Module 1: Data Preparation Using Tableau Prep

Demo Document I

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Demo I

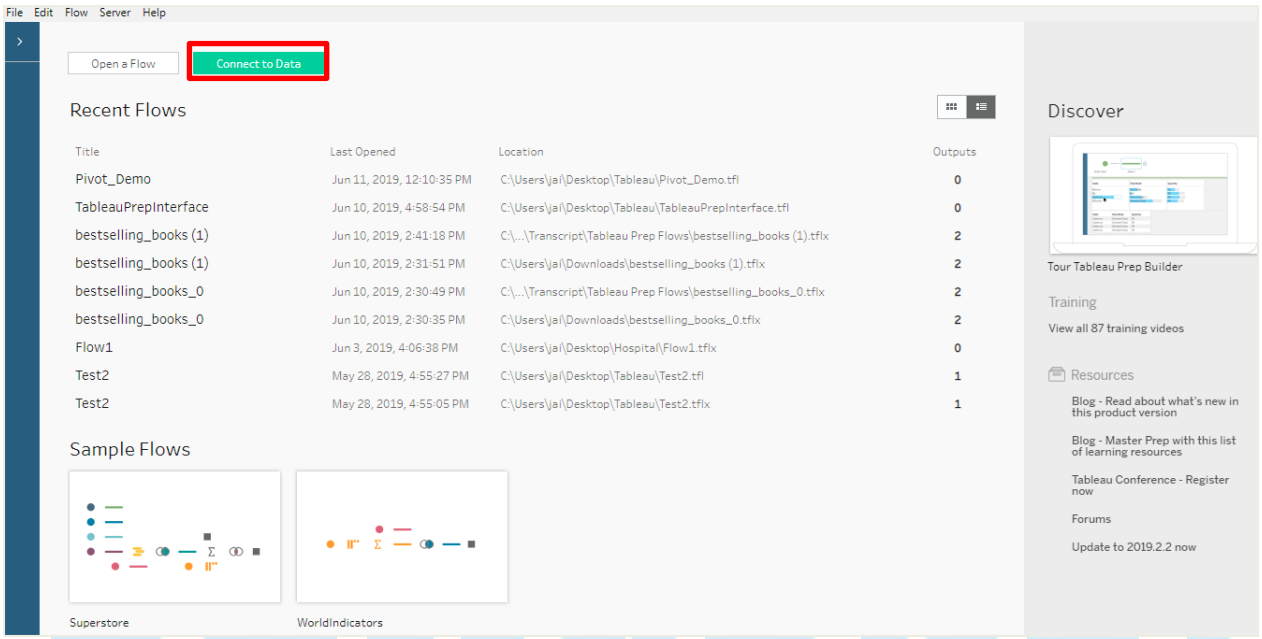
Group and Replace Feature in Tableau Prep

A common issue that we come across in data cleanliness is the presence of multiple values instead of a single value, such as WI and West Indies. Group and Replace feature in Tableau Prep allows one to group multiple values and replace them with a single value. In this demo, we have a dataset named “Group and Replace” to illustrate the Group and Replace feature in Tableau Prep.

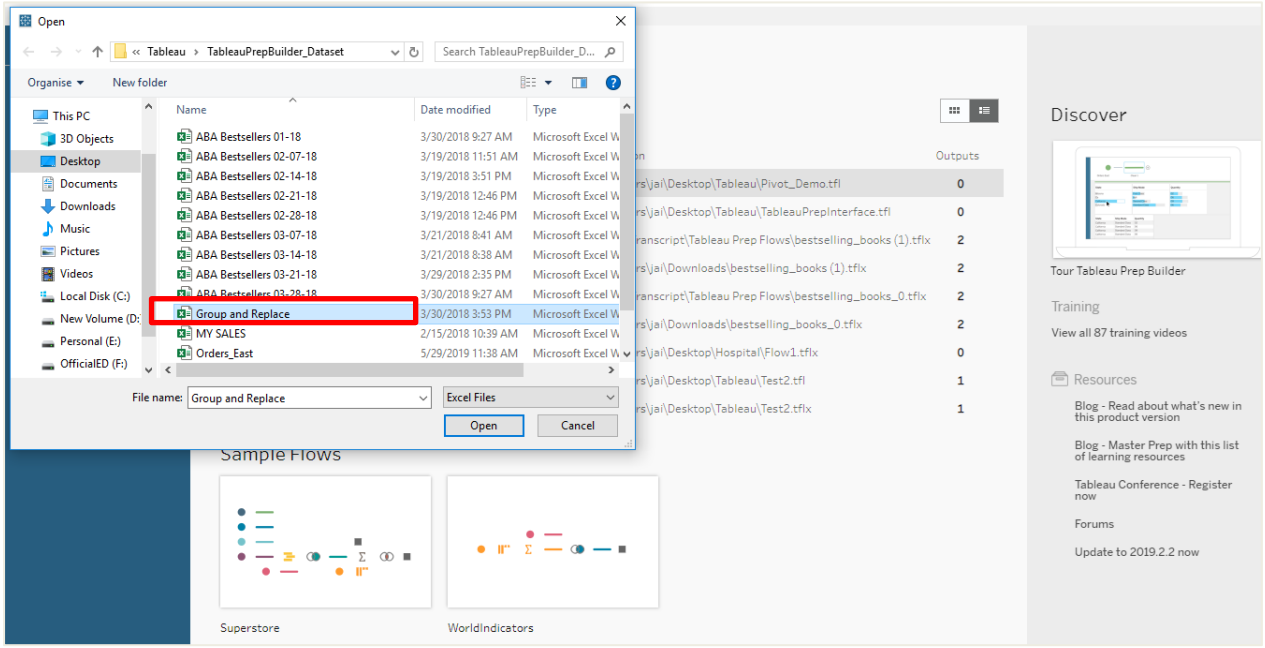
	A	B	C	D	E
1	**Airbnb listings**	**Misspellings**	**Name formatting**	**Typo**	
2	bedroom	beautiful	Dorian Gray	Tableau	
3	bedrooms	beautiful	Lord Henry Wotton	Tableau	
4	bd	beautiful	Basil Hallward	Tableau	
5	bed	beautiful	Sibyl Vane	Tableau	
6	BED	beautiful	James Vane	Tableau	
7	bed	beautiful	Campbell, Alan	Tableau	
8	bedrm	beautiful	Lady Agatha	Tableau	
9	bdrm	beautiful	Lord Fermor	Tableau	
10	bedroom	beautiful	Victoria Wotton	Tableau	
11	bedroom	beautiful	Duchess of Monmout	Tableau	
12	bedroom	beautiful	Gray, Dorian	Tabelau	
13	bedroom	beautiful	Dorian Gray	Tableau	
14	bedroom	beautyfull	Dorian Gray	Tableau	
15	bedroom	beautyfull	Dorian Gray	Tableau	
16	bedroom	beautyfull	Dorian Gray	Tableau	
17	bedroom	beautyfull	Dorian Gray	Tableau	
18	bedroom	beatiful	Dorian Gray	Tableau	
19	bed	beatiful	Dorian Gray	Tableau	

