Cleaning & Modeling Data in Power BI

Contents

[Part 1: Clean the Data 1](#_Toc172171189)

[Part II: Set Up Relationships Between Tables 1](#_Toc172171190)

[To create the new relationship 2](#_Toc172171191)

[Sources: 4](#_Toc172171192)

# Part 1: Clean the Data

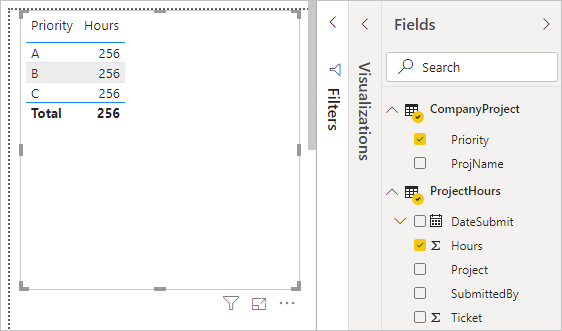
Load data from both projectdata tables (CompanyProject and ProjectHours) into PowerBI. Click on Transform data in home tab to go back to Power Query Editor after that and clean the data. Specifically…

* **Remove** **the AssignedBy column** in the *CompanyProject* table.
* **Remove duplicates** in the *ProjectHours* table.
* **Fill down the missing value** in the date column of ProjectHours or otherwise infer a value for that cell.
* Look at the **column profile** for the *project* column in *ProjectHours* and see that there is a misspelling of *Green* (Greeen). Use **Replace Values** to correct that.
* Close and Apply to load the data.

# Part II: Set Up Relationships Between Tables

Go to report view and select priority from the CompanyProject table and Hours from the ProjectHours table.

If we look at our table in the report canvas, you’ll see the number of hours is 256 for each project, which is also the total.



Clearly this number isn’t correct. Why? It’s because we can’t calculate a sum total of values from one table (**Hours** in the **Project** table), sliced by values in another table (**Priority** in the **CompanyProject** table) without creating a relationship between these two tables.

So, let’s create a relationship between these two tables.

Remember those columns we saw in both tables with a project name, but with values that look alike? We'll use these two columns to create a relationship between our tables.

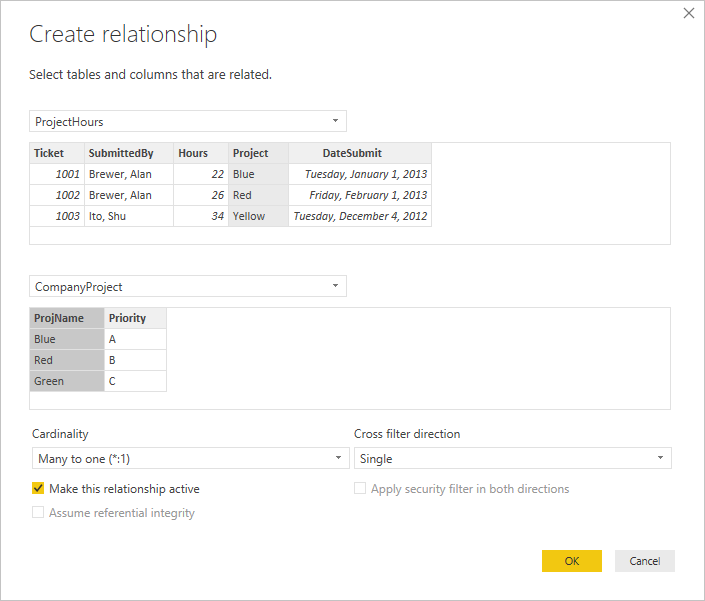
Why these columns? Well, if we look at the **Project** column in the **ProjectHours** table, we see values like Blue, Red, Yellow, Orange, and so on. In fact, we see several rows that have the same value. In effect, we have many color values for **Project**.

If we look at the **ProjName** column in the **CompanyProject** table, we see there’s only one of each of the color values for the project name. Each color value in this table is unique, and that’s important, because we can create a relationship between these two tables.

In this case, a *many-to-one* relationship. In a many-to-one relationship, at least one column in one of the tables must contain unique values. There are some more options for some relationships, which we'll look at later. For now, let’s create a relationship between the project columns in each of our two tables.

## To create the new relationship

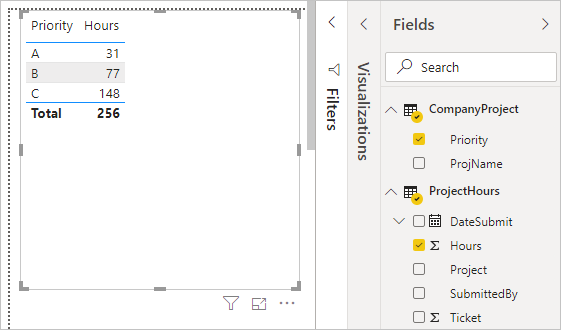
1. Select **Manage relationships** from the **Modeling** tab.
2. In **Manage relationships**, select **New** to open the **Create relationship** dialog box, where we can select the tables, columns, and any other settings we want for our relationship.
3. In the first drop-down list, select **ProjectHours** as the first table, then select the **Project** column. This side is the *many* side of our relationship.
4. In the second drop-down list, **CompanyProject** is preselected as the second table. Select the **ProjName** column. This side is the *one* side of our relationship.
5. Accept the defaults for the relationship options, and then select **OK**.



1. In the **Manage relationships** dialog box, select **Close**.

In the interest of full disclosure, you just created this relationship the hard way. You could have selected **Autodetect** in the **Manage relationships** dialog box. In fact, autodetect would have automatically created the relationship for you when you loaded the data if both columns had the same name.

Now, let’s look at the table in our report canvas again.



That looks a whole lot better, doesn’t it?

When we sum up hours by **Priority**, Power BI Desktop looks for every instance of the unique color values in the **CompanyProject** lookup table, looks for every instance of each of those values in the **ProjectHours** table, and then calculates a sum total for each unique value.

With autodetect, you might not even have to do that much.

# Sources:

Part 2 of these instructions (and the data as well, though it is modified to be dirty at the start) is shared from Microsoft. Here is the link: <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-create-and-manage-relationships>