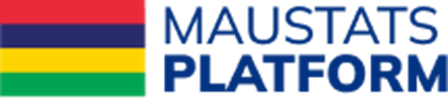
**User Guide**

**DATA MANAGEMENT**



USER GUIDE – Data MANAGEMENT

*Modern Statistics Platform*

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Date | Reasons of Changes | Version |
| iTM | 08th December 2023 | The first draft of Data Management MauStats. | V1 |
| iTM | 17th June 2024 | The second draft of Data Management MauStats. | V2 |

Table of Contents

[1.0 INTRODUCTION 4](#_Toc169606984)

[2.0 DATA MANAGEMENT 5](#_Toc169606985)

[2.1 TOPIC AND SUBTOPIC 6](#_Toc169606986)

[2.1.1 EXPORT 7](#_Toc169606987)

[2.1.2 IMPORT 9](#_Toc169606988)

[2.1.3 ADD 10](#_Toc169606989)

[2.2 PRIMARY DATA 12](#_Toc169606990)

[2.2.1 INGEST 13](#_Toc169606991)

[2.2.2 FILTER 17](#_Toc169606992)

[2.2.3 CLEAN 21](#_Toc169606993)

[2.2.4 CALCULATE 26](#_Toc169606994)

[2.2.5 COMPLETE 30](#_Toc169606995)

[2.3 DATA MAPPING 31](#_Toc169606996)

[2.3.1 IMPORT RULES 31](#_Toc169606997)

[2.3.2 IMPORT MAPPING 33](#_Toc169606998)

# 1.0 INTRODUCTION

The Modern Statistics (MauStats) Platform is developed as a web-based platform with central database to ensure consistency, security and accessibility. The platform is used to will help to automate and integrate data lifecycle from acquisition to dissemination using Generic Statistical Business Process Model (GSBPM) framework. The platform complies to manage data requirements for Census, SEE, and SBR processed data in this phase. It is also feasible for data integration based on SM requirements. This web-based platform consists of the two applications – User Interface and Data Management System (DMS).

In this user guide, we will be learning on how to use the Data Management module of the DMS application.

# 2.0 DATA MANAGEMENT

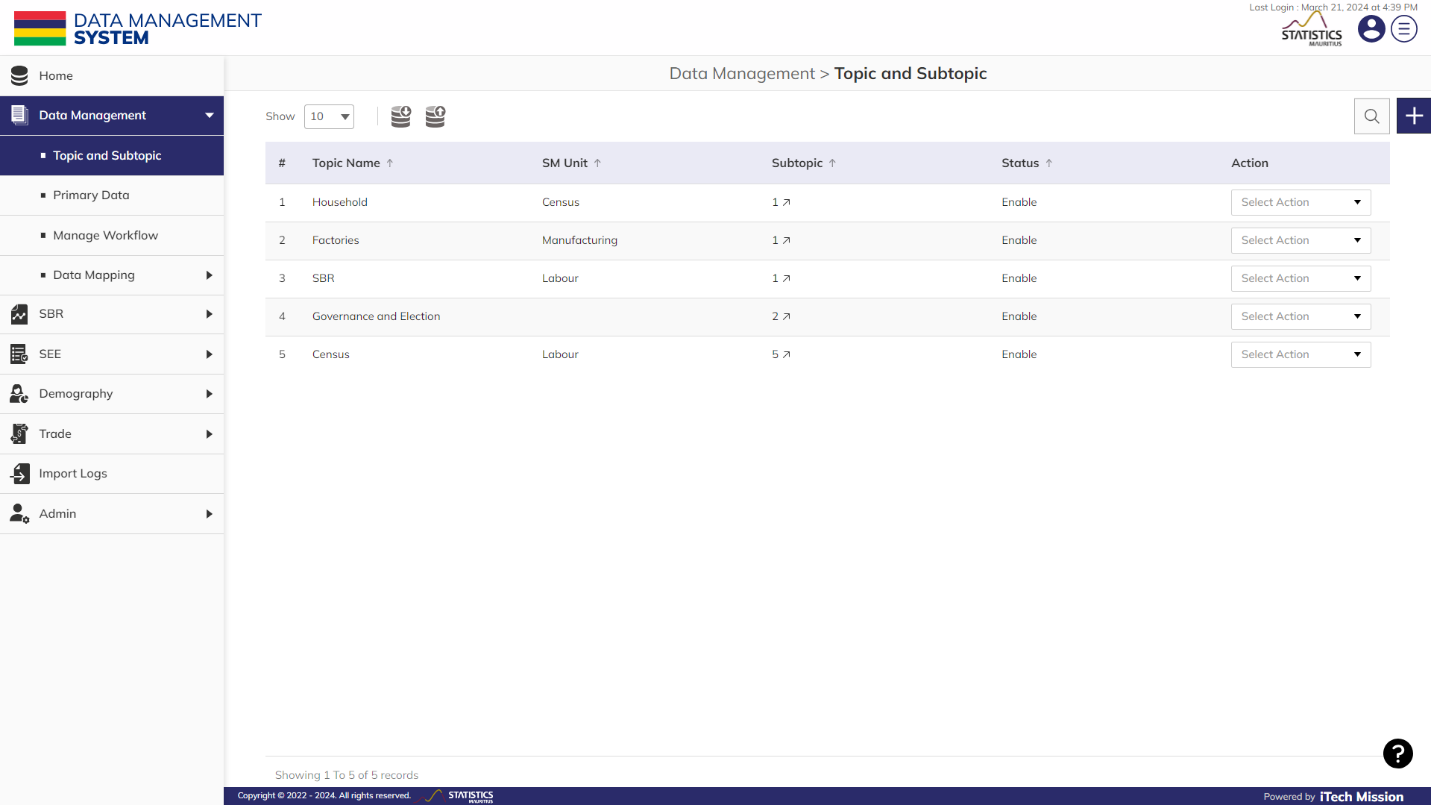
The Data Management module comprises of three submodules which allows management and administration of topics and subtopics, primary data, import rules and import mapping. Below are the sub-modules of the data management–

* Topic and Subtopic
* Primary Data
* Data Mapping

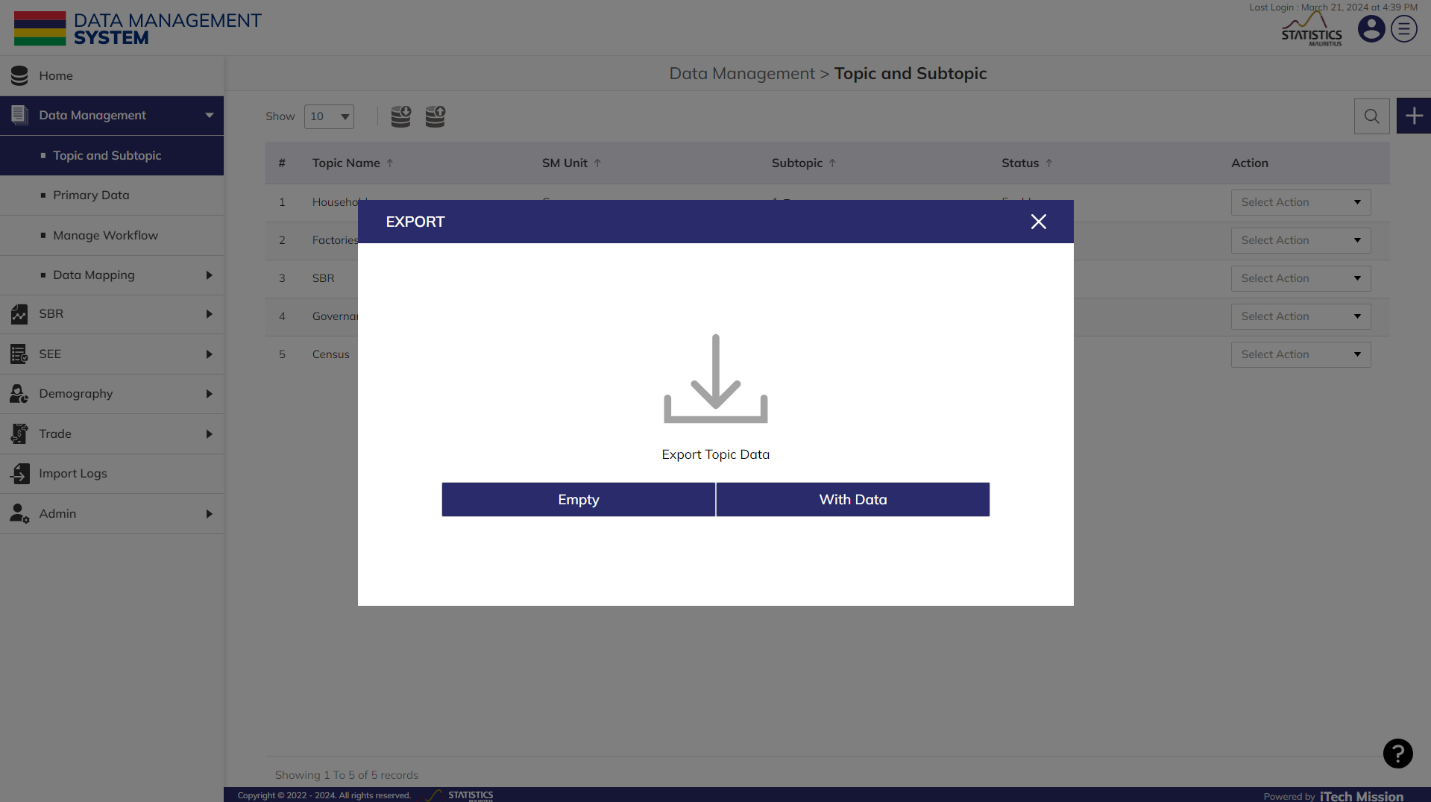
Let us learn how to use these sub-modules in detail.

## 2.1 TOPIC AND SUBTOPIC

This submodule allows the authorized users to create and manage the topics and its subtopics master list.

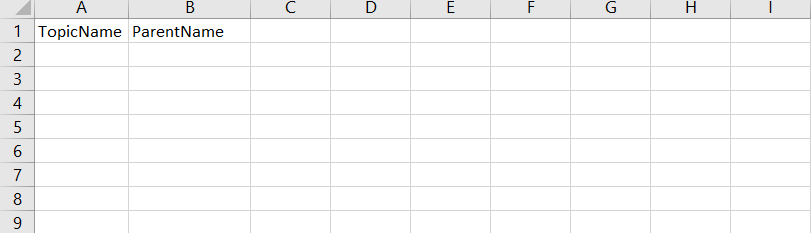
**Step 1:** Click on the **Topic and Subtopic** option available under Data Management to access this sub-module. You have the options to add, import and export, edit, delete and enable/disable the elements (see below figure).

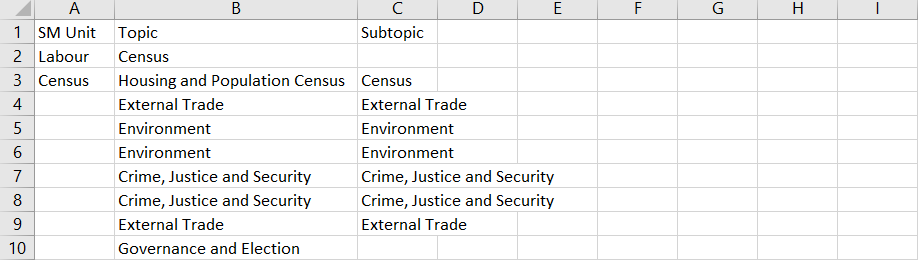
### 2.1.1 EXPORT

**Step 2:** Click on the **Export** button to download the empty topic data and the topic data master list available in the database (see below figure).

You can use this option to create a topic data template structure. Below is the structure that you will find in the enterprise in the topic data template.

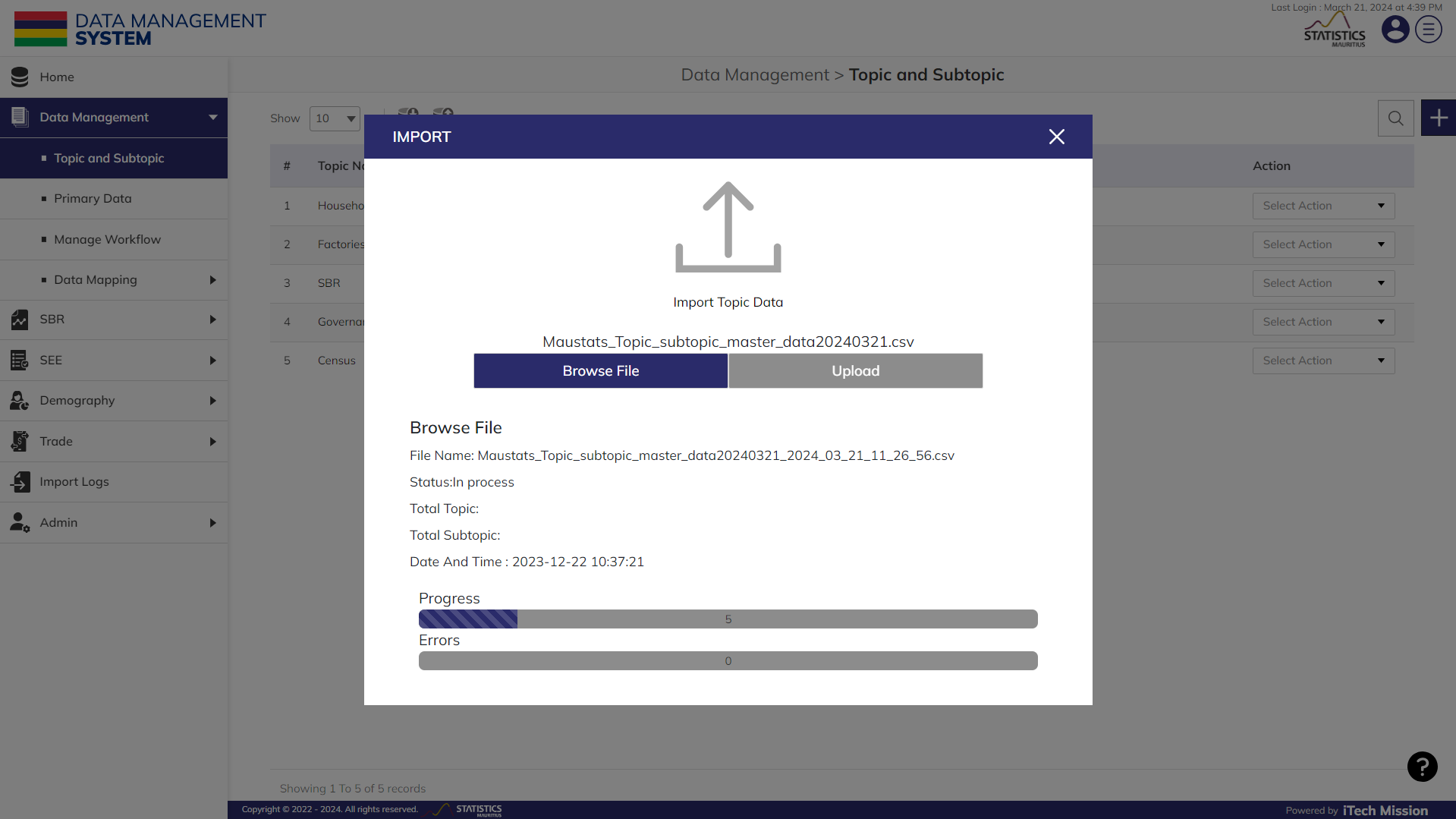
|  |  |
| --- | --- |
| **TopicName** | A unique subtopic identification that associates the subtopic to its parent topic. |
| **ParentName** | Name of the topic under which various subtopics are included. |

**Step 3:** Click on the **Empty** option to download an empty topic data template. The topic data list is imported and exported in the CSV (Comma Separated Value) file format.