Demo I – Solution

Step 1: Connect to the dataset

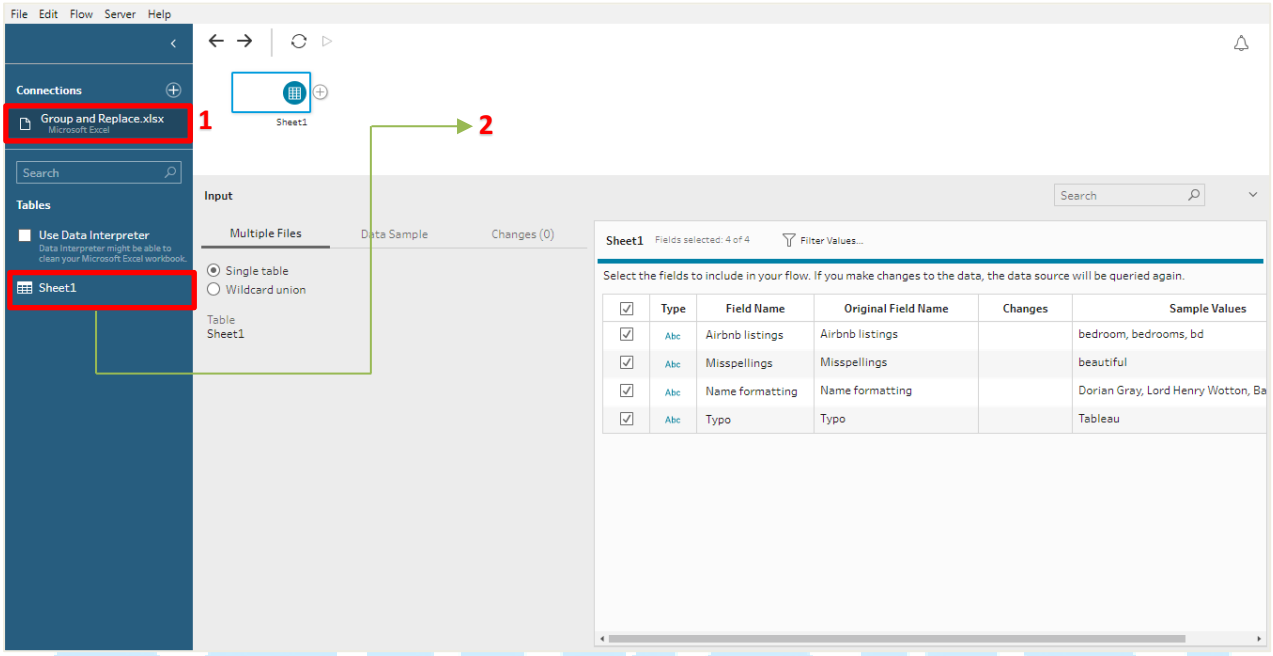


Step 2: Click on Microsoft Excel, and select “Group and Replace” dataset

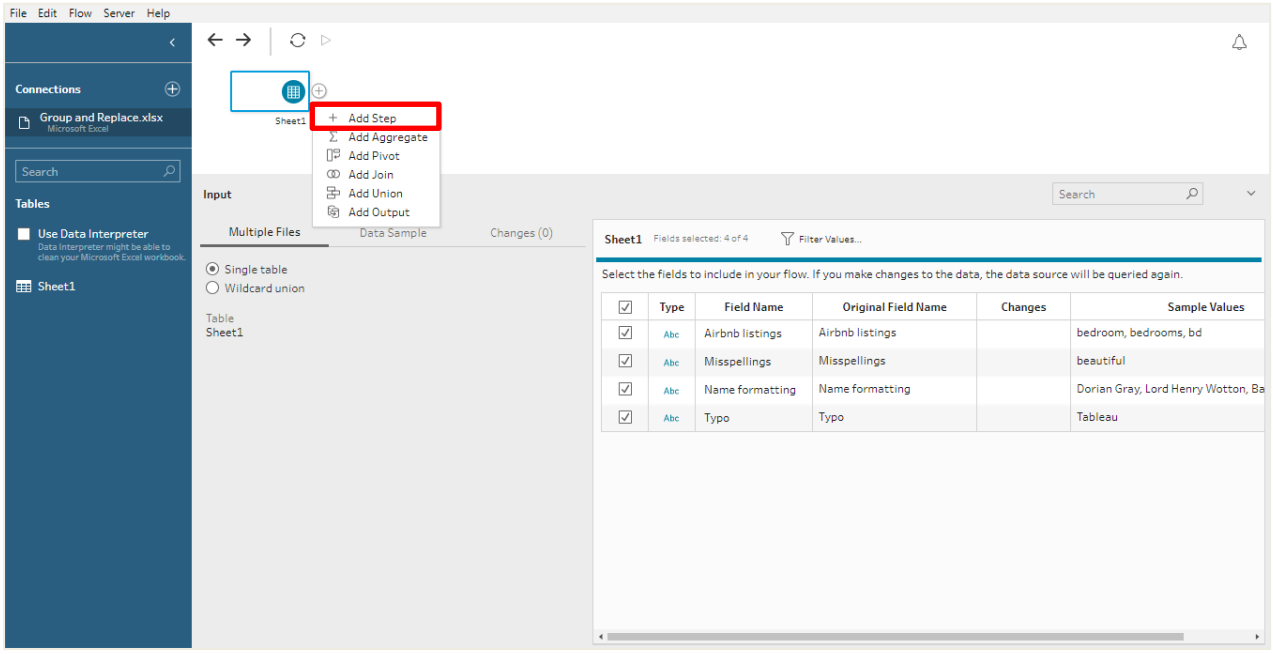


Step 3: 1 → Connection established with file: “Group and Replace”

