



A PRACTITIONER'S GUIDE TO

TABLEAU PREP BUILDER



2019.1

A Practitioner's Guide to Tableau Prep Builder

Published by

USEReady

1407 Broadway, Suite 721, New York, NY 10018

www.useready.com

Copyright © 2015 by USEReady New York

ISBN: 978-1-64516-475-3

Manufactured in the United States of America

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the USEReady, 1407 Broadway, Suite 721, New York, NY 10018 or online at www.useready.com

Limit of Liability / Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a specific purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies provided herein might not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting professional services. Neither the publisher nor the author is liable for damages arising because of information published in this book. The fact that an organization or a website is referred to in this book as a citation and/or potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that internet sites listed in this work may have changed or disappeared between when this book was written and when it is read.

Table of Contents

About USEReady	9
About Tableau.....	11
Letter from the CEO	13
Foreword (Jeff Black).....	15
Foreword (Jeremy Walsh).....	17
Preface	19
To the Reader	21
INTRODUCTION to TABLEAU PREP BUILDER.....	23
1.1 Tableau Prep Builder.....	25
1.2 Tableau Prep Builder Interface	27
1.3 Steps in Tableau Prep Builder.....	33
1.4 Features of a Workflow.....	34
INPUT and OUTPUT	53
2.1 Connect to Microsoft Excel.....	55
2.2 Data Interpreter.....	57
2.3 Configuration Window.....	60
2.4 Data Sampling.....	64
2.5 Connect to Tableau Data Extract File.....	69
2.6 Connect to Text File.....	71
2.7 Connect to Snowflake Server	72
2.8 Connect to Microsoft SQL Server	76
2.9 Other Features.....	79

DATA CLEANING	91
3.1 Profile Pane: An In-Depth Analysis	93
3.2 Functions	97
3.3 Tracking Changes.....	112
3.4 String Calculations	115
GROUP and REPLACE	123
4.1 Automatic Group and Replace Functions	125
4.2 Manual Grouping.....	130
4.3 Example	131
AGGREGATION and PIVOT	153
5.1 Aggregations.....	155
5.2 Aggregation Functions	156
5.3 Group By	160
5.4 Example 1	162
5.5 Pivot	168
5.6 Example 2	171
JOINS and UNIONS	177
6.1 Join	179
6.2 Types of Joins.....	181
6.3 Example 1	188
6.4 Union	210
6.5 Example 2	214
6.6 Example 3	224

TABLEAU PREP BUILDER CONDUCTOR	241
7.1 Tableau Prep Builder Conductor.....	243
7.2 Prerequisites	244
7.3 Publishing Workflows to Server.....	247
7.4 Scheduling Workflows on Server.....	251

About USEReady

USEReady's mission is to help users succeed with data.

We achieve this mission with fanatical customer centricity, humility, and integrity. At USEReady, we love to solve customer problems, contribute to community and continue to improve. We aim to build a strong learning culture and have fun doing it together.

We have harnessed modern Business Intelligence solutions with cutting edge Artificial Intelligence (AI) platforms to unlock the power of data. Our result-oriented solutions ensure that our customers are successful in adopting self-service technologies.

We have partnered with the best in class BI, Data and Cloud products such as Tableau, Snowflake, Alteryx, Informatica, Alation, Collibra, AWS and Microsoft. We have delivered proven success across Financial Services, Insurance, Retail and Media verticals.

We have been nominated and won several awards along this journey. Check us out at www.useready.com

About Tableau

Tableau Software, headquartered in Seattle, Washington, is a leader in data visualization and data analytics solutions. It was founded in 2003 to create software products focused on Business Intelligence with a revolutionary approach.

The company was recognized as the leader of the Gartner Quadrant for five consecutive years from 2012 to 2017 due to its remarkable Ability to Execute and Completeness of Vision.

Hundreds of thousands of data analysts around the world rely on Tableau Software products for their daily data preparation, data visualization, and data exploration needs. Tableau is also recognized as one of the top products in delivering powerful end-user interactivity supported by self-service capabilities.

In simple terms, Tableau Software helps people see and understand data.

For more details, please visit <https://www.tableau.com>.

Letter from the CEO



Tableau is an important BI partner and the key component to USEReady's goto market strategy. Tableau pioneered self-service BI and as BI is going self-driven, Tableau is emerging into a platform. I meet Francois Ajenstat, CPO, Tableau at least four times a year. During one of our meetings, I asked Francois "Why ETL product from Tableau? It is such a crowded space".

He said "Well, Tableau has over 80K customers and less than 10% are using a self-service ETL product. This means we have an unmet need for the rest of the 90%". I felt that is a pretty accurate view. Most Tableau creators rely on either IT to assist them with data prep or message data using a spreadsheet. Neither approach is ideal for rapid fire BI. With this new product, Tableau Prep Builder, Tableau customers have a tool that assists them with data shaping efforts.

At USEReady, we believe in helping organizations succeed with data. We are very proud of the fact that we carry this mission with a sense of customer centricity, community, integrity and humility. To serve the community we are part of, our consultants conduct several doctor sessions, webinars, author content, write blogs etc. Along with that strategy we have authored this book "A Practitioner's Guide to Tableau Prep Builder". This is the second book in our "Practitioner's Guide" series.

A total of ten practitioners have contributed their expertise to create this book, incorporating illustrations and hands-on exercises.

Finally, I am thankful to Vijai Narasimha, USEReady to lead the book initiative, Jeremy Walsh, Strategic Customer Success Manager, Tableau Software and Jeff Black, Customer Consulting Manager, Financial Services Tableau Software for reviewing and writing the foreword to this book.

I am quite confident that Tableau community is going to embrace Prep with the same love they bestowed to Tableau Desktop. You are welcome to share me your feedback or thoughts at udayh@useready.com

Uday Hegde
CEO
USEReady

Foreword



As data exponentially grows, continues to be created, and is stored in more places, the challenge of preparing data is only becoming more difficult. Often in the world of data analytics, 80% of a person's time is spent getting the data ready, leaving only 20% of time for analysis. Considering this ratio, if you have one hour to obtain data, prepare it, and answer any necessary questions, approximately 48 minutes of that time will be spent obtaining and preparing data. Merely 12 minutes are left to perform the value-add analysis. This ratio is a problem that needs balance.

Having worked with hundreds of enterprises and spending years in the financial services industry, this is one of the biggest challenges I see and have also experienced firsthand. As a former customer, bringing data together to answer questions about the business was a considerable challenge; many individuals do not possess the proper solution or skills to effectively resolve this challenge. At Tableau, I have spent a significant amount of time testing Tableau Prep Builder during its early alpha and beta stages. Since its release, I am now seeing customers balance this prep-to-analysis ratio with the product.

For those wanting to learn Tableau Prep Builder, this practitioner's guide—written by certified BI consultants—will build the foundation needed, whether you're an individual in business, a journalist who reports data findings, or a student at an educational institution.

Given their experience as a certified training partner and collaborating with enterprises on strategic data projects, USEReady is positioned to help you succeed with data.

The aim of this book is to help you learn Tableau Prep Builder, and in doing so, you will be better equipped to solve data challenges.

Jeff Black

Customer Consulting Manager, Financial Services

Tableau Software

Foreword



I have spent nearly a decade helping Tableau customers achieve self-sufficient analytics, most recently as a Strategic Customer Success Manager working closely with Tableau's largest customers on success planning, driving adoption, and enabling users on Tableau's Enterprise Platform.

