# 58同城二手车数据爬取可视化分析

## 数据采集

<https://gy.58.com/ershouche/?&PGTID=0d30001d-007d-f95d-deda-adfdf6bbadaa&ClickID=59>



共 70页，爬取每一页的每一年二手车的详细链接：



通过详细链接，访问详细信息，爬取页面：交易市场，车辆标题，表显里程，首次上牌，排量，变速箱，排行标准，二手价，新车报价，这些字段：



数据以csv形式存储，字段按照：交易市场，车辆标题，表显里程，首次上牌，排量，变速箱，排行标准，二手价，新车报价，链接，这样排列。

## 实现流程

分5个模块：

spiderMan：爬虫运行模块，爬虫逻辑

urlManager：管理 url 的模块

htmlDownload：请求访问模块

parseData：解析提取数据模块

dataOutput：数据输出存储模块

<https://blog.csdn.net/zjkpy_5/article/details/104717974> 字体反爬破解

## 数据分析

使用 R 进行可视化