**实验3 文件操作与数据存取**

**姓名：宁浩 学号：2312190617 班级：软件2301 成绩：**

**1. 实验目标**

本次实验旨在了解Python的文件数据操作方法，理解和掌握数据存取的各种方式。需掌握的相关知识点主要包括：

1) 文件数据读写：掌握文件的打开(open)以及打开模式（’r’,’w’和’a’）、读（read、readline、readlines）、写（write、writelines）方式，并学会with …. as句法，以及数据类型的转化、格式化处理和保存。

2) 文件管理：学会应用python的os、os.path、shutil等模块进行文件批量管理。

3) 数据永久存储：pickle模块的dump和load方法。注意此时文件对象的打开模式应为二进制形式(‘rb’或‘wb’).

请完成以下实验练习题，报告书写要求同第1次实验作业。注意，请在本报告中将自己的姓名、学号、班级书写正确。最后，请在规定时间内完成本次实验，并将报告的word文档(非pdf版本)以附件形式提交至学习通平台。

**2. 实验练习题**

**答题要求**：将正确运行的**源代码书写在题目下方**，并紧跟着贴上一份**源代码的截图**和运行结果的**截图**。答题格式同第1次实验作业。另外注意：如题目要求保存结果数据到文件，则该文件中的数据也算作运行结果，需截图粘贴。但提交报告时无需提供数据文件。

(1)文本文件“info\_stocks.txt”中每一行的name对应股票名称，shares对应多少股，price对应该只股票的单价。

(a) 请计算每只股票的总价，并按总价从大到小进行排序(程序中要求含lambda函数) ，将所有股票的名称、单价、总价依次打印输出。

(b) 要求使用Python内置函数filter过滤出单价大于80的股票有哪些? 并将符合条件的股票名称、单价、总价打印输出。

源代码：

(a)

import re

list1 = []

with open('C:\\Users\\12048\\Desktop\\python\_homework\\实验3 文件操作与数据读取\\info\_stocks.txt') as file:

lines = file.readline().replace('\'', '').replace(' ','').replace('\n','')

while lines !="":

line=re.split('[,:]',lines)

list1.append({

line[0]: line[1],

line[4]: line[5],

"total": float(line[3]) \* float(line[5])

})

lines = file.readline().replace('\'', '').replace(' ','').replace('\n','')

list1.sort(key=lambda x:x['total'],reverse=True)

print(list1)

for item in list1:

print(item)

(b)

import re

list1 = []

with open('C:\\Users\\12048\\Desktop\\python\_homework\\实验3 文件操作与数据读取\\info\_stocks.txt') as file:

lines = file.readline().replace('\'', '').replace(' ','').replace('\n','')

while lines !="":

line=re.split('[,:]',lines)

list1.append({

line[0]: line[1],

line[4]: line[5],

"total": float(line[3]) \* float(line[5])

})

lines = file.readline().replace('\'', '').replace(' ','').replace('\n','')

list1.sort(key=lambda x:x['total'],reverse=True)

list2=filter(lambda x:float(x['price'])>80,list1)

for i in list2:

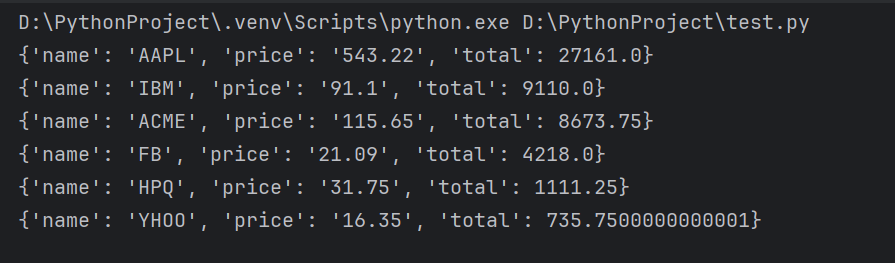
print(i)

源代码截图：

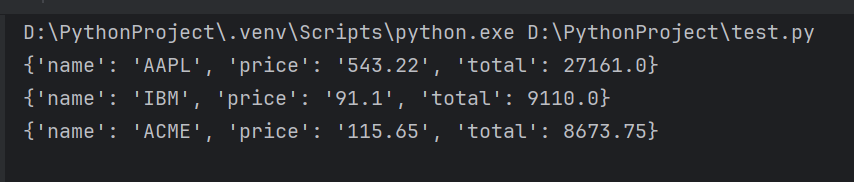




(a)运行结果截图：



(b)运行结果截图：



(2)创建文件，使用shutil模块中的copyfile()函数复制该文件，获取复制文件的创建时间。

源代码：

import os

import shutil

from os import mkdir

from datetime import datetime

src=r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\secret.txt'

dst=r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\secret\_copy.txt'

shutil.copyfile(src,dst)

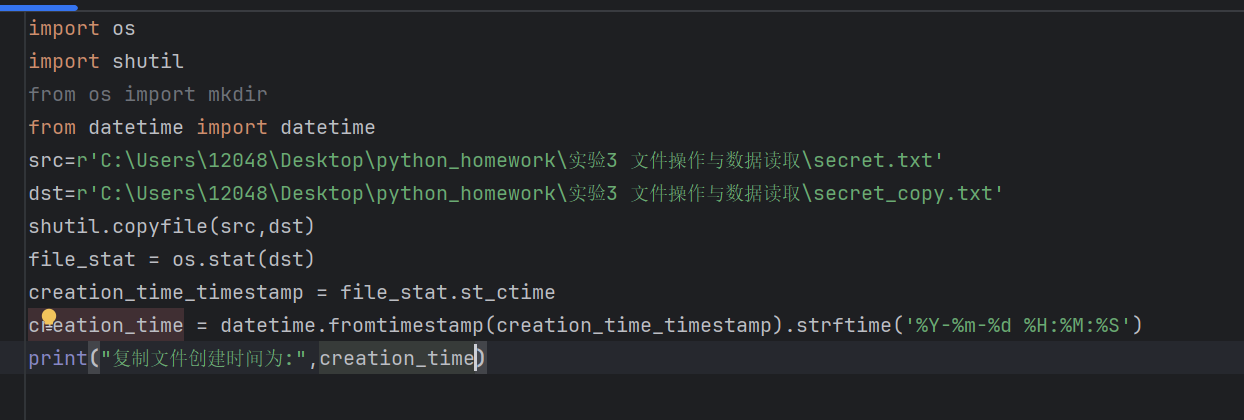
file\_stat = os.stat(dst)

creation\_time\_timestamp = file\_stat.st\_ctime

creation\_time = datetime.fromtimestamp(creation\_time\_timestamp).strftime('%Y-%m-%d %H:%M:%S')

print("复制文件创建时间为:",creation\_time)

源代码截图：



运行结果截图：



(3)对8~1000之间的每个正整数进行以下分析：  
①计算该数字的二进制表示  
②统计二进制中1的个数  
③判断是否为2的幂次方数

文件存储：将分析结果保存到"binary\_analysis.txt"文件中

文件内容格式可参考如下：

8 -> 1000 | 1的个数: 1 | 是2的幂次方

9 -> 1001 | 1的个数: 2 | 不是2的幂次方

...

