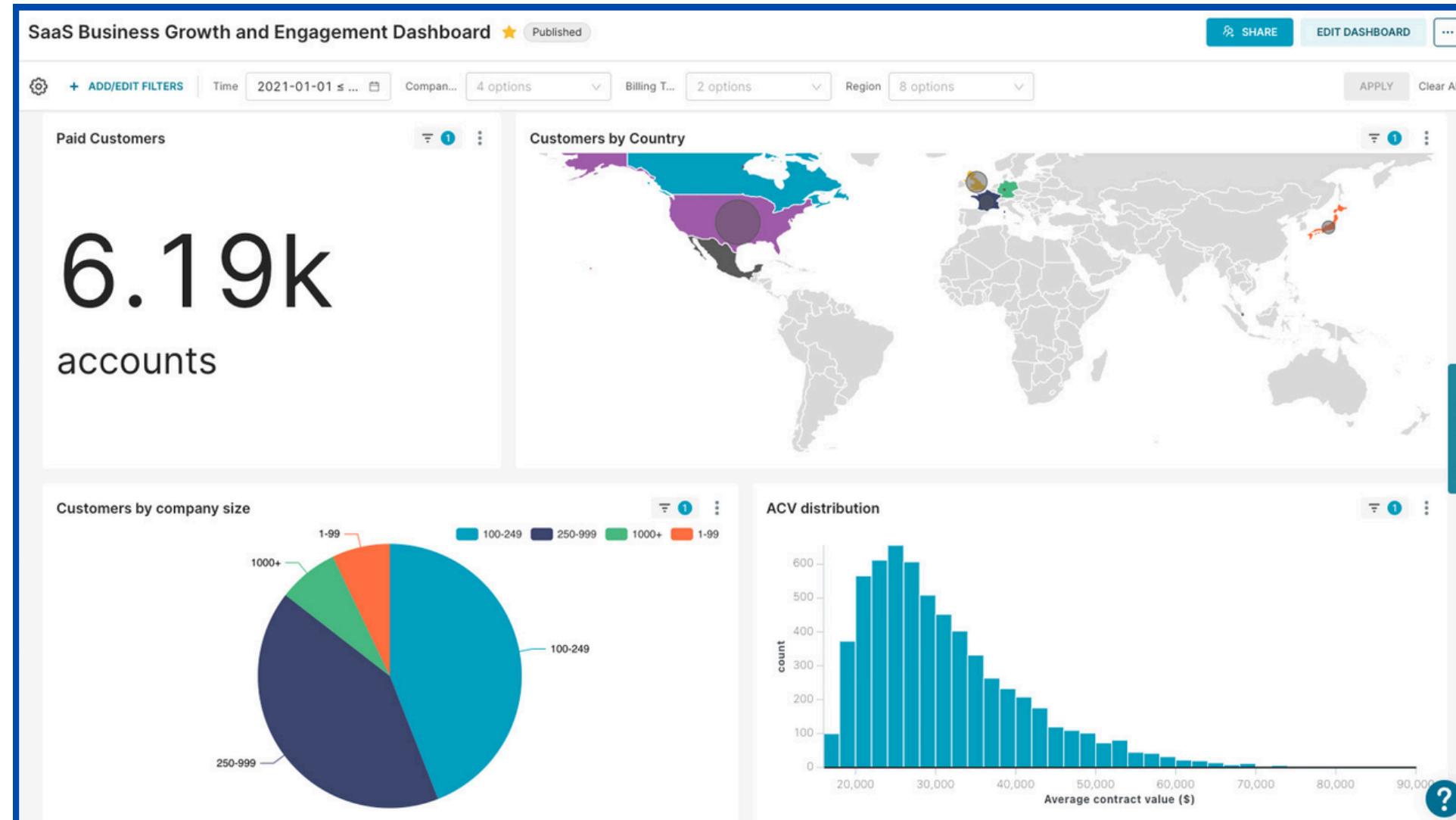


minicurso | 2024



Apache Superset

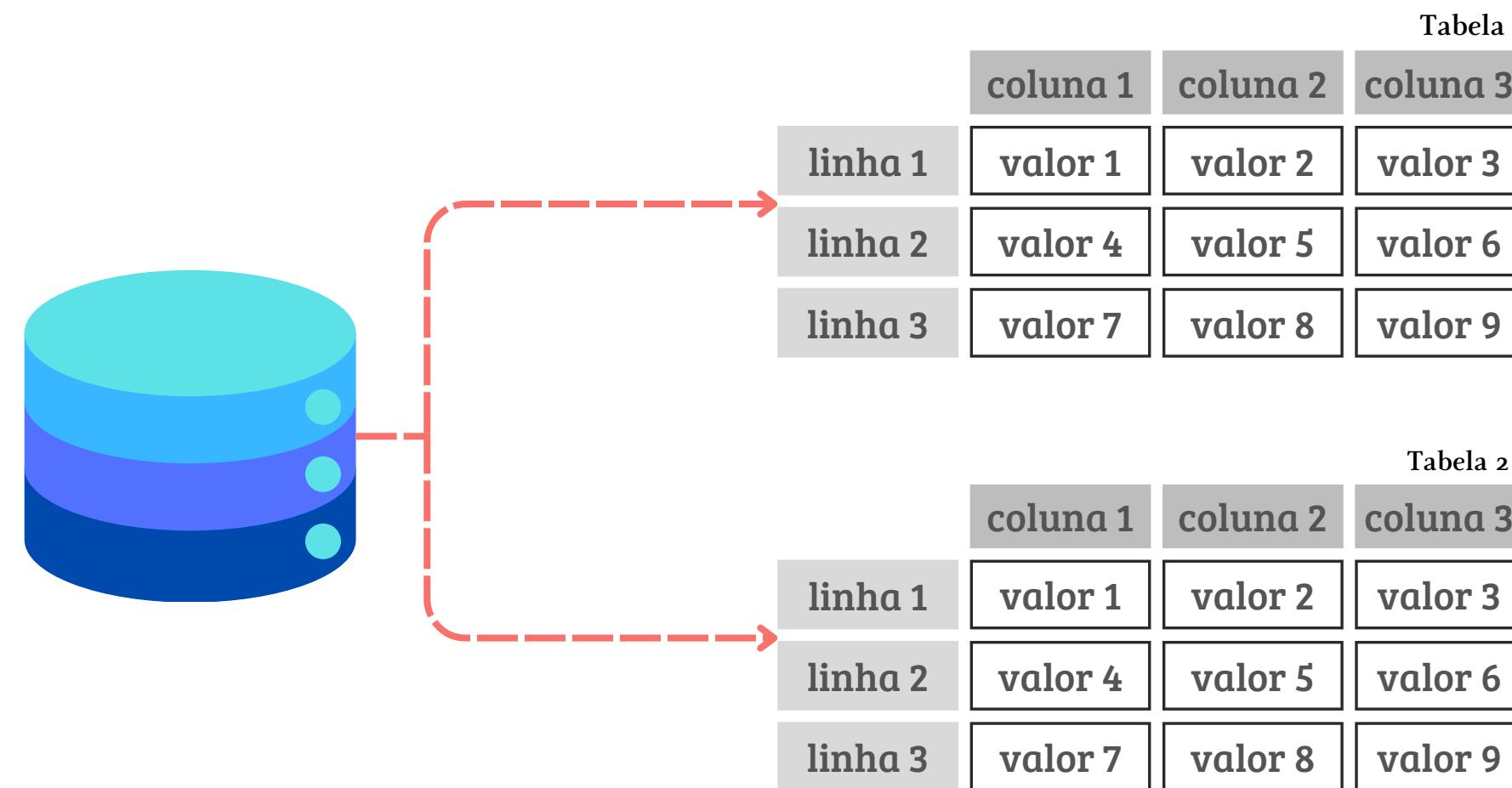
Marília Melo Favalesso



