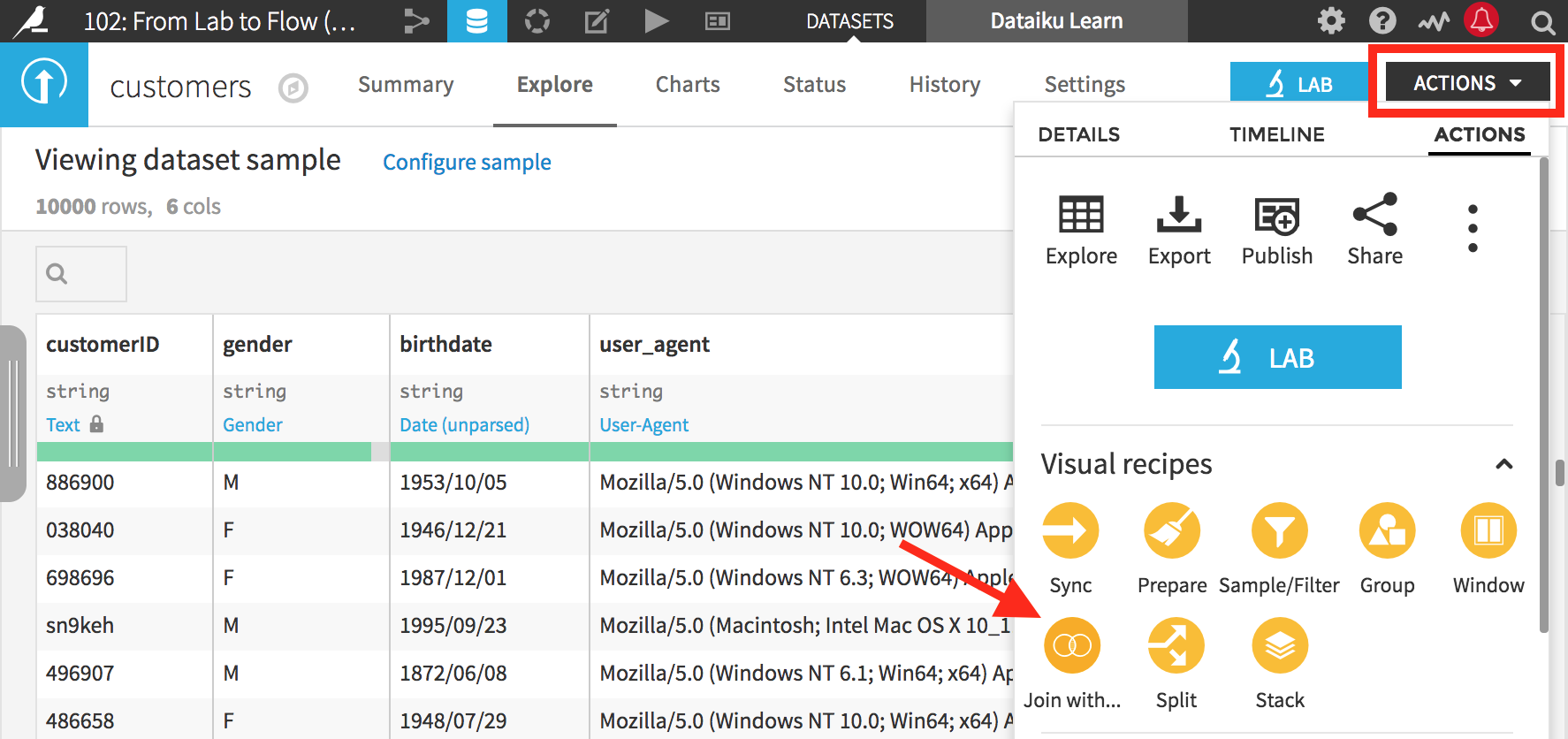
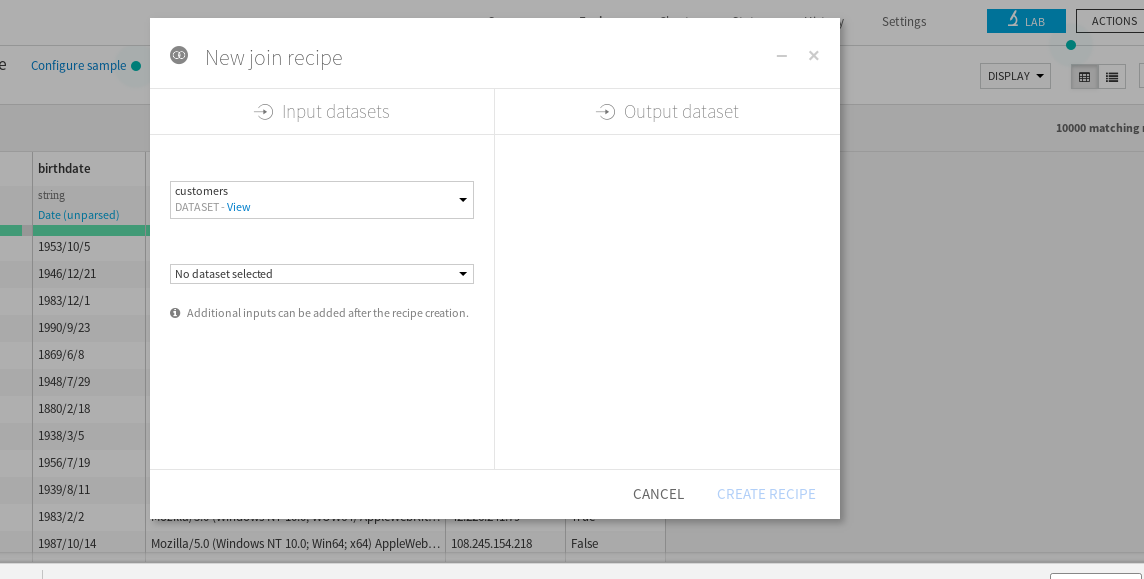
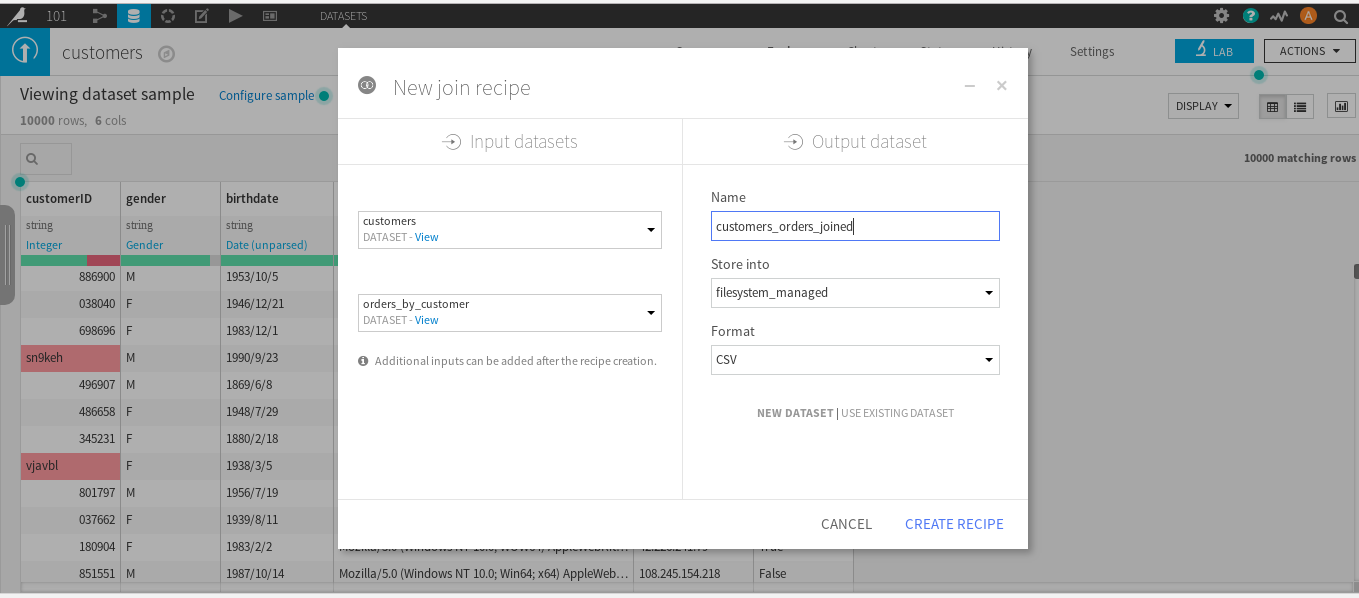
**1.我们有时候需要合并两张dataset**



