OUT PUT

Clean Data from Excel, CSV, PDF, and Google Sheets with Data Interpreter

Applies to: Tableau Cloud, Tableau Desktop, Tableau Server

When you track data in Excel spreadsheets, you create them with the human interface in mind. To make your spreadsheets easy to read, you might include things like titles, stacked headers, notes, maybe empty rows and columns to add white space, and you probably have multiple tabs of data too.

When you want to analyze this data in Tableau, these aesthetically pleasing attributes make it very difficult for Tableau to interpret your data. That's where Data Interpreter can help.

Tip: Though Tableau's Excel add-in is no longer supported, Data Interpreter can help you reshape your data for analysis in Tableau.

What does Data Interpreter do?

Data Interpreter can give you a head start when cleaning your data. It can detect things like titles, notes, footers, empty cells, and so on and bypass them to identify the actual fields and values in your data set.

It can even detect additional tables and sub-tables so that you can work with a subset of your data independently of the other data.

After Data Interpreter has done its magic, you can check its work to make sure it captured the data that you wanted and identified it correctly. Then, you can make any necessary adjustments.

After you select the data that you want to work with, you might also need to do some additional cleaning steps like pivoting your data, splitting fields, or adding filters to get the data in the shape you want before starting your analysis.

Note: If your data needs more cleaning than what Data Interpreter can help you with, try Tableau Prep(Link opens in a new window).

Turn on Data Interpreter and review results

1. From the **Connect** pane, connect to an Excel spreadsheet or other connector that supports Data Interpreter such as Text (.csv) files, PDF files or Google sheets.

2. Drag a table to the canvas (if needed), then on the **Data Source** page, in the left pane, select the **Use Data Interpreter** check box to see if Data Interpreter can help clean up your data.
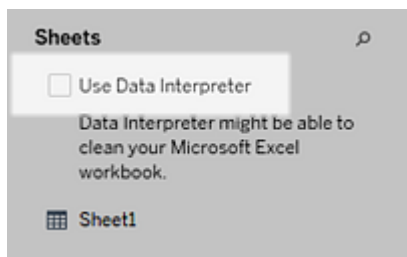
Note: When you clean your data with Data Interpreter, Data Interpreter cleans all the data associated with a connection in the data source. Data Interpreter does not change the underlying data.

3. In the Data pane, click the **Review the results** link to review the results of the Data Interpreter.
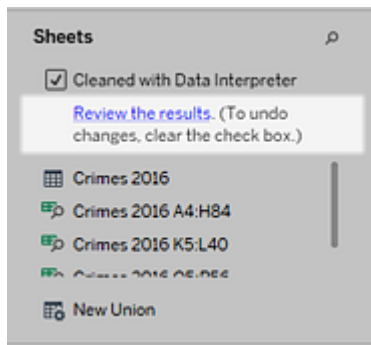

A copy of your data source opens in Excel on the **Key for the Data Interpreter** tab. Review the key to find out how to read the results.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2		Key for Understanding the Data Interpreter Results												
3														
4														
5		Use the key to understand how your data source has been interpreted.												
6		To view the results, click a worksheet tab.												
7		Note: Tableau never makes changes to your underlying data source.												
8														
9														
10														
11		Key:												
12			Data is interpreted as column headers (field names).											
13			Data is interpreted as values in your data source.											
14			Data derived from an Excel merged cell is interpreted as value in your data source.											
15			Data is ignored and not included as part of your data source.											
16			Data has been excluded from your data source.											
17			Note: To search for all excluded data, use CRTL +F on Windows											
18			or Command F on the Mac, and then type '***DATA REMOVED***'.											
19														
20														
21		If the Data Interpreter has interpreted the Tableau data source incorrectly, close the spreadsheet,												
22		and then clear the Cleaned with Data Interpreter check box from the Data Source page.												
23		If the Tableau data source continues to be interpreted incorrectly or for general information												
24		about why some data was removed by the Data Interpreter, refer to												
25		Resolving Common Issues with Data Interpreter Results												
26		Help Tableau improve the Data Interpreter by emailing your file to support@tableau.com												
27		or filing a support request with an attached file at:												
28														
29		http://tableau.com/support/request												
30														

Key for the Data Interpreter | Crimes 2016 | Crimes 2016_subtables | A04_H84 | K05_L40 | O05_P56

4. Click each tab to review how Data Interpreter interpreted the data source.

If Data Interpreter found additional tables, also called found tables or sub-tables, they are identified in the <sheet name>_subtables tab by outlining their cell ranges. A separate tab is also included for each sub-table, color coded to identify the header and data rows.