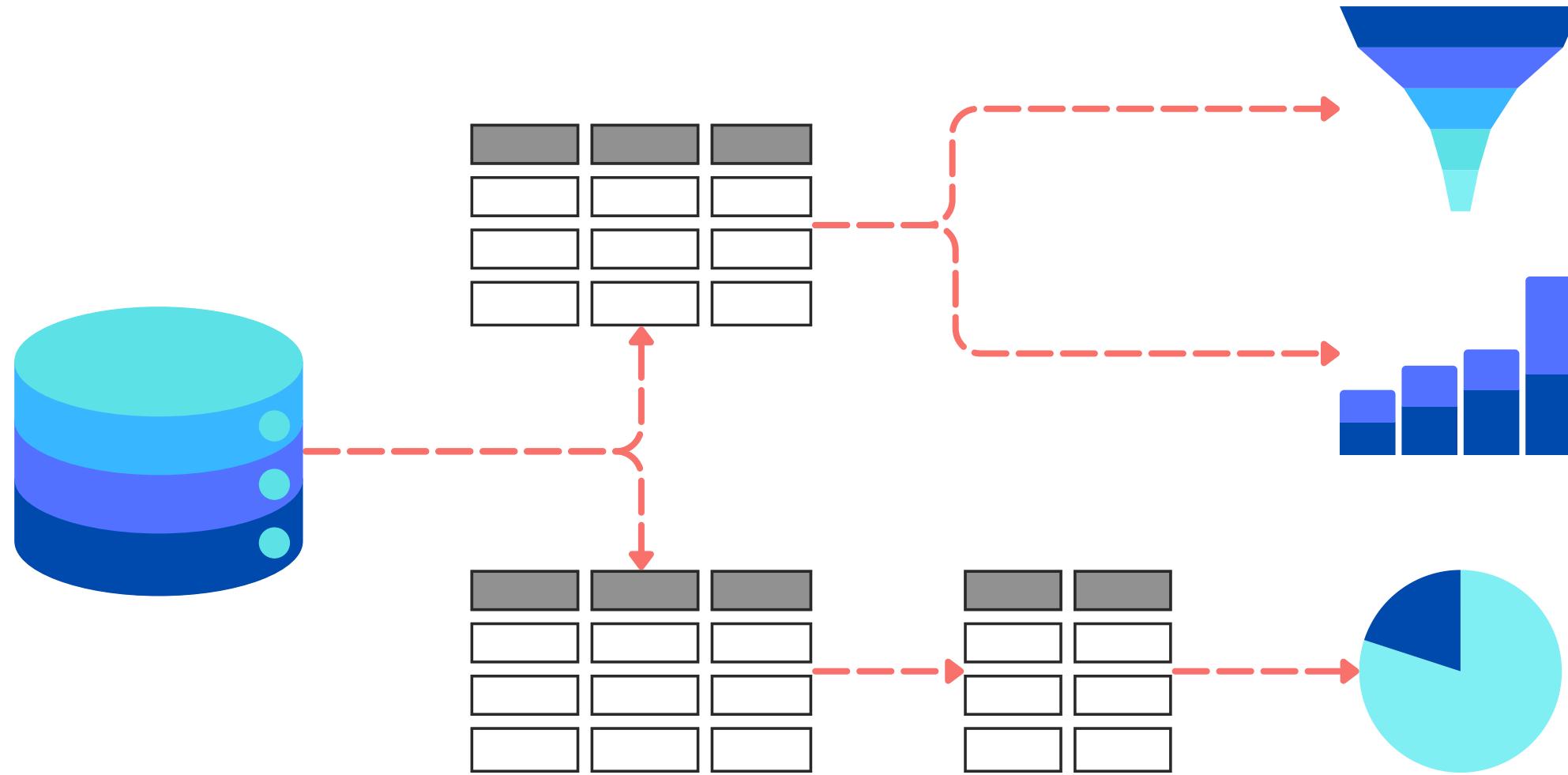
O Apache Superset é uma plataforma de análise de dados.

Oferece uma interface amigável para consultar dados, criar e acessar relatórios dinâmicos, produzir gráficos e fazer análises sem precisar de conhecimentos avançado em programação ou análise de dados.

