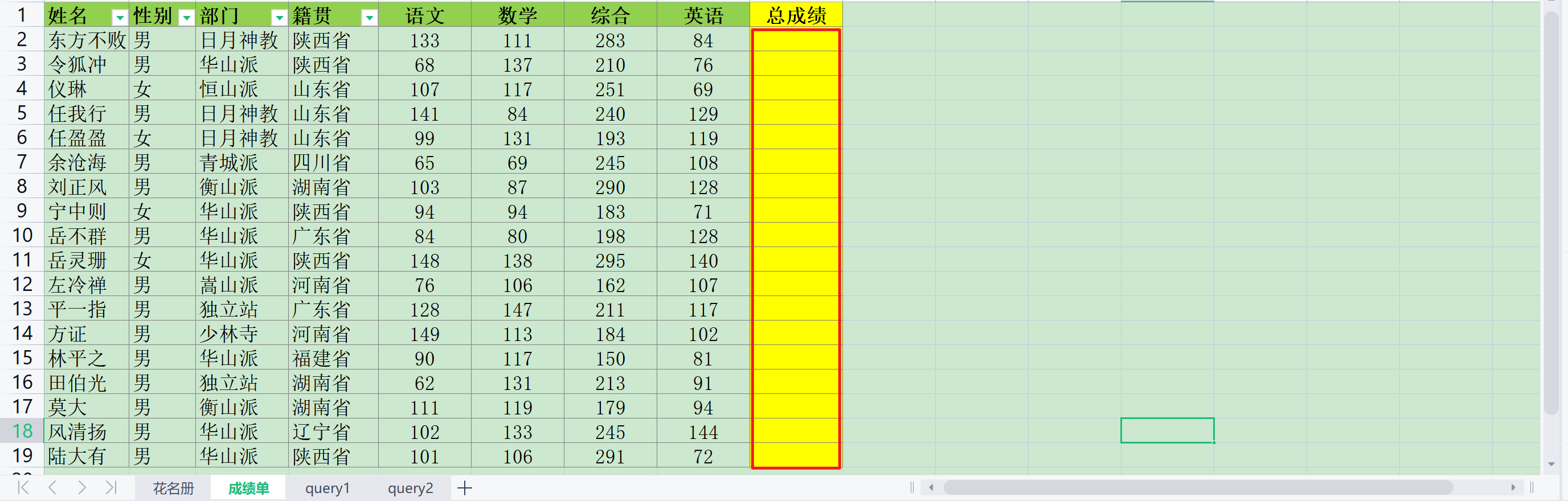
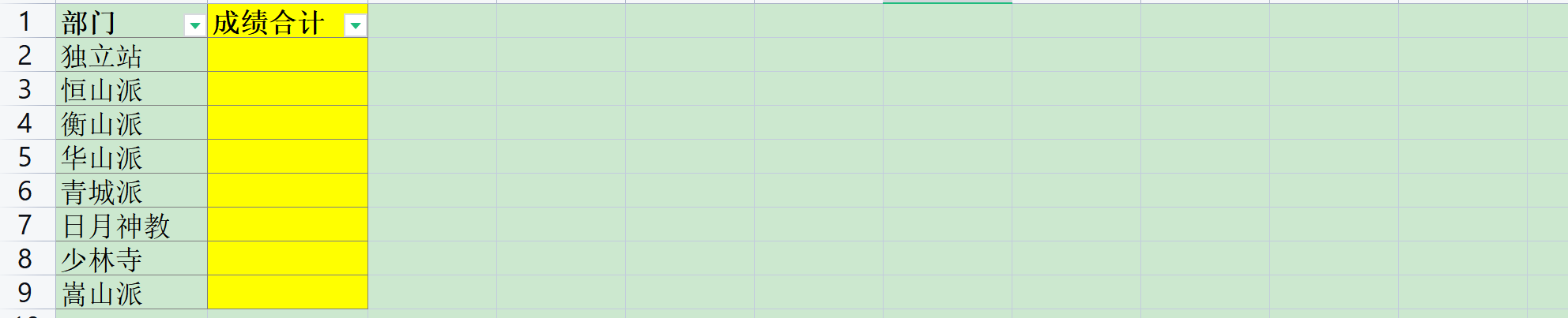
求和类函数SUM,SUMIFS详解：Excel+Python+SQL+Tableau四种工作实现方法：

摘要：求和函数是我们经常在用的，尤其是SUM,SUMIFS等函数，个人觉得SUMIFS的用法已经完整地包含了SUMIF函数，只学SUMIFS就可以，而这两种求和函数在EXCEL表格里,编程语言Python里,数据库语言SQL里，还有可视化软件Tableau里分别如何实现，我尽求详尽地写教程，演示数据只是极小的短短几行，实际中我们可能要处理百万，千万行甚至更多数据，而超过百万行以后EXCEL就明显力不从心了，因为首先行数在工作表里就装不下了，EXCEL格式.xlsx最多可以保存2的20次幂个行数（1048576行），2的14次幂个列数（16384列），超过这个量值的数据则会被EXCEL自动截断舍弃，所以在这种情况下我们处理大数据必然要用到后面的工具

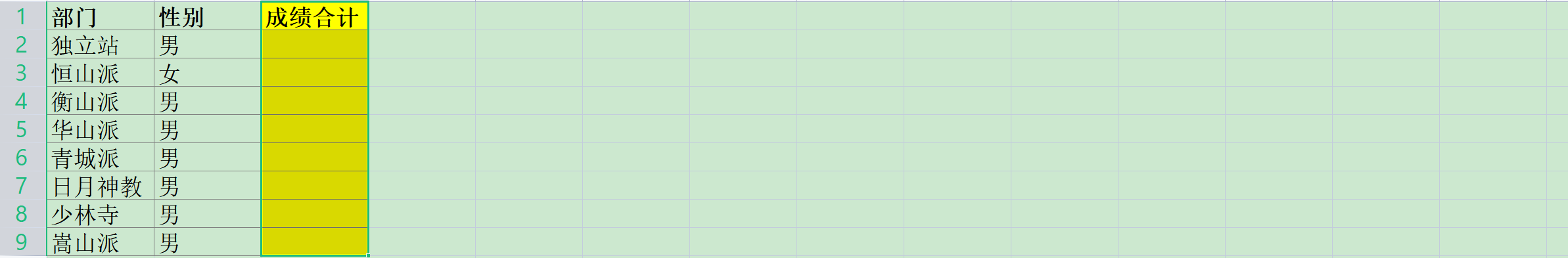
本例要求：1求出每个人的总成绩



2按部门条件进行成绩汇总求和



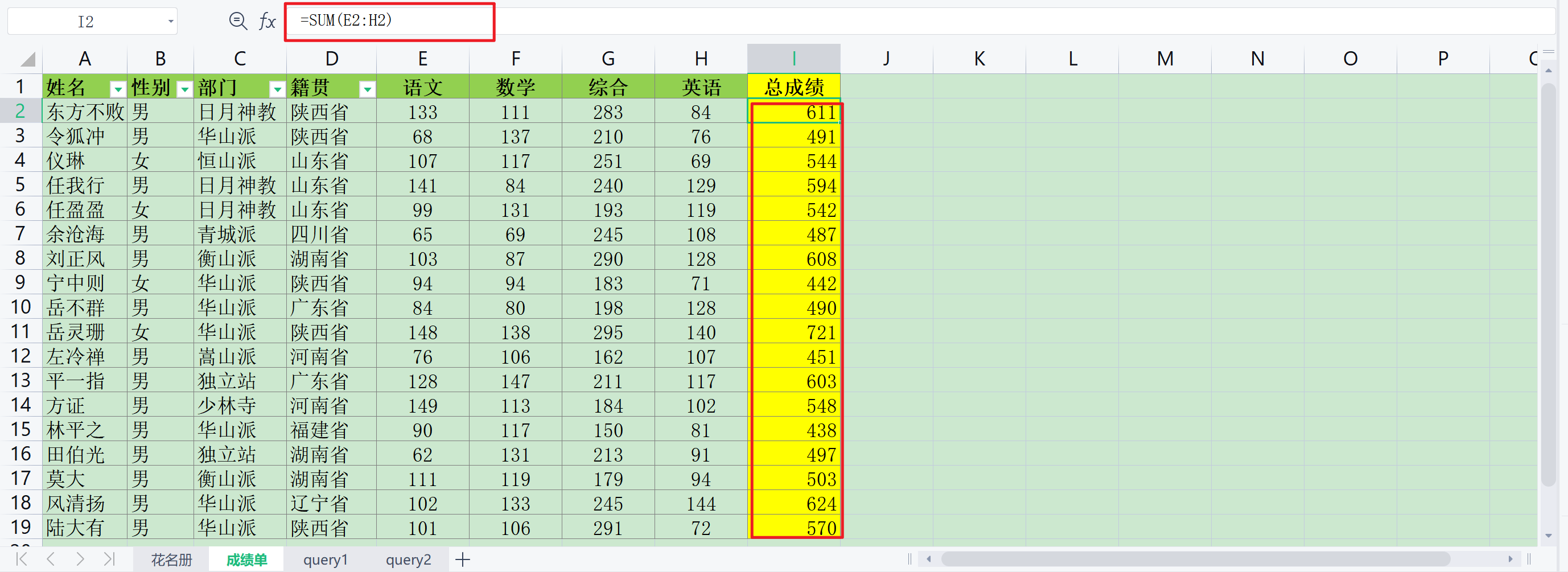
1. 按部门条件+性别条件进行成绩汇总求和



到这里有同学可能会说这也太简单了吧，是的，万事都是从简单开始，高手也要经过简单这个过程才能成为高手不是

一、Excel实现

1直接在I2单元格输入公式：=SUM(E2:H2) 向下填充即可（快捷键操作：选中I2到I19单元格，按ALT + = 可以快速生成结果）



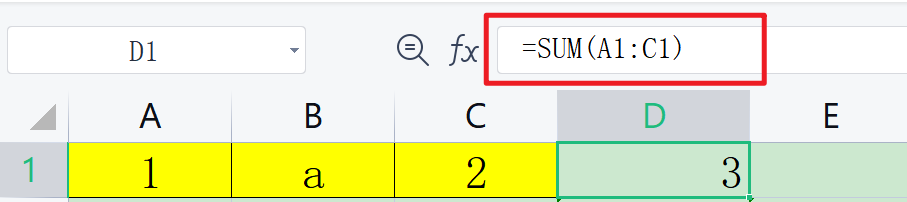
SUM()函数说明：sum单词就是总和的意思

SUM(number1,[number2],...)

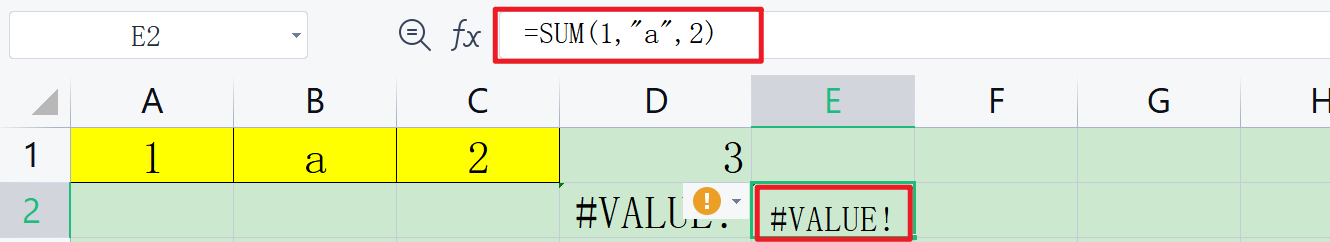
这个函数里面的参数可以是一个单元格，也可以是一个序列，或者一个区域，参数最多可以指定255个，每个参数的值可以包含多个单元格的值

这个函数参数里面如果是文本或者逻辑值，会自动忽略也就是不做运算

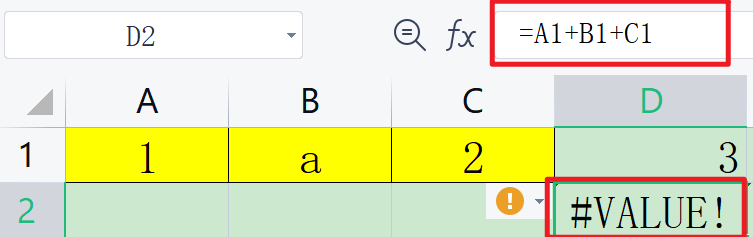
例：我们对1,a,2 三个值在单元格里用=sum（A1:C1)进行求和，这个函数自动忽略里面值为字母的单元格,结果为3



