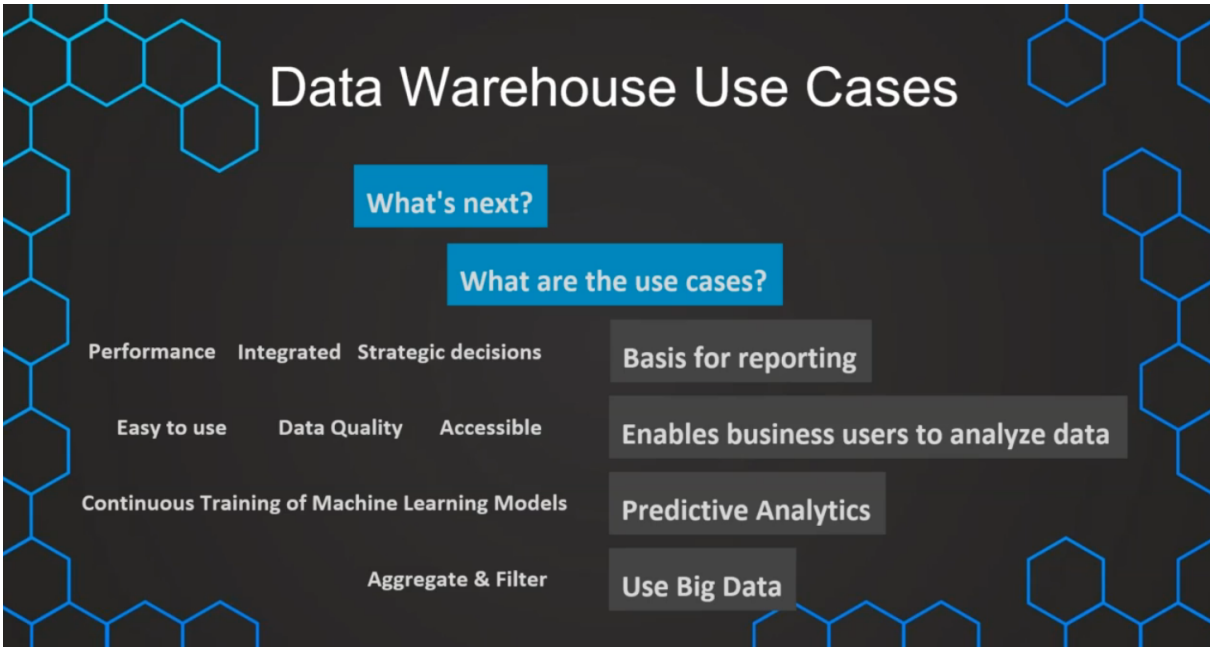
# 12. Using a Data Warehouse



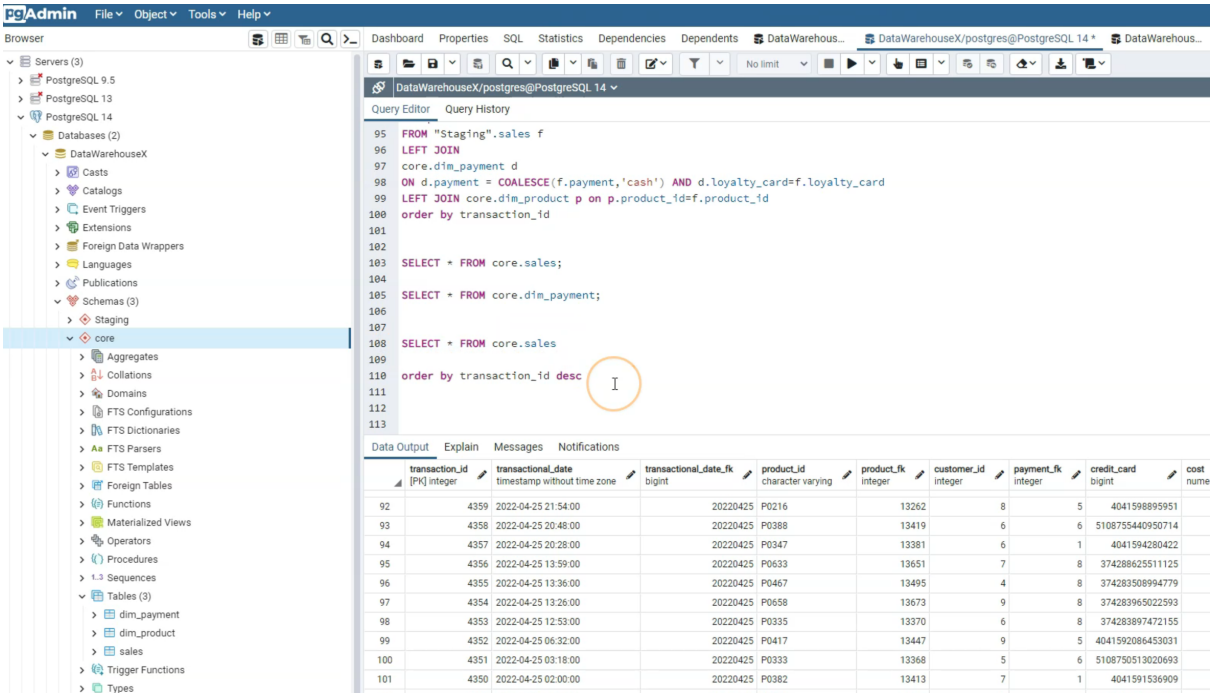

- So in general, we can say that a data warehouse is a centralized location for all of our integrated data that is optimized for the purpose of data analysis.
- So this is oftentimes, therefore, serving as a basis for our reporting, and our analysis because this can be now used, and this is one of the big use cases for a data warehouse to plan our strategic decision-making.
- We can analyze all of the different categories for a product. We can evaluate the performance, and find the right decisions that are now data-driven.
- And this is because now we have different data sources integrated together that help us to get now a clear, big picture, because all of the different data sources are integrated, we have a high performance, and this can be now all packaged together into our reporting.
- This is now very easy or at least much easier to create our reporting with this data warehouse. Of course, now, on top of that, even if we have some reportings that can be created, the business users are also now much more flexible to analyze the data.
- So they can, of course, go directly on the database to make their analysis, but they can also just use reporting tools, data visualization tools or any

other tools to quickly, and easily connect to our data warehouse, and then use this data to analyze it.

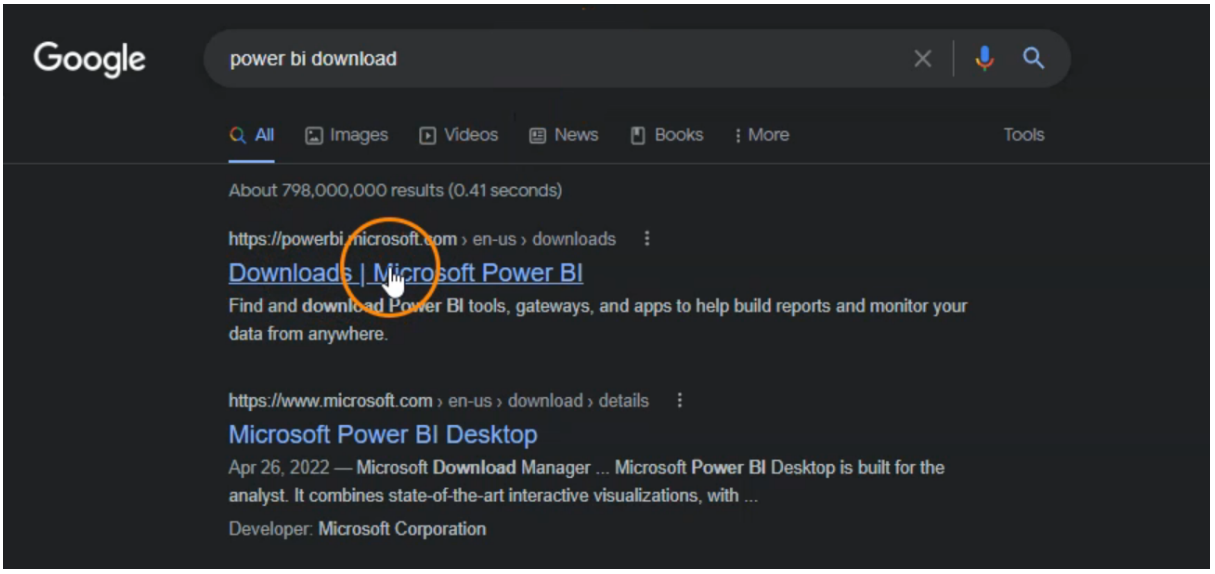
- And they don't need to be super technical, but they can be just business users that are now able to analyze the data. Because the data is now accessible, we have ensured the data quality and it is easy to use.
- But now, also, we can enable more advanced technology. So for example, we want to implement some predictive analytics using some machine learning models, because now, with our data warehouse, we have consolidated data that can just serve as a continuous data feed to continuously train our machine learning models, because they might require the data in a specific structure.
- And then with our ETL workflow, we can continuously ingest the data into these machine learning models. So this can be a great additional benefit, and you can already see and understand that basically our data warehouse just helps with all of the different data analysis options and capabilities.
- So this can be many, many different use cases. Also, regarding big data, we can get more out of it using a data warehouse, because usually, we don't want to store all of, let's say, if we have some internet of things data, from some devices that is just stored in JSON.
- We don't want to store that all in a database, or in our data warehouse. That's why we want to just aggregate maybe the data filter, what is relevant, restructure it, so that we can actually use that data and this is where our data warehouse can now help as well.

