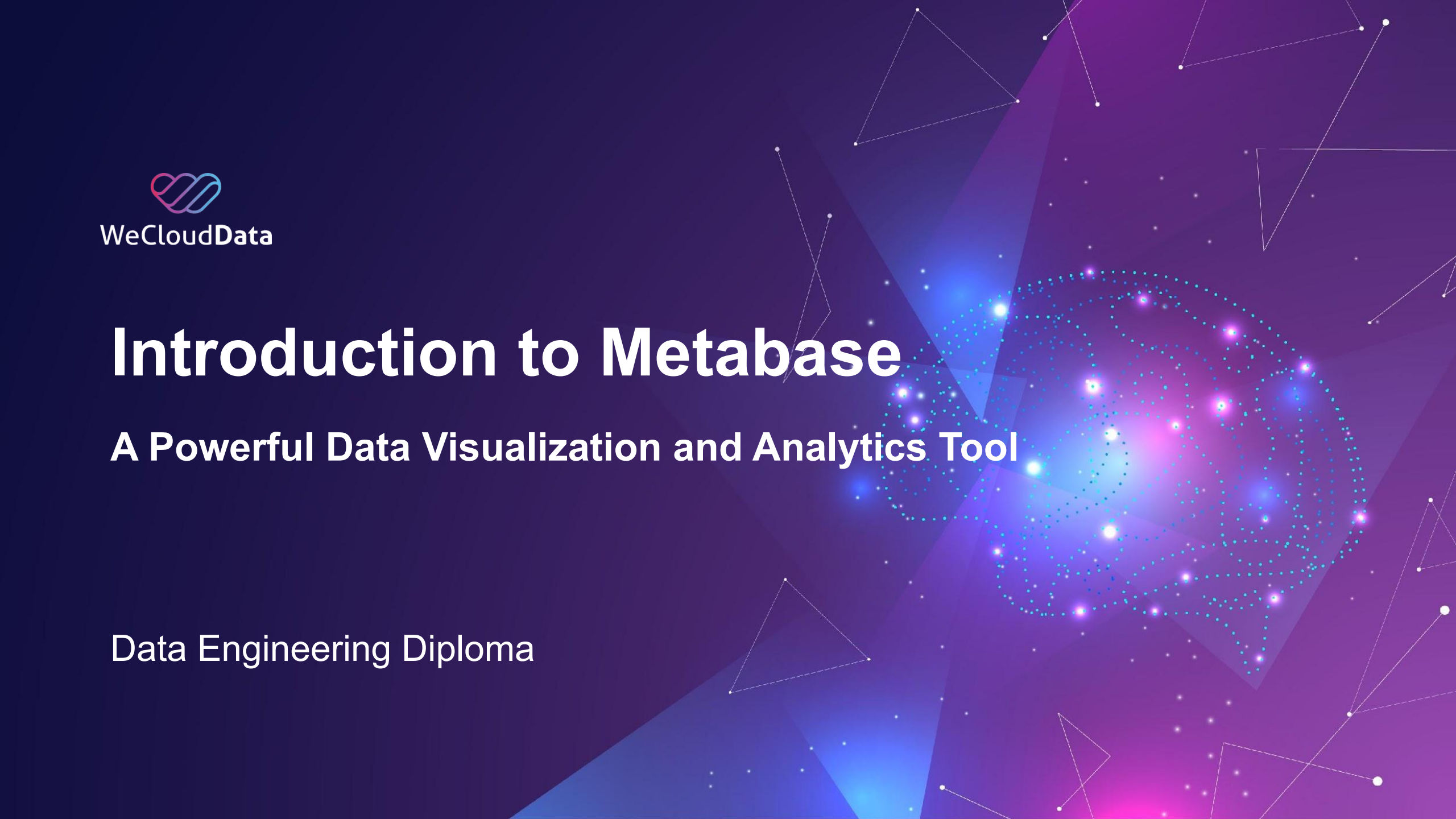



# Introduction to Metabase

## A Powerful Data Visualization and Analytics Tool

Data Engineering Diploma

- ❑ **What is Metabase?**
- ❑ Installing Metabase using Docker on Amazon EC2
- ❑ Connecting your Database to Metabase
- ❑ Data Fundamentals
- ❑ Creating Collections in Metabase
- ❑ Creating Questions in Metabase
- ❑ Creating a Dashboard in Metabase

# Agenda.

# What is Metabase?

## Metabase 

- an open-source business intelligence and analytics tool that is used by businesses and organizations of all sizes across industries like healthcare, finance, and e-commerce.
- popular among data analysts, business users, and developers who seek a powerful yet user-friendly tool for data exploration and visualization.

- ❑ What is Metabase ?
- ❑ **Installing Metabase using Docker on Amazon EC2**
- ❑ Connecting your Database to Metabase
- ❑ Data Fundamentals
- ❑ Creating Collections in Metabase
- ❑ Creating Questions in Metabase
- ❑ Creating a Dashboard in Metabase

# Agenda.

# Installing Metabase using Docker on Amazon EC2

- Launch an EC2 Instance
  - Log in to your AWS Management Console
  - Navigate to EC2 dashboard
  - Click “Launch Instances” and follow the wizard to select an Ubuntu AMI (which is commonly used for Docker) or another suitable AMI
- Connect to the EC2 Instance
  - Once your instance is running, connect to it using SSH
- Install Docker

# Installing Metabase using Docker on Amazon EC2

- Download and Run Metabase Docker Container
  - Pull the Metabase Docker image from Docker Hub
    ***docker pull metabase/metabase:latest***
  - Run the Metabase Docker container, mapping the container's port 3000 to the host's port 3000, and specifying a volume for data persistence
    ***docker run -d -p 3000:3000 --name metabase metabase/metabase***
    - Or, if you want to store the metadata of Metabase in your own database

```bash
docker run -d -p 3000:3000 \
  -e "MB_DB_TYPE=postgres" \
  -e "MB_DB_DBNAME=<metabasea-db>" \
  -e "MB_DB_PORT=5432" \
  -e "MB_DB_USER=<name>" \
  -e "MB_DB_PASS=<password>" \
  -e "MB_DB_HOST=<my-database-host>" \
   --name metabase metabase/metabase
```

# Installing Metabase using Docker on Amazon EC2

- Access Metabase Web Interface
  - Open a web browser and navigate to your EC2 instance's public IP address or DNS name, along with port 3000
    ***https://<your-instance-ip>:3000***
  - follow the prompts to configure Metabase (language selection, database setup, admin account)
- Configure Security Groups and Firewall
  - Ensure that your EC2 instance's security group allows incoming traffic on port 3000

✻ This guide provides a basic overview, and you may need to adjust steps based on your specific requirements. Always refer to the official documentation for Docker and Metabase for the latest and most detailed instructions.

- ❑ What is Metabase ?
- ❑ Installing Metabase using Docker on Amazon EC2
- ❑ **Connecting your Database to Metabase**
- ❑ Data Fundamentals
- ❑ Creating Collections in Metabase
- ❑ Creating Questions in Metabase
- ❑ Creating a Dashboard in Metabase

**Agenda.**

# Connecting your Database to Metabase

## Supported databases

The databases listed below have official drivers maintained by the Metabase team. Customers on **paid plans** will get official support.