如果写成=SUM(1,"a",2)这种形式则会报错，因为此时第二个参数并不是单元格引用，无法忽视



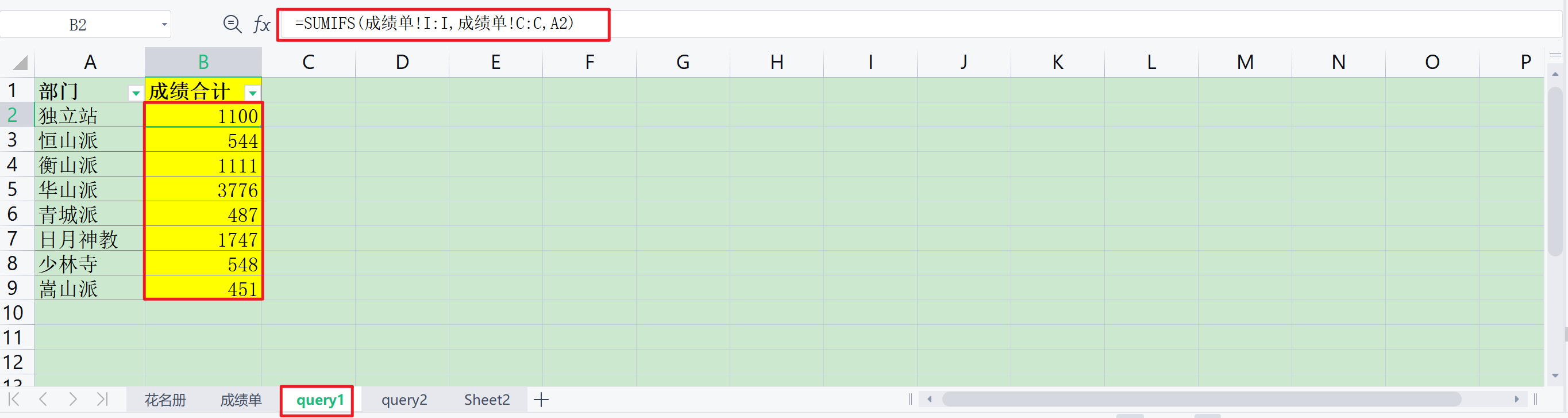
而如果我们用加号进行求和则会报错，字母a不能进行相加



回到最前面：曾经我在一本比较有名的EXCEL书籍里看到作者曾见过有人这样写求和公式：=SUM(A1+B1+C1)，作者本人表示哭笑不得，其实我的观点倒是无可厚非，只要能达到正确的结果就行，毕竟我们不能要求每个职场人都成为EXCEL高手，可能没必要理解函数的原理等等，只要我们会用或者不咋会用但能得到想要的结果就可以

2单条件求和，有同学可能会想到用SUMIF函数，我的观点是直接PASS这个函数（不做介绍），主观上我认为这是一个鸡肋的函数，因为SUMIFS函数涵盖它并且更好用

在query1表B2单元格输入公式：=SUMIFS(成绩单!I:I,成绩单!C:C,A2)，下拉填充完成查询



SUMIFS多条件求和用法：

SUMIFS(sum\_range, criteria\_range1, criteria1, [criteria\_range2, criteria2], ...)

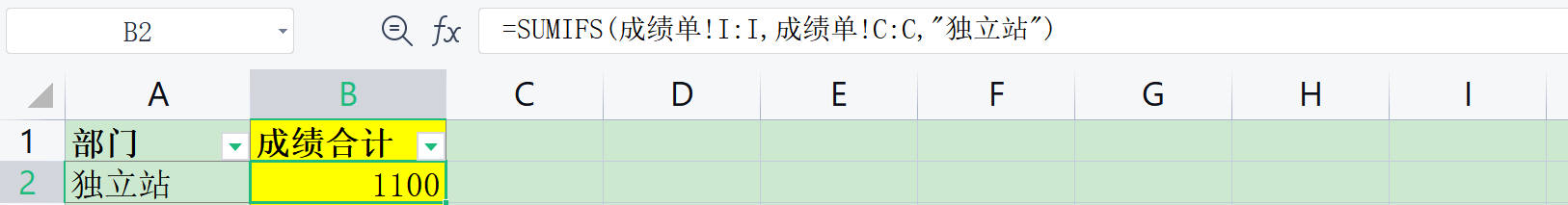
必填的是3个参数，后面可以选填更多条件

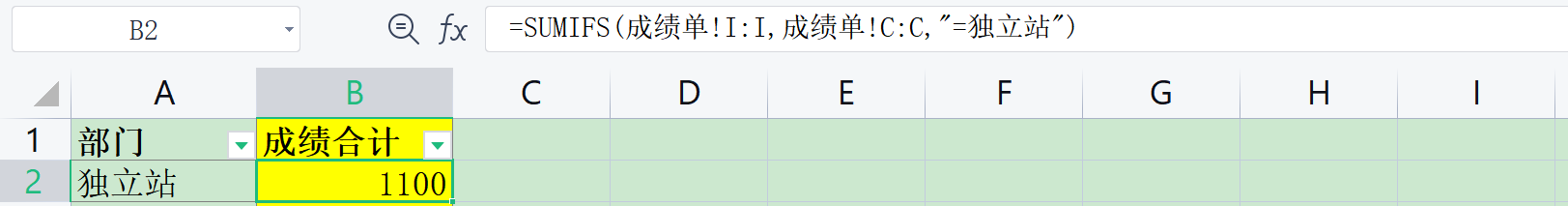
第一个参数：sum\_range，要求和的单元格区域，这里我们直接选成绩表里的总成绩I列，有同学会问那假如原表里没有总成绩这列，可不可以选择E列到H列的整个数据区域，答案是不行的，这种的话需要用到另一个数组求和公式，后面会讲



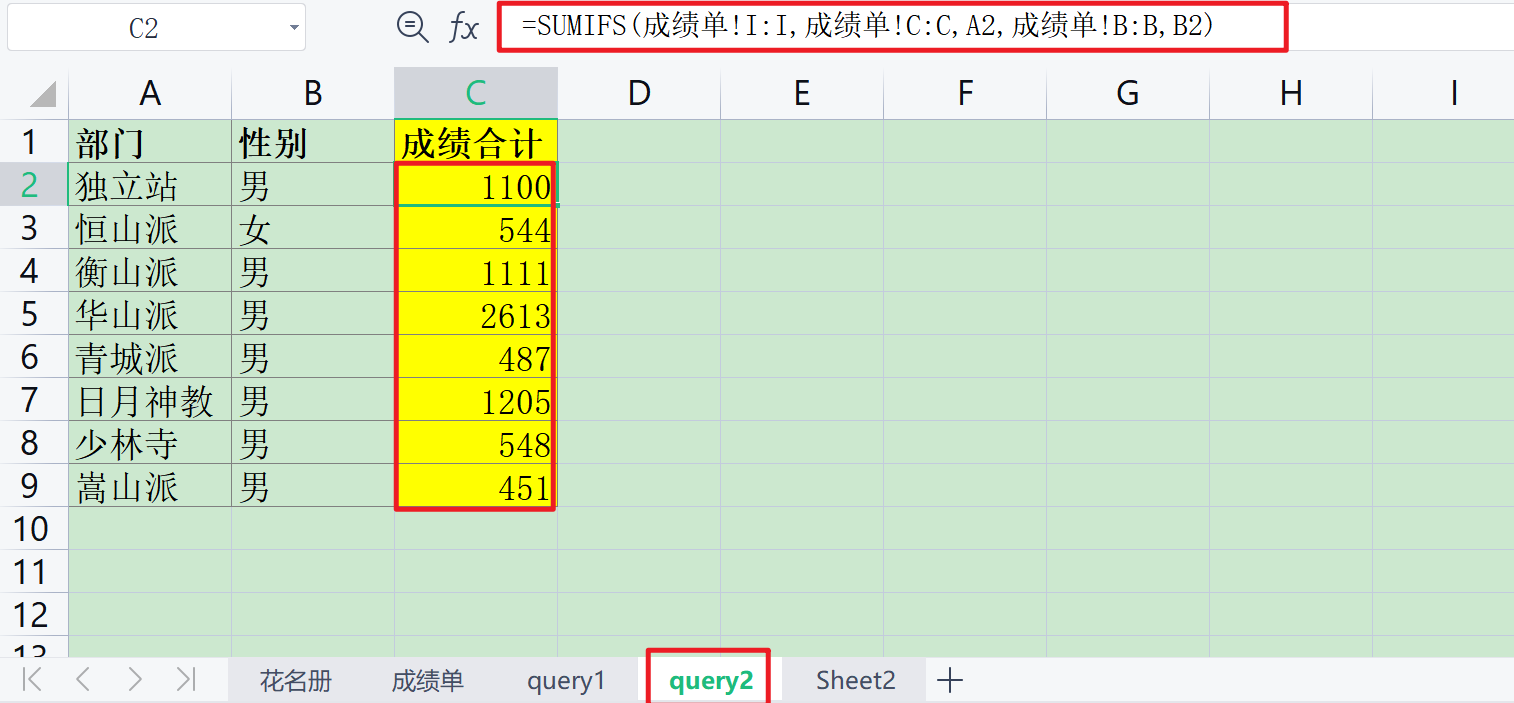
第二个参数criteria\_range1：criteria单词是条件的意思，这是条件区域1，也就是要查询的总成绩列所在表里面的条件列，我们是按部门进行求和，所以我们选择部门C列

第三个参数是criteria1，也就是条件1的意思，这里面我们选择query1表里面的A2列，也就是要查找求和的条件部门，其实这个条件是用的逻辑等于，B2单元格公式 = SUMIFS(成绩单!I:I,成绩单!C:C,A2) 等价于=SUMIFS(成绩单!I:I,成绩单!C:C,"=独立站")和=SUMIFS(成绩单!I:I,成绩单!C:C,"独立站")，这两种写法的结果是一样的，当然以此类推，还可以有不等于<>,大于条件>，小于条件<，大于等于条件>=，小于等于条件<=，通配符条件(\*,?)等





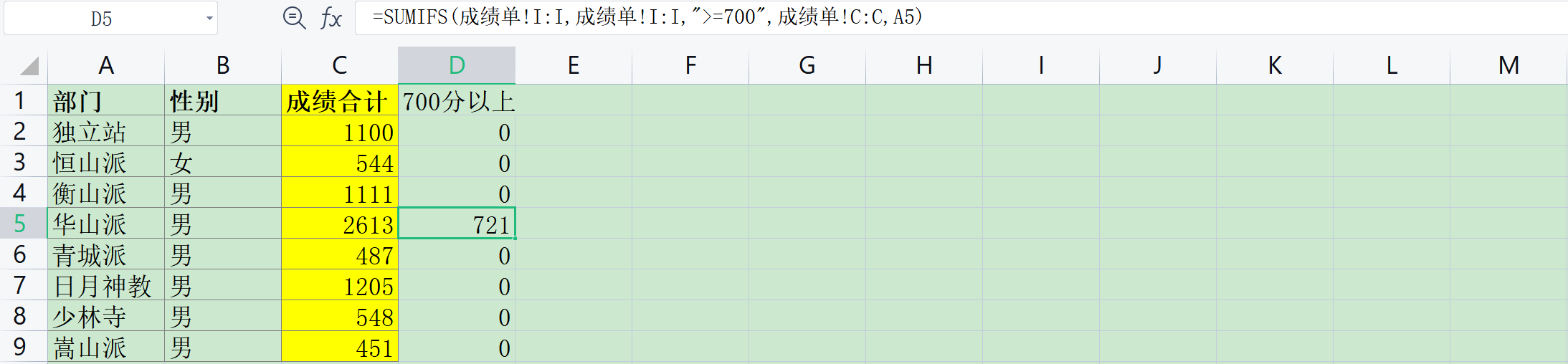
3多条件查询求和：我们在C2单元格输入公式：=SUMIFS(成绩单!I:I,成绩单!C:C,A2,成绩单!B:B,B2)向下拉填充即可