# Connecting Datawarehouse to Power BI

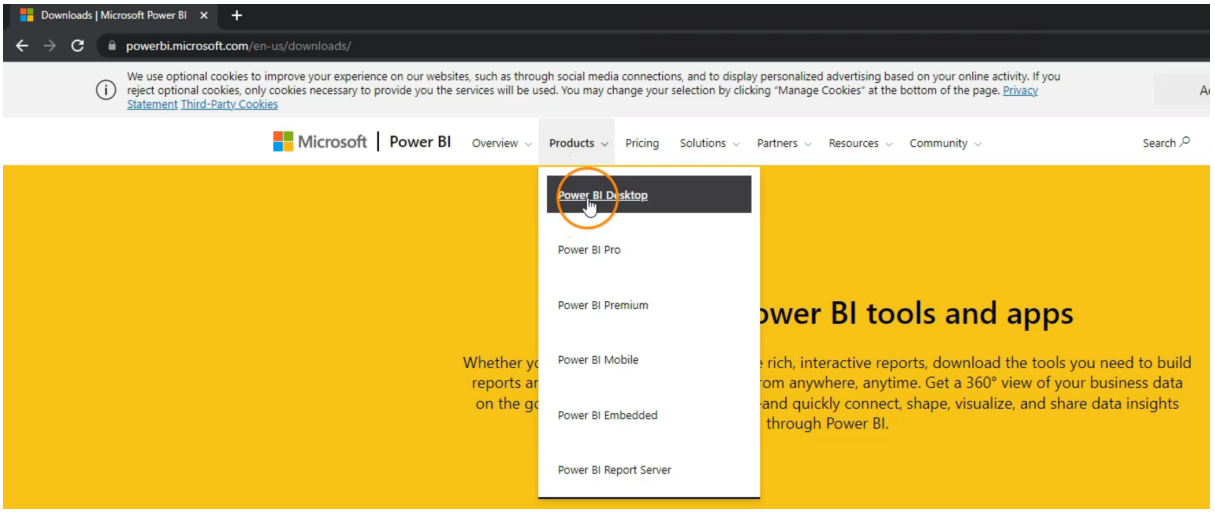
- Once we have created our data warehouse it can be used in multiple ways. One of the ways is of course, directly going onto the database and querying data from there.



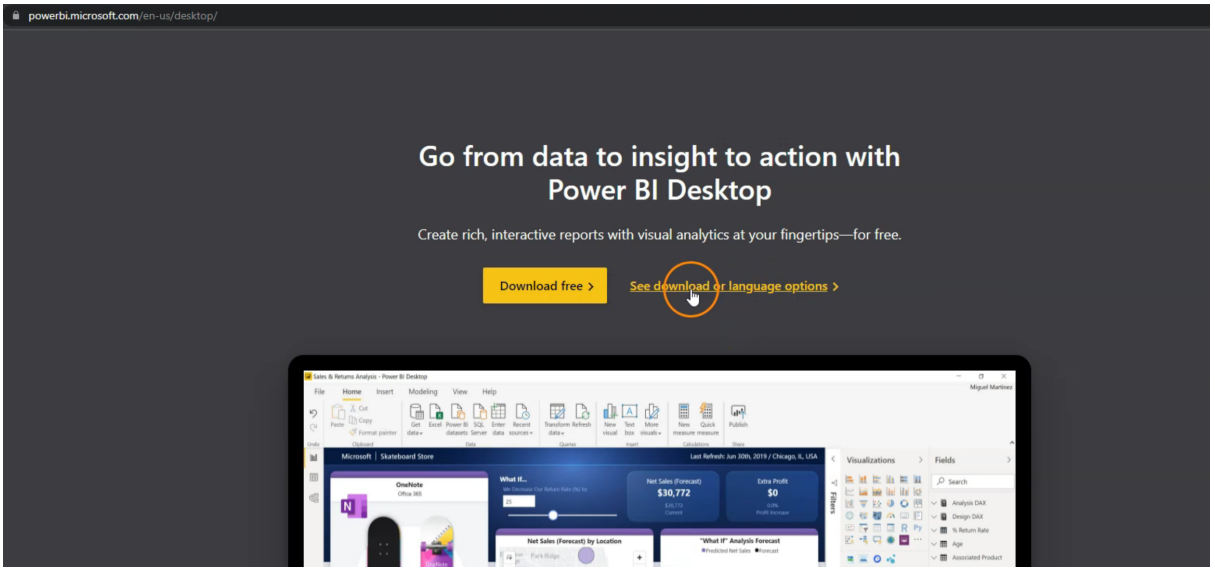

- So we can, for example just query the data as we have done in here. We can maybe group the data and perform certain SQL operations but oftentimes it's helpful to connect the data warehouse to a specific tool.
- And one of the most important ones as mentioned is to connect for example a data visualization or any other BI tool. And as this usually works the same way in all of the different tools, we therefore want to demonstrate how we can connect to a tool.
- And we want to demonstrate this with Power BI. So we can quickly just go through Power BI installation.
- So if you want to download Power BI just search in Google for Power BI download. And then the first link that is not an ad will be download Microsoft Power BI. So make sure it is the site PowerBI.microsoft.com.

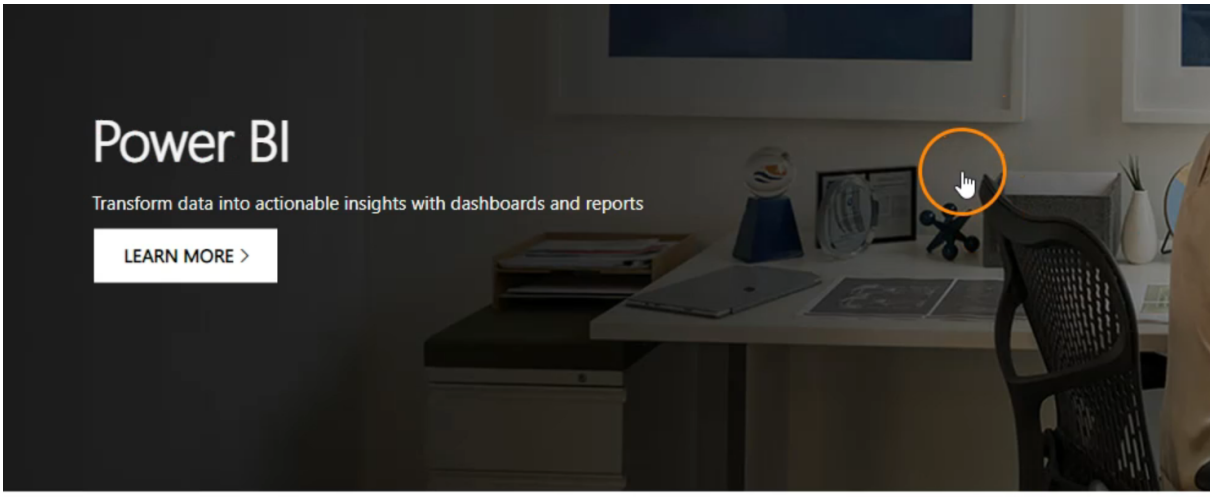



- And then you are also in the category downloads. In here on the products, you'll find Power BI desktop. You can just either click on download for free or just go also to see download options.




- And in here you can now also just click on download and we want to use the upper version. So that's just the 64-bit version. Of course, it's depending on your system. If you happen to have a very old system, so a 32-bit system then you can just go with the lower one. But for most systems, this top one will be the right choice.



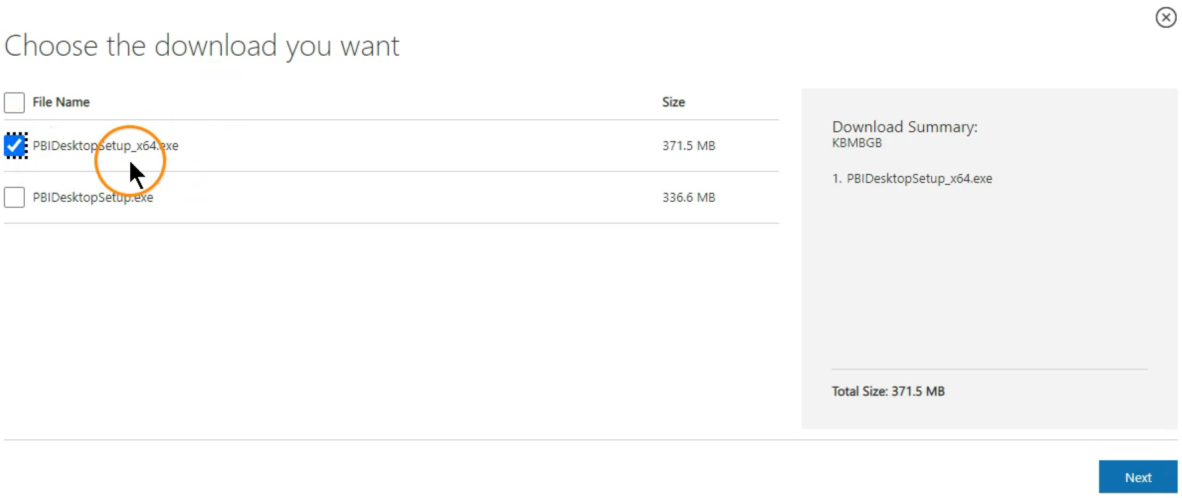




Microsoft Power BI Desktop

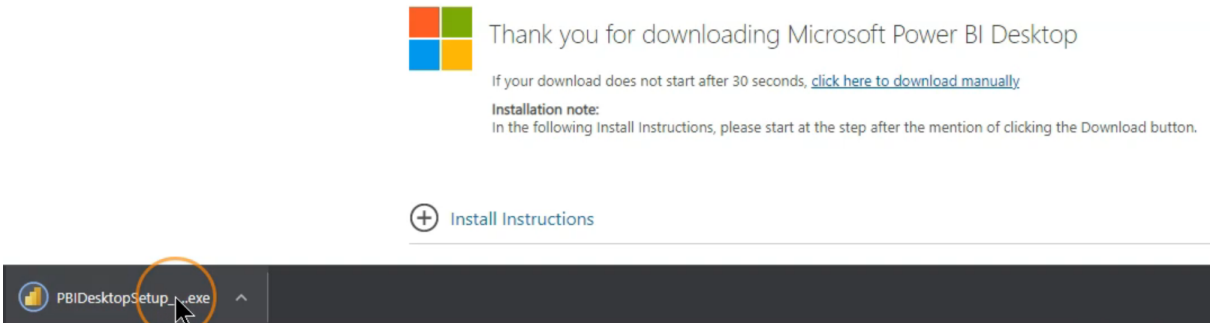
*Important!* Selecting a language below will dynamically change the complete page content to that language.

