

Module 1: Data Preparation Using Tableau Prep

Case Study I – Solution Guide

edureka!

edureka!

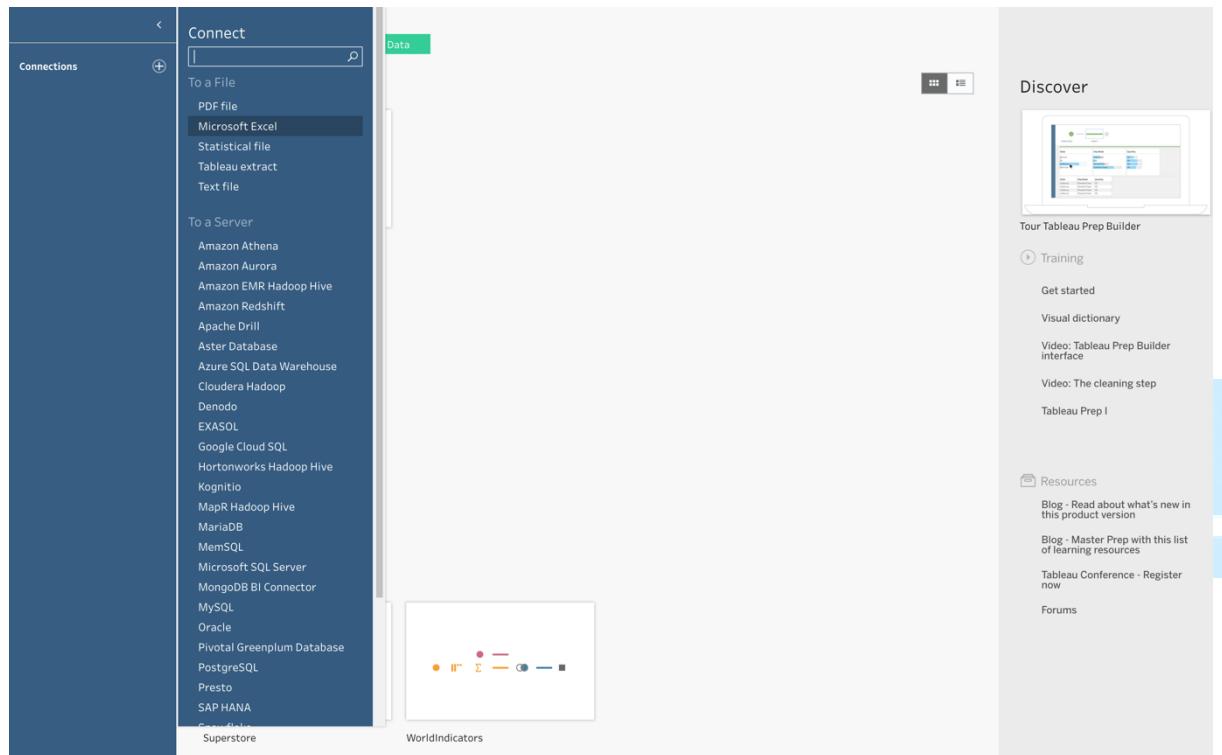
© Brain4ce Education Solutions Pvt. Ltd.

Solution Guide:

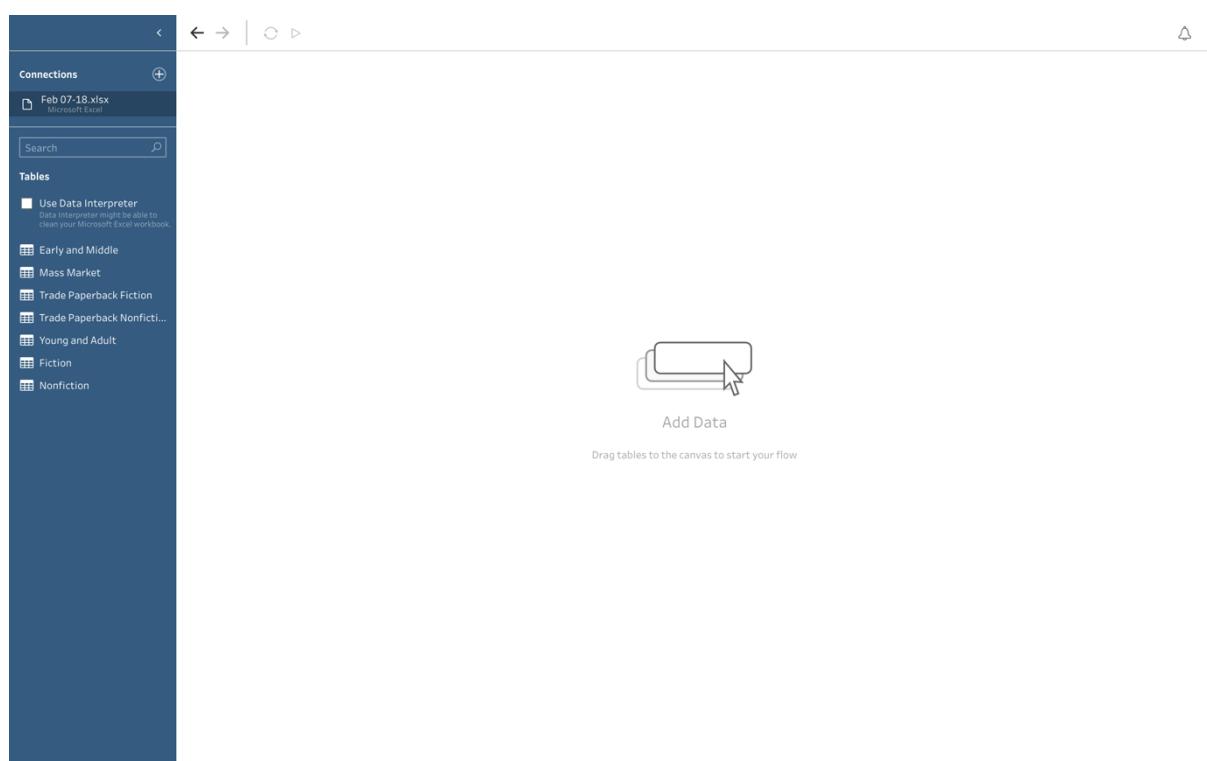
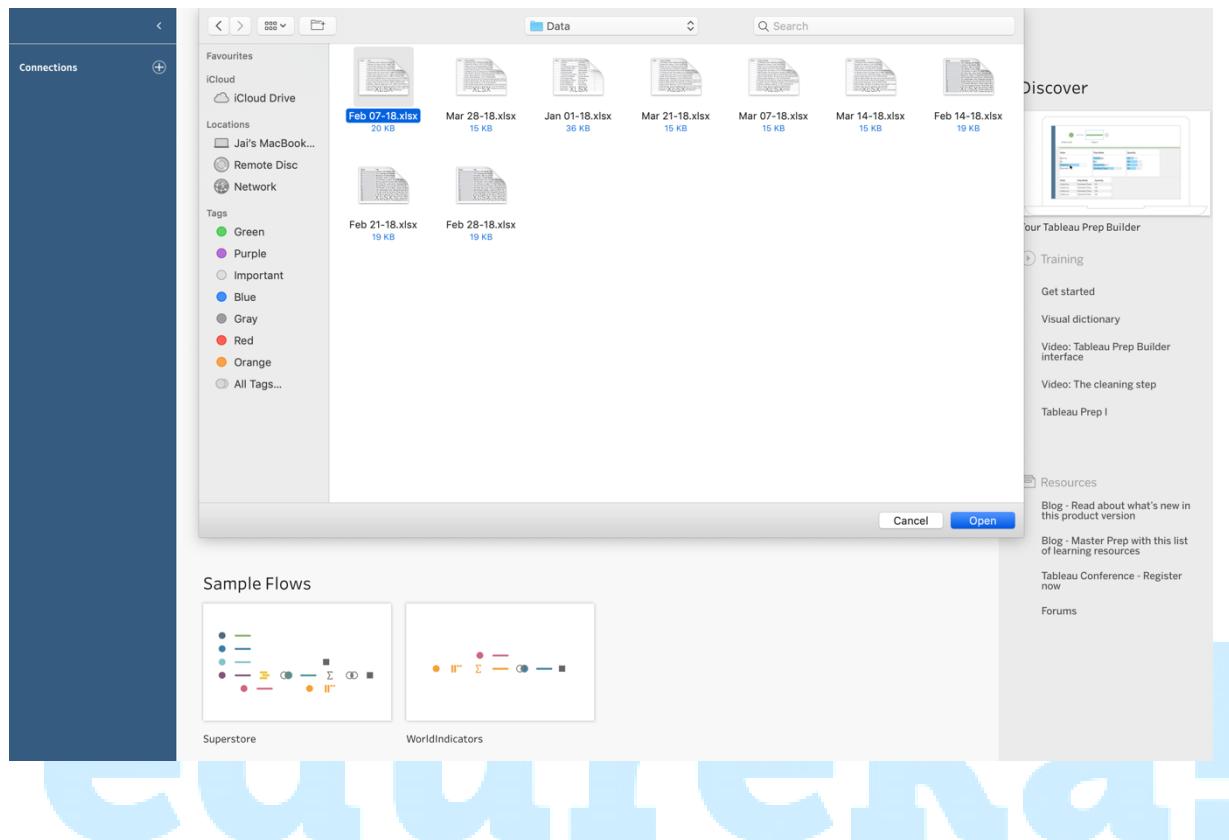
Q1. Connect to the data “Feb 07-18” and drag the Nonfiction table to the view and manipulate its columns.

