Module 1: Data Preparation Using Tableau Prep

Demo Document III

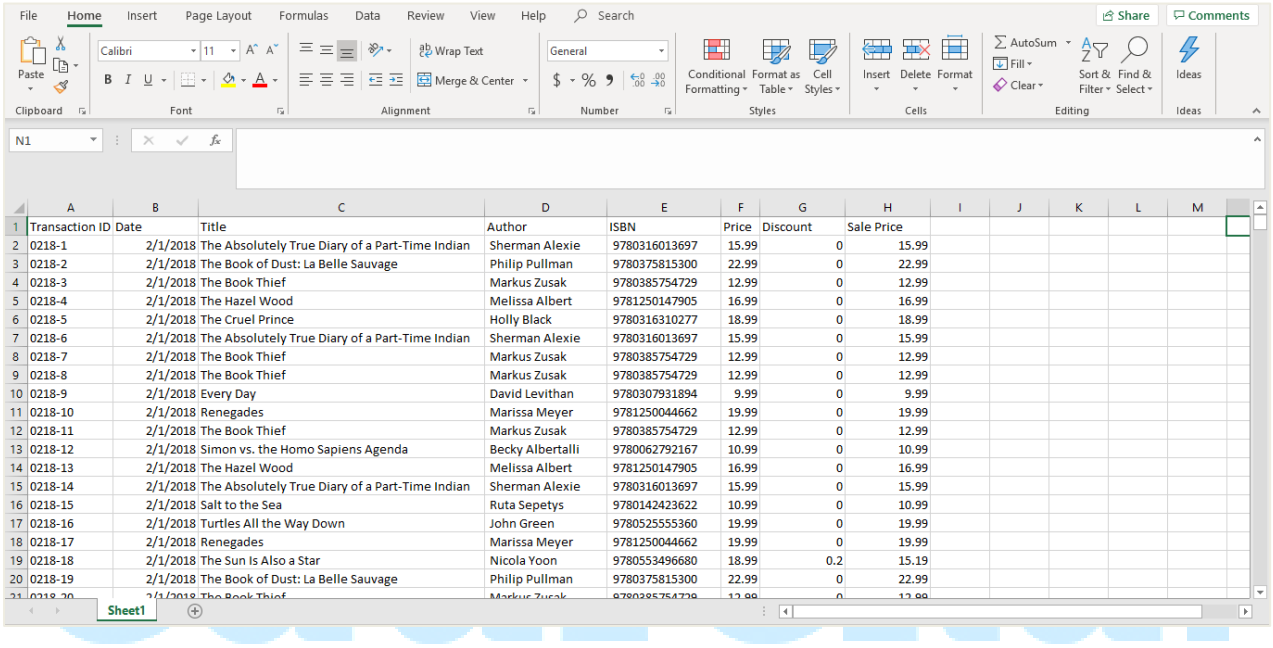
edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Demo III

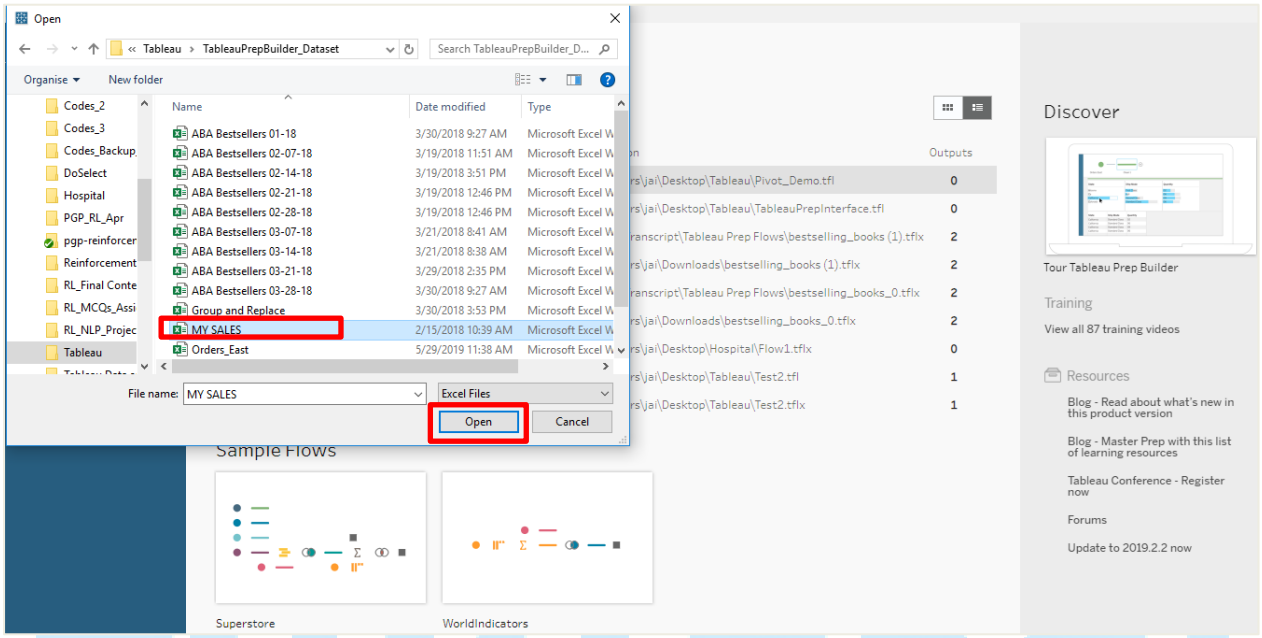
Aggregation is used to change the level of detail of data so that it is less granular, often in preparation for being combined with other data at a higher level of aggregation. Here, we are working with transactional data about sales from a book store. The sales data is a row per transaction. We will aggregate the sales data to be book level rather than transaction level.



