



TABLEAU PREP

FOR BUSINESS INTELLIGENCE



With Best-Selling Tableau Instructor Dustin Cabral



Course Structure



This is a **project-based** course, for students looking for a practical, hands-on, and highly engaging approach to learning Tableau Prep for business intelligence

Course resources include:

- ★ **Downloadable Ebook** to serve as a helpful reference when you're offline or on the go
- ★ **Quizzes & Homework Exercises** to reinforce key concepts, with step-by-step solutions
- ★ **Bonus Projects** to test your abilities and apply the skills developed throughout the course

Course Outline

1

Intro to Tableau Prep

Download Tableau Prep, explore the Tableau Prep visual dictionary, and discover community features

2

Connecting to Data

Start and open a workflow, connect to data sources, manage field metadata and properties, perform wildcard unions and merge fields

3

Examining & Filtering

Build and organize your flow, review data types and size, and filter your data using values and calculations

4

Operations & Calculations

Leverage value and field operations like group, clean, convert, and split, and create custom calculations (LODs and more)

5

Combining & Pivoting

Combine and pivot your various data by leveraging aggregate, join, union and pivot tools

6

Sharing & Updating

Create full and incremental refresh outputs to deliver data to Tableau Server, databases or local flat files

Introducing the Course Project

THE **SITUATION**

You've just been hired by **Maven Charter Schools**, an up-and-coming private education institution. They have a wealth of public and private school data, but need help cleaning and transforming it in order to expose meaningful patterns and insights.

THE **BRIEF**

Maven Charter Schools would like you to **examine, clean, shape, combine and share** competitive education data from the Massachusetts education market.

All you've been given is a folder of excel/csv files containing information about **teacher pay and performance**, student **SAT scores**, **pupil expenditures**, and **graduation rates** by school and district.

THE **OBJECTIVE**

Use Tableau Prep to:

- Connect to multiple data sources
- Examine and filter your data
- Clean and shape fields
- Combine and aggregate data
- Share and update curated data sources



Setting Expectations

1 This course is designed to get you up & running with Tableau Prep

- *Our goal is to provide a deep foundational understanding of Tableau Prep Builder; we won't cover advanced topics like R/Python or Tableau Prep Server integration in depth*

2 What you see on your screen **may not always match mine**

- *Tableau Prep updates on a **monthly** basis for minor releases and **quarterly/yearly** for major releases, so features and functionality may change over time*

3 This course is primarily geared towards **data cleansing and preparation**

- *Data visualization is another key component of the analytics and business intelligence workflow, which we cover in depth in separate courses (**Tableau Desktop for Beginners** and **Advanced Tableau Desktop**)*

4 We will not cover **Tableau Prep Conductor** as part of this course

- *This course will focus on **Tableau Prep Builder** specifically; online flow automation and collaboration features will be reviewed at a high-level only*

Introducing Tableau Prep

Meet Tableau Prep

Tableau Prep is a self-service data preparation tool, providing users with visual and intuitive tools to **combine**, **shape**, and **clean** raw data for analysis



The screenshot shows the Tableau Prep interface with a data flow diagram on the left and a preview of the output data on the right.

Data Flow Diagram:

- Four data sources (Orders (East), Orders (West), Orders (Central), Orders (South)) are connected to a central node labeled "All Orders".
- "All Orders" is connected to "Orders + Returns".
- "Orders + Returns" is connected to "Split Customer".
- "Split Customer" is connected to "Create 'All Order...'".
- "All Orders" is also connected to "Returns (All)".
- "Returns (All)" is connected to a node with a red bar.
- "Quota" is connected to "Pivot Quotas".
- "Pivot Quotas" is connected to "Quota + Orders".
- "Quota + Orders" is connected to "Create 'Annual ...'".
- "Quota + Orders" is also connected to an "Aggregate" node, which then connects to "Create 'Annual ...'".

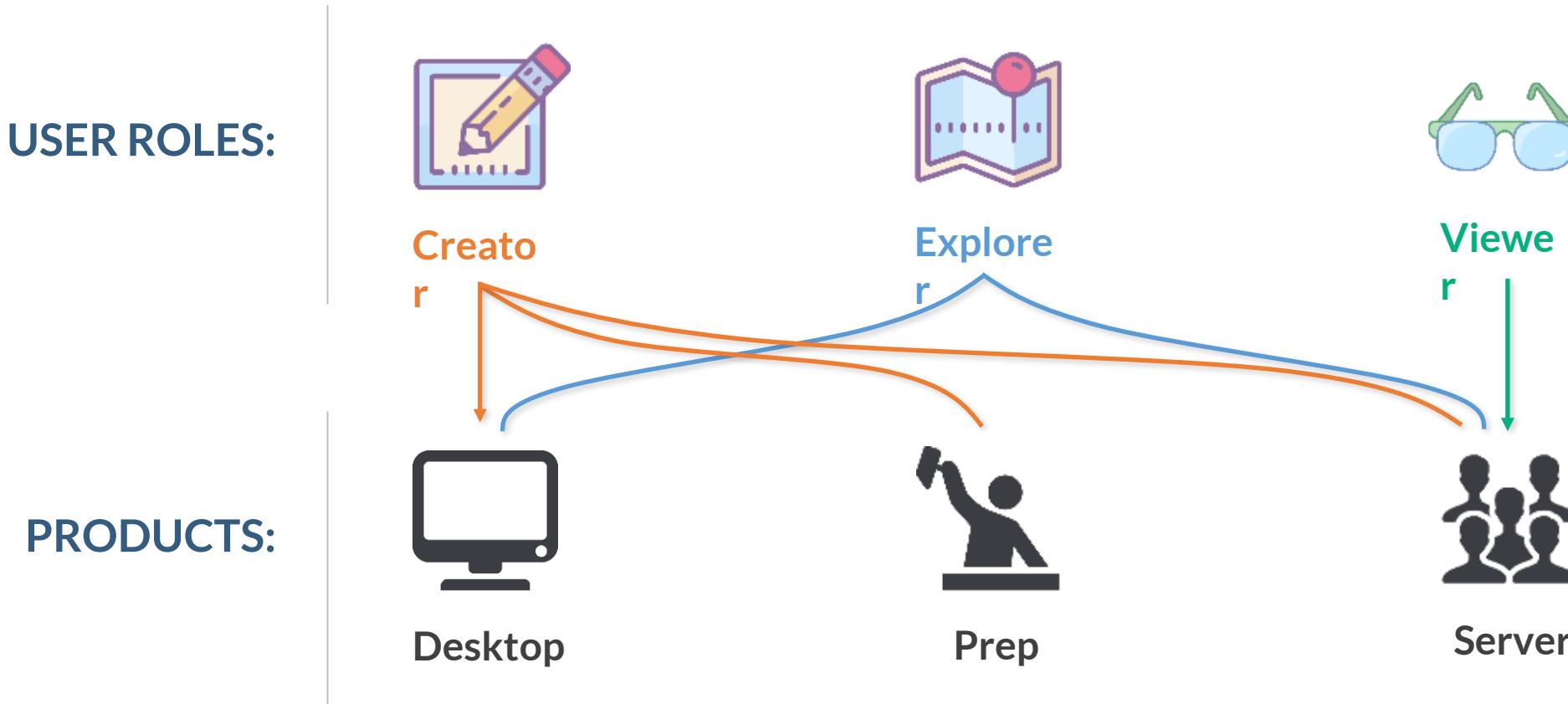
Output Preview:

The preview shows a table titled "Create 'Annual Performance' 9 Fields" with the following data:

Discount	Year of Sale	Quota	Year	Region-1	Region	Profit	Quantity	Sales
0.2649356223176	2,015	100,000	2,015	Central	Central	539.5533999999997	1,726	103,838.16459999986
0.15949119373777	2,015	125,000	2,015	East	East	16,644.189499999982	1,962	128,092.69
0.12048710601719	2,015	100,000	2,015	South	South	11,879.12000000003	1,336	103,845.8435
0	2,015	115,000	2,015	West	West	79,148.82230000001	7,029	433,226.42400000008
0.24063025210084	2,016	105,000	2,016	Central	Central	11,716.802000000014	1,815	102,874.22199999997

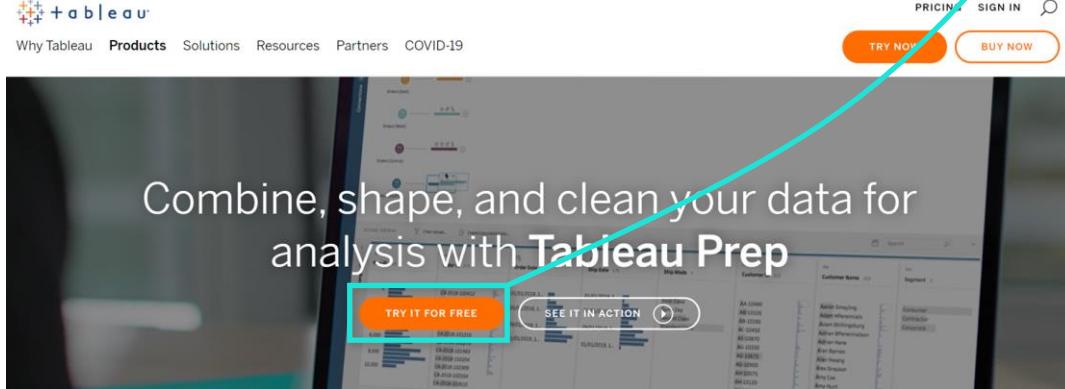
Tableau User Roles

Tableau Prep is included as part of the **Tableau Creator** role, which includes Tableau Prep Builder, Tableau Desktop, and one license of Tableau Server or Tableau Online

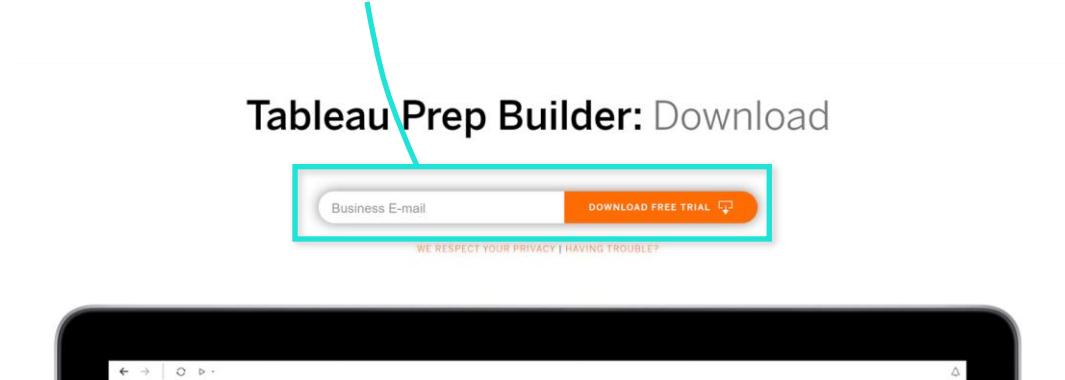


Downloading Tableau Prep [Trial/Paid]

1) Go to [tableau.com/products/prep](https://www.tableau.com/products/prep) and click



2) Enter a **business email** to start a **14-day free trial**



By downloading a trial, you'll get **14 days free** before starting a paid monthly subscription

- If you start a paid subscription, we recommend the **Tableau Creator [For Individuals]** option

Tableau Prep Builder

Registration

Please complete all fields for the registered user

First name

Last name

Business email

Organization

Department

Job Role

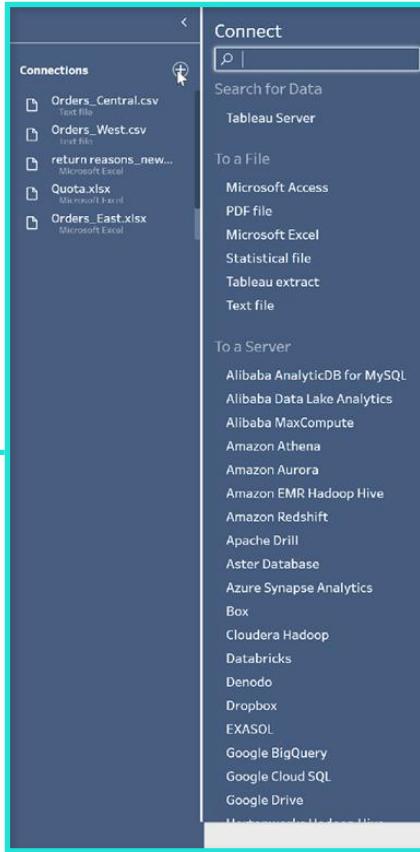
Country/Region

United States

Tableau Prep Workspace

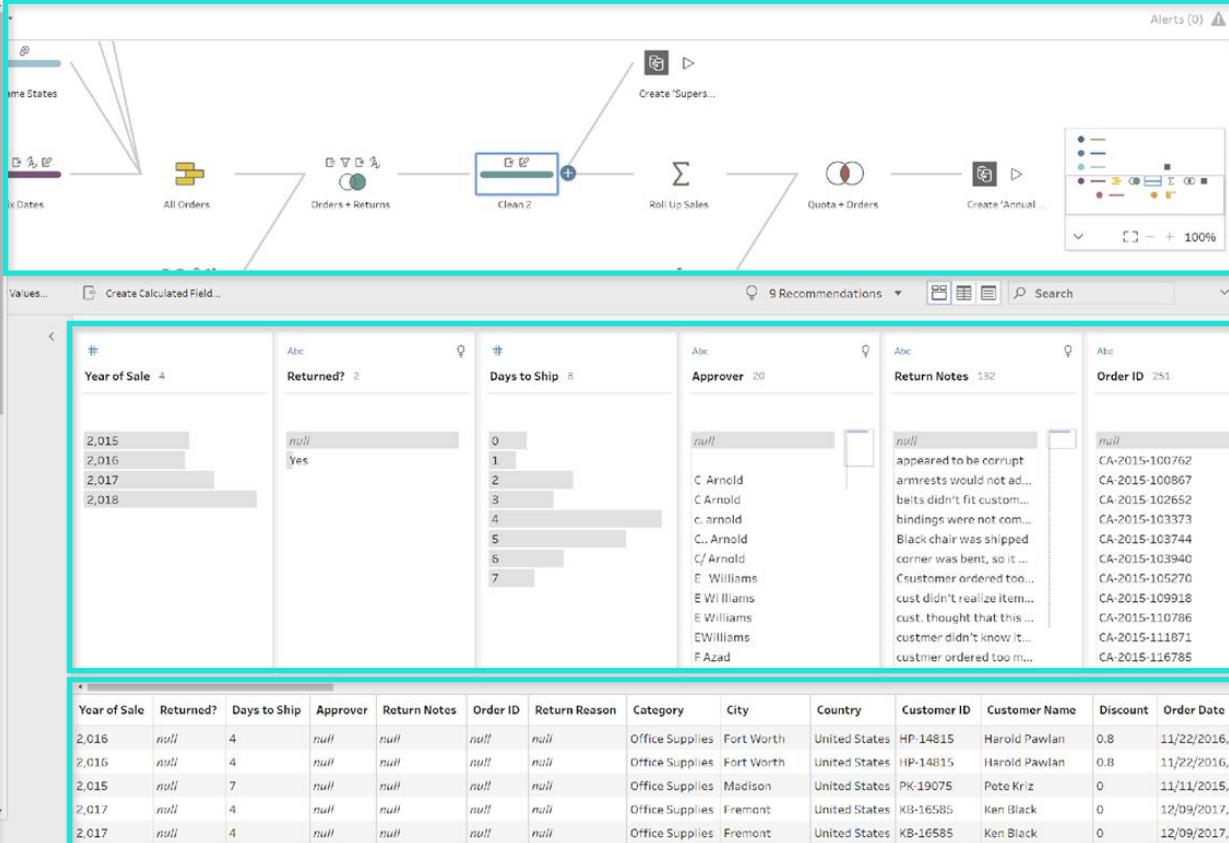
Connections Pane

Connect to local, server, or published data sources



Flow Pane

A visual representation of each operation or step in the data preparation process



Data Grid

Displays a preview of the rows and columns in your source data

Profile Pane

Displays a summary of each field in your data sample

Visual Dictionary

Tableau Prep uses visual indicators to represent **steps**, **field types**, and **notifications** within a flow; familiarizing yourself with these indicators will help you interpret exactly how a flow functions

Input Steps

Icons in flow pane shows data source type

	Data Source
	Data Source with Wildcard Union
	Excel
	Excel with Wildcard Union
	CSV
	CSV with Wildcard Union
	Tableau Extract

Clean Steps, Changes Pane & Toolbars

Icons track changes made to data

	Calculated Field
	Change Data Type
	Edit Value
	Exclude Values
	Filter Values
	Group Values
	Keep Only
	Hide Profile Pane
	Show Profile Pane
	Merge Fields
	Remove Fields
	Rename Field
	Search
	Split Fields

Join Steps

Icons define join types between data sources

	Full Anti Join
	Inner Join
	Left Inner Join
	Left Outer Join
	Full Outer Join
	Right Inner Join
	Right Outer Join

Aggregate Steps

Aggregated data shown with Sigma icon

	Aggregate Data
---	----------------

Pivot Steps

Icon represents data pivoting columns to rows

	Pivot Data
---	------------

Union Steps

Icon shows where data sources are combined

	Union Data
---	------------

Visual Dictionary

Tableau Prep uses visual indicators to represent **steps**, **field types**, and **notifications** within a flow; familiarizing yourself with these indicators will help you interpret exactly how a flow functions

Profile Cards

Icons identify data types and field transformations

 Calculated Field	 Rename Field
 Change Data Type	 Search
 Edit Value	 Split Fields
 Exclude Values	 Boolean Data Type
 Filter Values	 Date Data Type
 Group Values	 Date Time Data Type
 Keep Only	 Numeric Data Type
 Merge Fields	 Text Data Type
 Remove Field	

Output Steps

Icons identify data output types and running flow

 CSV File
 Published Data Source
 Local Tableau Data Extract
 Run Flow

Profile Pane

Summary of row count and data sample indicator

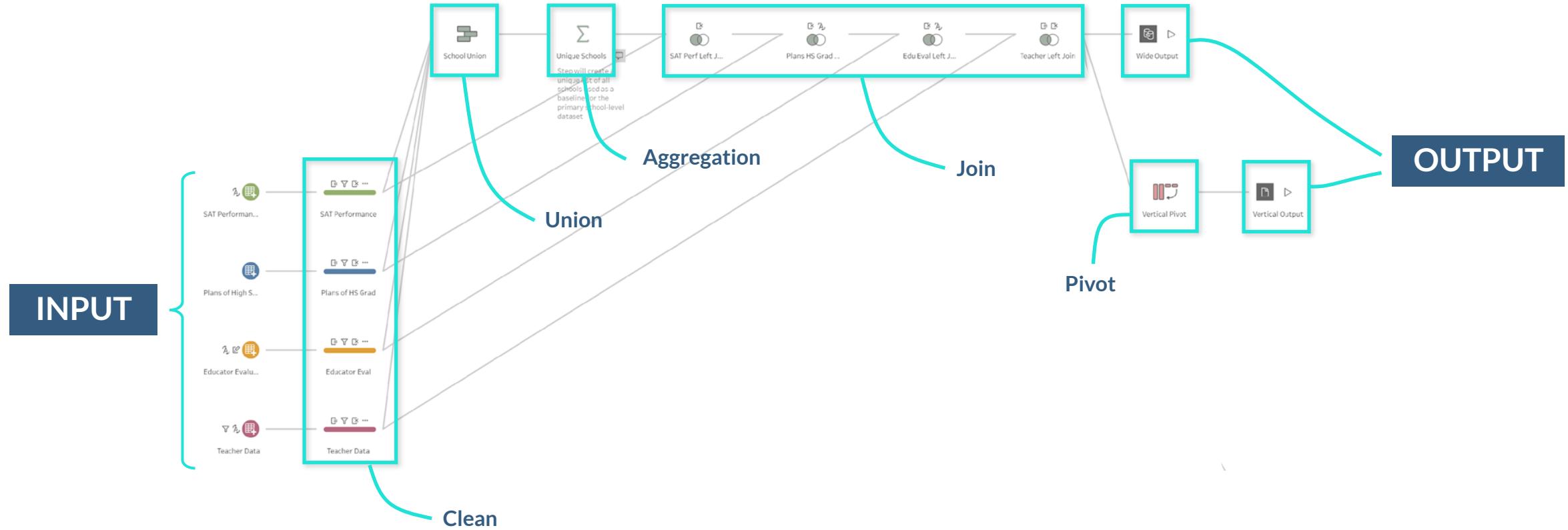
 Sampled	Shows when data is sampled
 Hover to show exact row count	Hover to show exact row count

Notifications

Identify problems, errors or alerts

 No Notifications
 Notification Alert
 Error in the Step

Example Flow Diagram



PRO TIP: Data Design

It's important to think about **data design** *before* you begin to clean or transform your data, as design needs will vary based on your **audience**, **use case**, and **performance needs**



Who is the *end user* or *audience* consuming the data?

