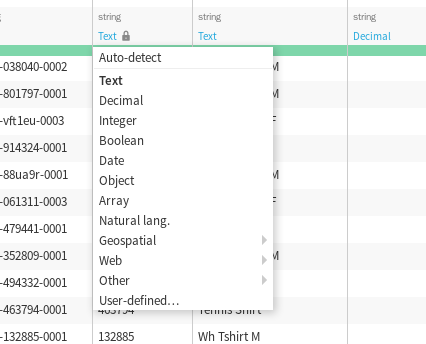
参照：

<https://www.dataiku.com/learn/guide/tutorials/basics.html>

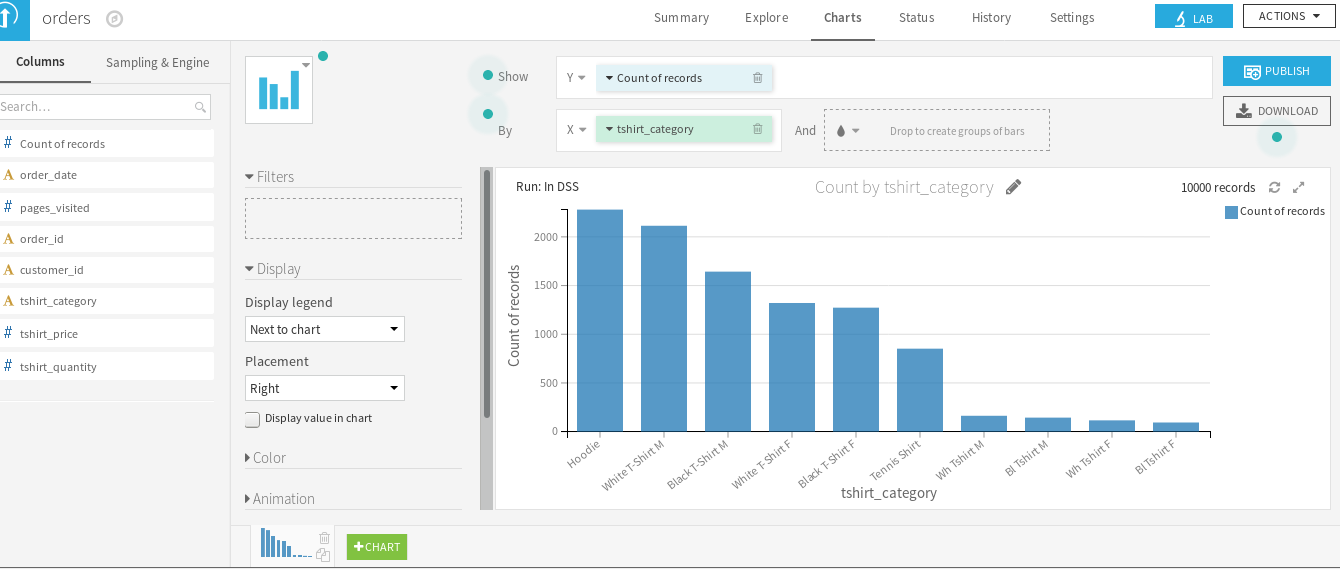
前提：

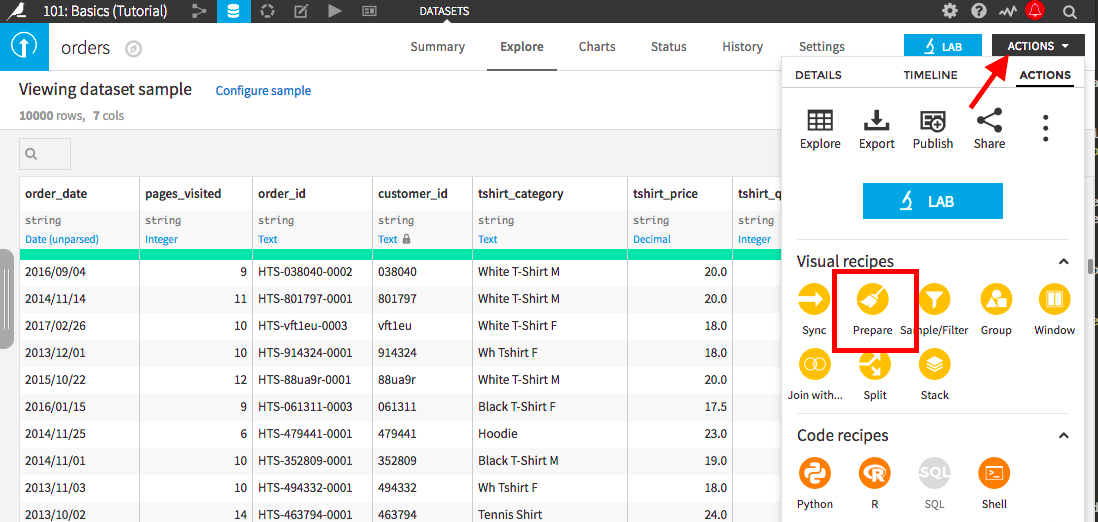
安装好dataiku

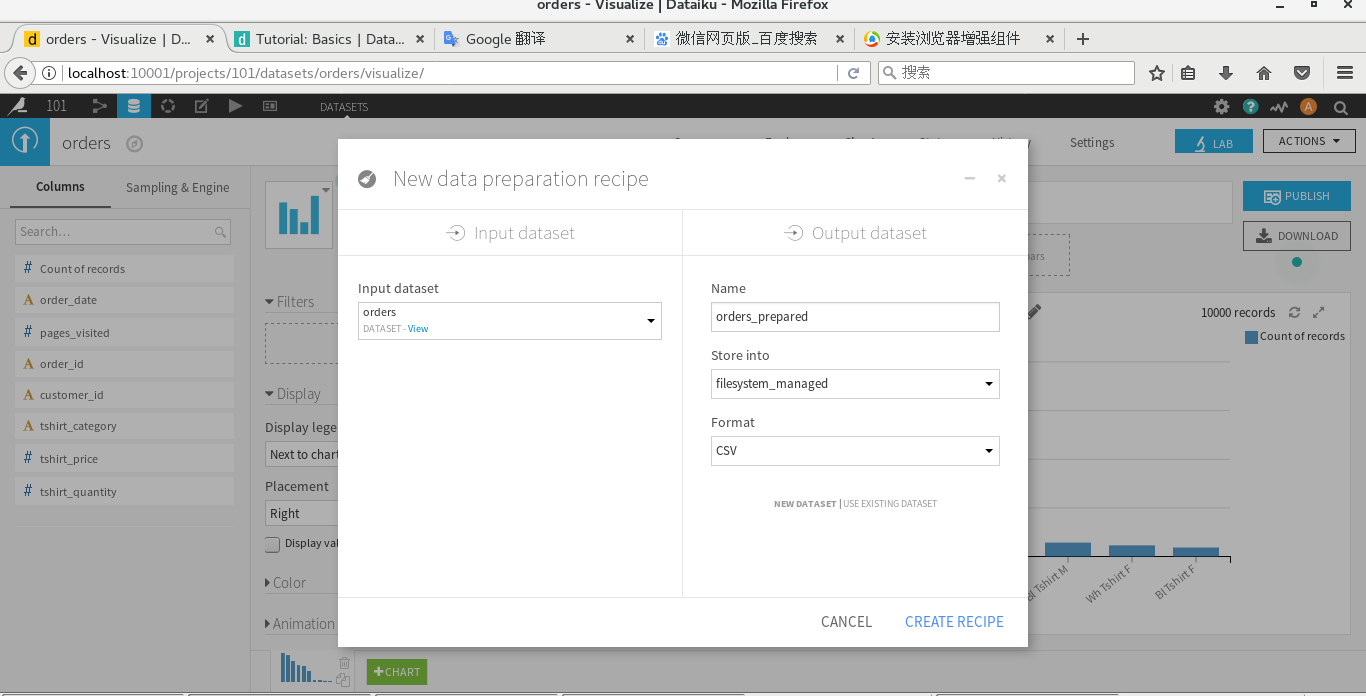