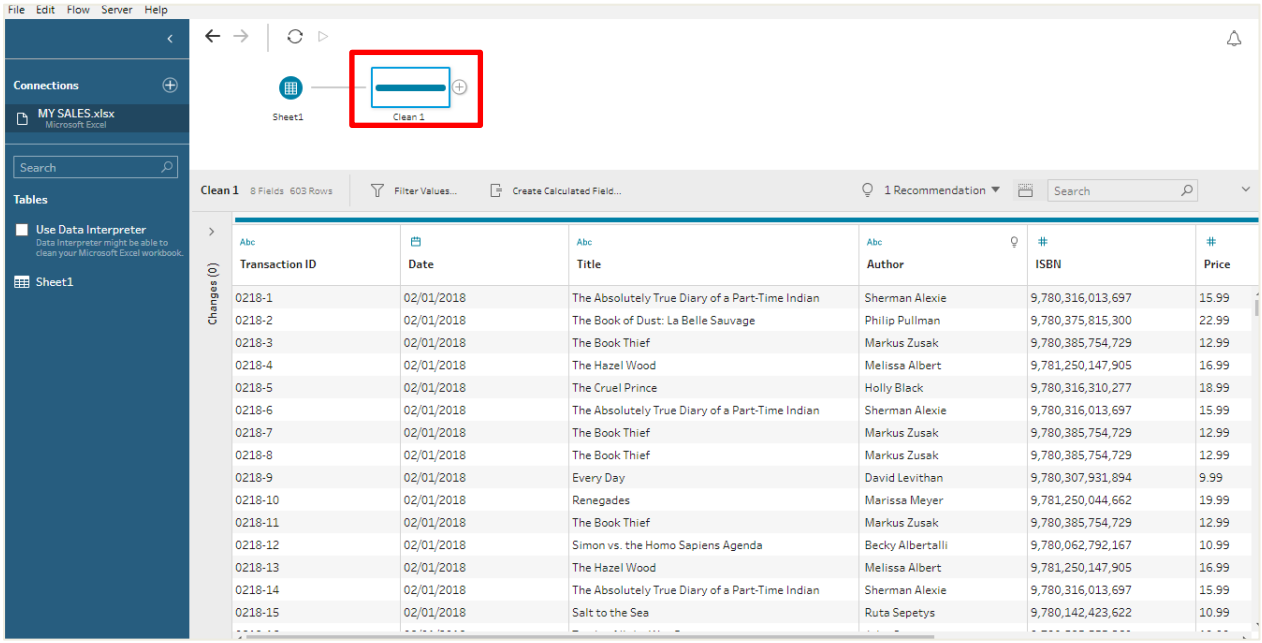
Transaction ID	Date	Title	Author	ISBN	Price	Discount	Sale Price
0218-1	2/1/2018	The Absolutely True Diary of a Part-Time Indian	Sherman Alexie	9780316013697	15.99	0	15.99
0218-2	2/1/2018	The Book of Dust: La Belle Sauvage	Philip Pullman	9780375815300	22.99	0	22.99
0218-3	2/1/2018	The Book Thief	Markus Zusak	9780385754729	12.99	0	12.99
0218-4	2/1/2018	The Hazel Wood	Melissa Albert	9781250147905	16.99	0	16.99
0218-5	2/1/2018	The Cruel Prince	Holly Black	9780316310277	18.99	0	18.99
0218-6	2/1/2018	The Absolutely True Diary of a Part-Time Indian	Sherman Alexie	9780316013697	15.99	0	15.99
0218-7	2/1/2018	The Book Thief	Markus Zusak	9780385754729	12.99	0	12.99
0218-8	2/1/2018	The Book Thief	Markus Zusak	9780385754729	12.99	0	12.99
0218-9	2/1/2018	Every Day	David Levithan	9780307931894	9.99	0	9.99
0218-10	2/1/2018	Renegades	Marissa Meyer	9781250044662	19.99	0	19.99
0218-11	2/1/2018	The Book Thief	Markus Zusak	9780385754729	12.99	0	12.99
0218-12	2/1/2018	Simon vs. the Homo Sapiens Agenda	Becky Albertalli	9780062792167	10.99	0	10.99
0218-13	2/1/2018	The Hazel Wood	Melissa Albert	9781250147905	16.99	0	16.99
0218-14	2/1/2018	The Absolutely True Diary of a Part-Time Indian	Sherman Alexie	9780316013697	15.99	0	15.99
0218-15	2/1/2018	Salt to the Sea	Ruta Sepetys	9780142423622	10.99	0	10.99
0218-16	2/1/2018	Turtles All the Way Down	John Green	9780525555360	19.99	0	19.99
0218-17	2/1/2018	Renegades	Marissa Meyer	9781250044662	19.99	0	19.99
0218-18	2/1/2018	The Sun Is Also a Star	Nicola Yoon	9780553496680	18.99	0.2	15.19
0218-19	2/1/2018	The Book of Dust: La Belle Sauvage	Philip Pullman	9780375815300	22.99	0	22.99

Demo III – Solution

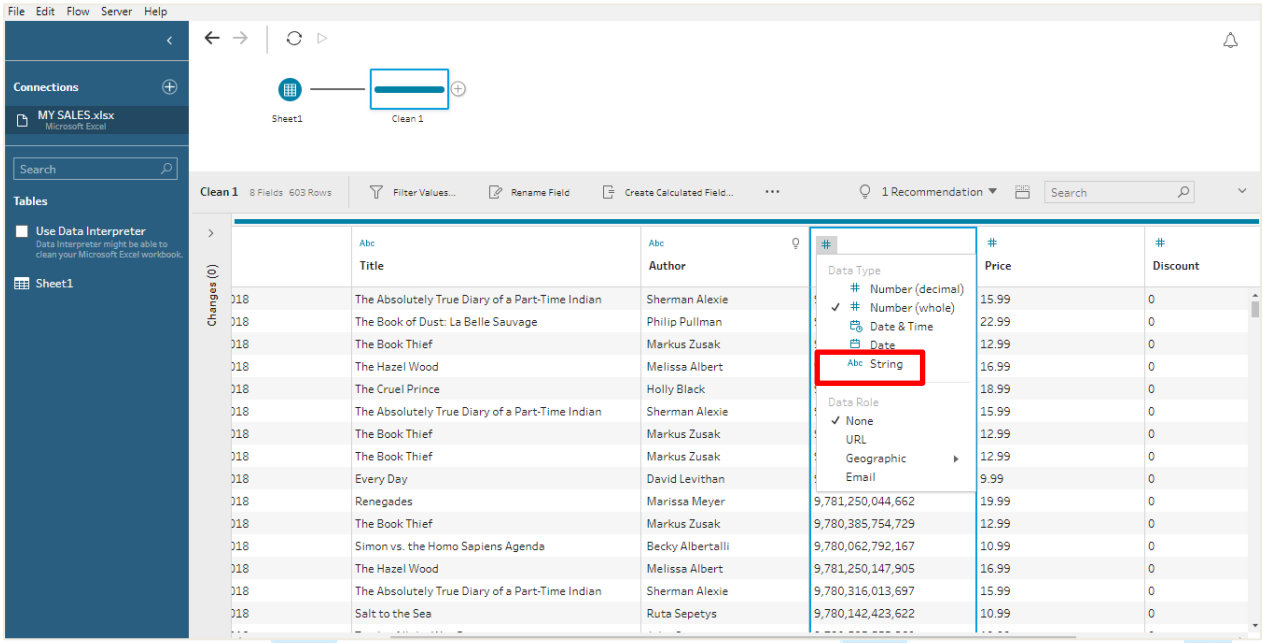
Step 1: Connect to the dataset and select “MY SALES” dataset and open the selected file.



Step 2: Add a Cleaning step.

