

轻松搞定 电力设备大数据分析

—— Tableau Desktop 在电力系统中的应用

南方电网科学研究院有限责任公司 组编

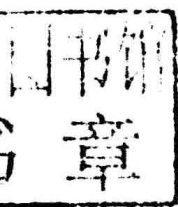


中国电力出版社
CHINA ELECTRIC POWER PRESS

轻松搞定 电力设备大数据分析

—— Tableau Desktop 在电力系统中的应用

南方电网科学研究院有限责任公司 组编



中国电力出版社
CHINA ELECTRIC POWER PRESS

内 容 提 要

本书共分为 5 章，分别包括 Tableau Desktop 软件简介、数据连接与融合、常用图表制作、图表常用功能和案例介绍。本书着重介绍 Tableau Desktop 软件在电力企业资产管理中的应用，书中所用到的样本数据均进行了处理。本书可操作性强，对 Tableau Desktop 软件中的常用图表、常用功能的使用步骤均进行了详细介绍。

本书主要基于电力设备数据进行编写，可供电力设备资产管理人员阅读，同时本书实用性较强，也可供有数据可视化分析需求的人员使用。

图书在版编目（CIP）数据

轻松搞定电力设备大数据分析：Tableau Desktop 在电力系统中的应用 / 南方电网科学研究院有限责任公司组编. —北京：中国电力出版社，2017.11

ISBN 978-7-5198-1182-2

I. ①轻… II. ①南… III. ①电力设备—应用软件 IV. ①TM4-39

中国版本图书馆 CIP 数据核字（2017）第 237930 号

出版发行：中国电力出版社

地 址：北京市东城区北京站西街 19 号（邮政编码 100005）

网 址：<http://www.cepp.sgcc.com.cn>

责任编辑：马 青（010-63412784） 贾丹丹

责任校对：王开云

装帧设计：张俊霞 左 铭

责任印制：邹树群

印 刷：北京大学印刷厂

版 次：2017 年 11 月第一版

印 次：2017 年 11 月北京第一次印刷

开 本：787 毫米×1092 毫米 16 开本

印 张：11

字 数：220 千字

印 数：0001—1500 册

定 价：40.00 元

版 权 专 有 侵 权 必 究

本书如有印装质量问题，我社发行部负责退换

本书编写人员

主 编 袁 耀

副主编 李锐海 李志强 喇 元 李战鹰

参 编 王 颂 陈 曦 赵林杰 彭在兴

程建伟 张 曦 邹 林 易 林

刘 芹 杨家辉 金 虎 刘 旭

王帅兵

前言

Tableau Desktop 是一款商业智能数据可视化分析软件，在全球拥有 10000 多家企业或组织客户，遍及各个行业的各类规模企业。使用 Tableau Desktop，可方便地连接到各类数据源，用户只需采用拖放的方式就可快速地创建出交互、美观、智能的视图和仪表盘，方便快捷地开展数据可视化分析。

中国南方电网有限责任公司（简称南方电网公司）资产管理部门从 2015 年开始使用 Tableau Desktop 软件，并在日常工作报表、资产数据分析、资产管理辅助决策等工作中广泛地使用该软件开展相关工作。与传统数据分析软件相比，Tableau Desktop 极大地提高了数据分析效率，并为资产管理辅助决策提供实时技术支持。

为了进一步推广 Tableau Desktop 软件在电力企业中的应用，我们组织编写了《轻松搞定电力设备大数据分析——Tableau Desktop 在电力系统中的应用》，总结了近年来使用 Tableau Desktop 软件的一些经验和技巧，并分享了相关应用案例。

本书包括 5 章，主要介绍了 Tableau Desktop 软件的数据连接与融合、常用图表制作、图表常用功能以及案例分享。本书着重介绍 Tableau Desktop 软件在电力企业资产管理中的应用，书中所用到的样本数据均进行了处理。

本书由袁耀主编，李锐海、李志强、喇元、李战鹰担任副主编，王颂、陈曦、赵林杰、彭在兴、程建伟、张曦、邹林、易林、刘芹、杨家辉、金虎、刘旭、王帅兵等也参与了本书的编写工作。

本书所提供的各章节实例素材文件都是基于 Tableau Desktop 10.3 版本制作的，读者可以通过本书提供的百度网盘二维码下载样本数据和案例，提取密码为 2zq8。

