

单独录屏
录音

Tableau超越你的业务理解

喜乐君 2021/4

1

业务分析与Tableau实现——2021年版本

- Part 1 可视化分析与业务决策 分析面向决策 | 分析是两步走 | 可视化的价值
- Part 2 从Excel、SQL到Tableau：本质与差异
 - 问题分析方法：结构与过程 | 问题的三种结构 | 聚合的过程 | 聚合函数（+计算）
 - 可视化分析方法：结构 | 六种问题类型 | 空间、标记与意义
 - 字段分类 | 第一字段分类 | 第二字段分类
- Part 3 从问题到图形（上）
 - 初级可视化：图形的构成、选择与基本图形 不同的图形类型
 - 中级可视化：基于标记、坐标轴、参考线和简单计算的可视化改进
 - 高级可视化：{实例}绝对坐标轴到相对坐标轴的转化
- Part 4 样本控制与交互设置
 - 筛选器及相互优先级 | 集 | 参数 | 集动作 | 参数动作 | 分层结构、组、排序
- Part 5 函数使用与问题分析——广义LOD表达式
 - 行级别计算与聚合计算、聚合函数、LOD表达式、表计算
- Part 6 数据准备、数据合并和数据模型

© 2020 喜乐君

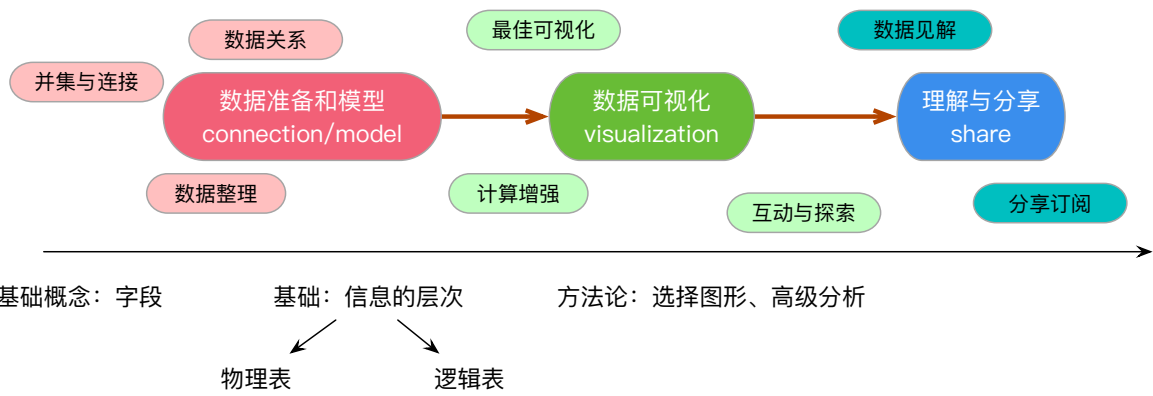
+ a b l e a u 2

2

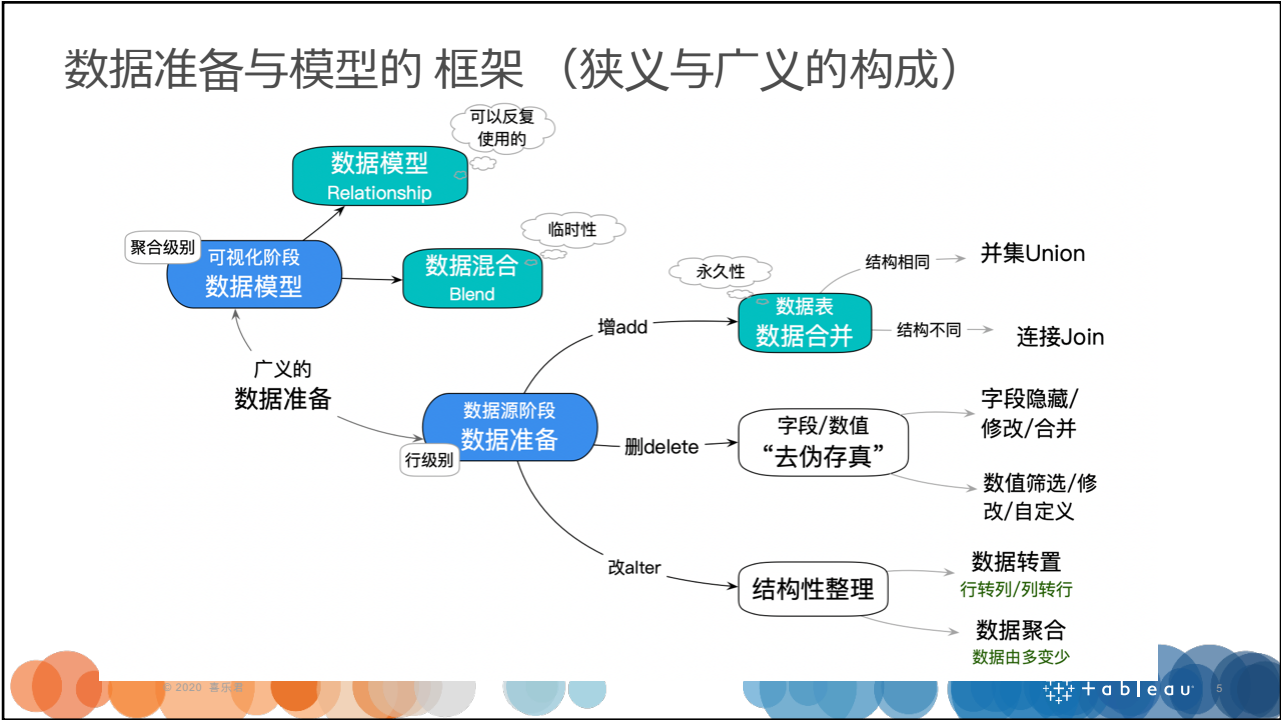
6.1 数据准备的构成内容

3

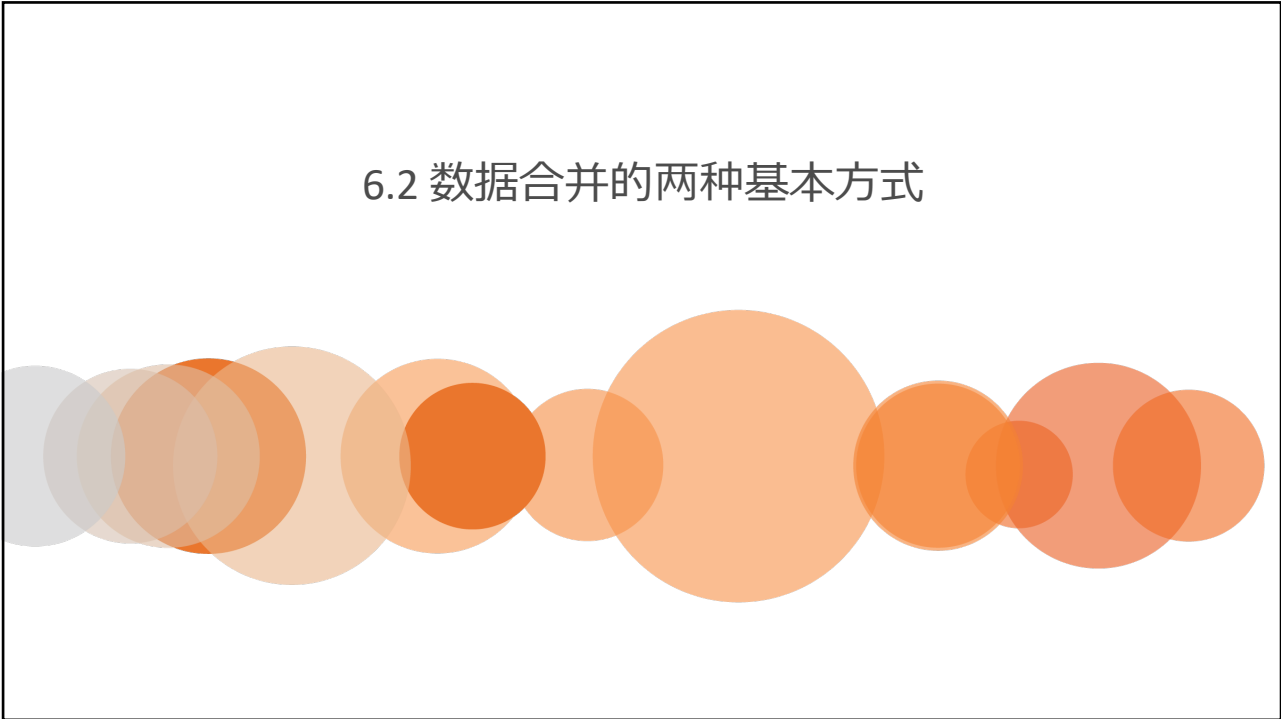
数据准备和模型，是数据可视化的先行



4



5



6

思考过去：Excel中典型的数据合并场景

手动的方式适合临时查询，难以长期使用——不能通往模型化、自动化

- 1、多个sheet 需要，不断的复制、粘贴，合并到一起——不断的向下追加
 - Exg. 超市东区销售表、西区销售表、南区销售表……
- 2、多个sheet，根据某个特定的字段做关联查询——vlookup函数
 - Exg. 从工作表B中，查询每个客户的标签，然后添加到工作表A中
 - Vlookup (B1, sheet B, 3,1)