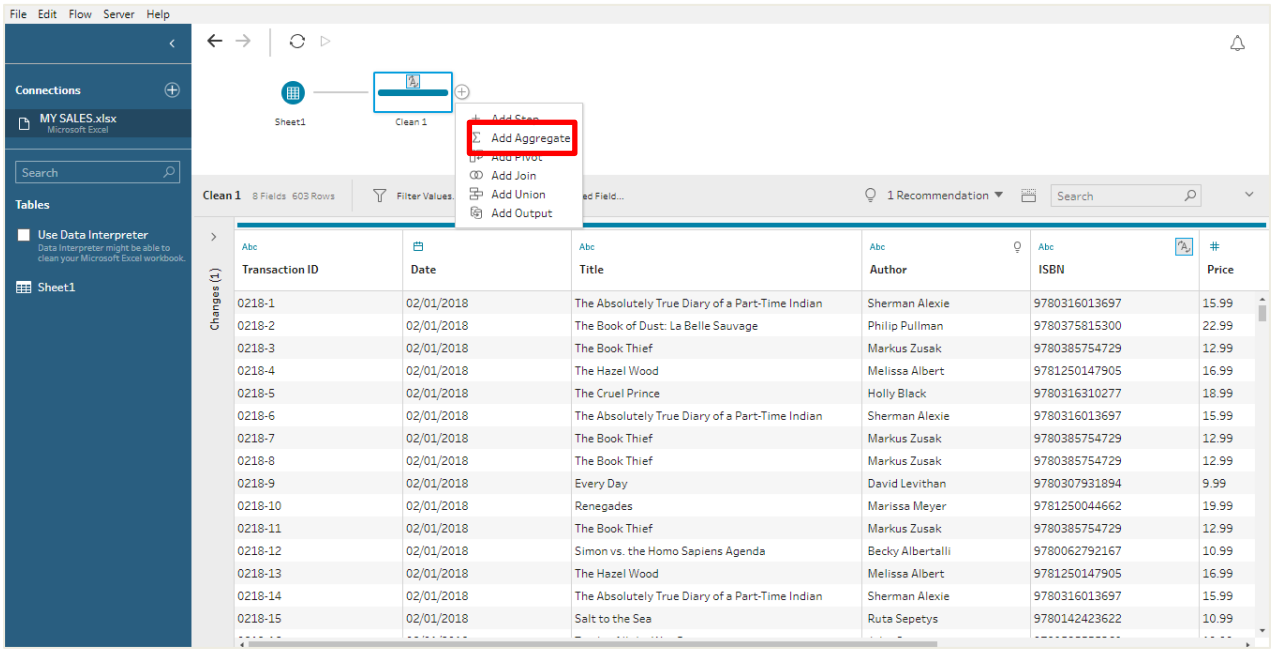


Step 3: Right click on “ISBN” card in the table and choose “String” data type for that field.



edureka!

Step 4: Click the “+” button and Select “Add Aggregate”.

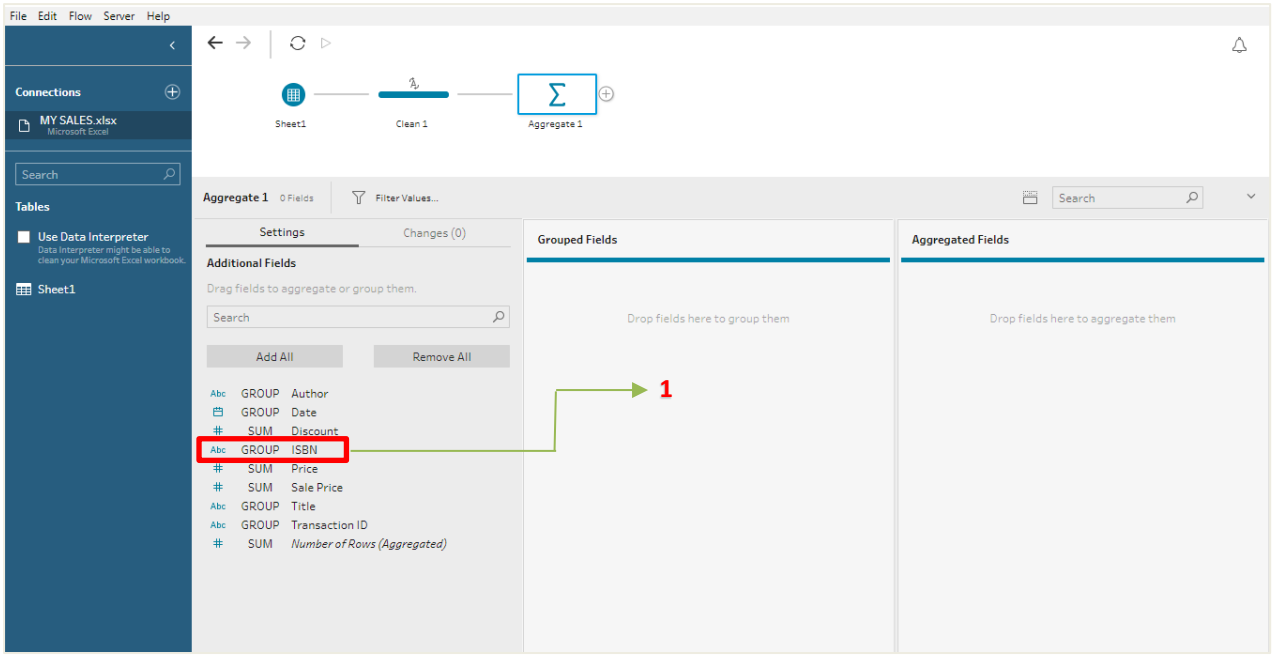


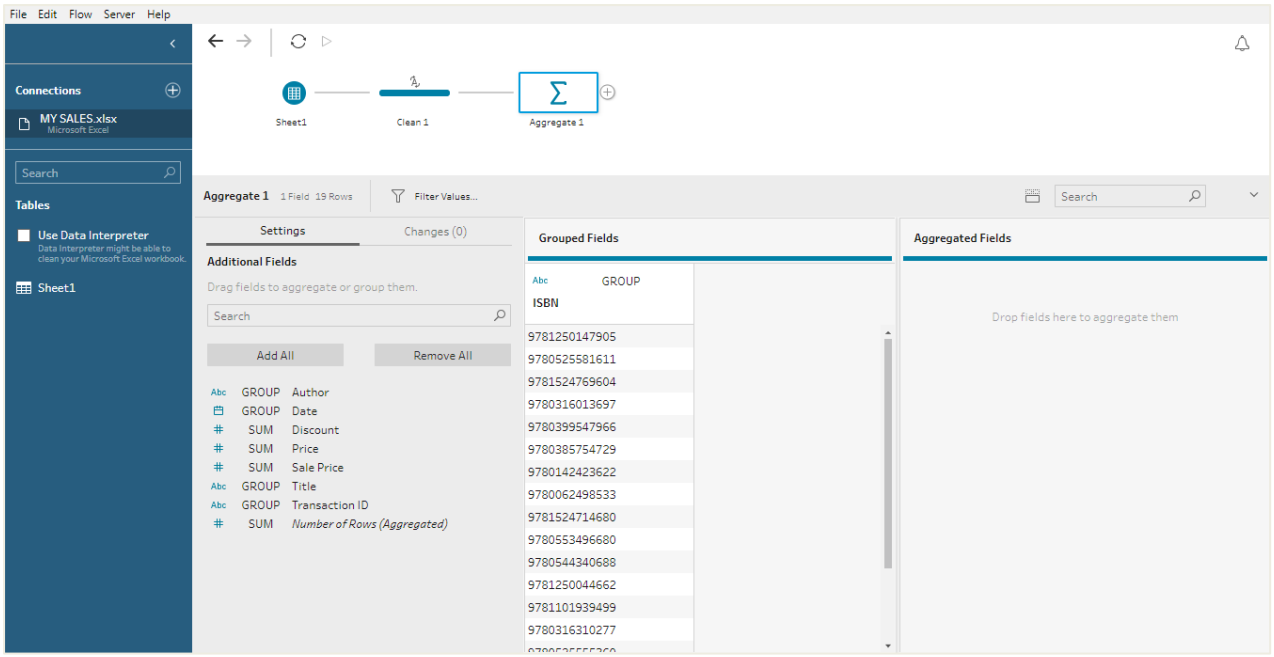
Step 5: 1 → Drag and Drop the "ISBN" field to the Grouped Fields area since it is more specific.

