**Tableau Prep Project on Gender Pay Gap**

Submitted in partial fulfillment of the requirement

for the award of the Advanced Diploma of

**DATA SCIENCE AND APPLICATION**

Of

**METRO COLLEGE OF TECHNOLOGY**

**2020-2021**

Submitted to

**Prof.Hamid Rajaee**

Submitted by

**Mr. Sajith Vellappillil**



**About Tableau Prep**

Tableau Prep Builder is a tool in the Tableau product suite designed to make preparing data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

**Using Tableau Prep**

Start by connecting to the data from a variety of files, servers, or Tableau extracts. Next, connect to and combine data from multiple data sources.

Drag and drop or double-click to bring the tables into the flow pane, and then add flow steps where you can then use regular operations such as filter, split, rename, pivot, join, union and more to clean and shape your data.

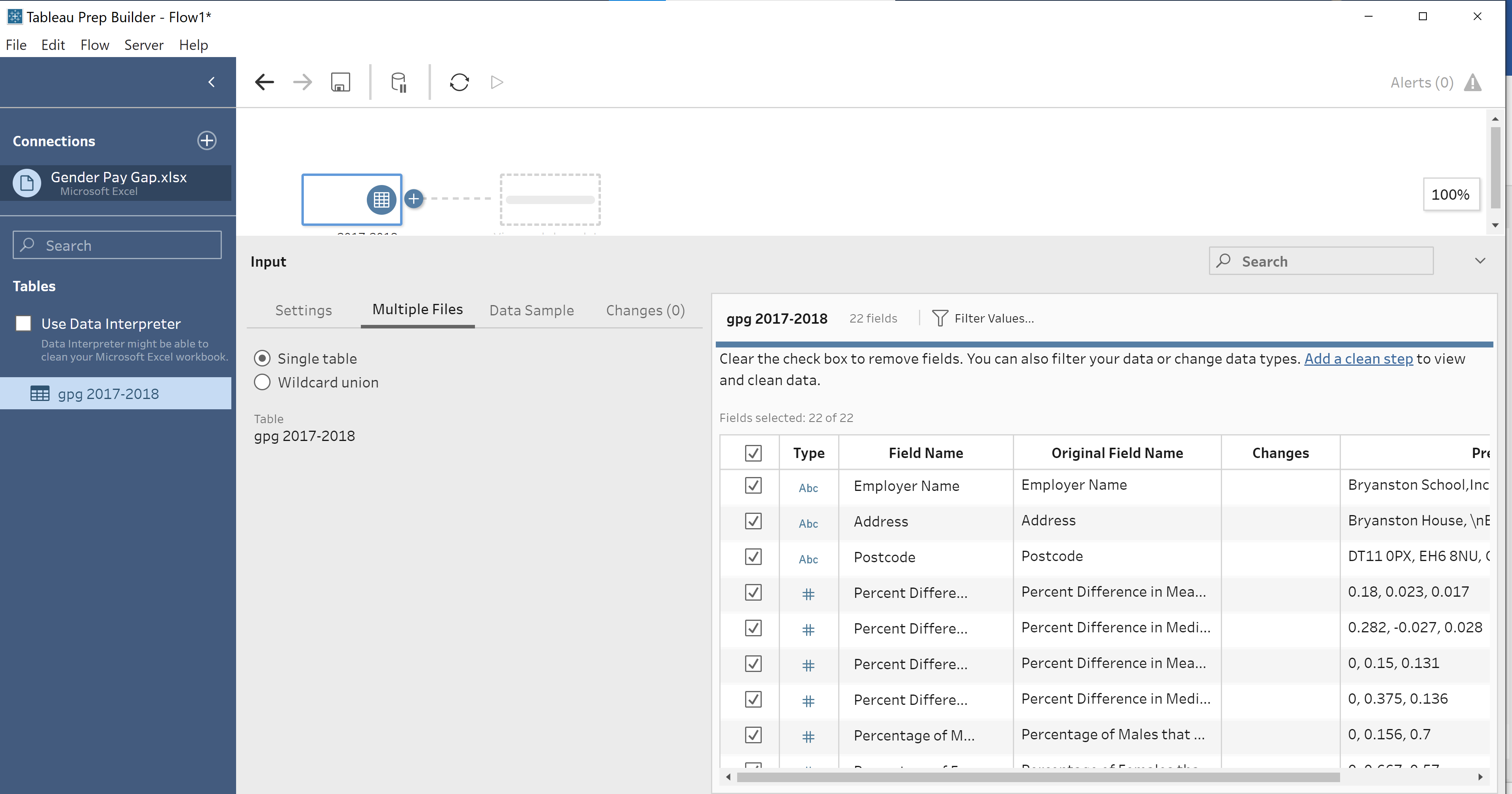
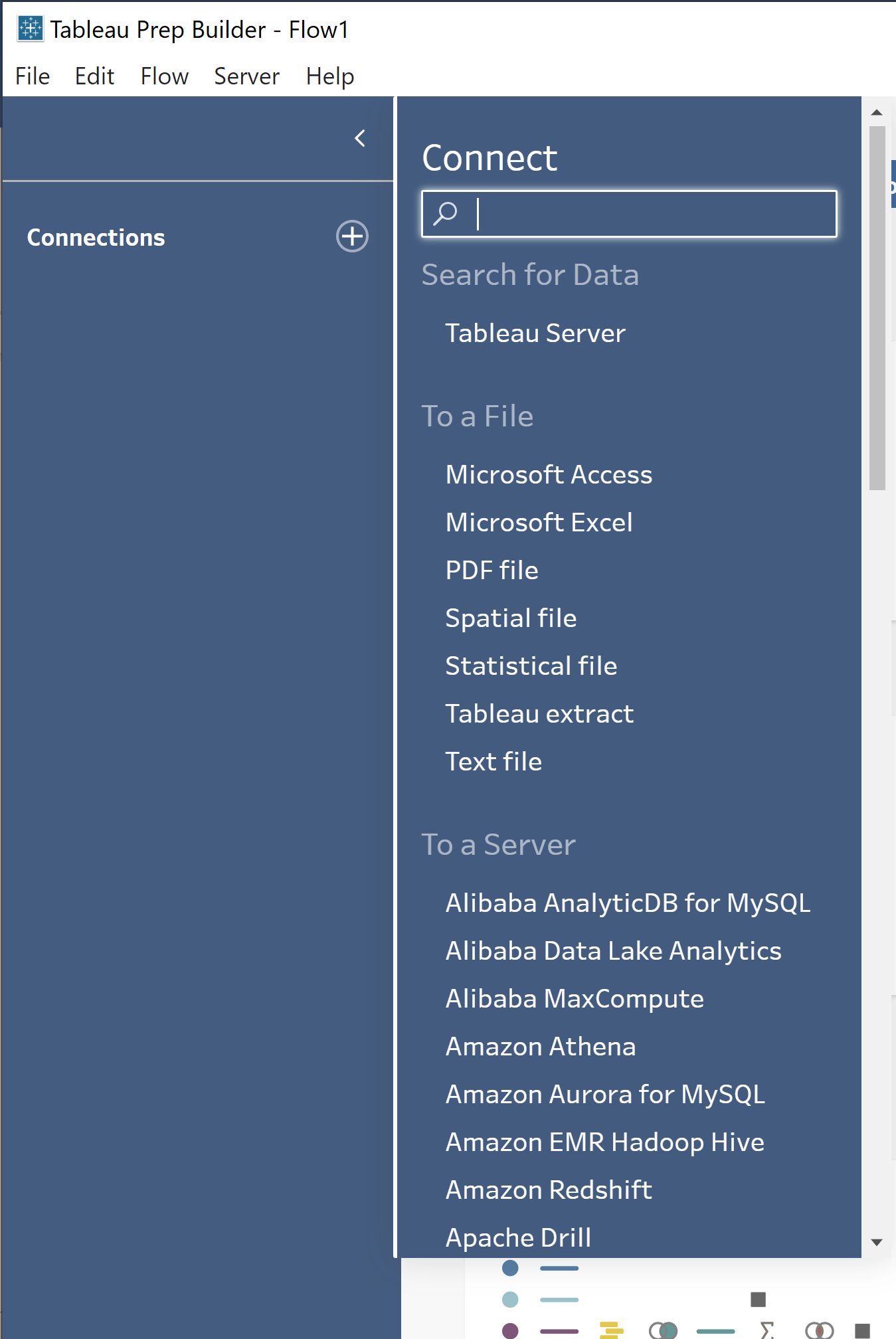
Each step in the process is represented visually in a flow chart that you create and control. Tableau Prep tracks each operation so that you can check your work and make changes at any point in the flow.

Once done with your flow, run it to apply the operations to the entire data set.

Tableau Prep works seamlessly with other Tableau products. You can create an extract of your data at any point in your flow. For example, publish your data source to Tableau Server or Tableau Online, publish your flow to Tableau Server or Tableau Online to continue editing on the web or refresh your data using a schedule. You can also open Tableau Desktop directly from within Tableau Prep Builder to preview your data.

