**Power BI爬虫技巧：如何获取网页中的链接？**

对于网页中可见的规范化数据，提取很简单，但对于网页中可以再次点击的链接，怎么提取呢？

在抓取之前，选择菜单“文件”下面的“选项和设“选项”，点击“预览功能”，将“新web表推理”选中。

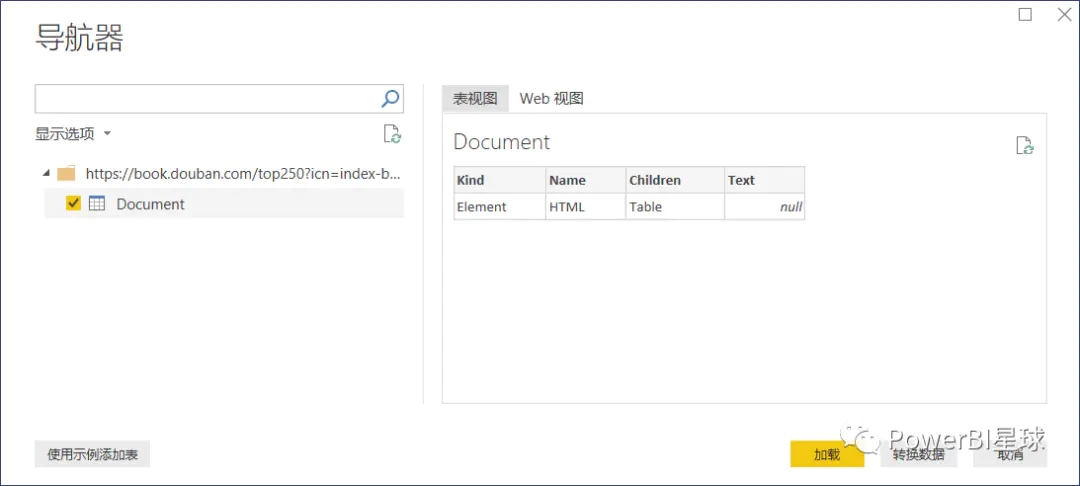


接下来本文以豆瓣读书中的TOP250 为例：

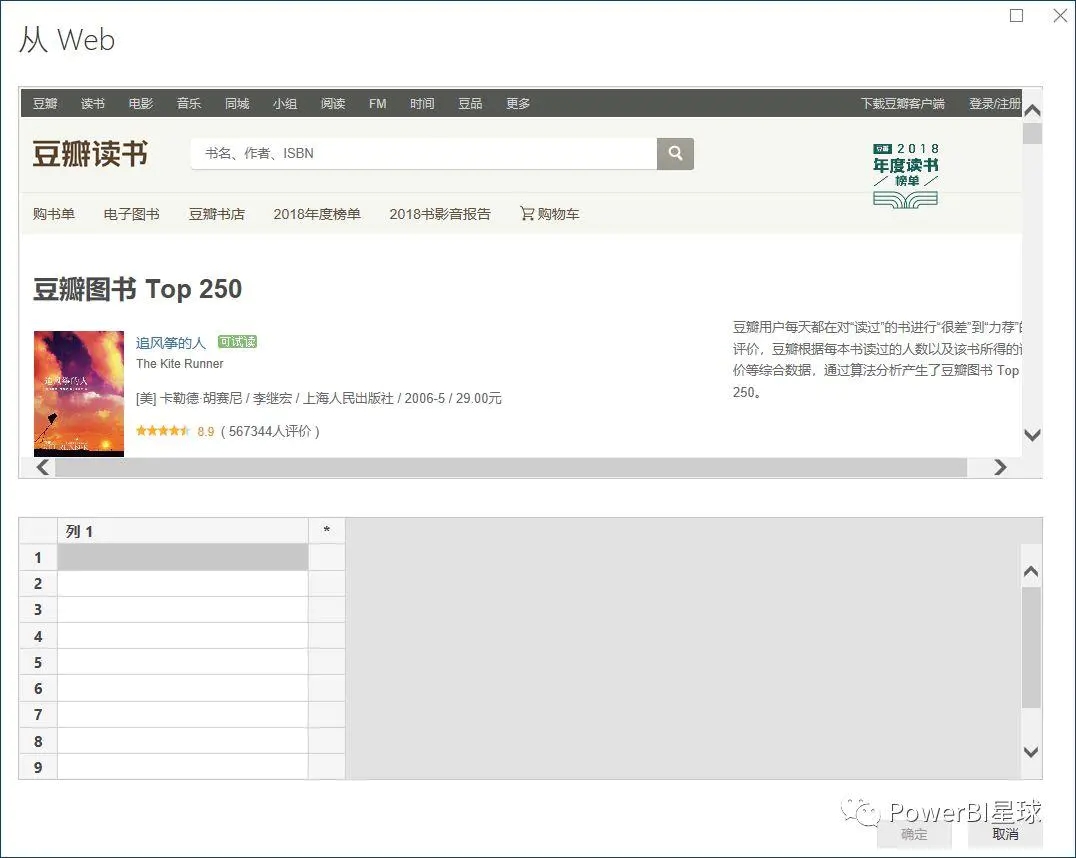
