

Tableau Prep  
By 喜乐君

ETL  
E 数据连接与轻处理  
T-Cleaning 数据清洗  
T-Transform 数据转化  
L- 保存与使用

Extract-transform-load

ETL是将业务系统的数据经过抽取、清洗转换之后加载到数据仓库的过程。目的是将企业中的分散、零乱、标准不统一的数据整合到一起，为企业的决策提供分析依据。

1 数据连接

- 到文件
  - 本地数据
    - 文本设置
    - 适配符并集
  - 到服务器
    - 服务器

2 设置样本

- 选择数据量
  - 默认示例量
  - 使用所有数据
  - 固定行数: <= 1,000,000
- 抽样方法
  - 快速选择
  - 随机抽样(更全面,但可能会影响性能)

3 数据整理

- 建立筛选器
- 更改「数据类型」
- 移除字段 (Row ID)
- 移除字段
- 字段重命名 (右键/双击)
- 给字段赋予应有的数据角色
  - 更改数据类型
  - 赋予「地理角色」
  - 赋予「URL」特征
  - 赋予「邮箱」特征
- 基于字段的常见操作
  - 字段删除
  - 复制字段 Duplicate (方便备份或再处理,比如调整数据分组)
  - 仅保留选中的字段 Keep only (相当于「移除字段」的反向操作)

字段保健

- 数据类型的选择
- 数据角色的选择

数据手术

- 1 双击-修改数据
- 2 筛选数据
  - 筛选器
  - 计算
  - 固定值
  - 通配符匹配
  - Null值
- 3 以分组方式批量修改数据
- 4 特定字符清理
  - 字母改大写/小写
  - 移除字母/数字/标点符号
  - 移除空格
- 5 拆分值
  - 自动拆分
  - 自定义拆分(Split)函数

计算字段

- 所有的操作本质上都是计算
- 超越上面的快捷方式完成整理
- 全部
- 字符串
- 数字
- 日期
- 类型转换
- 逻辑

T-Transform 数据转化

- 整理后的加工【并集和连接】
  - 并集--数据结构完全一致
  - 连接--不同数据做匹配连接
- 更改数据层次【数据聚合】
  - 数据粒度
  - 按需定制「数据仓库」--生成多个数据粒度的数据
- 更改数据结构【数据转置】
  - 行转列
  - 列转行

L- 保存与使用

- 输出
  - 保存为文件
  - 作为数据源发布
  - 输出格式
    - Tableau数据提取 .hyper
    - Tableau数据提取 .tde
    - 文本文件 .csv
- 添加分支
  - 随时添加分支,可以多次输出
  - 随时在Desktop中预览