UNDERSTANDING BASIC JOINS

Joins combine tables by adding more columns of data across similar row structures.
This can cause data loss or duplication if tables are at different levels of detail and joined data sources must be fixed before analysis can begin.

When **joining** tables, the fields that are joined on must be of the **same data type**.
If we change the data type after performing the join on the tables, the join will break.

Fields used in the join clause cannot be removed without breaking the join

UNDERSTANDING BASIC JOINS

Data blending simulates a **traditional left join**.
The **main difference** between a **Join** and **Blend** is when the **aggregation** is performed.
A **join** first combines the **data at a row level** and then carries out the **aggregation**.
A **blend aggregates** and **then combines the data**.

In most scenarios, traditional table joins are the best choice as they tend to be most performant as the work is done by the database rather than the local machine.
Table joins are better when tables have a 1:1 relationship (i.e., there is only one record for each value in the linking fields in each table).
If the tables have a 1:many or many: many relationship this creates duplicated data after the join
Duplicated data can be resolved after the table join is created, but the increased amount of data may cause performance issues.

DETAILS OF THE DATA SOURCE USED FOR JOINS

We will be using a fictious spreadsheet Understanding_Joins.xlsx having 2 tables viz Projections and Actuals for explaining joins

Projections Table

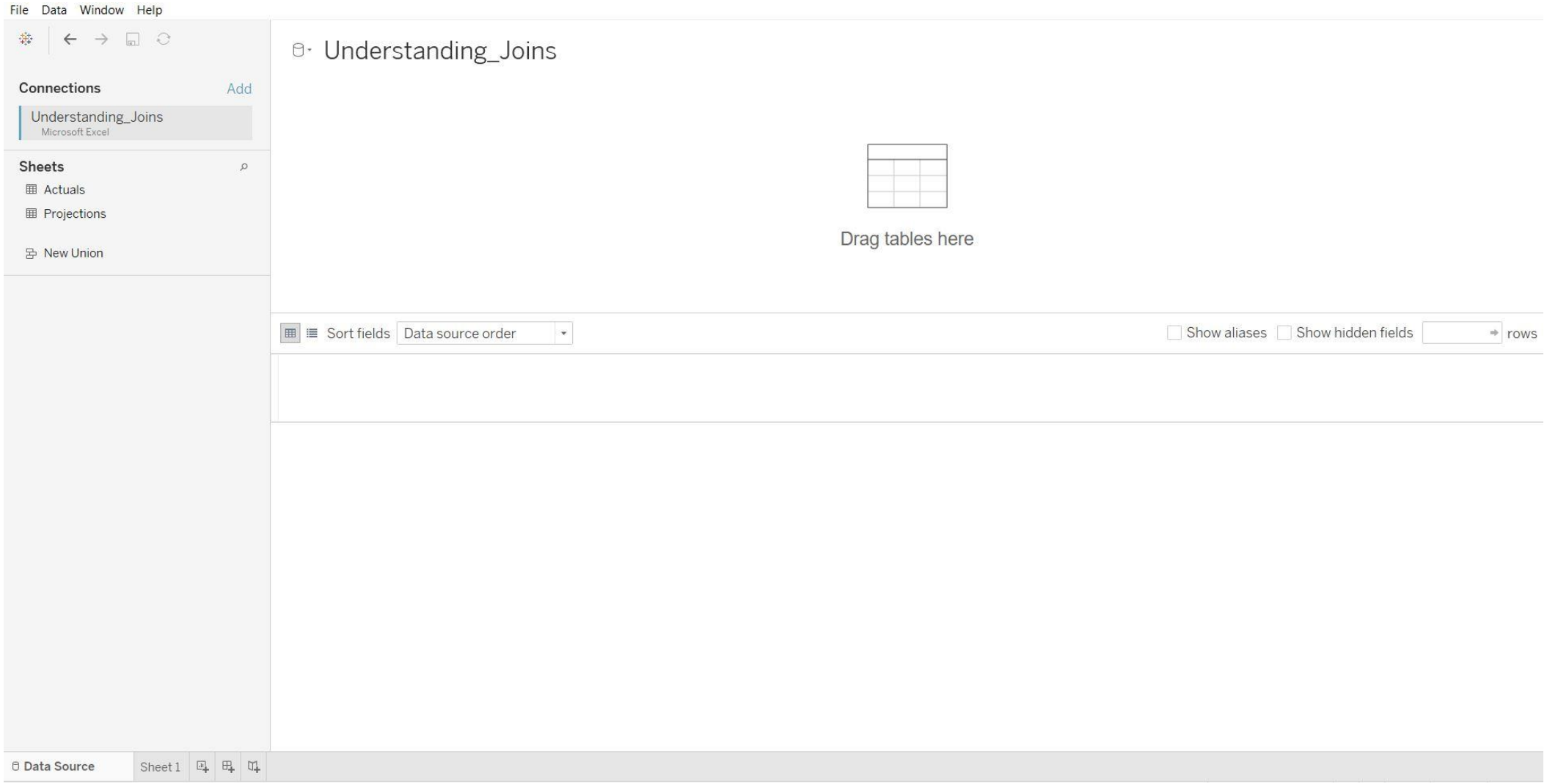
Team	Place	Projected Score
Team A	Place A	225
Team B	Place B	345
Team C	Place C	175
Team D	Place D	267
Team E	Place E	300
Team H	Place H	145
Team I	Place I	254