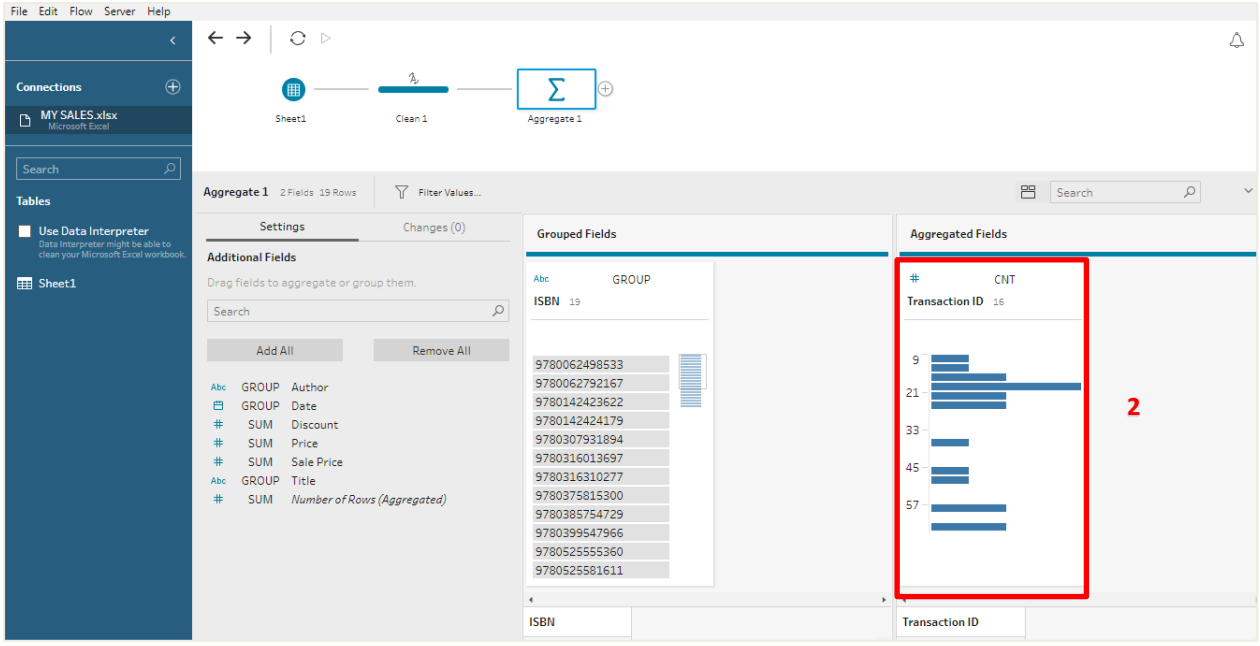
2 → Drag and Drop "Transaction ID" to the Aggregated Fields area. Transaction ID is of string data type, and the default aggregation is Count.

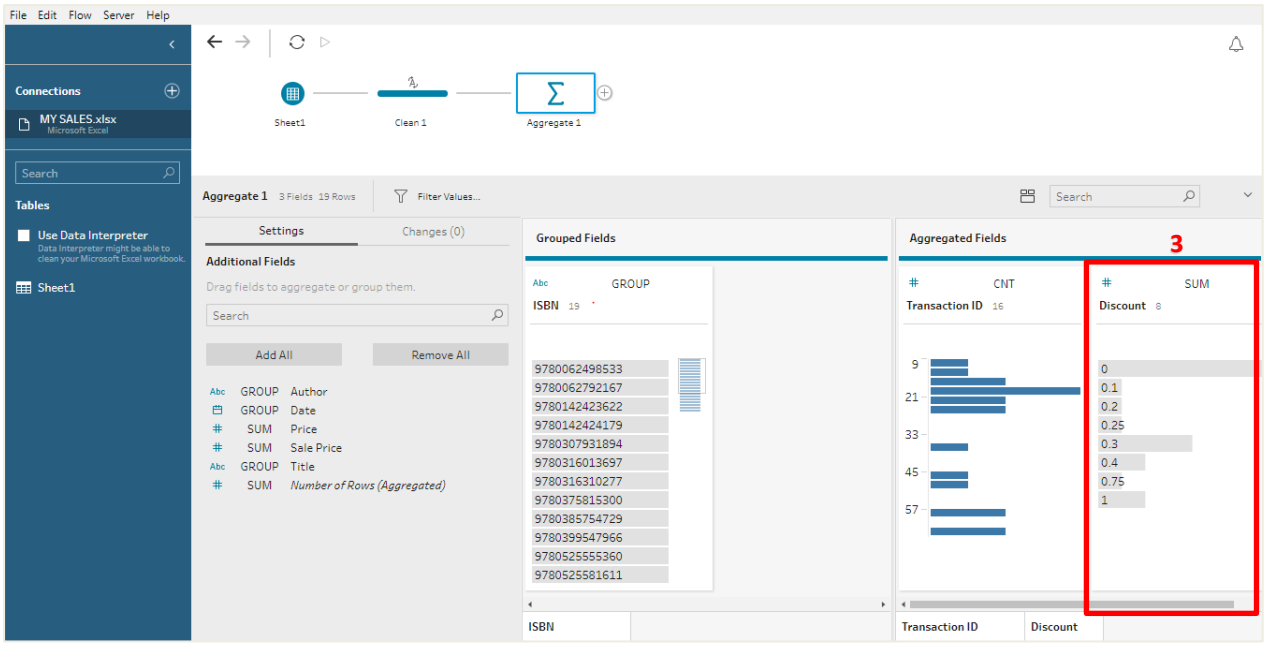
3 → Drag Discount to Aggregated Fields area.

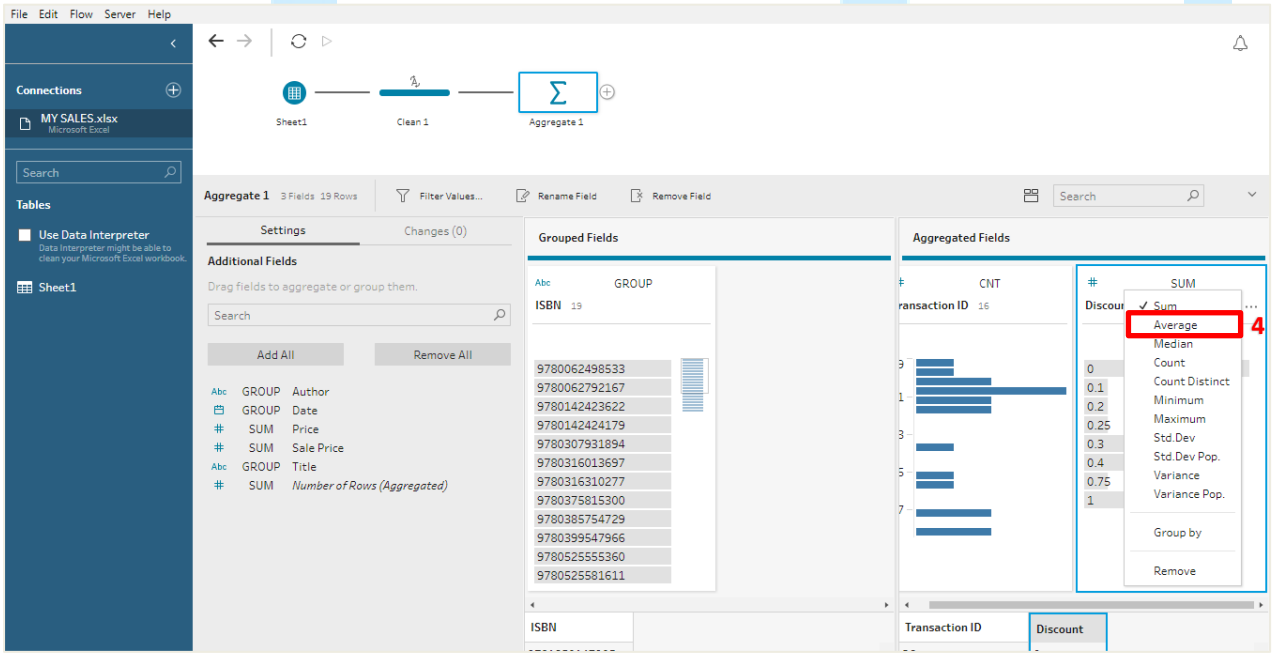
4 → Default Aggregation is SUM. Change it to AVERAGE by selecting it from the Dropdown menu.