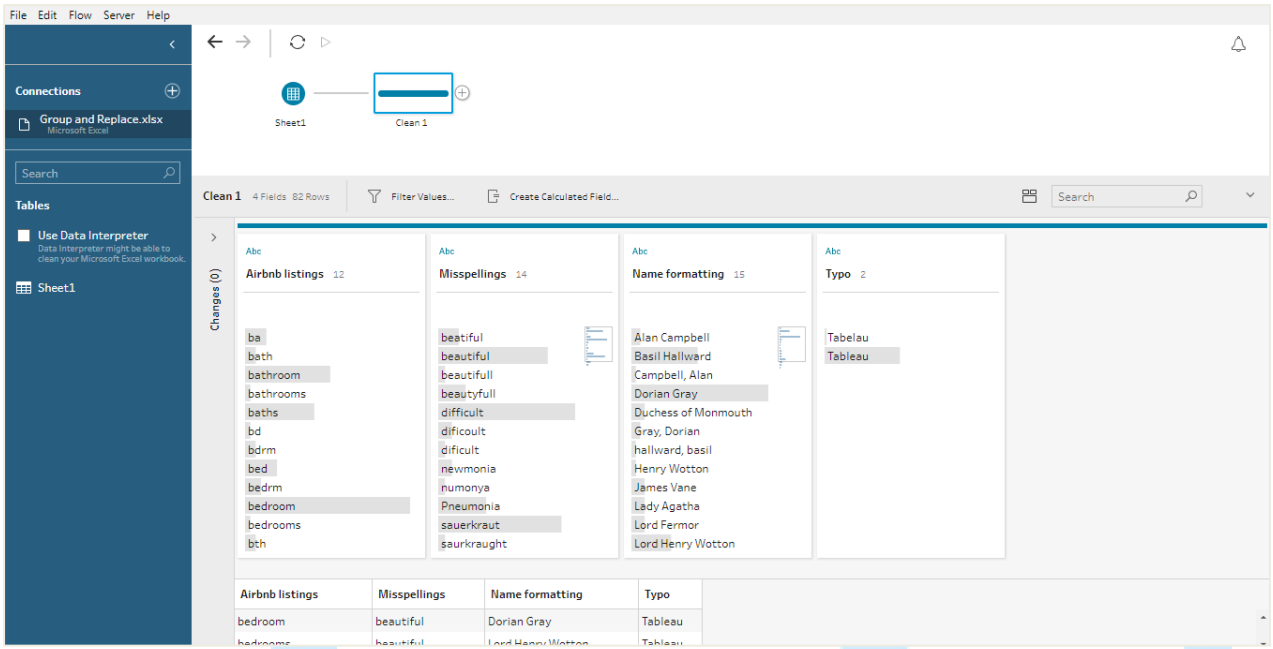
2 → Drag and drop the excel sheet/table here



Step 4: Click on “+” button and add a Cleaning step

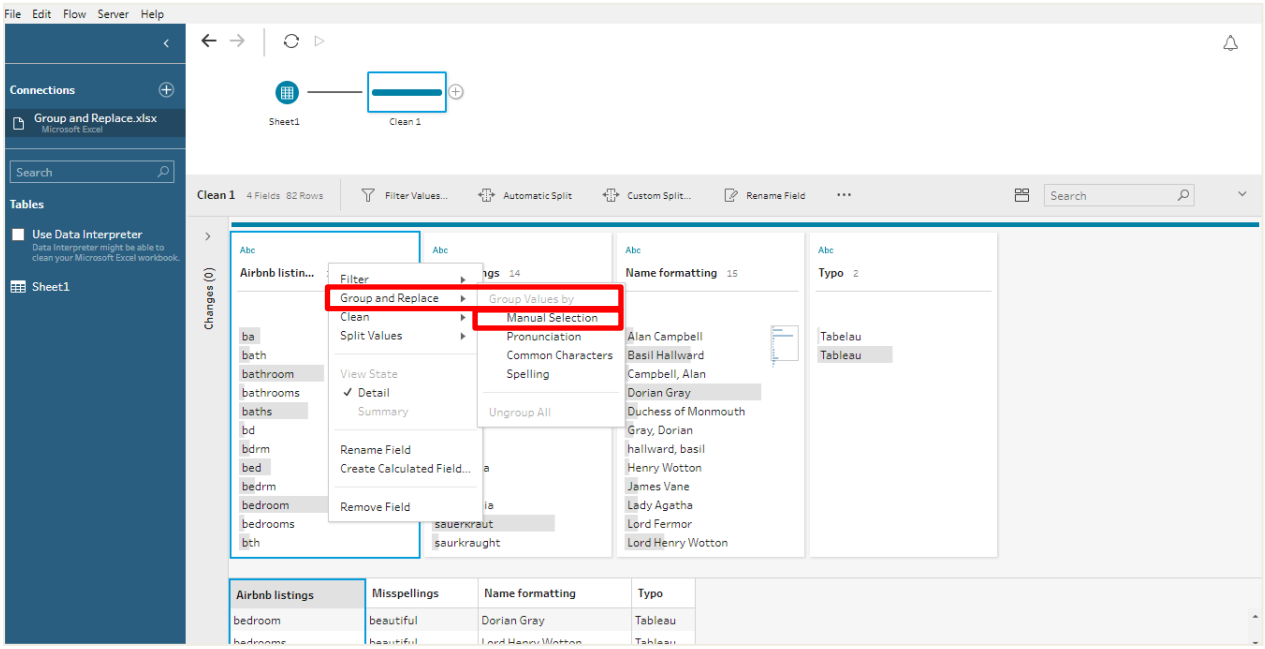


Here, we can observe that there exists a data cleanliness issue in the “Airbnb listings” field.



edureka!

Step 5: Right click on the card for the “Airbnb listings” field. Under “Group and Replace” option, choose the “Manual Selection” option.

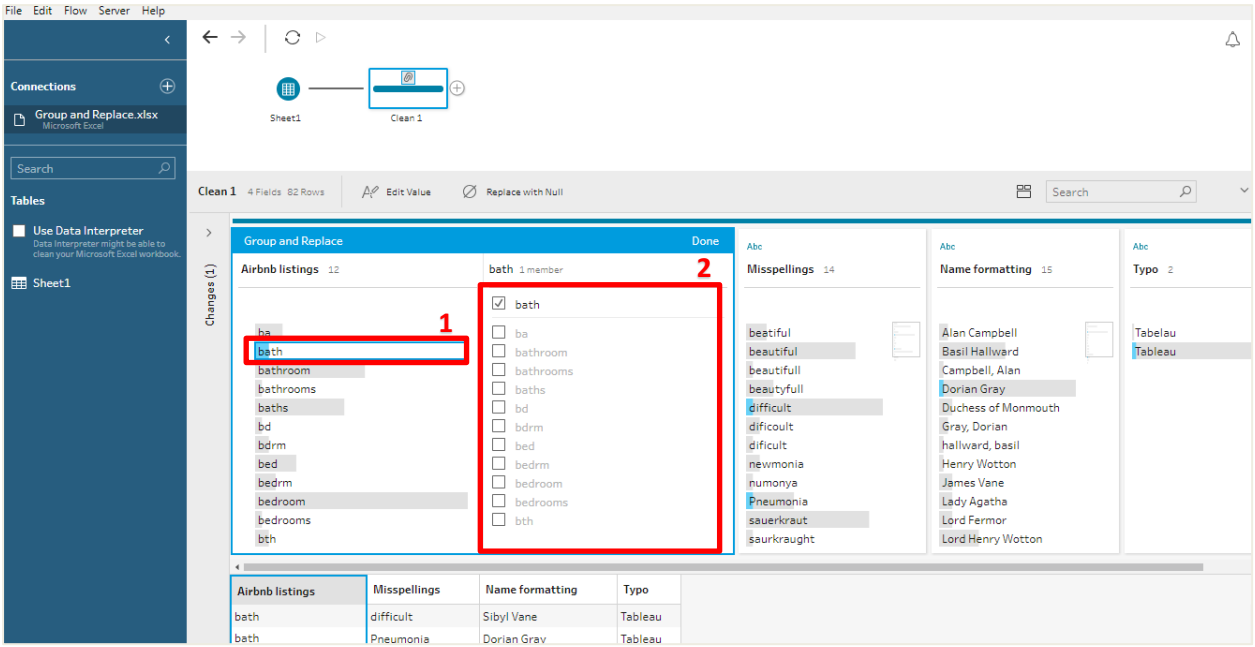


Step 6: 1 → The value that we choose first becomes the replacing value.

