9组智能交通(AI) 194835 徐玥

2020.1.22

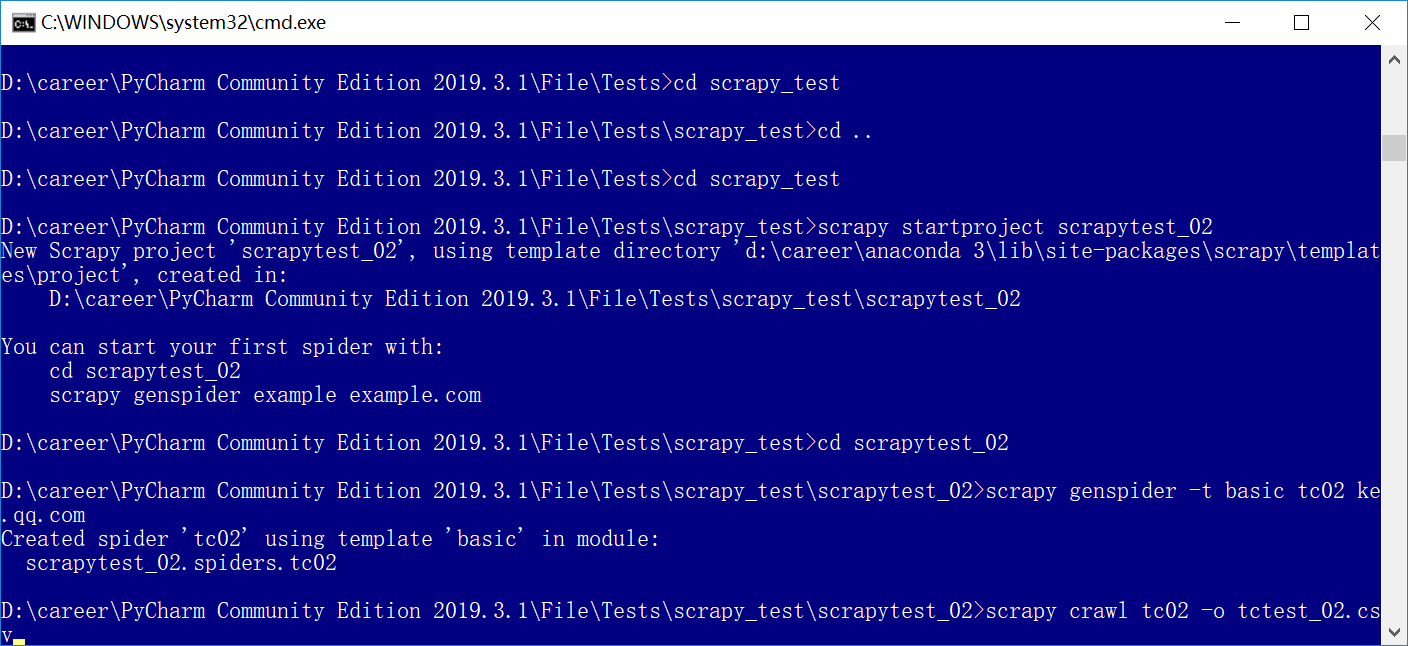
作业要求：