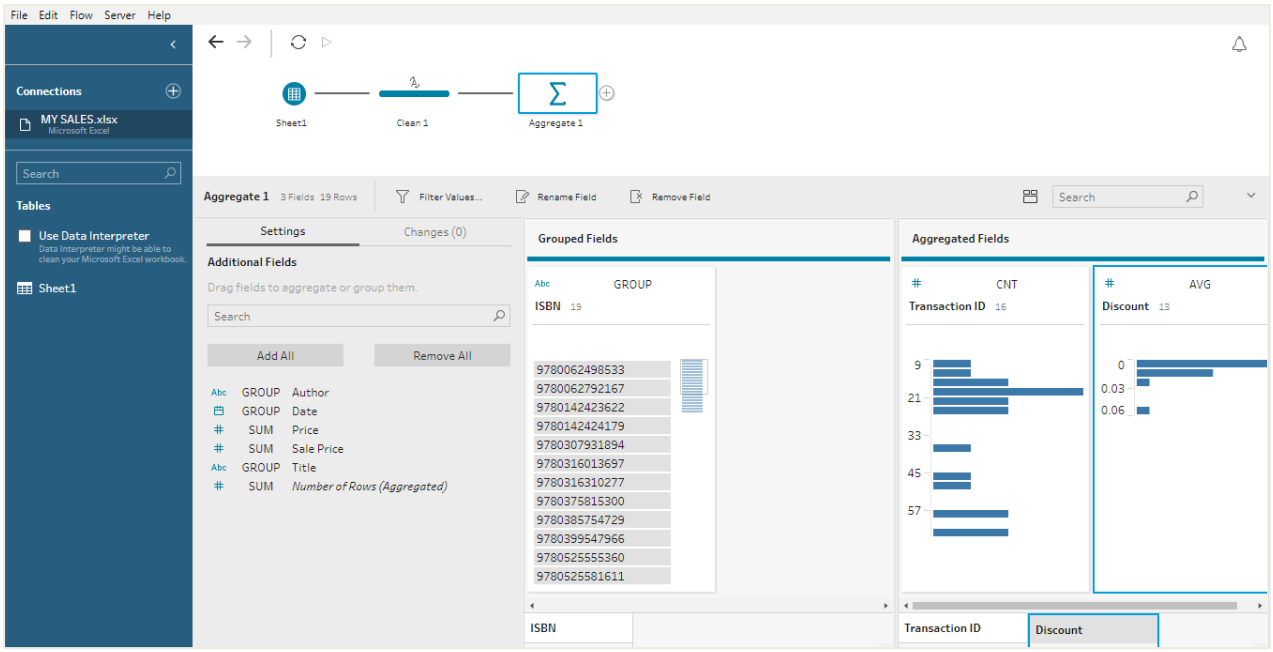



File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Aggregate 1 1 Field 19 Rows
Filter Values...
Search
Settings
Changes (0)
Additional Fields
Drag fields to aggregate or group them.
Search
Add All
Remove All
Abc GROUP Author
GROUP Date
SUM Discount
SUM Price
SUM Sale Price
Abc GROUP Title
Abc GROUP Transaction ID
SUM Number of Rows (Aggregated)
Grouped Fields
Abc GROUP
ISBN
9781250147905
9780525581611
9781524769604
9780316013697
9780399547966
9780385754729
9780142423622
9780062498533
9781524714680
9780553496680
9780544340688
9781250044662
9781101939499
9780316310277
Aggregated Fields
Drop fields here to aggregate them


File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Aggregate 1 2 Fields 19 Rows
Filter Values...
Search
Settings
Changes (0)
Additional Fields
Drag fields to aggregate or group them.
Search
Add All
Remove All
Abc GROUP Author
GROUP Date
SUM Discount
SUM Price
SUM Sale Price
Abc GROUP Title
SUM Number of Rows (Aggregated)
Grouped Fields
Abc GROUP
ISBN 19
9780062498533
9780062792167
9780142423622
9780142424179
9780307931894
9780316013697
9780316310277
9780375815300
9780385754729
9780399547966
9780525555360
9780525581611
Aggregated Fields
CNT
Transaction ID 16
9
21
33
45
57
2
ISBN
Transaction ID