Select Language: English

Download

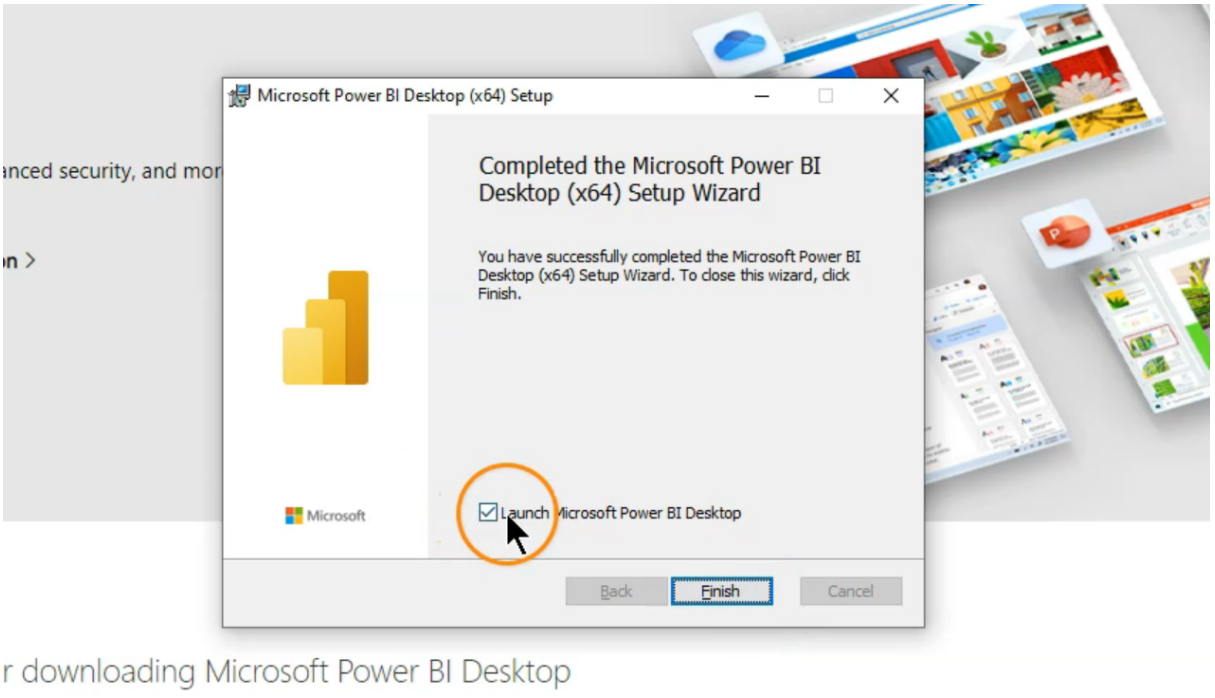



- So in our case, I have a 64-bit machine, and in case of doubt, you probably also have a 64-bit machine. So just select this one and then we can click on next. And then this file will be downloaded to your machine.

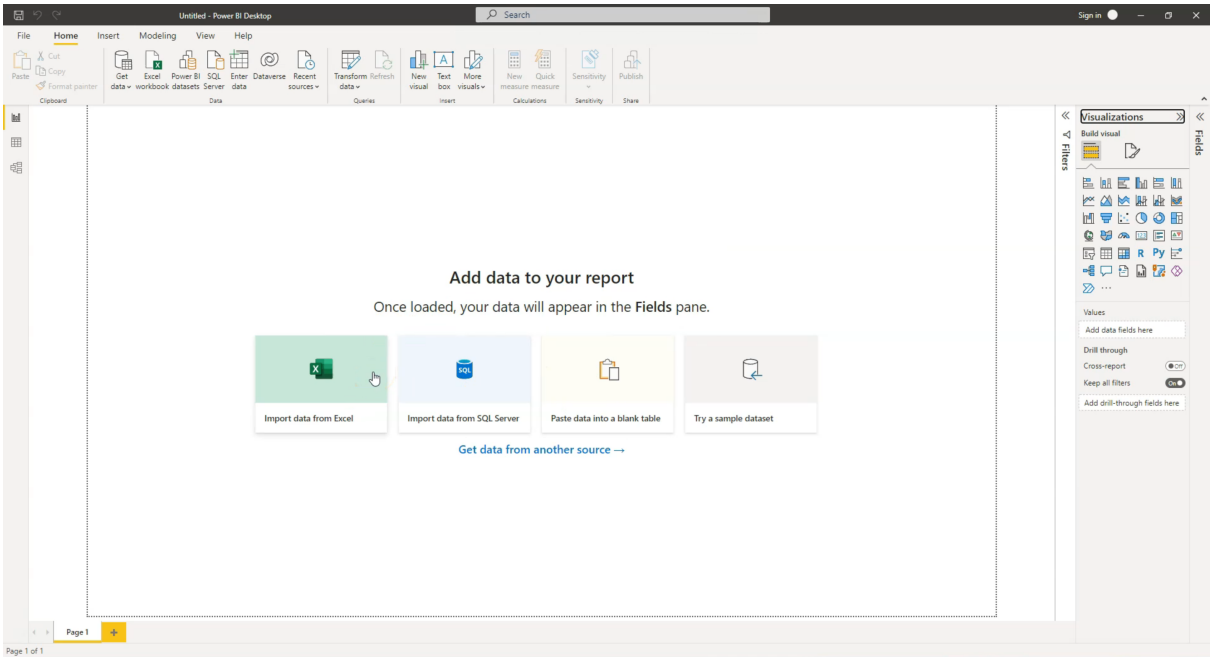



- Once this has been downloaded, we can just start and execute this installation file.

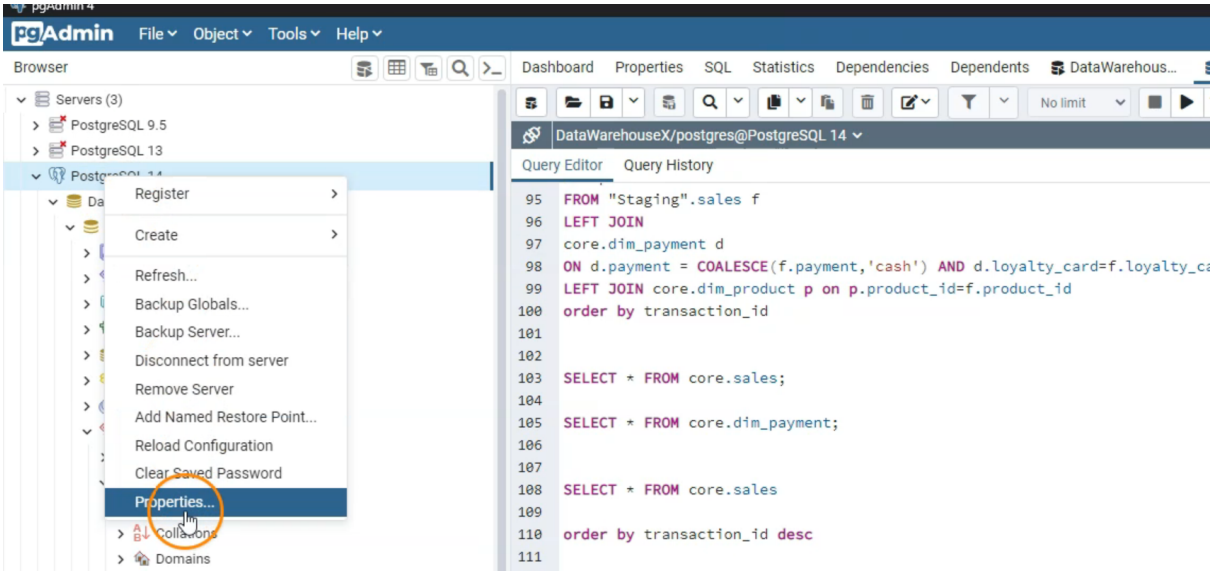

- We can just click on next. And from here we can just go through the very simple installation process. So we can just click through this installation process and then we can install Power BI.
- And then afterwards, this should be installed successfully. So we can also just immediately launch Power BI. So we can have this box checked and just click on finish.



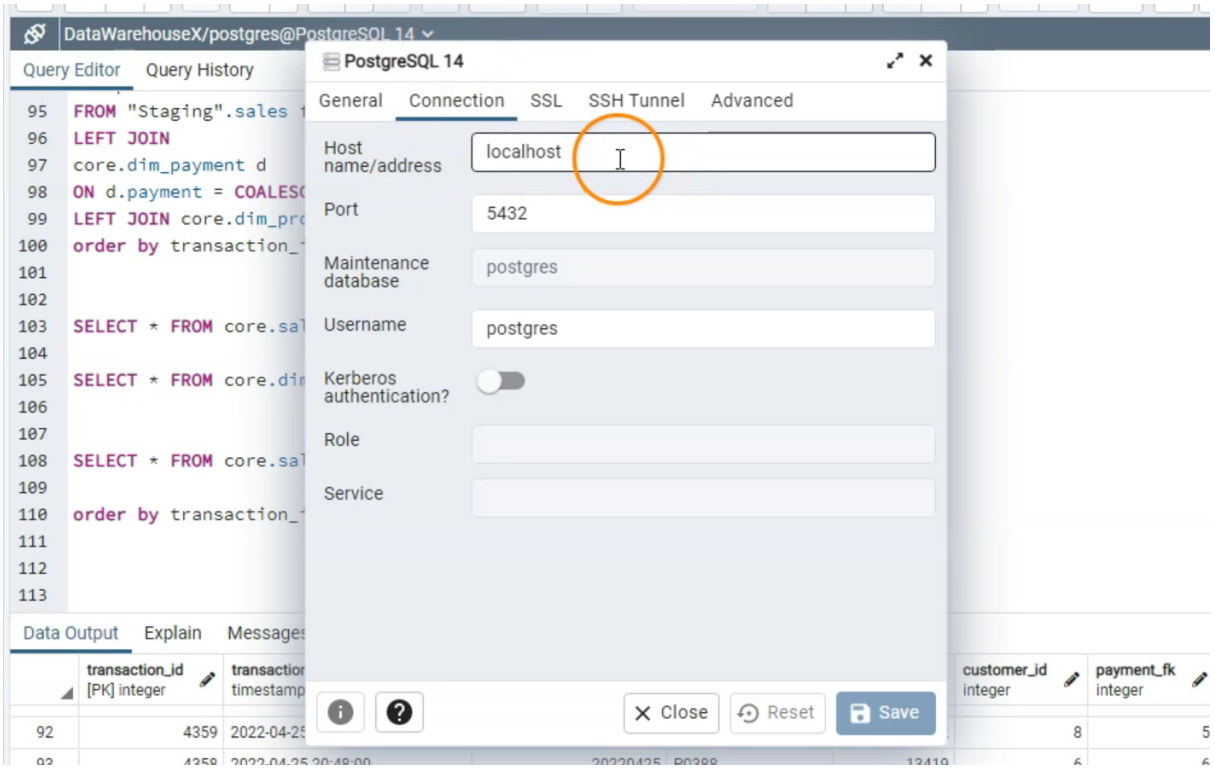
- And afterwards, Power BI should open up, we can just close those pop up windows out. And now this is the Power BI interface.



