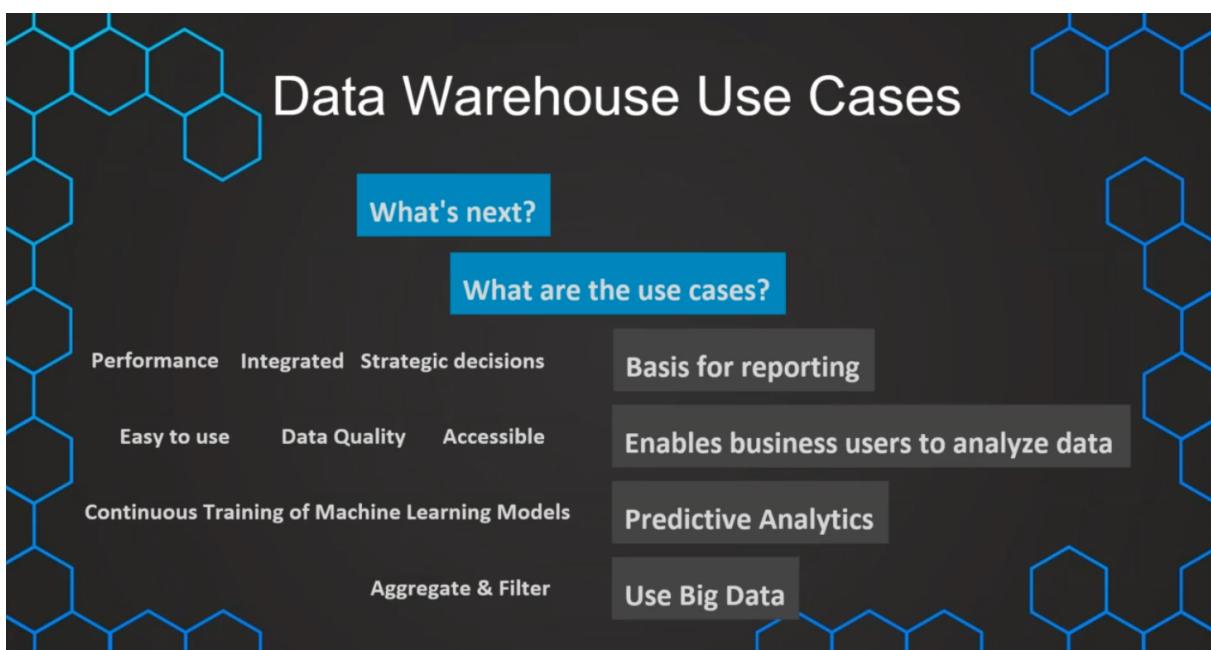


12. Using a Data Warehouse



- 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.

```

95   FROM "Staging".sales f
96   LEFT JOIN
97     core.dim_payment d
98   ON d.payment = COALESCE(f.payment,'cash') AND d.loyalty_card=f.loyalty_card
99   LEFT JOIN core.dim_product p on p.product_id=f.product_id
100  ORDER BY transaction_id
101
102
103  SELECT * FROM core.sales;
104
105  SELECT * FROM core.dim_payment;
106
107
108  SELECT * FROM core.sales
109
110  ORDER BY transaction_id DESC
111
112
113

```

transaction_id	transactional_date	transactional_date_fk	product_id	product_fk	customer_id	payment_fk	credit_card	cost
92	4359 2022-04-25 21:54:00	20220425_P0216	13262	8	5	4041598895951		
93	4358 2022-04-25 20:48:00	20220425_P0388	13419	6	6	5108755440950714		
94	4357 2022-04-25 20:28:00	20220425_P0347	13381	6	1	4041594280422		
95	4356 2022-04-25 19:59:00	20220425_P0633	13651	7	8	37428862551125		
96	4355 2022-04-25 13:36:00	20220425_P0467	13495	4	8	374283508994779		
97	4354 2022-04-25 13:26:00	20220425_P0658	13673	9	8	374283965022593		
98	4353 2022-04-25 12:53:00	20220425_P0335	13370	6	8	374283897472155		
99	4352 2022-04-25 06:32:00	20220425_P0417	13447	9	5	4041592086453031		
100	4351 2022-04-25 03:18:00	20220425_P0333	13368	5	6	5108750513020693		
101	4350 2022-04-25 02:00:00	20220425_P0382	13413	7	1	4041591536909		

- 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.

A screenshot of a Google search results page for the query "power bi download". The top result is a link to "Downloads | Microsoft Power BI" (https://powerbi.microsoft.com/en-us/downloads). A mouse cursor is hovering over this link, which is highlighted with a yellow circle. Below the link, a snippet of text reads: "Find and download Power BI tools, gateways, and apps to help build reports and monitor your data from anywhere." The snippet is also circled in yellow.

- 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.

A screenshot of the "Downloads | Microsoft Power BI" page. The URL in the browser address bar is https://powerbi.microsoft.com/en-us/downloads/. A dropdown menu is open under the "Products" heading, with "Power BI Desktop" highlighted by a yellow circle and a cursor pointing at it. Other options in the dropdown include "Power BI Pro", "Power BI Premium", "Power BI Mobile", "Power BI Embedded", and "Power BI Report Server". To the right of the dropdown, there is a section titled "Power BI tools and apps" with a brief description.

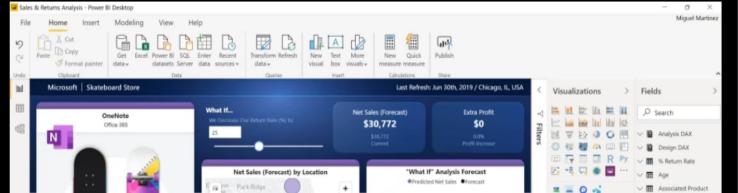
- And in here you can now also just click on download and we want to use the upper version. So that's just the 64-bit version. Of course, it's depending on your system. If you happen to have a very old system, so a 32-bit system then you can just go with the lower one. But for most systems, this top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

powerbi.microsoft.com/en-us/desktop/

Go from data to insight to action with Power BI Desktop

Create rich, interactive reports with visual analytics at your fingertips—for free.

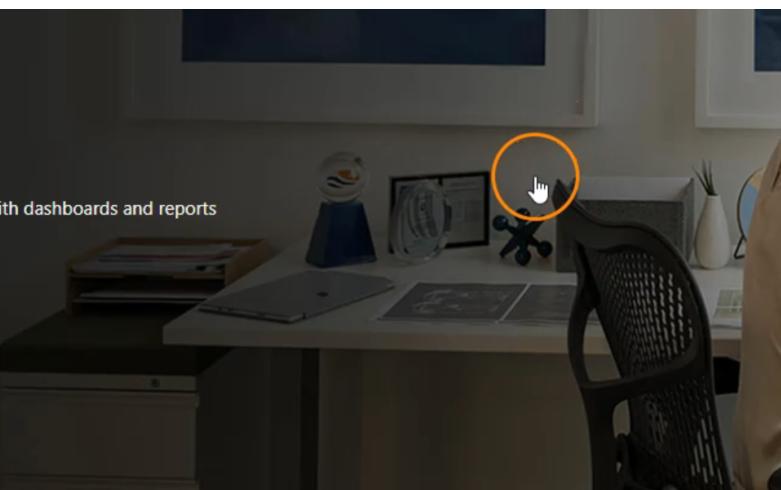
[Download free >](#) [See download or language options >](#)



Power BI

Transform data into actionable insights with dashboards and reports

[LEARN MORE >](#)



Microsoft Power BI Desktop

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

Select Language:



Choose the download you want

File Name	Size
PBIDesktopSetup_x64.exe	371.5 MB
PBIDesktopSetup.exe	336.6 MB

Download Summary:
KBMGB
1. PBIDesktopSetup_x64.exe

Total Size: 371.5 MB

[Next](#)

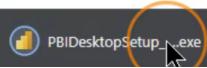
- So in our case, I have a 64-bit machine, and in case of doubt, you probably also have a 64-bit machine. So just select this one and then we can click on next. And then this file will be downloaded to your machine.