**Step 4:** Click on the **With Data** option to download the topic data master list.

### 2.1.2 IMPORT

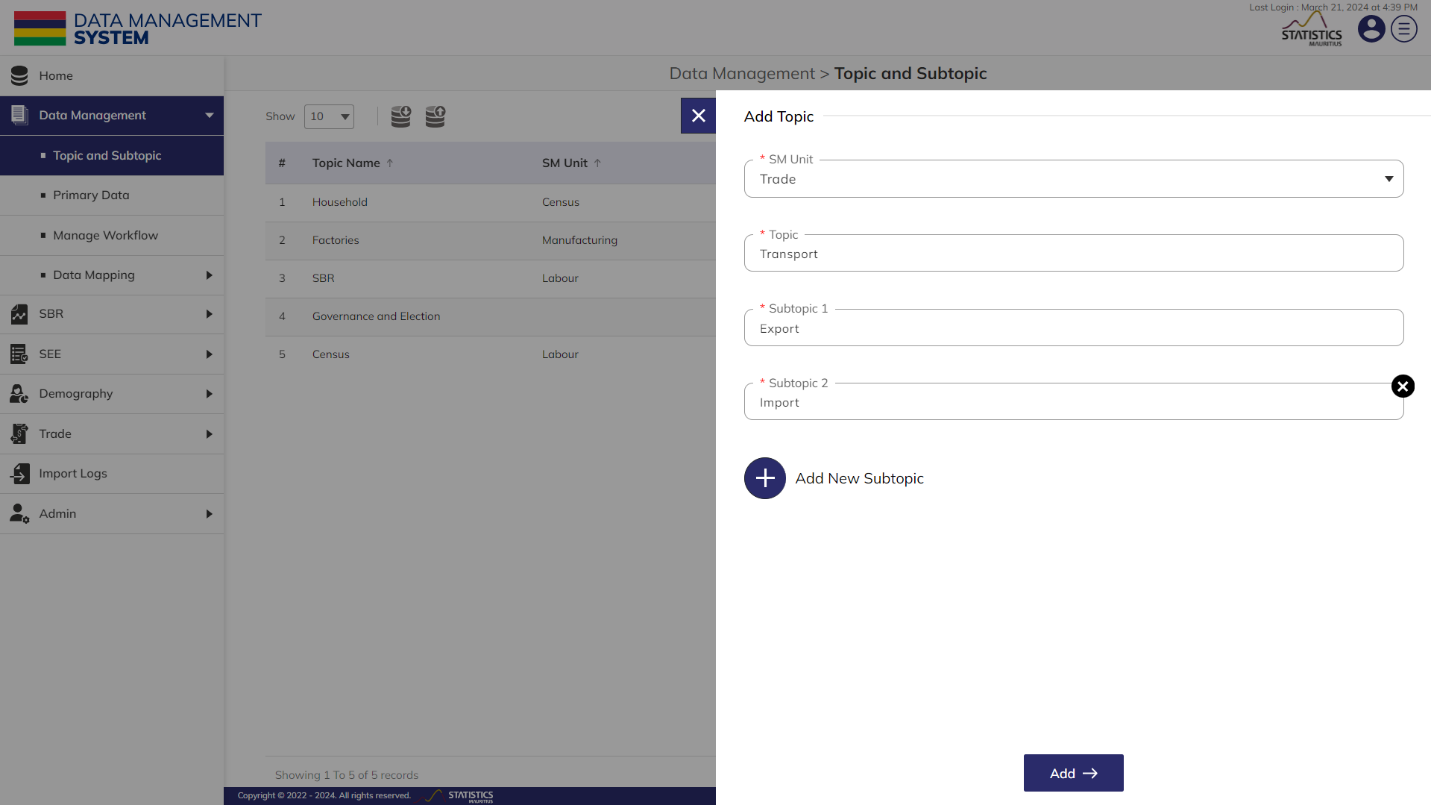
**Step 5** Click on the **Import** button to browse the file and upload the topic data master list into the database.

To upload the topic data master list into the database, first enter the topic data structure in the template as explained above and then click on the Browse File button to select the template. Now, click on the Upload button to start the import process (see below figure).

During the import process, the records which contains blank entries, entries with special characters and duplicate entries are not imported in the database. A Last Upload summary will appear after the import process which provides the count of successfully imported records out of the total records. An error log will be generated to report on unsuccessful imported records which you can access from Log module.

### 2.1.3 ADD

**Step 6:** Click on the **Plus** button available to the right side of the page to add a new topic (see figure below). Enter the following details to add a new element.

* Select **SM Unit**
* Enter **Topic**
* Enter **Subtopic** (By clicking on Add New Subtopic, more than one subtopic can also be added to the same topic)

All details that are mandatory to be entered.Click on the **Add** button to save and confirm. You can now view the new topic added in the topic and subtopic list.

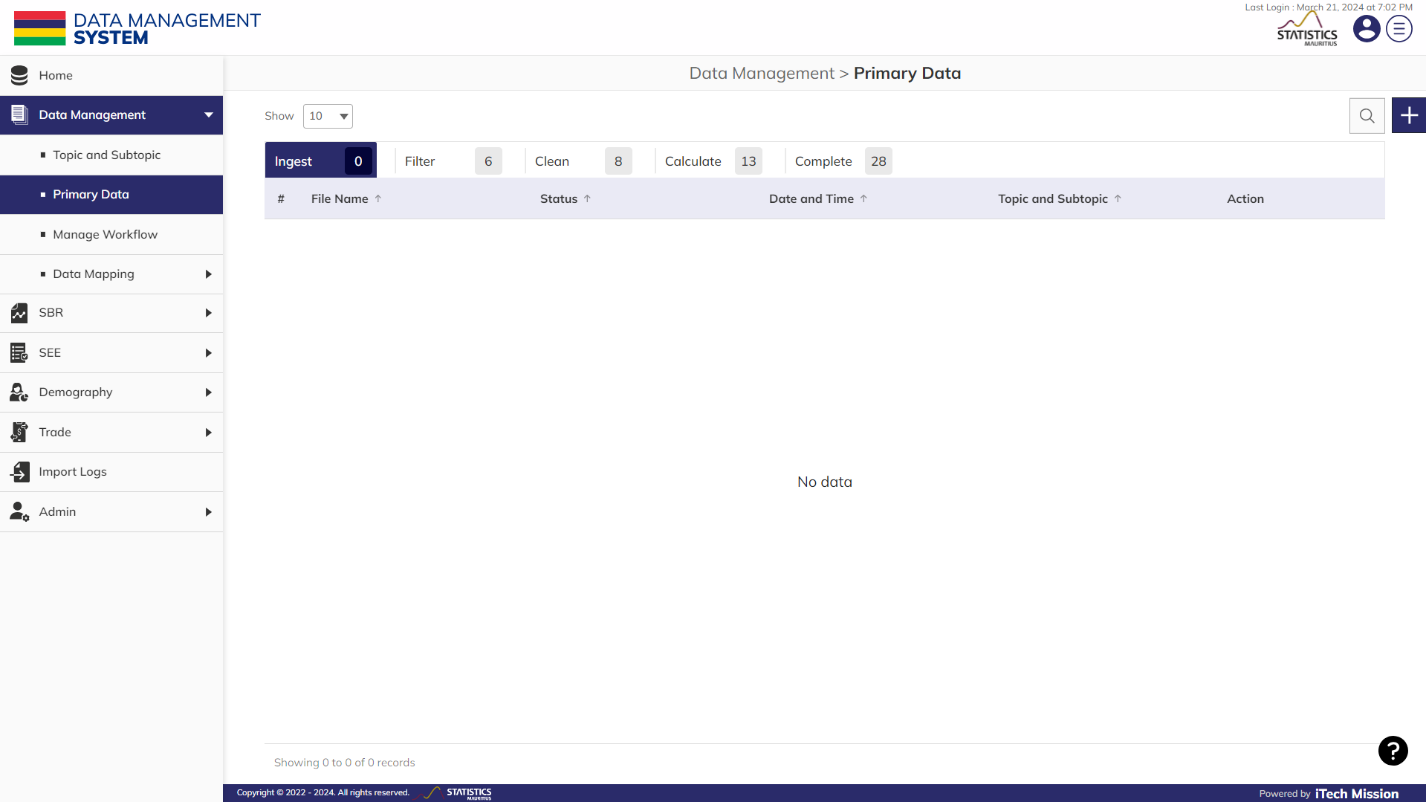
**Step 7:** Click to select **Edit** option available in the action dropdown to edit the details of the selected element. Edit the required detail and click on update button to save and confirm.

**Step 8:** Click to select **Enable/Disable** option available in the action dropdown to show or hide the specific element. The hidden element will not be shown throughout the application.

**Step 9:** Click to select **Delete** option available in the action dropdown to delete the selected element. A confirmation popup will appear, click yes to confirm deletion of selected element.

## 2.2 PRIMARY DATA

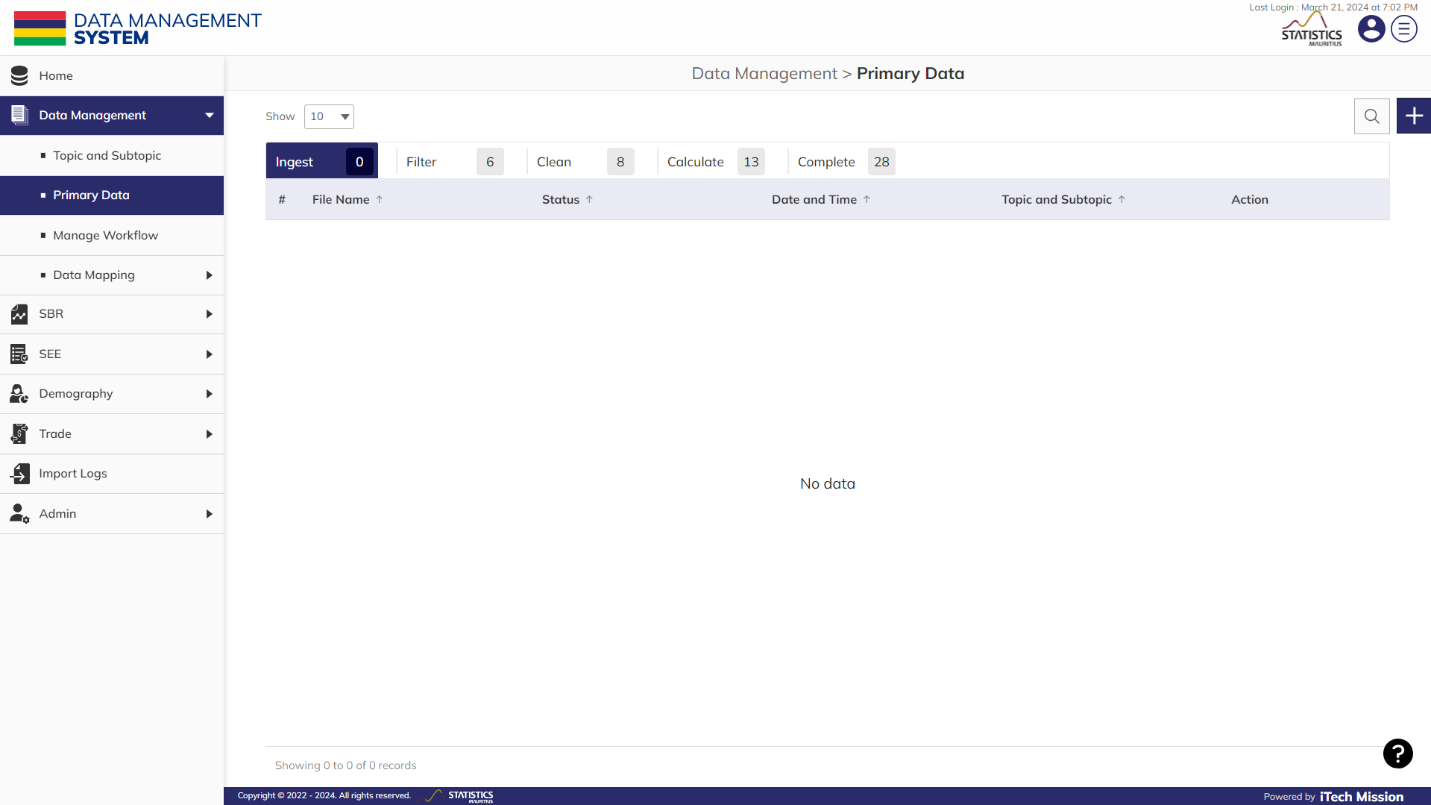
**Step 1:** Click on the **Primary Data** under Data Management option from the left menu to access this module. This module has five-tab views – **Ingest**, **Filter**, **Clean**, **Calculate** and **Complete** (see below figure). Each tab view has various set of actions that one can perform on the ingested data.

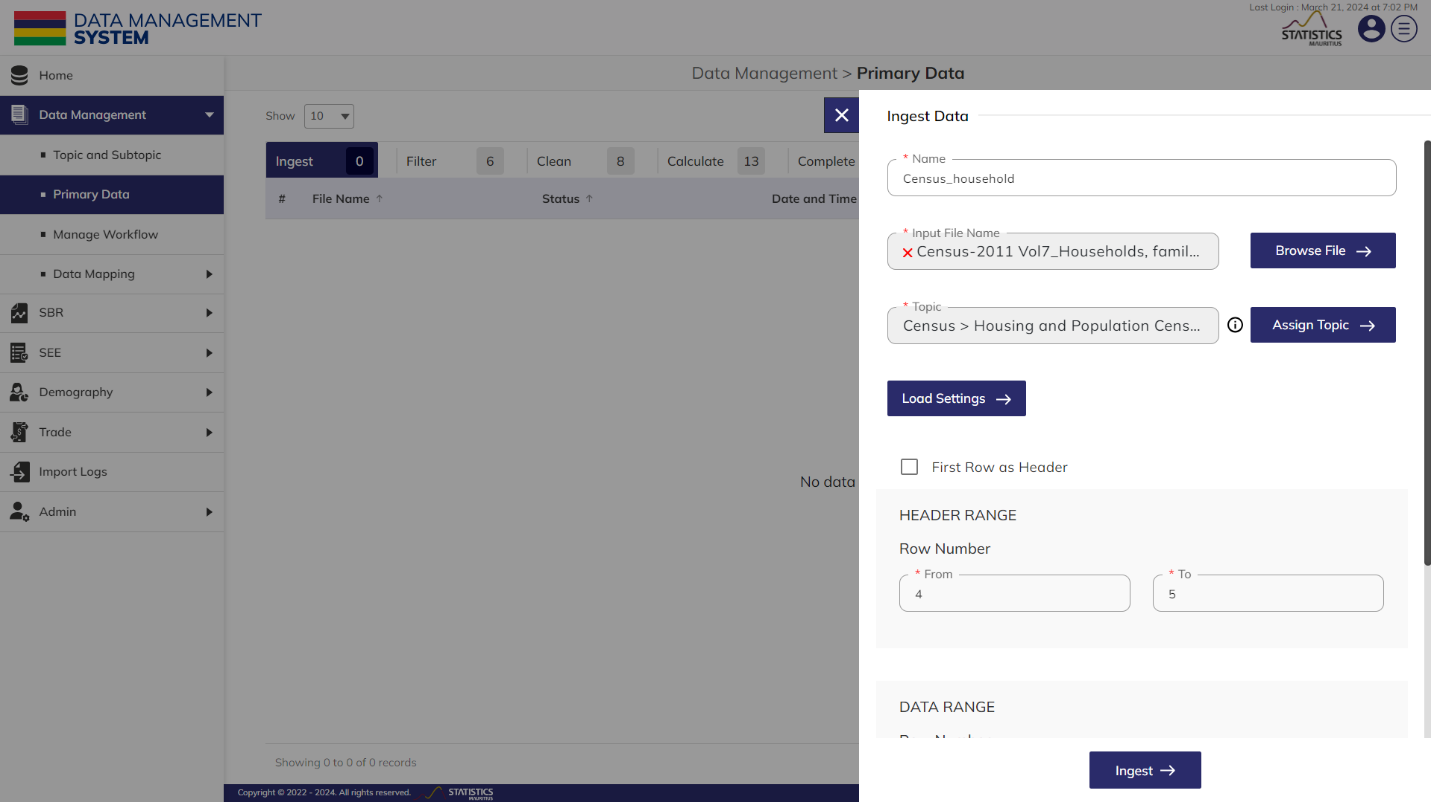
Let us learn about each tab view in detail.

### 2.2.1 INGEST

The **Ingest** tab allows you to read and process the primary unstructured data files in distinct formats which are copied/uploaded in the designated cloud storage location by the respective SM Unit. This tab facilitates the ingestion of primary or raw data into the platform. Let us take an example to ingest a census data file.

Below is the step-by-step guide on how to ingest this data.

**Step 2**: Click on the Plus button available to the right side of the page to ingest a new file (see figure below). Enter the following details to add a new file: -

* Enter **Name**
* Click on the “**Browse file**” option (see below figure).

Search the required file using the search bar and click on the **checkbox** to select it. Click on the **Ok** button to confirm.