Actuals Table

Team Name	Place Name	Actual Score	Entered By
Team A	Place A	175	Person 1
Team B	Place B	300	Person 2
Team C	Place C	225	Person 3
Team D	Place D	190	Person 4
Team E	Place E	256	Person 5
Team F	Place F	335	Person 6
Team G	PlaceG	195	Person 7

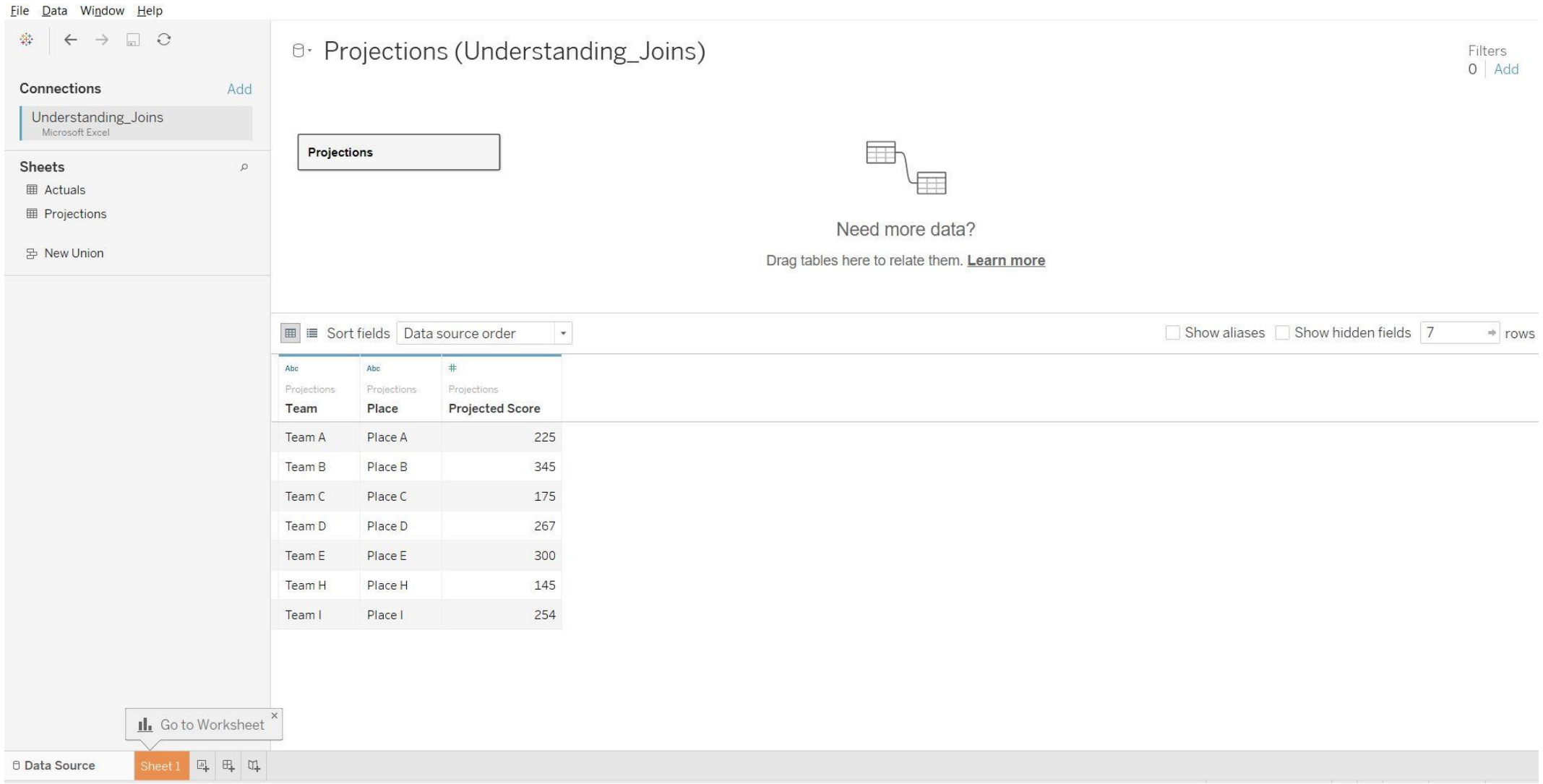
STEPS TO CREATE A JOIN

Step 1: Connect to the relevant data source or sources e.g.: Understanding_Joins.xlsx