 Thank you for downloading Microsoft Power BI Desktop

If your download does not start after 30 seconds, [click here to download manually](#)

Installation note:
In the following Install Instructions, please start at the step after the mention of clicking the Download button.



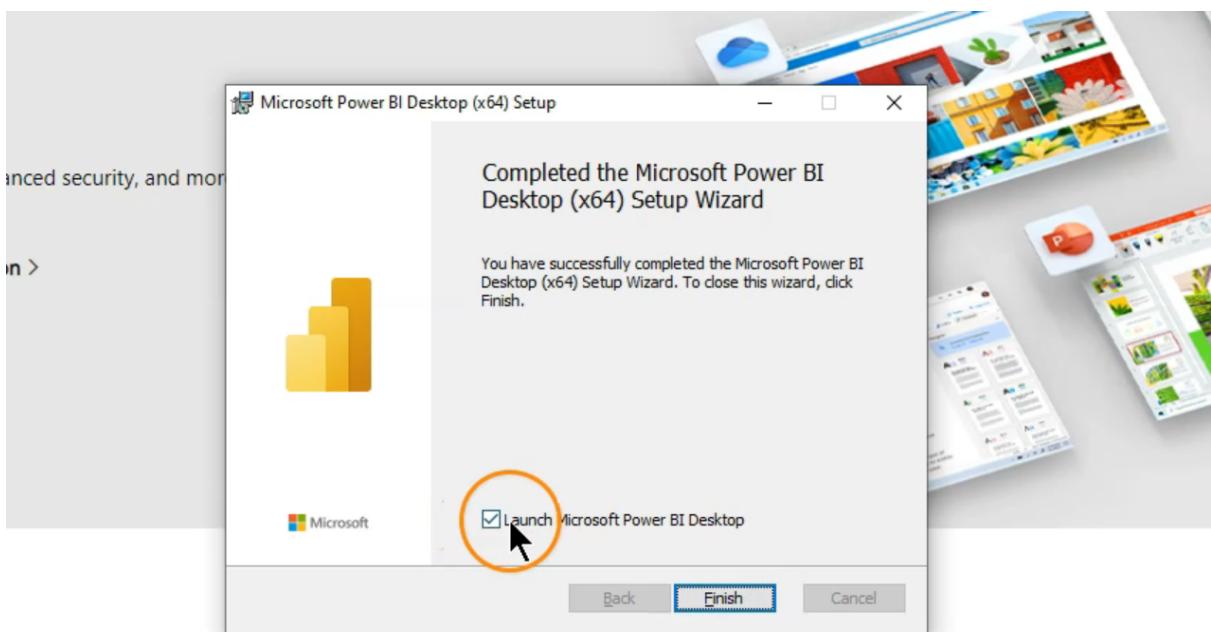


- Once this has been downloaded, we can just start and execute this installation file.



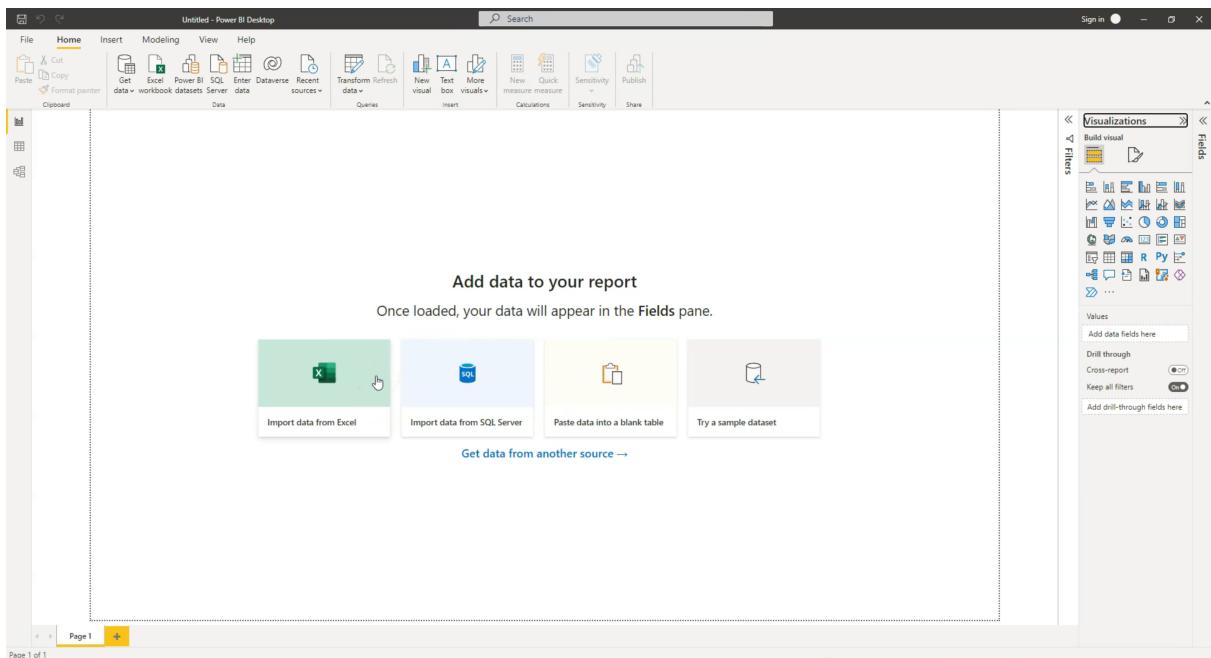
or downloading Microsoft Power BI Desktop

- 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.

