# 大众点评抓取

## 网址链接

<http://www.dianping.com/shoplist/shopRank/pcChannelRankingV2?rankId=fce2e3a36450422b7fad3f2b90370efd71862f838d1255ea693b953b1d49c7c0>

通过观察每个城市的链接主要区别于ranKld，每个城市有特定的ID，因此先获取到相应城市的ID，便可进行后续抓取。

获取到的城市ID为：

["上海","fce2e3a36450422b7fad3f2b90370efd71862f838d1255ea693b953b1d49c7c0"],

["北京","d5036cf54fcb57e9dceb9fefe3917fff71862f838d1255ea693b953b1d49c7c0"],

["广州","e749e3e04032ee6b165fbea6fe2dafab71862f838d1255ea693b953b1d49c7c0"],

["深圳","e049aa251858f43d095fc4c61d62a9ec71862f838d1255ea693b953b1d49c7c0"],

["天津","2e5d0080237ff3c8f5b5d3f315c7c4a508e25c702ab1b810071e8e2c39502be1"],

["杭州","91621282e559e9fc9c5b3e816cb1619c71862f838d1255ea693b953b1d49c7c0"]

,["南京","d6339a01dbd98141f8e684e1ad8af5c871862f838d1255ea693b953b1d49c7c0"],

["苏州","536e0e568df850d1e6ba74b0cf72e19771862f838d1255ea693b953b1d49c7c0"],

["成都","c950bc35ad04316c76e89bf2dc86bfe071862f838d1255ea693b953b1d49c7c0"],

["武汉","d96a24c312ed7b96fcc0cedd6c08f68c08e25c702ab1b810071e8e2c39502be1"],

["重庆","6229984ceb373efb8fd1beec7eb4dcfd71862f838d1255ea693b953b1d49c7c0"],

["西安","ad66274c7f5f8d27ffd7f6b39ec447b608e25c702ab1b810071e8e2c39502be1"]

## 抓取页面



## 抓取分析

通过浏览器分析可发现该网站通过Ajax请求，所有数据来源于：

<http://www.dianping.com/mylist/ajax/shoprank?rankId=fce2e3a36450422b7fad3f2b90370efd71862f838d1255ea693b953b1d49c7c0>

该链接同之前请求一样，只需要替换rankId便可进行多城市数据获取。最终抓取的数据只需要解析json边可获得所需字段，由于大众没有特别反爬限制，只需要不断轮换userAgent便可绕过反爬。

我们对上海，北京，广州，深圳，天津，杭州，南京，苏州，成都，武汉，重庆，西安等城市的前100家商铺进行数据获取，并分析最终所获取数据集，见《大众点评数据分析》

# 请求头

USER\_AGENT\_LIST = [

"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.1 (KHTML, like Gecko) Chrome/22.0.1207.1 Safari/537.1",

"Mozilla/5.0 (X11; CrOS i686 2268.111.0) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11",

"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.6 (KHTML, like Gecko) Chrome/20.0.1092.0 Safari/536.6",

"Mozilla/5.0 (Windows NT 6.2) AppleWebKit/536.6 (KHTML, like Gecko) Chrome/20.0.1090.0 Safari/536.6",

"Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.1 (KHTML, like Gecko) Chrome/19.77.34.5 Safari/537.1",

"Mozilla/5.0 (X11; Linux x86\_64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.9 Safari/536.5"]

head = {

'User-Agent': '{0}'.format(random.sample(USER\_AGENT\_LIST, 1)[0]) # 随机获取

}

代码展示：见daZhongFood/spider/dazhongfood.py

最终获取结果存储至MySQL。(完整数据集见daZhongFood/data)

## 最终结果