- **Amazon Athena**
- **BigQuery** (Google Cloud Platform)
- **Druid**
- **MongoDB (version 4.2 or higher)**
- **MySQL (version 5.7 or higher, as well as MariaDB version 10.2 or higher)**
- **Oracle**
- **PostgreSQL**
- **Presto**
- **Redshift (Amazon Web Services)**
- **Snowflake**
- **SparkSQL**
- **SQL Server**
- **SQLite**
- **Vertica**

* If you don't see your database listed here, see **partner and community drivers**.
* As of version 46.6.4, Metabase **no longer supports H2 connections**. But Metabase still ships with an H2 database to include an embedded application database, as well as to provide some sample data out of the box.



Reference: https://www.metabase.com/docs/latest/databases/connecting#connecting-to-supported-databases

# Connecting your Database to Metabase

## Adding a database connection

- To add a database connection, click on the gear icon  in the top right, and navigate to Admin settings > Databases > Add a database.
- Fill out the fields for that database, and click Save changes at the bottom.
    - The connection settings vary from database to database. For the purpose of learning, we are using the Snowflake database.
- After entering the connection details, click the “Save” button to ensure that the Metabase can connect to your database successfully.

Reference: https://www.metabase.com/docs/latest/databases/connecting#adding-a-database-connection

- ❑ What is Metabase ?
- ❑ Installing Metabase using Docker on Amazon EC2
- ❑ Connecting your Database to Metabase
- ❑ **Data Fundamentals**
- ❑ Creating Collections in Metabase
- ❑ Creating Questions in Metabase
- ❑ Creating a Dashboard in Metabase

# Agenda.

# Data Fundamentals

If you are planning to dive into Metabase, having a solid grasp of some essential database terminology can be incredibly useful.

- **Tables**, Columns, Rows
  - Databases consist of one or more tables. Each table features at least one column and one row. Rows are composed of cells, and each cell holds a value that aligns with its respective column.
- **Columns and Fields**
  - In Metabase, the terms 'Columns' and 'Fields' are commonly used *interchangeably*, particularly when a column contains data from a single field.
  - Column: typically a list of uniform values, often sourced from a single field.
    - Example: the 'product_id' column holds ids.
  - Field: represents a specific data element within a table.
    - Example: the 'product_id' field stores individual identification codes.
- Keys
  - A primary key serves as a unique identifier for each row within a table.

Reference: https://www.metabase.com/learn/databases/database-basics

# Data Fundamentals

**Common Data Types**

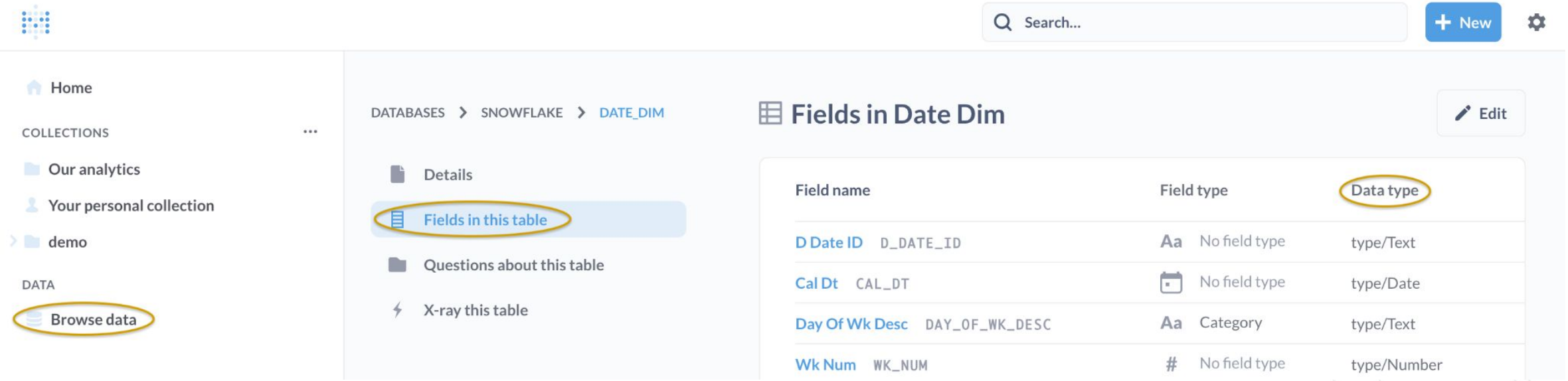
- String Types (e.g., text, char, vchar)
  - These fields hold text snippets, commonly called "strings". They may include letters, numbers, and special characters.
- Numerical Types (e.g., integer, float, decimal)
  - These fields contain numerical data. Integers hold whole numbers, while float and decimal types can store fractional numbers.
- Temporal Types (e.g., timestamp, date, time)
  - These specialized fields store date and time information, often referred to as "timestamps".
- Boolean Types
  - These fields can have only one of two values, generally True or False.

Reference: https://www.metabase.com/learn/databases/data-types-overview

# Data Fundamentals

## Data Types

In Metabase, the data type for each field can be viewed by going to the Data Browser. There, you can click on the gray book icon beside the table name to open the Data Reference page. Once on this page, select 'Fields in this table' from the left sidebar. The data types are displayed in the third column for each field.