Now more than ever, there is a demand to empower users to clean, combine, aggregate, and prepare data for analysis. USEReady is a leader in this discipline and has once again created a training manual not only to help users adopt Tableau Prep Builder quickly and easily but to further Tableau's mission: Help People See and Understand Data.

USEReady has created a book designed for those who are just beginning their journey on data cleansing and preparation. Tableau Prep Builder is a new tool in the Tableau Product suite designed to make data prep easy and intuitive. By the time you finish working the examples and real world use cases in this easy-to-follow training aid, you will know how to connect to a variety of files or servers, combine data from multiple data sources, easily bring tables into a "flow pane" and perform operations such as validate, filter, pivot, union, and join to clean and prepare, or shape, your data.

Tableau disrupted the BI space in 2003 with the introduction of Tableau Desktop with its "drag and drop" concept that thousands of customers have come to know and love today. With Tableau Prep Builder, Tableau has again found a way to disrupt the BI space, helping those thousands of customers continue their data journey through data preparation.

USEReady fully appreciates Tableau's mission and for many years has helped customers achieve self-service analytics through faster insight, and faster time-to-market with Tableau's Enterprise solution, which now includes Tableau Prep Builder.

Jeremy Walsh
Strategic Customer Success Manager
Tableau Software

Preface

The world of data management has changed a lot in the last decade. Companies are amassing huge amounts of data every time the clock ticks. Maintaining such volumes of data and making sense of it are becoming more challenging every day. But the bigger challenge is using this data for reporting needs.

Data can come in various forms and shapes. There is a constant need for data preparation—a series of steps or processes to combine, shape, clean, and organize the data for easy usage.

Whenever data preparation is the topic of discussion, the words *Extract*, *Transform*, and *Load* (ETL) come into play. ETL simply says this: **extract** data from different sources, **transform** data into the required format, and **load** data into a system that is ready for consumption.

In the end, people who can control their data and use a combination of best Business Intelligence tools will have the right business insights and make better data-driven decisions.

In the world of Tableau, a widely used tool is Tableau Desktop, one of the best data visualization tools for ease of use, adaptability, and attractive features for end users. Tableau developers have faced issues developing the infographics attributed to the data, ranging from the shape of the data to missing information, unwanted and redundant data, mismatched elements, and the need for a basic cleanup. That limited Tableau developers' creativity, and they ended up depending on the data team. But with the introduction of Tableau Prep Builder, a new addition to the Tableau software suite, many

of those issues were resolved, and the Tableau Desktop developers breathed a sigh of relief.

It was a great move on Tableau's part to introduce this simple and fantastic tool with very powerful capabilities needed by Tableau analyst's day in and day out.

To the Reader

This book is written as a hands-on learning manual. Its aim is to cover most of the aspects and functionalities of Tableau Prep Builder. It is very understandable to users who have minimal experience with ETL tools and demonstrates the features through fundamental concepts.

Illustrations are provided throughout the book for each step, so the user can easily follow along as the chapters progress. At the same time, the book also provides great reading material. It is divided into seven chapters, with the first six focusing on content related to data preparation. The last chapter demonstrates the leverage of Tableau Prep Builder Conductor on Tableau Server. All the chapters are treated as use cases with data sets prepared and modified for training purposes.

The exercises will use data that can be download from <https://resources.useready.com/publications/a-practitioners-guide-to-tableau-prep-builder-version-2019-1/> by following the instructions on the website to unpackage the file.

CHAPTER 4

GROUP and REPLACE

As discussed in the previous chapter, data cleaning can be one of the most important steps of the workflow. It can also be a labor-intensive job for any analyst. Human error will often lead to data field domains becoming needlessly large and field aggregations becoming entirely inaccurate.

Fortunately, thanks to both the automatic and manual Group and Replace functions in Tableau Prep Builder, a lot of this busywork is virtually eliminated.

4.1 Automatic Group and Replace Functions

There are three automatic Group and Replace functions in Tableau Prep Builder — *Spelling*, *Common Characters*, and *Pronunciation*. These three functions uniquely analyze the domain of a given field and automatically group similar fields together when they believe a mistake has been made. Although each of these functions are very powerful, using one of these functions is often not enough. Sometimes, we may need a combination of these.

Let us start by examining the *Spelling* and *Common Characters* functions. They are the most similar of the three algorithms. Consider this example: We have a data set with a field called *Month*. The domain consists of five members: *September*, *March*, *August*, *January*, and a new month named *Septembur*. Now clearly, this is a mistake that needs to be corrected.


By selecting *Spelling* or *Common Characters*, Tableau Prep Builder automatically groups these two values together and replaces every instance of *Septembur* in the dataset with *September*.

September (1)
March
August
January
Septembur (2)

Table 4.1 – Misspelled words

September
March
August
January

Table 4.2 – Grouping (1) and (2)

Once we have grouped and replaced within a field, the default value of grouped members will change to the most frequently repeating value within that group. When the group is created, a small paper clip  appears next to the new group in the domain. We can edit the name of this group by right-clicking the group name in the left column and selecting *Edit Value*. Or if we believe there has been a mistake and one of the values cannot be grouped together, we can just right-click the incorrect group and select *Ungroup*.

Even though the *Spelling* and *Common Characters* functions can sometimes have similar outputs, they still have their own unique use cases and methodologies. The *Spelling* function, for instance, is best used in fields where the domain consists of actual words.

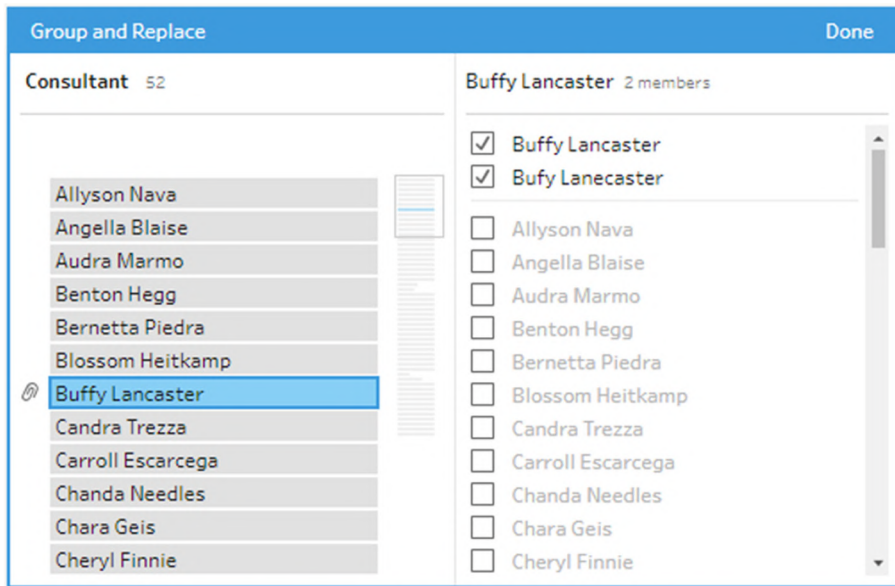


Figure 4.1 – Grouping Preview

Names, locations, categories, and other dimension fields are great examples of when to use this specific function.