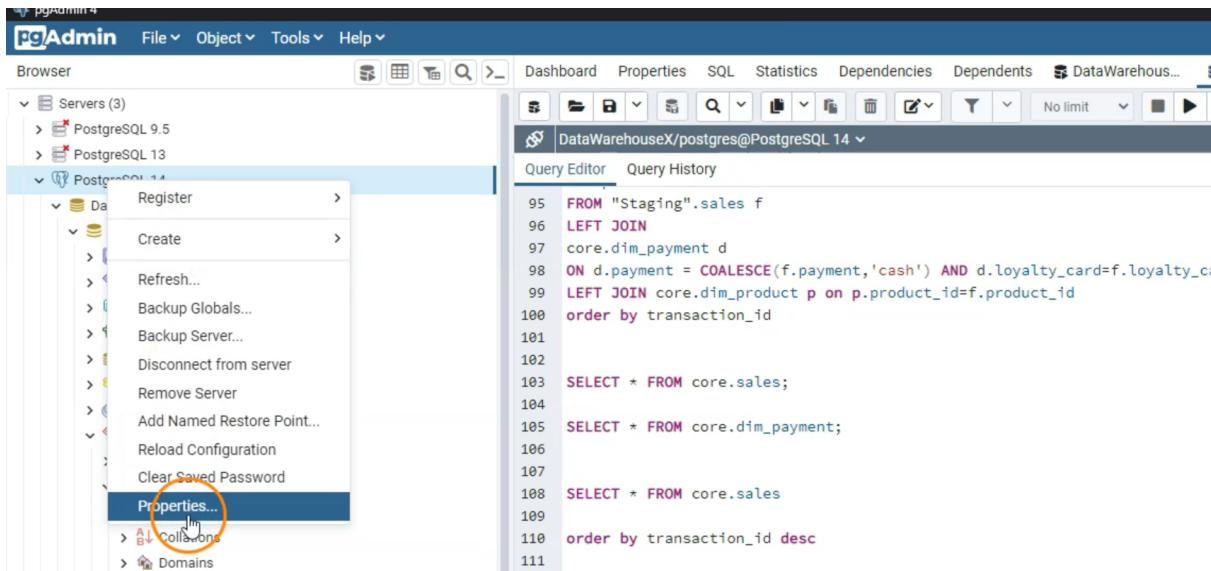


or downloading Microsoft Power BI Desktop

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



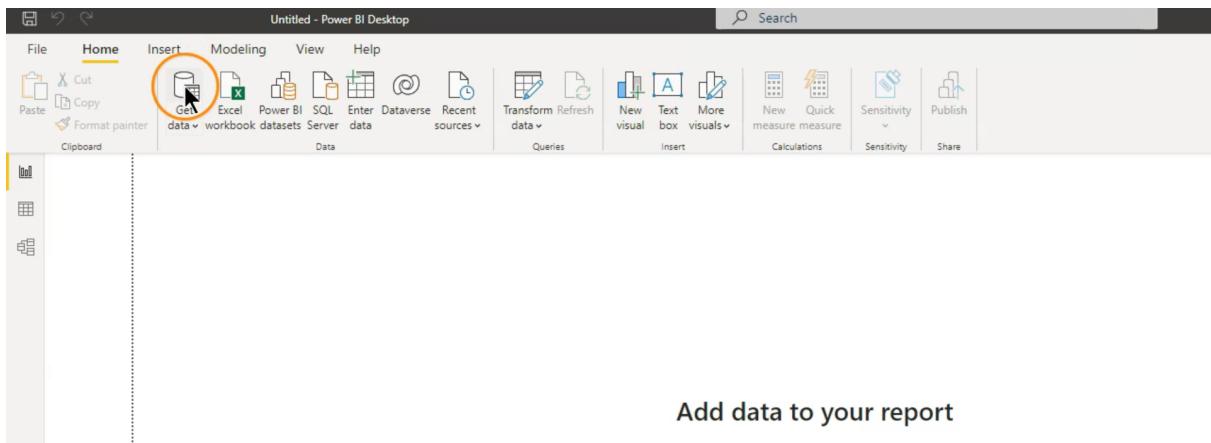
- Of course, again, it's not supposed to be a Power BI in depth course, but just quickly we want to see how we can now work with our data warehouse in Power BI as well because this is probably one of the big use cases.
- So if we have now already set up our data warehouse we can in PG admin or in our 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.



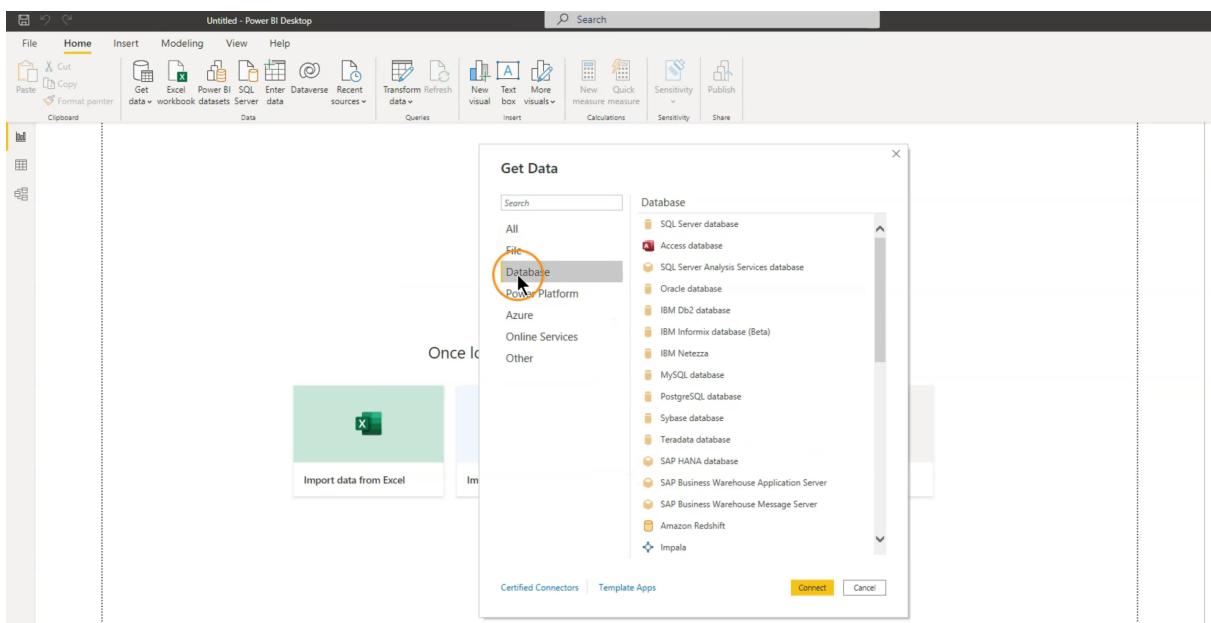
- And here we find the connection. So we have the host name, we have the port and of course the username. And we should also remember the password. So our data warehouse in the end is just a database.

The screenshot shows the 'Connection' tab of the PostgreSQL 14 connection dialog. The 'Host name/address' field contains 'localhost' and is circled in orange. Other fields include 'Port' (5432), 'Maintenance database' (postgres), 'Username' (postgres), and 'Kerberos authentication?' (disabled). Below the connection tab, the 'Data Output' tab is active, showing a table with two rows of data. The first row has columns: transaction_id [PK] integer, transaction timestamp, customer_id integer, and payment_fk integer. The second row has values: 4359, 2022-04-25 20:49:00, 8, and 5.

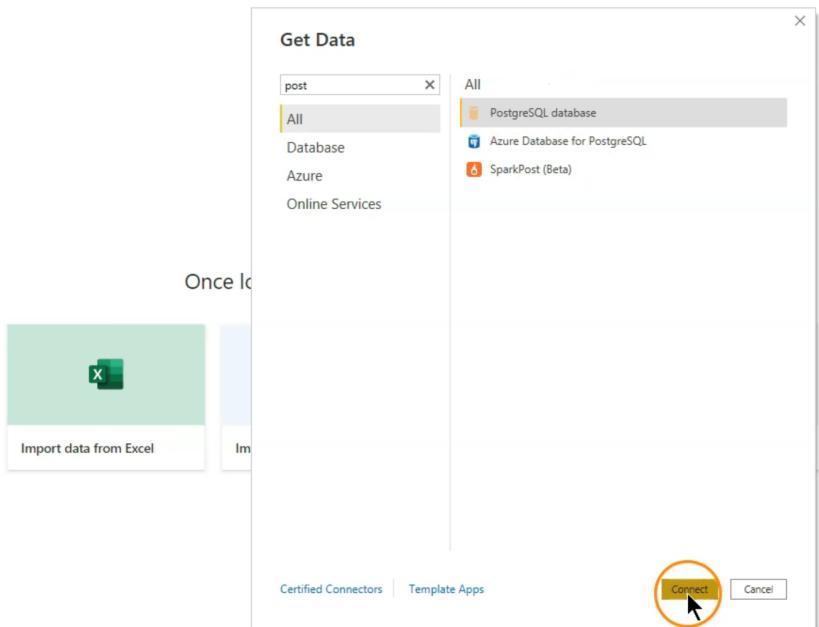
- That's why in Power BI or in any other tool we need to connect to our database system. So in our case, we can connect to a data source just by going to this connections field. And this will then open up all of the different connectors.



- 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.