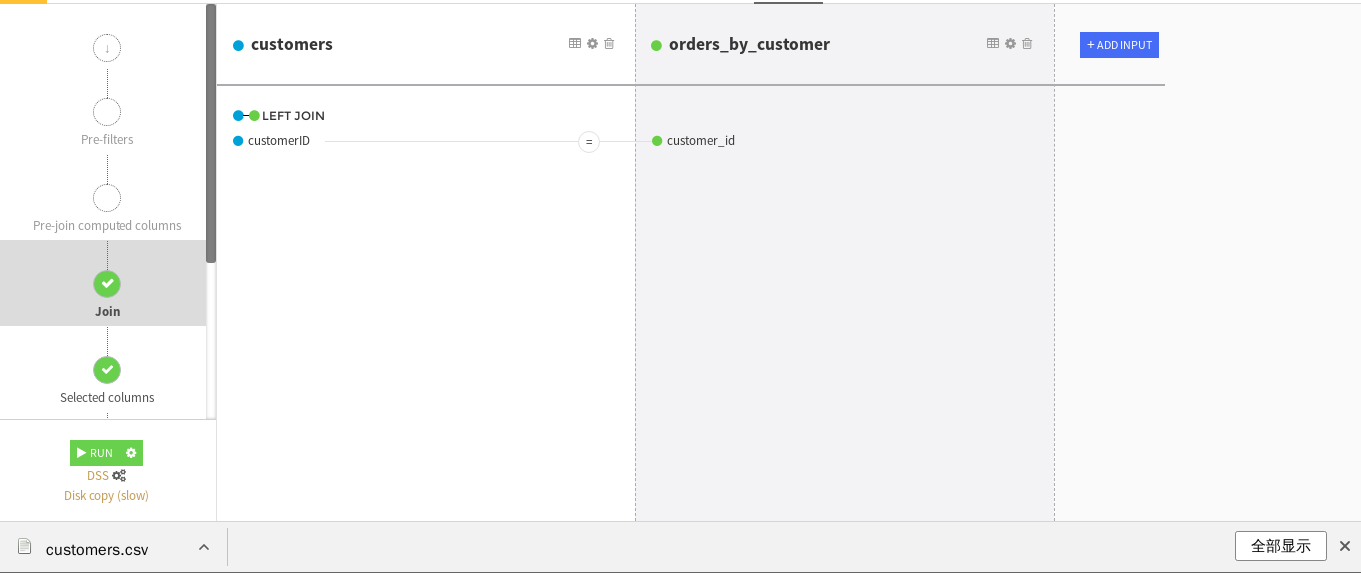
选择要合并的dataset



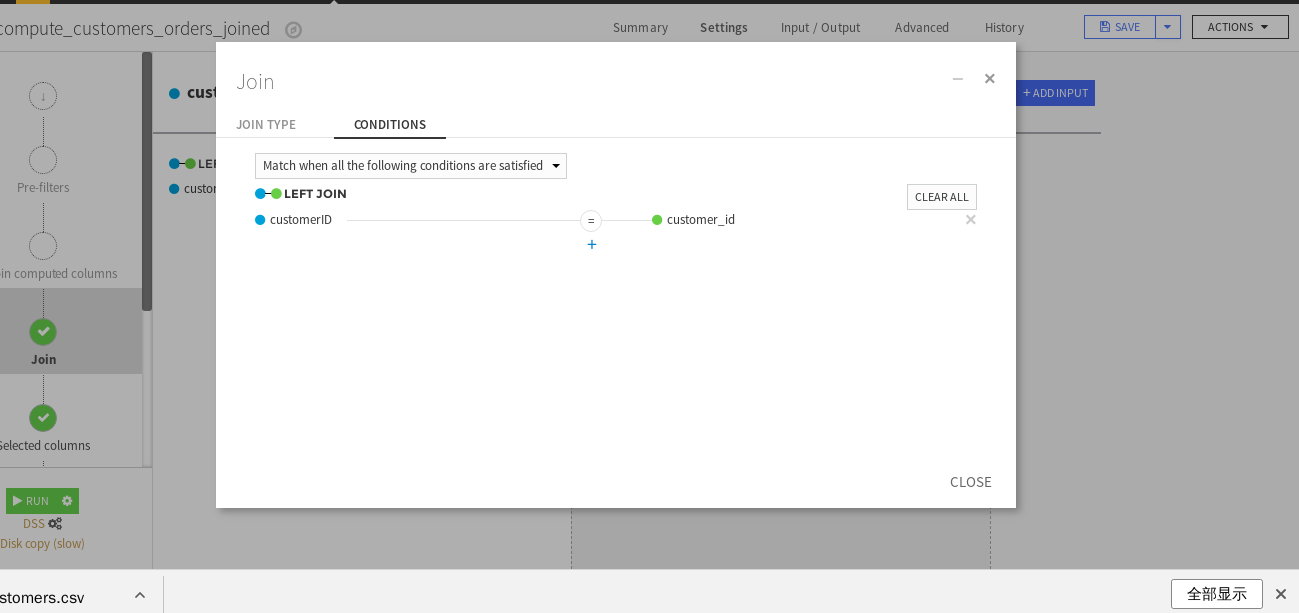
根据需要修改合并的表名



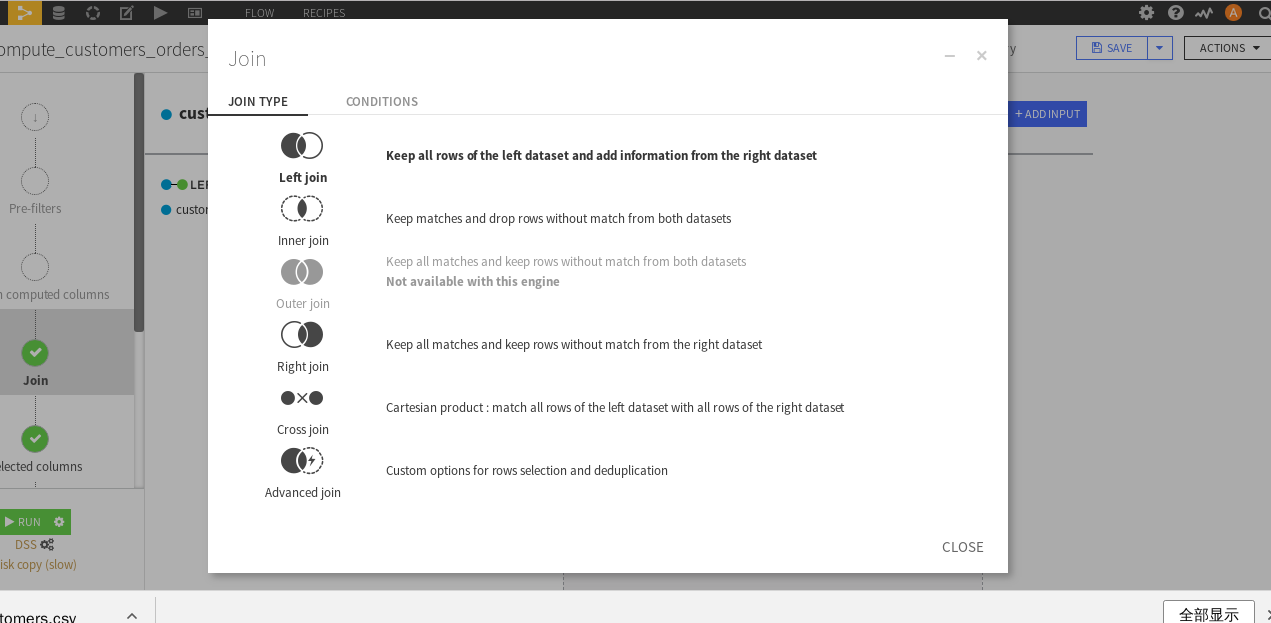
该系统默认的连接方式是左连接，点击LEFT JOIN图标



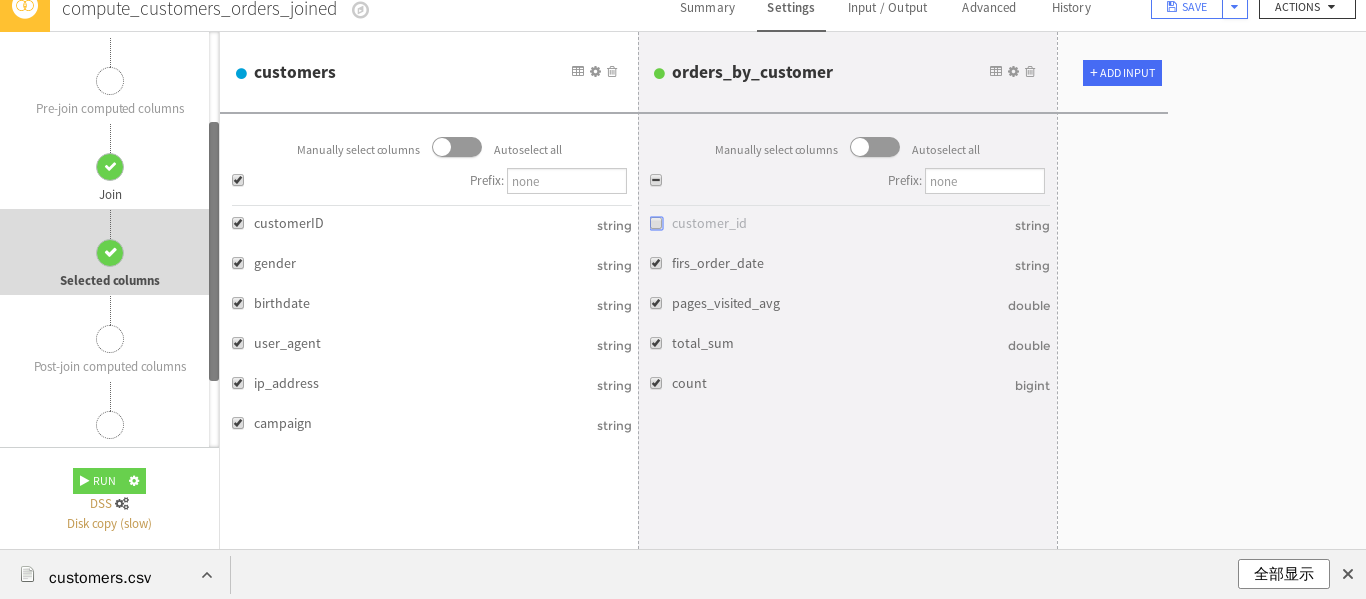
点击JOIN TYPE



根据需要选择连接的方式