虽然作者在本书的编写中力求叙述准确、完善，但由于水平有限，书中难免有欠妥之处，希望读者和同仁能及时指正。

编者

2017 年 10 月



目 录

前言

1	Tableau Desktop 软件简介	1
2	数据连接与融合	6
2.1	连接 Excel 文件	6
2.2	数据源操作	8
3	常用图表制作	14
3.1	柱状图	14
3.2	饼图	15
3.3	曲线图与区域图	15
3.4	树地图	17
3.5	突出显示表	17
3.6	堆叠图	18
3.7	并排图	19
3.8	文本表	19
3.9	填充气泡图	20
3.10	散点图	21
3.11	地图	21
3.12	组合图	23
3.13	瀑布图	24
3.14	拟合曲线图	26

3.15	柱状对比图	29
3.16	背景图片	32
3.17	空心饼图	35
3.18	帕累托图	38
4	图表常用功能	44
4.1	排序	44
4.2	筛选	52
4.3	参数	69
4.4	计算	88
4.5	合并同类字段	116
4.6	仪表板	119
5	案例介绍	127
5.1	变电设备台账缺陷数据分析案例	127
5.2	设备厂商运行质量指标分析案例	142
5.3	设备制造厂家缺陷率分析案例	151
5.4	变电设备厂商运行评价分析案例	161
	参考文献	168

Tableau Desktop 软件简介

近年来，通过数据分析辅助企业管理决策已越来越受到人们的重视，特别是对于电力等传统行业来说，如何从基于经验的管理向数据驱动的管理转变，显得至关重要。在数据驱动管理的理念中，数据分析是非常重要的一环，但从传统的商务智能（business intelligence, BI）软件来看，由于其开发难度较高，对专业能力有较高要求，需要 IT 人员、业务人员互相配合，无法实时响应，因此并不能很好地适用于当前数据驱动管理的需求。正是在基于以上背景的情况下，美国斯坦福大学的数据科学家开发了 Tableau Desktop 软件，使得非 IT、数据库人员可以方便快速地连接访问数据，开展自助式的数据分析，大大提高了数据分析效率。

Tableau Desktop 软件是一款数据可视化的商务智能展现工具，主要面向对象为普通商业用户，而非专业开发人员，因此它简单易学、易用、可视化。下面，就 Tableau Desktop 软件的一些基本功能作简要介绍。

1. 能连接到各类主要数据库

Tableau Desktop 可以直接与任何数据库、数据体、数据仓库、文件以及电子表格相联接，如 Excel、Access、Oracle、Amazon Athena、Microsoft OneDrive 等。图 1-1 展示了 Tableau Desktop 软件目前可连接的主要数据库类型。

目前，分析人员可以在无需编程的情况下，使用 66 种连接器连接超过 75 种的数据源，其中包括 10.3 版本中新增的 PDF 连接器，它可以通过一键式操作，将 PDF 表直接导入 Tableau 中。

2. 操作简便

Tableau Desktop 最大特点仅通过拖放技术便可以快速连接到数据进行查询、分析，易于学习，方便使用，无需编写代码，是一款人人都能使用的自助式分析软件。它分析数据的方法是基于数据可视化，而不是传统的数据查询、数据统计与数据作图。

同时，Tableau Desktop 内嵌了多个快速智能显示图表，如柱状图、饼图、曲线图、树状图、甘特条形图、文本图等，可以方便地切换不同视图。Tableau Desktop 支持地图数据分析，只需一些位置数据，如城市名称、邮政编码、纬度经度等，便可以方便、快



图 1-1 Tableau Desktop 可连接的数据库

速地创建地图视图。此外，Tableau Desktop 还支持创建函数的方式建立视图，如通过创建正弦函数建立雷达图、玫瑰图等。Tableau Desktop 可以动态筛选数据，分析不同类别的趋势，同时也可以将数据放在交互式地图上，所有操作都无需编写任何代码。通过拖放操作，可以生成各种类型的图表，方便查看数据的统计汇总，同时软件还具有趋势分析、回归、关联等众多功能。