Sol.:

Step 1: Open Tableau Prep and click on the “Connect to Data” option, select “Microsoft Excel” since our data is in .xlsx format.



Step 2: In the dialogue box given below, navigate or browse to the desired file that is “Feb 07-18”. Select the file and click “Open”.



Step 3: To manipulate Nonfiction table, drag the table to the view.

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' pane displays a single connection named 'Feb 07-18.xlsx' from Microsoft Excel. Below it, the 'Tables' pane lists several categories: Early and Middle, Mass Market, Trade Paperback Fiction, Trade Paperback Nonficti..., Young and Adult, Fiction, and Nonfiction. The 'Nonfiction' table is selected and highlighted with a blue border. The main workspace is titled 'Nonfiction' and shows a sample of four fields: 'Type', 'Field Name', 'Original Field Name', 'Changes', and 'Sample Values'. The sample values are: 'Abc', 'Last Week/ Week...', 'Last Week/ Weeks on List', '3/5, 2/40, 3/44'; '#', 'Rank', 'Rank', '1, 2, 3'; '#', 'Rank Duplicate', 'Rank Duplicate', '1, 2, 3'; and 'Abc', 'Info', 'Info', 'Fire and Fury | Michael Wolff \$30.00 | ISBN: 97'. A large blue decorative banner with the word 'CONTRACT' is visible at the bottom of the screen.

Type	Field Name	Original Field Name	Changes	Sample Values
Abc	Last Week/ Week...	Last Week/ Weeks on List		3/5, 2/40, 3/44
#	Rank	Rank		1, 2, 3
#	Rank Duplicate	Rank Duplicate		1, 2, 3
Abc	Info	Info		Fire and Fury Michael Wolff \$30.00 ISBN: 97

Step 4: Notice that the Rank field is duplicated. Remove it by unchecking the Rank field check box.

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' sidebar lists 'Feb 07-18.xlsx' (Microsoft Excel) as the active connection. Below it, the 'Tables' sidebar lists various categories: Early and Middle, Mass Market, Trade Paperback Fiction, Trade Paperback Nonficti..., Young and Adult, Fiction, and Nonfiction. The 'Nonfiction' table is selected, showing 4 Fields and 0 Rows. The main workspace displays the 'Nonfiction' table details. A modal window titled 'Nonfiction' shows 'Fields selected: 3 of 4'. It includes a 'Filter Values...' button and a table with columns: Type, Field Name, Original Field Name, Changes, and Sample Values. The table data is as follows:

Type	Field Name	Original Field Name	Changes	Sample Values
<input checked="" type="checkbox"/>	Abc	Last Week/ Week...	Last Week/ Weeks on List	1/5, 2/40, 3/44
<input checked="" type="checkbox"/>	#	Rank	Rank	1, 2, 3
<input type="checkbox"/>	#	Rank Duplicate	Rank Duplicate	1, 2, 3
<input checked="" type="checkbox"/>	Abc	Info	Info	Fire and Fury Michael Wolff \$30.00 ISBN: 97

Q2. Split the Info field into 4 different columns. Rename them and remove the unnecessary field(s) as well.

Sol.: Click on the plus (+) icon and select “Add Step”.

Module 1- Data Preparation Using Tableau Prep

The screenshot shows the Tableau Prep interface. On the left, there's a sidebar with 'Connections' (Feb 07-18.xlsx) and 'Tables' (Early and Middle, Mass Market, Trade Paperback Fiction, Trade Paperback Nonfiction, Young and Adult, Fiction, Nonfiction). A 'Use Data Interpreter' checkbox is checked. In the center, a 'Nonfiction' step is selected, showing '4 Fields 0 Rows'. A context menu is open over the step, with 'Add Step' highlighted. Below the step, there are tabs for 'Multiple Files', 'Data Sample', and 'Changes (1)'. Under 'Changes (1)', it says 'Select the fields to include in your flow. If you make changes to the data, the data source will be queried again.' A table lists fields: Type, Field Name, Original Field Name, Changes, and Sample Values. Fields include 'Abc' (Last Week/Week...), '#', 'Rank', 'Rank Duplicate', and 'Info'.

