

第一章

数据可视化及 Tableau

【学习目标】

1. 了解数据与数据可视化之间的联系与差别以及常用的数据可视化软件。
2. 掌握 Tableau 软件的操作界面。

【能力目标】

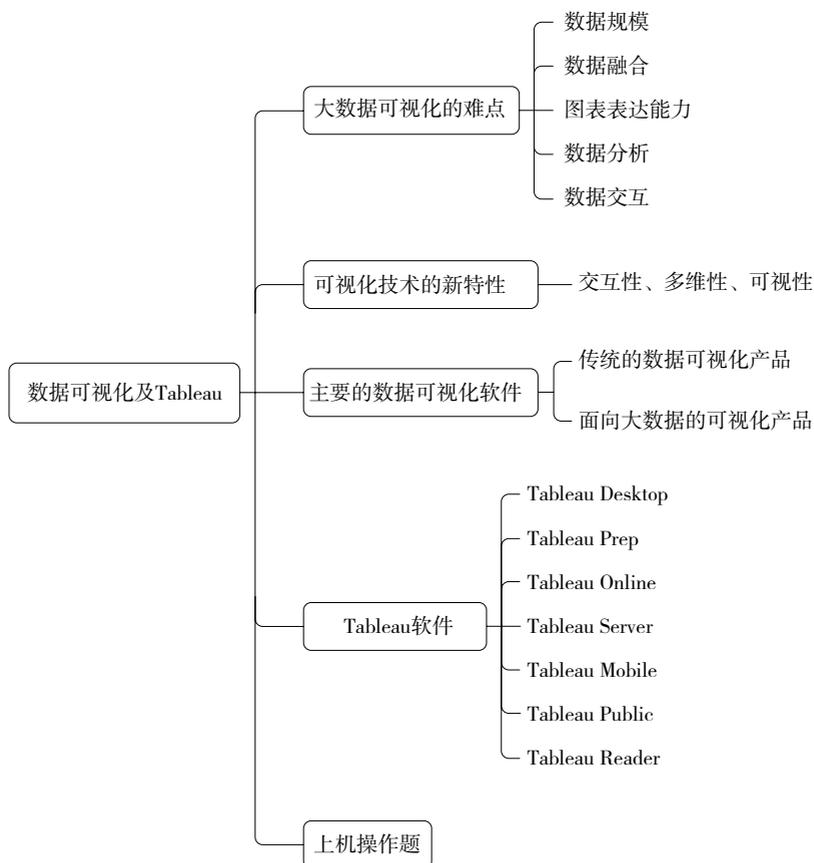
1. 了解数据与数据可视化之间的联系、差别以及常用的数据可视化软件，提高信息素养。
2. 掌握 Tableau 软件的操作界面，培养自学能力。

【思政目标】

1. 了解数据与数据可视化之间的联系与区别以及常用的数据可视化软件，培养学术辩证思维能力。
2. 掌握 Tableau 软件的操作界面，培养审美观。



【思维导图】



第一节 数据可视化

2016年9月，中国在G20峰会上第一次给出了数字经济的官方定义：数字经济是指以使用数字化的知识和信息作为关键生产要素、以现代信息网络作为重要载体、以信息通信技术的有效使用作为效率提升和经济结构优化的重要推动力的一系列经济活动。大数据的核心不是“大”，也不是“数据”，而是蕴含在其中的商业价值。作为挖掘数据背后潜在价值的重要手段，商业智能和分析平台成为大数据部署中的关键环节。同时，获取价值的难点不在于数据分析应用的部署，而在于专业数据分析人才的缺失。

长久以来，企业的大量数据主要为 IT 部门所掌管，业务部门或者分析人员若需进行数据探索、分析或简单的报表操作，则要联动其他部门配合，这大大降低了工作效率。但数据恰恰在业务和分析人员手中才有价值产出，在此背景下，人们需要一种工具能帮助更多的人读懂数据，并释放其中的潜能。

一、大数据可视化的难点

现阶段，数据可视化在大数据场景应用中仍面临诸多难点，包括数据规模、数据融合、图表表达能力、数据分析与数据交互等。

1. 数据规模

大数据规模大、价值密度低，而屏幕空间所能显示的数据量有限，因此为了有效展示使用者所关注的的数据及其特征，需要采用有效的数据压缩方法。目前已有的方法通常只是针对数据本身进行采样或聚合，多未考虑数据可视化的显示特性。近期一些学者提出了针对特定可视化场景的数据压缩方法，但是依然缺少通用的、面向可视化的数据压缩方法，也缺少实际应用的产品。

2. 数据融合

大数据的另一个表现是数据类型多样，常常分布于不同的数据库。融合不同来源及不同类型的数据、为使用者提供统一的可视化视角、支持可视化的关联探索与关系挖掘等都是重要的问题。这涉及数据关联的自动发现、多类型数据可视化、构建知识图谱等多个具体技术问题。

3. 图表表达能力

随着数据来源的增加，数据类型也不断增加，数据使用者对数据的交互需求越来越多，已有的数据可视化产品无法满足使用者的可视化需求，时常出现可视化形式的产品不支持或支持不够等问题。这就对系统的图表表达能力提出了更高的要求，同时使用者对系统支持的个性化定制提出了新的要求。

4. 数据分析

传统的 BI 工具主要聚焦于数据筛选、聚合及可视化等功能，其已经不能满足大数据分析的需求，所以盖特纳公司（Gartner）提出了“增强分析”的概念，其认为数据可视化只有结合丰富的大数据分析方法，将数据的探索式分析形成一个闭环，才能实现完整的大数据可视化产品，有效帮助使用者理解数据。预测性分析是大数据发展的趋势，数据可视化有效地结合预测方法将有助于使用者进行决策。

5. 数据交互

数据可视化的使用者需要通过可视化与图表背后的数据和处理逻辑进行交互，由此反映使用者的个性化需求，帮助其用一种交互迭代的方式理解数据。在传统的交互手段基础上，更加自然的交互方式将有助于使用者与数据更好地交互，也有助于拓展大数据可视化产品的使用范围与应用场景。

数据可视化技术与产品面临主要挑战的同时也面对新的机遇，如 Yu 等人提出的面向数据流式可视化的自然语言交互接口可以通过自然语言与可视化常见操作的映射实现交互。微软公司的 Excel 软件也集成了自然语言交互，其中的 AnnaParser 算法能够对数据表进行抽象并结合表格知识理解和实现语义。

二、可视化技术的新特性

(1) 交互性。可视化分析是获取数据、单向表示数据、注意结果和提出后续问题的过程，后续问题可能需要向下钻取、向上钻取、筛选、引入新数据或创建数据的其他视图。

(2) 多维性。数据可视化必须足够灵活以便说明各种问题，可以按每一维的值对数据进行分类、排序、组合和显示。

(3) 可视性。数据可以用图像、曲线、二维图形、三维体和动画来显示，并可通过多种模式和关联关系被应用于可视化分析。

三、主要的数据可视化软件

1. 传统的数据可视化产品

1) PowerBI

作为微软推出的数据可视化产品，PowerBI 在 2019 年的 GartnerBI 象限中排在首位。其优点在于易用性，交互方式类似 Excel；缺点在于性能相对较弱，缺少数据准备与清洗工具。

2) Tableau

基于关系型代数理论研发的 Tableau 是目前使用最为广泛的数据可视化产品之一。其优点在于支持基于拖放的交互方式、丰富的功能以及支持 Hadoop 和 Google BigQuery 等大数据平台；其缺点是仅支持结构化数据，大数据实时响应较慢，权限约束有限。

3) QlikView

QlikView 为新兴的数据可视化产品,也越来越得到广泛应用。其优点在于数据关联查询与钻取能力,图表绘制快速;其缺点在于易用性不足,作为内存型的数据可视化产品,其数据处理速度依赖内存,对硬件要求较高。