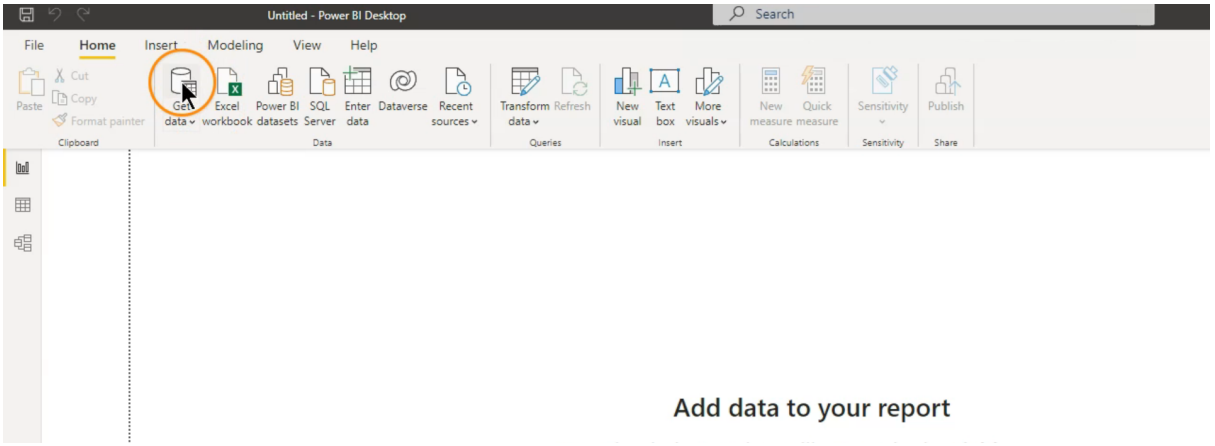

- Of course, again, it's not supposed to be a Power BI in depth course, but just quickly we want to see how we can now work with our data warehouse in Power BI as well because this is probably one of the big use cases.
- So if we have now already set up our data warehouse we can in PG admin or in our database management system just go and find those connection credentials.
- In here, we can go to the server that we are having our data warehouse in and we can right click on it in our case, PostGreSQL 14. And we can there go to properties.



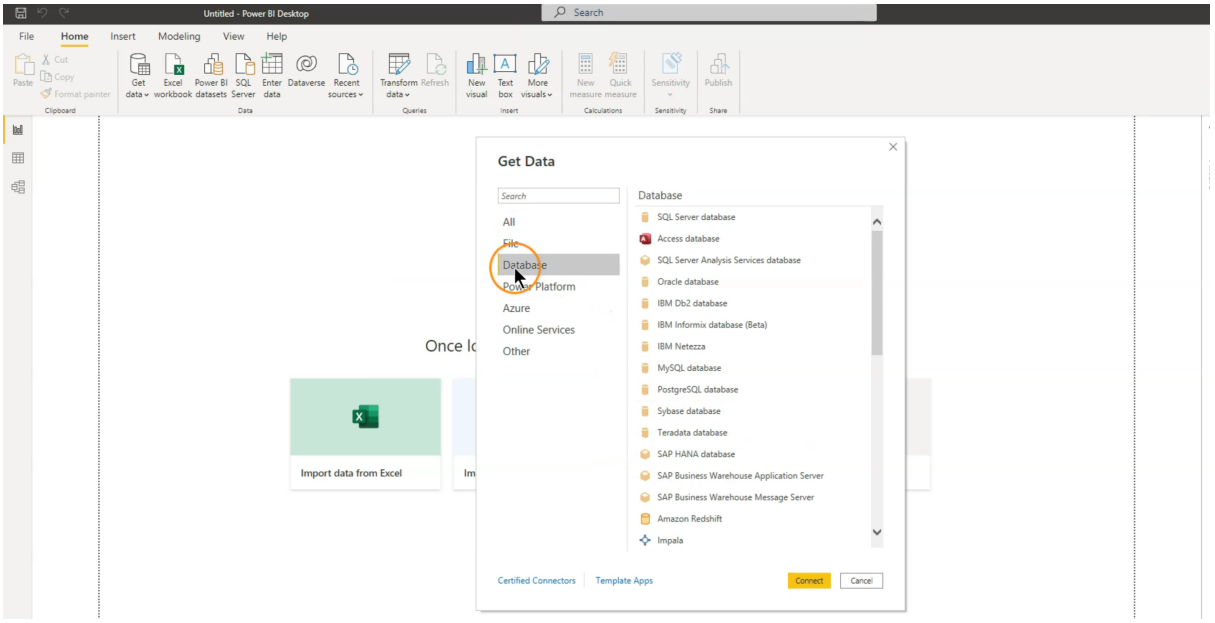

- And here we find the connection. So we have the host name, we have the port and of course the username. And we should also remember the password. So our data warehouse in the end is just a database.



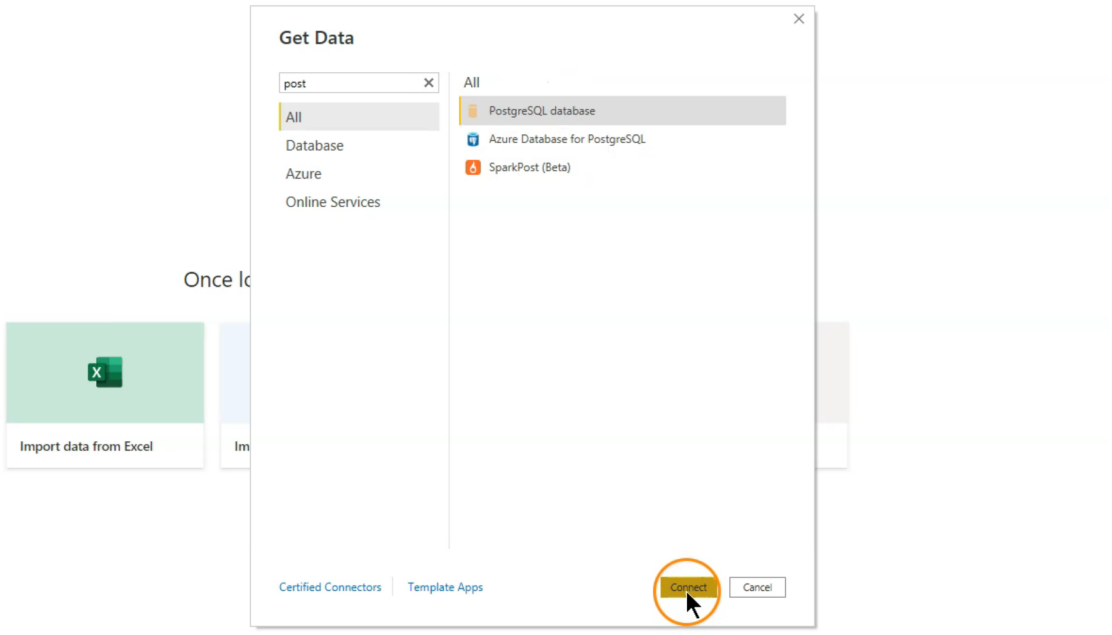

- That's why in Power BI or in any other tool we need to connect to our database system. So in our case, we can connect to a data source just by going to this connections field. And this will then open up all of the different connectors.




- A data warehouse is just a database, so we need to go to database and there find our database system.



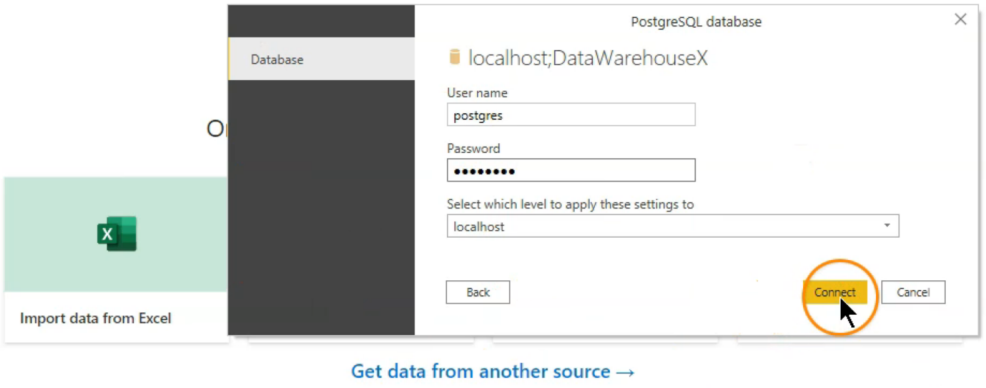

- So I can, for example, search for Postgres I find it in here, and we can then simply click on connect.



