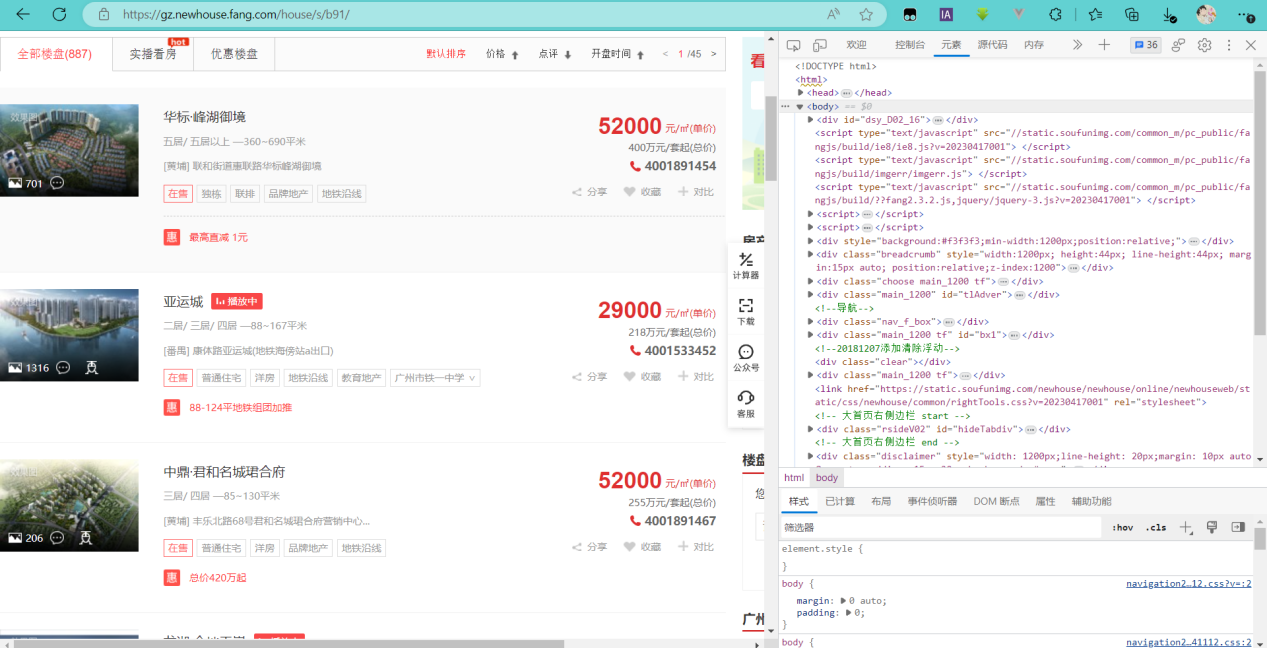
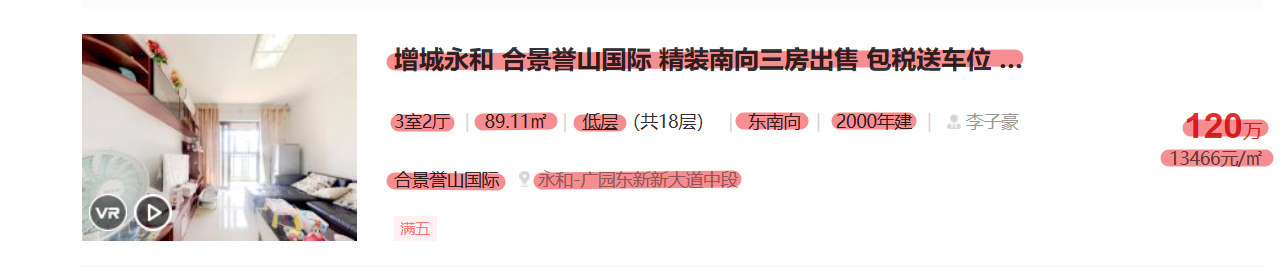
**课程设计**

