

SOP: Data extraction and analysis from DWH using Superset

[Last update: July 2023]

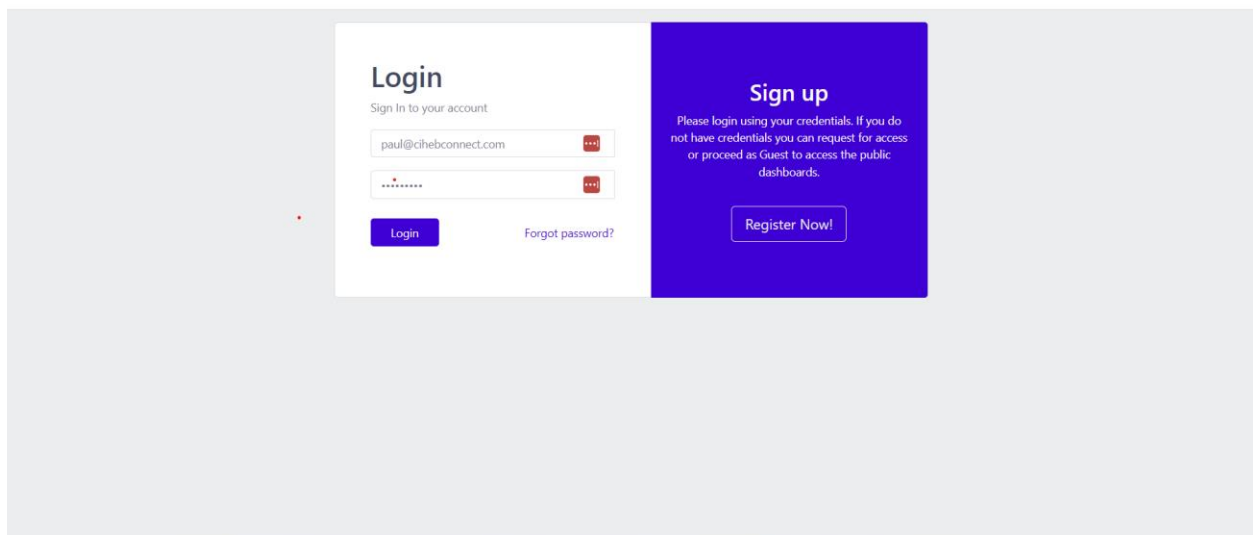
Goal: Facilitate self-service initiative in data extraction from DWH and analysis using Metabase.

Objective:	<i>Provide guidance on how to perform data extraction from DWH platform and analysis using Superset.</i>
Target group:	<i>SI, Health Informatics, M&E, Data Managers</i>
Requirements:	<i>Internet connectivity, Valid login credentials to DWH</i>

Overview

This document acts as a guide to exploring and visualizing the data warehouse database using Apache Superset. Metabase is an open-source self-service platform that one can install and host locally. Alternatively, you can access our hosted installation and can be accessed via <https://dwh.kenyahmis.org/>

Note: Your access to datasets within Superset is tied to your selected registration Partner/Organization.



The screenshot shows the Superset login and sign up interface. On the left, there is a 'Login' section with the text 'Sign In to your account'. Below this, there are two input fields: one for the email address (containing 'paul@cihebconnect.com') and one for the password (masked with dots). A 'Login' button is positioned below the password field, and a 'Forgot password?' link is to its right. On the right side, there is a blue 'Sign up' section with the text 'Please login using your credentials. If you do not have credentials you can request for access or proceed as Guest to access the public dashboards.' Below this text is a 'Register Now!' button.

Operations

The following sections provides guide on various operations with Apache Superset. They include Connection to Data source, Database querying, Data Visualizations, Data exports among others.

Connect to data source.

