

Tableau Prep Project
on
Gender Pay Gap
DATA SCIENCE AND APPLICATION
Of
METRO COLLEGE OF TECHNOLOGY

Submitted to
Prof.Hamid Rajaee

Submitted by
Janah Vijayarathnam



Tableau Prep

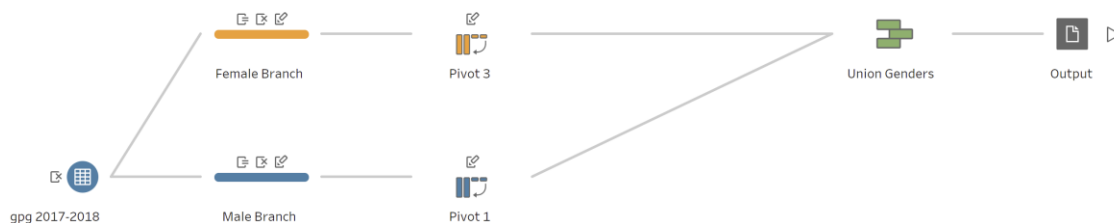
Tableau Prep Builder is a tool in the Tableau product suite designed to make preparing data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

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.

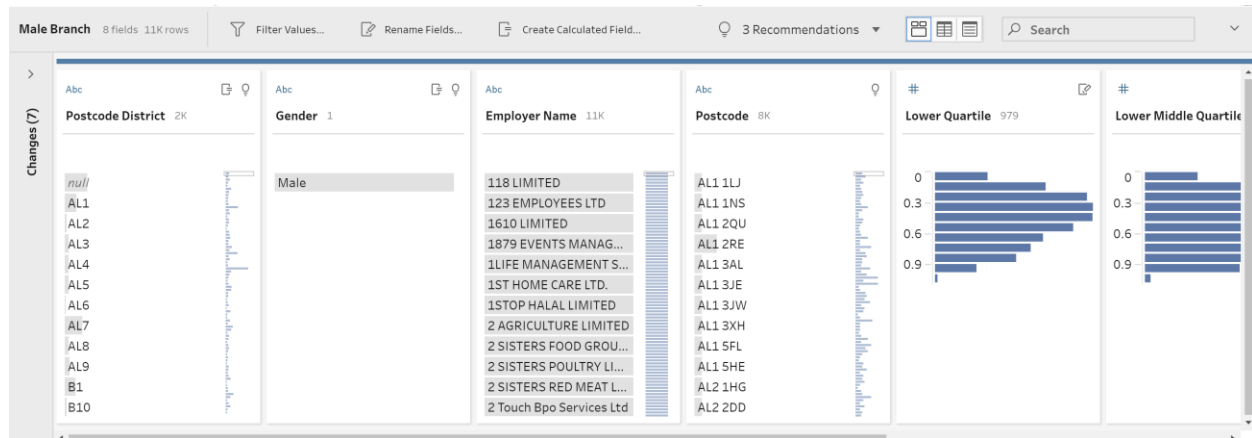
Flow pane

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.



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.



About the dataset:

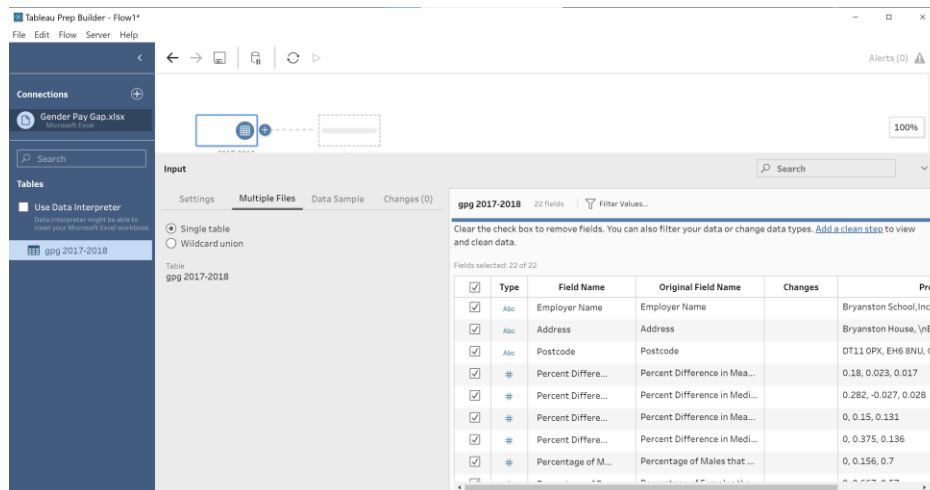
For processing in Tableau prep I have selected an excel file which will be split into 2 parts as

- A. Male Branch and
- B. Female.

