



E



# 中国主要城市公共开放数据 数据质量评测报告

总报告  
(广州、深圳、杭州)

北京华矩管理咨询有限公司  
2024年5月29日

## 版权声明

本报告版权属于北京华矩管理咨询有限公司所有，并受法律保护。转载、摘编或利用其它方式使用本报告内容或者观点的，应注明“版权所有方及来源：北京华矩管理咨询有限公司《中国主要城市公共开放数据数据质量评测报告总报告（广州、深圳、杭州）》”。

违反上述声明者，版权所有方将依法追究其法律责任。

# 目 录

结论 .....	4
一、 项目概述 .....	8
1.1 测试目的 .....	8
1.2 测试样本数据情况 .....	8
二、 测试方法 .....	10
2.1 测试方法 .....	10
2.1.1 规范性测试 .....	10
2.1.2 完整性测试 .....	10
2.1.3 准确性测试 .....	10
2.1.4 一致性测试 .....	11
2.1.5 时效性测试 .....	11
2.1.6 可访问性测试 .....	11
2.2 选择的维度和指标 .....	11
2.3 数据质量度量方法 .....	13
2.3.1 数据质量度量指标计算方法 .....	13
2.3.2 总体评分计算方法 .....	13
三、 数据质量测试结果 .....	14
3.1 待测数据表清单 .....	14
3.1.1 广州市待测数据表清单 .....	14
3.1.2 深圳市待测数据表清单 .....	14
3.1.3 杭州市待测数据表清单 .....	15
3.2 广州市数据质量测试结果 .....	16
3.2.1 基础元数据测试结果-表基本属性 .....	16
3.2.2 完整性测试结果 .....	16
3.2.3 一致性测试结果 .....	21
3.2.4 准确性测试结果 .....	22
3.2.5 规范性测试结果 .....	24
3.2.6 时效性测试结果 .....	25
3.2.7 可访问性测试结果 .....	25
3.2.8 质量维度统计结果 .....	25
3.2.9 数据质量测试各局委办得分 .....	26
3.3 深圳市数据质量测试结果 .....	28
3.3.1 基础元数据测试结果-表基本属性 .....	28
3.3.2 完整性测试结果 .....	29
3.3.3 一致性测试结果 .....	32
3.3.4 准确性测试结果 .....	33
3.3.5 规范性测试结果 .....	35
3.3.6 时效性测试结果 .....	36
3.3.7 可访问性测试结果 .....	36

3.3.8 质量维度统计结果 .....	36
3.3.9 数据质量测试各局委办得分 .....	37
3.4 杭州市数据质量测试结果 .....	39
3.4.1 基础元数据测试结果-表基本属性 .....	39
3.4.2 完整性测试结果 .....	40
3.4.3 一致性测试结果 .....	43
3.4.4 准确性测试结果 .....	44
3.4.5 规范性测试结果 .....	46
3.4.6 时效性测试结果 .....	47
3.4.7 可访问性测试结果 .....	47
3.4.8 质量维度统计结果 .....	47
3.4.9 数据质量测试各局委办得分 .....	48
3.5 总体数据质量测试结果 .....	50
四、评测问题示例 .....	52

## 声 明

1. 本测试报告仅适用于报告中明确指出的被测样品及版本。
2. 本测试报告可作为对被测样品进行质量评价的参考依据。
3. 当被测样品出现版本更新或其它任何改变时，本测试报告不再适用。
4. 本报告中提供的质量问题解决建议需要在测试环境验证通过之后，才能在实际运行环境中实施，否则会带来一定风险。
5. 本报告版权属于北京华矩管理咨询有限公司所有，并受法律保护。转载、摘编或利用其它方式使用本报告内容或者观点的，应注明“版权所有方及来源：北京华矩管理咨询有限公司《中国主要城市公共开放数据数据质量评测报告总报告（广州、深圳、杭州）》”。违反上述声明者，版权所有方将依法追究其法律责任。

## 结论

在广州公共数据开放平台、深圳市政府数据开放平台、杭州市数据开放平台抽取样本数据，分别对交通管理局、市场监管局、生态环境局抽取样本数据共计 54 张表，612 个字段，1318502 条记录进行数据质量测试。测试结果如下：

### 一、整体描述

三个城市的 data 质量整体评分：95.16 分

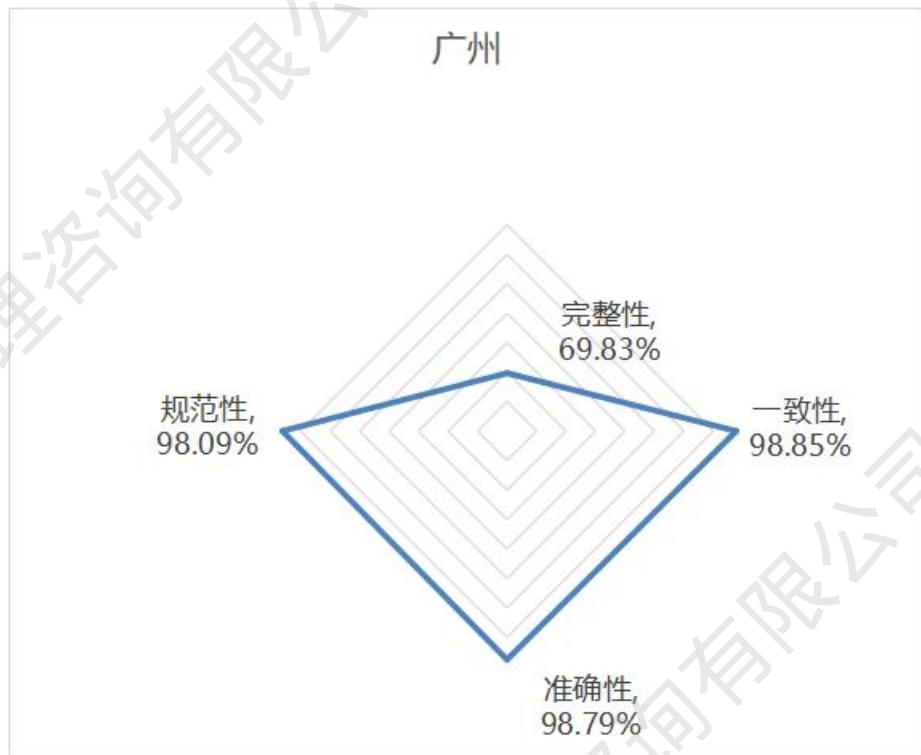


质量维度	完整性	一致性	准确性	规范性
质量维度评分	83.87%	99.45%	98.22%	99.12%
质量维度权重	25.00%	25.00%	25.00%	25.00%

## 二、广州整体数据质量维度评分

抽取广州市交通管理局 10 张表、广州市市场监督管理局 4 张表、广州生态环境局 4 张表共计 18 张表，200 个字段，111933 条记录进行数据质量测试：

广州市数据质量整体评分：91.39 分。

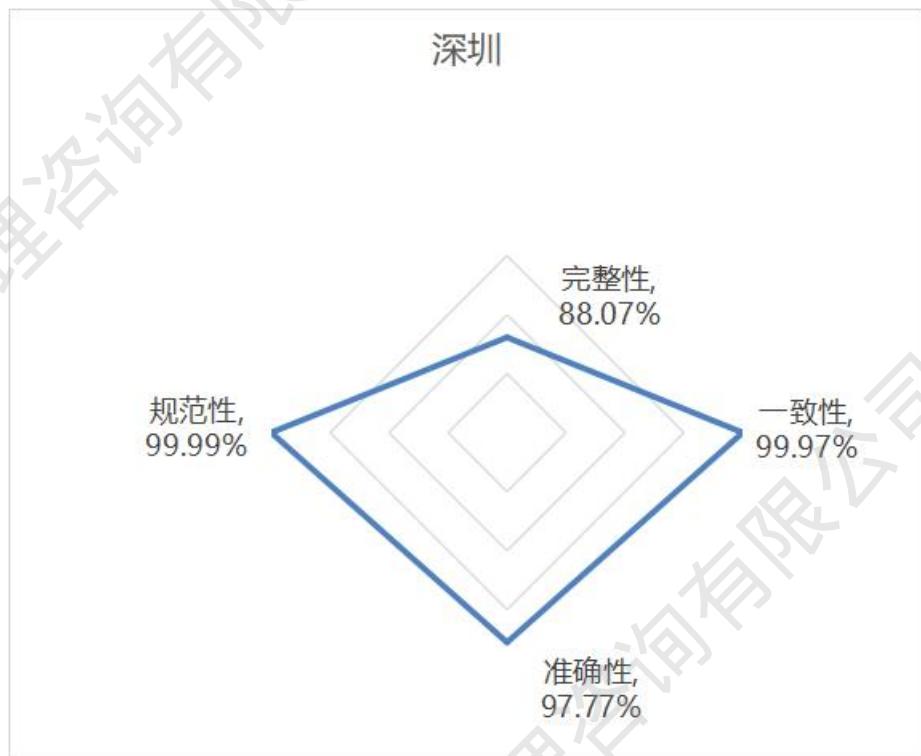


质量维度	完整性	一致性	准确性	规范性
质量维度评分	69.83%	98.85%	98.79%	98.09%
质量维度权重	25.00%	25.00%	25.00%	25.00%

### 三、深圳整体数据质量维度评分

抽取深圳市交通管理局 10 张表、深圳市市场监督管理局 4 张表、深圳生态环境局 4 张表，共计 18 张表，198 个字段，259066 条记录进行数据质量测试：

深圳市数据质量整体评分：96.45 分。



质量维度	完整性	一致性	准确性	规范性
质量维度评分	88.07%	99.97%	97.77%	99.99%
质量维度权重	25.00%	25.00%	25.00%	25.00%

#### 四、杭州整体数据质量维度评分

抽取杭州市交通管理局 10 张表、杭州市市场监督管理局 4 张表、杭州生态环境局 4 张表，共计 18 张表，214 个字段，947503 条记录进行数据质量测试：

杭州数据质量整体评分：97.65 分。



质量维度	完整性	一致性	准确性	规范性
质量维度评分	93.71%	99.52%	98.09%	99.30%
质量维度权重	25.00%	25.00%	25.00%	25.00%

# 一、项目概述

## 1.1 测试目的

高质量数据是数据应用的基础，而当下影响数据质量的提升既有观念问题，也有技术问题，既有时间问题，也有成本问题，鉴于此，北京华矩管理咨询有限公司对当前政府公共开放数据进行数据质量评测，并形成政府公共开放数据的数据质量指数，为政府公共开放数据的建设及数据质量提升提供有益的建议和参考，并以此带动全社会对数据质量的重视，对数据质量评测技术及方法的认识，从而提升全社会的数据质量，为数据要素 X 及新质生产力献策献力。