点击selected column,根据需要选择不要要保留的列

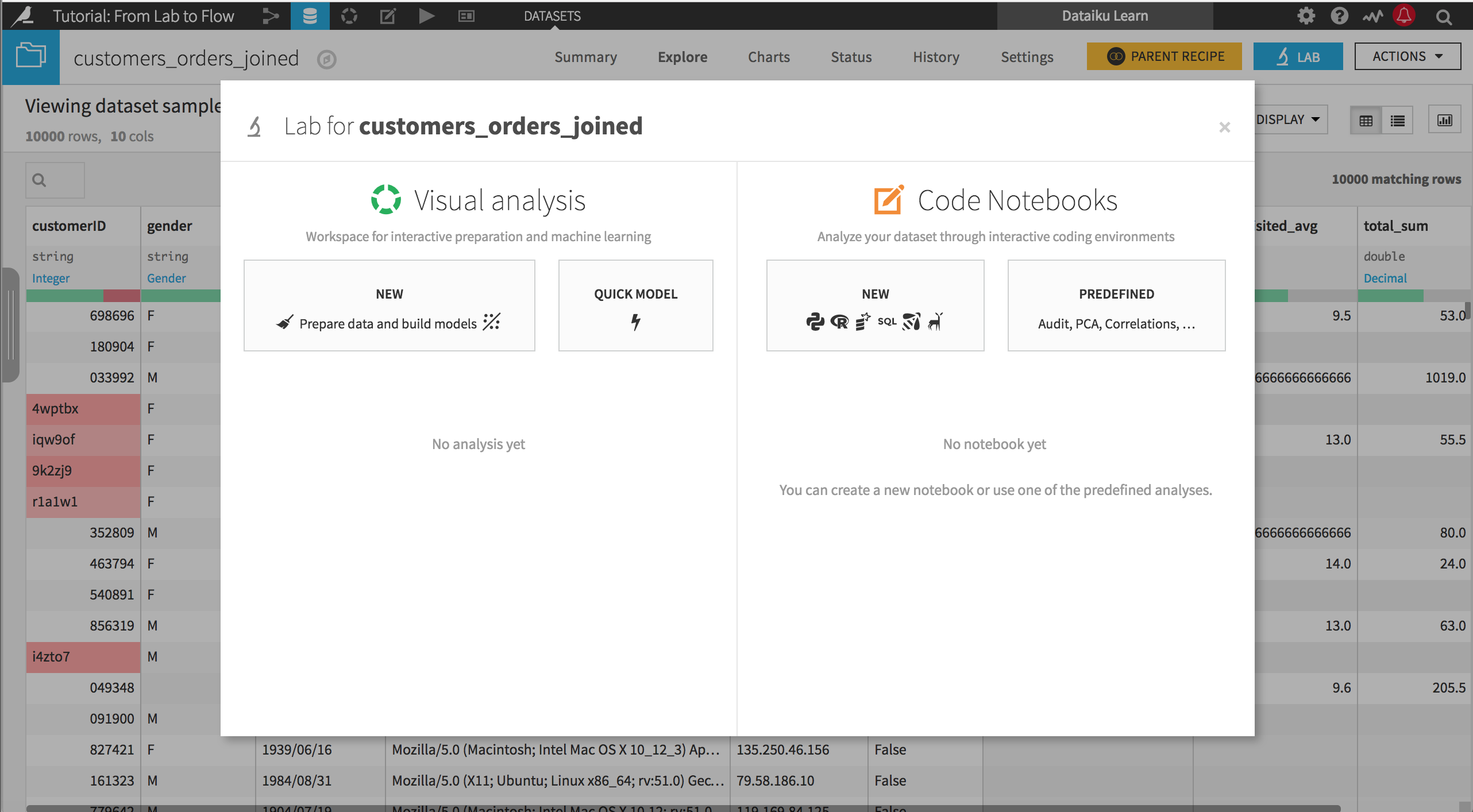


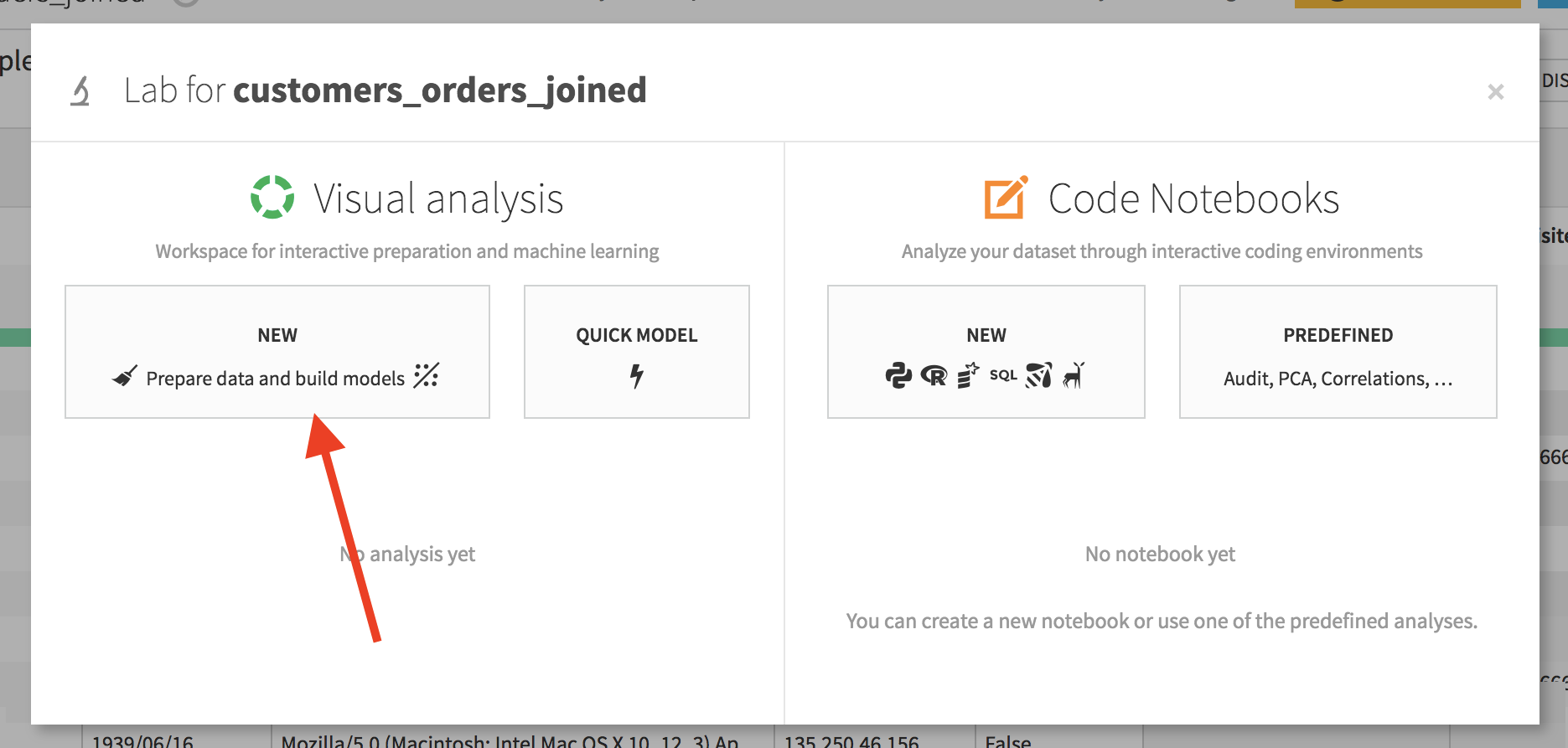
点击run等

2**.这个平台提供实验室，在实验室中我们可以模拟对数据的处理，处理结果暂时对项目无效，**

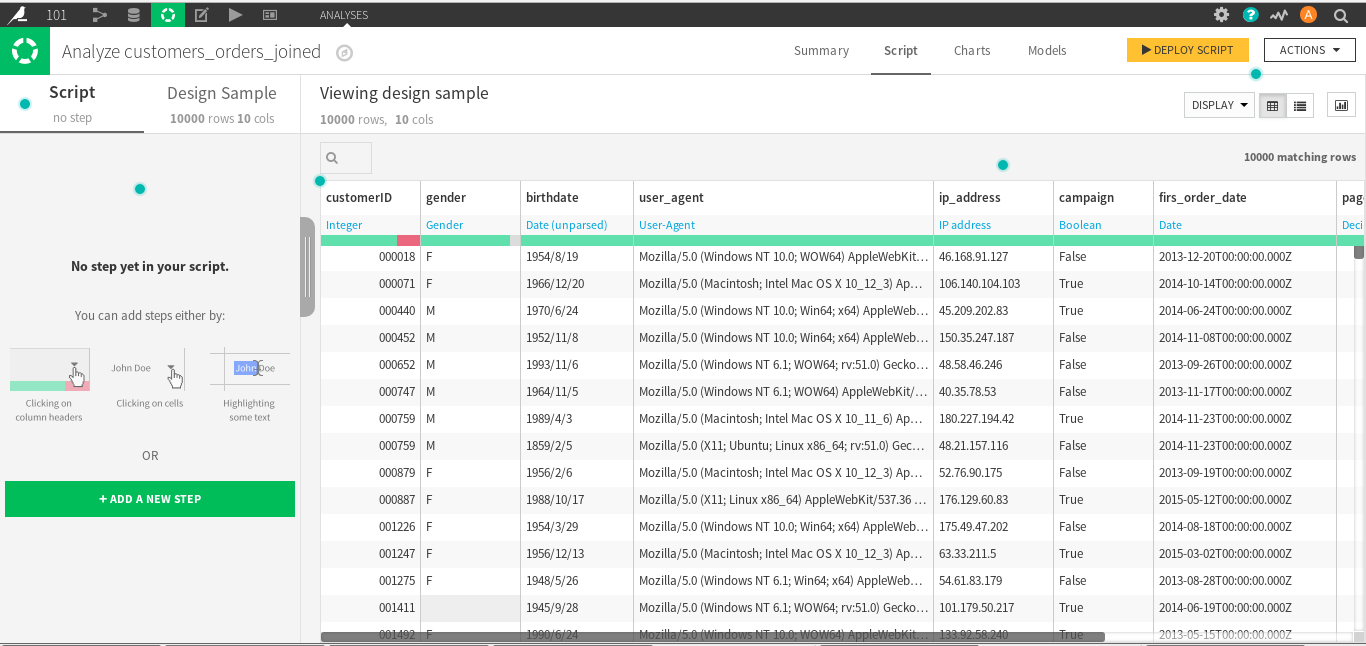
**你需要部署才行**

点击右上角的LAB

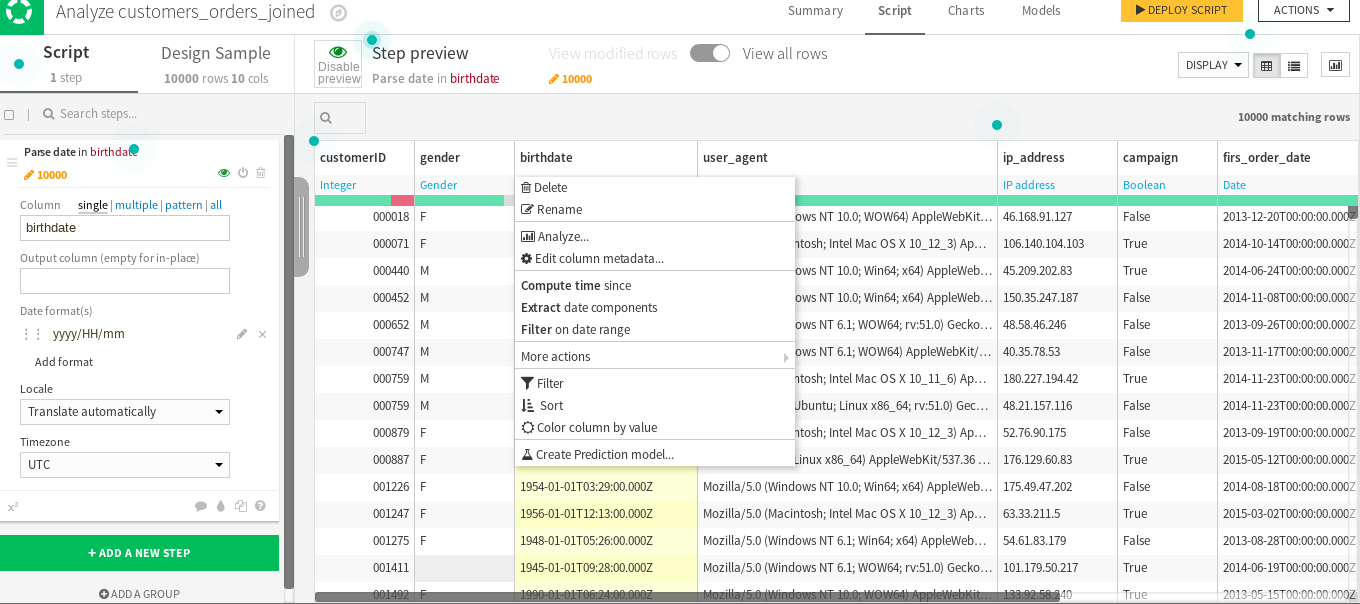




