是它没考虑到长度是0的情况么？

因为要是没有输入数据的话就不存在找中位数这一说

所以我觉得可以不用考虑不等于0的情况

要是考虑进去也很完美，返回一个错误是最好不过的了

问题2：

我的每日一题问到查看一列数据类型我觉得可以用.info函数，在结果中找到对应列,后面就会有int float or object显示；或者用.dtypes，比如.dtypes['hour']这样就只显示这一列的类型。

注意代码的中
文符号错误问
题

对哒两种都可以，一般要是列不是特别多我们会用info

要是列比较多pandas会折叠，就用type

然后又可能因为xx这边是用手机打的，记得敲代码的时候注意中文的引号是不能识别的哦

我还想问一下如果自己下载了anaconda安装了jupyter notebook在自己电脑上 像numpy panda等这些包是不是就自动安装上了？用的时候就只需import就可以了？

对的，jupyter notebook直接吧常用的包打包安排好了哦

直接调用就可以了

jupyter
notebook安装
问题

那比如要读取一个数据文件 如果是ipynb和数据文件在一个文件夹下面 就可以通过数据文件名称直接读取那个数据文件，不然就要写清楚路径对吧？

jupyter notebook安装问题

那比如要读取一个数据文件 如果是ipynb和数据文件在一个文件夹下面 就可以通过数据文件名称直接读取那个数据文件，不然就要写清楚路径对吧？

只要你的数据集跟ipynb在一个文件夹，导入数据的时候就不需要写路径，只是不同类型的数据集导入的方式不同，比如excel就不能用df.read_csv了，后面入门正式课程会详细介绍的。

问题3：

拖拽回答问题

习题 1/2

请确定下面的数据类型为数值还是分类？

分类

分类

数值

数值

数值

| 变量 | 数据类型 |
|------------------|------|
| 邮政编码 | |
| 年龄 | |
| 收入 | |
| 婚姻状况 (单身、已婚、离异等) | |

请问这个是怎么做的

我点击都没反应

要拖拽哦

你试试拖拽呢

点中蓝框框拖到相应位置试试呢

问题4：

练习：销售总额

在下面这道练习中，你需要更改输入和输出数据的类型来获得你希望的结果。

请根据提供的数据计算并打印一周的销售总额。你需要打印一个格式为，`"This week's total sales: xxx"` 的字符串中，其中 xxx 将是实际销售总额的数字。你需要更改输入数据的类型才可以计算销售总额。

销售总额练习问题

```
1 mon_sales = "121"
2 tues_sales = "105"
3 wed_sales = "110"
4 thurs_sales = "98"
5 fri_sales = "95"
6
7 #TODO: Print a string with this format: This week's total sales: xxx
8 # You will probably need to write several lines of code before the print statement.
9 a=float(mon_sales)+float(tues_sales)+float(wed_sales)+float(thurs_sales)+float(fri_sales)
10
11 print("This week's total sales:",str(a))
```

That doesn't match my answer. Check that your string is formatted correctly and that you calculated the right total.