这个查询里面我们多加了一个条件：性别，就是在前面三个参数的基础上，后面再加一个条件（两个参数），理论上这个可以无限进行增加条件

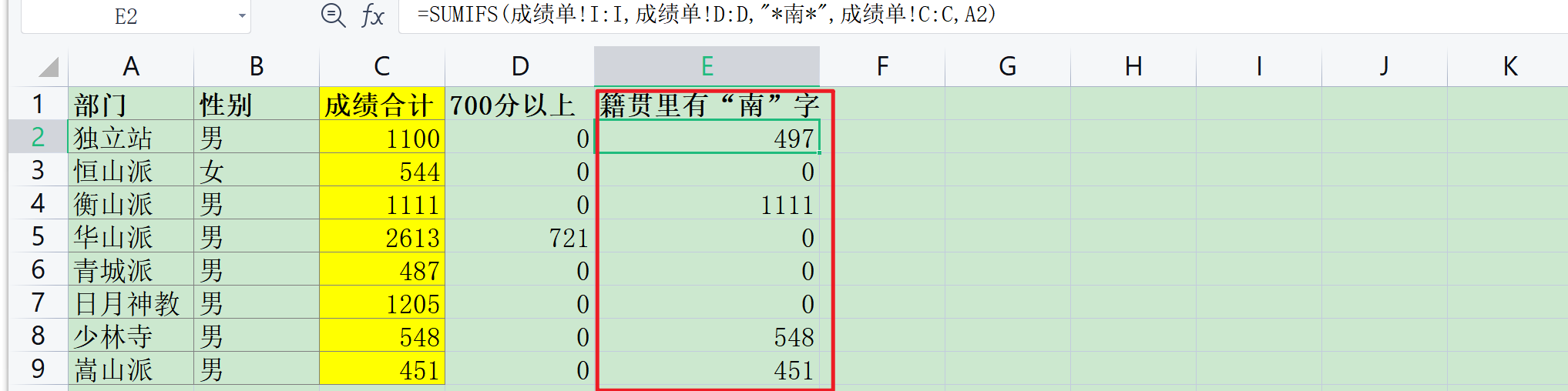
举几个多条件查询的例子:

1. ：查询总成绩在700分及以上的同学的成绩并按部门进行求和，公式为：SUMIFS(成绩单!I:I,成绩单!I:I,">=700",成绩单!C:C,A2)

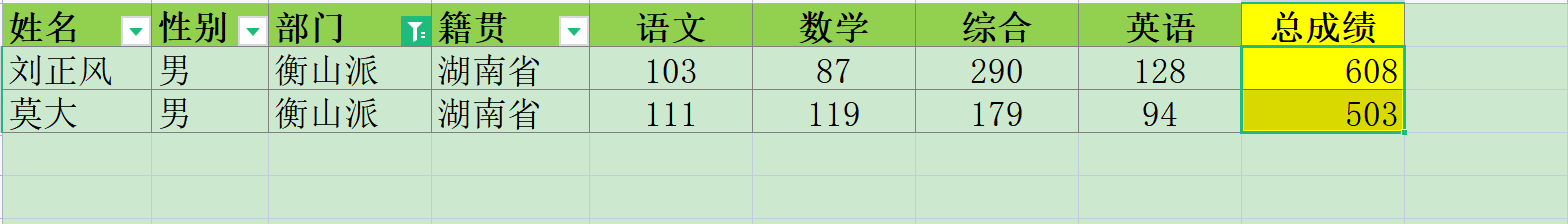


只有华山派的岳灵珊考了721分，大于700，所以只有华山派有值，其他则为0

1. 查询成绩表籍贯里有“南”字的同学的成绩并按部门进行求和：=SUMIFS(成绩单!I:I,成绩单!D:D,"\*南\*",成绩单!C:C,A2)



这里查询条件用了通配符\*，即只要在籍贯这列里面出现‘南’字，我们则选取再按部门进行求和



如衡山派一共两个人，籍贯里面都有南字，成绩合计等于1111

1. PYTHON实现：

完整代码：

import pandas as pd #1

df = pd.read\_excel("c:/study\_note/xiao.xlsx",sheet\_name = "成绩单") #2

df['总成绩'] = df.sum(axis=1,numeric\_only=True) #3

df1 = df.groupby(["部门"])["总成绩"].sum() #4

df2 = df[((df['部门']!='恒山派')&(df['性别']=='男') | (df['部门']=='恒山派')&(df['性别']=='女'))] #5

df2 = df2.groupby(['部门','性别'])['总成绩'].sum() #6

with pd.ExcelWriter("c:/study\_note/xiao.xlsx",mode="a",engine="openpyxl") as writer: #7

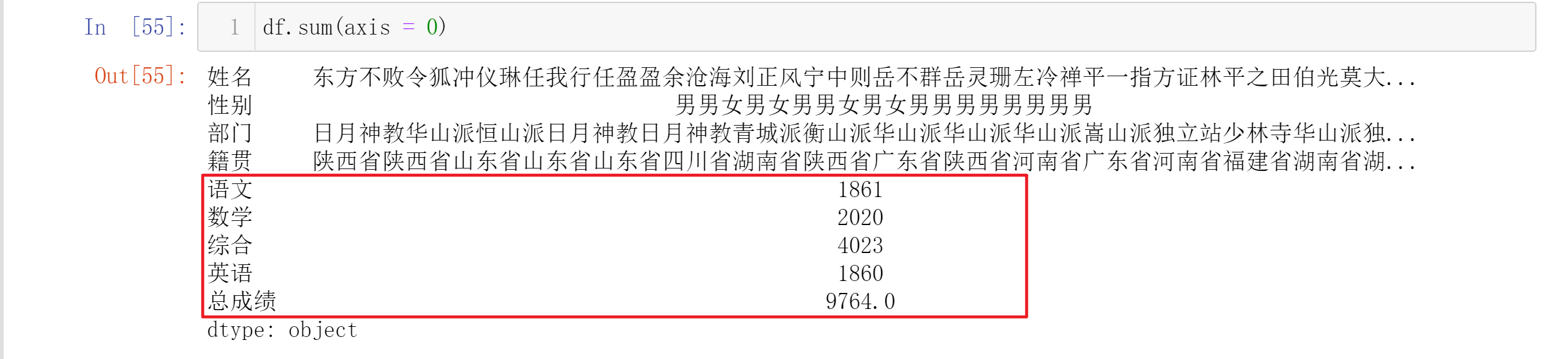
df.to\_excel(writer,sheet\_name="成绩表2",index=False) #8

df1.to\_excel(writer,sheet\_name='query11') #9

df2.to\_excel(writer,sheet\_name='query22') #10

代码说明：这个代码还是用pandas库完成我们要查询的结果，以在原工作簿上新增加表的形式完成我们的工作，并没有修改原表中的数据

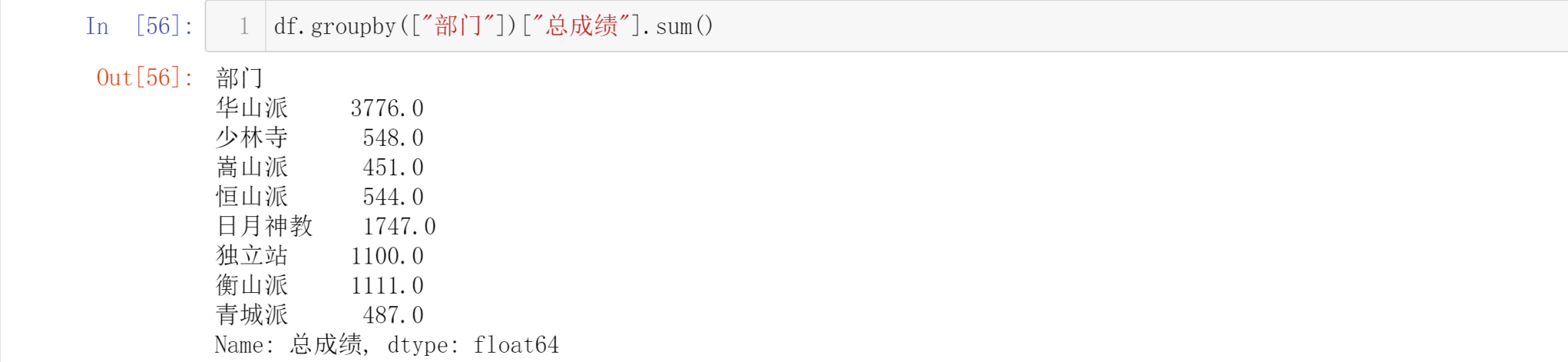
1. 导入第三方模块pandas
2. 读取成绩单表数据并保存在变量df中
3. 进行求和并赋值操作，df.sum()函数表示对表进行求和，如果参数axis = 0，表示按列求和，求和的结果是把各科成绩向下相加，如下：



所以我们这里要按行进行横向求和，所以参数axis =1 (axis英文是坐标轴的意思）

参数 numeric\_only = True表示只对行里面的数值格式进行求和，而不对其他类型进行求和

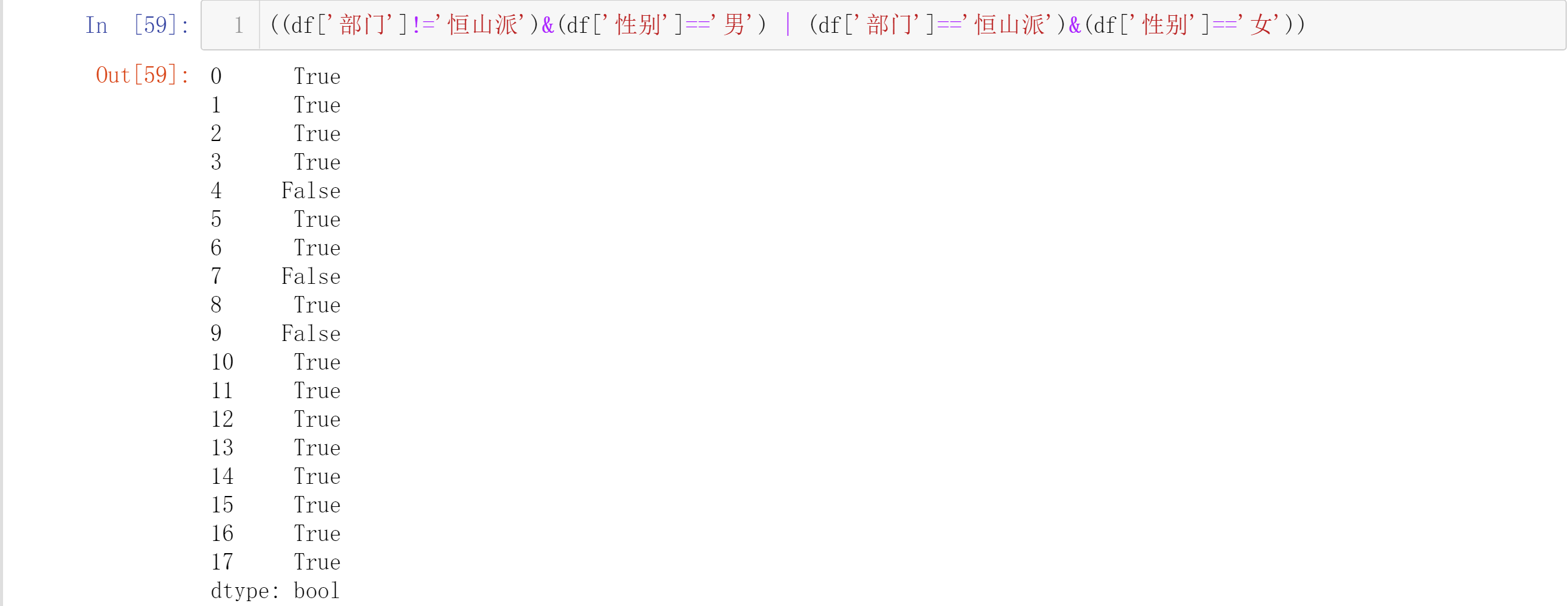
1. 这行代码实现的是单条件求和，其中groupby函数叫作分组函数，这里相当于分组求和复合函数，先用groupby对整个数据进行按‘部门’进行分组，这时部门变为索引列，然后方括号["总成绩"]表示只取一列的数据，最后面是求和sum函数，表示以部门作为条件，取总成绩列的合计值，如下：



5这个条件有点复杂，因为前面我是用EXCEL操作直接引用，没想到后面给自己挖了个坑，因为多条件求和的时候，指定的条件里是恒山派只取性别为女的，其他部门只选性别为男的，这行代码的结果就相当于进行了多条件筛选，如下：