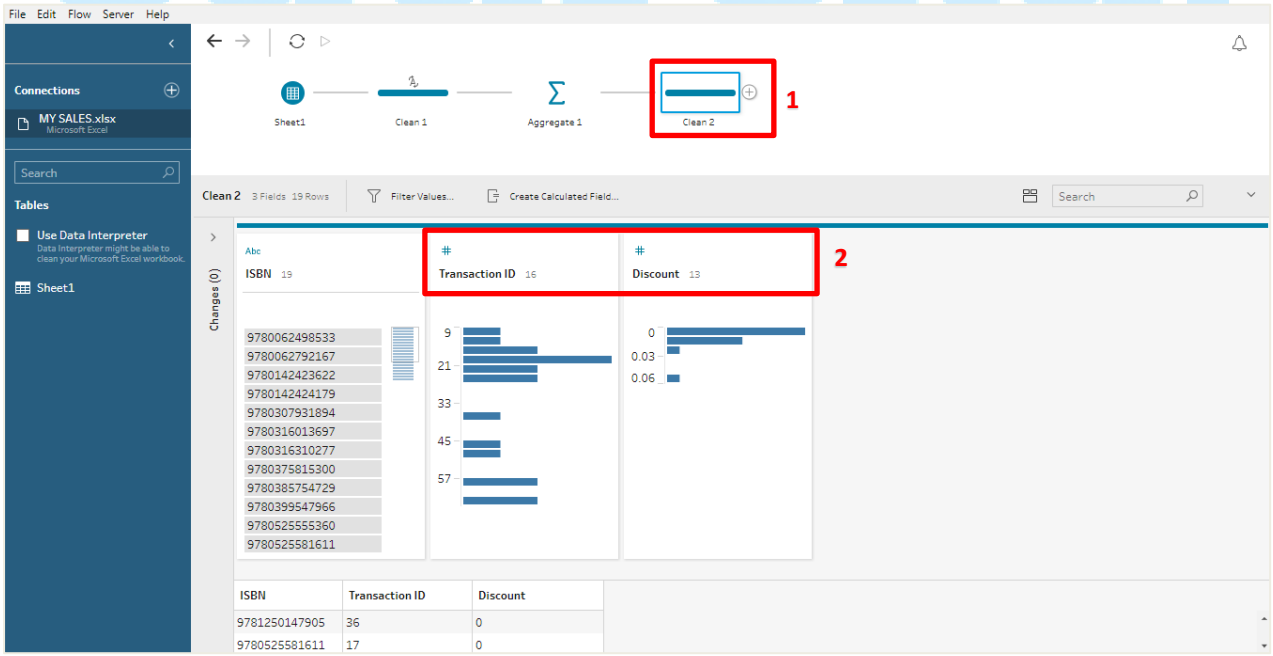

File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Aggregate 1 3 Fields 19 Rows
Filter Values...
Search
Settings
Changes (0)
Additional Fields
Drag fields to aggregate or group them.
Search
Add All
Remove All
Abc GROUP Author
GROUP Date
SUM Price
SUM Sale Price
Abc GROUP Title
SUM Number of Rows (Aggregated)
Grouped Fields
Abc GROUP
ISBN 19
9780062498533
9780062792167
9780142423622
9780142424179
9780307931894
9780316013697
9780316310277
9780375815300
9780385754729
9780399547966
9780525555360
9780525581611
Aggregated Fields
3
CNT
Transaction ID 16
SUM
Discount 8
9
21
33
45
57
0
0.1
0.2
0.25
0.3
0.4
0.75
1
ISBN
Transaction ID
Discount

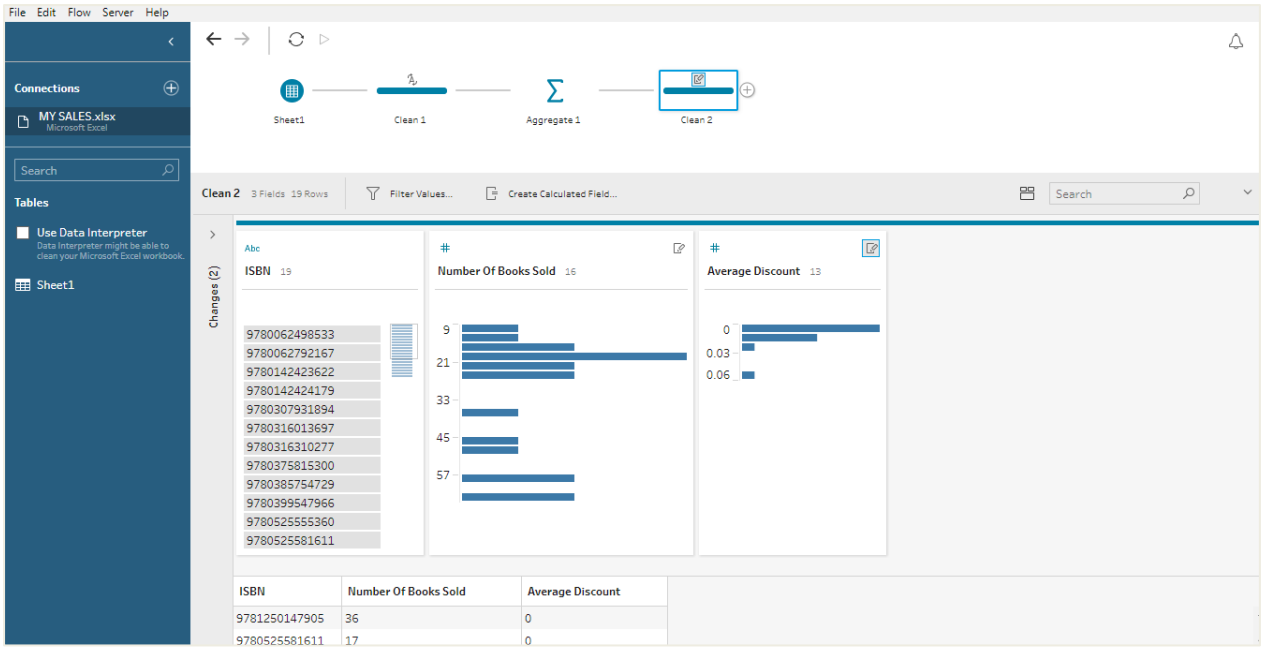

File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Aggregate 1 3 Fields 19 Rows
Filter Values...
Rename Field
Remove Field
Search
Settings
Changes (0)
Additional Fields
Drag fields to aggregate or group them.
Search
Add All
Remove All
Abc GROUP Author
GROUP Date
SUM Price
SUM Sale Price
Abc GROUP Title
SUM Number of Rows (Aggregated)
Grouped Fields
Abc GROUP
ISBN 19
9780062498533
9780062792167
9780142423622
9780142424179
9780307931894
9780316013697
9780316310277
9780375815300
9780385754729
9780399547966
9780525555360
9780525581611
Aggregated Fields
CNT
ransaction ID 16
SUM
Discou
Sum
Average
Median
Count
Count Distinct
Minimum
Maximum
Std.Dev
Std.Dev Pop.
Variance
Variance Pop.
Group by
Remove
4
0
0.1
0.2
0.25
0.3
0.4
0.75
1
ISBN
Transaction ID
Discount



Step 6: **1** → Add a Cleaning step

2 → Rename these fields





Step 7: **1** → Add a Cleaning step.

2 → Observe the Contextual bar. There are 3 fields and 19 rows.

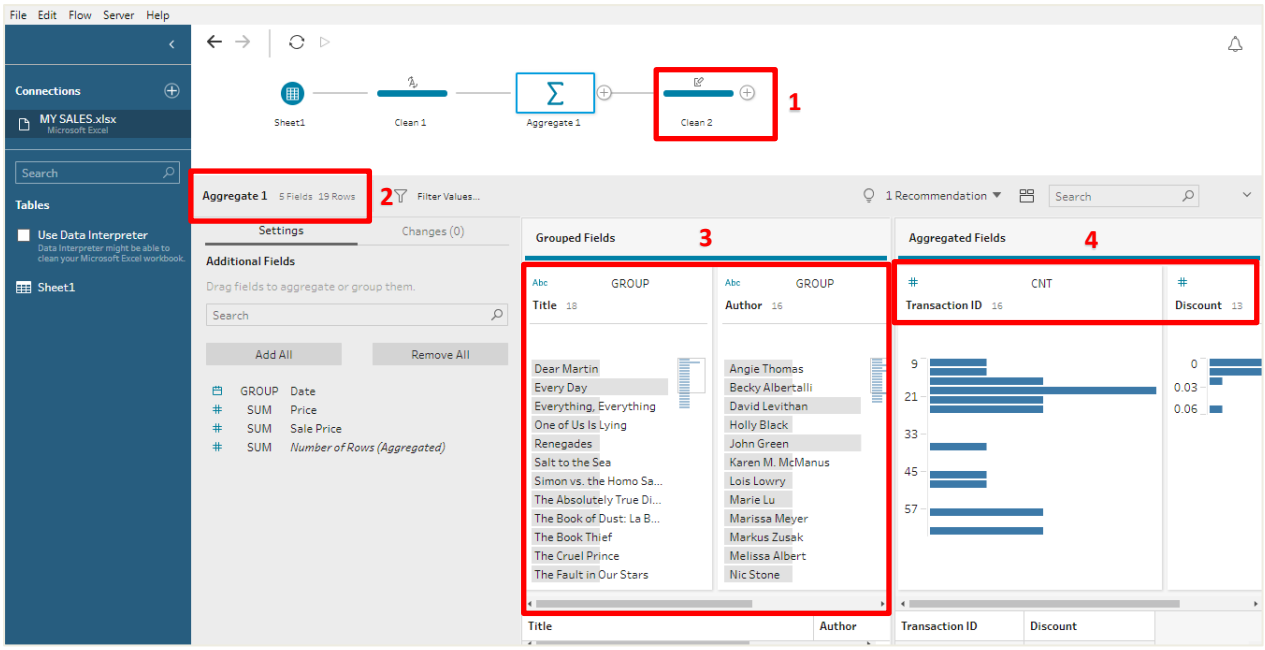
3 → Click on the Aggregate step, drag and drop Title and Author to the Grouped fields.