| | 1 | 2 | 3 | 4 |
|---|------------|----------|----------|-----------|
| 2 | Fruit | Sun | Mon | Tue |
| 3 | Apple | \$ 80.00 | \$260.00 | \$ 155.00 |
| 4 | Banana | \$291.00 | \$200.00 | \$ 264.00 |
| 5 | Peach | \$103.00 | \$166.00 | \$ 288.00 |
| 6 | Lychee | \$116.00 | \$129.00 | \$ 156.00 |
| 7 | Mango | \$122.00 | \$282.00 | \$ 112.00 |
| 8 | Watermelon | \$112.00 | \$155.00 | \$ 274.00 |

7

6.3.1 基于业务模式和数据表结构，多表合并的两种方式

数据并集：相同数据结构（相同业务场景）的数据，行级别，上下追加

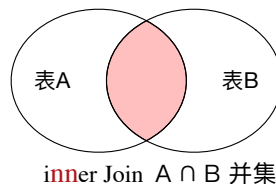
数据连接：按照指定连接字段和连接方式（不同业务场景），行级别，左右相接

并集：多个结构相同的文本文件



可视化分析

内连接



可视化分析

并集和连接的结果，是一个完整的数据源整体

8

6.3.2 并集：业务场景完全相同数据表的数据合并方法

Excel、SQL、Tableau Desktop | Prep Builder

- **Excel**
 - 复制/粘贴——字段的标题内容和次序一致
- **SQL UNION 操作符**
 - UNION 操作符用于合并两个或多个 SELECT 语句的结果集。
 - 请注意，UNION 内部的 SELECT 语句**必须拥有相同数量的列**。列也必须拥有相似的数据类型。同时，每条 SELECT 语句中的**列的顺序必须相同**。
 - SELECT E_Name FROM Employees_China
 - UNION / Union ALL
 - SELECT E_Name FROM Employees_USA

9

6.3.2 并集：业务场景完全相同数据表的数据合并方法

Desktop 演示 | Prep Builder单独

手动并集与 通配符并集



10

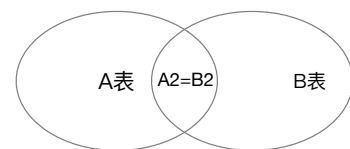
6.3.3 连接Join：业务场景不同的数据表的数据合并

- **Excel：**
- Vlookup函数
- **SQL：**
- Join函数

查找范围

vlookup (A2, A\$1:G\$50 , 2, 0)

If 字段A2=查找范围的首行 返回查找范围的第二列字段



如果字段A2=B2,
则该行两侧所有数据首尾相连

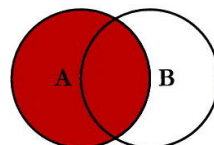
11

6.3.3 连接Join：业务场景不同的数据表的数据合并

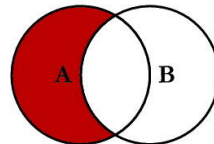
SQL的语法和支持的join范围

SELECT *
FROM Table_A
INNER JOIN Table_B
ON Table_A.id= Table_B. id
Where 条件

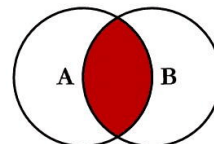
SQL JOINS



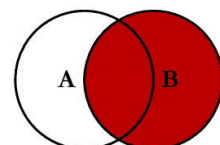
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key



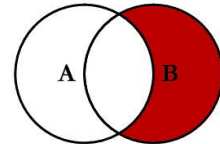
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL



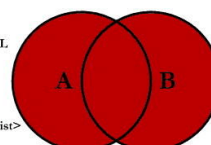
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key



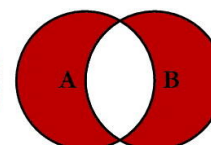
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key



SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL



SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key



SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL

12

Tableau Desktop

13

在数据连接时，使用自定义字段。

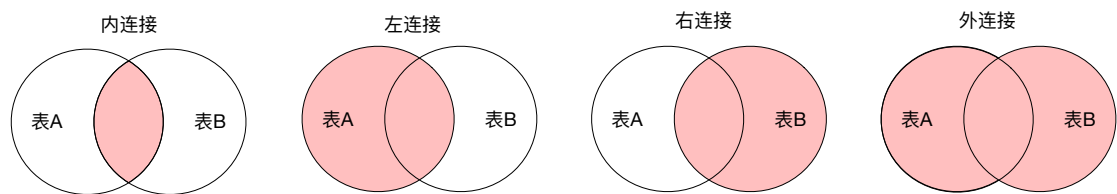
14

Tableau Desktop默认支持的四种基本连接规则

思考：如何实现像SQL中一样的更多连接结果呢？ Left only/ right only/ full outer not inner

By 喜乐君

关系型数据 连接的四种基本类型



* 仅为基本类型，其他类型见后面高级部分

实例：仅仅成功销售出去的订单交易分析——剔除退货交易

解答：join与数据源筛选器的匹配

· Left only/ right only/ full outer not inner

- 订单与商品APAC.csv
- 订单与商品EMEA.csv
- 订单与商品LATAM.csv
- 订单与商品USCA.csv
- 退货商品明细.csv

☐ Left Only交易明细Union+排除退货交易



Practice

- 使用超市的数据表 完成
 - 各地区的销售额数据合并
 - 仅保留正常交易的订单明细
- 使用书店的数据，尝试完成
 - 不同图书的作者、图书、出版社和系列

17

6.3 小结：数据合并与层次理论

关键：如何用层次框架，理解union、join、blend、relationship多种数据合并方式？

在202009“两种层次及其与分析的关系”基础上修改

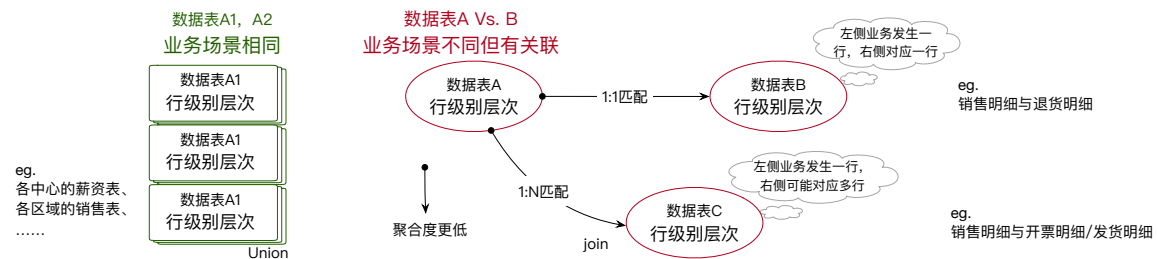
18

多表合并的本质，是业务场景的匹配

数据表层次是聚合的依据，也是数据合并的依据

