



# CONNECTING TO DATA

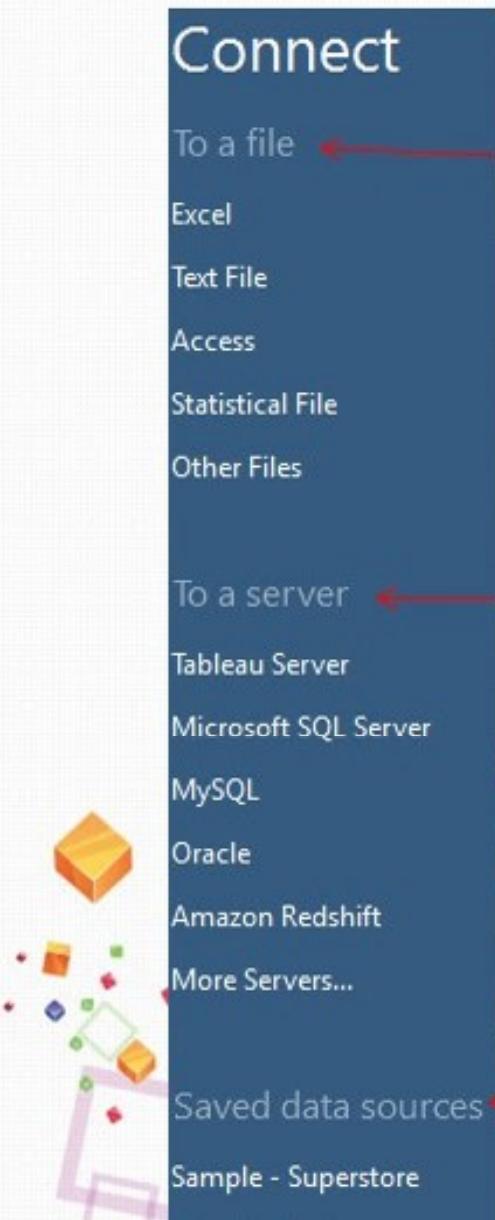
# BASIC CONCEPT

To begin building a view and analyzing your data, first connect Tableau to your data. Your data can be as simple as a Microsoft Excel workbook, or as elaborate as a SQL Server or Oracle data warehouse.

After connecting your data and setting up the data source in Tableau Desktop, the data fields become available on the Data Source page and in the Data pane on the left side of the sheet. This section describes the types of data supported and how to create and maintain a basic connection.



# TABLEAU CAN BE CONNECTED TO:



To A File like Excell, CSV, etc. This option is for connecting to locally stored data or file based data.

To A Server like Tableau Server, MySQL, SQL Server, etc. This option links to data stored in a database or a cloud service.

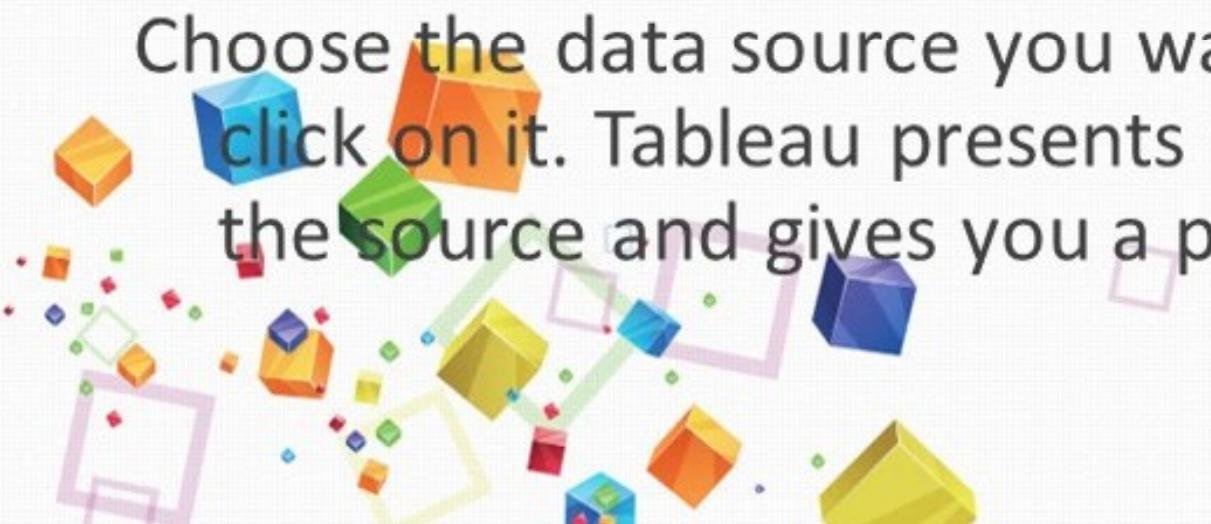
Saved Data source Files (.tds) are found on computer's Hard disk in the data sources directory under the My Tableau Repository. It will have all the alterations you have made in the data source like filters, etc.

# WHY CONNECT TO A DATASOURCE?

To do any kind of data analytics you need to have a question- for which you are seeking an answer and the data- to assist.

To build views of your data or to perform analysis, you must first connect Tableau to a data source.

Choose the data source you want to connect to and click on it. Tableau presents the different tables in the source and gives you a platform to work on.



# TABLEAU WORKSPACE- DESIGN

The screenshot shows the Tableau workspace design interface. At the top, there's a menu bar with File, Data, Server, Window, Help, and some icons. Below the menu, the title bar shows "My Product Excell". A red oval highlights this title, and an arrow points from it to the text "The file name you have connected to." To the left, under "Workbook", it says "Connected to Excel" and lists "My Product Excell.xlsx". Under "Sheets", there's a text input field "Enter sheet name" and a list of four sheets: "ActualData", "Planned Data", "PRODUCTData", and "Store". A red oval highlights the "Store" sheet, and an arrow points from it to the text "The different sheets in the data source". In the center, there's a large white area with the text "Drag the sheets here to join the tables" and "Drag sheets here" below it. At the bottom of this area, there are buttons for "Sort Fields", "Data source order", "Show aliases", "Show hidden fields", and a "Rows" dropdown. At the very bottom, there's a preview grid labeled "Data Grid".