1，**调整列的类型**



2.点击chart,把一个左边的列值拖到Y，一个拖到X，可以查看Y随X的变化

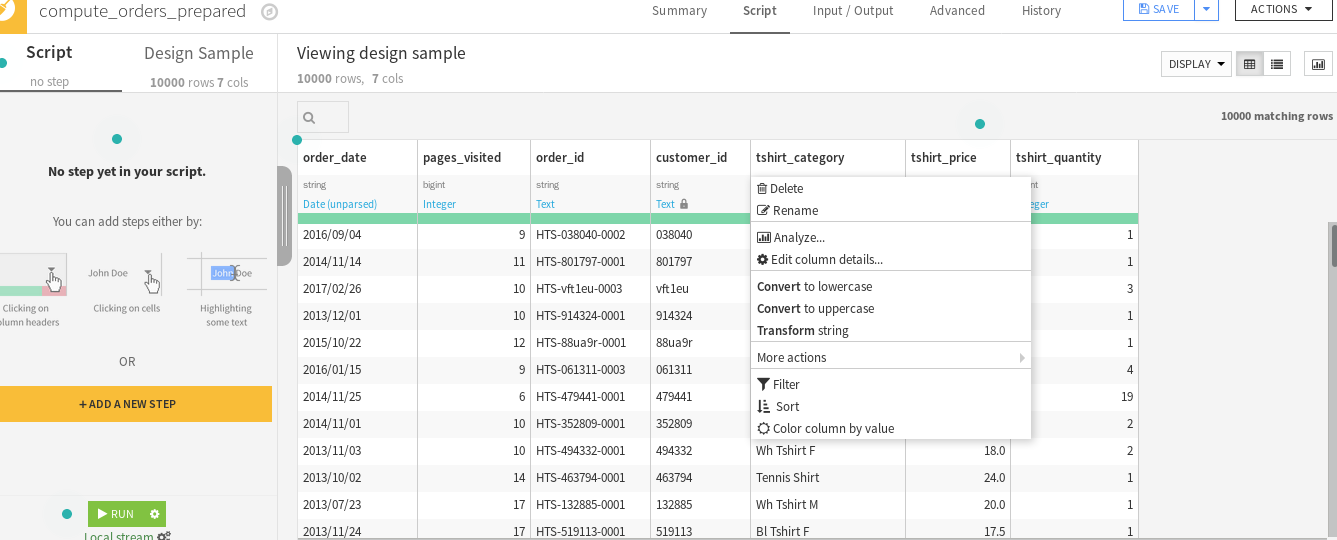




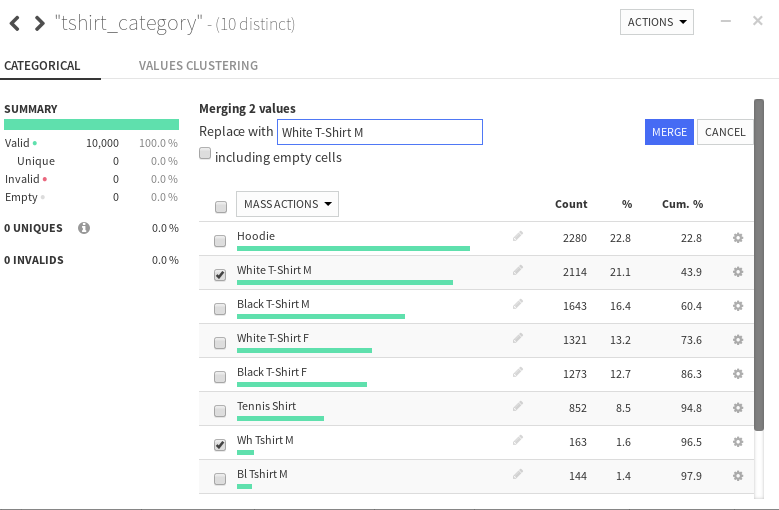


点击create RECIDE,点击列名

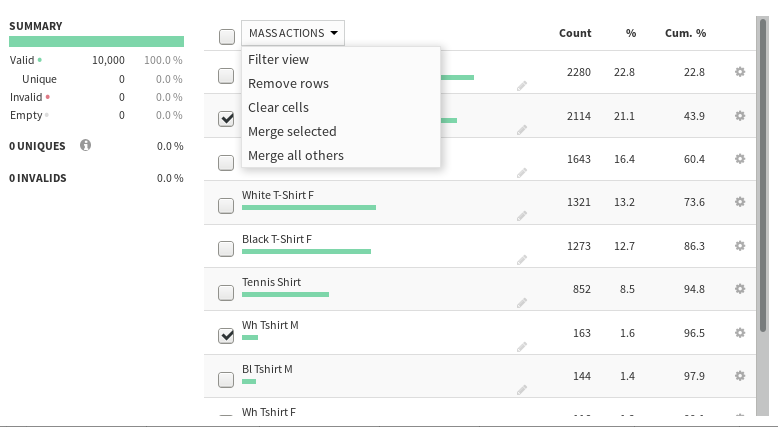
3.**有时候在一个列中用几个值是代表同一个意思（比如红或红色），我们可以把它改为红色。**