相同的业务场景，数据表结构一致，使用union并集数据；

不同的业务场景，不同的数据结构，通过行级别字段左右匹配，join数据

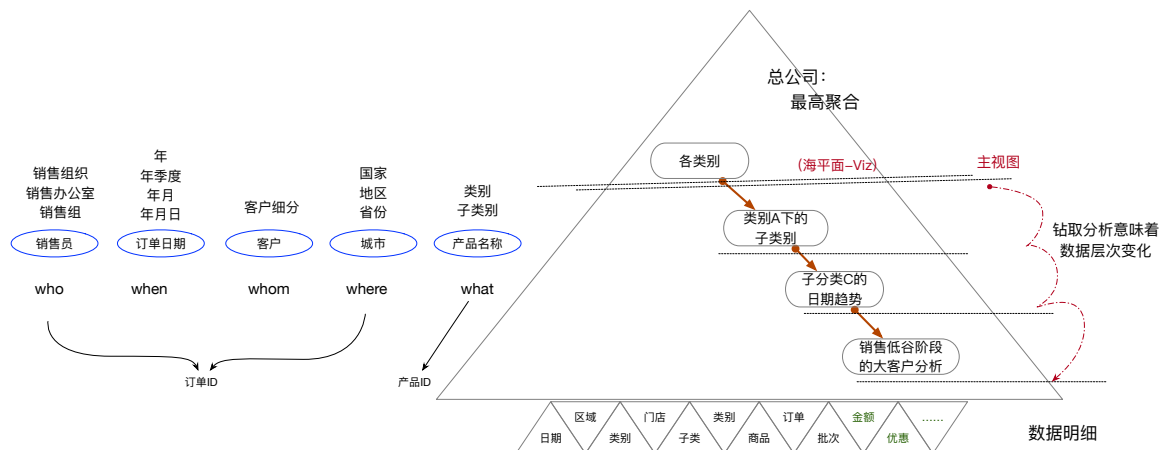


问题：如何识别 数据表的行级别层次？

19

每个数据表都对应真实的业务：识别 数据表行级别层次字段

用一句话描述业务逻辑 → 用最少的字段组合描述数据表（数据表的层次字段）



最少的字段组合描述数据表——可以代表每一行的唯一性（与IT的“主键”略有不同）

20

大数据分析中，层次（LOD）是关键

“最后，我找到了从Excel分析到Tableau大数据分析的根本性差异，即层次。

客观的数据层次用于描述数据结构和颗粒度，**主观的视图层次**用于描述业务问题及其相关性，并通过计算的多种分类把二者融为一体。

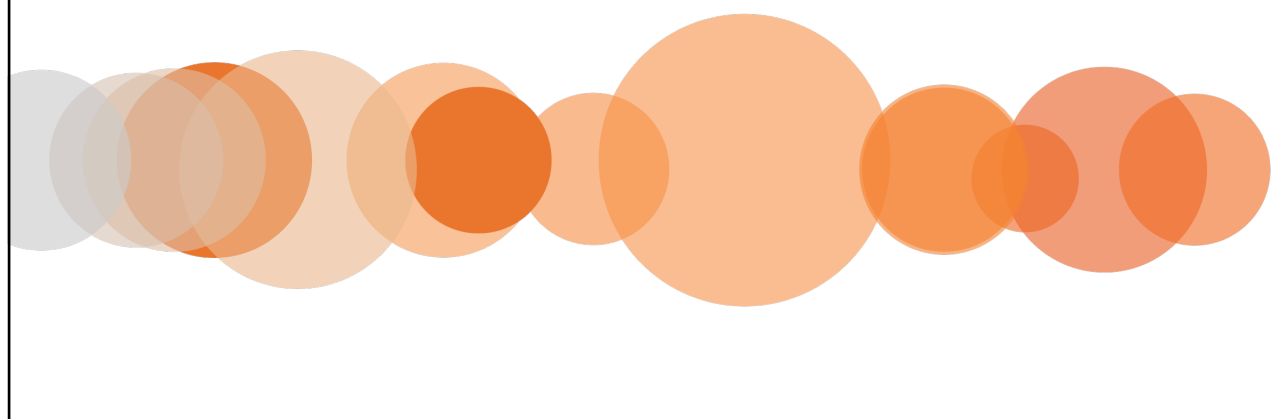
全书都贯穿了“层次分析”的思路，并在高级计算部分得以升华——高级计算的实质就是多层次问题分析。”

· ——《tableau原理与实践》自序



21

6.4 超越明细，走向模型：以聚合为桥梁的匹配（非合并）



22

问题：如何合并 销售明细表，与各月目标KPI？

- 数据表：门店每日的销售明细表（**订单ID*产品ID**）
- 数据表：公司类别、细分的每日销售目标
- 完成：各年、各月的各类别的销售额与目标达成情况（**年月*类别**）

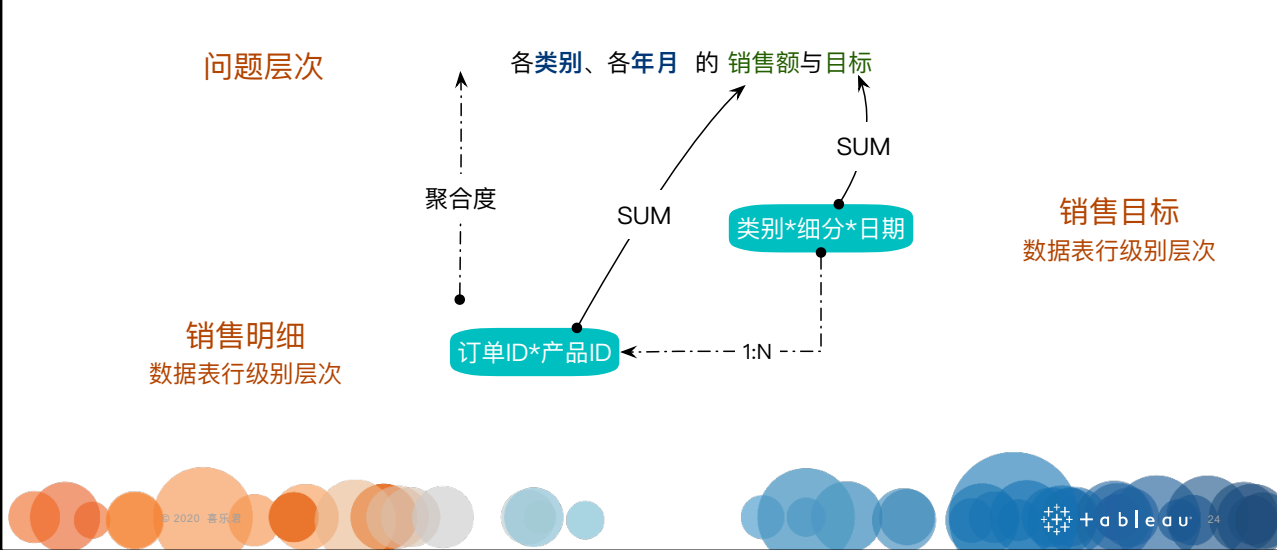
| B | C | G | L | M | N | O | P | | A | B | C | D |
|-----------|----------|----|----------|------|-----|------------|--------|---|------|-----|----------|-------|