Reference: https://www.metabase.com/learn/databases/data-types-overview

# Data Fundamentals

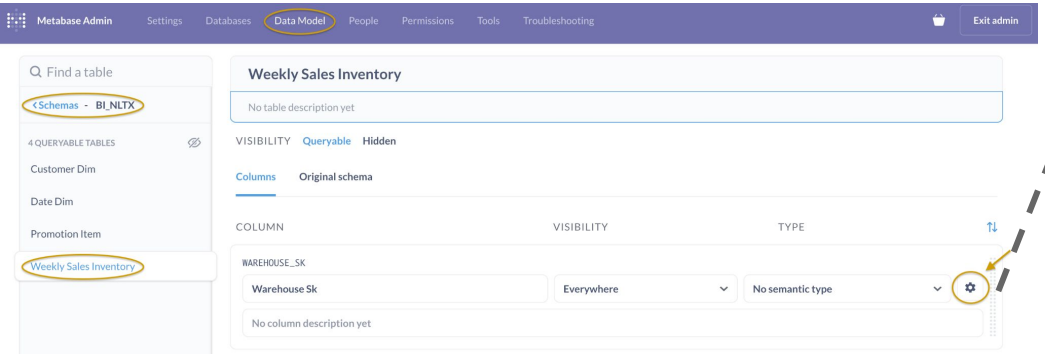
## Metadata

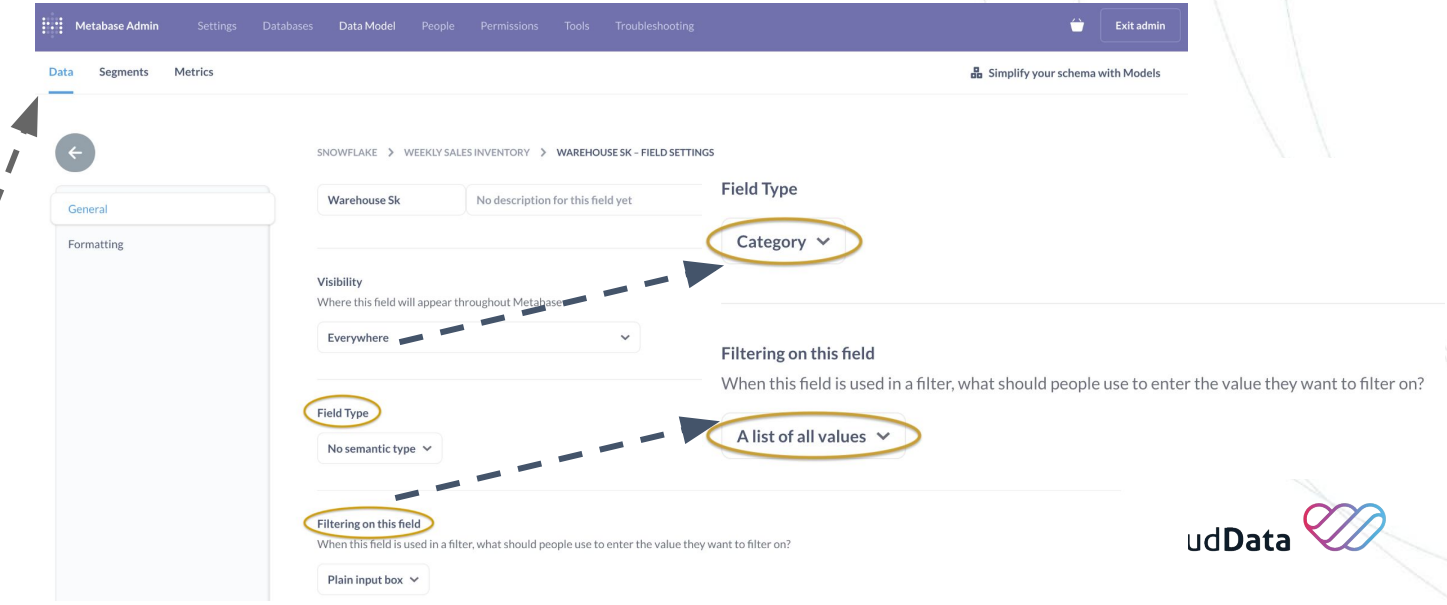
In Metabase, **admins** have the ability to modify the display names, descriptions, and semantic types (commonly referred to as field types) for fields. These adjustments give users additional context regarding the purpose of each field and guide Metabase on how to interpret various fields.

Properly classifying field types allows Metabase to automatically choose the appropriate chart format, generate maps from geographical data, or present URLs as clickable links. A list of available Metabase field types can be found on this page.

## Casting to a Specific Data Type

- Click on the gear icon in the top right, and navigate to Admin settings > Data Model.
- Locate your database and table.
- Click on the gear icon at the right of a column's settings box.
- Scroll to cast to a specific data type.
- Select a casting option.




Reference: https://www.metabase.com/docs/latest/data-modeling/metadata-editing#field-type

udData

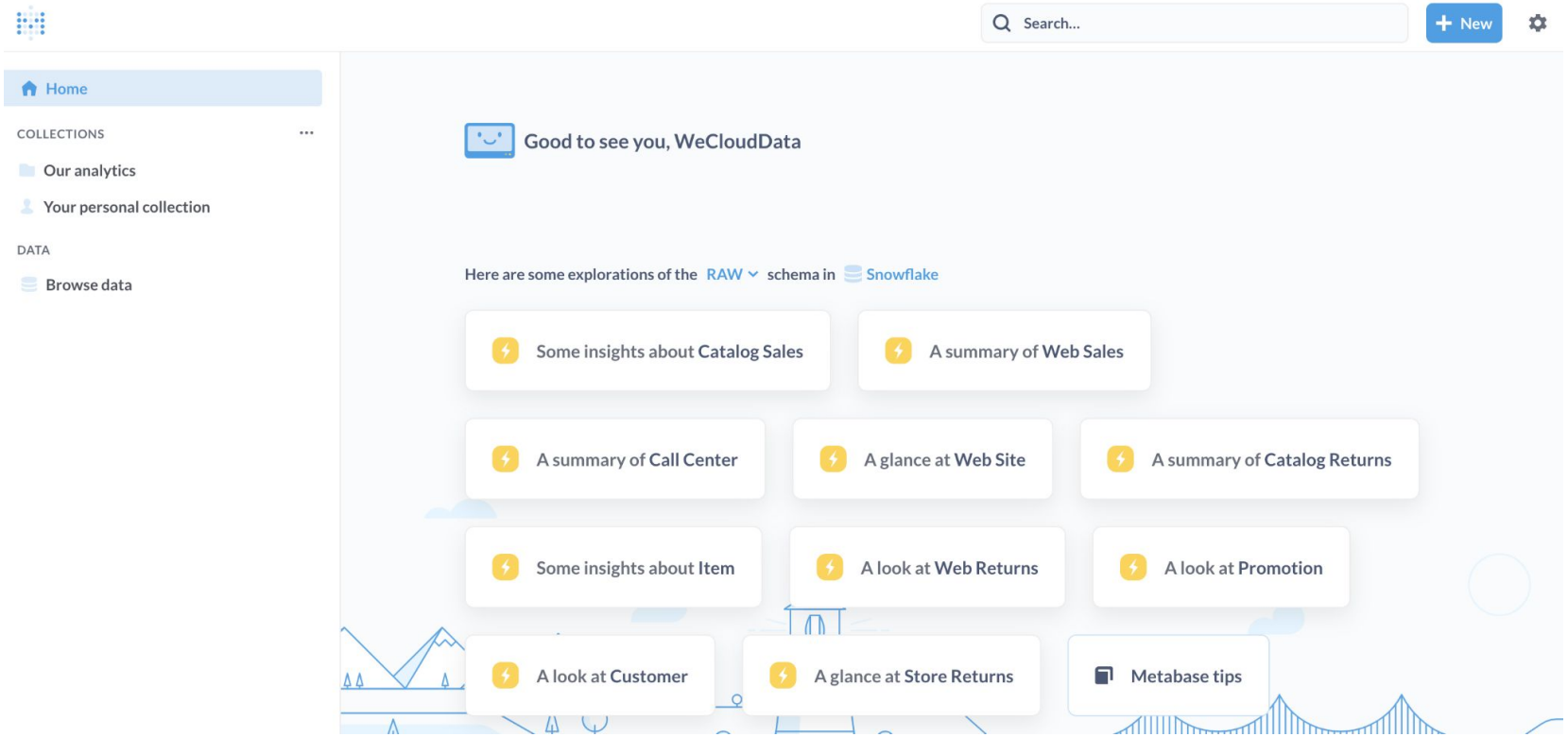
- ❑ What is Metabase ?
- ❑ Installing Metabase using Docker on Amazon EC2
- ❑ Connecting your Database to Metabase
- ❑ Data Fundamentals
- ❑ **Creating Collections in Metabase**
- ❑ Creating Questions in Metabase
- ❑ Creating a Dashboard in Metabase

# Agenda.

# Creating Collections in Metabase

Congratulations! If you have successfully completed all the steps, you are now ready to begin creating queries and dashboards by utilizing this database as a data source. When creating queries, you will be able to choose the database connection you have just established.