The screenshot shows the Tableau Prep interface. On the left, there's a sidebar with 'Connections' (Feb 07-18.xlsx) and 'Tables' (Early and Middle, Mass Market, Trade Paperback Fiction, Trade Paperback Nonfiction, Young and Adult, Fiction, Nonfiction). A 'Use Data Interpreter' checkbox is checked. In the center, a 'Clean 1' step is selected, showing '3 Fields 15 Rows'. A context menu is open over the step, with 'Add Step' highlighted. Below the step, there's a 'Changes (0)' section. It contains three tables: 'Last Week/ Week...' (15 rows), 'Rank' (15 rows), and 'Info' (15 rows). Each table shows sample data rows.

Step 1: Right-click on the Info field and select Split Values, from the Drop-down menu, and further choose the Custom Split option.

The screenshot shows the Tableau Prep interface with a connection to 'Feb 07-18.xlsx'. A flow is visible from 'Nonfiction' to 'Clean 1'. In the 'Tables' section, 'Clean 1' is selected, showing 3 fields and 15 rows. The 'Info' column is highlighted. A context menu is open over the 'Info' column, with 'Split Values' selected. The 'Custom Split...' option is highlighted in the submenu. Below the table, a preview shows the original 'Info' column and its split results.

Last Week/ Weeks on List	Rank	Info
/5	1	Fire and Fury Michael Wolff \$30.00 ISBN: 97812501
2/40	2	Astrophysics for People in a Hurry Neil deGrasse Tyson
3/44	3	The Subtle Art of Not Giving A F*ck Mark Manson \$24.00
4/16	4	Leonardo da Vinci Walter Isaacson \$35.00 ISBN: 97814197
15/31	5	Killers of the Flower Moon David Grann \$28.95 ISBN: 97814767
10/4	6	When Daniel H. Pink \$28.00 ISBN: 9780735210622
/1	7	The Monk of Mokha Dave Eggers \$28.95 ISBN: 97814767
5/6	8	The Last Black Unicorn Tiffany Haddish \$26.00 ISBN: 97814767
9/11	9	Principles: Life and Work Ray Dalio \$30.00 ISBN: 97814767

In the dialogue box, make the following changes: -

- Use “|” as Separator.
- Change the split off option to All.

Module 1- Data Preparation Using Tableau Prep

The screenshot shows the Tableau Prep interface. A flow is set up from a 'Nonfiction' connection to a 'Clean 1' stage. In the 'Clean 1' stage, a 'Changes (0)' section is open, showing a 'Custom Split' dialog. The dialog asks 'How should this data be split?' and 'Use the separator |'. Below the dialog, a preview shows three rows: 'The Gentle Art of Swed...', 'The Last Black Unicorn ...', and 'The Monk of Mokha | D...'. The main table below shows a list of books with their ranks and details.

Last Week/ Weeks on List	Rank	Info
1/5	1	Fire and Fury Michael Wolff \$30.00 ISBN: 9781250158062
2/40	2	Astrophysics for People in a Hurry Neil deGrasse Tyson \$18.95
3/44	3	The Subtle Art of Not Giving A F*ck Mark Manson \$24.99
4/16	4	Leonardo da Vinci Walter Isaacson \$35.00 ISBN: 9781501139154
15/31	5	Killers of the Flower Moon David Grann \$28.95 ISBN: 9780393609394
10/4	6	When Daniel H. Pink \$28.00 ISBN: 9780735210622
/1	7	The Monk of Moksha Dave Eggers \$28.95 ISBN: 9781250158062
5/6	8	The Last Black Unicorn Tiffany Haddish \$26.00 ISBN: 9781501181825
9/11	9	Principles: Life and Work Ray Dalio \$30.00 ISBN: 9781501139154

Step 2: Even after the split, notice that the column is divided into 3 fields. This is because we need to use “\$” sign as the separator. Right-click on the Info - Split 2 field and go to Split Values, and within that select the Automatic Split option.

The screenshot shows the Tableau Prep interface with a more complex flow. A 'Nonfiction' connection feeds into a 'Clean 1' stage, which then splits into 'Info - Split 1', 'Info - Split 2', and 'Last Week/Weeks on List' stages. The 'Info - Split 2' stage has a context menu open over its 'Info' field, with 'Automatic Split' selected. The preview shows the data now correctly split into single columns. The main table below shows the final cleaned data.

Last Week/ Weeks on List	Rank	Info
1/5	1	Fire and Fury Michael Wolff \$30.00 ISBN: 9781250158062
2/40	2	Astrophysics for People in a Hurry Neil deGrasse Tyson \$18.95
3/44	3	The Subtle Art of Not Giving A F*ck Mark Manson \$24.99
4/16	4	Leonardo da Vinci Walter Isaacson \$35.00 ISBN: 9781501139154
15/31	5	Killers of the Flower Moon David Grann \$28.95 ISBN: 9780393609394
10/4	6	When Daniel H. Pink \$28.00 ISBN: 9780735210622
/1	7	The Monk of Moksha Dave Eggers \$28.95 ISBN: 9781250158062
5/6	8	The Last Black Unicorn Tiffany Haddish \$26.00 ISBN: 9781501181825
9/11	9	Principles: Life and Work Ray Dalio \$30.00 ISBN: 9781501139154

Module 1- Data Preparation Using Tableau Prep

The screenshot shows the Tableau Prep interface. At the top, there's a connection to 'Feb 07-18.xlsx' (Microsoft Excel) and a 'Nonfiction' source. Below that is a search bar and a 'Tables' section with various categories like 'Early and Middle', 'Mass Market', etc. The main area is titled 'Clean 1' and contains six columns labeled 'Info - Split 2 - Split 1', 'Info - Split 2 - Split 2', 'Info - Split 1', 'Info - Split 2', 'Info - Split 3', and 'Last Week/Week...'. Each column has 15 rows. The data in the columns is as follows:

Info - Split 2 - Split 1	Info - Split 2 - Split 2	Info - Split 1	Info - Split 2	Info - Split 3	Last Week/Week...
Daniel H. Pink	15.95	Astrophysics for People...	Daniel H. Pink \$28.00	ISBN: 9780062300546	/1
Dave Eggers	18.95	Fire and Fury	Dave Eggers \$28.95	ISBN: 9780062457714	1/5
David Grann	18.99	Grant	David Grann \$28.95	ISBN: 9780385534246	10/4
J.D. Vance	24.99	Hillbilly Elegy	J.D. Vance \$27.99	ISBN: 9780393609394	11/3
Margareta Magnusson	26.00	How Democracies Die	Margareta Magnusson...	ISBN: 9780399185045	12/8
Mark Manson	27.99	Killers of the Flower M...	Mark Manson \$24.99	ISBN: 9780735210622	13/17
Mary Beard	28.00	Leonardo da Vinci	Mary Beard \$15.95	ISBN: 9781101947319	15/31
Michael Wolff	28.95	Principles: Life and Work	Michael Wolff \$30.00	ISBN: 9781250158062	2/40
Neil deGrasse Tyson	30.00	The Book of Joy	Neil deGrasse Tyson \$...	ISBN: 9781501124020	3/44
Ray Dalio	35.00	The Gentle Art of Swed...	Ray Dalio \$30.00	ISBN: 9781501139154	4/16
Ron Chernow		The Last Black Unicorn	Ron Chernow \$40.00	ISBN: 9781501173240	5/6
Steven Levitsky	40.00	The Monk of Mokha	Steven Levitsky \$26.00	ISBN: 9781501181825	6/75
Info - Split 2 - Split 1					
Michael Wolff	30.00	Fire and Fury	Michael Wolff \$30.00	ISBN: 9781250158062	1/5
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	Neil deGrasse Tyson \$18.95	ISBN: 9780393609394	2/40
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	Mark Manson \$24.99	ISBN: 9780062457714	3/44
Walter Isaacson	35.00	Leonardo da Vinci	Walter Isaacson \$35.00	ISBN: 9781501139154	4/16
David Grann	28.95	Killers of the Flower Moon	David Grann \$28.95	ISBN: 9780385534246	15/3
Daniel H. Pink	28.00	When	Daniel H. Pink \$28.00	ISBN: 9780735210622	10/4
Dave Eggers	28.95	The Monk of Mokha	Dave Eggers \$28.95	ISBN: 9781101947319	1/
Tiffany Haddish	26.00	The Last Black Unicorn	Tiffany Haddish \$26.00	ISBN: 9781501181825	5/6
Ray Dalio	30.00	Divisive: Life and Death	Ray Dalio \$30.00	ISBN: 9781501181825	6/11

Step 3: Rename each column by double-clicking on its name. Rename the field as follows: -

- Info - Split 2 - Split 1 as Author
- Info - Split 2 - Split 2 as Price
- Info - Split 1 as Name of the Book
- Info - Split 3 as ISBN

Module 1- Data Preparation Using Tableau Prep

Clean 1 8 Fields 15 Rows

Author	Price	Name of the book	Info - Split 2	ISBN	Last Week/ Weeks on List
Michael Wolff	30.00	Fire and Fury	Michael Wolff \$30.00	ISBN: 9781250158062	1/5
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	Neil deGrasse Tyson \$18.95	ISBN: 9780393609394	2/40
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	Mark Manson \$24.99	ISBN: 9780062457714	3/44
Walter Isaacson	35.00	Leonardo da Vinci	Walter Isaacson \$35.00	ISBN: 9781501139154	4/16
David Grann	28.95	Killers of the Flower Moon	David Grann \$28.95	ISBN: 9780385534246	15/31
Daniel H. Pink	28.00	When	Daniel H. Pink \$28.00	ISBN: 9780735210622	10/4
Dave Eggers	28.95	The Monk of Mokha	Dave Eggers \$28.95	ISBN: 9781101947319	/1
Tiffany Haddish	26.00	The Last Black Unicorn	Tiffany Haddish \$26.00	ISBN: 9781501181825	5/6
Ron Chernow	40.00	The Monk of Mokha	Ron Chernow \$40.00	ISBN: 9781501173240	6/75
Steven Levitsky			Steven Levitsky \$26.00	ISBN: 9781501181825	

Step 4: Manually remove Info - Split 2 and Info columns from the table. Right-click the column card and select Remove.

Clean 1 8 Fields 15 Rows

Author	Price	Name of the book	Info - Split 2	ISBN	Last Week/ Weeks on List
Michael Wolff	30.00	Fire and Fury	Michael Wolff \$30.00	ISBN: 9781250158062	1/5
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	Neil deGrasse Tyson \$18.95	ISBN: 9780393609394	2/40
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	Mark Manson \$24.99	ISBN: 9780062457714	3/44
Walter Isaacson	35.00	Leonardo da Vinci	Walter Isaacson \$35.00	ISBN: 9781501139154	4/16
David Grann	28.95	Killers of the Flower Moon	David Grann \$28.95	ISBN: 9780385534246	15/31
Daniel H. Pink	28.00	When	Daniel H. Pink \$28.00	ISBN: 9780735210622	10/4
Dave Eggers	28.95	The Monk of Mokha	Dave Eggers \$28.95	ISBN: 9781101947319	/1
Tiffany Haddish	26.00	The Last Black Unicorn	Tiffany Haddish \$26.00	ISBN: 9781501181825	5/6
Ron Chernow	40.00	The Monk of Mokha	Ron Chernow \$40.00	ISBN: 9781501173240	6/75
Steven Levitsky			Steven Levitsky \$26.00	ISBN: 9781501181825	

- Perform the same for the Info field.

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30.00	Fire and Fury	ISBN: 9781250158062	/1	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	ISBN: 9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	ISBN: 9780062457714	3/44	3
Walter Isaacson	35.00	Leonardo da Vinci	ISBN: 9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	ISBN: 9780385534246	15/31	5
Daniel H. Pink	28.00	When	ISBN: 9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	ISBN: 9781101947319	/1	7
Tiffany Haddish	26.00	The Last Black Unicorn	ISBN: 9781501181825	5/6	8
Ray Dalio	30.00	Principles: Life and Work	ISBN: 9781501124020	9/11	9

Q3. Remove all the punctuations and letters from the ISBN field.

Sol.:

Step 1: Right-click on ISBN and go to Clean and select the “Remove Punctuation” option.

The screenshot shows the Tableau Prep interface with a 'Nonfiction' dataset connected to a 'Clean 1' step. The 'ISBN' column is highlighted, and a context menu is open over it. The 'Clean' option is selected, and a submenu is displayed with 'Remove Punctuation' as one of the choices. The main workspace shows a grid of data with columns for Author, Price, Name of the book, ISBN, Last Week/ Weeks on List, and Rank.