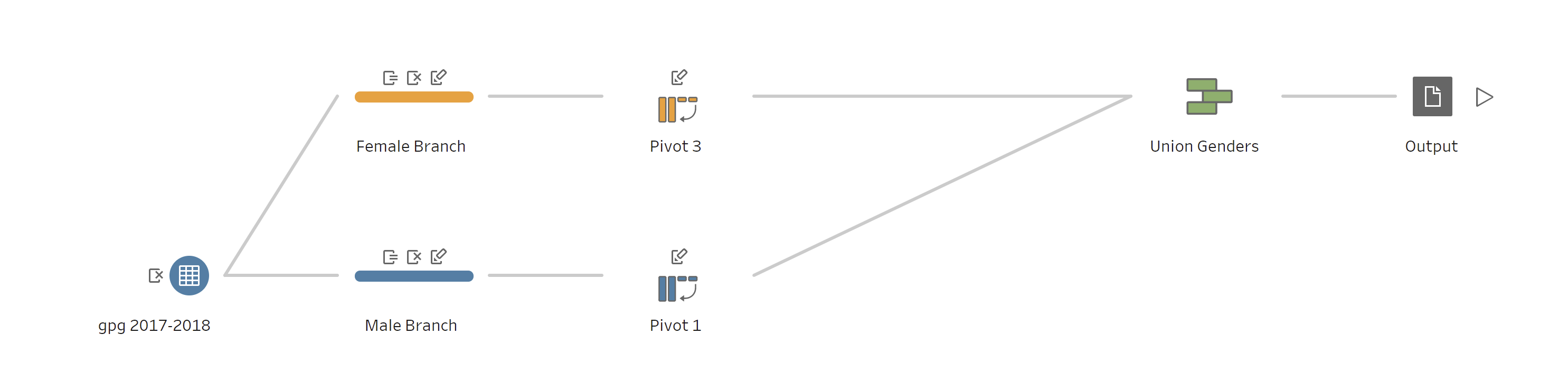
**Connections pane**

On the left side of the workspace is the Connections pane, which shows the connected databases and files. Add connections to one or more data sources and drag the tables you want to work with into the Flow pane.



**Flow pane**

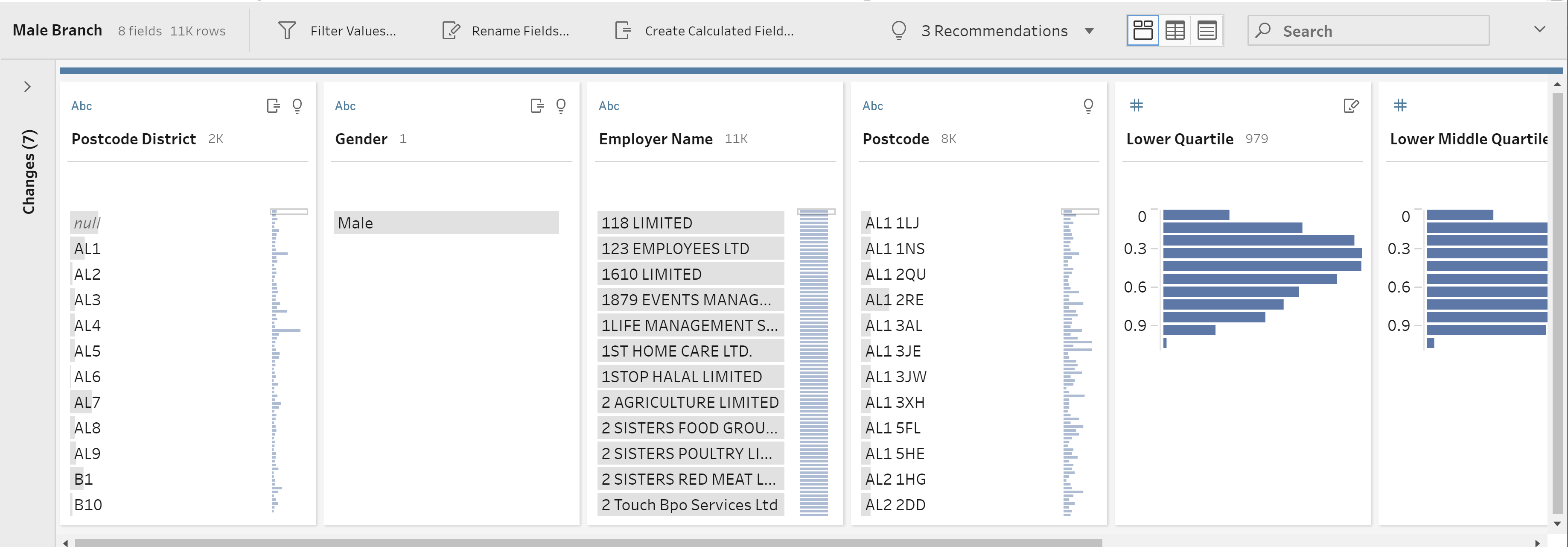
At the top of the workspace is the Flow pane. This is where you'll build your flow. As you connect to, clean, shape, and combine your data, steps appear in the Flow pane and align from left to right along the top. These steps tell you what kind of operation is being applied, in what order, and how your data is affected by it. For example, the Join step shows you which join type you've applied, the join clauses, recommended join clauses, and the fields of the tables that are included in the join.



You start your flow by dragging tables into the Flow pane. Here you can add additional data sets, pivot your data, union or join data, create aggregations, and generate output files in the form of .tds files, Hyper extract (.hyper) files, or published data sources that you can use in Tableau. You can even write your output data to a database. For more information about generating output files.

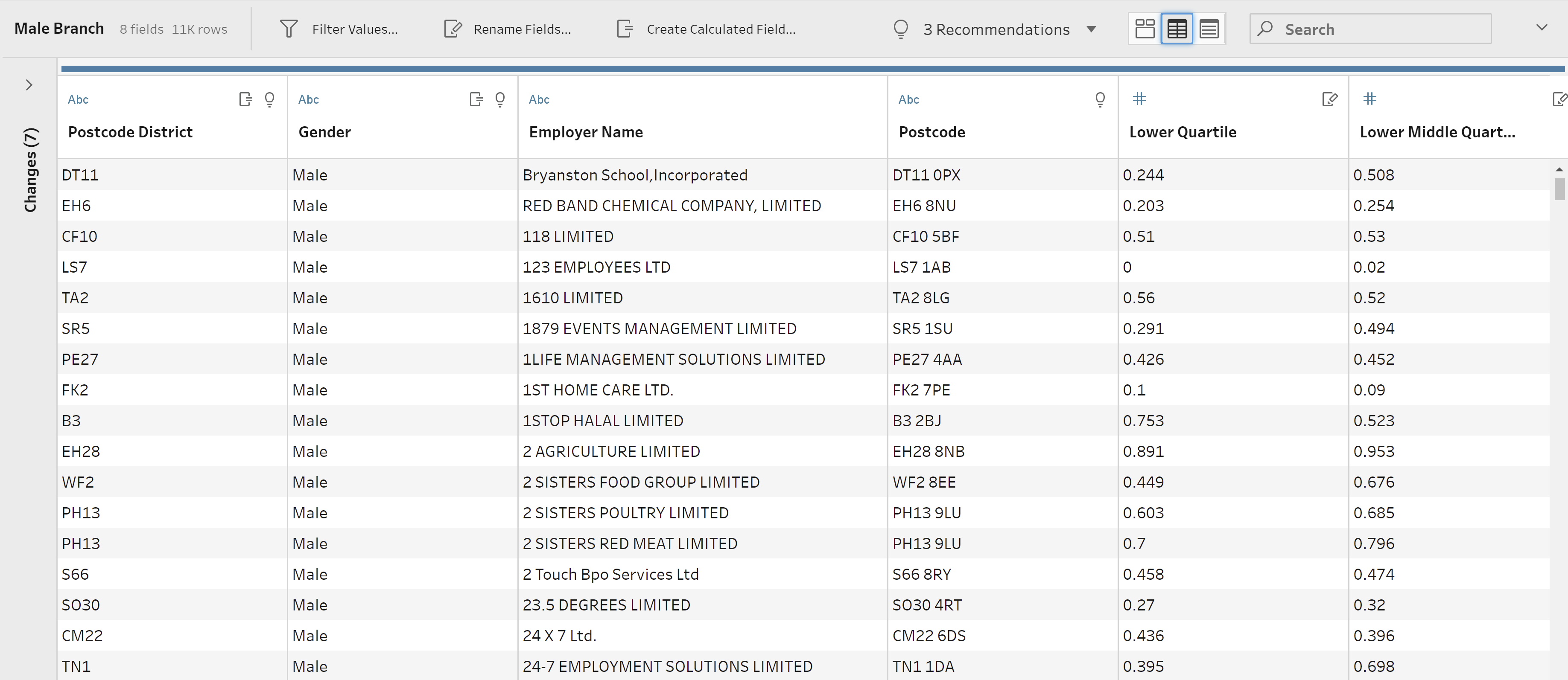
**Profile pane**

In the center of the workspace is the Profile pane. The Profile pane shows you the structure of your data at any point in the flow. The structure of your data can be represented in different ways depending on the operation you want to perform on your data or the step that you select in the Flow pane.