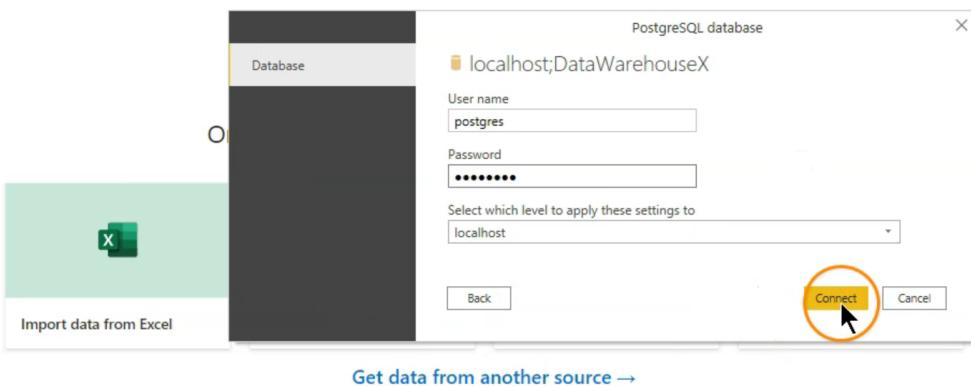


- Now we have remembered that the server is just local host because we have just hosted that locally on our machine.
- And now the database name is, in our case we have this called DataWarehouseX. So that's why I will also enter it in here DataWarehouseX, and then we can decide whether we want to import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



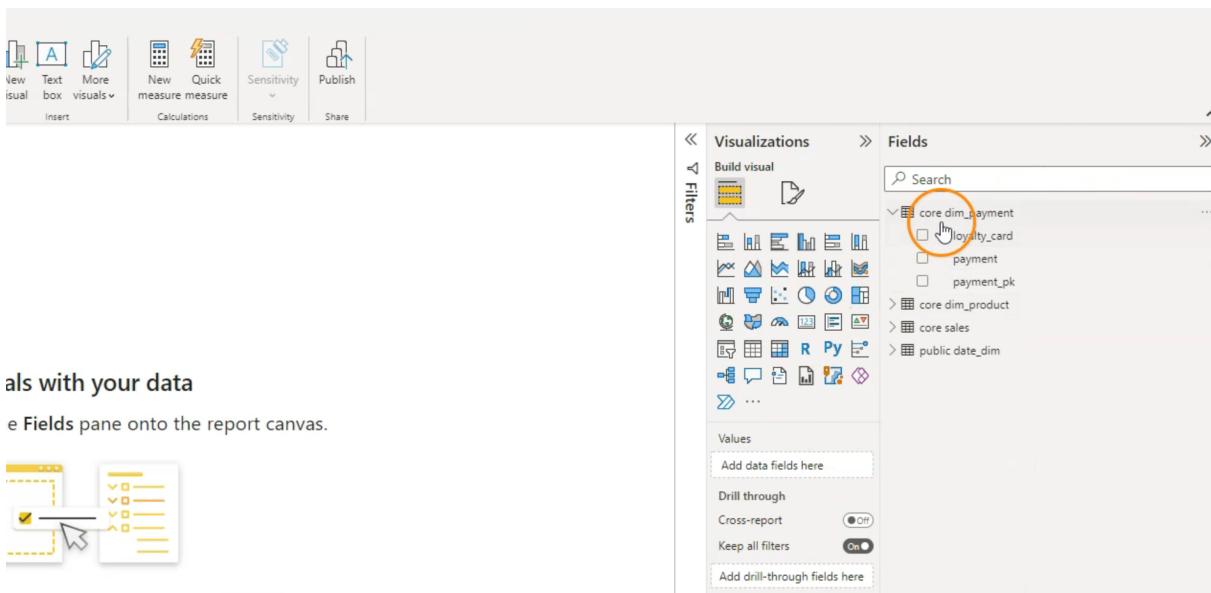
- We can also click here on okay. And then we can now of course just see all of those tables. So we want to of course check all of the boxes for our core layer. So we have two dimensions and then our sales fact and also we can use that date dimension.

The screenshot shows the Power BI Navigator window. On the left, there's a tree view of data sources under 'localhost: DataWarehouseX [10]'. Several tables are checked: 'core.dim_payment', 'core.dim_product', and 'core.sales'. The 'core.sales' node is highlighted with a yellow circle. To its right is a preview grid for the 'core.sales' table, which contains 23 rows of transactional data. At the bottom are buttons for 'Select Related Tables', 'Load' (highlighted in yellow), 'Transform Data', and 'Cancel'.

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

The screenshot shows the Power BI ribbon with the 'Data' tab selected. A message at the top says 'There are pending changes in your queries that haven't been applied.' Below the ribbon, a 'Load' dialog box is open, listing several tables: 'core.dim_payment', 'core.dim_product', 'core.sales', and 'public.date_dim'. At the bottom of the dialog is a 'Get data from another source →' link. The background shows the Power BI workspace with two data import options: 'Import data from Excel' and 'Import data from SQL'.