源代码：

def save(dst,string):

with open(dst,'a',encoding='utf-8') as f:

f.write(string)

def binary(n):

result=""

while n>0:

result=str(n%2)+result

n=n//2

return result

def pd(n):

if (n&(n - 1)) == 0:

return True

else:

return False

def handler(src):

for num in range(8,1001):

b=binary(num)

sum1=0

for i in b:

if int(i)==1:

sum1=sum1+1

if pd(num):

save(src,str(num)+"->"+b+"| 1的个数:"+str(sum1)+"| 是2的幂次方\n")

else:

save(src,str(num)+"->"+b+"| 1的个数:"+str(sum1)+"| 不是2的幂次方\n")

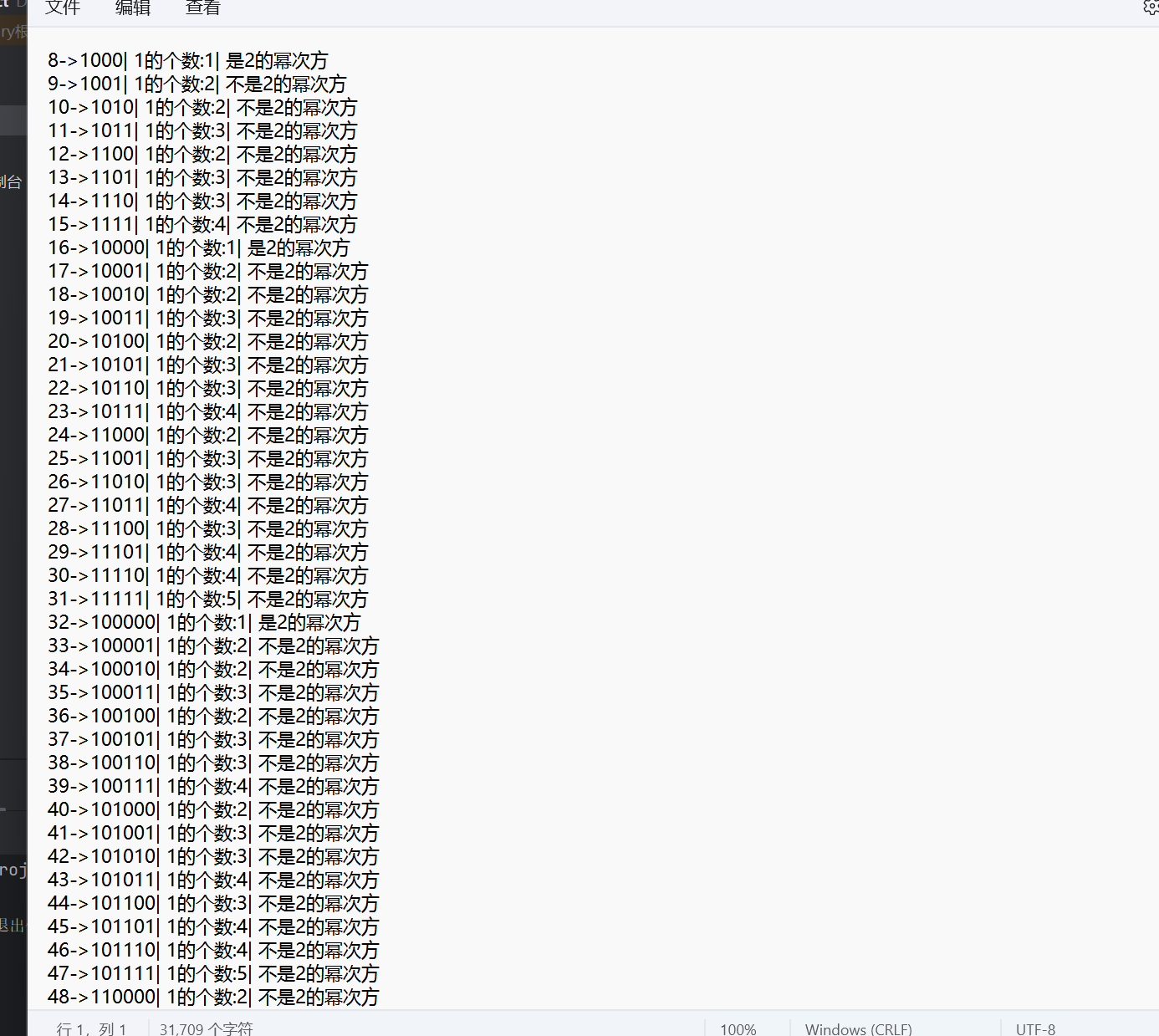
src =r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\binary\_analysis.txt'

handler(src)

源代码截图：



文件内容截图：



1. 判断当前路径中是否有oldfilename.txt文件，如果存在则按行输出内容，否则创建该文件。  
   源代码：

import os.path

if not os.path.exists(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\oldfilename.txt'):

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\oldfilename.txt', 'w') as f:

print("不存在oldfilename.txt，现已经创建")

else:

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\oldfilename.txt','r') as f:

while True:

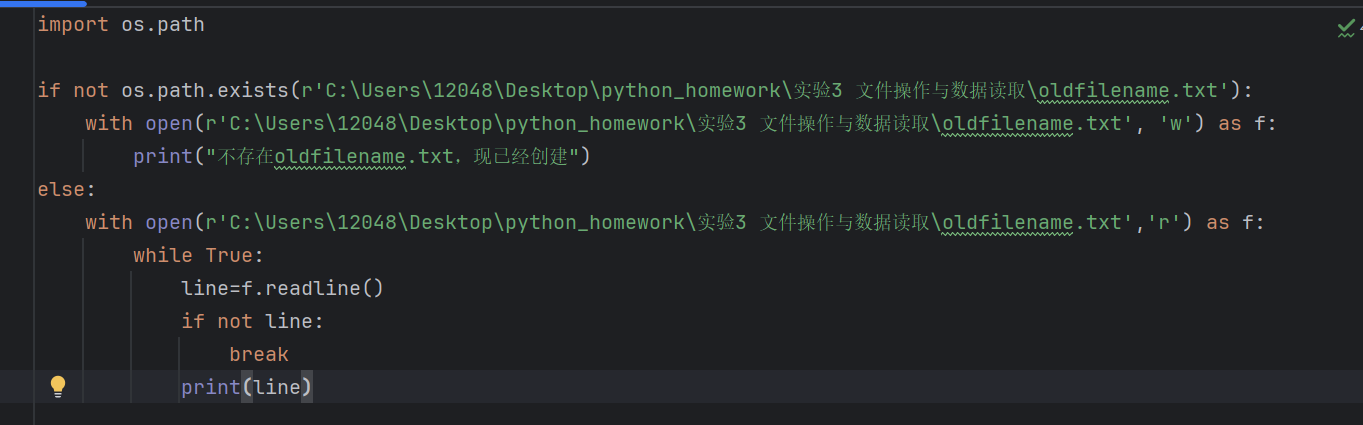
line=f.readline()

if not line:

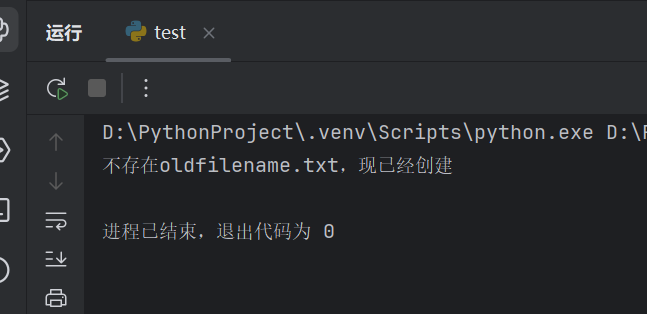
break

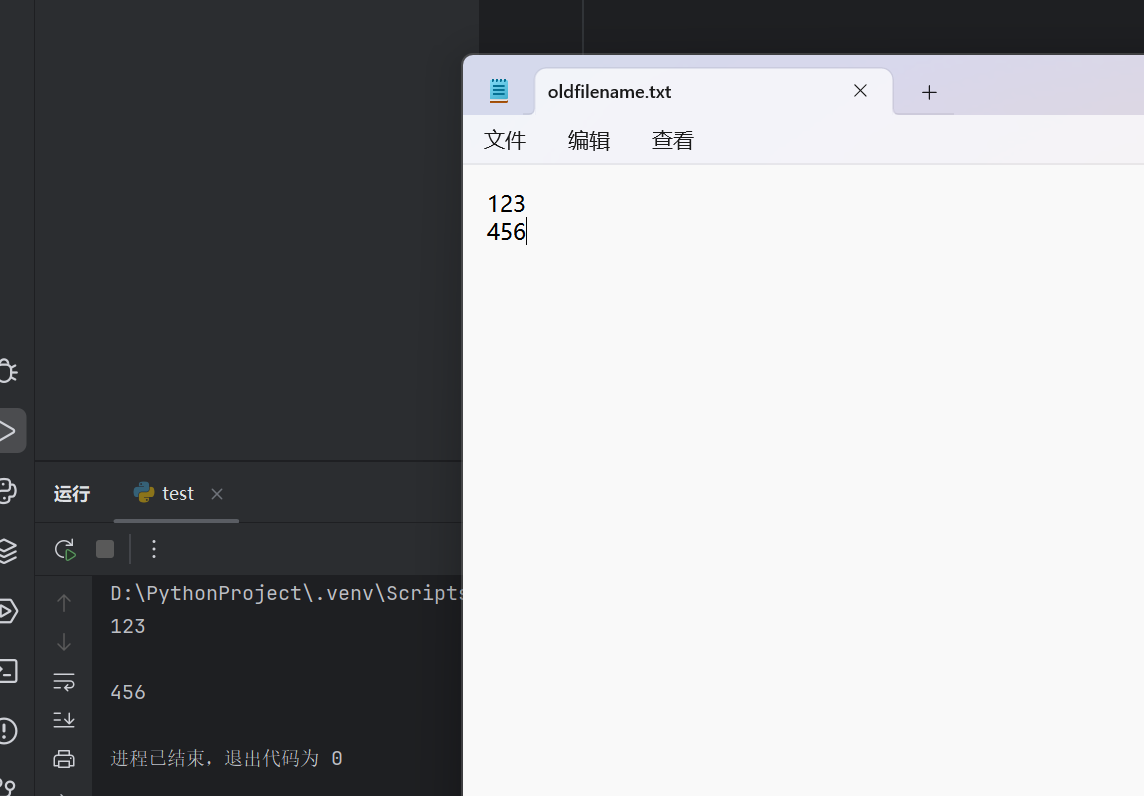
print(line)

源代码截图：

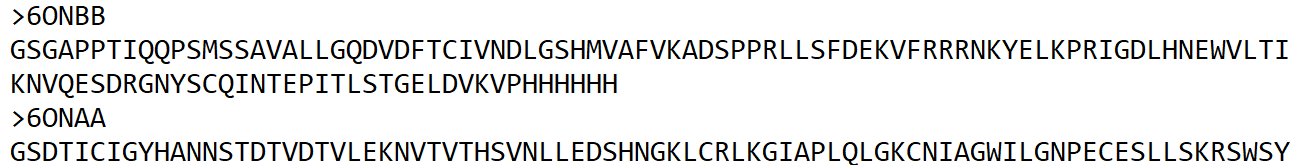


运行结果截图：





(5)文件” seqs\_fasta.txt”保存有符合fasta格式的100多个蛋白质氨基酸序列，截取部分如下图所示：

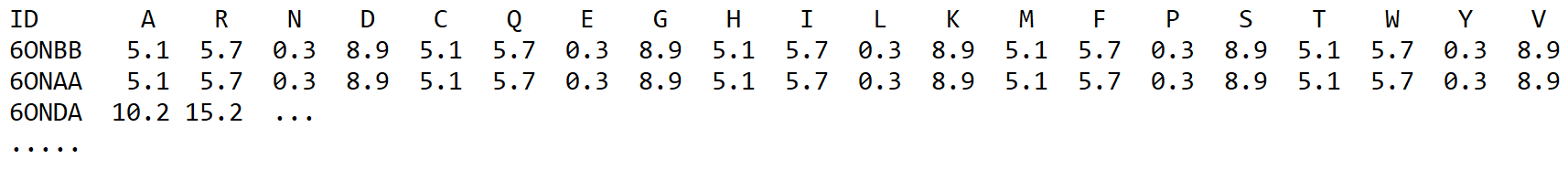


其中每条序列以“>”开头，紧接的“6ONBB”表示蛋白质序列ID，两行“>6ONBB”与“>6ONAA”之间展示的是ID为6ONBB的氨基酸序列。

自然界中，任何天然蛋白质主要由20种标准的氨基酸构成，可与字母表aa20={'A', 'R', 'N', 'D', 'C', 'Q', 'E', 'G', 'H', 'I', 'L', 'K', 'M', 'F', 'P', 'S', 'T', 'W', 'Y', 'V'}中的英文字母一一对应。同一种氨基酸在一条蛋白质序列中可能出现多次。

请计算文件“seqs\_fasta.txt”中每条蛋白质序列的氨基酸组分，即字母表aa20中的每个英文字母在该蛋白质序列中出现的频率（百分比%），并将统计结果保存至一个新的文件（创建的新文件可取名为’result6.txt’）。

注：要求氨基酸出现的频率值保留小数点后1位，且每个氨基酸所在列的频率值数据右对齐保存，保存文件的格式参考如下：



源代码：

aa20=['A', 'R', 'N', 'D', 'C', 'Q', 'E', 'G', 'H', 'I', 'L', 'K', 'M', 'F', 'P', 'S', 'T', 'W', 'Y', 'V']

b=[0]\*26

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\seqs\_fasta.txt','r') as f:

file=f.read().split('>')

list1=[]

with open(r"C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\result6.txt", "a") as f:

f.write(format('ID',' <10'))

for obj in aa20:

f.write(format(obj,' >10'))

f.write('\n')

for content in file:

if content=='':

continue

else:

list1.append([content[0:6].replace('\n',''),content[6:].replace('\n','')])

for item in list1:

for char in item[1]:

if char in aa20:

b[ord(char)-65]+=1

with open(r"C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\result6.txt","a") as f:

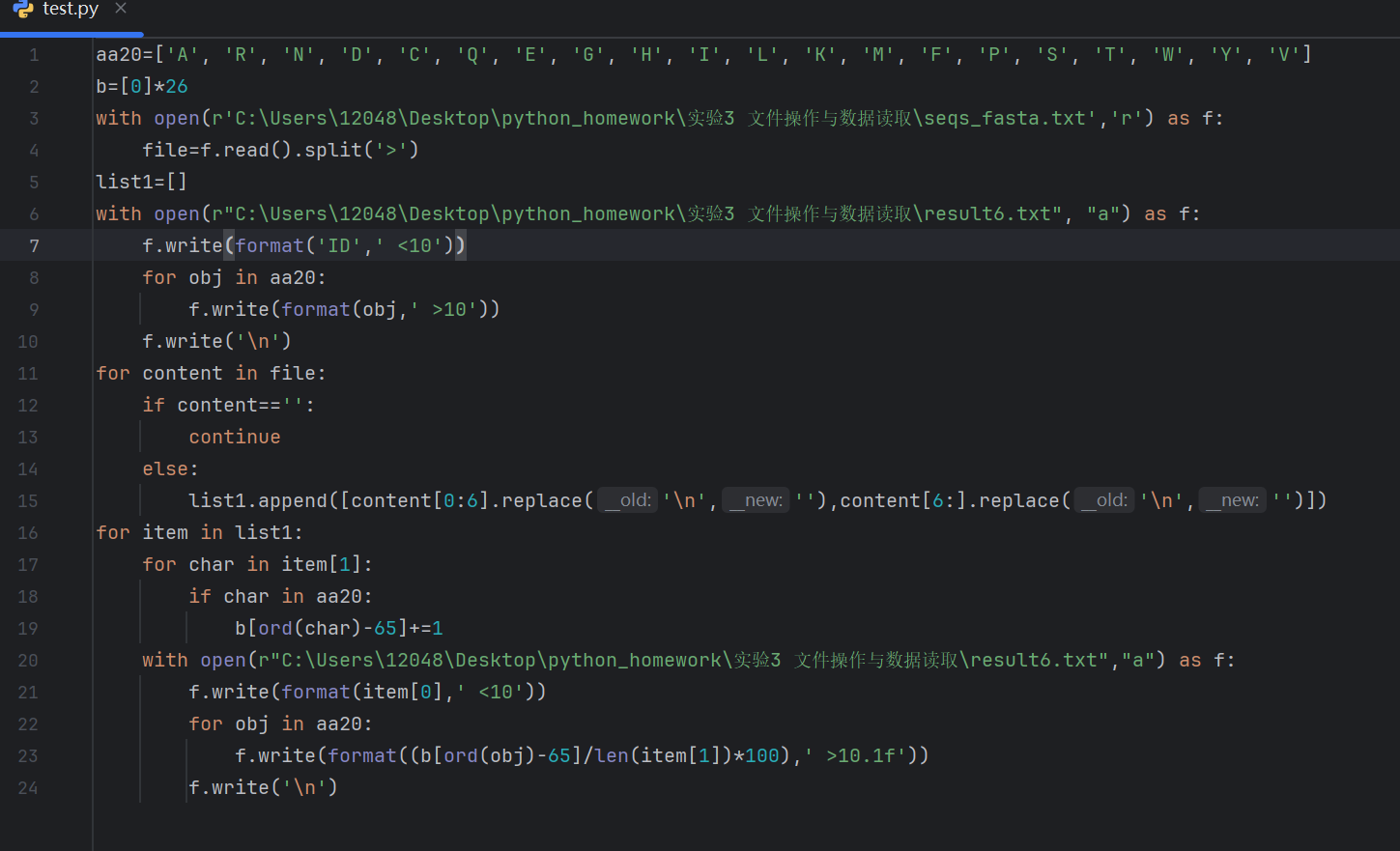
f.write(format(item[0],' <10'))

for obj in aa20:

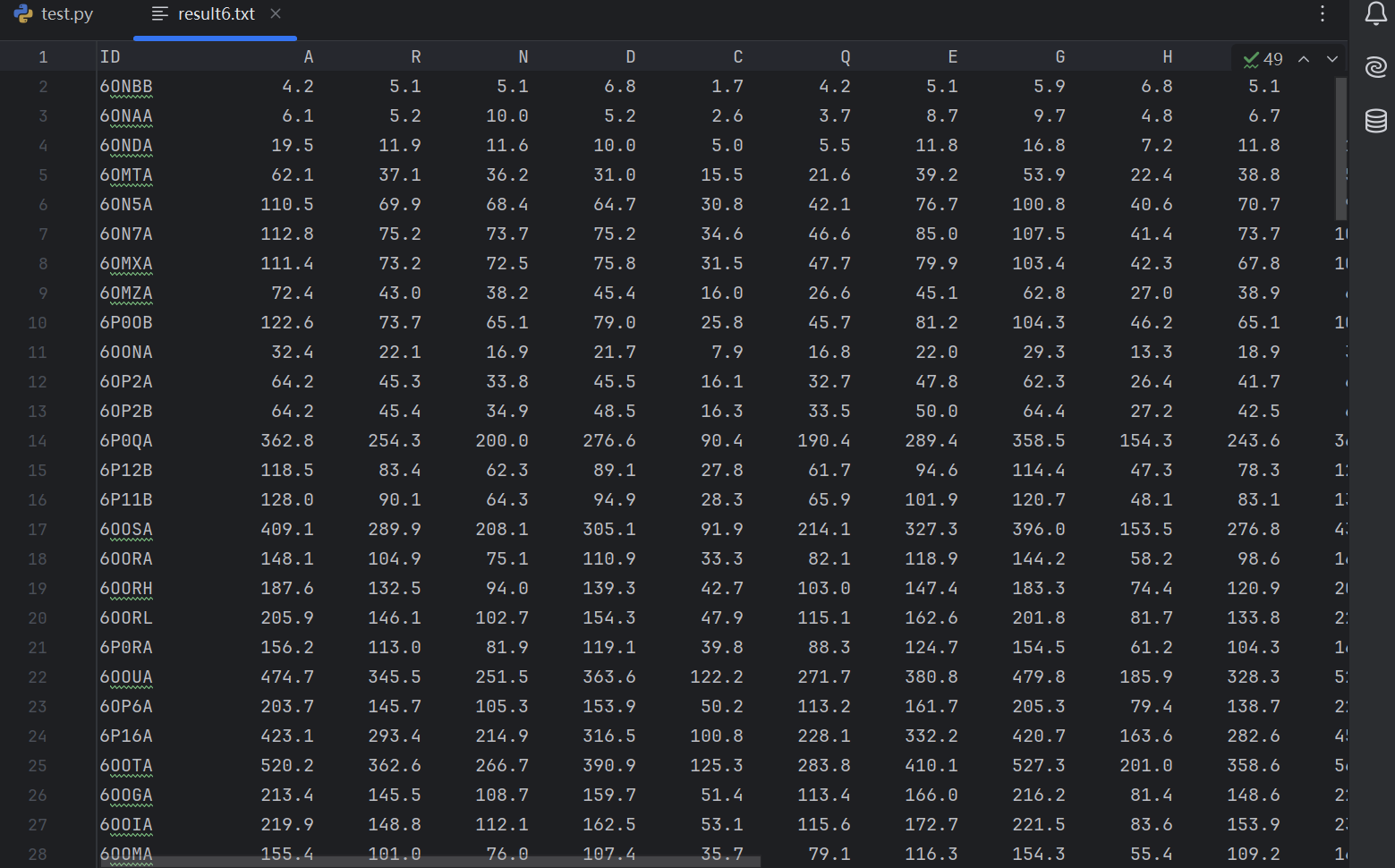
f.write(format((b[ord(obj)-65]/len(item[1])\*100),' >10.1f'))

f.write('\n')

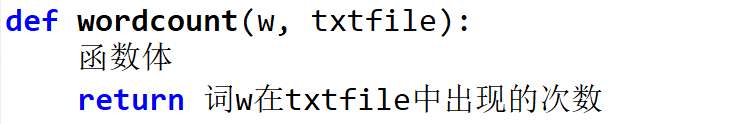
源代码截图：



文件内容截图：



(6)首先，定义一个名为“wordcount”的函数，功能为统计中文文本中某个关键字出现的次数，函数原型如下：

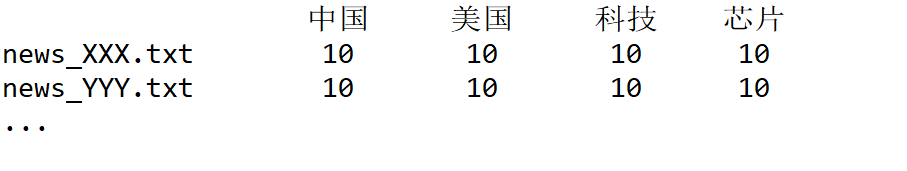


其中w和txtfile均为字符串。

其次，在存放本次实验材料的文件夹中，利用os.mkdir()创建一个新的文件夹，取名“mydir”；同时，自动识别出以“news\_”开头的所有文本文件，利用shutil模块将其移动至新建的文件目录“mydir”中(注：需编程自动实现文件的移动，不是手工拷贝或移动)。

进一步，利用pickle模块将函数wordcount以及识别出的以“news\_”开头的所有文本文件名组合成一个列表，永久保存至文件“wc.pkl”，并存储在文件夹“mydir”中。

最后，再次利用pickle模块将保存在“wc.pkl”中的列表数据载入，获得函数wordcount，并调用wordcount计算四个关键字“中国”、“美国”、“科技”和“芯片”在以“news\_”开头的所有文本文件中出现的次数，打印输出，格式参考如下：



源代码：

import os

import pickle

import shutil

def wordcount(w,txtfile):

sum=0

for i in range(0,len(txtfile)-len(w)+1):

if txtfile[i:i + len(w)] == w:

sum+=1

return sum

src=r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取'

des=r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\mydir'

os.mkdir(des)

for item in os.listdir(src):

if item.startswith('news\_'):

srce=os.path.join(src,item)

dest=os.path.join(des,item)

shutil.move(srce,dest)

print("所有以news\_开头的文件已移置指定位置")

list1= [wordcount]

for name in os.listdir(des):

list1.append(name)

with open(os.path.join(des,'wc.pkl'),'wb') as f:

pickle.dump(list1,f)

with open(os.path.join(des,'wc.pkl'),'rb') as f:

list1\_data=pickle.load(f)

us=list1\_data[0]

print(format('中国','>30'),end='')

print(format('美国','>10'),end='')

print(format('科技','>10'),end='')

print(format('芯片','>10'))

for i in list1\_data[1:]:

list2=[]

with open(os.path.join(des,i),'r') as f:

flie1=f.read()

for w in ['中国','美国','科技','芯片']:

list2.append(us(w,flie1))