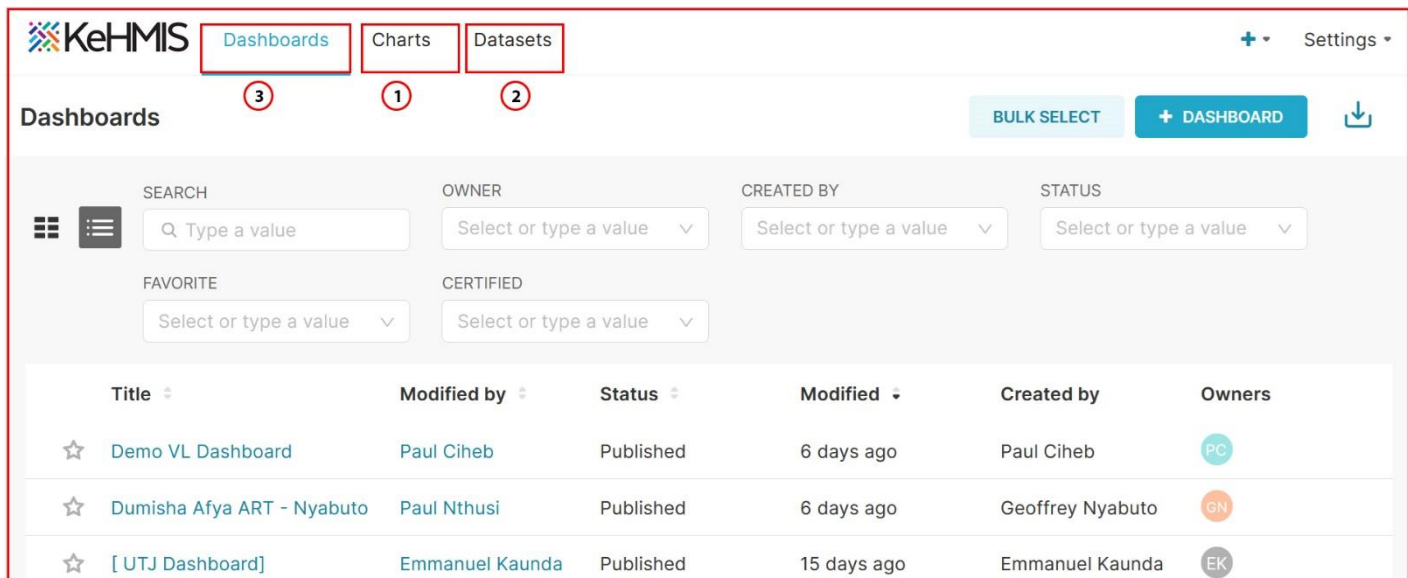
The first thing you'll want to do is connect to a data source that you want to explore:

Steps:

1. Log into Apache Superset using your provided credentials and navigate to the Home page.

You will be able to access three main pages:

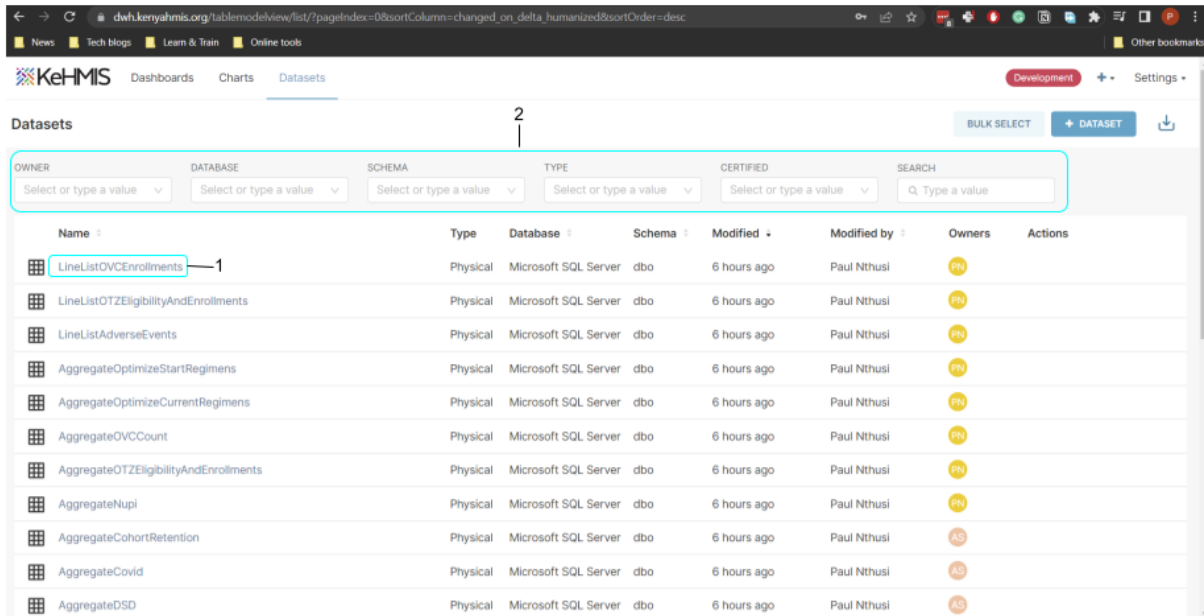
- Charts: Visualizations in the form of tables, graphs etc. that are derived from the provided datasets.
- Datasets: Raw data that has been exposed from the data warehouse
- Dashboards: Saved collection of charts