**Data grid**

At the bottom of the workspace is the **Data** grid, which shows you the row level detail in your data. The values displayed in the **Data** grid reflect the operations defined in the **Profile** pane. You can perform the same cleaning operations here as you can in the Profile pane if you prefer to work at a more detailed level.



**About the data**

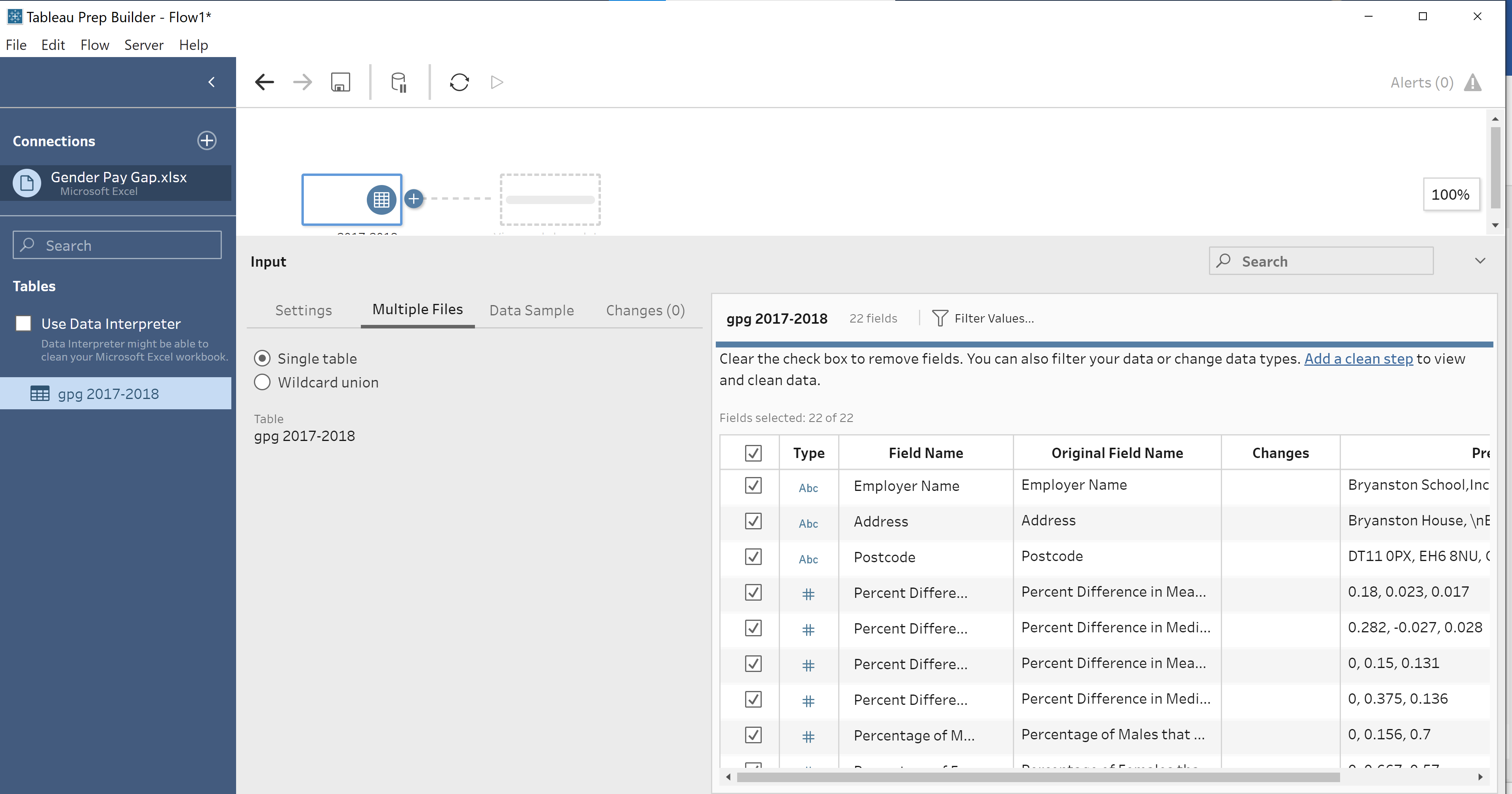
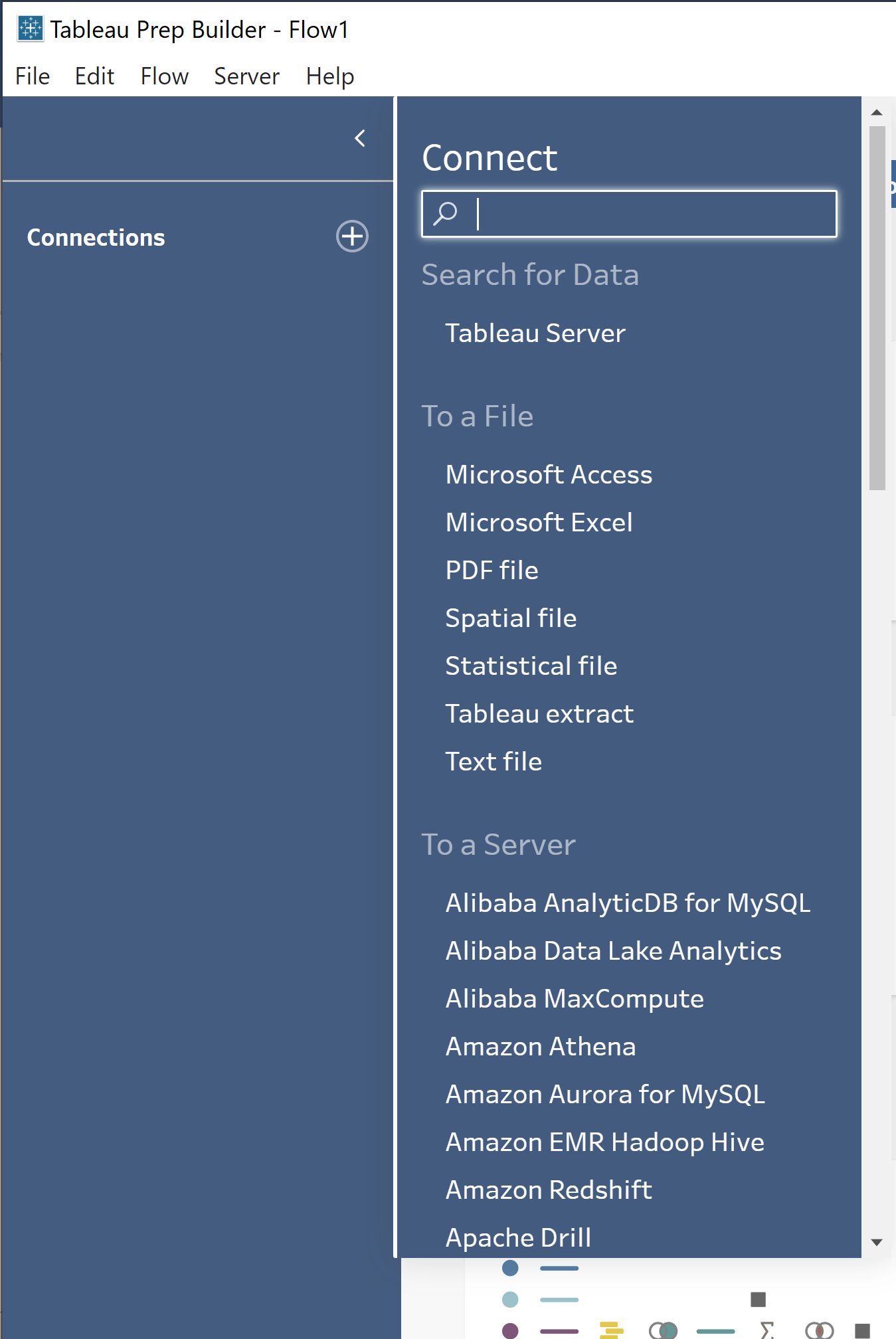
For processing in Tableau prep I have selected an excel file which will be split into 2 parts as