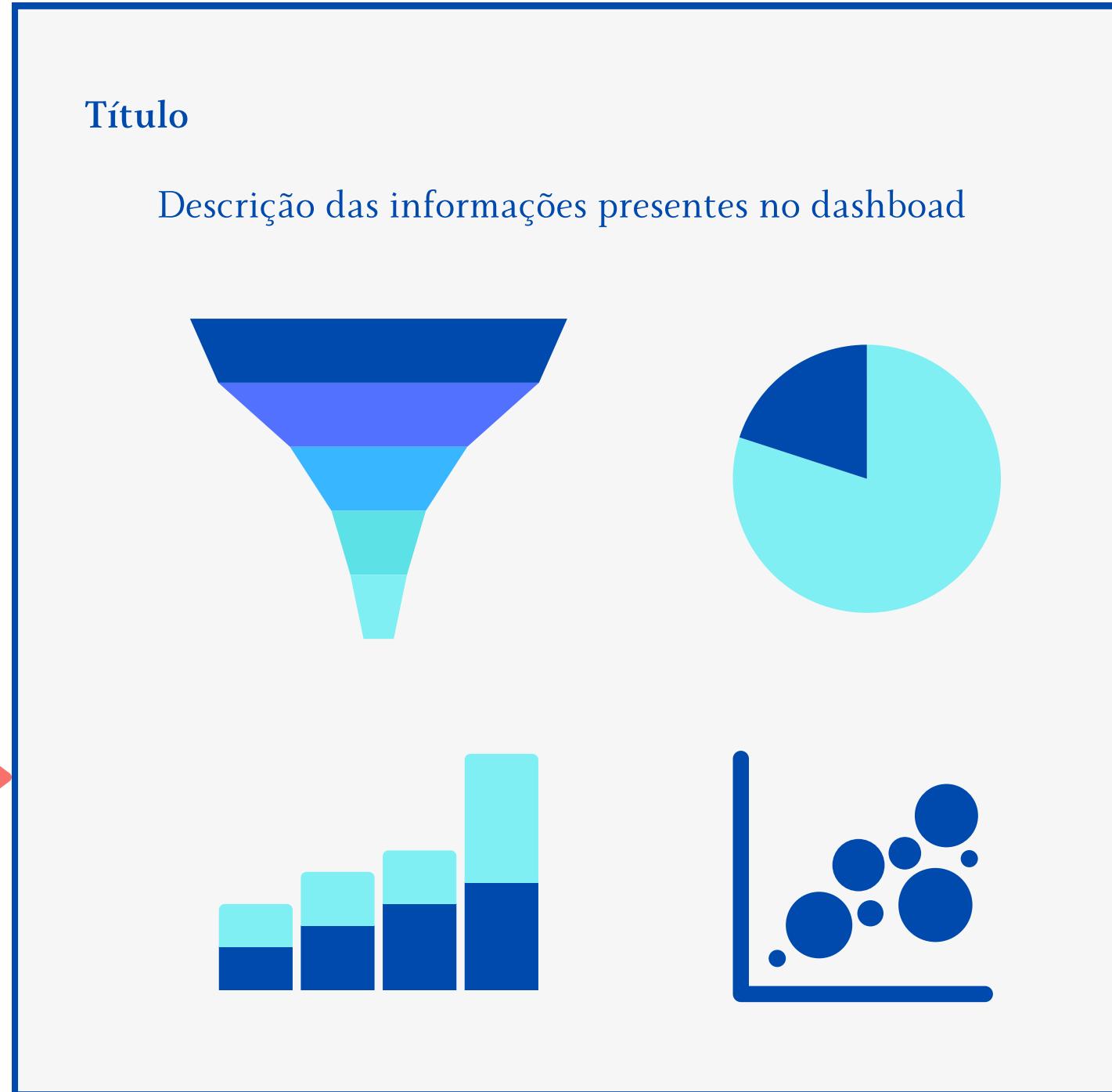
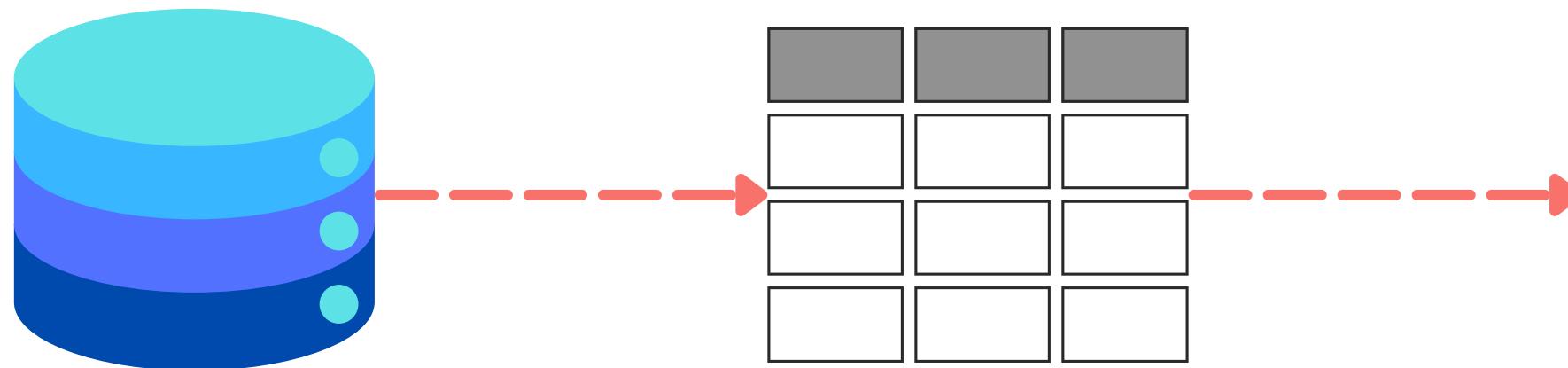
No Apache Superset, a análise e visualização de dados ocorrem em conjuntos de **dados estruturados** (tabelas