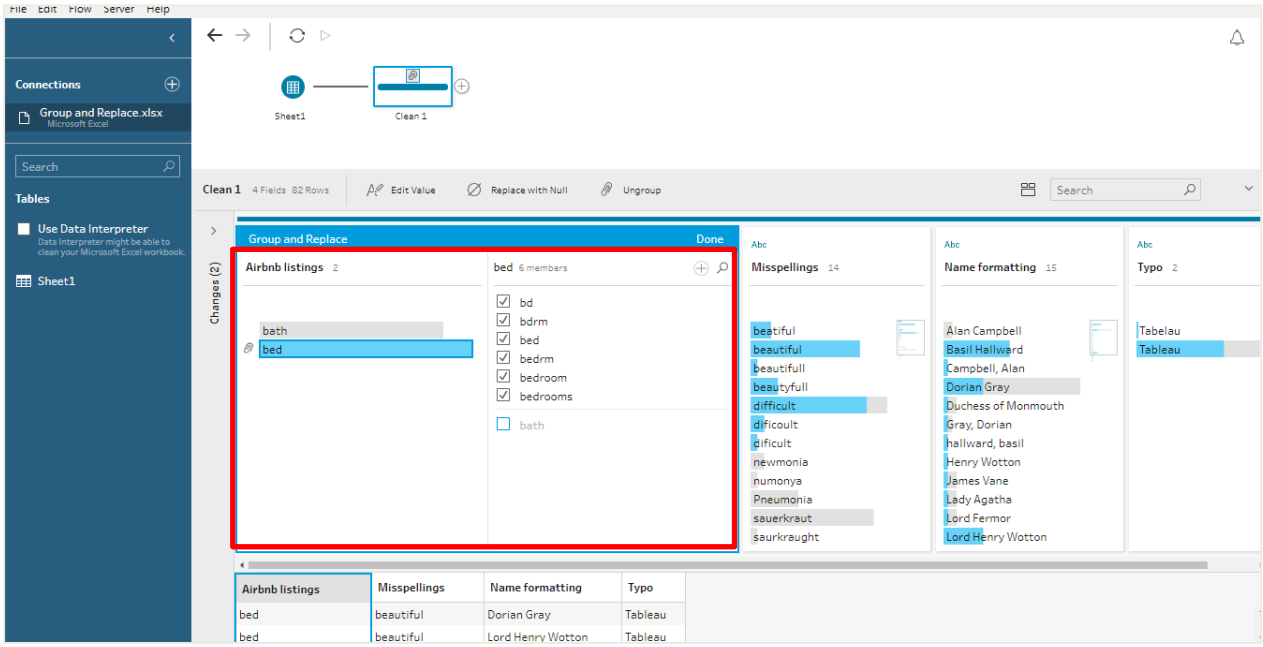
Select the word “bath”.

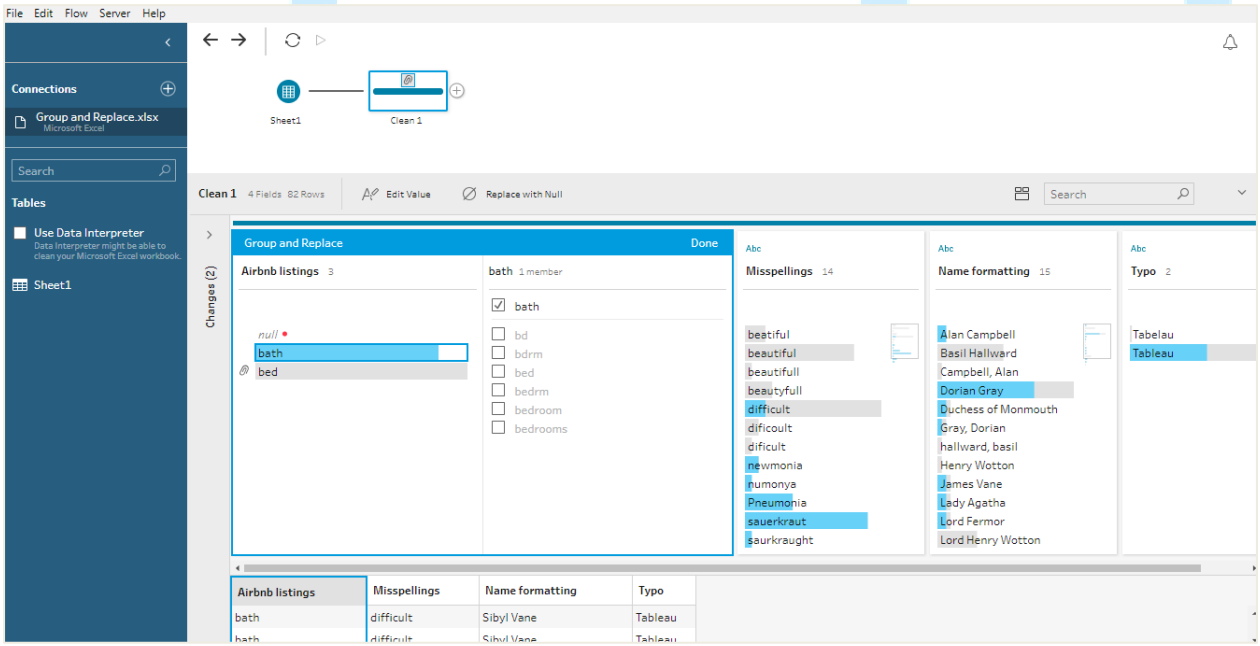
2 → With the replacing value selected, the remaining values are displayed here.