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

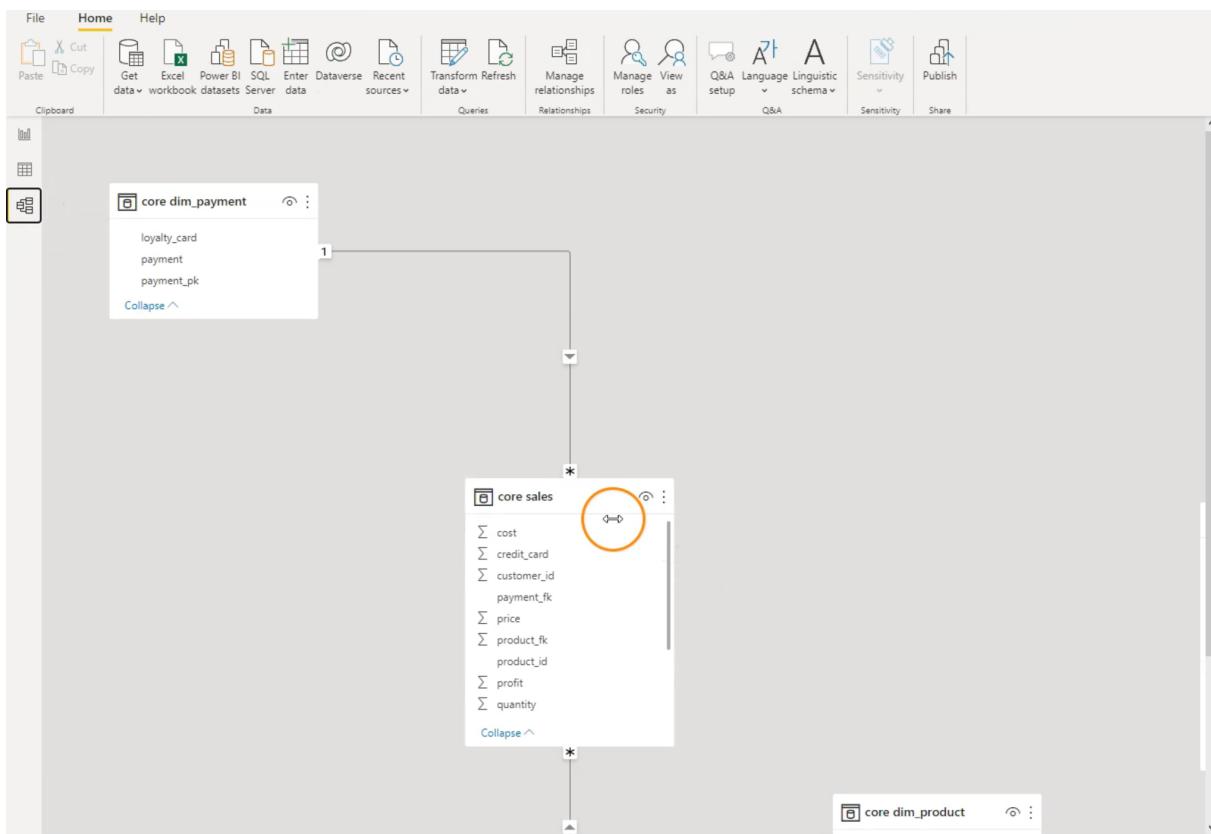


als with your data

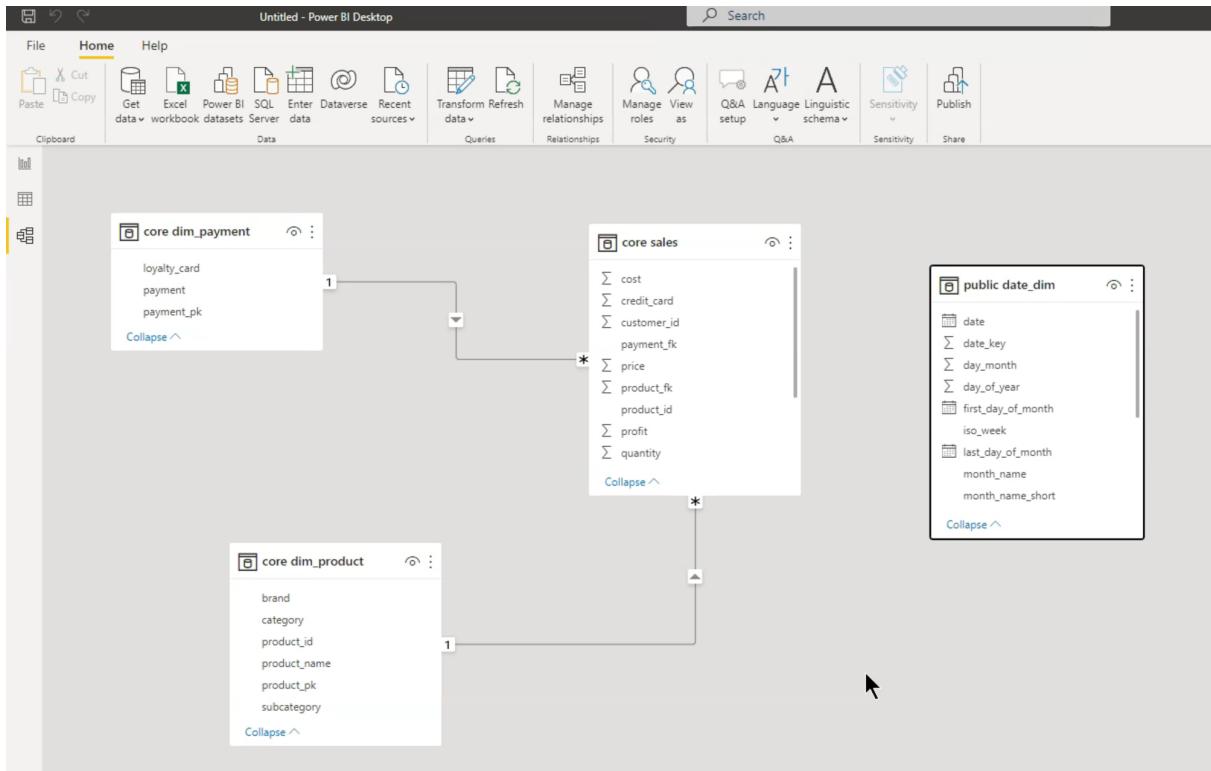
the Fields pane onto the report canvas.



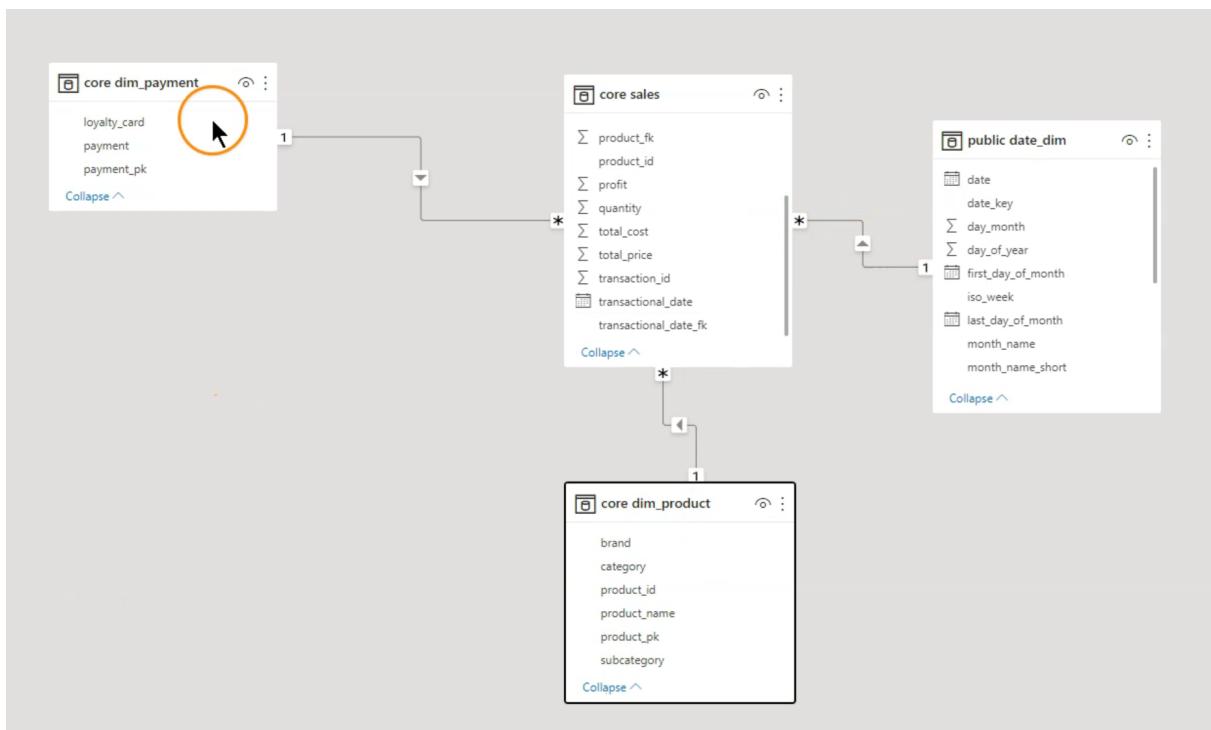
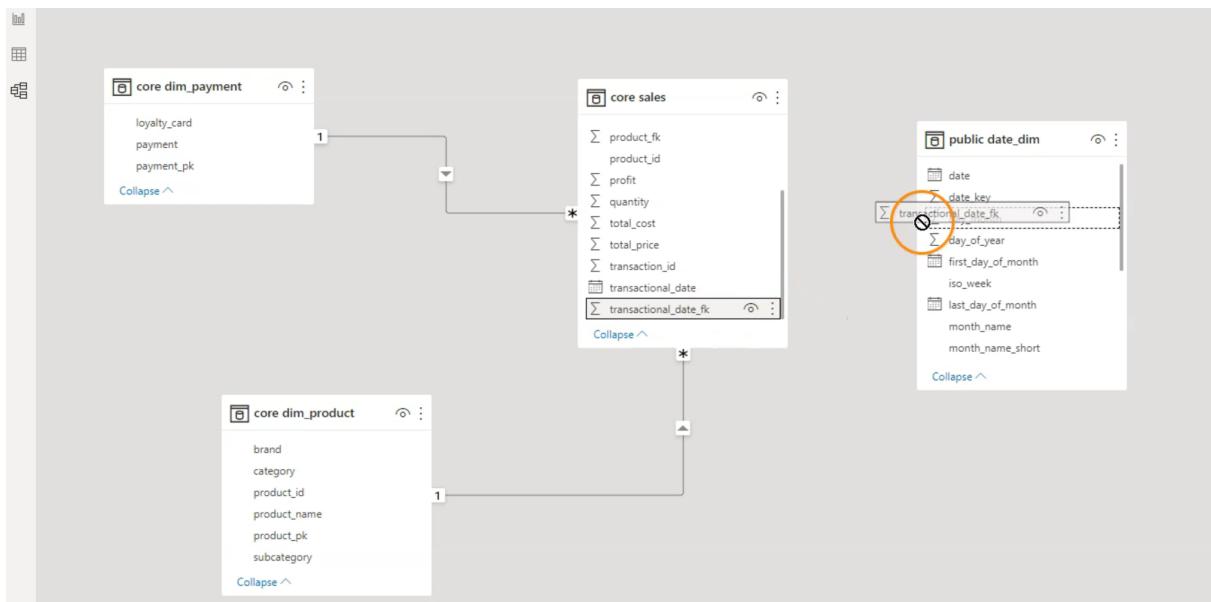
- But we want to also have a look at the model view. And this is where we can see our star schema because we have talked about that a lot.