Title	Modified by	Status	Modified	Created by	Owners
☆ Demo VL Dashboard	Paul Ciheb	Published	6 days ago	Paul Ciheb	PC
☆ Dumisha Afya ART - Nyabuto	Paul Nthusi	Published	6 days ago	Geoffrey Nyabuto	GN
☆ [UTJ Dashboard]	Emmanuel Kaunda	Published	15 days ago	Emmanuel Kaunda	EK

Datasets

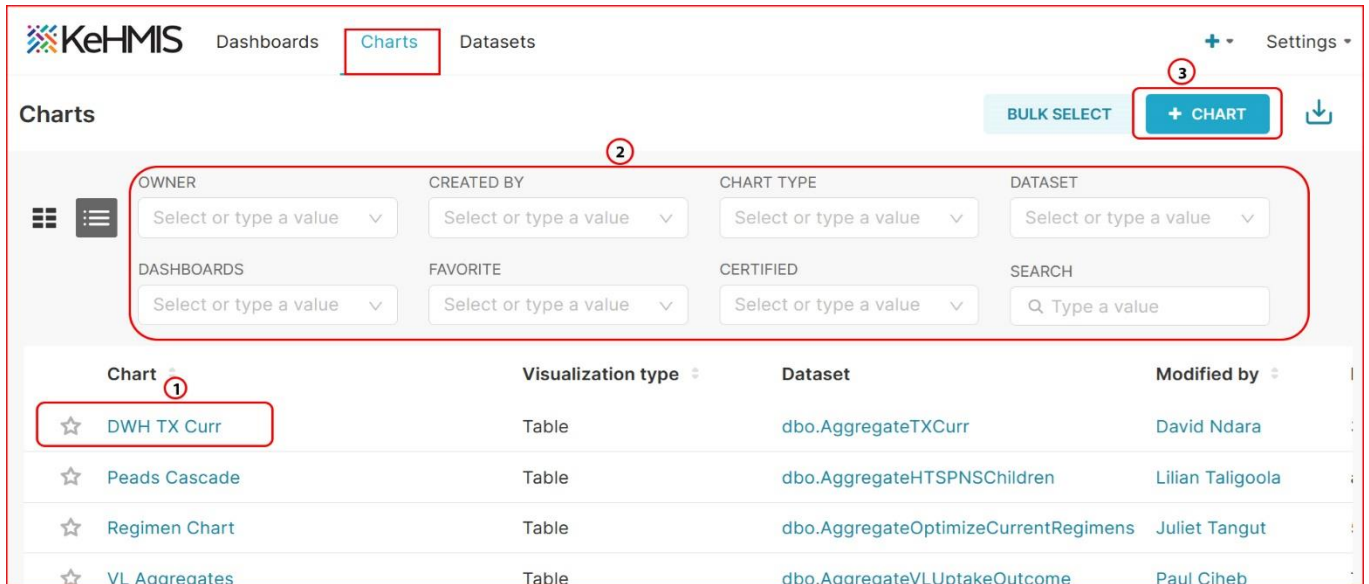
Explore Datasets



- The datasets page lists all the available data provided from the data warehouse.
- You can filter that dataset using drop downs as shown in step 2 of **Figure 3**.
- Clicking on a dataset (**Step 1**) will take you to the create chart page where you can create your visualization from the dataset.

Charts

Explore Charts