3. 可创建交互式仪表板

Tableau Desktop 软件另一个重要特点是可以迅速地创建美观、交互的仪表板。它可以将多个视图整合在交互式仪表板中，可以嵌入到网页和文档，用于进一步探索和分析。和视图一样，仪表板可以与动态数据连接，通过仪表板可以让用户多维度地查看数据间的关系，挖掘数据潜在价值信息。通过创建交互式仪表板，分析人员可非常方便地钻取到底层数据，无需编写特制脚本或者预先设置。图 1-2 为 Tableau Desktop 制作的关于国内各城市空气质量的可视化仪表板，通过该仪表板，可以方便地对数据进行动态关联与钻取，实现数据可视化分析。总之，Tableau Desktop 可支持多种方式创建视图并与数据进行可视化的动态交互，为分析数据提供了极大的便利。

4. 复杂计算

Tableau 也提供了由函数库支持的功能强大的语言，可用于表达复杂逻辑。借助计算字段，用户可以轻松执行算术运算、表达条件逻辑，或对特定数据类型进行特殊化操作。此外，在 Tableau 中，通过简单的操作即可实现时间序列以及预测性分析，为处理时间型数据提供了便利。

5. 与 R、Python 等编程语言连接

Tableau 与 Python 和 R 直接集成，从而让用户可以使用现有模型并从全球统计社区获得帮助。Tableau 将脚本和数据发送给 RServe 和 TabPy 之类的外部预测服务机构。

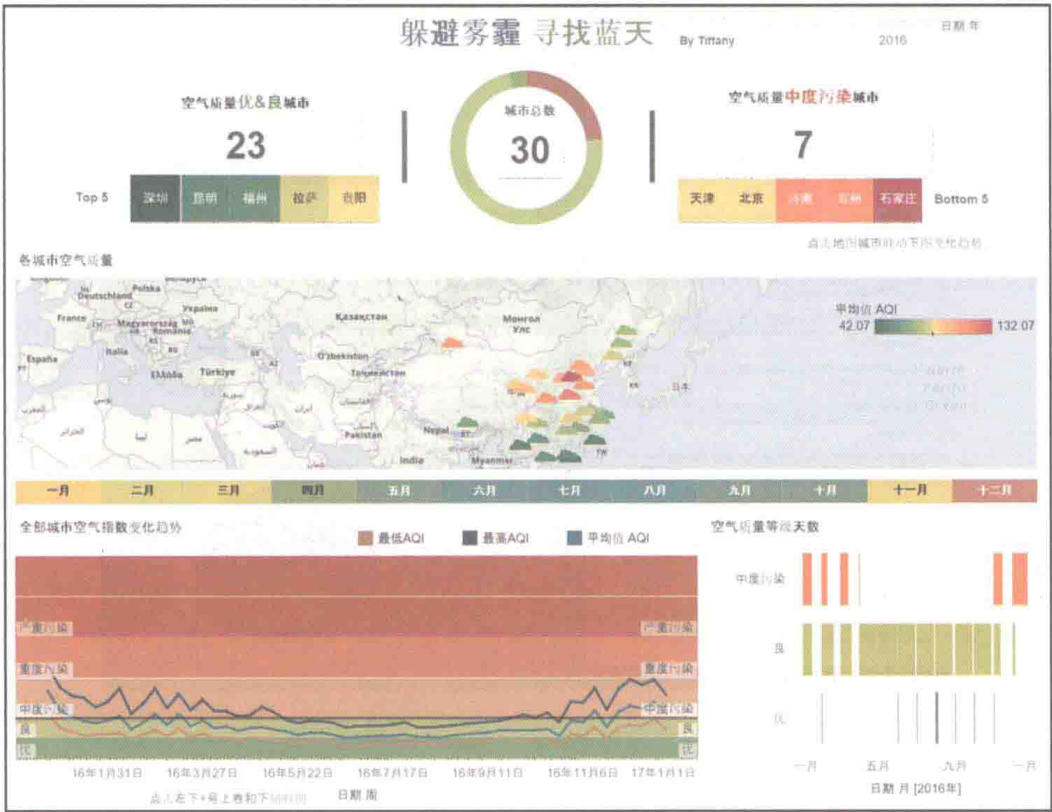


图 1-2 Tableau Desktop 可视化仪表板示例

随后，结果会返回到 Tableau，供 Tableau 可视化引擎使用。因此，Tableau 用户可以针对 Tableau 中的数据调用任何可用的 R 或 Python 函数，并对使用 Tableau 在这些环境中创建的模型实施操作。