The file name you have connected to.

Connected to Excel

Workbook

My Product Excell.xlsx

Sheets

Enter sheet name

ActualData

Planned Data

PRODUCTData

Store

Drag the sheets here to join the tables

Drag sheets here

Show aliases

Show hidden fields

Rows

Data Grid

The different sheets in the data source

# WORK ON SINGLE TABLE

## Workbook

My Product Excell.xlsx

## Sheets

Enter sheet name

- ActualData
- Planned Data
- PRODUCTData
- Store

ActualData

#	Abc	#
ActualData	ActualData	ActualData
Shops	Productid	Week Of S
2,071	FO3590981	
2,071	FS3486711	
2,071	FO3590981	
2,071	FO3590981	
2,071	FO3346481	
2,071	FS3351881	

The different columns and rows  
in the sheet i.e the preview of  
the data.

Data Source

Sheet1

Workbooks to create visualization

# Types of Data Mapping in Tableau

- Relationship
- Join
- Blend
- Cross Database Join
- Union



# Create Relationship in Tableau

- With the recent Tableau 2020.2 release, relationships have brought new capabilities in Tableau. Relationship is a flexible and easy way to combine data from different tables for your analysis. Relationships are defined on matching fields, so that during analysis, Tableau brings in the right data from the right tables at the right aggregation—handling level of detail for the analysis. A data source with relationships acts like a custom data source for every viz, but you only build it once.



A vibrant, abstract illustration featuring numerous 3D-style geometric shapes like cubes, spheres, and pyramids in various colors (blue, red, yellow, green, orange) scattered across a white background. Some shapes are larger and more prominent, while others are smaller and appear to be falling or floating. The overall effect is dynamic and modern.

# JOINING TABLES IN TABLEAU

# CREATING JOINS IN TABLEAU

Seldom will your data source include every bit of information you need in a single table. As long as data resides in a single spreadsheet or database and each table includes unique identifiers that tie the table or tab together, you can perform join of these tables within Tableau.

Database joins can be complex, but the basic concept is to bring together related information in your view.

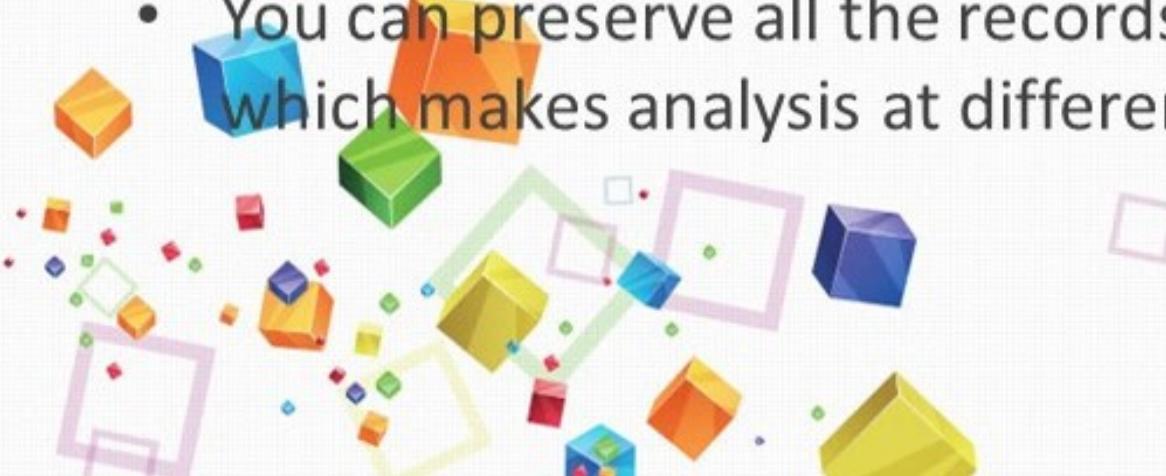
In Tableau you can define join when you make initial connection or add them later.

A default Inner Join is always created in Tableau.