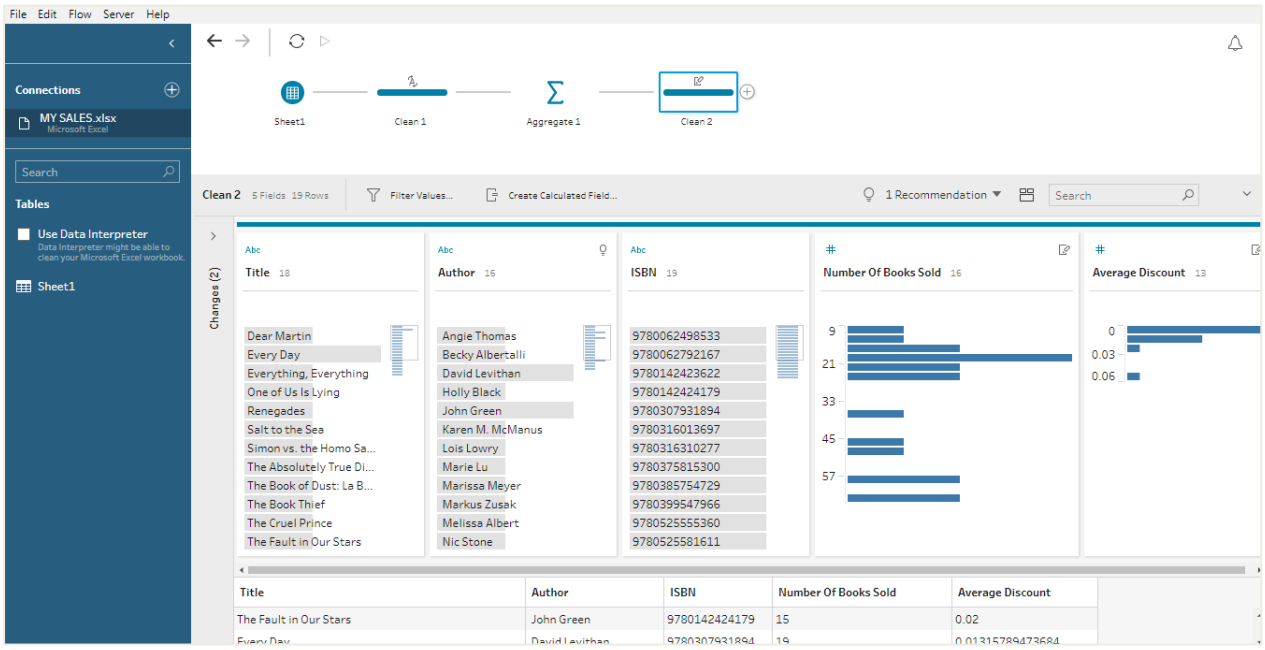
We observe that it still shows 19 rows. These two fields did not change the granularity

of the data since they are already accounted for in the ISBN.

4 → Rename these fields

edureka!