2. 面向大数据的可视化产品

1) Apache Superset

基于 Flask-AppBuilder 构建的开源数据可视化系统,其为 B/S 架构,集成了地图、折线图、饼图等可视化方法,并提供了方便的看板定制功能。Apache Superset 的优点是系统具有可扩展性与权限控制机制;其缺点是系统稳定性较差和大数据处理能力不足。

2) Apache Zeppelin

Apache Zeppelin 拥有面向大数据的交互式数据分析与协作记事本工具,为开源项目,同为 B/S 架构。其优点是具有与不同于大数据框架的集成能力与系统可扩展性;其缺点是需要编程、不支持异步数据,大规模数据处理效率略低,客户端可能需要等待较长的时间。

第二节 Tableau 软件

Tableau 公司成立于 2003 年,主要面向企业用户提供数据可视化服务,是一家商业智能软件提供商。运用 Tableau 授权的数据可视化软件对数据进行处理和展示的多为企业,但 Tableau 的产品并不局限于企业,其他机构以及个人均可使用 Tableau 软件进行数据分析。数据可视化是数据分析的完美结果,其能够让枯燥的数据以简单友好的图表形式展现出来。Tableau 在抢占细分市场(也就是大数据处理末端的可视化市场)上具有一定的优势,同时,其还为客户提供解决方案服务。

一、Tableau Desktop

“所有人都能学会的业务分析工具”,这是 Tableau 官方网站对 Tableau Desktop 的描述。Tableau Desktop 的使用者不需要精通复杂的编程和统计原理,只需要把数据直接拖放到工具簿中,通过一些简单的设置就可以得到想要的可视化图形。Tableau Desktop 是一款完全的数据可视化软件,如图 1-1 所示。

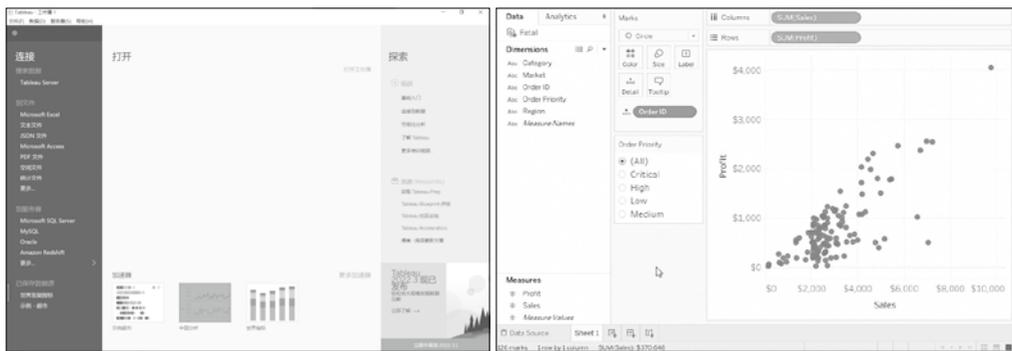


图 1-1 Tableau Desktop 的开始页面

Tableau Desktop 专注于结构化数据的快速可视化，使用者可以快速进行数据可视化并构建交互界面，用来辅助人们进行可视化思考。快速、易用的可视化是 Tableau Desktop 最大的特点，其能够满足大多数企业、政府机构的数据分析和展示需求，也能满足部分大学、研究机构的可视化项目需求。简单、易用的同时，Tableau Desktop 还拥有高效的数据引擎；Tableau Desktop 还具有完美的数据整合能力，可以将两个数据源整合在同一层，甚至可以将一个数据源筛选为另一个数据源，并将之在数据源中突出显示。Tableau Desktop 这种强大的数据整合能力具有很大的实用性。另外，Tableau Desktop 还有一项独具特色的数据可视化功能——嵌入地图，使用者可以用经过自动地理编码的地图呈现数据，这对企业进行产品市场定位、制定营销策略等有非常大的帮助。

二、Tableau Prep

Tableau Prep 是一款简单易用的数据处理工具（部分 ETL 工作），其是一种更方便的、按需搭建数据模型的工具，可以与 Tableau Desktop、Tableau Server 和 Tableau Online 进行无缝衔接，并随时随地在 Tableau Prep 中进行数据提取、将数据源发布到 Tableau Server 或 Tableau Online。用户还可以直接从 Tableau Prep 中打开 Tableau Desktop 进行数据预览。

Tableau Prep 保持了与 Tableau Desktop 一致的蓝色基调 UI，但不支持多语言选择，其界面分为 3 部分，左侧第一部分用于数据链接，中间是最近使用过的操作流程及预设的展示操作流程，右侧则是一些教学资源，如图 1-2 所示。

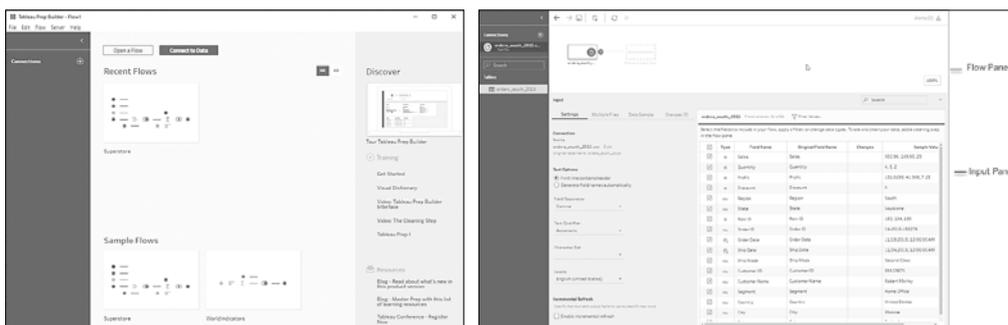


图 1-2 Tableau Prep 的界面

三、Tableau Online

Tableau Online 是 Tableau Server 的软件及服务托管版本，它让商业分析更加快速轻松。操作者可以利用 Tableau Desktop 发布仪表板，然后将之与同事、合作伙伴或客户共享，利用云商业智能技术随时随地、快速地找到答案。Tableau Online 的页面如图 1-3 所示。

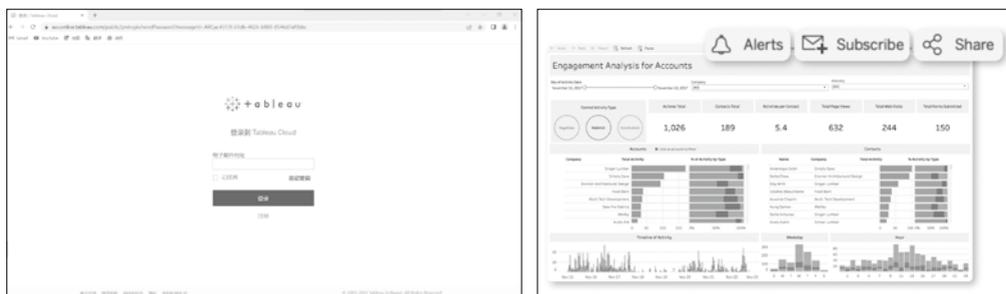


图 1-3 Tableau Online 的页面

Tableau Online 可连接云端数据和办公室内的数据，还可连接其他托管在云端的数据源（如 Salesforce 和 Google Analytics）并按计划安排数据刷新，或从公司内部向 Tableau Online 推送数据，从而让团队轻松访问并按设定的计划刷新数据，在数据连接发生故障时获得警报。

四、Tableau Server

Tableau Server 是一种新型的商业智能服务，其通过企业服务器安装 Tableau Server，并由管理员进行管理，允许管理员将需要访问 Tableau Server 的人员添加为用户（无论是发布、浏览还是管理）。传统的商业智能系统往往很笨重、复杂，需

