

# 数据说明

## 1.上海-用户数据

字段	意义
user_id	用户ID
gender	性别： 1=男， 2=女， 0=未知
following_cnt	粉丝数量
contribution	贡献度
register_time	注册时间
city	城市
age	年龄
tag	用户特征标签
malicious	用户是否异常
malicious_reason	用户异常的原因

## 1.上海-用户活动

字段	意义
user_id	用户ID
Event_id	活动ID
Business_id	商家ID
Quota	活动名额
response	活动页面下的回应个数
Timestamp	活动时间
Participants_id	所有参与者的ID
Participants_gender	所有参与者的性别
Is_chosen	是否被选中

注意：用户活动数据中城市信息有部分不正确。处理数据时需要利用business\_id进行确认(商家数据暂无)。

## 1.上海-用户评论

字段	意义
user_id	用户ID
Business_id	商家ID
Timestamp	评论时间
Text	评论的文本内容
Rate	用户对商家的打分
Star	其他用户给评论的电子
Response	该评论的回复数
Pic_list	评论中图片ID

## 2.dianping\_user\_business文件夹

该文件夹下包含了用户和商家的详细信息。

### shanghai\_useful.csv

该文件包含了25985个商家的详细信息。文件中每一行代表了一个商家的数据。具体格式略。

### shanghai\_useful\_detail\_addr.json

该文件包含了25985个商家的详细地质信息以及POI的详细信息，是通过将shanghai\_useful.csv中商家的经纬度提取出来并在百度地图api中进行逆地理解析得到的。其中每一行是对应一个商家的地理信息。信息中还包括了商家周围POI的详细地理信息。具体数据结构略。

### user\_with\_valid\_bid.json

该文件中包含了用户和已经爬取的商家之间的对应关系，具体结构如下：

```
{
  "Uid_1":[
    "bid_1","bid_2",...
  ],
  ...
}
```

## 2.shanghai\_business\_info.json

商家信息文件，key值为商家id。

字段	意义
review_count	评论数量
longitude	经度
is_open	是否营业
stars	评分
address	地址
latitude	纬度
neighborhood	行政区

注意： 1.和2.中的数据虽然都是商家点评数据，但前后商家信息并不是完全对应的。学生需分别进行用户侧写。