If Data Interpreter does not provide the expected results, clear the **Cleaned with Data Interpreter** check box to use the original data source.
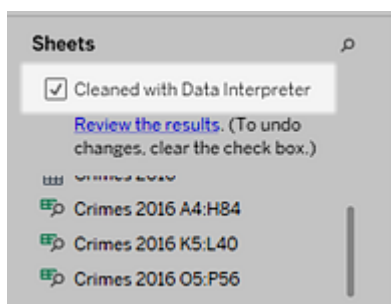
5. To replace the current table with any of the found tables, drag the current table off the canvas and then drag the found table that you want to use to the canvas.

If Data interpreter has misidentified the range of the found table, after you drag the found table to the canvas, click the drop-down arrow on that table, and then select **Edit Found Table** to adjust the corners of the found table (the top-left cell and bottom-right cell of the table).
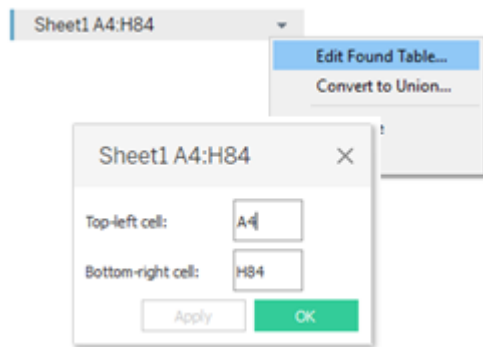
6. After you have the data that you want to work with, you can apply any additional cleaning operations to your data so that you can analyze it.

Data Interpreter Example

In this example we are connecting to an Excel spreadsheet with violent crime data by city and state for the year 2016. This spreadsheet includes multiple tables on one

sheet and some extra formatting.

A. Title

B. Merged header cells

C. Extra white space

D. Sub-tables

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Violent Crimes in 2016 In the United States by City and State																
2																	
3																	
4	Location	Location	Months	Months	Months	Months	Months	Months									Header
5	city	state	Apr	Jun	Jul	Aug	Sep	Oct			state	Total Crimes 2016			State	Populatio	Header
6	Albuquer	New Mexico					46				Alabama	12			Alabama	4860545	Data
7	Anaheim	California		4							Alaska	26			Alaska	741522	Data
8	Anchorag	Alaska	1				26				Arizona	132			Arizona	6908642	Data
9	Arlington	Texas				17					California	515			Arkansas	2988231	Data
10	Atlanta	Georgia					85				Colorado	64			California	39296476	Data
11	Aurora	Colorado					16				D.C.	105			Colorado	5530105	Data
12	Austin	Texas				28					Florida	210			Connectic	3587685	Data
13	Bakersfiel	California		22							Georgia	85			Delaware	952698	Data
14	Baltimore	Maryland						230			Hawaii	9			District of	684336	Data
15	Boston	Massachusetts					28				Illinois	536			Florida	20656589	Data
16	Buffalo	New York					38				Indiana	151			Georgia	10313620	Data
17	Chandler	Arizona					3				Kansas	10			Hawaii	1428683	Data

The extra formatting in this spreadsheet makes it difficult for Tableau to determine what the field headers and values are.

Instead, it reads the data vertically and assigns each column the default value F1, F2, F3 (Field 1, Field 2, Field 3) and so on. Blank cells are read as null values.

To see if Data Interpreter can help clean this data set, we select **Use Data Interpreter**.

Data Interpreter detected the proper headings for the fields, removed the extra formatting and found several sub-tables. The sub-tables are listed in the **Sheets** section in the Data pane and are named using the original sheet name and the cell ranges for each sub-table.

In this example there are three sub-tables: **Crimes 2016 A4:H84**, **Crimes 2016 K5:L40**, and **Crimes 2016 O5:P56**.

To examine the results of the Data Interpreter more closely, we click the **Review the results** link in the Data pane to view an annotated copy of the spreadsheet.

Here we see a copy of the original data, color coded to identify which data was

identified as header data and which data was identified as field values.

The next tab shows us the sub-tables that Data Interpreter found, outlined by the cell

Location (city)	Location (state)	Months Apr	Months Jun	Months Jul	Months Aug	Months Sep	Months Oct		state	Total Crimes 2016		State	Population 2016
Albuquer	New Mexico					46			Alabama	12		Alabama	4860545
Anaheim	California		4						Alaska	26		Alaska	741522
Anchorag	Alaska	1				26			Arizona	132		Arizona	6908642
Arlington	Texas				17				California	515		Arkansas	2988231
Atlanta	Georgia					85			Colorado	64		California	39296476
Aurora	Colorado					16			D.C.	105		Colorado	5530105
Austin	Texas				28				Florida	210		Connectic	3587685
Bakersfiel	California		22						Georgia	85		Delaware	952698
Baltimore	Maryland						230		Hawaii	9		District of	684336
Boston	Massachusetts					28			Illinois	536		Florida	20656589
Buffalo	New York					38			Indiana	151		Georgia	10313620
Chandler	Arizona					3			Kansas	10		Hawaii	1428683
Charlotte-	North Carolina		25						Kentucky	95		Idaho	1680026
Chicago	Illinois						536		Louisiana	127		Illinois	12835726
Chula Vist	California	2			1				Maryland	230		Indiana	6634007
Cincinnati	Ohio					50			Massachu	28		Iowa	3130869
Cleveland	Ohio					89			Michigan	221		Kansas	2907731
Colorado :	Colorado			15					Minnesot	26		Kentucky	4436113
Columbus	Ohio					70			Missouri	223		Louisiana	4686157
Corpus Ch	Texas		9						Nebraska	29		Maine	1330232
Dallas	Texas				118				Nevada	128		Maryland	6024752
Denver	Colorado				33				New Jerse	86		Massachu	6823721
Detroit	Michigan	5				221			New Mexi	46		Michigan	9933445
Durham	North Carolina						30		New York	290		Minnesot	5525050
El Paso	Texas					14			North Car	82		Mississipp	2985415
Fort Wayn	Indiana					34			Ohio	217		Missouri	6091176

