**ASSIGNMENT 2**

**QUERY EDITOR BASICS**

**CONNECT MULTIPLE DATA SOURCES**

Power BI Desktop ships with one of the Core Tool, the Query Editor.

In this module, we'll have a look at what the Query Editor actually is.

How we do whole data preparation, data cleansing, data cleaning.

We learn how we can connect Power BI Desktop to different source types.

How we can edit rows and columns.

Transformations, i.e.,

How can we change the layout of existing tables?

How can we work with the formatting of the different columns in our tables?

How can we handle errors?

We will focus in the query editor.

Steps we perform:

1. Source file connection

2. Working with this information, cleaning data, cleansing data, transforming data.

3. So, all that you have to know about preparing information that you can finally load into the data model later on.

Let’s see what data cleansing, or data cleaning, data preparation is?

We learnt that we have different source types, different source files available, which can be connected to power BI Desktop.

Power BI Desktop establishes a connection to these source files, and we call each of these connections a so-called query.

So far, we saw a single query, a single connection only, but we basically can have an unlimited amount.

Of course, there are limitations at some point, but generally you can have a lot of different queries, a lot of different connections in power BI Desktop.

This just means extracting information from the source files and loading it into the memory of power BI Desktop.

Remember you are not able to change the actual source files.

We can only read the information from the source files and load that information into power BI Desktop.

So, this is the **first step** querying data.

The **second step** now is the real cleaning of the data.

And, cleaning has several aspects, several topics that can be summarized under this topic.

For example, cleaning data can mean that:

1. We remove duplicate or unrequired data.

2. Duplicates occur in data and source files. This is pretty common.

3. You have unrequired information, unrequired roles, or unrequired columns, which might contain information that is important in the actual source file, but for your analysis, this information might not be required. These are few things you want to get rid of, because keep in mind that our goal is to only load the information into our data model that is actually required for our data analysis. It's all about efficiency and keeping the amount of loaded data as small as possible.

4. Other things that can be summarized under cleaning data is fixing errors, is fixing topics like missing values, which might be added. So, you can also add new rows if required. And, you can also clean empty fields, because fields that don't contain information are not required in our later analysis.

5. The same applies to the format of our data. For the later analysis, it's very important that power BI Desktop understands whether we deal with text here, with numbers, with dates, and so on. So, after all the first cleaning steps, we should have only one specific data type, one specific data format in the corresponding column. Then, we can format that data.

And, with that, we have a clean structure and power BI Desktop can identify the data type.

6. Even more advanced topics when it comes to cleaning data is the combination of multiple data sources. We can establish connections to multiple source files.

We learned that, but you might want to create a table, for example, that contains information from different source files. For this, you need to perform certain operations that allow you to combine rows or columns from different tables, and this is also something that you do as part of the data cleaning process.

As a final result of all these steps, and, as I said, we have more topics to cover in this part of the query editor, we should have one or multiple tables with the final data that we need.

Then, we arrive at the **next step** which is not part of this into following modules, the actual **data analysis or visualization**, this is the part where the actual deep dive occurs where we think about what we want to analyze based on that clean information, and also what great visuals we could create based on that data.

The **last step** is not required for the next modules. As I said, we'll focus onto the first two topics here.

We have already prepared our empty project template here.

You can use the empty template file to create a new Power BI Desktop project.

In case you didn't follow the steps, use the one I shared with Google Classroom

You can download the **empty-template** and also the **three source files.**

These files will be our source data for the next steps in this module.

These are **population data for different periods**. We have data from **2010 up to 2040**

We have **three different file types**, **one or two CSV files** and **one Excel file** once again.

If you for example, have a look at the Excel file here You see that we have, **multiple columns and multiple rows.**

Here, we have a column for the **corresponding country**, we have a column for the **corresponding year**, where the **population was counted**. Then we have **different age groups**, so here zero to four, and this of course also increases also in this database. And we have **information about the male population, about the female population**, and about the **total population**.

**Remember:**

The empty-template, i.e. PBIX file is also to be downloaded and save all in the folder of your choice with the three source files.

Rename the empty template to Ass-2-query-editor-basics

It makes it easier for us to identify the project file we are currently working on.

Let's then open this basics file.

Here the regular popup does not appear as we saved this file earlier with the corrected settings.

**Step 1:**

Get started with connecting Power BI Desktop to these three source files.

Go to **Get Data** option -> Either click the button in the **upper part** or in the **lower part**.

The **lower part** here opens the **most common data sources**.

We need the **Excel workbook and the text CSV file to connect**.