1. Male Branch and
2. Female.

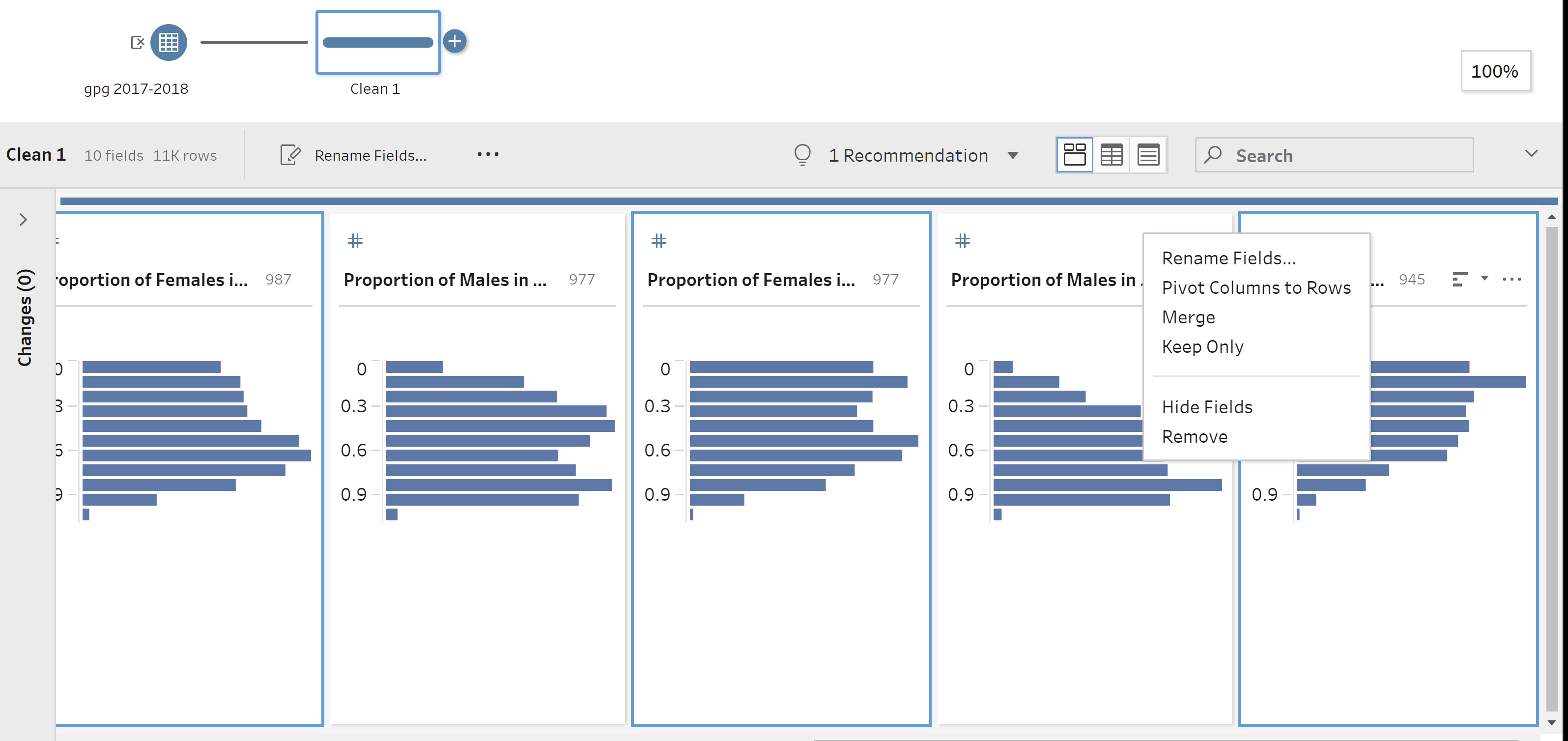
The data is about the Gender pay gap.