要专业技术人员（通常为企业的工厂部门员工）操作和维护，但 IT 技术人员通常缺乏商业背景，这导致了系统利用的低效率和时间滞后。Tableau Server 为用户分配许可级别，按照许可级别管理和分配不同的权限，被许可的用户可以将自己在 Tableau Desktop（只支持专业版）中完成的数据、可视化内容、报告与工作簿发布到 Tableau Server 中与人共享。他人可以查看这些被共享的数据并进行交互，通过共享的数据源以极快的速度工作。

五、Tableau Mobile

Tableau Mobile 可以帮助用户随时掌握数据（其需要搭配 Tableau Online 或 Tableau Server 账户才能使用），可以让用户快速流畅地查看数据，提供快捷、轻松的数据处理途径。Tableau Mobile 的主要功能有随处编写和查看、脱机快照、订阅、灵活展示图表、增强内容的安全性、共享并与团队轻松协作等。

六、Tableau Public

Tableau Public 可以连接数据、创建交互式数据可视化内容，并将其直接发布到自己的网站，通过发现数据的内在含义引导读者，让他们与数据互动，发掘新的见解，这一切不用编写代码即可实现。

七、Tableau Reader

Tableau Reader 是一款免费桌面应用程序，其可让用户与 Tableau Desktop 中生成的可视化数据进行交互。利用 Tableau Reader，用户可以筛选、向下钻取和查看数据明细，使数据详细到用户需要的程度。

上机操作题

- （1）了解数据可视化技术，熟悉 Tableau 的操作界面。
- （2）熟悉 Tableau 各产品的特性，并思考数据可视化产品之间的关系。

第二章

Tableau Desktop



【学习目标】

1. 了解 Tableau Desktop 的新增功能。
2. 熟悉数据类型、运算符以及运算符之间的优先级关系。
3. 掌握 Tableau Desktop 的软件安装流程、软件界面的布局及其可以导出的文件类型。



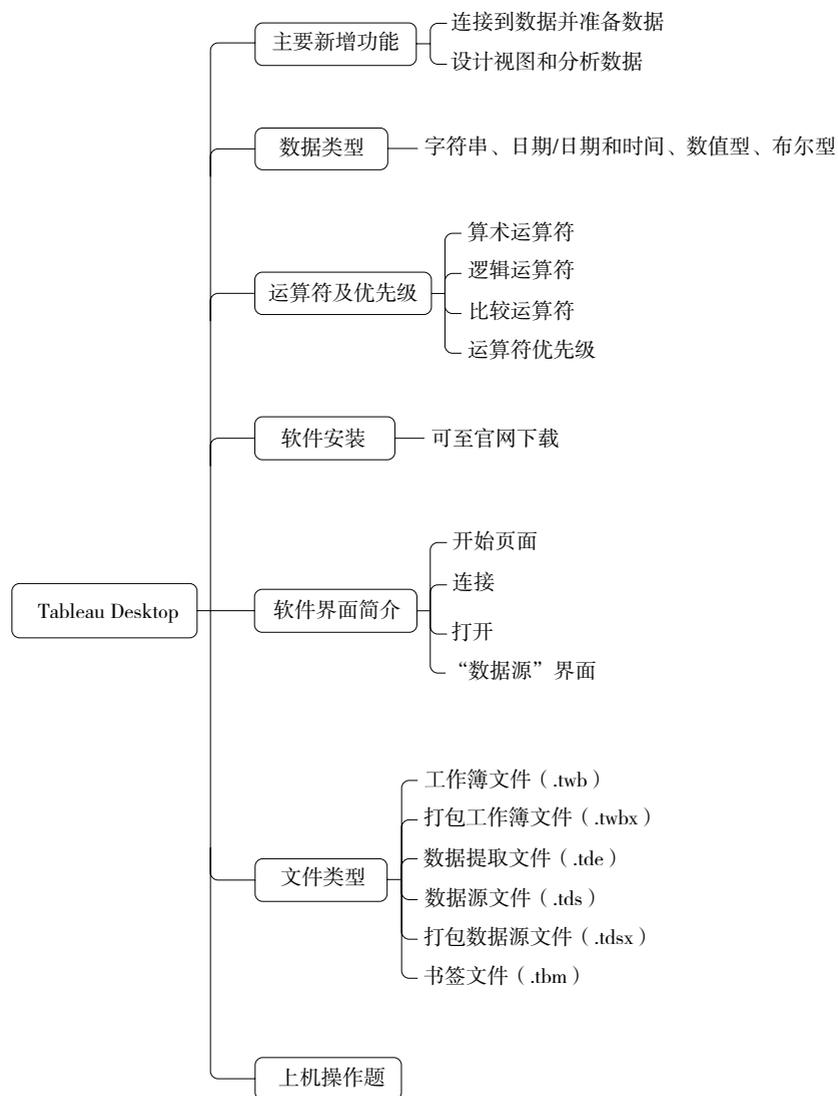
【能力目标】

1. 了解 Tableau Desktop 的新增功能，培养独立思考能力。
2. 熟悉数据类型、运算符以及运算符之间的优先级关系，培养逻辑思维能力。
3. 掌握 Tableau Desktop 的软件安装流程、软件界面的布局及其可以导出的文件类型，培养实操能力。



【思政目标】

1. 了解 Tableau Desktop 的新增功能，培养创新精神。
2. 熟悉数据类型、运算符以及运算符之间的优先级关系，培养严谨的科学精神。
3. 掌握 Tableau Desktop 的软件安装流程、软件界面的布局及其可以导出的文件类型，培养主观能动性。

 【思维导图】


第一节 主要新增功能

用户可以在 Tableau 的官方网站下载最新版本的免费试用软件。截至 2021 年 7 月, Tableau 软件的最新版本是 Tableau Desktop 2020.4, 本书也是基于该版本编写的, 为了后续能更好地学习本书内容, 建议读者下载和安装该版本软件。

Tableau Desktop 2020.4 的主要新增功能如下。

(1) 连接到数据并准备数据。主要包括：使用关系为多表数据的分析合并；通过 Snowflake 代理配置连接数据；通过 Azure Synapse Analytics 连接器连接数据；连接到 Esri ArcGSI Server 服务器；连接到 Oracle 数据库中的空间字段。

(2) 设计视图和分析数据。主要包括：添加集控件（其允许用户快速修改集的成员）；通过与可视化项之间直接交互在集内添加或移除值；通过“数据解释”功能控制用于分析的字段。

第二节 数据类型

数据源中的所有字段都属于一种数据类型。数据类型反映了该字段所存储信息的种类，如整数、日期和字符串。字段的数据类型在“数据”窗格中由图标标识，Tableau Desktop 的主要数据类型如图 2-1 所示。

图标	数据类型
Abc	文本（字符串）值
☐	日期值
🕒	日期和时间值
#	数字值
T/F	布尔值（仅限关系数据源）
📍	地理值（用于地图）

图 2-1 Tableau Desktop 的主要数据类型

下面介绍 Tableau 支持的几种数据类型。

1. 字符串

字符串（string）是由 1 个或多个字符组成的序列。例如，“Wisconsin” “ID-44400” 和 “Tom Sawyer” 都是字符串。字符串通过单引号或双引号识别，引号本身可以被重复包含在字符串中，如 “OHanrahan”。

2. 日期 / 日期和时间

日期 / 日期和时间（date/date time）表示的信息如 “January 23, 2020” 或 “January 23, 2020 12: 32: 00AM”。如果要将以长型格式编写的日期解释为日期 / 日期和时间，就要在日期两端放置 “#” 符号。例如，“January 23, 2020” 会被视为字符串数

据类型，而 # January 23, 2020# 会被视为日期 / 日期和时间数据类型。

3. 数值型

Tableau 中的数值 (numeric) 可以为整数或浮点数。浮点数计算的结果可能并非完全符合预期。例如，当 SUM 函数的返回值为“-1.42e-14”时，求和结果正好为“0”，出现这种情况的原因是程序中的数字以二进制格式存储，有时会以极高的精度级别被舍入。