- *Is the data to be utilized by analysts, managers or executives? How many users need access?*



What *purpose* or *use case* is the data designed to support?

- *Is the data intended for ad-hoc data pulls, deep dive dashboards, or executive-level KPI reporting?*

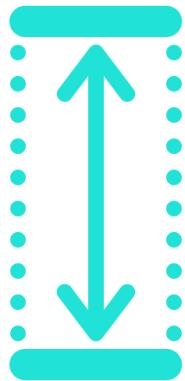


Are there *speed* or *performance* implications to consider?

- *What are the expectations regarding query performance, refresh frequency, and data depth?*

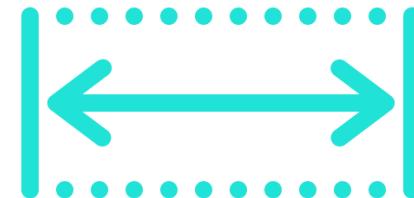
PRO TIP: Data Design

Vertical Views



- Row-heavy data which is the **most flexible** structure for Tableau Desktop
- Ideal combo of **good performance & dynamic aggregation**
- Commonly used with **transactional data**

Wide Views



- Highly **dimensional data with many columns**
- Allows for **deep analysis** and many “cuts” of data
- Most common with **survey data** and **unique record data sets**

Aggregated Views



- **Highly aggregated** and curated views for best performance
- Ideal for **executive-level visualizations** and specific high-level use cases

Connecting to Data

Connecting to Data

Tableau Prep enables users to **connect**, **clean**, and **configure** raw data from virtually any source

Connect



- **Connect** to local files, databases or published sources
- **Enhance** connections with wildcard unions, SQL and more

Clean



- **Clean** your data upfront with tools like data interpreter
- **Filter** initial data down before the main flow

Configure



- **Configure** field names, data types, text settings, etc.
- Choose which fields to **include** or **exclude** from the flow

Data Connection Types

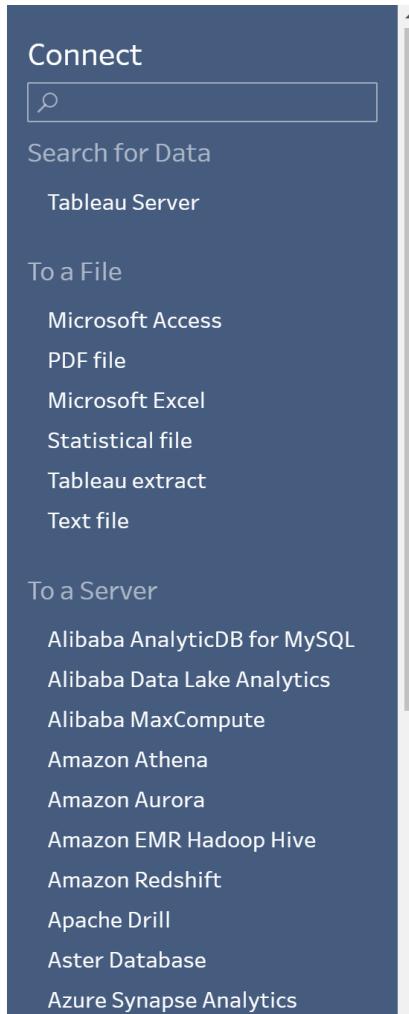


Tableau Prep Builder enables users to connect to many data sources and platforms, including:

- **Flat Files (xlsx, csv, access, pdf, .hyper, etc.)**
- **Servers (SQL Server, Salesforce, Hadoop, Snowflake, Postgres, etc.)**
- **Published Data Sources (Tableau Server / Online Sources)**

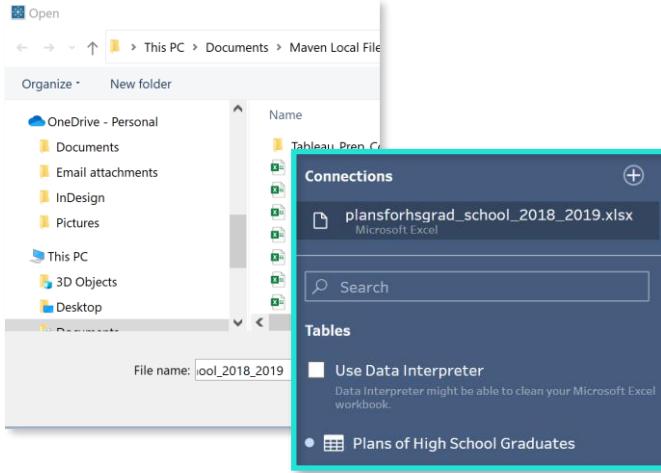
The interface shows a list of published data sources:

Search for Data				
Content Type:	Databases and Files (21)			Search for data
①	Type	↑ Name	Workbooks	Server / Path
	Table	postgres	0	localhost



PRO TIP: Don't have local credentials? Leverage published Tableau Server connections as data inputs!

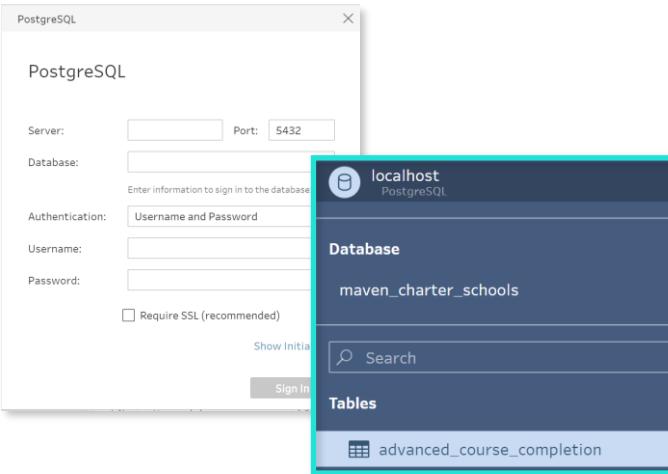
Data Connection Examples



Local Files

When you connect to local flat files, prep builder will show **tabs** for excel files and can union files within a given directory

NOTE: Data Interpreter is available for text/csv files



Databases

When you connect to a **database**, you must enter credentials in order to access the schemas, tables and views available

NOTE: Data Interpreter is NOT available for database connections

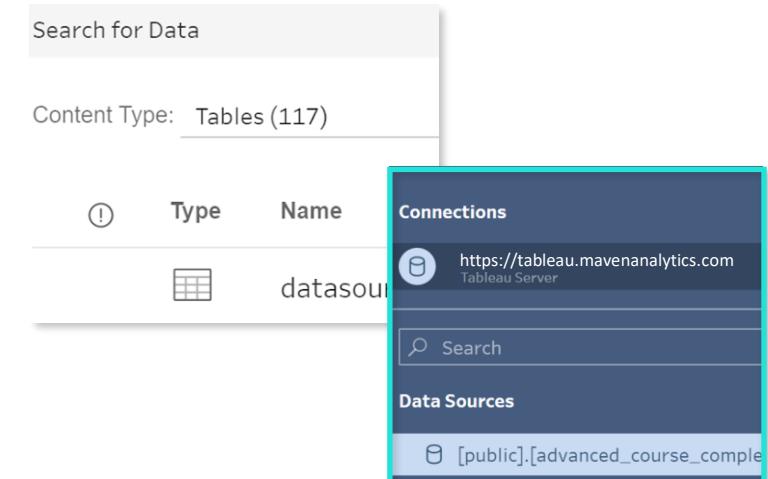


Tableau Server

When you connect to **tableau server**, enter your server credentials to view all published data sources, tables, and files

NOTE: Data Interpreter is NOT available for tableau server connections

Wildcard Unions

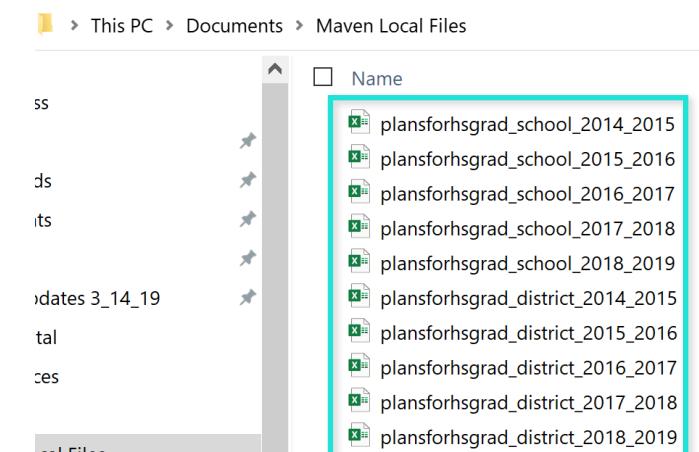
Wildcard unions allow you to combine files or tables within a folder or directory at the input stage

Search In
Select the directory/schema to use to find files/tables for the union

Include Subfolders
Includes files contained in subdirectories of the parent folder

Files, Sheets & Tables
Include or exclude files, sheets or tables using these dropdowns

Included Files & Tables
Previews the files or tables matched based on the wildcard settings



Matching Pattern
Includes only files, sheets or tables which contain specific text (*), or leave blank to union all files



PRO TIP: CSVs union automatically in the same directory, as well as sheets in Excel workbooks!

PRO TIP: Input Joins

Joins can also be created at the input stage for certain database connections; if table relationships are present, **Linked Keys** will be available to specify which fields to use for the join

The screenshot illustrates the process of creating an input join between two tables: 'sat_performance' and 'advanced_course_completion'.

Left Panel (Database View): Shows the 'localhost PostgreSQL' database 'maven_charter_schools' with tables 'advanced_course_completion' and 'sat_performance'.

Middle Panel (Input Stage): The 'sat_performance' table is selected. The 'Input' tab is active, showing the 'Multiple Tables' section. A 'Related Tables' dropdown menu is open, listing 'advanced_course_completion' with a 'Related fields' icon. A callout box highlights this icon with the text 'Unique Identifier (Primary Key)'.

Top Center (Linked Keys): A callout box highlights the 'Unique Identifier (Primary Key)' icon, which is also present in the 'Related Fields (Foreign Key)' and 'Unique and Related Fields' sections.

Bottom Center (Table Relationships): A callout box highlights the 'Related fields' and 'Unique identifier' sections, which are connected by a dashed line to the 'Related Tables' dropdown in the input stage.

Right Panel (Join Preview): Shows the 'Join 1' configuration. It displays the join type as 'inner', the join clause as 'sat_performance School_Code = advanced_course_c... School_Code', and the summary of join results showing 377 included and 0 excluded values for 'sat_perfo...', and 377 included and 24 excluded values for 'advanced...'.

Input Cleaning

Change Data Type

Click type to change from given drop-down menu

	Type	Field Name	Original Field Name	Count
1	Abc	School Name	School Name	13
2	#	School Code	School Code	13
3	#	4 Year Private Co...	4 Year Private College	13
4	#	4 Year Public Coll...	4 Year Public College	13
5	#	2 Year Private Co...	2 Year Private College	13
6	#	2 Year Public Coll...	2 Year Public College	13
7	#	Other Post Seco...	Other Post Secondary	13
8	#	Apprenticeship	Apprenticeship	13
9	#	Work	Work	13
10	#	Military	Military	13
11	#	Other	Other	13
12	#	Unknown	Unknown	13
13	#	Total Count	Total Count	13

Remove Field

Uncheck fields to remove them from the flow

Filter Values

Click Filter icon and create filter from calculation window

Calculation: ABS(number)
Must return a boolean value
Reference: All

Calculation is valid: 13.4, 19.2, 10.5

ABS(number)
Returns the absolute value of the given number.
Example: ABS(-7) = 7

Search: ABS

Calculation is valid: 0, 1.9, 10.5

0, 2.9

1.2, 0

2.4, 1, 0

0, 5.8

1.2, 1, 0

82, 104, 38

Apply Save

Rename Field

Double-click fields to enter a new field name

Text Configuration

Text files require additional configuration in the **Settings** tab to determine how they will be ingested

Input

Settings Multiple Files Data Sample

Connection
Text file
TeacherSalaries_District_2017_2018.csv Edit
Original Table Name: TeacherSalaries_District_2017_2018

Text Options
 First line contains header
 Generate field names automatically

Field Separator
Comma

Text Qualifier
Automatic

Character Set
UTF-8

Locale
English (United States)

Incremental Refresh
Specify the input and output fields to use to identify new rows.
 Enable incremental refresh

First Line Contains Header is the default, and pulls the first row as headers

Generate Field Names Automatically will generate generic headers (F1, F2, etc.)

Field Separator gives a character dropdown to choose a field delimiter

- NOTE:** Choosing “Other” will allow for a custom delimiter

Text Qualifier selects the character that encloses the values in a file

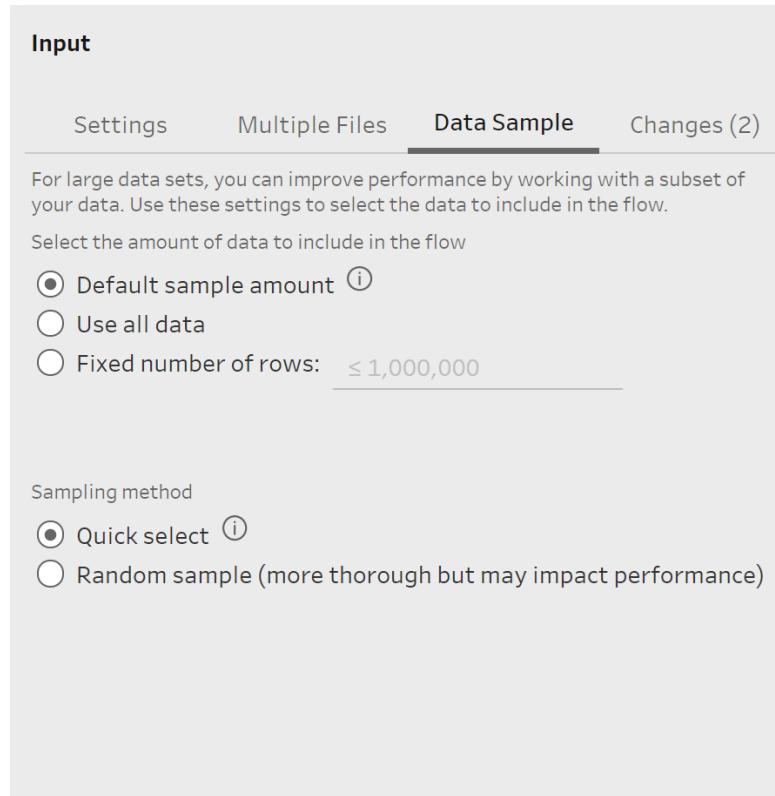
- NOTE:** This defaults to automatic and gives ‘, “, and “none” as options

Character Set selects the character set that describes the file encoding (UTF-8, etc.)

Locale sets the geographic location to parse the file (important for dates, currency, decimals/thousands separators, etc.)