https://book.douban.com/top250?start=0

在这个网页中，不仅显示书名、评分、作者等信息列表，还可以通过点击封面或者书名，进入该书的详情页，下面看看如何提取这个链接。

在Power BI Desktop中，选择用web获取数据，会看到这个导航器，



在表视图中你看不到可以提取的数据，没有关系，你可以点击左下角的“**使用示例添加表**”，然后你就能看到这个网页了，



在这里，只要手动输入前两条信息，PowerBI就会判断你要提取的字段，并自动把该网页中剩余的同类数据添加进来，比如输入前两个书名，

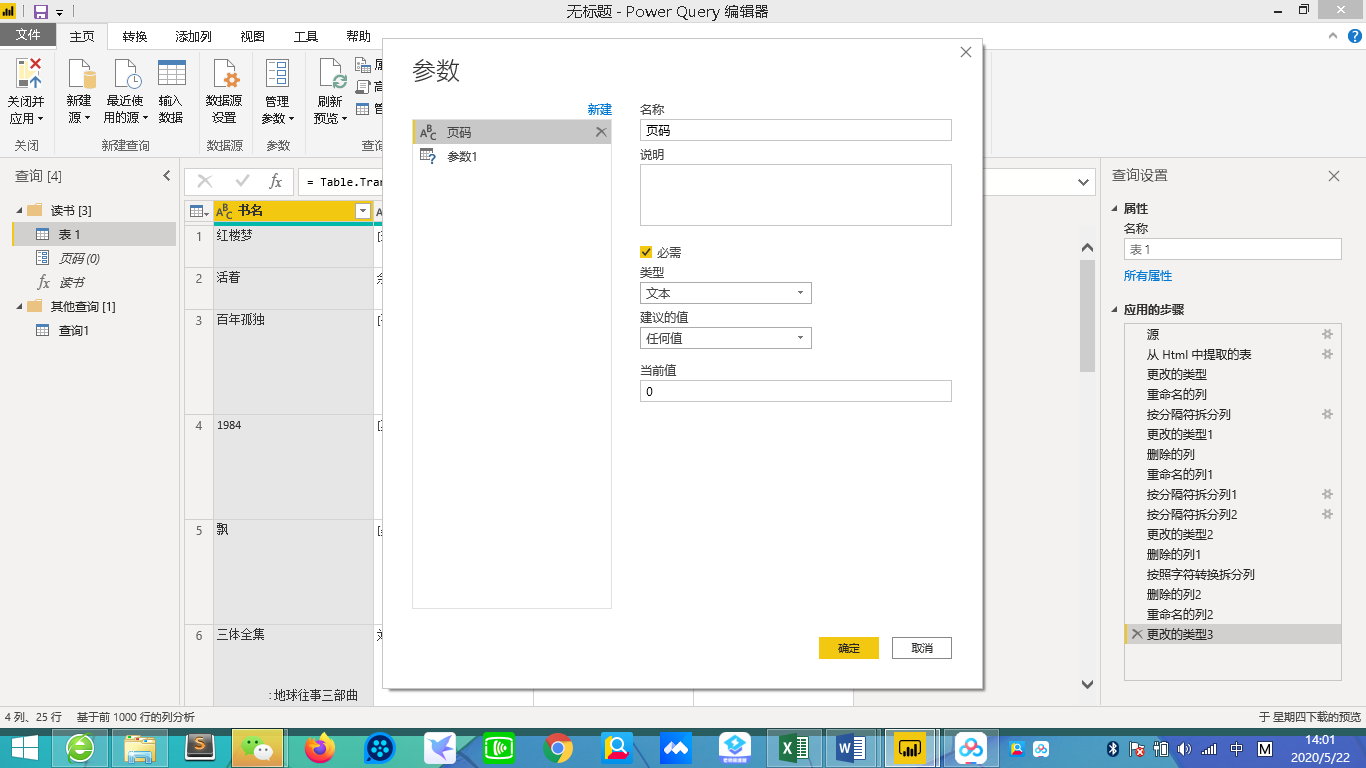
同样的方式，你也可以提取评分、作者、出版社等信息。

网址在这个网页是不可见的，你没法直接输入前两个网址，这种情况下就没有示例，难道要一个一个手工复制粘贴吗？

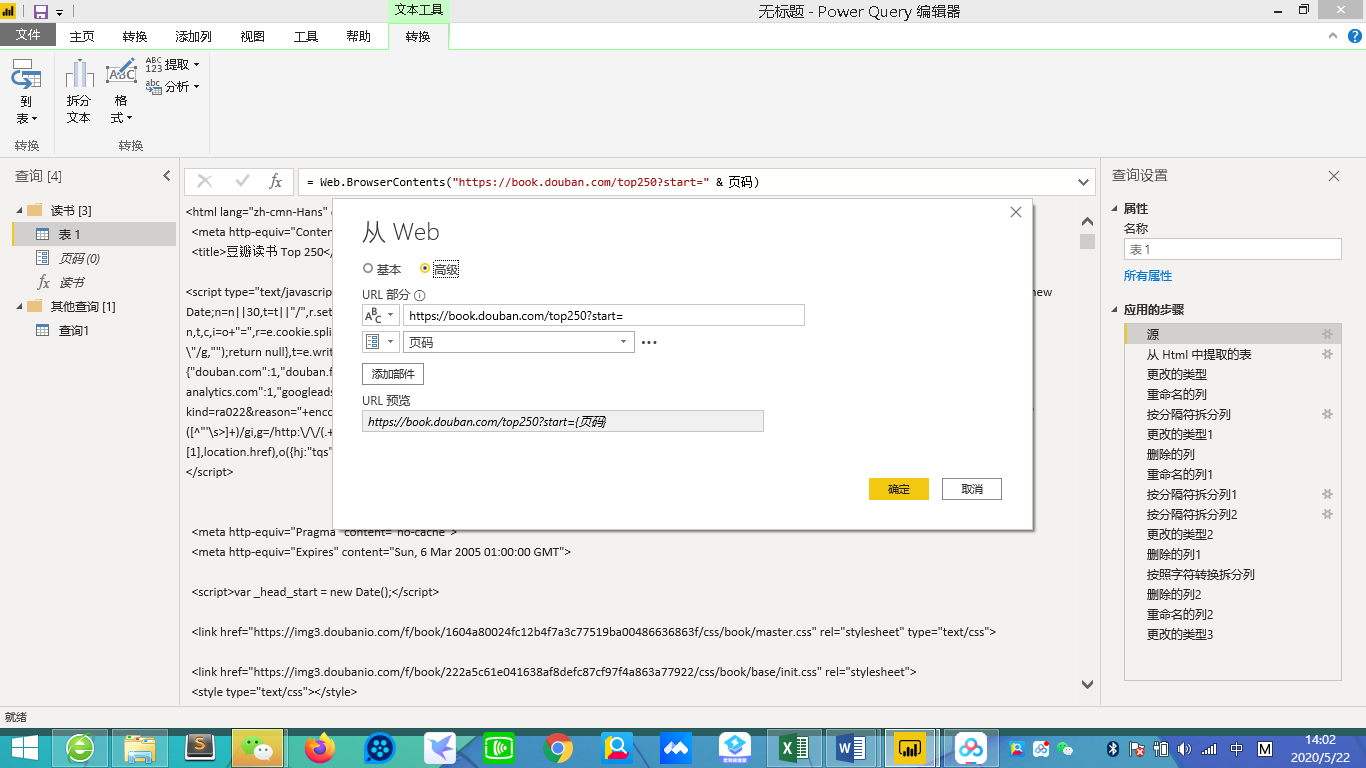
当然不用，虽然这里看不见，但可以把链接打开，不就知道网址是什么了吗？