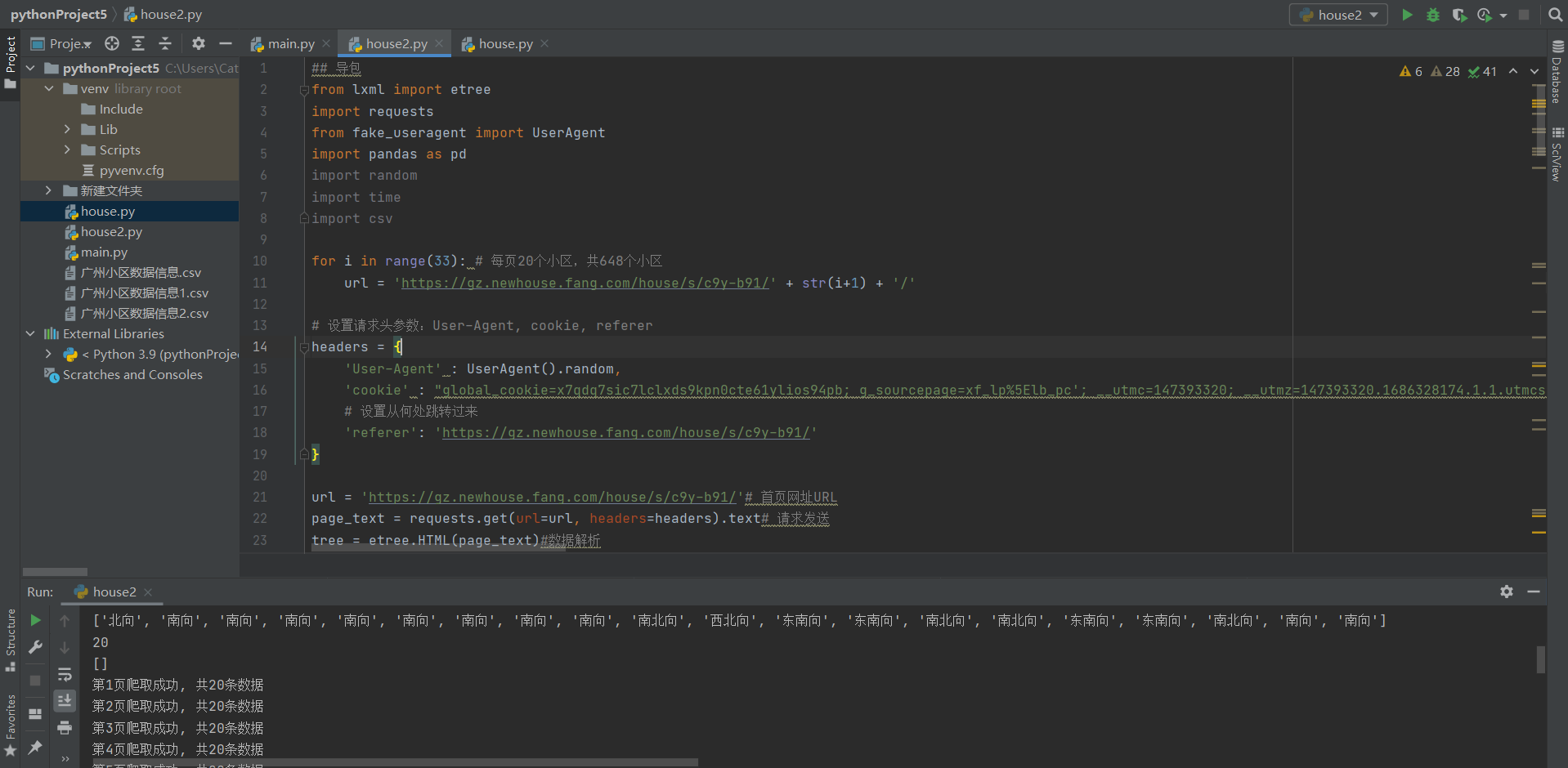
1. 观察要爬取的“房天下”广州地区的房屋信息，主要包括（'标题', '楼盘', '单价', '总价', '地址', '户型','面积','高度','朝向','年限'）数据。