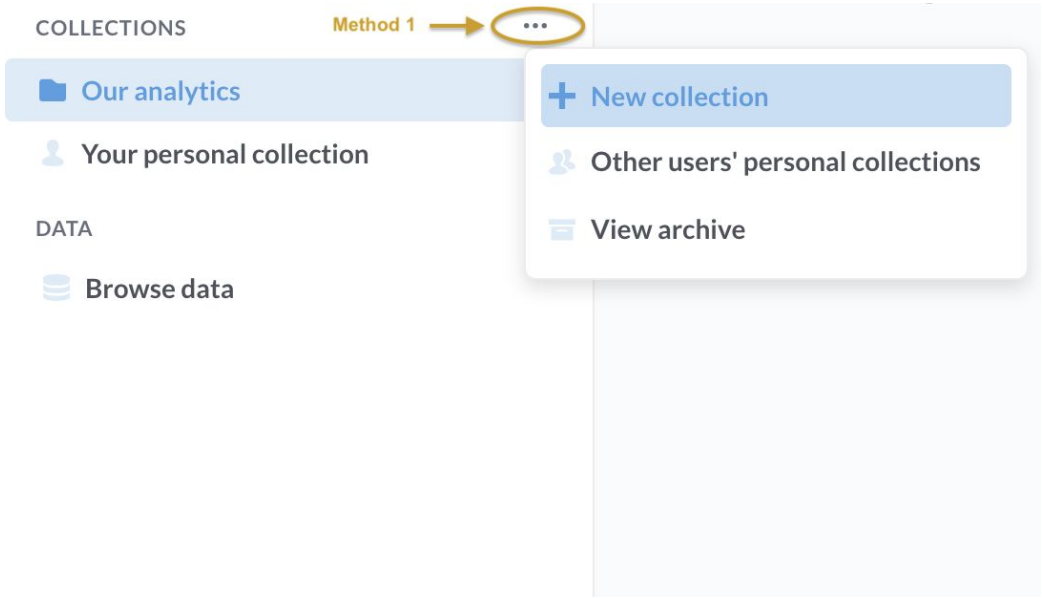
If you are currently on the Admin settings page, click on the “Exit admin” option in the top right corner.

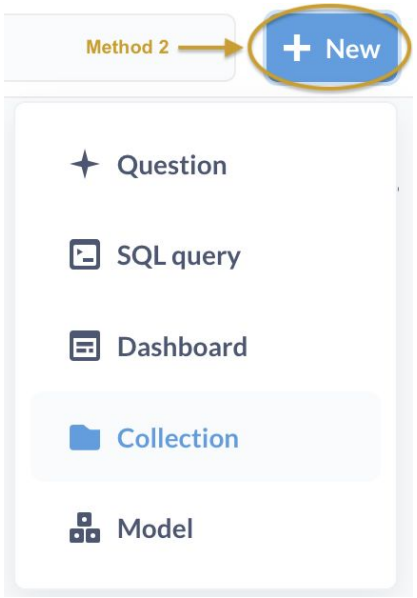


# Creating Collections in Metabase

In Metabase, collections are organizational structures used to group related dashboards, questions (queries), and other content together for easier management and navigation. Collections allow you to locally organize your analytics and data visualization assets in a way that makes sense for your specific use case.

For the purpose of learning, our initial step involves creating a collection within 'Our analytics' titled 'TPC-DS project'. Within this collection, we'll proceed to generate two additional collections, known as 'Dashboards' and 'Questions (Queries)', both nested under the 'TPC-DS project' collection.



- ❑ What is Metabase ?
- ❑ Installing Metabase using Docker on Amazon EC2
- ❑ Connecting your Database to Metabase
- ❑ Data Fundamentals
- ❑ Creating Collections in Metabase
- ❑ **Creating Questions in Metabase**
- ❑ Creating a Dashboard in Metabase

# Agenda.

# Creating Questions in Metabase

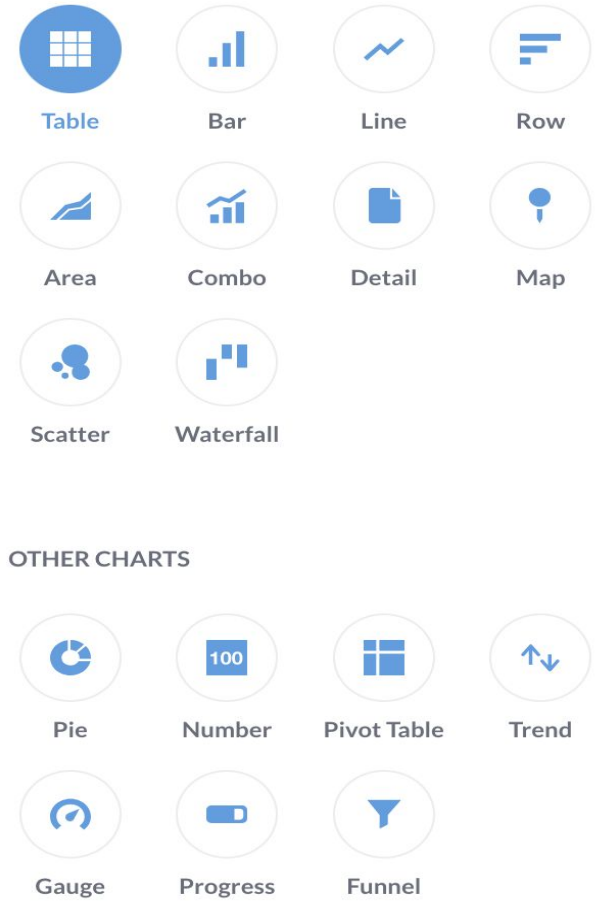
To ask a question in Metabse, click the '+New' button in the top right corner of the main navigation bar, and select either

- **Question (aka. the query builder)**
- **SQL query**

You can initiate a question from one of the following sources:

- **Raw Data**
  - specify both the database and the table within that database
- **A Saved Question**
  - The results of any previously saved question can serve as the foundation for a new inquiry.
- **A Model**
  - A model is a uniquely crafted saved question designed as an ideal foundation for further questions. These are occasionally referred to as derived tables, as they often aggregate data from several raw tables.

✻ Kindly observe that with dbt transformations, building models in Metabase is no longer necessary.




✻ please check this page for more information.

# Creating Questions in Metabase

**Question (aka. the query builder)**

- The query builder's notebook editor has three default steps:
  - Picking Data
  - Filtering
  - Summarizing and Grouping By

**SQL query**

In Metabase's **Query Builder**, you can apply filters directly as you construct queries, while in **SQL queries**, you must explicitly build and map filter conditions to table columns using SQL syntax.

***Field Filter***

**Distinguishing Field Filters from simple Text, Number, and Date variables**

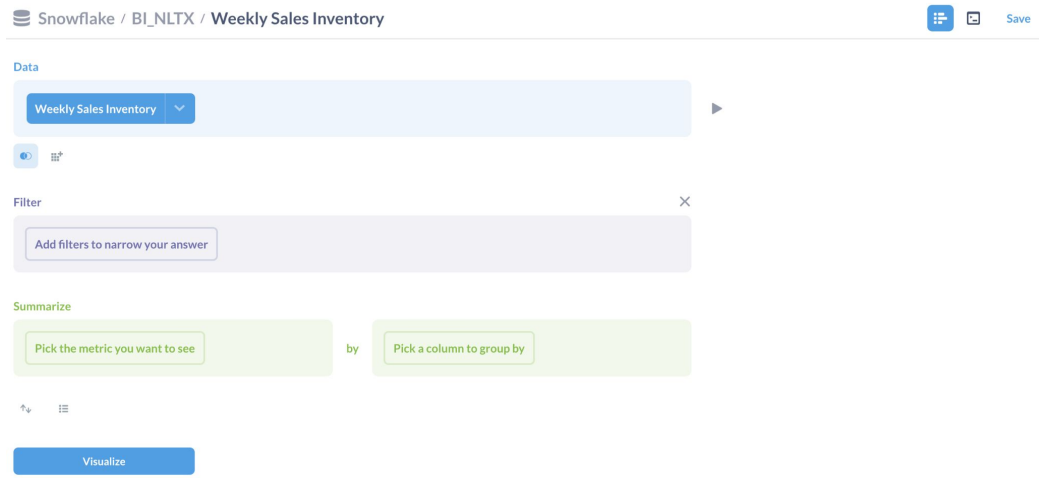
- Field Filters are optional by default. If no value is given, the SQL query will run as if the Field Filter didn't exist. You do, however, have the option to require a value.
- ***Field Filters won't work with table aliases.*** Since Field Filters rely on metadata about columns in your tables (and the specific names of those tables), the filters can't "know" that you've aliased a table. And depending on the database you're using, you may need to include the full schema path in the **FROM** clause.
- Field Filters use a special syntax so they can handle SQL code behind the scenes. You simply supply a Field Filter to a **WHERE** clause (without a column or operator) and the Field Filter will manage the SQL code for you. This allows the code to account for multiple selections people make in the filter widget.