**依次打开把前两名的链接，并分别将网址复制粘贴到示例的前两行中，如此等等就可以把所见的信息下载下来。**

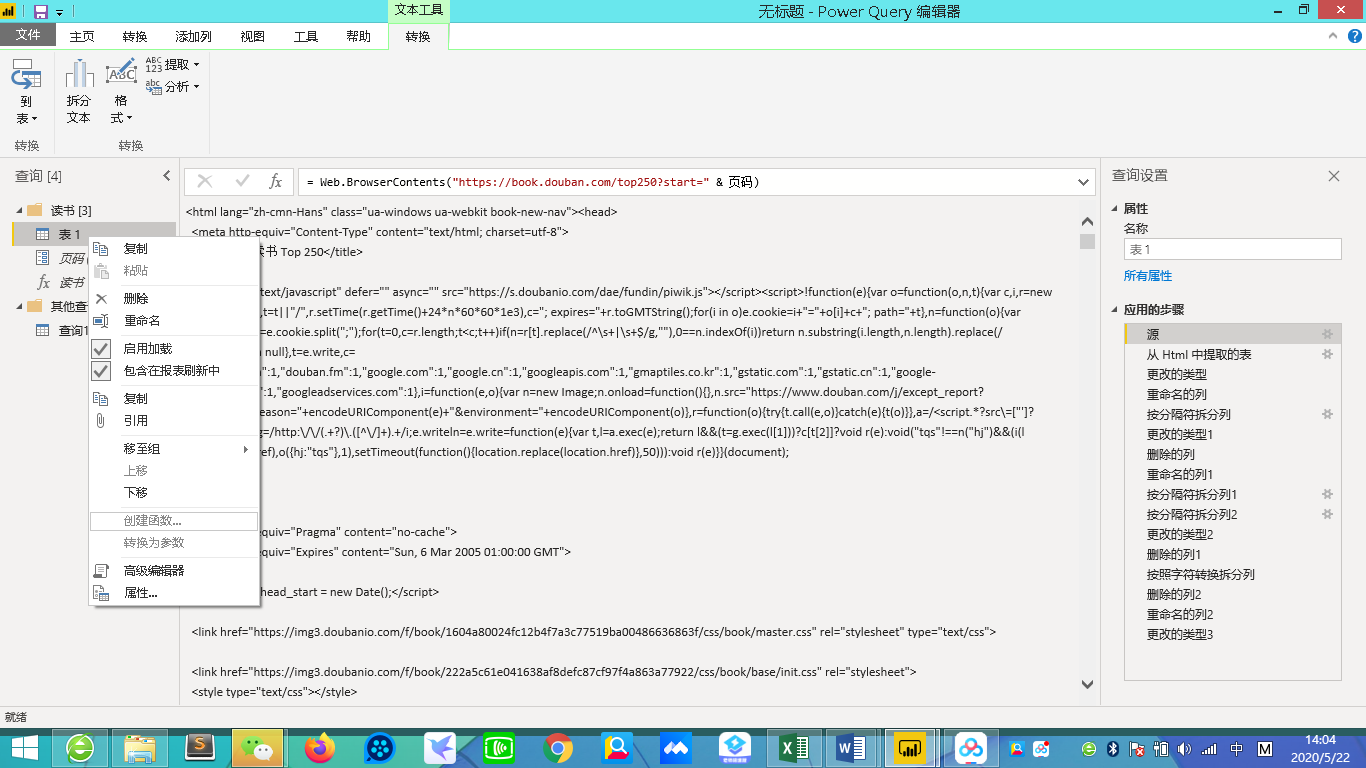
**选择“转换数据”，不选择“加载”进入Power Query界面。**

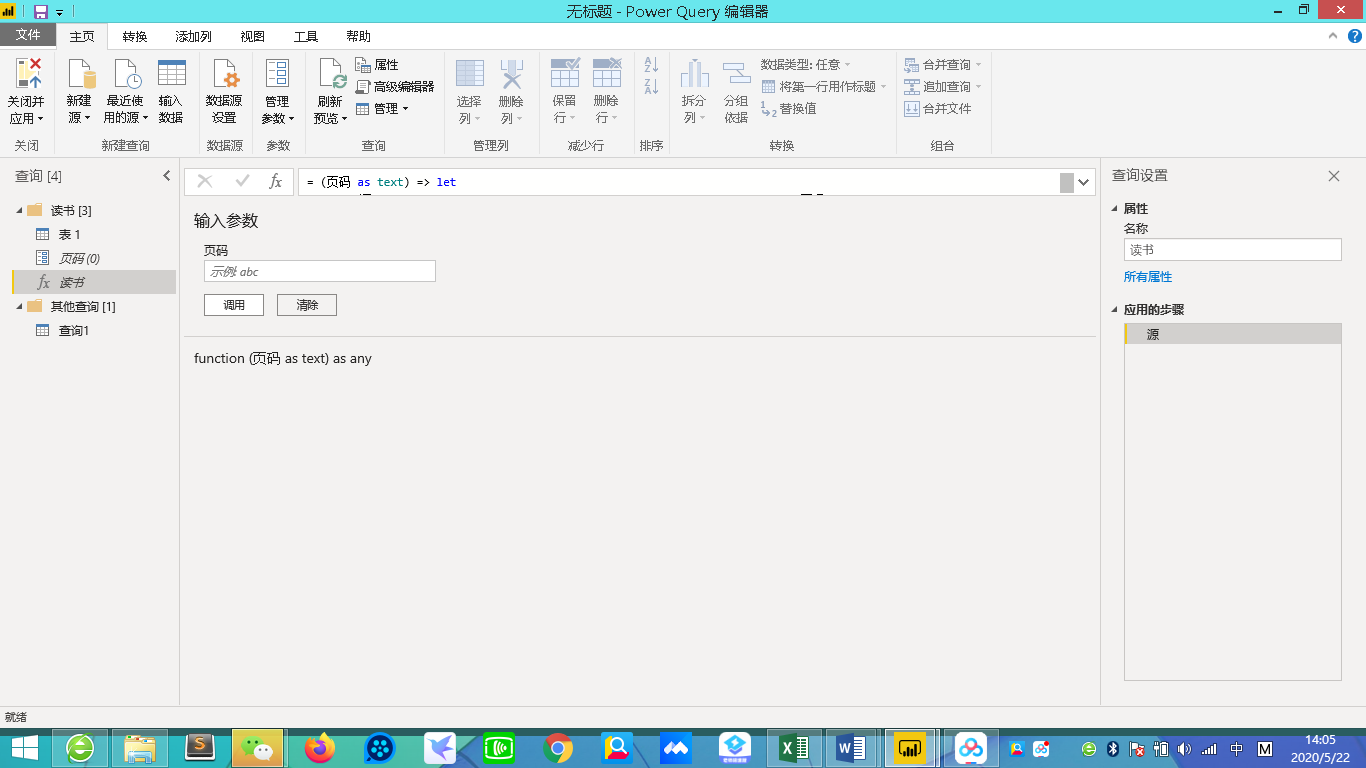
**选择“主页”的“管理参数”下面的“新建参数”，建立一个“页码”的参数，类型为“文本”，初始值为“0”。，见下图。**