| 订单 ID | 订单日期 | 细分 | 产品 ID | 类别 | 子类别 | 产品名称 | 销售额 | 1 | 类别 | 细分 | 订单日期 | 销售目标 |
| GB-2018-1 | 2018/2/7 | 公司 | OFF-ST-1 | 办公用品 | 收纳具 | Fellowes F | 79.2 | 2 | 办公用品 | 公司 | 2016/1/1 | 3470 |
| GB-2018-1 | 2018/2/7 | 公司 | TEC-AC-1 | 技术 | 配件 | SanDisk N | 388.92 | 3 | 家具 | 公司 | 2016/1/1 | 5077 |
| GB-2018-1 | 2018/2/7 | 公司 | OFF-LA-1 | 办公用品 | 标签 | Avery Leg | 35.19 | 4 | 技术 | 公司 | 2016/1/1 | 3599 |
| GB-2018-1 | 2018/2/7 | 公司 | OFF-ST-1 | 办公用品 | 收纳具 | Fellowes F | 50.94 | 5 | 办公用品 | 公司 | 2016/1/3 | 7611 |
| GB-2018-1 | 2018/2/7 | 公司 | TEC-AC-1 | 技术 | 配件 | Memorex M | 307.44 | 6 | 办公用品 | 消费者 | 2016/1/3 | 434 |
| GB-2018-1 | 2018/2/7 | 公司 | OFF-ST-1 | 办公用品 | 收纳具 | Rogers She | 122.4 | 7 | 家具 | 消费者 | 2016/1/3 | 408 |
| GB-2018-1 | 2018/2/7 | 公司 | TEC-PH-1 | 技术 | 电话 | Apple Sign | 413.82 | 8 | 技术 | 公司 | 2016/1/3 | 1950 |
| GB-2016-5 | 2016/8/2 | 公司 | TEC-CO-1 | 技术 | 复印机 | Canon Per | 428.22 | 9 | 技术 | 消费者 | 2016/1/3 | 23622 |