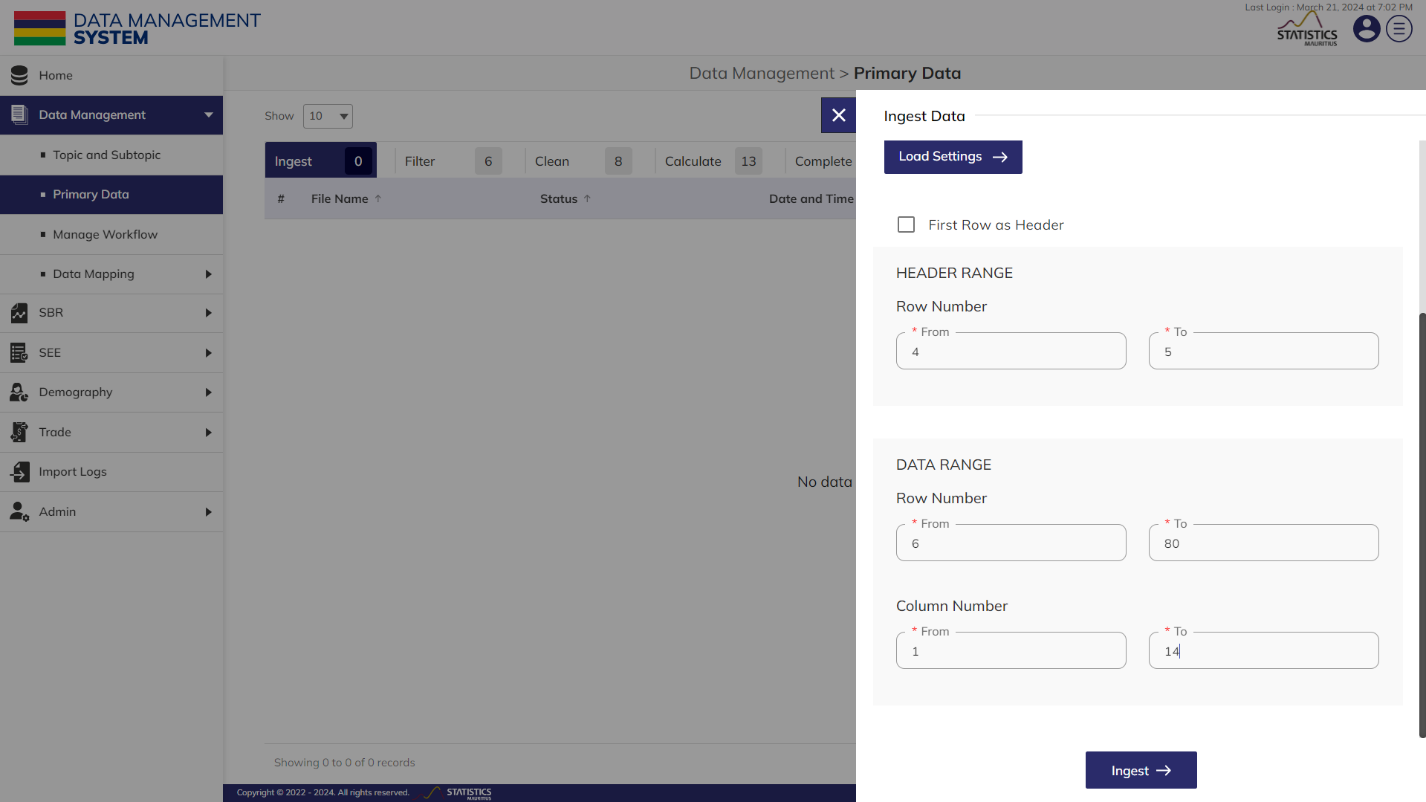
**NOTE:**

If any data file with similar structure already ingested, you have the option to copy its settings in the current file. To proceed with this, refer to step 5; otherwise, you can skip this step if you are importing a new data file.

* Click on the "**Assign Topic** " option to assign the data file to the subtopic and topic.
* Click on the "**Load Settings**" option. Select the required file and click on the **Ok** button to apply settings.
* Enter the **Header Range** - from row number “**4**” to row number “**5**”.

**NOTE:**

If your data file has header in single row, it should be marked by checking the “**First row as header**” option. Since our data file has header in multiple rows, we have to enter the header range as per the data file.

* Enter the **Data Range** - from row number “**6**” to row number “**40**” and from column number “**1**” to column number “**14**” (see below figure).

**NOTE:**

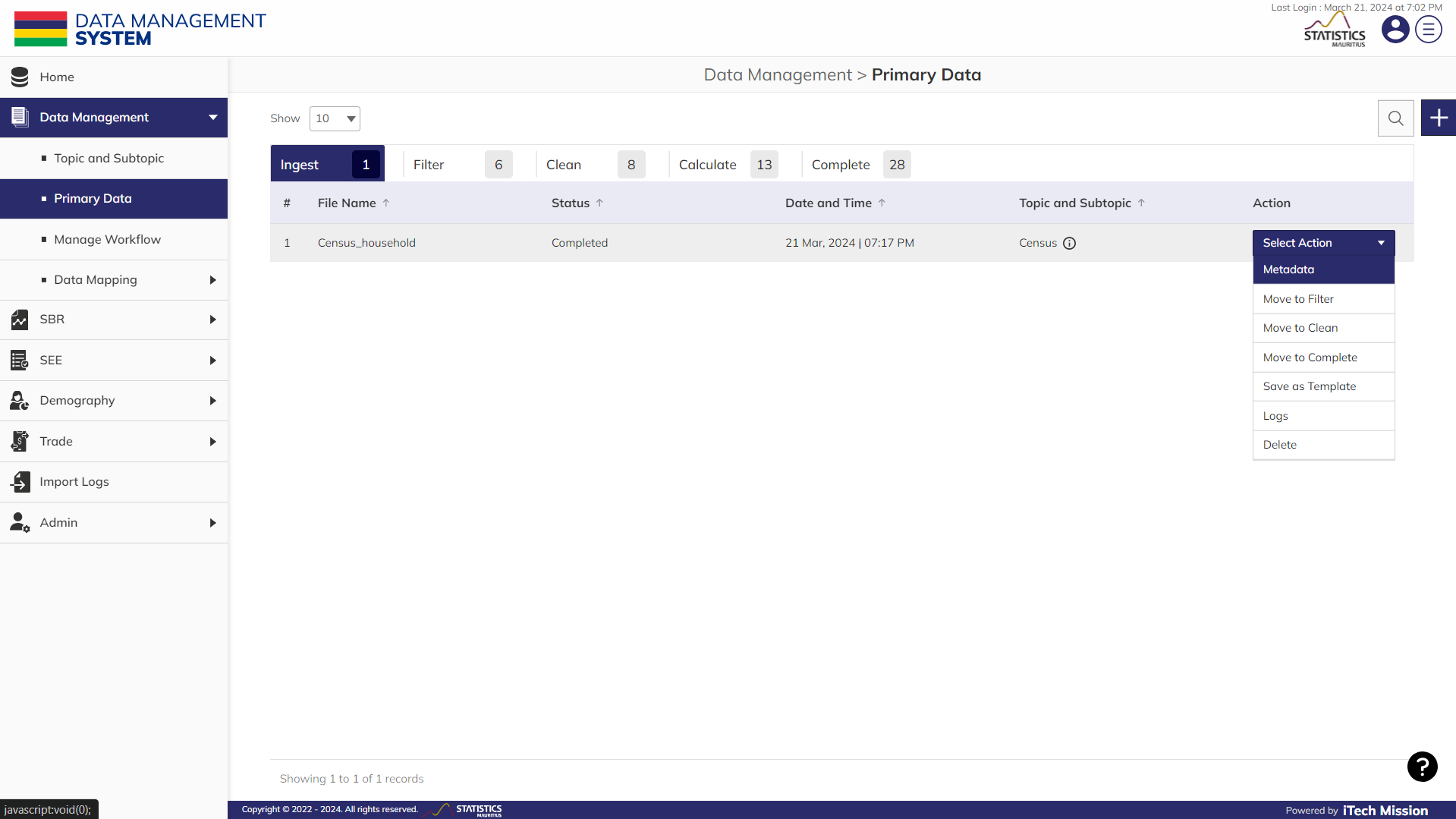
The data range is defined as the range of rows and columns that contains the data in our data file. For instance, in the given data file, the range of the rows in which the data exist is from “6” to “80” and range of column is from “A” to “N” that means “1” to “14”.

**Step 3**: Click on the “**Ingest”** button to start the ingestion process.

**NOTE:**

As the ingestion process initiates, the status will show "In-progress" during the ongoing ingestion process and users will have the following options available in the dropdown under the action column – “Delete”, “View Metadata and Logs”. Additional options will appear once the ingestion status is “Completed”.

There will be following actions available in the dropdown on the “Ingest” tab –

* Delete
* Metadata
* Move to filter
* Move to clean
* Move to complete
* Save as Template
* Logs

**Step 4**: Click on the “**Move to filter”** option from the dropdown available under action column to move the data to the “**Filter**” tab view

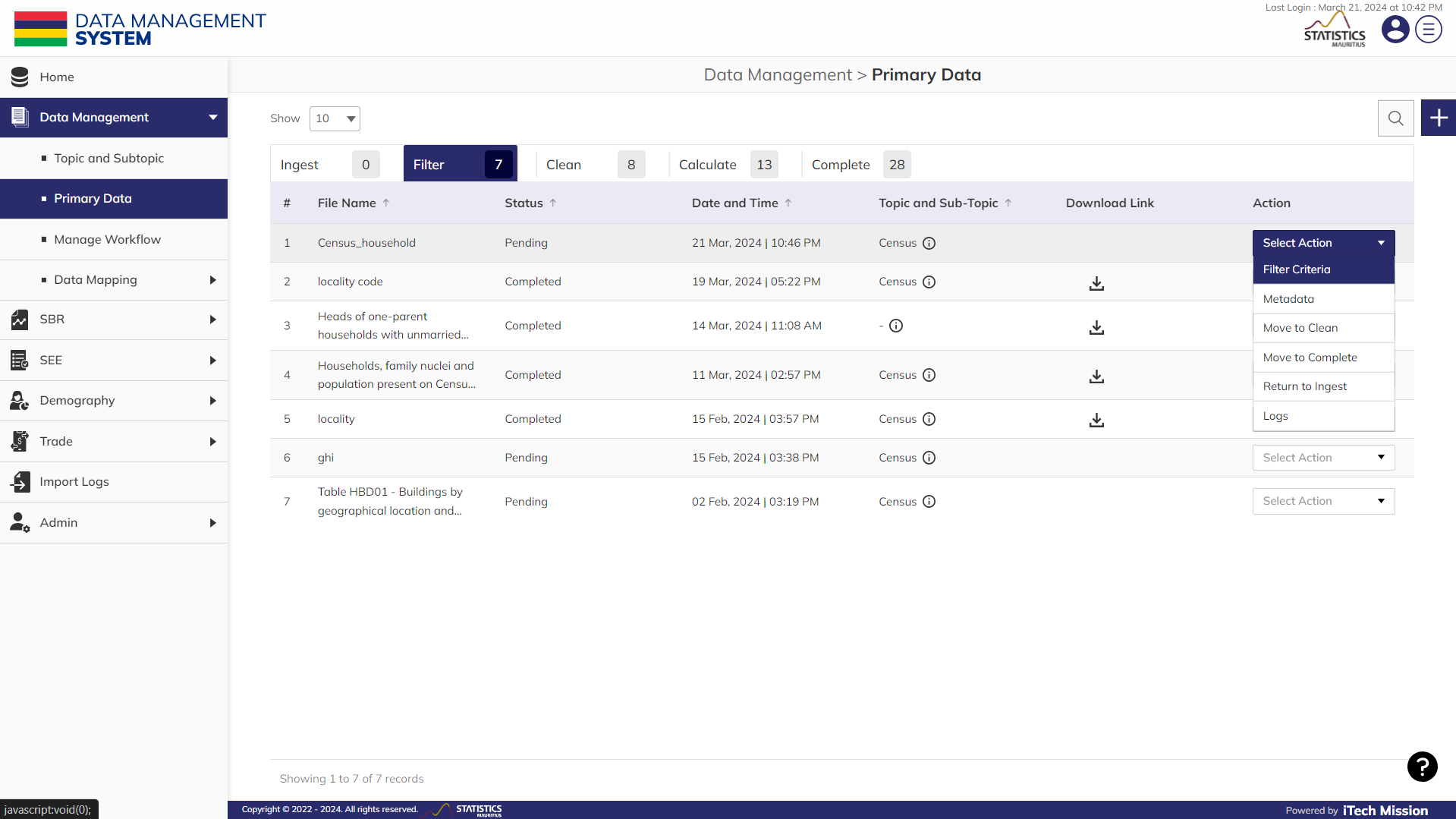
### 2.2.2 FILTER

The **Filter** tab view provides you the option to identify outliers, invalid or incorrect data before proceeding to the cleaning step. This tab has the option to filter data based on specified criteria and download the data locally in .xlsx format.

**NOTE:**

The status of the data will be "Pending" by default when it is moved to the filter tab. The status will change to “Completed” once the filtration is done and the download link is generated for the same.

There are following actions available in the dropdown on the “Filter” tab –

* Filter Criteria
* Metadata
* Move to clean
* Move to complete
* Return to ingest
* Logs

Below is the step-by-step guide on how to filter and download the ingested data.

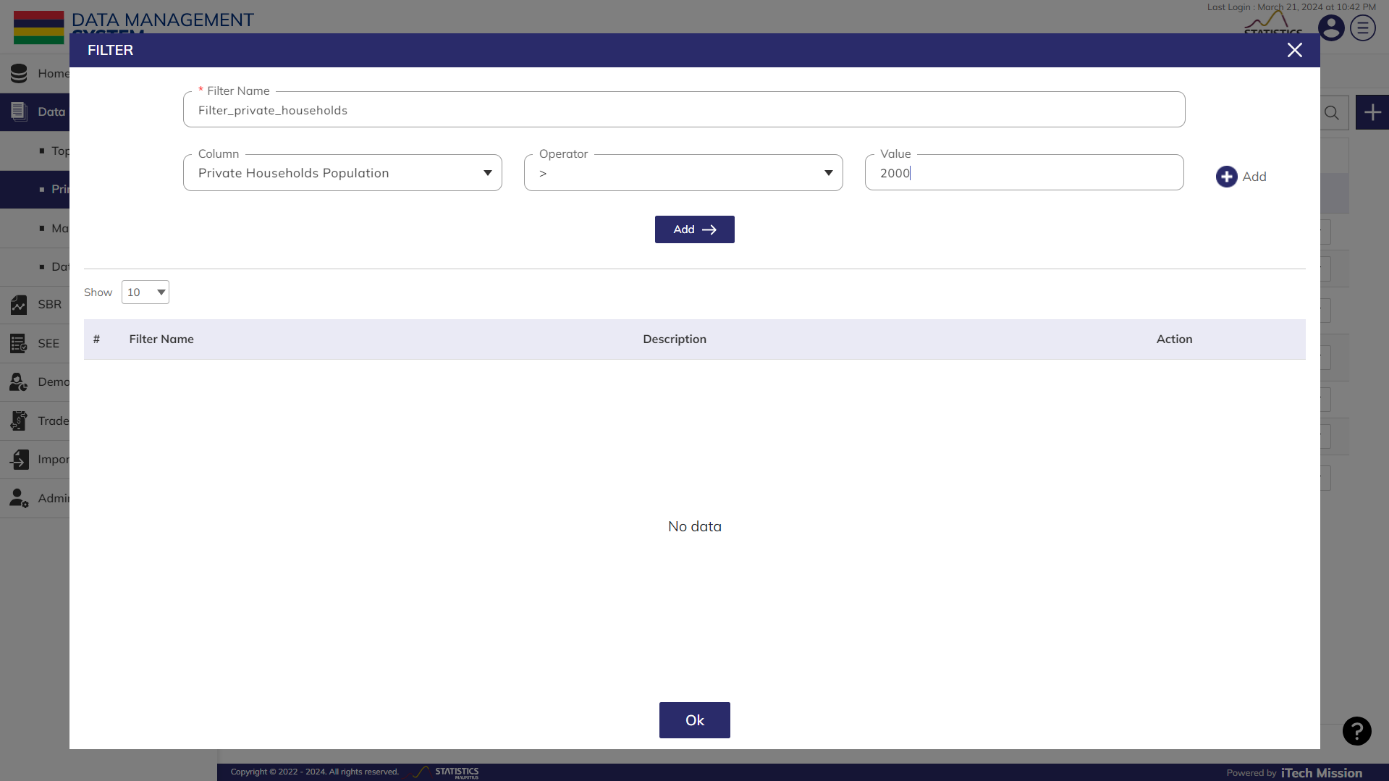
**Step 5**: Click on the “**Filter criteria**” option from the dropdown available under action column.