**Steps in data processing :**

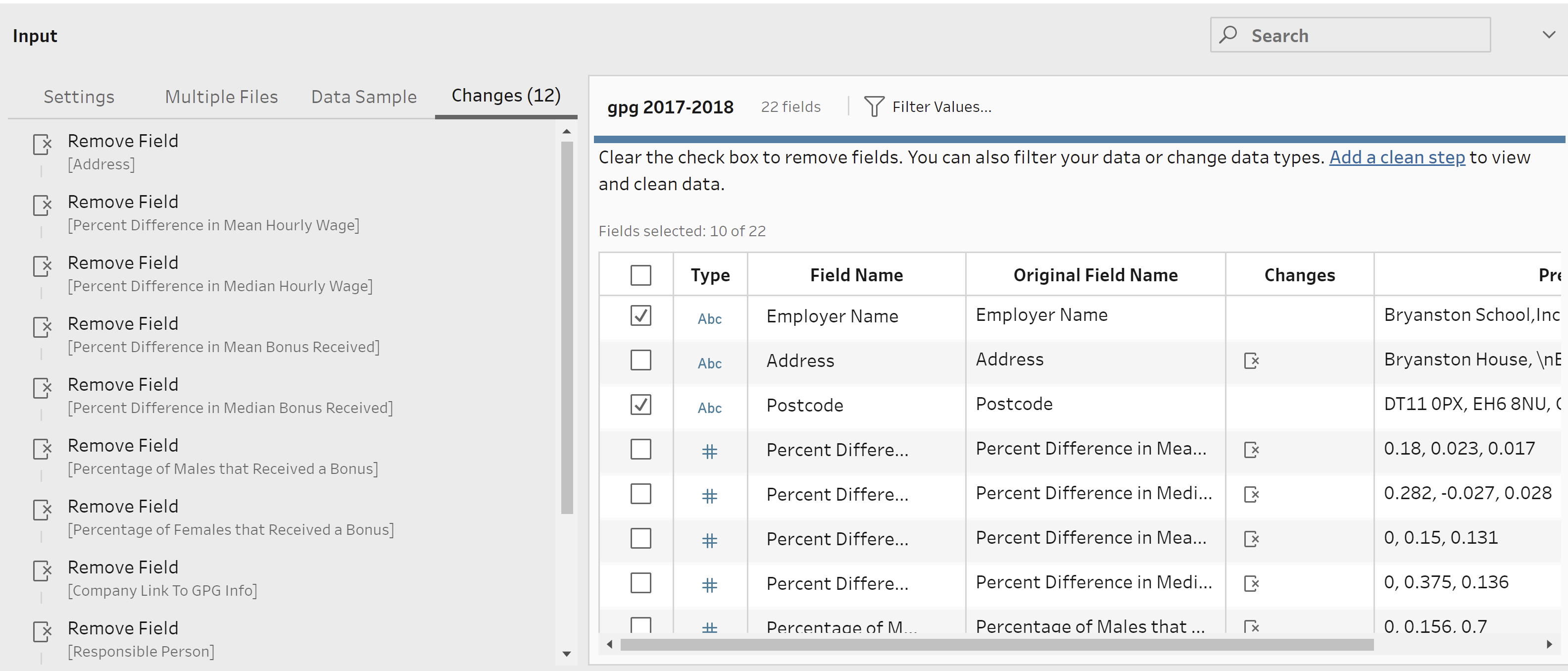
Connecting to the data



Now we need to clean the data

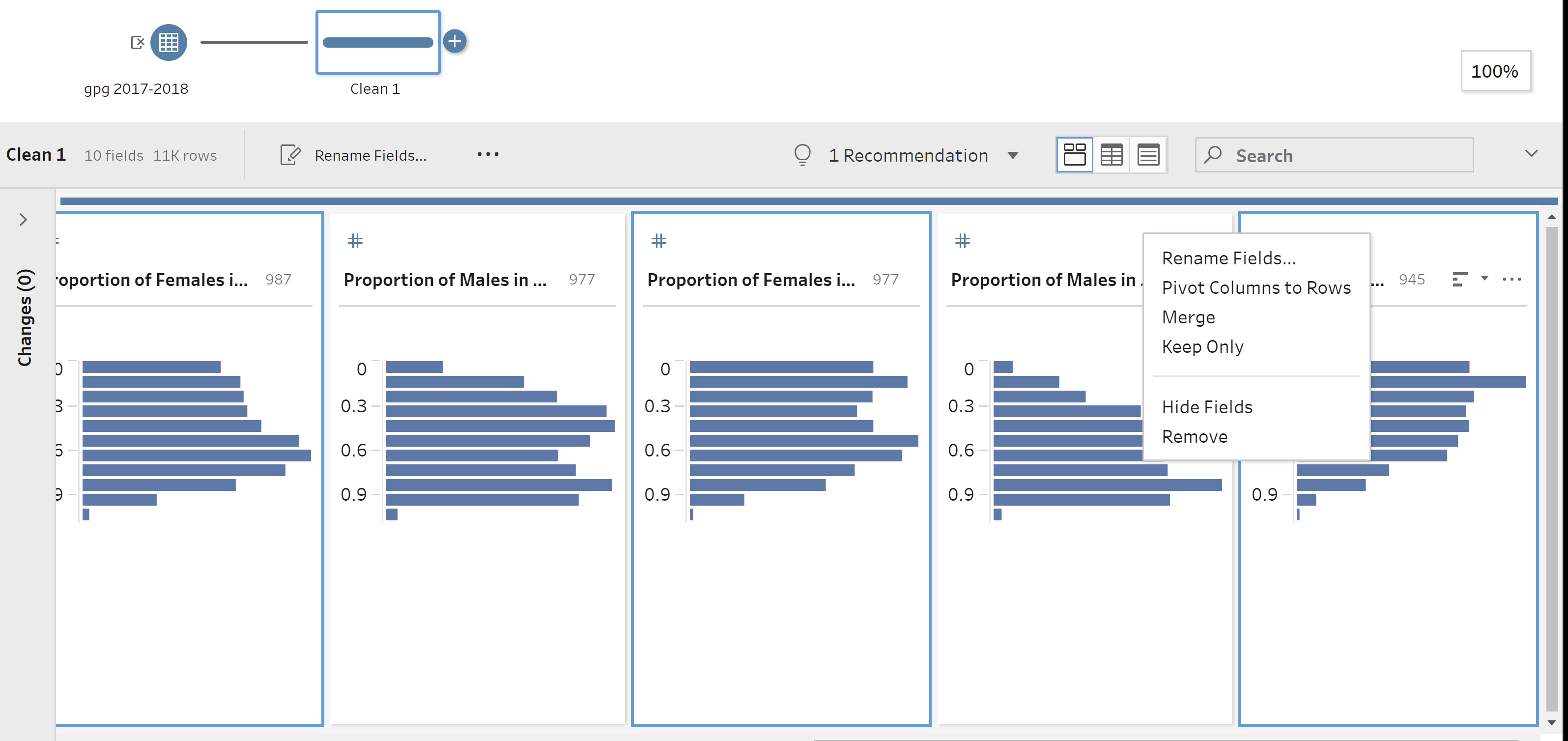


Removing unwanted columns

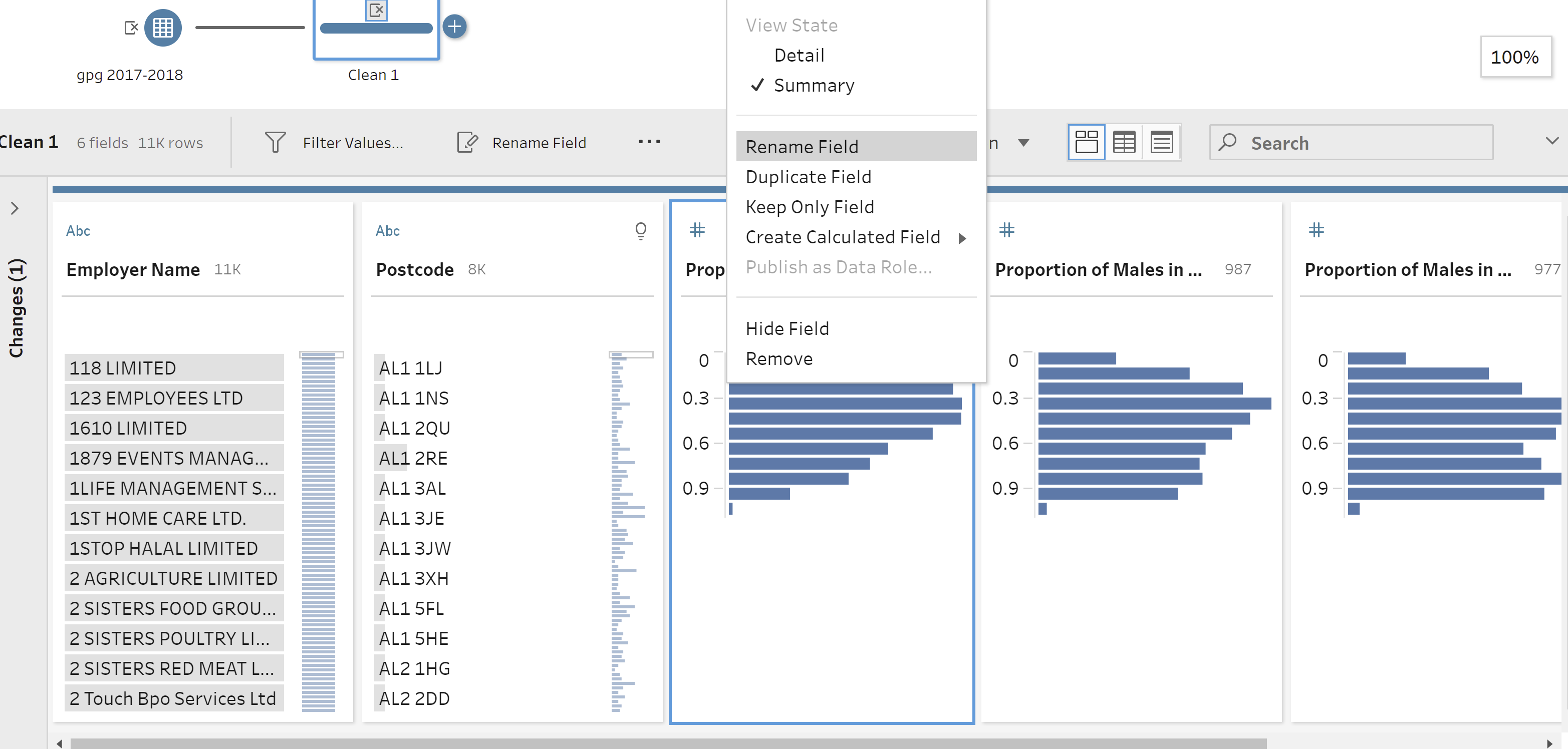


1. **MALE BRANCH**

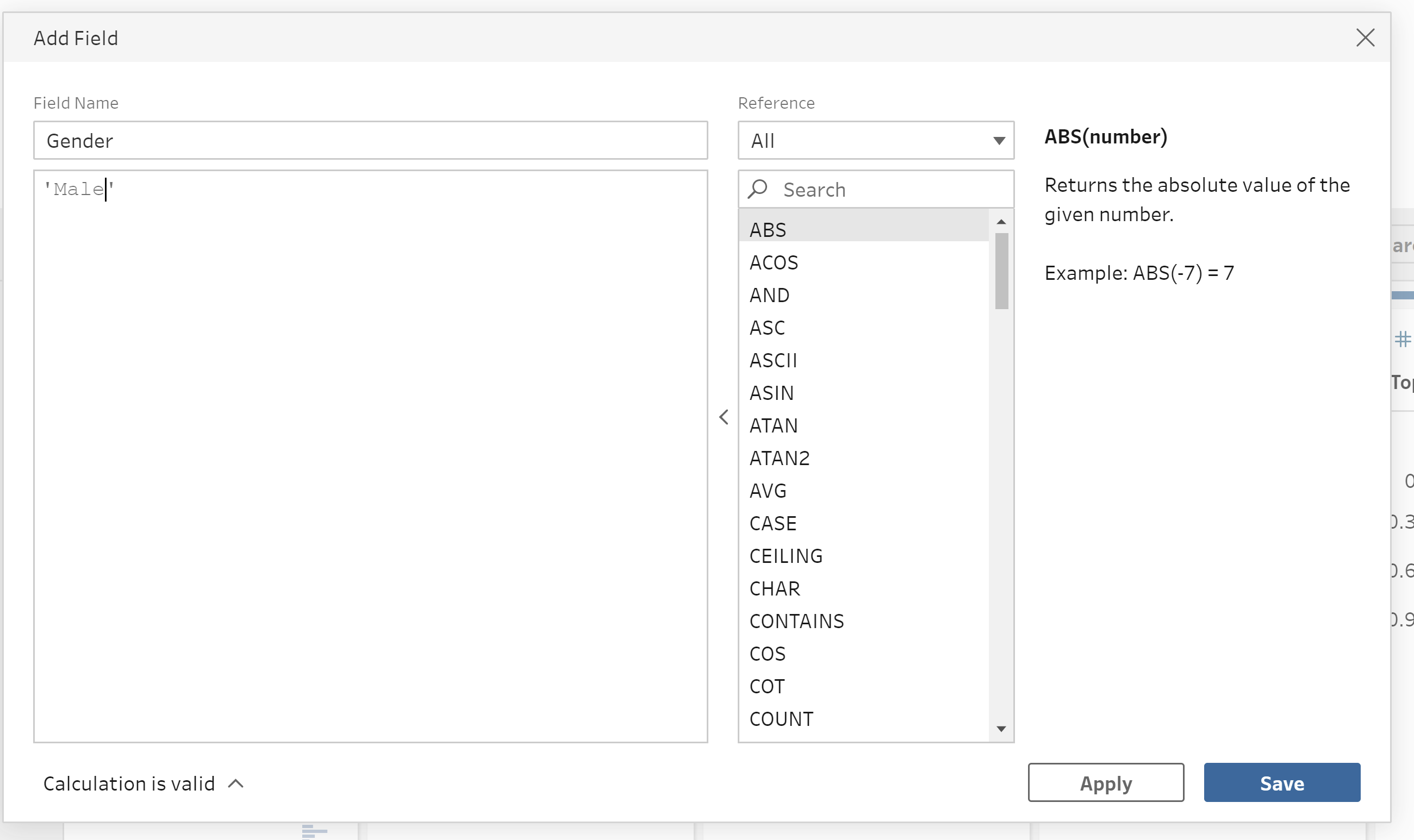