© 2020 喜尔格

tableau

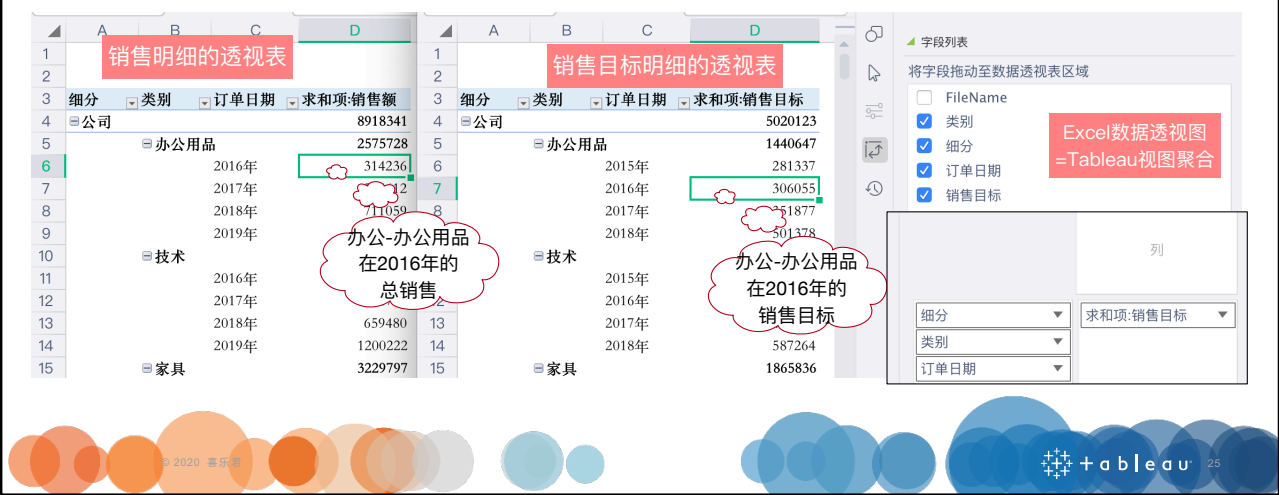
23

如何用层次的方法理解上面的需求和聚合关系



用Excel的数据透视表理解上述过程——聚合级别的匹配

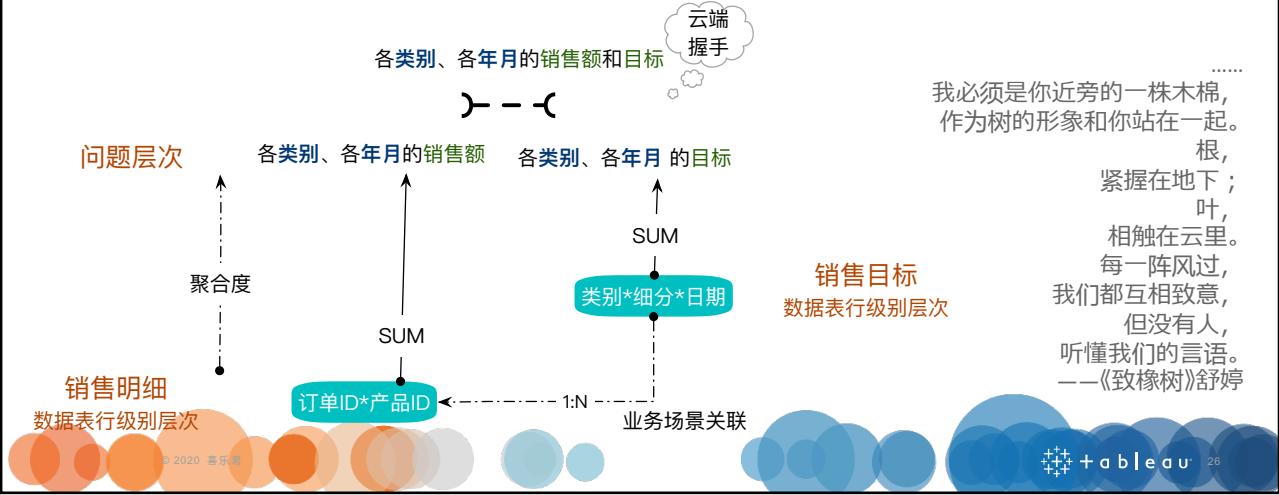
——使用excel的透视图理解数据混合



25

问题：union和join都是行级别的匹配，此时无能为力

- 问题的关键是：如何仅仅在问题所在的层次匹配聚合的数据结构，而无需事先合并？



26

“云端握手”解决方案：先做聚合，再做匹配，而非相反

- 数据混合Data Blend
- 在可视化视图阶段，在指定层次的数据匹配
- 举例：各类别、各年月的销售额总和、销售目标达成
 - Excel 的方法
 - Desktop 的方法
 - SQL的方法

27

实例：Desktop的数据混合

a-切换到“销售目标”数据源

b-相同的维度字段自动匹配

c-双击把度量字段加到视图

销售额、销售目标_从“销售目标”中查询主表各个维度对应的销售目标值
(基于“示例-超市”的销售额聚合与基于“销售目标”的销售目标聚合的聚合连接)

| 细分 | 类别 | 订单日期 年 | 销售额 | 销售目标 |
|-----|------|--------|-----------|---------|
| 公司 | 办公用品 | 2016 | ¥ 284,585 | 306,055 |
| | | 2017 | ¥ 318,610 | 351,877 |
| | | 2018 | ¥ 367,229 | 501,378 |
| | | 2019 | ¥ 514,172 | |
| | 技术 | 2016 | ¥ 312,094 | 414,586 |
| | | 2017 | ¥ 435,011 | 404,567 |
| | | 2018 | ¥ 411,686 | 587,264 |
| | | 2019 | ¥ 605,784 | |
| | 家具 | 2016 | ¥ 379,681 | 419,855 |
| | | 2017 | ¥ 426,088 | 432,802 |
| | | 2018 | ¥ 440,519 | 645,925 |
| | | 2019 | ¥ 657,335 | |
| 消费者 | 办公用品 | 2016 | ¥ 481,471 | 477,197 |
| | | 2017 | ¥ 487,787 | 676,283 |

28

补充：SQL的方法——揉合了join和聚合匹配的特征

- Select s.类别、s.年、s.月、sum(s.销售额), t.类别, t.年, t.月, sum(t.销售目标)
- From superstore s, target t
- Join on s.类别= t.类别 and on s.类别= t.类别
- 注意
 - SQL, 是以join的方式, 在聚合上完成的匹配, 结果是一个物理表
 - 这个过程, 就是Tableau desktop数据混合视图中的查询过程

29

数据混合的优缺点

- 优点：
 - 相对于union和join, 可以在聚合阶段实现数据匹配, 有助于提高性能, 避免数据重复
 - 保持了各个数据表的独立性
- 缺点：
 - 混合关系不能作为模型发布, 因此不能重复使用
 - VS join 的结果可以反复使用
 - 混合关系是“左连接”, 有主表 和辅助表之分, 因此, 辅助表的数据会被排除
 - Join可以通过设置left、right、outer保留所有数据
 - 混合是一次性的
- 看上去, 混合就像是一个临时解决问题的方法.....