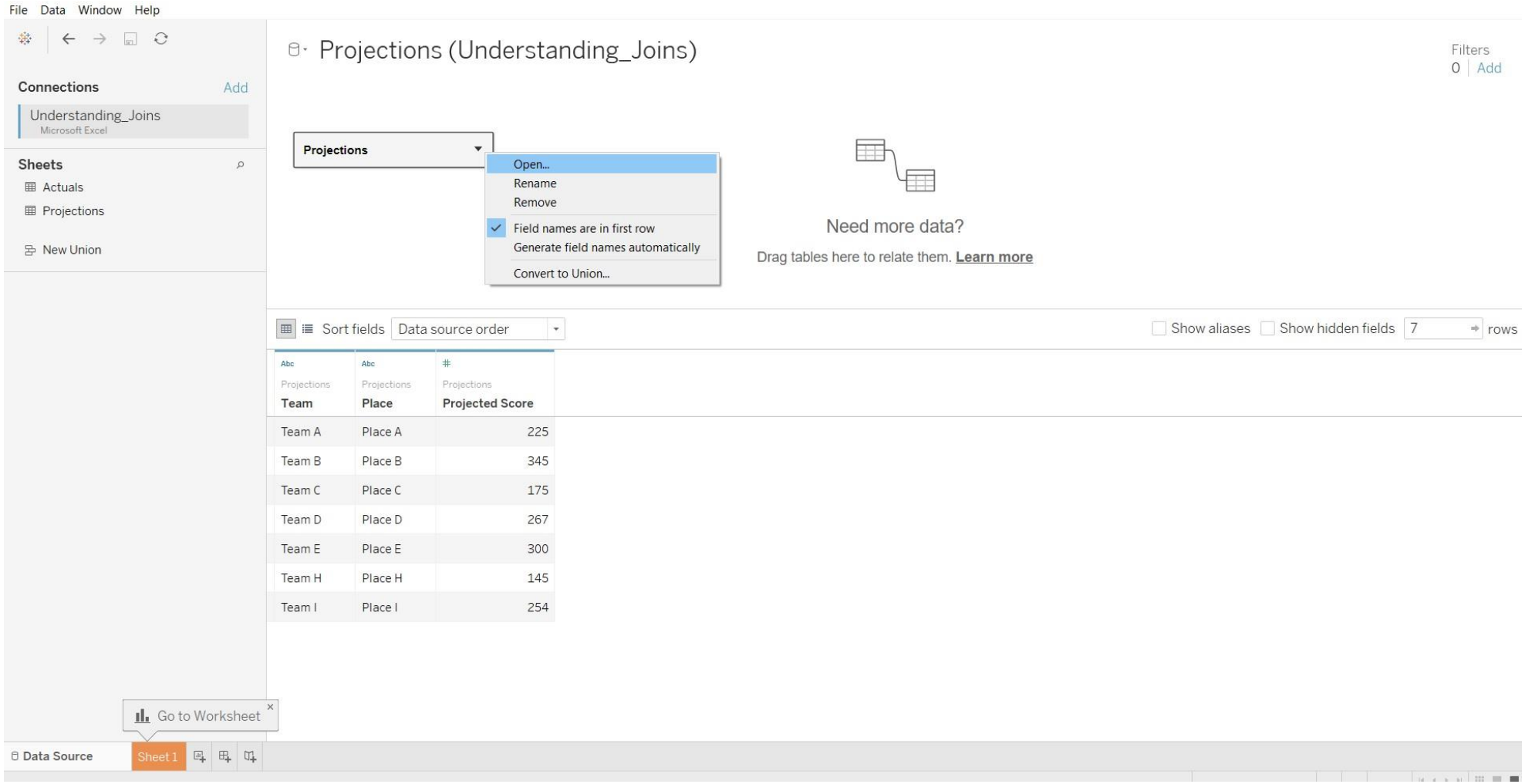
STEPS TO CREATE A JOIN

Step 2: Drag the first table **Projections** to the canvas



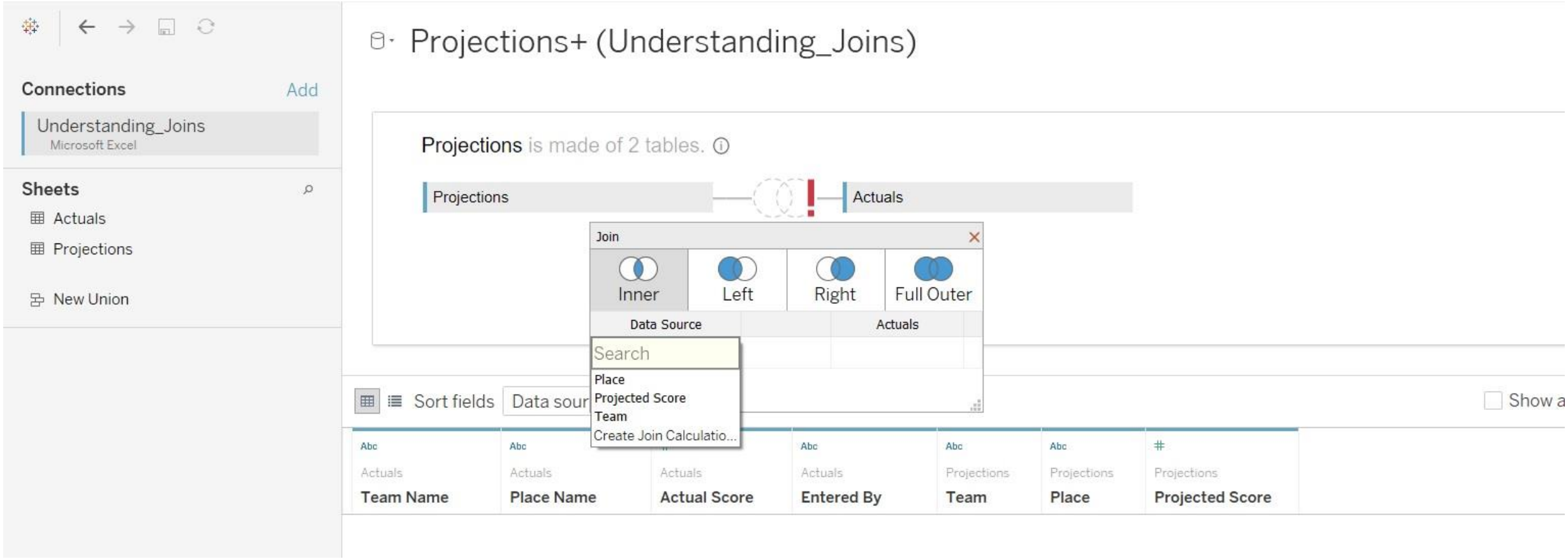
STEPS TO CREATE A JOIN

Step 3: Select **Open** from the menu or **double-click** the first table to open the join canvas (**physical layer**)