如图红色标记处

1. 编写python爬虫程序，爬取数据

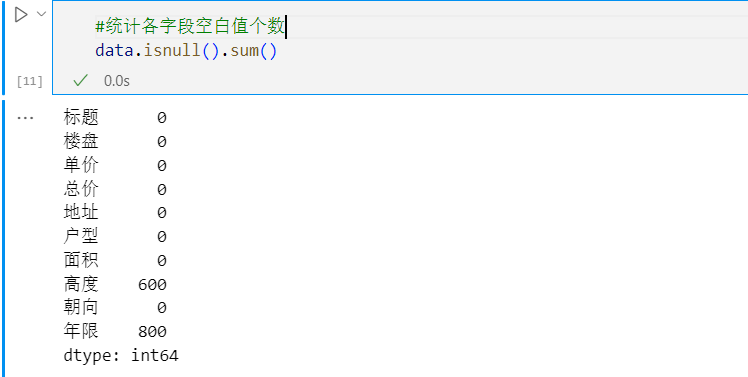


获取csv文件



1. 对数据进行查看和清洗





各字段空白值统计可得到’高度’、’年限’字段空白值个数分别为600、800。

3-1填充空值，将空白值设置成“未知”



3-2删除单价中的元/㎡和面积中的㎡

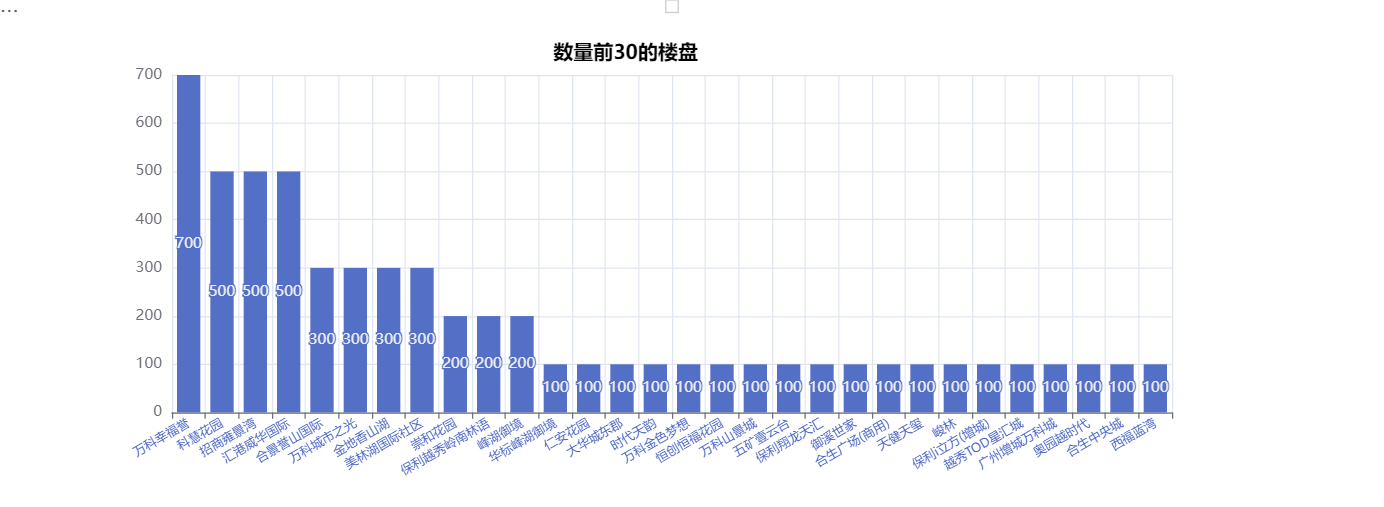




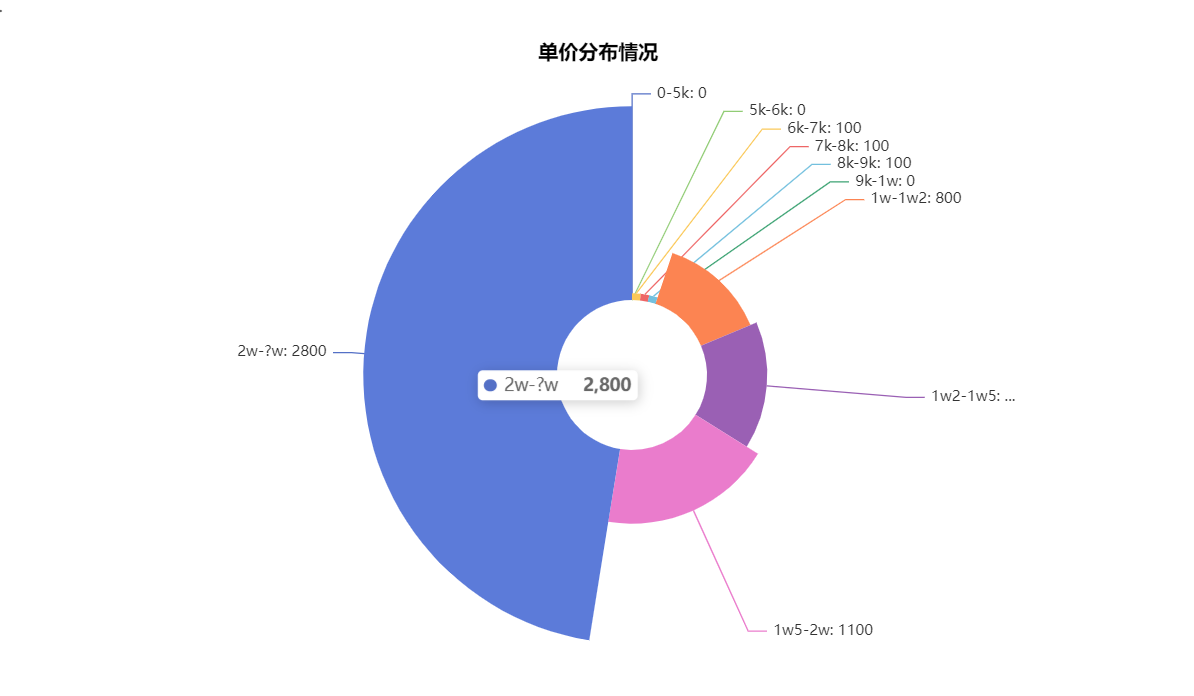