Data Sampling

To optimize performance, Tableau Prep **samples large data sets** and returns a subset of records



Default sample amount: Prep Builder determines number of rows to return

Use all data: Retrieves all rows regardless of size (*can cause performance issues*)

- **NOTE:** Data will still limit to **1 million rows** (Aggregate/Union) and **3 million** (Join/Pivot)

Fixed number of rows: Select custom number of rows (*recommended <1 million*)

Quick select (default): Sample is returned as quickly as possible, using N number of rows or cached data available from a prior query

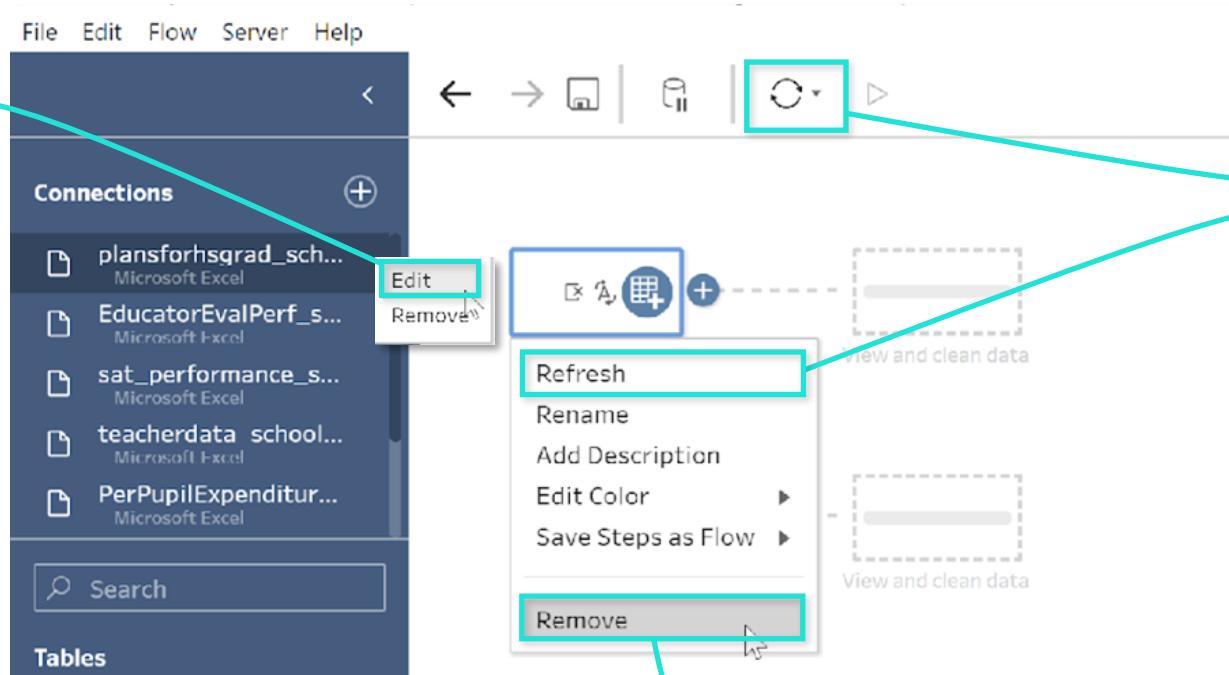
Random sample: Returns the number of rows requested, but looks at all records and returns a representative sample (*may impact performance prior to cache*)

Refreshing Data

If data changes while building a flow, you can **refresh** during the input stage using several methods:

OPTION 2: Edit Connection

*Edit the data connection
and return to the flow*



OPTION 1: Refresh

For File Inputs, refresh using the refresh icon or the input step

OPTION 3: Remove & Re-Add

Completely remove the input step, re-connect, and drag the table back into the flow

HOMEWORK: Connecting to Data

THE **SITUATION**

Happy Hipsters, a lifestyle apparel company, wants to analyze **World Happiness** data to support an upcoming marketing campaign, and has enlisted your help

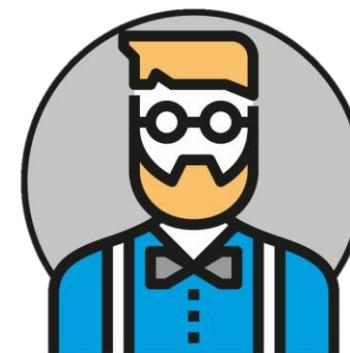
THE **BRIEF**

The Happy Hipsters team has asked you to help **clean and consolidate their raw data into a single source**, which will enable them to explore and analyze key global happiness metrics for their new campaign

THE **OBJECTIVE**

Use Tableau Prep to:

- Connect to source data
- Use a wildcard union to combine files
- Clean data upon input
- Configure and refresh data sources



**Happy
Hipsters**

Examining & Filtering

Examining & Filtering

After connecting to sources, users can **examine & filter** data using Tableau Prep's visual interface; it's important to conduct these steps before making any major changes to your data in the flow!

Examine



- *Profile* your data by looking at field **value distributions**
- Review **data types**, **data size**, and find specific **fields or values**
- *Sort & Highlight* values in your fields to find gaps or deficiencies

Filter



- *Reduce* the data being pulled, using various **filtering tools**
- *Organize* your flow's tools and settings for optimal **performance** and clear documentation

Data Types & Sizes

One of the first steps in evaluating data is to examine **data size**, **field types** and **unique values**; this can be done at several stages, but the simplest approach is to add a **clean step**

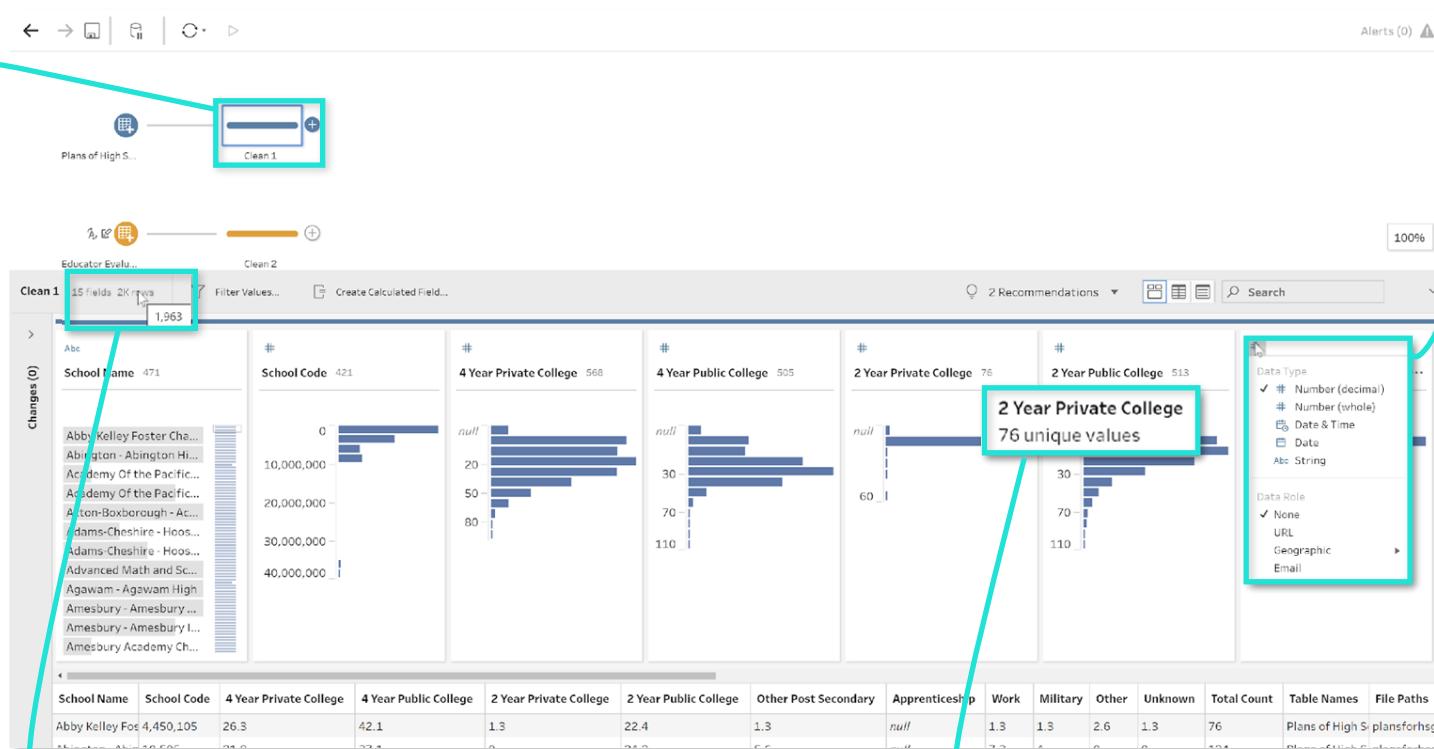
Adding a Clean Step

Clean Steps can be added to a flow in two distinct ways:

- Automatic
- Manual (+)

Automatic: Select gray outline

Manual (+): Insert: + Clean Step



Field Data Type

Data types can be modified by selecting from the header

Data Size

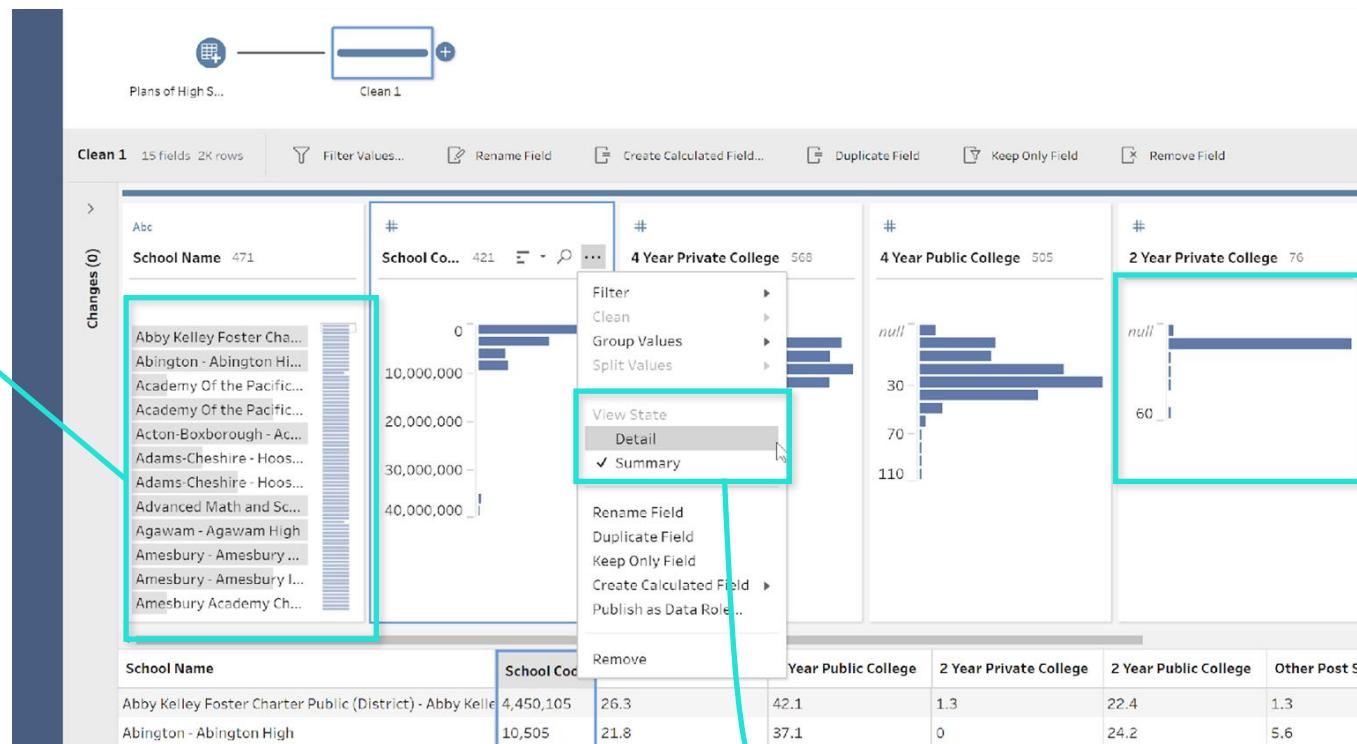
Shows the **number of fields** and **row count** (hover to see exact count)

Unique Values

Displays the **distinct values** in each field

Value Distribution

The profile pane allows you to visualize the **distribution** of your data, by plotting the frequency of each distinct value as **bins** in a histogram; this is a great way to identify outliers and null values!



Finding Fields & Values

Use the toolbar search or field search options to find specific fields or values in your data

The screenshot shows the Alteryx workspace interface. At the top, there are navigation icons (back, forward, search, etc.) and an 'Alerts (0)' indicator. Below the toolbar, a search bar contains the text 'school'. A magnifying glass icon is positioned next to the search bar. A callout bubble labeled 'Search for Values' points to this icon. A large callout bubble labeled 'Search for Fields' points to the search bar. The main workspace shows a data table with columns 'School Name' and 'School Code'. The 'School Name' column contains a list of school names, and the 'School Code' column shows their corresponding codes. A 'Search Options' panel is open, showing radio buttons for 'Contains' (selected), 'Starts with', 'Ends with', 'Exact match', and 'Does not contain'. Below this are buttons for 'Create Filter from Search' (Keep Only and Exclude).

Search for Values
Search for values using various match options (contains, starts with, exact match, etc.) or click (...) for advanced options or to filter found values

Search for Fields

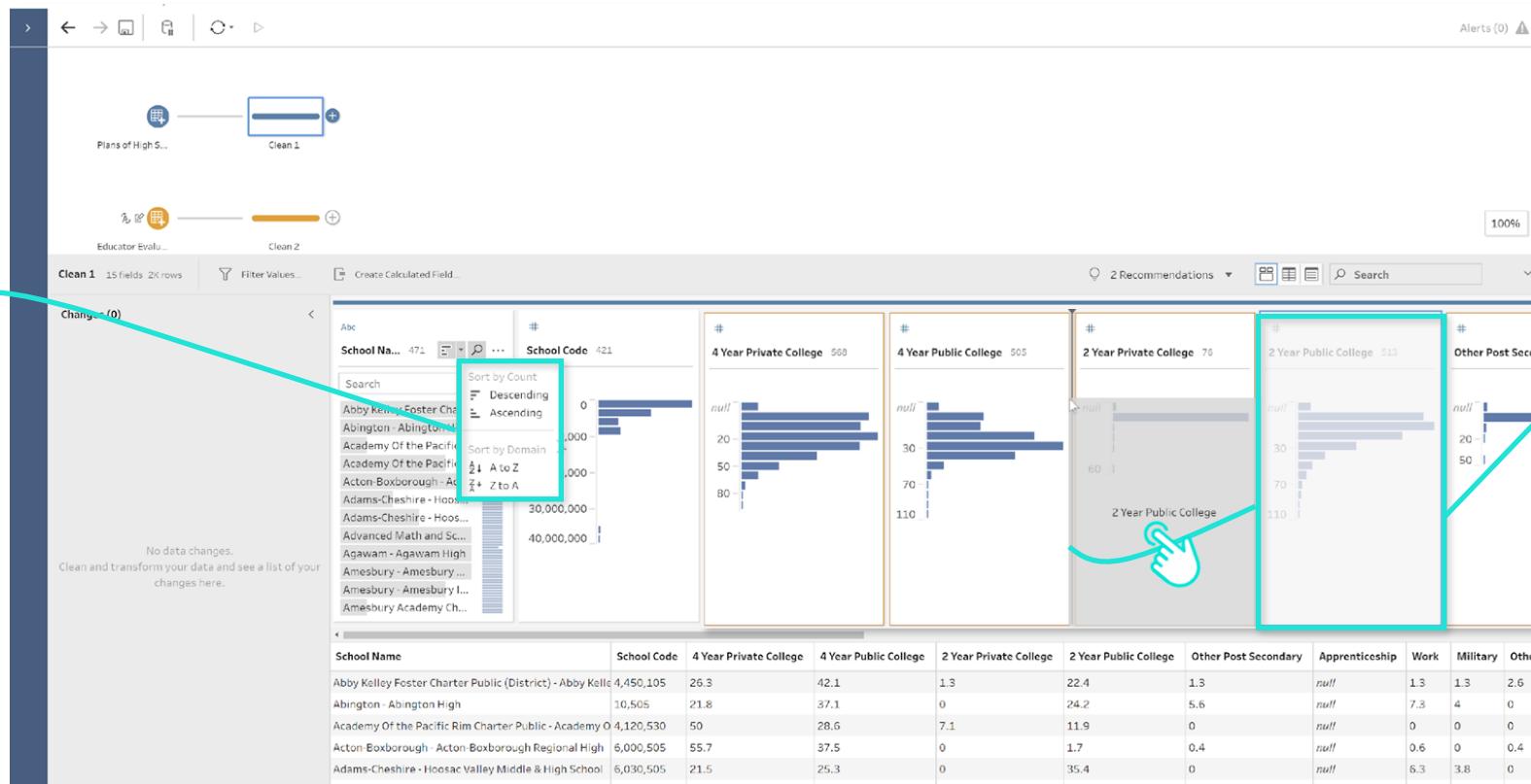
Enter a full or partial search to return matching fields

Sorting & Moving Profile Cards

Within the profile pane, you can **sort bins** by either frequency or alphabetical order (ascending or descending), or click to drag and rearrange profile cards

Sort Bins & Fields

Sort by **count** (frequency) or **domain** (alphabetical)



Move Cards

Reorganize profile cards by dragging until a black line appears

Highlighting

Highlighting is a quick way to trace fields back through flow steps, see related values across fields, and pinpoint identical values in your data

Trace Fields

Select a field to trace where it was used or modified within your flow

Related Values

Highlight related values by selecting a value/bin in the profile pane

Identical Values

Select a value in the data grid to highlight all identical values

School Name	School Code	Tests Taken	Reading / Writing	Math	Reading	Writing	Table Names	File Paths
Belmont - Belmont High	00260505	271	null	614	590	581	SAT Performance Report	sat_performance_school_2014_2015.xls
Brookline - Brookline High	00460505	420	null	616	589	581	SAT Performance Report	sat_performance_school_2014_2015.xls
Lincoln-Sudbury - Lincoln-Sudbury Regional High	06950505	346	null	622	590	586	SAT Performance Report	sat_performance_school_2014_2015.xls
Newton - Newton North High	02070505	383	null	604	587	588	SAT Performance Report	sat_performance_school_2014_2015.xls
Quabbin - IB School of Quabbin	07530515	11	null	600	613	597	SAT Performance Report	sat_performance_school_2014_2015.xls
Sharon - Sharon High	02660505	243	null	603	583	579	SAT Performance Report	sat_performance_school_2014_2015.xls

Filtering Methods

There are several **filtering** methods in Tableau Prep, based on the field type and step chosen:

Keep or Exclude

Keeps or removes selected value or field (available for all field types; String, Number, Date, Date Time, etc.)