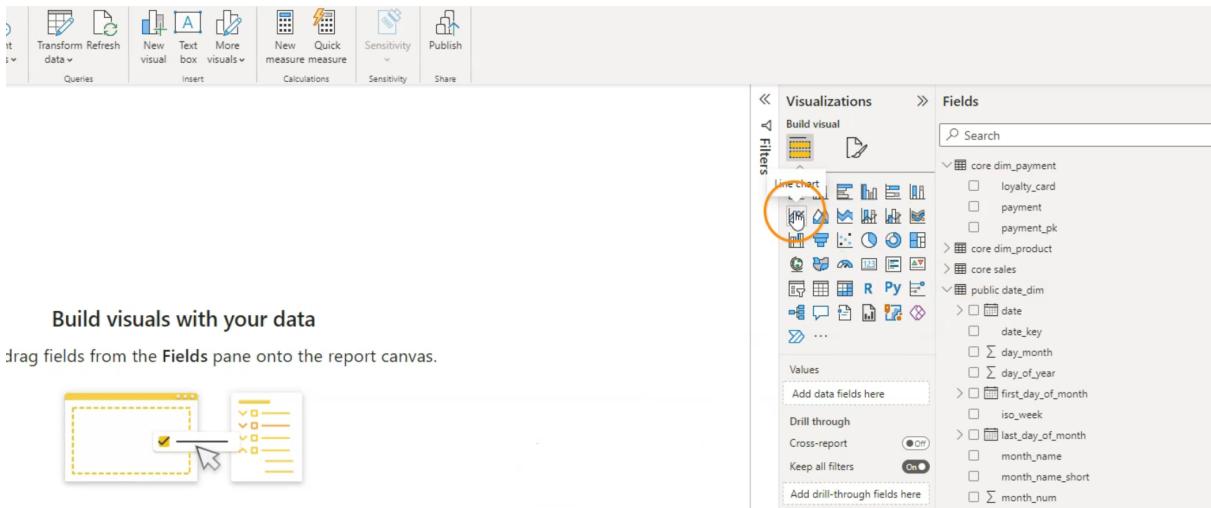
- And we want to see that now also live in action. So this is basically now also kind of serving as a data mart where we are just getting some specific tables that we need for our use case.



- 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.

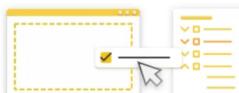


- And with that, we have now basically set up our star schema with the fact table in the middle and all of the dimensions around. So this is now really nice, and we want to see how does the data look like now for the end users and how is it to visualize the data?
- For example, we can just click on this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

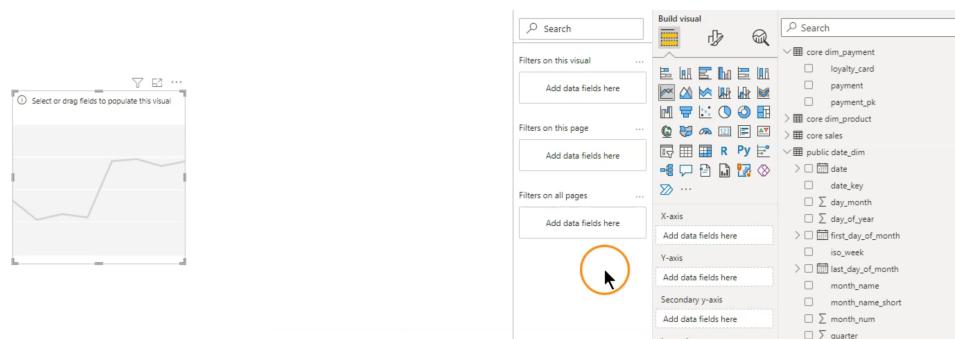


Build visuals with your data

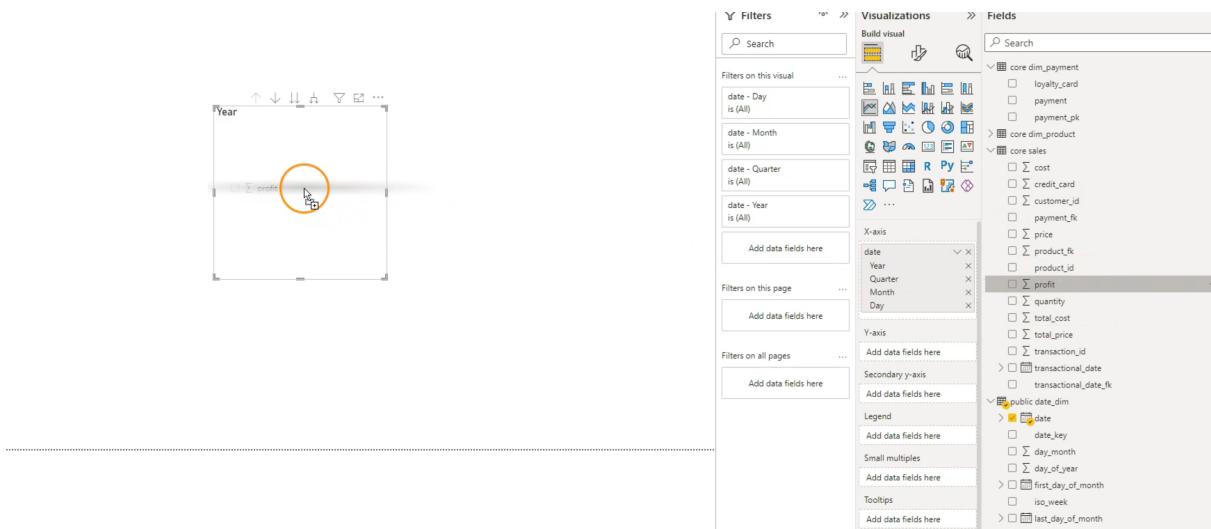
drag fields from the Fields pane onto the report canvas.