重设练习

测试答案

提交答案

请问这个是什么原因测试不通过

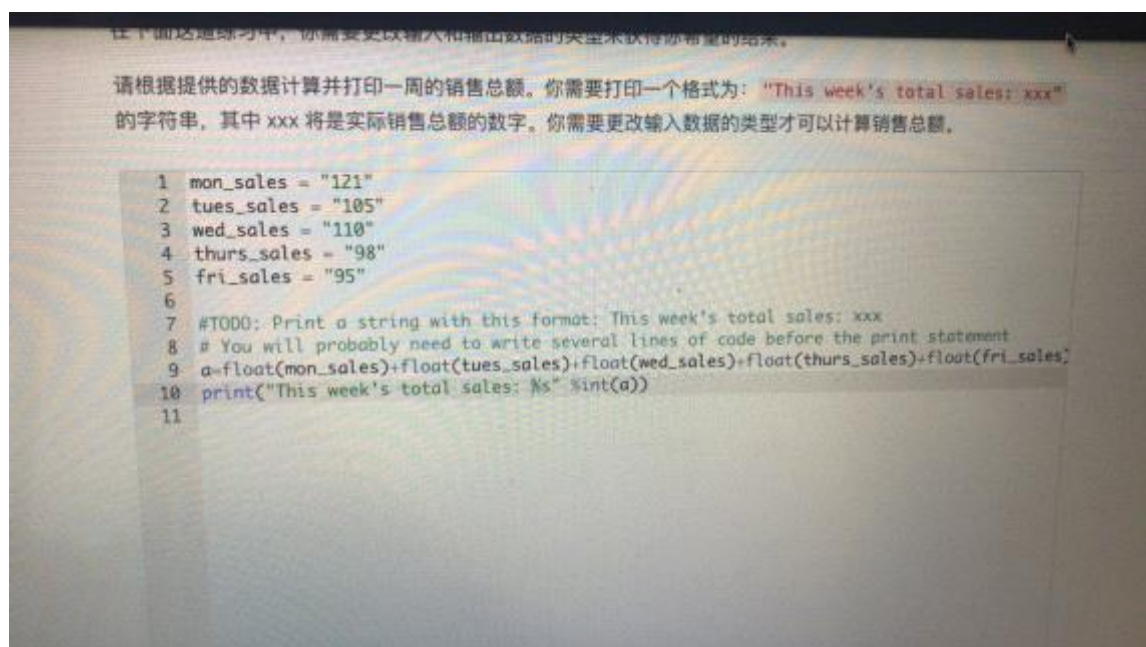
是这样字

它这个可能设计的有点死板，你必须一点不差的对上去才行

"This week's total sales: xxx"

首先输出的一定是没有.0的也就是说只有前三位数字

其次冒号后面有一个空格，不知道徐海看到了嘛，这个也是需要有的



第一张是测试给的标准答案，第二个是我以为它想要求用格式化输出，其实不用，你参考第一张图就行

问题5：

import含义

```
In [1]: import csv
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn
%matplotlib inline
```

in[1] 是input的意思吧

python使用import来导入需要使用的模块哦~

<http://www.iplaypy.com/jinjie/import.html>

可以百度查看一下不太懂的语法

问题6：

练习：列表的总和

请在下面的练习中定义一个函数 `list_sum`，该函数将一个列表作为参数，并返回列表中元素的总和。你可以使用 `for` 循环迭代列表。

```
1- def list_sum(input_list):
2-     sum = 0
3-     for element in input_list:
4-         sum += element
5-     return sum
6-
7- #These test cases check that list_sum works correctly
8- test1 = list_sum([1, 2, 3])
9- print("expected result: 6, actual result: {}".format(test1))
10-
11- test2 = list_sum([-1, 0, 1])
12- print("expected result: 0, actual result: {}".format(test2))
```

列表总和练习
问题

为什么要有两种result？

你是问为什么测试两次么？

这个是测试用的哦~

它是为了多验证几次

看你的函数是否正确

那actual result 后面为啥是用{}

这个是.format的用法哦

<https://www.digitalocean.com/community/tutorials/how-to-use-string-formatters-in-python-3>