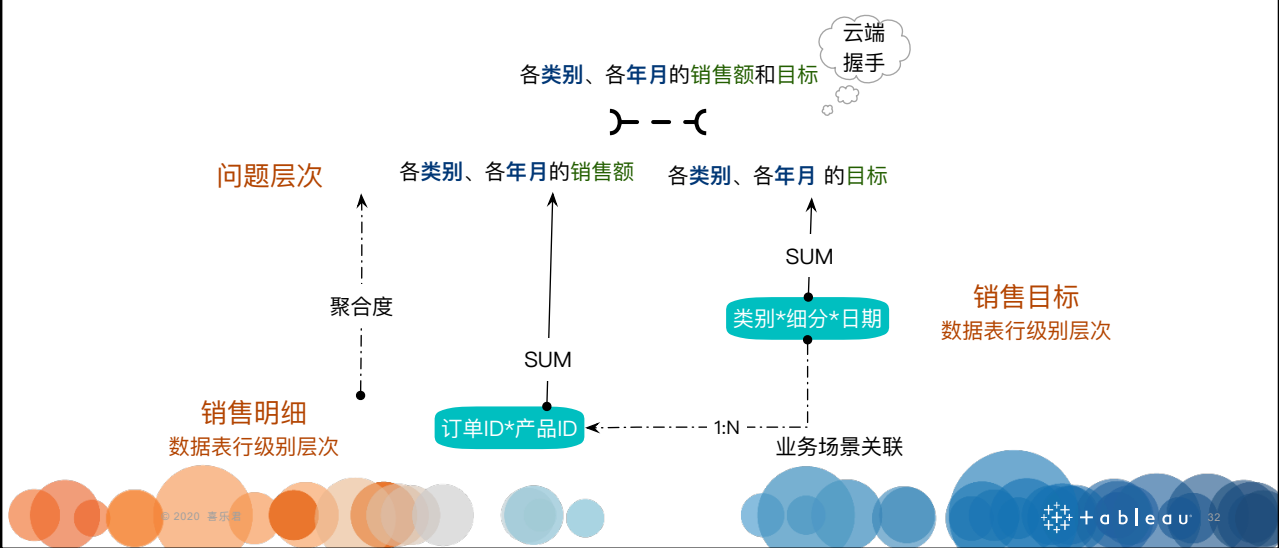
30

6.5 union并集、join 连接、blend混合的区别 ——迈向数据关系

31

6.5.1 从层次角度理解union和join

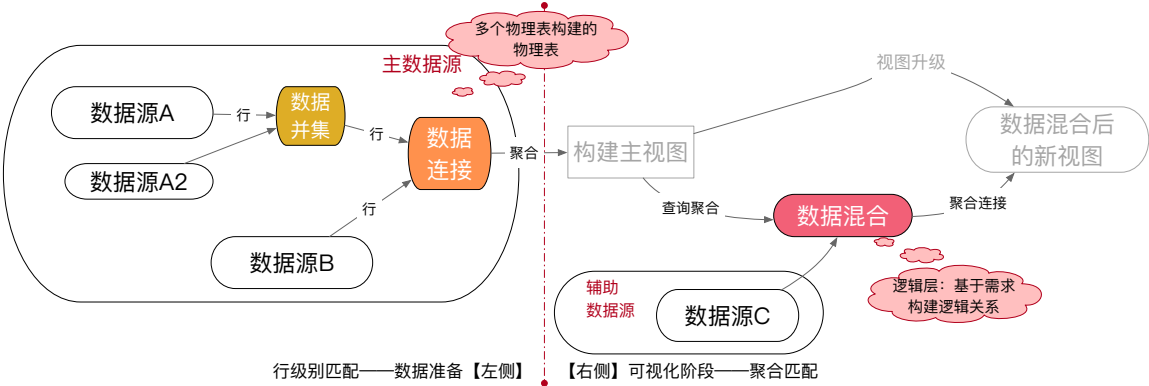
Union和join是在数据表行级别的匹配合并，而 数据混合blend是在问题级别（聚合）的匹配，并无实质合并



32

6.5.2 一朝顿悟：Union、join和blend的关系图

数据连接和并集创建物理表，数据混合是逻辑关系匹配
博客“如何合并数据，顿悟后的究竟指南”<https://yupengwu.com/2019/08/30/tableau-data-combinaton/>



连接的稳定+混合的灵活/性能 == ? ?

6.5.3 取长补短、更上一层楼

“数据关系”

· ——将数据混合的聚合匹配理念，在数据源阶段实现

6.5.4 如何把行级别匹配和聚合匹配整合在一起呢？

关键是 物理层 和 逻辑层的双层结构设计