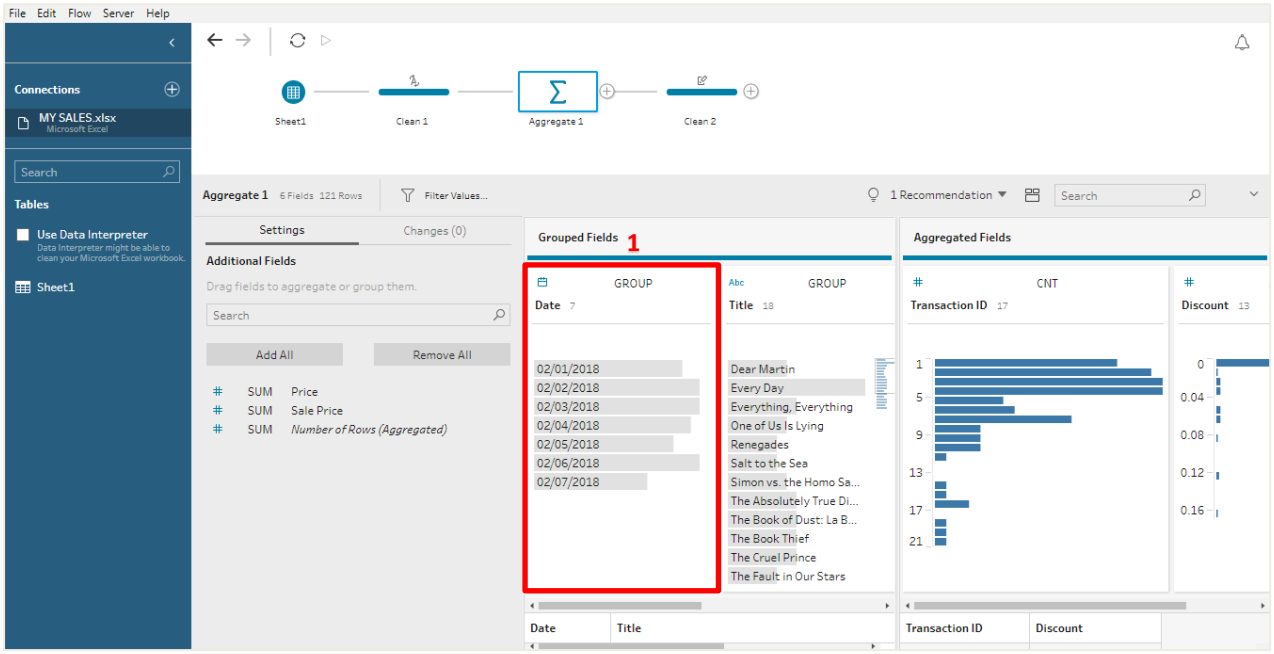


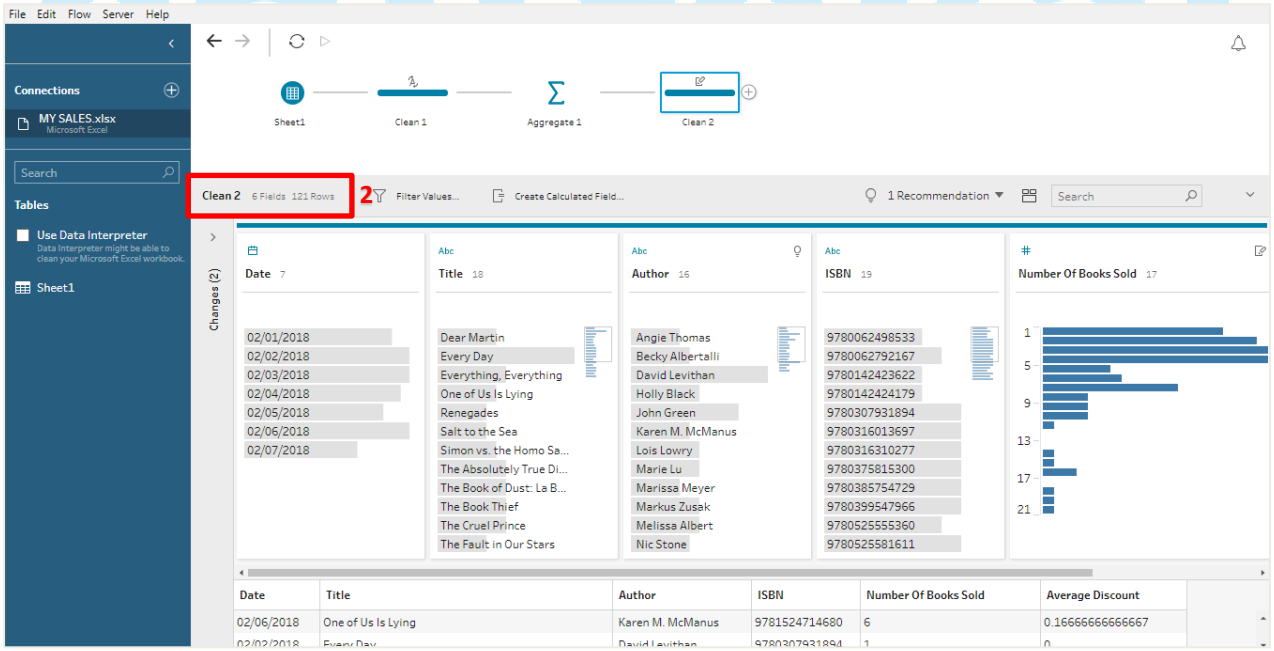


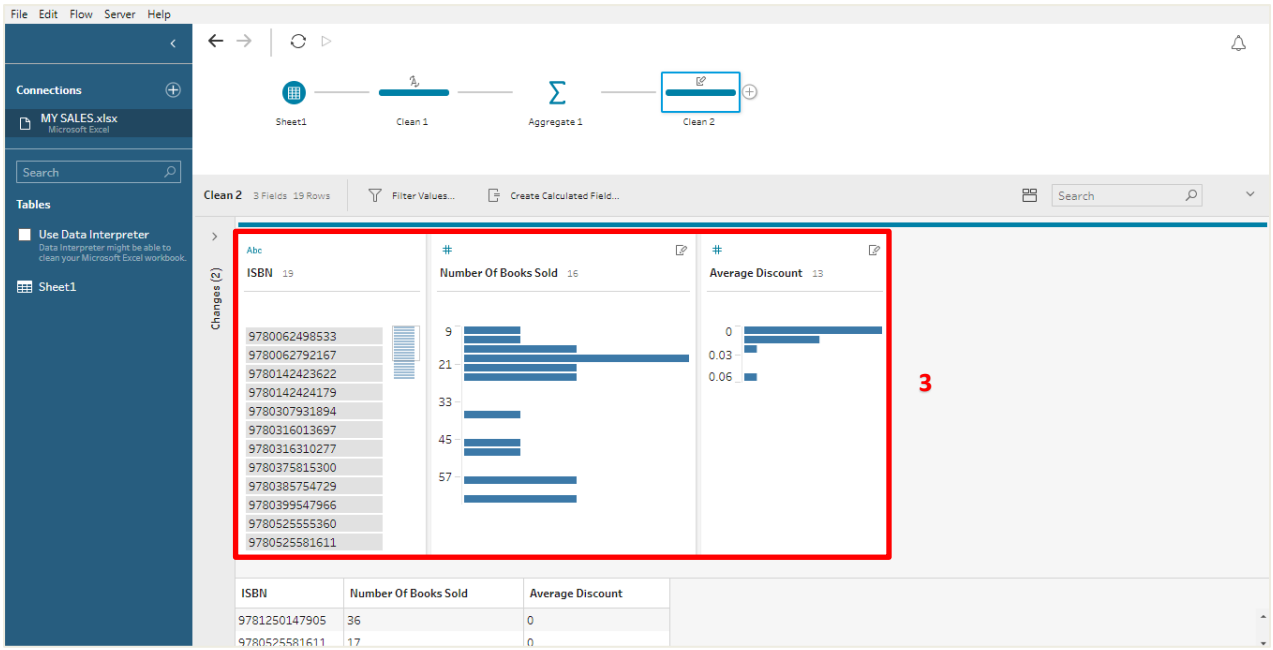
Step 8: **1** → Go to Aggregate step and drag and drop “Date” field.

2 → Observe the Contextual bar. We now have 121 rows since the granularity of what makes up a row now is Date and ISBN

3 → Go back to the original 3 fields and 19 rows


File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Clean 2
Aggregate 1 6 Fields 121 Rows
Filter Values...
1 Recommendation
Search
Settings
Changes (0)
Additional Fields
Drag fields to aggregate or group them.
Search
Add All
Remove All
SUM Price
SUM Sale Price
SUM Number of Rows (Aggregated)
Grouped Fields 1
GROUP
Date 7
02/01/2018
02/02/2018
02/03/2018
02/04/2018
02/05/2018
02/06/2018
02/07/2018
GROUP
Title 18
Dear Martin
Every Day
Everything, Everything
One of Us Is Lying
Renegades
Salt to the Sea
Simon vs. the Homo Sa...
The Absolutely True Di...
The Book of Dust: La B...
The Book Thief
The Cruel Prince
The Fault in Our Stars
Aggregated Fields
CNT
Transaction ID 17
Discount 13
Date
Title
Transaction ID
Discount


File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Clean 2
Clean 2 6 Fields 121 Rows
2
Filter Values...
Create Calculated Field...
1 Recommendation
Search
Changes (2)
Date 7
02/01/2018
02/02/2018
02/03/2018
02/04/2018
02/05/2018
02/06/2018
02/07/2018
Title 18
Dear Martin
Every Day
Everything, Everything
One of Us Is Lying
Renegades
Salt to the Sea
Simon vs. the Homo Sa...
The Absolutely True Di...
The Book of Dust: La B...
The Book Thief
The Cruel Prince
The Fault in Our Stars
Author 16
Angie Thomas
Becky Albertalli
David Levithan
Holly Black
John Green
Karen M. McManus
Lois Lowry
Marie Lu
Marissa Meyer
Markus Zusak
Melissa Albert
Nic Stone
ISBN 19
9780062498533
9780062792167
9780142423622
9780142424179
9780307931894
9780316013697
9780316310277
9780375815300
9780385754729
9780399547966
9780525555360
9780525581611
Number Of Books Sold 17
Date
Title
Author
ISBN
Number Of Books Sold
Average Discount
02/06/2018
One of Us Is Lying
Karen M. McManus
9781524714680
6
0.16666666666667


File Edit Flow Server Help
Connections
MY SALES.xlsx
Microsoft Excel
Search
Tables
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Sheet1
Sheet1
Clean 1
Aggregate 1
Clean 2
Clean 2 3 Fields 19 Rows
Filter Values...
Create Calculated Field...
Search
Changes (2)
ISBN 19
Number Of Books Sold 16
Average Discount 13
9780062498533
9780062792167
9780142423622
9780142424179
9780307931894
9780316013697
9780316310277
9780375815300
9780385754729
9780399547966
9780525555360
9780525581611
9
21
33
45
57
0
0.03
0.06
3
ISBN
Number Of Books Sold
Average Discount
9781250147905
36
0

edureka!