1. 对数据进行可视化

4-1数量前30的楼盘画柱状图

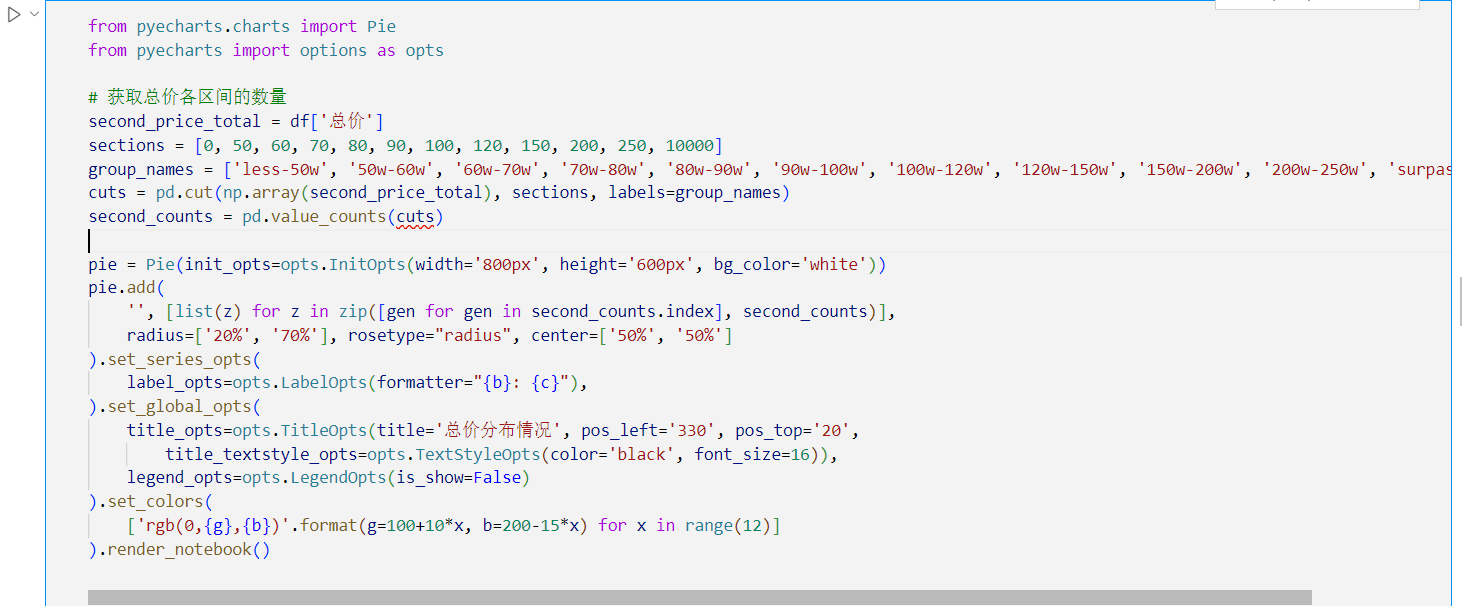


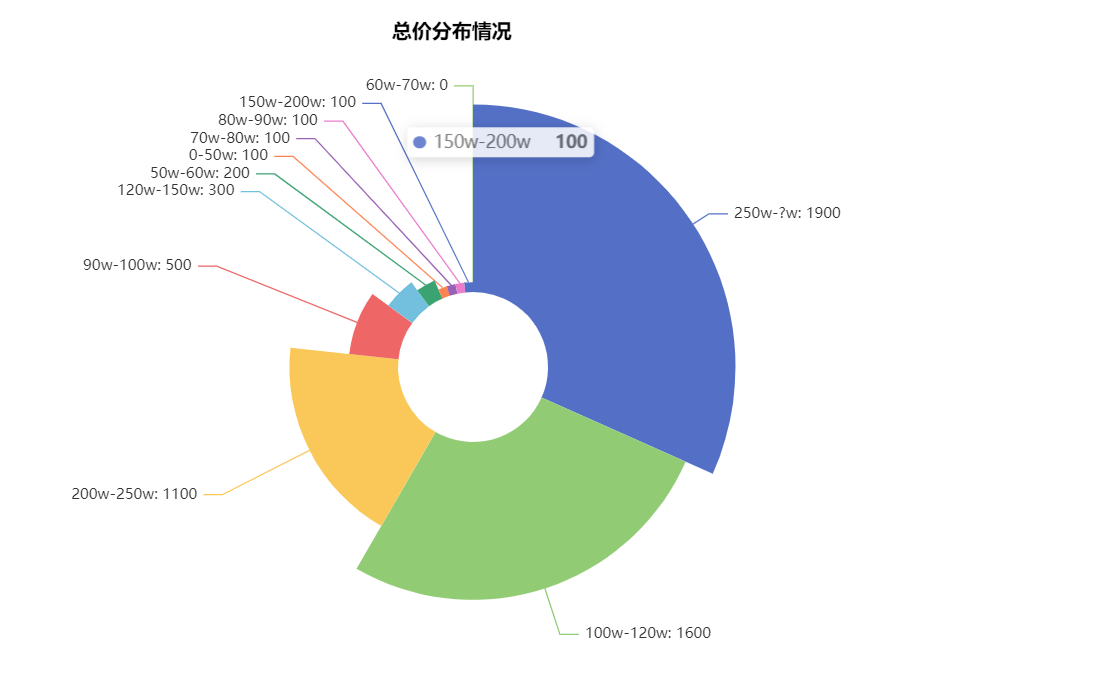


