# 数据说明

## 数据展示

**原始数据：**

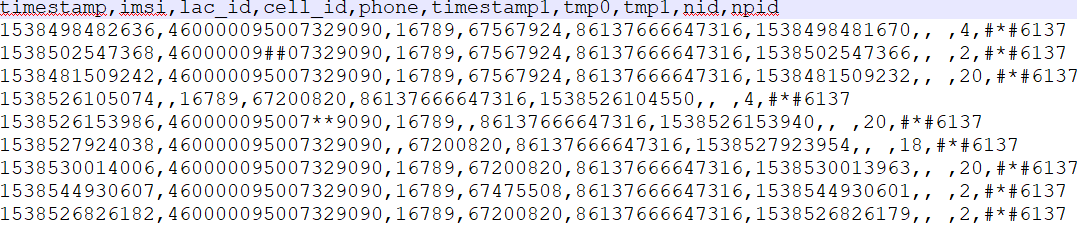
timestamp 信息记录开始时间 imsi 用户识别码（唯一id）

lac\_id 基站位置区编码 cell\_id 扇区编号

phone 电话号 timestamp1 信息记录结束时间

tmp0 一级行政区编号 tmp1 二级行政区编号

nid 信息类别1 npid 信息类别2



原始数据含有脏数据，需要数据清洗，清洗后样例：

20181003061550,460000095074424637,16655,235035926

数据清洗规则（参考）：

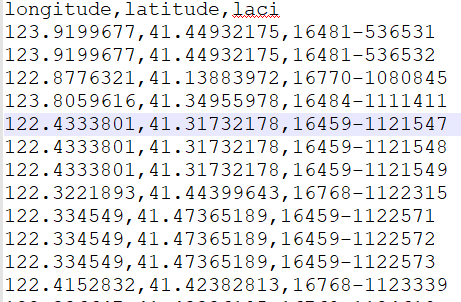
1. 抽取timestamp,imsi,lac\_id,cell\_id 四个字段
2. 去除imsi中，包含特殊字符的数据条目（‘#’,’\*’,’^’）
3. 去除空间信息残缺的记录条目（imsi、lac\_id、cell\_id中为空）
4. timestamp时间戳转换格式 ‘20190603000000’--年月日时分秒
5. 去除干扰数据条目（不是2018.10.03当天的数据）
6. 去除两数据源关联后经纬度为空的数据条目
7. 以人为单位，按时间正序排序

**基站经纬度数据：**

longitude 经度

latitude 纬度

laci 基站信息

****

原始数据中的lac\_id、cell\_id与基站经纬度数据中的laci 关联后可获取经纬度信息

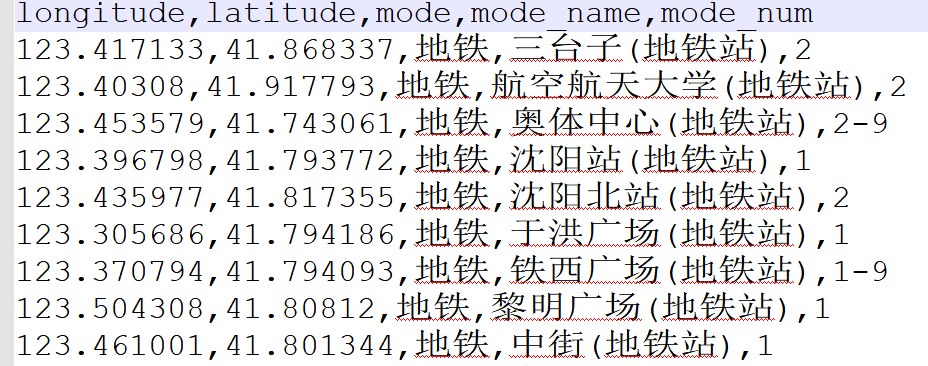
**出行方式静态数据：**

longitude 经度 latitude 纬度

mode 方式（数据中包括地铁、公交）

mode\_name 站名

mode\_num 线路



**人群密度：**

人群密度模块数据从原始数据中获取