4. 布尔型

布尔型 (boolean) 的包含值为 true 或 false 的字段，当结果未知时会出现未知值。例如，表达式 $7 > \text{Null}$ 会生成未知值，并被自动转换为 Null。

此外，Tableau 中还支持地理数据，该类型的字段可以根据用户需要将省市数据转换为具有经度、纬度坐标的字段，这是地图可视化分析的前提。

在日常工作中，Tableau 可能会将字段标识为错误的数据类型。例如，它可能将包含日期的字段标识为整数而不是日期，用户可以在“数据源”界面上更改原始数据源字段的数据类型。

在“数据源”界面单击字段的“字段类型”按钮，从下拉列表中选择一种新数据类型，如图 2-2 所示。



图 2-2 在“数据源”界面中更改数据类型

提取数据时，需要确保之前已经更改了所有必要的数据类型，否则数据会不准确。例如，Tableau 会把原始数据源中的浮点数字段解释为整数，生成的浮点数字段的部分精度会被截断。

如果要在“数据”窗格中更改字段的数据类型，可以单击字段的“字段类型”按钮，然后从下拉列表中选择一种新数据类型，如图 2-3 所示。

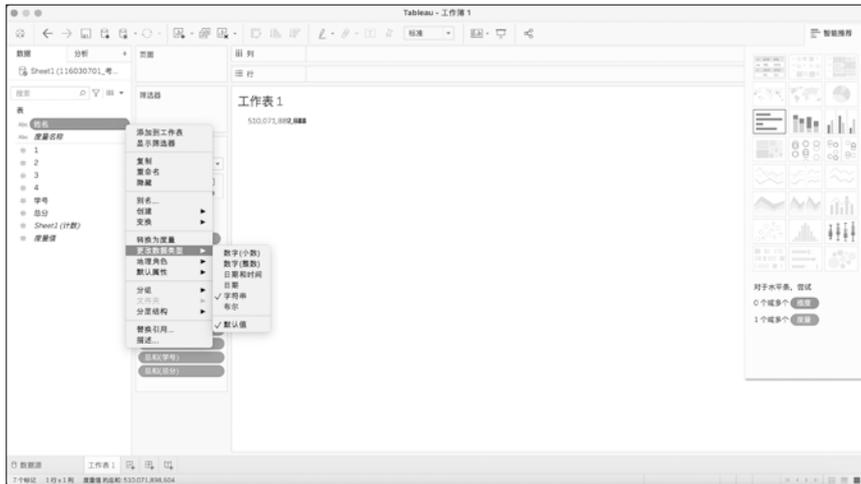


图 2-3 在“数据”窗格中更改数据类型

若要在视图中更改字段的数据类型，则要在“数据”窗格中右击需要更改数据类型的字段，选择“更改数据类型”，然后选择需要的数据类型，如图 2-4 所示。

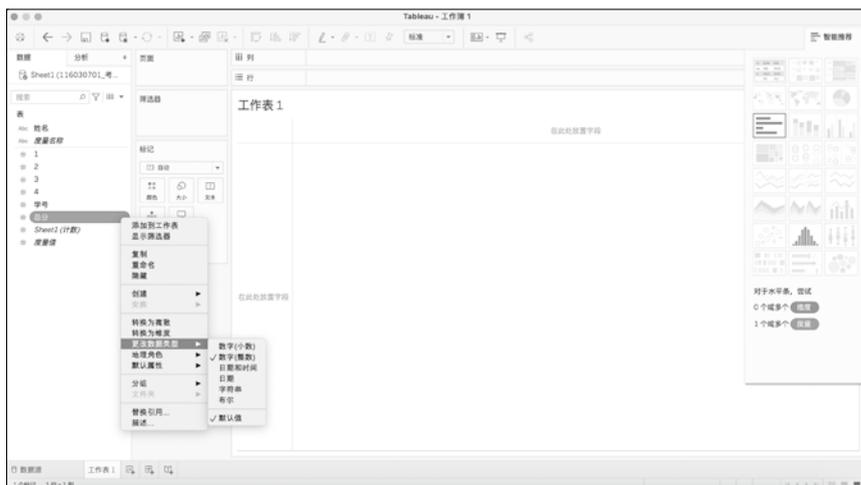


图 2-4 在视图中更改数据类型

此外，由于数据库中的数据比 Tableau 中数据建模的精度更高，因此将这些值添加到视图中时，状态栏右侧将显示一个精度警告对话框。

第三节 运算符及优先级

运算符用于执行程序代码运算，其通常针对一个及以上的操作数进行运算。例如，“2+3”的操作数是“2”和“3”，运算符是“+”。Tableau 支持的运算符有算术运算符、逻辑运算符、比较运算符等。

1. 算术运算符

(1) “+”（加号）。此运算符应用于数字时表示相加；应用于字符串时表示串联；应用于日期时可将天数与日期相加。例如，'abc'+def='abcdef'、#April 15, 2004#+15=#April 30, 2004#。

(2) “-”（减号）。此运算符应用于数字时表示相减；应用于表达式时表示求反；应用于日期时可从日期中减去天数，还可计算两个日期之间的天数差异。例如，17-3=14、-(7+3)=-10、#April 15, 2004#-#April 8, 2004#=7。

(3) “*”（乘号）。此运算符表示计算数字之积，如 5*4=20。

(4) “/”（除号）。此运算符表示计算数字之商，如 20/4=5。

(5) “%”（求余）。此运算符表示计算数字余数，如 5%4=1。

(6) “^”（乘方）。此符号等效于 POWER 函数，用于计算数字的指定次幂，如 6^3=216。

2. 逻辑运算符

(1) AND。逻辑运算且，两侧必须为表达式或布尔值。

例如，IIF (Profit=100 AND Sales=1000, "High", "Low")，如果两个表达式都为 true，那么结果就为 true；如果任意一个表达式为 unknown，那么结果就为 unknown；其他情况结果都为 false。

(2) OR。逻辑运算或，两侧必须为表达式或布尔值。

例如，IIF (Profit=100 OR Sales=1000, "High", "Low")，如果任意一个表达式为 true，那么结果就为 true；如果两个表达式都为 false，那么结果就为 false；如果两个表达式都为 unknown，那么结果就为 unknown。

(3) NOT。逻辑非运算符，此运算符可用于对另一个布尔值或表达式求反。

例如，IIF (NOT (Sales=Profit), "Not Equal", "Equal")，如果 Sales 等于 Profit，那么结果为 Equal，否则结果为 Not Equal。

3. 比较运算符

Tableau 提供了较多的比较运算符,有“==”或“=”(等于)、“>”(大于)、“<”(小于)、“>=”(大于等于)、“<=”(小于等于)、“!=”或“<>”(不等于)等,这些运算符用于比较两个数字、日期或字符串,并返回布尔值(true 或 false)。

4. 运算符优先级

所有运算符都应按特定顺序计算,如 $2*1+2$ 等于 4 而不等于 6,因为“*”运算符的优先级比“+”运算符高。如表 2-1 所示为运算符的优先级,其按 1~8 依次递减。如果两个运算符具有相同优先级,则按照从左向右的顺序进行计算。

表 2-1 各运算符的优先级

优先级	运算符	优先级	运算符
1	- (求反)	5	==、>、<、>=、<=、!=
2	^ (乘方)	6	NOT
3	*、/、%	7	AND
4	+、-	8	OR