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30.00	Fire and Fury	ISBN 9781250158062	1/5	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	ISBN 9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	ISBN 9780062457714	3/44	3
Walter Isaacson	35.00	Leonardo da Vinci	ISBN 9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	ISBN 9780385534246	15/31	5
Daniel H. Pink	28.00	When	ISBN 9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	ISBN 9781101947319	/1	7
Tiffany Haddish	26.00	The Last Black Unicorn	ISBN 9781501181825	5/6	8
Ray Dalio	30.00	Principles: Life and Work	ISBN 9781501124020	9/11	9

Step 2: Right-click on ISBN, choose the Clean option and select the “Remove Letters” option from the drop down.

Module 1- Data Preparation Using Tableau Prep

Connections: Feb 07-18.xlsx Microsoft Excel

Tables: Early and Middle, Mass Market, Trade Paperback Fiction, Trade Paperback Nonfiction, Young and Adult, Fiction, Nonfiction

Use Data Interpreter: Data Interpreter might be able to clean your Microsoft Excel workbook.

Clean 1: 6 Fields, 15 Rows

Changes (12):

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30.00	Fire and Fury	ISBN 9781250158062	1/5	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	ISBN 9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	ISBN 9780062457714	3/44	3
Walter Isaacson	35.00	Leonardo da Vinci	ISBN 9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	ISBN 9780385534246	15/31	5
Daniel H. Pink	28.00	When	ISBN 9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	ISBN 9781101947319	/1	7
Tiffany Haddish	26.00	The Last Black Unicorn	ISBN 9781501181825	5/6	8
Ray Dalio	30.00	Principles: Life and Work	ISBN 9781501124020	9/11	9



9780062300546	
9780062457714	
9780385534246	
9780393609394	
9780399185045	
9780735210622	
9781101947319	
9781250158062	
9781501124020	
9781501139154	
9781501173240	
9781501181825	

Notice that all the punctuations and letters have now been removed and the data type is still String.

Q4. Change the data type of Price from String to Float.

Sol.: Click the “Abc” icon located at the top left corner in the Price card.

The screenshot shows the Tableau Prep interface in 'Clean' mode. On the left, there's a sidebar with various connection and transformation options. The main workspace shows a flow from a 'Nonfiction' source to a 'Clean 1' target. In the center, the 'Changes' pane displays a list of 13 changes made to the data. One change is highlighted with a dropdown menu open over the 'Price' column, which is currently set to 'String'. The dropdown menu includes options like 'Number' (decimal), 'Number' (integer), and 'Text'. Other columns shown in the changes list include 'Author', 'Name of the book', 'ISBN', 'Last Week/ Weeks on List', and 'Rank'. Below the changes list is a preview table with the same columns and data as the changes list. The preview table shows the following data:

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30.00	Fire and Fury	9781250158062	1/5	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	9780062457714	3/44	3
Walter Isaacson	35.00	Leonardo da Vinci	9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	9780385534246	15/31	5
Daniel H. Pink	28.00	When	9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	9781101947319	1/1	7
Tiffany Haddish	26.00	The Last Black Unicorn	9781501181825	5/6	8
Ray Dalio	30.00	Principles: Life and Work	9781501124020	9/11	9

- In the drop-down menu, select “Number” (decimal) and change the datatype to float.

Module 1- Data Preparation Using Tableau Prep

The screenshot shows the Tableau Prep interface. On the left, there's a sidebar with 'Tables' and a 'Nonfiction' connection. The main area shows a flow from 'Nonfiction' to 'Clean 1'. In 'Clean 1', the 'Author' field is selected, and its data type is being changed from 'String' to 'Number (decimal)'. The 'Price' column is highlighted in blue. The data grid below shows various books with their authors, prices, ISBNs, and ranks.

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30.00	Fire and Fury	9781250158062	1/5	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	9780062457714	3/44	3
Walter Isaacson	35.00	Leonardo da Vinci	9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	9780385534246	15/31	5
Daniel H. Pink	28.00	When	9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	9781101947319	/1	7
Tiffany Haddish	26.00	The Last Black Unicorn	9781501173240	5/6	8
Ray Dalio	30.00	Principles: Life and Work	9781501124020	9/11	9

- Once selected, notice that the data type has changed.

This screenshot is identical to the one above, but it shows the state after the 'Author' field has been converted to 'Number (decimal)'. The 'Price' column remains highlighted in blue. The data grid shows the same 15 books with their updated price values.

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30	Fire and Fury	9781250158062	1/5	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	9780062457714	3/44	3
Walter Isaacson	35	Leonardo da Vinci	9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	9780385534246	15/31	5
Daniel H. Pink	28	When	9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	9781101947319	/1	7
Tiffany Haddish	26	The Last Black Unicorn	9781501173240	5/6	8
Ray Dalio	30	Principles: Life and Work	9781501124020	9/11	9

Q5. Split the “Last Week/Weeks on List” into two separate columns. Once done, remove the unnecessary field.

Sol.: Right-click on “Last Week/Weeks on List” and go to “Split Values” and select “Automatic Split”.

The screenshot shows the Tableau Prep interface with the 'Clean 1' stage selected. The stage contains four tables: 'Changes (14)', 'Price (11)', 'Name of the book (15)', and 'ISBN (15)'. The 'Changes (14)' table has a context menu open over the 'Last Week/ Weeks on List' column, which includes options like 'Filter', 'Group and Replace', 'Clean', 'Split Values', and 'Automatic Split...'. The 'Automatic Split...' option is highlighted. The other three tables show sample data for price, book names, and ISBNs respectively.

Author	Price	Name of the book	ISBN	Last Week/ Weeks on List	Rank
Michael Wolff	30	Fire and Fury	9781250158062	1/5	1
Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	9780393609394	2/40	2
Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	9780062457714	3/44	3
Walter Isaacson	35	Leonardo da Vinci	9781501139154	4/16	4
David Grann	28.95	Killers of the Flower Moon	9780385534246	15/31	5
Daniel H. Pink	28	When	9780735210622	10/4	6
Dave Eggers	28.95	The Monk of Mokha	9781101947319	1/1	7
Tiffany Haddish	26	The Last Black Unicorn	97815011139154	5/6	8
Ray Dalio	30	Principles: Life and Work	9781501124020	9/11	9

Module 1- Data Preparation Using Tableau Prep

The screenshot shows the Tableau Prep interface with the 'Clean 1' step selected. The left sidebar lists tables such as Early and Middle, Mass Market, Trade Paperback Fiction, Trade Paperback Nonfiction, Young and Adult, Fiction, and Nonfiction. The main workspace has four panes: 'Changes (16)' (showing a bar chart of counts for 'Last Week/Week...'), 'Last Week/Week...', 'Author', and 'Price'. Below these are two tables: 'Last Week/ Weeks on List - Split 1' and 'Last Week/ Weeks on List - Split 2'. The 'Last Week/Week...' pane contains a bar chart of counts for 'Last Week/Week...'. The 'Author' and 'Price' panes show lists of authors and their prices. The 'ISBN' table at the bottom right lists book titles, authors, and ISBN numbers.