& 表示逻辑，|表求逻辑或，(df['部门']!='恒山派')&(df['性别']=='男') 表求部门不等于恒山派且性别为男，(df['部门']=='恒山派')&(df['性别']=='女'))]表示部门等于恒山派且性别为女，然后用|或这个条件把这两个条件进行’打包封装“传递给前面的变量df,所传递的参数值是布尔值，也就是TRUE 或者FALSE,如果是TRUE则进行选取，如果是FALSE则不选取

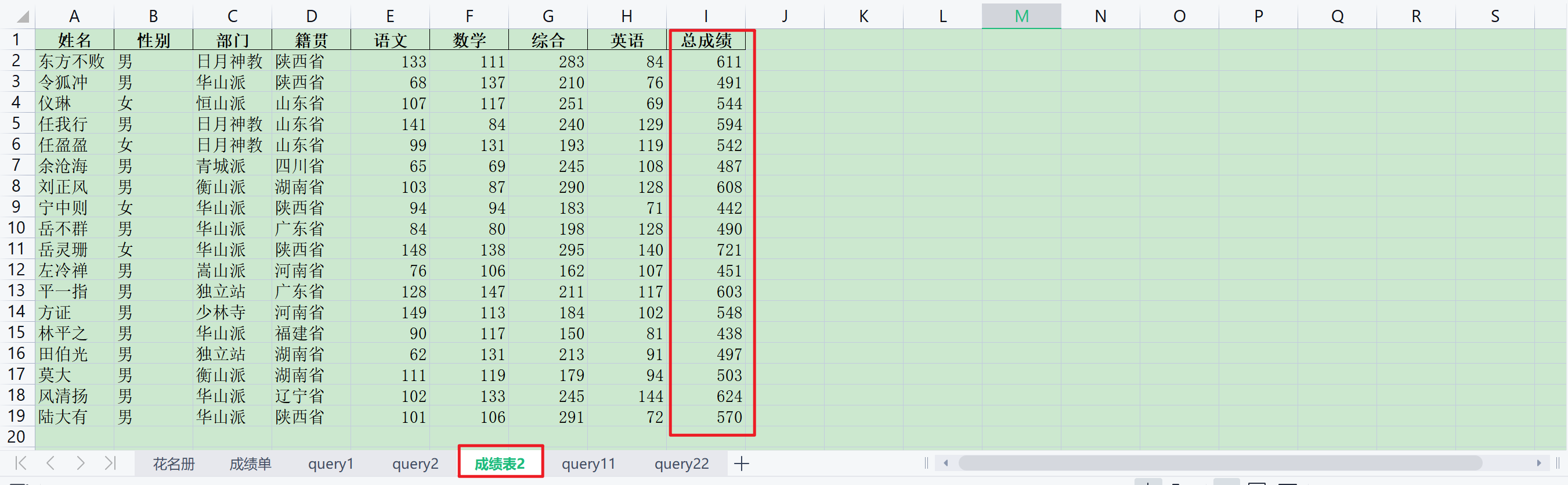


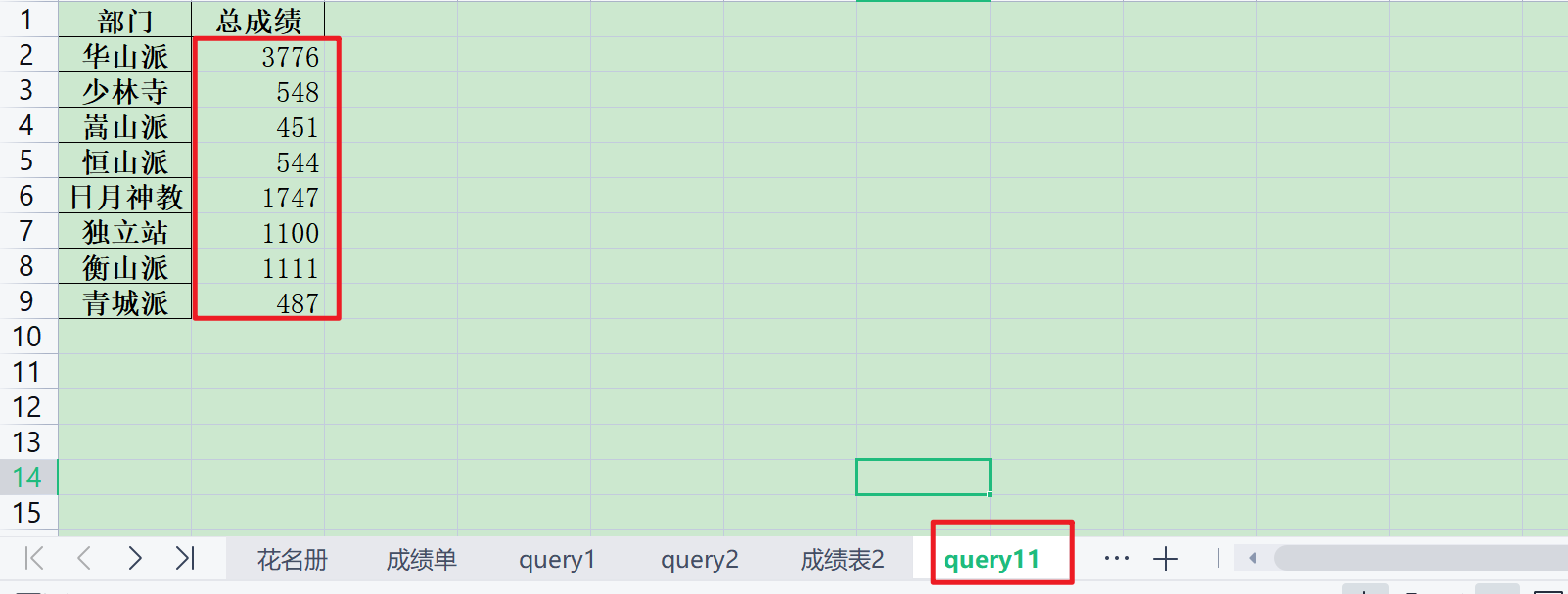
6表示对前一条进行多条件筛选后的数据进行多条件分组，第一个条件是部门，第二个条件是性别，除了在groupby()函数里面多了一个参数，其他的和前面单条件分组是完全一样的，结果如下：

7写入函数ExcelWriter,因为我们不改变原表，只是把新形成的表增加到原工作簿上，所以mode=”a”,增加模式，引擎用openpyxl

8-10:把用PYTHON得到的三个表数据增加写到EXCEL工作簿里面，sheet\_name=””为新表的命名

运行结果：在xiao工作簿新增加3个表格（成绩表2，query11,query22)，即为我们要求的结果







到此python完成SUM求和，SUMIFS单条件求和，SUMIFS多条件求和

附：如果直接在原表上修改，如何用代码实现，个人是不建议初学者直接更改原来，因为难免”代码一乱，损失百万”，一定要事先做好表格备份，不然如果代码出现一点偏差，可能运行后原表数据就没了

本例代码如下：

import pandas as pd

import xlwings as xw

df = pd.read\_excel("c:/study\_note/xiao.xlsx",sheet\_name = "成绩单")

df['总成绩'] = df.sum(axis=1,numeric\_only=True)

df1 = df.groupby(["部门"])["总成绩"].sum()

df2 = df[((df['部门']!='恒山派')&(df['性别']=='男') | (df['部门']=='恒山派')&(df['性别']=='女'))]

df2 = df2.groupby(['部门','性别'])['总成绩'].sum()

app = xw.App(visible=False,add\_book=False)

wb = app.books.open("c:/study\_note/xiao.xlsx")

wb.sheets[1].range("I1").options(index=False).value = df['总成绩']

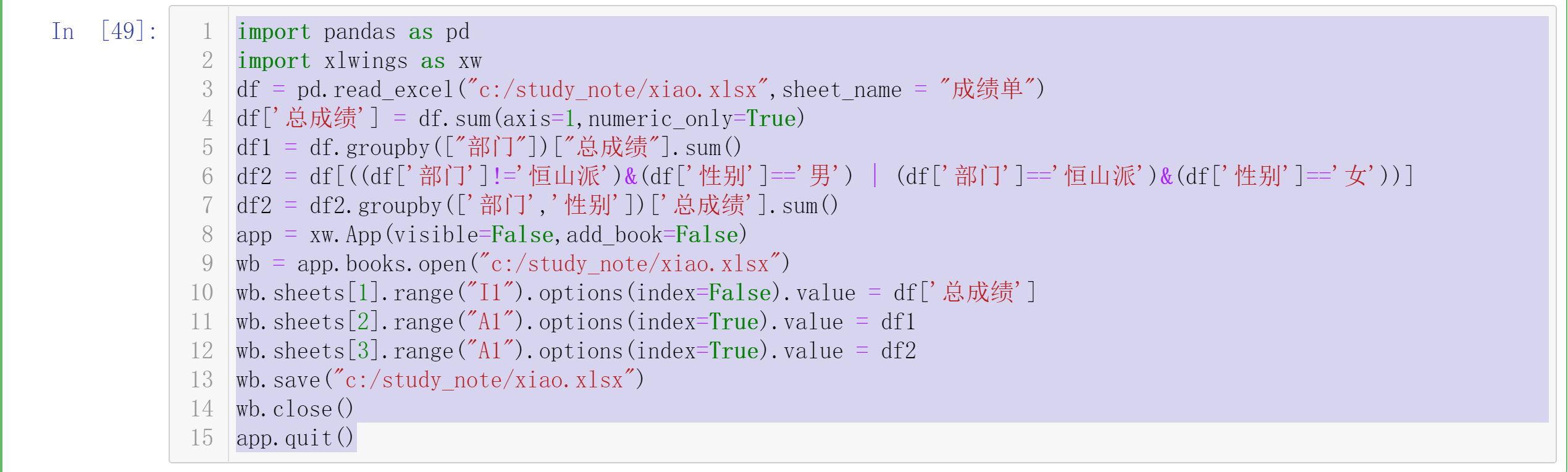
wb.sheets[2].range("A1").options(index=True).value = df1

wb.sheets[3].range("A1").options(index=True).value = df2

wb.save("c:/study\_note/xiao.xlsx")

wb.close()

app.quit()



代码说明：这里面是用第三方模块xlwings实现修改原表，它是一个专门处理EXCEL的python第三方库，wings是翅膀的意思，xl代表表格，这个库名本义可能就是让表格飞起来，看这个官方logo很明显是这个意思



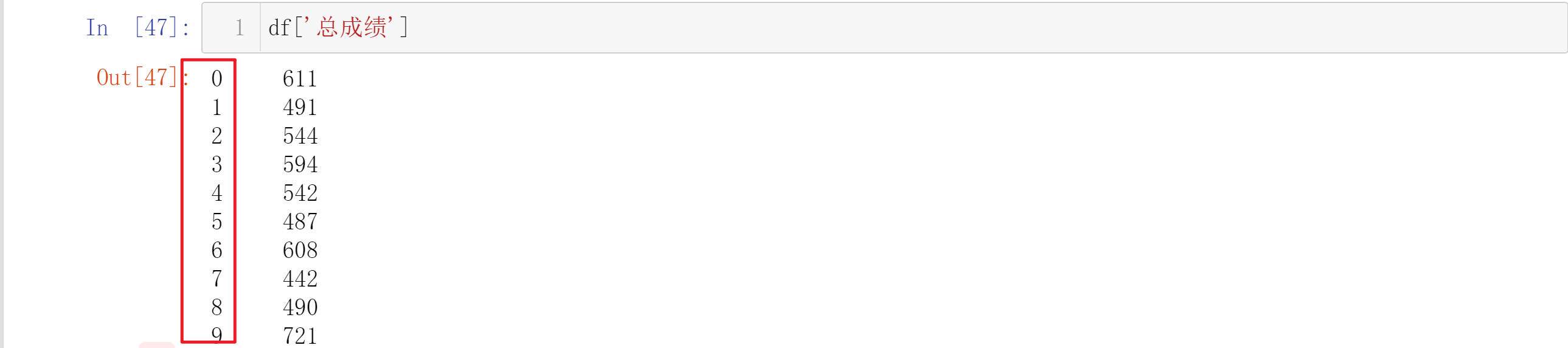
这个库的作者还写了一本excel + python的书籍，非常推荐大家可以看一看