用户可以根据需要使用括号,括号中的运算符在计算时优先于括号外的运算符,即从括号内部开始向外计算,如 $[1+(2*2+1)*(3*6/3)]=31$ ”。

第四节 软件安装

Tableau 是一款非常出色的可视化工具,其软件操作简单而高效,读者可以前往 Tableau 的官方网站下载 (<https://www.tableau.com/zh-cn/support/releases/desktop/2020.4>) Tableau Desktop 2020.4 版本。需要说明的是,Tableau 是一款商业软件,用户可以免费试用 14 天,在试用期间,其所有功能都是免费开放的。以下是 Tableau 软件的安装步骤。

下载 Tableau 软件之后,双击程序图标即可开始软件的安装。首先进入的是安装引导页面,单击“继续”按钮,如图 2-5 所示。

进入软件用户协议条款界面,单击“Continue”按钮,如图 2-6 所示。

在接下来的页面单击“Agree”按钮,如图 2-7 所示。

选择软件的安装路径(也可直接单击“安装”按钮),如图 2-8 所示。



图 2-5 软件安装引导页面

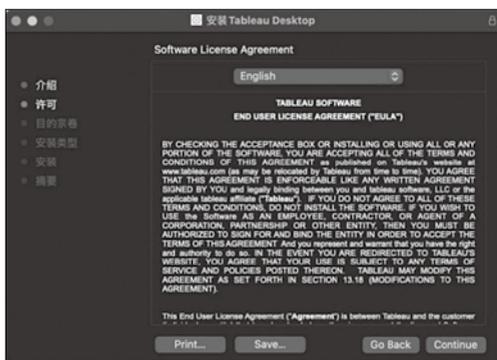


图 2-6 软件用户协议条款

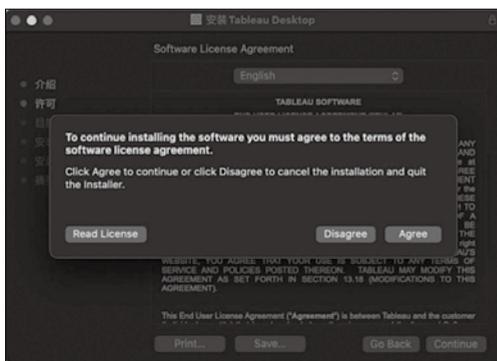


图 2-7 软件用户协议条款



图 2-8 安装路径页面

在设置好相关的安装路径后，单击“安装”按钮，开启 Tableau 的安装进程，该进程需要几分钟的时间完成。在软件安装成功后打开软件，用户需要选择试用或激活，如果是首次使用或者没有购买 Tableau，可以填写相关信息以选择试用 14 天。如果拥有 Tableau 的产品密钥，则可以通过该密钥激活 Tableau，如图 2-9 所示。



图 2-9 选择试用或激活

第五节 软件界面简介

1. 开始页面

Tableau Desktop 的工作簿与 Excel 工作簿十分类似，均包含一个或多个工作表，可以是普通工作表、仪表板或故事。用户通过这些工作簿文件可以对运算结果进

行组织、保存和共享。打开 Tableau 时，程序会自动创建一个空白工作簿，用户也可以自己创建新工作簿，方法是执行“文件”→“新建”命令。

Tableau Desktop 的开始页面主要由“连接”和“打开”两个区域组成，用户可以从连接数据、访问最近使用的工作簿等，如图 2-10 所示。



图 2-10 Tableau Desktop 的开始页面

2. 连接

(1) 连接“到文件”。可以连接存储在 Microsoft Excel 中的文件、文本文件、JSON 文件、Microsoft Access 文件、Tableau 数据提取文件和统计文件等文件中的数据源。

(2) 连接“到服务器”。可以连接存储在数据库中的数据，如 Tableau Server、Microsoft SQL Server、Oracle 或 MySQL 等。

(3) 已保存数据源。可以快速打开之前保存到“我的 Tableau 存储库”文件夹中的数据源，默认情况下这里会显示一些已保存的数据源的实例。

3. 打开

在“打开”窗格中可以执行以下操作。

(1) 访问最近打开的工作簿。首次打开 Tableau Desktop 时此窗格为空；创建和保存新工作簿后，此处将显示最近打开过的工作簿。

(2) 锁定工作簿。可单击工作簿缩略图左上角的“锁定”按钮，将工作簿固定在开始页面中。

4. “数据源”界面

连接数据源“成绩单”（获取资源请扫描右侧二维码）。

在建立数据的初始连接后，Tableau 将引导用户进入“数据源”界面（用户也可以单击“显示开始页面”按钮返回开始页面，并重新连接数据源），如图 2-11 所示。



图 2-11 单击“显示开始页面”按钮

Tableau Desktop 的界面外观和可用选项会因连接的数据类型而异。“数据源”界面通常由 3 个主要区域组成，即左侧窗格、画布和网格，如图 2-12 所示。



图 2-12 “数据源”界面

1) 左侧窗格

“数据源”界面的左侧窗格用于显示有关 Tableau Desktop 连接数据的详细信息。对基于文件的数据，左侧窗格中可能显示数据来源的文件名和文件中的工作表；对关系数据，左侧窗格中可能显示来源服务器、数据库或架构、数据库中的表。

2) 画布

连接若干关系数据和基于文件的数据后，用户可以将一张或多张表拖曳到画布区域的顶部以设置数据源。在连接多维数据集的数据后，“数据源”界面的顶部会显示可用的目录或要从中进行选择的查询和多维数据集。

3) 网格

使用网格后用户可以查看数据源中的字段和前 1 000 行数据，还可以对数据源进行修改，如排序、隐藏、重命名、重置名称、创建计算、更改列 / 行排序或添加别名。

此外，根据连接的数据类型单击“管理元数据”按钮可以导航到元数据网格中。元数据网格会将数据源中的字段显示为行，以便快速检查数据源的结构并执行日常管理操作，如重命名字段或一次性隐藏多个字段等，如图 2-13 所示。

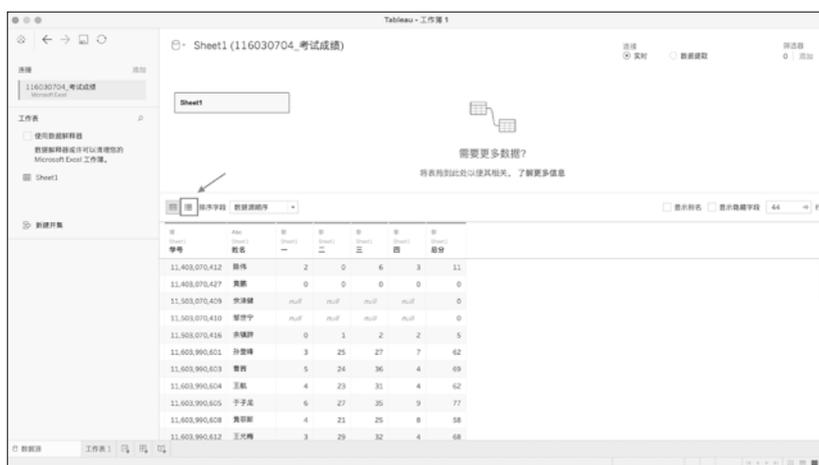


图 2-13 元数据网格

第六节 文件类型

结束数据可视化分析后，用户可以使用多种不同的 Tableau 专用文件类型保存分析结果，如保存工作簿、打包工作簿、数据提取、数据源、打包数据源和书签等。

(1) 工作簿文件 (.twb)。Tableau 的工作簿文件具有 .twb 文件扩展名，工作簿文件中将包含一个或多个工作表，还可以有若干仪表板和故事。

(2) 打包工作簿文件 (.twbx)。Tableau 的打包工作簿文件具有 .twbx 文件扩展名，打包工作簿文件是一个 .zip 文件，其将包含一个工作簿及任何提供支持的本地文件数据源和背景图像，适合与不能访问该数据的人共享。

(3) 数据提取文件 (.tde)。Tableau 的数据提取文件具有 .tde 文件扩展名，是部分或整体数据源的一个本地副本文件，可用于共享数据、脱机工作和增强数据库性能。

