**京东评论提取**

**提前安装工具：**

语言：python3

库：requests, json； 如果要生成评论词云，需安装jieba, wordcloud

**包含程序文件：**

获取评论等信息：jingdong.py

生成词云：comment.py

**获取评论：**

1、win + R 运行 cmd

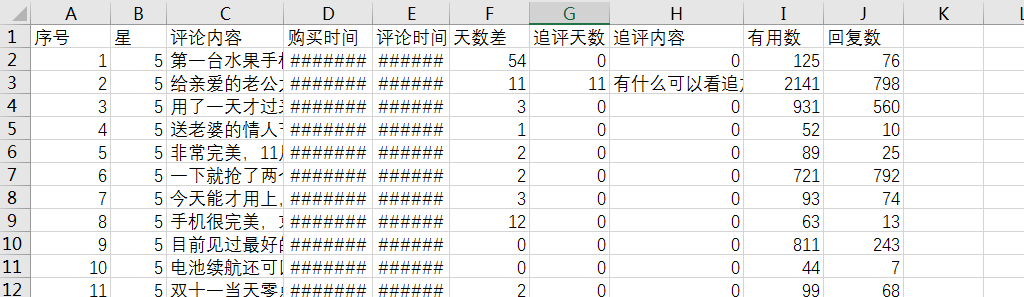
2、进入程序文件所在目录

3、输入python jingdong.py 运行，输入产品ID，评论页数，问答页数，保存文件名

4、成功后，在程序目录下得到\*\*\*\*.xls文件，用excel打开即可看到爬取内容。检查内容，对格式不规范的进行人工修正。

**生成词云：**

1、对格式修正后的xls文件内容进行删减，只保留评论相关内容，最终格式如图所示。



2、对上一步完成的文件另存为.txt格式，保存在程序文件所在目录

3、在程序所在目录新建词云忽略词文件，txt格式，内容为空即可。

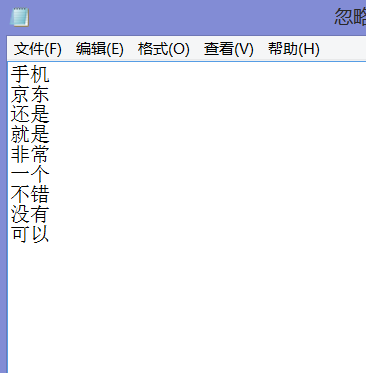
4、win + R 运行 cmd

5、进入程序文件所在目录

6、输入python comment.py 运行，输入第2步保存的txt文件名，输入第3步新建的忽略词txt文件名，输入最终词云图片名，输入词云显示词语数

7、得到\*\*\*\*.png图片文件，即生成的词云

8、根据生成的词云，找到不需要的词语，在第3步的忽略词文件中加入，格式如图。



9、重新从第6步或第4步开始运行程序，得到不包含忽略词的词云图片

10、重复第4至第9步，直至词云图片符合要求