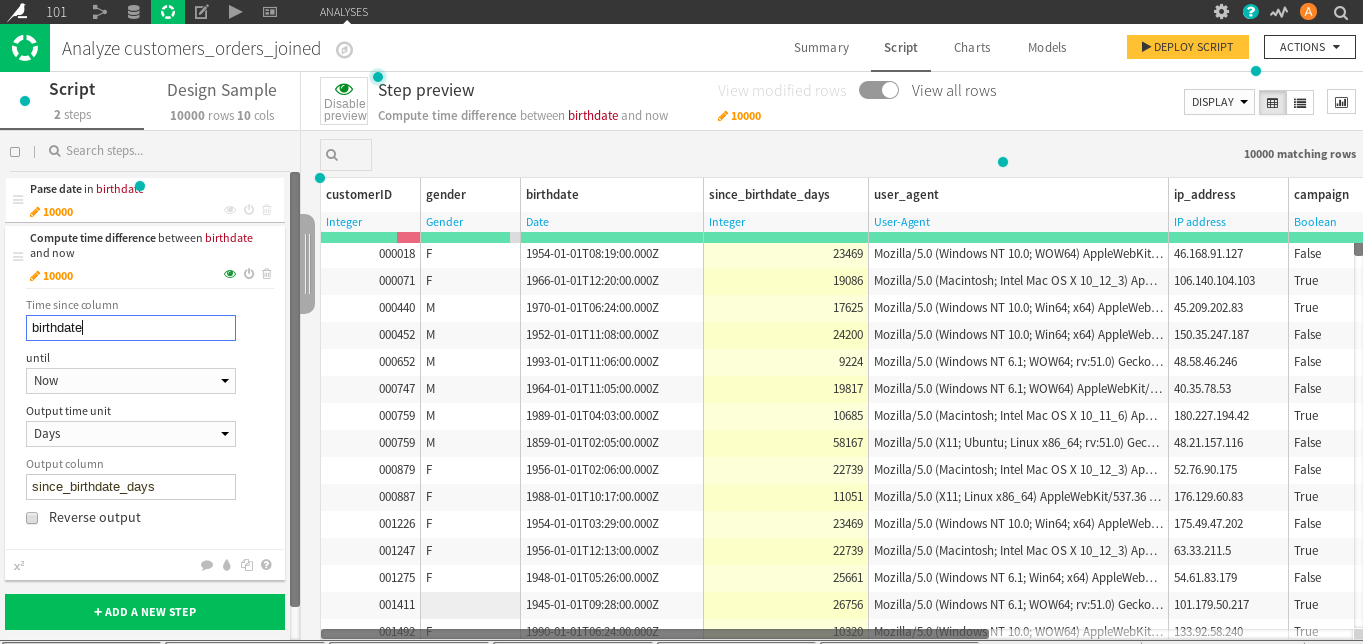
3.我们有时候想要得到两个日期间隔的时间



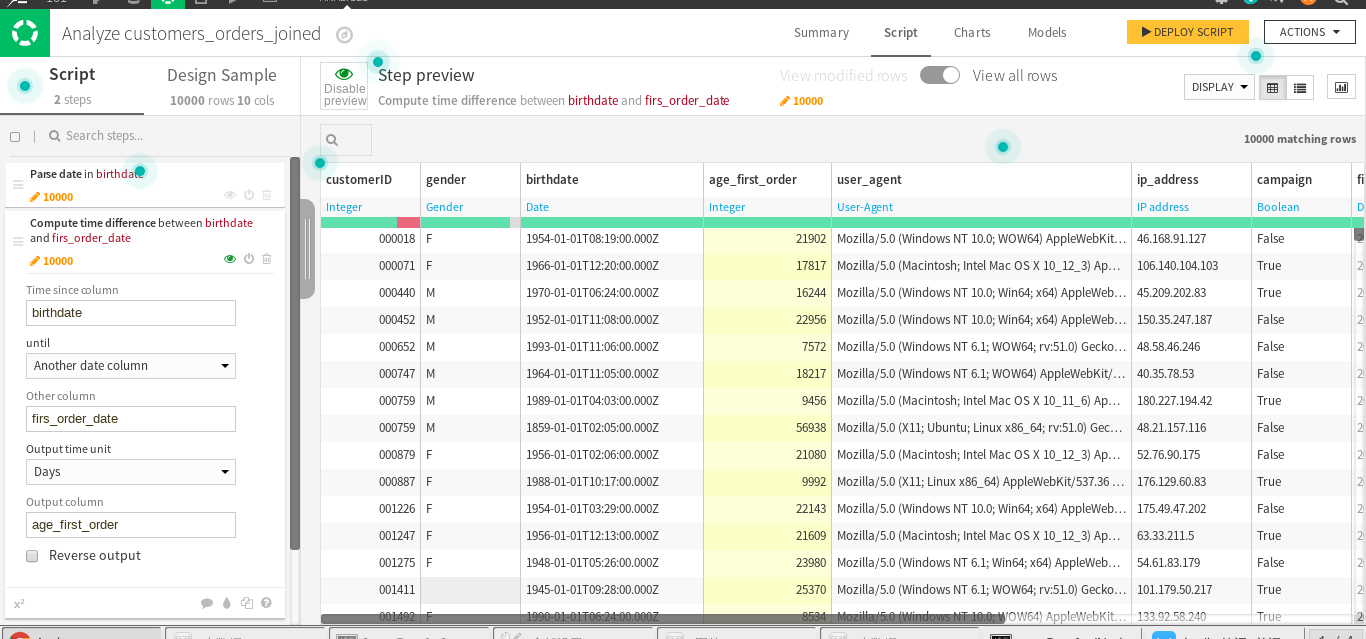
根据需要点击日期列，选择compute time since



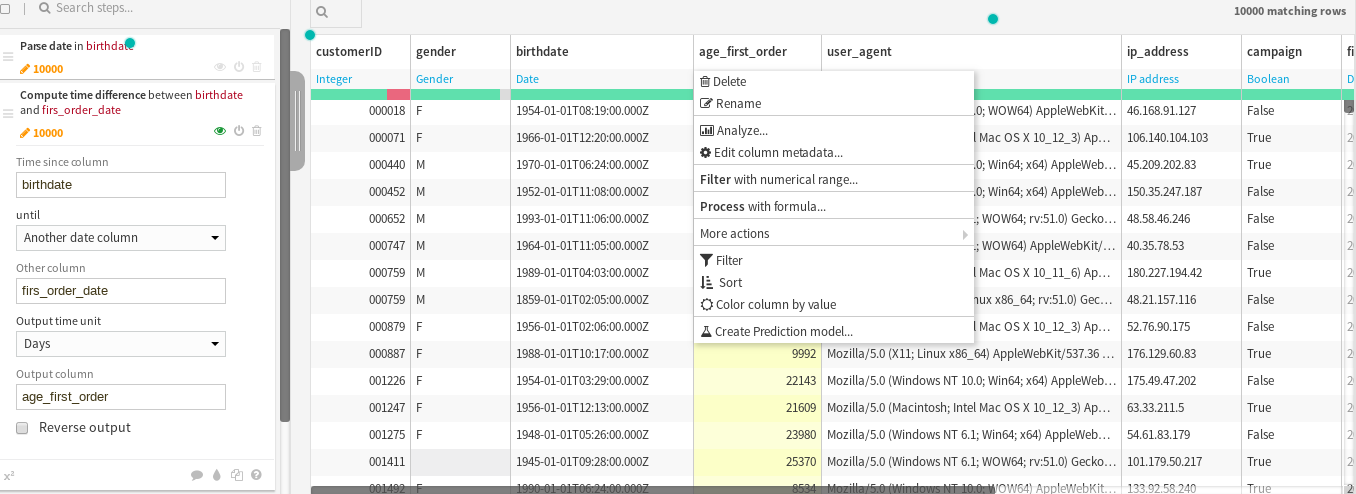
在左侧的util下拉框选择Another date column

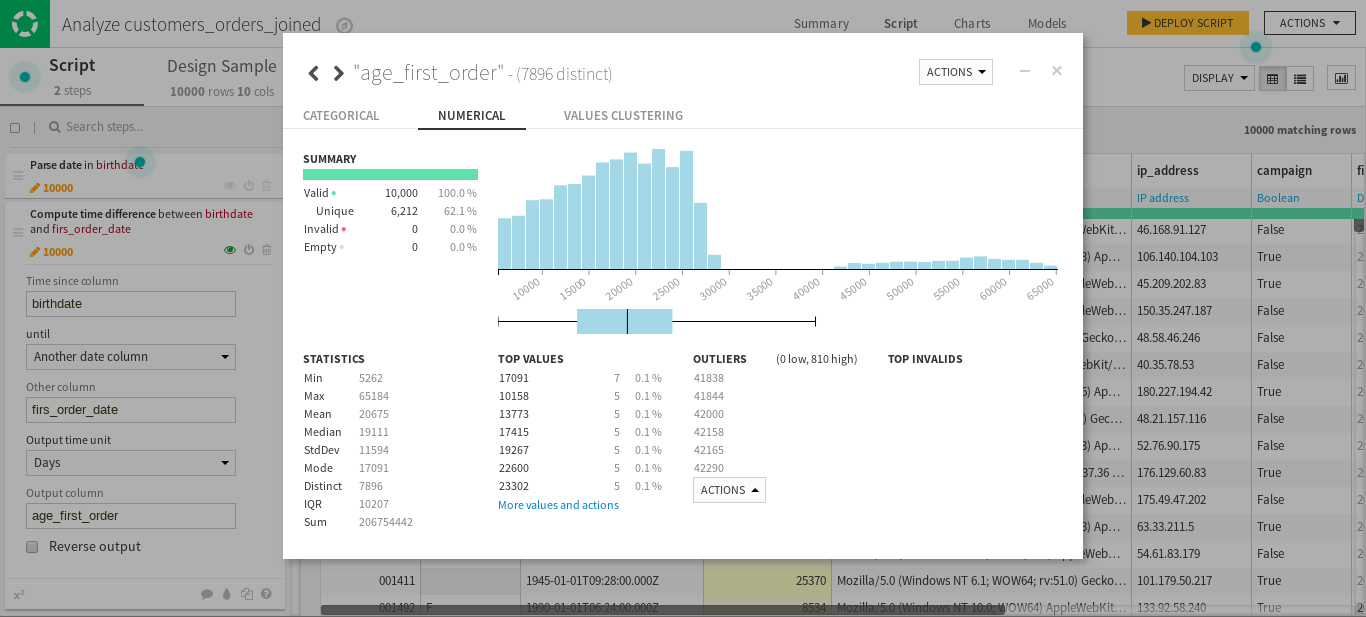