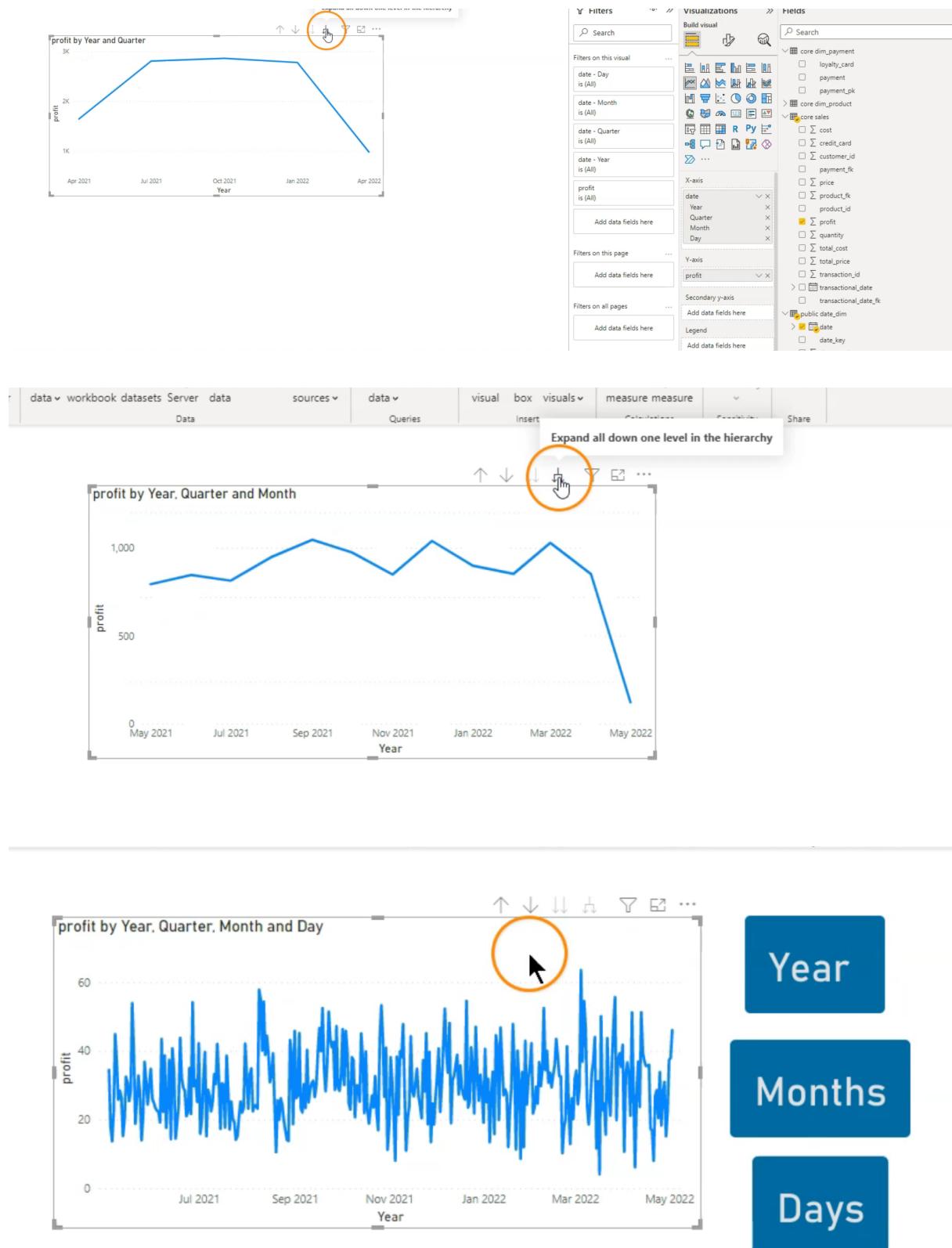
- So we can now just drag and drop the date into this visual from the date table. And of course, we can now also just drag in from our fact table some measures.



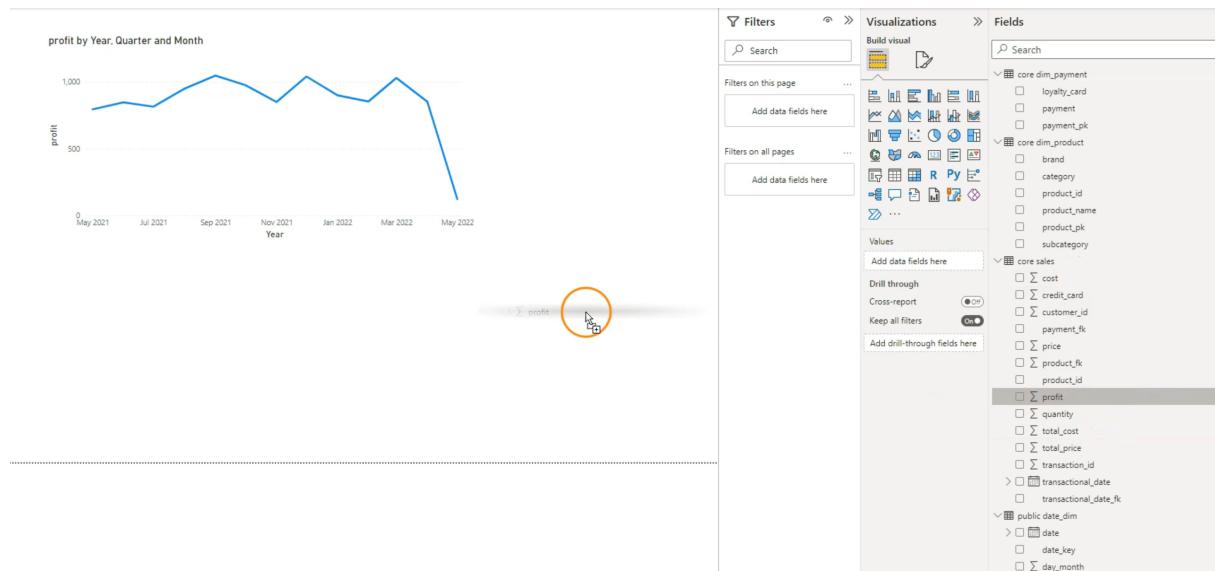
- So for example, the profit. We can again just drag and drop it and now we can just make this a little bit larger and we can now also click on this top button to drill down into the different levels.



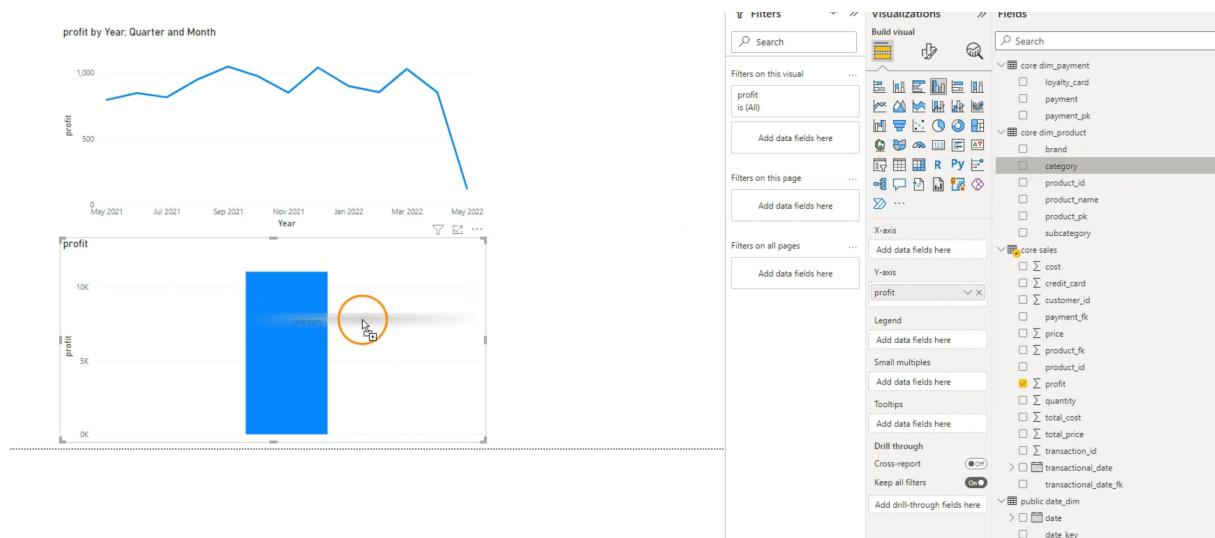
- So like this we can see exactly how the performance is over time. And of course the same, we can now also easily do with the product category.