*Let us suppose you want to filter the data where the “****Private Households Population****” is “****greater than****” <****2000****> as per the ingested census file.*

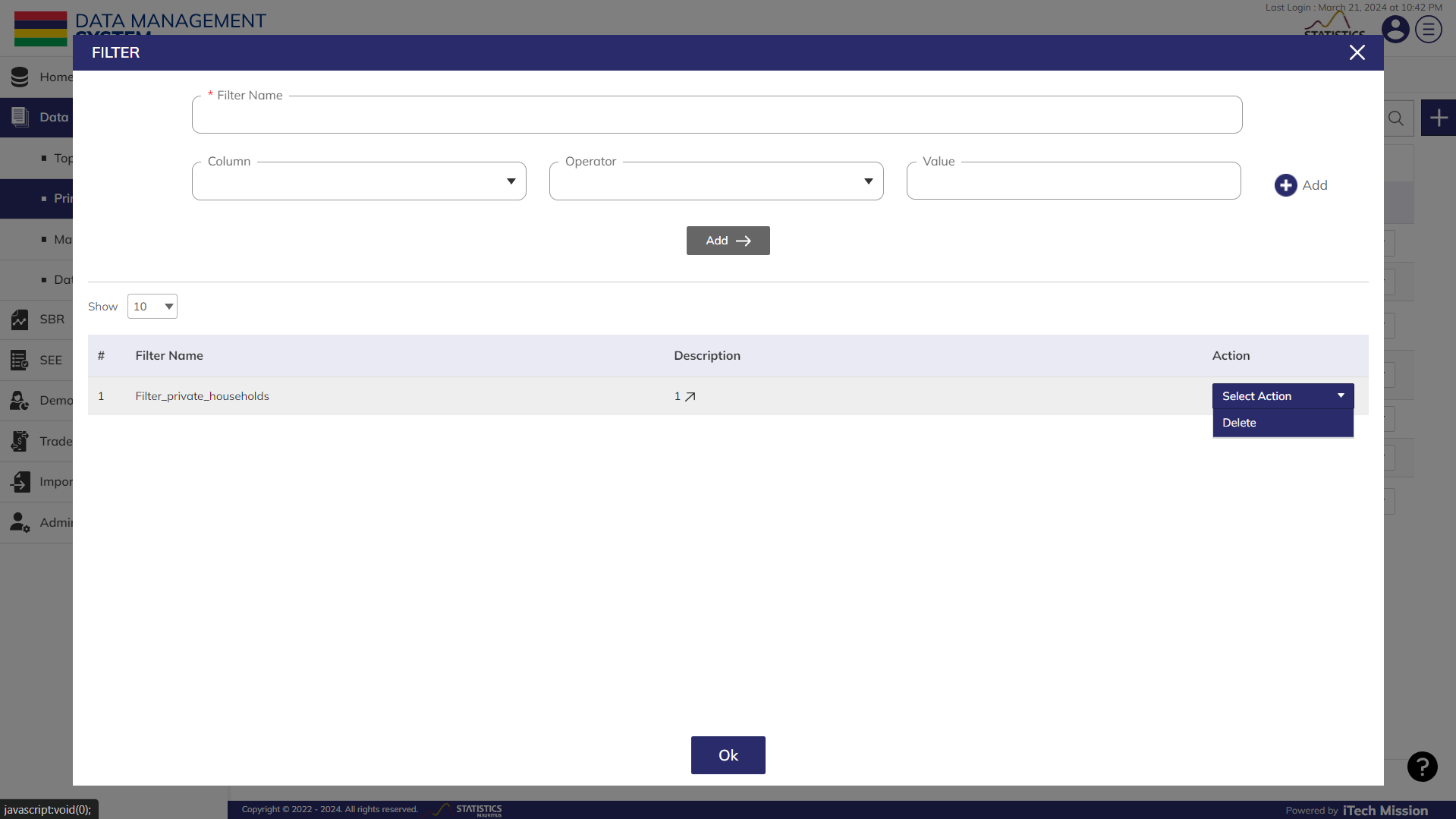
The above statement comprises of:

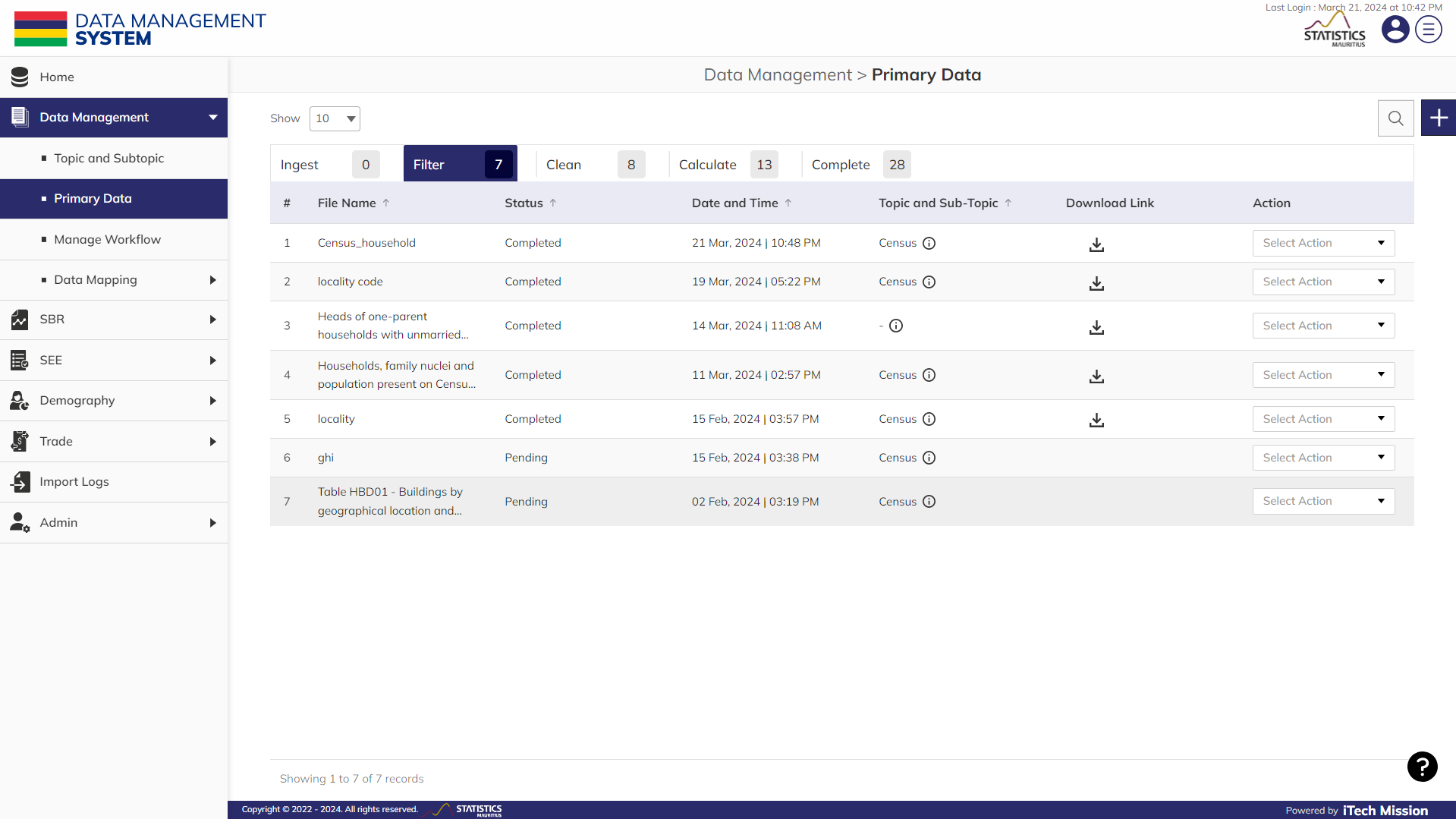
1. “*Private Households Population* “as **Column**
2. “Greater than” as **Operator**
3. <*2000>* as **Value**

**Step 6**: Enter the following details to add filter criteria: -

* Enter the **Filter Name.**
* Select the “***Private Households Population****” option* from the **Column** dropdown.
* Select the “*>*” option from the **Operator** dropdown.
* Enter the **Value** as “2000”.

**Step 7**: Click on the **Add** button to add the filter criteria in the tabular grid below

****Step 8**: Click on the **Ok** button to request for the download link of the filtered data (see below figure).

**Step 9**: Click on the **Download** icon under download link column to download the filtered data on your local system (see below figure).

**Step 10**: Click on the “**Move to clean**” option from the dropdown available under action column to move the data to the “**Clean**” tab view.

**NOTE:**

The download button will appear under the “Download Link” column when the status of file changes to “Completed” and file will be generated based on the specified filter criteria.

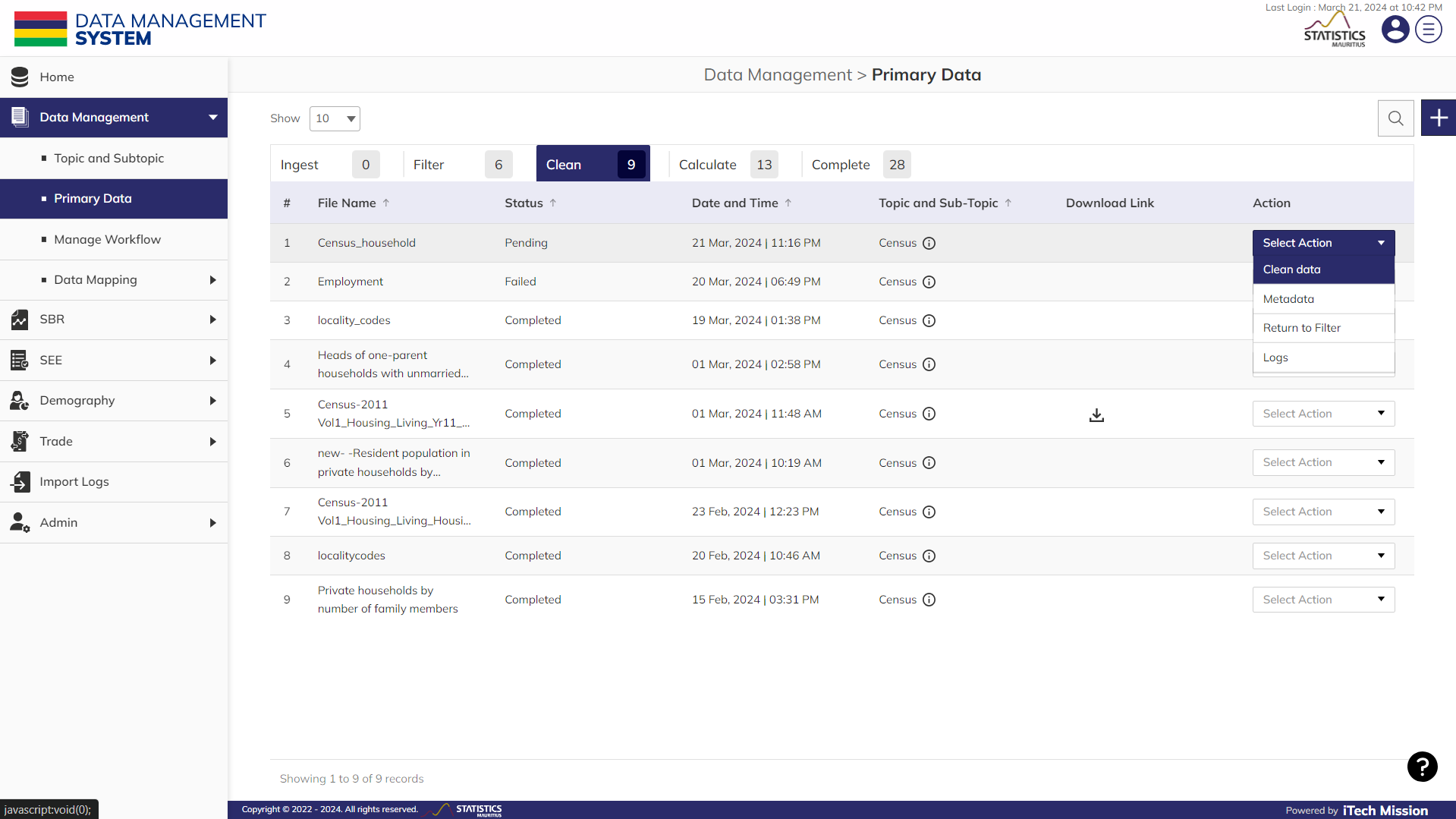
### 2.2.3 CLEAN

The **Clean** tab allows you to clean the ingested data by applying the appropriate business rules, including identifying duplicates, blank entries, and entries with special characters. You have options to delete selected columns or rows, rename columns, replace any data value, and remove duplicates.

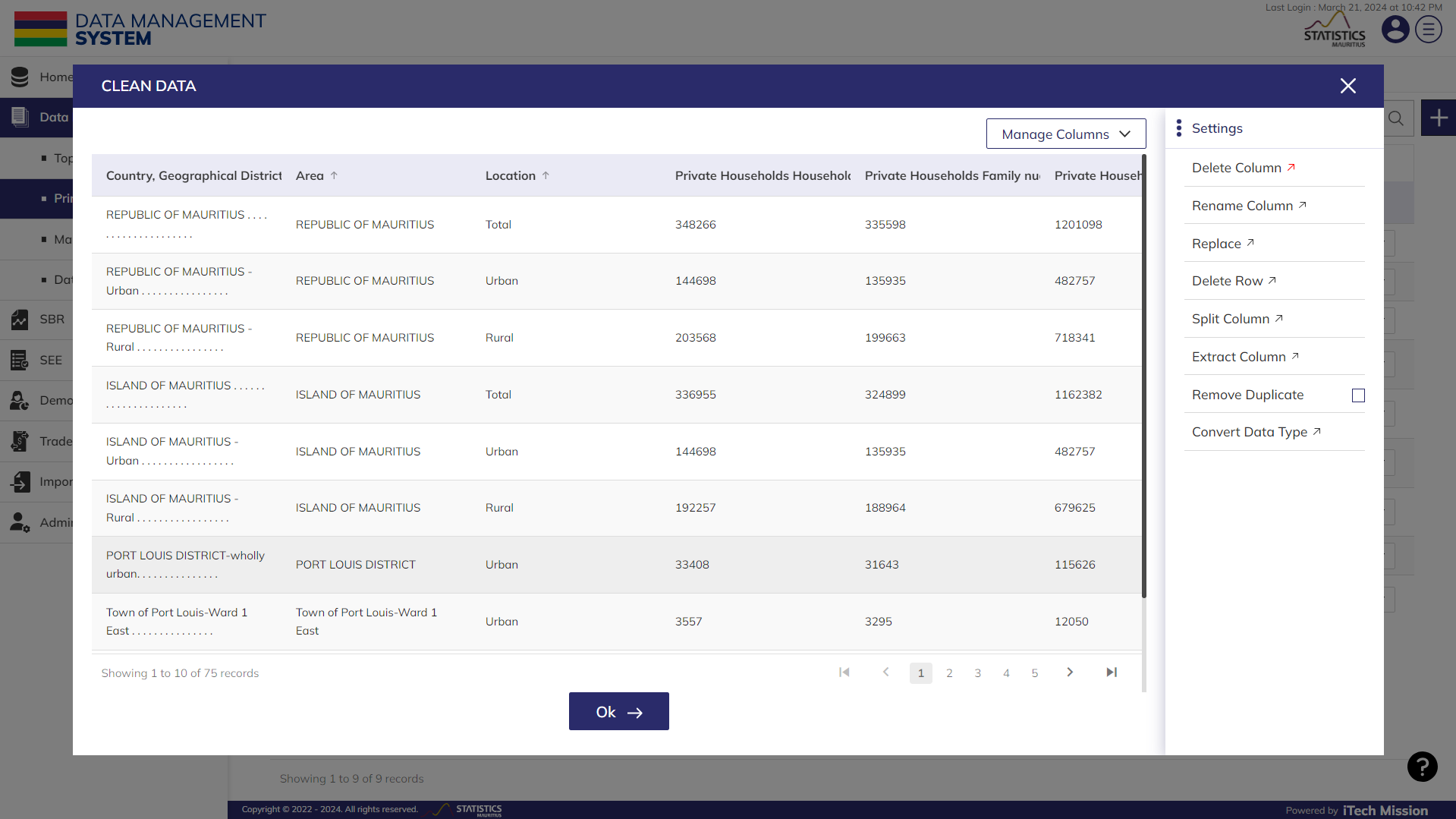
**NOTE:**

The status of the data will be "Pending" by default when it is moved to the clean tab. The status will change to “In-progress” once the cleaning process is done.

Below is the step-by-step guide on how to clean the ingested data.

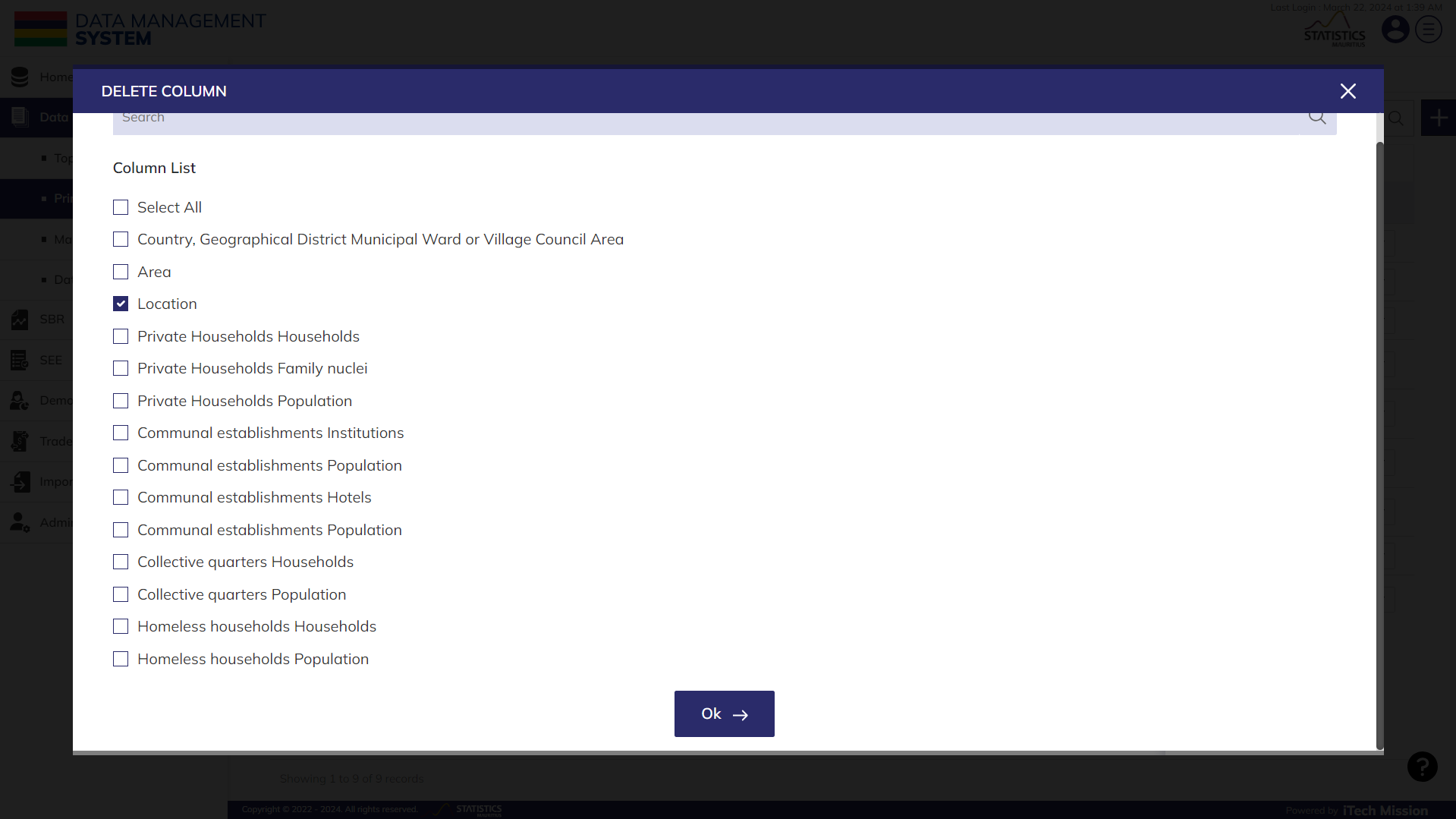
**Step 11**: Click on the “**Clean data**” option from the dropdown available under action column (see below figure).