Calculation Filter

Filters values based on calculated field condition (available for all field types)

Selected Values Filter

Chooses values to keep or exclude even if they aren't in the data source (available for all field types)

Range of Values Filter

Filters by minimum and maximum value parameters (available for Number field type)

Range of Dates Filter

Filters by minimum and maximum date value parameters (available for Date and Date Time field types)

Wildcard Match Filter

Filters by partial or whole matching text (available for String field type)

Null Values Filter

Keeps only Null or Non-Null Values (available for all field types)

Filtering Methods



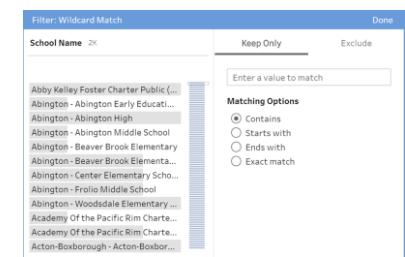
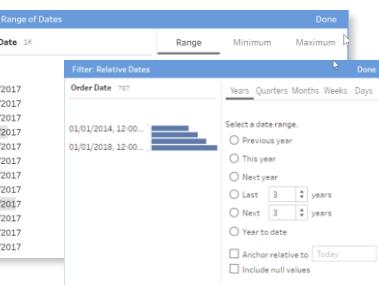
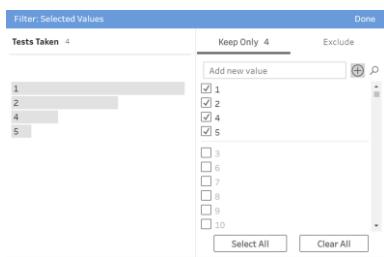
Calculation Filter

Condition must be **Boolean**
(only filter available in steps other than clean step)

Keep Only
Exclude

Keep Only/Exclude

Single or multi-select values from the profile card to keep or exclude



Selected Values

Manually select values to keep/exclude (keyed values can be added even if not in data)

Range of Values

Filter numeric values within a specified lower/upper limit

Range/Relative Dates

Range of dates (upper/lower) or time period relative to today or an anchor date

Null Values

Filter to only null or non-null values

Wildcard Match

Keep/exclude values based on a pattern (filter results display on left pane)

HOMEWORK: Examining & Filtering

THE SITUATION

Your brother-in-law Sai just started his first business venture: a food truck specializing in Indian desserts called **Bengali Sweet Treats**. As the family's resident data nerd, you've been enlisted to help him analyze popular Indian dishes to help him perfect his menu.

THE BRIEF

Sai needs you to examine a spreadsheet containing hundreds of Indian dishes, and profile their ingredients, prep time, regional origin, and flavor profile.

You'll need to **connect**, **profile**, and **filter** the data to give Sai some ideas for his award-winning food truck!

THE OBJECTIVE

Use Tableau Prep to:

- Examine data types and sizes
- Profile value distribution across fields
- Sort, move and highlight relevant data
- Filter values to pinpoint key records

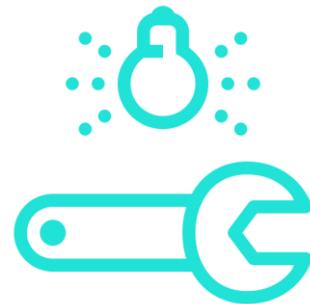


Operations & Calculations

Operations & Calculations

Tableau Prep includes a range of tools for cleaning and transforming data, including **value & field operations** (*grouping, cleaning, converting, splitting, etc.*) and **calculations** (*analytic, logical, LOD, etc.*)

Value & Field Operations



- *Clean & transform data using a range of value and field operations (group, filter, split, etc.)*
- *NOTE: Cleaning steps can be performed in multiple flow steps (except output)*

Calculated Fields



- *Perform logical, string, aggregate or level of detail calculations to create new fields*
- *Apply analytic functions (i.e. rank) across tables or partitions*

Value & Field Operations

Common value & field operations fall into three main categories based on the scope of impact (*records, fields and values*) and can be accessed from multiple flow steps

	Operation	Input	Clean	Aggregate	Pivot	Join	Union
Records	Filter	✓	✓	✓	✓	✓	✓
	Keep / Remove Field	✓	✓	✓	✓	✓	✓
	Rename Field	✓	✓	✓	✓	✓	✓
	Duplicate Field		✓		✓	✓	✓
	Calculated Field		✓		✓	✓	✓
	Clean		✓		✓	✓	✓
	Convert Dates		✓	✓	✓	✓	✓
	Edit Values		✓		✓	✓	✓
	Group Values		✓		✓		✓
	Split Values		✓		✓	✓	✓
Fields	Change Data Type	✓	✓	✓	✓	✓	✓
	Filter						
	Keep / Remove Field						
	Rename Field						
	Duplicate Field						
	Calculated Field						
	Clean						
	Convert Dates						
	Edit Values						
	Group Values						
Values	Split Values						
	Change Data Type						
	Filter						
	Keep / Remove Field						
	Rename Field						
	Duplicate Field						
	Calculated Field						
	Clean						
	Convert Dates						
	Edit Values						

Clean Step Layouts

Cleaning Operations

Accessible via the *profile pane* or drop-down menu

The screenshot shows the 'Clean 1' step in a data pipeline. The top navigation bar includes 'Plans of High S...', 'Clean 1', 'Filter Values...', 'Automatic Split...', 'Custom Split...', 'Rename Field...', 'Create Calculated Field...', 'Duplicate Field...', and a '...' button. The main area displays a list of schools with their names and codes, and a histogram for '2 Year Private College' counts. A context menu is open over the 'School Name' column, with 'Clean' selected. The 'Clean' submenu includes 'Filter', 'Clean', 'Group Values', 'Split Values', 'View State' (Detail, Summary), 'Rename Field', 'Duplicate Field', 'Keep Only Field', 'Create Calculated Field', and 'Publish as Data Role...'. The bottom section shows a data grid with columns: School Name, School Code, 2 Year Private College, 2 Year Public College, 4 Year Private College, 4 Year Public College, Other Post Secondary, Apprenticeship, Work, Military, Other, and Unknown.

School Name	School Code	2 Year Private College	2 Year Public College	4 Year Private College	4 Year Public College	Other Post Secondary	Apprenticeship	Work	Military	Other	Unknown
Abby Kelley Foster Charter Public (District) - Abby Kel...	4,450,105	1.3	22.4	31.6	31.6	5.3	null	1.3	1.3	5.3	0
Abington - Abington High	10,505	2.7	19.1	32.7	38.2	1.8	null	3.6	0	1.8	0
Academy Of the Pacific Rim Charter Public (District) - A...	4,120,530	0	16.7	46.3	37	0	null	0	0	0	0
Acton-Boxborough - Acton-Boxborough Regional High	6,000,505	0	3.8	49.7	41.4	0.4	null	2.3	0.8	0.8	0.6

Layout Options:

Profile Pane (default)

Shows profile pane + data grid

Data Grid

Shows detailed data view

List View

Shows columns in list form

PRO TIP: Pausing Data Updates

Pause data updates to optimize performance during flow development (NOTE: the view will automatically switch over to *list view* while data updates are paused)

The screenshot shows the Power BI Data Flow interface for a dataset named 'Clean 1'. The top navigation bar includes back, forward, and search icons. A message 'Updates are paused. Click Resume to see your results. [Learn more](#)' is displayed, with a 'Resume' button highlighted with a red box. The main workspace shows two data flows: 'Plans of High S...' and 'Educator Evalu...'. Below the workspace is a toolbar with options: Filter Values..., Rename Field, Create Calculated Field..., Duplicate Field, Keep Only Field, Remove Field, and a grid icon. A callout bubble points to the 'Resume' button in the message bar with the text 'Pause/Resume Updates'.

Pause/Resume Updates

Options to pause or resume updates

Changes (1)

Type	Field Name	Changes
Abc	School Name	...
#	School Code	Filter Clean Group Values Split Values
#	2 Year Private College	
#	2 Year Public College	
#	4 Year Private College	
#	4 Year Public College	
#	Other Post Secondary	
#	Apprenticeship	
#	Work	
#	Military	
#	Other	
#	Unknown	

Custom Calculation
Fixed LOD
Rank

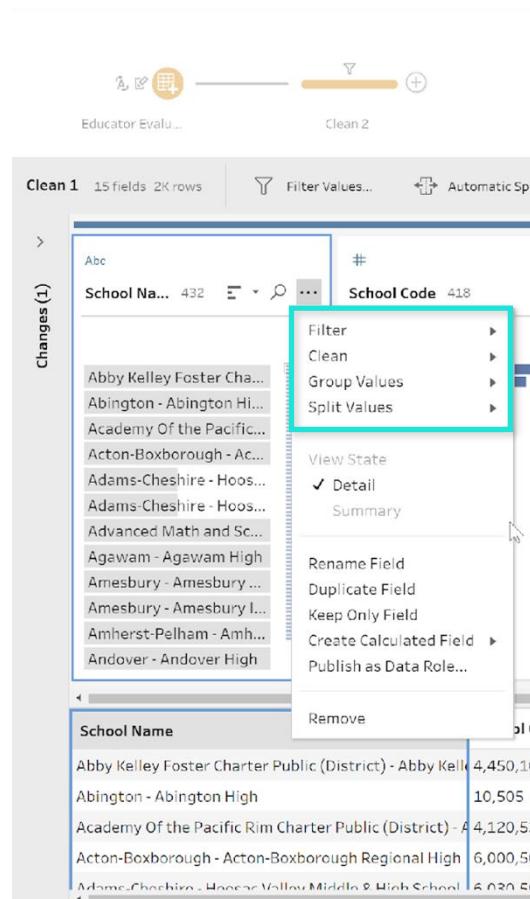
Updates are paused. Click Resume to use this feature.
[Learn more](#)

Limited Features

Features which require visual representation of values (splitting, filtering, grouping, etc.) are disabled while updates are paused

Value Operations

Value operations can be used to filter, clean, group or split values inside fields



The screenshot shows the Tableau Prep interface with the 'Clean 1' step selected. A context menu is open over the 'School Name' field, with the 'School Code' field highlighted. The menu options include 'Filter', 'Clean', 'Group Values', and 'Split Values', which are highlighted with a red box. Other options like 'View State', 'Detail', and 'Summary' are also visible.

Filter allows you to reduce the number of records using various filter criteria

Clean provides a list of quick cleaning operations which apply to all values in the field

Group Values replaces individual or multiple values with new a group value

Split Values parses values using an automatically detected or custom-defined delimiter



PRO TIP: Use Tableau Prep's recommendations (light bulb) to automatically clean your data

Value Operations | Clean

Use **cleaning tools** to change text case, remove specific characters, or trim spaces from strings

The screenshot shows the Power BI Data Flow interface. A data source 'SAT Performance' is connected to a 'Clean 3' step. The 'Clean 3' step is expanded to show two columns: 'School Name' (451 rows) and 'School Code' (408 rows). A context menu is open over the 'School Code' column, with the 'Clean' option selected. A sub-menu is displayed with the following options: Make Uppercase, Make Lowercase, Remove Letters, Remove Numbers, Remove Punctuation, Trim Spaces, Remove Extra Spaces, and Remove All Spaces. A teal bracket on the right side of the slide groups these eight options together.

Make Uppercase changes text case to upper

Make Lowercase changes text case to lower

Remove Letters removes all letter characters from a string

Remove Numbers removes all number characters from a string

Remove Punctuation removes all forms of punctuation

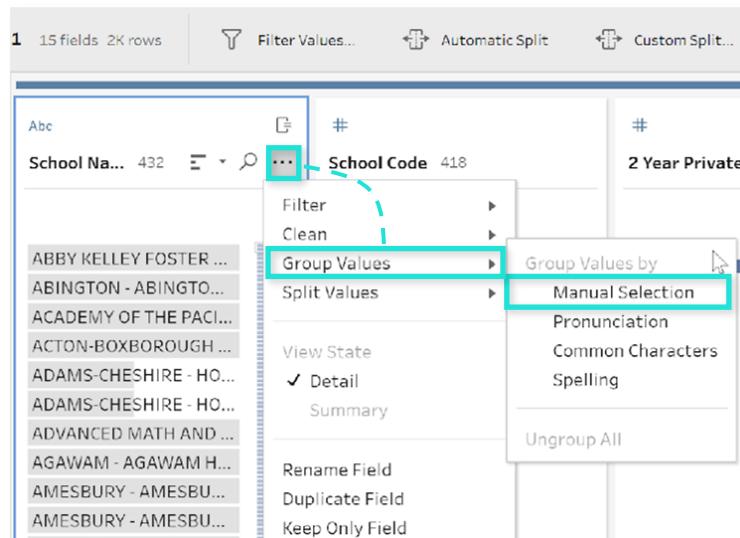
Trim Spaces removes leading or trailing spaces

Remove Extra Spaces removes extra spaces (when >1)

Remove All Spaces removes any spaces contained in the string

Value Operations | Manual Grouping

Manually group text values using multi-select or checkbox selections



1 15 fields 2K rows

Filter Values... Automatic Split... Custom Split...

School Name 432

ABC

ABBY KELLEY FOSTER ...
ABINGTON - ABINGTO...
ACADEMY OF THE PACI...
ACTON-BOXBOROUGH ...
ADAMS-CHESHIRE - HO...
ADAMS-CHESHIRE - HO...
ADVANCED MATH AND ...
AGAWAM - AGAWAM H...
AMESBURY - AMESBU...
AMESBURY - AMESBU...

...

School Code 418

#

2 Year Private C...

...

Group Values

Filter

Clean

Group Values by

Manual Selection

Split Values

Pronunciation

Common Characters

Spelling

View State

Detail

Summary

Ungroup All

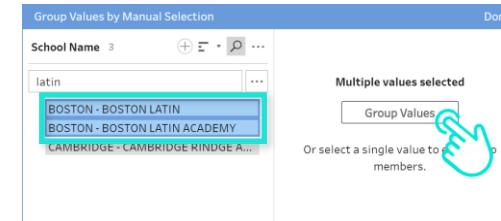
Rename Field

Duplicate Field

Keep Only Field



Manual Grouping



Group Values by Manual Selection

School Name 3

latin

BOSTON - BOSTON LATIN

BOSTON - BOSTON LATIN ACADEMY

CAMBRIDGE - CAMBRIDGE RINDGE A...

Done

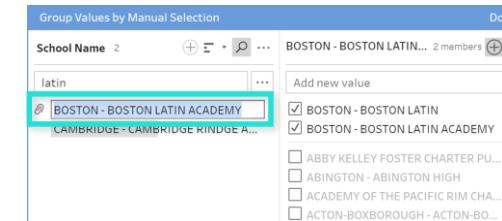
Multiple values selected

Group Values

Or select a single value to add to members.

Multi-Select

Search for a matching string and use Ctrl/Cmd to select values to group



Group Values by Manual Selection

School Name 2

latin

BOSTON - BOSTON LATIN

BOSTON - BOSTON LATIN ACADEMY

CAMBRIDGE - CAMBRIDGE RINDGE A...

Done

Add new value

BOSTON - BOSTON LATIN

BOSTON - BOSTON LATIN ACADEMY

ABBY KELLEY FOSTER CHARTER PU...

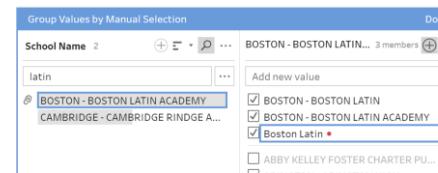
ABINGTON - ABINGTON HIGH

ACADEMY OF THE PACIFIC RIM CHA...

ACTON-BOXBOROUGH - ACTON-BO...

Checkbox Selection

Use checkboxes to add/remove values from a group



Group Values by Manual Selection

School Name 2

latin

BOSTON - BOSTON LATIN ACADEMY

CAMBRIDGE - CAMBRIDGE RINDGE A...

Boston Latin

Done

Add new value

BOSTON - BOSTON LATIN

BOSTON - BOSTON LATIN ACADEMY

Boston Latin

ABBY KELLEY FOSTER CHARTER PU...



PRO TIP: To add new values which do not currently exist in the data set, select an existing group and manually type in the value (shown with a red asterisk)

Value Operations | Automatic Grouping

Automatically group text values using fuzzy matching algorithms based on pronunciation, common characters or spelling

The screenshot shows a data visualization interface with a table of school names. The 'School Code' field is selected, and a context menu is open. The 'Group Values' option is highlighted with a teal box. A large teal bracket on the right side groups three callout boxes that expand the 'Group Values by' options:

- Pronunciation:** Shows a grouping interface for 'District Name 336'. A slider is set to 'Grouping'. Items listed include ACTON, ACTON-BOXBOROUGH, ABBY KELLEY FOSTER CHARTER, ABINGTON, ACADEMY OF THE PACIFIC RIM C..., and ACTON-BOXBOROUGH.
- Common Characters:** Shows a grouping interface for 'District Name 394'. A slider is set to 'Grouping'. Items listed include BOSTON PREPARATORY CHARTER, BOSTON DAY AND EVENING AC..., BOSTON GREEN ACADEMY HOR..., BOSTON PREPARATORY CHART..., BOSTON RENAISSANCE CHART..., and BOSTON.
- Spelling:** Shows a grouping interface for 'District Name 335'. A slider is set to 'Grouping'. Items listed include BRISTOL COUNTY AGRICULTURAL, BRIDGEWATER-RAYNHAM, BRIMFIELD, BRISTOL COUNTY AGRICULTURAL, BRISTOL-PLYMOUTH REGIONAL..., and BROCKTON.

Pronunciation

Find and group values which **sound alike**, and move threshold slider to the left or right to adjust strictness (**left** = fewer groups, **right** = more groups)

Common Characters

Find and group values with letters and/or numbers in common (i.e. "John Smith" and "Smith, John" likely refer to the same person)

Spelling

Find and group values which are spelled alike, and move threshold slider to the left or right to adjust strictness (**left** = fewer groups, **right** = more groups)

Value Operations | Split Values

Split text based values on automatic or custom-defined delimiters

The screenshot illustrates the Power BI Data Editor interface for performing data cleaning and splitting operations. On the left, a table view shows a list of school names. A context menu is open for the 'School Name' column, with the 'Split Values' option highlighted. This menu also includes other options like 'Filter', 'Clean', 'Group Values', 'View State' (set to 'Detail'), and 'Remove'. To the right of the table, a 'Tests' pane displays the count of rows (450) and fields (408). A 'Custom Split' dialog box is open, prompting the user to define how data should be split using a separator (e.g., '-'). The dialog includes fields for 'Split off' (set to 'First') and '1' fields. A preview pane shows the original data and the results of the split operation, with the first row being 'ABBY KELLEY FOSTER ...' and the second row being 'ABINGTON'. A large callout on the right side of the interface highlights the 'Custom Split' dialog and the 'Edit Field' section below it, which shows the formula `TRIM(SPLIT([School Name], "-", 1))` for the 'School Name - Split 1' field.

