



---

# How to prepare your data for Tableau with Analytics Canvas

## Get through the clutter

Data in the real world is dirty. It can be incomplete, containing errors and outliers, and it can be inconsistent. To better visualize your data, you need to remove the clutter and uncover the relevant data. And each time you work with a new dataset, you have to repeat the data preparation process. It is a time-consuming and error-prone task.

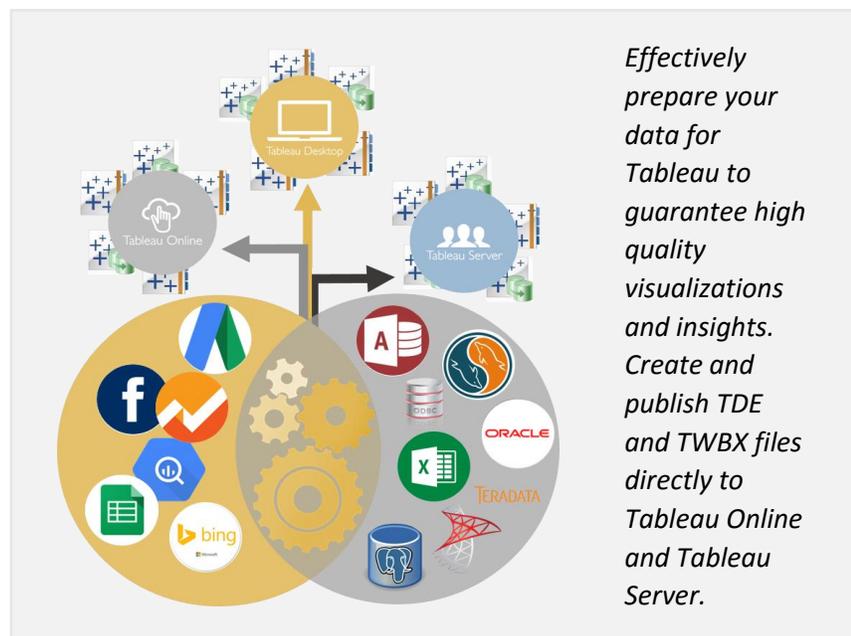
If you are like most analysts, you really need to improve this process, so you can focus on what you are hired to do - analyze and visualize the data, to produce insights!

In this whitepaper we will tackle the following issues:

1. Exploring and normalizing data
2. Blending datasets
3. Transforming datasets
4. Removing and correcting missing values, and working with inconsistent and bad data

A few words on [Analytics Canvas](#), just in case you are not familiar with it. It is a data preparation (ETL) tool that helps to connect you to various data sources, extract data, transform it, and prepare it for analysis. Analytics Canvas:

- Provides a visual interface for data preparation.
- Automates the process of data preparation when the dataset changes.
- Provides tools that ensure the data is updated consistently, eliminating errors.



*“Analytics Canvas with its Tableau connector brings out the best in the visual thinking culture. Less tinkering and wasting time validating numbers, more data discovery and insightful conclusions. Business intelligence has never been so simple and affordable too!”*

Analytics Canvas User

## Exploring and normalizing data

### Example 1. Data exploration

Some visualizations in Tableau, such as the scatter plot, are sensitive to the scale of the data. If you have some outliers with range spans in the millions, while the rest of the values are in the dozens, then the larger scale will influence the outcome. If discovering outliers is not the purpose of your visualization, you may want to quickly review the dataset to identify outliers and decide what to do.



*In Analytics Canvas’s Data Profiler tool you can rapidly explore large datasets to identify any problems with the data. Just click on the Data Profiler icon to find out the minimum and maximum values, averages, and cardinality. You can also find out if you have any missing rows or null rows, and get an idea of overall data quality.*

Data Type	Null Rows	Missing Rows	Populated Rows	Completeness	Cardinality	Uniqueness	Min
Text	0	0	100000	100%	100000	100%	

### Example 2. Data normalization

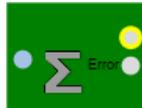
Often, variables measured on different scales do not contribute equally to the visualization. For example, a variable that ranges between 0 and 100 will outweigh a variable that ranges between 0 and 1. Visualizing this data without normalization provides little value. But transforming data to comparable scales can solve this problem.