Following are the types of cleaning actions available that you can perform on the data –

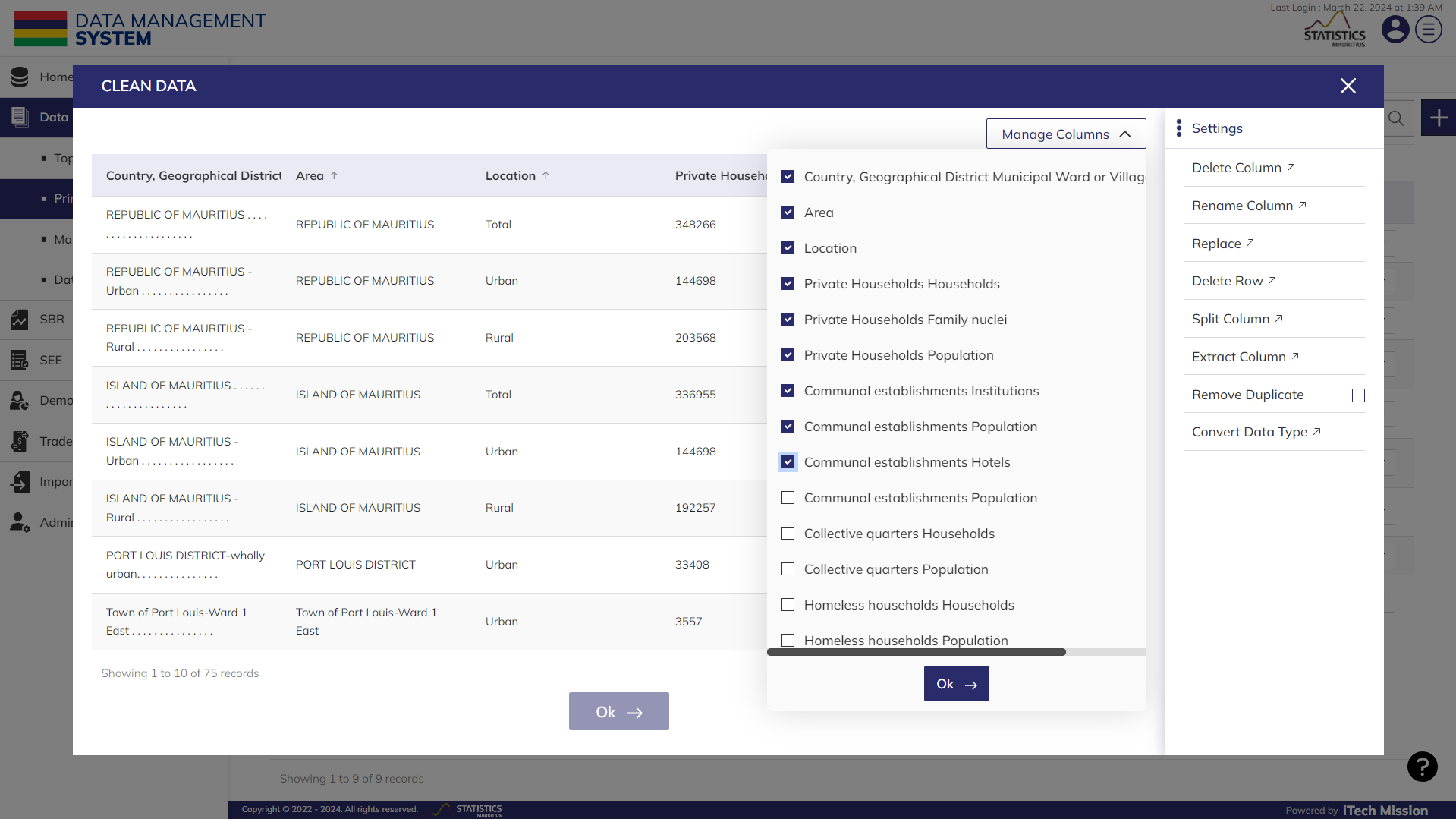
* Delete column
* Rename column
* Replace
* Delete Row
* Split Column
* Extract Column
* Remove duplicate
* Convert Data Type

*Let us suppose you want to delete a column as per the ingested file.*

**Step 12**: Click on the “**arrow**” button available with the “Delete column” option to the right side of the pop-up window.

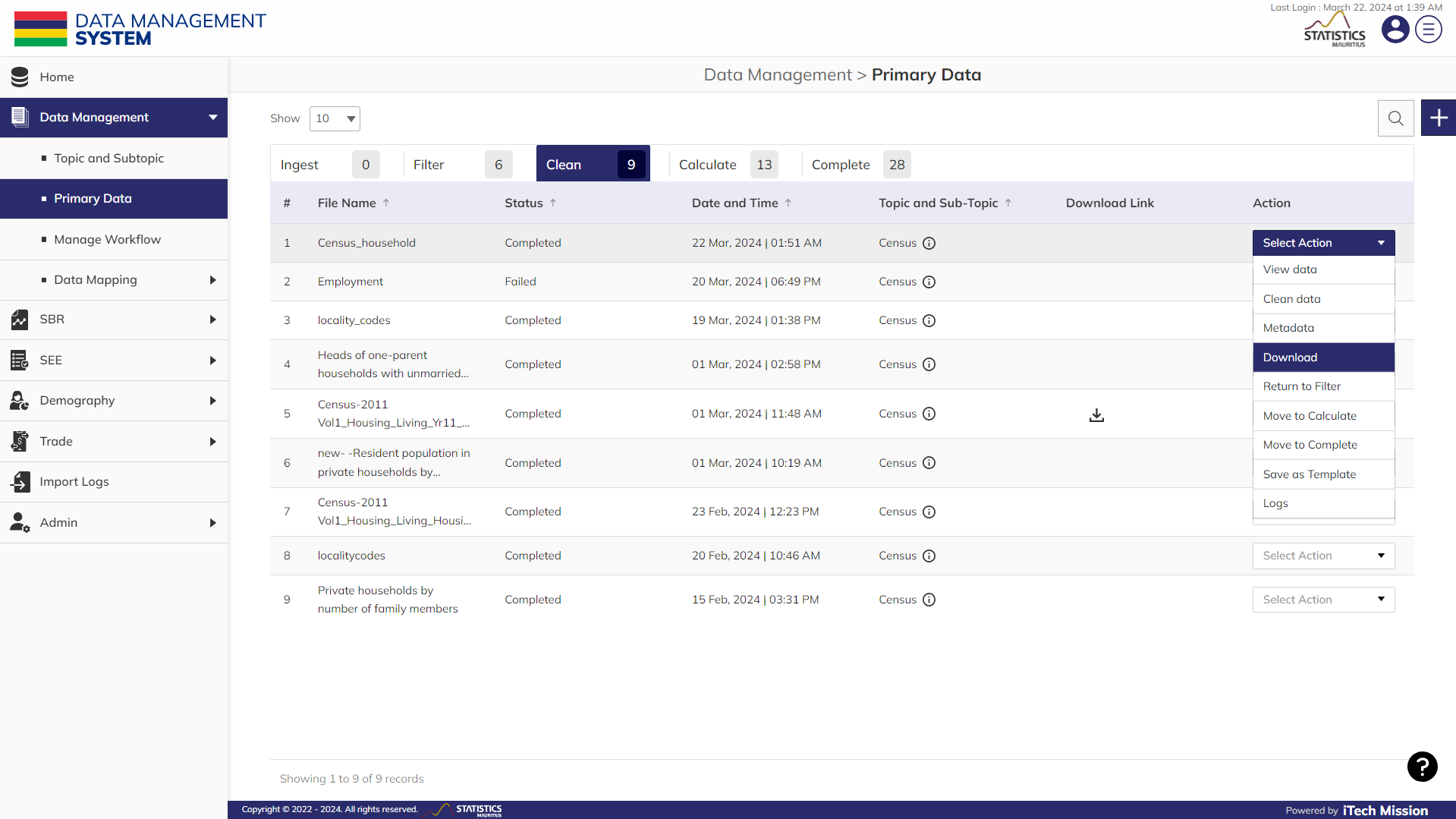
**Step 13**: Search the **Column Name** and click on the **Checkbox** to select it (see below figure).

**Step 14**: Click on the **Ok** button to confirm.

**Step 15**: Click on the **“Manage Column**" option located in the top right corner of the grid to display or hide the columns in the table belowandclick on the **Ok** button available in the dropdown to confirm**.**

**Step 16**: Click on the **Ok** button available at the bottom of the clean data pop up window to confirm the cleaning process.

There are following actions available in the dropdown of the “Clean” tab –

* View Data
* Clean Data
* Metadata
* Download
* Return to filter
* Move to calculate
* Move to complete

**NOTE:**

The status will show "In-progress" during the ongoing cleaning process. Once the cleaning process is finished, the status will be changed to “Completed”. The actions in the dropdown will appear as per the status.

**Step 17**: Click on the **Download** option available under action column to download the cleaned data on your local system.

**Step 18**: Click on the “**Move to calculate”** option from the dropdown available under action column to move the data to the “**Calculate**” tab view.

### 2.2.4 CALCULATE

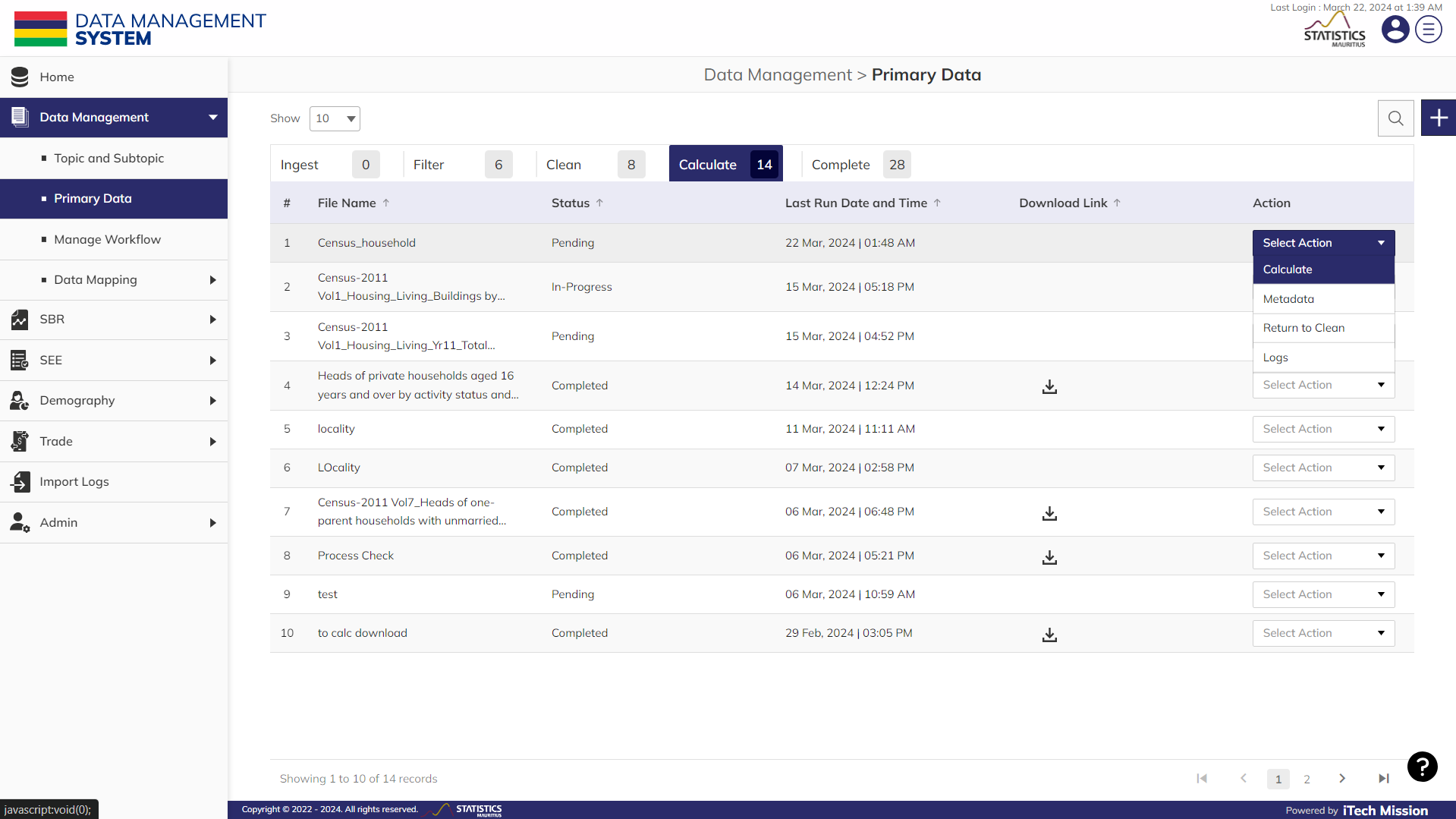
The **Calculate** tab allows you to create and manage arithmetic expressions that will be applied on the datasets to calculate data as per the requirement. There are options available to add, edit, delete and execute the arithmetic expressions and store the calculated data in the datasets for further usage.

**NOTE:**

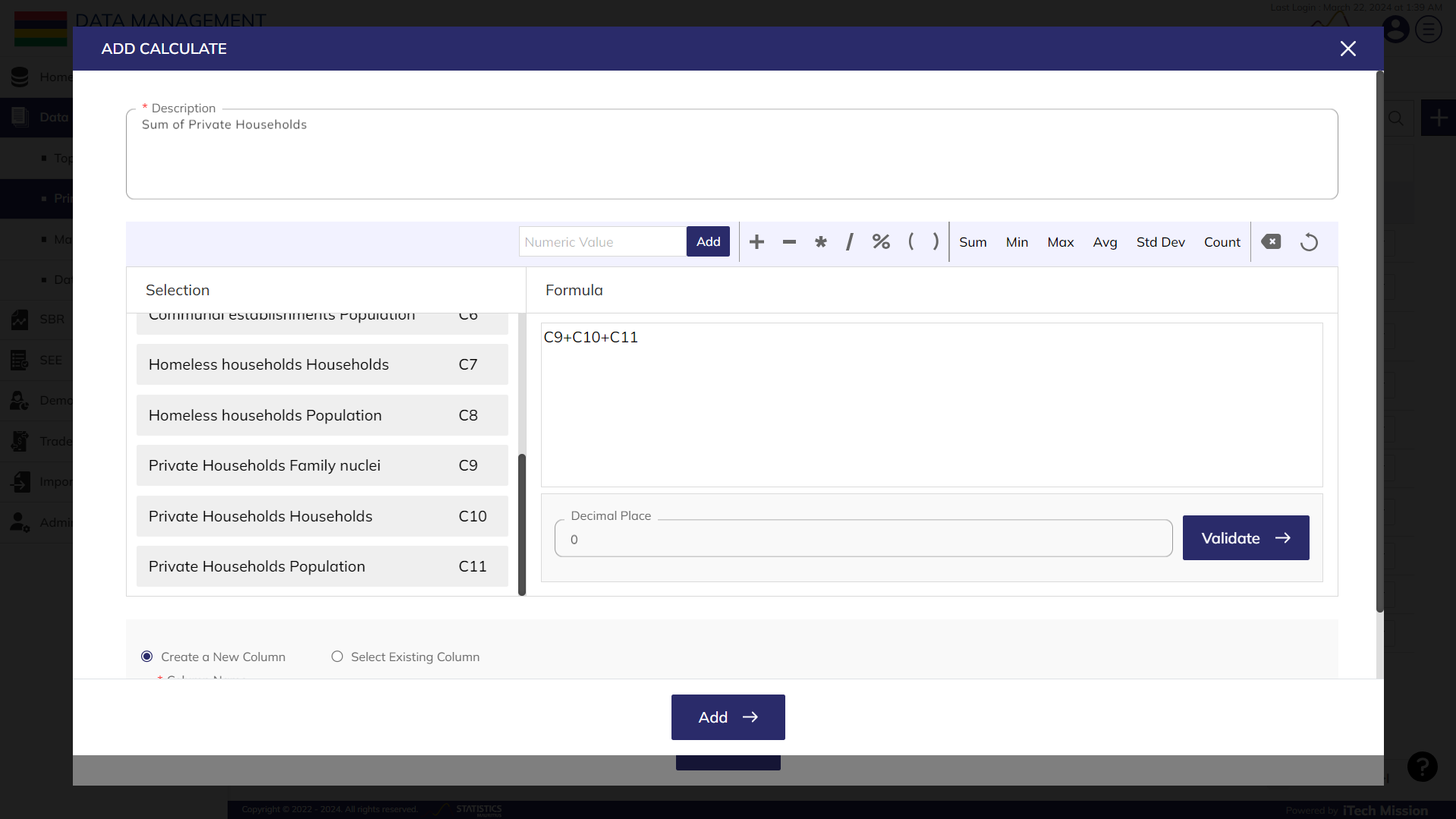
The status of the data will be "Pending" by default when it is moved to the calculate tab.