Step 1: Rename each field as follows: -

- Last Week/Weeks on List - Split 1 as Last Week.
- Last Week/Weeks on List - Split 2 as Weeks on List.

This screenshot shows the Tableau Prep interface with the 'Clean 1' step selected, similar to the previous one but with renamed fields. The 'Changes (18)' pane now shows a bar chart for 'Last Week'. The 'Last Week/Week...' pane shows a bar chart for 'Weeks on List'. The 'Author' and 'Price' panes remain the same. The 'ISBN' table at the bottom right lists book titles, authors, and ISBN numbers.

Step 2: Right-click on Last Week/Weeks on List and select the “Remove” option.

The screenshot shows the Tableau Prep interface with the 'Clean 1' step selected. On the left, there's a sidebar with various tables listed under 'Tables'. In the main area, there are four columns: 'Name of the book', 'ISBN', and 'Price' are grouped under 'Abc', while 'Last Week / ...' is grouped under '#'. A context menu is open over the '#/1' entry in the 'Last Week / ...' column, with the 'Remove' option highlighted. The menu also includes other options like 'Filter', 'Group and Replace', 'Clean', 'Split Values', 'View State', 'Detail', 'Summary', 'Rename Field', 'Duplicate Field', 'Keep Only Field', and 'Create Calculated Field...'. Below the columns, a table shows the data with columns: 'Last Week', 'Weeks on List', 'Author', 'Price', 'Name of the book', and 'ISBN'.

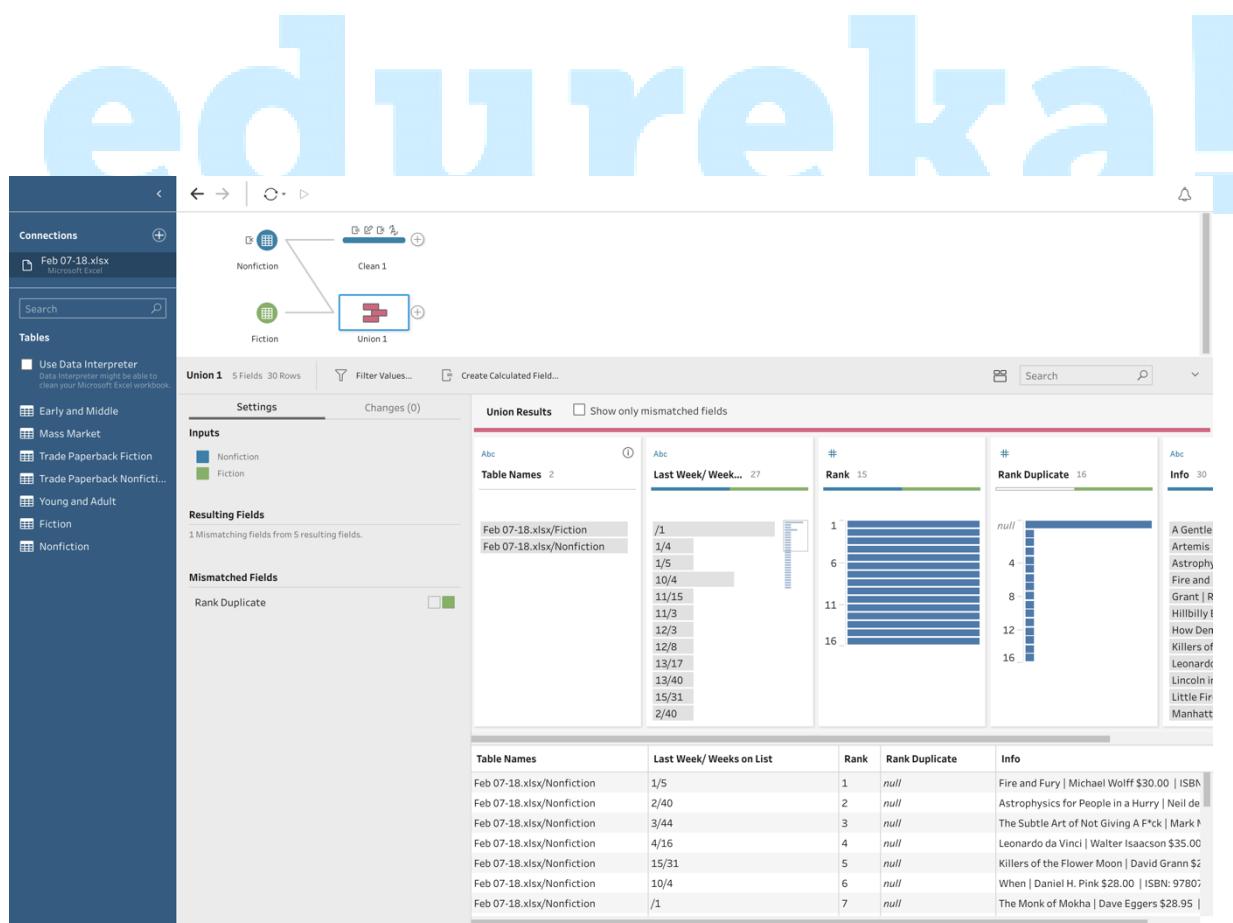
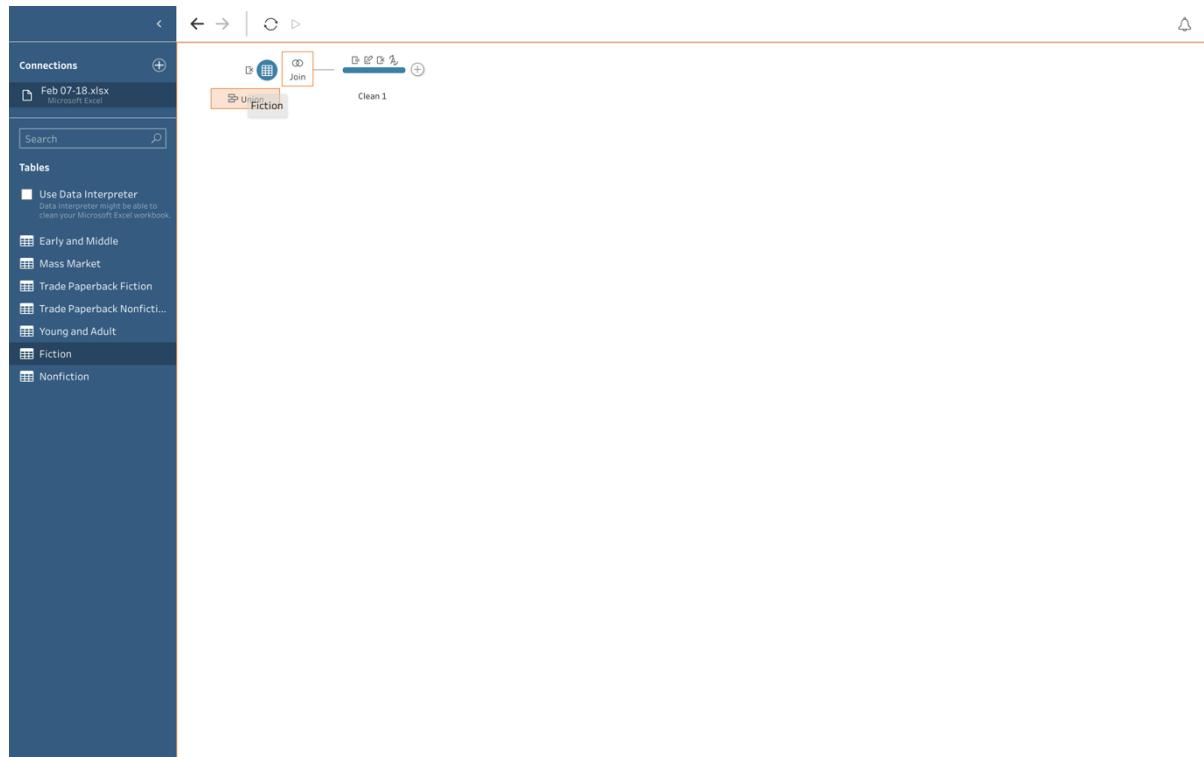
Last Week	Weeks on List	Author	Price	Name of the book	ISBN
1	5	Michael Wolff	30	Fire and Fury	9781250158062
2	40	Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry	9780393609394
3	44	Mark Manson	24.99	The Subtle Art of Not Giving A F*ck	9780062457714
4	16	Walter Isaacson	35	Leonardo da Vinci	9781501139154
15	31	David Grann	28.95	Killers of the Flower Moon	9780385534246
10	4	Daniel H. Pink	28	When	9780735210622
null	1	Dave Eggers	28.95	The Monk of Mokha	9781101947319
5	6	Tiffany Haddish	26	The Last Black Unicorn	9781501181825
9	11	Ray Dalio	30	Principles: Life and Work	9781501124020