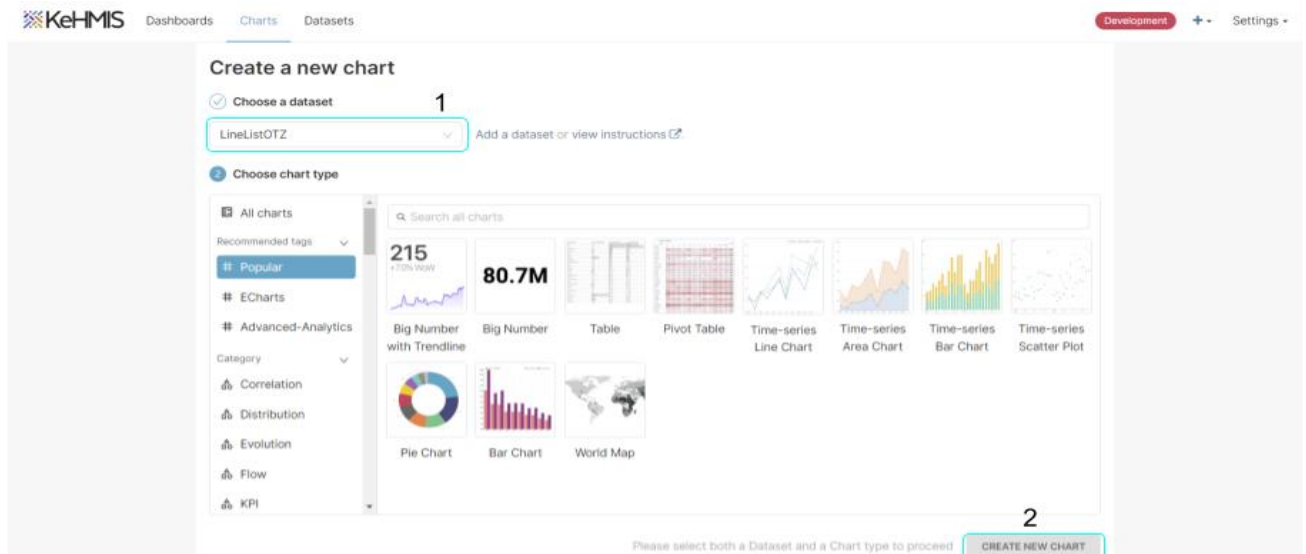


The screenshot shows the KeHMIS interface with the 'Charts' tab selected. At the top right, there is a '+ CHART' button (Step 3) and a 'BULK SELECT' button. Below this is a filter section (Step 2) with dropdown menus for OWNER, CREATED BY, CHART TYPE, DATASET, DASHBOARDS, FAVORITE, CERTIFIED, and a search bar. The main area displays a table of charts (Step 1):

Chart	Visualization type	Dataset	Modified by
DWH TX Curr	Table	dbo.AggregateTXCurr	David Ndara
Peads Cascade	Table	dbo.AggregateHTSPNSChildren	Lilian Taligoola
Regimen Chart	Table	dbo.AggregateOptimizeCurrentRegimens	Juliet Tangut
VL Aggregates	Table	dbo.AggregateVL UptakeOutcome	Paul Ciheb