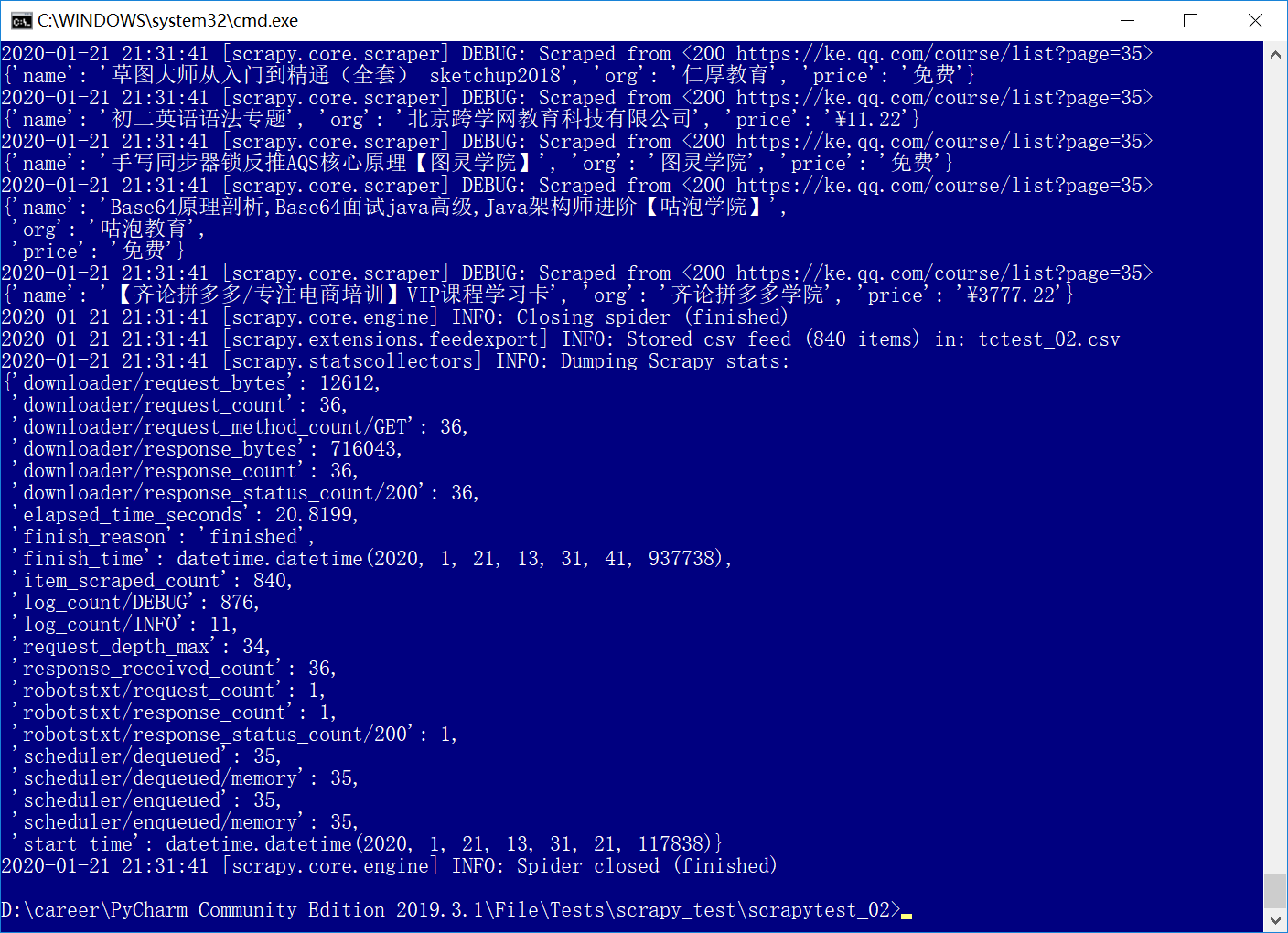
实现一个爬虫程序爬取腾讯课堂所有开课信息；

分析课程数据并可视化；