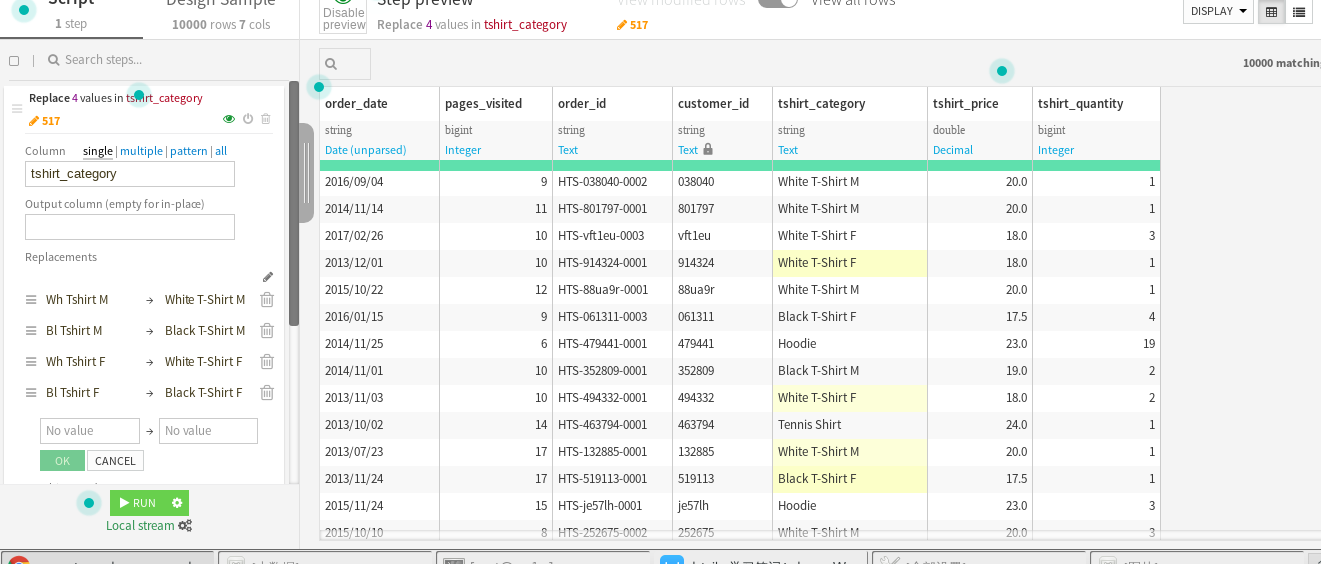


选择analyze



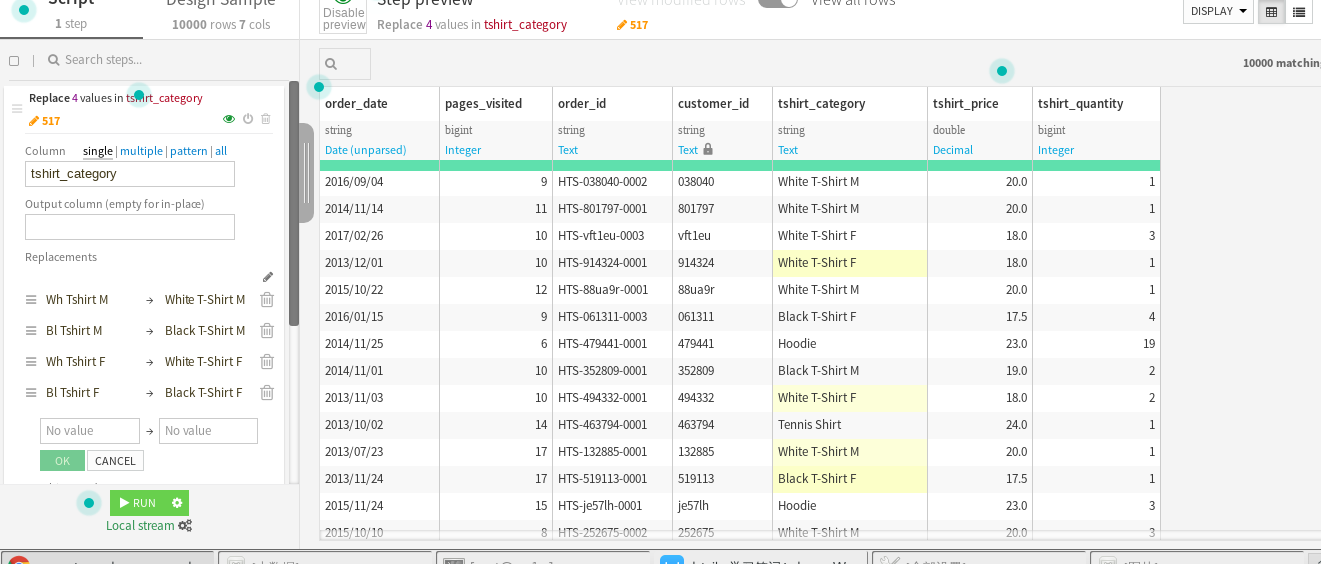
选择merge



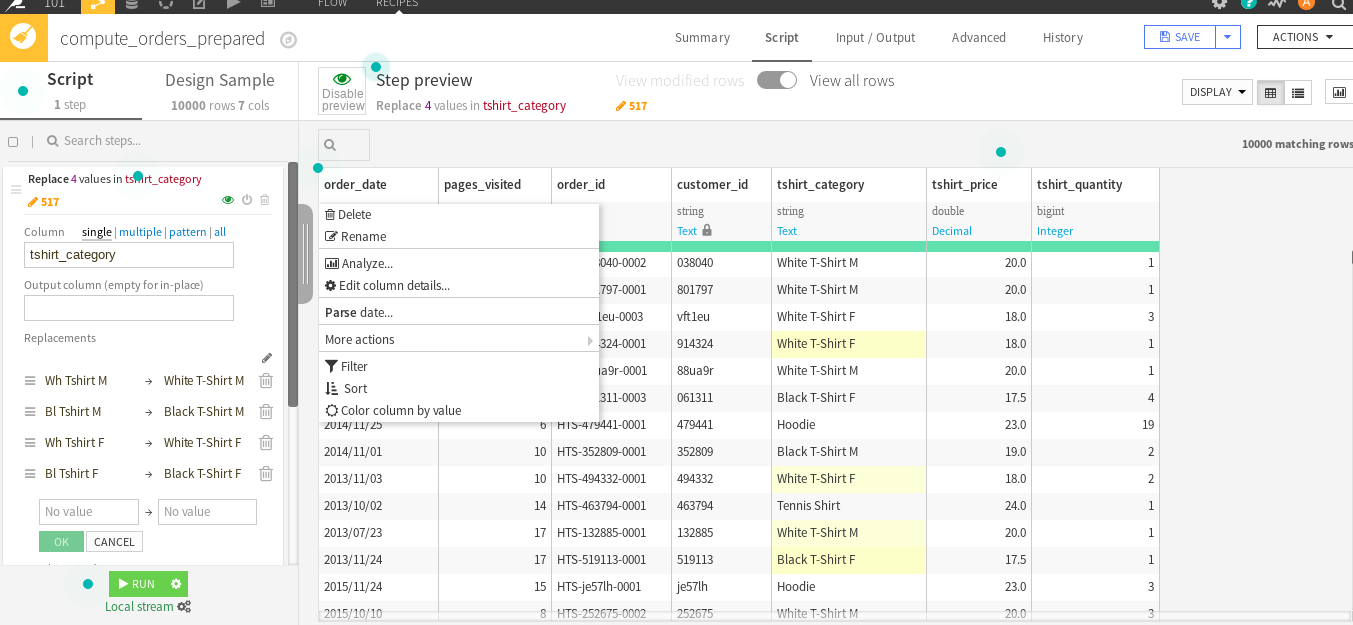


退出后左侧出现预处理的脚本

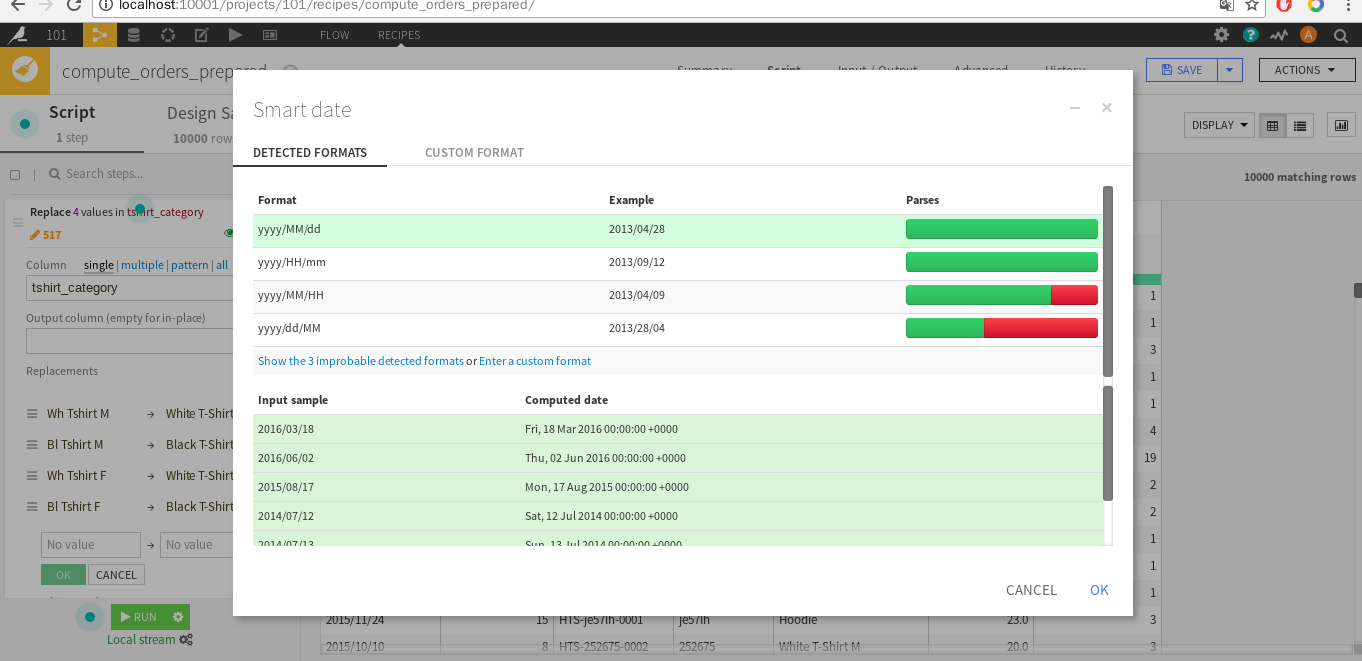
4.**我们一开始从表中输入的时间是无法解析的，所以我们需要设置时间的解析格式**