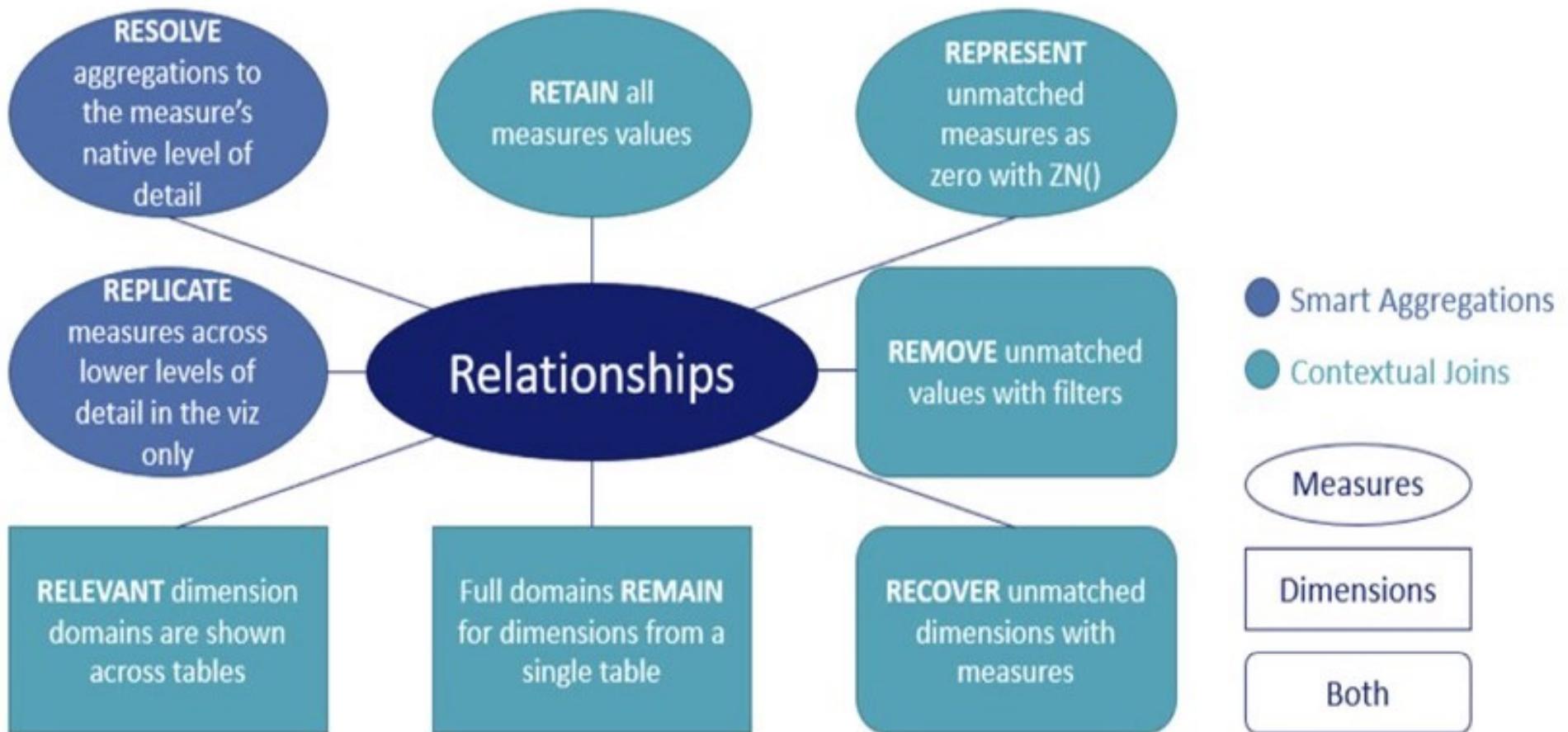


## RELATIONSHIPS HELP IN FOLLOWING WAYS:

- Less data preparation required in the initial stages; like in custom SQL or database view.
- Relationships always preserve all measures. Important values can never go missing. And unlike joins, relationships won't double your trouble by duplicating data stored at different levels of detail.
- You can preserve all the records at different level of Detail, which makes analysis at different level of granularity far easy.



# THE 8 R'S FROM TABLEAU ON RELATIONSHIP



# JOIN- DESIGN

## Workbook

ActualSales\_BMC\_Planned\_Store 3.xlsx

## Sheets

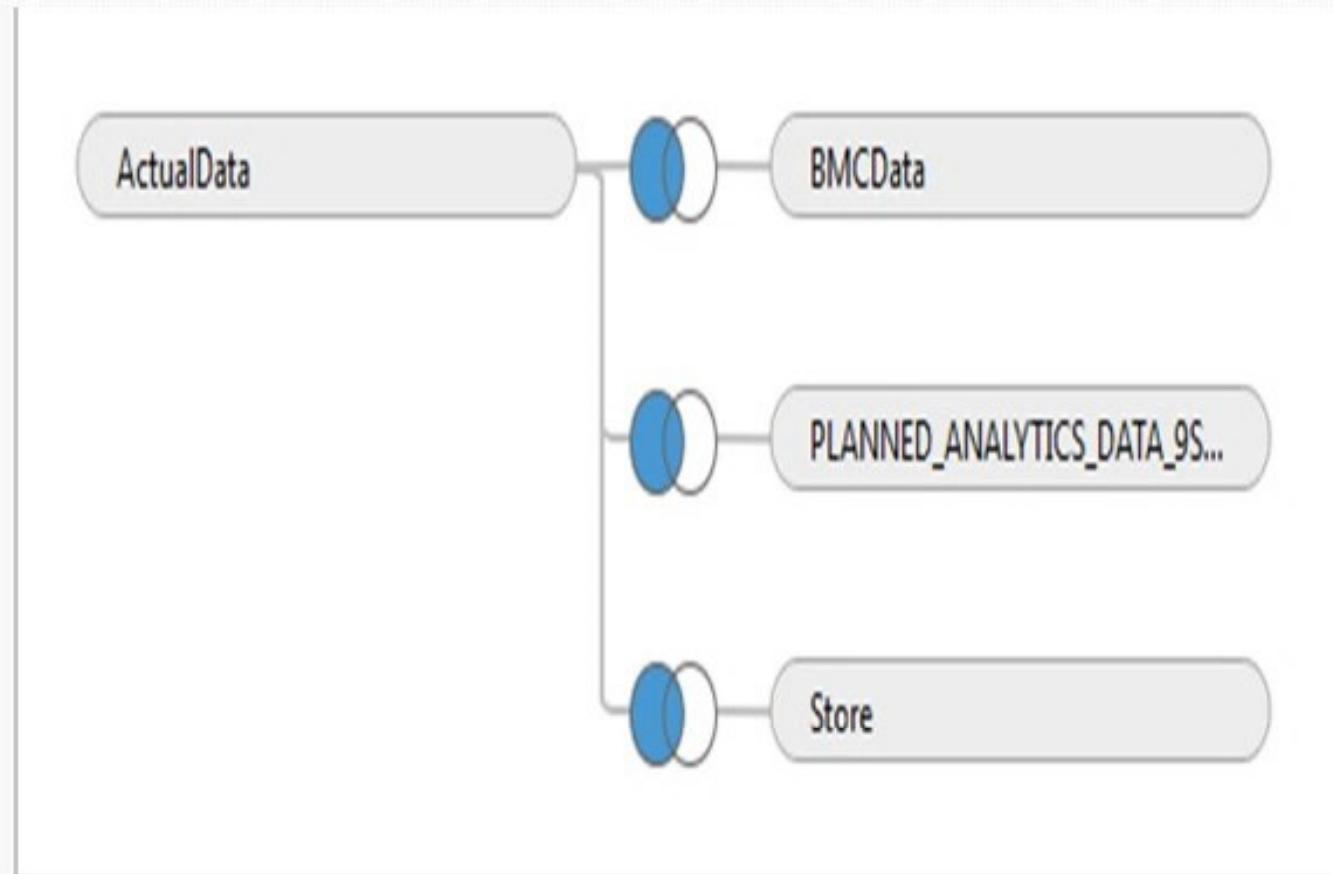
Enter sheet name

ActualData

BMCData

PLANNED\_ANALYT...\_DATA\_9STORE\_1

Store



Sort fields

Data source order



# TYPES OF JOINS

Custom\_SQL\_Query

Join

Inner      Left      Right      Full Outer

Data Source      products

prod\_id      =      Prod Id (Products)