# Creating Questions in Metabase

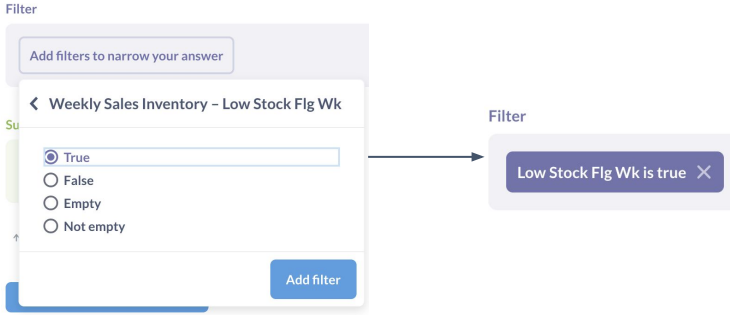
## Question (aka. the query builder)

***Demo #1***: Detect items experiencing low stock levels, along with their corresponding week and warehouse numbers, marked as "True".

- Once you are in the query builder interface, choose the data source from which you want to query. Then select the specific table against which you want to build your query. You'll see something like this:

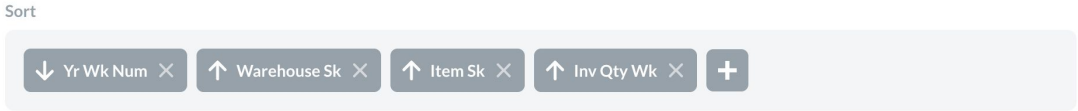



- Filter on conditions where LOW_STOCK_FLG_WK is "True".




- Summarize the data based on YR_WK_NUM, WAREHOUSE_SK, ITEM_SK, INV_QTY_WK, WKS_SPLY, and LOW_STOCK_FLG_WK.


- Order the results based on the most recent YR_WK_NUM, and sort the values of WAREHOUSE_SK, ITEM_SK, INV_QTY_WK in ascending order.




- Utilize data as a table with conditional formatting based on the ITEM_SK column.


| Yr Wk Num | Warehouse Sk | Item Sk | Inv Qty Wk | Wks Sply | Low Stock Flg Wk |
|---|---|---|---|---|---|
| 202,337 | 1 | 2,600 | 0 | 0 | true |
| 202,337 | 1 | 3,554 | 0 | 0 | true |
| 202,337 | 1 | 6,134 | 0 | 0 | true |
| 202,337 | 1 | 6,967 | 0 | 0 | true |
| 202,337 | 1 | 8,696 | 0 | 0 | true |

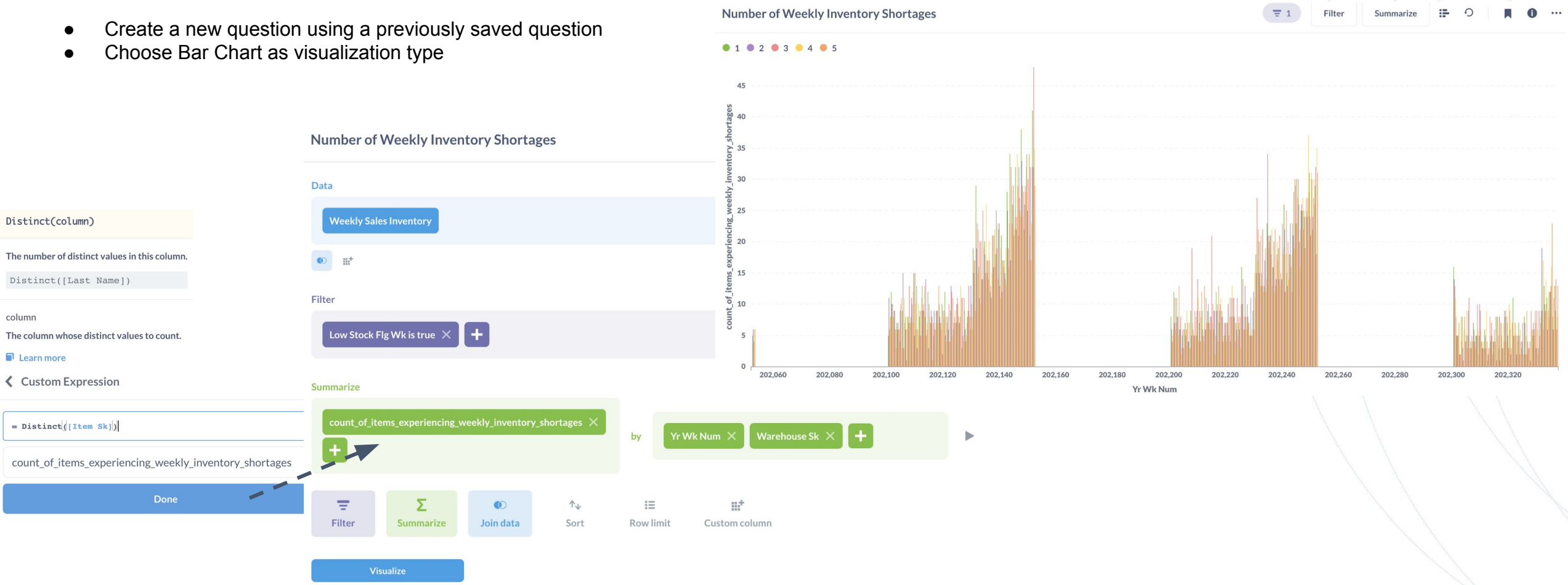
Reference: https://www.metabase.com/docs/latest/questions/query-builder/introduction , https://www.metabase.com/learn/questions/custom-expressions#further-reading

# Creating Questions in Metabase

## Question (aka. the query builder)

***Demo #2***: Find the total count of items experiencing weekly inventory shortages, along with their corresponding week and warehouse numbers.

- Create a new question using a previously saved question
- Choose Bar Chart as visualization type




Reference: https://www.metabase.com/docs/latest/questions/query-builder/introduction , https://www.metabase.com/learn/questions/custom-expressions#further-reading

WeCloudData

# Creating Questions in Metabase

**Question (aka. the query builder)**

- The query builder's notebook editor has three default steps:
  - Picking Data
  - Filtering
  - Summarizing and Grouping By

**SQL query**

In Metabase's **Query Builder**, you can apply filters directly as you construct queries, while in **SQL queries**, you must explicitly build and map filter conditions to table columns using SQL syntax.