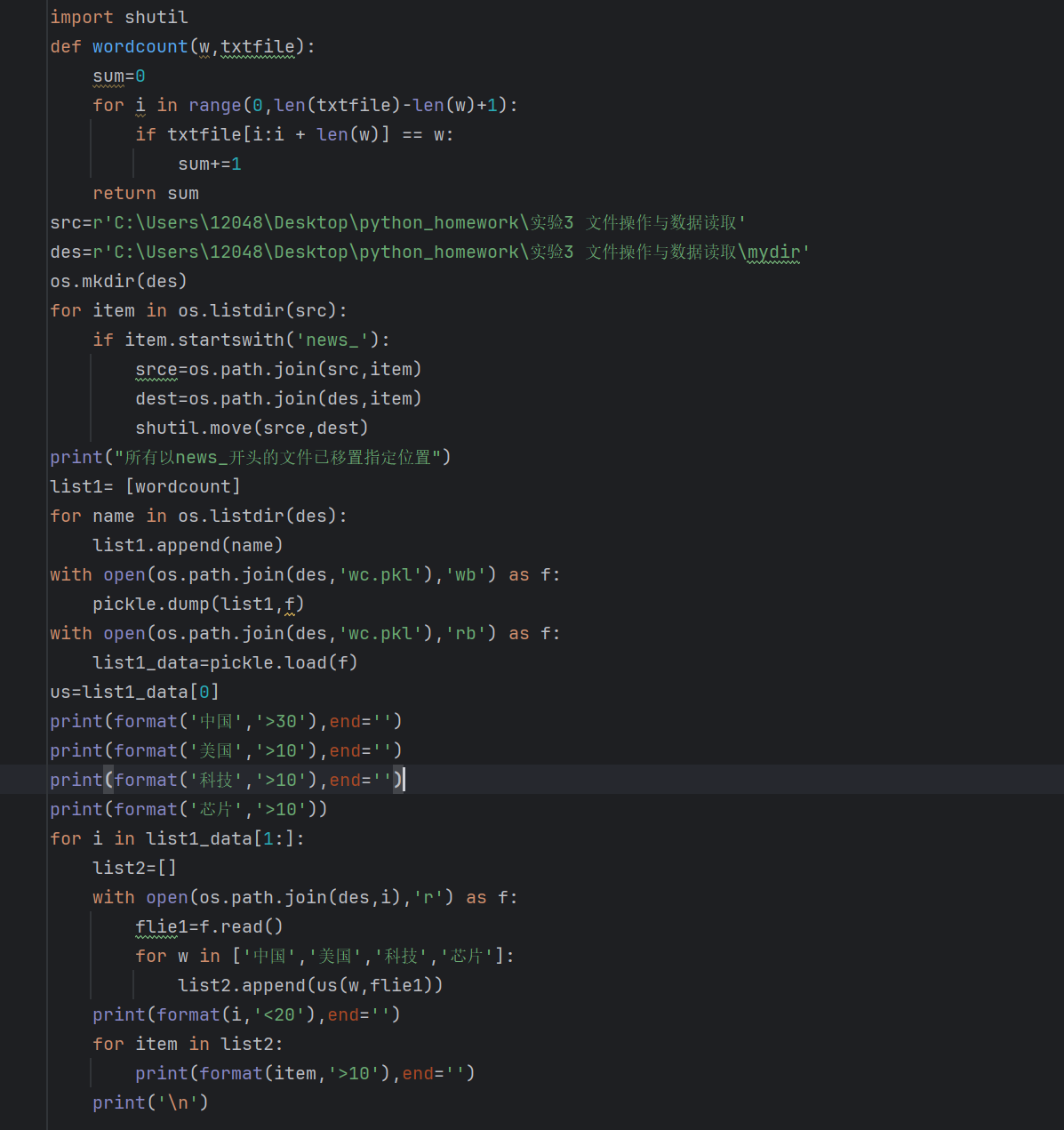
print(format(i,'<20'),end='')

for item in list2:

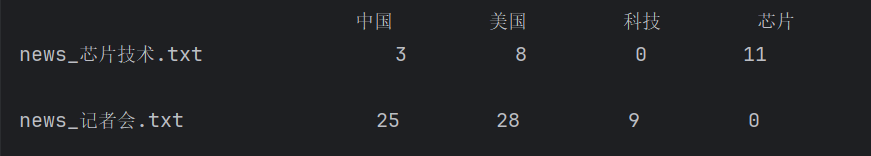
print(format(item,'>10'),end='')

print('\n')

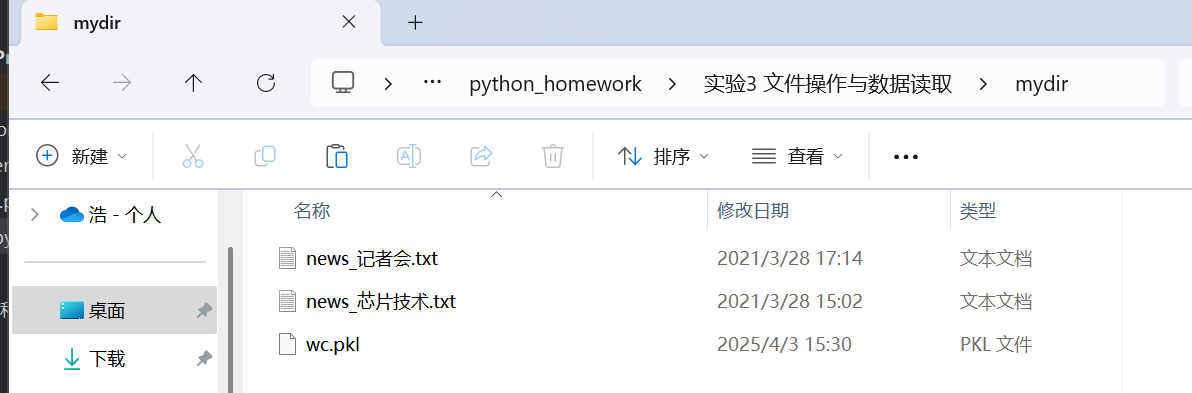
源代码截图：



运行结果截图：



文件夹mydir截图：



(7)从键盘输入一些字符，逐个把它们写到指定的文件out.txt中，直到输入一个@为止。

源代码：

while True:

with open(r"C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\out.txt","a") as f:

char=input('输入:')

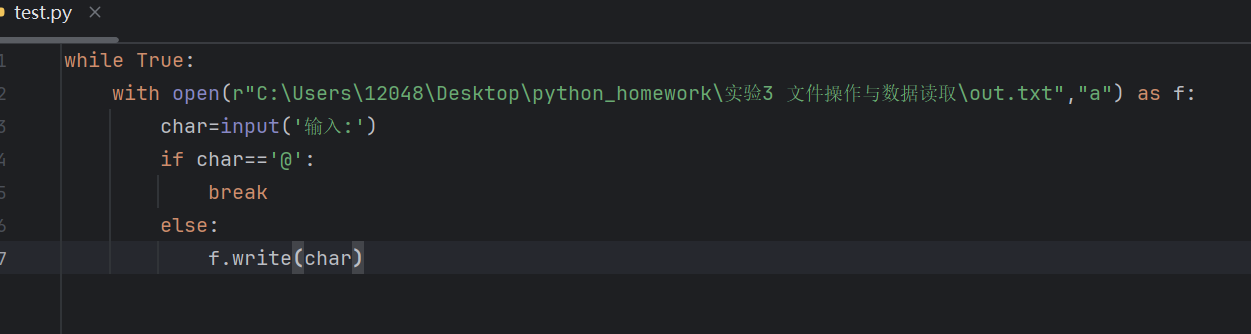
if char=='@':

break

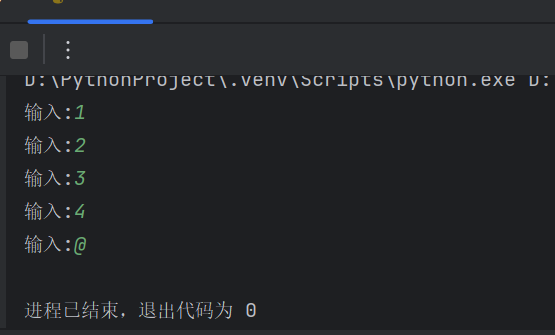
else:

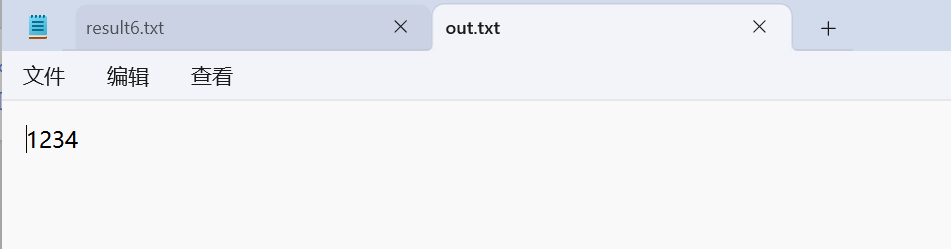
f.write(char)

源代码截图：



运行结果截图：





(8)读取学生成绩文件stu\_grade.txt，文件第一列表示学生学号，第二列表示学生姓名，第三列表示学生成绩，按学生成绩排序并统计及格率后，输出到新文件stu\_grade\_output.txt中。

源代码：

list=[]

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\stu\_grade.txt','r',encoding='utf-8') as f:

while True:

line=f.readline()

if not line:

break

else:

list.append(line.replace('\n','').split(','))

list.sort(key=lambda x:int(x[2]),reverse=True)

sum1=0

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\stu\_grade\_output.txt','a',encoding='utf-8') as fo:

for i in list:

if int(i[2])>=60:

sum1+=1

for j in range(3):

fo.write(i[j])

if j!=2:

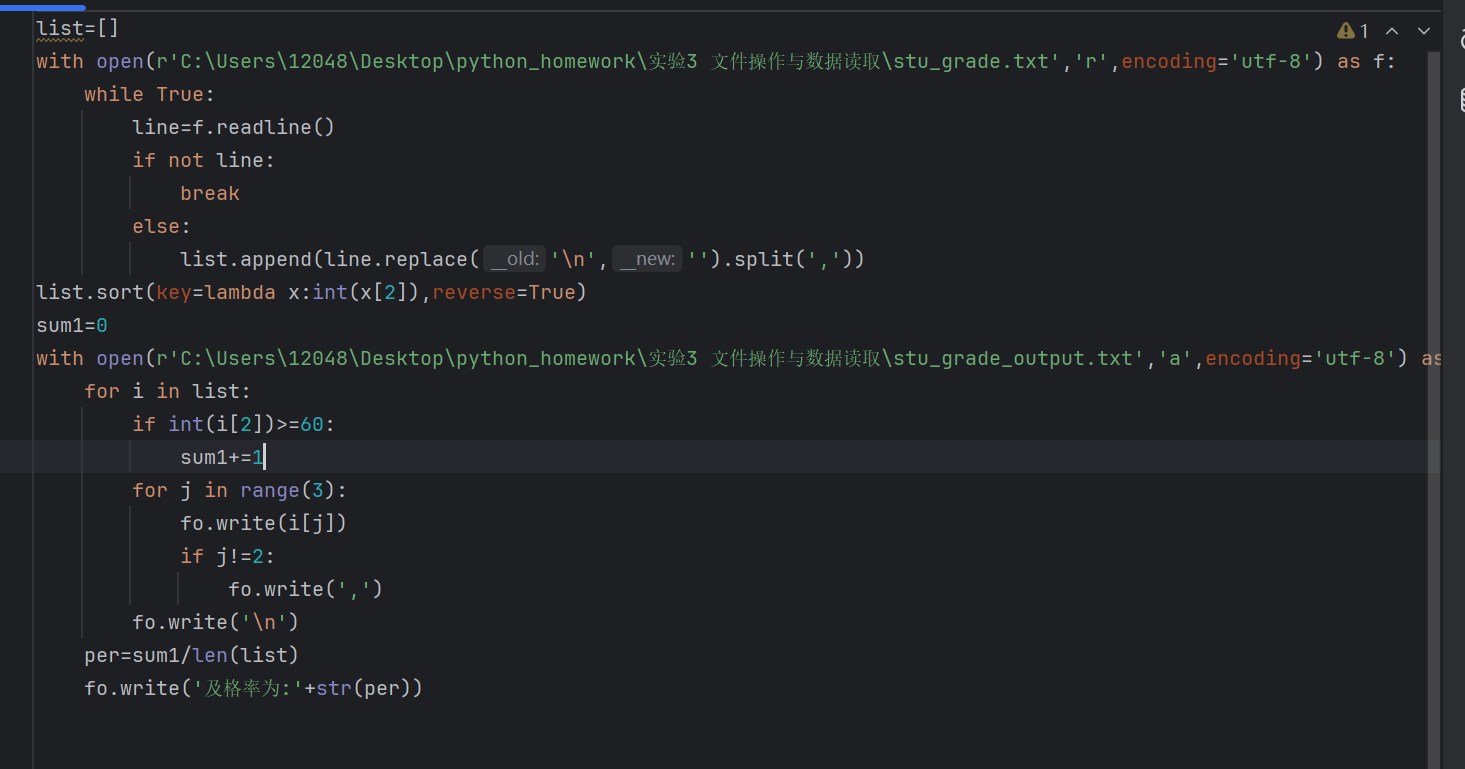
fo.write(',')

fo.write('\n')

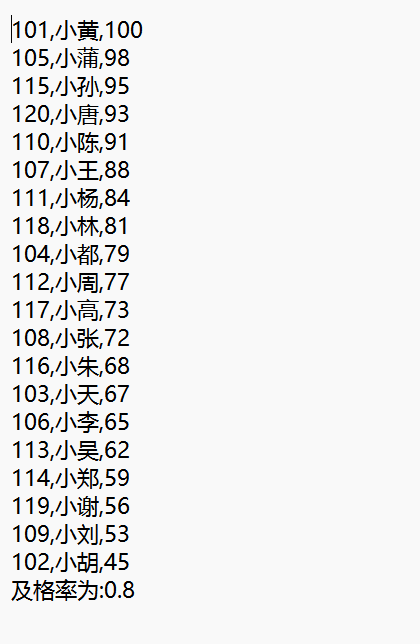
per=sum1/len(list)

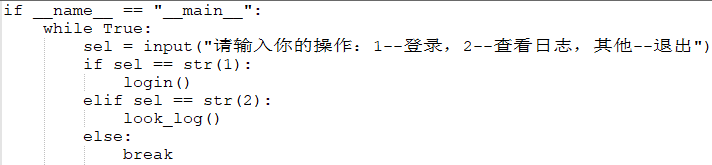
fo.write('及格率为:'+str(per))

源代码截图：



运行结果截图：



(9)创建一个叮咚客服管理系统的界面，每次登录时，将用户的登录日志写入文件中（包含登录时间），并且在程序中查看用户的登录日志。假定系统中只有一个用户（用户名： "张三" ，密码"123456"）可以登录成功；请记录用户是否登录成功以及登录时间。  
 下图为系统的主函数，请写出login()函数和look\_log()函数，实现上述要求。

源代码：

import time

def login():

admin=input("请输入用户名:")

password=input("请输入密码:")

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\daily.txt','a',encoding='utf-8') as f:

ti = time.strftime('%Y-%Y-%m-%d %H:%M:%S', time.localtime(time.time()))

if admin=="张三" and password=="123456":

f.write("本地时间:"+ti+"--用户:"+admin+"--登录成功\n")

else:

f.write("本地时间:"+ti+"--用户:"+admin+"--出现异常登录信息,登录失败\n")

def look\_log():

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\daily.txt', 'r', encoding='utf-8') as f:

print(f.read())

if \_\_name\_\_ == '\_\_main\_\_':

while True:

sel = input("请输入你的操作：1--登录，2--查看日志，其他--退出")

if sel == str(1):

login()