*Let us suppose you want to* ***add*** *the column values.*

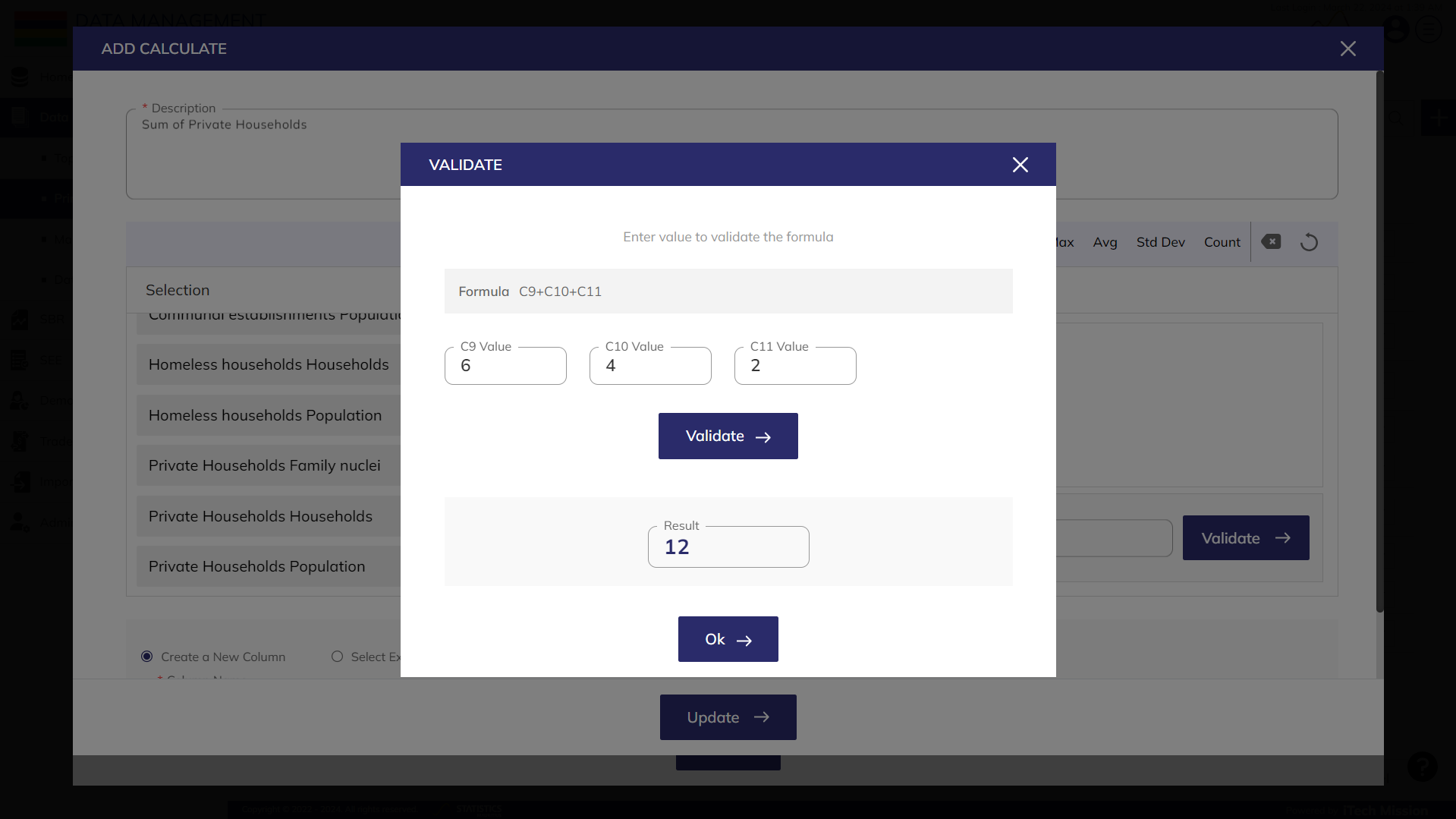
Below is the step-by-step guide on how to calculate this data.

**Step 19**: Click on the “**Calculate**” option from the dropdown available under action column (see below figure).

**Step 20**: Click on the **Plus** button available to the top right corner of the add calculate form to add calculation. Enter the following details to add a new calculation: -

* Enter the **Description** of the calculation.
* Select the **column 9** from section list.
* Select the operator **“+”** from the available operators.
* Now select the **column 10** from section list.
* Select the operator **“+”** from the available operators.
* Select the **column 11** from section list.
* **Create a new column** or **Select Existing column**. Enter **Column Name** if a new column is created or select existing column from the dropdown if other option is selected.

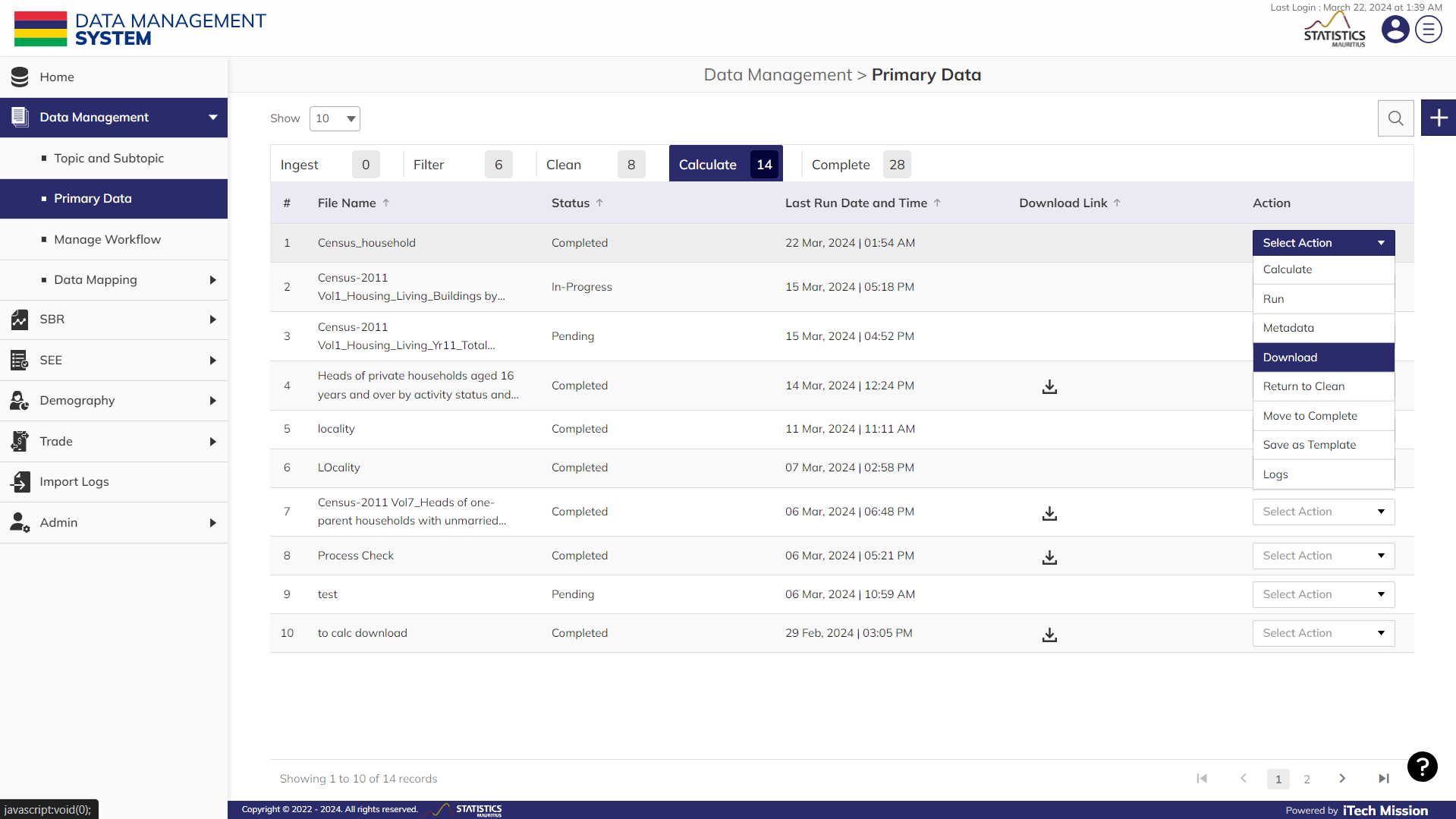
**Step 21**: Click on the “**Validate”** button to check whether the created expression is arithmetically correct.

**Step 22**: After the expression is validated, click on the **Ok** button to confirm.

**Step 23**: Click on the “**Add**” button to create the calculation.

There are following actions available in the dropdown in the “Calculate” tab once the status of the data changes to “**Completed**”–

* Calculate
* Run
* Metadata
* Download
* Return to clean
* Move to complete
* Save as Template
* Logs

**Step 24**: Click on the “**Run**” option from the dropdown available under action column to run the selected calculation.

**NOTE:**

The status will show "In-progress" during the ongoing calculation process. Once the calculation process is finished, the status will be changed to “Completed”. The actions in the dropdown will appear as per the status.

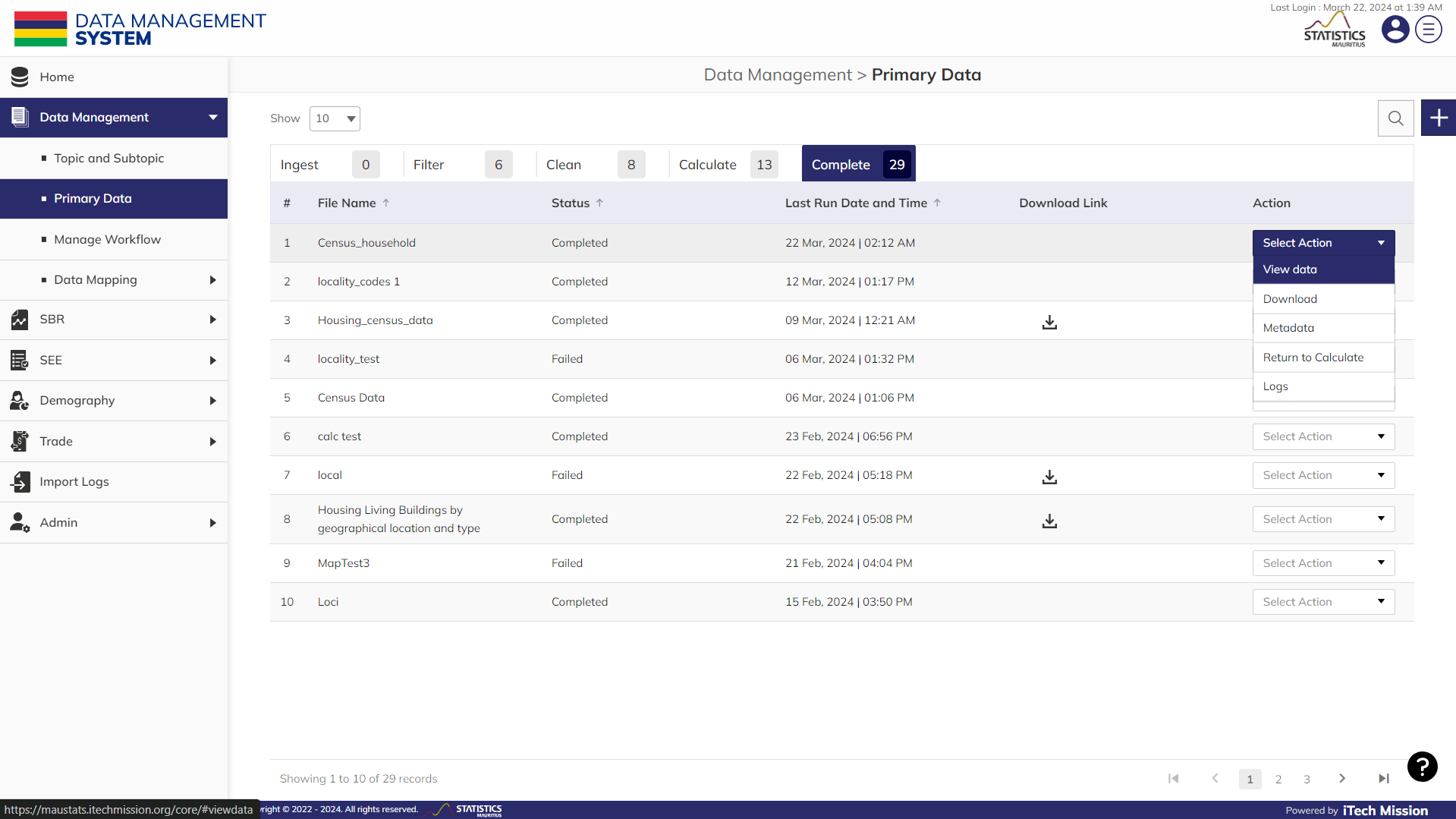
**Step 25**: Click on the **Download** option from the action dropdown to download the calculated data on your local system.

**Step 26**: Click on the “**Move to complete”** option from the dropdown available under action column to move the data to the “**Complete**” tab view.

### 2.2.5 COMPLETE

This is the last tab view and all the data which are moved to completed tab can be viewed on the **Data Dissemination** module.

There are following actions available in the dropdown in the “Complete” tab –

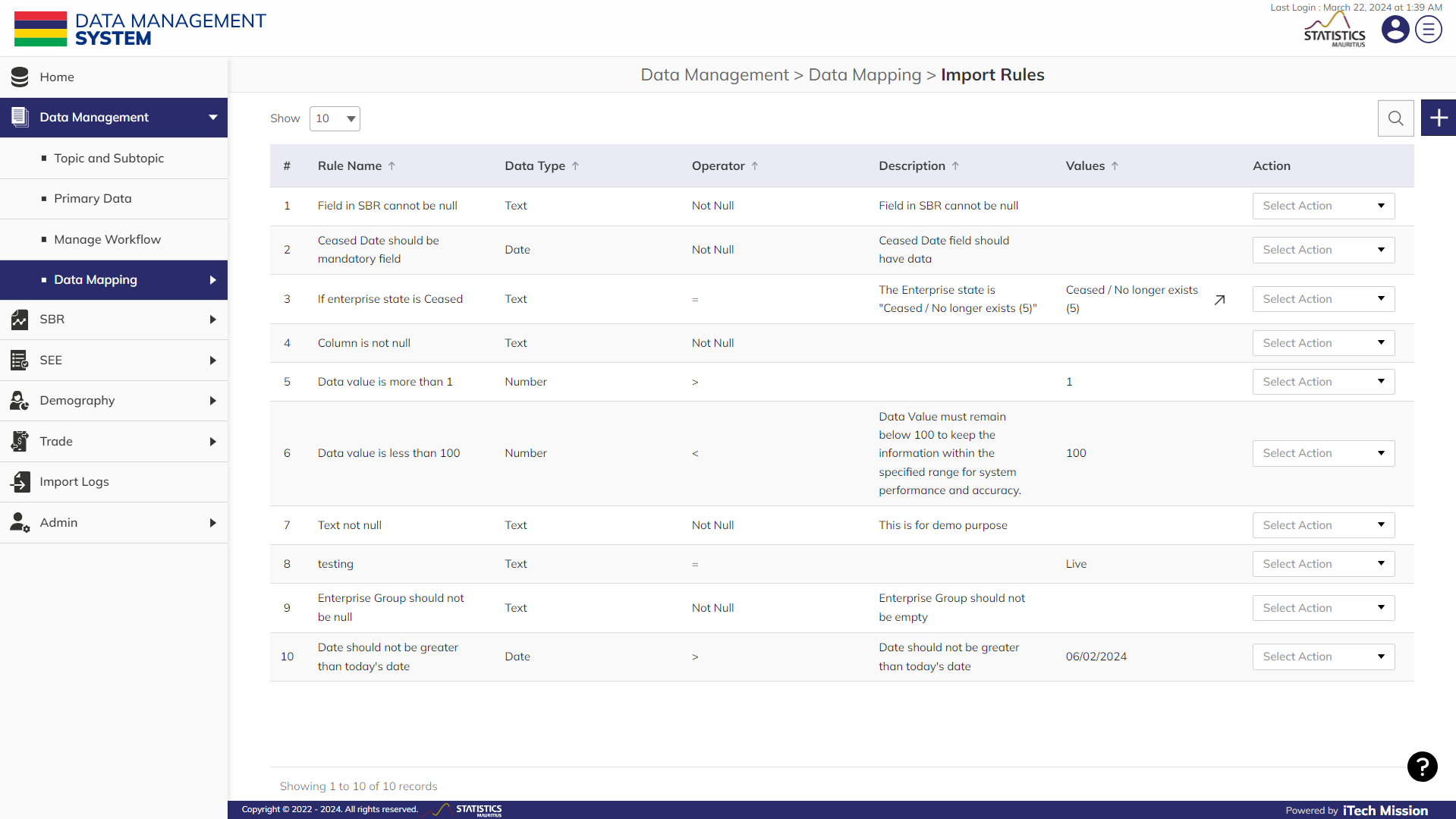
* View data
* Download
* Metadata
* Return to Calculate
* Logs

**Step 27**: Click on the **Download** option from the action dropdown to download the completed data on your local system.