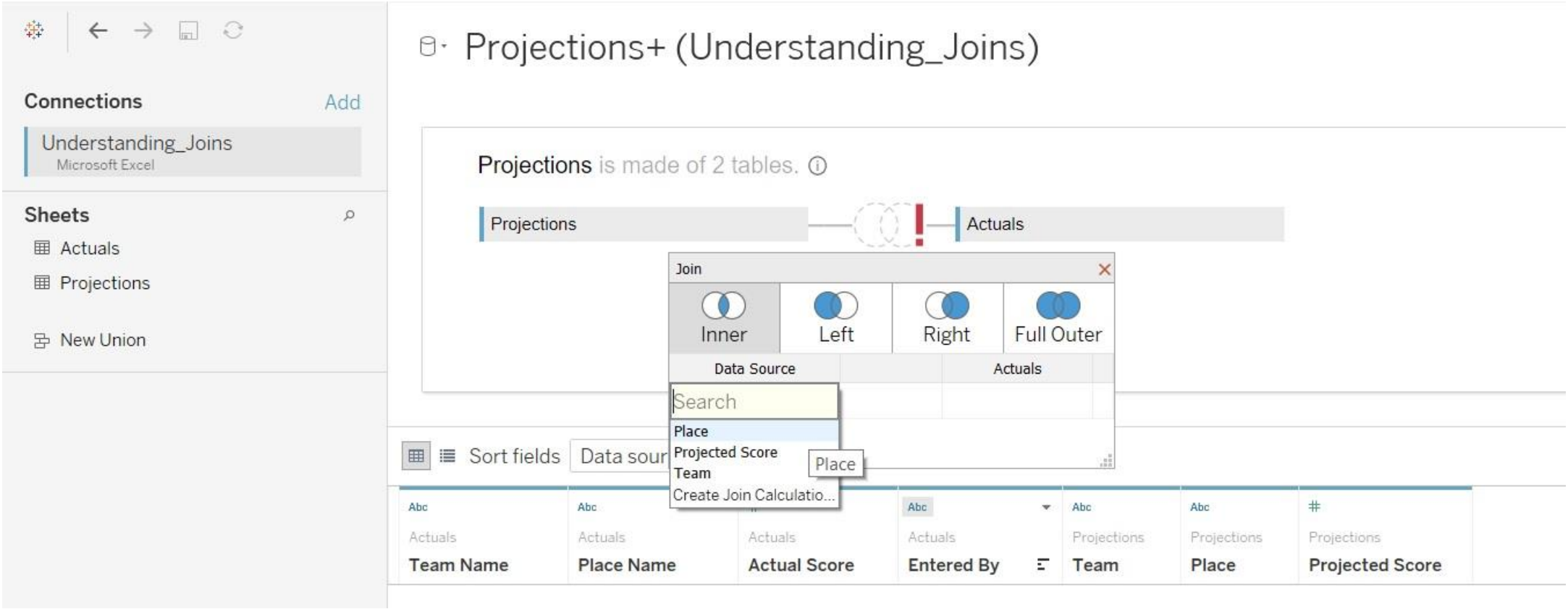
STEPS TO CREATE A JOIN

Step 4: Double-click or drag the **Actuals** table to the join canvas



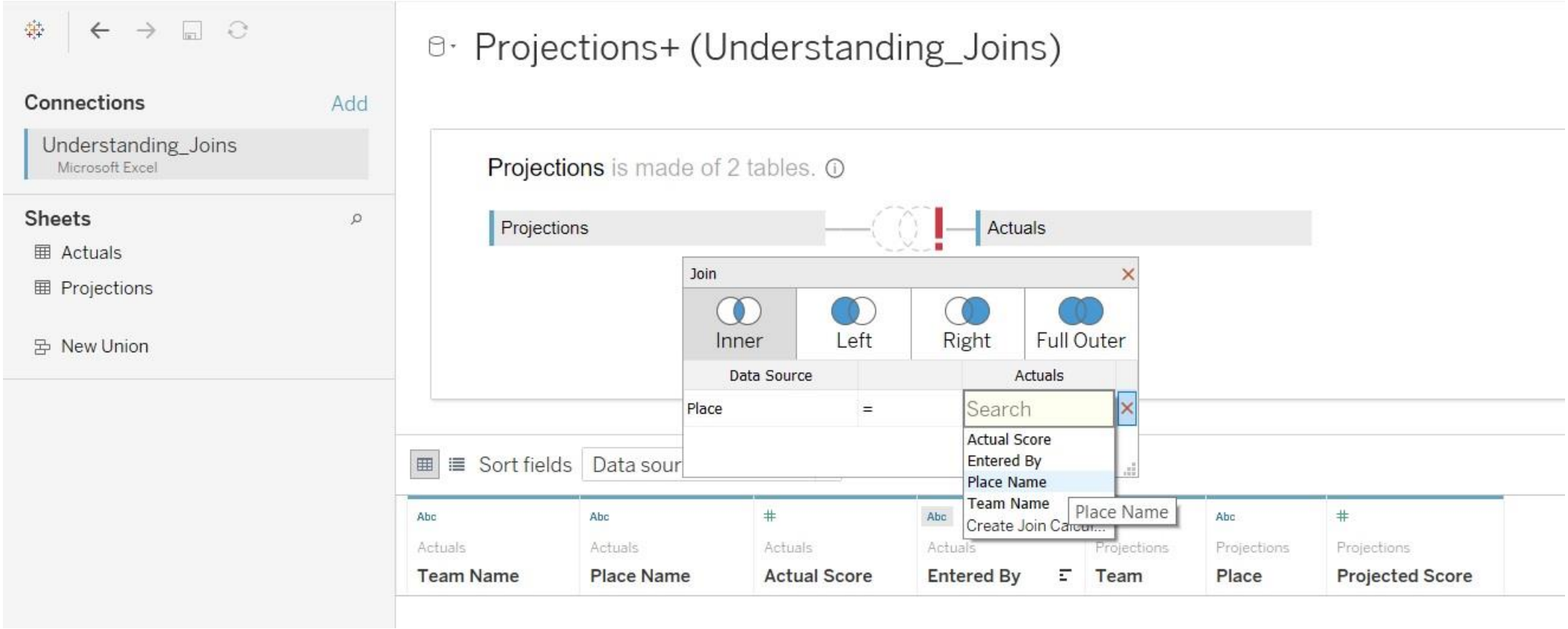
STEPS TO CREATE A JOIN

Step 5: Select the connecting field i.e., **Place** from the drop-down of Left side table