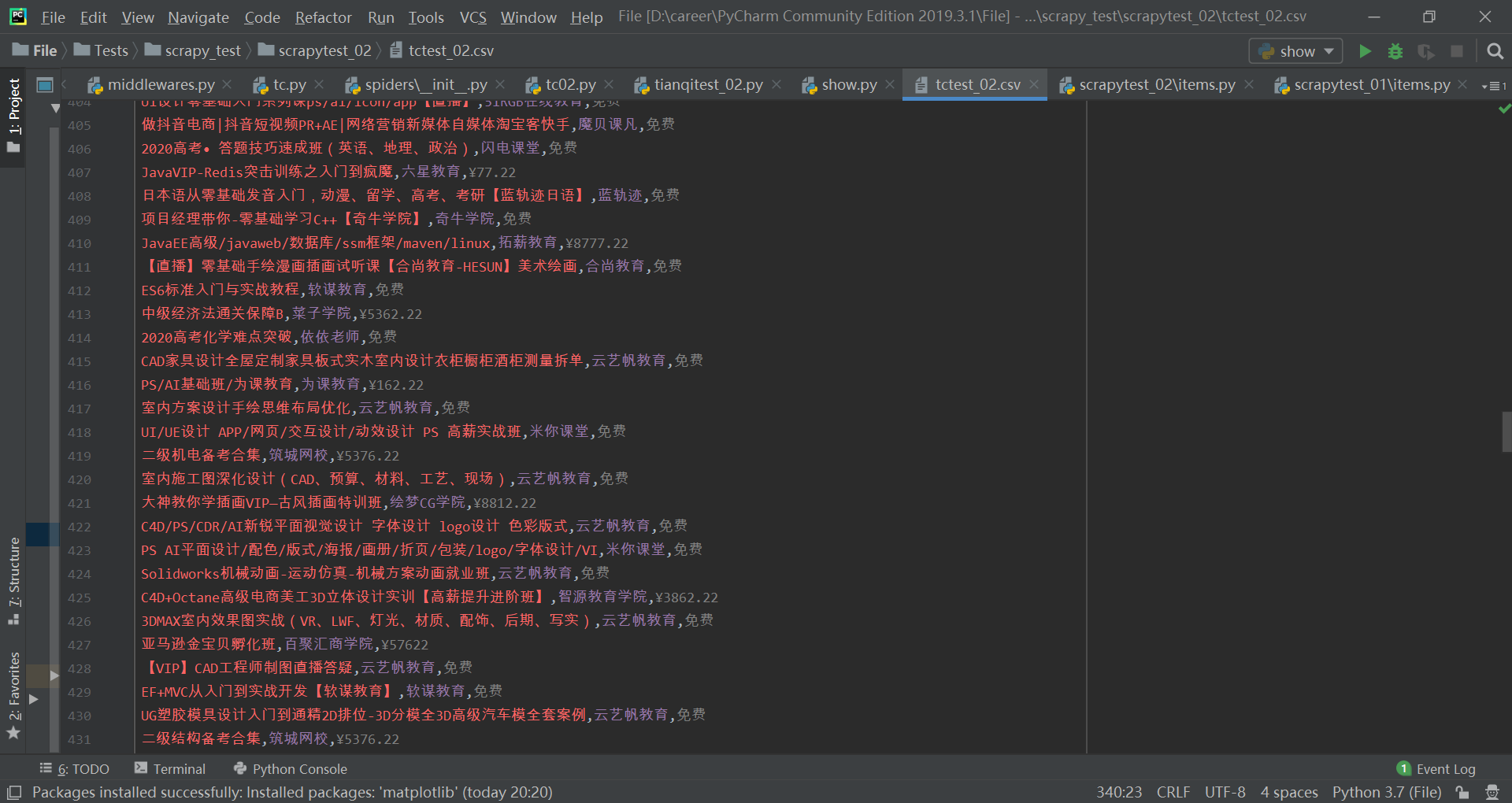
作业记录：

使用scrapy爬取数据(项目名：scrapytest\_02 爬虫名：tc02)