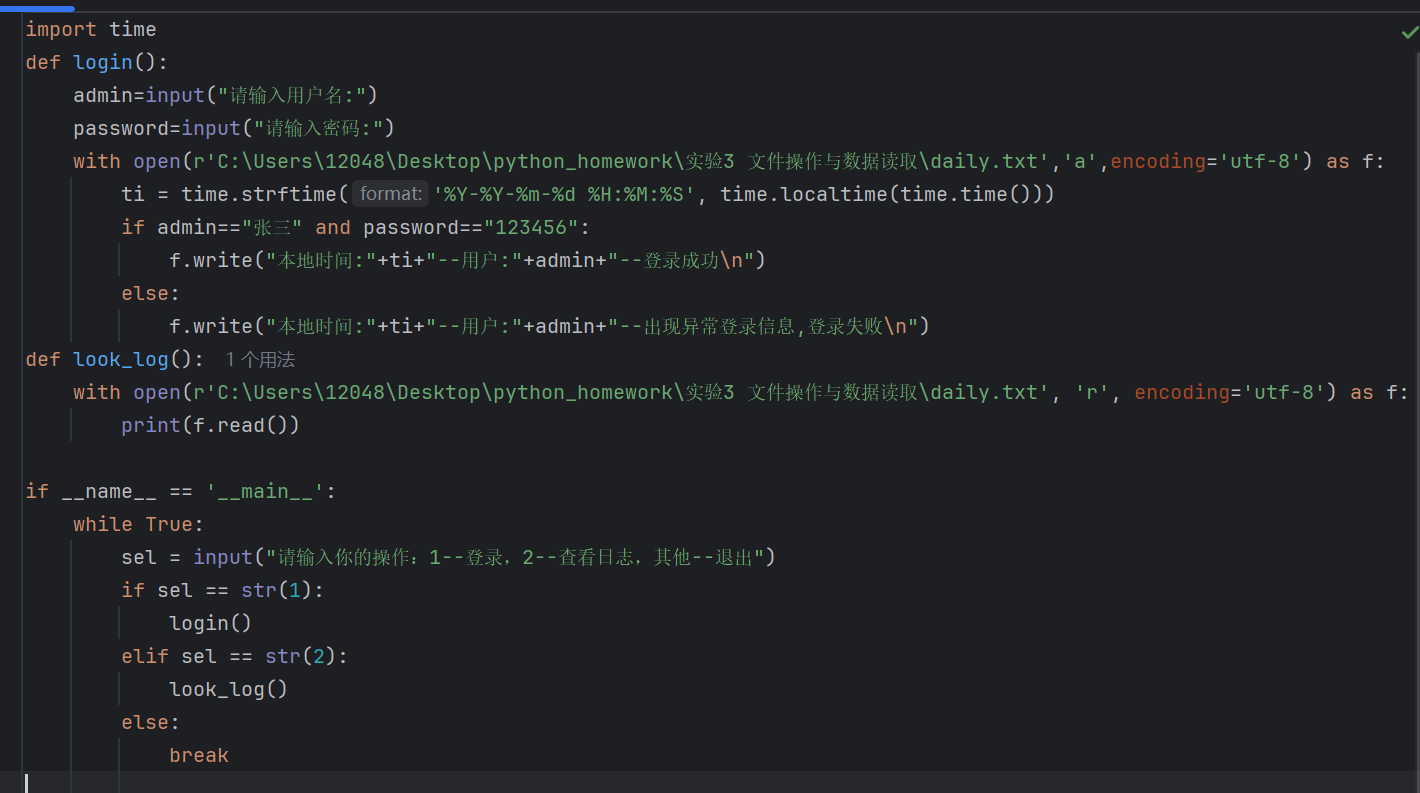
elif sel == str(2):

look\_log()

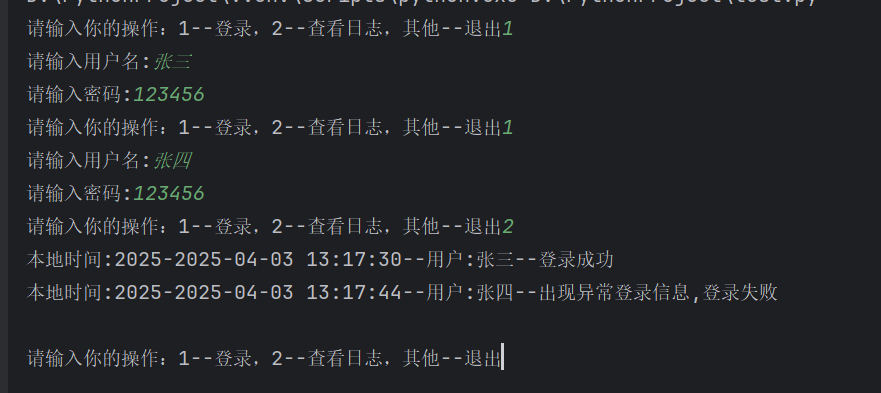
else:

break

源代码截图：

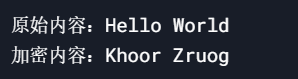


运行结果截图：



(10)编写一个文件加密程序，读取"secret.txt"文件内容，使用凯撒密码（位移+3）进行加密，将加密结果保存到"secret\_encrypted.txt"，再编写解密程序读取加密文件并还原，加密解密函数要使用pickle保存为"cipher.pkl"

文件格式示例：



源代码：

import pickle

def ec(src,des): #加密程序

with open(src,'r',encoding='utf-8') as f1:

with open(des, 'a', encoding='utf-8') as f2:

while True:

line = f1.readline()

if not line:

break

else:

for item in line:

if item.isupper():

f2.write(chr((ord(item)-ord('A')+3)%26+ord('A')))

elif item.islower():

f2.write(chr((ord(item)-ord('a')+3)%26+ord('a')))

else:

f2.write(item)

f2.write('\n')

def dc(src,des): #解密程序

with open(src,'r',encoding='utf-8') as f1:

with open(des, 'a', encoding='utf-8') as f2:

while True:

line = f1.readline()

if not line:

break

else:

for item in line:

if item.isupper():

f2.write(chr((ord(item) - ord('A') - 3) % 26 + ord('A')))

elif item.islower():

f2.write(chr((ord(item) - ord('a') - 3) % 26 + ord('a')))

else:

f2.write(item)

f2.write('\n')

if \_\_name\_\_ == "\_\_main\_\_":

src=r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\secret.txt'

des=r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\secret\_encrypted.txt'

with open('cipher.pkl', 'wb') as f: # 保存加密解密函数

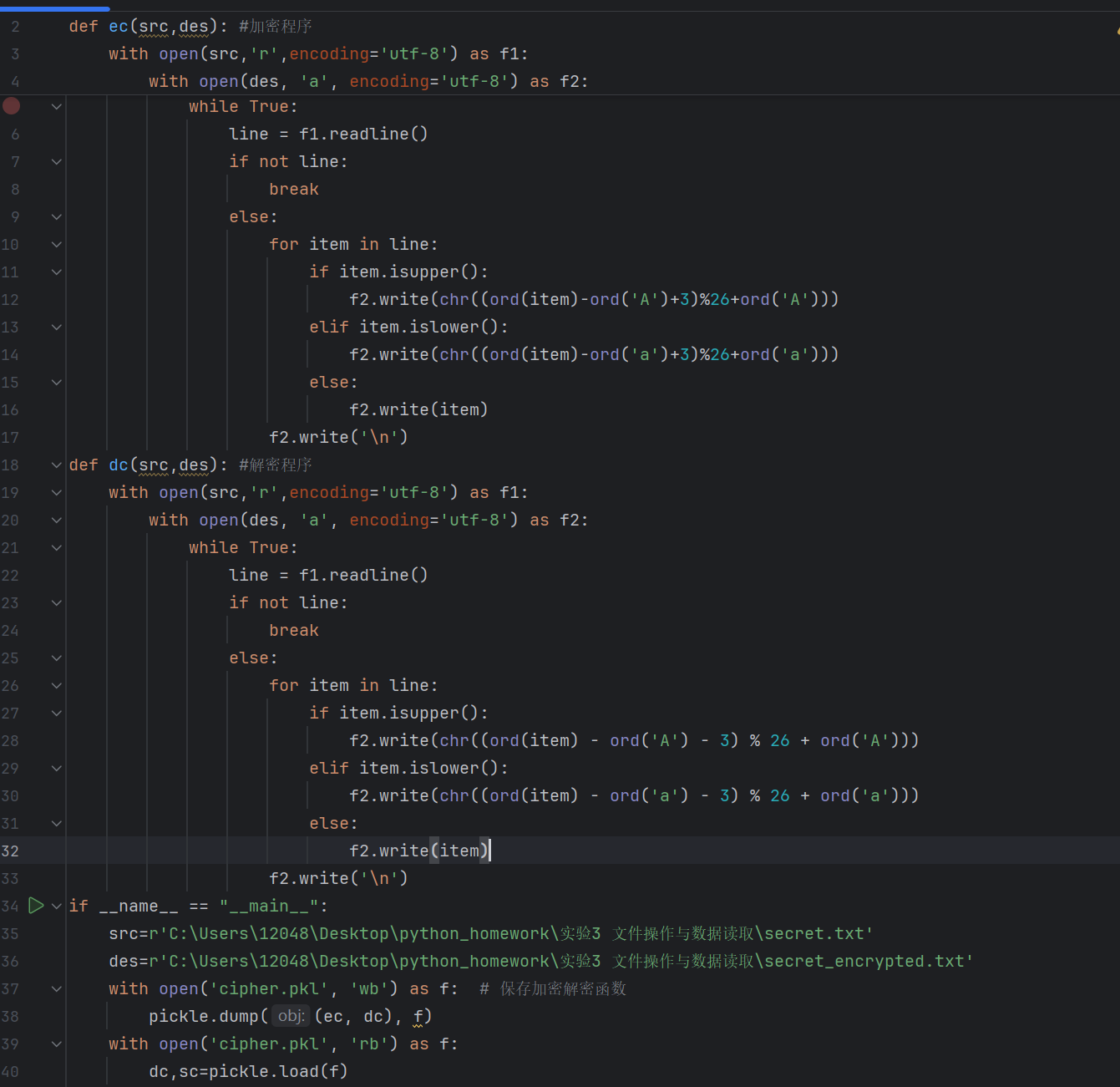
pickle.dump((ec, dc), f)