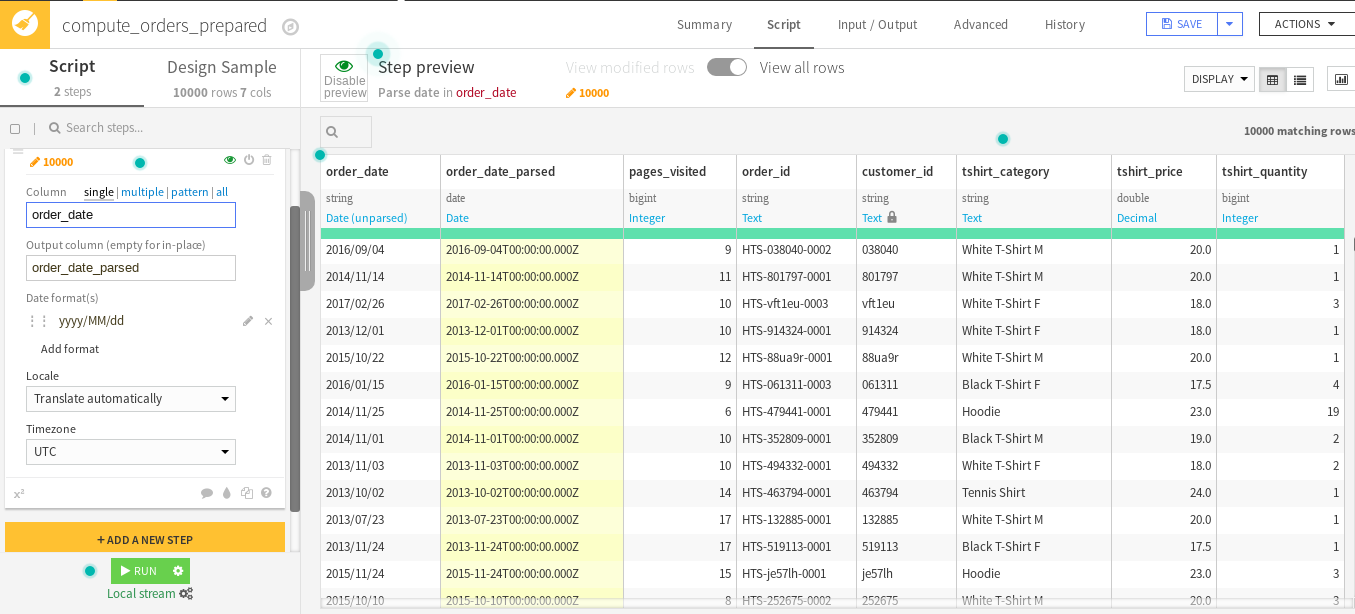
点击order\_data列



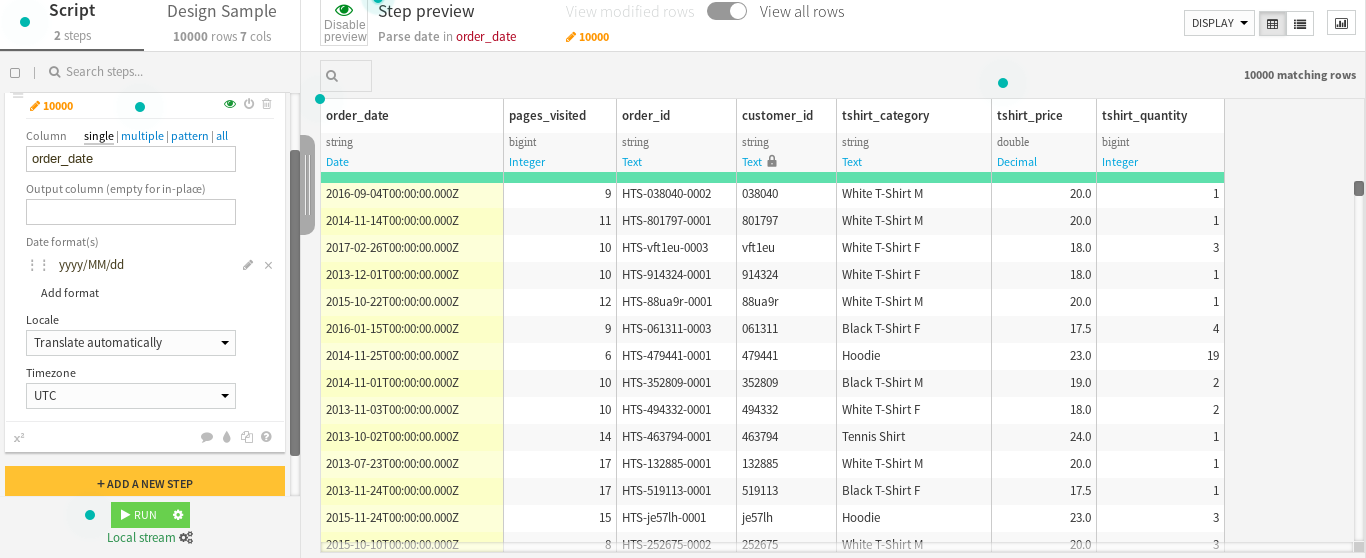
点击parse date



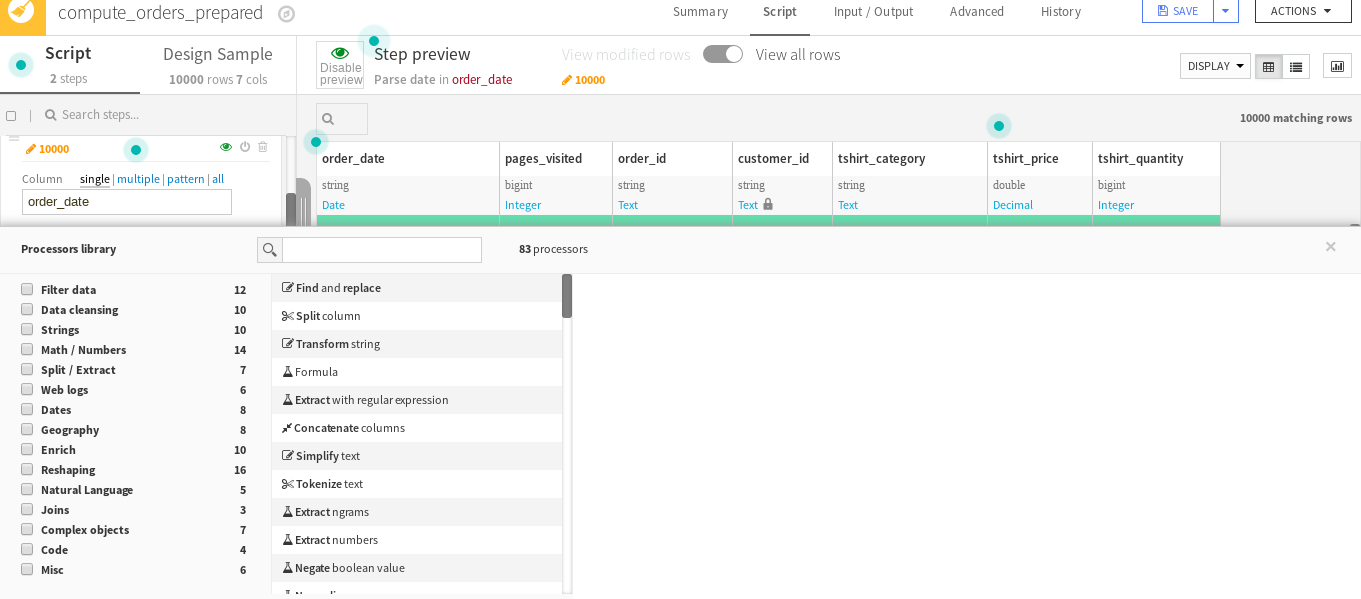
根据需要选择格式，这里选择第一个



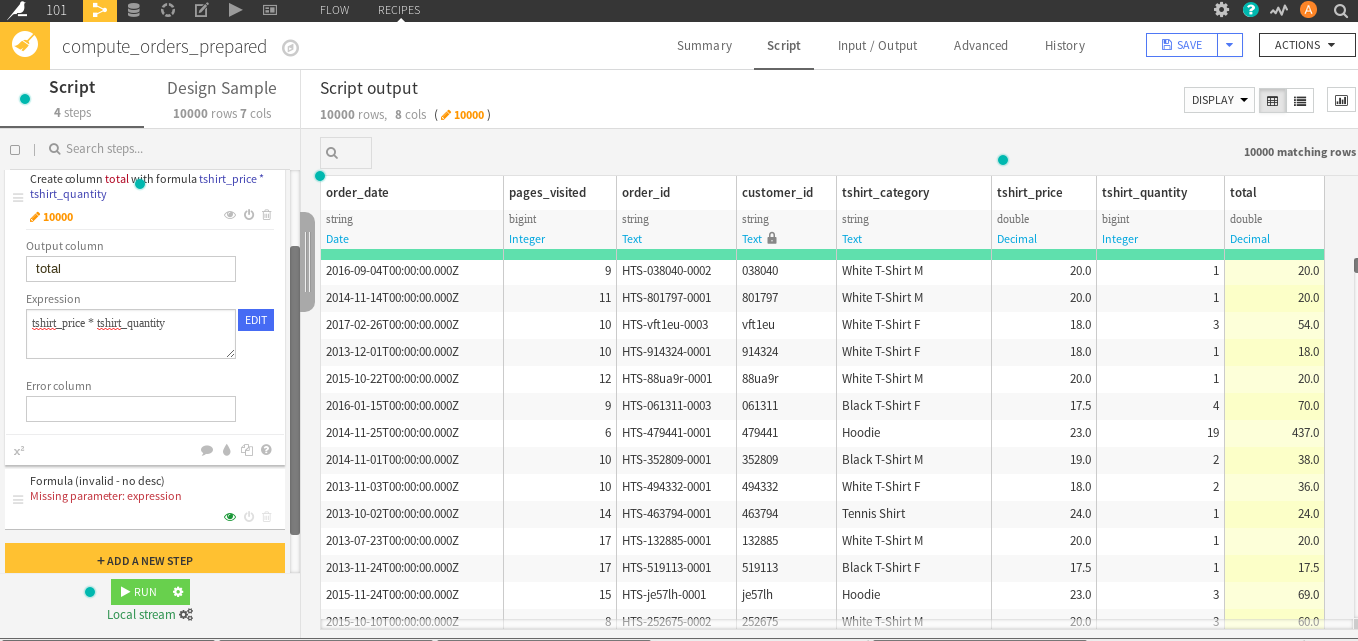
出现了新的一列order\_date\_parsed，左侧的output\_column清空，就会直接修改order\_date