Automatic Split

Splits values automatically using common delimiters

Custom Split

Define the delimiter and number of columns for the split

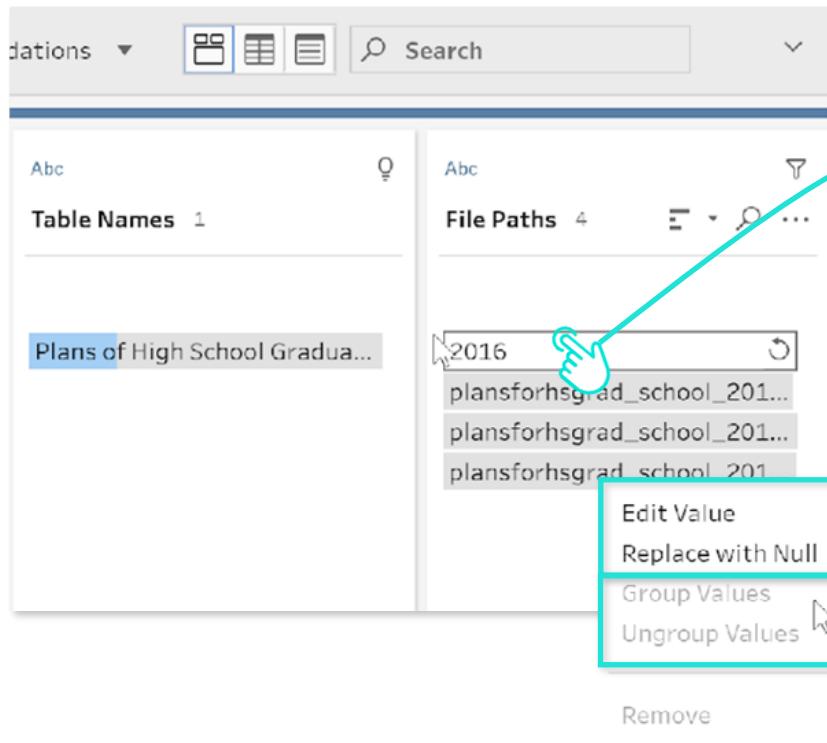
Calculated Split

Split text using a custom calculated field

NOTE: Calculations are automatically generated when either split type (automatic or custom) is performed

Value Operations | Edit

Values can be **edited individually** or as a **group** to correct inaccuracies or standardize variations



Double-Click

Double-click a value in the profile pane to edit it directly (field turns into a group after the first try)

Right-Click

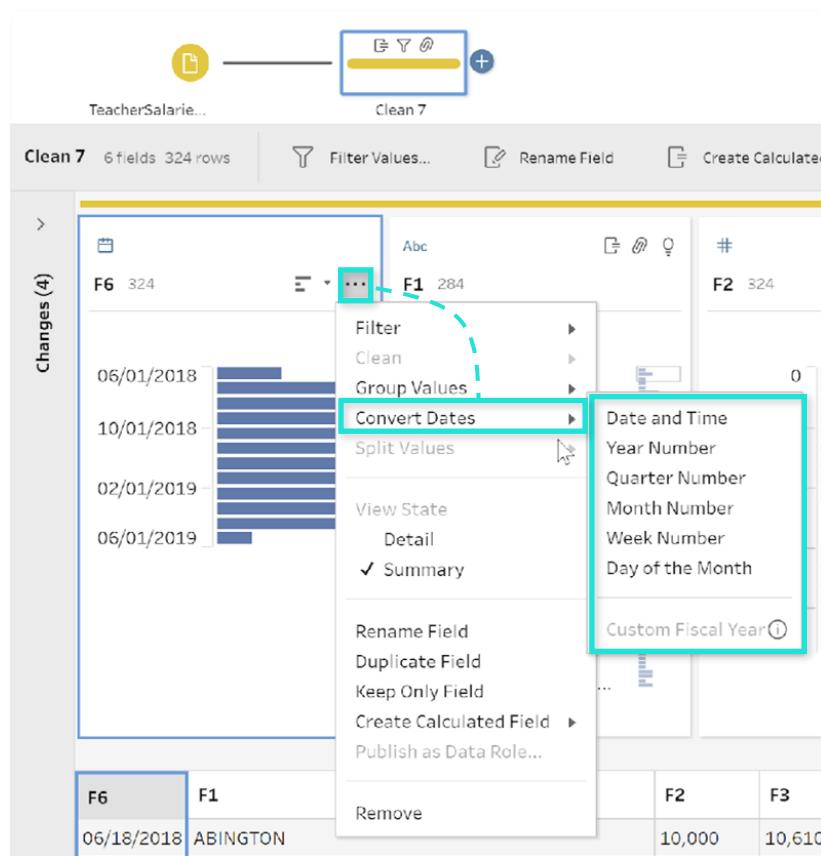
Right-click and choose "Edit Value" to edit or replace the value with null

Edit Multiple Values (Group Values)

Ctrl/Cmd click to manually group multiple values

Value Operations | Convert Dates

Convert dates to modify formats without the need for calculated fields or parsing functions



The screenshot shows the Power BI 'Clean 7' data view. A date column 'F1' is selected, and a context menu is open. The 'Convert Dates' option is highlighted with a red box and a callout line pointing to the 'Convert Dates' section of the 'Custom Fiscal Year' dialog. The dialog shows various date conversion options: Date and Time, Year Number, Quarter Number, Month Number, Week Number, Day of the Month, and Custom Fiscal Year. The 'Custom Fiscal Year' section is expanded, showing a dropdown for 'Select a date option' with 'Year Number' selected, and a dropdown for 'Fiscal year start' with 'January' selected.

Date and Time

Convert date field to **datetime** format (ex. 1/23/2020, 11:14:02 PM)

Year Number

Convert date field to **year number** format (ex. 2010, 2015, 2020)

Quarter Number

Convert date field to **quarter number** format (ex. 1, 2, 3, 4).

Month Number

Convert date field to **month number** format (ex. 1, 2, 3, 4 ... 11, 12)

Week Number

Convert date field to **week number** format (ex. 1, 2, 3, 4 ... 52, 53)

Day of the Month

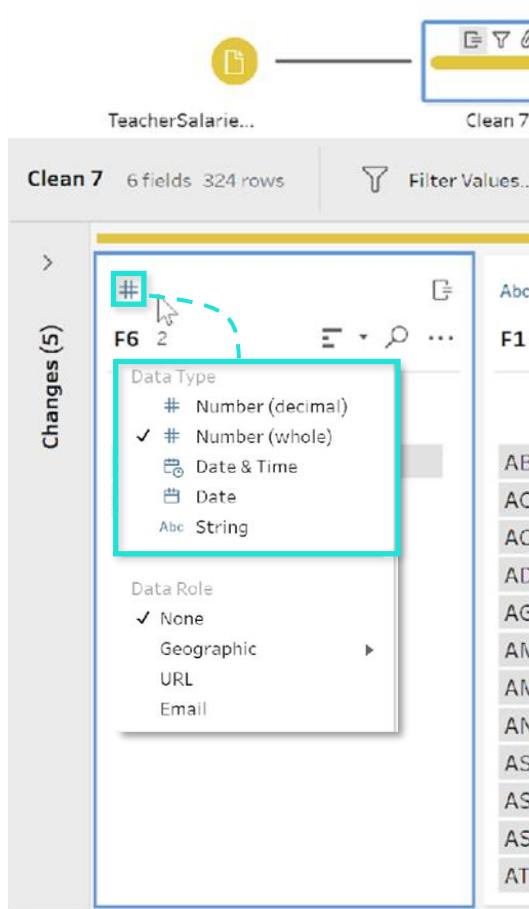
Convert date field to **day of month** format (ex. 1, 2, 3, 4 ... 31)

Custom Fiscal Year

Convert date field based on a **custom fiscal calendar**

Field Operations | Field Types

Field types can be customized in every flow step except the output, and are used to assign fields as numbers (decimal or whole values), dates (date or datetime) or text strings



Number (decimal)

Numeric value with decimal values (best for exact values like dollars, ratios, etc.)

Number (whole)

Numeric value with no decimal (best for quantity, date parts, ID fields, etc.)

Date & Time

Date and Time in the same field (best for exact time needs - where parts of a day matter)

Date

Date fields (best when date filtering and date calculations are needed - datediff, dateadd, etc.)

String

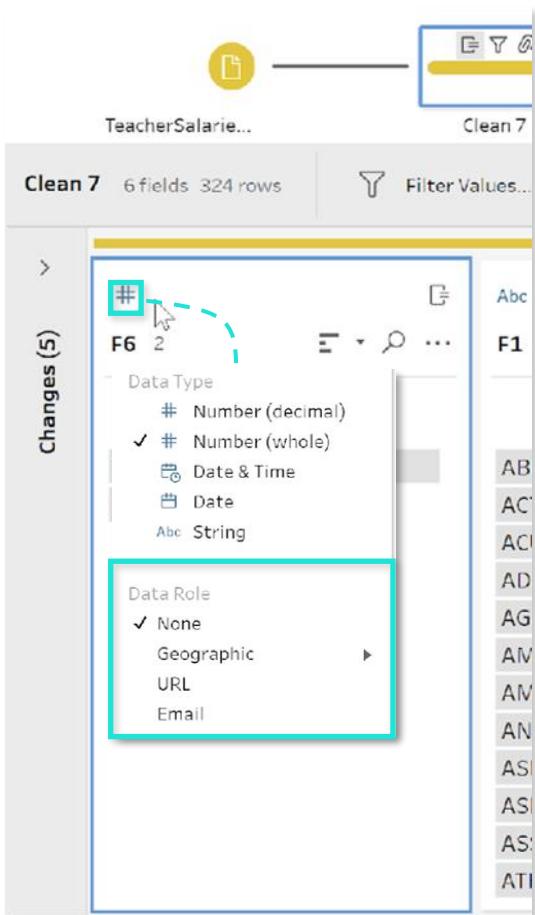
String fields (best for most dimensional values, text values that should be parsed, etc.)



Note that data types not only impact how fields are used in Tableau Prep, but also how **data visualization tools** interact with data and users

Field Operations | Data Roles

Data roles represent standard sets of values, which can be used to validate the values within a field



A screenshot of the Tableau Prep Builder interface. On the left, a sidebar shows 'Changes (5)'. In the center, a table is being edited with a row for 'F6'. The 'Data Role' dropdown for this row is open, showing options: 'None' (selected), 'Geographic', 'URL', and 'Email'. A cyan bracket on the right side of the slide groups this 'None' option with the 'None (default)' section. The top of the screen shows 'TeacherSalarie...' and 'Clean 7'.

None (*default*)

The default role for each field (no role assigned)

Geographic

Geospatial roles based on the same domains as Tableau Desktop

- Airport
- Area Code
- CBSA/MSA
- City
- Congressional District (US)
- Country/Region
- County
- NUTS Europe
- State/Province
- Zip code/Postal Code

URL

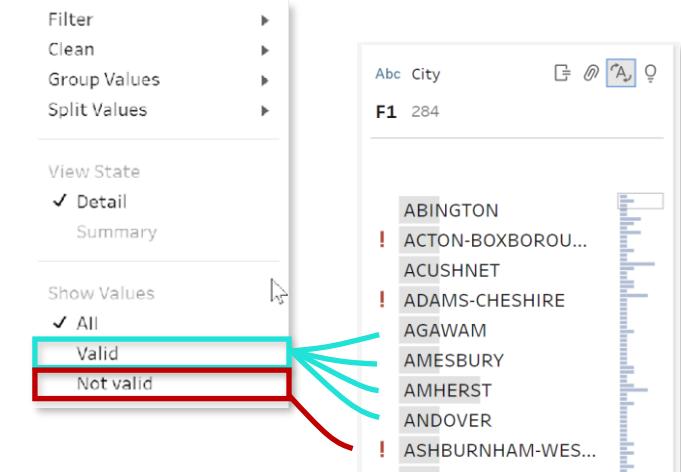
Web link-based role / URL fields

Email

Email role fields

Show Values (Valid/Not Valid)

Once applied, developers can view valid and not valid values and use value editing to correct potential issues



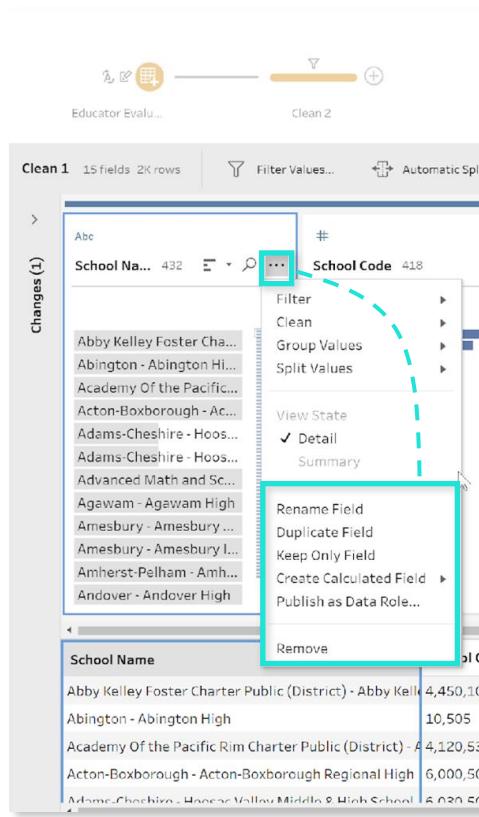
A screenshot of the Tableau Prep Builder interface. On the left, a sidebar shows 'Changes (5)'. In the center, a table is being edited with a row for 'F1'. The 'Data Role' dropdown for this row is open, showing 'Valid' (selected) and 'Not valid'. A cyan bracket on the right side of the slide groups this 'Valid' option with the 'Show Values (Valid/Not Valid)' section. The top of the screen shows 'Abc City' and 'F1 284'. To the right, a preview pane shows a list of city names: ABINGTON, ACTON-BOXBOROU..., ACUSHNET, ADAMS-CHESHIRE, AGAWAM, AMESBURY, AMHERST, ANDOVER, ASHBURNHAM-WES...

Published Data Roles are used in Prep Builder in conjunction with **Prep Conductor** (*not covered in this course*) to compare values in your flow against published standardized data values

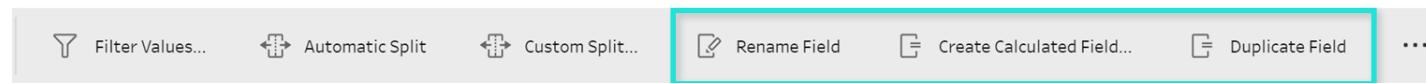


Field Operations | Cleaning

Field cleaning operations can be used to **modify, add or remove** fields from the flow



The screenshot shows a 'Clean' step in Power Automate. The step is titled 'Clean 1' and contains 15 fields and 2K rows. The 'School Name' field is selected, showing a list of school names and their counts. A context menu is open over the 'School Name' field, with the 'Keep Only Field' option highlighted by a red box. Other options in the menu include 'Filter', 'Clean', 'Group Values', 'Split Values', 'View State' (Detail checked), 'Rename Field', 'Duplicate Field', 'Create Calculated Field', and 'Publish as Data Role...'. A 'Remove' option is also visible at the bottom of the menu.



Rename Field changes the field name referenced (*can double-click name as well*)

Duplicate Field creates a copy of the field (*and adds a “-1” to the name*)

Keep Only Field keeps only the selected field(s) in the flow.

- **NOTE:** Use *Ctrl* or *Cmd* to select more than one field to keep

Create Calculated Field creates a new calculated field with the selected field referenced

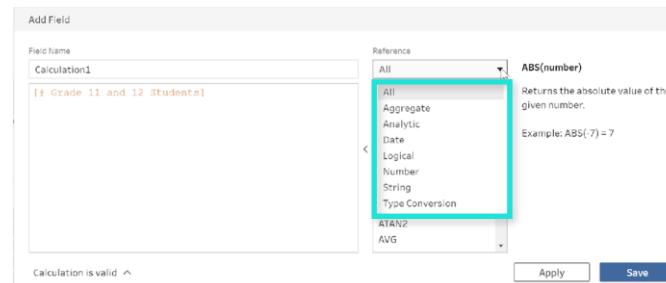
- **NOTE:** We'll cover calculations in depth later in this section!

Remove removes the selected field(s) from the flow

Calculated Fields

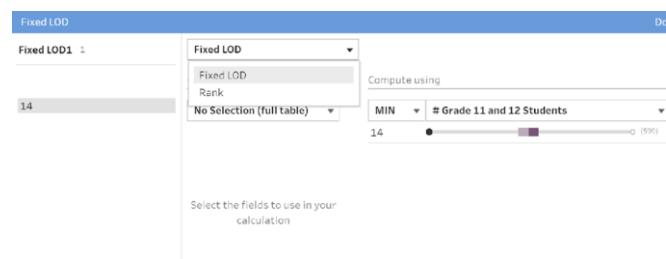
Calculated fields can be created via **standard editor** or **visual editor**, depending on the function

The screenshot shows the Tableau interface with a data source named 'Advanced Cour...'. Two calculated fields are being created: 'Clean 6' and 'Clean 7'. The 'Create Calculated Field...' button is highlighted. A context menu is open on a field, with 'Create Calculated Field' also highlighted. A dropdown menu shows options: 'Custom Calculation', 'Fixed LOD', and 'Rank'. The 'Custom Calculation' option is highlighted.