Common Characters, on the other hand, is most suited for fields where the domain consists of strings that don't have a spelling but rather an order of specific characters. Its greatest utility lies in correcting ID fields, project codes, or other non-sequential character strings.

The third and final automatic grouping function is *Pronunciation*. Unlike the previous two functions, which look at the grammatical or sequential structure of domain members, this function analyzes how they are broken up phonetically. Like the Spelling

function, it is best used on fields that actually have a pronunciation. Again, names, categories, and other dimension fields are most suitable.

Now let us look at the underlying logic that Tableau Prep Builder uses to make these automatic groups. Understanding these processes will help us make better decisions on which function to use and when.

There can be situations where more than one of the above-described functions may be needed for the same field.

The *Spelling* function is based on the **Levenshtein distance** algorithm and works on all eight supported Tableau Prep Builder languages—English, French, Portuguese, Chinese, Korean, Japanese, Spanish, and German. What the algorithm does is determine which correctly spelled word has the shortest Levenshtein distance with the misspelled word. The Levenshtein distance is the number of individual characters that would either need to be inserted, substituted, or deleted from a misspelled word in order for it to match a correctly spelled word within the existing domain.

Using our September vs. Septembur example, these two words would have a Levenshtein distance of one. Since this is the lowest possible number, *Septembur* will be replaced with *September* and group the values together in the field.

The *Common Characters* function, on the other hand, uses **n-gram fingerprint** method in order to establish groups and patterns. This algorithm turns every member of a field's domain into a key of their unique characters after removing white spaces, punctuation, and so forth. Members with similar or identical keys are then grouped with one another, and the values are replaced.

The *Pronunciation* function employs the **Metaphone 3** algorithm produced and licensed by Lawrence Philips. The basic premise of the algorithm is that it breaks all words down into their phonetic groupings. It then compares the phonetic keys across all members of a domain, and similar ones are grouped together once they meet a certain threshold of commonality.

Using our *September* vs. *Septembur* example, we can see how the two words have similar phonetic makeups and how the comparison between the two could be drawn.

When choosing which one of these automatic grouping functions to use, it is important to keep in mind their underlying methodologies. Which one makes the most sense in the context of a certain field? When examining the domain, which types of errors are present? In which order should the automatic functions be enacted? It may take a few tries to get the correct combination of grouping functions, but these are the types of questions we should ask ourselves in order to get on the right track.

4.2 Manual Grouping

Oftentimes, we'll find that the automatic grouping functions aren't enough to get our data set where we want it to be. They will do a great job of getting 99% of the desired groups but will require a manual touch in order to get the data ready to be analyzed. That is where Tableau Prep Builder's *Manual Grouping* feature comes in handy.

The need for manual grouping typically stems from a couple of different scenarios:

- There are too many typos to the point where none of the algorithms can establish a pattern. This is a rare occurrence but does happen.
- A pattern is established between two domain members that are meant to remain separate.

For example, A *Name* field contains *Freddy Vargas* and *Freddie Vargas* in its domain. They are two different people but could be accidentally grouped together with the Pronunciation function.

- There is a desire to combine two domain members that share little to no commonality. That can often result from formatting differences.

For example, a *Client* field contains *Dana Sherman* and *Marla Feld* in its domain. They both work for the same company called *Sutton Systems*, and we want to have their data linked together. We can then manually group these and rename them both *Sutton Systems*.

As another example, a field is supposed to be binary, but due to formatting inconsistency, there are three values instead of two. The domain includes these values: *True*, *False*, and *Yes*. There must have been a miscommunication regarding formatting, so we must now group the *True* and *Yes* together in order to rectify this. The new domain values are now *True* and *False*.

There are many reasons why someone may need to use manual grouping, but these seem to be the most common. Automatic grouping functions should normally be where we start when trying to combine domain members, but they should always be double-checked.

4.3 Example

Now we are ready to try an exercise. The *Unorganized & Associates.xlsx* file required for this example is in Chapter 4. This data set is for a fictional consulting company called *Unorganized & Associates*. As introduced in Chapter 2, the data set is about a professional services firm and its consultants which has information about Resource Utilization.

Requirement: We need to help this consulting firm clean up their data using automatic and manual Group and Replace functions.

Our first step is to connect to the data source. We need to select *Microsoft Excel* from the connections pane and navigate to the data source *Unorganized and Associates* in the Chapter 4 folder.

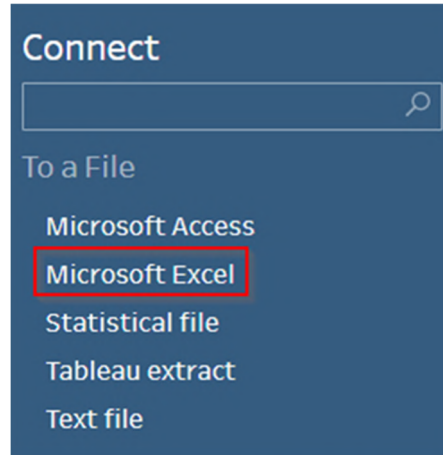


Figure 4.2 – Connections Pane

Next, we need to drag *Number of Hours* into the workflow pane from the Tables pane.

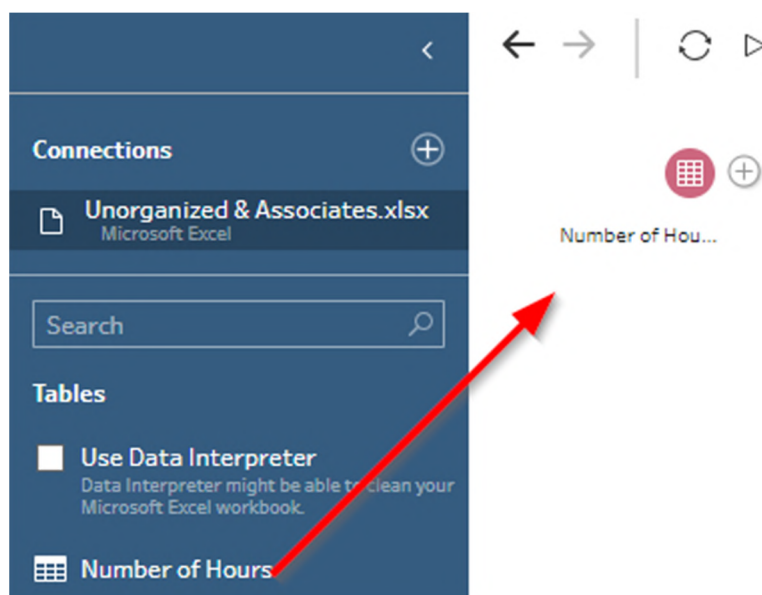


Figure 4.3 – Table selection

Then we need to add a cleaning step so we can both examine and repair the domains of our fields.

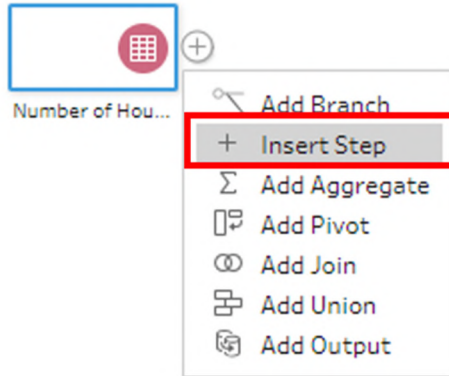


Figure 4.4 – Adding a Cleaning step

Now let us look at the domains for our *Consultant* and *Client* fields. The firm should only have 51 employees and 74 clients, so there has clearly been some inconsistency in terms of reporting. Let us see if we can get those numbers down. Let us start by trying to group the *Consultant* field with Spelling. We can click on the ellipses or More options dropdown.