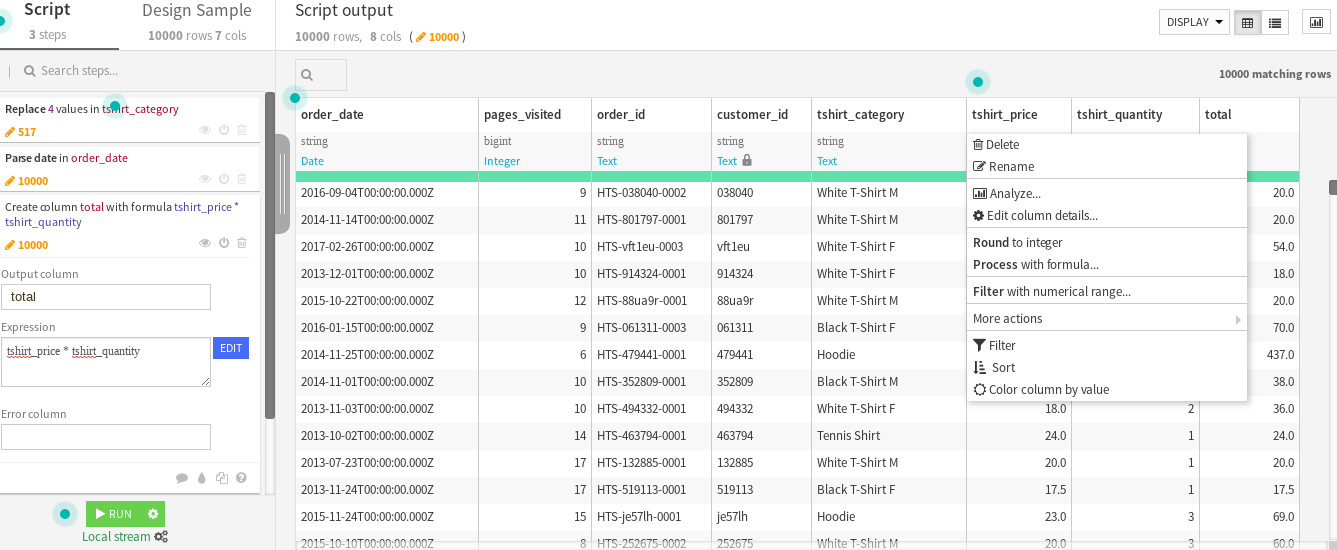
5.**有时候我们需要根据其他列的值得到一个新的列**，点击ADD A NEW STEP



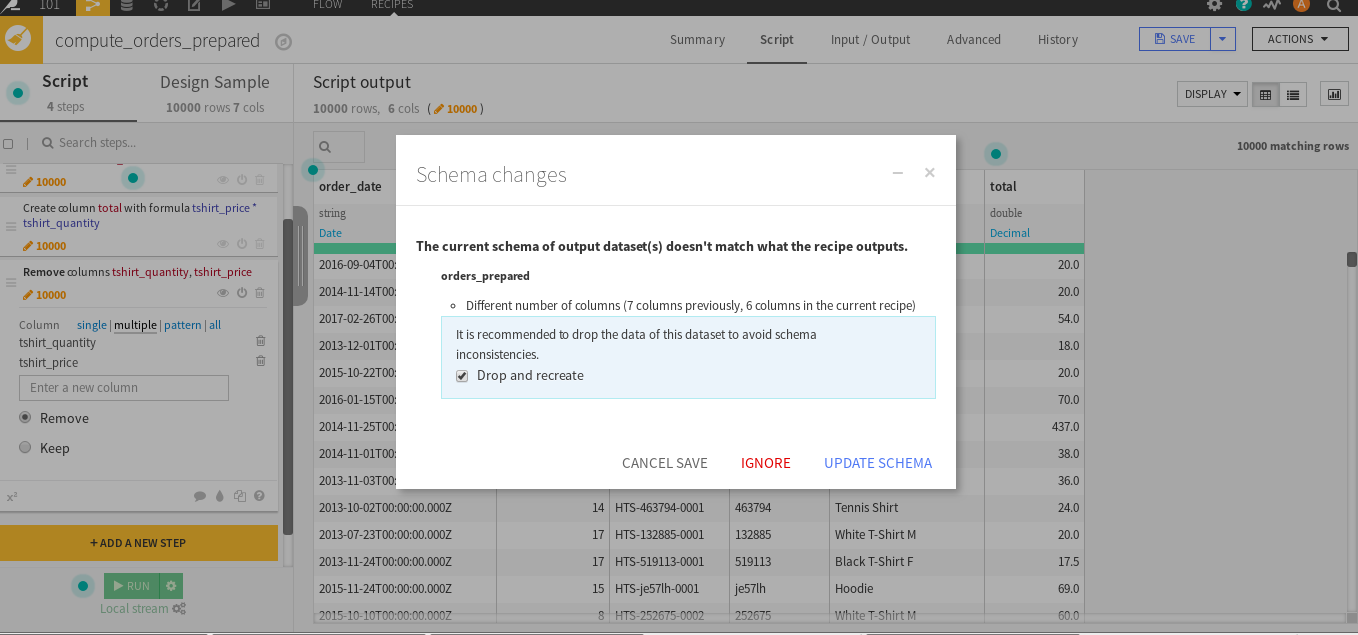
选择formula,在左边修改output column的值，输入有其他列组成的表达式