**建立参数后，将参数引入网页。具体操作如下：点击右边的“源”的齿轮，弹出web对话框，选择“高级”，将网页中最后的数字去掉，在下一行选择“参数”页码，确定。**

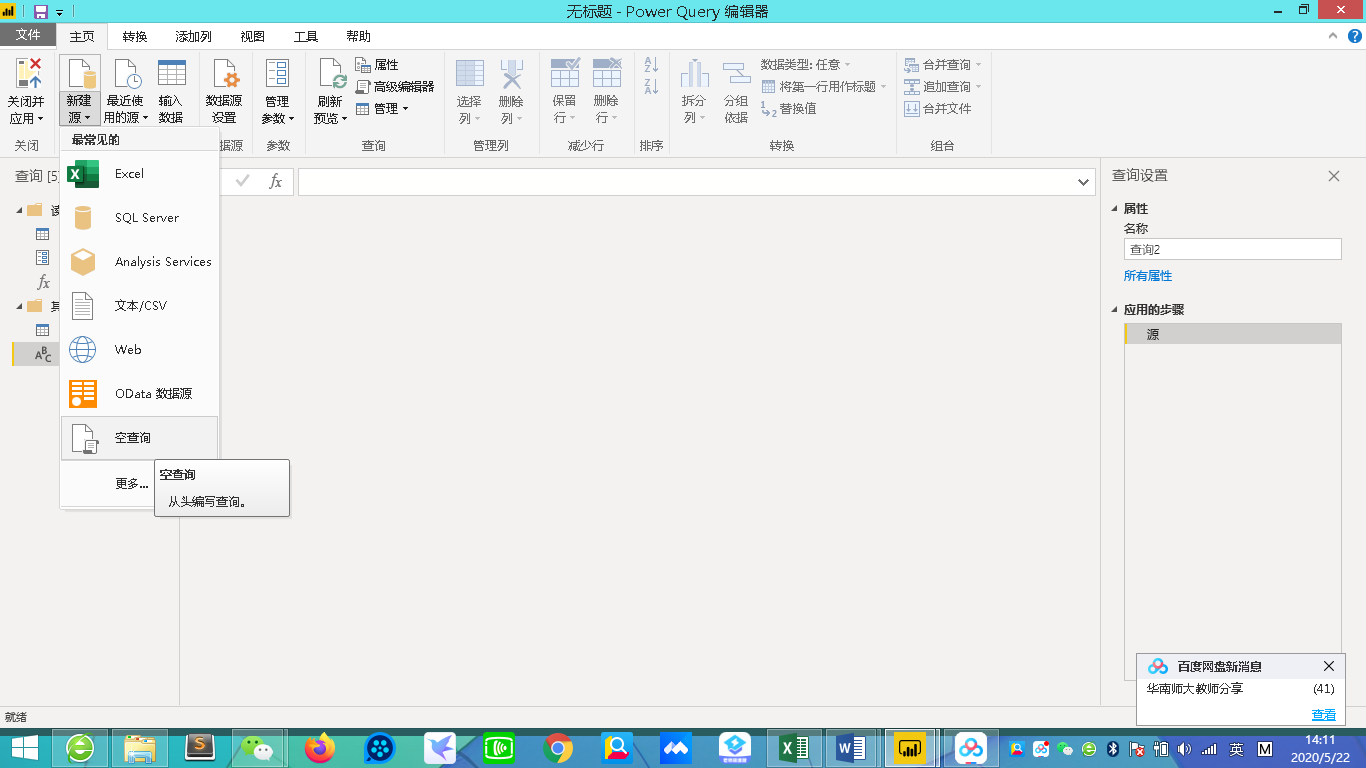


**右键点击左边的“表1”，选择“创建新函数”。**





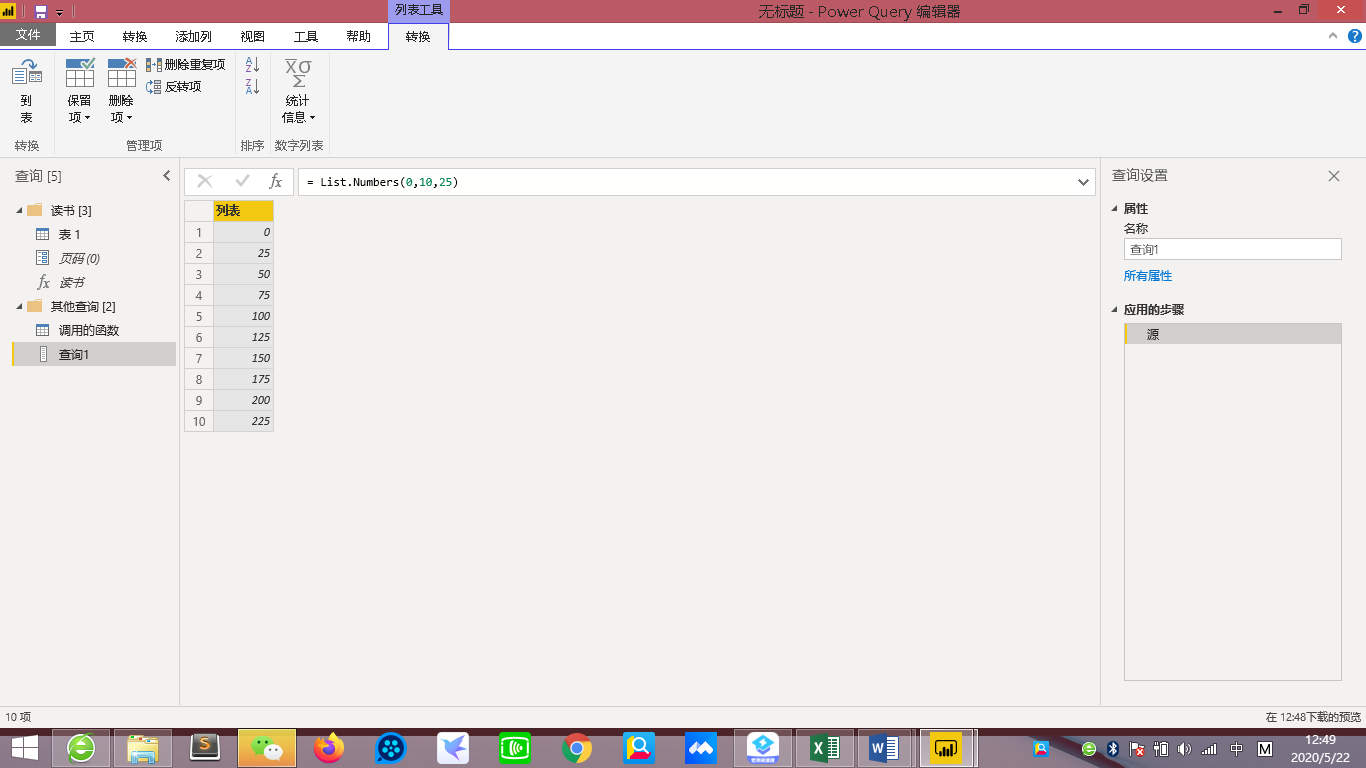