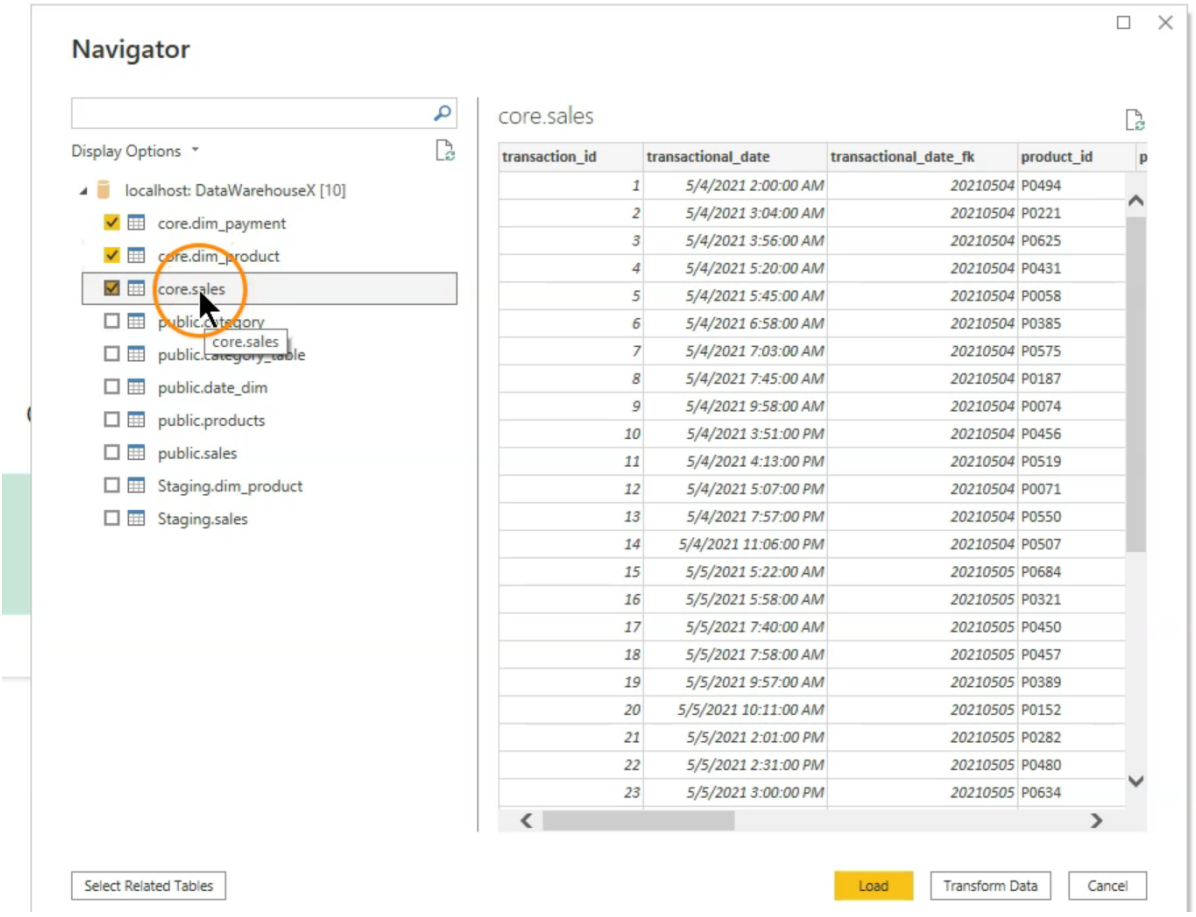

- Now we have remembered that the server is just local host because we have just hosted that locally on our machine.
- And now the database name is, in our case we have this called DataWarehouseX. So that's why I will also enter it in here DataWarehouseX, and then we can decide whether we want to import all of the data from the data warehouse.


- And then Power BI is basically processing the data in memory as well. Or alternatively, we can also just directly connect to the database and then the data will be queried from the database when it's needed.

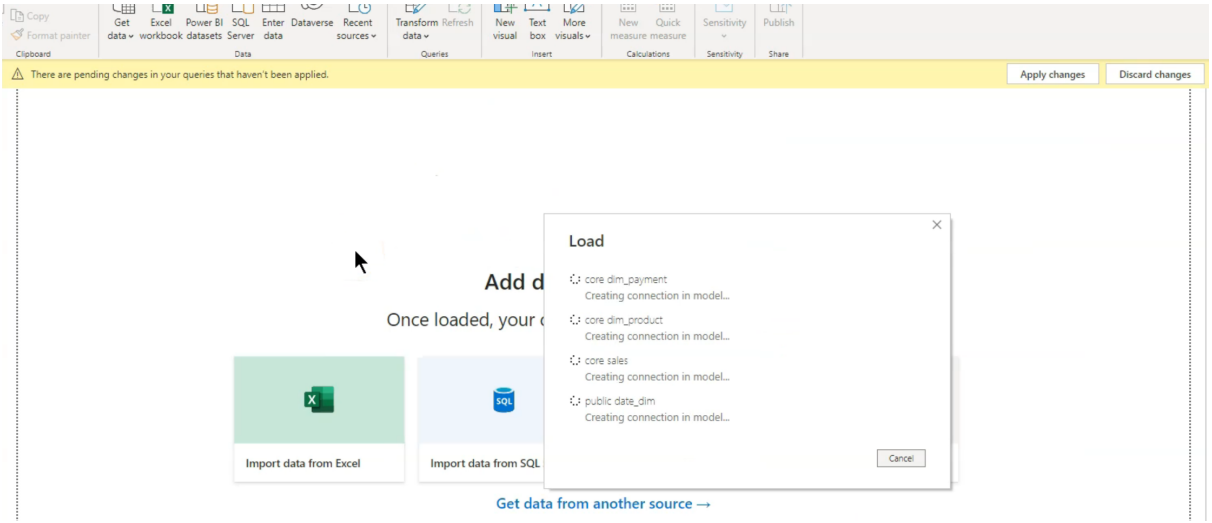
- So when we have a database or a data warehouse with a very high performance, we can just directly query the data also from there. And then we use the performance of our data warehouse.
- In our case though, we can just leave it at import and then the data will be just imported into the Power BI file and it will be processed then again in memory in Power BI.
- So we have insured also a high performance like that. Now of course, we need to get access to this database.
- So usually we can set up a user with read access to this database. In our case, this is just a username Postgres and the password that we have set up when we have set up our SQL system. So this is all that we need to do and then we can simply click on connect.



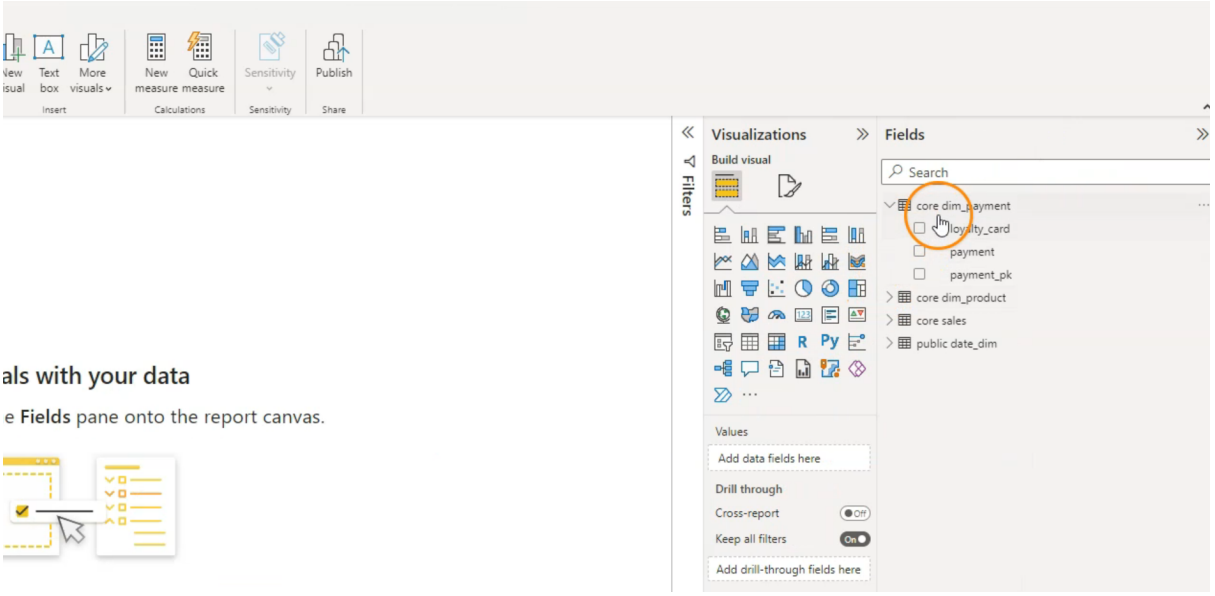

- We can also click here on okay. And then we can now of course just see all of those tables. So we want to of course check all of the boxes for our core layer. So we have two dimensions and then our sales fact and also we can use that date dimension.



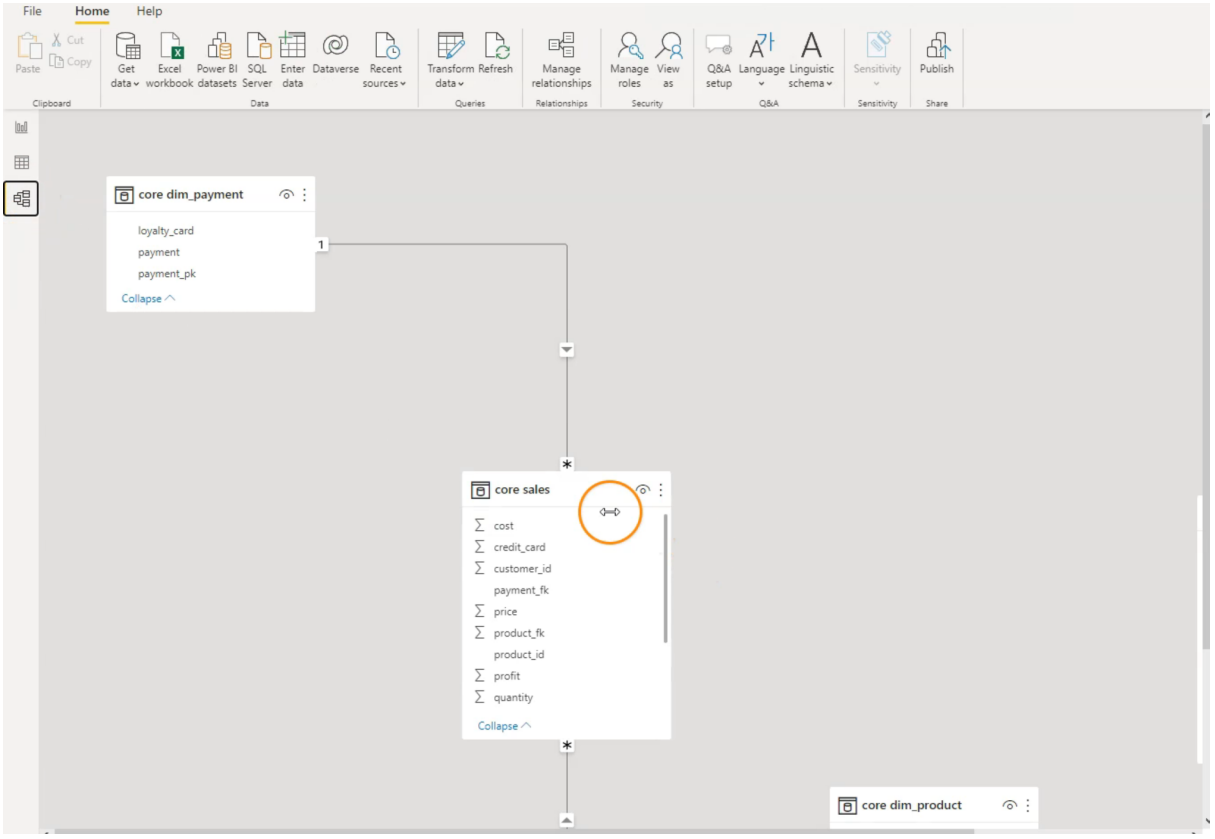

- And then once we have checked them, we can just click on load and then the data will be loaded into Power BI.