根据需要选取要对比的列，修改输出列的名字，和时间单位

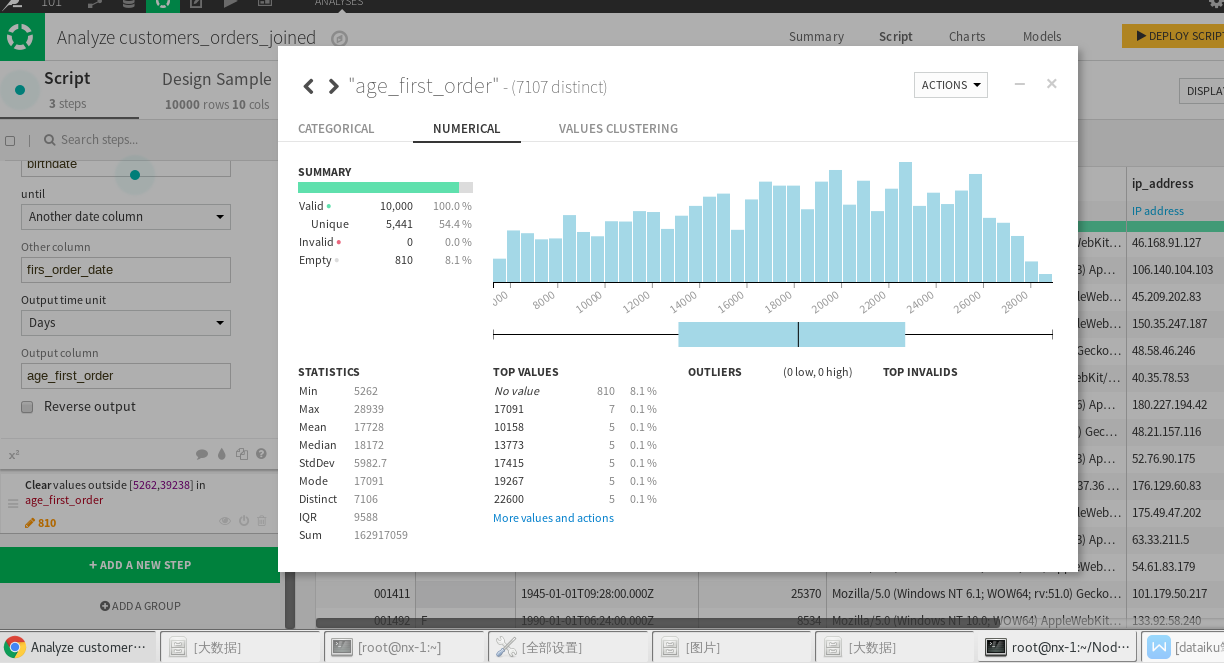


1. **我们有时候需要清理不合理的数据**，点击要相应的列，在下拉框选择analyze

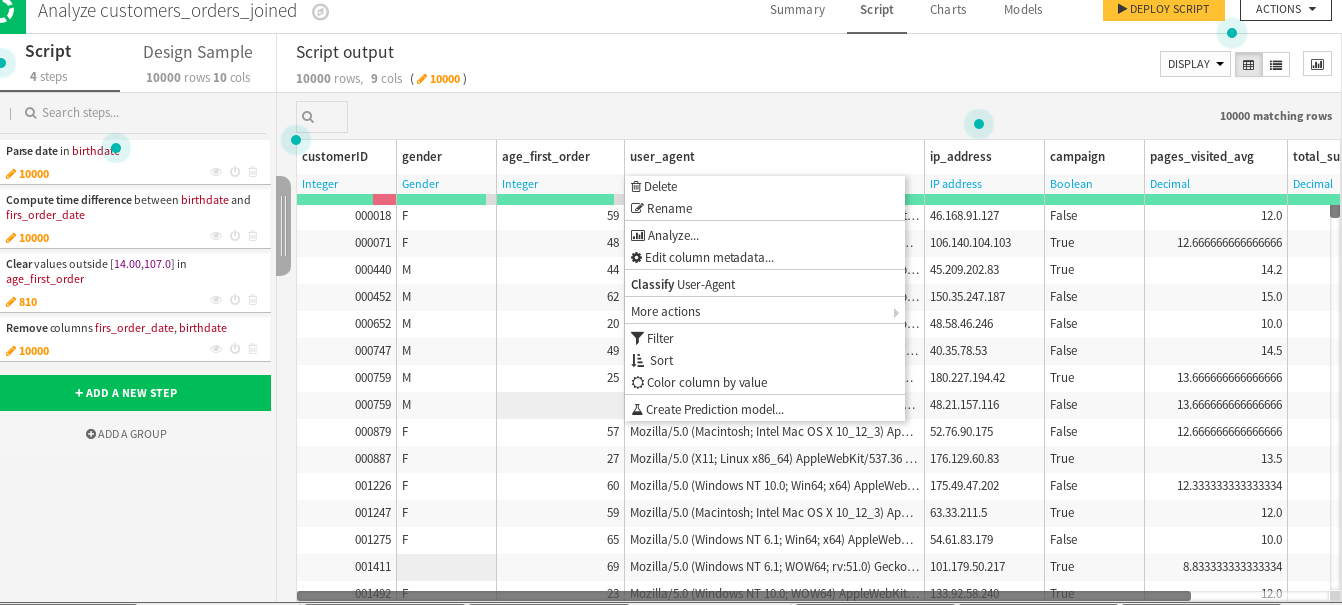




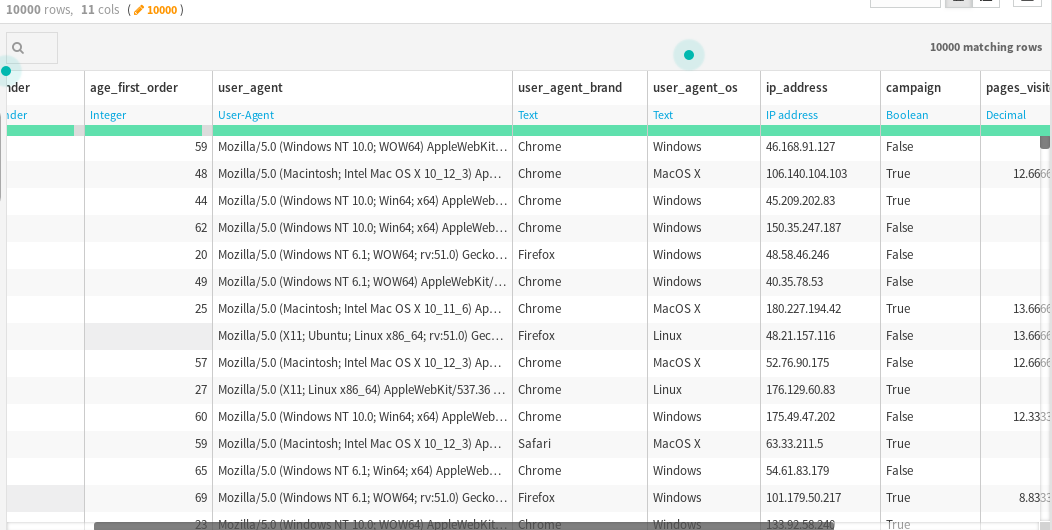
在图中我们发现一些数据实在不合理，所以我们点击下面的Actions,根据需要从下拉框中选择清洗数据的范围