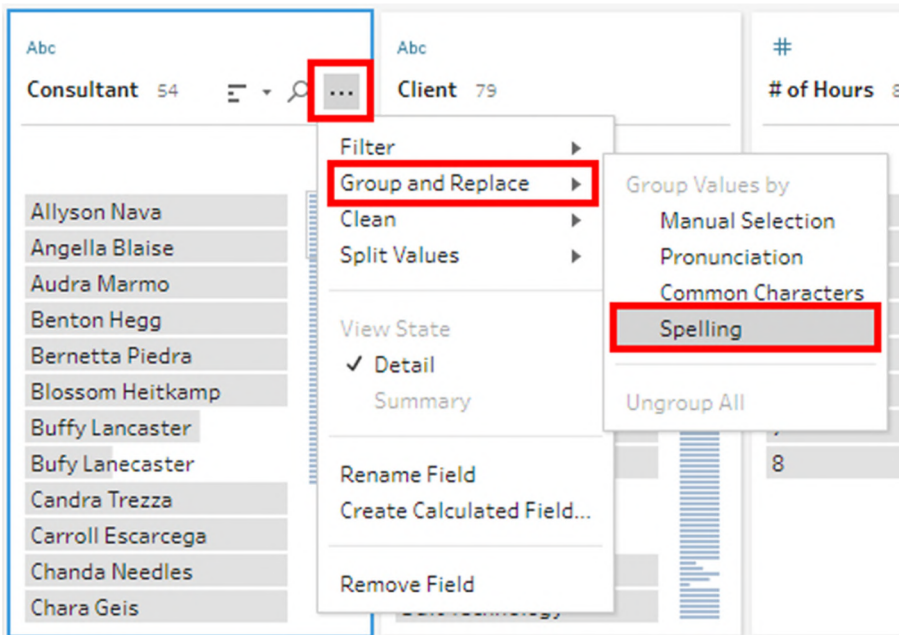


Figure 4.5 – Group on Spelling

After that is applied, the *Consultant* domain should now shrink to 52 members. We can see that Tableau Prep Builder has created two groups: *Buffy Lancaster* and *Marylin Stoddard*. As soon as the group is created, we can see the *paperclip* indicator. We will use other tools later on to trim it down further.

***Note:** The Changes pane as seen in Chapter 3 where all changes performed in this step are logged. As we use more Group and Replace functions, more changes will be displayed here.

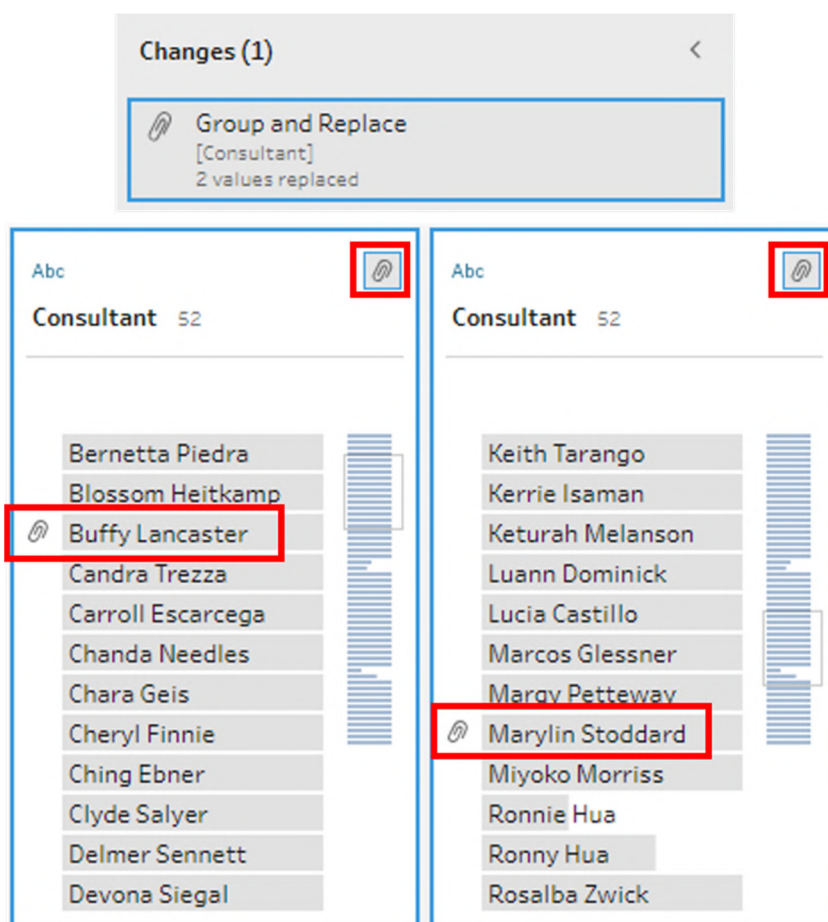


Figure 4.6 – Consultant field – grouped changes

Now let us look at the *Client* field. Again, the domain has too many members. There should only be 74 clients, but there are currently 79.

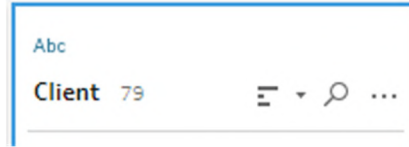


Figure 4.7 – Client field

Let us start by selecting the *Common Characters* function from the options menu and see how much the domain is reduced.

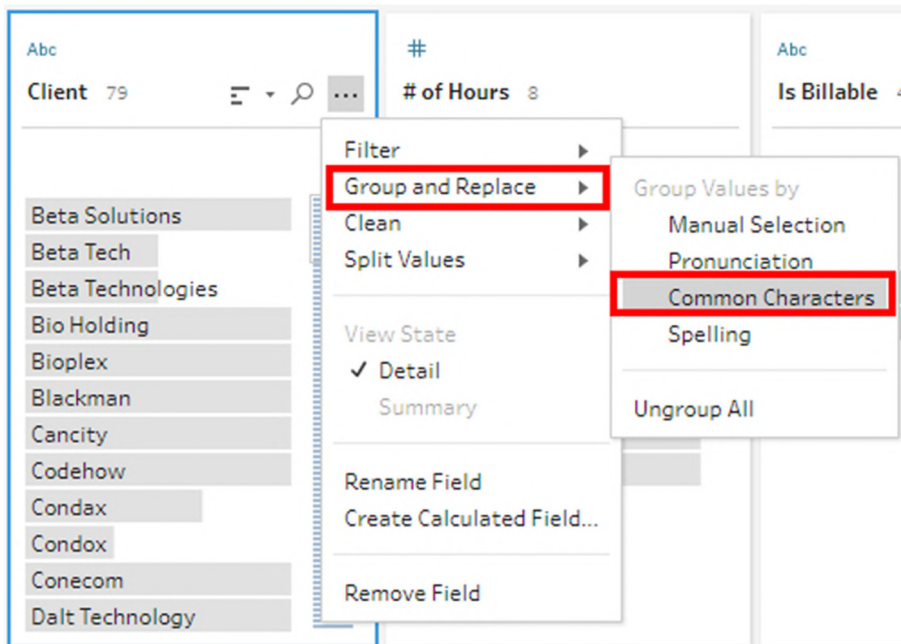


Figure 4.8 – Group on Common Characters

The domain for Clients has been reduced from 79 to 78, and a new item has been added to the Change pane highlighted below.