com linhas e colunas) **provenientes** da fonte de dados escolhida.

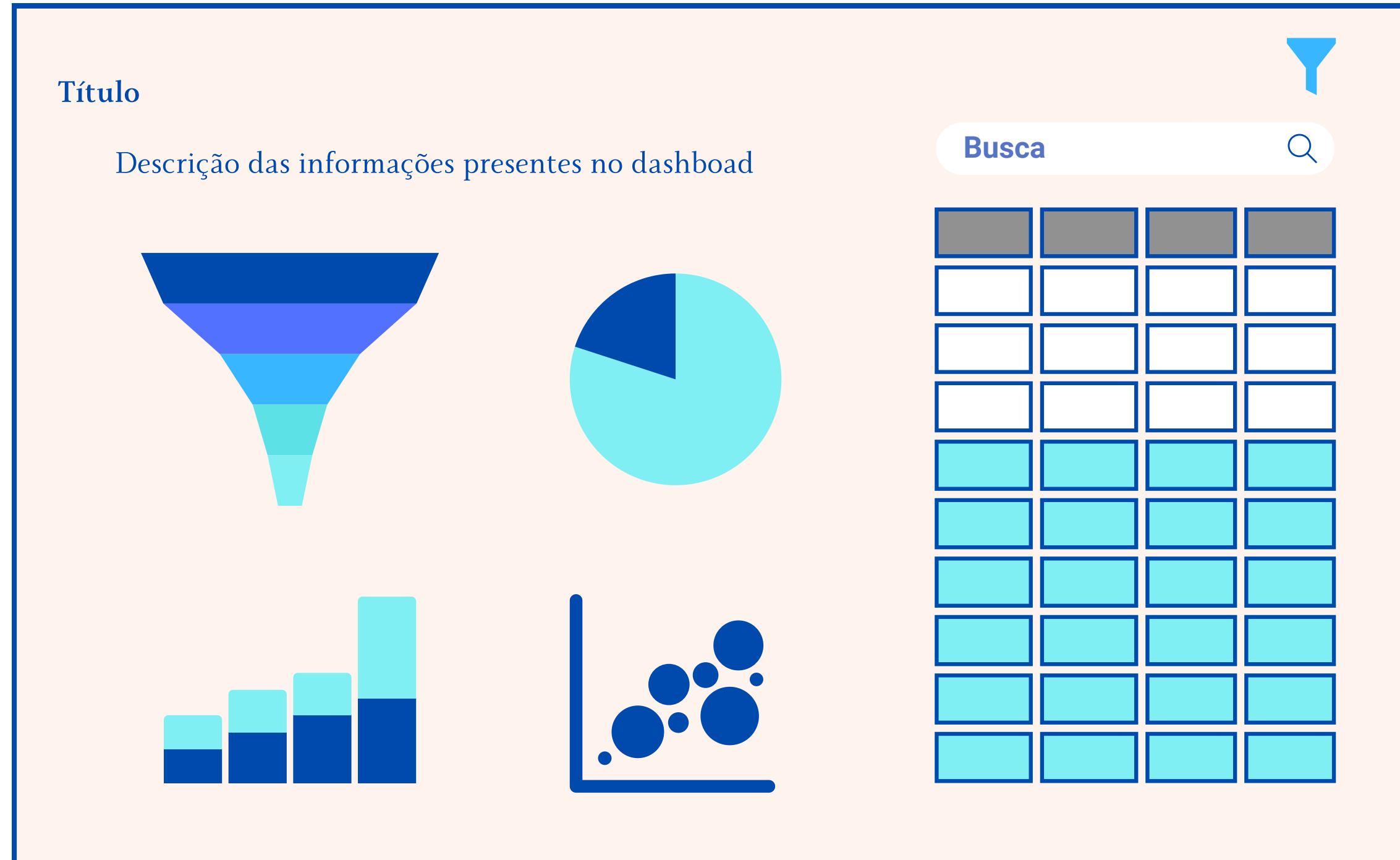


Os dados provenientes das tabelas podem ser manipulados e