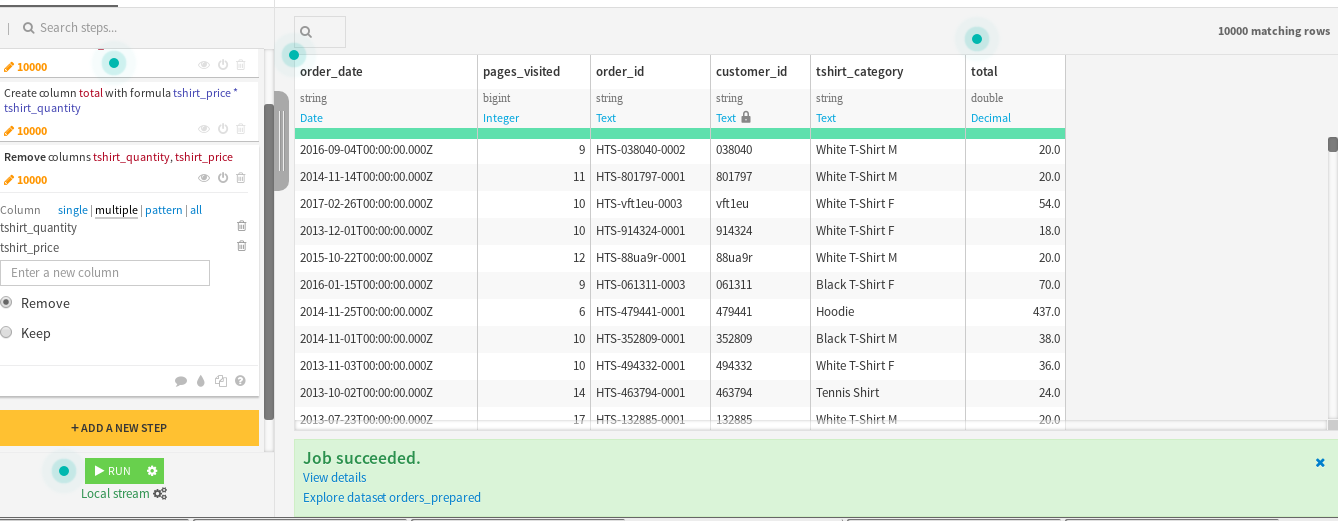
1. **我们很可能需要删除不要的列**，点击列名，点击删除



**7.然而这些修改只是模拟修改，我们需要实际运行**，点击Run

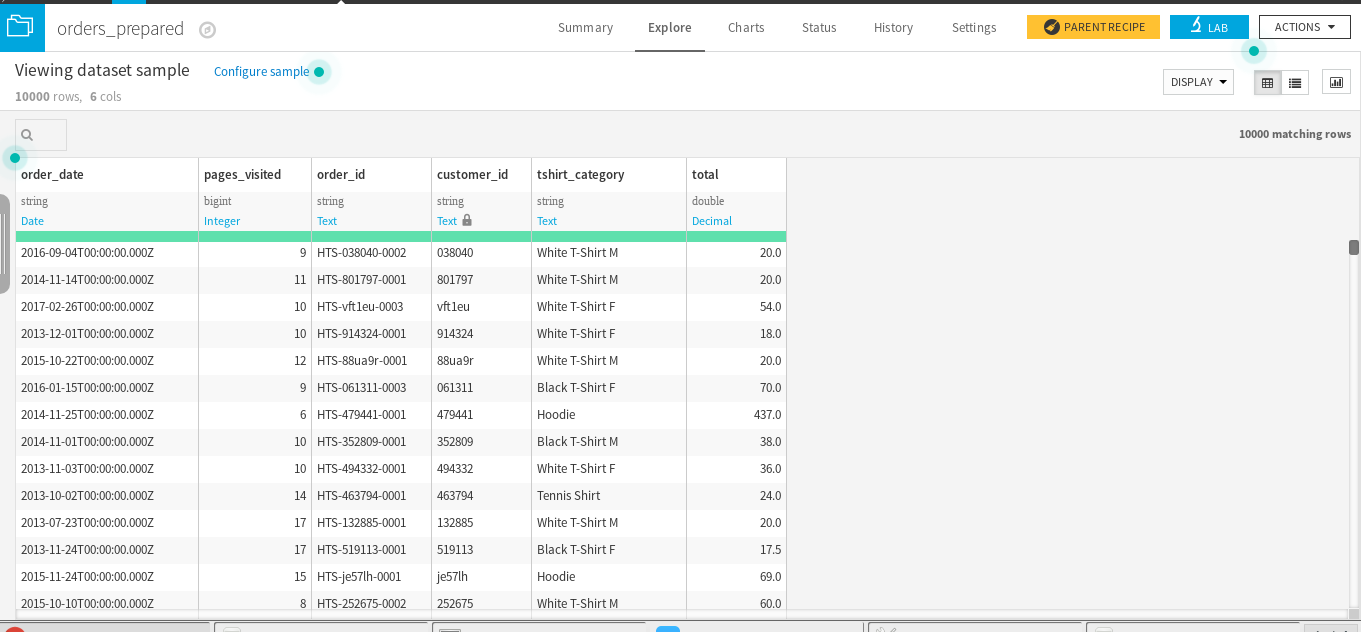


点击update schema

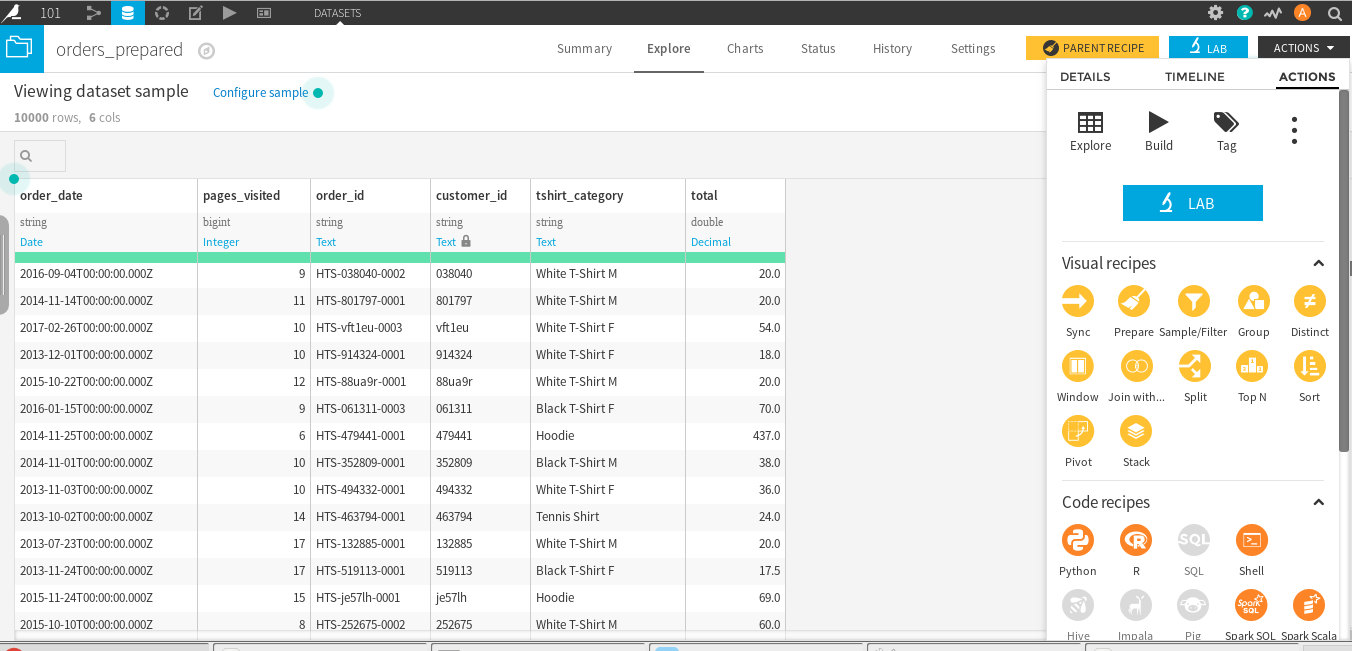
se

点击 explore dataset ...