图 1-3 中的仪表板使用名为“梯度提升”的集成式机器学习技术，根据历史癌症病例，针对乳腺肿瘤提供诊断建议。此模型经过训练，并由数据科学家在 Python 编程环境中对其准确性进行统计学评估，随后在 TabP（Tableau Python Server）上发布。此仪表板提供表格输入字段，使用户可以为不同的医学指标键入值，这些值被传送到托管的机器学习模型并实时转化为诊断建议。这样，众多分析师和仪表板可以扩充共享的机器学习资产，而数据科学家可以使用自己喜欢的工具，在同一个地点轻松管理这些资产。

Tableau 和 Python 及 R 的组合具有极其强大的功能。它可以将复杂的建模结果表达为美观而有趣的可视化，让各种技能水平的用户都能使用这些结果，并通过与仪表板中内嵌的控件进行交互，提出假设问题并评估假想方案。

此外，Tableau Desktop 中还内置了聚类分析等功能。Tableau 中聚类分析是一个使用无人监督的机器学习来分割数据的选项。在图 1-4 中，用户可以看到使用联合国发展指标来对各国/地区进行聚类分析获得的结果。该算法建议使用 3 群集解决方案，3 个群集分别对应不发达国家/地区、发展中国家/地区和发达国家/地区。

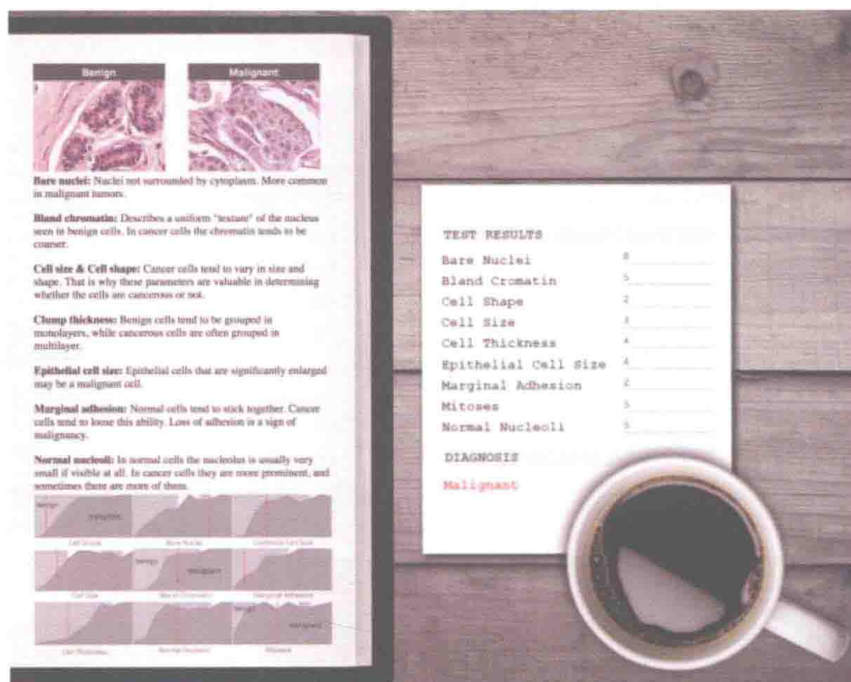


图 1-3 集成式机器学习模型及其输入参数

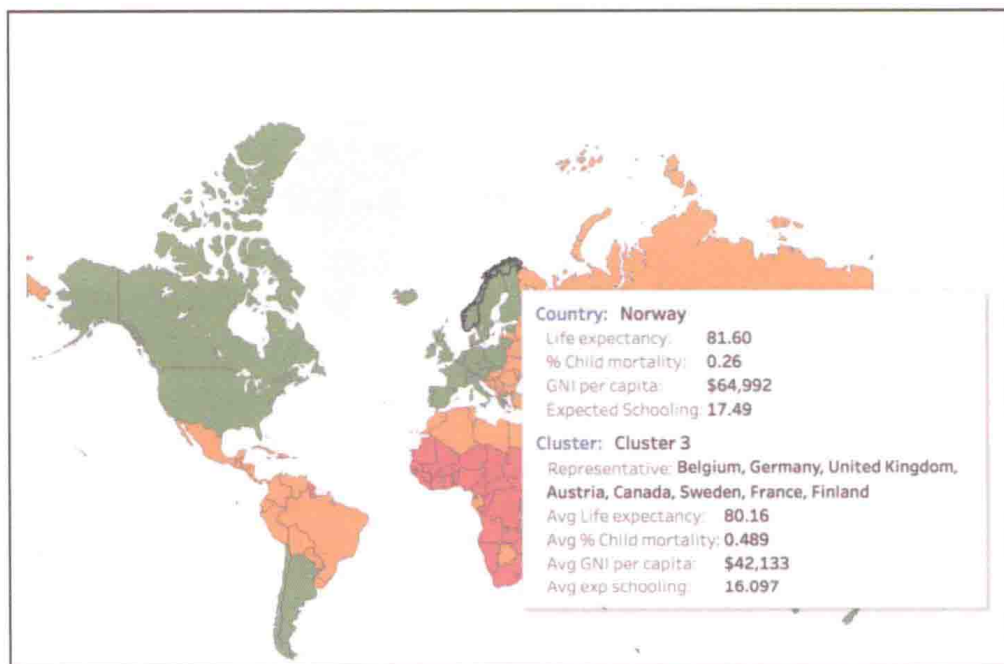


