2019年年终报告数据提取说明

# 数据来源：

|  |  |  |
| --- | --- | --- |
| 数据分类 | 数据表 | 数据表说明 |
| 驾驶行为数据: | ghy.car\_driving\_behavior\_version3 | 用户历史行程记录 |
| ghy.car\_driving\_number\_24h\_distribution | 用户历史24小时驾驶记录 |
| ghy\_dy.car\_daily\_s\_e\_evmile\_mile\_version1 | 每日始末里程、能耗数据 |
| yun\_can\_year.car\_status | 实时车况数据 |
| byd\_wnj.months\_max\_speed\_year\_report | 1-9月最大车速 |
| yun\_can\_minutes.extract\_can | 10-12月原始总线数据 |
| 充电行为数据 | ghy.car\_charging\_behavior\_version3 | 用户历史充电记录 |
| 云服务相关数据 | yun.foot\_auto | 用户云服务使用记录 |
| PAD相关数据 | car\_behavior\_clean.pad\_app\_data | 用户PAD中APP使用记录 |
| 运营商流量 | data\_statistics.cmcc\_cucc\_2019xx(01、...、12) | 1-12月运营商流量数据 |
| 车辆相关数据 | data\_statistics.dms\_link\_info | 销服接入记录 |



# 数据提取：

脚本路径：/user/yu.zhigang/SQL\_SCRIPT

|  |  |  |
| --- | --- | --- |
| workflow/coordinates | 文件名（数据表/表名为脚本名称） | 表说明 |
| annual\_report\_2019\_drive | annual\_report.r2019\_vehicle\_info | 2019年云服务年报车辆信息 |
| annual\_report.r2019\_daily\_city |  |
| annual\_report.r2019\_daily\_city\_clean |  |
| annual\_report.r2019\_city\_city\_distance |  |
| annual\_report.r2019\_drive\_part11 | 实时车况数据 |
| annual\_report.r2019\_drive\_part1 |  |
| annual\_report.r2019\_drive\_part2 |  |
| annual\_report.r2019\_drive\_part3 |  |
| annual\_report\_2019\_charge | annual\_report.r2019\_p2\_charge |  |
| annual\_report\_2019\_cloud | annual\_report.r2019\_cloud\_function\_daily\_counts |  |
| annual\_report.r2019\_cloud\_function\_daily\_counts\_rank |  |
| annual\_report.r2019\_cloud\_part1 |  |
| annual\_report.r2019\_cloud\_part2 |  |
| annual\_report\_2019\_pad | annual\_report.r2019\_pad\_app\_daily\_stats |  |
| annual\_report.r2019\_pad\_part1 |  |
| annual\_report.r2019\_pad\_part2 |  |
| r2019\_speed\_months | annual\_report.r2019\_speed\_months | 最近3个月最大车速定时任务 |
| r2019\_speed\_months1 | annual\_report.r2019\_speed\_months1 | 24小时车速数据定时任务 |
| annual\_report\_2019\_speed | annual\_report.r2019\_p3\_speed | 最大车速、平均车速 |

