**Lecture (Data transformation)**

It is time to do some transformations on that data to just clean it up and make sure it’s in a sensible format.

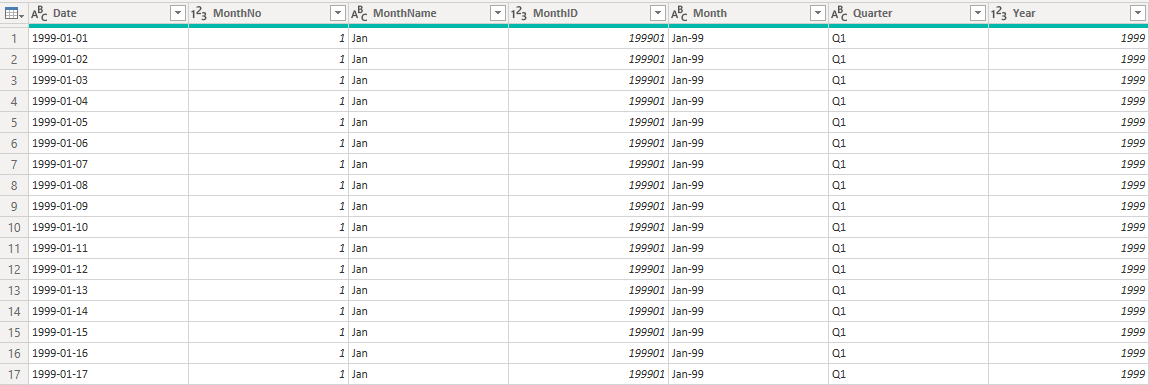
I connected to the data source. Now this means these are the different transformations or steps that have been applied from connecting to that data source, navigating to a particular table and then as I start to do things like changing the data type or splitting columns up or renaming them, they’ll get added as steps here.

So, like I said this date, table, I need to do some simple things in here so let’s just rename, sorry, duplicate, this date column to start with, when I right click on a column I get a whole bunch of different options in terms of transformations that I can apply here. You can see I can do things like moving a particular column or removing everything else if I’ve got lots and lots of columns. We are going to go through a few more of these other ones, but I am just going to duplicate this for now. And you can see a new column gets added to the right-hand side and a new step gets applied on this right-hand pane. Now I wanted to duplicate this because the first thing I am going to do is actually change the data type of this. So, you can see, back on our home tab that this data got pulled in just as a text field. Well, actually it’s a date so let’s go and change the data type to Date and you can see the Power BI Desktop has parsed this in to a date format. I can rename this column here as well, just by double clicking and now I can start to do some more transformations and operations on this. Power BI has got an inherent understanding of how dates work and on this add column tab you can see I’ve got some different things that I can use to extract different bits of information from this date column. So if I wanted look at, you know to find the date that contained the start date of the month that each row was in, or I could do something like extracting the day of the week for this particular date, let’s take that. We will create a new column you can see; we have got an index from null to 6 for all of the different days of the weeks so we can see that the 1st January 1999 was a Thursday, Saturday rather. We’re zero indexing this so the second was a Sunday and then 1/3 was on Monday, etc. etc. I might also go and extract something similar, like quarters or which week of the year was it so those first two were in the first week of the year running up until Sunday and then the second week started on the 3rd of January. So, I can do some clever, different types of transformations using dates here. I’ve got some more tables that I want to go and change so actually our geography table is just fine. I am just going to rename that one, rename it correctly.

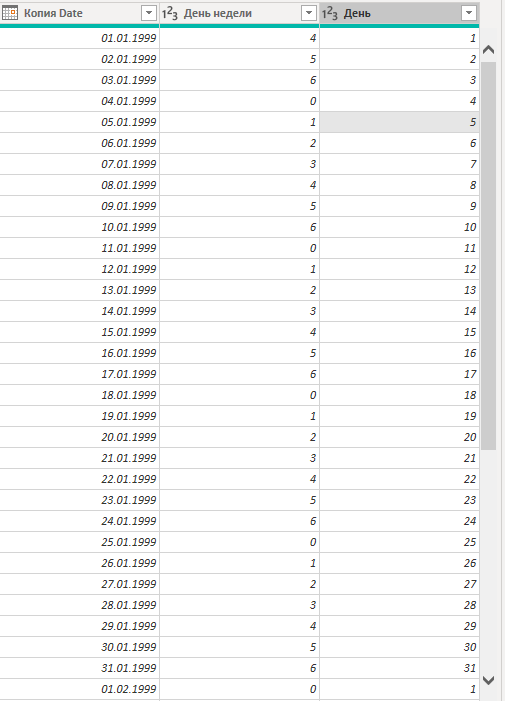
I am going to rename manufacturer table again and we have only got a couple of columns here so not very much to change and then our product table. Let’s give that a new name as well and there is a simple transformation that I want to do here as well. So, I’ve got an ID and then a product name. Now this product name contains the manufacturer and then some code and actually we already have got a column for the manufacturer so this is a little bit redundant. Let’s extract the product code into its own column. So again, I am going to duplicate this column first because I want to keep the old one around and now what I want to do is actually split this column up. I am going to right click on the column. See, there is an option to split columns and this is one of the most common operations that we see people doing. I’ve got a choice, I can either do it by a fixed number of characters if I know that the product code is always going to be a certain number of characters long. Or I can do it by a delimiter. So, delimiter is, saying, I want to split this on a comma or a semicolon or space in our case. So, I am going to choose from here the space and then I can choose, do I want to split it once at the first instance of the space, once at the right most of the last delimiter, or at each occurrence of the delimiter. So, if I’ve got a long string of text that’s got many spaces or commas in it, it would create a different column for each one of those. In this case I actually just want to do it once and I’ll hit OK. So now you can see these columns have been split up. I’ve got the manufacturer name and then the product code separate. This is obviously a bit redundant given that I’ve got the manufacturer column right next to it and I can rename this one ‘this is our product code’. So now I have been able to enhance this dataset with a little bit more information that might make it easier to build reports and do further analysis.

So, my final table here is my fact sales table, so let’s just rename this one ‘Sales Fact’ and you can see in here I’ve got keys to the products and dates when these things, these sales were made. We can see the zip code that it was made in and I’ve got just some numbers about the units and the revenue that were used and actually, there is nothing else that I need to do for that query either. So once I am happy and I’ve done all the transformations to the queries that I need, back on my home tab I’ve got this option to close and apply those changes. Now what will happen is, we’ll go and create the connection back to that data source and we’ll start loading all of that data. Previously, we were just looking at a preview, but now you can see we have pulled in the rows from these dates, geography, manufacturers tables and we are loading that Sales Fact table as well. And we are applying all of those transformations all of those different steps that we had defined

earlier and pulling that data into the Power BI Desktop. Now I am not going to wait for this to complete. We have got quite a few million rows in here and you can see it's pulling through them fairly quickly. When you go through the lab yourself, you can see how long that takes and when you connect into your own data source you can see how long that takes, too. So, let’s pick this up in the next segment.



**Date table**



**New columns**