* Removing Female related columns



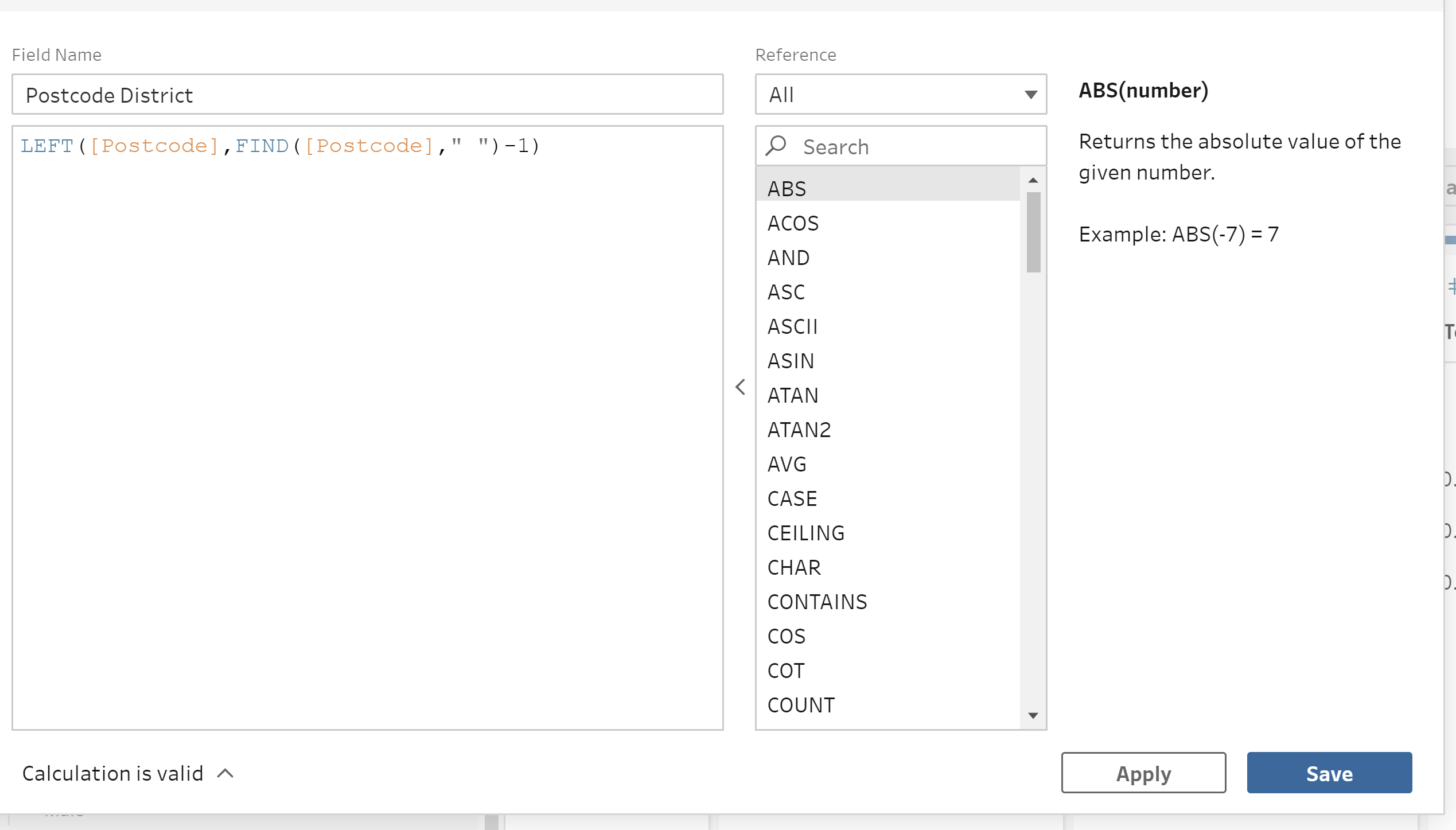
* Now we are renaming few columns



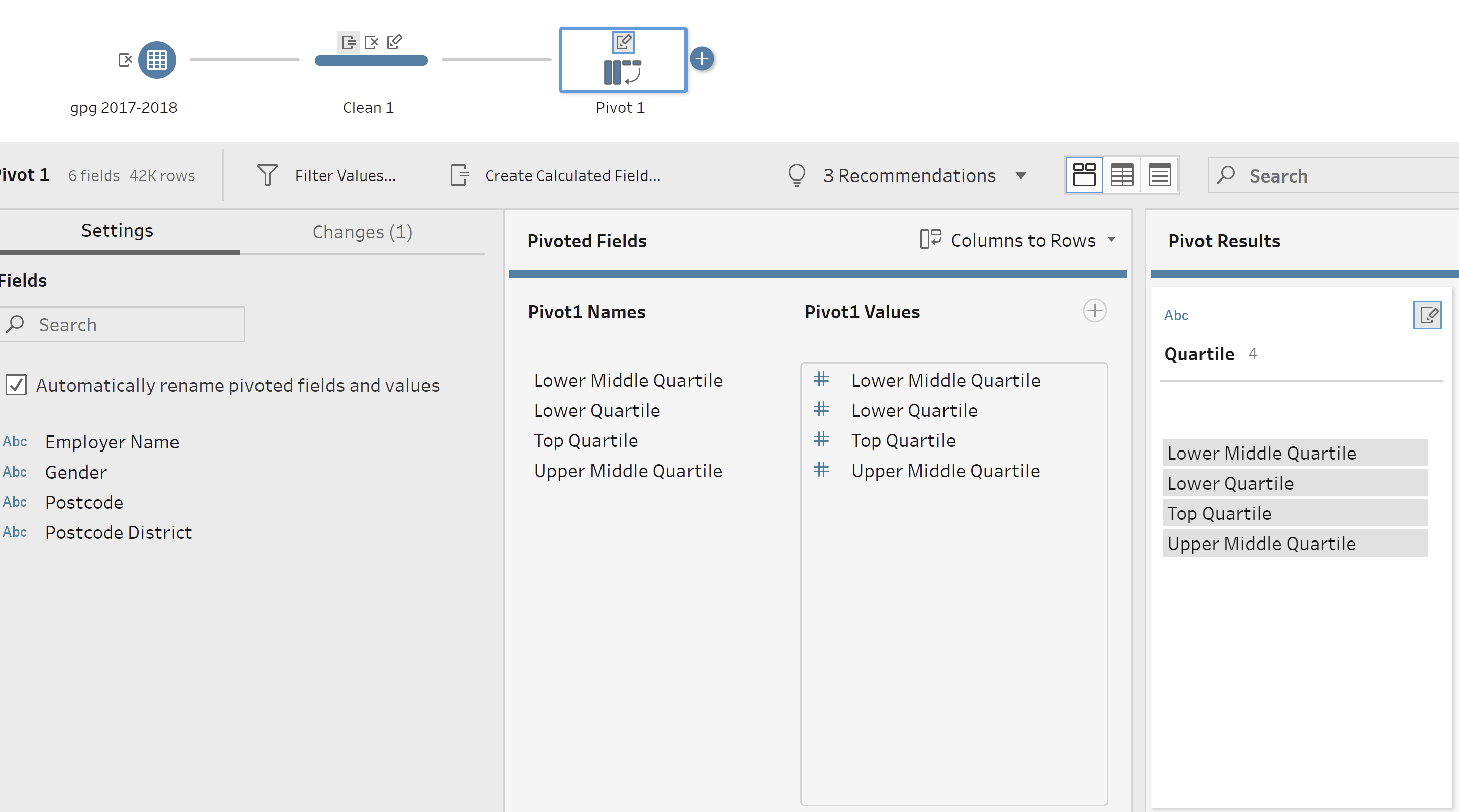
* Creating Gender-Male Column



* Creating new 'Postcode District' column by taking part data from existing column 'Postcode.'