Add new join clause

The screenshot shows a data integration interface. At the top, a grey rounded rectangle labeled "Custom\_SQL\_Query" is connected by a line with two overlapping circles to another grey rounded rectangle labeled "products". Below this, a white rectangular box titled "Join" contains four options: "Inner" (two overlapping circles), "Left" (one blue circle overlapping one white circle), "Right" (one white circle overlapping one blue circle), and "Full Outer" (two overlapping circles). To the right of the "Join" box is a red "X". Below the join types is a table with two columns: "Data Source" and "products". In the "Data Source" column, there is a small icon of a computer monitor and the text "prod\_id". In the "products" column, there is a small icon of a bar chart and the text "Prod Id (Products)". At the bottom left, there is a decorative graphic of colorful 3D cubes.

# INNERJOIN

When you join two tables Tableau defaults to “Inner Join” type.

Using an **“Inner Join”** returns only records that match in both left and right tables.



# LEFT JOIN & RIGHT JOIN

The **LEFT JOIN** will give all the records of **left table** and the matching records of the **right table**.

The **RIGHT JOIN** will give all the records of **right table** and the matching records of the **left table**.

When a value in one table doesn't have a corresponding match in another table, you see a **null** value in the data grid.



# OUTER JOIN

This type of Join returns **all the records of both the primary table and the joined table.**

All the unmatched rows will be filled be 'null' values.



# Quiz

Q.- How to bring various tabs of excel together for visualization?

Q.- Which is the default join in Tableau?

Q.- Which are the types of join in Tableau?



# DATA BLENDING

Data Blending is when you blend data from multiple data source to a single worksheet. The data are blended on a common dimension.

Data Blending does not create a row level join and is not a way to add new dimensions or rows to your data.

Data Blending should be used when you have related data in multiple data sources that you want to analyze in a single view.

Data Blending creates a default left join, but if the two dimensions do not have the same name, then, you can define a custom relationship that creates correct mapping between fields.



## Primary Data source

Data	
	Airline Delayed Arrivals in Dat...
<b>Dimensions</b>	
	Airline
	AirlineBik
	CancellationCode
	Cancelled
	Carrier
	CarrierDelay
	Date
	Departure Delay
	DepDel2S
	Dest
	DestAirportSeq
	DestCityName
	DestState

## Secondary Data source

Data	
	Airline Delayed Arrivals in Dat...
	Orders (staples - 6.xls)
<b>Dimensions</b>	
	Call Center
	Customer Name
	Customer Segment
	Customer State
	Date
	Employee Name
	Line Id
	Order Id
	Order Priority
	Order Status
	Prod Type1
	Prod Type2
<b>Measures</b>	
	Customer Balance
	Discount
	Quantity

## Blended Data

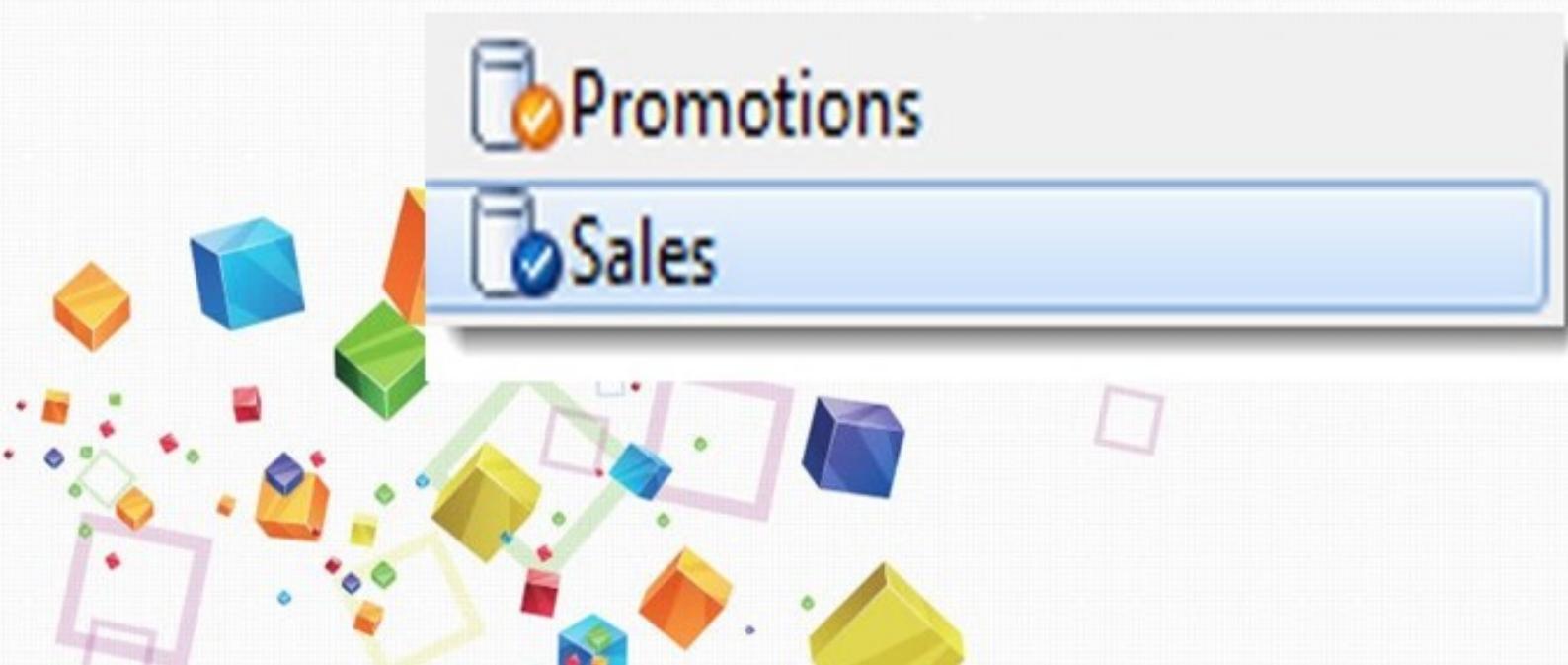
Customer Segment is an attribute of State.