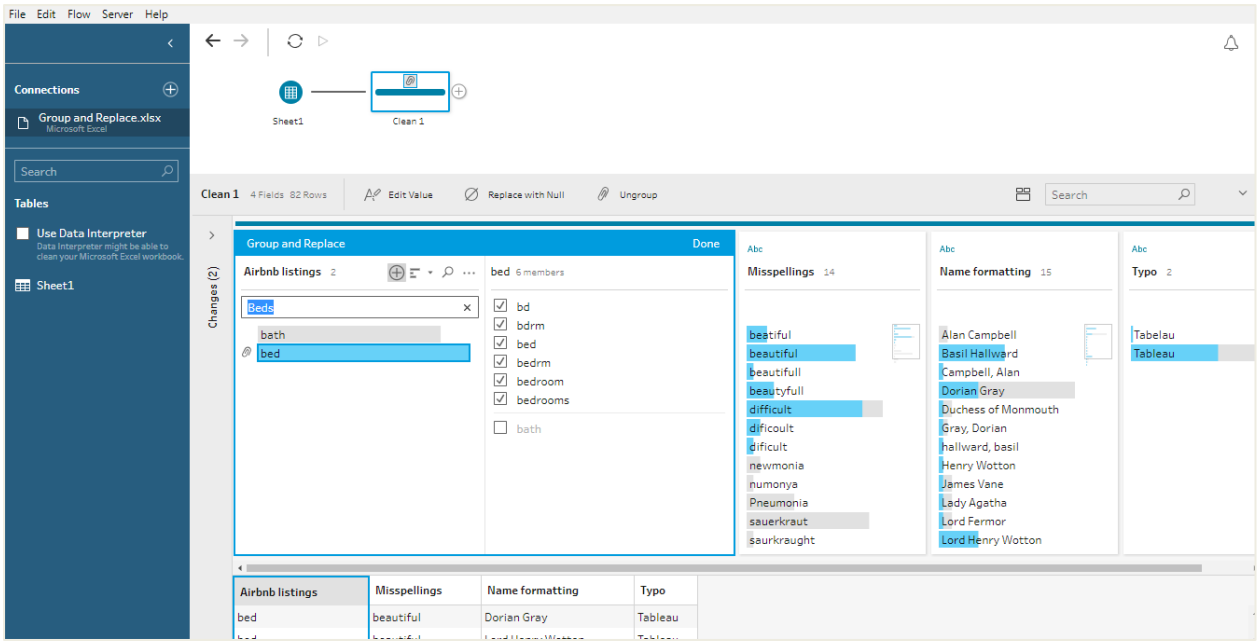
Select all the values that should be grouped under the replacing value. Left side of the editor gives a preview of the new values for the field with a grouping paperclip icon besides the word “bath”.

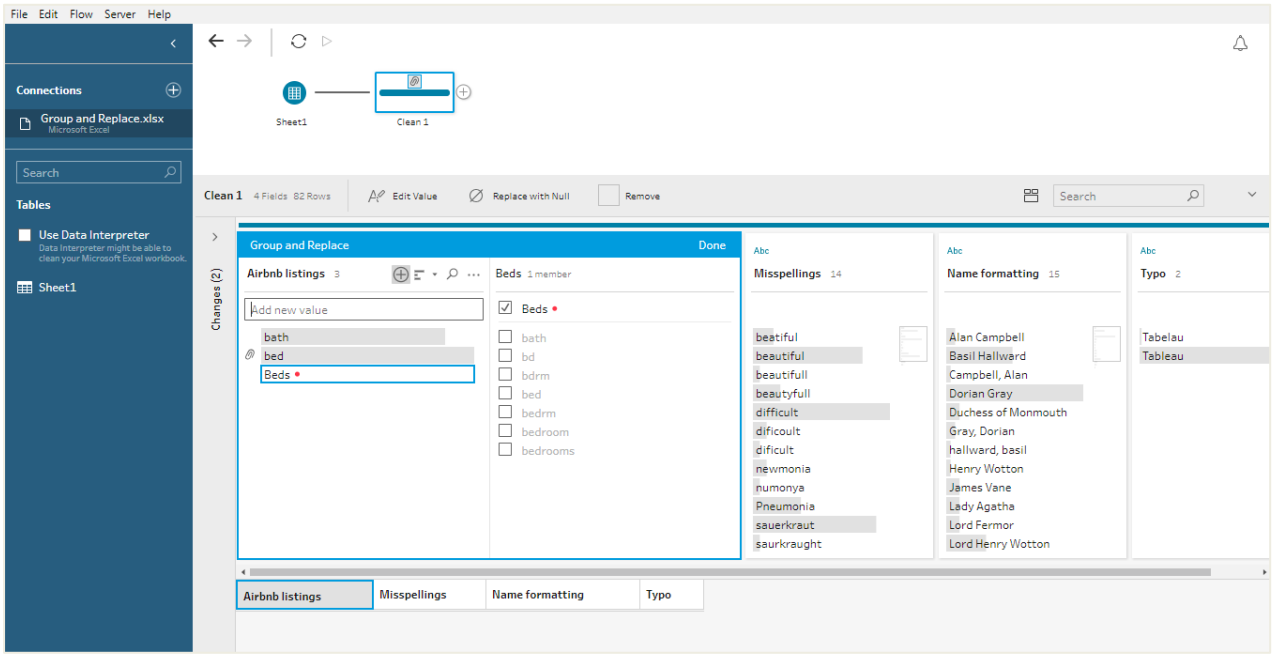


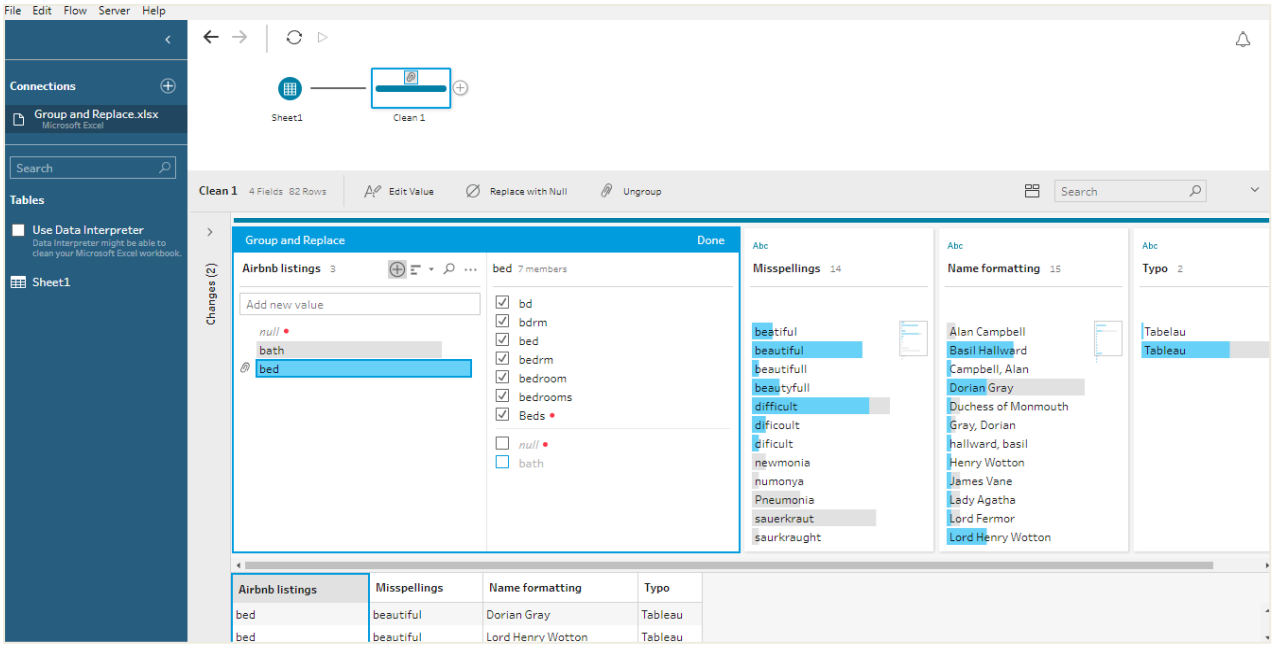
Step 7: Repeat Step 6 for the word “bed” in the “Airbnb listings” field.