Standard Editor

Standard calculation editor available for **all functions**



Visual Editor

Modified calculation editor for **Fixed LOD** and **Rank**

Aggregate

Summarize or change the **level of granularity** of your data

Analytic

Perform calculations across **tables or partitions**

Date

Create, modify, and calculate **date/time** fields

Logical

Determine if a **conditional statement** is true or false

Number

Computation-based functions used on **numerical fields**

String

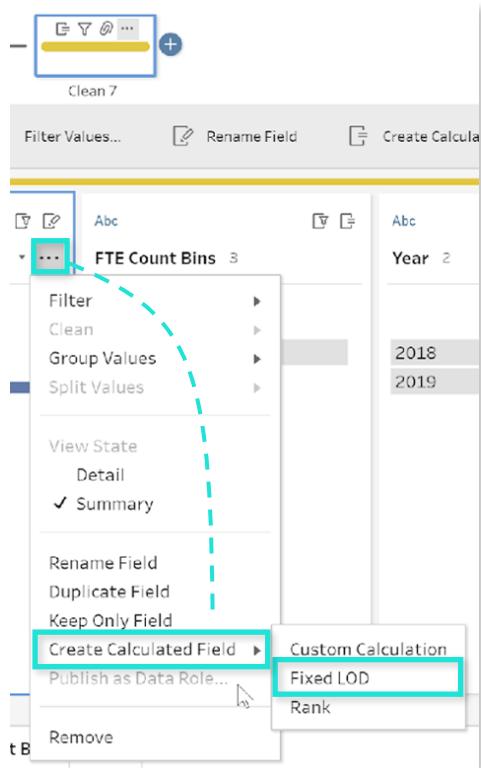
Manipulation **text-based data**

Type Conversion

Convert **fields** from one data type to another

Level of Detail Calculations

Level of detail (LOD) calculations are used to perform aggregations at different grains of data



LOD Expression Syntax:

```
{ FIXED [Year] : MEDIAN( [% Students Completing Advanced] ) }
```

Level of Detail Element

FIXED is the only option in Prep

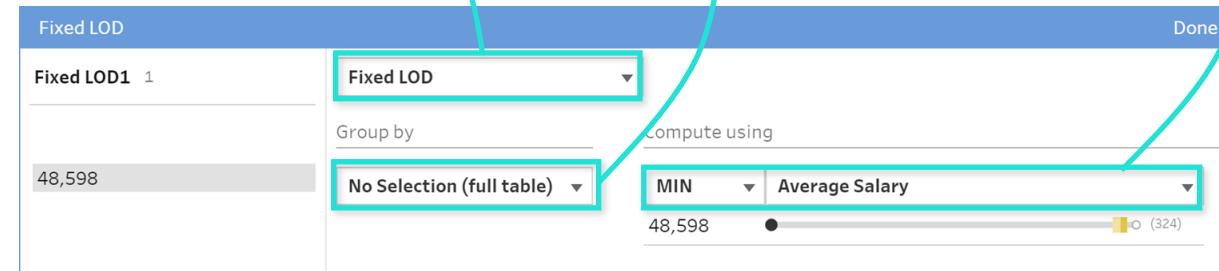
Dimension Declaration

Grain at which data is aggregated

Aggregate Expression

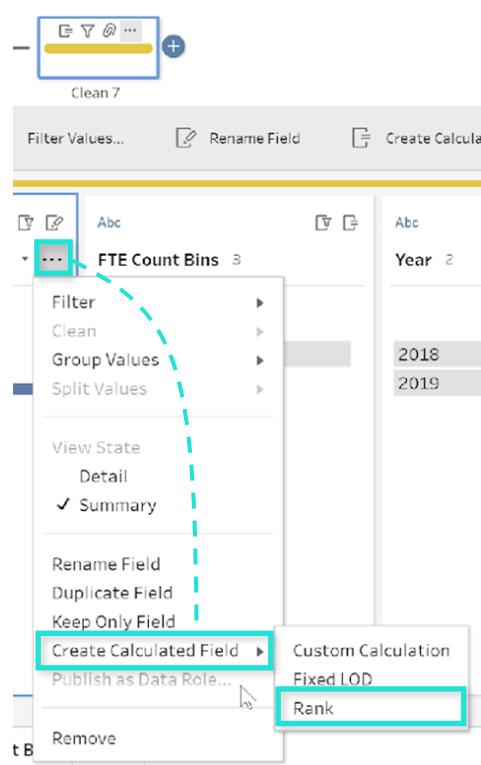
Calculation to be performed

LOD Visual Editor:



Analytic Calculations | Rank

Rank calculations are a subset of analytic calculations which can be applied across an entire table or a subset of rows (partition)



Analytic Calculation Editor:

```
{ PARTITION [Year] : { ORDERBY [% Students Completing Advanced] ASC: RANK() }
```

Partition

Designates rows to which calculation will be applied

Order by

Specifies field to generate sequence for ranking

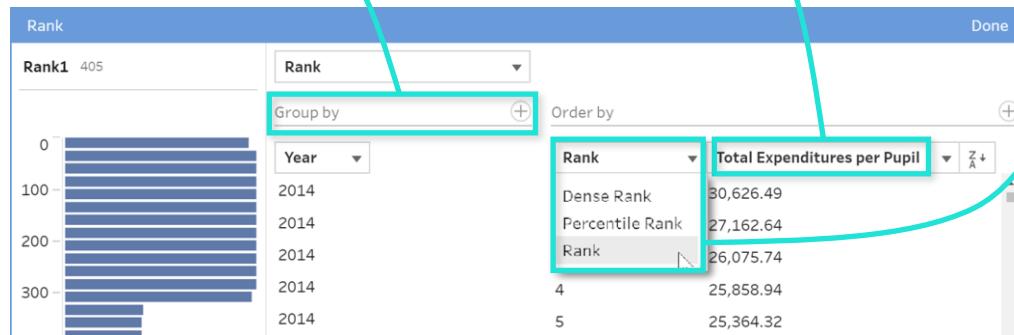
Rank Calculation

Rank or row number calculation, with optional sort order (DESC by default)

Options include:

- RANK()
- RANK_DENSE()
- RANK_MODIFIED()
- RANK_PERCENTILE()
- ROW_NUMBER()

Visual Editor:

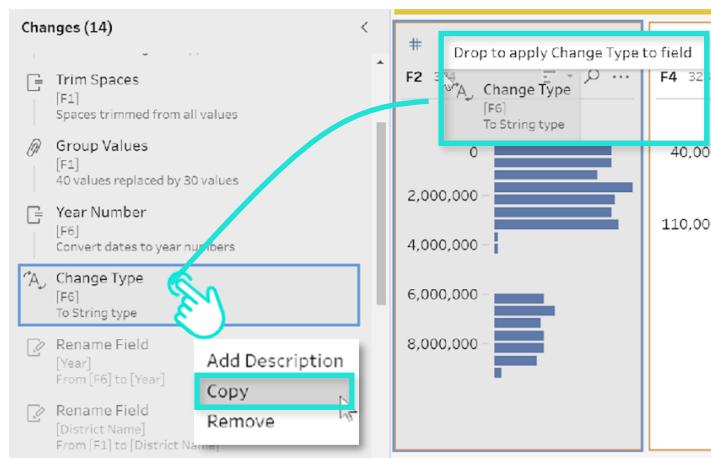


PRO TIP: Copy & Paste

Copy and paste individual elements within flows, including cleaning operations, fields or steps

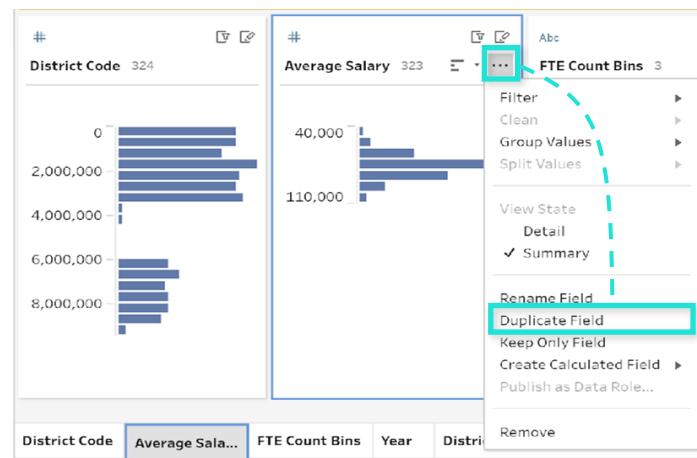
Copy Cleaning Operations

Drag from the changes pane onto another field, or right-click to copy within the same flow



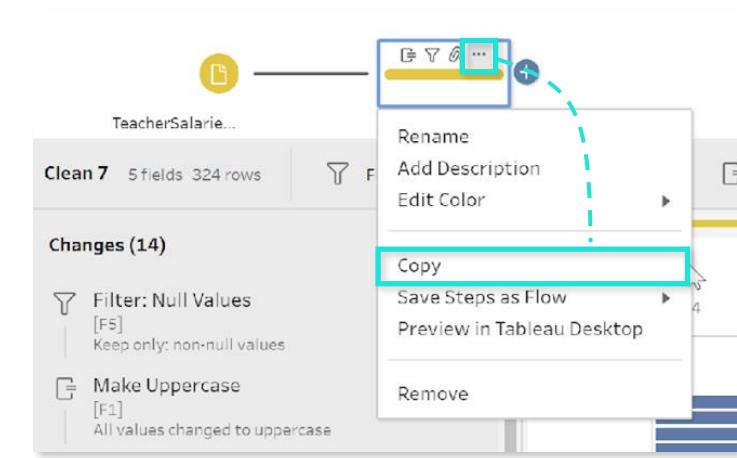
Duplicate Fields

Copy fields using the “Duplicate Field” options



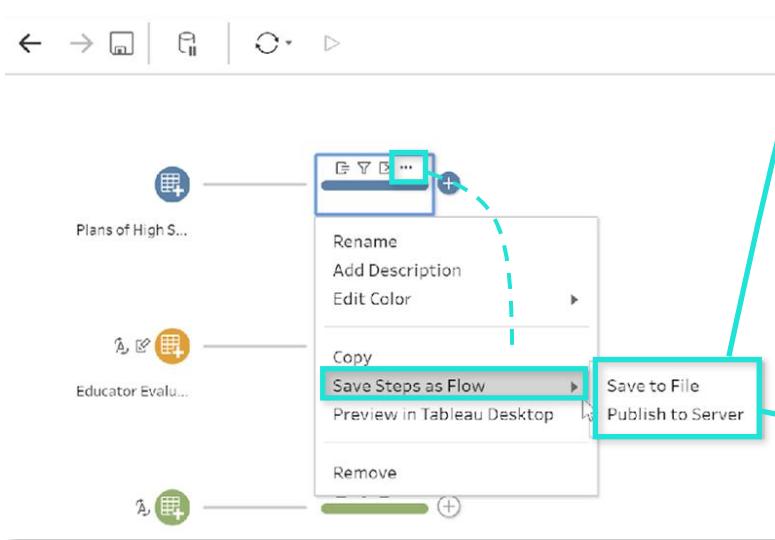
Copy Flow Steps

Copy individual steps into different parts of the flow (or whitespace)



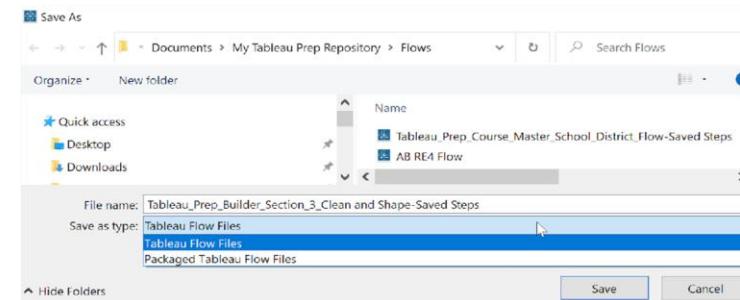
PRO TIP: Reusable Flow Steps

Reusable flow steps can be created, saved and imported into other flows, and are commonly used for steps which are used frequently or leveraged by other users



Save to File

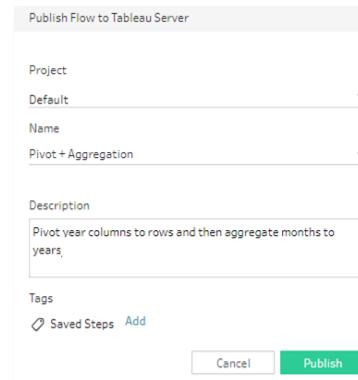
Save a file locally as a *Flow* or *Packaged Flow*



Publish to Server

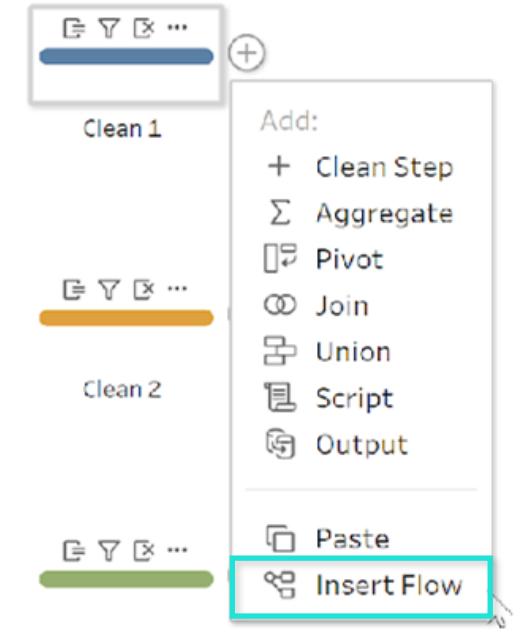
Publish a flow to Tableau Server using publisher credentials

NOTE: Published flows which utilize file-based input steps are not yet supported



Insert Flows

Insert flows using the “*Insert Flow*” option from any step or whitespace



HOMEWORK: Operations & Calculations

THE **SITUATION**

Your old boss at **Tech Data Talent (TDT)** contracted you for some data prep assistance. You'll need to use your Tableau Prep skills to make sure the TDT team is working with clean and accurate data.

THE **BRIEF**

Your task is to **clean survey response data** to help the team accurately analyze mental health trends in the tech industry. The key will be to **clean and organize** the data in a way that will allow TDT's analytics group to easily analyze and visualize patterns.

THE **OBJECTIVE**

Use Tableau Prep to:

- Clean and manipulate values
- Modify and customize fields
- Create calculated fields
- Export flow steps

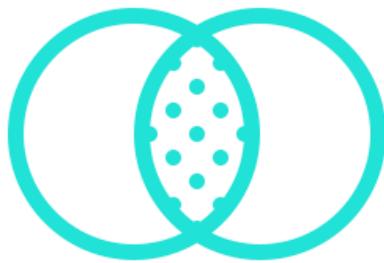


Combining & Pivoting

Combining & Pivoting

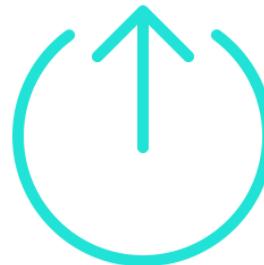
Data can be transformed and combined using several types of flow steps in Tableau Prep, including **Union**, **Join**, **Aggregate** and **Pivot**

Union & Join



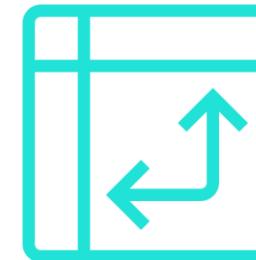
- **Union** and **join** are used to blend data together to create combined tables
- **Union** stacks records from common columns, and **joining** adds related fields from another table

Aggregate



- Change the **granularity** of your data using **aggregate** (i.e. daily to monthly)
- **Group** data by fields in your table to control the **level of aggregation**

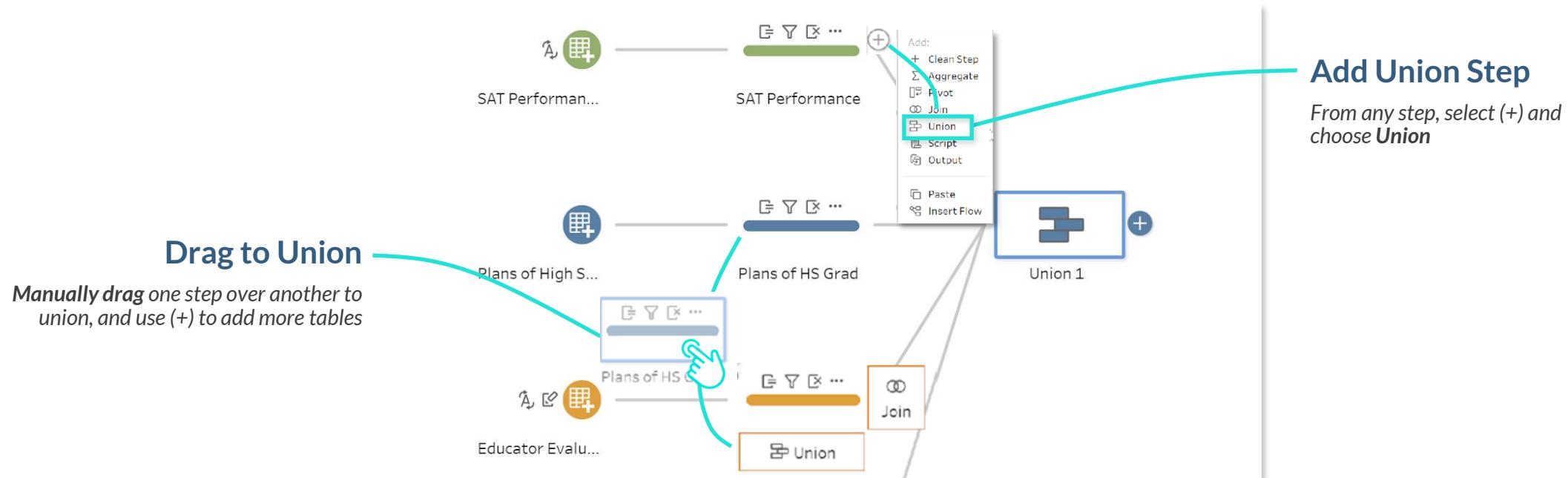
Pivot



- Transpose **rows to columns** (or columns to rows) using a **pivot** step
- Set up data outputs for **optimal consumption** using different table layouts

Combine Data | Union

The **union** step appends (or “stacks”) records from multiple tables, based on matching columns



PRO TIP: If you need to union 10+ tables, try using **wildcard unions** in the input step!

Union Results & Common Issues

Review the **union results** in the profile pane to identify and resolve common union issues, including data type differences and mismatched fields

Inputs
Color-coded list of tables included in union

Resulting Fields
Count of total and mismatched fields

Mismatched Fields
List of fields which did not union (may be unique to source or truly missed during union)

Data Type Differences
Columns with the same name but different types will automatically default to strings

[Year] is combined from different types of data. During Union, the different types are assigned to string to align fields by name.