## 1.2 测试样本数据情况

待测数据情况	城市	交通运输局			市场监管局			生态环境局			总计		
		表	字段	记录	表	字段	记录	表	字段	记录	表	字段	记录
	广州	10	116	55205	4	49	37244	4	35	19484	18	200	111933
	深圳	10	108	93119	4	46	130000	4	44	35947	18	198	259066
	杭州	10	109	339038	4	52	405868	4	53	202597	18	214	947503
	总计										54	612	1318502
数据采集基本描述	1) 数据来源于广州市公共数据开放平台（网址： <a href="https://data.gz.gov.cn/">https://data.gz.gov.cn/</a> ）、深圳市政府数据开放平台（网址： <a href="https://opendata.sz.gov.cn/">https://opendata.sz.gov.cn/</a> ）、杭州市数据开放平台（网址： <a href="https://data.hangzhou.gov.cn/">https://data.hangzhou.gov.cn/</a> ）												
	2) 抽取样本数据为每个城市交通管理局(10 张表)、市场监督管理局(4 张表)、生态环境局(4 张表)；所抽取样本数据为 csv 文件数据集，数据量过大时，抽取为网站所提供该表抽取时最新一个 csv 文件。												

	<p>3) 数据下载时间，广州数据为 4 月 29 日至 5 月 14 日期间下载，深圳数据为 4 月 29 日至 5 月 14 日期间下载，杭州数据为 4 月 29 日至 5 月 26 日期间下载。</p> <p>4) 样本数据涉及交通、企业、环保等民生数据。</p> <p>5) 测试流程包含获取数据、制定方案、导入数据、实施分析、统计汇总和出具报告。</p>
待测数据包含的相关领域数据	<p>1. <input type="checkbox"/> 个人隐私信息</p> <p>2. <input type="checkbox"/> 国家法律保护的数据</p> <p>3. <input checked="" type="checkbox"/> 政府公共开放数据</p> <p>4. <input type="checkbox"/> 其他（特别说明）</p>
数据所属阶段	政府对外公开数据

表 1-1: 测试样本数据情况

## 二、 测试方法

### 2.1 测试方法

本次测试参照《GB/T 36344-2018 信息技术 数据质量评价指标》中要求，将通过工具扫描的方式分析待测数据的基础元数据，通过基础元数据的了解数据基本情况，然后对各个质量维度有针对性的深度扫描分析，结合业务需求及系统设计来分析判断待测数据的数据质量情况。

#### 2.1.1 规范性测试

规范性即数据符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度。评价指标为数据标准、数据模型、元数据、业务规则、权威参考数据及安全规范。

#### 2.1.2 完整性测试

完整性即按照数据规则要求，数据元素被赋予数值的程度。本次测试存在空值、空格值、NULL 值、“-”等，均认为存在完整性问题。评价指标为数据元素完整性和数据记录完整性。

#### 2.1.3 准确性测试

准确性即数据准确表示其所描述的真实实体（实际对象）真实值的程度。评价指标为数据内容正确性、数据格式合规性、数据重复率、数据唯一性和脏数据出现率。

## 2.1.4 一致性测试

一致性即数据与其他特定上下文中使用的数据无矛盾的程度。评价指标为相同数据一致性和关联数据一致性。

## 2.1.5 时效性测试

时效性即数据在时间变化中的正确程度。评价指标为基于时间段的正确性、基于时间点及时性和时序性。

## 2.1.6 可访问性测试

可访问性即数据能够被访问的程度。评价指标为可访问和可用性。

## 2.2 选择的维度和指标

质量维度	指标编号	指标名称	指标描述	是否选择
规范性	101	数据标准	<p>数据符合数据标准的度量。</p> <p>注 1：评价数据质量时需要收集数据在命名、创建、定义、更新和归档时遵循的标准，包括国际标准、国家标准、行业标准、地方标准或相关规定等。</p> <p>注 2：和数据归档一样甚至更重要，在一个完整的数据规则中旧数据的销毁一般也有一个比较详细且具有可执行性的规定</p>	是
规范性	102	数据模型	<p>数据符合数据模型的度量。</p> <p>注 1：数据模型是一种直观描述组织数据结构的手段，是数据表达的规范。</p> <p>注 2：评价数据质量时需要检查是否存在清晰可理解的数据模型定义以及这些数据的组织形式</p>	不适用
规范性	103	元数据	<p>数据符合元数据定义的度量。</p> <p>注：元数据标注、描述或刻画其他数据、以使检索、或使用信息更容易。评价数据质量时需要检查是否提供可解读的元数据文档。</p> <p>示例：包含各字段名称、描述、类型值域等内容的数据字典为一种元数据文档</p>	是

规范性	104	业务规则	数据符合业务规则的度量。 注1：业务规则是一种权威性原则或指导方针，用来描述业务交互，并建立行动和数据行为结果及完整性的规则。 注2：评价数据质量时需要检查是否存在良好归档的业务规则	不适用
规范性	105	权威参考数据（权威参考源）	参考数据是系统、应用软件、数据库、流程、报告及交易记录和主记录用来参考的数值集合或分类表。 注：评价数据质量时需要收集参考数据列表。 示例：一张用于一个特定字段的有效值列表为一种参考数据类型	不适用
规范性	106	安全规范	安全规范是安全和隐私方面的规则。包括数据权限管理、数据脱敏处理等	不适用
完整性	201	数据元素完整性	按照业务规则要求，数据集中应被赋值的数据元素的赋值程度	是
完整性	202	数据记录完整性	按照业务规则要求，数据集中应被赋值的数据记录的赋值程度	是
准确性	301	数据内容正确性	数据内容是否是预期数据	是
准确性	302	数据格式合规性	数据格式（包括数据类型、数值范围、数据长度、精度等）是否满足预期要求。 示例：性别一栏不能出现男/女以外的内容；身份证号不能出现标点符号；以及对字符编码的一些限制。都需要通过规定内容的格式来实现	是
准确性	303	数据重复率	特定字段、记录、文件或数据集意外重复的度量	是
准确性	304	数据唯一性	特定字段、记录、文件或数据集唯一性的度量	是
准确性	305	脏数据出现率	正确字段、记录、文件或数据集之外无效数据的度量。 示例：事务发生回滚时由于回滚机制不健全或不完善导致可能出现脏数据	是
一致性	401	相同数据一致性	同一数据在不同位置存储或被不同应用或用户使用时。数据的一致性；数据发生变化时。存储在不同位置的同一数据被同步修改	是
一致性	402	关联数据一致性	根据一致性约束规则检查关联数据的一致性	是
时效性	501	基于时间段的正确性	基于日期范围的记录数或频率分布符合业务需求的程度	不适用
时效性	502	基于时间点及时性	基于时间戳的记录数、频率分布或延迟时间符合业务需求的程度	不适用
时效性	503	时序性	数据集中同一实体的数据元素之间的相对时序关系	不适用
可访问性	601	可访问	数据在需要时的可获取性	不适用
可访问性	602	可用性	数据在设定有效生存周期内的可使用性	不适用

表 2-1: 质量维度与指标

## 2.3 数据质量度量方法

### 2.3.1 数据质量度量指标计算方法

报告中涉及的数据质量指标计算方法如下：

**单字段：**

单字段在某一质量维度，存在的问题记录占比  $A = \text{问题记录数} / \text{总记录数}$ 。

单字段在某一质量维度的得分为  $B = 1 - A$ 。

**单表：**

1. 完整性 (%)：单表中各字段完整性得分的平均值。
2. 规范性 (%)：单表中各字段规范性得分的平均值。
3. 一致性 (%)：单表中各字段一致性得分的平均值。
4. 准确性 (%)：单表中各字段准确性得分的平均值。

**总数据集：**

总数据集各维度的得分为各表相应维度的得分的平均值。

### 2.3.2 总体评分计算方法

整体评分根据权重来进行加权计算。本次测试对于所测试的维度权重平均分配。则总体评分计算如下：总分=完整性得分\*25%+规范性得分\*25%+一致性得分\*25%+准确性得分\*25%

### 三、 数据质量测试结果

#### 3.1 待测数据表清单

##### 3.1.1 广州市待测数据表清单

序号	城市	局委办	表名	字段数	记录数
1	广州	交通运输局	国省道公路基本信息表	8	89
2	广州	交通运输局	广州公交企业信息	6	104
3	广州	交通运输局	广州公交线路信息	5	4696
4	广州	交通运输局	广州市交通运输局行政处罚信用信息公示	30	10000
5	广州	交通运输局	广州市交通运输局行政许可信用信息公示	28	10000
6	广州	交通运输局	广州市危运企业情况	7	717
7	广州	交通运输局	广州市客运企业情况	7	5781
8	广州	交通运输局	广州市机动车维修企业资料及分布查询	7	10000
9	广州	交通运输局	广州市经营性停车场基础信息	7	8213
10	广州	交通运输局	路政许可公示信息	11	5605
11	广州	市场监管局	企业基本信息	14	10000
12	广州	市场监管局	企业异常名录详细信息	7	10000
13	广州	市场监管局	广州市市场主体农合年报基本信息	13	10000
14	广州	市场监管局	食品生产企业信息	15	7244
15	广州	生态环境局	广州市城市集中式生活饮用水水源地水质信息	7	690
16	广州	生态环境局	广州市重点排污单位名录	6	6174
17	广州	生态环境局	空气质量日报	6	10000
18	广州	生态环境局	辐射安全许可证核发	16	2620

表 3-1-1: 广州市待测表数据清单

##### 3.1.2 深圳市待测数据表清单

序号	城市	局委办	表名	字段数	记录数
1	深圳	交通运输局	公交线路信息	13	257
2	深圳	交通运输局	出租车企业信息	5	1633
3	深圳	交通运输局	国内航运企业名录	5	400
4	深圳	交通运输局	市交通运输局 行政处罚信息	27	10000
5	深圳	交通运输局	市交通运输局 行政许可信息	21	10000

6	深圳	交通运输局	普通货运企业基本信息	6	10000
7	深圳	交通运输局	深圳公交线路轨迹表	7	57363
8	深圳	交通运输局	港口经营企业名录	5	31
9	深圳	交通运输局	道路包车 机动 客运经营许可证明	10	2344
10	深圳	交通运输局	道路客运企业版块行政检查汇总公示明细	9	1091
11	深圳	市场监管局	经营异常名录	8	10000
12	深圳	市场监管局	药品经营许可证	15	10000
13	深圳	市场监管局	隶属企业信息	7	100000
14	深圳	市场监管局	食品经营许可证	16	10000
15	深圳	生态环境局	深圳市辐射安全许可证公示名单	3	757
16	深圳	生态环境局	空气质量日报	9	32970
17	深圳	生态环境局	重点排污单位废水监测报告	24	1629
18	深圳	生态环境局	饮用水质月度情况	8	591