**创建新函数后，选择“主页”的“新建源”下面的“空查询”。**



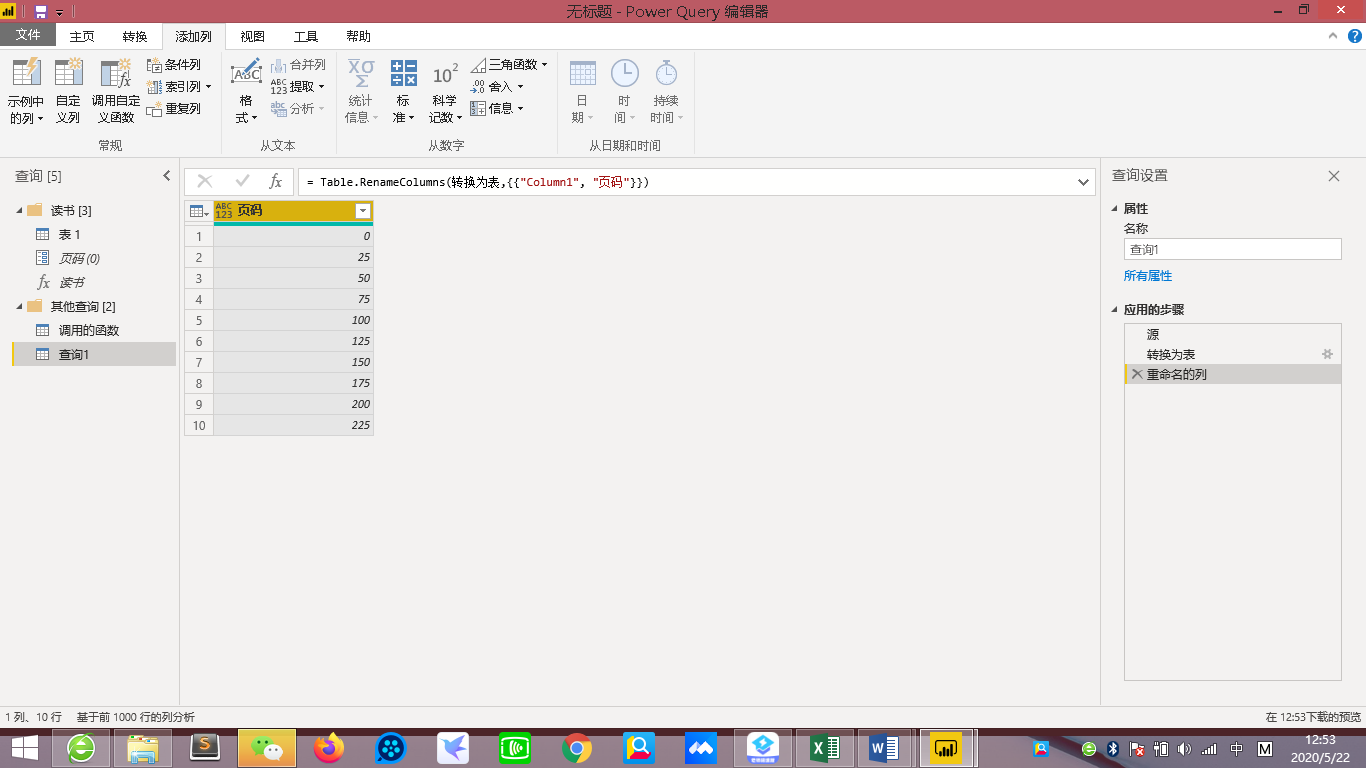
**在函数的空白栏内填入：= List.Numbers(0,10,25)，回车后会生成一个列表，点击“主页”的“到表”，将列表转换为表，然后将列的数据类型改为“文本”。**