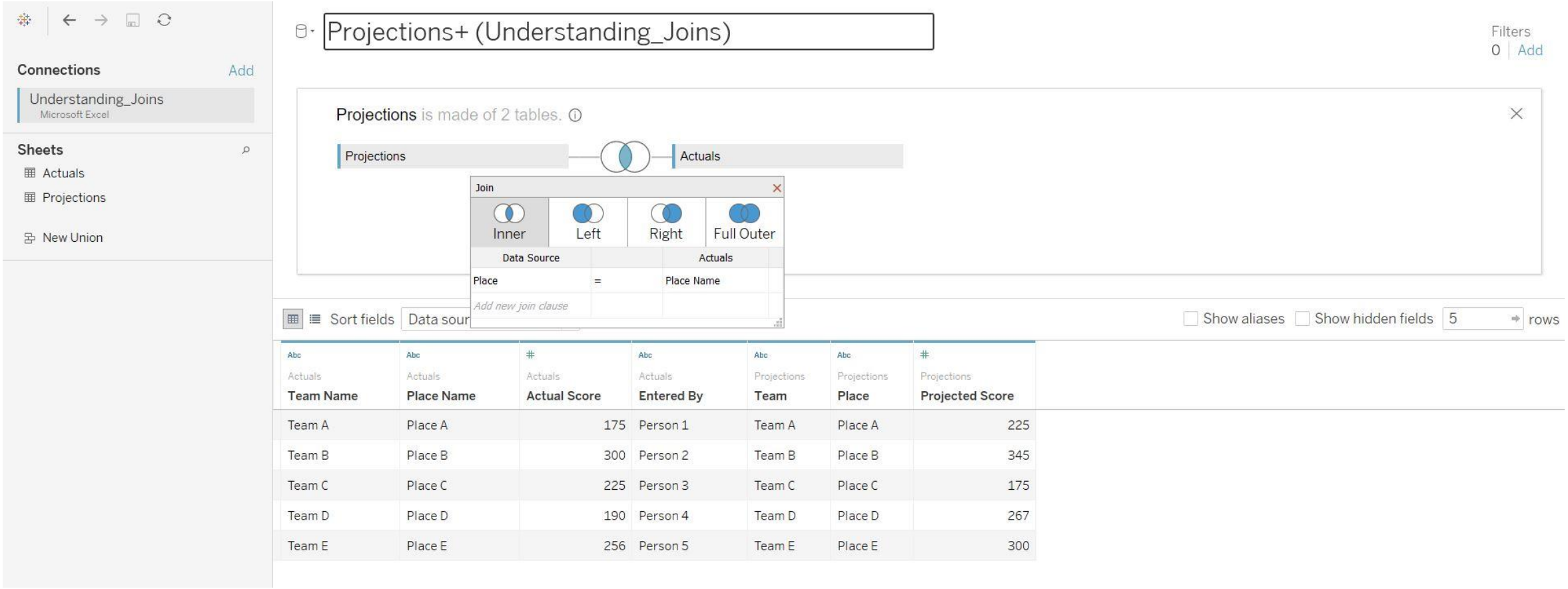
STEPS TO CREATE A JOIN

Step 6: Select the connecting field i.e., **Place Name** from the drop-down of Right-side table