Because there are multiple segments per state, an asterisk displays.

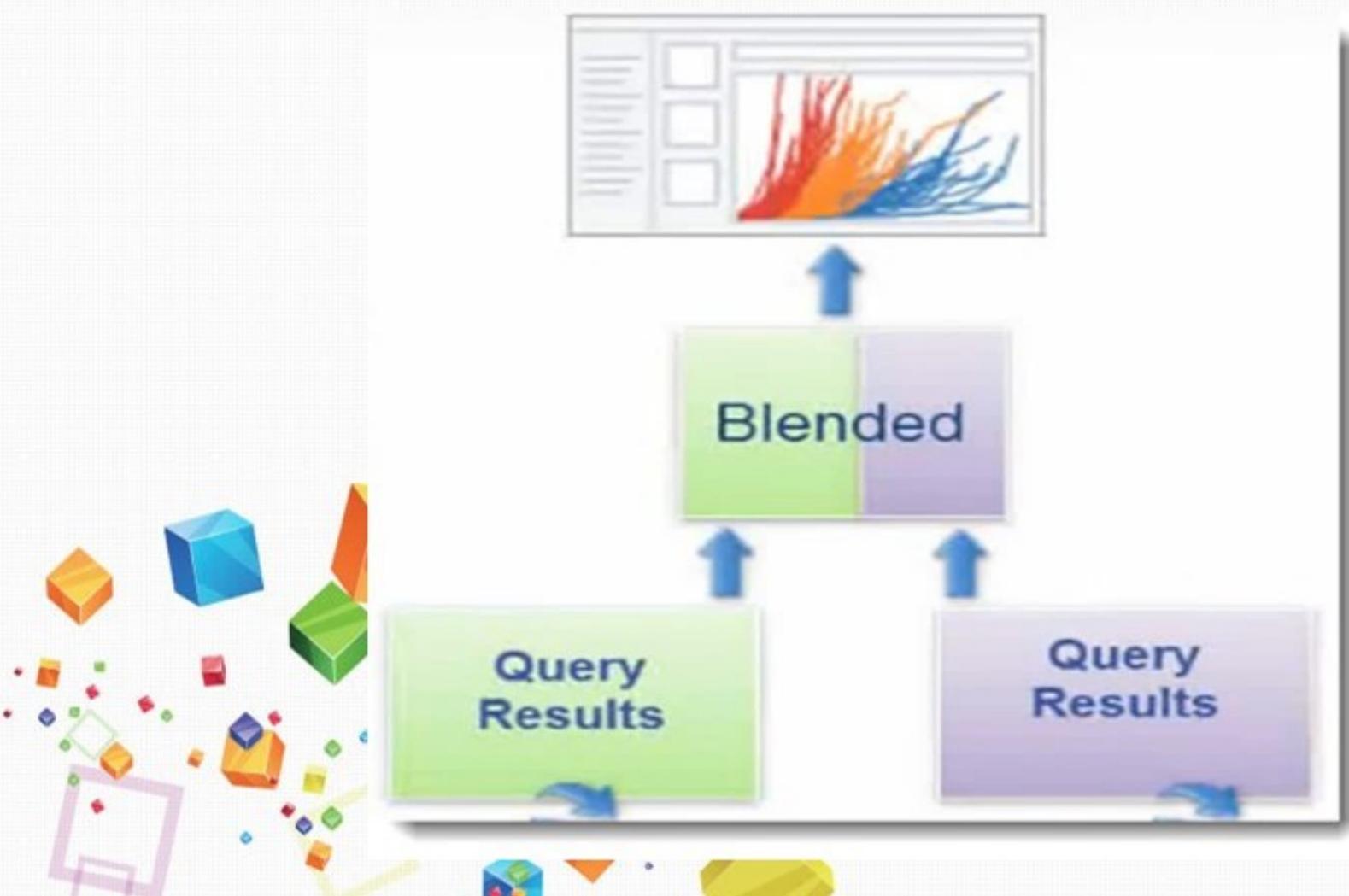
Data		
	Airline Delayed Arrivals in Dat...	
	Orders (staples - 6.xls)	
<b>Dimensions</b>		
	Abo Call Center	
	Abo Customer Name	
	Abo Customer Segment	
	Customer State	
	Date	
	Abo Employee Name	
	Line Id	
	Order Id	
	Abo Order Priority	
	Abo Order Status	
	Drop	
<b>Columns</b>		
Rows	DestState	
	Customer Segment	
DestState	Customer Segment	Drop
AK	Null	Abc
AL	*	Abc
AR	Null	Abc
AZ	*	Abc
CA	*	Abc
CO	*	Abc
CT	*	Abc
DC	*	Abc

# TIPS FOR DATA BLENDING

1. In the data window, the blue tick indicates the Primary data source (Sales in the image below). An orange tick indicates a secondary data source (Promotions).