Create Charts

- The Charts page above, lists all the visualizations that have been created by the user.
- You can filter the charts using the filter options (**Step 2**)
- Clicking on one of the charts (**Step 1**) takes you to the chart edit page where you can make changes to the visualization.
- You can also create a new chart by clicking on the new chart button (**Step 3**) which takes you to the create chart page (**Figure 5**)

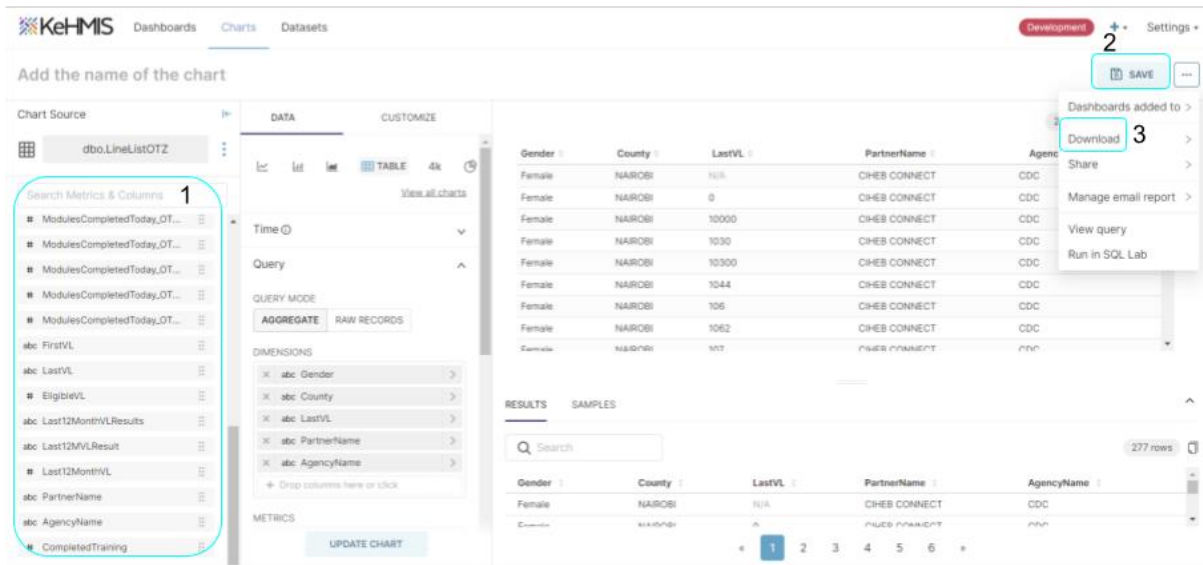


The screenshot shows the 'Create a new chart' page. Step 1 is 'Choose a dataset', where 'LineListOTZ' is selected. Step 2 is 'Choose chart type', which shows a grid of chart options including Big Number, Table, Pivot Table, Time-series Line Chart, Time-series Area Chart, Time-series Bar Chart, Time-series Scatter Plot, Pie Chart, Bar Chart, and World Map. A 'CREATE NEW CHART' button is at the bottom right.

When creating a chart, you first have to select a dataset which is the source of the data (**Step 1**)

You can then select the visualization type (bar chart, pie chart, table etc.)
 Lastly, click the Create new chart button (**Step 2**) that will take you to the chart edit page.