4-2对’单价’数据进行可视化

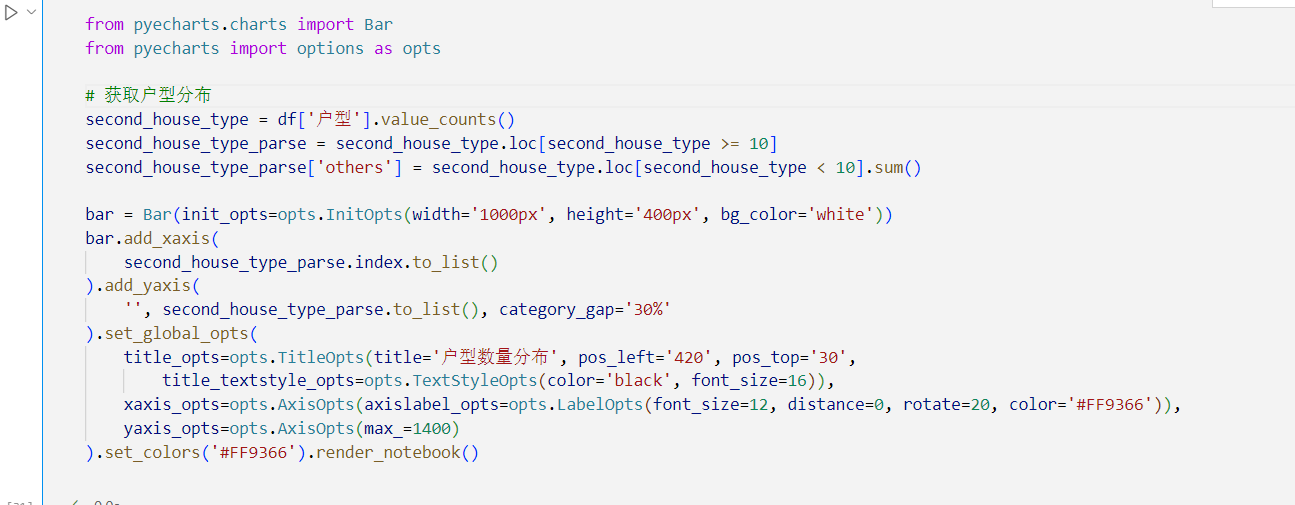


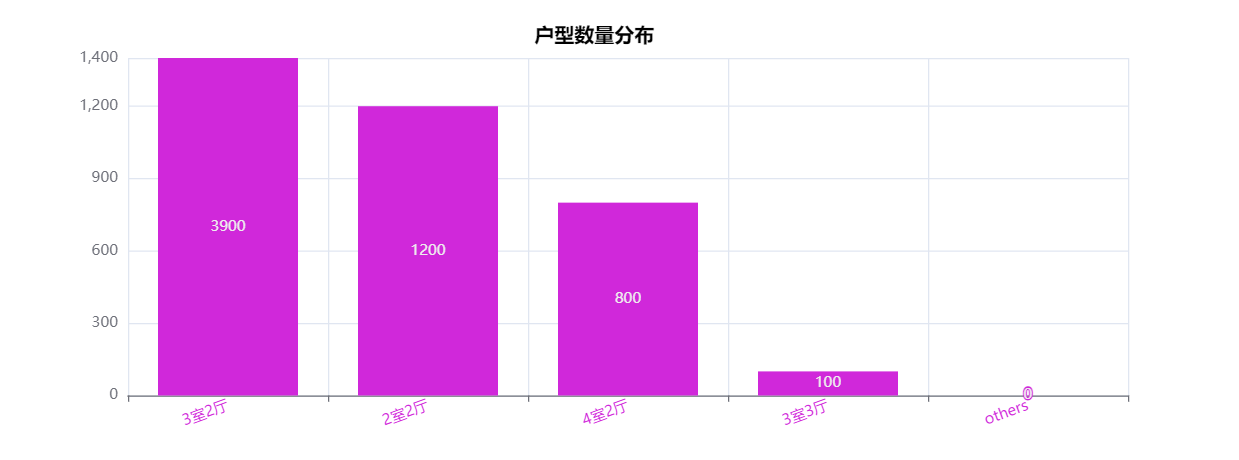
4-3对总价数据分布进行可视化