我们直接从第8行代码说：

8表示启动EXCEL程序，visible=False,表示在后台运行，不在前台显示，,add\_book=False表示不新建工作簿

9打开工作簿xiao.xlsx,并把这个打开的工作簿对象赋值给变量wb，如下图，可以看到这个变量代表什么：

10.wb.sheets[1] 这个表示工作簿wb里面的第二个表格，从左到右，我们这里取成绩单表，排在工作表里面第二个，同理后面两行代码表示第3个，第4个的意思，range("I1")表示要修改表数据的区域，如果改的是一个区域，则从最左上角向下和向右进行填充，这里我们选I1单元格，是对应取总成绩那列的首个单元格，options(index=False)这个选项参数是忽略索引的意思，如果不忽略的话，修改进来就会把数据里面的默认索引编号也带进来，如下图，最后.value 就是默认赋值语法，最终的结果是把df['总成绩'] 这一列值写入到第二个表里的从I1单元格开始向下的列里面



1. 代码同第10行的用法一样，唯一不同的是这里参数options(index=True)表示取索引的意思，因为这个赋值的数据源是由函数GROUPBY生成的数据，前面的条件会作为索引列显示，如下图，所以我们需要这个索引列作为我们的查询结果，还有一种方法就是参数依旧不选索引，先把数据源里的索引进行还原为普通列，麻烦一点，实现结果一样



12代码作用同第11行

13保存工作簿到指定路径和指定名称，相当于在EXCEL界面点保存按钮

14关闭工作簿

15退出EXCEL程序

1. SQL实现方法：

首先我们把成绩单这个表在数据库里进行创建：

**CREATE** **TABLE** score

(

姓名 **VARCHAR**(5),

性别 **CHAR**(1),

部门 **VARCHAR**(4),

籍贯 **VARCHAR**(8),

语文 **INT**,

数学 **INT**,

综合 **INT**,

英语 **INT**,

总成绩 **INT**

);

然后用图形化界面把EXCEL里的原始数据导入进数据库里，得到如下结果：



总成绩即为我们要求和的区域

然后再创建两个条件求和的表格，分别命名为query2,query3,代码如下：

**CREATE** **TABLE** query2

(

部门 **VARCHAR**(4),

总成绩 **INT**

);

**INSERT** **INTO** query2 部门 **SELECT** **DISTINCT** 部门 **FROM** score;

**CREATE** **TABLE** query3

(

部门 **VARCHAR**(4),

性别 **CHAR**(1),

总成绩 **INT**

);

**INSERT** **INTO** query3 (部门,性别)

**SELECT** **DISTINCT** 部门,性别 **FROM** score **WHERE** (部门='恒山派' **AND** 性别='女' )**OR** ( 部门!='恒山派' **AND** 性别="男" );

得到如下两张表：





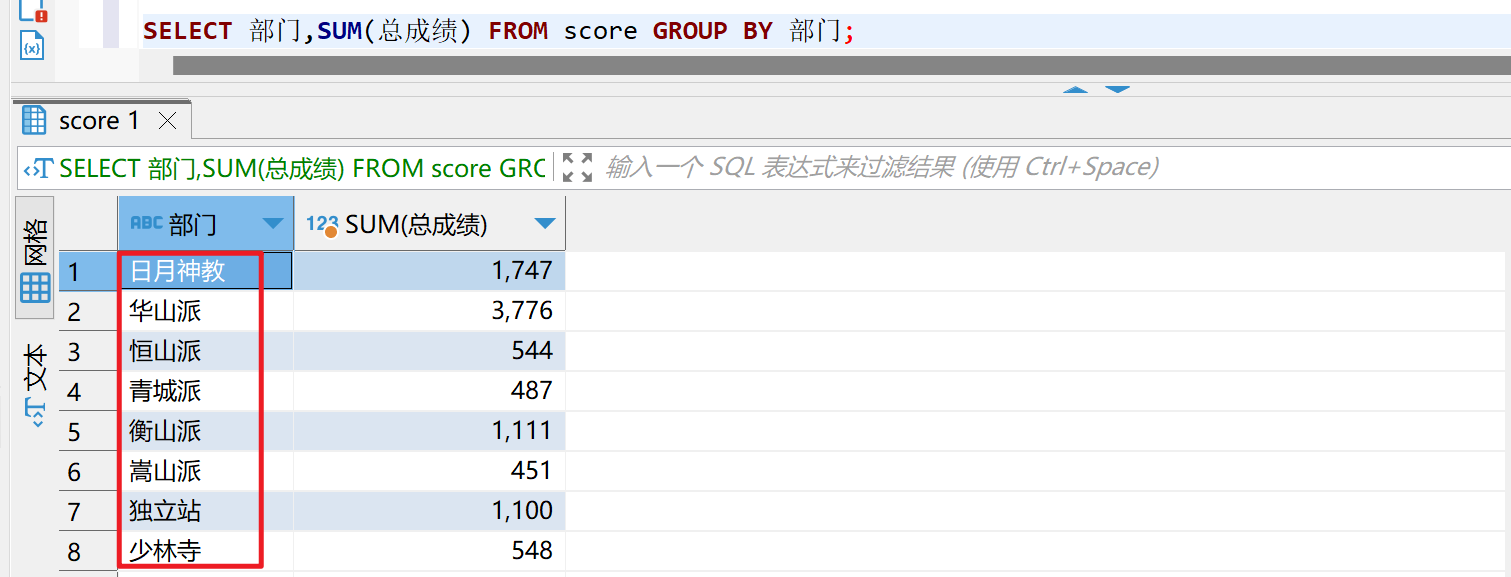
空白区域即为我们要填入的条件求和区域

1按行求和代码：完成总成绩求和

**UPDATE** score **SET** 总成绩 = 语文+数学+综合+英语;



第二问题：单条件求和代码：**SELECT** 部门,**SUM**(总成绩) **FROM** score **GROUP** **BY** 部门;



先将这个查询结果保存到一个临时表，然后用代码更新到我们的查询表query2里面

**CREATE** **TABLE** temp1

(

部门 **VARCHAR**(4),

总成绩 **INT**

);

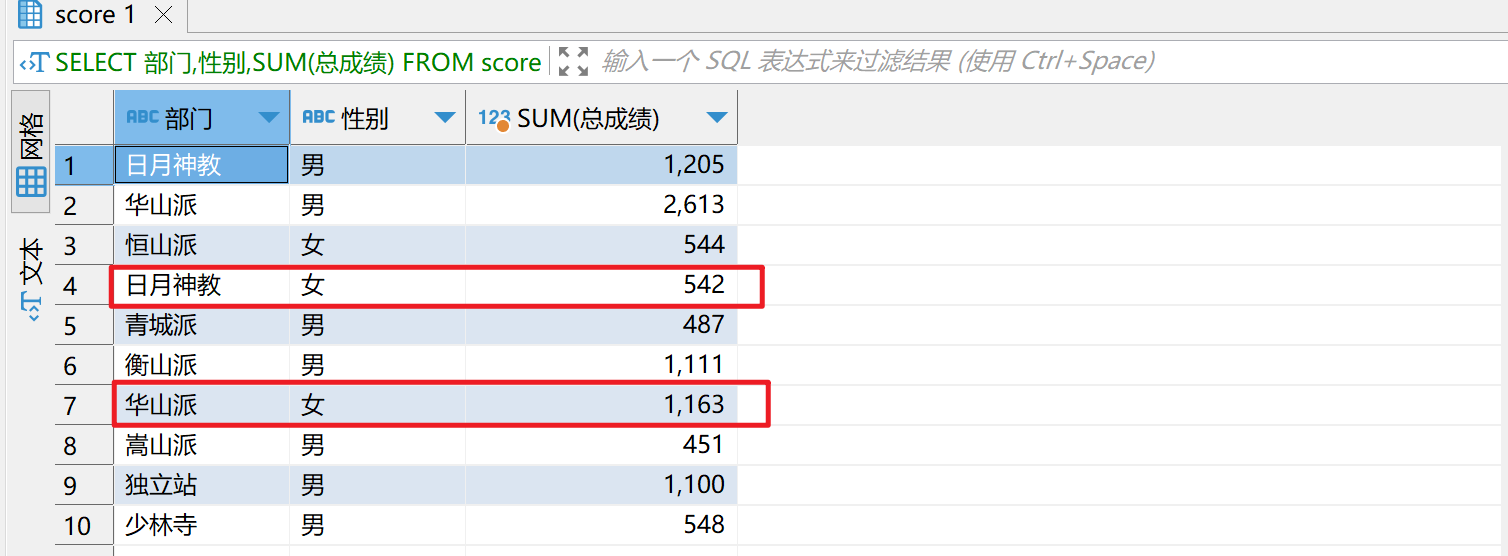
**SELECT** \* **FROM** temp1 ;

**INSERT** **INTO** temp1 **SELECT** 部门,**SUM**(总成绩) **FROM** score **GROUP** **BY** 部门;

**UPDATE** query2,temp1 **SET** query2.总成绩 = temp1.总成绩 **WHERE** query2.部门 = temp1.部门;

第三个问题：多条件求和,代码：

**SELECT** 部门,性别,**SUM**(总成绩) **FROM** score **GROUP** **BY** 部门,性别;



我们把这个结果保存到query3表里面，用取巧的办法可以把其他两个门派里面的性别为女的行删除即可，或者把查询结果保存到一个临时表，然后用UPDATE把值更新到query3表

代码：

**CREATE** **TABLE** temp2

(

部门 **VARCHAR**(4),

性别 **CHAR**(1),

总成绩 **INT**

);

**INSERT** **INTO** temp2 **SELECT** 部门,性别,**SUM**(总成绩) **FROM** score **GROUP** **BY** 部门,性别;

**UPDATE** query3,temp2 **SET** query3.总成绩 = temp2.总成绩 **WHERE** query3.部门 = temp2.部门 **AND** query3.性别 = temp2.性别;

**SELECT** \* **FROM** query3;

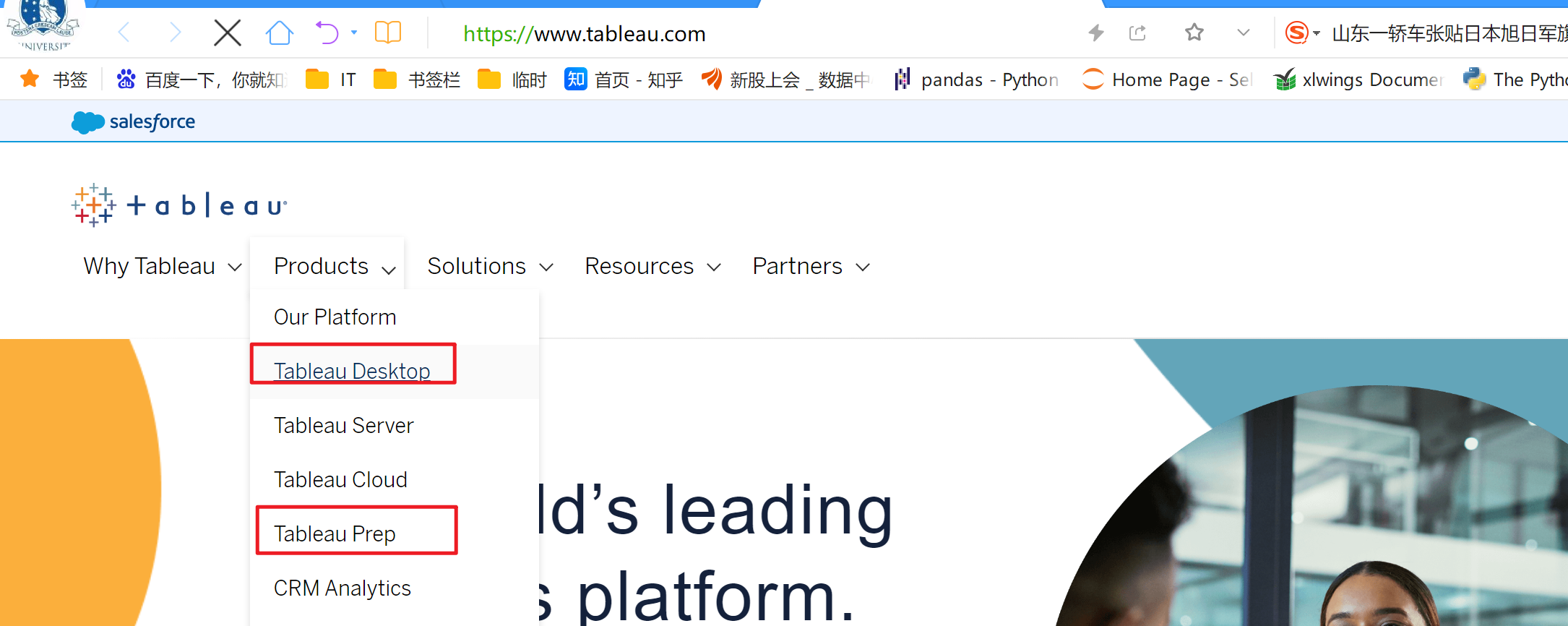


至此，SQL完成任务

1. Tableau软件实现：

Tableau是一个非常有名的可视化软件，可对海量数据进行可视化输出，有的同学可能会有疑问，这个软件不是生成可视化图表的吗，还能进行数据处理？答案是可以的，用星爷电影里的台词讲就是“表面上是一个吹风机，实际上是剃须刀”，只有想不到，没有做不到，Tableau完全可以对表格数据处理并输出结果