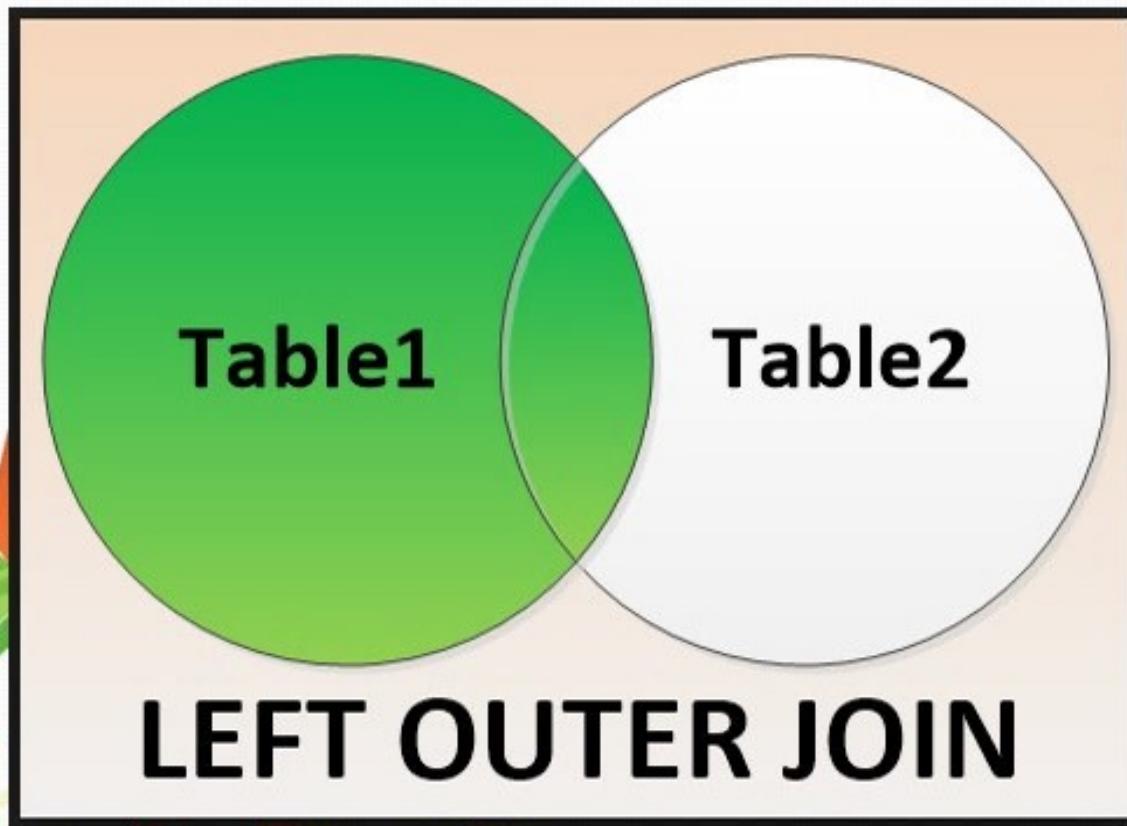
2. Blending runs a query on each data source and joins the aggregated results of those queries.



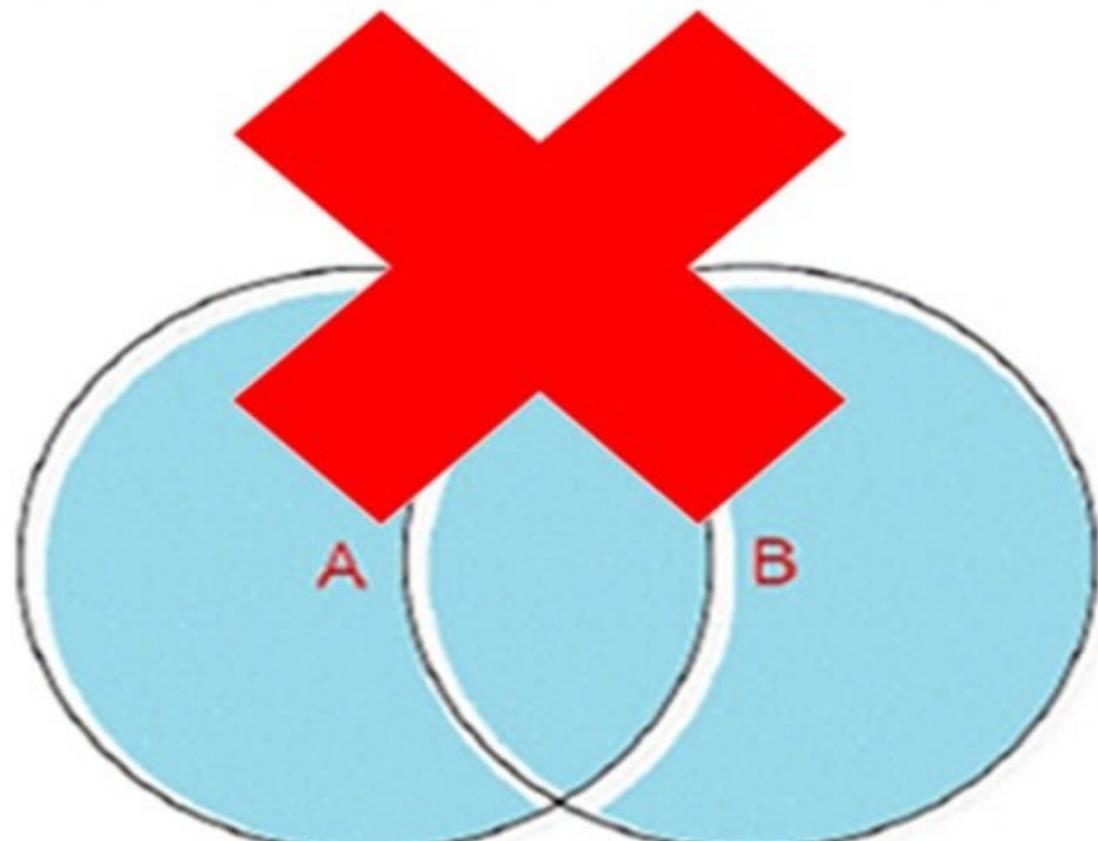
3. The first field you drop onto a worksheet defines the primary data source. Any other data source you use will be a secondary data source. The secondary source fields are shown on shelves with the orange tick marks.