(4) 数据源文件 (.tds)。Tableau 的数据源文件具有 .tds 文件扩展名，是连接经常使用的数据源的快捷方式，其并不包含实际数据，只包含连接到数据源所必需的信息和用户“数据”窗格中所做的修改。

(5) 打包数据源文件 (.tdsx)。Tableau 的打包数据源文件具有 .tdsx 文件扩展名，是一个 .zip 文件包，其包含数据源文件 (.tds) 和本地文件数据源。用户可使用此文件格式创建一个文件，以便将之与不能访问该数据的其他人共享。

(6) 书签文件 (.tbn)。Tableau 的书签文件具有 .tbn 文件扩展名，其中包含单个工作表，是快速分享所做工作的简便方式。

上机操作题

- (1) 访问 Tableau 的官方网站，下载并安装 Tableau Desktop。
- (2) 打开软件，进入开始页面，导入一个 Excel 文件并查看数据。
- (3) 比较 Tableau 的运算符及其优先级与其他软件的异同。

第三章

连接数据源

【学习目标】

1. 了解 Microsoft Excel 等文件数据、MySQL 的功能及运用领域。
2. 掌握 Tableau 与外部数据源连接以及使用 Tableau Desktop 连接到主要数据源的方法。

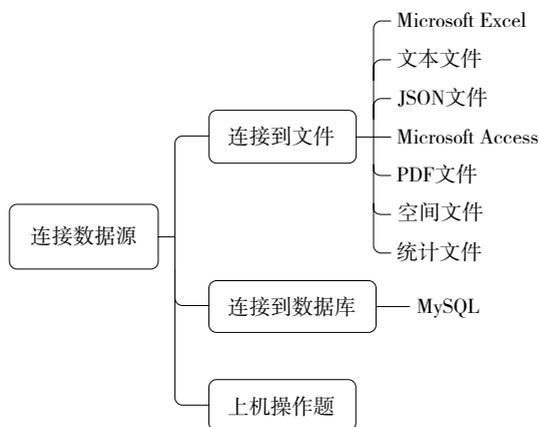
【能力目标】

1. 了解 Microsoft Excel 等文件数据、MySQL 的功能及运用领域，培养自主学习的能力，领略与数据有关的前沿技术。
2. 掌握 Tableau 与外部数据源连接以及用 Tableau Desktop 连接主要数据源的方法，培养多角度分析问题的思维能力。

【思政目标】

1. 了解 Microsoft Excel 等文件数据、MySQL 的功能及运用领域，培养创新精神和科学精神。
2. 掌握 Tableau 与外部数据源连接以及用 Tableau Desktop 连接到主要数据源的方法，培养数据安全意识。

【思维导图】



数据无处不在，分析始于连接。不管是本地数据还是数据库数据、云端数据，Tableau 均可以方便、迅速地与之建立连接，Tableau 支持的各类数据源包括以下几种。

- (1) 文件数据：Excel 文件、文本文件、JSON、统计文件等。
- (2) 关系型数据库：MySQL、Microsoft SQL Server、Oracle、DB2 等。
- (3) 云端数据库：Windows Azure、Google BigQuery、阿里云数据库等。
- (4) 其他数据源：配置 ODBC 驱动器实现各类数据源的连接。
- (5) 组合数据源：支持定义多个连接并将之连接到文件和关系型数据库。

第一节 连接到文件

一、Microsoft Excel

Microsoft Excel 是微软办公软件中用于进行各种数据处理、统计分析和辅助决策的软件，其被广泛应用于管理、统计、金融等领域。

(1) 打开“连接”界面，单击“Microsoft Excel”选项，如图 3-1 所示。

(2) 在浏览文件的页面选择需要导入的 Excel 文件，这里选择的文件是“某公司销售数据.xlsx”，选定之后单击“打开”



图 3-1 连接 Excel

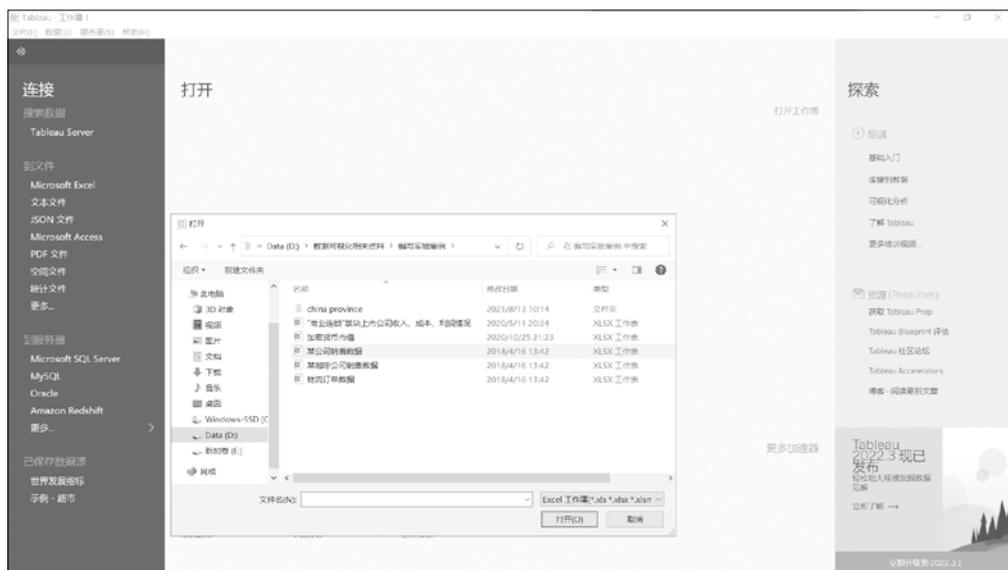


图 3-2 浏览、选择并导入 Excel 文件

按钮，如图 3-2 所示。

(3) 如图 3-3 所示，进入“数据源”界面后，“工作表”标签下是 Excel 中的多个工作表，用户可以根据需求进行选择，这里选择“全国订单明细（某公司销售数据）”工作表，双击“全国订单明细”或将之拖曳到指定位置即可导入数据，如图 3-4 所示。

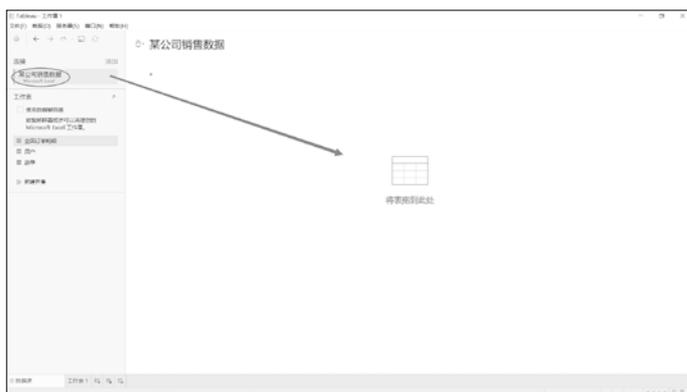


图 3-3 双击 / 拖曳表格

注：在数据量不是很大的情况下，一般选择“实时连接”。



图 3-4 导入数据

(4) 导入之后软件默认显示前 1 000 行数据，用户可以自行修改显示行数，导入的 Excel 示例数据如图 3-5 所示。



排序字段	数据源顺序	显示别名	显示隐藏字段	1,000	行								
订单号	订单日期	顾客姓名	订单等级	订单数量	销售额	折扣点	运输方式	利润额	单价	运输成本	区域	省份	城市
3	2014/10/13	李鹏晨	低级	6	261.54	0.040000	火车	-213.25	38.94	35.000	华北	河北	
6	2016/2/20	王勇民	其它	2	6.00	0.010000	火车	-4.64	2.08	2.550	华南	河南	
32	2015/7/15	姚文文	高级	26	2,808.08	0.070000	火车	1,054.82	107.53	5.810	华南	广东	
32	2015/7/15	姚文文	高级	24	1,761.40	0.090000	大卡	-1,748.56	70.89	89.300	华北	内蒙古	
32	2015/7/15	姚文文	高级	23	160.23	0.040000	火车	-85.13	7.59	5.030	华北	内蒙古	
32	2015/7/15	姚文文	高级	15	140.56	0.040000	火车	-128.38	8.46	8.990	东北	辽宁	
35	2015/10/22	高亮平	其它	30	140.56	0.030000	火车	60.72	9.11	2.250	东北	吉林	
35	2015/10/22	高亮平	其它	14	1,892.85	0.010000	火车	48.99	155.99	8.990	华南	湖北	