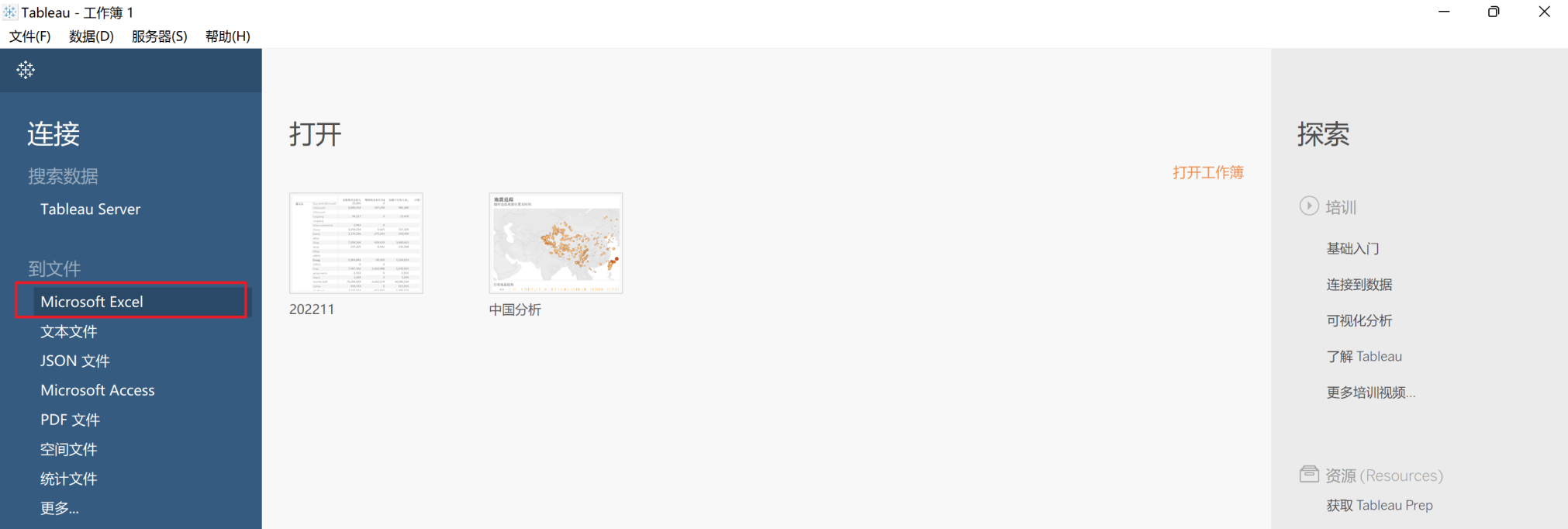
先到官网下载 Tableau Desktop软件，这个有15天的免费试用期，如果是在校学生或者有在职学历教育的同学可以申请免费使用一年，上传学生证及身份证等，届时官方会用邮件发过来一个KEY码，激活即可用一年，如果过了试用期还想用，那就上某宝找卖家



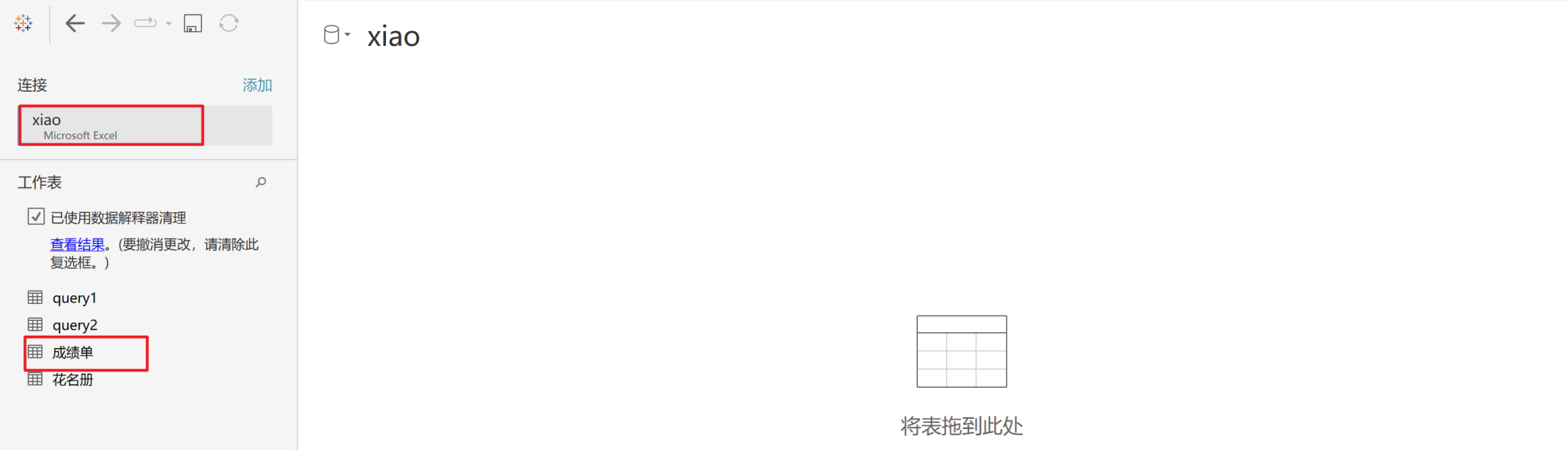
找到官网地址，下载DESKTOP，如果有想进一步学习的同学可以顺手把Prep也一起下载，这是一个为Tableau先预处理数据源的工具，安装好后打开软件：那个白色的图标软件



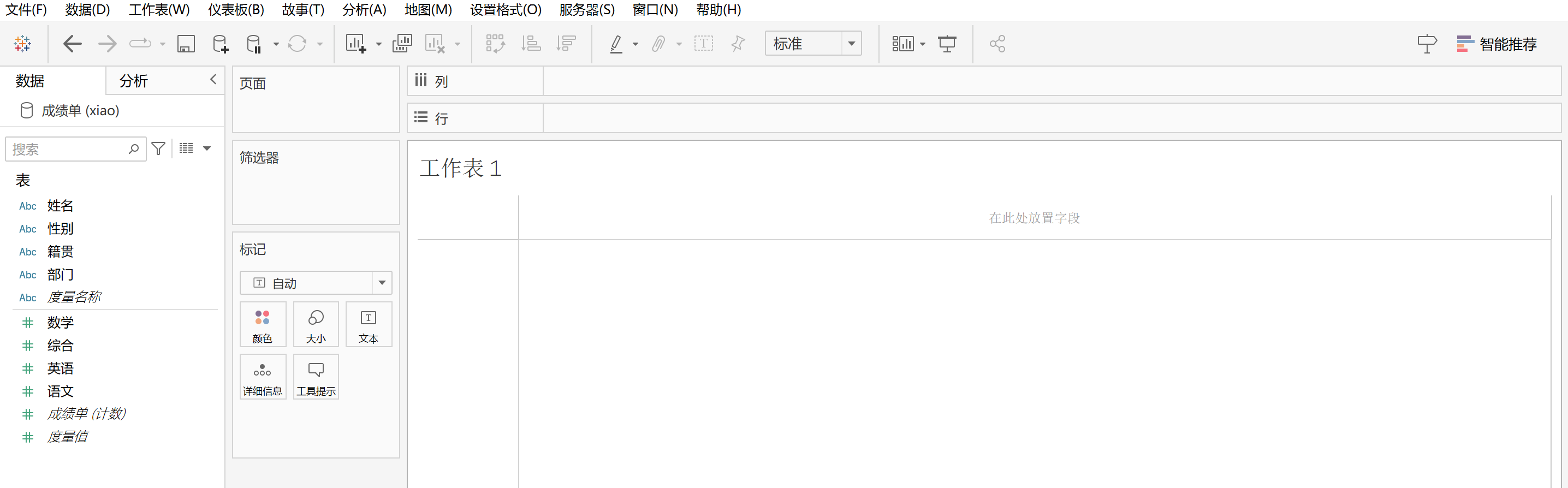
选择左边连接- 到文件 - EXCEL 那项，然后在电脑里找到我们要处理的成绩单表格(注：要先把原始数据里面要求和的总成绩字段删掉，我们后面在Tableau里面用创建计算字段的形式进行实现）