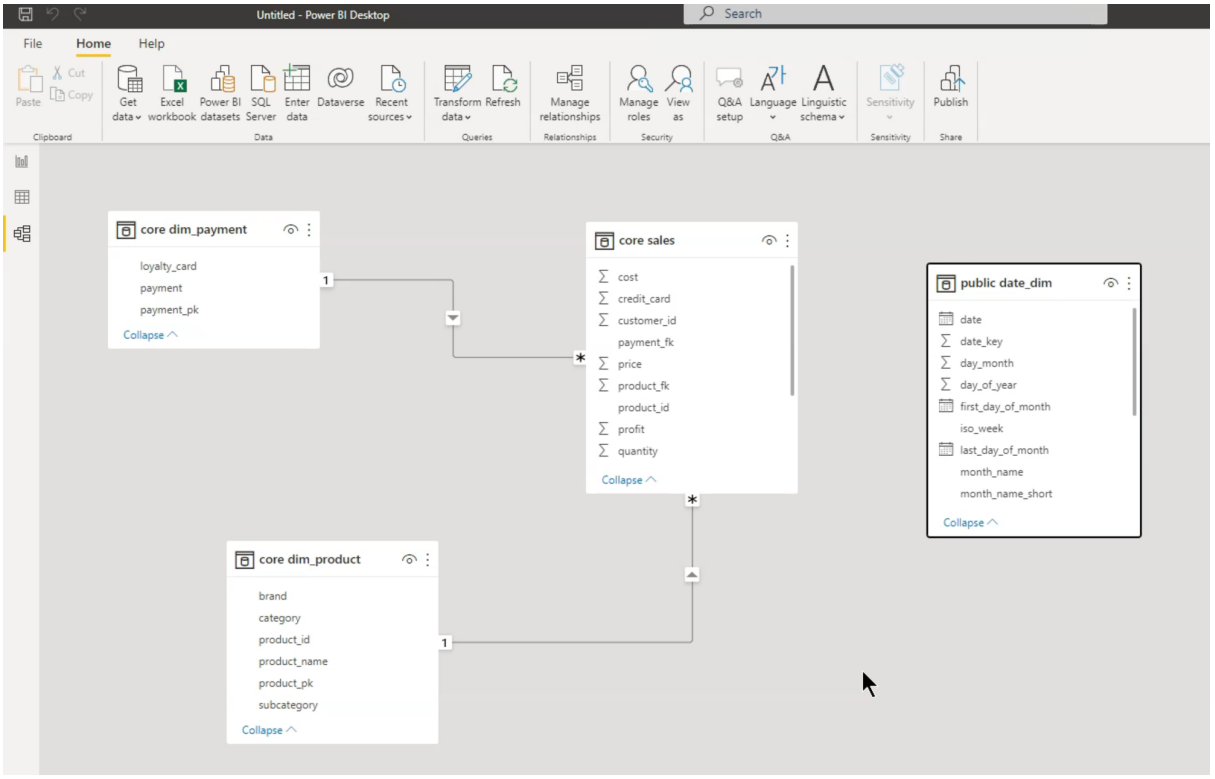

- And now we have those tables available and we can now use this data to create our reports and visualize the data.



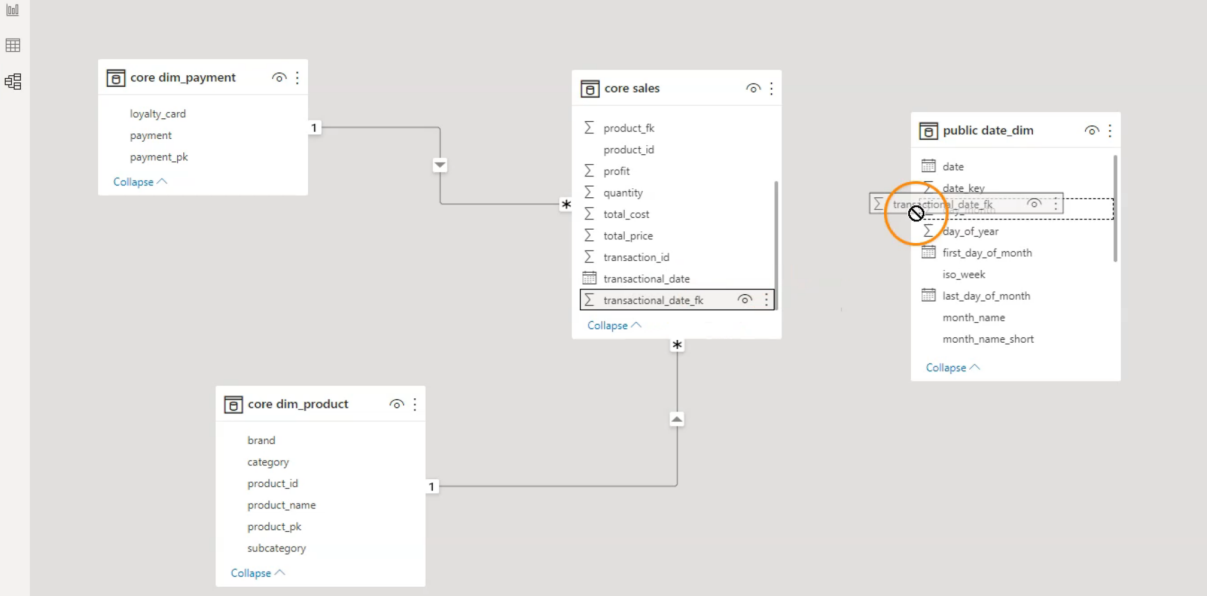

- But we want to also have a look at the model view. And this is where we can see our star schema because we have talked about that a lot.

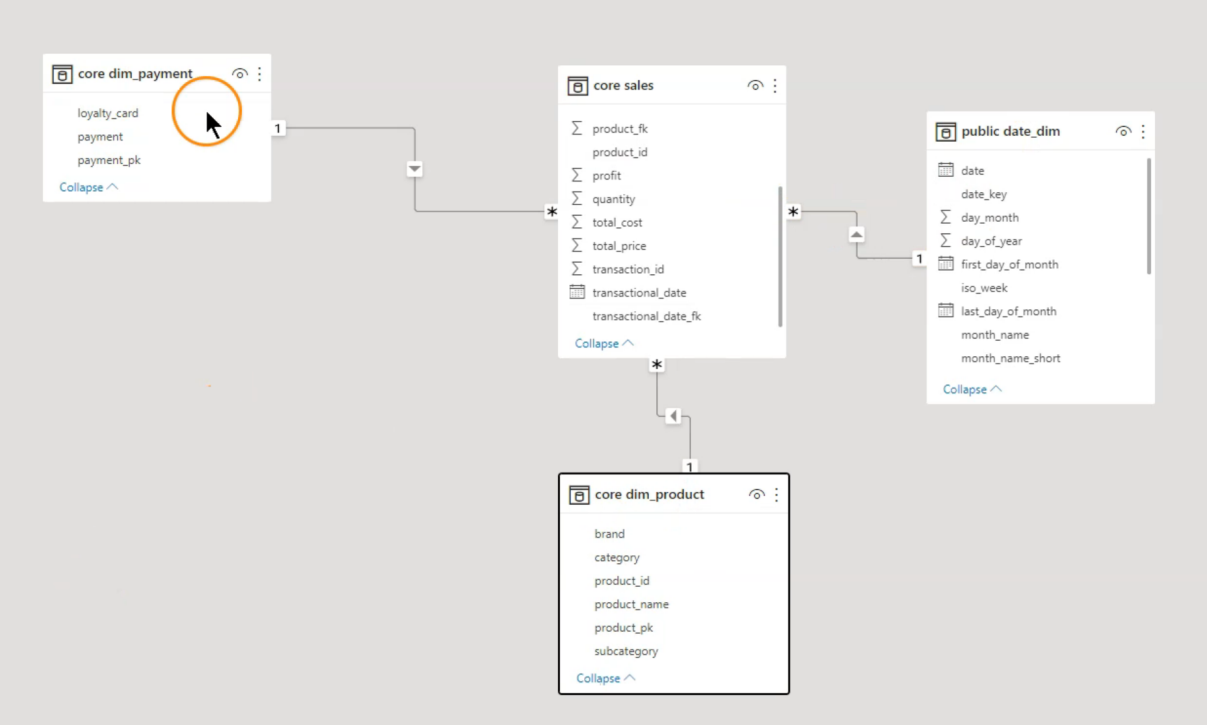



- And we want to see that now also live in action. So this is basically now also kind of serving as a data mart where we are just getting some specific tables that we need for our use case.

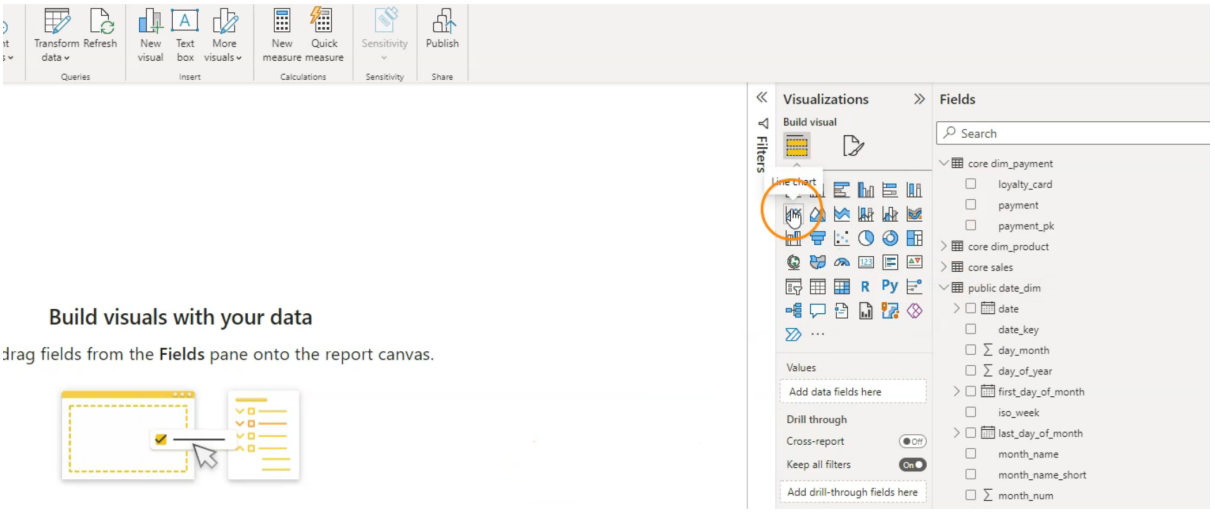



- And then in Power BI we can also set up this task schema where we can just create these connections between the primary key in the dimension and the foreign key in our fact table.
- And then this connection can be easily used and the users don't need to worry about making any joints. They can simply use these tables very easily together. And we'll see that in action in just a second.
- Also, we just want to make sure that we also connect the date dimension with the sales table. So therefore we want to search for transaction date foreign key and drag and drop that onto the date key. And then we also have this connection established.