Q6. Here you have been asked to connect to Fiction table and perform the following changes: -

- Create a Union for Fiction and Non-fiction table.
- Apply the same edits on the Fiction table.

Sol.: Drag the Fiction table over Non-fiction and select Create Union.

Module 1- Data Preparation Using Tableau Prep



Step 1: Right-click on the line that connects Non-fiction and Clean 1, and select Remove.

The screenshot shows the Tableau Prep interface with the 'Union 1' step selected. A context menu is open over the connection line between the 'Nonfiction' and 'Clean 1' nodes. The 'Remove' option is highlighted. The 'Union Results' pane displays the following data:

Table Names	Last Week/ Weeks on List	Rank	Rank Duplicate	Info
Feb 07-18.xlsx/Nonfiction	1/5	1	null	Fire and Fury Michael Wolff \$30.00 ISBN
Feb 07-18.xlsx/Nonfiction	2/40	2	null	Astrophysics for People in a Hurry Neil de
Feb 07-18.xlsx/Nonfiction	3/44	3	null	The Subtle Art of Not Giving A F*ck Mark H
Feb 07-18.xlsx/Nonfiction	4/16	4	null	Leonardo da Vinci Walter Isaacson \$35.00
Feb 07-18.xlsx/Nonfiction	15/31	5	null	Killers of the Flower Moon David Grann \$2
Feb 07-18.xlsx/Nonfiction	10/4	6	null	When Daniel H. Pink \$28.00 ISBN: 9780
Feb 07-18.xlsx/Nonfiction	/1	7	null	The Monk of Mokha Dave Eggers \$28.95

The screenshot shows the Tableau Prep interface with the 'Union 1' step selected. The 'Connections' pane indicates that the 'Clean 1' node is now connected directly to the 'Union 1' node, as the intermediate connection line between 'Nonfiction' and 'Clean 1' has been removed. An error icon is present in the top right corner.

Step 2: Drag Union over Clean 1 and select the Add option.

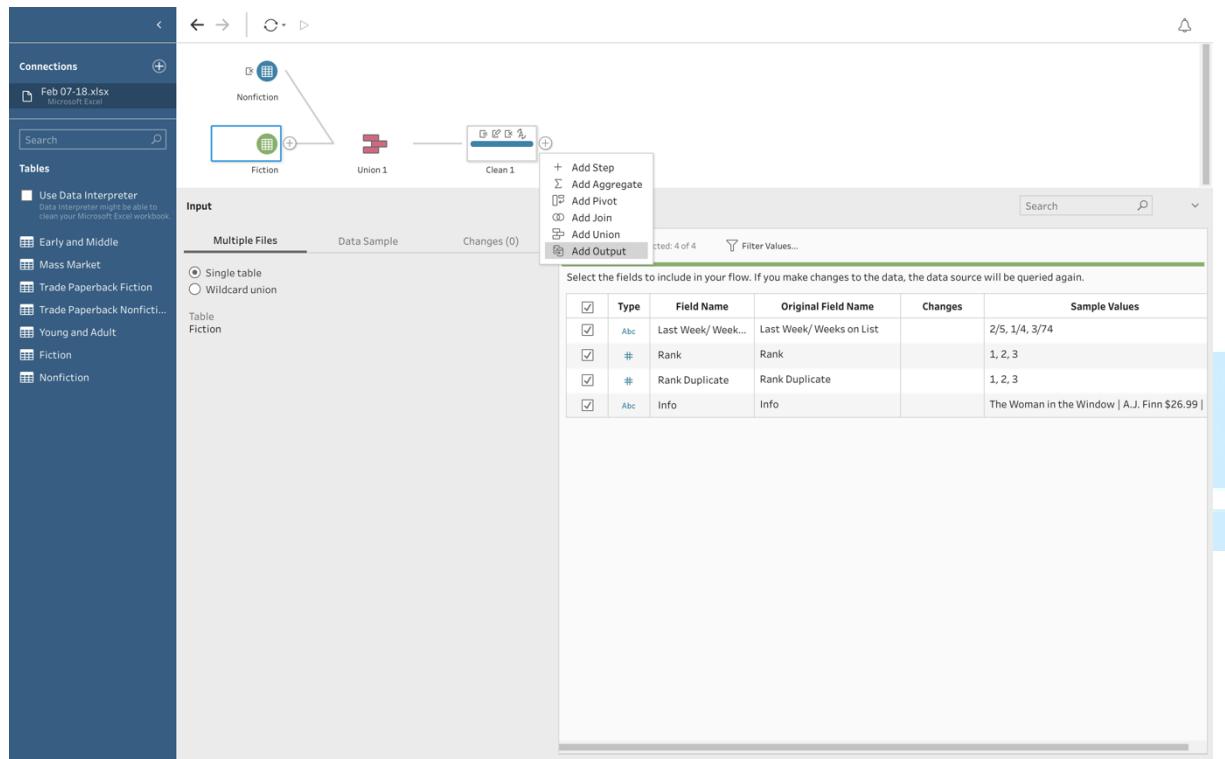
The screenshot shows the Tableau Prep interface with the 'Clean 1' step selected. The top navigation bar includes 'Connections' (Feb 07-18.xlsx), 'Search', and 'Tables'. The left sidebar lists various data sources like 'Early and Middle', 'Mass Market', etc. The main workspace shows a flow diagram with 'Nonfiction' and 'Fiction' inputs merging into 'Union 1', which then connects to 'Clean 1'. The 'Changes' pane on the left details numerous cleaning operations, such as calculated fields for splitting 'Info' into three parts ('Split 1', 'Split 2', 'Split 3') and renaming fields like 'Author' and 'Price'. The bottom pane displays the resulting clean data with columns: '#', 'Last Week', 'Weeks on List', 'Author', 'Price', 'Name of the book', 'ISBN', and 'Last Week/Weeks on List'.