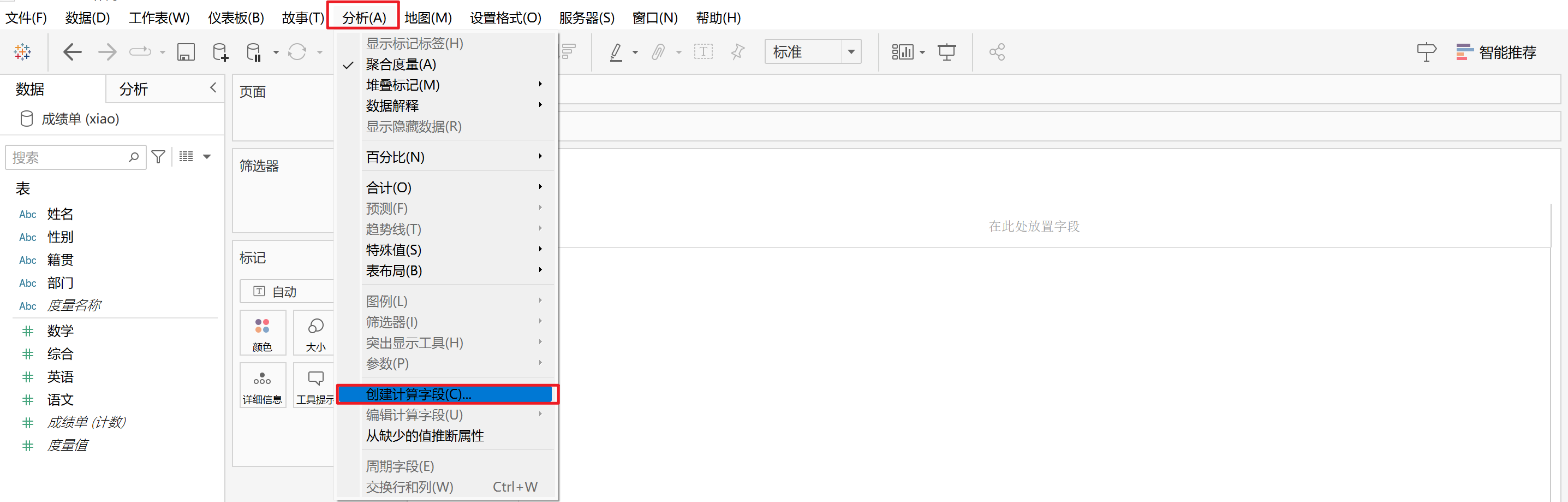
建立好连接，我们把要处理的成绩单拖到右边空白区域



下面是数据预览，连接模式选实时即可，然后用鼠标点左下角“工作表1”，生成空白主界面，左边是数据字段项，分为度量名称（维度），和度量值（数值）

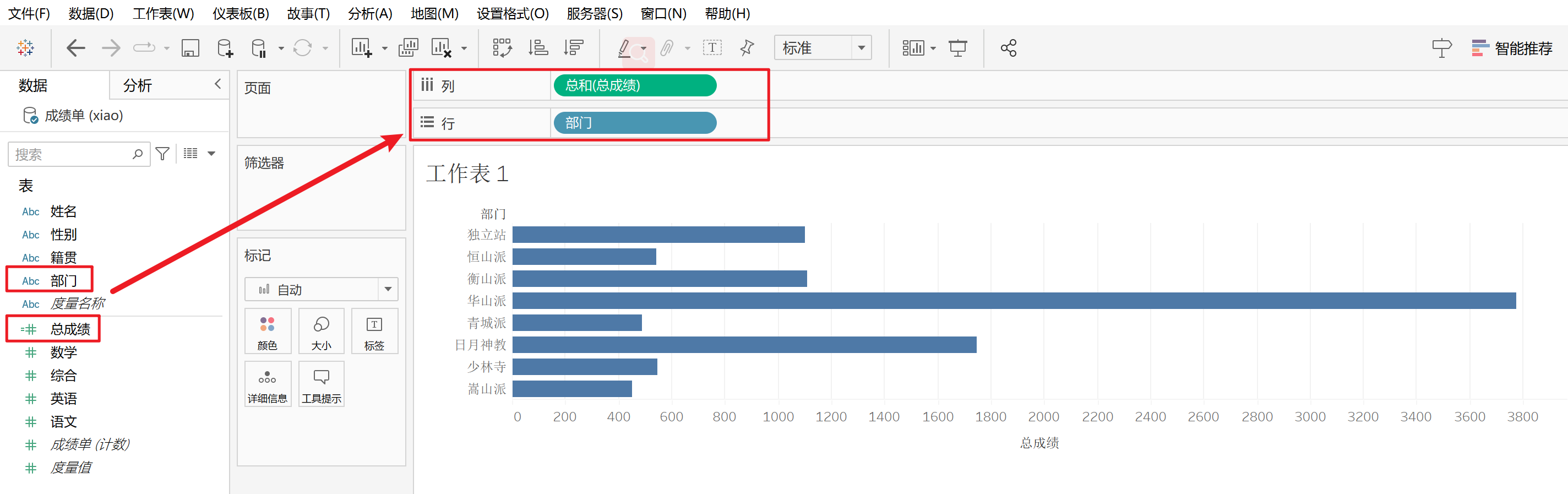


因为这里没有总成绩字段，我们需要在分析- 创建计算字段，创建一个总成绩字段

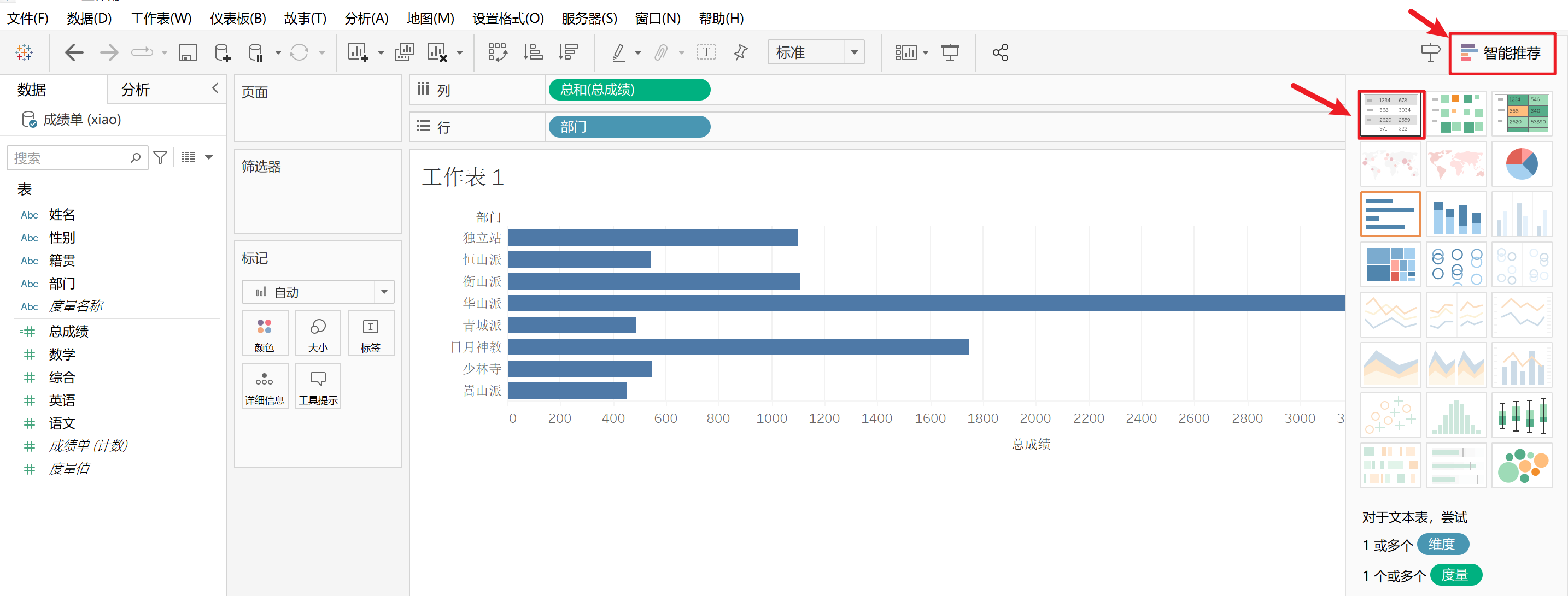


新字段命名为总成绩，然后把左边字段里的数值字段名用鼠标拖到右边的计算框里，用加号相连，下面显示计算有效，点应用

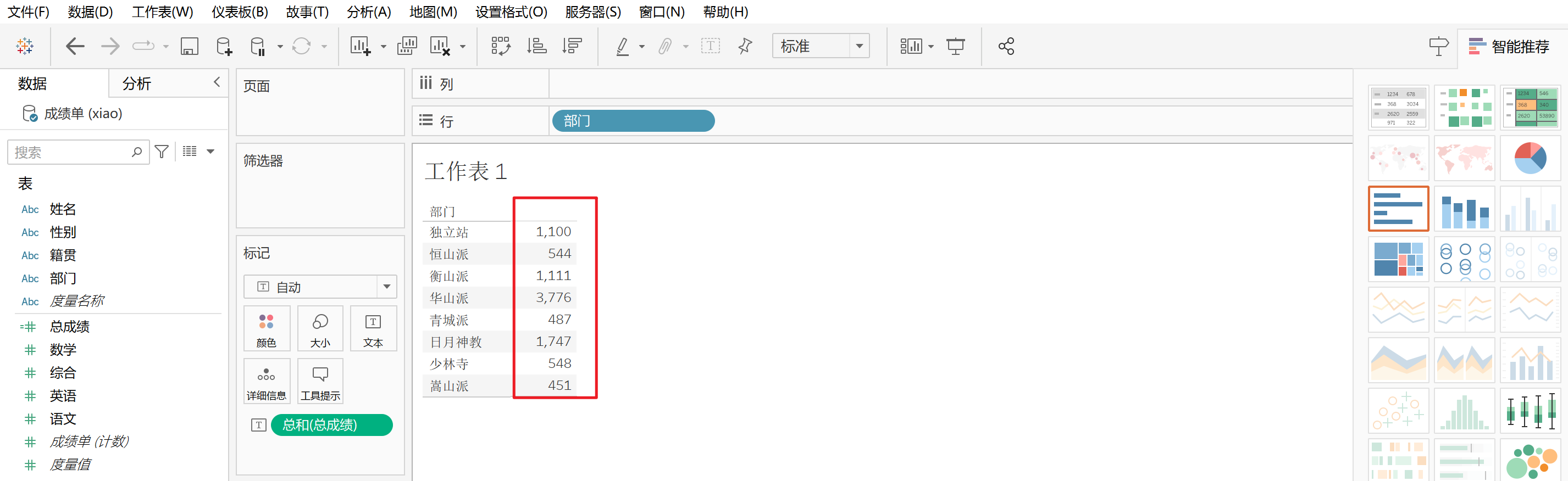
我们把新生成的总成绩字段拖到正上方的列空格里，把部门拖到行空格里



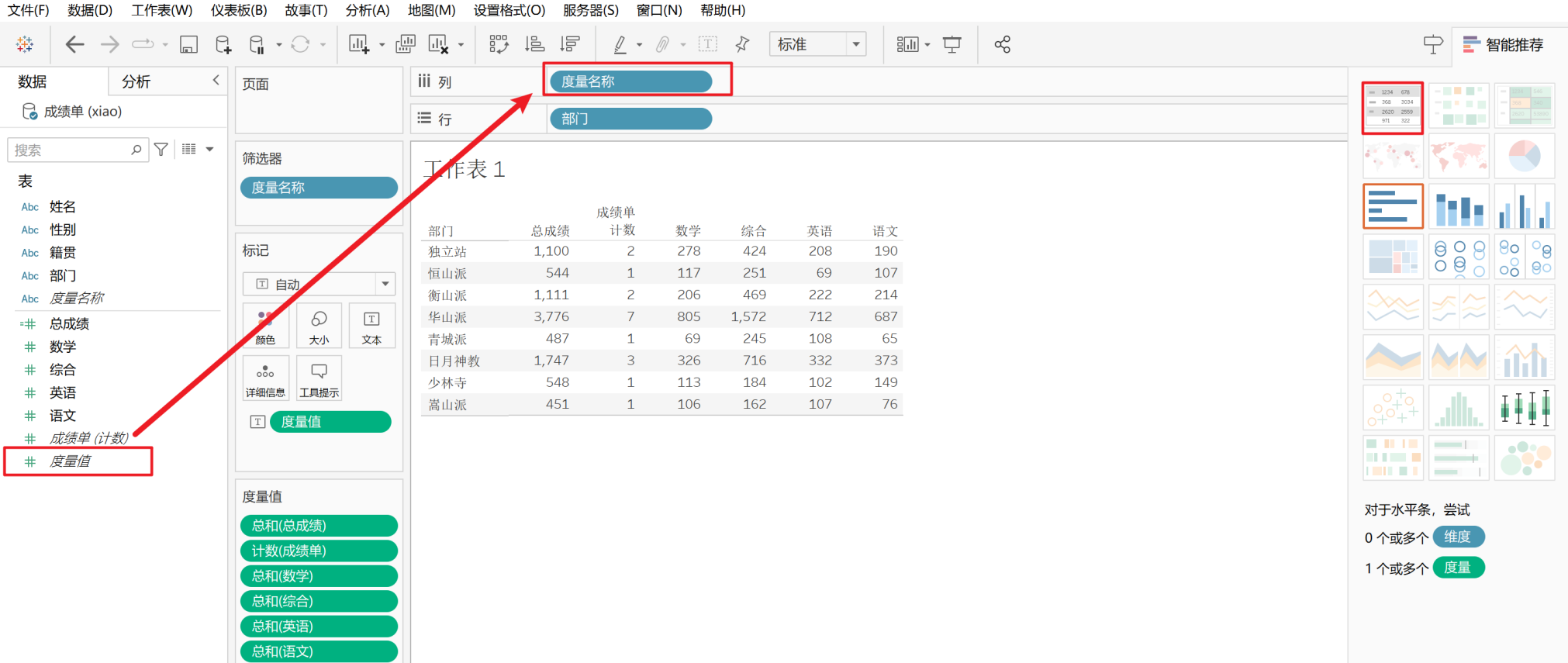
因为上面生成的是自动匹配的可视化图表类型-条形图，我们需要让它显示为表格形式，鼠标点右上角‘智能推荐’，然后选左上角那个表格形式