The data is about the Gender pay gap.

Steps in data processing :

Step 1: Connecting to the data



Now we need to clean the data



Step 2: Removing unwanted columns

Input

Settings Multiple Files Data Sample **Changes (12)**

gpg 2017-2018 22 fields Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. [Add a clean step](#) to view and clean data.

Fields selected: 10 of 22

<input type="checkbox"/>	Type	Field Name	Original Field Name	Changes	Preview
<input checked="" type="checkbox"/>	Abc	Employer Name	Employer Name		Bryanston School, Inc
<input type="checkbox"/>	Abc	Address	Address		Bryanston House, nE
<input checked="" type="checkbox"/>	Abc	Postcode	Postcode		DT11 0PX, EH6 8NU, C
<input type="checkbox"/>	#	Percent Differe...	Percent Difference in Mea...		0.18, 0.023, 0.017
<input type="checkbox"/>	#	Percent Differe...	Percent Difference in Medi...		0.282, -0.027, 0.028
<input type="checkbox"/>	#	Percent Differe...	Percent Difference in Mea...		0, 0.15, 0.131
<input type="checkbox"/>	#	Percent Differe...	Percent Difference in Medi...		0, 0.375, 0.136
<input type="checkbox"/>	#	Percentage of M...	Percentage of Males that ...		0.0156, 0.7

A. MALE BRANCH

- Removing Female related columns

gpg 2017-2018 Clean 1 100%

Clean 1 10 fields 11K rows Rename Fields... 1 Recommendation Search

Changes (0)

Proportion of Females in ... 987

Proportion of Males in ... 977

Proportion of Females in ... 977

Proportion of Males in ... 945

Rename Fields...
Pivot Columns to Rows
Merge
Keep Only
Hide Fields
Remove

-
- The screenshot displays the Qlik Sense interface with a data model and two bar charts. A context menu is open over the 'Postcode' field in the data model, listing actions like 'Rename Field', 'Duplicate Field', 'Keep Only Field', 'Create Calculated Field', 'Publish as Data Role...', 'Hide Field', and 'Remove'. The charts show 'Proportion of Males in ...' for different categories.

- Add Field

Field Name

Gender

'Male'

Reference

All

Search

ABS

ACOS

AND

ASC

ASCII

ASIN

ATAN

ATAN2

AVG

CASE

CEILING

CHAR

CONTAINS

COS

COT

COUNT

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid

Apply

Save

Field Name: Postcode District

Reference: All

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid ^

Apply Save

- Creating Pivot table by selecting required columns:

gpg 2017-2018 Clean 1 Pivot 1

ivot 1 6 fields 42K rows Filter Values... Create Calculated Field... 3 Recommendations Search

Settings Changes (1)