STEPS TO CREATE A JOIN

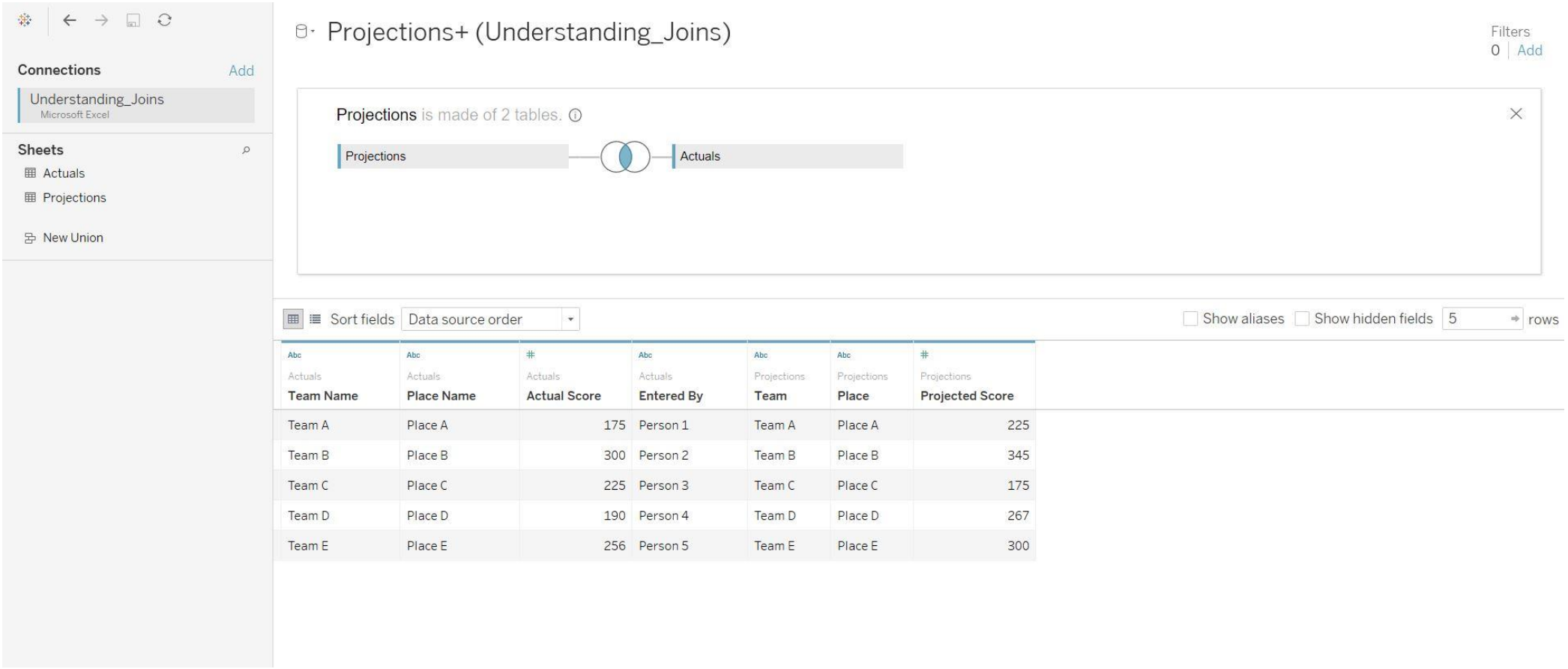
Step 7: When finished, close the join dialog and join canvas.