表 3-1-2: 深圳市待测表数据清单

### 3.1.3 杭州市待测数据表清单

序号	城市	局委办	表名	字段数	记录数
1	杭州	交通运输局	杭州公路客运周转量信息	9	29
2	杭州	交通运输局	杭州公路客运量信息	9	29
3	杭州	交通运输局	杭州公路货运周转量信息	9	29
4	杭州	交通运输局	杭州公路货运量信息	9	29
5	杭州	交通运输局	杭州水路客运量信息	9	29
6	杭州	交通运输局	杭州货运车辆基本信息	10	119001
7	杭州	交通运输局	杭州高速及国省道路基本信息	7	38
8	杭州	交通运输局	省回流 网络预约出租汽车车辆信息	8	199481
9	杭州	交通运输局	省回流 货运企业信息	9	14235
10	杭州	交通运输局	省回流 道路运输经营许可证	30	6138
11	杭州	市场监管局	企业经营异常名录信息	9	150000
12	杭州	市场监管局	在册市场主体基本信息	13	100000
13	杭州	市场监管局	杭州食品经营许可证	10	150000
14	杭州	市场监管局	药品经营许可证信息	20	5868
15	杭州	生态环境局	杭州市放射源信息	14	1618
16	杭州	生态环境局	水质自动站小时采样数据信息	21	150000
17	杭州	生态环境局	污染源基础信息	12	39510
18	杭州	生态环境局	重点排污单位公开目录信息	6	11469

表 3-1-3: 杭州市待测表数据清单

## 3.2 广州市数据质量测试结果

### 3.2.1 基础元数据测试结果-表基本属性

名称	字段数	记录数	数值	行最小长度	行最大长度
广州_交通运输局_国省道公路基本信息表	8	89	102	48	75
广州_交通运输局_广州公交企业信息	6	104	248	53	78
广州_交通运输局_广州公交线路信息	5	4696	12438	29	77
广州_交通运输局_广州市交通运输局行政处罚信用信息公示	30	10000	26467	466	779
广州_交通运输局_广州市交通运输局行政许可信用信息公示	28	10000	22603	222	2966
广州_交通运输局_广州市危运企业情况	7	717	1438	44	351
广州_交通运输局_广州市客运企业情况	7	5781	14145	54	349
广州_交通运输局_广州市机动车维修企业资料及分布查询	7	10000	12442	63	201
广州_交通运输局_广州市经营性停车场基础信息	7	8213	23039	52	165
广州_交通运输局_路政许可公示信息	11	5605	8998	141	2394
广州_市场监管局_企业基本信息	14	10000	50408	170	1631
广州_市场监管局_企业异常名录详细信息	7	10000	10652	84	99
广州_市场监管局_广州市市场主体农合年报基本信息	13	10000	9990	113	161
广州_市场监管局_食品生产企业信息	15	7244	31761	141	390
广州_生态环境局_广州市城市集中式生活饮用水水源地水质信息	7	690	100	42	55
广州_生态环境局_广州市重点排污单位名录	6	6174	1984	46	105
广州_生态环境局_空气质量日报	6	10000	293	49	83
广州_生态环境局_辐射安全许可证核发	16	2620	372	126	129

表 3-2-1: 表基本属性

### 3.2.2 完整性测试结果

数据完整性检查，发现以下问题：

字段记录中存在空值、空格值、NULL 值或者是“-”的现象。在测试的 18 张表中通过数据完整性测试的表有 0 个，占总表数的 0%，测试的每张表中都存在完整性问题。发现完整性问题 87 个，涉及 18 张表，完整性问题主要为数据元素完整性。

样本数据中存在完整性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
广州市交通运输局行政许可信用信息公示	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在空值和 null 值	9998	99.980%
路政许可公示信息	许可决定日期	字段空值率(单表)	完整性	数据元素完整性	存在空值	3	0.054%
国省道公路基本信息表	市域范围内公路里程公里	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	89	100.000%
国省道公路基本信息表	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	89	100.000%
广州公交企业信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	104	100.000%
广州公交线路信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	4696	100.000%
广州市交通运输局行政处罚信用信息公示	备注	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	暂扣或吊销证照名称及编号	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	没收违法所得没收非法财物的金额万元	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	罚款金额万元	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	证件类型	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	法定代表人证件类型	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	3765	37.650%
广州市交通运输局行政处罚信用信息公示	行政相对人代码6社会组织登记证号	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	行政相对人代码5事业单位证书号	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	行政相对人代码4税务登记号	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	行政相对人代码3组织机构代码	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	行政相对人代码2工商注册号	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政处罚信用信息公示	行政相对人代码1统一社会信用代码	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%

广州市交通运输局行政处罚信用信息公示	更新时间	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政许可信用信息公示	许可证书名称	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	63	0.630%
广州市交通运输局行政许可信用信息公示	证件类型	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	9998	99.980%
广州市交通运输局行政许可信用信息公示	法定代表人证件类型	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	9674	96.740%
广州市交通运输局行政许可信用信息公示	行政相对人代码 6 社会组织登记证号	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	9998	99.980%
广州市交通运输局行政许可信用信息公示	行政相对人代码 5 事业单位证书号	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政许可信用信息公示	行政相对人代码 4 税务登记号	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政许可信用信息公示	行政相对人代码 3 组织机构代码	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政许可信用信息公示	行政相对人代码 2 工商注册号	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市交通运输局行政许可信用信息公示	行政相对人代码 1 统一社会信用代码	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	9996	99.960%
广州市交通运输局行政许可信用信息公示	行政相对人类别	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	2	0.020%
广州市危运企业情况	更新时间	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	717	100.000%
广州市危运企业情况	经营地址	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	117	16.318%
广州市危运企业情况	所在辖区	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	25	3.487%
广州市客运企业情况	更新时间	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	5781	100.000%
广州市客运企业情况	所在辖区	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	63	1.090%
广州市机动车维修企业资料及分布查询	联系电话	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市机动车维修企业资料及分布查询	更新时间	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市机动车维修企业资料及分布查询	所属辖区	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	5677	56.770%
广州市机动车维修企业资料及分布查询	企业类型	字段空值率（单表）	完整性	数据元素完整性	存在 null 值	4318	43.180%
广州市经营性停车场基础信	更新时间	字段空值	完整性	数据元素	存在 null	8213	100.000%

息		率(单表)		完整性	值		
路政许可公示信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	5605	100.000%
企业基本信息	数据汇总时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
企业基本信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
企业异常名录详细信息	列入经营异常名录原因类型 中文	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
企业异常名录详细信息	移出经营异常名录原因 中文	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
企业异常名录详细信息	数据汇总时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
企业异常名录详细信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市市场主体农合年报基本信息	数据汇总时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
广州市市场主体农合年报基本信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
食品生产企业信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	7244	100.000%
广州市城市集中式生活饮用水水源地水质信息	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	690	100.000%
广州市重点排污单位名录	统一社会信用代码 组织机构代码	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	6174	100.000%
广州市重点排污单位名录	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	6174	100.000%
空气质量日报	Aqi	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
空气质量日报	空气质量类别	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	172	1.720%
空气质量日报	首要污染物	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	172	1.720%
空气质量日报	更新时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	10000	100.000%
辐射安全许可证核发	法定代表人姓名	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	许可内容	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	行政相对人统一社会信用代码	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	许可机关	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	更新时间	字段空值	完整性	数据元素	存在 null	2620	100.000%

		率(单表)		完整性	值		
辐射安全许可证核发	行政相对人名称	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	审批类别	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	项目名称	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	许可截止期	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	当前状态	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	行政相对人组织机构代码	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	许可决定日期	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
辐射安全许可证核发	行政许可决定书文号	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2620	100.000%
广州市市场主体农合年报基本信息	销售额或营业收入	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2432	24.320%
广州市市场主体农合年报基本信息	金融贷款	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2432	24.320%
广州市市场主体农合年报基本信息	成员人数	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	80	0.800%
广州市市场主体农合年报基本信息	成员人数中农民人数	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	80	0.800%
广州市市场主体农合年报基本信息	本年度新增成员人数	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	80	0.800%
广州市市场主体农合年报基本信息	本年度退出成员人数	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	80	0.800%
广州市市场主体农合年报基本信息	获得政府扶持资金补助	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2432	24.320%
广州公交企业信息	机构类型	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	1	0.962%
广州市客运企业情况	开业时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	486	8.407%
广州市危运企业情况	开业时间	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	465	64.854%
企业基本信息	营业期限	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	9902	99.020%
食品生产企业信息	签发日期	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	3	0.041%
食品生产企业信息	有效期至	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	3	0.041%
企业异常名录详细信息	移出日期	字段空值	完整性	数据元素	存在 null	6995	69.950%

		率(单表)		完整性	值		
辐射安全许可证核发	备注	字段空值率(单表)	完整性	数据元素完整性	存在 null 值	2619	99.962%
食品生产企业信息	法定代表人负责人	字段空值率(单表)	完整性	数据元素完整性	存在“-”值	1	0.014%
食品生产企业信息	社会信用代码	字段空值率(单表)	完整性	数据元素完整性	存在“-”值	1	0.014%
广州市城市集中式生活饮用水水源地水质信息	超标指标及超标倍数	字段空值率(单表)	完整性	数据元素完整性	存在“-”值	690	100.000%

表 3-2-2: 一致性测试结果

### 3.2.3 一致性测试结果

数据一致性检查，发现以下问题：

在测试的 18 张表中通过通过单表数据一致性测试的表有 15 个，占总表数的 83.3%；发现单表中一致性问题 3 个，涉及 3 张表。单表中关联数据一致性，存在逻辑关联的字段，前后不一致，需要排查不一致的原因。

在数据集中也发现跨表一致性问题，涉及 2 张表 2 个字段，为相同数据一致性方面，相同的数据在不同的表中出现的内容不一致，存在问题，没有统一的标准。

样本数据中存在单表一致性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
广州公交线路信息	线路名称-线路编码	单表不同字段逻辑冲突	一致性	关联数据一致性	相同的线路名称，出现了不同的线路编码。	91	1.938%
广州公交企业信息	机构简码-机构名称	单表不同字段逻辑冲突	一致性	关联数据一致性	相同的机构简码，机构名称不同。	6	5.769%
食品生产企业信息	许可证号，签发日期，证书状态-企业名称	单表不同字段逻辑冲突	一致性	关联数据一致性	相同的许可证号，签发日期，证书状态；企业名称不同。	6	0.083%

表 3-2-3: 单表一致性测试结果

样本数据中存在跨表一致性问题的字段如下：

表	问题字段 /说明	问题类别	质量维度	指标名称	问题描述
企业基本信息	注册号	数据规范不一致-格式（跨表）	一致性	相同数据一致性	不同表中，注册号的长度不一致。
广州市市场主体农合年报基本信息		数据规范不一致-格式（跨表）	一致性	相同数据一致性	不同表中，注册号的长度不一致。

表 3-2-4: 跨表一致性测试结果

### 3.2.4 准确性测试结果

数据准确性检查，发现以下问题：

在测试的 18 张表中通过准确性测试的表有 8 个，占总表数的 44%。存在数据准确性问题 14 个，涉及 10 张表。样本数据中存在准确性问题有数据格式合规性和数据内容正确性，数据格式合规性为多个字段存在编码的格式不一致的问题，数据内容正确性多为出现不符合该字段的值出现，数据重复率则是多张表存在复制行。

