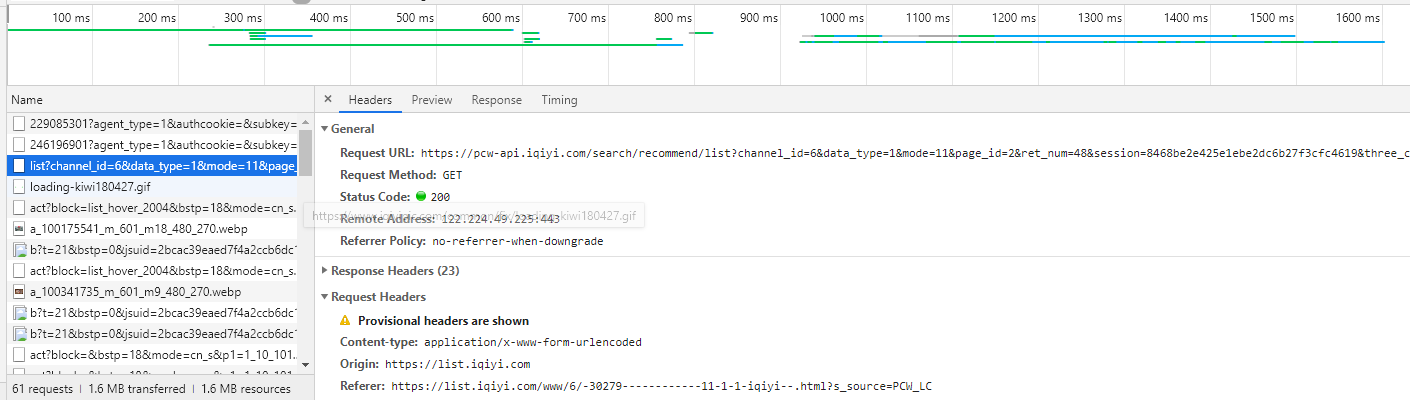
# 目录结构



运行爱奇艺--综艺爬虫可视化文件，即可实现爬虫+数据清洗+可视化

# 程序设计步骤

1. 首先手动访问爱奇艺综艺网页：<https://list.iqiyi.com/www/6/-30279------------11-1-1-iqiyi--.html?s_source=PCW_LC>
2. 打开开发者工具，滚动右边滚动条，查看翻页请求的链接地址：



1. 在找到链接后，持续翻页，观察链接变化情况，发现只有page\_id在变化。所以可以修改url中的page\_id来实现翻页。
2. 查看url的响应结果，可以看到具体数据



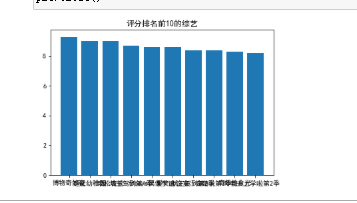
1. 编码实现上述逻辑，实现爬虫
2. 数据预处理、数据清洗，因为拿到很多综艺节目暂时没有评分，所以评分字段为空，使用pandas的dropna方法删除空数据
3. 对评分一列数据进行排序
4. 将排序结果的前10个综艺节目进行可视化

# 遇到的问题及解决办法

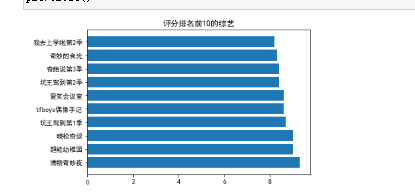
问题1：爬虫在做数据提取的时候，无法直接对ret.text进行提取，打印观察到返回的数据是json格式，但是数据类型是<class 'str'>

解决1：使用json.loads()将数据转换成<class 'dict'>后，再进行数据的提取

问题2：数据可视化的时候，可视化结果x轴上的字显示有相互覆盖的情况

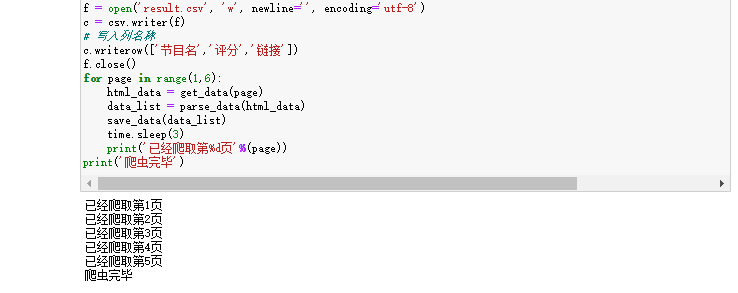


解决2：使用plt.barh()方法，交换xy轴，在x轴刻度名较长的时候，使得图片更加美观

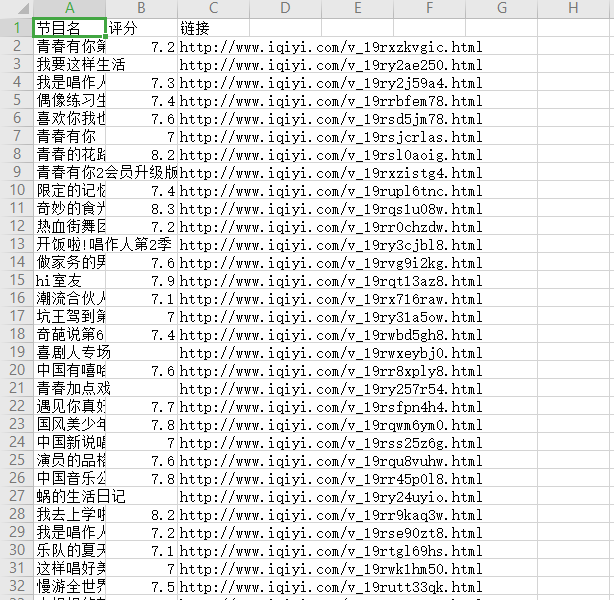


# 程序运行截图：

1. 爬虫



1. 数据保存结果



1. 可视化结果