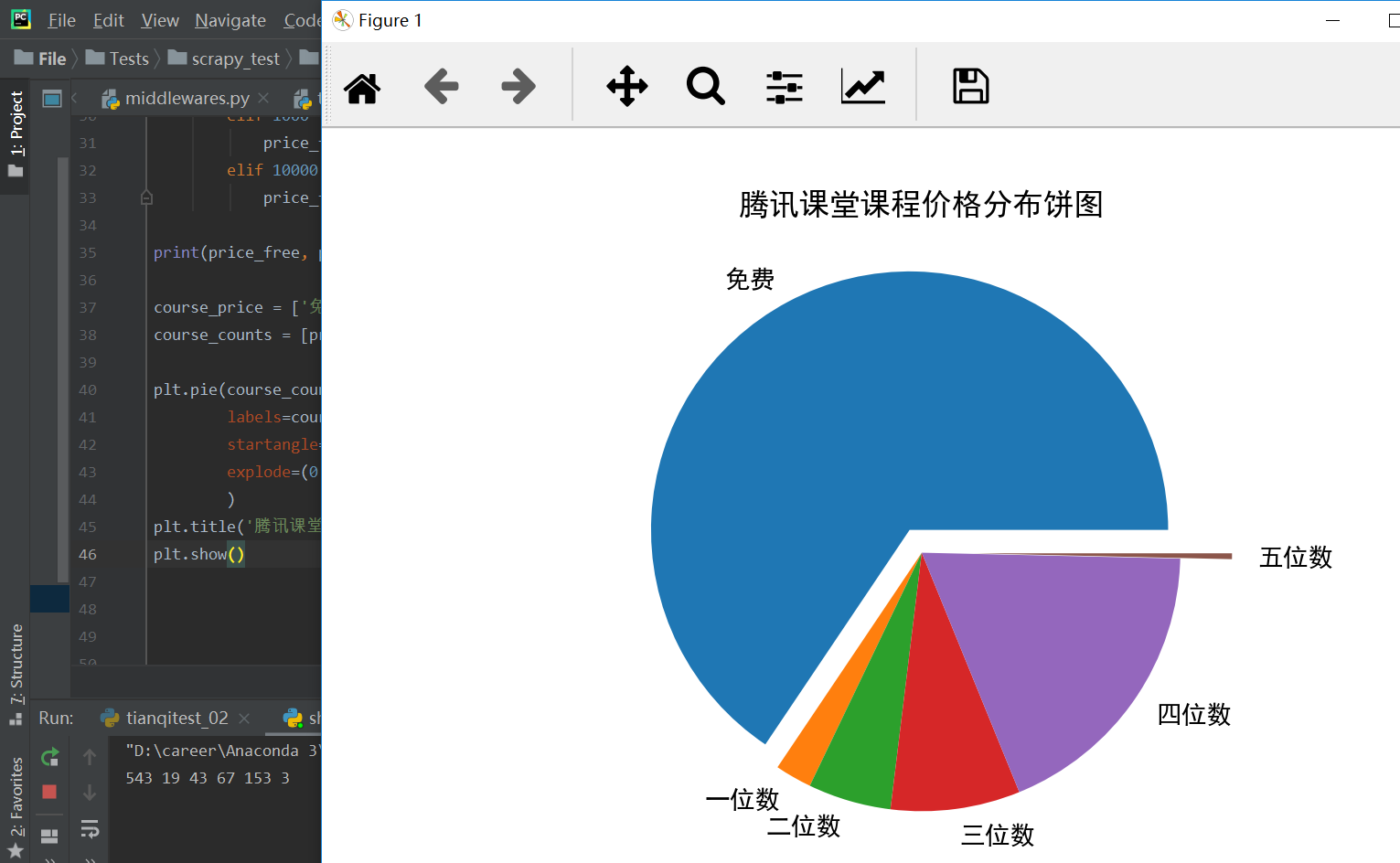
爬取结果使用csv文件保存(文件名 tctest\_02.csv)



说明一下，

考虑了一下开课信息能怎么数据可视化……感觉和数值相关的只有价格，所以处理了一下，做了不同价格区间的饼图，其中突出了免费的部分(0.1)与五位数的部分(0.2)并输出了数值(543 19 43 67 153)

*问题：查到饼图有百分值的选项(autopct='%1.0f%%')，但是不知道为什么运行时会出错提示不存在该参数……*



另：可爬取数据还有很多部分，如下图所示显示“129人最近报名”、“645人购买”和“9人购买”的部分('div/span[@class="line-cell item-user custom-string"/text()')以及显示“共19节”和“随到随学”的部分('div/span[@class="line-cell item-task"]/text()')和相同位置但是内容不同的“课程2”部分('div/span[@class="line-cell item-pkg-num"]/text()')等;但是内容类型较多且不大好统一命名就没有爬……



另外因为老师说可以爬别的……为了绘制一下常见的折线图没有用scrapy写了个小型爬虫(爬的天气数据 量不大)