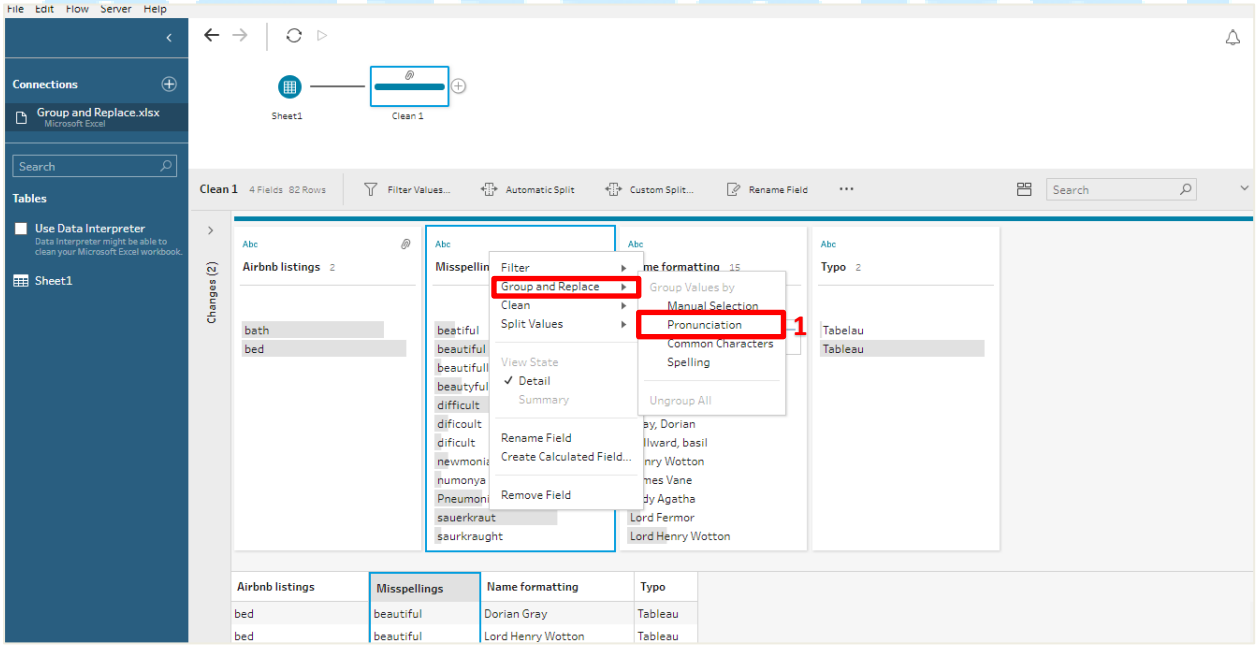



File Edit Flow Server Help
Connections
Group and Replace.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Clean 1 4 Fields 82 Rows
Edit Value
Replace with Null
Ungroup
Search
Changes (2)
Group and Replace
Done
Airbnb listings 2
bed 6 members
Beds
bath
bed
bd
bdrm
bed
bedrm
bedroom
bedrooms
bath
Misspellings 14
beatiful
beautiful
beautifull
beautyfull
difficult
dificoult
dificult
newmonia
numonya
Pneumonia
sauerkraut
saurkraught
Name formatting 15
Alan Campbell
Basil Hallward
Campbell, Alan
Dorian Gray
Duchess of Monmouth
Gray, Dorian
hallward, basil
Henry Wotton
James Vane
Lady Agatha
Lord Fermor
Lord Henry Wotton
Typo 2
Tabelau
Tableau
Airbnb listings
Misspellings
Name formatting
Typo
bed
beautiful
Dorian Gray
Tableau


File Edit Flow Server Help
Connections
Group and Replace.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Clean 1 4 Fields 82 Rows
Edit Value
Replace with Null
Remove
Search
Changes (2)
Group and Replace
Done
Airbnb listings 3
Beds 1 member
Add new value
bath
bed
Beds
Beds
bath
bd
bdrm
bed
bedrm
bedroom
bedrooms
Misspellings 14
beatiful
beautiful
beautifull
beautyfull
difficult
dificoult
dificult
newmonia
numonya
Pneumonia
sauerkraut
saurkraught
Name formatting 15
Alan Campbell
Basil Hallward
Campbell, Alan
Dorian Gray
Duchess of Monmouth
Gray, Dorian
hallward, basil
Henry Wotton
James Vane
Lady Agatha
Lord Fermor
Lord Henry Wotton
Typo 2
Tabelau
Tableau
Airbnb listings
Misspellings
Name formatting
Typo



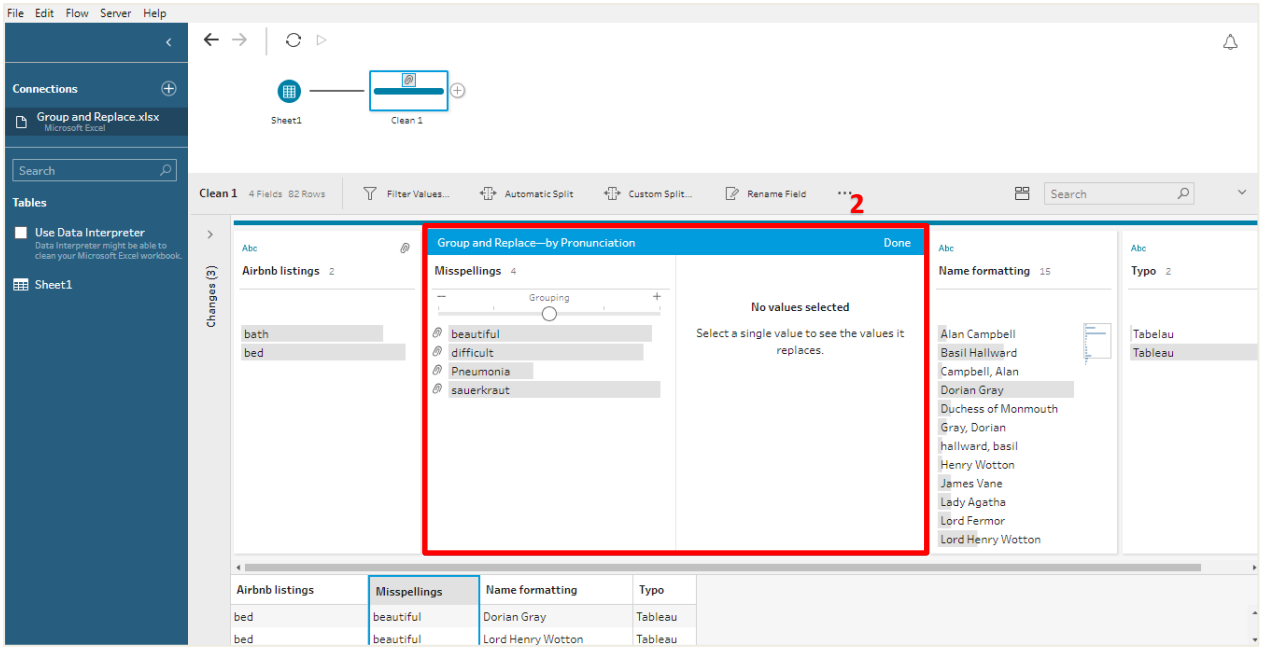
Step 8: 1 → Select the Pronunciation option under "Group and Replace" menu, which automatically addresses the misspelled words in the Misspellings column.