Merge Fields
Drag fields over each other to merge them into one (in case union didn't identify them as a match)

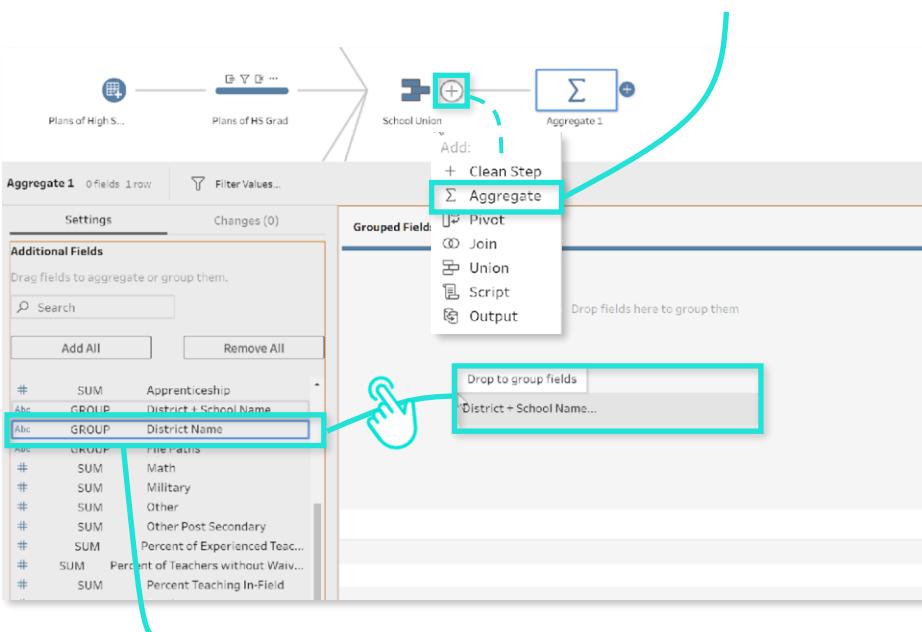
District + School Name	District Name	School Name	Table Names	% Proficient - 1
HUDSON - HUDSON HIGH	HUDSON	HUDSON HIGH	teacherdata_school_2014_2015.xlsx/Teacher Data	null
WAYLAND - WAYLAND HIGH SCHOOL	WAYLAND	WAYLAND HIGH SCHOOL	teacherdata_school_2014_2015.xlsx/Teacher Data	null

Combine Data | Aggregate

Aggregate allows you to change the granularity of your data by summarizing values at higher levels

Add Aggregate Step

Select the (+) icon next to your existing step and choose “Aggregate”

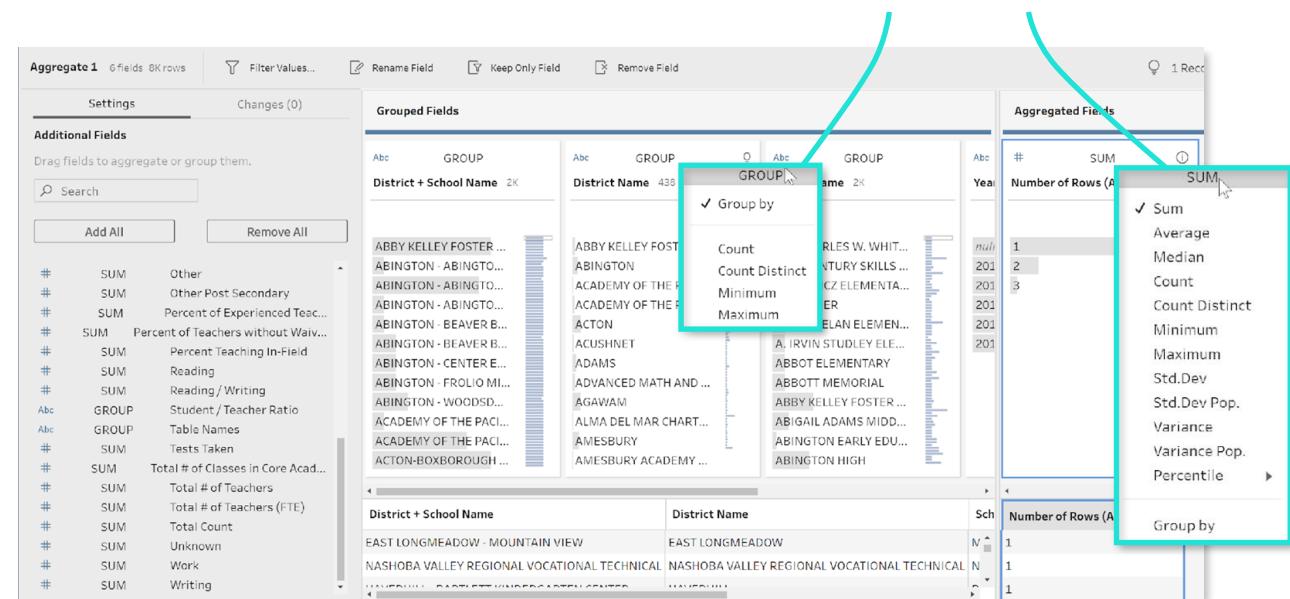


Additional Fields

Drag fields to the “Grouped Fields” or “Aggregated Fields” panes
(*NOTE:* fields not selected will not pass through this step)

Adjust Grouping & Aggregation

Select the card headers to update grouping or aggregation logic



PRO TIP: Use “Group By” with no aggregation to create a unique list of dimensions

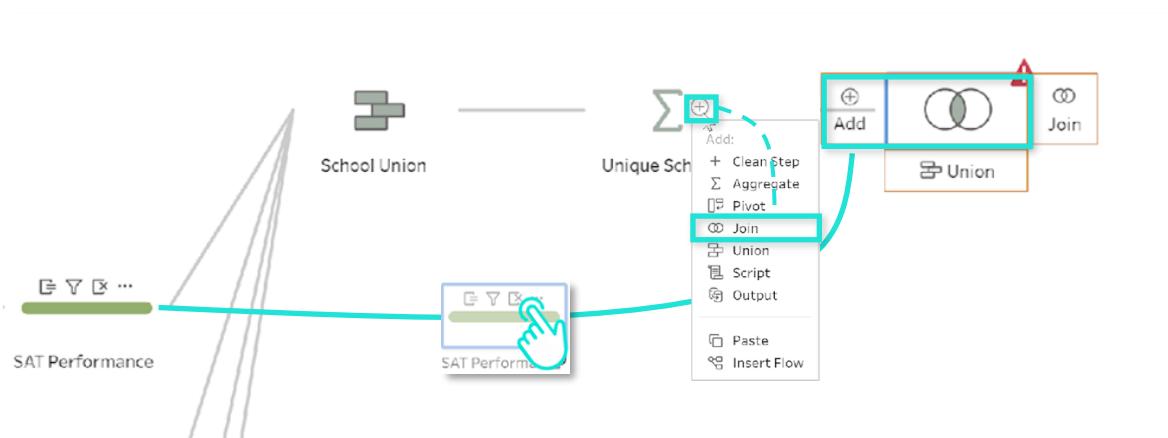
Combine Data | Join

Join is used to **combine data between tables** which share common or related fields

Join Types

	Inner	For each row, includes values that have matches in both tables
	Left	Include all values from left table and matches from right table
	Right	Include all values from right table and matches from left table
	Left (only)	Include only values from left table and no matches to right
	Right (only)	Include only values from right table and no matches to left
	Outer	Includes all values from right and left that don't match
	Full Outer	Includes all values from both tables, non-matches are null

Creating a Join



Click to Join

Select (+) from a step and choose "Join"

Drag to Join

Add more sources to the join by dragging step to the "(+)" Add icon in the Join step

Join Results & Common Issues

Review the **join results** in the profile pane to **identify and resolve common issues**, including mismatched values or incorrect join types or clauses

The screenshot shows the Tableau interface with the following components highlighted:

- Add Join Clauses**: A callout points to the "Join Clauses" pane, which lists fields for joining "Unique Schools" and "SAT Performance".
- Applied Join Clauses**: A callout points to the "Applied Join Clauses" pane, showing the selected join clauses: "School Code = School Code" and "Year = Year".
- Join Type**: A callout points to the "Join Type" pane, which displays a Venn diagram for "Unique Schools" and "SAT Performance" and a summary of join results.
- Summary of Join Results**: A callout points to the "Summary of Join Results" pane, showing a bar chart of included and excluded values: Unique Schools (8,161), SAT Performance (1,845), and Join Result (8,161).
- Join Clause Recommendations**: A callout points to the "Join Clause Recommendations" pane, listing recommended join clauses: "District + School Name = District + School Name", "School Name = School Name", and "District Name = District Name".
- Join Results**: A callout points to the "Join Results" pane, which displays a list of joined data with black text for correct matches and red text for mismatches.
- Join Clauses Pane**: A callout points to the bottom pane of the interface, which displays the "Join Clauses" section of the Tableau interface.

Pivot

Pivoting transposes rows to columns (or vice versa), allowing you to create “wide” or “tall” tables

Pivot Names & Values
Assign Name and Value columns for pivoted fields

Pivot Type Selector
Columns to Rows or Rows to Columns

Add Pivot Step
Select “Pivot” from latest step (+) dropdown

Add Fields
Select and drag fields into the “Pivoted Fields” pane

Wildcard Search
Use wildcard search to find all fields containing specific text and pivot automatically

Pivot 1 42 fields 8K rows Filter Values... Create Calculated Field...

Fields

- Automatically rename pivoted fields and values
- # Evaluated
- # of Educators to be Evaluated
- % 4 Year College
- % Evaluated
- % Evaluated - Overall Median
- % Exemplary
- % Needs Improvement
- % of Core Academic Classes Taught by Teachers

Pivoted Fields

Pivot1 Names Pivot1 Values

Drop fields here to pivot them
Or
Use wildcard search to pivot

Evaluated...

Pivot Type Selector

✓ Columns to Rows
 Rows to Columns

Pivot Results

% 4 Year College 1K
Total SAT 503

0 500 1,000 1,500

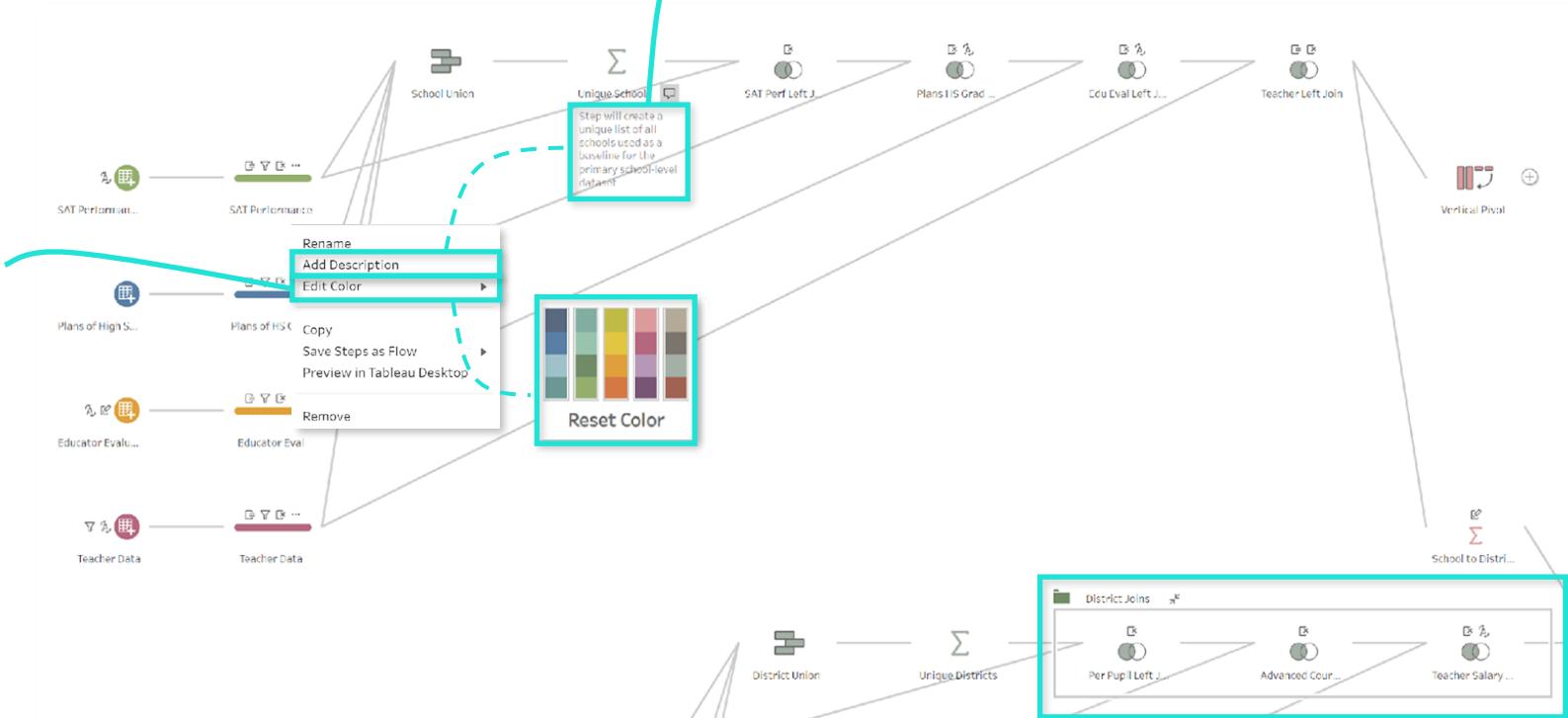
Add:

- + Clean Step
- Σ Aggregate
- Pivot**
- Join
- Union
- Script
- Output

PRO TIP: Organize Your Flow

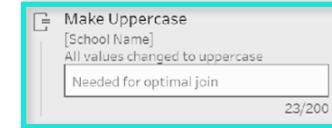
Color Scheme

Customize colors to identify related steps in the flow



Descriptions

Add descriptive notation to steps to provide details and clarity



Group Steps

Use groups to organize and compress large collections of flow steps to make them easier to digest and share



HOMEWORK: Combining & Pivoting

THE SITUATION

As a leader of your local F1 racing fan club, you're in charge of preparing data for the club's upcoming annual F1 fantasy draft.

THE BRIEF

You've been asked to gather data to help members accurately analyze **driver stats**, **lap times**, and **race results**. The key will be to combine raw data into a **centralized source** that combines all historical race data as well as peripheral driver and result information.

THE OBJECTIVE

Use Tableau Prep to:

- Join & Union multiple sources
- Aggregate and Group data
- Pivot columns to rows
- Organize flow steps

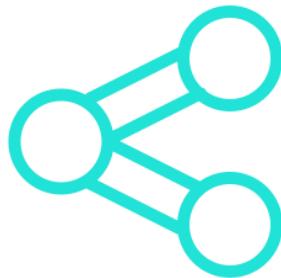


Sharing & Updating

Sharing & Updating

Tableau Prep allows you to configure options for **sharing data outputs** and **updating flows**

Share



- Share data outputs as **local files**, **published data sources**, or **updated tables in databases**
- Preview your data in **Tableau Desktop** prior to automating your flow to ensure your success criteria have been met

Update



- Refresh your flow and configure incremental update options
- Learn about the benefits of using **Prep Conductor** to fully automate prep flows

Saving Flows

Save your flow locally to retain steps, bundle local data sources, and share flows with other users

Save Flow
Manually save your flow as a **.tfl** file to retain your work

Save As
Use **Save As** to choose the type of flow file saved

Export Packaged Flow
Exports a packaged version of the flow directly as a **.tflx** file

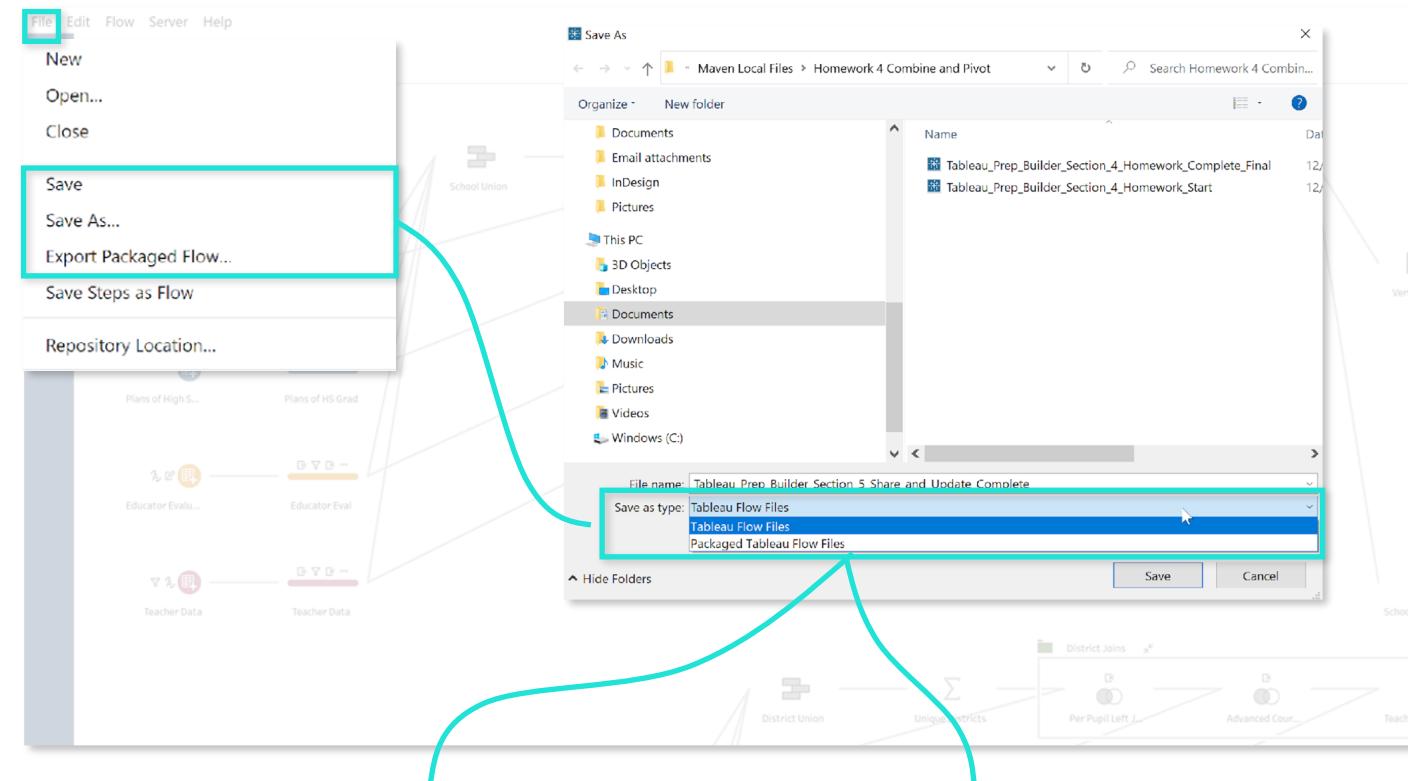


Tableau Flow File

The standard flow as a **.tfl** file (no data retained)