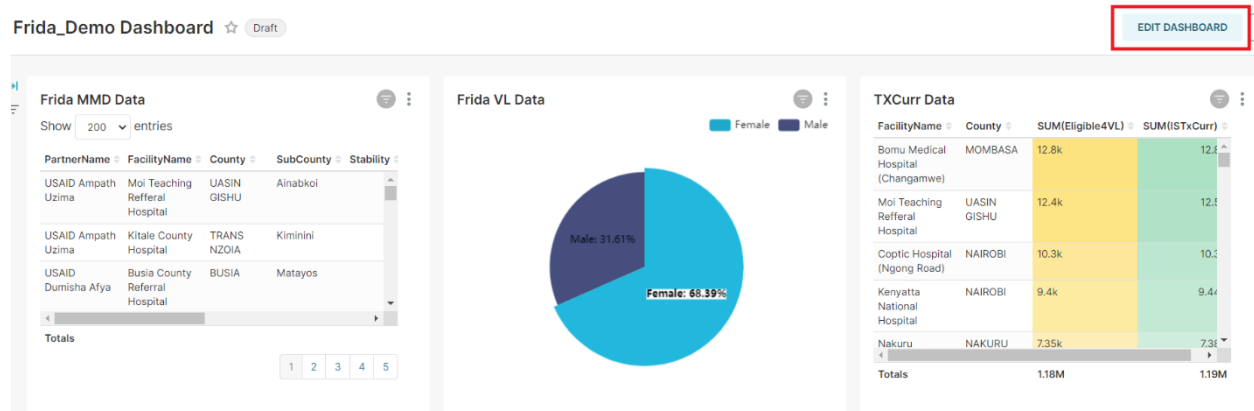
Edit, Export Chart Data



- The Chart edit page above allows you to customize your chart visualizations. You can select the columns you'd like to use as chart dimensions (**Step 1**) and customize the appearance of the chart by using the options in the "Customize" tab.
- Once you are done editing the chart, you can save your work by clicking on the Save button (**Step 2**). This will also provide you with an option to add it to a dashboard.
- You can also export your data by clicking on the Download option (**Step 3**). This will provide options for csv, json and image format.

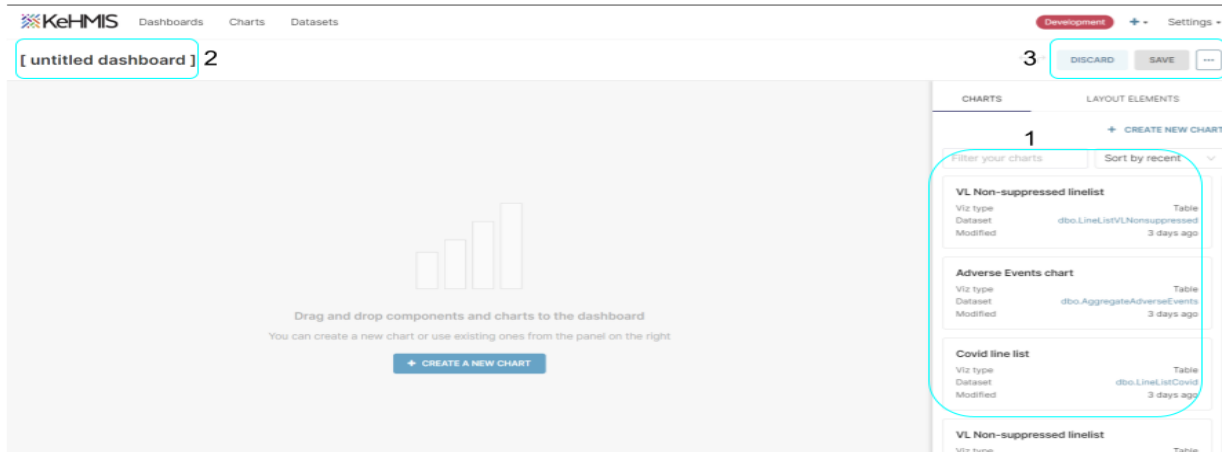
Dashboard

Explore Dashboard



The dashboards page lists all the dashboards created by the user. You can click on the dashboard name (**Step 1**) to view and edit the dashboard or click on the new dashboard button (**Step 2**) to create one.

Create Dashboard



- The Create dashboard page above, allows you to drag created charts (**Step 1**) into the working space so as to create a dashboard.
- You can label the dashboard by setting the name in title section (**Step 2**).
- Finally, you can either save or discard changes made to the dashboard by clicking on the save, discard buttons (**Step 3**).

For more information on creating dashboards, refer to [Creating Your first dashboard](#) guide

END