图 3-5 导入的 Excel 示例数据

转到工作表，出现如图 3-6 所示界面，这样 Tableau 就连接到了数据源。操作界面的左侧功能区菜单分别有“维度”列表框和“度量”列表框，这是 Tableau 自动识别数据表中的字段后提供的分类，“维度”一般是定性的数据，“度量”一般是定量的数据。

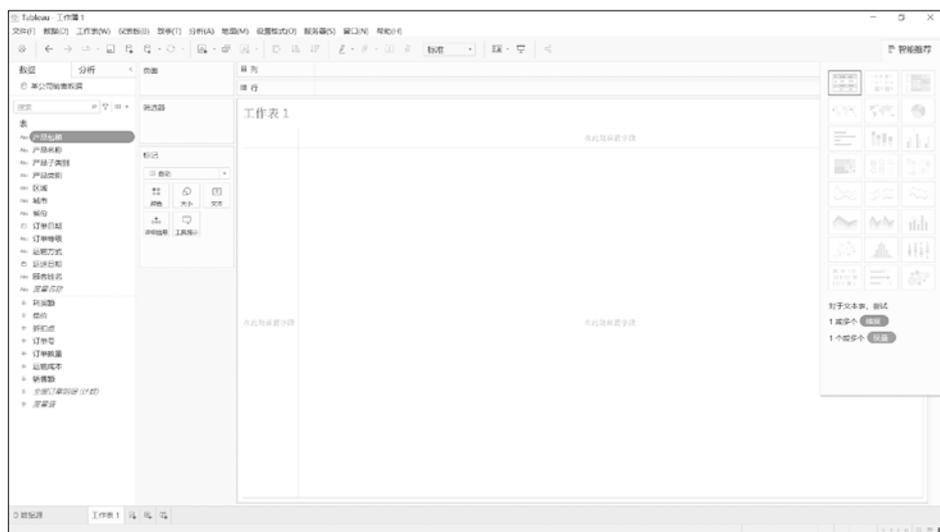


图 3-6 数据窗口

注：有时，某个字段并不是“度量”，但由于它的变量值是定量的数据形式，所以也会出现在“度量”中。例如，表中的“订单号”被分在“度量”中，但其数值不具有实际的量化意义，此时只需将其拖曳到“维度”列表框即可。

二、文本文件

文本文件是指以 ASCII 编码（即文本）方式存储的文件。更确切地说，以英文、数字等字符存储的是 ASCII，以中文字存储的是机内码，文本文件最后一行末尾通常放置的是文件的结束标志。

在开始页面的“连接”窗格中单击“文本文件”选项，如图 3-7 所示。在弹出的对话框中选择要连接的“考试分数.txt”文件，单击“打开”按钮，如图 3-8 所示。



图 3-7 单击“文本文件”选项



图 3-8 选择要连接的文本文件

注：Tableau 默认情况下会自动生成字段名称，但由于选择的文本文件中已经有每个字段的名称，因此这里需要选择“字段名称位于第一行中”选项，如图 3-9 所示。



图 3-9 设置字段名称

三、JSON 文件

JSON 是一种轻量级的数据交换格式，适用于服务器与 JavaScript 等脚本语言之间的数据交互，具有读写更加容易、易于被机器解析和生成、支持 Java 等多种语言的特点。

在开始页面的“连接”窗格中单击“JSON 文件”选项，如图 3-10 所示。在弹出的对话框中选择要连接的“全国城市数据.json”文件，如图 3-11 所示。



图 3-10 单击“JSON 文件”选项

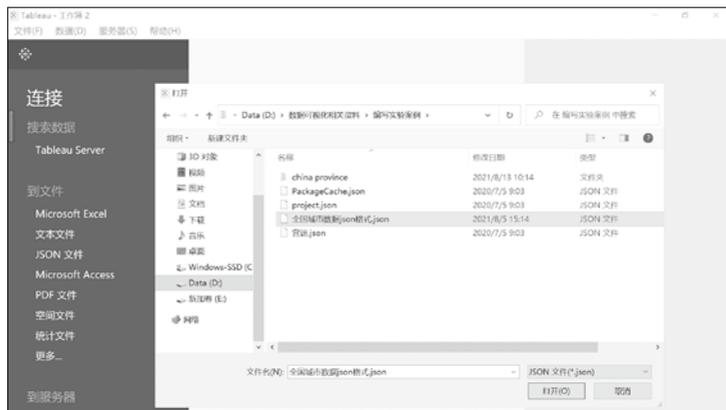


图 3-11 选择要连接的 JSON 文件

单击“打开”按钮，打开“选择架构级别”对话框，其作用是确定用于分析的维度和度量，如图 3-12 所示。



图 3-12 “选择架构级别”对话框

如果架构没有错误，单击“确定”按钮即可完成“全国城市数据.json”数据文件的导入，如图 3-13 所示。

单击“打开”按钮后，可以看到 Access 数据库中的所有表。例如，这里需要分析“净利润/净资产收益率”的相关数据，于是将其拖曳到右侧画布区域即可，如图 3-16 所示。



图 3-16 打开 Access 数据文件

五、PDF 文件

Tableau 可以读取 PDF 文件中的数据。在开始页面的“连接”窗格中单击“PDF 文件”选项，如图 3-17 所示。

选择要连接的“某公司销售数据 .pdf”文件，单击“打开”按钮，如图 3-18 所示。



图 3-17 单击“PDF 文件”选项



图 3-18 选择要连接的 PDF 文件

在“扫描 PDF 文件”对话框中指定要扫描的数据所在页面，Tableau 可以扫描全部页面、单个页面或一定范围内的页面，扫描时会将文件的第 1 页计为数据第 1 页，并忽略文件中使用的页面编号，如图 3-19 所示。

单击“确定”按钮，Tableau 将读取“某公司销售数据 .pdf”文件中的数据，如图 3-20 所示。



图 3-19 “扫描 PDF 文件”对话框



图 3-20 打开 PDF 数据文件

六、空间文件

Tableau 可以读取空间文件中的数据。在连接之前，需要确保以下文件位于同一个文件夹下。

- (1) 对于 ESRI Shapefile 文件：文件夹中必须包含 .shp、.shx 和 .dbf 文件。
- (2) 对于 MapInfo 表：文件夹中必须包含 .tab、.dat、.mapid 或 .mid/.mlf 文件。
- (3) 对于 KML 文件：文件夹中必须包含 .kml 文件，不需要包含其他文件。
- (4) 对于 GeoJSON 文件：文件夹中必须包含 .geojson 文件，不需要包含其他文件。

在开始页面的“连接”窗格中单击“空间文件”选项，如图 3-21 所示，即可在弹出的对话框中选择需要连接的全国各个省份的 ESRI Shapefile 地图文件，这里选择“china province.shp”文件，如图 3-22 所示。