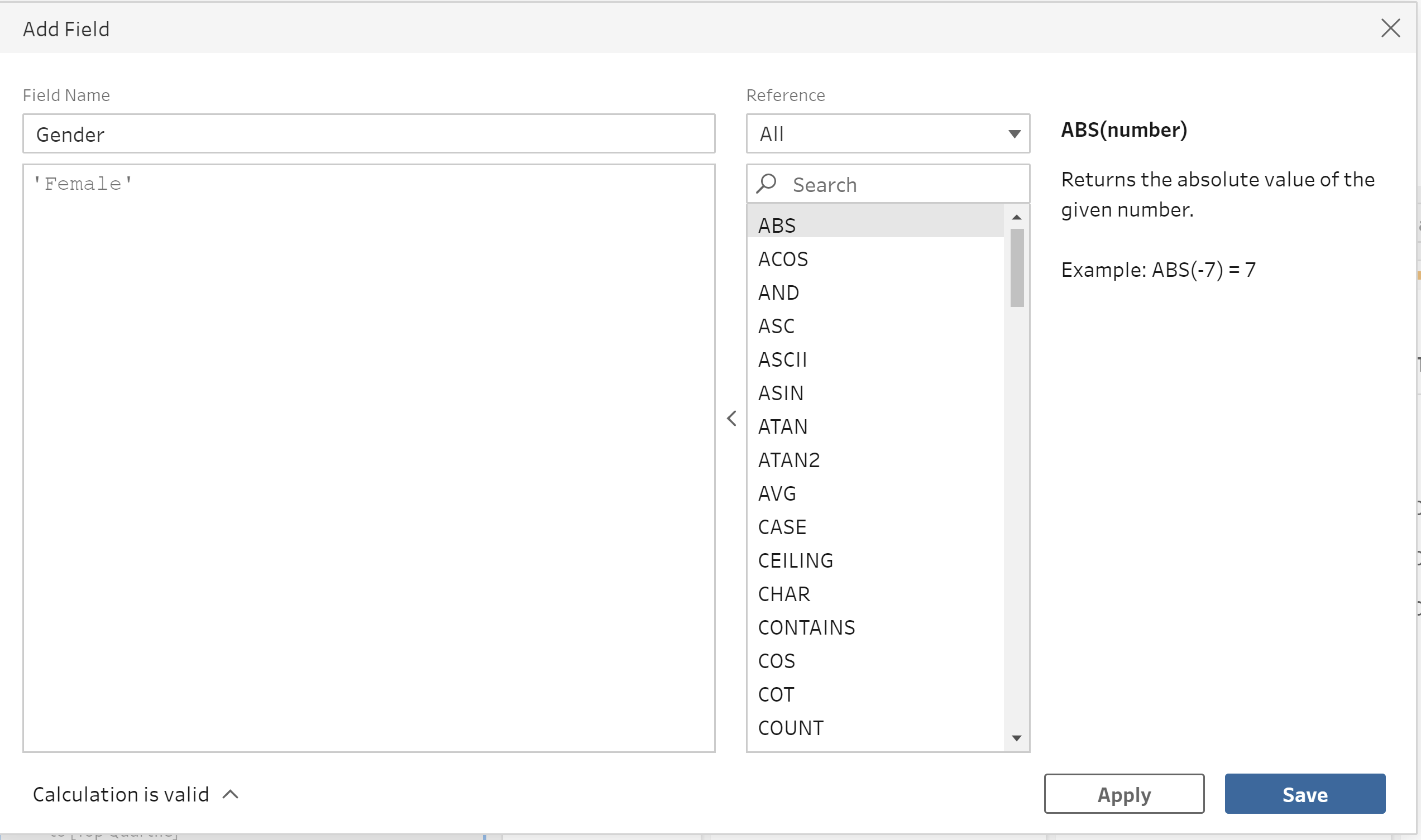
* Creating Pivot table by selecting required columns:



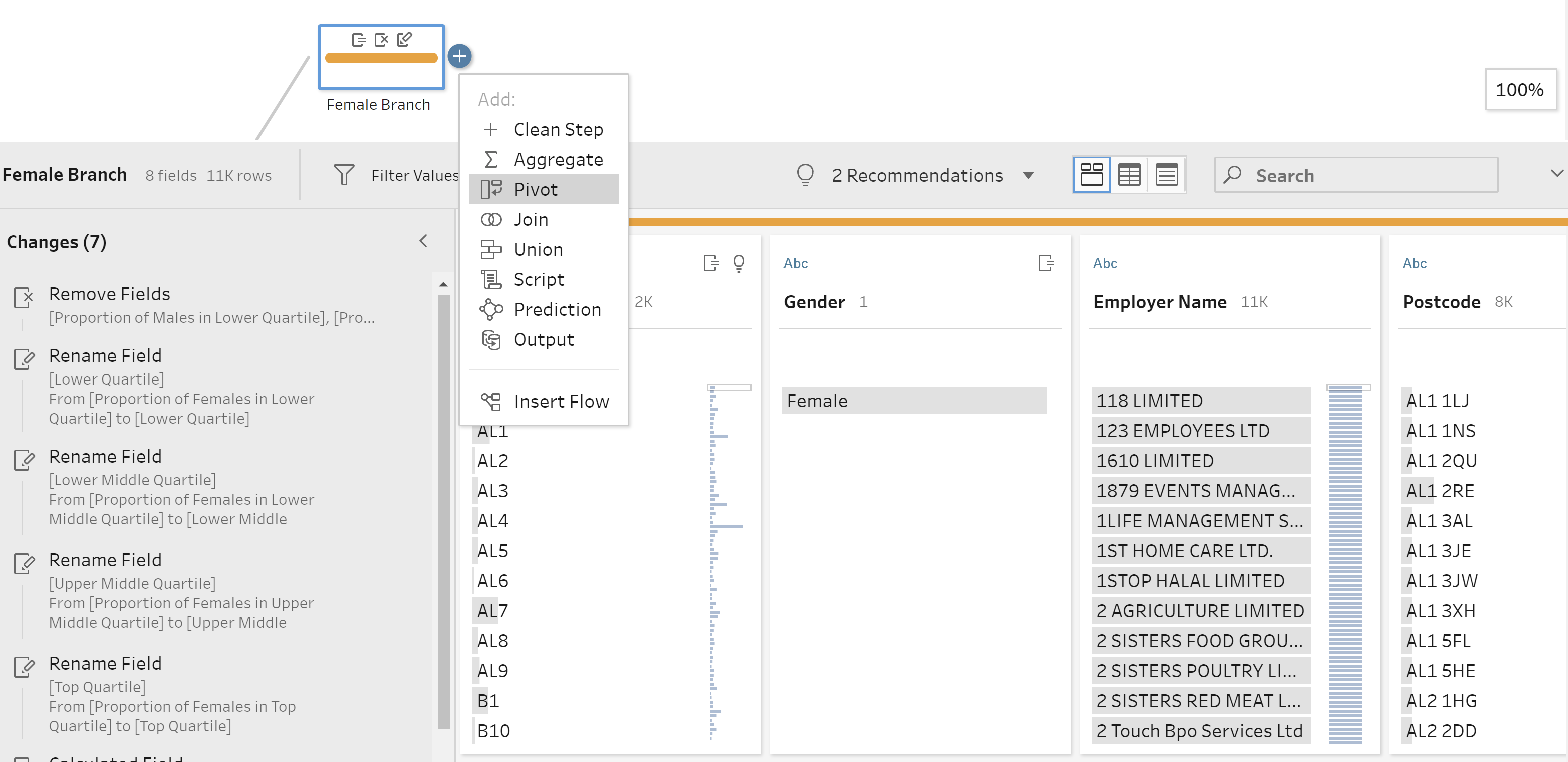
1. FEMALE BRANCH

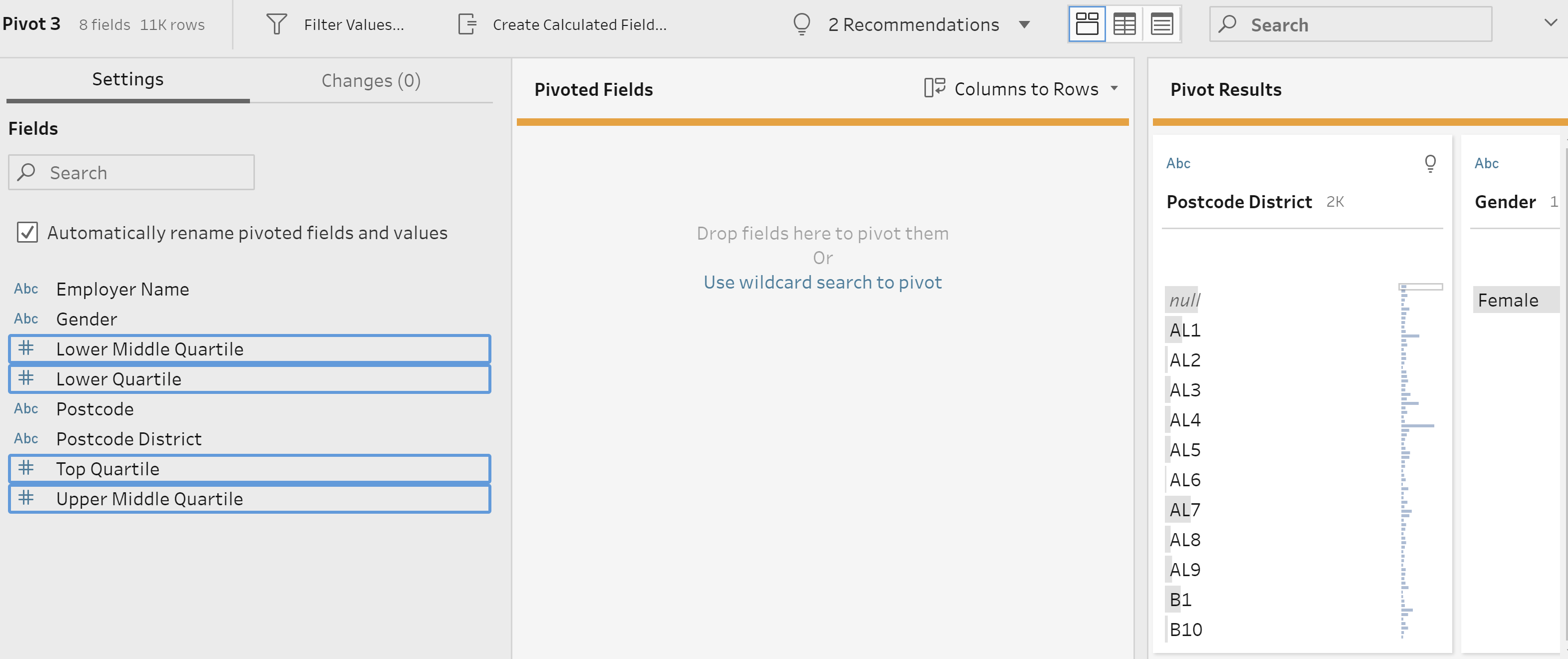
Repeat the steps:

* Removed Male related columns
* Renamed Columns
* Created Gender-Female Column

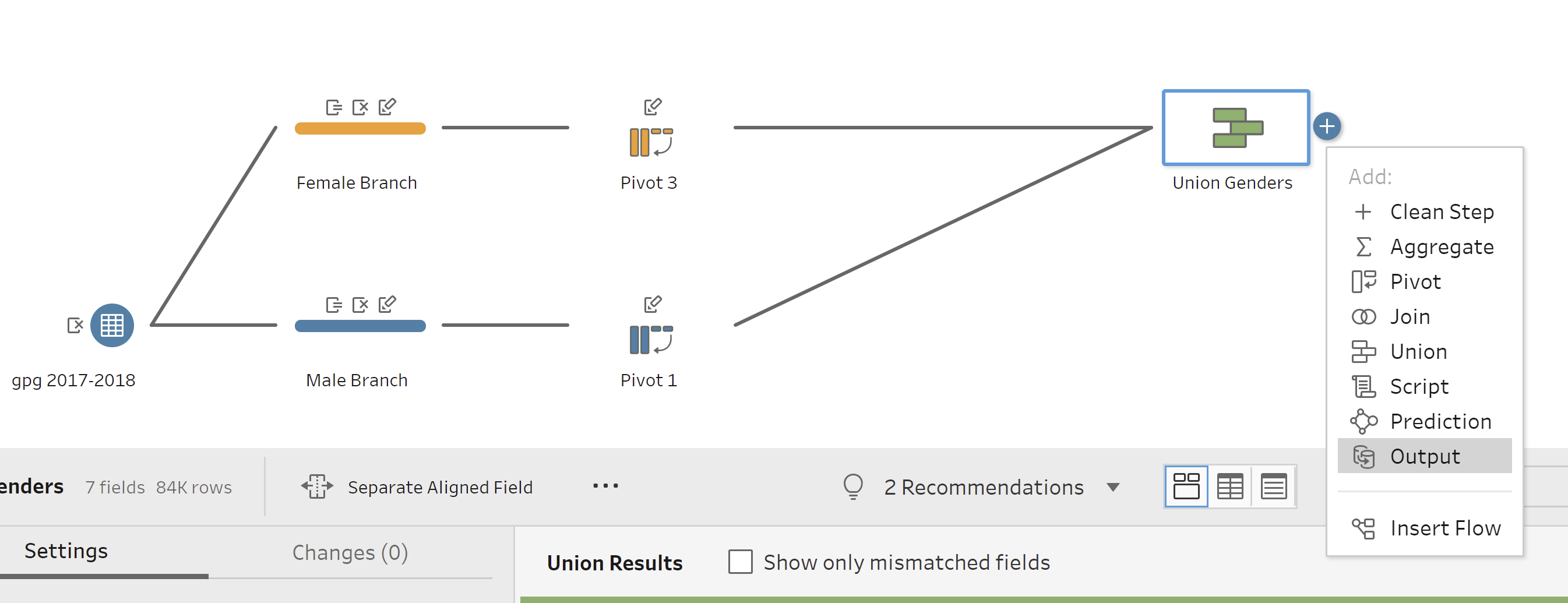


* Created new 'Postcode District' column by taking part data from existing column 'Postcode.'
* Created Pivot table by selecting required columns

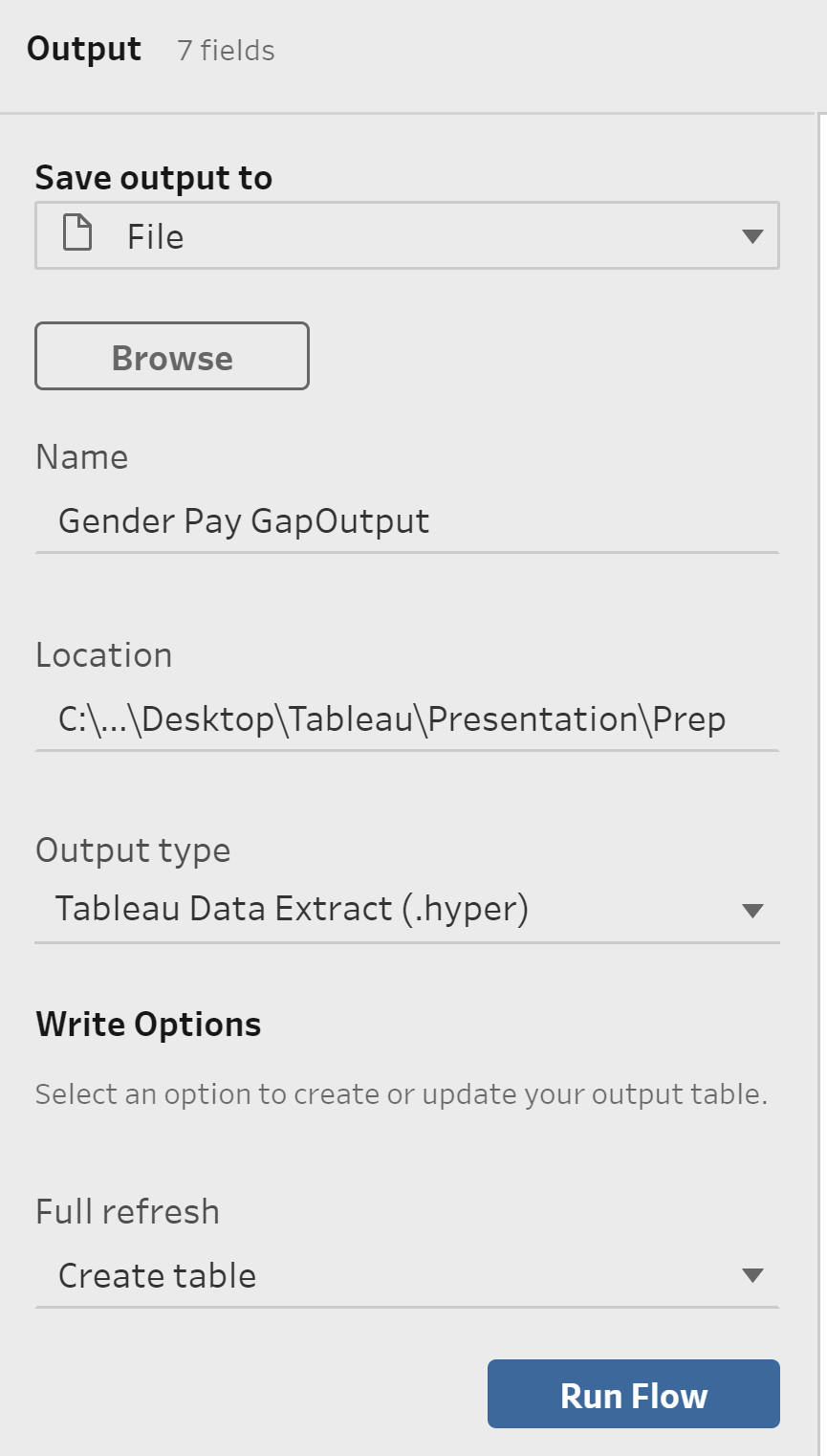




Now we need to create a union between the Male and Female Pivot tables:



Now that we have completed the data processing, we can now export this data for visualization or any other purpose.



Complete workflow in Tableau prep builder:

