

# 数据质量评估

(模板 1.0)



## 目录

1	基础评估.....	2
2	数据定量评估.....	3
2.1	数据表评估.....	3
2.2	数据字段评估.....	3
2.3	数据分布统计.....	4
2.4	单表相关性分析.....	5
2.5	多表相关性分析.....	5
2.6	数据量统计.....	5
2.7	样本数据.....	5
2.8	异常数据.....	6
2.9	失效数据.....	6
3	数据非定量评估.....	6
3.1	数据目的.....	6
3.2	数据用途.....	6
3.3	数据日志.....	6
3.4	数据可靠性.....	6
3.5	可访问性.....	7
4	评估结论.....	7

## 1 基础评估

评估时间：20211118

表名（英文）：FACT\_JS\_BUS\_CUST

表名（中文）：久事公交线路刷卡客流量表

变量数：2

数据量：50G

缺失数量：0

缺失百分比：0

重复数量：0

重复百分比：0%

所占空间大小：50G

所占空间中平均记录的大小：60 Byte

数值类型字段：CUST\_CNT

字典字段：无

字符串：LINE\_NAME

时间字段：RQ

未确定类型：无

## 2 数据定量评估

### 2.1 数据表评估

元数据	内容
表中文名	久事公交线路刷卡客流量表
表英文名	FACT_JS_BUS_CUST
主数据	否
数据颗粒度	日
更新频次	每日
字段数量	3

### 2.2 数据字段评估

元数据	内容
字段名 (中文)	日期、线路名称、客流量

字段名 (英文)	RQ、LINE_NAME,CUST_CNT
业务分类	公共类别
敏感级别	\
主数据	否
合格数量	3
合格率	100%
缺失数量	0
缺失率	0%
异常数量	0
异常率	0%
原子模型	
唯一性	
相关字段数量	3
时效性	
离散值个数	
离散分布	
最大值	
最小值	
可访问数量	全部
可访问率	100%
可获取数据量	205 万
可达率	100%
是否存在参照数据	否
准确数量	205 万
准确率	100%

## 2.3 数据分布统计

## 2.4 单表相关性分析

客流量

线路

日期

日期

线路

客流量

## 2.5 多表相关性分析

## 2.6 数据量统计

205 万条

## 2.7 样本数据

列出 10 行

主键	日期	线路名称	客流量
1	2021-11-4	71 路	32971
2	2021-11-4	189 路	17281
3	2021-11-4	虹桥枢纽 4 路	16701
4	2021-11-4	110 路	16511
5	2021-11-4	21 路	14661
6	2021-11-4	66 路	13570
7	2021-11-4	812 路	13569

8	2021-11-4	741 路	13488
9	2021-11-4	25 路	13181
10	2021-11-4	71 路区间	13007

## 2.8 异常数据

无

## 2.9 失效数据

无

## 3 数据非定量评估

数据质量非定量指标提供综述性的、非定量的质量信息。

### 3.1 数据目的

仅用于科学研究、规划设计的使用目的,且仅限于需求方使用,不得共享、转让、出境。

### 3.2 数据用途

仅用于科学研究、规划设计的使用。

### 3.3 数据日志

数据更新周期存量加更新,视需求方的需求而定。

### 3.4 数据可靠性

基于交通出行的明细数据,按需求汇总,数据真实、可靠。

### 3.5 可访问性

可访问性：用户通过接口访问的方式可以快速的获取到所需要的数据。

安全性：对敏感信息进行脱敏处理，同时对数据的访问获取有严格的限制。

## 4 评估结论

本次挂牌数据产品符合国内外相关规范、标准，具有实体完整性、引用完整性和域完整性，记录信息不存在异常或错误，记录符合值域一致性、格式一致性等。

上海公共交通卡股份有限公司

2021/11/22