Actuals Team Name	Actuals Place Name	Actuals Actual Score	Actuals Entered By	Projections Team	Projections Place	Projections Projected Score
Team A	Place A	175	Person 1	Team A	Place A	225
Team B	Place B	300	Person 2	Team B	Place B	345
Team C	Place C	225	Person 3	Team C	Place C	175
Team D	Place D	190	Person 4	Team D	Place D	267
Team E	Place E	256	Person 5	Team E	Place E	300

STEPS TO CREATE A JOIN

Step 8: Now we see the final join created



Actuals Team Name	Actuals Place Name	Actuals Actual Score	Actuals Entered By	Projections Team	Projections Place	Projections Projected Score
Team A	Place A	175	Person 1	Team A	Place A	225
Team B	Place B	300	Person 2	Team B	Place B	345
Team C	Place C	225	Person 3	Team C	Place C	175
Team D	Place D	190	Person 4	Team D	Place D	267
Team E	Place E	256	Person 5	Team E	Place E	300

JOIN TYPES

In general, there are four types of joins that we can use in Tableau: **inner, left, right, and full outer**.

If we aren't sure what join type, we want to use to combine data from multiple tables, it is preferred to use relationships.

Not all databases support all join types. If an option is unavailable in the join dialog, it is likely due to a constraint from the data source.

Join Type	Result
Inner 	When you use an inner join to combine tables, the result is a table that contains values that have matches in both tables. When a value doesn't match across both tables, it is dropped entirely.
Left 	When you use a left join to combine tables, the result is a table that contains all values from the left table and corresponding matches from the right table. When a value in the left table doesn't have a corresponding match in the right table, you see a null value in the data grid.
Right 	When you use a right join to combine tables, the result is a table that contains all values from the right table and corresponding matches from the left table. When a value in the right table doesn't have a corresponding match in the left table, you see a null value in the data grid.
Full outer 	When you use a full outer join to combine tables, the result is a table that contains all values from both tables. When a value from either table doesn't have a match with the other table, you see a null value in the data grid.
Union 	Though union is not a type of join, union is another method for combining two or more tables by appending rows of data from one table to another. Ideally, the tables that you union have the same number of fields, and those fields have matching names and data types. For more information about union

SUMMARY OF JOINS

Given below are some of the important points w.r.t Joins:

1. The **visual cue** for a **join** is a **Venn Diagram**
2. We need **at least 1 data source** to create a **join**
3. We can have **multiple connections** while creating **joins**
4. The **structure** created during a join is a **new combined form**
5. **Tables** are **combined** via a **Physical join**
6. Numbers are **aggregated** based on the **join level granularity**