This screenshot is identical to the one above, but the 'Union' icon is now highlighted with a red box, indicating it has been successfully added to the 'Clean 1' step. The rest of the interface, including the flow diagram, 'Changes' pane, and data preview, remains the same.

Q7. Export the data in a CSV format and save the flow in .tflx format.

Sol.: To save the flow, we will need to add an Output step.

Step 1: Click the “+” button and select the Add Output option.



- Now change the Output type to CSV and click on Run flow.

Module 1- Data Preparation Using Tableau Prep

The screenshot shows the Tableau Prep Builder interface. On the left, there's a sidebar with 'Connections' (Feb 07-18.xlsx) and 'Tables' (Early and Middle, Mass Market, etc.). A 'Union 1' stage is connected to a 'Clean 1' stage, which then leads to an 'Output' stage. The 'Output' stage has 10 fields and 30 rows. A modal window titled 'Save to Output.csv' is open, showing a preview of the data with columns: Table Names, Last Week, Weeks on List, Author, Price, and Name of the book. The data includes entries like 'Feb 07-18.xlsx/Nonfiction' by Michael Wolff and 'Feb 07-18.xlsx/Fiction' by A.J. Finn.

- The data is now stored as a CSV file.

Step 2: Go to File and select the Save As option.

This screenshot shows the 'File' menu open in Tableau Prep Builder. The 'Save As...' option is highlighted. The rest of the interface is similar to the previous screenshot, showing the 'Output' stage and the 'Save to Output.csv' modal with the same data preview.

Step 3: In the dialogue box, change the format to Packaged Tableau Flow Files.

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' pane lists a connection to 'Feb 07-18.xlsx'. Below it, the 'Tables' pane shows tables like 'Early and Middle', 'Mass Market', 'Trade Paperback Fiction', 'Trade Paperback Nonfiction', 'Young and Adult', 'Fiction', and 'Nonfiction'. A flow diagram on the right consists of three nodes: 'Nonfiction' (blue), 'Fiction' (green), and 'Union 1' (red). A 'Save As' dialog box is open over the flow, prompting to save as 'Data Flow.tffx' with 'Format: Packaged Table...'. In the bottom right corner, a preview window displays a table titled 'Save to Output.csv' with columns: Table Names, Last Week, Weeks on List, Author, Price, and Name of the book. The table lists various books from the 'Nonfiction' and 'Fiction' tables, such as 'Fire and Fury', 'Astrophysics for People in a Hurry', 'The Subtle Art of Not Giving A F*ck', etc.

Table Names	Last Week	Weeks on List	Author	Price	Name of the book
Feb 07-18.xlsx/Nonfiction	1	5	Michael Wolff	30	Fire and Fury
Feb 07-18.xlsx/Nonfiction	2	40	Neil deGrasse Tyson	18.95	Astrophysics for People in a Hurry
Feb 07-18.xlsx/Nonfiction	3	44	Mark Manson	24.99	The Subtle Art of Not Giving A F*ck
Feb 07-18.xlsx/Nonfiction	4	16	Walter Isaacson	35	Leonardo da Vinci
Feb 07-18.xlsx/Nonfiction	15	31	David Grann	28.95	Killers of the Flower Moon
Feb 07-18.xlsx/Nonfiction	10	4	Daniel H. Pink	28	When
Feb 07-18.xlsx/Nonfiction	null	1	Dave Eggers	28.95	The Monk of Mokha
Feb 07-18.xlsx/Nonfiction	5	6	Tiffany Haddish	26	The Last Black Unicorn
Feb 07-18.xlsx/Nonfiction	9	11	Ray Dalio	30	Principles: Life and Work
Feb 07-18.xlsx/Nonfiction	6	75	J.D. Vance	27.99	Hillbilly Elegy
Feb 07-18.xlsx/Nonfiction	7	2	Steven Levitsky	26	How Democracies Die
Feb 07-18.xlsx/Nonfiction	12	8	Mary Beard	15.95	Women & Power
Feb 07-18.xlsx/Nonfiction	8	46	The Dalai Lama	26	The Book of Joy
Feb 07-18.xlsx/Nonfiction	11	3	Margareta Magnusson	18.99	The Gentle Art of Swedish Death Cleaning
Feb 07-18.xlsx/Nonfiction	13	17	Ron Chernow	40	Grant
Feb 07-18.xlsx/Fiction	2	5	A.J. Finn	26.99	The Woman in the Window
Feb 07-18.xlsx/Fiction	1	4	Chloe Benjamin	26	The Immortalists
Feb 07-18.xlsx/Fiction	3	74	Amor Towles	27	A Gentleman in Moscow
Feb 07-18.xlsx/Fiction	5	18	Jesmyn Ward	26	Sing, Unburied, Sing
Feb 07-18.xlsx/Fiction	4	21	Celeste Ng	27	Little Fires Everywhere
Feb 07-18.xlsx/Fiction	null	1	Jojo Moyes	27	Still Me
Feb 07-18.xlsx/Fiction	6	18	Dan Brown	29.95	Origin