4. A default blend is equivalent to a left outer join. However, by switching which data source is primary, or by filtering nulls, it is possible to emulate left, right and inner joins.



5. You cannot emulate a full outer join using blending.

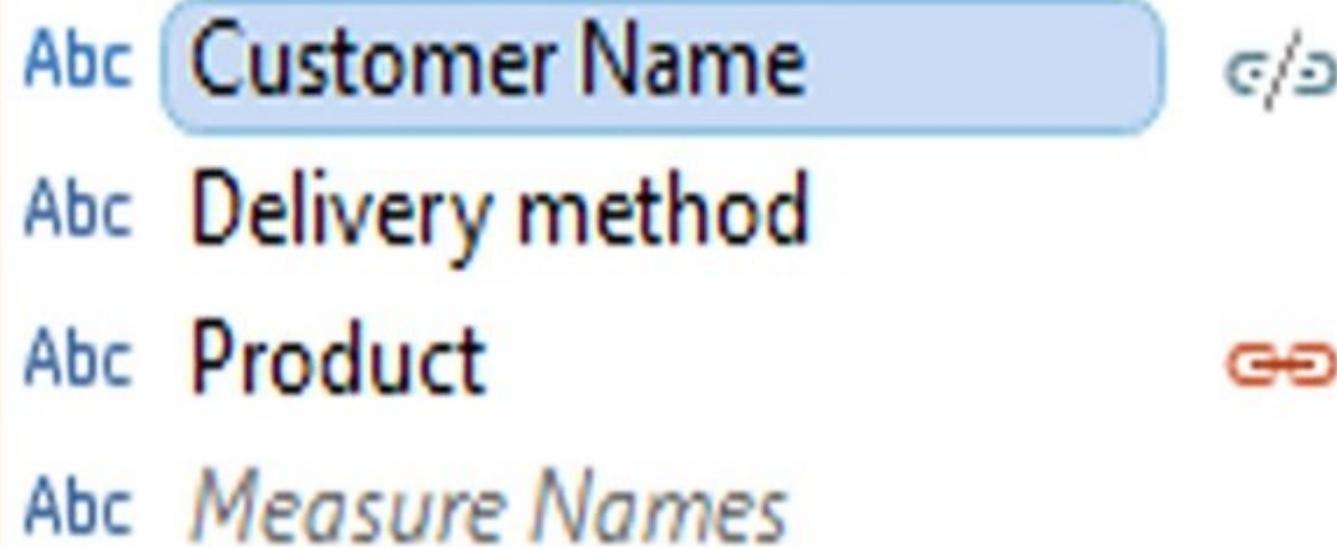


6. Sometimes you see NULL values when you blend. This is when there is no corresponding row based on the join key in the secondary data source. In the example below, Patient is in the primary data source. Illness is from the secondary data source. Sam and Will have no data in the secondary data source.



Illness	Patient	
Null	Sam	04/01/2014
	Will	05/01/2014
Asthma	Sarah	03/01/2014
Coeliac disease	Celine	03/01/2014
Diabetes	Douglas	03/01/2014
	Matt	02/01/2014
Ischaemic hear..	Morgan	04/01/2014

7. If you do not override the default choice, Tableau will automatically switch the blend field to the most appropriate for the view.



Abc **Customer Name** C/D

Abc **Delivery method**

Abc **Product** GO

Abc **Measure Names**

8. The \* means that the primary and secondary data sources do not share data that enables you to create the view you have attempted to draw. The \* is an indication that your data sources do not contain enough information to blend and display the view you have created.



# BLENDING vs. JOINS

When trying to determine whether to join tables or use data blending, consider the following:

- \* Where your data is coming from
- \* How many connections you have in your workbook
- \* The number of records you have in the data



# GENERAL GUIDELINES

If a workbook uses **data from more than one data source**, you can use **data blending**. Blends are more fragile than joins. They persist only on the worksheet on which they are created.

If a workbook uses **two data connections from the same data source, joining the data** tables can improve performance and filtering control. If you are adding data to an existing data source or a table in the same data connection, using a join is better.



# Cross database join

- When related data is stored in tables across different databases, you can use a cross-database join to combine the tables.
- To create a cross-database, you must create a multi-connection Tableau data source. You do so by adding and then connecting to each of the different databases (including Excel and text files) before you join.



# If cross data base join is not Proper

- Before you join tables using a cross-database join, make sure that the data types of the fields in the join key match. If the data types of the fields in the join key don't match, the join breaks, which is indicated by a red exclamation point..



# How to fix a broken join in Cross DB Join

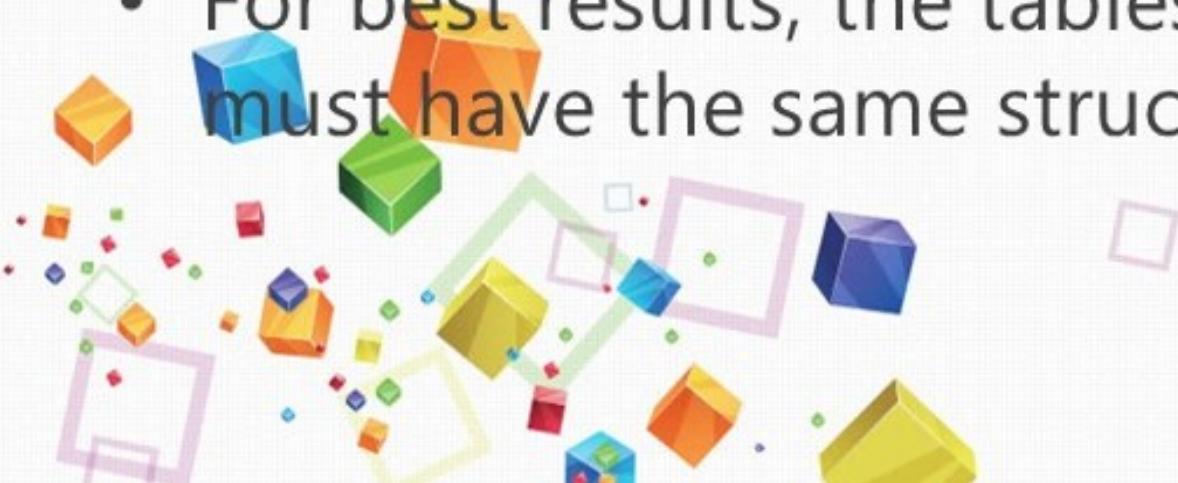
To fix a broken join, take one of the following steps:

- Use a calculation to resolve type and format mismatches between fields in the join key.
- For text or Excel-based data, you can modify the data type of one of the text or Excel fields in the join key using the data type menu in the data grid.
- For most connections, you can use the **Convert to custom SQL** option to change the data type of one of the fields in the join key. The **Convert to custom SQL** option is available only when the data source contains only one connection. In this case, remove the second connection and then select **Data > Convert to custom SQL**



# New Union

- You can union your data to combine two or more tables from Excel workbook or text file data.
- The workbooks or text files must be in the same folder and come from the same connection.
- You union tables by appending values (rows) from one table to another.
- For best results, the tables that you combine must have the same structure



## New Union: Quick Review

- A union table can be used in a join.
- A union table can be used in a join with another union table.
- The fields generated by a union, **Sheet name and Table name, can be used as the join key**.
- If a named range is used in union, null values display under the **Sheet name field**.



## New Union: Quick Review (cont.)

- The field generated from a merge can be used in a pivot or split.
- The field generated from a merge can be used as a join key.
- The data type of the field generated from a merge can be changed.
- Union tables from within the same connection.

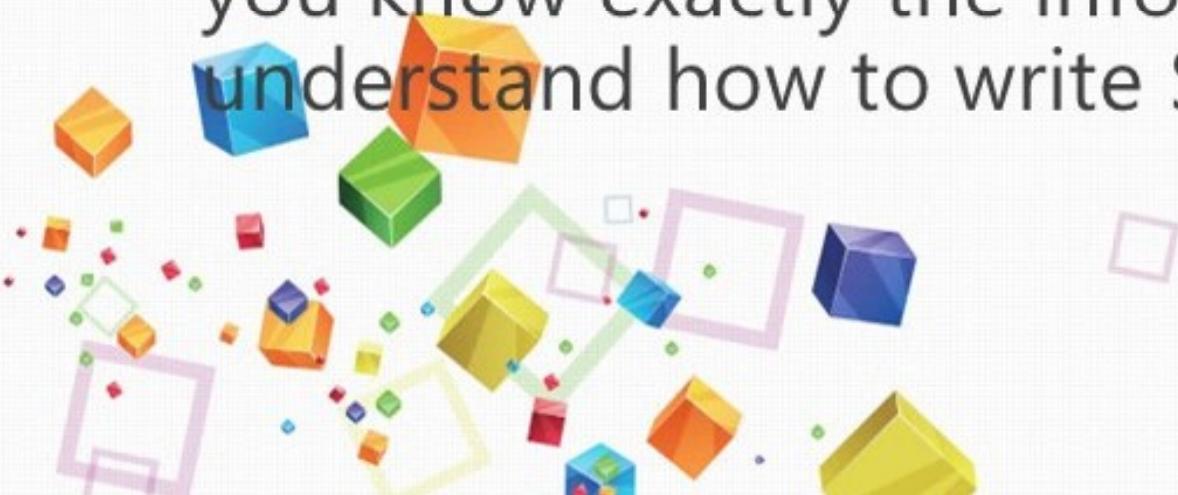


# QUIZ

- Q.- What are the various connectivity that Tableau offers?
- Q.- How to bring various tabs of excel together for visualization?
- Q.- Which is the default join in Tableau?
- Q.- Which is the default join in Blend?
- Q.- What is the color of Primary Data source and Secondary Data source?
- Q.- To connect to a multiple tables in a single data source at one time, \_\_\_\_\_ must be specified.
- 

# New Custom SQL

- For most databases, you can connect to a specific query rather than the entire data set.
- Because databases have slightly different SQL syntax from each other, the custom SQL you might use to connect to one database might be different from the custom SQL you might use to connect to another.
- However, using custom SQL can be useful when you know exactly the information you need and understand how to write SQL queries.



# Uses of Custom SQL

- **Combine your tables vertically (union)**
- If you need to append data to each other, you can use the union option directly in Tableau.
- In some cases your database does not support this option, so you can use custom SQL instead.



# Uses Of Custom SQL (cont.)

- **Change data type of a field to do a cross-database join**
- When you want to perform a join between two tables, the data type of the fields you join on must be the same.
- In cases when the data type of the fields are not the same, you can use custom SQL to change the data type of the field before performing the join.



# Uses of Custom SQL (cont.)

- **Reduce the size of your data**
- When working with very large data sets, sometimes you can save time while working with your data if you reduce its size first.



# Uses Of Custom SQL (cont.)

- **Restructure your data (pivot)**
- In some cases, you might be working with a table that needs to be restructured before analysis.
- Though this type of task can be done directly in Tableau by using options like pivot, your database might not support it. In this case, you can use custom SQL instead.



# Uses Of Custom SQL (cont.)

- **Combine joins & aggregate your data**
- If you need to combine tables and aggregate your data, you can use both a join and default aggregation type options directly in Tableau. In some cases you might need to use custom SQL instead.



# Error when duplicate columns are referred in Custom Query

- If your custom SQL query references duplicate columns, you may get errors when trying to use one of the columns in your analysis in Tableau. This will happen even if the query is valid.
- For eg. If a common field exists in both the tables that you are referring to, Tableau will connect to the query but you will get an error when you use this common field. This is because, Tableau doesn't know which table you are referring to as the same field is present in both the tables.