## 2.3 DATA MAPPING

**Data mapping** modules have two sub-modules – Import rules, Import mapping. Below is the step-by-step guide on how to use the **Import rules** and **Import mapping** sub-module in detail.

### 2.3.1 IMPORT RULES

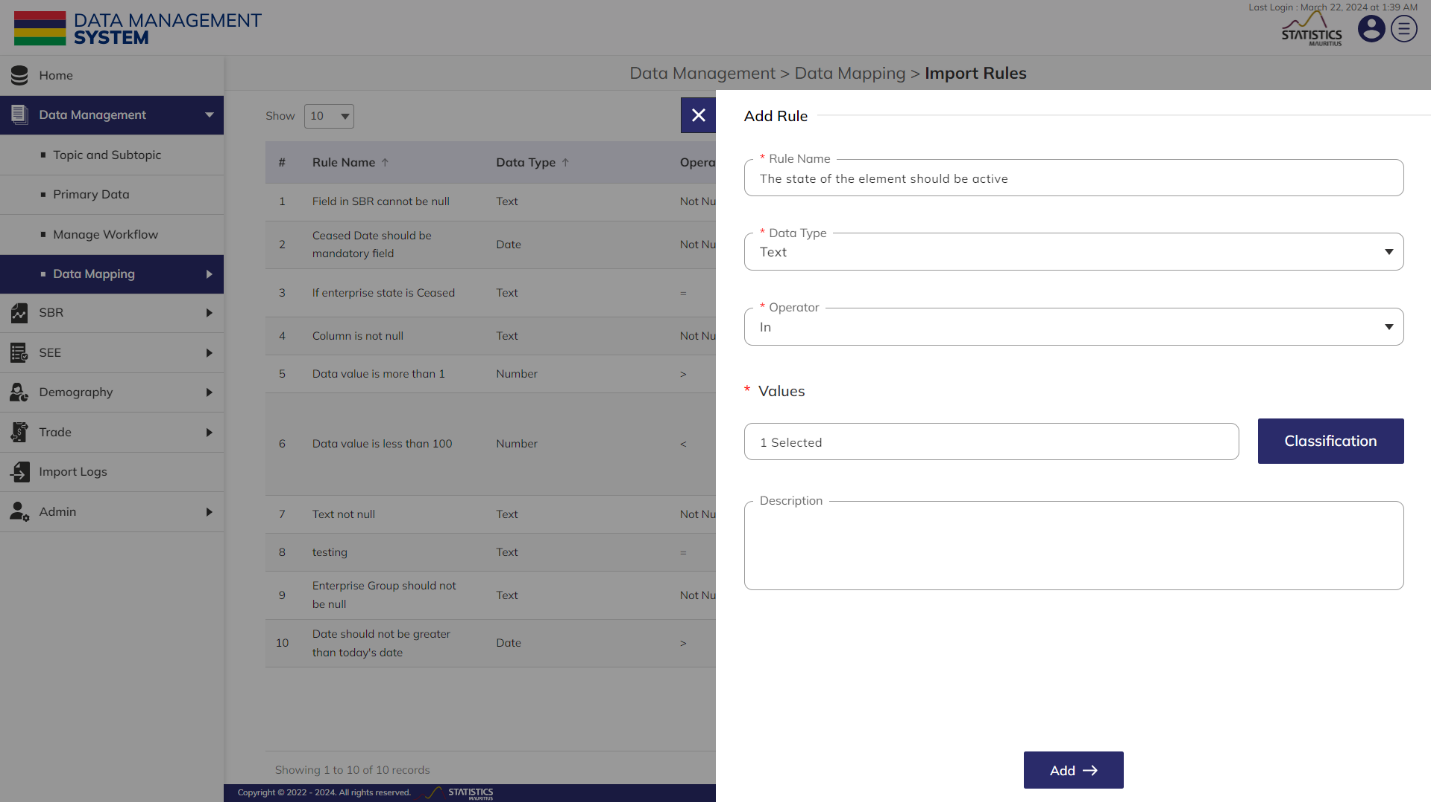
**Step 1:** Click on the **Import Rules** option available under the Data **Mapping** module from the left panel to access this sub-module (see below figure). This sub-module allows you to create rules for importing data into specific fields.

Below are the steps to create and manage the import rules.

**NOTE:**

The field “**Value**” while adding an import rule will appear based on the selected operator. For example, selecting “Unique” as operator does not require to specify value.

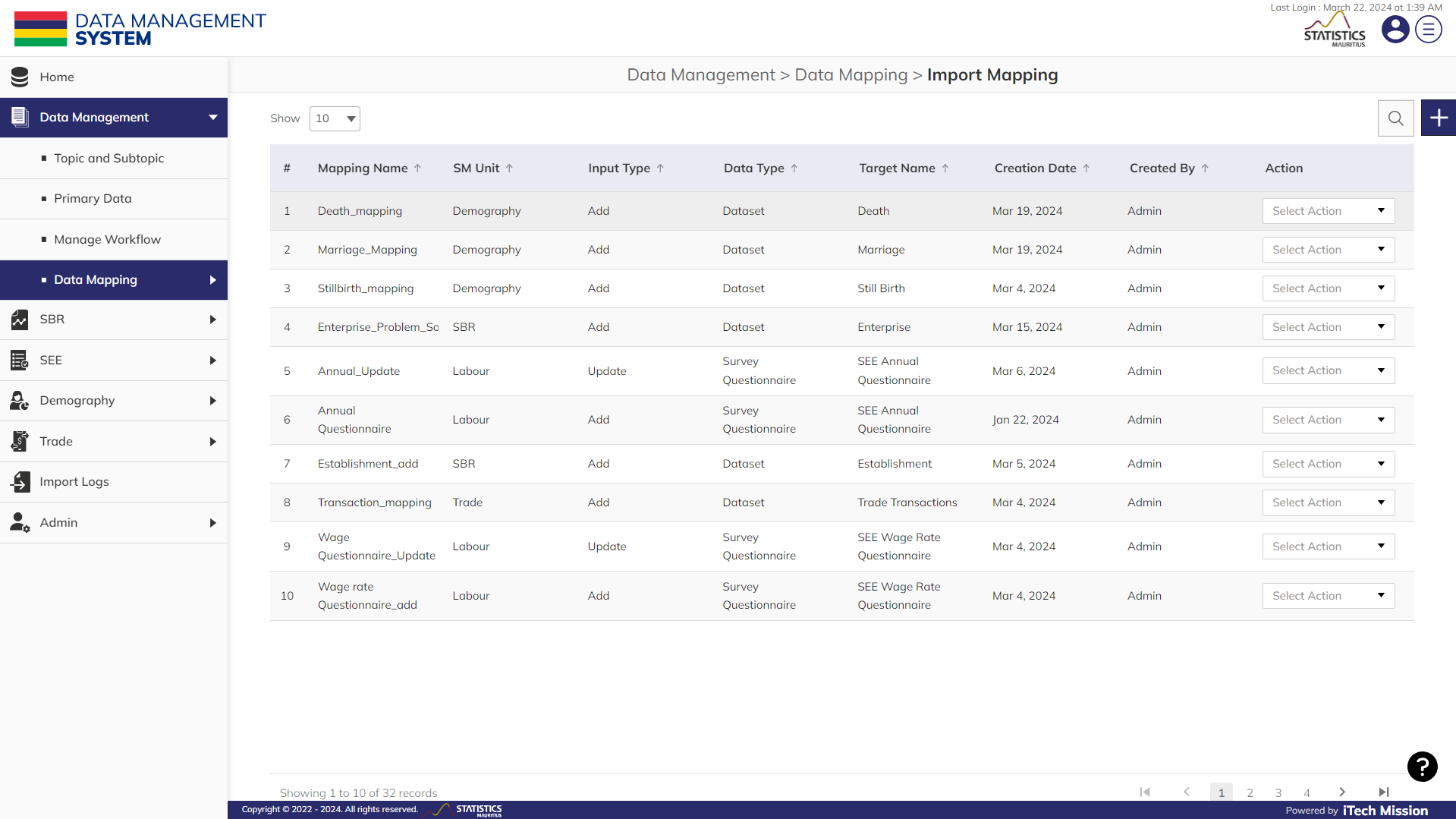
**Step 2:** Click on the Plus button available to the right side of the page to add a new import rule (see figure below). Enter the following details to add a new rule:-

* Enter the **Rule name**.
* Select **Data type**.
* Select **Operator**.
* Select or Enter **Values** (If required).
* Enter **Description** of the import rule.

All the details that are mandatory to be entered.Click on the **Add** button to save and confirm. You can now view the new import rule added in the list.

**Step 3:** Click to select **Edit** option available in the action dropdown to edit the details of the selected element. Edit the required detail and click on the update button to save and confirm.

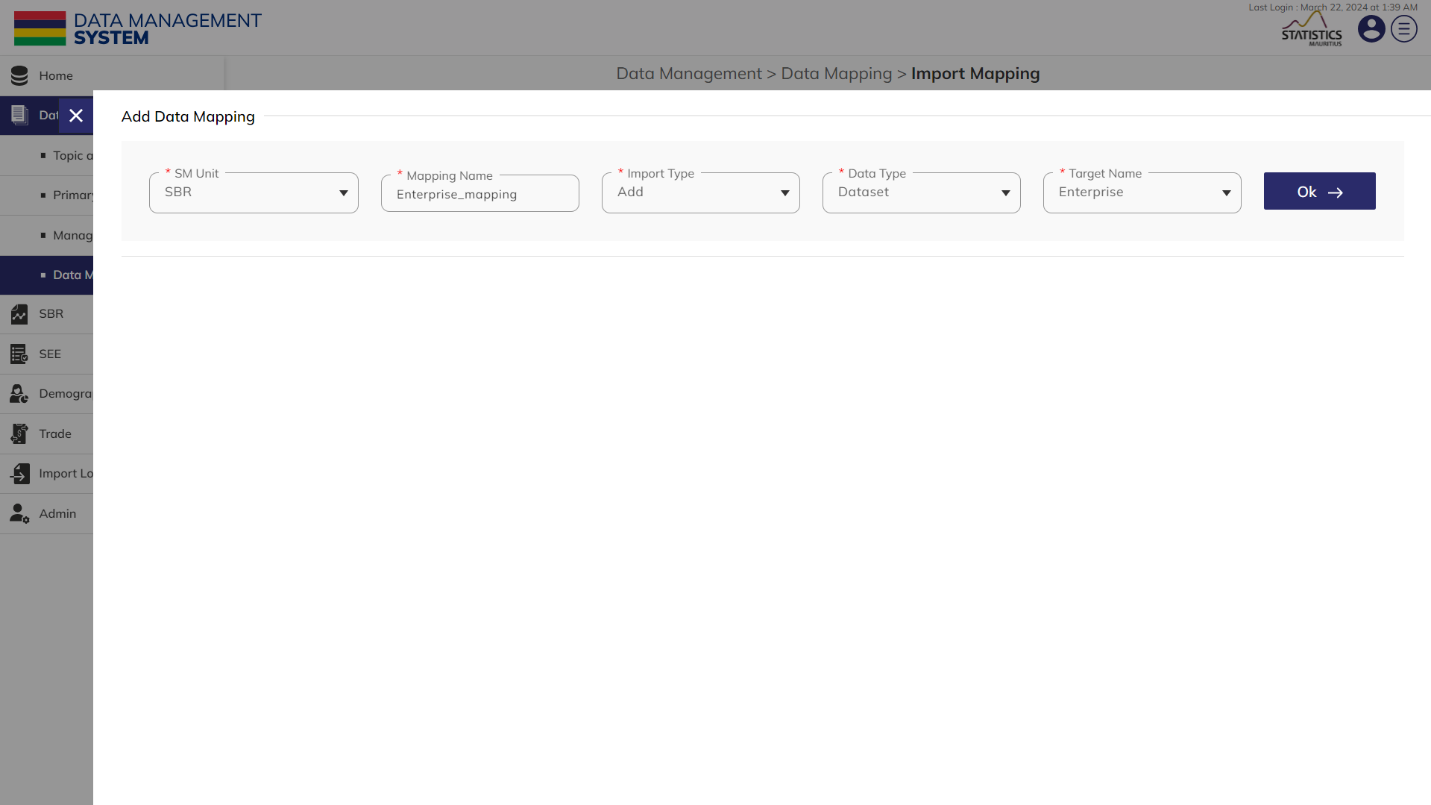
### 2.3.2 IMPORT MAPPING

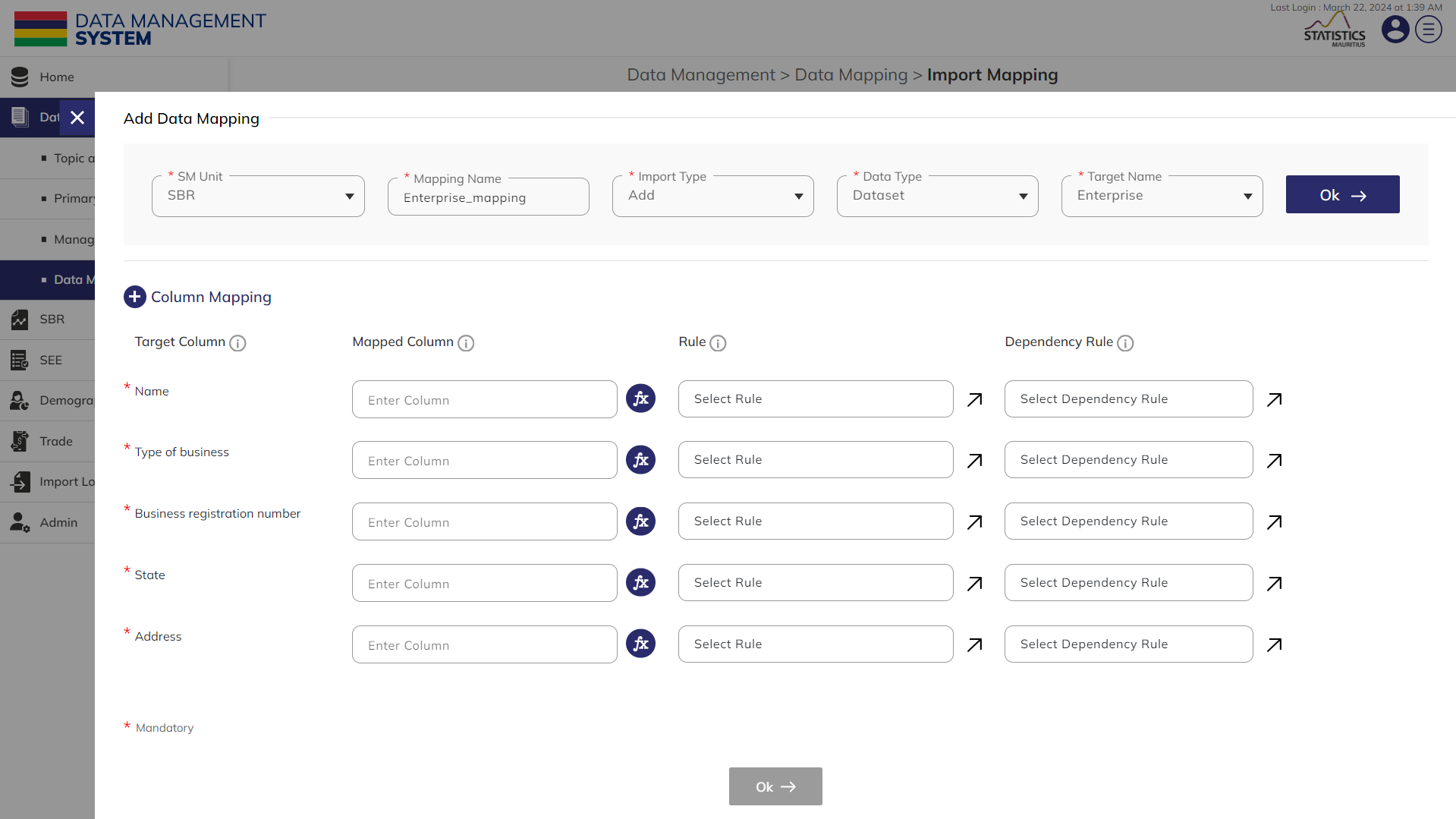
Click on the **Import Mapping** option available under the **Data Mapping** module from the left panel to access this sub-module (see below figure). This submodule allows the authorized users to create and manage the data mapping utility for adding and updating the data. Users can create the mapping files by specifying the correspondence between the target column and the mapped columns to import the respective data.