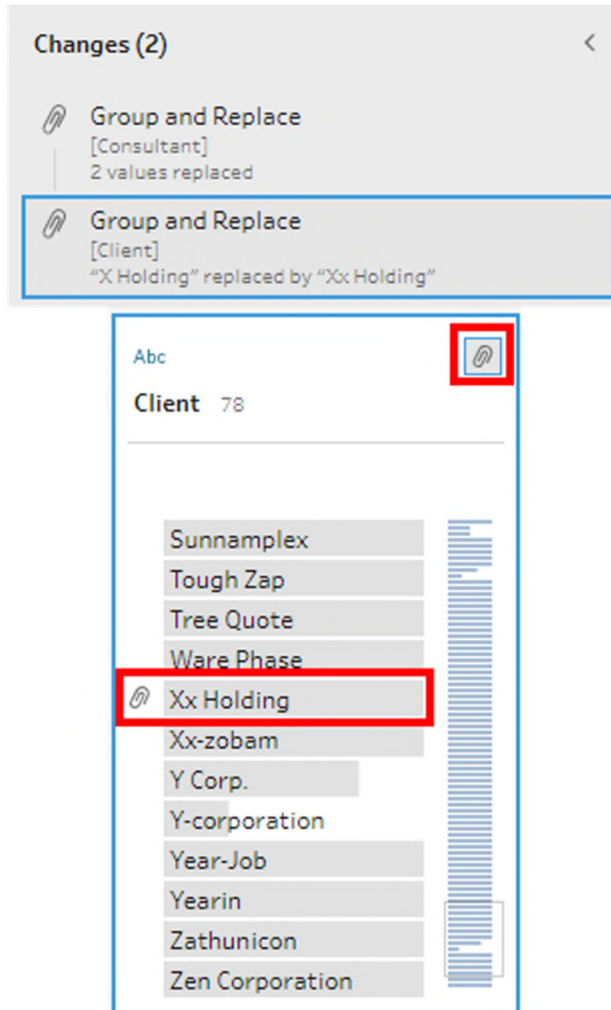


Figure 4.9 – Client field – grouped changes

If we try grouping with Spelling as well, we will reduce the domain of the field from 78 to 74 and create some new groups like Condax, Doncon etc.

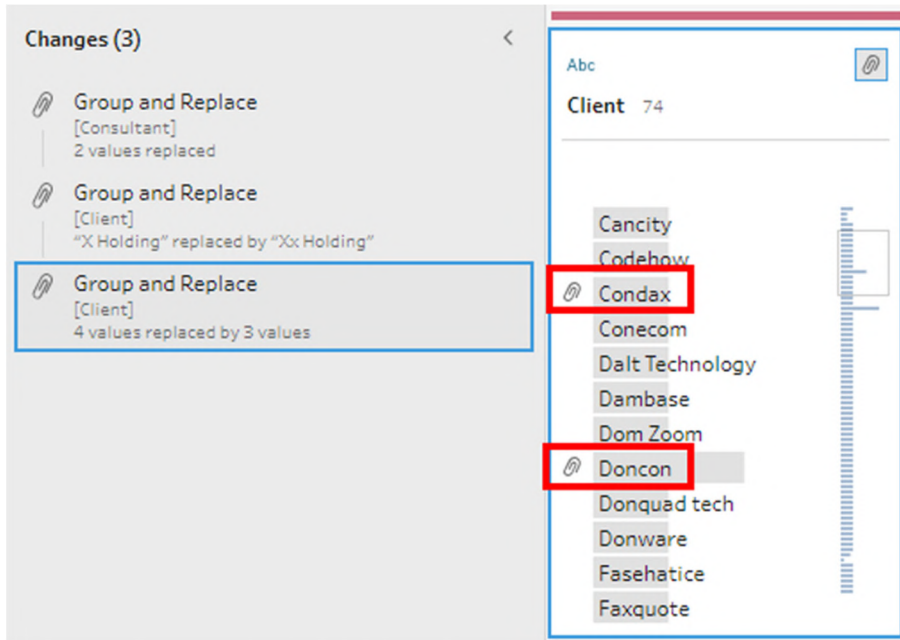


Figure 4.10 – Client field – Group on spelling

Now let us look at our next Group and Replace function.

Until now, we've only been using the Spelling and Common Characters functions to clean our data. Now let us use Pronunciation to further reduce the size of the domains. Select Pronunciation from the list of Group and Replace functions under the *Consultant* field.

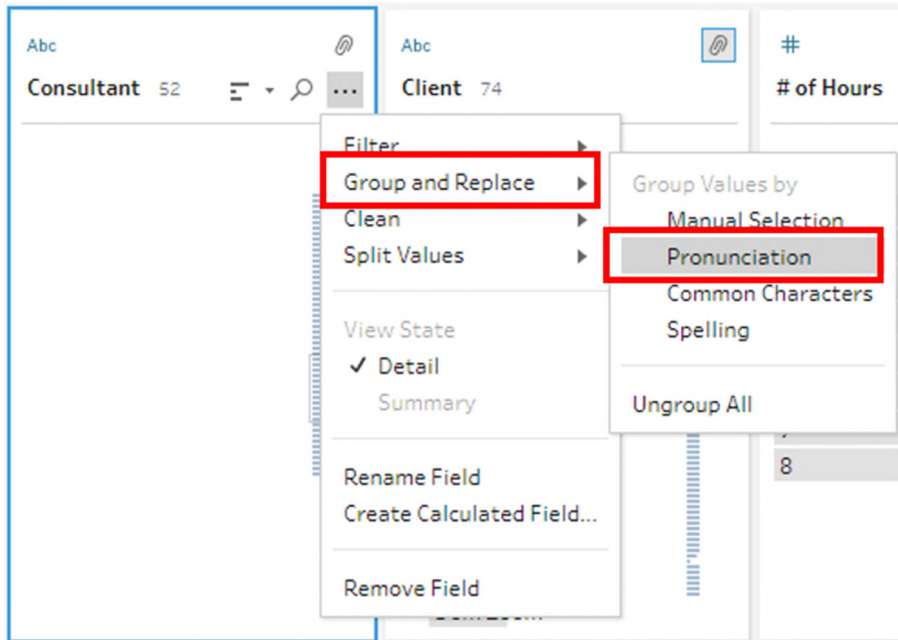


Figure 4.11 – Group on Pronunciation

After performing this action, our new domain for the *Consultants* field has been reduced from 52 members to 50. One of the groups created is *Frederick Ferrari III*. During data entry, someone wrote *Frederick Ferrari the Third* instead of *Frederick Ferrari III*. The other group, *Ronny Hua*, was created from a spelling error.

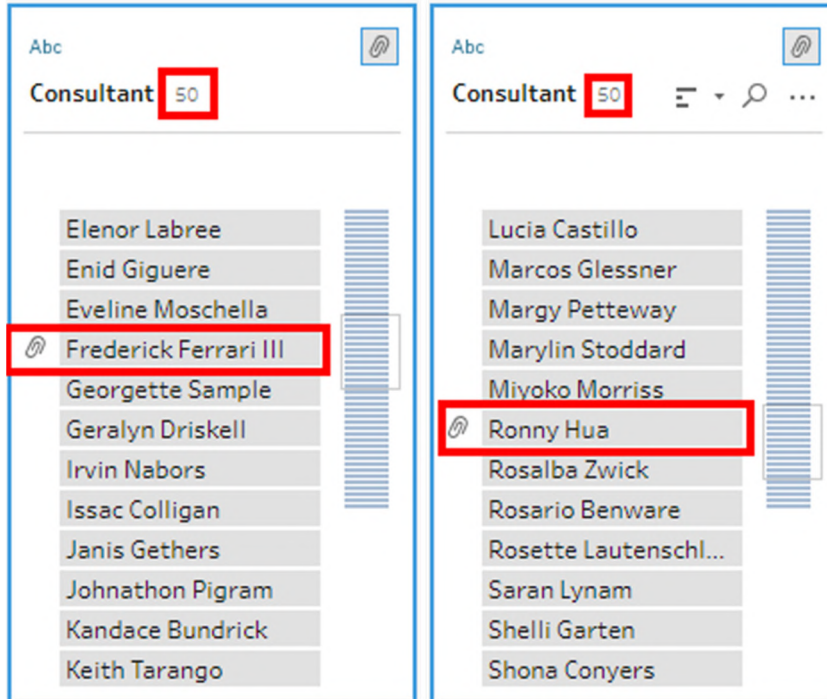


Figure 4.12 – Consultant field - Group on Pronunciation

The Pronunciation function accidentally over-reduced the size of the *Consultants* domain. Let us ungroup *Ronny Hua* in order to fix that mistake. We can right-click on the member and select the *Ungroup* option.

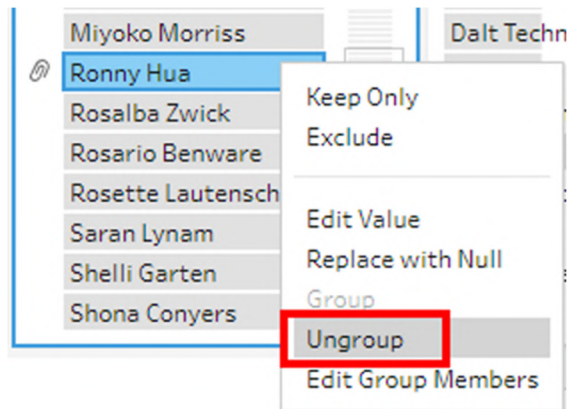


Figure 4.13 – Ungrouping

The *Consultant* field has now been fully cleaned. Let us move on to fixing the *Client* field's domain. We can perform the same Pronunciation function on this field as well.

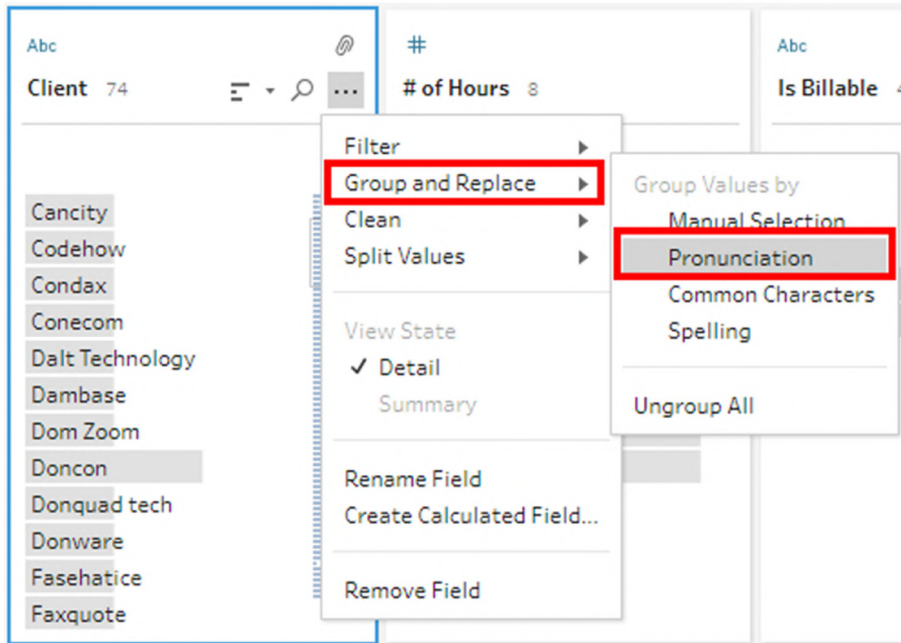


Figure 4.14 – Group on Pronunciation

After grouping we have 72 members in the *Client* domain. Two new groups have been created under the names *Y Corp.* and *Beta Technologies*, which have rectified multiple formatting issues. The *Beta Technologies* group combined *Beta Technologies* and *Beta Tech.*, while the *Y Corp.* group combined *Y Corp.* and *Y Corporation*. This is an example of how the Pronunciation function can recognize and interpret abbreviations in a data set.

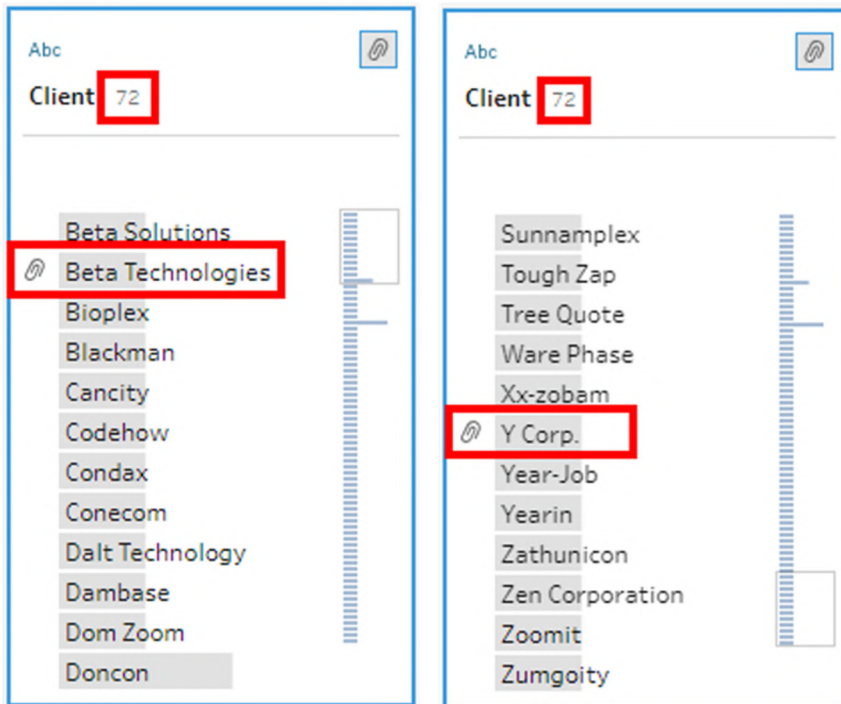


Figure 4.15 – Client field - Group on Pronunciation

This field is almost completely normalized now. We have just one more correction to make. It appears that we have two different divisions of the same company listed on the client sheet. We need to combine *Beta Solutions* and *Beta Technologies* into *Beta Corporation*. We can click the *Manual Selection* option under *Group and Replace*.