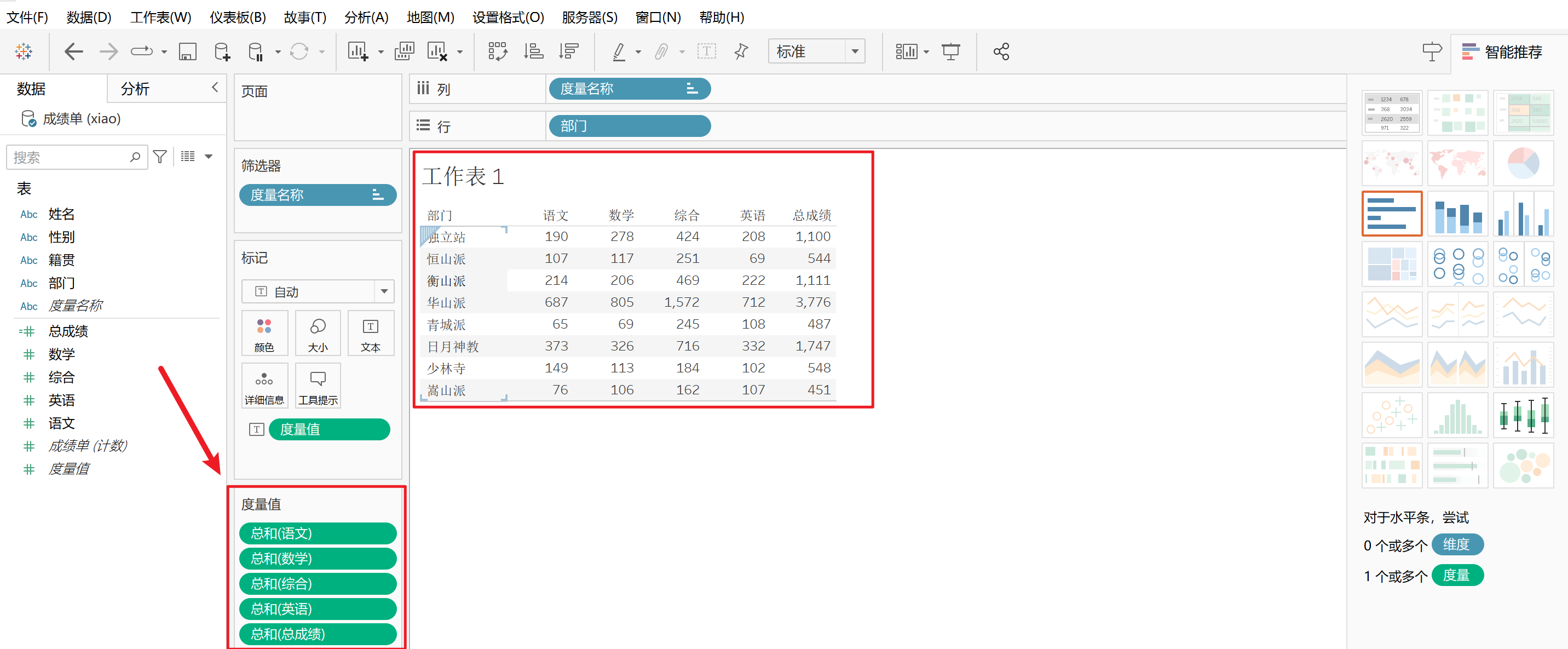
最终生成的图像是



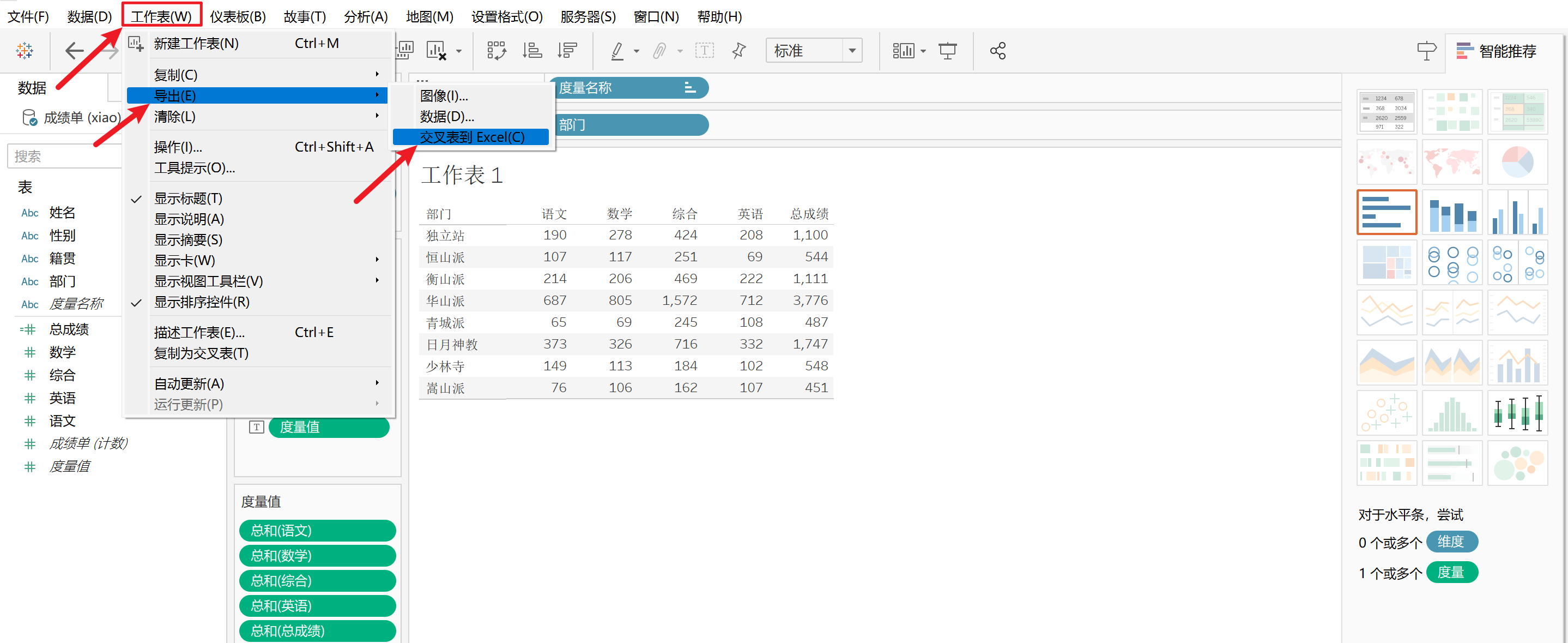
这个第二列虽然是我们的求和结果，但是并没有显示字段名，效果不佳，我们改善一下，左下角的度量值拖到列空白处，然后再选表格形式

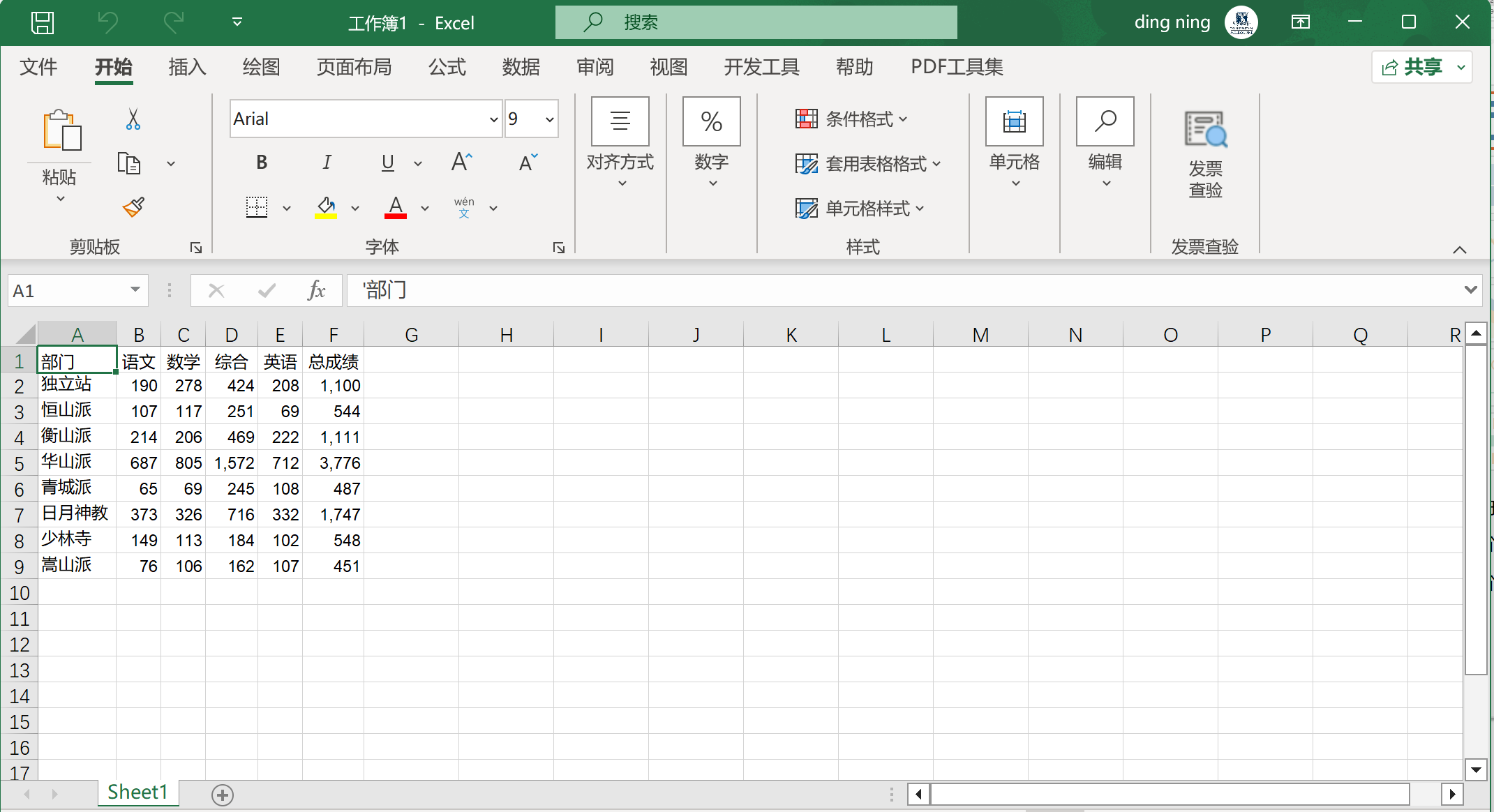


我们把左下角方框里面的成绩单删除，然后用鼠标拖动里面的名称改变一下顺序，最终结果如下：

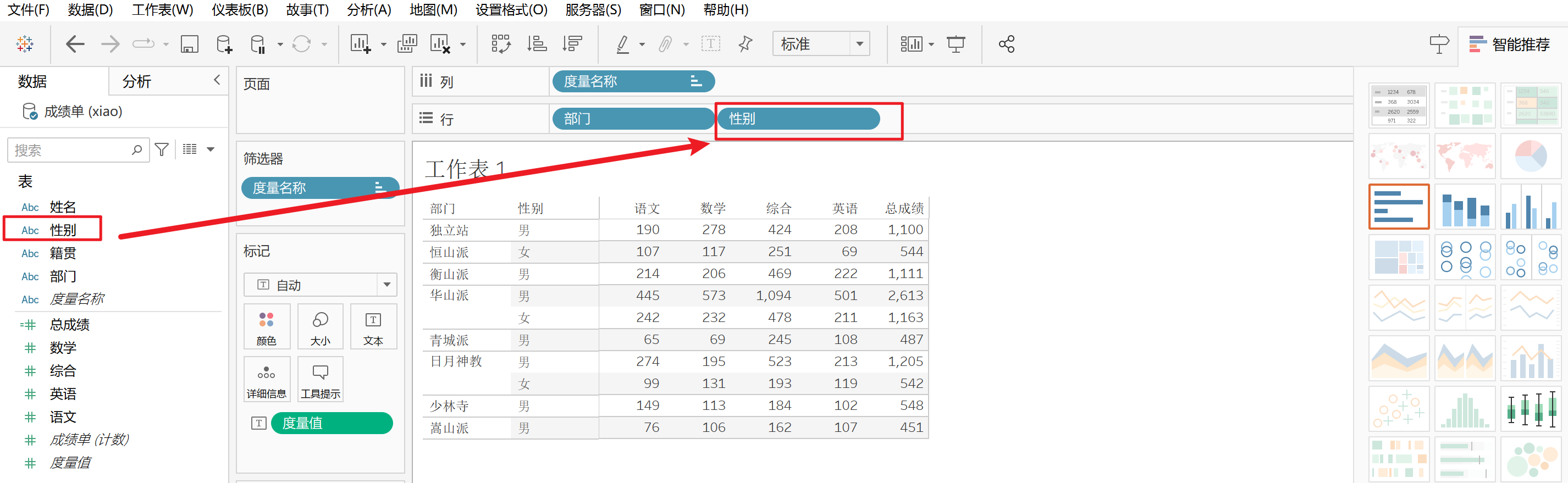


最后一步导出到EXCEL表格，选工作表 - 导出 - 交叉到EXCEL 即可实现导出





第三个问题多条件查询,就是把左边性别字段移到行空格部门后面，即可实现，然后导出，直接在表格里进行筛选操作即可



补充一下前面第一个问题，其实就是把里面的名称字段全都拖到行空格处，如下图，这个表就是我们第一张对每个同学进行成绩求和的表，然后导出为表格即可完成