To have an **overview of all available connectors**, you can either press **More,** or you click the button in the upper part.

Both of these options will open the, well, Get data window. It gives an overview of all connectors Power BI ships with. The default option is the All menu.

Here you see a list of all connectors that we have in Power BI Desktop.

There are different file connectors, which are pretty common ones like Excel, Text or also JSON.

If you want to connect Power BI Desktop to a database, you can also explore options like the Power Platform or also Azure, the Microsoft cloud service and others.

If you know the source type already, you can also use the search bar up here.

Type csv, then the option is displayed here to the right. Select it now as our first use source files are CSV files **Press Connect**.

Now navigate into the folder where you just stored the source files and select the first one, **population-2010-2019**.

With pressing **Open**, we establish a connection to the source file.

We are again here in this so-called **navigator**, which now allows us to preview the actual content of that source file, and to also define some other options.

One of these options up here is **the Delimiter.**

The Delimiter here defines the different columns that we have here in our project file.

**TRY:**

If I would set this to a **COMMA** for example, then our file would kind of break, because as you see each individual column here is separated by a comma in our source file.

You can also check this manually in this CSV file. By the way, I'll skip this step here as we have to preview already in Power BI Desktop.

So, make sure to change this to **SEMI COLON** here to get the correct preview, and therefore also the correct loading logic for our project.

In case comma is not working for you here, make sure to use the template file I just talked about or use colon delimiter that applies to your specific setting.

Here, we have the **Data Type Detection**, which here is Based on the first 200 rows. We could also enhance this to the entire dataset or to **Do not detect data types**.

The last option is actually the one that we set in our settings. If you remember, there we set the automatic type detection to zero.

**Extract Table Using Examples** is something we'll skip for the moment.

We'll, again, focus onto **Load and Transform Data**.

**Load:** Loading means that we load the data directly into the data model.

**Transform:** We'll focus onto the query-editor and we can directly access the query-editor from this navigator with this Transform Data option down here, which we'll now also select.

Here, we are now in the query-editor.

**Check:**

In Query Settings column here, we only see one **applied step** here, just Source.

Earlier, we saw that we had some automatically applied steps.

This was due to the fact that the settings were the default ones.

**Try:**

In case you cannot see this Query Settings column here,

Go to the **View ribbon**, and there **activate or deactivate** the Query Settings.

Suggested: You should keep it activated as we use this column quite a lot.

**TRY:**

Go to **Home ribbon.**

Check: In **the Query Settings**, we can also **rename** our queries if you want to.

We have this option to the right also, to give queries a different name if you want to.

Now with the first connection being established, we can also add the two other files.

For this, we of **don't have to leave the query-editor**.

Stay here -> Go to **Home ribbon** -> go to **New Source**, as we saw it in the actual data model.

In the lower part of this button, we have the most common sources.

Click it in the upper part, you again get an overview of all connectors to all source types Power BI Desktop ships with.

**The second file we want to establish a connection to is also a CSV file**.

Let's again select this option.

Press **Connect,** and let's now select **the population-2020- 2029** file.

With Open, we again see the **navigator,** but now **with different** options.

Here at the bottom, the rest remains the same that we saw for the initial file.

**Remember**:

Here, we don't have the choice between loading the data into the data model or to going to the query-editor because we are in the query-editor already.

Therefore, with **OK,** the second query is available. Here, we can now switch between these queries by simply selecting the corresponding query we want to work with.

**Again, go to New Source**.

Now select the **Excel Workbook as our first source file is, well, an Excel file.**

Now select the **population-2030-2040** file.

Press **Open** once again.

Now the process is actually the same, but you see a big difference here.

The **navigator looks different now for the Excel file**.

For the CSV file, we don't have the choice between different sheets here.

For the **Excel file, we do have this choice, therefore we don't get the full preview** immediately.

Here, as we saw it before, we simply can select the corresponding sheet.

**Tick the box** here to make sure that we **select this sheet** and load it into the query-editor, and confirm this **with OK.**

And here we now see that we also establish the connection to the third query.

We can **rename this query** now, either here in the Query Settings or as you know it from other, well, tools.

We can simply **right-click** it here and select **Rename**, and now named this **population-2030-2040**.

By clicking here, we confirm this change.

**We successfully established a connection to three different source files in Power BI Desktop.**

To make sure we don't lose this progress, you can now of course save this file here with this option up here.

So, press **Save**, and these pending changes now can be **applied later.**

**Because applying changes would mean that we load that information into the data model.**

As we didn't change anything in the query-editor so far, we can apply these changes later.

The file will still be saved though, which is the case right now.