5.我们有时候发现在一个列中有太多信息了，我们需要把他们分成多个列，点击列

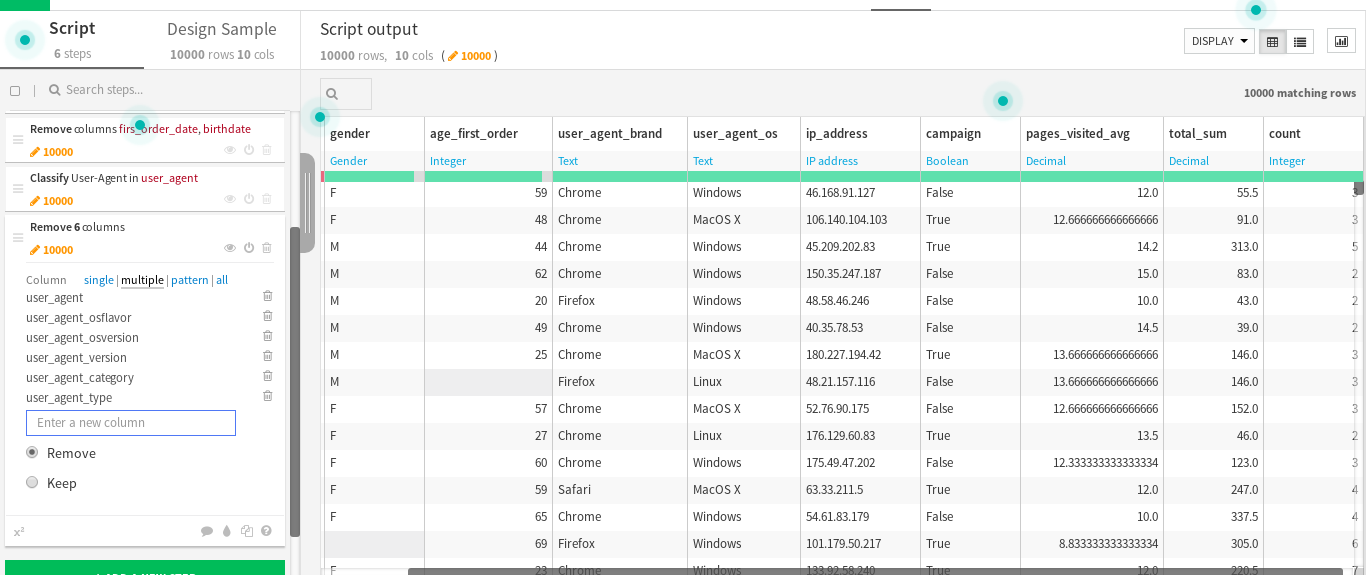


选择 Classify...

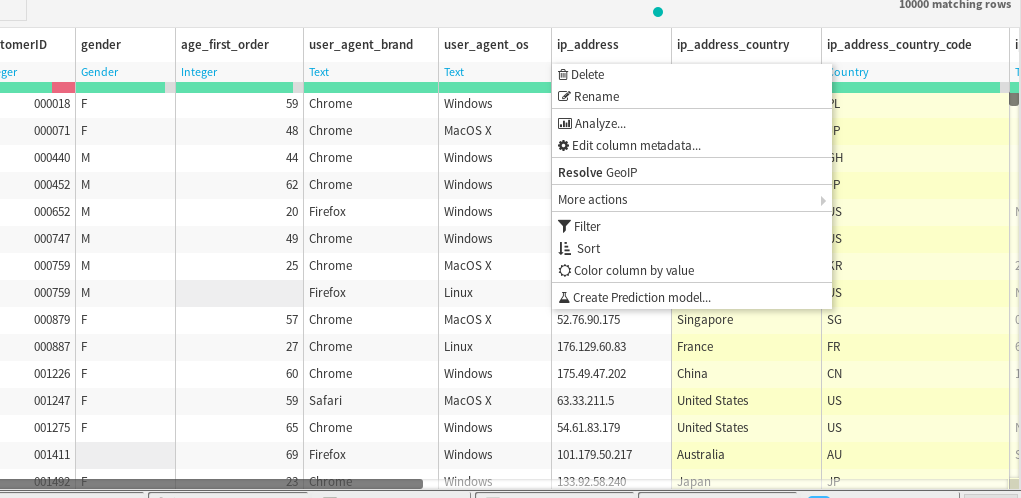


删除不需要的列

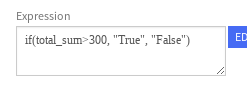
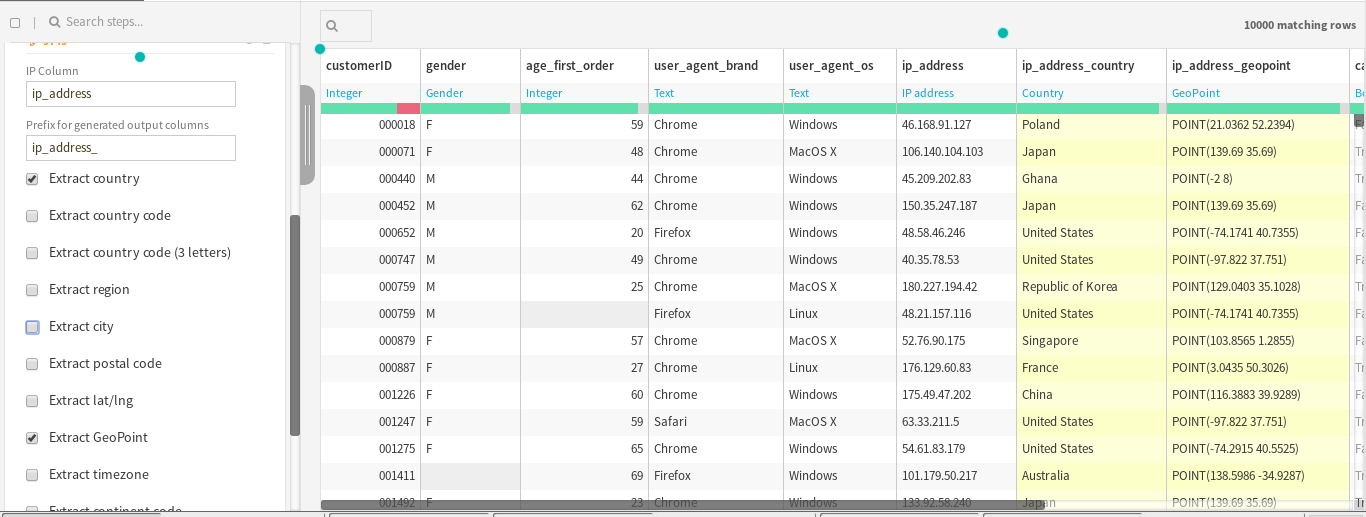
6.我们有时候需要从ip地址提取一些信息，点击ip列