Key for the Data Interpreter | Crimes 2016 | **Crimes 2016_subtables** | A04_H84 | K05_L40 | O05_P56

In this example the first sub-table, **Crimes 2016 A4:H84**, has the main data that we want to work with. To use this table as our data table, we can simply drag the original table off the canvas and then drag the new table to the canvas.

Once we have the data that we want to work with in the canvas, we can do some additional clean up on the data. For example we can:

- Change the field names so that they represent city, state, and month names.
- Pivot the months fields.
- Drag in the third sub-table **Crimes 2016 05:P56** and join it to our first sub-table on the **State** field to include state populations for our analysis.
- Hide any duplicate fields that were added as a result of the join.

The results might look something like this:

Now we are ready to start analyzing our data in Tableau.

When Data Interpreter is not available

The Data Interpreter option might not be available for the following reasons:

- **The data source is already in a format that Tableau can interpret:** If Tableau Desktop doesn't need extra help from Data Interpreter to handle unique formatting or extraneous information, the Data Interpreter option is not available.
- **Many rows or many columns:** The Data Interpreter option is not be available when your data has the following attributes:
    - Data contains more than 2000 columns.
    - Data contains more than 3000 rows and more than 150 columns.
- **The data source is not supported:** Data Interpreter is only available for Microsoft Excel, Text (.csv) files, PDF files and Google Sheets. For Excel, your data must be in the .xls or .xlsx format…

Tableau - Book1 - Tableau license expires in 14 days

File   Data   Server   Window   Help

Connections   Add

Consumption (1)
Text file

Files

☑ Cleaned with Data Interpreter
Review the results. (To undo changes, clear the check box.)

▦ Consumption (1).csv
▦ Consumption.csv

New Union
New Table Extension

▱· Consumption (1)

Connection ◉ Live   ○ Extract

Filters 0 | Add

Consumption (1).csv

Consumption (1).csv ▾   6 fields 16599 rows   100 → rows

States	Regions	Latitude	Longitude	Dates	Usage
Abc Consumption (1).csv	Abc Consumption (1).csv	⊕ Consumption (1).csv	⊕ Consumption (1).csv	▤ Consumption (1).csv	# Consump
Punjab	NR	31.5200	75.9800	02-01-2019	
Haryana	NR	28.4500	77.0200	02-01-2019	
Rajasthan	NR	26.4500	74.6400	02-01-2019	
Delhi	NR	28.6700	77.2300	02-01-2019	
UP	NR	276000	78.0500	02-01-2019	

Name
Consumption (1).csv

Fields

Type	Field Name	Physical Table	Remote Fiel...
Abc	States ▾	Consumption (1).csv	States

Data Source   Sheet 1

Marked.Consumption (1).csv.2126184443616400 - Microsoft Excel

Home   Insert   Page Layout   Formulas   Data   Review   View

Calibri 11 General

R14

	A	B	C	D	E	F	G
		Key for Understanding the Data Interpreter Results					
		Use the key to understand how your data source has been interpreted.					
		To view the results, click a worksheet tab.					
		Note: Tableau never makes changes to your underlying data source.					

Key:
Data is interpreted as column headers (field names).
Data is interpreted as values in your data source.
Data derived from an Excel merged cell is interpreted as value in your data source.
Data is ignored and not included as part of your data source.
Data has been excluded from your data source.
Note: To search for all excluded data, use CRTL +F on Windows
  or Command F on the Mac, and then type '***DATA REMOVED***'.

If the Data Interpreter has interpreted the Tableau data source incorrectly, close the spreadsheet,
  and then clear the Cleaned with Data Interpreter check box from the Data Source page.
If the Tableau data source continues to be interpreted incorrectly or for general information
  about why some data was removed by the Data Interpreter, refer to
  Resolving Common Issues with Data Interpreter Results
Help Tableau improve the Data Interpreter by emailing your file to support@tableau.com
  or filing a support request with an attached file at:

  http://tableau.com/support/request

Key for the Data Interpreter | Consumption _1__csv