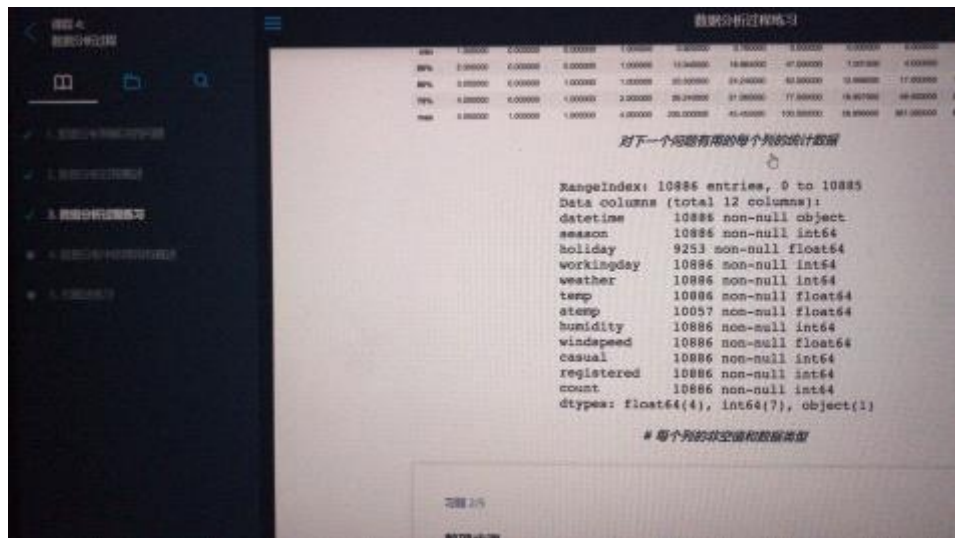
可以参考这篇文章

问题7：

课程4的练习题

旭龙老师，请问“课程4-3”的练习是需要另外学习相关知识点的吧？我看了几遍视频里面是没有具体介绍的。就是课程4里第三个步骤，数据分析过程练习

练习题是吧，方便截图给我看下么

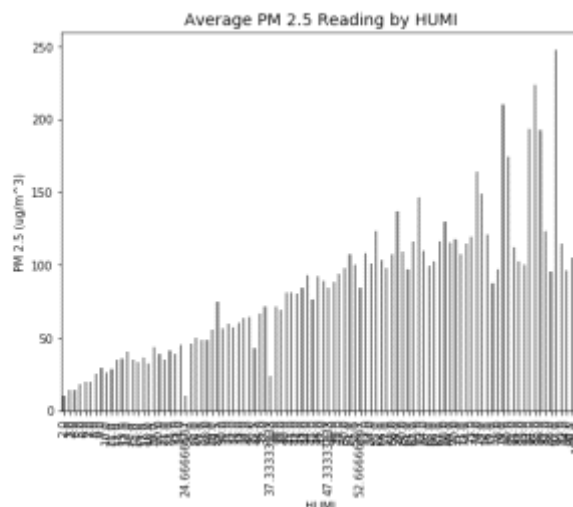


语音电话内容：这个练习是对一个美国共享单车的项目进行的练习提问，问题步骤是按照前面课程来的，提问步骤，整理，探索，得出结论，传达结论。学员们可以按照自己对每一步的理解还有项目的理解吓出自己的答案，课程只有五个步骤的含义，没有介绍这个项目哦，要是需要参考答案的话，可以问助教取阅。

问题8：

```
In [105]: # TO DO:
univariate_plot(df_sxt, 'HUMI', 'grey')
```

做可视化注意
数据类型适合
什么样的可视
化图形



这个横坐标可以调整一下间距之类的吗？

你研究humi最好用散点图哦，因为HUMI是连续变量，柱状图是用来展示分类变量的哦，连续变量可以使用散点图或者折线图。

问题9：

昨天中位数那个课程习题有个问题：

```

def median(numbers):
    numbers.sort()
    if len(numbers) % 2:
        middle_index = int(len(numbers)/2)
        return numbers[middle_index]
    else:
        a=int(len(numbers)/2)
        return (numbers[a-1]+numbers[a])/2

```

中位数练习题

为什么if len(numbers) % 2:后面没有>0？

首先建议可以先百度一下if条件语句的用法哦，if条件语句的核心就是if后面接的是为TRUE的内容，要是运行结果为False就会跳到else中去，所以在这个里面，取了%2有数、也就是奇数的情况下，就相当于TRUE所以这个>0是可以省略的，当然你留下>0是没有问题的，只不过省略掉更加简洁。

len()出来的结果不是已经是int类型了吗，为什么还要再外面套一层int？而且我试了一下删掉了的话还会报错。

```
]]: 1 4/2  
]: 2.0
```

python只要是除法一定是精确的哦，就是还是要取int的。

问题10：

与降雨相关的
项目问题需要
注意是否限定
降雨量



能够通过代码成功筛选数据，筛选出的数据能够回答提出的两个问题。

你的第一个问题是：2010-2015年北京PM2.5浓度与降水量/时的关系，"precipitation = 0"也应该算上呀

问题3b: 请使用上面给出的reading_stats函数来调用你所需要的数据，请在下面填写合适的条件（conditions）。提示：对应此函数的第二个参数。

答案:

第一个问题: df1 = reading_stats(df_all_cities,["city == 'Beijing'", "precipitation > 0"])

第二个问题: df2 = reading_stats(df_all_cities,["city == 'Beijing'", "year == 2015", "month == 3", "day == 25"])

降水量每时把0加进去不就把所有数据都取进去了么？

审阅的意思是你没取到等于=0的数据哦

把=0加上不就是北京五年的所有数据了么，他的意思是没有降水和有降水也要对比因为你的问题是浓度与降雨量/时的关系，而不是有没有下雨的关系哇

你的第一个问题是：2010-2015年北京PM2.5浓度与降水量/时的关系，"precipitation = 0"也应该算上呀

问题3b: 请使用上面给出的reading_stats函数来调用你所需要的数据，请在下面填写合适的条件（conditions）。提示：对应此函数的第二个参数。

答案:

第一个问题: df1 = reading_stats(df_all_cities,["city == 'Beijing'", "precipitation > 0"])

第二个问题: df2 = reading_stats(df_all_cities,["city == 'Beijing'", "year == 2015", "month == 3", "day == 25"])

你要是研究的是下雨与pm2.5浓度的关系就可以只去>0的明白了。。概念有偏差。。