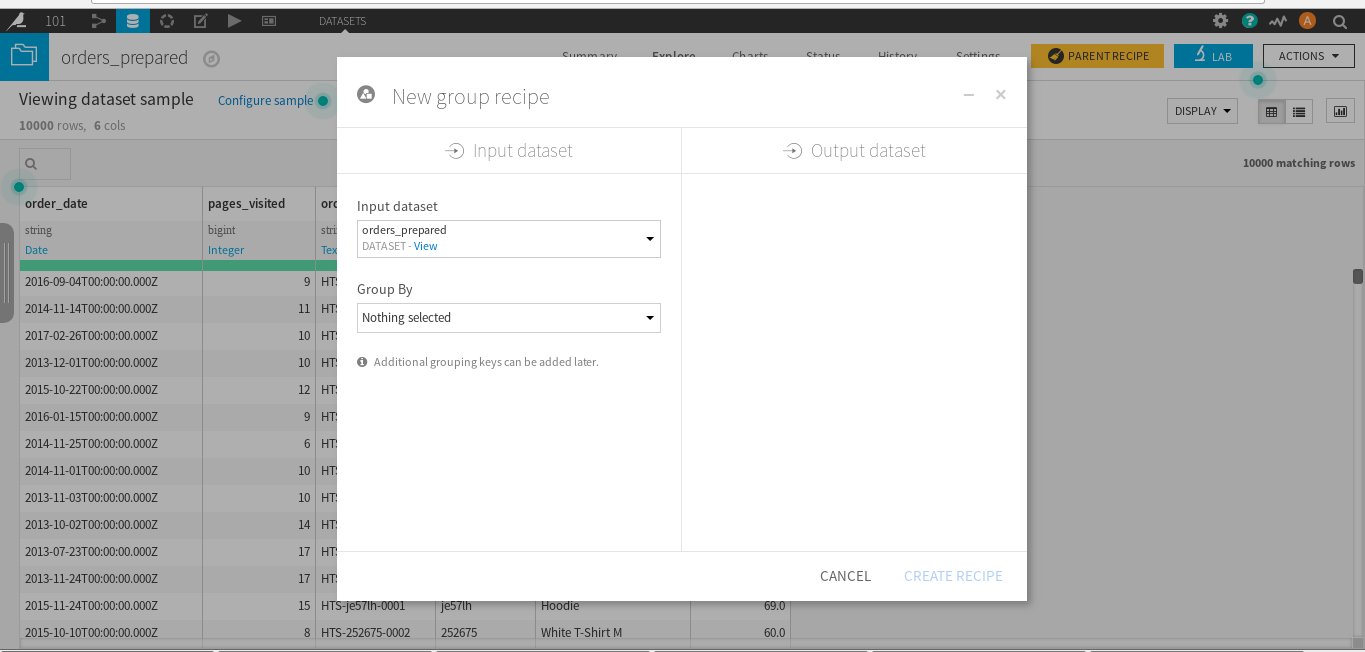
**8.我们有时候需要对数据进行分组**



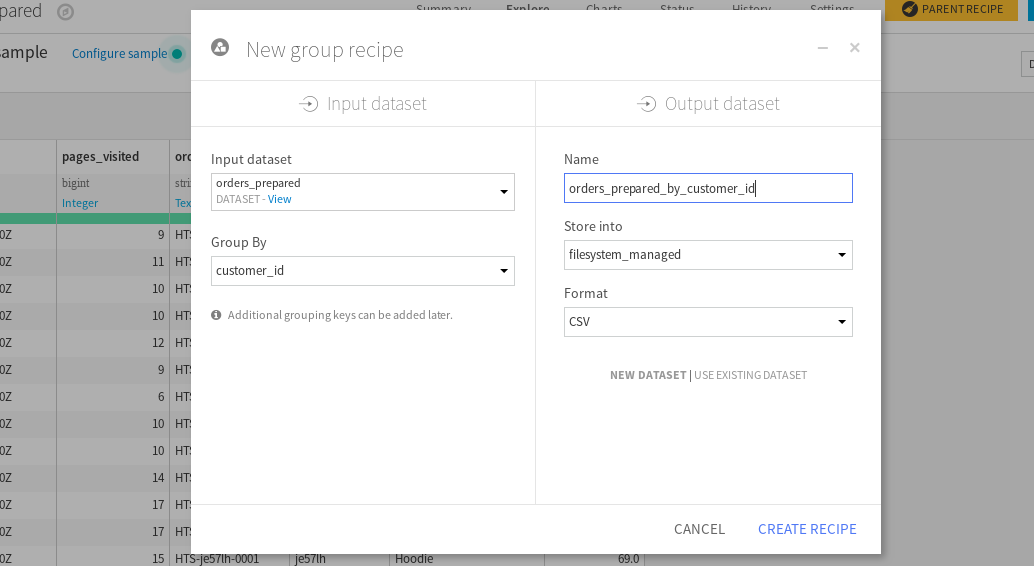
点击actions



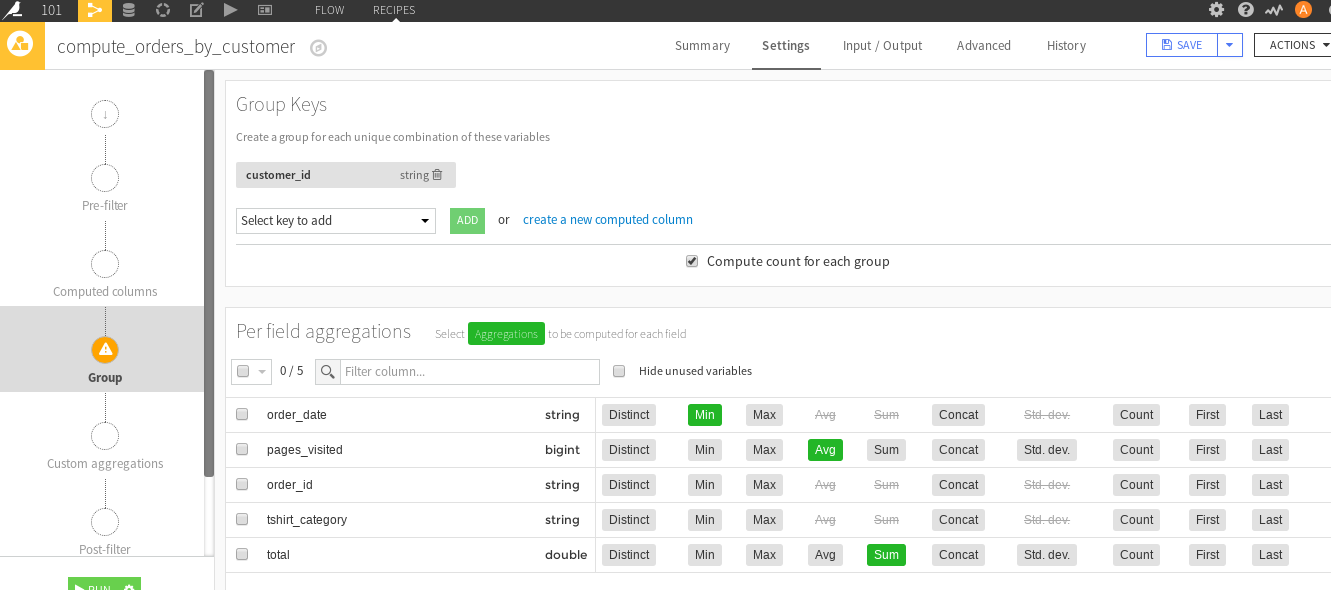
点击group

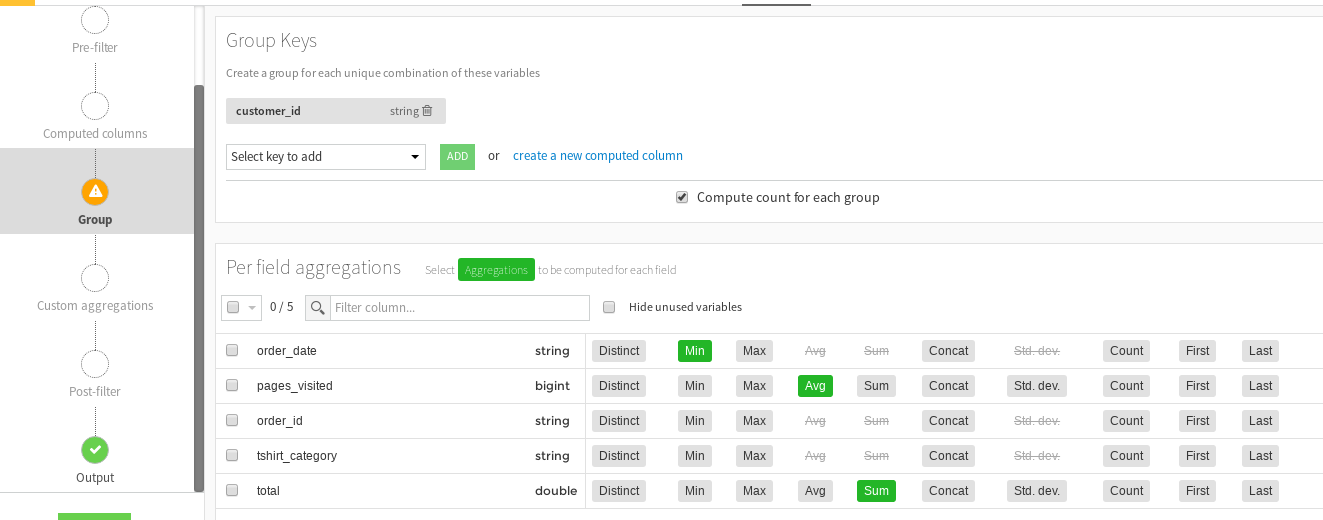


输入分组的依据，并且更改默认的dataset的名字

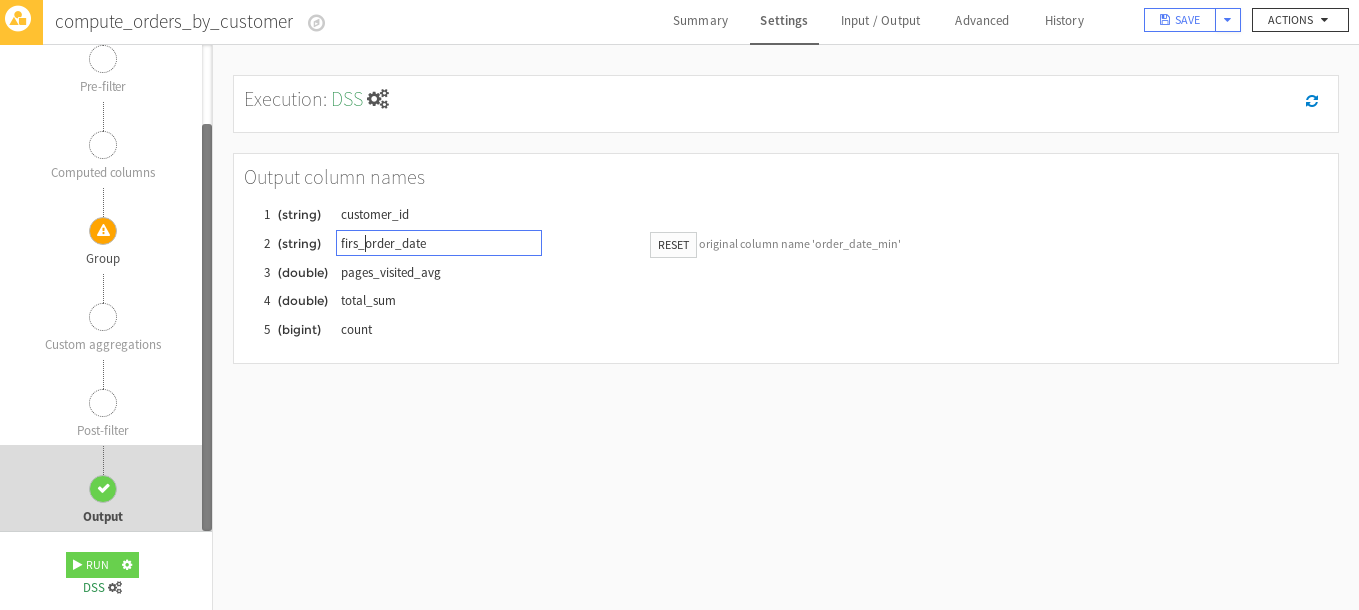


根据需要修改其他属性，比如取最小值

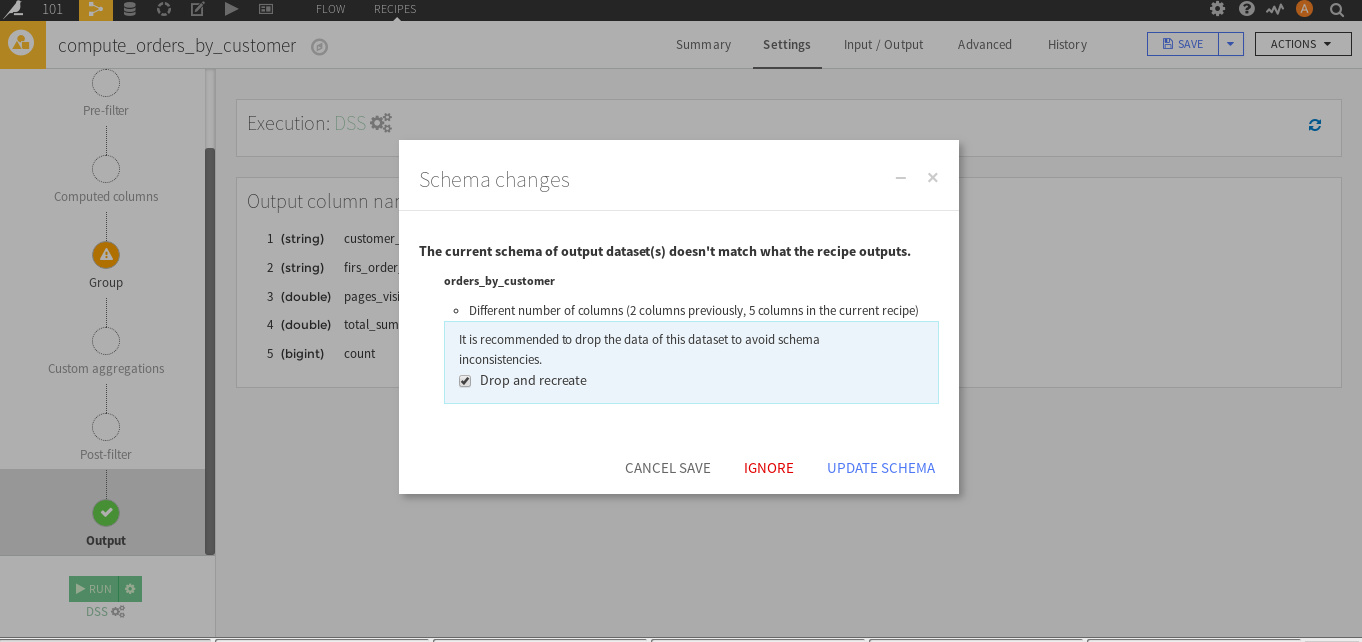




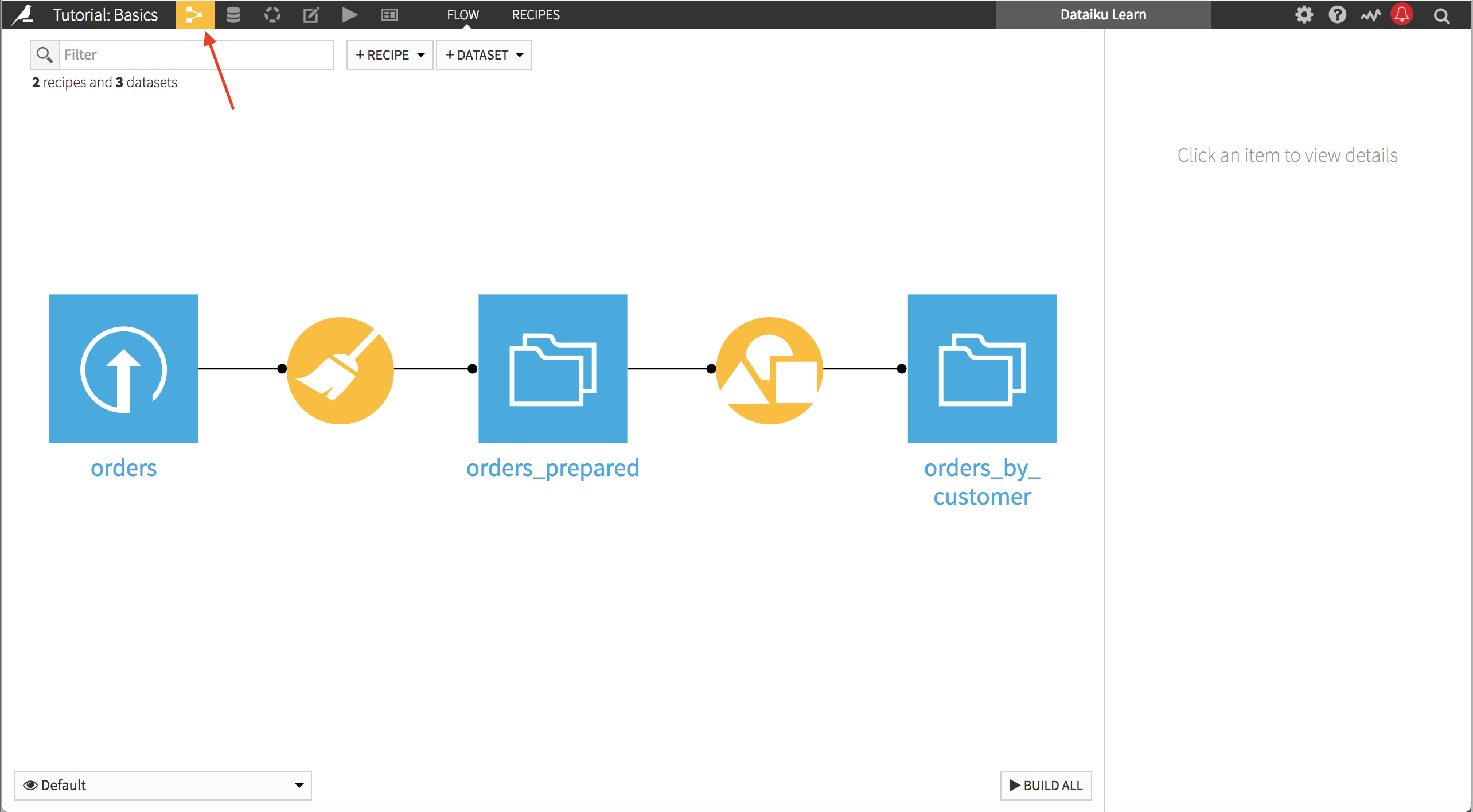
点击output,根据需要修改列名



点击run



其他的和上一个目的类似



Transformation 转型

Prepackaged 预包装

eamlessly 无缝

Collaborate 合作

Perform 演出

wide-ranging 广泛

Distributed 分散式

Infrastructure 基础设施

Various 各个

Cleansing 净化

recode 重新编码

Recipe：代表处理数据的一系列操作，比如可以合并代表相同意思的数据（比如颜色列中有红，红色两种类型的数据，红，红色其实是一个意思，我们可以合并成红色），这个操作可以通过图形化界面实现也可以通过代码实现。

explicitly 明确地

intuitive 直观的

Shortcut 捷径

Registered 注册