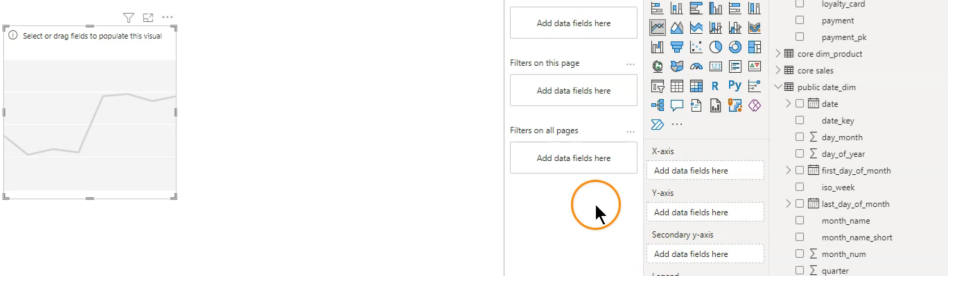


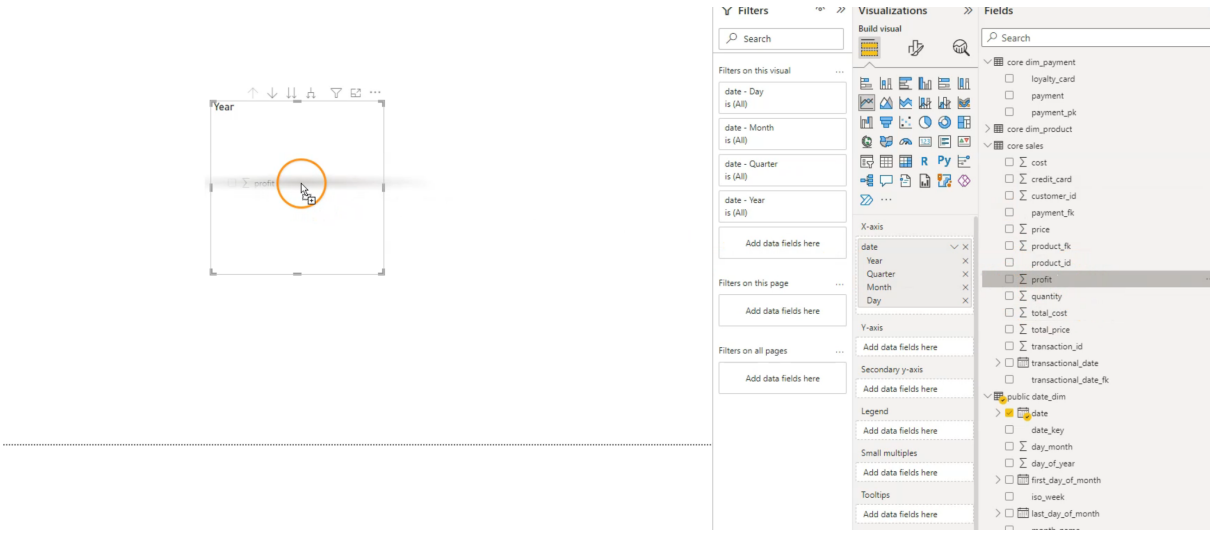

- And with that, we have now basically set up our star schema with the fact table in the middle and all of the dimensions around. So this is now really nice, and we want to see how does the data look like now for the end users and how is it to visualize the data?
- For example, we can just click on this line chart to set up a chart with the performance of our sales over time.



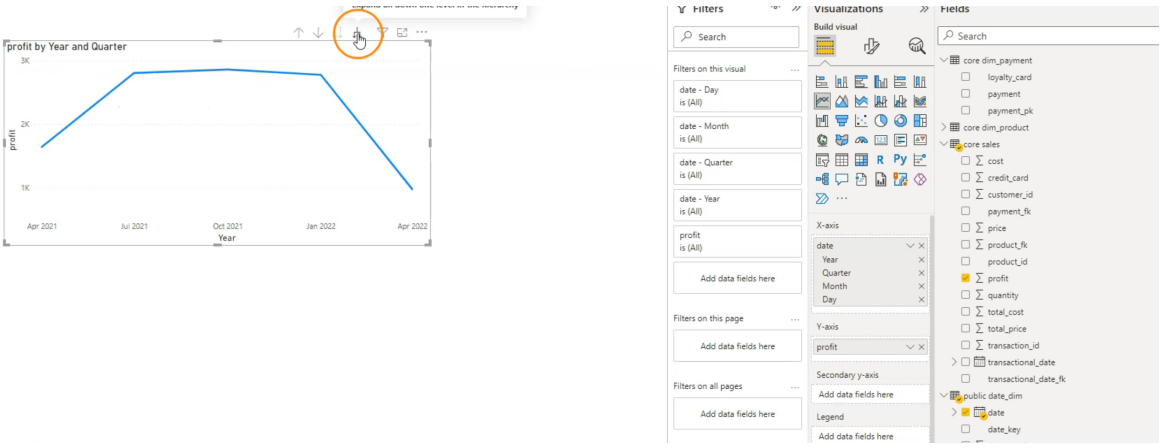

- So we can now just drag and drop the date into this visual from the date table. And of course, we can now also just drag in from our fact table some measures.



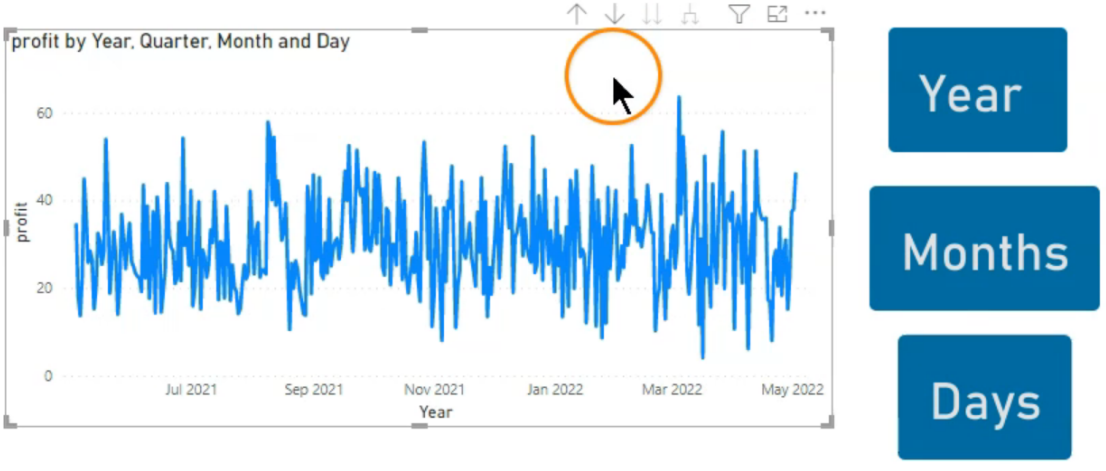

- So for example, the profit. We can again just drag and drop it and now we can just make this a little bit larger and we can now also click on this top button to drill down into the different levels.



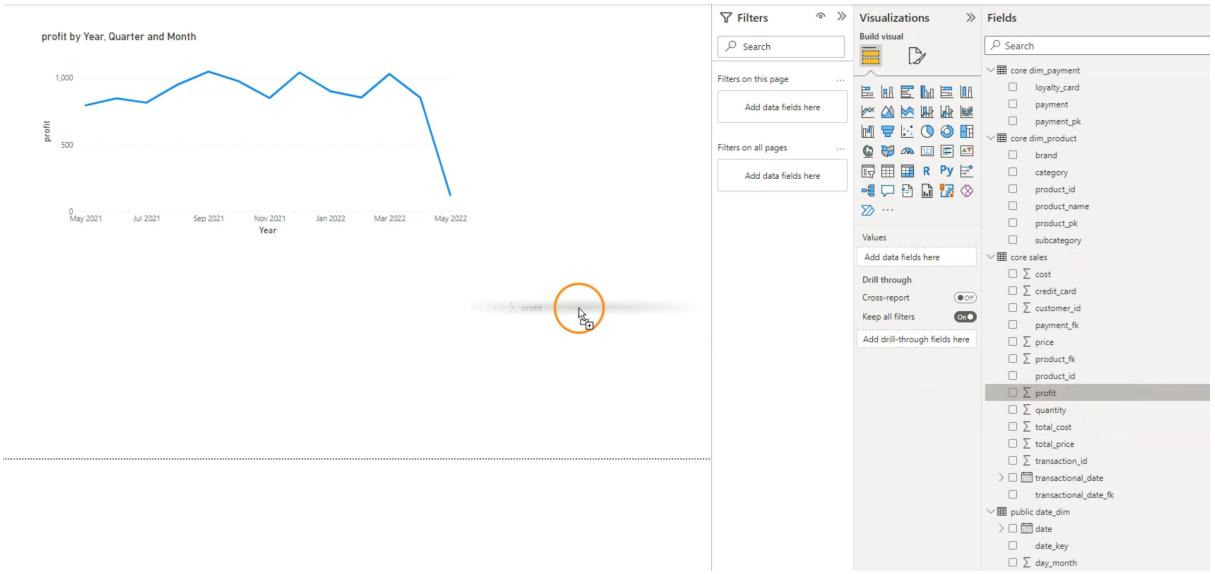
- So like this we can see exactly how the performance is over time. And of course the same, we can now also easily do with the product category.