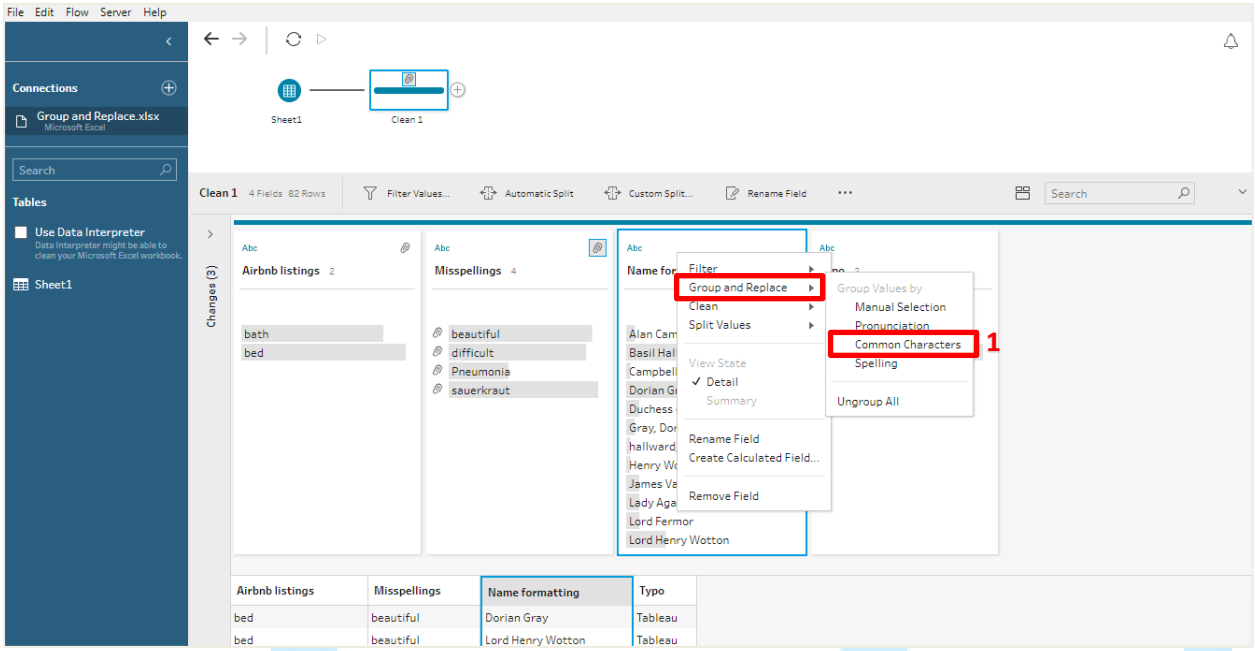
2 → Grouping is automatic here.

- Towards the left, the new groups are displayed.

- If the replaced value does not match the desired value, right click on the new group and select “Edit Value” and modify it to the desired value.
- When you click on a group, right side of the editor exposes the values in the group.
- Manually add to or remove from the group if desired.

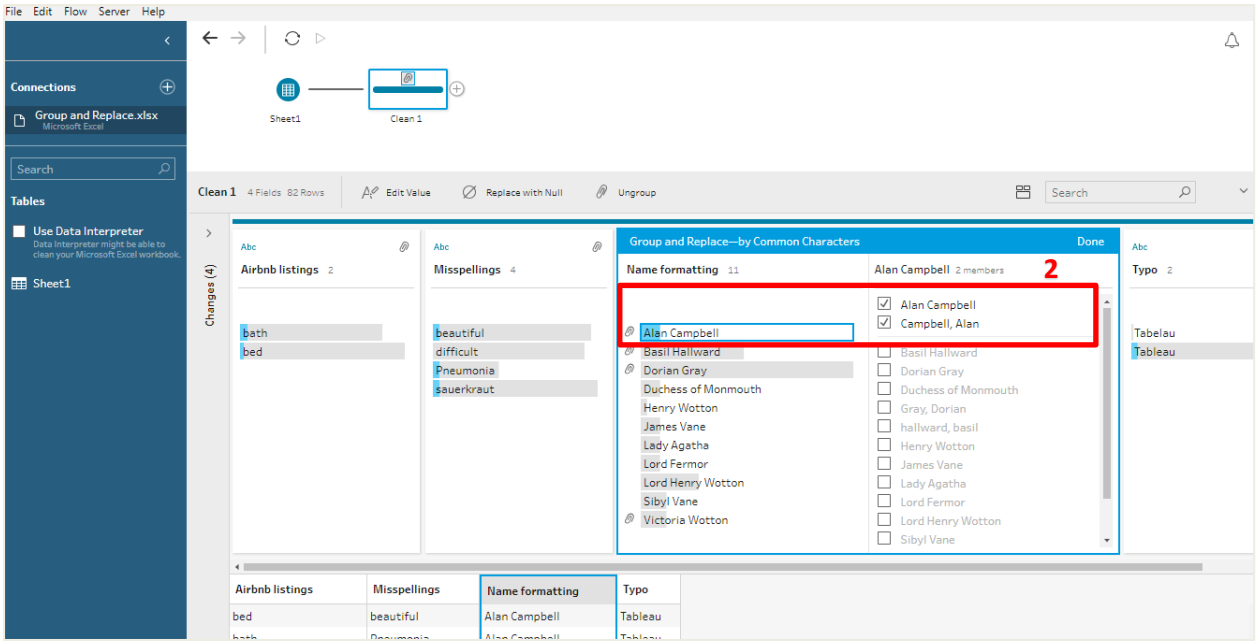


Step 9: 1 → Select the Common Characters option under the "Group and Replace" menu.