图 1-4 联合国发展指标 Tableau 聚类分析示例

借助 Tableau 的灵活前端，业务用户可以在无需进行编码，也无需理解数据库的情况下提出问题。Tableau 还可以凭借其分析深度成为数据科学家手中的一件强大武器。利用复杂计算、R 和 Python 集成、以及预测功能，数据科学家可以在 Tableau 中完成复杂分析并轻松共享可视化结果。无论使用 Tableau 的目的是进行数据探索和质量控制，



还是进行模型设计和测试，Tableau 的灵活性都可以让整个分析节省大量时间。Tableau 可以让各级别的分析更加易于访问并缩短分析所需时间，从而在整个企业推动关键协作、提高决策质量。

由于 Tableau Desktop 数据可视化分析软件的优越性，目前已有包括国家电网、华为、中国东方航空等国内大型企业开始使用 Tableau Desktop 软件。南方电网公司资产管理部门从 2015 年开始逐渐使用 Tableau Desktop 软件，并将其应用到资产管理的各个环节之中，显著提高了数据分析实时性和管理效率。

数据连接与融合

本章主要介绍了 Tableau 连接数据的基本操作，包括数据表之间的连接方式，如左侧连接、右侧连接、内部连接和外部连接，通过 Tableau 提供的表连接功能，可以快速地从数据库中创建数据中间表。此外，本章还介绍了多张数据表的合并方法，通过该功能可以直接在 Tableau 中合并同类表格，提高了数据合并效率。