Packaged Tableau Flow File

Packaged flow, which bundles Excel, text, and Tableau extracts with the flow as a **.tflx** file

PRO TIP: Preview in Desktop

Use the Preview in Tableau Desktop option to preview the output while developing a flow

Tableau Prep:

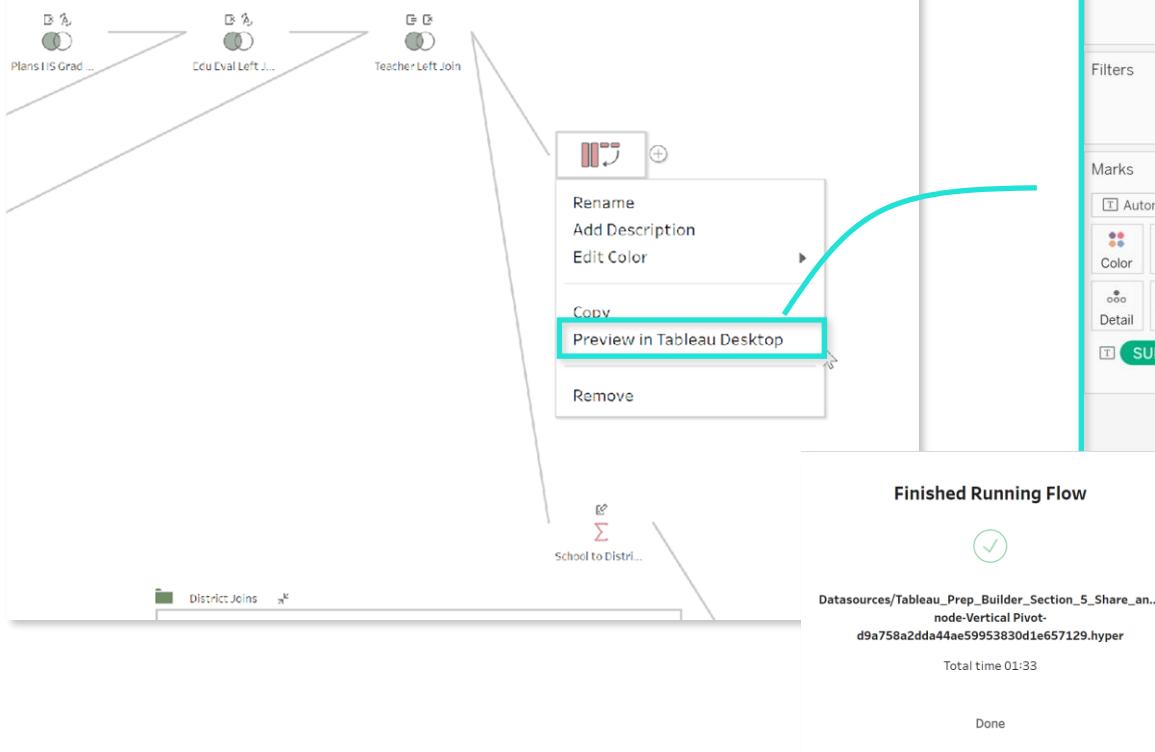


Tableau Desktop:

Metric Names	2015	2016	2017	2018	2019	2020			
	Year	246	300	197	282	8,631	7,795	8,056	7,933
2 Year Private College		10,107	10,061	10,096	9,885	11,010	11,587	11,253	11,416
2 Year Public College		82,835	83,608	85,361	89,681	88,754	89,855	89,681	88,754
# Evaluated		226	233	223	209	167,294	167,968	170,633	167,294
# of Educators to be Evaluated		181,048	180,745	182,668	180,745	21,747	23,461	25,252	21,747
% 4 Year College		7,194	6,304	5,933	6,304	28,007	27,979	28,824	28,007
% Evaluated		127,242	125,357	124,798	125,357	126,620	126,471	126,168	127,242
% Evaluated - Overall Median		126,620	126,471	126,168	126,471	719	483	524	126,620
% Exemplary		21,747	23,461	25,252	23,461	7,194	6,304	5,933	7,194
% Needs Improvement		149	149	183	149	170,964	170,799	183,307	183,307
% of Core Academic Classes Taught by T...		124,798	124,798	124,798	124,798	Math	170,964	170,799	186,930
% of Teachers Licensed		126,168	126,168	126,168	126,168	Military	753	774	805
% of Teachers Licensed in Teaching Assi...		524	524	524	524	Other	383	579	690
% Proficient		149	149	183	149	Other Post Secondary	733	705	872
% Proficient - Divisional Median		126,168	126,168	126,168	126,168				
% Unsatisfactory		719	483	524	524				
Apprenticeship		149	149	183	149				
Math		124,798	124,798	124,798	124,798				
Military		126,168	126,168	126,168	126,168				
Other		524	524	524	524				
Other Post Secondary		183	183	183	183				

Create Local Extracts

Create local extracts in Tableau Prep to output as either .csv or .hyper file formats

Save Output To

Choose output target (file, published data source, or database table)

Name

Name the output extract

Location

Choose a location to save the output

Output Type

Choose an output type (.hyper or .csv for local files)

Write Options

Choose local write options (create table or append to table)

Run Flow

Execute the flow on full data

The screenshot shows the Tableau Prep interface with a data flow and an 'Output' step. The 'Output' step is highlighted with a red box. A tooltip for this step says: 'Step will create a unique list of all schools used as a baseline for the primary school-level dataset'. The 'Add' menu for this step is open, showing options like 'Clean step', 'Aggregate', 'Pivot', 'Join', 'Union', 'Script', and 'Output'. A red box highlights the 'Output' option in this menu. The 'Output' dialog box is also open, showing options for 'File', 'Published data source', and 'Database table'. A red box highlights the 'File' option. The 'Output type' dropdown in the dialog shows 'Tableau Data Extract (.hyper)' and 'Comma Separated Values (.csv)'. A red box highlights the 'Tableau Data Extract (.hyper)' option. Below these, a tooltip for 'Create table' says: 'If the table already exists, it will be deleted.' A tooltip for 'Append to table' says: 'Add data to the existing table. If the table doesn't exist, it's created when the flow is first run.' The 'Run Flow' button at the bottom of the dialog is also highlighted with a red box. To the right of the dialog, a 'Data Grid / List View' window is open, showing a list of output field details. A red box highlights the 'Tableau Data Extract (.hyper)' option in the 'Output' dialog.

Add Output

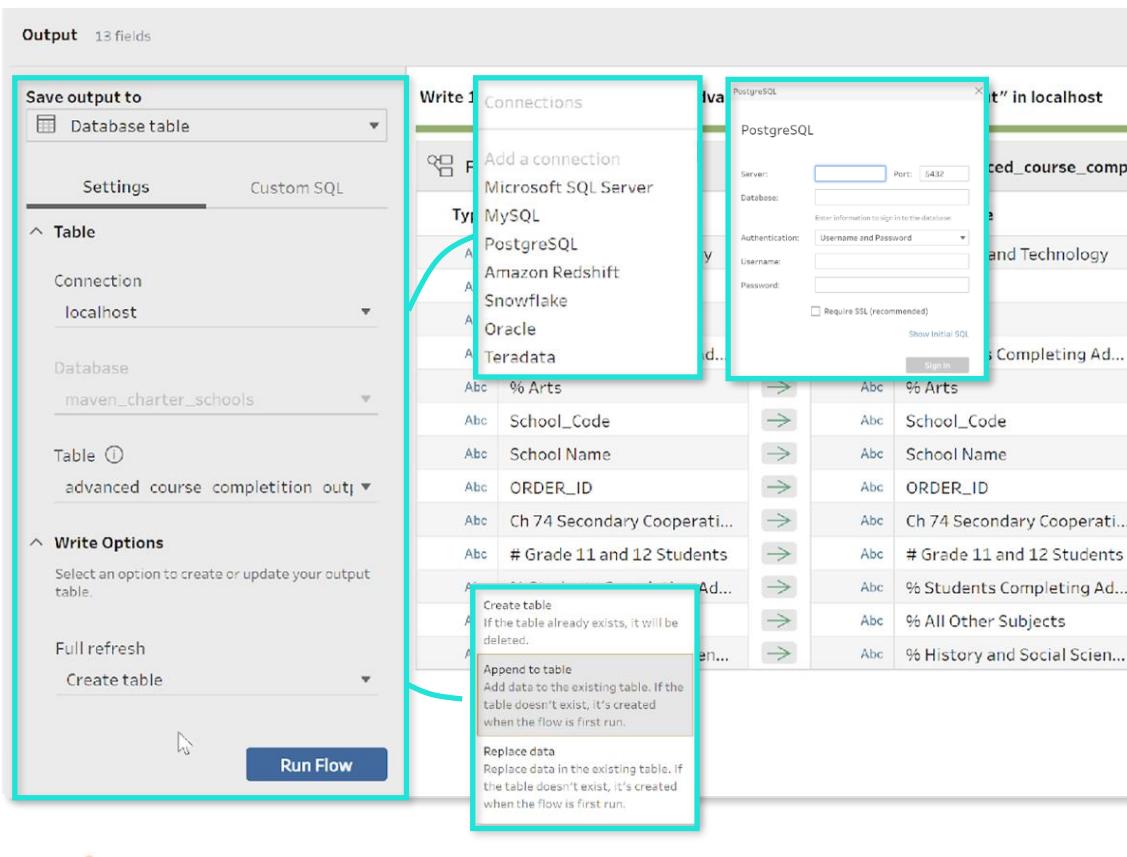
Select (+) > Output to create an output step

Data Grid / List View

Choose list view for output field details

Save to External Databases

Prep can write to external databases as a new table or append/replace data in an existing table



Save Output To
Choose output target (file, published data source, or database table)

Connection
Select database type and enter credentials

Database
Select a database schema

Table
Select an existing table or create a new one

Write Options
Create, append, or replace table data

Run Flow
Execute the flow on full data

Custom SQL
Embed custom SQL code to execute before or after flow has written data to database

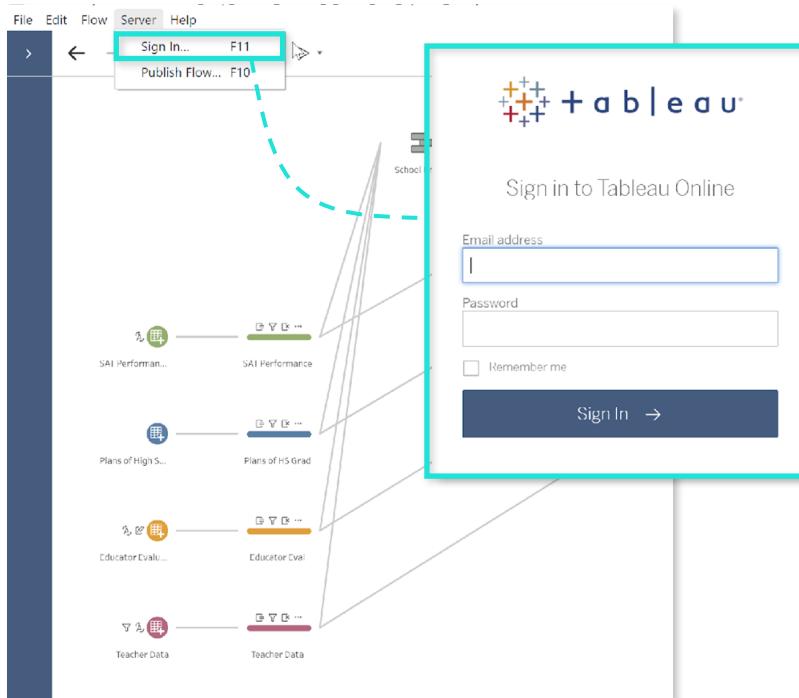
PRO TIP: Select “enable incremental refresh” on input/output to only add new data!

Create Published Data Sources

Publish data sources to Tableau Server to grant user access to data and enable automated refresh

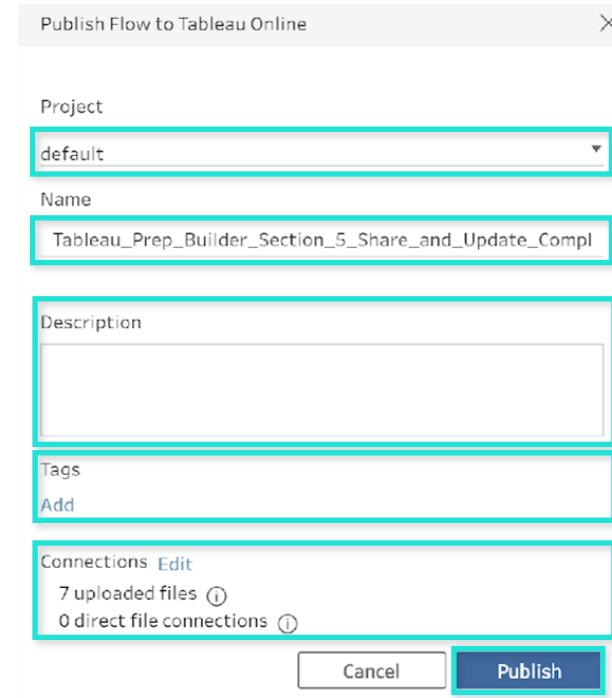
Sign-In to Tableau Server

Use credentials or SSO to log into Tableau Server / Tableau Online



Publish Flow

Configure the flow's publishing options



Project

Select the project where your flow will be located

Name

Name your flow

Description

Give a brief description of what your flow does

Tags

Make the flow searchable on server using tags

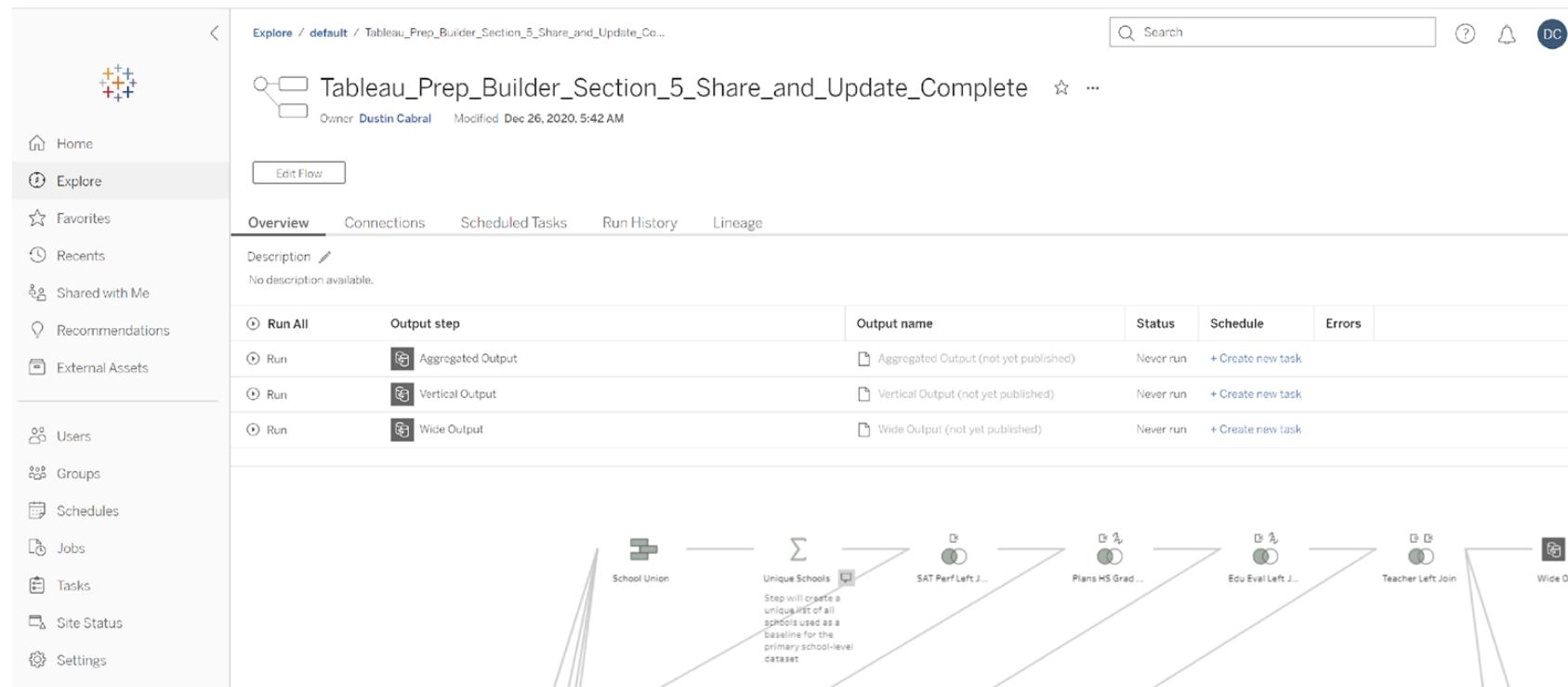
Connections

Edit connections to embed credentials; local files need to be uploaded (flat) or use direct connection (refreshed on regular basis)

NOTE: Direct connection requires that Tableau Server is granted access to the share / database

PREVIEW: Tableau Prep Conductor

Tableau Prep Conductor can be used to automate and optimize flows in Tableau Server / Online



Schedule Flows

Schedule flows to automatically run on a set day or at a specified refresh time

Create / Edit Flows

Create and edit flows in your browser, and run flows manually on-demand

Administration

View performance and scheduling to optimize flow runs

Alerts

Configure alerts and email notifications to notify you of failed flows

HOMEWORK: Sharing & Updating

THE SITUATION

Your friend Anna is a Director at **Maven Financial**, a local bank branch, and needs your help extracting customer data from Tableau Prep.

THE BRIEF

Anna has asked you to **set up outputs for various stakeholders**, utilizing various file formats and platforms. Your job is to deliver the data in a **predictable and efficient way**, to enable the business to use it going forward.

THE OBJECTIVE

Use Tableau Prep to:

- Save flows locally
- Preview flows in Tableau Desktop
- Output flow data to local sources
- Output flow data to a database and Tableau Server*