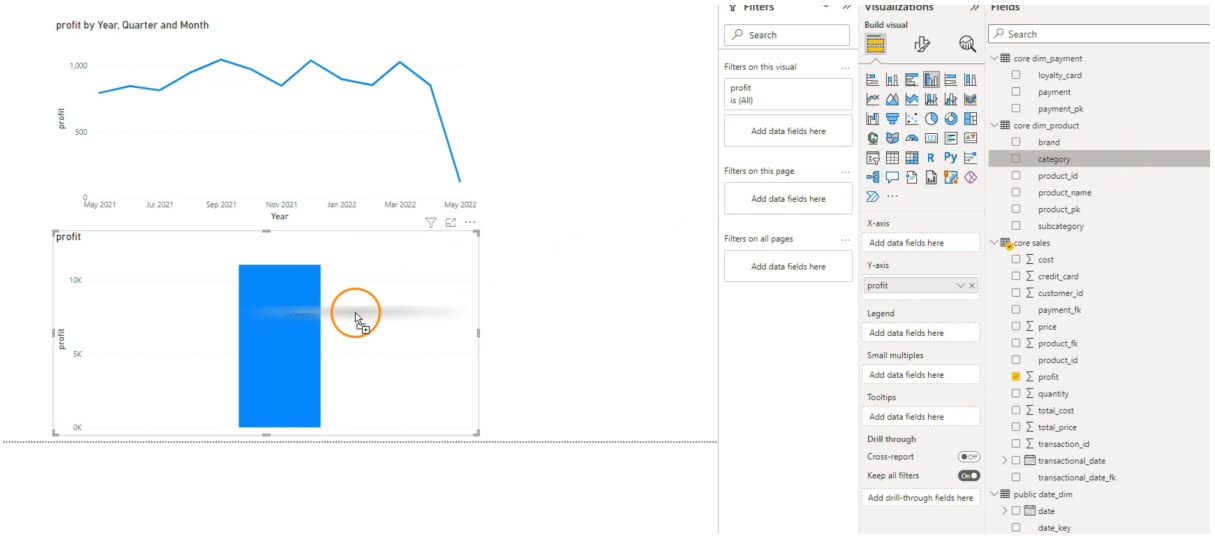


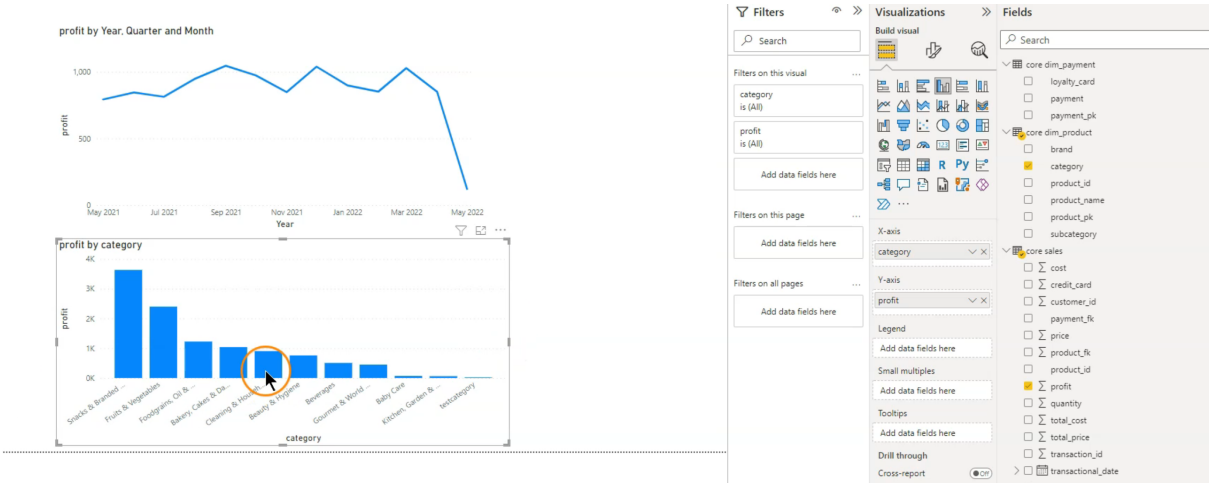


- So for example, we want to see what is the profit. We can just drag it into this open canvas.

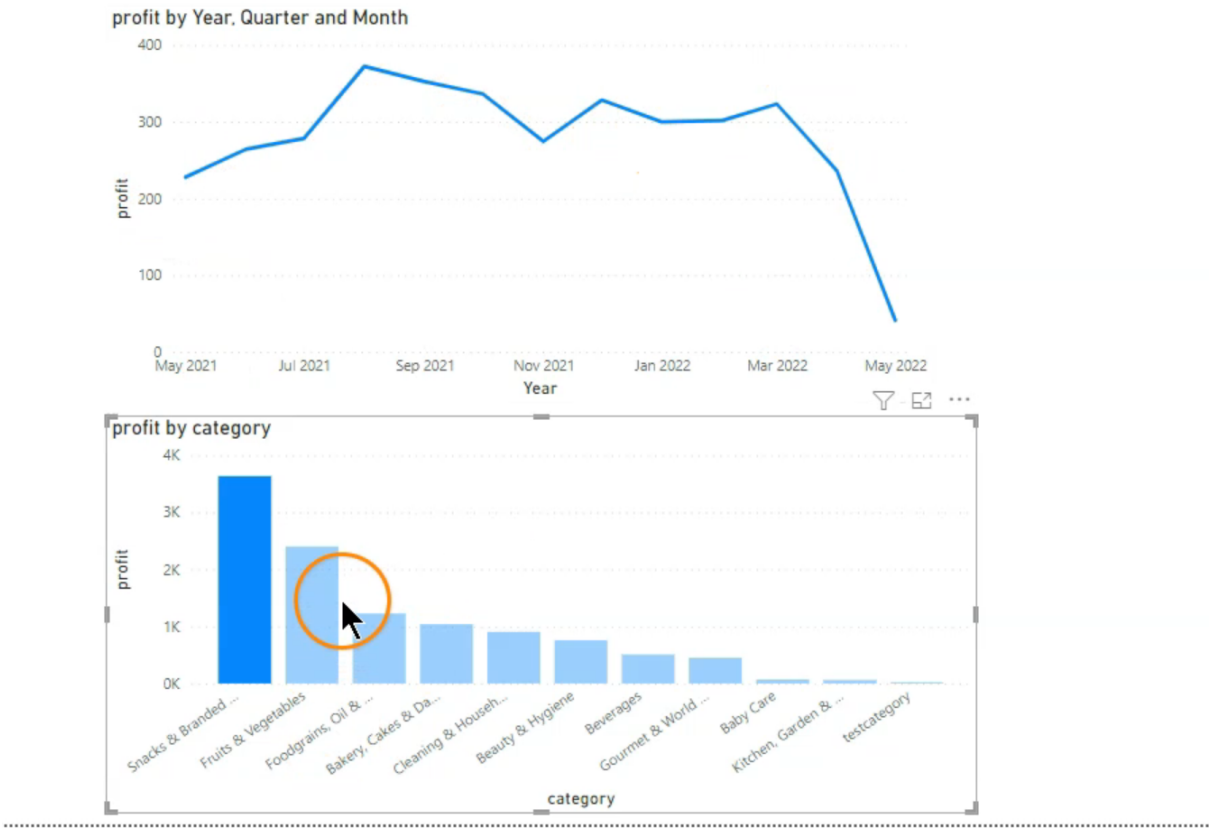


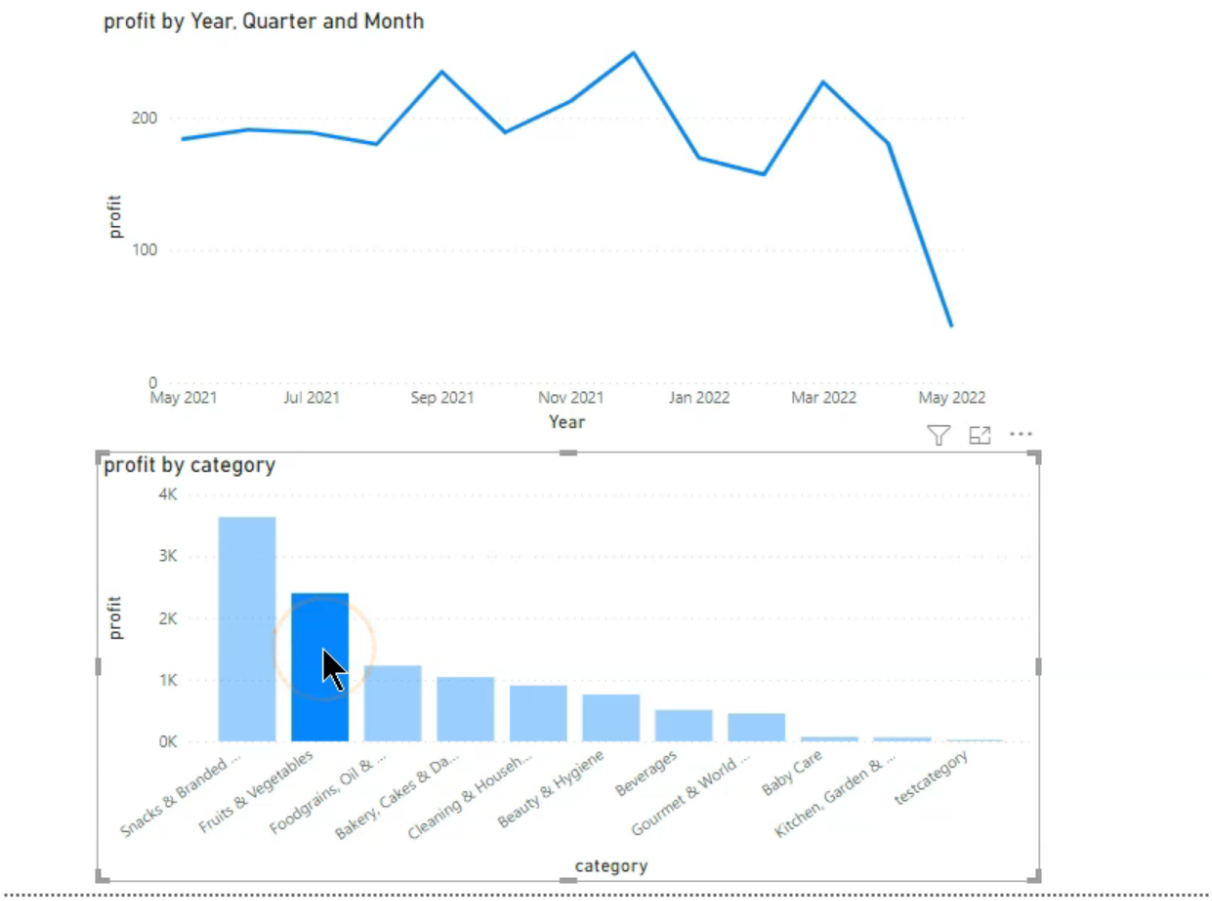

- And then we can also just drag in the category again, just into this visual. And we can like this easily now analyze also the profit by category, and we can now even use it very interactively.






- So for example, we can just click on snacks and see now also the development over time for snacks or also maybe for fruits and vegetables.






- So all of that is now very easily possible with a high performance, with a high usability thanks to our data warehouse.
- I hope this was a quick and helpful demonstration of how our data warehouse can be then used also in other tools and demonstrate how useful it can be.