tratados para compor visualizações personalizadas. O Apache Superset conta com **40+ tipos de visualizações**.

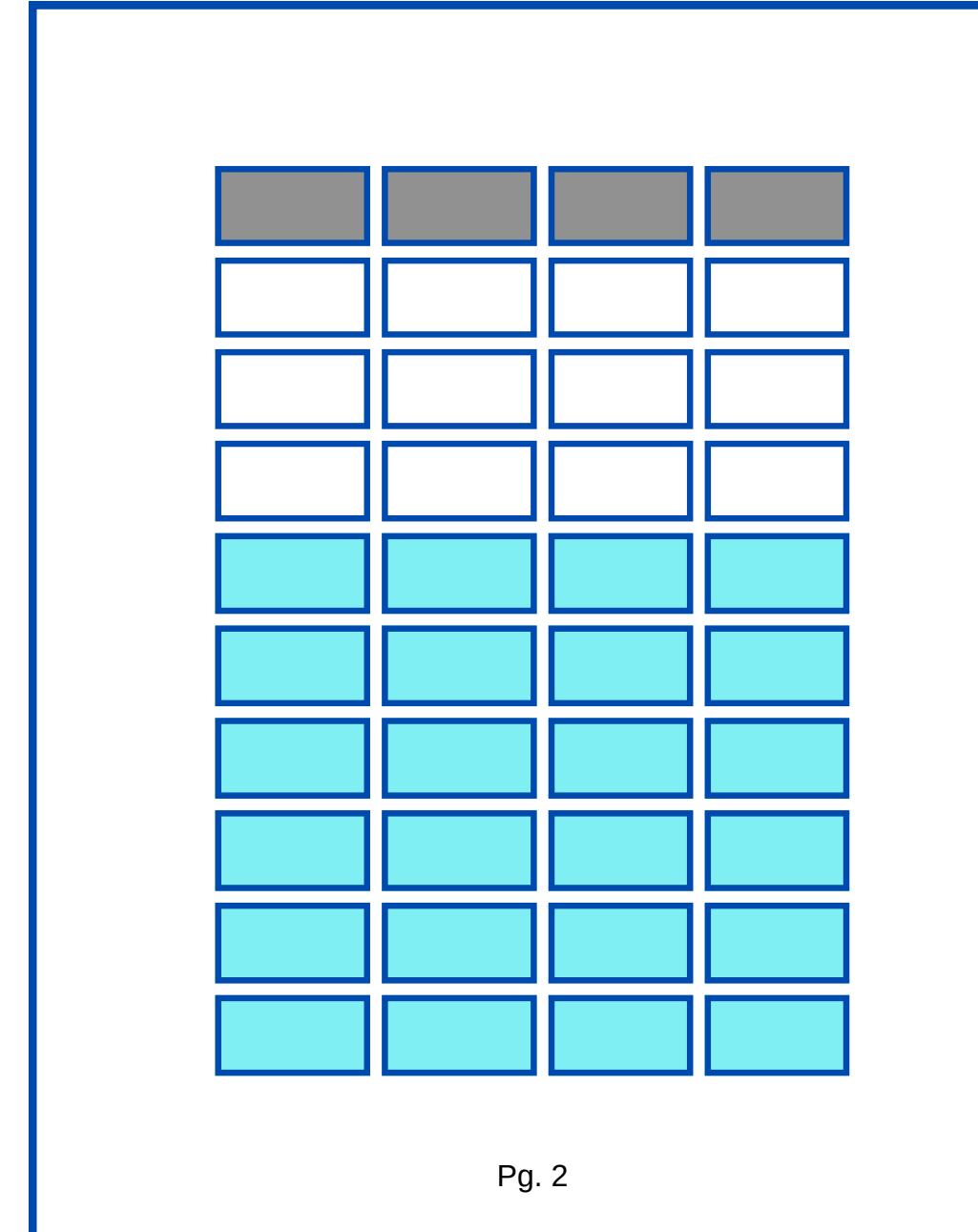
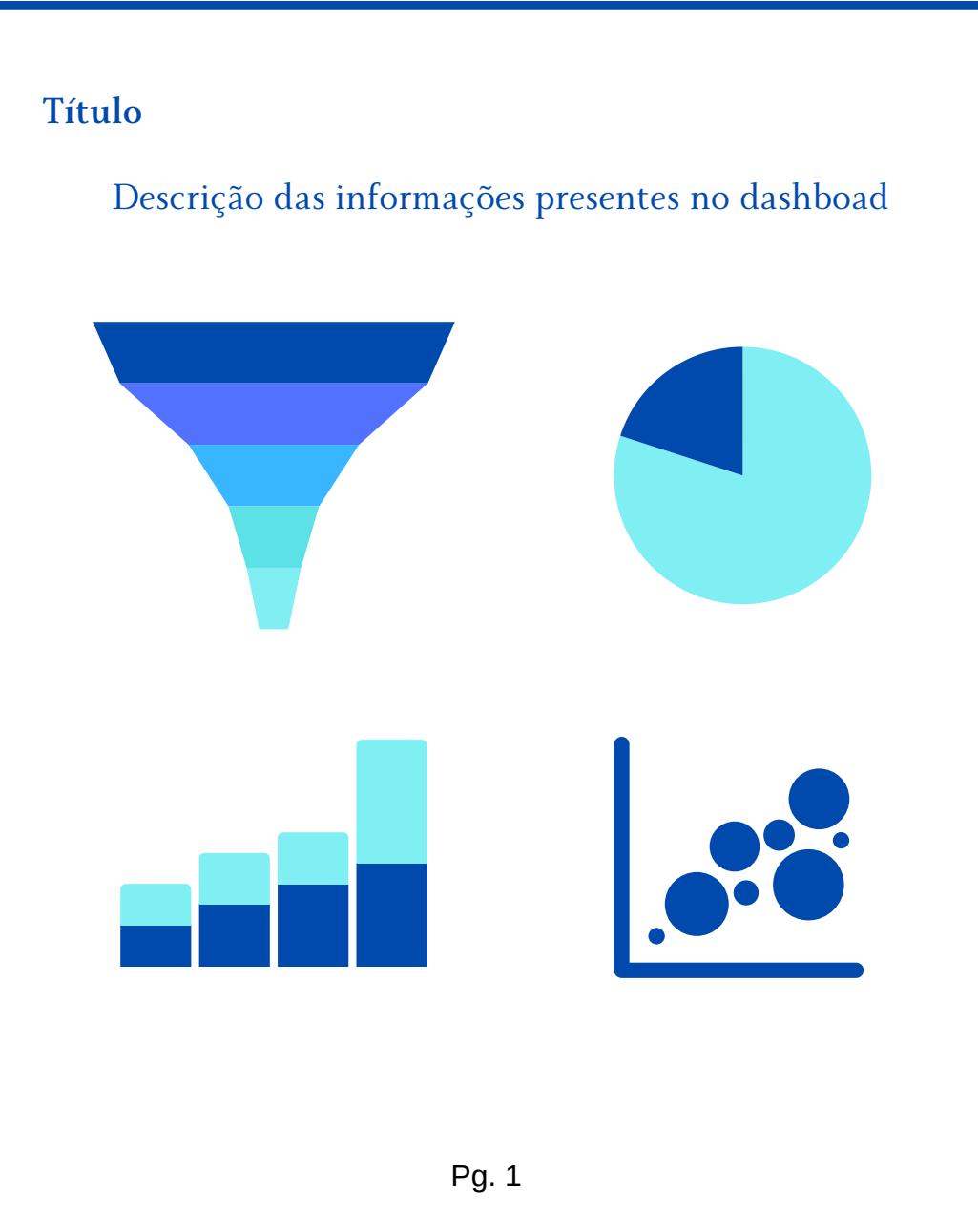


As visualizações, por sua vez, são combinados em **dashboard** (painéis ou relatórios analíticos/dinâmicos) que fornecem uma visão abrangente e interativa dos dados.