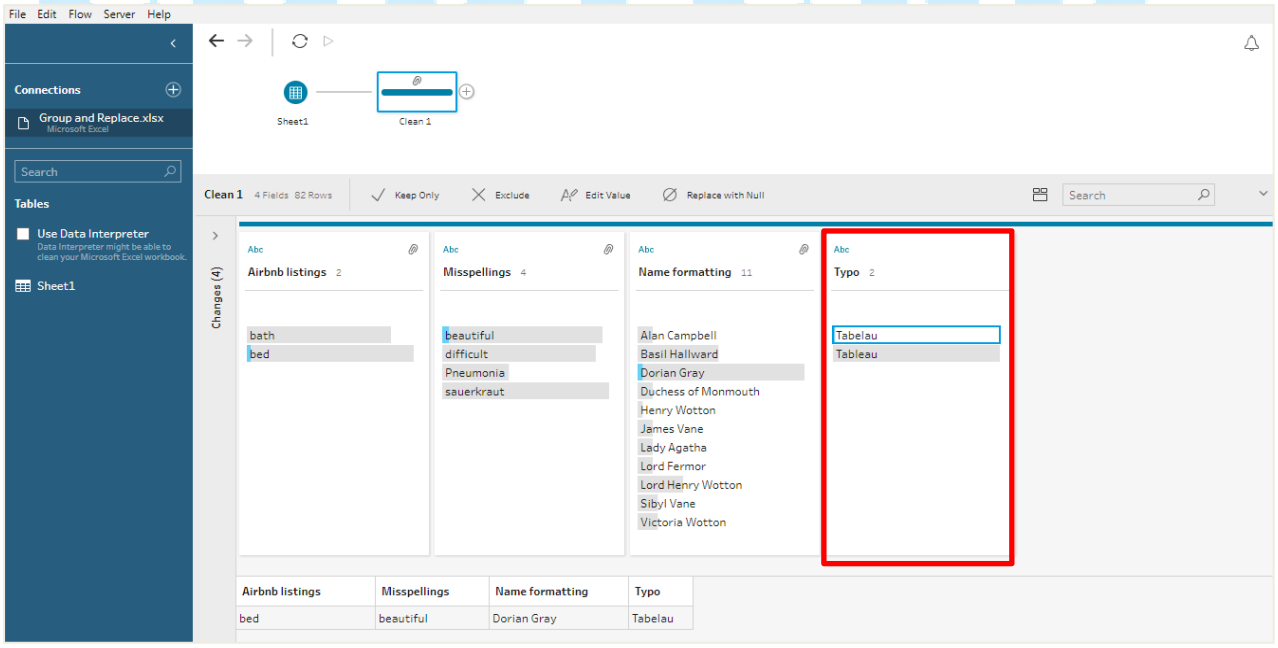


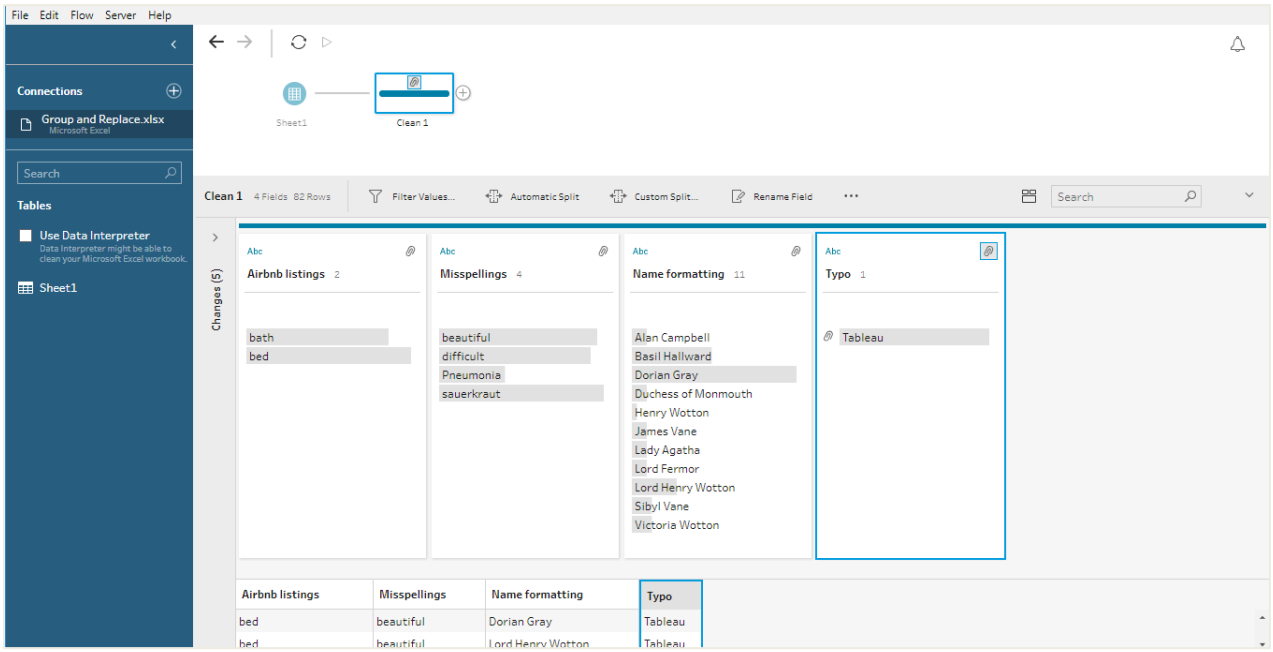
2 → In the Name formatting column, few names are listed as "last name, first name" rather than the more common "first name, last name".

- Since there exists two values for the same person, it should be grouped.
- Since the grouping is automatic, the value that occurs very often in the group becomes the replaced value.



Step 10: Address a typo or other small-scale issue by directly editing a value. Here, we fix the misspelled word “Tabelau” by right clicking and editing the value which in turn groups with the correct value.




File Edit Flow Server Help
Connections
Group and Replace.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Clean 1 4 Fields 82 Rows
Filter Values...
Automatic Split
Custom Split...
Rename Field
Search
Changes (5)
Abc
Airbnb listings 2
bath
bed
Abc
Misspellings 4
beautiful
difficult
Pneumonia
sauerkraut
Abc
Name formatting 11
Alan Campbell
Basil Hallward
Dorian Gray
Duchess of Monmouth
Henry Wotton
James Vane
Lady Agatha
Lord Fermor
Lord Henry Wotton
Sibyl Vane
Victoria Wotton
Abc
Typo 1
Tableau
Airbnb listings
Misspellings
Name formatting
Typo
bed
beautiful
Dorian Gray
Tableau
bed
beautiful
Lord Henry Wotton
Tableau

edureka!