***Field Filter***

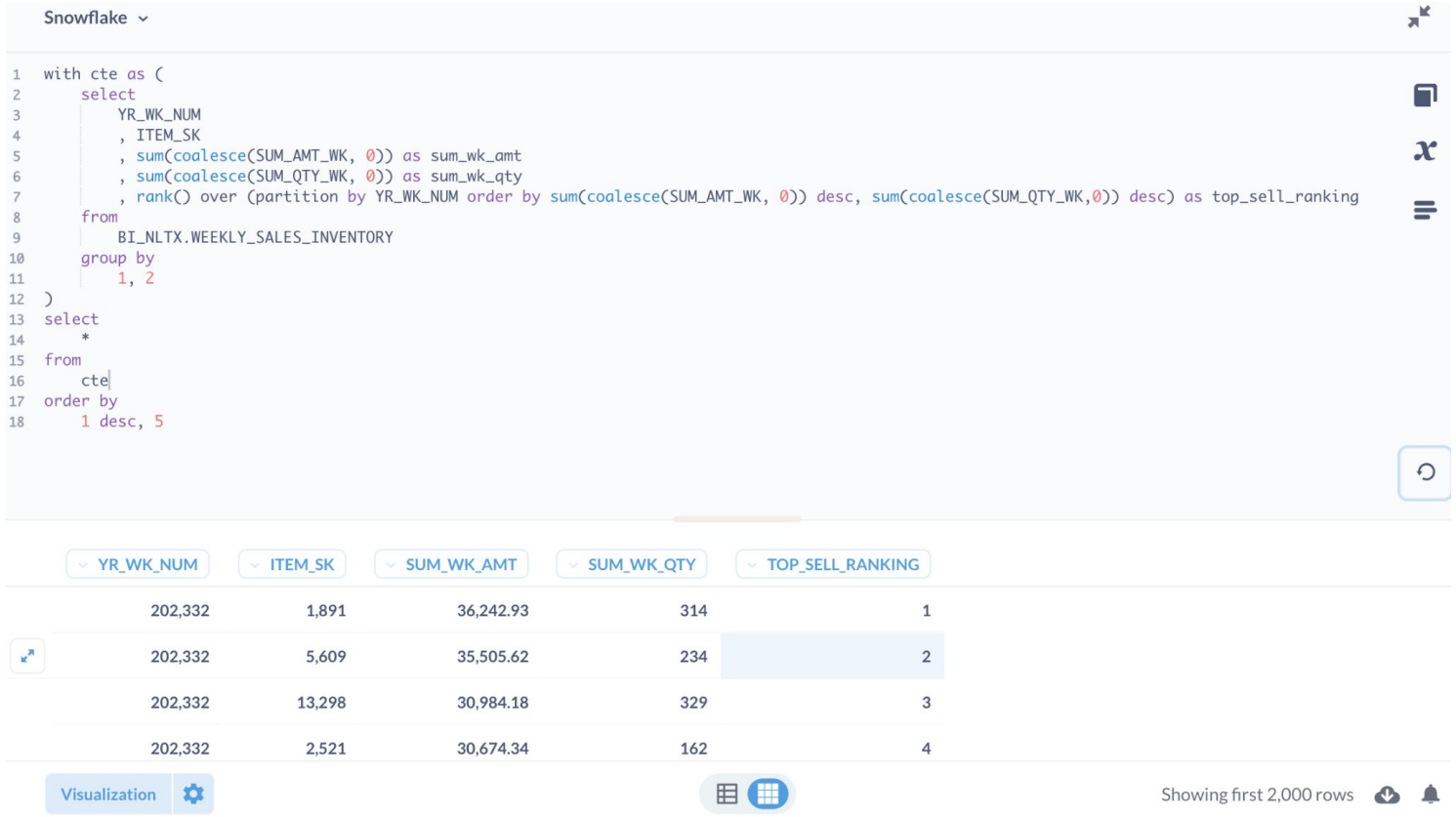
**Distinguishing Field Filters from simple Text, Number, and Date variables**

- Field Filters are optional by default. If no value is given, the SQL query will run as if the Field Filter didn't exist. You do, however, have the option to require a value.
- ***Field Filters won't work with table aliases.*** Since Field Filters rely on metadata about columns in your tables (and the specific names of those tables), the filters can't "know" that you've aliased a table. And depending on the database you're using, you may need to include the full schema path in the **FROM** clause.
- Field Filters use a special syntax so they can handle SQL code behind the scenes. You simply supply a Field Filter to a **WHERE** clause (without a column or operator) and the Field Filter will manage the SQL code for you. This allows the code to account for multiple selections people make in the filter widget.

# Creating Questions in Metabase

**SQL query**
After clicking 'SQL query', you'll see an editor where you can write and run queries in SQL or your database's native querying language.

***Demo***: Identify the highest performing items of the week by analyzing sales amounts and quantities. Order the results based on the most recent YR_WK_NUM, and sort the value for the number of TOP_SELL_RANKING in ascending order.

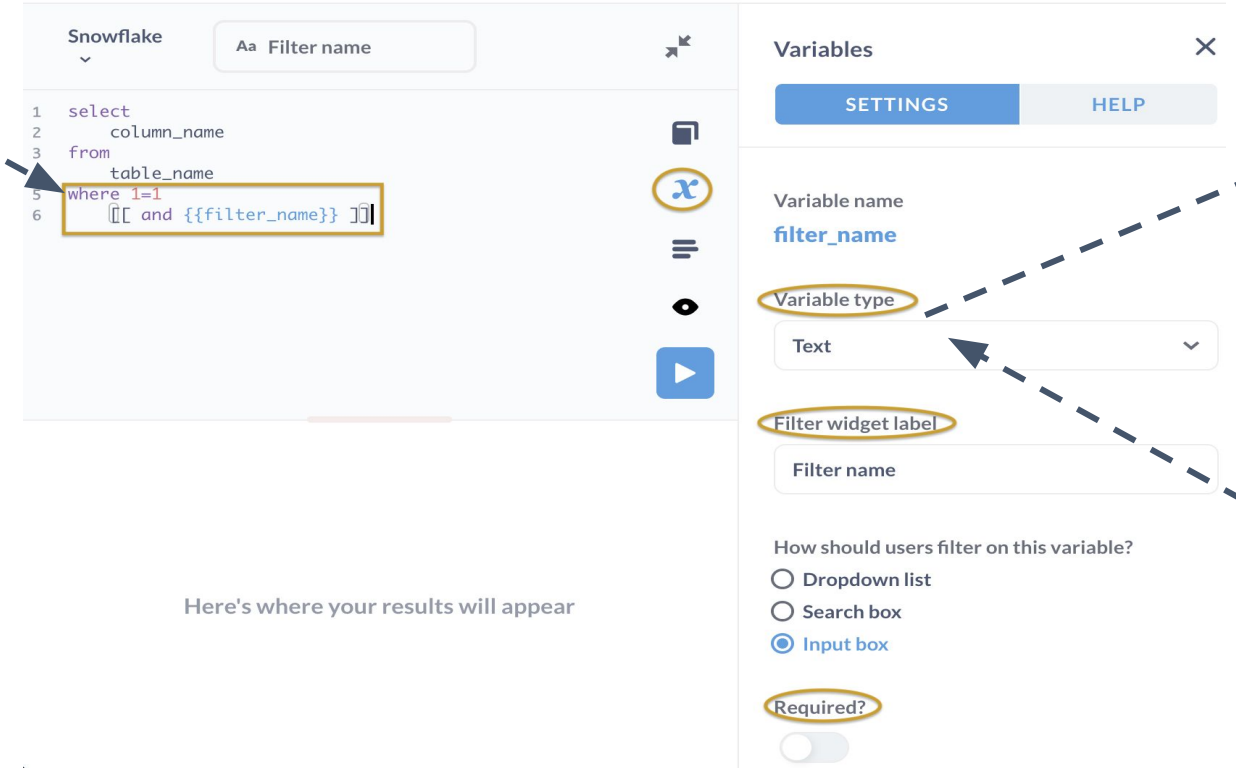



WeCloudData

# Creating Questions in Metabase

## How To Use Optional Filters in an SQL Query?

To make a filter optional,
you can include it in your SQL query within a WHERE clause using **[[ ... ]]** to denote that it is optional.