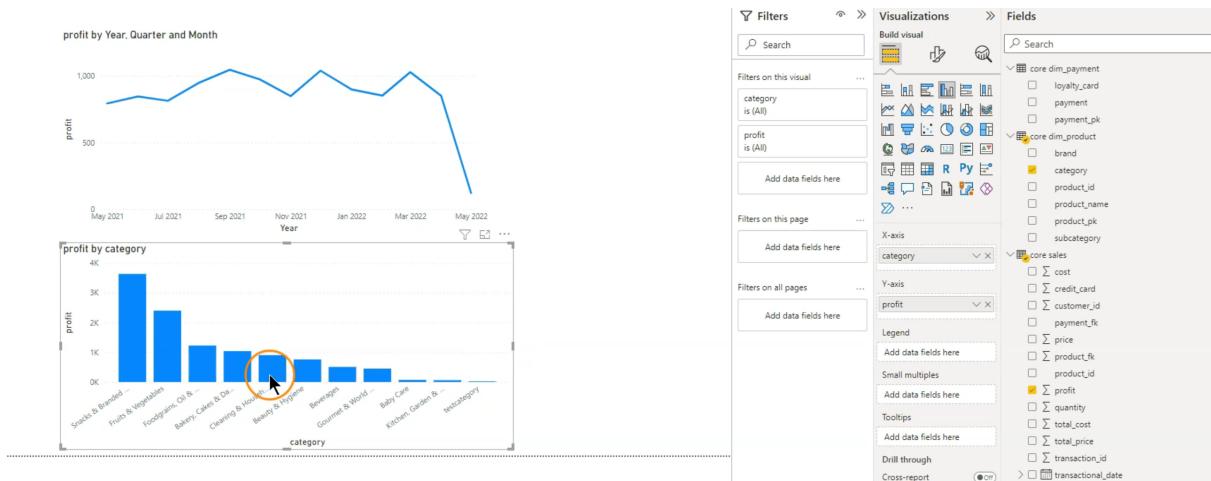


- So for example, we want to see what is the profit. We can just drag it into this open canvas.

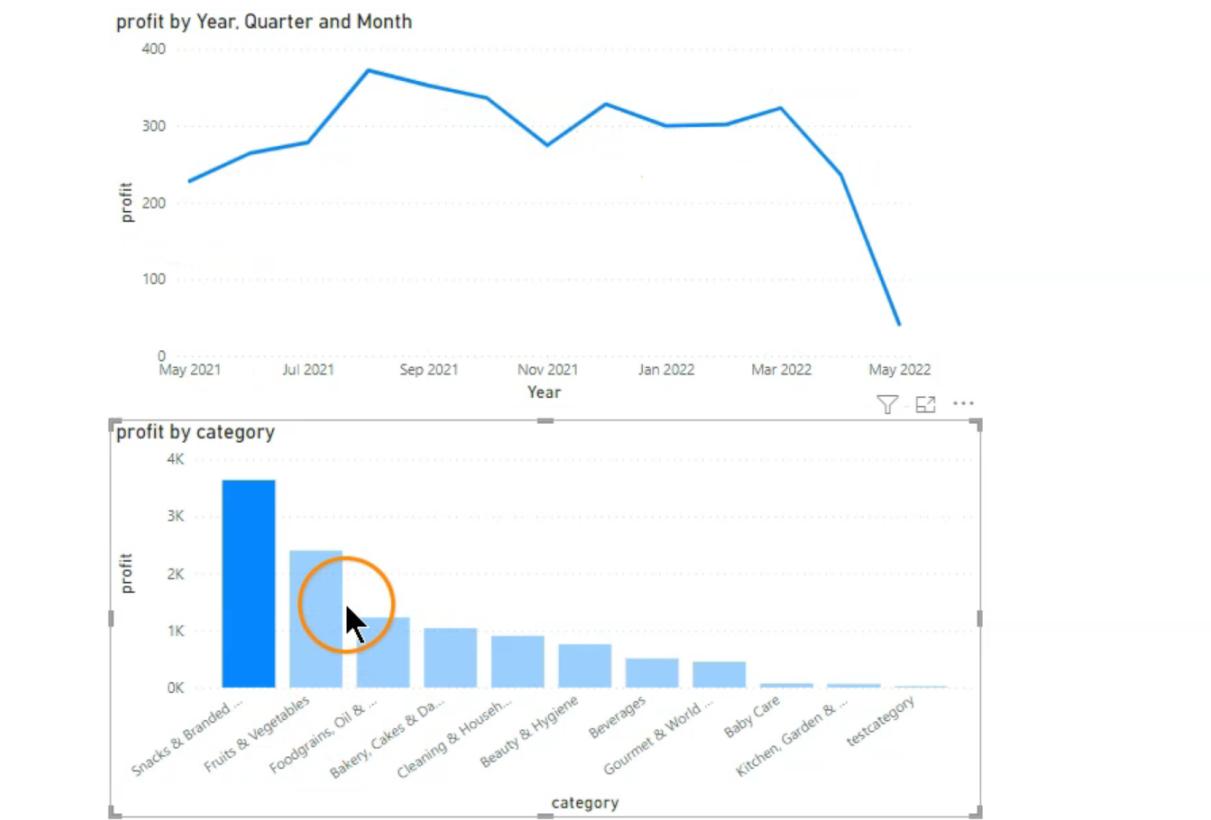


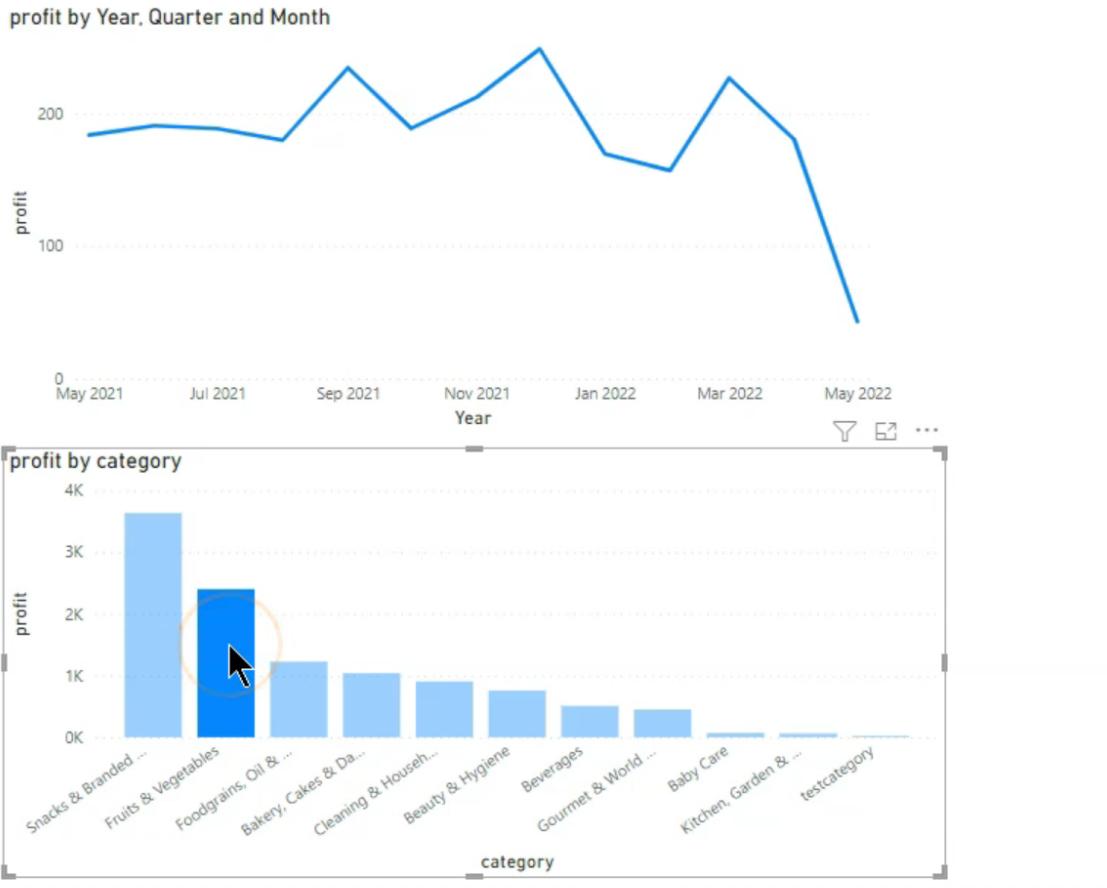
- And then we can also just drag in the category again, just into this visual. And we can like this easily now analyze also the profit by category, and we can now even use it very interactively.





- 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.