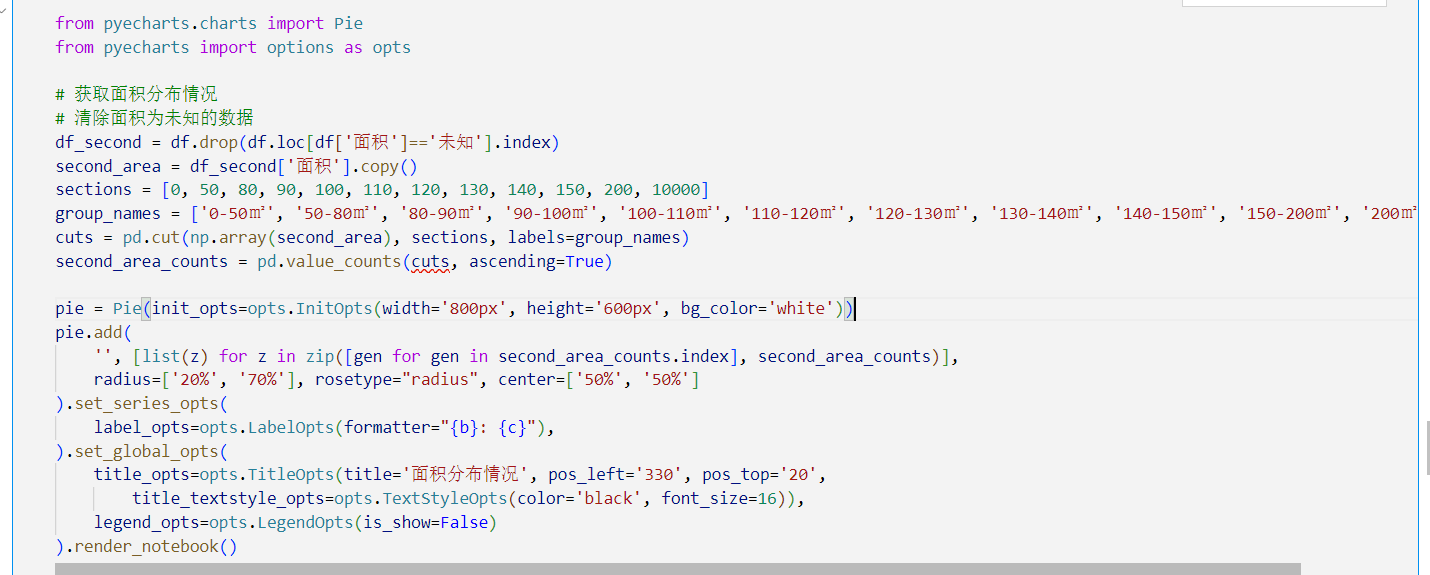


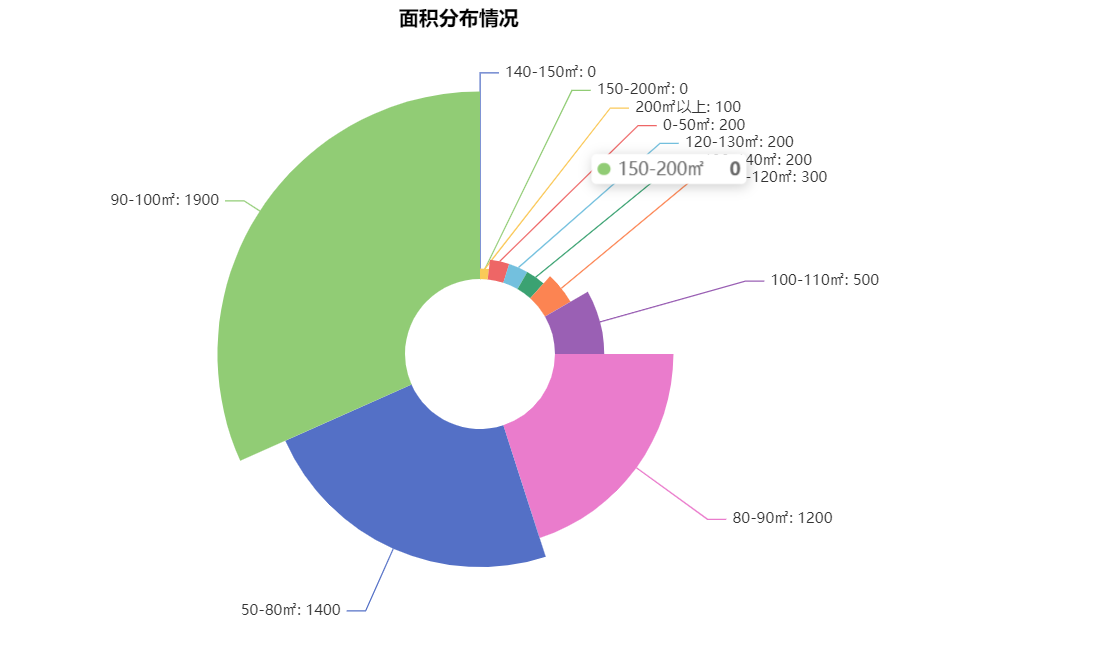
4-4对户型数量进行可视化





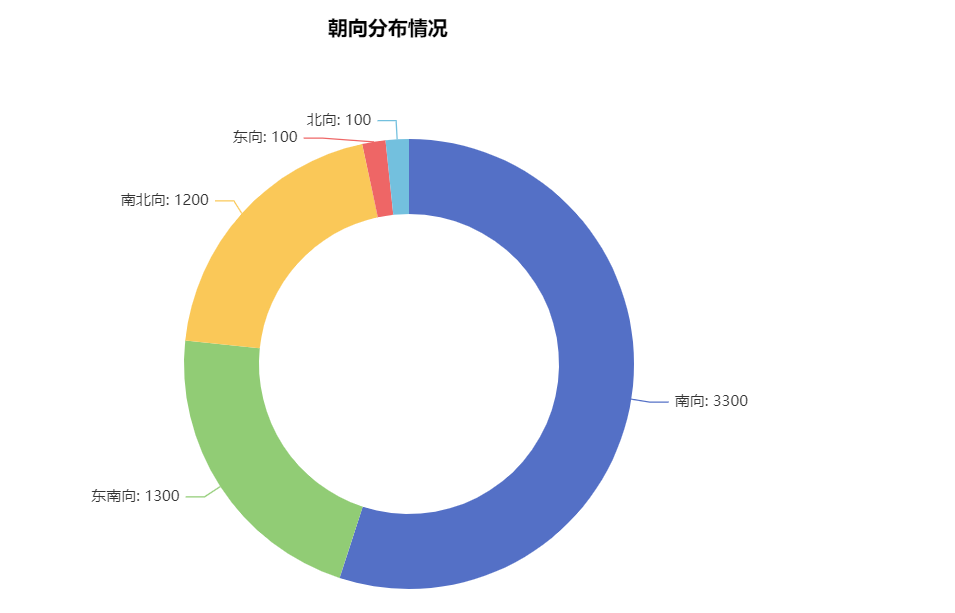
4-5对面积分布数量进行可视化





4-6对朝向数量分布进行可视化





由于在使用爬虫爬取数据时，“朝向”一列有“年限”一列的数据被爬进来了，因此需要对错误数据进行剔除再进行可视化。

4-7对楼房的标题进行词云可视化