with open('cipher.pkl', 'rb') as f:

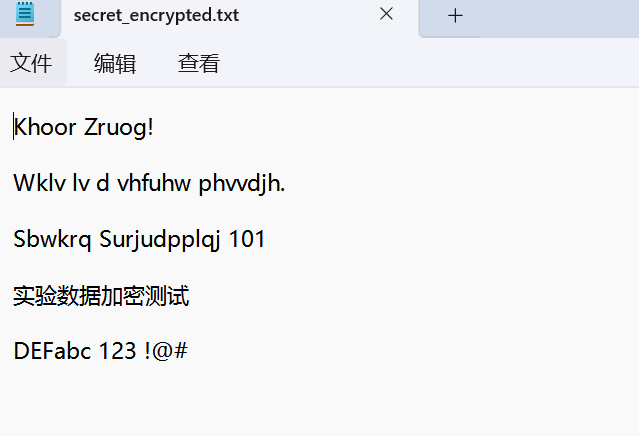
dc,sc=pickle.load(f)

dc(src,des)

源代码截图：



运行结果截图：



（11）读取"exp\_data.csv"文件（格式：日期,温度,湿度,光照强度）

计算平均温度、最高湿度、光照总和，找出温度超过30℃的异常数据，将统计结果保存到"report.txt"。

源代码：

import csv

list\_er=[]

temperature=[]

humidity=[]

light=[]

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\exp\_data.csv', 'r', encoding='utf-8') as f:

reader = csv.reader(f)

flag=True

for row in reader:

if flag:

flag=False

continue

date = row[0]

temp = float(row[1])

hum = float(row[2])

lig = float(row[3])

temperature.append(temp)

humidity.append(hum)

light.append(lig)

if temp >30:

list\_er.append([date,temp])

average\_temperature = sum(temperature)/len(temperature)

max\_humidity = max(humidity)

sum\_light = sum(light)

with open(r'C:\Users\12048\Desktop\python\_homework\实验3 文件操作与数据读取\report.txt', 'w', encoding='utf-8') as f:

f.write('平均温度为:'+str(average\_temperature) + '\n')

f.write('最高湿度为:'+str(max\_humidity) + '\n')

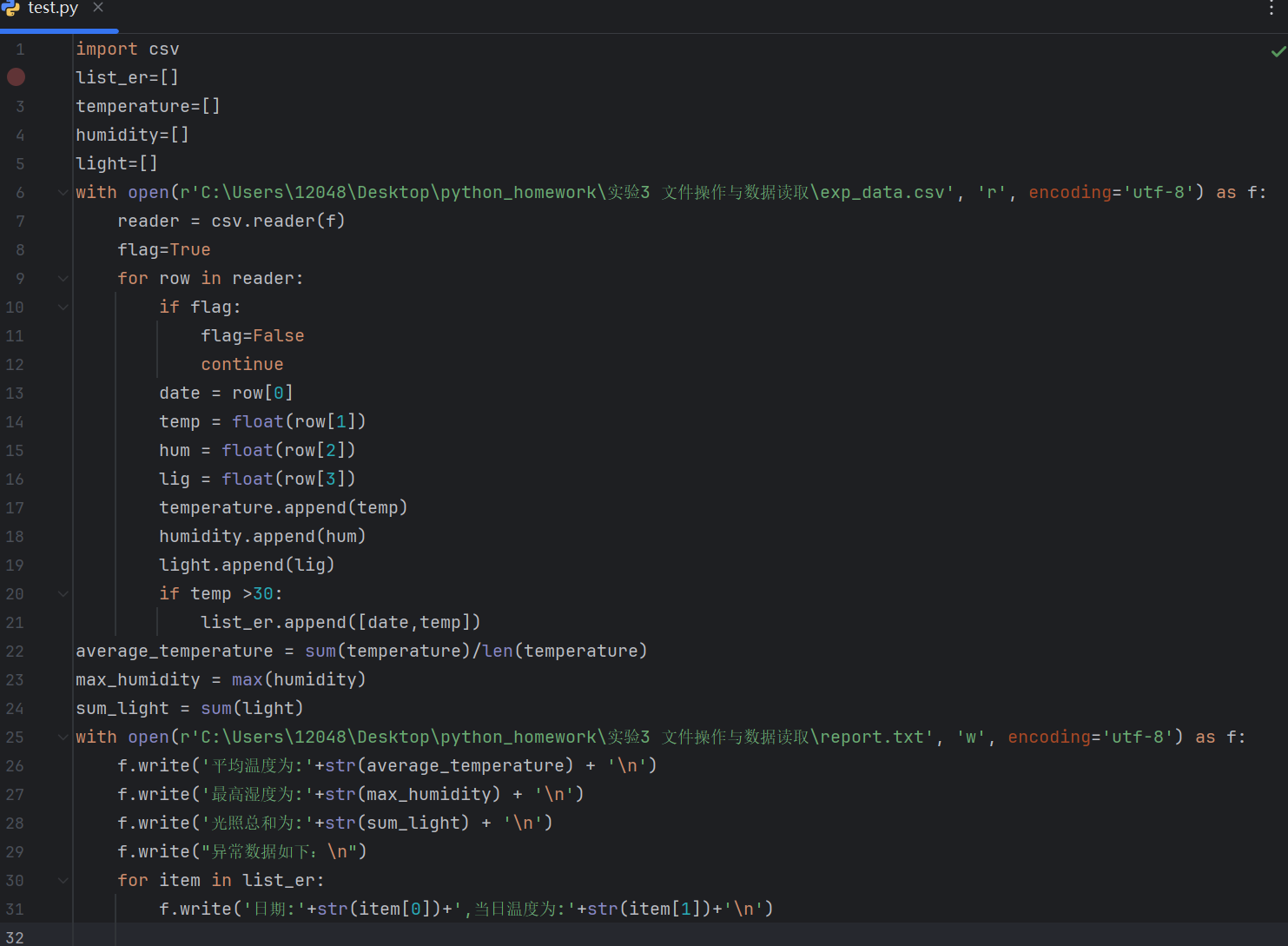
f.write('光照总和为:'+str(sum\_light) + '\n')

f.write("异常数据如下：\n")

for item in list\_er:

f.write('日期:'+str(item[0])+',当日温度为:'+str(item[1])+'\n')

源代码截图：



运行结果截图：