样本数据中存在准确性问题的字段如下：

表	问题字段 /说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
广州公交线路信息	线路编码	数据规范不一致-格式（单表）	准确性	数据内容合规性	线路编码格式不一致，出现八种格式。绝大多数编码为 5 位，出现三条记录不为 5 位的，可能存在一定问题。	3	0.064%
路政许可公示信息	法定代表人姓名	字段值异常（最大值及内容）（单	准确性	数据内容正确性	法人姓名出现测试值、证件号码、null 值等。	16	0.285%

		表)					
路政许可公示信息	企业名称	字段值异常(最大值及内容)(单表)	准确性	数据内容正确性	企业名称出现人的姓名，可能存在一定问题。	7	0.125%
食品生产企业信息	企业名称	字段值异常(最大值及内容)(单表)	准确性	数据内容正确性	企业名称出现全是问号的情况，还出现编码值。	7	0.097%
路政许可公示信息	负责人姓名	字段值异常(最大值及内容)(单表)	准确性	数据内容正确性	负责人姓名中记录的是通行证的类型，并不是负责人的名字。	5605	100.000%
广州市交通运输局行政许可信用信息公示	许可证书名称	字段值异常(最大值及内容)(单表)	准确性	数据内容正确性	许可证书名称中出现的并不是许可证的相关类型		
广州市市场主体农合年报基本信息	年报年度	数据规范不一致-格式(单表)	准确性	数据内容合规性	年报年度的值存在空格和点，如“2014.”.	117	1.170%
食品生产企业信息	所属街道	字段值异常(最大值及内容)(单表)	准确性	数据内容正确性	所属街道出现代码，可能是未对街道代码进行转换造成，存在一定问题。	40	0.552%
国省道公路基本信息表	复制行	单表中重复记录(2个(含2条)以上完全一样的记录)	准确性	数据重复率	存在复制行	65	73.034%
广州市危运企业情况	复制行	单表中重复记录(2个(含2条)以上完全一样的记录)	准确性	数据重复率	存在复制行	17	2.371%
广州市交通运输局行政处罚信用信息公示	复制行	单表中重复记录(2个(含2条)以上完全一样的记录)	准确性	数据重复率	存在复制行	8	0.080%
广州市经营性停车场基础信息	复制行	单表中重复记录(2个(含2条)以上完全一样的记录)	准确性	数据重复率	存在复制行	2	0.024%
食品生产企业信息	复制行	单表中重复记录(2个(含2条)以上完全一样的记录)	准确性	数据重复率	存在复制行	17	0.235%
空气质量日报	复制行	单表中重复记录(2个(含2条)以上完全一样的记录)	准确性	数据重复率	存在复制行	32	0.320%

表 3-2-5: 准确性测试结果

### 3.2.5 规范性测试结果

数据规范性检查，发现以下问题：

在测试的 18 张表中通过规范性测试的表有 11 个，占总表数的 61%。

样本数据中存在数据规范性问题有 7 个，涉及 7 张表。多为数据标准不一致，单字段未有统一标准。

样本数据中存在规范性问题的表和字段如下：

表	问题字段 /说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
广州市危运企业情况	所在辖区	数据规范不一致-内容（单表）	规范性	数据标准	所在辖区的标准不统一，比如“南沙”和“南沙区”出现了同意不同词现象。	/	/
广州市客运企业情况	所在辖区	数据规范不一致-内容（单表）	规范性	数据标准	存在标准不统一的情况，如“天河”和“广州天河”，且绝大部分都是城市+区的标准。	127	2.197%
食品生产企业信息	证书状态	数据规范不一致-内容（单表）	规范性	数据标准	标准不统一，同时里面的值存在特殊符号。	13	0.179%
广州市交通运输局行政许可信用信息公示	数据来源单位	数据规范不一致-内容（单表）	规范性	数据标准	来源单位的值存在问题，带有空格，可能会导致进行数据统计会少计算该值数目。	129	1.290%
广州市市场主体农合年报基本信息	获得政府扶持资金补助	数据规范不一致-内容（单表）	规范性	数据标准	该字段金额单位方面可能存在一定问题，最小值小于 1，有部分过小的值单位可能是万元，部分比较大的值，如最大值 500000，单位可能是元。所以推测该字段的单位不统一。	/	/
广州市城市集中式生活饮用水水源地水质信息	水质类别	数据规范不一致-内容（单表）	规范性	数据标准	水质类别出现同意不同词现象，标准并未统一。	/	/
广州市重点排污单位名录	分类	数据规范不一致-内容（单表）	规范性	数据标准	出现同意不同词的现象，同时存在错别字。	/	/

表 3-2-6: 规范性测试结果

### 3.2.6 时效性测试结果

本次测试数据不适用时效性测试。

### 3.2.7 可访问性测试结果

本次测试数据不适用可访问性测试。

### 3.2.8 质量维度统计结果

#### 含义解释

- 1) 通过测试的字段: 如 A 表 (单表) age 字段, 存在完整性问题, 认为 age 字段未通过完整性测试; 反之则说明, age 字段通过了完整性测试。
- 2) 通过测试的总字段个数: 指对整个数据集(18 张表)中各表通过相应维度测试的字段个数之和。
- 3) 通过测试的表字段个数: 指对单个表中通过测试的字段数的统计。

以上三个含义解释都是针对样本数据, 在本次测试中的结果的统计。

#### 3.2.8.1 通过测试的总字段个数

总字段	完整性	一致性	准确性	规范性
200	113	197	186	193

表 3-2-7: 通过测试的总字段个数

### 3.2.8.2 通过测试的表字段个数

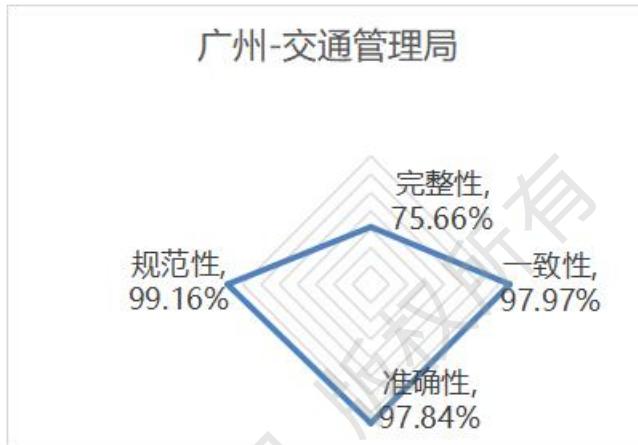
表名	总字段数	完整性	一致性	准确性	规范性
广州_交通运输局_国省道公路基本信息表	8	6	8	7	8
广州_交通运输局_广州公交企业信息	6	4	5	6	6
广州_交通运输局_广州公交线路信息	5	4	4	4	5
广州_交通运输局_广州市交通运输局行政处罚信用信息公示	30	17	30	29	30
广州_交通运输局_广州市交通运输局行政许可信用信息公示	28	17	28	27	27
广州_交通运输局_广州市危运企业情况	7	3	7	6	6
广州_交通运输局_广州市客运企业情况	7	4	7	7	6
广州_交通运输局_广州市机动车维修企业资料及分布查询	7	3	7	7	7
广州_交通运输局_广州市经营性停车场基础信息	7	6	7	6	7
广州_交通运输局_路政许可公示信息	11	9	11	8	11
广州_市场监管局_企业基本信息	14	11	14	14	14
广州_市场监管局_企业异常名录详细信息	7	2	7	7	7
广州_市场监管局_广州市市场主体农合年报基本信息	13	4	13	12	12
广州_市场监管局_食品生产企业信息	15	10	14	12	14
广州_生态环境局_广州市城市集中式生活饮用水水源地水质信息	7	5	7	7	6
广州_生态环境局_广州市重点排污单位名录	6	4	6	6	5
广州_生态环境局_空气质量日报	6	2	6	5	6
广州_生态环境局_辐射安全许可证核发	16	2	16	16	16
总计	200	113	197	186	193

表 3-2-8: 通过测试的表字段个数

### 3.2.9 数据质量测试各局委办得分

- 广州市交通管理局整体数据质量维度评分（10 张表，112 个字段，55205 条记录）

广州市交通管理局数据质量评分：92.66 分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	75.66%	97.97%	97.84%	99.16%
质量维度权重	25.00%	25.00%	25.00%	25.00%

2. 广州市市场监督管理局整体数据质量维度评分（4张表，49个字段，37244条记录）

广州市市场监督管理局数据质量评分：92.21分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	70.90%	99.90%	99.96%	98.07%
质量维度权重	25.00%	25.00%	25.00%	25.00%

3. 广州市生态环境局整体数据质量维度评分（4张表，35个字

段，19484 条记录）

广州市生态环境局数据质量评分：87.40 分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	54.17%	100.00%	99.99%	95.43%
质量维度权重	25.00%	25.00%	25.00%	25.00%

### 3.3 深圳市数据质量测试结果

#### 3.3.1 基础元数据测试结果-表基本属性

名称	字段数	记录数	数值	行最小长度	行最大长度
深圳_交通运输局_公交线路信息	13	257	1502	81	135
深圳_交通运输局_出租车企业信息	5	1633	226	19	76
深圳_交通运输局_国内航运企业名录	5	400	1136	24	104
深圳_交通运输局_市交通运输局 行政处罚信息	27	10000	49999	302	553
深圳_交通运输局_市交通运输局 行政许可信息	21	10000	25222	172	2109
深圳_交通运输局_普通货运企业基本信息	6	10000	13828	34	51
深圳_交通运输局_深圳公交线路轨迹表	7	57363	140882	147	156
深圳_交通运输局_港口经营企业名录	5	31	77	48	58
深圳_交通运输局_道路包车 机动 客运经营许可证明	10	2344	7962	106	267
深圳_交通运输局_道路客运企业版块行政检查汇总公示明细	9	1091	3878	97	768
深圳_市场监管局_经营异常名录	8	10000	55038	90	152
深圳_市场监管局_药品经营许可证	15	10000	29101	178	403

深圳_市场监管局_隶属企业信息	7	100000	229423	50	231
深圳_市场监管局_食品经营许可证	16	10000	20735	162	292
深圳_生态环境局_深圳市辐射安全许可证公示名单	3	757	2121	19	83
深圳_生态环境局_空气质量日报	9	32970	33413	23	115
深圳_生态环境局_重点排污单位废水监测报告	24	1629	4898	72	225
深圳_生态环境局_饮用水质月度情况	8	591	825	36	6929

表 3-3-1: 表基本属性

### 3.3.2 完整性测试结果

数据完整性检查，发现以下问题：

字段记录中存在空值、空格值、NULL 值或者是“-”的现象。在测试的 18 张表中通过数据完整性测试的表有 1 个，占总表数的 5.67%。发现完整性问题 55 个，涉及 17 张表，完整性问题主要为数据元素完整性。