Below are the steps to create and manage the map file for CBRD enterprise registration data.

**Step 4:** Click on the Plus button available to the right side of the page to add a new import mapping (see figure below). Enter the following details to add a new element.

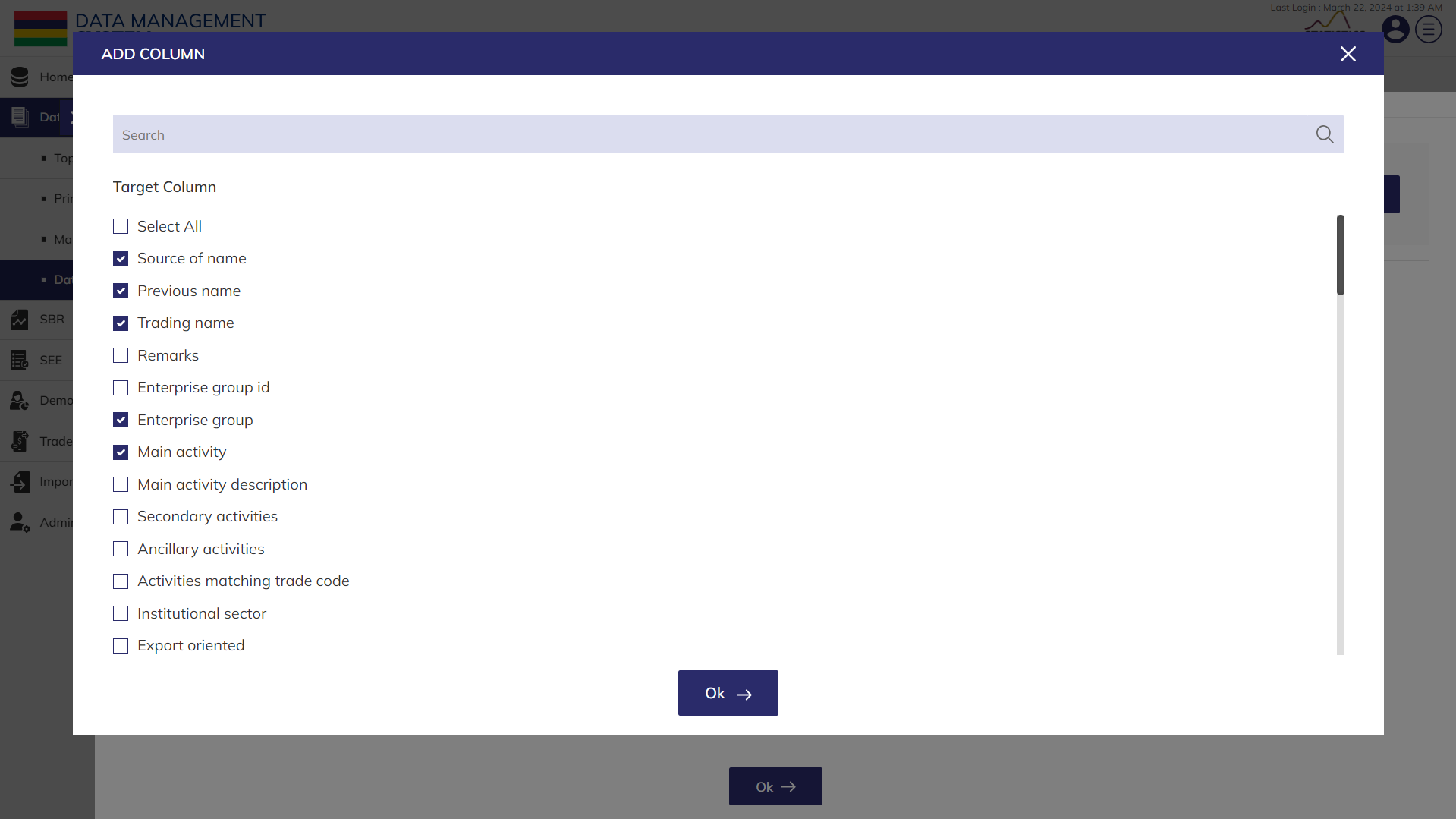
* Select **SM Unit**
* Enter the **Mapping Name**
* Select the **Import Type** as Add
* Select the **Data Type** as **Dataset**
* Select the **Target name**

**Step 5:** Click on the **Ok** button.

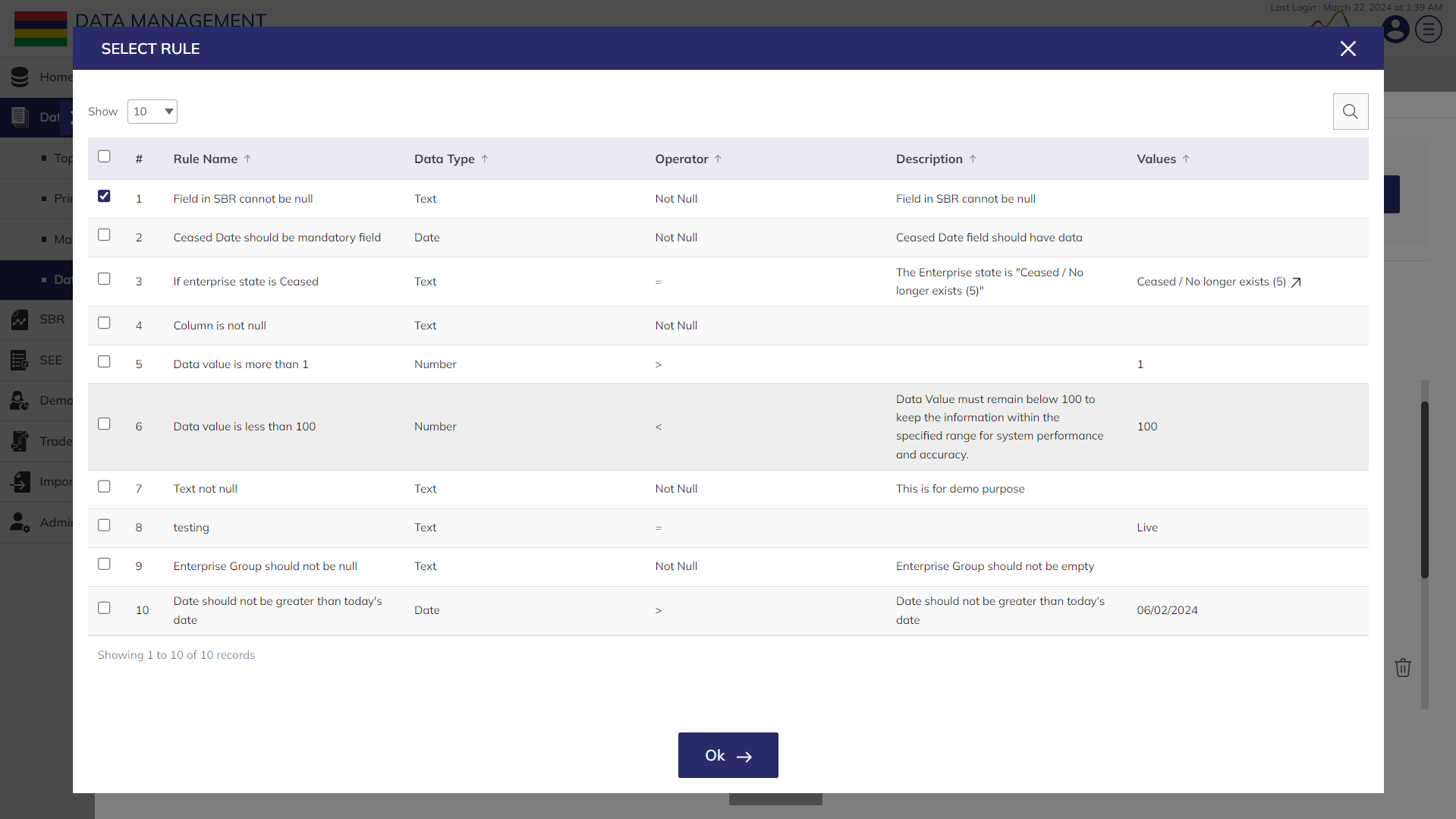
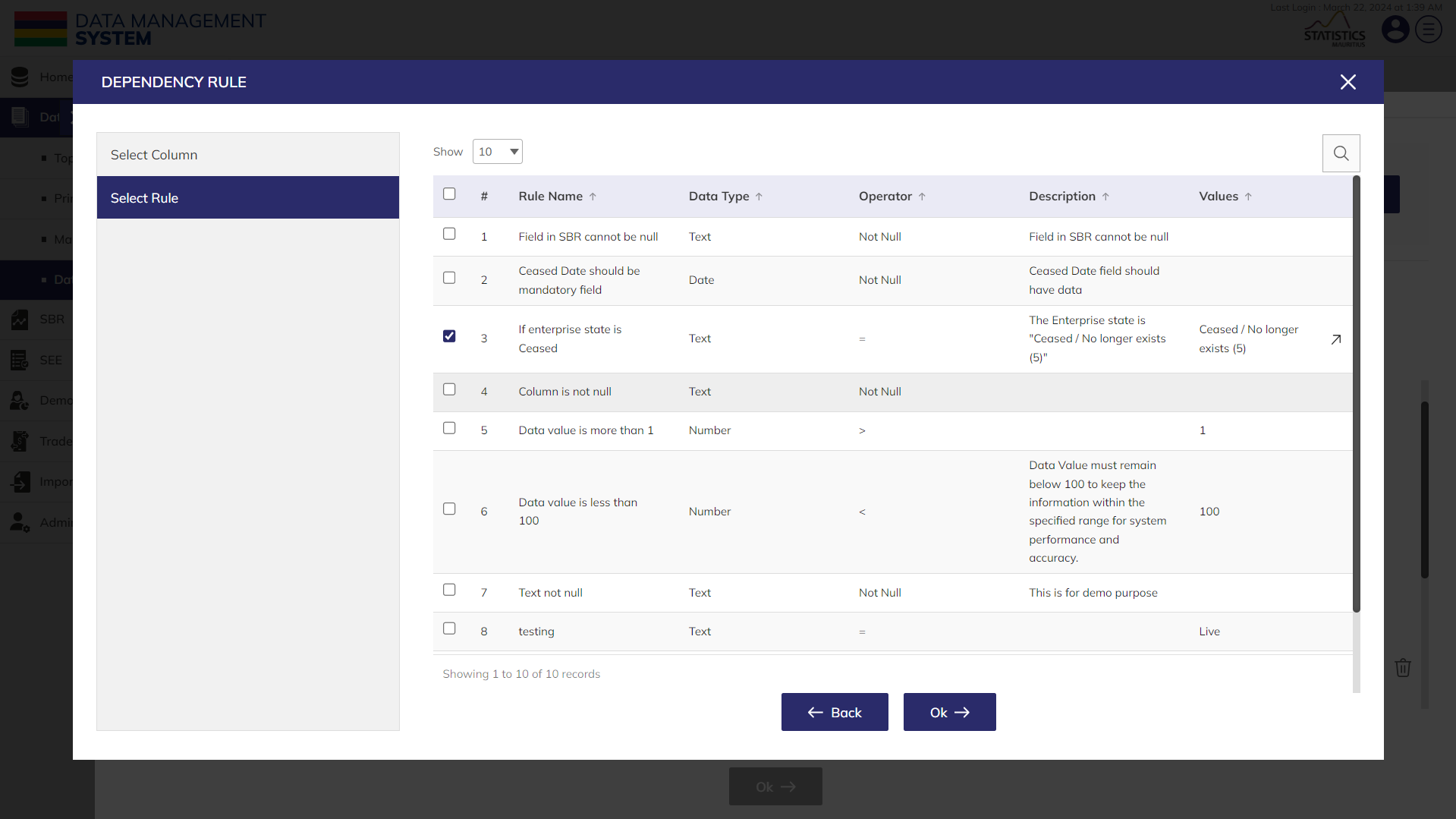
The mandatory field will appear below with the option to map columns and apply rules to each field (see below figure).

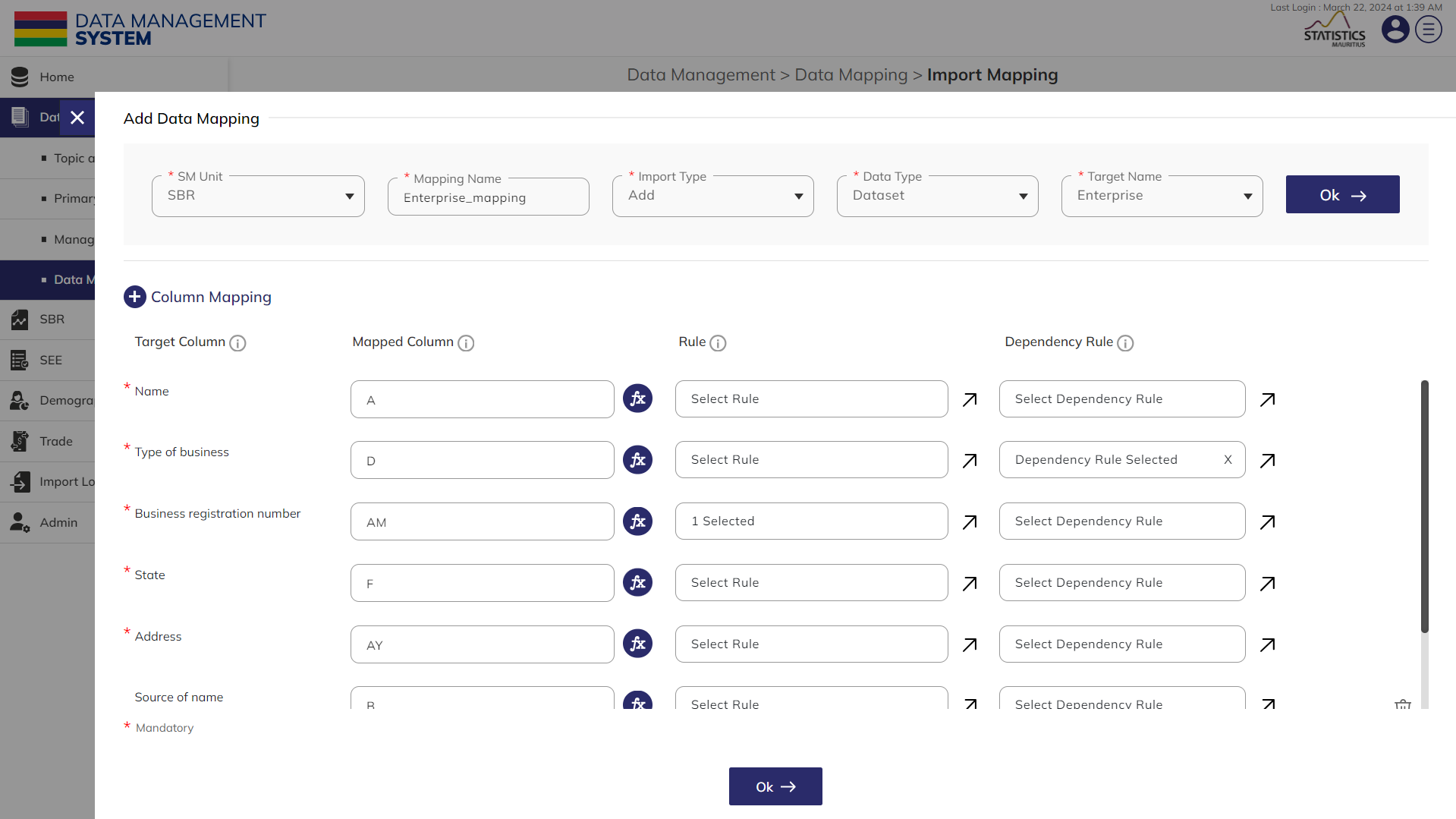
**Step 6:** **Map** the mandatory **Columns** of the selected **Target**

**Step 7:** Click on the **Plus** sign available with the mapping column to add additional columns (if required).

**Step 8**: Click on the **Checkbox** available with the field name to select it. Click on the **Ok** button confirm (see below figure)

**Step 9:** Click on the **Arrow** button available with the filed under **Rule** and select the required rule. Click on the **Ok** button to apply (see below figure).

**Step 10:** Click on the **Arrow** button available with the filed under **Dependency Rule** and select the **Target Column** first and click on **Next** to select the required **Rule**. Click on the **Ok** button to apply (see below figure).

**Step 11:** Click on the **Ok** button to confirm creating the map file for importing CBRD enterprise registration data (see below figure).

You can now use this map file for importing the enterprise registration data using the Import option available in the **Enterprise** module.

**Step 12:** Click to select **Edit** option available in the action dropdown to edit the details of the selected element. Edit the required detail and click on update button to save and confirm.

**Step 13:** Click to select **Delete** option available in the action dropdown to delete the selected element. A confirmation popup will appear, click yes to confirm deletion of selected element.



PRODUCT DOCUMENT