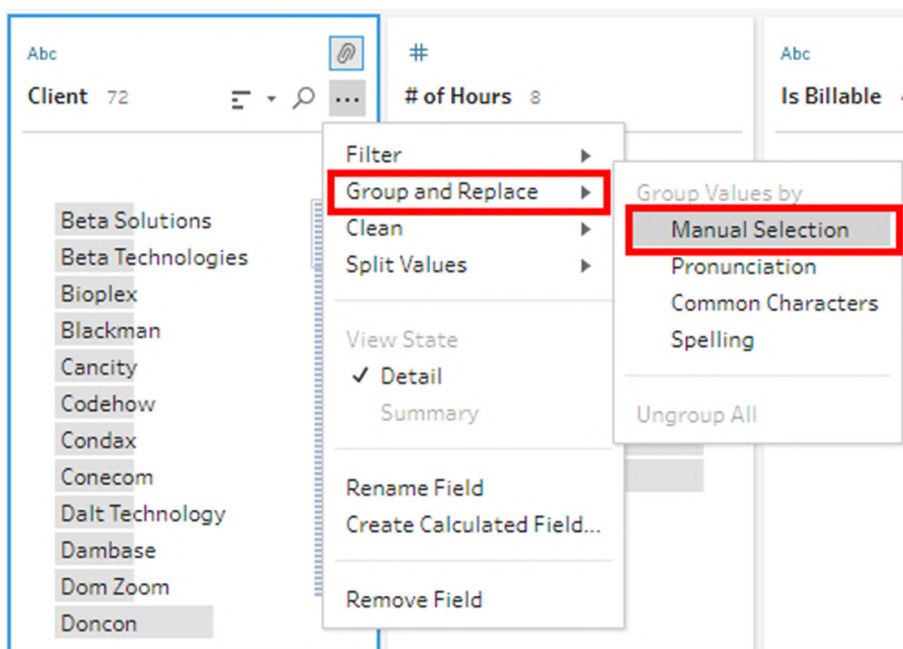


Figure 4.16 – Group on Manual Selection

Now we can select *Beta Solutions* and check the box next to *Beta Technologies* in the pane to the right. Then click *Done* in the top right corner.

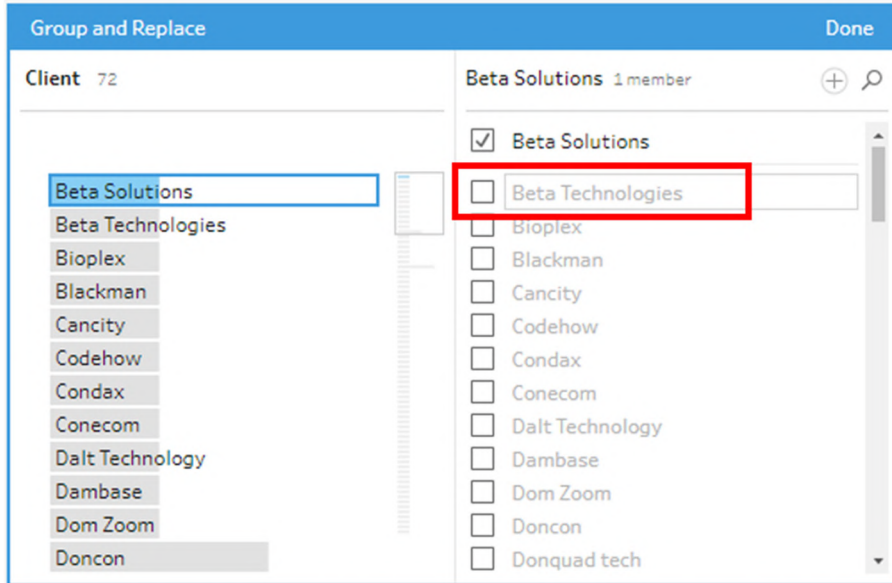


Figure 4.17 – Beta Corporation – Manual Selection

The name now needs to be changed. We can right-click on the new group we just created and select the *Edit Value* option. Type in *Beta Corporation* and press Enter.

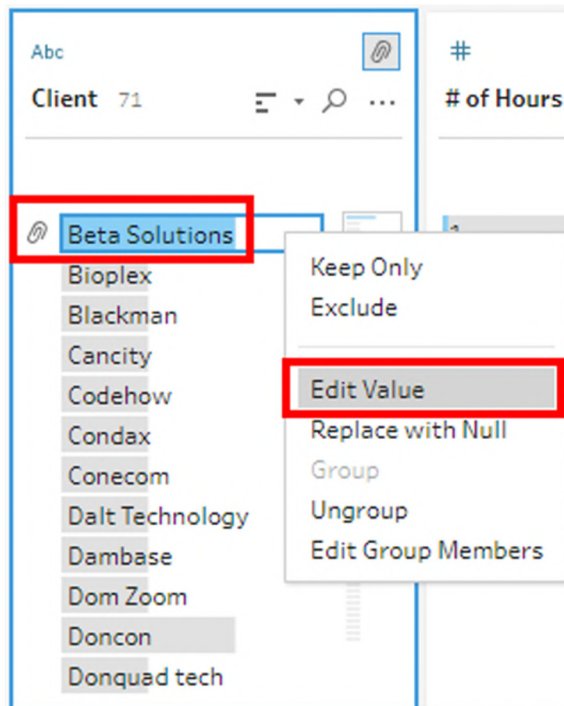


Figure 4.18 – Edit Value – Creating Alias

If we examine this change in the *settings* pane, there is now a red dot next to *Beta Corporation*. That's because this value is not present in the original data source, and Tableau Prep Builder is making us aware of that. This way any new member will get added into this group automatically when the data updates.

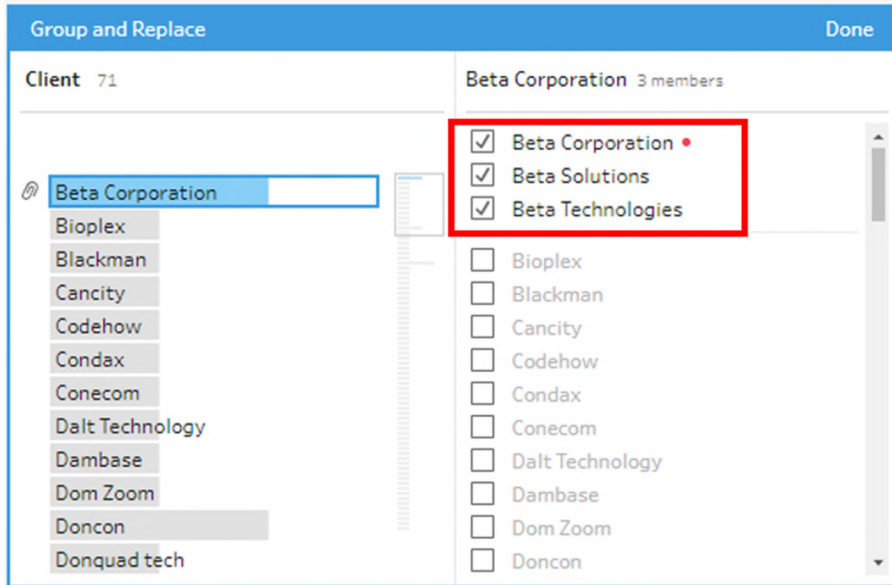


Figure 4.19 – New Value placeholder

Now the *Client* field has been completely repaired. The only field left to clean is the *Is Billable* field. That is supposed to be a binary field with the only two values in the domain being *True* and *False*. Due to a data entry error, the values *Yes* and *No* are also included in the domain. We again have to use *Manual Selection* in order to correct this mistake. Select the *Manual Selection* option from the dropdown menu.