2.1 连接 Excel 文件

连接 Excel 文件的步骤如下：

(1) 启动 Tableau 后，在开始界面（见图 2-1）中，单击“Excel”，并在弹出的对话框中找到相应 Excel 文件即可。

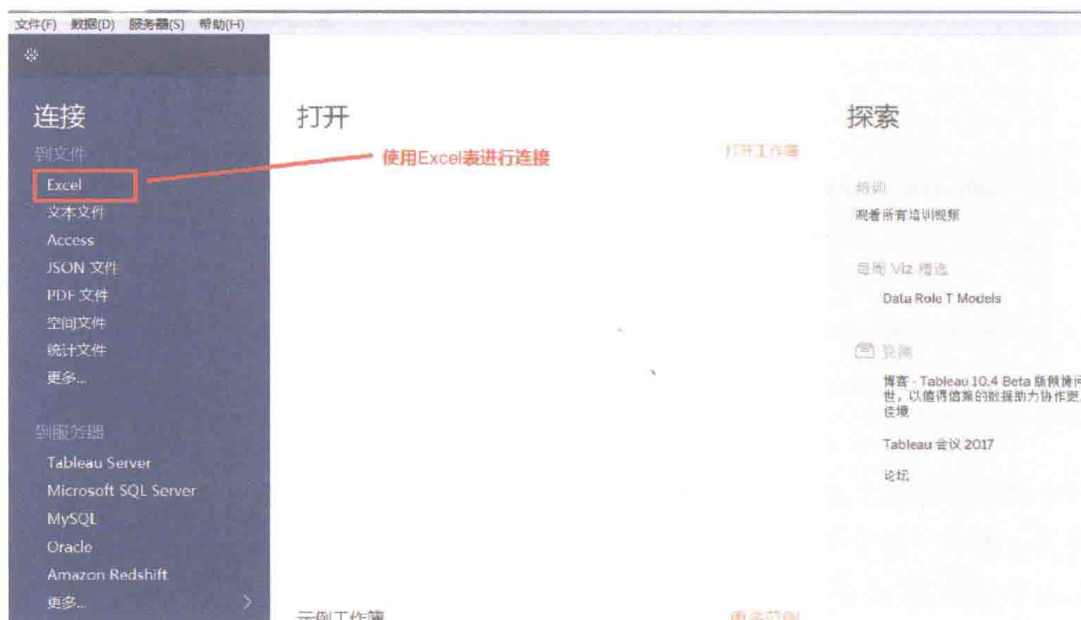


图 2-1 Tableau 开始界面



(2) 图 2-2 为连接 Excel 文件后的界面图，即数据表操作界面。在该界面下，左边红色框 1 中显示的为 Excel 表中的各工作表名称。为了分析某一张工作表数据，只需将对应工作表拖拽到图 2-2 所示的红色框 2 区域。完成拖拽后，工作表中的数据会在图中红色框 3 中显示。最后，单击“工作表”，切换到操作界面。



图 2-2 数据表操作界面

Tableau 操作界面如图 2-3 所示。整个界面可分为以下几个功能区：



图 2-3 工作表操作界面

红色框 1、2、3 表示维度、度量和参数区：该功能区显示数据表中的字段名称和用户创建的参数名称。其中，维度框中字段数据为定性数据，不可以进行数学运算；度量框中的字段数据为定量数据，可以进行数学运算。

红色框 4、5、6 表示筛选器、标记、图例区。筛选器区可添加筛选字段。标记区可



对绘图区中的图表进行格式设置，主要包含颜色、大小、标签等。颜色图例区显示绘图区中的颜色图例。

红色框 7 表示横纵坐标区。制图时，只需将“维度和度量区”中的字段拖拽到该区域中的“行”和“列”，Tableau 即可自动生成相应图形。

红色框 8 表示绘图区，即图、表显示区域。

红色框 9 表示智能图表显示框，包含了 Tableau 内置的 24 种展示图表形式，通过该功能区可以快速地选择需要展示的图形格式。

红色框 10 表示筛选器区中的筛选字段。

2.2 数据源操作

2.2.1 数据表连接

Tableau 中，可以对多张数据表进行连接，形成一张宽表。例如，在对变电设备生产数据中的台账数据和缺陷数据进行分析时，可以将设备台账表和缺陷表通过 Tableau 的连接功能进行整合，方便数据的分析。

Tableau 中的数据表连接类型分为内部、左侧、右侧和完全外部连接，如图 2-4 所示。



图 2-4 数据表连接

下面以样本数据中的“数据表连接示例数据 1”和“数据表连接示例数据 2”为例，说明各种连接类型的作用。

表 2-1、表 2-2 分别为“数据表连接示例数据 1”和“数据表连接示例数据 2”中的数据，表 2-1 表示台账数据，表 2-2 表示缺陷数据。通过两张表中的公共字段“设备 ID”，

可以在 Tableau 中进行关联。

表 2-1 数据表连接示例数据 1

设备 ID	设备电压等级 (kV)	省级	地市级单位	设备型号	投运年份
1	500	A	A1	3AP2 FI	2007
2	500	A	A1	3AP2 FI	2007
3	500	A	A1	3AP2 FI	2013
4	500	A	A1	LW10B	2013
5	500	A	A2	LW10B	2013
6	500	A	A2	3AP2 FI	2009
7	500	A	A2	LW15	2010
8	500	A	A2	LW15	2010

表 2-2 数据表连接示例数据 2

设备 ID	缺陷年份	缺陷原因	缺陷部位	严重等级
1	2014	元器件损坏	机构机械部件	紧急
1	2015	接触不良	机构二次部件	重大
5	2015	过热	本体载流部件	重大
7	2015	元器件损坏	机构二次部件	重大
7	2015	元器件损坏	机构二次部件	重大
9	2015	元器件损坏	机构机械部件	重大

无论采用何种形式的连接，Tableau 中都会将连接表中的所有字段整合到一张宽表中，不同的是当采用不同连接时，各连接表中的数据会因为连接形式的不同而形成不同的宽表。