**注：语法：**List.Numbers(**start** as number, **count** as number, optional **increment** as nullable number) as list

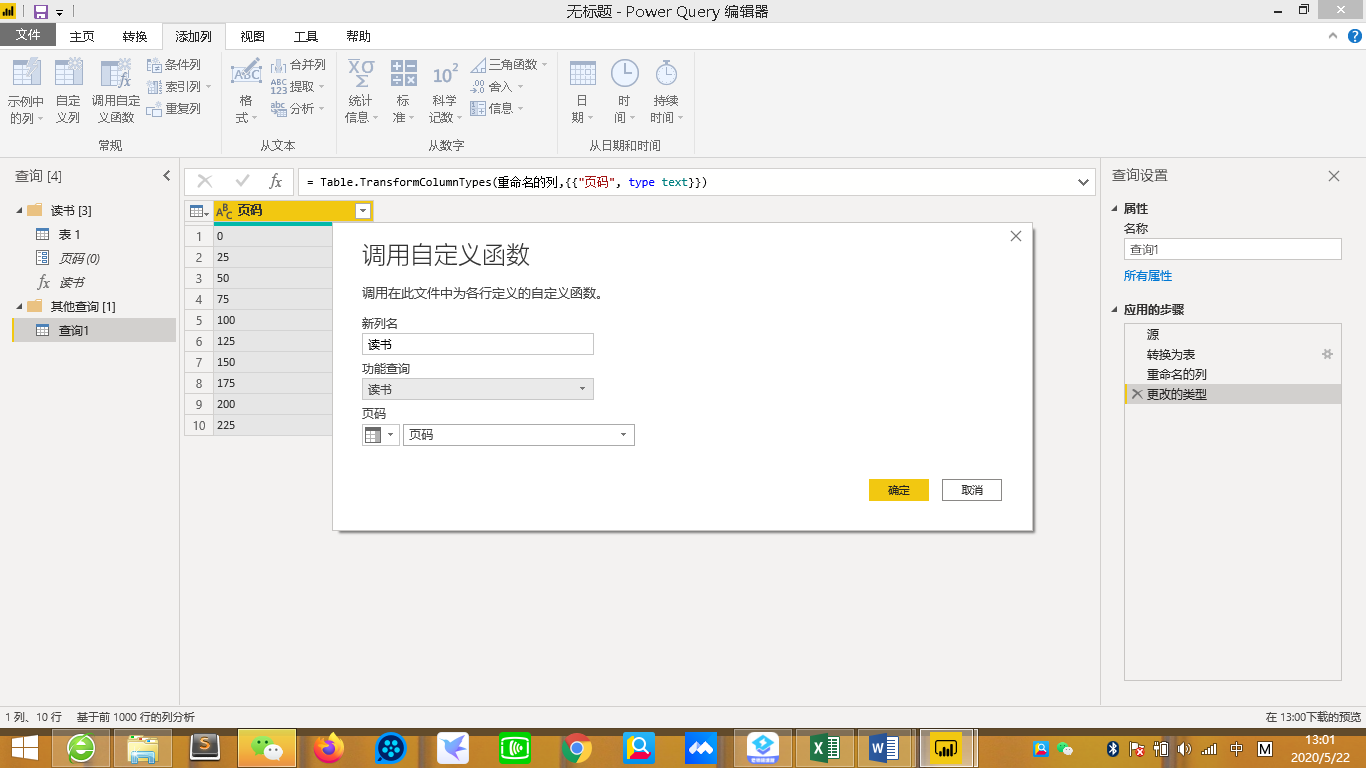
返回给定了初始值、计数和可选增量值的数值列表。 默认增量值为 1。

* start：列表中的初始值。
* count：要创建的值的数目。
* increment：[可选] 按该值递增。 如果省略，值按 1 递增。