Calculation Block

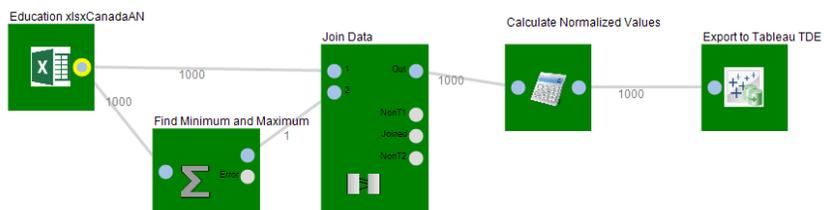


*Using a Calculation block in Analytics Canvas, add a new calculation and standardize the data by converting it to, for example, percentile values.*

Summarize 1



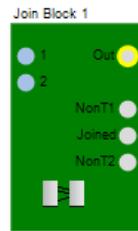
*You can also use a Summarization block to calculate minimum and maximum values, and rescale the data using the formula  $x_{new} = x - x_{min} / x_{max} - x_{min}$*



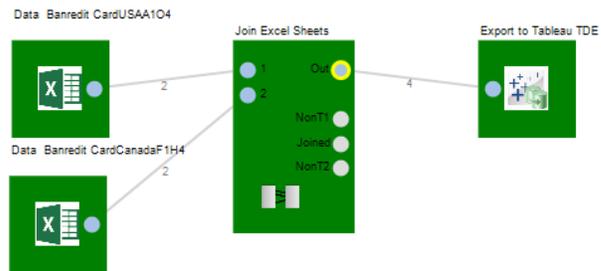
## Blending datasets

### Example 1. Joining tables

Sometimes data stored in different tables really belongs together in one table, in a single column.

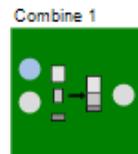


*You can use a Calculation block to give the columns the same name and a Join block to join data in one table.*

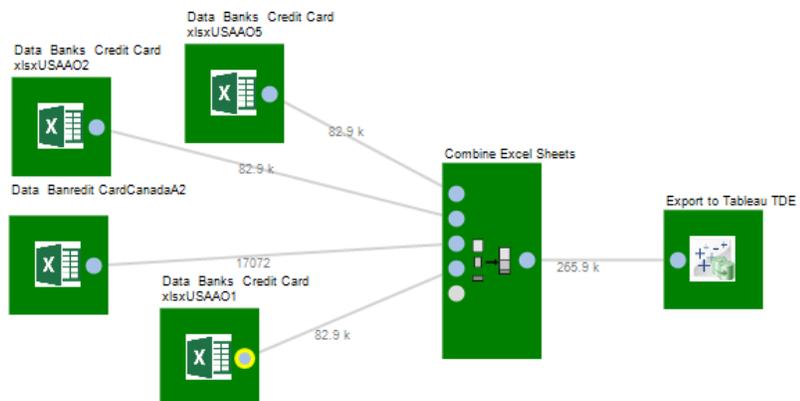


### Example 2. Combining datasets that don't have a common key

Usually you can join datasets by a common key, such as the customer ID or date. However, in some cases you will have no common key to join by.



*In Analytics Canvas, you can combine the datasets using a Combine block. Missing data will be filled with null values.*

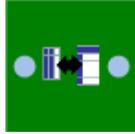


## Transforming datasets

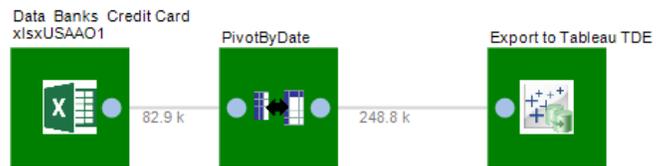
### Example 1. Transposing and pivoting columns and rows

It's not uncommon to see two very different datasets which have a different structure. Sometimes you need to make the rows into columns, so you can make sense of the data in Tableau.