图 3-21 单击“空间文件”选项

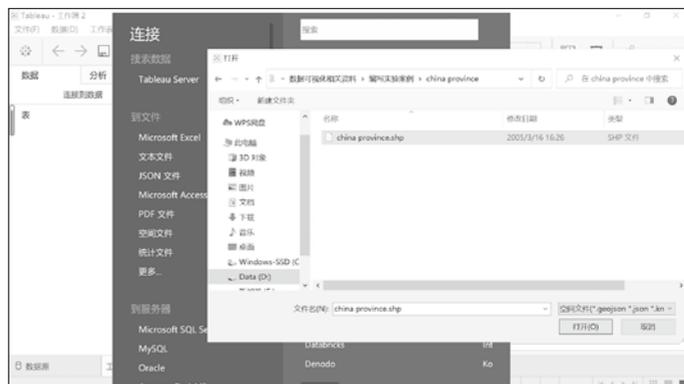


图 3-22 选择要连接的空间文件

单击“打开”按钮，Tableau 将读取“china province.shp”空间文件中的数据，如图 3-23 所示。



图 3-23 打开空间数据文件

七、统计文件

统计文件是指从 SAS、SPSS 和 R 等统计分析软件中导出的数据文件。Tableau 对各类统计分析软件具有很好的兼容性，可以直接导入 SAS (.sas7bdat)、SPSS (.sav) 和 R (.rdata、.rda) 等类型的数据文件。

在开始页面的“连接”窗格中单击“统计文件”选项，如图 3-24 所示。此时若要导入 SPSS 格式的数据文件，可以在弹出的对话框中选择“酒店数据.sav”文件，如图 3-25 所示。



图 3-24 单击“统计文件”选项

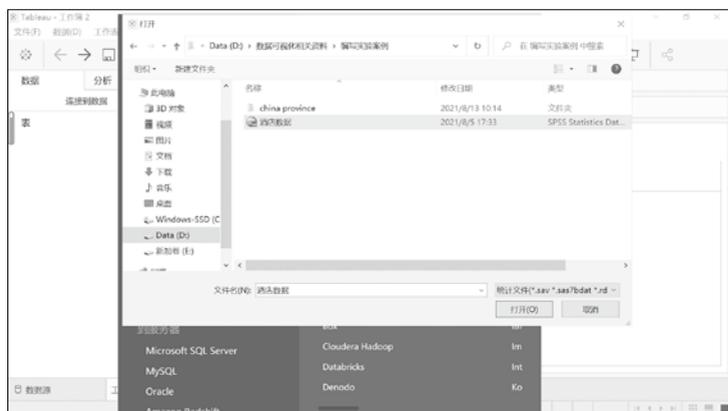


图 3-25 选择要连接的统计文件

单击“打开”按钮，“酒店数据.sav”文件中的数据就被导入 Tableau 中了，如图 3-26 所示。

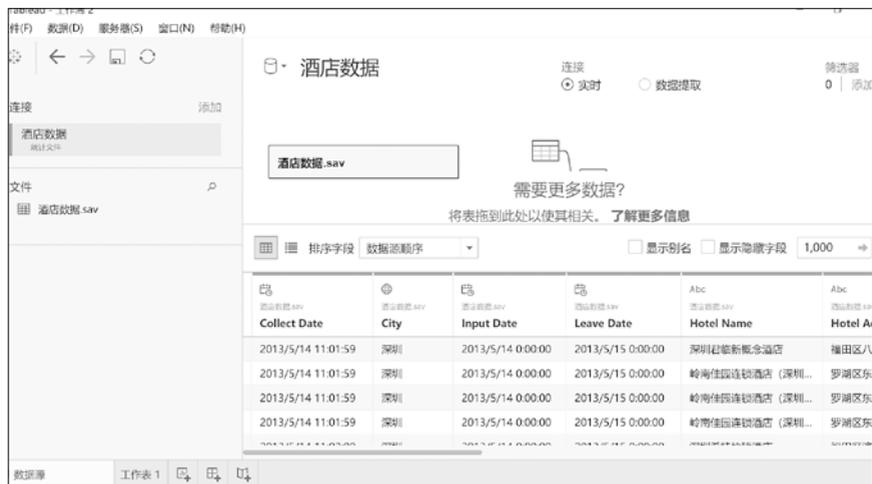


图 3-26 打开统计文件

第二节 连接到数据库

MySQL 是一个典型的关系型数据库管理系统，且开源免费。关系型数据库可以将数据保存在不同的表中，而不是将数据放在一个大“仓库”内，这样可以增加数据的读取速度并提高其灵活性。MySQL 所使用的 SQL 语言是用于访问数据库的最常用的标准化语言。

MySQL 采用双授权政策，其软件系统分为社区版和商业版。在连接 MySQL 数据库之前，首先需要到 MySQL 数据库的官方网站下载对应版本的 Connector ODBC 驱动程序，然后安装。安装过程比较简单，参数配置保持默认即可。安装完成后，在 Tableau 的开始页面的“连接”窗格中单击“MySQL”选项，然后输入数据库的服务器地址、用户名和密码等，单击“登录”按钮即可登录到数据库系统中。

注意：当连接的数据库服务器是 SSL 服务器时，在“连接”窗格中还需要勾选“需要 SSL”选项。如果连接不成功，用户需要检查用户名和密码是否正确。如果确认无误后仍然连接失败，那么使用者就需要联系网络管理员或数据库管理员进行处理。

成功登录服务器后，使用者可选择需要连接的数据库和表，这里将选择需要的数据库，再将数据库中的表拖曳到右侧画布区域中。

使用 Tableau 连接数据库，首先要选择目标数据库的类型，这里选择 MySQL，打开如图 3-27 所示的对话框。

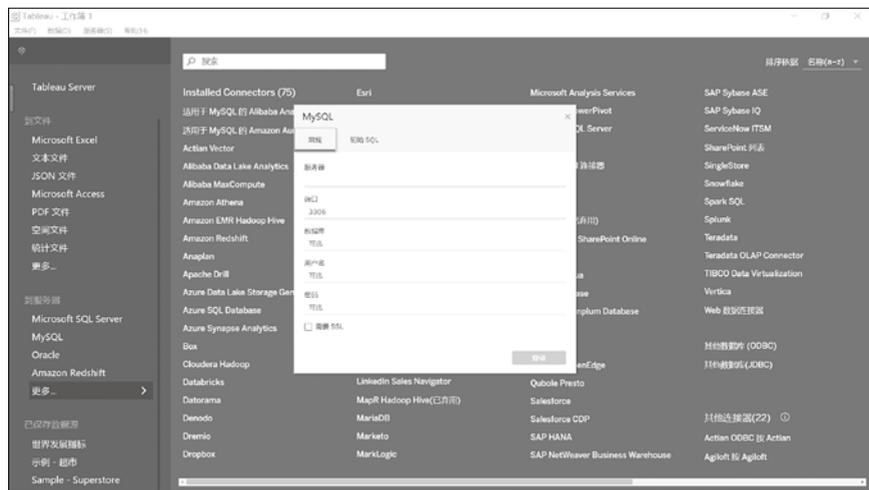


图 3-27 弹出对话框

具体操作步骤如下。

- (1) 输入服务器名称和端口号。
- (2) 输入服务器的用户名和密码。
- (3) 单击“确定”按钮以进行连接测试。
- (4) 在建立连接后，选择服务器上的某个数据库。
- (5) 选择数据库中一个或多个数据表，或者使用 SQL 语言查询特定的数据表。
- (6) 给连接到的数据库定义一个名称，以便其能在 Tableau 中显示。

经过以上步骤之后，单击“确定”按钮，完成连接数据库操作，后续就可以使用数据库中的数据进行分析了。

注：这里连接的是本地服务器，请读者根据各自服务器的情况输入相关信息。若要在步骤（5）中使用 SQL 语句查询特定的字段，则只需选中“特定”选项。

上机操作题

- (1) 使用 Tableau 分别导入 Excel、PDF、Microsoft Access 等类型的数据。
- (2) 使用 Tableau 连接 MySQL 数据库中的表。