Fields

Automatically rename pivoted fields and values

Abc Employer Name
Abc Gender
Abc Postcode
Abc Postcode District

Pivoted Fields

Pivot1 Names

Lower Middle Quartile
Lower Quartile
Top Quartile
Upper Middle Quartile

Pivot1 Values

Lower Middle Quartile
Lower Quartile
Top Quartile
Upper Middle Quartile

Pivot Results

Abc

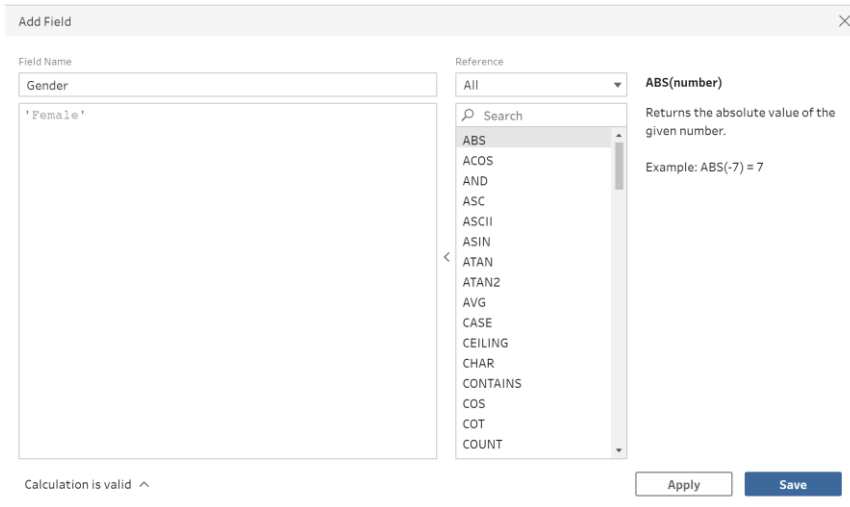
Quartile 4

Lower Middle Quartile
Lower Quartile
Top Quartile
Upper Middle Quartile

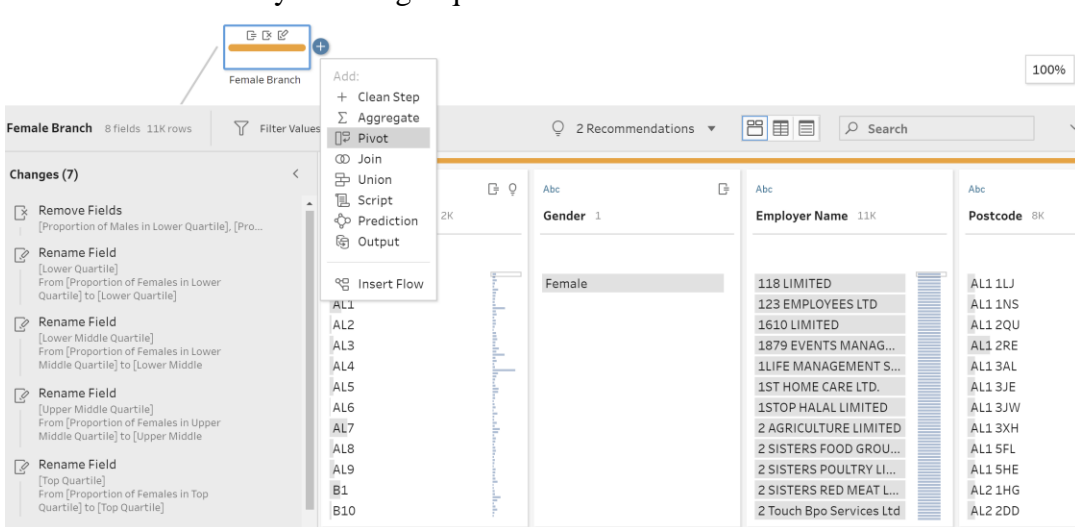
B. FEMALE BRANCH

Repeat the steps:

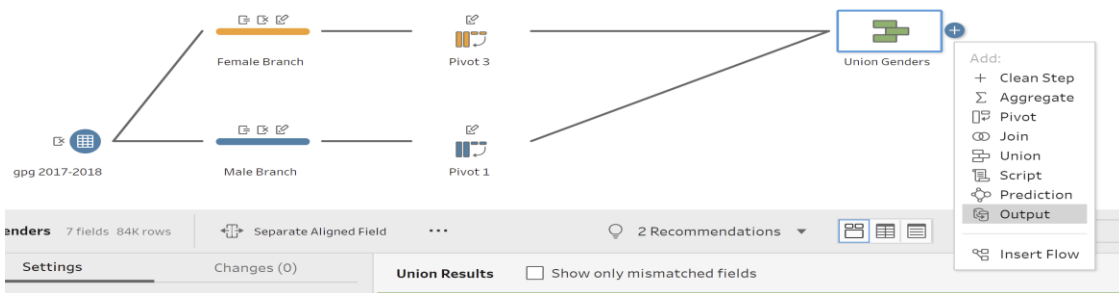
- Removed Male related columns
- Rename Columns
- Created Gender-Female Column



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



Step 4: 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:
-