样本数据中存在完整性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
公交线路信息	Companyid	字段空值率(单表)	完整性	数据元素完整性	存在空值	1	0.389%
公交线路信息	Id	字段空值率(单表)	完整性	数据元素完整性	存在空值	1	0.389%
国内航运企业名录	Id	字段空值率(单表)	完整性	数据元素完整性	存在空值	1	0.250%
国内航运企业名录	Sjc	字段空值率(单表)	完整性	数据元素完整性	存在空值	1	0.250%
市交通运输局行政处罚信息	备注	字段空值率(单表)	完整性	数据元素完整性	存在空值	10000	100.000%
市交通运输局行政许可信息	备注	字段空值率(单表)	完整性	数据元素完整性	存在空值	8351	83.510%
道路客运企业版块行政检查汇总公示明细	备注	字段空值率(单表)	完整性	数据元素完整性	存在空值	1091	100.000%
道路客运企业版块行政检查汇总	处置措施	字段空值率(单表)	完整性	数据元素完整性	存在空值	6	0.550%

公示明细							
饮用水质月度情况	Symc	字段空值率 (单表)	完整性	数据元素完整性	存在空值	1	0.169%
经营异常名录	Jdwsh	字段空值率 (单表)	完整性	数据元素完整性	存在空格值和"-"	6799	67.990%
经营异常名录	Ycsy	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	6587	65.870%
经营异常名录	Tyshxydm	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	283	2.830%
经营异常名录	Zzjgdm	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	283	2.830%
隶属企业信息	Entregno	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	1898	1.898%
隶属企业信息	Entname	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	2	0.002%
隶属企业信息	Addr	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	131	0.131%
隶属企业信息	Subattcode	字段空值率 (单表)	完整性	数据元素完整性	存在空格值和"-"	1295	1.295%
重点排污单位废水监测报告	Wrshxyl	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	1	0.061%
重点排污单位废水监测报告	Sd	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	1	0.061%
出租车企业信息	Zcdz	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	1	0.061%
饮用水质月度情况	Cbwrw	字段空值率 (单表)	完整性	数据元素完整性	存在"-"和 "null"	591	100.000%
空气质量日报	Aqi	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	679	2.059%
重点排污单位废水监测报告	Ph	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	653	40.086%
市交通运输局 行政处罚信息	社会组织登记证号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	9645	96.450%
市交通运输局 行政处罚信息	事业单位证书号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	9645	96.450%
重点排污单位废水监测报告	Fhw	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	1394	85.574%
饮用水质月度情况	Jcsj	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	50	8.460%
道路客运企业版块行政检查汇总公示明细	检查时间	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	163	14.940%
空气质量日报	Jcsj	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	30353	92.062%

出租车企业信息	Yzbm	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	1630	99.816%
市交通运输局 行政处罚信息	工商注册号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	9995	99.950%
深圳市辐射安全许可证公示名单	Dwmc	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	49	6.473%
市交通运输局 行政处罚信息	法定代表人	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	2940	29.400%
出租车企业信息	Qyjc	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	8	0.490%
普通货运企业基本信息	Ssqy	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	2	0.020%
空气质量日报	Jkxyzk	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	2159	6.548%
空气质量日报	Ysjb	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	2159	6.548%
食品经营许可证	Jyxm	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	32	0.320%
空气质量日报	Sywrw	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	11098	33.661%
空气质量日报	Zslb	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	2159	6.548%
国内航运企业名录	Xkzh	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	339	84.750%
市交通运输局 行政许可信息	许可证书名称	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	1650	16.500%
隶属企业信息	Subatt	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	1295	1.295%
道路包车 机动车客运经营许可证证明	Cllx	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	136	5.802%
空气质量日报	Zsjb	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	2159	6.548%
普通货运企业基本信息	Gsdz	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
普通货运企业基本信息	Sffgs	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
港口经营企业名录	Sjc	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	31	100.000%
药品经营许可证	仓库地址	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
市交通运输局 行政处罚信息	税务登记号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
市交通运输局 行	行政相对人	字段空值率	完整性	数据元素完整	存在"-"	10000	100.000%

政许可信息	代码 2 工商注册号	(单表)		性			
市交通运输局 行政许可信息	行政相对人代码 3 组织机构代码	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
市交通运输局 行政许可信息	行政相对人代码 4 税务登记号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
市交通运输局 行政许可信息	行政相对人代码 5 事业单位证书号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%
市交通运输局 行政许可信息	行政相对人代码 6 社会组织登记证号	字段空值率 (单表)	完整性	数据元素完整性	存在"-"	10000	100.000%

表 3-3-2: 完整性测试结果

### 3.3.3 一致性测试结果

数据一致性检查，发现以下问题：

在测试的 18 张表中通过通过单表数据一致性测试的表有 17 个，占总表数的 94.4%；发现单表中一致性问题 1 个，涉及 1 张表。单表中关联数据一致性，存在逻辑关联的字段，前后不一致，需要排查不一致的原因。

在数据集中也发现跨表一致性问题，涉及 5 张表 5 个字段，为相同数据一致性方面，相同的数据在不同的表中出现的内容不一致，存在问题，没有统一的标准。

样本数据中存在单表一致性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
出租车企业信息	QYMC-Z CDZ	单表不同字段逻辑冲突	一致性	关联数据一致性	相同的企业名称，注册地址不同，可能存在一定问题。	44	2.694%

表 3-3-3: 单表一致性测试结果

样本数据中存在跨表一致性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述
港口经营企业名录	QYMC(企业名称)	数据规范不一致-内容(跨表)	一致性	相同数据一致性	企业名称在不同的表中，格式不一致，有些出现数字与编码，未应用同一标准，缺乏相应的企业名单主数据。
经营异常名录		数据规范不一致-内容(跨表)	一致性	相同数据一致性	企业名称在不同的表中，格式不一致，有些出现数字与编码，未应用同一标准，缺乏相应的企业名单主数据。
重点排污单位废水监测报告		数据规范不一致-内容(跨表)	一致性	相同数据一致性	企业名称在不同的表中，格式不一致，有些出现数字与编码，未应用同一标准，缺乏相应的企业名单主数据。
国内航运企业名录		数据规范不一致-内容(跨表)	一致性	相同数据一致性	企业名称在不同的表中，格式不一致，有些出现数字与编码，未应用同一标准，缺乏相应的企业名单主数据。
出租车企业信息		数据规范不一致-内容(跨表)	一致性	相同数据一致性	企业名称在不同的表中，格式不一致，有些出现数字与编码，未应用同一标准，缺乏相应的企业名单主数据。

表 3-3-4: 跨表一致性测试结果

### 3.3.4 准确性测试结果

数据准确性检查，发现以下问题：

在测试的 18 张表中通过准确性测试的表有 9 个，占总表数的 50%。存在数据准确性问题 17 个，涉及 9 张表。样本数据中存在准确性问题有数据格式合规性和数据内容正确性，数据格式合规性为多个字段存在编码的

格式不一致的问题，数据内容正确性多为出现不符合该字段的值出现，数据重复率则是多张表存在复制行。

样本数据中存在准确性问题的字段如下：

表	问题字段 /说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
市交通运输局-行政许可信息	有效期至	数据规范不一致-格式 (单表)	准确性	数据内容合规性	时间的格式不一致，83.5%精确到日，16.5%精确到秒。	1649	16.490%
市交通运输局-行政许可信息	许可编号	数据规范不一致-格式 (单表)	准确性	数据内容合规性	许可编号格式不一致，存在纯数字，也存在汉字加数字的组合。		
深圳公交线路轨迹表	lon	数据规范不一致-格式 (单表)	准确性	数据内容合规性	经度的精确位数不一致，小数点后精度有7、8、9、13、14位的	8463	14.753%
药品经营许可证	有效期至	字段值异常(最大值及内容) (单表)	准确性	数据内容正确性	有效期至存在过大值(9999-12-31)，与第二大值(2028-11-22)相差较大，可能存在问题	703	7.030%
经营异常名录	注册号	数据规范不一致-格式 (单表)	准确性	数据内容合规性	注册号的格式不一致，样本数据中98.5%的为15位数字，其他格式可能存在一定问题。	149	1.490%
空气质量日报	AQI	字段值异常(数值正负值) (单表)	准确性	数据内容正确性	空气质量污染指数(AQI)出现负数-999，可能存在一定问题。	3	0.009%
重点排污单位废水监测报告	日期	数据规范不一致-格式 (单表)	准确性	数据内容合规性	日期格式不一致，99.693%为NNNN-NN-NN型格式。	5	0.307%
空气质量日报	YSJB	字段值异常(最大值及内容) (单表)	准确性	数据内容正确性	空气指数颜色(YSJB)中有些值并不是颜色相关值。		
重点排污单位废水监测报告	LSY	数据规范不一致-格式 (单表)	准确性	数据内容合规性	磷酸盐(LSY)字段中出现的值有带单位的，但单位还不同；还有不带单位的纯数字；还有带大于小于号的，所记录的内容含义不明。		
重点排污单位废水监测报告	色度	数据规范不一致-格式 (单表)	准确性	数据内容合规性	色度的标准不一致，存在多种格式。		
重点排污单位废水监测报告	FHW	数据规范不一致-格式 (单表)	准确性	数据内容合规性	氟化物(FHW)字段中的值，是否带单位，单位是否统一，		

深圳市辐射安全许可证公示名单	XH	数据规范不一致-格式（单表）	准确性	数据内容合规性	存在问题。 序号中格式不一致，部分值并不是序号值，采用了辐射安全许可证号进行填充。	168	22.193%
市交通运输局-行政处罚信息	复制行	单表中重复记录（2个（含2条）以上完全一样的记录）	准确性	数据重复率	该表存在复制行	14	0.140%
市交通运输局-行政许可信息	复制行	单表中重复记录（2个（含2条）以上完全一样的记录）	准确性	数据重复率	存在复制行	6	0.060%
出租车企业信息	复制行	单表中重复记录（2个（含2条）以上完全一样的记录）	准确性	数据重复率	该表存在大量复制行	1626	99.571%
空气质量日报	复制行	单表中重复记录（2个（含2条）以上完全一样的记录）	准确性	数据重复率	存在复制行	90	0.273%
市交通运输局-行政许可信息	编号	键值（字段）不唯一（单表）	准确性	数据唯一性	编号值出现重复，密钥不唯一。	80	0.800%

表 3-3-5: 准确性测试结果

### 3.3.5 规范性测试结果

数据规范性检查，发现以下问题：

在测试的 18 张表中通过规范性测试的表有 15 个，占总表数的 83%。

样本数据中存在数据规范性问题有 3 个，涉及 3 张表。多为数据标准不一致，单字段未有统一标准。

样本数据中存在规范性问题的表和字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)

市交通运输局 -行政处罚信息	处罚类别	数据规范不一致- 内容（单表）	规范性	数据标准	处罚类别标准不一致， 存在一定问题。	/	/
饮用水质月度情况	sz1b	数据规范不一致- 内容（单表）	规范性	数据标准	水质类别出现同意不同词现象，标准并未统一。	2	0.338%
重点排污单位 废水监测报告	dq	数据规范不一致- 内容（单表）	规范性	数据标准	地区字段出现同意不同词现象，标准未统一。	1	0.061%

表 3-3-6: 规范性测试结果

### 3.3.6 时效性测试结果

本次测试数据不适用时效性测试。

### 3.3.7 可访问性测试结果

本次测试数据不适用可访问性测试。

### 3.3.8 质量维度统计结果

#### 含义解释

- 1) 通过测试的字段：如 A 表（单表）age 字段，存在完整性问题，认为 age 字段未通过完整性测试；反之则说明，age 字段通过了完整性测试。
  - 2) 通过测试的总字段个数：指对整个数据集(18 张表)中各表通过相应维度测试的字段个数之和。
  - 3) 通过测试的表字段个数：指对单个表中通过测试的字段数的统计。
- 以上三个含义解释都是针对样本数据，在本次测试中的结果的统计。

### 3.3.8.1 通过测试的总字段个数

总字段	完整性	一致性	准确性	规范性
198	143	197	181	195

表 3-3-7: 通过测试的总字段个数

### 3.3.8.2 通过测试的表字段个数

表名	总字段数	完整性	一致性	准确性	规范性
深圳_交通运输局_公交线路信息	13	11	13	13	13
深圳_交通运输局_出租车企业信息	5	2	4	4	5
深圳_交通运输局_国内航运企业名录	5	2	5	5	5
深圳_交通运输局_市交通运输局 行政处罚信息	27	21	27	26	26
深圳_交通运输局_市交通运输局 行政许可信息	21	14	21	17	21
深圳_交通运输局_普通货运企业基本信息	6	3	6	6	6
深圳_交通运输局_深圳公交线路轨迹表	7	7	7	6	7
深圳_交通运输局_港口经营企业名录	5	4	5	5	5
深圳_交通运输局_道路包车 机动车 客运经营许可证明	10	9	10	10	10
深圳_交通运输局_道路客运企业版块行政检查汇总公示明细	9	6	9	9	9
深圳_市场监管局_经营异常名录	8	4	8	7	8
深圳_市场监管局_药品经营许可证	15	14	15	14	15
深圳_市场监管局_隶属企业信息	7	2	7	7	7
深圳_市场监管局_食品经营许可证	16	15	16	16	16
深圳_生态环境局_深圳市辐射安全许可证公示名单	3	2	3	2	3
深圳_生态环境局_空气质量日报	9	2	9	6	9
深圳_生态环境局_重点排污单位废水监测报告	24	20	24	20	23
深圳_生态环境局_饮用水水质月度情况	8	5	8	8	7
总计	198	143	197	181	195

表 3-3-8: 通过测试的表字段个数

### 3.3.9 数据质量测试各局委办得分

- 深圳市交通管理局整体数据质量维度评分（10 张表、108 个字段、931198 条记录）

深圳市交通管理局数据质量评分：95.52 分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	84.82%	99.95%	97.32%	99.98%
质量维度权重	25.00%	25.00%	25.00%	25.00%

2. 深圳市市场监督管理局整体数据质量维度评分(4张表、46个字段、130000条记录)

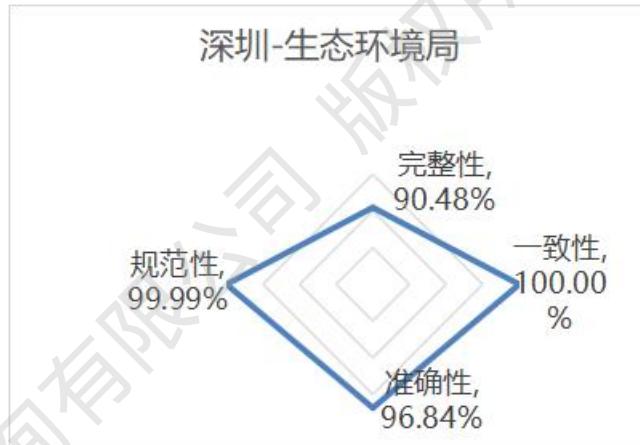
深圳市市场监督管理局数据质量评分：98.41分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	93.80%	100.00%	99.84%	100.00%
质量维度权重	25.00%	25.00%	25.00%	25.00%

3. 深圳市生态环境局整体数据质量维度评分（4张表、44个字段、35947条记录）

深圳市生态环境局数据质量评分：96.83分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	90.48%	100.00%	96.84%	99.99%
质量维度权重	25.00%	25.00%	25.00%	25.00%

### 3.4 杭州市数据质量测试结果

#### 3.4.1 基础元数据测试结果-表基本属性

名称	字段数	记录数	数值	行最小长度	行最大长度
杭州_交通管理局_省回流 货运企业信息	9	14235	52176	82	185
杭州_交通运输局_杭州公路客运周转量信息	9	29	159	1	67
杭州_交通运输局_杭州公路客运量信息	9	29	163	1	59
杭州_交通运输局_杭州公路货运周转量信息	9	29	159	1	69
杭州_交通运输局_杭州公路货运量信息	9	29	159	1	61
杭州_交通运输局_杭州水路客运量信息	9	29	159	1	56
杭州_交通运输局_杭州货运车辆基本信息	10	119000	8617	1	196
杭州_交通运输局_杭州高速及国省道路基本信息	7	38	218	1	61
杭州_交通运输局_省回流 网络预约出租汽车车辆信息	8	199481	215983	1	104
杭州_交通运输局_省回流 道路运输经营许可证	30	6138	69667	853	1620
杭州_市场监管局_企业经营异常名录信息	9	150000	386988	1	237

杭州_市场监管局_在册市场主体基本信息	13	100000	479933	110	995
杭州_市场监管局_杭州食品经营许可证	10	150000	738932	129	295
杭州_市场监管局_药品经营许可证信息	20	5868	38827	158	966
杭州_生态环境局_杭州市放射源信息	14	1618	4099	1	120
杭州_生态环境局_水质自动站小时采样数据信息	21	150000	152389	131	201
杭州_生态环境局_污染源基础信息	12	39510	249446	67	260
杭州_生态环境局_重点排污单位公开目录信息	6	11469	32904	25	100

表 3-4-1: 表基本属性

### 3.4.2 完整性测试结果

数据完整性检查，发现以下问题：

字段记录中存在空值、空格值、NULL 值或者是“-”的现象。在测试的 18 张表中通过数据完整性测试的表有 3 个，占总表数的 16.67%。发现完整性问题 58 个，涉及 15 张表，完整性问题主要为数据元素完整性。

样本数据中存在完整性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
重点排污单位公开目录信息	上报时间	字段空值率(单表)	完整性	数据元素完整性	存在空值	11350	98.962%
药品经营许可证信息	仓库地址	字段空值率(单表)	完整性	数据元素完整性	存在空值	4412	75.187%
重点排污单位公开目录信息	企业类型	字段空值率(单表)	完整性	数据元素完整性	存在空值	1749	15.250%
在册市场主体基本信息	住所	字段空值率(单表)	完整性	数据元素完整性	存在空值	33	0.033%
重点排污单位公开目录信息	单位地址	字段空值率(单表)	完整性	数据元素完整性	存在空值	7	0.061%
药品经营许可证信息	发证机关	字段空值率(单表)	完整性	数据元素完整性	存在空值	1138	19.393%
水质自动站小时采样数据信息	固定评价标记	字段空值率(单表)	完整性	数据元素完整性	存在空值	80210	53.473%
水质自动站小时采样数据信息	所属流域	字段空值率(单表)	完整性	数据元素完整性	存在空值	191	0.127%
药品经营许可证信息	日常监督管理	字段空值率	完整性	数据元素完	存在空值	720	12.270%

	机构	(单表)		完整性			
水质自动站小时采样数据信息	是否超标	字段空值率(单表)	完整性	数据元素完整性	存在空值	80313	53.542%
杭州食品经营许可证	有效期至	字段空值率(单表)	完整性	数据元素完整性	存在空值	2	0.001%
水质自动站小时采样数据信息	水质状况	字段空值率(单表)	完整性	数据元素完整性	存在空值	80570	53.713%
水质自动站小时采样数据信息	水质类别	字段空值率(单表)	完整性	数据元素完整性	存在空值	80570	53.713%
水质自动站小时采样数据信息	水质类别标识	字段空值率(单表)	完整性	数据元素完整性	存在空值	80313	53.542%
污染源基础信息	污染源地址	字段空值率(单表)	完整性	数据元素完整性	存在空值	76	0.192%
污染源基础信息	污染源编号	字段空值率(单表)	完整性	数据元素完整性	存在空值	843	2.134%
污染源基础信息	法人代表姓名	字段空值率(单表)	完整性	数据元素完整性	存在空值	101	0.256%
药品经营许可证信息	法定代表人	字段空值率(单表)	完整性	数据元素完整性	存在空值	119	2.028%
在册市场主体基本信息	法定代表人 经营者姓名	字段空值率(单表)	完整性	数据元素完整性	存在空值	220	0.220%
省回流 货运企业信息	注册地址	字段空值率(单表)	完整性	数据元素完整性	存在空值	18	0.126%
污染源基础信息	环保联系人	字段空值率(单表)	完整性	数据元素完整性	存在空值	20017	50.663%
水质自动站小时采样数据信息	监测对象名称	字段空值率(单表)	完整性	数据元素完整性	存在空值	150000	100.000%
水质自动站小时采样数据信息	监测对象编号	字段空值率(单表)	完整性	数据元素完整性	存在空值	150000	100.000%
污染源基础信息	纬度	字段空值率(单表)	完整性	数据元素完整性	存在空值	67	0.170%
污染源基础信息	经度	字段空值率(单表)	完整性	数据元素完整性	存在空值	67	0.170%
杭州食品经营许可证	经营场所	字段空值率(单表)	完整性	数据元素完整性	存在空值	547	0.365%
在册市场主体基本信息	经营范围	字段空值率(单表)	完整性	数据元素完整性	存在空值和空格值	224	0.224%
省回流 货运企业信息	经营范围	字段空值率(单表)	完整性	数据元素完整性	存在空值	2266	15.919%
杭州食品经营许可证	经营项目	字段空值率(单表)	完整性	数据元素完整性	存在空值	168	0.112%
药品经营许可证信息	统一社会信用代码	字段空值率(单表)	完整性	数据元素完整性	存在空值	1261	21.489%
污染源基础信息	联系电话	字段空值率	完整性	数据元素完	存在空值	20010	50.645%

		(单表)		完整性			
在册市场主体基本信息	营业期限止	字段空值率 (单表)	完整性	数据元素完整性	存在空值	24435	24.435%
在册市场主体基本信息	营业期限起	字段空值率 (单表)	完整性	数据元素完整性	存在空值	3019	3.019%
药品经营许可证信息	许可证有效日期起	字段空值率 (单表)	完整性	数据元素完整性	存在空值	3	0.051%
药品经营许可证信息	质量负责人	字段空值率 (单表)	完整性	数据元素完整性	存在空值	1268	21.609%
水质自动站小时采样数据信息	超标污染物及超标倍数	字段空值率 (单表)	完整性	数据元素完整性	存在空值	141047	94.031%
杭州货运车辆基本信息	经营范围名称	字段空值率 (单表)	完整性	数据元素完整性	存在空格值	7	0.006%
企业经营异常名录信息	作出决定机关移出	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	60722	40.481%
杭州市放射源信息	放射源照片	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值和空值	1618	100.000%
企业经营异常名录信息	移出经营异常名录原因	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	60722	40.481%
杭州市放射源信息	纬度	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值和空值	1618	100.000%
杭州市放射源信息	经度	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值和空值	1618	100.000%
杭州水路客运量信息	月度同比	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州公路货运量信息	月度同比	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州公路货运周转量信息	月度同比	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州公路客运周转量信息	月度同比	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	3	10.345%
杭州公路客运量信息	月度同比	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州水路客运量信息	本月水路客运量	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州公路客运量信息	本月公路客运量	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州公路货运量信息	本月公路货运量	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州公路客运周转量信息	本月公路客运周转量	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	1	3.448%
杭州公路货运周转量信息	本月公路货运周转量	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	2	6.897%
杭州市放射源信息	放射源类别	字段空值率	完整性	数据元素完	存在 null	5	0.309%

		(单表)		完整性	值		
杭州市放射源信息	状态	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	3	0.185%
杭州市放射源信息	国家地区	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	422	26.082%
杭州市放射源信息	出厂日期	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	456	28.183%
企业经营异常名录信息	移出日期	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	60722	40.481%
企业经营异常名录信息	统一社会信用代码	字段空值率 (单表)	完整性	数据元素完整性	存在 null 值	7803	5.202%

表 3-4-2: 完整性测试结果

### 3.4.3 一致性测试结果

数据一致性检查，发现以下问题：

在测试的 18 张表中通过通过单表数据一致性测试的表有 14 个，占总表数的 77.78%；发现单表中一致性问题 4 个，涉及 4 张表。单表中关联数据一致性，存在逻辑关联的字段，前后不一致，需要排查不一致的原因。

在数据集中也发现跨表一致性问题，涉及 4 张表 4 个字段，为相同数据一致性方面，相同的数据在不同的表中出现的内容不一致，存在问题，没有统一的标准。

样本数据中存在单表一致性问题的字段如下：

问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
污染源编号-污染源名称	单表不同字段逻辑冲突	一致性	关联数据一致性	相同污染源编号，污染源名称不同，存在一对多现象。	1040	2.632%
测点编号-测点名称	单表不同字段逻辑冲突	一致性	关联数据一致性	相同的测点编号，测点名称不同。	1376	0.917%
有效开始时间-有效截止时间	单表不同字段逻辑冲突	一致性	关联数据一致性	存在 5 条记录开始时间大于截止时间，可能存在一定问题。	5	0.003%
营业期限起-营业期限止	单表不同字段逻辑冲突	一致性	关联数据一致性	存在 5 条记录营业期限起大于营业期限止	5	0.005%

表 3-4-3: 单表一致性测试结果

样本数据中存在跨表一致性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述
在册市场主体基本信息	企业名称	数据规范不一致-内容（跨表）	一致性	相同数据一致性	不同表见的企业名称出现不一致，有出现编码的情况。缺乏相应的企业名单主数据。
药品经营许可证信息		数据规范不一致-内容（跨表）	一致性	相同数据一致性	不同表见的企业名称出现不一致，有出现编码的情况。缺乏相应的企业名单主数据。
杭州食品经营许可证		数据规范不一致-内容（跨表）	一致性	相同数据一致性	不同表见的企业名称出现不一致，有出现编码的情况。缺乏相应的企业名单主数据。
企业经营异常名录		数据规范不一致-内容（跨表）	一致性	相同数据一致性	不同表见的企业名称出现不一致，有出现编码的情况。缺乏相应的企业名单主数据。

表 3-4-4: 跨表一致性测试结果

### 3.4.4 准确性测试结果

数据准确性检查，发现以下问题：

在测试的 18 张表中通过准确性测试的表有 10 个，占总表数的 56%。

存在数据准确性问题 17 个，涉及 8 张表。样本数据中存在准确性问题有数据格式合规性和数据内容正确性，数据格式合规性为多个字段存在编码的格式不一致的问题，数据内容正确性多为出现不符合该字段的值出现，数据重复率则是多张表存在复制行。

样本数据中存在准确性问题的字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
省回流 网络预约出租汽车车辆信息	车牌颜色	字段值异常（最值及内容）（单表）	准确性	数据内容准确性	车牌颜色字段中出现值并不是颜色相关的值。	199481	100.000%

杭州货运车辆基本信息	车牌颜色	字段值异常(最值及内容)(单表)	准确性	数据内容正确性	车牌颜色字段中出现的并不是颜色，如果是颜色代码，该字段应该命名为车牌颜色代码更为妥当。	119001	100.000%
杭州货运车辆基本信息	车辆类型名称	字段值异常(最值及内容)(单表)	准确性	数据内容正确性	该字段的值并不是车辆类型名称，出现的可能是车辆类型代码，并没有将代码转换为对应的名称。	119001	100.000%
杭州高速及国省道路基本信息	起点名称	字段值异常(最值及内容)(单表)	准确性	数据内容正确性	起点名称中出现值“OK”，可能存在一定问题。	1	2.632%
杭州高速及国省道路基本信息	止点名称	字段值异常(最值及内容)(单表)	准确性	数据内容正确性	止点名称中出现英文字母和数字值，可能存在一定问题。	2	5.263%
省回流 网络预约出租汽车车辆信息	有效开始时间	字段值异常(最值及内容)(单表)	准确性	数据内容正确性	有效开始时间出现不合理值。	2	0.001%
杭州市放射源信息	放射源编码	数据规范不一致-格式(单表)	准确性	数据内容合规性	放射源编码的格式不一致，95%以上为12位字符编码，其他不为12位的，可能存在一定问题。	21	1.298%
杭州市放射源信息	使用许可证编号	数据规范不一致-格式(单表)	准确性	数据内容合规性	许可证编号中存在不是编号的值，使用的是另外的字段(用途)进行填充。	9	0.556%
污染源基础信息	序号	数据规范不一致-格式(单表)	准确性	数据内容合规性	序号的格式不一致，2%的数据为1-4位的数字编码，其余的都是22位的字符编码。	807	2.045%
污染源基础信息	行政区划	数据规范不一致-格式(单表)	准确性	数据内容合规性	行政区划编码格式不一致，约98%为12位行政区划，2%为6位行政区划，精确的颗粒度不一致。	808	2.045%
污染源基础信息	污染源编号	数据规范不一致-格式(单表)	准确性	数据内容合规性	污染源编号格式不一致，有3条记录为18位字符编码，其余都是22位字符编码。	3	2.141%
杭州货运车辆基本信息	复制行	单表中重复记录(2个(含2条))	准确性	数据重复率	存在复制行	2142	1.800%

		以上完全一样的记录)						
污染源基础信息	经度	数据规范不一致-格式(单表)	准确性	数据内容合规性	格式不一致，出现中文汉字。	5	0.013%	
污染源基础信息	纬度	数据规范不一致-格式(单表)	准确性	数据内容合规性	格式不一致，出现中文汉字。	5	0.013%	
杭州食品经营许可证	主键	数据规范不一致-格式(单表)	准确性	数据内容合规性	存在 15、19、21、32 位字符的不同格式。		0.000%	
在册市场主体基本信息	营业期限止	字段值异常(最值及内容)(单表)	准确性	数据内容准确性	存在过小值 "0100-01-01 00:00:00" 和 "1900-01-01 00:00:00"，可能存在一定问题。	2	0.002%	
水质自动站小时采样数据信息	测点代码	数据规范不一致-格式(单表)	准确性	数据内容合规性	测点代码的格式不一致。		0.000%	

表 3-4-5: 准确性测试结果

### 3.4.5 规范性测试结果

数据规范性检查，发现以下问题：

在测试的 18 张表中通过规范性测试的表有 16 个，占总表数的 89%。

样本数据中存在数据规范性问题有 2 个，涉及 2 张表。多为数据标准不一致，单字段未有统一标准。

样本数据中存在规范性问题的表和字段如下：

表	问题字段/说明	问题类别	质量维度	指标名称	问题描述	问题记录数	占比(问题记录数/总记录数)
省回流 网络预约出租汽车车辆信息	厂牌	数据规范不一致-内容(单表)	规范性	数据标准	厂牌的标准不统一，同时颗粒度也不同。	/	/

药品经营许可证信息	许可证状态	数据规范不一致-内容（单表）	规范性	数据标准	状态中存在不是字段说明中的枚举值。	195	3.323%
-----------	-------	----------------	-----	------	-------------------	-----	--------

表 3-4-6: 规范性测试结果

### 3.4.6 时效性测试结果

本次测试数据不适用时效性测试。

### 3.4.7 可访问性测试结果

本次测试数据不适用可访问性测试。

### 3.4.8 质量维度统计结果

#### 含义解释

- 1) 通过测试的字段：如 A 表（单表）age 字段，存在完整性问题，认为 age 字段未通过完整性测试；反之则说明，age 字段通过了完整性测试。
- 2) 通过测试的总字段个数：指对整个数据集（18 张表）中各表通过相应维度测试的字段个数之和。
- 3) 通过测试的表字段个数：指对单个表中通过测试的字段数的统计。

以上三个含义解释都是针对样本数据，在本次测试中的结果的统计。

#### 3.4.8.1 通过测试的总字段个数

总字段	完整性	一致性	准确性	规范性
214	156	210	197	212

表 3-4-7: 通过测试的总字段个数

### 3. 4. 8. 2 通过测试的表字段个数

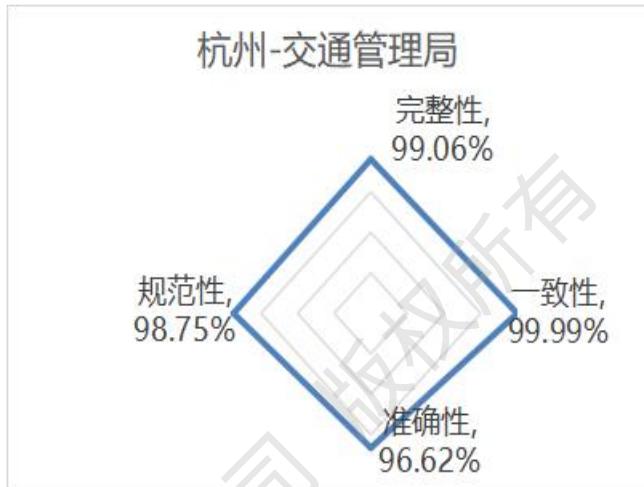
表名	总字段数	完整性	一致性	准确性	规范性
杭州_交通运输局_杭州公路客运周转量信息	9	7	9	9	9
杭州_交通运输局_杭州公路客运量信息	9	7	9	9	9
杭州_交通运输局_杭州公路货运周转量信息	9	7	9	9	9
杭州_交通运输局_杭州公路货运量信息	9	7	9	9	9
杭州_交通运输局_杭州水路客运量信息	9	7	9	9	9
杭州_交通运输局_杭州货运车辆基本信息	10	9	10	7	10
杭州_交通运输局_杭州高速及国省道路基本信息	7	7	7	5	7
杭州_交通运输局_省回流 网络预约出租汽车车辆信息	8	8	7	6	7
杭州_交通运输局_省回流 货运企业信息	9	7	9	9	9
杭州_交通运输局_省回流 道路运输经营许可证	30	30	30	30	30
杭州_市场监管局_企业经营异常名录信息	9	5	9	9	9
杭州_市场监管局_在册市场主体基本信息	13	8	12	12	13
杭州_市场监管局_杭州食品经营许可证	10	7	10	9	10
杭州_市场监管局_药品经营许可证信息	20	13	20	20	19
杭州_生态环境局_杭州市放射源信息	14	7	14	12	14
杭州_生态环境局_水质自动站小时采样数据信息	21	12	20	20	21
杭州_生态环境局_污染源基础信息	12	5	11	7	12
杭州_生态环境局_重点排污单位公开目录信息	6	3	6	6	6
总计	214	156	210	197	212

表 3-4-8: 通过测试的表字段个数

### 3. 4. 9 数据质量测试各局委办得分

1. 杭州市交通管理局整体数据质量维度评分（10 张表，109 个字段，339038 条记录）

杭州市交通管理局数据质量评分：98.60 分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	99.06%	99.99%	96.62%	98.75%
质量维度权重	25.00%	25.00%	25.00%	25.00%

2. 杭州市市场监督管理局整体数据质量维度评分(4张表, 52个字段, 405868条记录)

杭州市市场监督管理局数据质量评分：98.50分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	94.03%	100.00%	100.00%	99.96%
质量维度权重	25.00%	25.00%	25.00%	25.00%

### 3. 杭州市生态环境局整体数据质量维度评分（4 张表，53 个字段，202597 条记录）

杭州市生态环境局数据质量评分：94.44 分

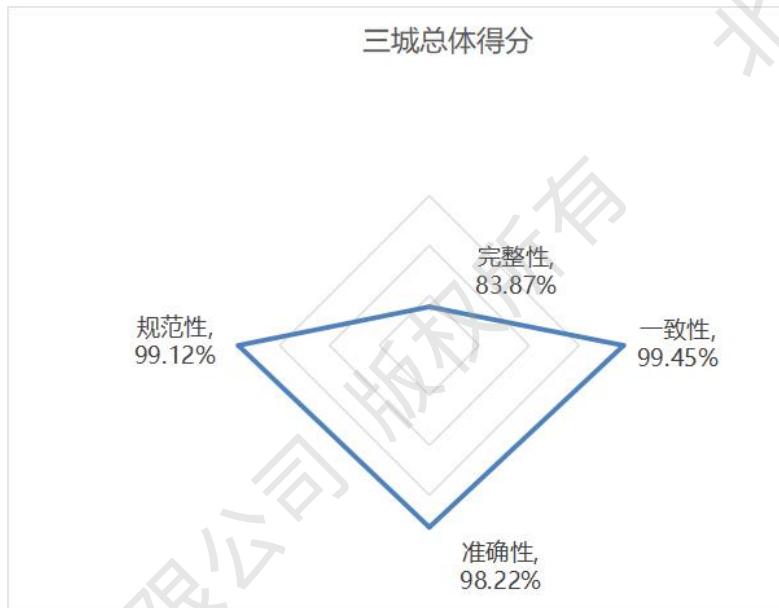


质量维度	完整性	一致性	准确性	规范性
质量维度评分	80.04%	97.89%	99.84%	100.00%
质量维度权重	25.00%	25.00%	25.00%	25.00%

## 3.5 总体数据质量测试结果

在广州公共数据开放平台、深圳市政府数据开放平台、杭州市数据开放平台抽取样本数据，分别对交通管理局、市场监管局、生态环境局抽取样本数据共计 54 张表，612 个字段，1318502 条记录进行数据质量测试。测试结果如下：

三个城市的 data 质量整体评分：95.16 分



质量维度	完整性	一致性	准确性	规范性
质量维度评分	83.87%	99.45%	98.22%	99.12%
质量维度权重	25.00%	25.00%	25.00%	25.00%

## 四、评测问题示例

示例一：

### 公共数据质量问题剖析

三城总体得分

完整性	一致性	准确性	规范性
99.12%	98.22%	99.45%	99.87%

名称：深圳-市场-企业信息\_完整性-机构代码02  
描述：机构代码为空或用“~”表述  
类型：已过滤的数据行 (深圳-市场监管-隶属企业信息 2920001303581) 19628 行

名称：深圳-市场监督-食品生产许可\_完整性  
描述：多个字段的空值率>50%  
类型：各实体成员属性 16

名称：深圳-环保-辐射-数据完整性-企业名称  
描述：关键字段企业名称完整性问题  
类型：数值 709

名称：深圳-环保-饮用水-数据完整性  
描述：关键字段完整性问题-检测时间(jys)  
类型：数值 51

**问题源自：**  
数据采集- 数据迁移- 数据加工

华矩科技

示例二：

### 公共数据质量问题剖析

三城总体得分

完整性	一致性	准确性	规范性
99.12%	98.22%	99.45%	99.87%

**问题源自：**

- 系统设计
- 跨部门协调
- 数据治理

名称 = 广州跨局委办数据分析 (三个局委办) 书签由 tssadmin  
名称 实体名称  
所在辖区 广州-交通-客运企业情况  
所属行政区域 广州-环保-重点排污单位名录  
行政区划 广州-市场监管-食品生产企业信息

名称 = 广州跨局委办数据分析 (三个局委办) 书签由 tssadmin  
名称 实体名称  
社会信用代码 广州-市场监督-食品生产企业信息  
统一社会信用代码 广州-市场监督-企业基本信息  
统一社会信用代码 组织机构代码 广州-环保-重点排污单位名录

名称 = 杭州不同局委办数据质量分析 书签由 tssadmin 在 2024/05/28 16:18:53 上创造。  
名称 实体名称 数值 最小值 最大值 最小长度 最大长度  
统一社会信用代码 杭州-环保-污染源基础信息 20240430 38470 119.256590 上海涛洁汽车服务有限责任公司杭州分公司 2 19  
统一社会信用代码 杭州-市场监管-在册市场主体基本信息 20240526ht 100000 91330100E94374215A 93330185698069638 18 18

名称 = 深圳跨局不同部门数据质量分析 书签由 tssadmin 在 2024/06/04 16:38:05 上创造。  
名称 实体名称 数值 模式 最小值 最大值 最小长度 最大长度  
检查时间 深圳-交通-道路客运企业按块行政检查汇总公示明细 2920000403604 511 2 2020-01-09 - 1 10  
Fzrq 深圳-市场监督-食品生产许可证 2920001300349 464 6 01-5月-21 12:00:00.0000000000 上午 MB2C92739 9 65881  
UFzrq 深圳-市场监督-食品生产许可证 2920001300349 463 5 01-5月-21 12:00:00.0000000000 上午 2039 4 31  
Rq 深圳-环保-重点排污单位废水监测报告 2920001003581 306 5 2002-10-08 44068.2020-07-1577778 9 21

华矩科技

### 示例三：

**公共数据质量问题剖析**

**问题源自：**

- 数据迁移
- 多头管理
- 数据集成

企业名称	统一社会信用代码
杭州丝绸服饰中心	91330103MA2GDHEL3A
杭州丝绸服饰中心	91330103MA2GDME21E
杭州尼丹卡贸易有限公司	91330103MA2GDJUQ19
杭州尼丹卡贸易有限公司	91330100MA2GEU0Y1K
杭州中润贸易有限公司	91330103MA2GDKD46U
杭州中润贸易有限公司	913301026970699122
杭州伊麦尔食品有限公司	91330103MA2GDLFJ4N
杭州伊麦尔食品有限公司	91330103MA2GETNG8U
杭州国兴贸易有限公司	91330103MA2GDL0K7U
杭州国兴贸易有限公司	91330106682904559J
杭州森峰娱乐厅	91330103MA2GDMKG4J
杭州森峰娱乐厅	91330103MA2GER4A2M
浙江省富力工贸实业总公司	91330103MA2GDL2W8K
浙江省富力工贸实业总公司	91330100MA2GEXTA8N
杭州庆宏物资有限公司	91330103MA2GDJ9H45
杭州庆宏物资有限公司	91330105MA2GETD62Q

**九宫格数据® HUGONGGE-DATA**

**问题源自：**

一些关键字段的处理不当（或缺失），如这里的更新时间，导致这批数据不可理解，表现出来就是一致性问题。完整性导致一致性问题。

**华矩科技**

### 示例四：

**公共数据质量问题剖析**

**问题源自：**

- 局部标准
- 组织协调

名称	描述	类型	行
广州-同局-不同部门数据-字段名称不一致	企业地址字段名称不一致	基线分析项目属性	3
广州-同局-不同部门数据-数据不一致	辖区的字段名不一致	基线分析项目属性	3
广州-同局-不同部门数据-数据不一致-辖区	同局三个表的所在辖区的内容表达不规范	基线分析项目属性	3
广州-跨局委办-数据规范-字段名不一致	统一社会信用代码字段名不一致	基线分析项目属性	3
杭州-同局-不同部门数据-数据不一致	市场主体基础信息与药品经营许可信息中同一企业但不同的法人	数据行	25
杭州-同局-不同部门数据-数据不一致2	市场主体基础信息与药品经营许可信息中统一企业信用代码但企业名称不一样	数据行	1
杭州-同局-不同部门数据-数据不一致3	市场主体基础信息与药品经营许可信息中同一企业但不同注册地址	数据行	24
杭州-同局-不同部门数据-数据结构不一致	环保两个不同的系统的数据的经纬度数据结构不一致	基线分析项目属性	4
杭州-跨局委办-数据规范-企业名称不一致	三个局委办关于企业主体名称的字段名都不一样	基线分析项目属性	3

名称	参考	实体名称	数值	模式	最小值	最大值	最小长度	最大长度
所在辖区	4	广州-交通-危运企业情况	27	3	null	黄埔区	2	4
所在辖区	4	广州-交通-客运企业情况	61	4	null	黄埔	2	5
所属辖区	5	广州-交通-机动车维修企业资料及分布查询	12	3	null	黄埔区(开发区)	3	8

**九宫格数据® HUGONGGE-DATA**

**华矩科技**



需要获得更多信息，请关注华矩科技公众号

咨询热线：400 038 1778

邮箱地址：[contact@hjbis.com](mailto:contact@hjbis.com)

更多内容可以通过 [www.biz-insight.com/](http://www.biz-insight.com/) 点击“案例及下载”获取