You'll need to **select the type of variable** you want to use from among text, number, date, or field filter.

If you opt for a field filter as the variable type, you'll also be required to specify the database table and column to which it should be mapped.

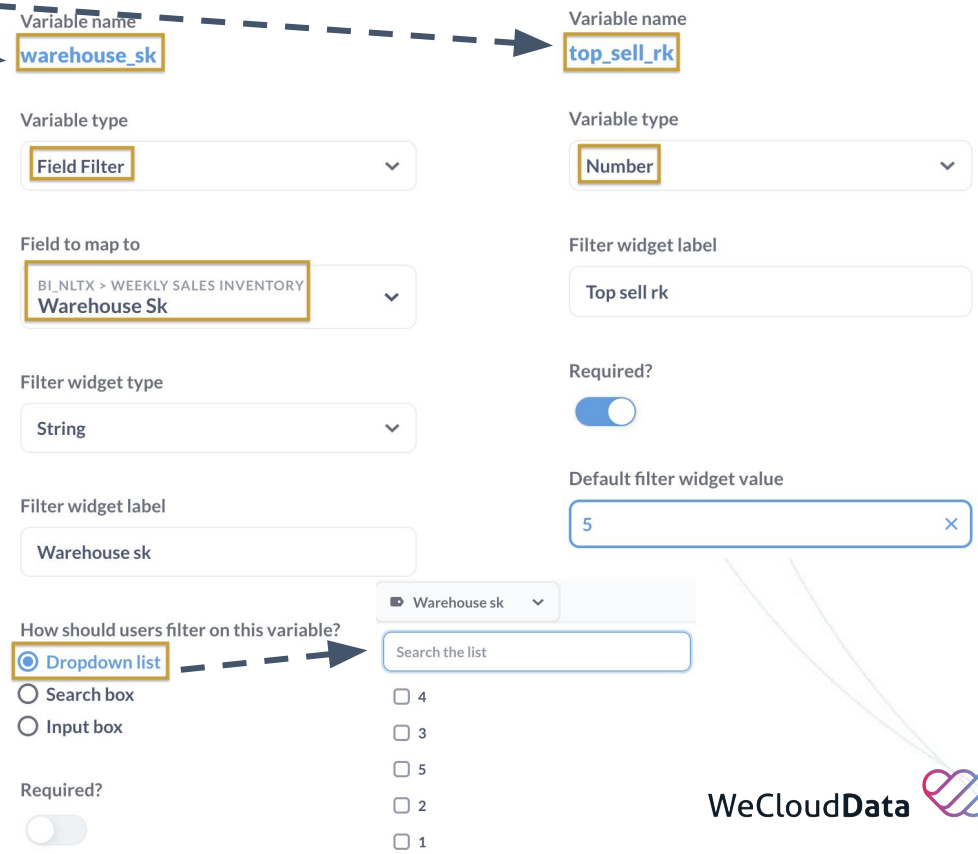
# Creating Questions in Metabase

**SQL query (cont.)**

***Demo***: Identify the highest performing items of the week by analyzing sales amounts and quantities. Order the results based on the most recent YR_WK_NUM, and sort the value for the number of TOP_SELL_RANKING in ascending order.

Snowflake | Warehouse sk | Top sell rk: 5

```
with cte as (
    select
        YR_WK_NUM
        , ITEM_SK
        , sum(coalesce(SUM_AMT_WK, 0)) as sum_wk_amt
        , sum(coalesce(SUM_QTY_WK, 0)) as sum_wk_qty
        , rank() over (partition by YR_WK_NUM order by sum(coalesce(SUM_AMT_WK, 0)) desc, sum(coalesce(SUM_QTY_WK,0)) desc) as top_sell_ranking
    from
        BI_NLTX.WEEKLY_SALES_INVENTORY
    where 1=1
        [[and {{warehouse_sk}}]]
    group by
        1, 2
)
select
    *
from
    cte
where 1=1
    [[and top_sell_ranking<={{top_sell_rk}}]]
order by
    1 desc, 5
```



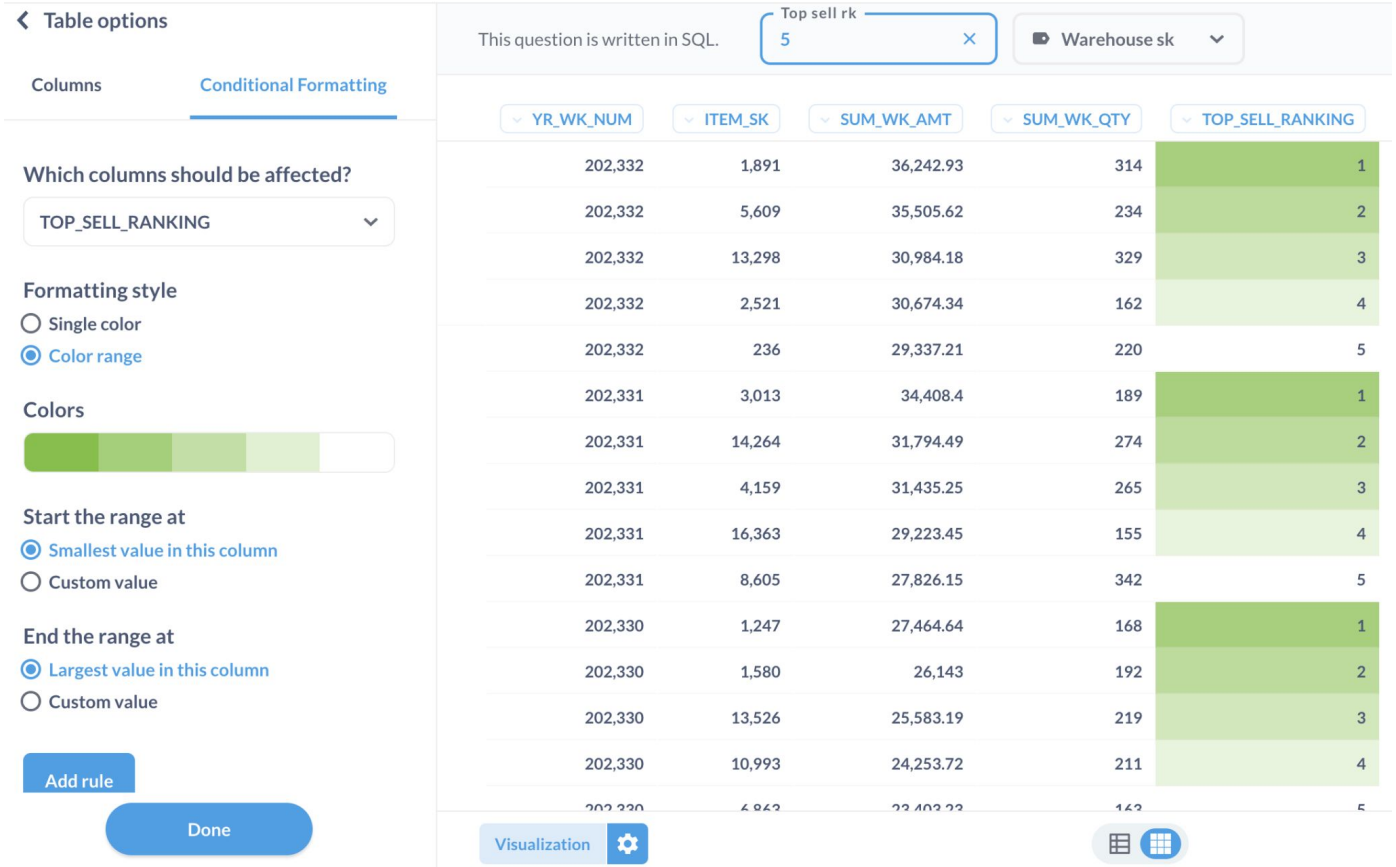

WeCloudData

# Creating Questions in Metabase

## SQL query (cont.)

***Demo***: Identify the highest performing items of the week by analyzing sales amounts and quantities. Order the results based on the most recent YR_WK_NUM, and sort the value for the number of TOP_SELL_RANKING in ascending order.