(1) 内部连接。内部连接指对连接表之间的数据取交集。以表 2-1 和表 2-2 数据为例，在 Tableau 中选择内部连接，并指定连接字段为“设备 ID”，连接后的结果如图 2-5 所示。

设备ID	设备电压等级...	省级	地市级单位	设备型号	投运年份	设...	缺陷年份	缺陷原因	缺陷部位	严重等级
1	500	A	A1	3AP2 FI	2,007	1	2,014	元器件损坏	机构机械部件	紧急
5	500	A	A2	LW10B	2,013	5	2,015	过热	本体载流部件	重大
7	500	A	A2	LW15	2,010	7	2,015	元器件损坏	机构二次部件	重大
1	500	A	A1	3AP2 FI	2,007	1	2,015	接触不良	机构二次部件	重大
7	500	A	A2	LW15	2,010	7	2,015	元器件损坏	机构二次部件	重大

图 2-5 内部连接

由图 2-5 可看出，连接后 Tableau 形成了一张新的宽表，表头中的字段为表 2-1 和



表 2-2 字段的整合。而宽表数据则取表 2-1 和表 2-2 数据的交集。需要注意的是，Tableau 在对表 2-1 和表 2-2 取交集时，保留了两张表中共有设备 ID 的所有数据，而将非公共的设备 ID 过滤掉。

(2) 左侧连接。左侧连接指以左侧的表为主表，与被连接表连接时，Tableau 会保留主表的所有数据，以及被连接表中与主表有交集的数据。仍以表 2-1、表 2-2 数据为例，在 Tableau 中以表 2-1 作为主表，表 2-2 作为连接表，并设置“设备 ID”为连接字段，结果如图 2-6 所示。

设备ID	设备电压等级...	省级	地市级单位	设备型号	投运年份	设...	缺陷年份	缺陷原因	缺陷部位	严重等级
1	500	A	A1	3AP2 FI	2,007	1	2,014	元器件损坏	机构机械部件	紧急
2	500	A	A1	3AP2 FI	2,007	null	null	null	null	null
3	500	A	A1	3AP2 FI	2,013	null	null	null	null	null
4	500	A	A1	LW10B	2,013	null	null	null	null	null
5	500	A	A2	LW10B	2,013	5	2,015	过热	本体载流部件	重大
6	500	A	A2	3AP2 FI	2,009	null	null	null	null	null
7	500	A	A2	LW15	2,010	7	2,015	元器件损坏	机构二次部件	重大
8	500	A	A2	LW15	2,010	null	null	null	null	null
1	500	A	A1	3AP2 FI	2,007	1	2,015	接触不良	机构二次部件	重大
7	500	A	A2	LW15	2,010	7	2,015	元器件损坏	机构二次部件	重大

图 2-6 左侧连接

由图 2-6 可看出，使用左侧连接后，Tableau 形成的宽表表头包含了表 2-1 和表 2-2 的表头。此外，从数据条数来看，宽表中的数据包含了表 2-1 的所有数据，同时还包含了表 2-2 中与表 2-1 有交集的数据。在图 2-6 中，由于表 2-1 台账数据中的设备编号 2、3、4、6、8 未发生缺陷，因此形成宽表后，对应缺陷表头字段的内容 Tableau 用 Null 值进行填充。

(3) 右侧连接。与左侧连接原理类似，右侧连接以右侧的表为主表，左侧的表为辅表，进行连接。右侧连接如图 2-7 所示。

设备ID	设备电压等级...	省级	地市级单位	设备型号	投运年份	设...	缺陷年份	缺陷原因	缺陷部位	严重等级
1	500	A	A1	3AP2 FI	2,007	1	2,014	元器件损坏	机构机械部件	紧急
5	500	A	A2	LW10B	2,013	5	2,015	过热	本体载流部件	重大
7	500	A	A2	LW15	2,010	7	2,015	元器件损坏	机构二次部件	重大
1	500	A	A1	3AP2 FI	2,007	1	2,015	接触不良	机构二次部件	重大
7	500	A	A2	LW15	2,010	7	2,015	元器件损坏	机构二次部件	重大
null	null	null	null	null	null	9	2,015	元器件损坏	机构机械部件	重大

图 2-7 右侧连接