Visualizing 可视化

Cancel 取消

Processor 处理器

At this point 在此刻

notoriously difficult 非常困难

Extract 提取

可以调整日期格式，比如yyyy-mm-dd

Compute 计算

Formula 式

Reference 引用

bring up 培养

advanced 高级的

可以删除不要的属性列，求乘积

Recall 回顾

Plus 加

Identical 相同

Storage 存储

Categories 类别

可以把一个属性当成Y，另一个属性当成X,看看他们之间的关系，平均值，或在和等

Overall 总体

Customer 顾客

Spending 开销

Aggregation 聚合

Perspective 透视

Valuable 有价值的

Order 订购

Ultimate 最终

Group 组

Interaction 相互作用

Descriptive 描述的

可以进行根据属性进行分组，对分组后的属性进行处理，比如求平均值，最小值等

Pipeline 管道

Broom 扫把

Preparation 预备

Action按钮有很多功能，包括数据的处理，分组等

基本流程

导入数据->展现数据->对数据进行清理->进行分组等得到有效数据->展现数据

展现数据：

装换数据的类型

默认显示

图表显示

两个属性的X—Y显示

对数据进行处理

可以合并代表相同意思的数据（比如颜色列中有红，红色两种类型的数据，红，红色其实是一个意思，我们可以合并成红色）

删除某一列

增加某一列（取两个列的和）

进行分组等得到有效数据

对分组后的属性进行处理，比如求平均值，最小值等