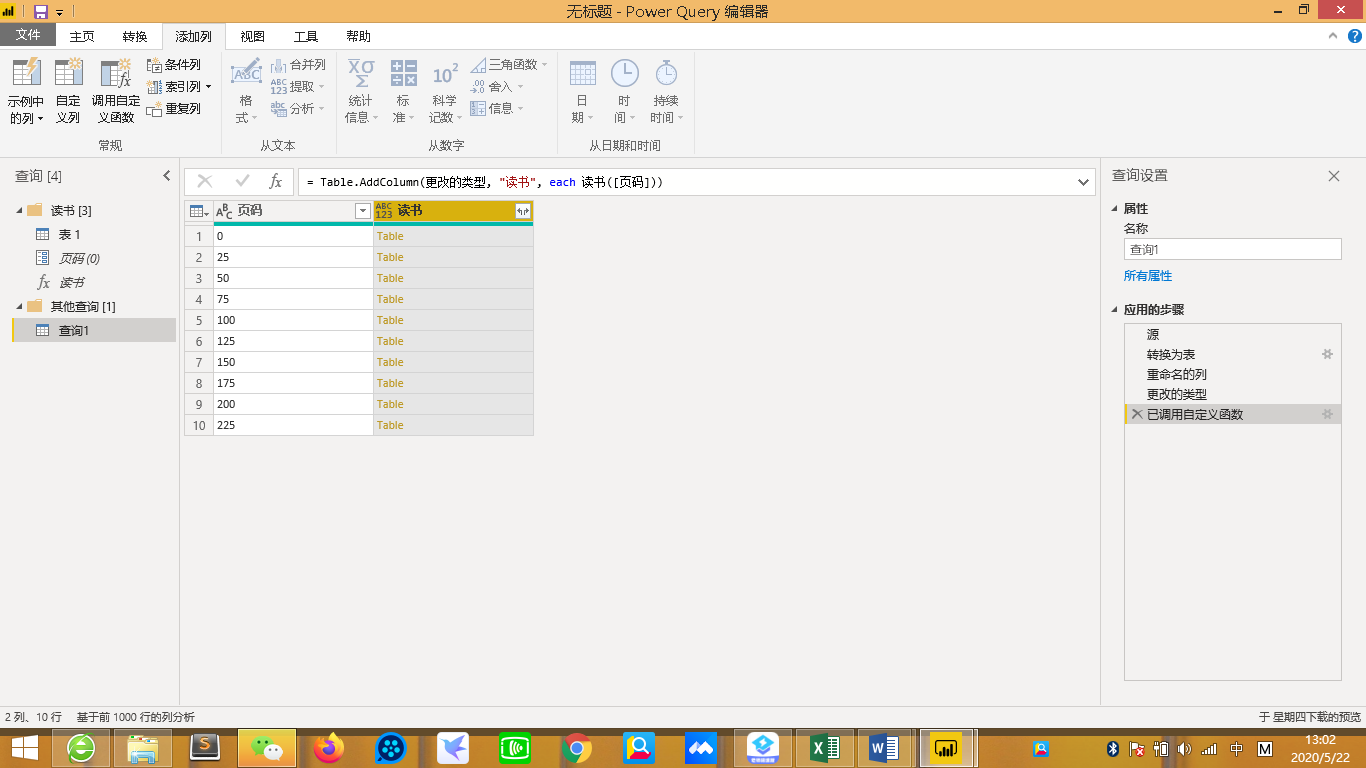


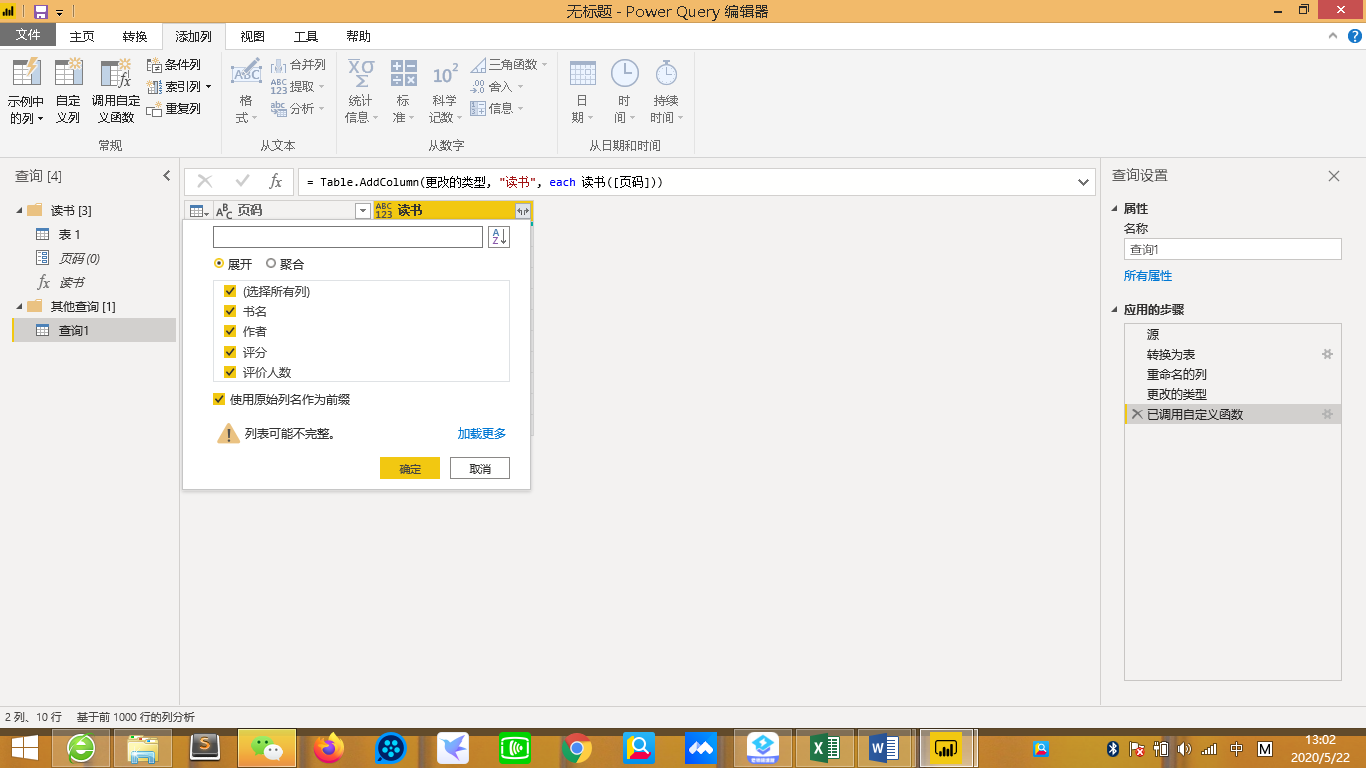


选择“添加列”下面的“调用自定义函数”，进行选择（只有唯一选择）。

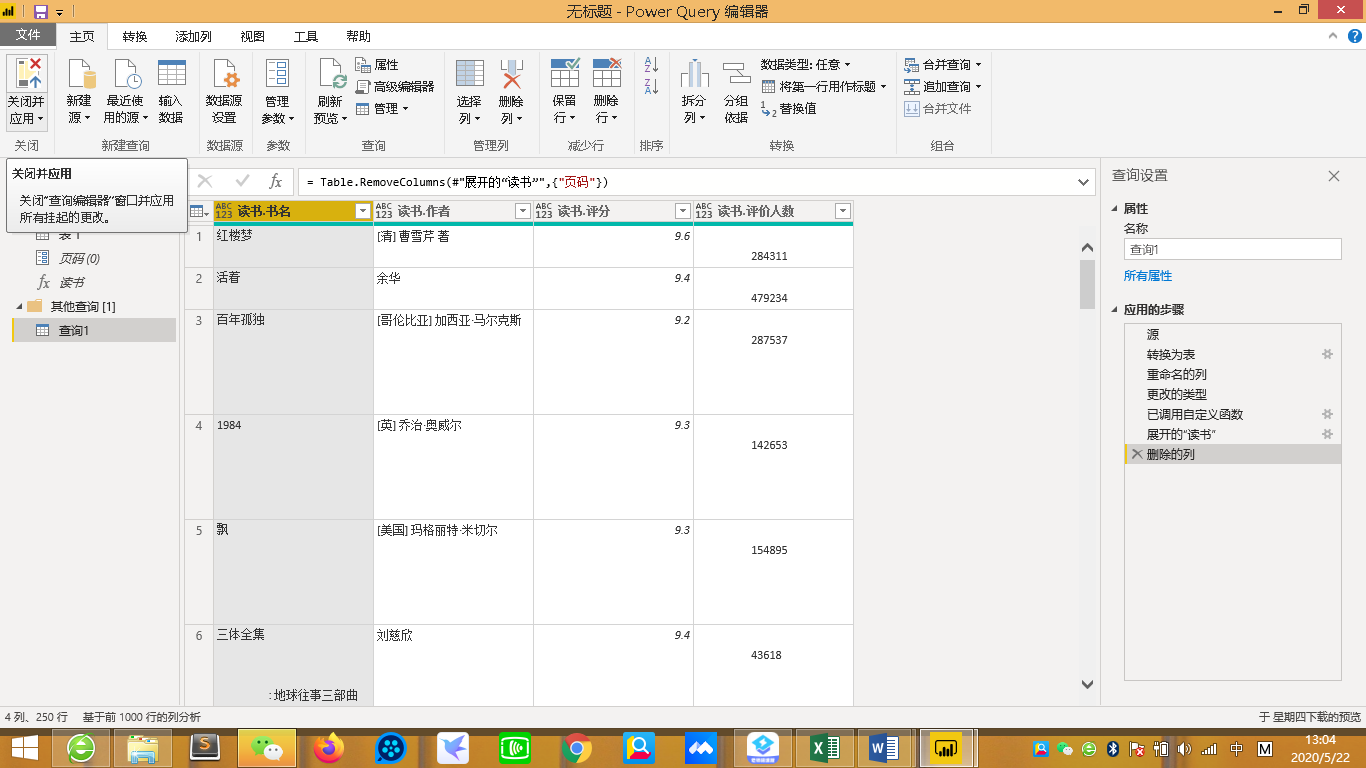


当表内容与页码相连接后，展开表格内容，，选择相应的列。





进行适当调整后，选择“主页”下的“关闭并应用”。



进入power bi界面，点击左边“表格”图标，展示表格内容，点击“报表”图标进行可视化。