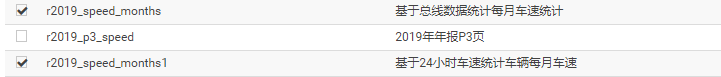
结果表：annual\_report.r2019

|  |  |  |
| --- | --- | --- |
| **页面** | **字段** | **释义** |
| 基本信息 | user\_id | 云服务用户ID |
| vin | 车架号 |
| power\_type | 动力类型 |
| model\_series | 车型 |
| isdilink | 是否DiLink车型 |
| 标签前置页 | tag\_main1 | 方案一：称号标签 |
| tag\_main11 | 方案一：欠缺标签 |
| tag\_main2 | 方案二标志 |
| 页面1 | total\_mile | 年度行驶总里程 |
| total\_mile\_over\_per | 年度行驶总里程超过云服务车主比例 |
| tag\_total\_mile | 年度行驶总里程标签 |
| total\_evmile | 年度行驶总EV里程 |
| total\_fuel | 年度耗油量 |
| total\_elect | 年度耗电量 |
| fuel100 | 年度百公里油耗 |
| fuel100\_over\_per | 年度百公里油耗超过燃油车主比例 |
| tag\_fuel100\_over\_per | 百公里油耗超过车主比例标签 |
| elect100 | 年度百公里电耗 |
| save\_money | 年度节省费用 |
| save\_co2 | 减排CO2 |
| save\_meat | 年度节省费用相当于多少斤猪肉 |
| p1\_isvalid | P1页面有效标志 |
| 页面2 | charge\_counts | 年度充电次数 |
| charge\_avg\_time | 平均单次充电时长 |
| charge\_over\_per | 充电次数超过N%的车主 |
| tag\_charge1 | 充电次数超过N%的车主标签 |
| tag\_charge2 | 充电次数超过N%的车主标签文案 |
| charge\_most\_timezone | 最常充电时段 |
| p2\_isvalid | P2页面有效标志 |
| 页面3 | max\_speed | 最高车速 |
| avg\_speed | 平均车速 |
| tag\_max\_speed | 最高车速标签 |
| p3\_isvalid | P3页有效标志 |
| 页面4 | drive\_time | 行驶时长 |
| voice\_counts | 智能语音助手使用次数（唤醒小迪次数） |
| tag\_voice1 | 智能语音助手使用次数标签 |
| tag\_voice2 | 智能语音助手使用次数标签文案 |
| avg\_voice\_counts | 平均每个车主唤醒次数 |
| p4\_isvalid | 页面有效标志 |
| 页面5 | tag\_time\_routine1 | 年度出行次数最多的时段标签 |
| tag\_time\_routine2 | 年度出行次数最多的时段标签文案 |
| p5\_isvalid | P5页是否有效 |
| 页面6 | most\_control | 年度最常用远程控制 |
| most\_control\_counts | 年度最常用远程控制次数 |
| tag\_most\_control1 | 年度远程控制最多的类型标签 |
| tag\_most\_control2 | 年度远程控制最多的类型标签文案 |
| least\_control | 年度最少使用远程控制 |
| tag\_least\_control1 | 年度最少使用远程控制功能推荐标签文案 |
| p6\_isvalid | 全部有效 |
| 页面7 | date\_city | 离家最远的日期,城市 |
| date\_mile | 行驶里程最长的日期，里程 |
| date\_night | 深夜三更行程日期 |
| date\_city\_hight | 最高海拔日期，城市，海拔 |
| date\_air\_open | 远程开空调最多的日期，开空调次数 |
| date\_door\_unlock | 车门解锁最多的日期，车门解锁次数 |
| date\_vehicle\_status | 查看实时车况最多的日期，实时车况次数 |
| date\_energy | 能耗排名最多的日期, 查看能耗排名次数 |
| date\_whistle | 闪灯鸣笛最多日期，闪灯鸣笛次数 |
| date\_music | 听歌最久的日期，时长（分钟） |
| date\_game | 玩游戏最久的日期，时长（分钟） |
| date\_learn | 学习APP最久的日期，时长（分钟） |
| 页面8 | avg\_month\_flow | PAD端APP流量(平均每月流量，单位：GB） |
| avg\_flow\_vin | 所有车主月平均流量（GB） |
| p8\_isvalid | P8页面有效 |
| 页面9 | app\_counts | 使用过的APP数量 |
| app\_time | 所有app使用时长，单位：h |
| avg\_app\_time | 所有dilink车主平均使用时长 |
| top\_app\_list | 使用时长最常的2个APP |
| tag\_app\_time1 | APP使用时长标签 |
| tag\_app\_time2 | APP使用时长标签文案 |
| p9\_isvalid | 页面有效性 |
| 页面10 | score\_techno | 科技得分（60-100） |
| score\_energy | 节能得分（60-100） |
| score\_routine | 出行得分（60-100） |
| score\_envir | 环保得分（60-100） |
| score\_intell | 智能得分（60-100） |
| score\_entert | 娱乐得分（60-100） |
| score | 综合得分（60-100） |
| tag\_score1 | 综合得分标签 |
| tag\_score2 | 综合得分标签文案 |
| p10\_isvalid |  |
| 结束页 | enroll\_days | 云服务陪伴天数 |

## 驾驶数据



## 车速数据

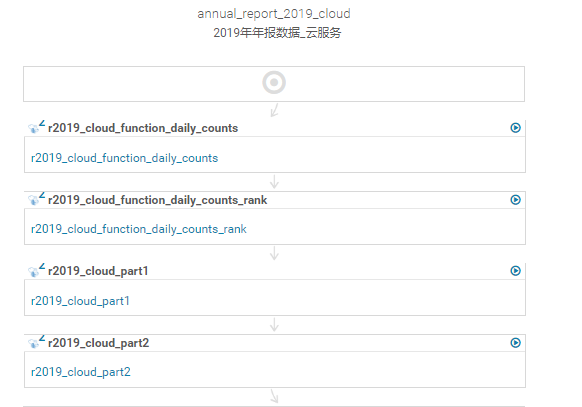




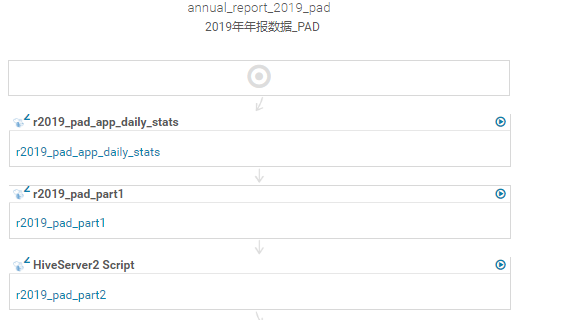
## 充电数据



## 云服务数据



## PAD数据



## 年报汇总数据



依次输出三张表：r2019\_v1、r2019\_v1\_rank、r2019\_v2

# 数据验证

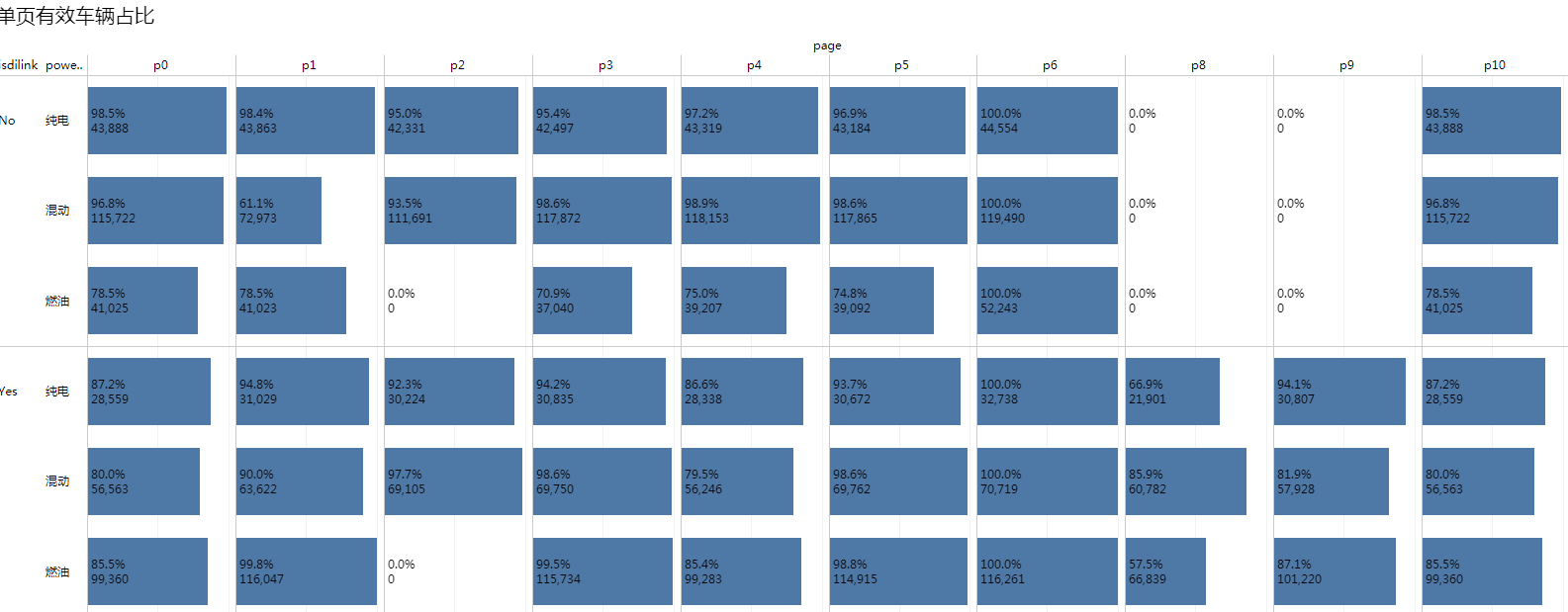
年报数据及验证结果路径：<http://10.9.34.41/#/views/2019_/sheet0?:iid=1>

## 数据有效范围检验

见附录：个人年报数据\_终版

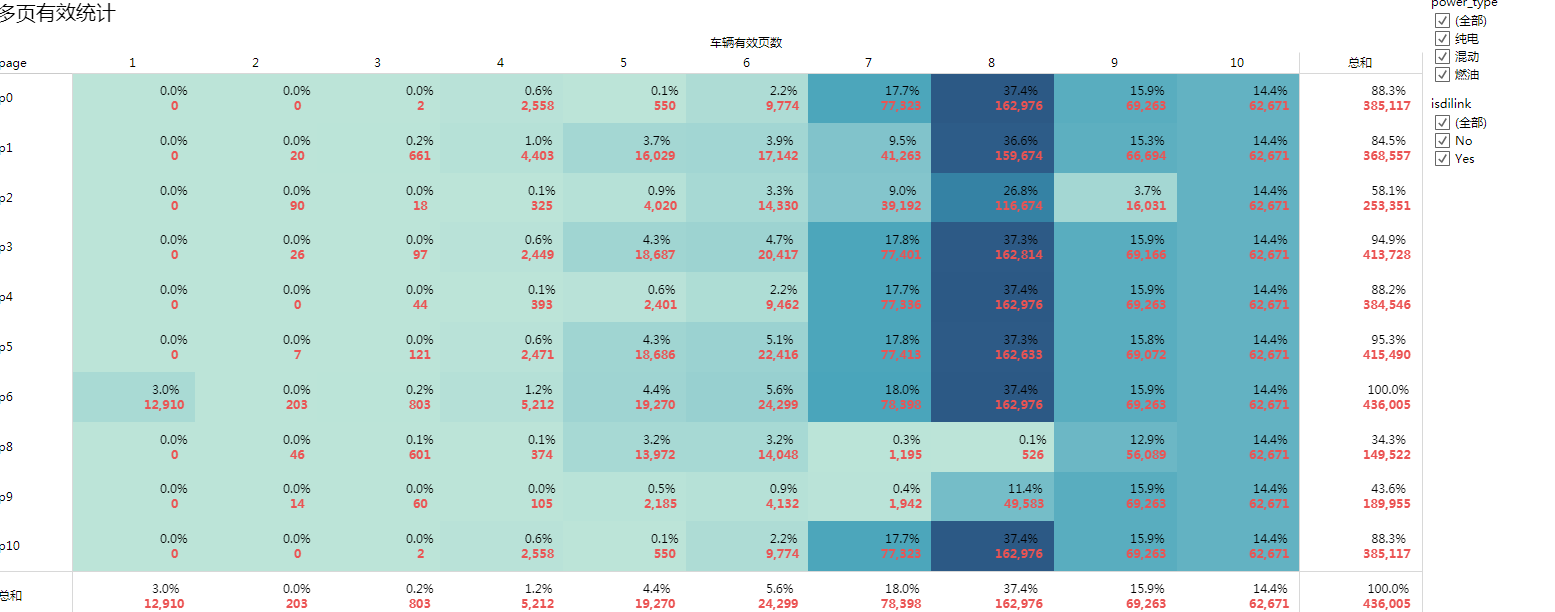
## 单页有效性统计

检验页面数据设定是否合理



## 用户有效页数统计

辅助决定年报用户群体



## 自定义规则页数验证

辅助决定年报用户群体

## 标签分布

检验标签定义是否合理、是否存在数据泄露（标签覆盖不全）

## 缺失页数据核查

辅助决定年报用户群体

## 车辆抽查

从用户角度抽查年报数据准确性情况

# 附录