选择Resolve GeoIP

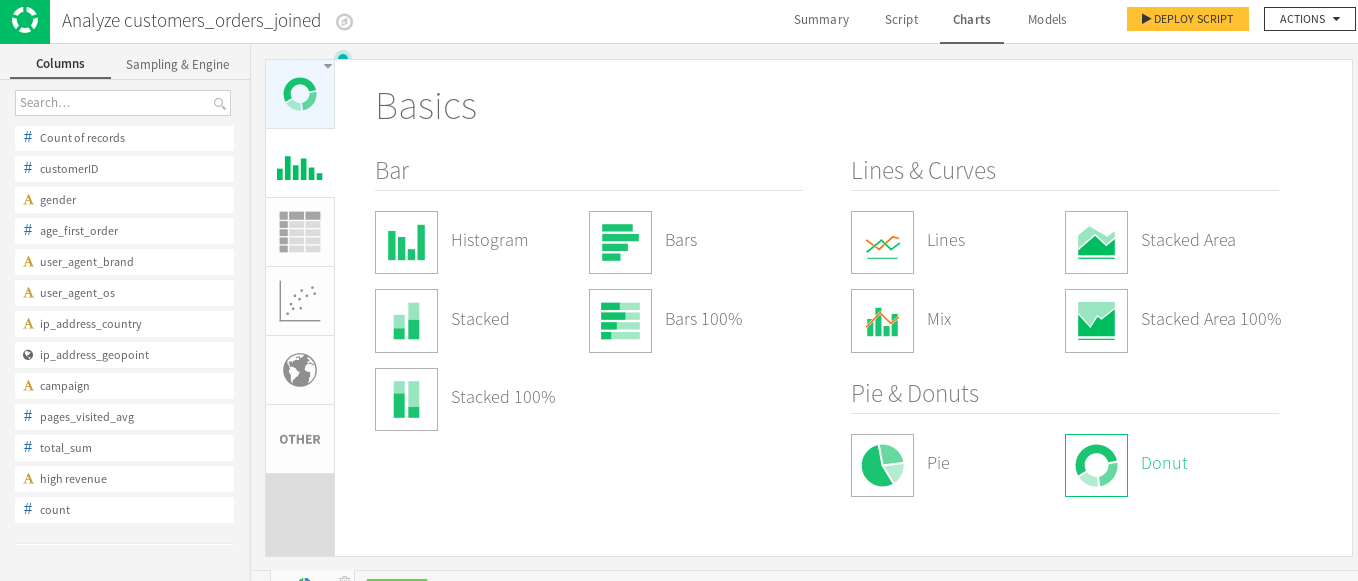


点击左侧最下面的脚本（绿色按钮前面），根据需要去掉不需要的列

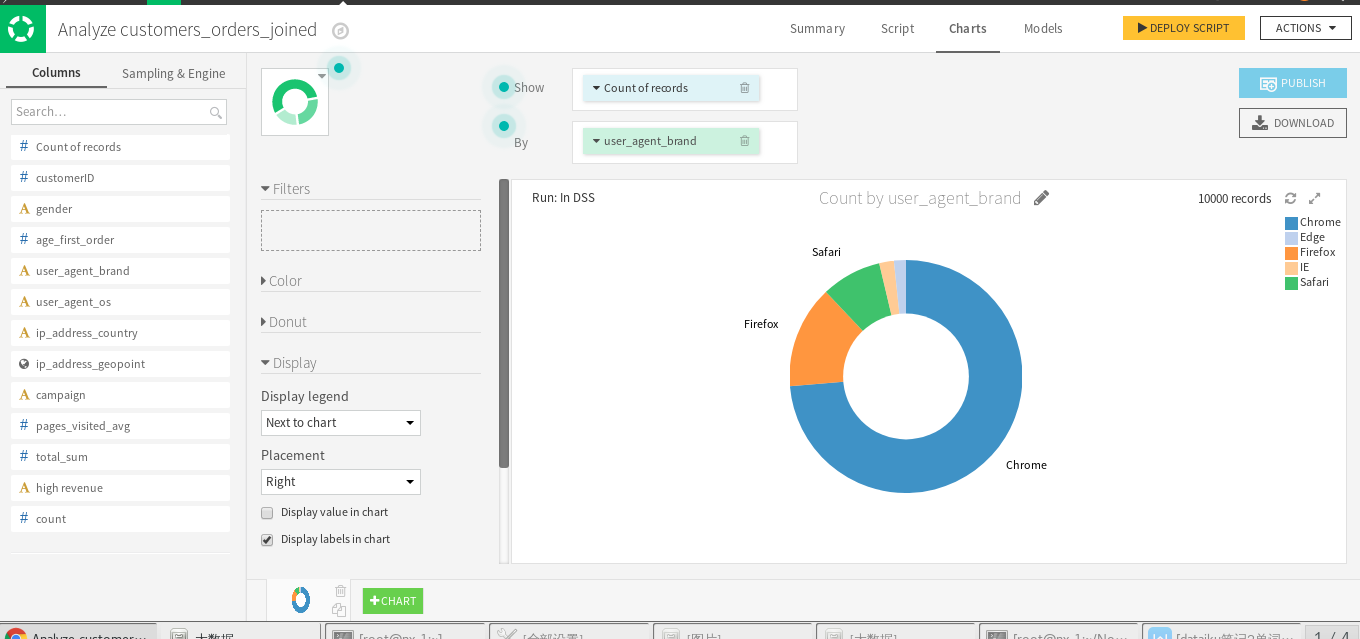


1. 我们有时候需要查看一个列值随着一个列值的变化，点击

，根据需要选择图形

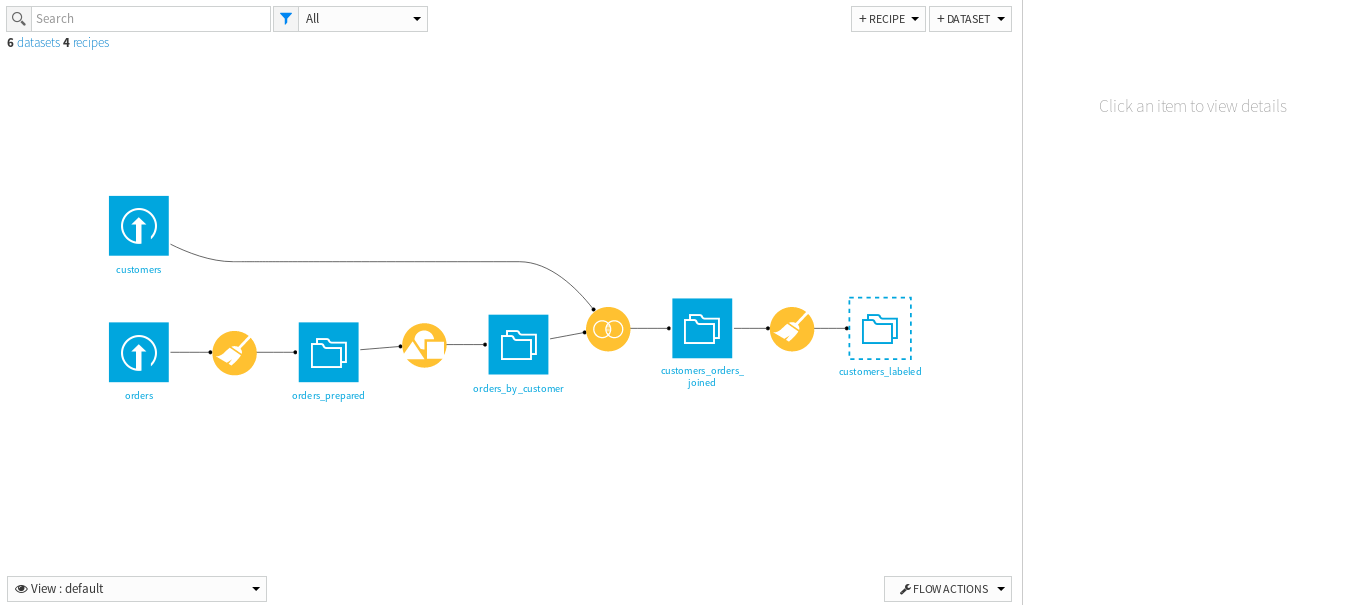


根据需要把左侧的列拖过去

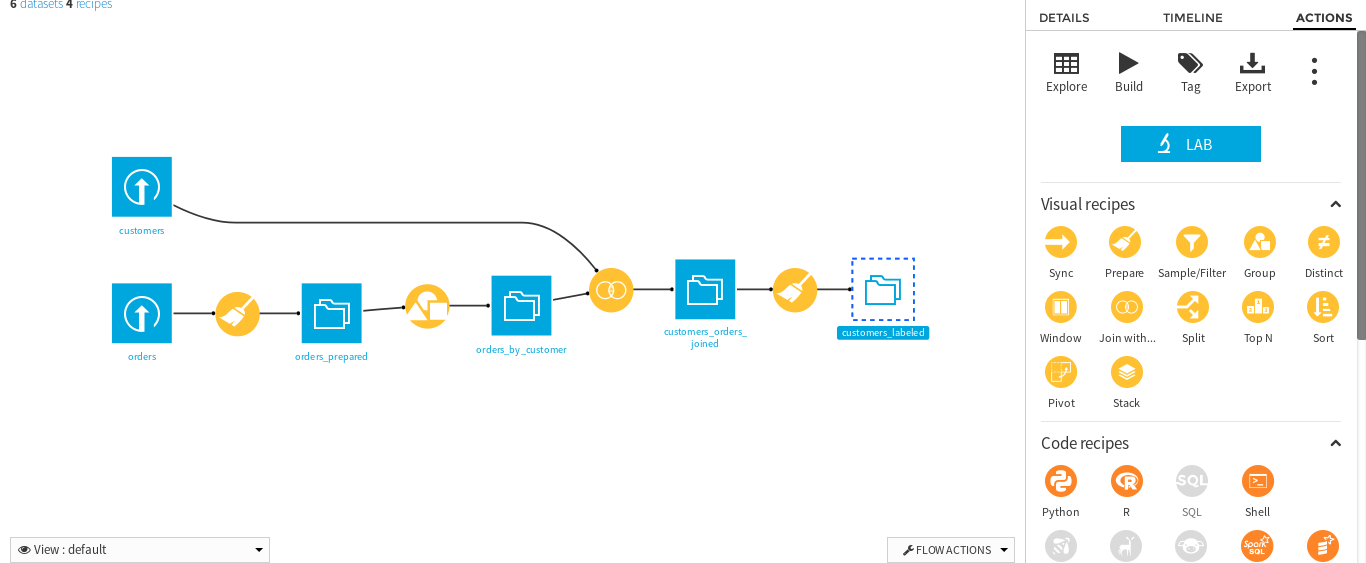


1. 当我们在实验室处理完之后，但实际的项目是不受影响的，所以我们需要部署

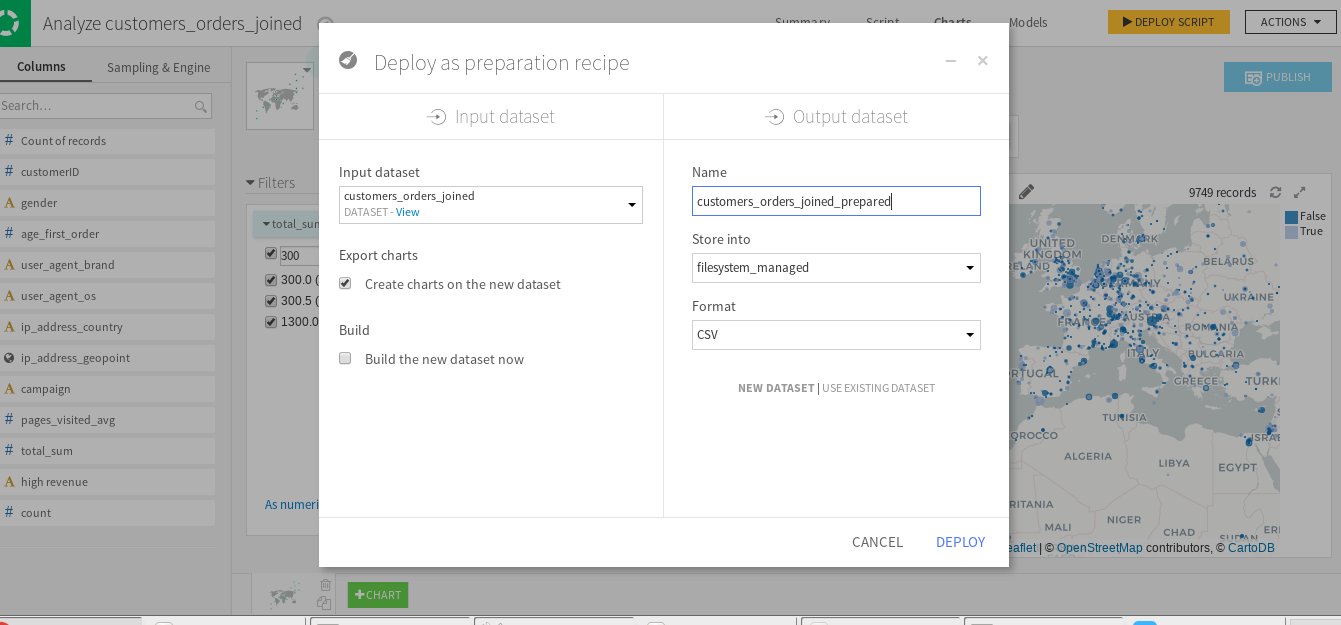
点击2018-04-05 09-00-27 的屏幕截图，点击最有侧白色的方框

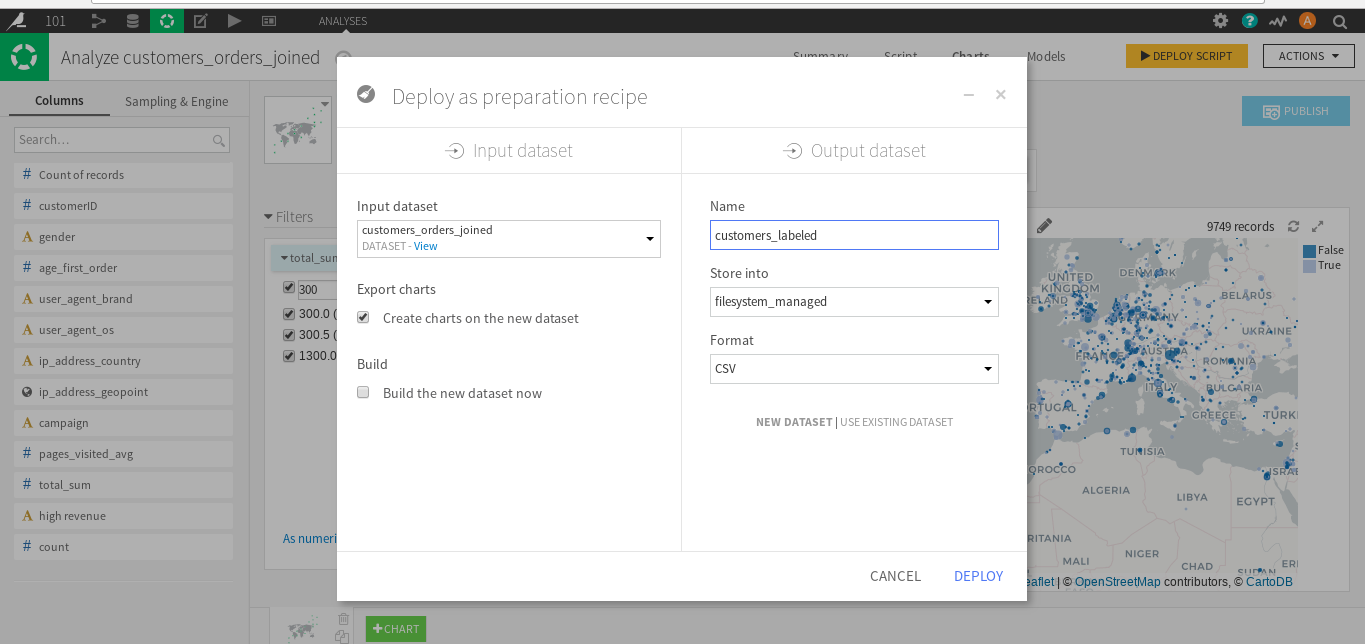


点击右侧的Build

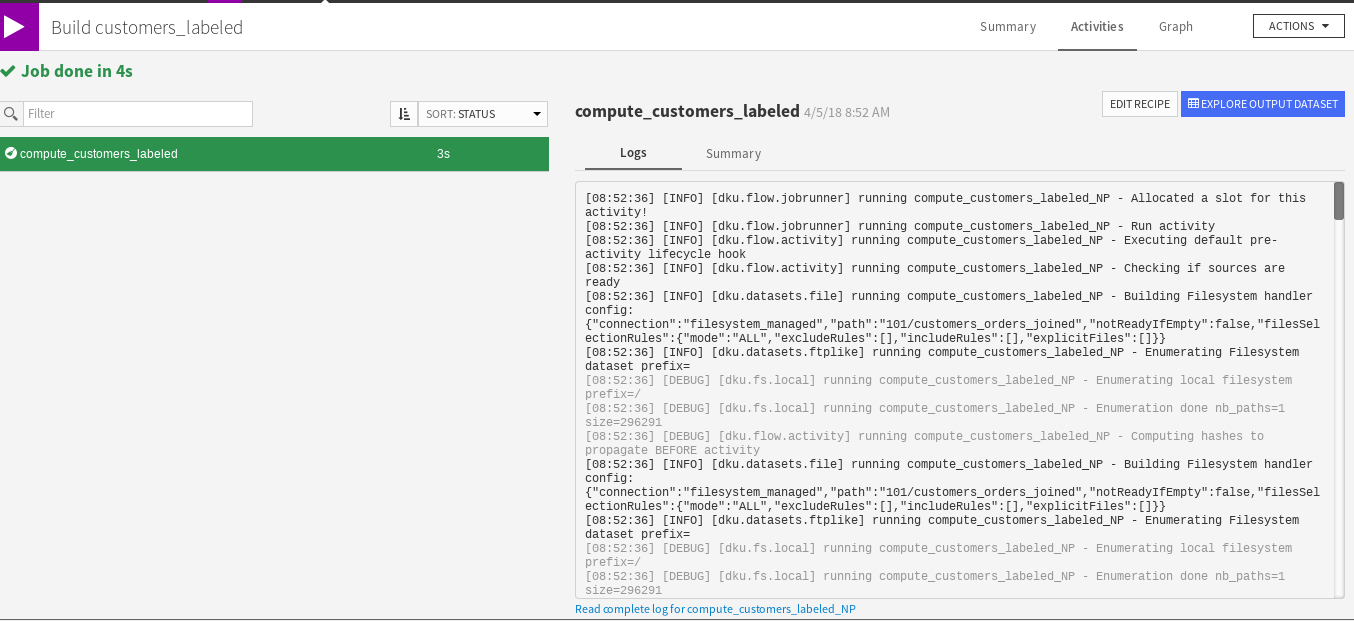


根据需要修改导出表的名字





点击右上角的explore ...



能够合并两张dataset,连接方式有左连接、右连接、外连接等。可以选择要合并的列。

内连接：保留两张表中都匹配的数据

外连接：保留两张表中所有的数据，包括不匹配的

Lab 实验室

Preliminary 初步的

dedicated 专用

Drafting 制图

Complementarily 互补

Overcrowding 人满为患

Steady 稳定的

Reused 重用

Assess 评估

Satisfied 满意

Lab：实验室，模拟数据的处理等，该实验室中可以有repie,机器学习的算法等，不会生成dataset等，但可以把模拟的dataset等转换为实际的dataset。可以进行可视化分析。

go through 经过

Distribution 分配

be prompted to 被提示

Interactive 互动

Topic 话题

couple 一对

可以用两个列的值相减得到新的一列

distribution 分配

turns out 事实证明

Outliers 离群

Indicative 指示