Pivot 1



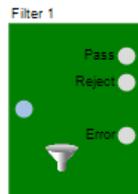
*In Analytics Canvas, you can use the Pivot block to transpose the entire table, or pivot columns.*



## Cleansing data

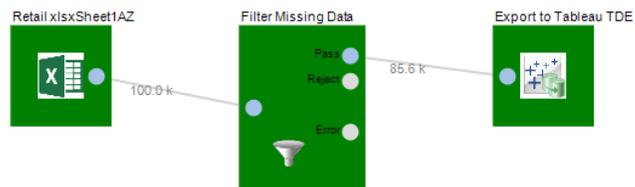
### Example 1. Missing values

Often, datasets can be missing values, which results in inaccurate visualizations. Values can be missing for various reasons. The default method for treating missing values is to include in the analysis only those cases without any missing values.



*In Analytics Canvas, it's easy to remove rows with the missing data. Using the Filter block, you can remove values that are missing in one easy step.*

However, removing missing values can impact the analysis. It is important to understand this impact and determine if conclusions from a subset will be valid or not.



Another practical method is filling missing values with some reliable value - for example, if cost and sales price values are missing from the dataset.

*In Analytics Canvas, using the Calculation block, you can replace null values with other values, such as an average, median, or mode. The principal consideration here is to get a valid statistical inference on the imputed data, not a generalization.*

Calculation Editor

#### Define the Calculation for the Column

Column Name

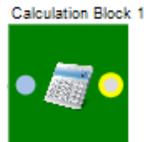


```
ISNULL([Original.Grand Total],[Average Grand Total])
```

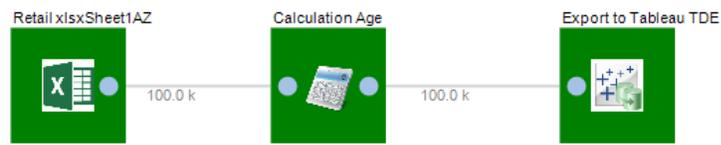


## Example 2. Inconsistent data

Often, datasets contain inconsistent data. This happens frequently in the joining of different datasets. For example, in one dataset you may have the customer's age, and in another, their birth date.



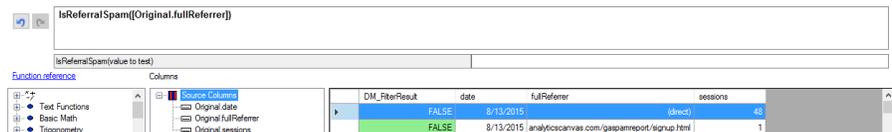
*In Analytics Canvas it's easy to compute new values, using functions. In this case, use the "year()" function in the Calculation block to find the year the customer was born, and then subtract one from the other to obtain a consistent column.*



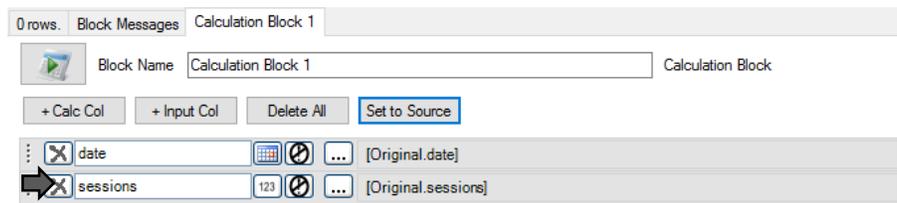
## Example 3. Bad data

Some data is simply not needed for analysis, and you may want to remove errors and outliers.

*Clean your data quickly with the Filter block to remove rows. For example, in web analytics, use the "IsReferralSpam()" function to filter out referral spam.*



*An unnecessary column can be removed by adding a Calculation block, and deselecting the column that needs to be removed.*



*“When you connect Google Analytics to Tableau directly, you soon realize there are certain limitations in analyzing the API data extracts of many profiles, and just like that, Analytics Canvas made all my headaches go away. Phenomenal data processing power, great usability with friendly tech support to boot, it facilitated data consolidation, simplified updates via automation and, most importantly, allowed me to seamlessly combine accurate insights with Tableau's data visualization prowess.”*

Analytics Canvas User

## **See Analytics Canvas in action**

---

For more information, please go to [analyticscanvas.com/tableau](https://analyticscanvas.com/tableau) and download a trial. Analytics Canvas comes with pre-build examples and corresponding Tableau dashboards that show data preparation process in action.

Don't let data preparation slow you down. Clean your data quickly and efficiently with Analytics Canvas.