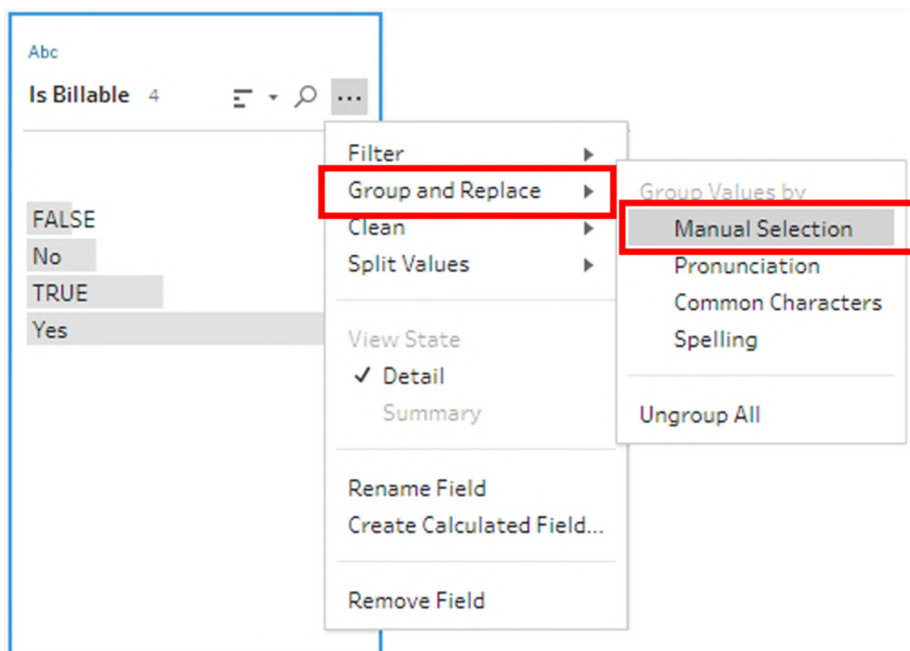


Figure 4.20 – Group on Manual Selection

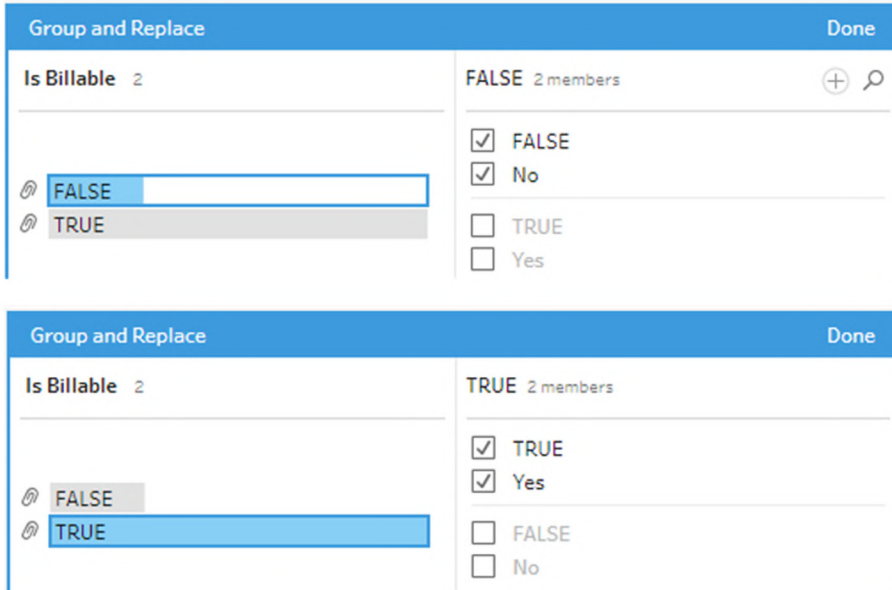


Figure 4.21 – Is Billable field - Group on Manual Selection

We can perform the same manual grouping as in the previous step and then add an Output step to the workflow. Set the Output type to Comma Separated Value (CSV) and select the destination. Now our data is clean, exported, and ready to be analyzed in Tableau. At any point in the process, we can visit the changes to track all the operations.

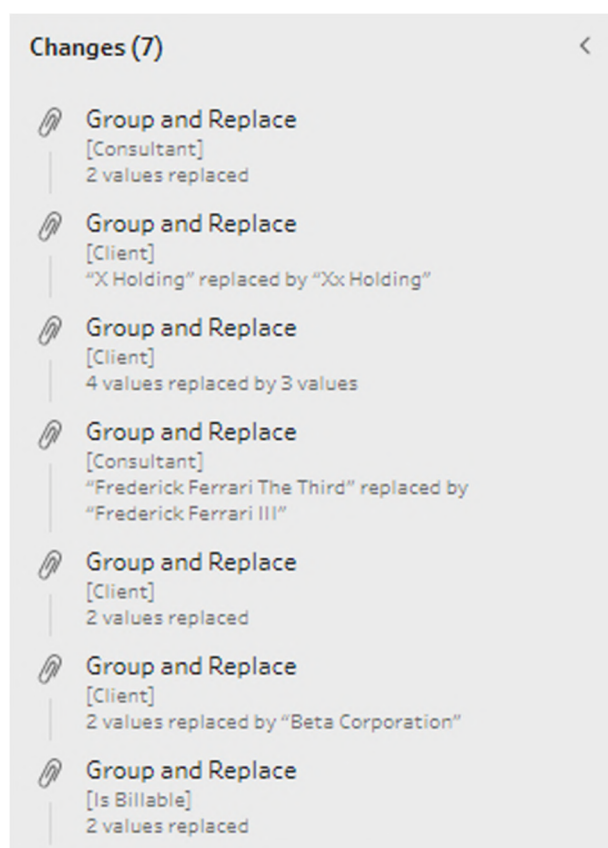


Figure 4.22 – All Changes



Figure 4.23 – Final workflow



Jeff Black, Tableau

For those wanting to learn Tableau Prep, this practitioner's guide—written by USEReADY Tableau consultants—will build the foundation needed, whether you're an individual in business, a journalist who reports data findings, or a student at an educational institution. Given their experience as a certified training partner and collaborating with enterprises on strategic data projects, USEReADY is positioned to help you succeed with data.



Jeremy Walsh, Tableau

Now, more than ever, there is a demand to empower users to clean, combine, aggregate, and prepare data for analysis. USEReADY is a leader in this discipline and has once again created a training manual not only to help users adopt Tableau Prep quickly and easily but to further Tableau's mission: Help People See and Understand Data.

Authored By

Akshay Venkatesh | Anirban Ghosh | Sachet Kashyap | Srini Ramanathan | Moyses Ramirez
Rahul Shetty | Rajesh Hegde | Anish Bhola | Fred Ferrari | Vijai Narasimha

GET IN TOUCH : www.useready.com | contact@useready.com

@2019 USEReADY. All rights reserved.

ISBN 978-1-64516-475-3