InterQuartile（**四分位数） 把数值从小到大分成四等份，分割点的数值就是四分位数，第一个分割点与第三个分割点的距离就是四分位距（InterQuartile Range）**

**Clip 清除**

**可以查看一个列的数据的分布，清除掉不合理数据，可以删除掉列属性**

**Leveraging 利用**

**Agent 代理人**

**Classify 分类**

**可以对一个列的数据进行分类**

**Recognize 认识**

**Extract 提取**

**deselect 取消**

**可以解析ip的所在的城市，国家等**

**Label 标签**

**revenue 收入**

**Formula 式**

**Validate 验证**

**可以对某一列添加表达式，并且产生一列判断它是true或false(true和false是可以自己设定的)**

**graphing 制图**

**Productive 生产的，丰富的**

**Loop 循环 套**

**a pie 一块**

**Donut 甜甜圈**

**可以以一个圈的方式看一个属性的分布（一个值随另一个值的影响）**

**relative share 相对份额**

**Enlightening 启发**

**approximately 大约**

**或在是一个值比如总数随着两个值的影响**

**Campaign 运动**

**Insight 眼光**

**Glean 收集**

**Moderately 适当**

**Valued 价值**

**可以画出一个属性和另一个属性的X-Y的散列表，还可以看到该点的其他细节**

**Extracted 提取**

**marketing 营销**

**Bound 界**

**Filter 过滤**

**可以把地理af放在地图中显示，并且可以显示每一点的地理信息**

**Reactive 反映**

**lineage 血统**

**Popup 弹出**

**在lab中处理之后，需要生成dataset**

**entire整个**

**push down 按下**

**Computation 计算**

**External 外部**

**可以把dataiku的dataset存储到数据库中**