In our demo, we want to utilize data as a table with conditional formatting based on the TOP_SELL_RANKING column.



- ❑ What is Metabase ?
- ❑ Installing Metabase using Docker on Amazon EC2
- ❑ Connecting your Database to Metabase
- ❑ Data Fundamentals
- ❑ Creating Collections in Metabase
- ❑ Creating Questions in Metabase
- ❑ **Creating a Dashboard in Metabase**

# Agenda.

# Creating a Dashboard in Metabase

### Create a New Dashboard

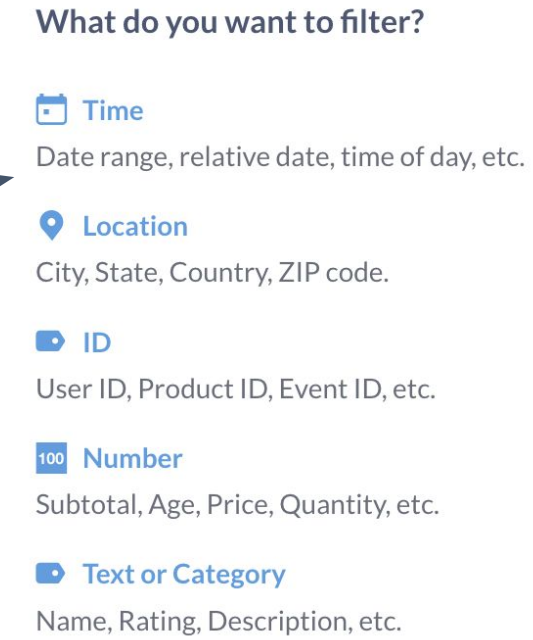
- Go to the Metabase homepage.
- Click on the '+ New' and then select 'Dashboard' button from the top navigation bar.
- Give your dashboard a name and description, and then click 'Create'.

### Add Saved Questions to Dashboard

- On your new dashboard page, click on the icon to edit the dashboard.
- Click on the ➕ and you will be taken to a page where you can select from your saved questions.
- Add the questions you want to include on your dashboard.

### Add Filters to Dashboard

- Click on the to add a filter.
- Choose the type of filter you want to add.
- After adding the filter, you will need to map the filter to the relevant fields in each question by clicking the gear icon beside it.



# Creating a Dashboard in Metabase

TPC-DS demo dashboard

warehouse number | Top-ranking | Lowe | Weeks Supply Nun

Greetings, this demonstration aims to provide you with a brief tutorial on crafting a dashboard within Metabase. We'll be utilizing both the notebook editor and SQL editor for this tutorial.

## Identify the highest and lowest performing items of the week by analyzing sales amounts and quantities.

Items with Highest Weekly Sales

| YR_WK_NUM | ITEM_SK | SUM_WK_AMT | SUM_WK_QTY | TOP_SELL_RANKING |
|---|---|---|---|---|
| 202,337 | 16,255 | 26,025.1 | 97 | 1 |
| 202,337 | 9,775 | 19,788.31 | 213 | 2 |
| 202,337 | 11,389 | 18,816.93 | 99 | 3 |
| 202,337 | 10,429 | 18,463.29 | 93 | 4 |
| 202,337 | 15,025 | 17,782.5 | 75 | 5 |
| 202,336 | 5,779 | 18,261.85 | 95 | 1 |
| 202,336 | 2,072 | 18,240 | 96 | 2 |
| 202,336 | 1,969 | 16,554.48 | 92 | 3 |
| 202,336 | 7,075 | 15,935.79 | 87 | 4 |

Rows 1-9 of 710

Items with Lowest Weekly Sales

| YR_WK_NUM | ITEM_SK | SUM_WK_AMT | SUM_WK_QTY | LEAST_SELL_RANKING |
|---|---|---|---|---|
| 202,337 | 2,659 | 0 | 9 | 1 |
| 202,337 | 9,074 | 0 | 24 | 2 |
| 202,337 | 17,264 | 0 | 26 | 3 |
| 202,337 | 6,530 | 0 | 30 | 4 |
| 202,337 | 4,573 | 0 | 32 | 5 |
| 202,337 | 5,342 | 0 | 70 | 6 |
| 202,337 | 1,220 | 0 | 93 | 7 |
| 202,337 | 2,191 | 0.96 | 1 | 8 |
| 202,337 | 8,594 | 4.36 | 4 | 9 |
| 202,337 | 2,456 | 4.56 | 3 | 10 |

Rows 1-10 of 1421

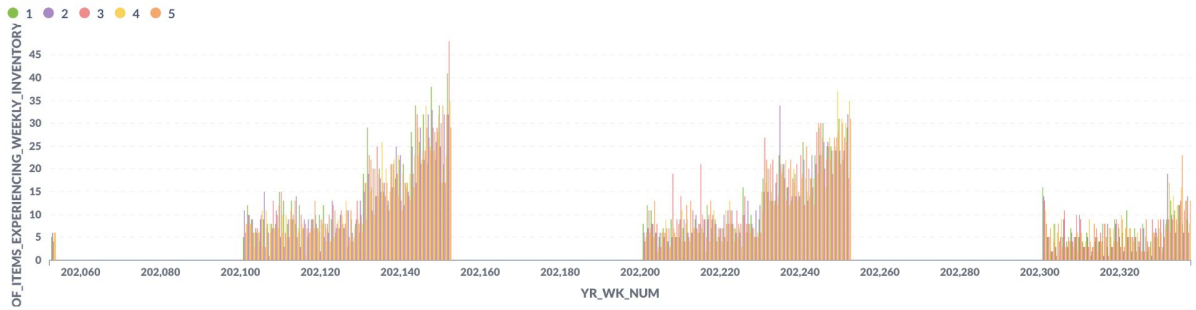
## Detect items experiencing low stock levels, along with their corresponding week and warehouse number, marked as "True."

Weekly Inventory Shortages_

| YR_WK_NUM | WAREHOUSE_SK | ITEM_SK | INV_QTY_WK | WKS_SPLY | LOW_STOCK_FLG_WK |
|---|---|---|---|---|---|
| 202,337 | 1 | 2,600 | 0 | 0 | true |
| 202,337 | 1 | 3,554 | 0 | 0 | true |
| 202,337 | 1 | 6,134 | 0 | 0 | true |
| 202,337 | 1 | 6,967 | 0 | 0 | true |
| 202,337 | 1 | 8,696 | 0 | 0 | true |
| 202,337 | 1 | 11,137 | 0 | 0 | true |
| 202,337 | 1 | 12,061 | 0 | 0 | true |

Rows 1-7 of first 2000

Number of Weekly Inventory Shortages_




WeCloudData

# Useful Resources

- https://www.metabase.com/learn/
- https://www.metabase.com/docs/latest/
- https://www.metabase.com/learn/sql-questions/field-filters
- https://www.metabase.com/learn/dashboards/custom-destinations