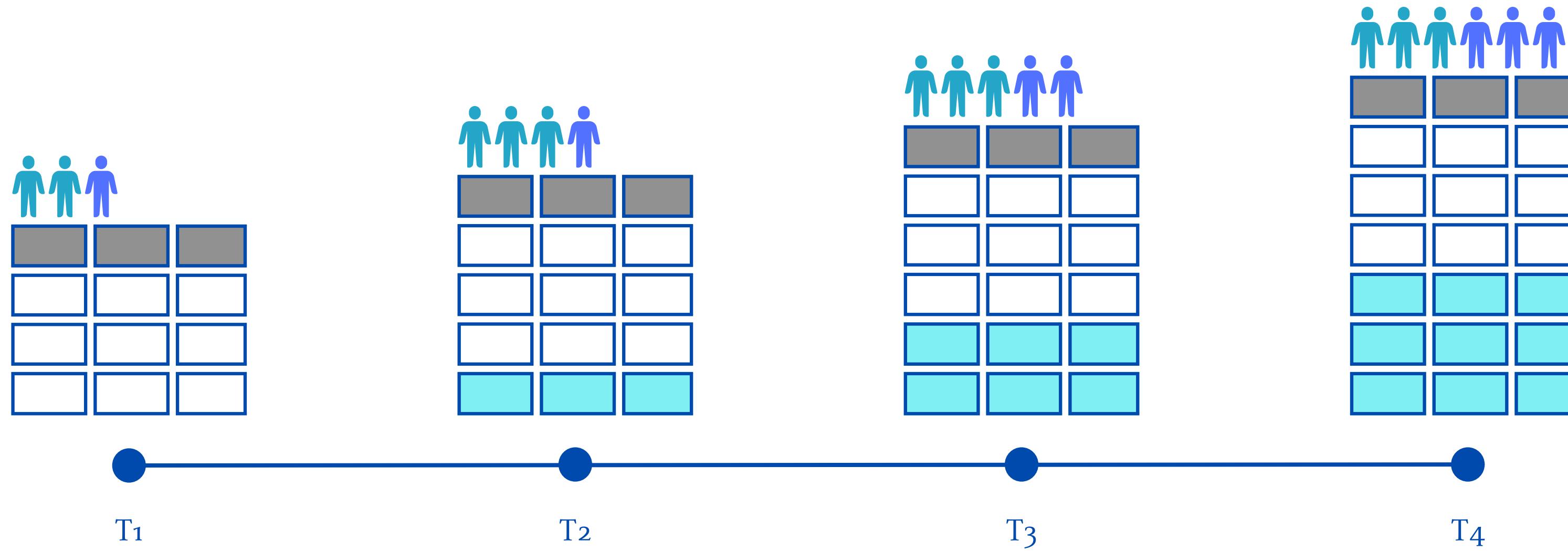


Os painéis podem incluir as **tabelas** usadas para criar os gráficos, seja em seu **estado original ou em versões personalizadas.**
(além de buscas e filtros)



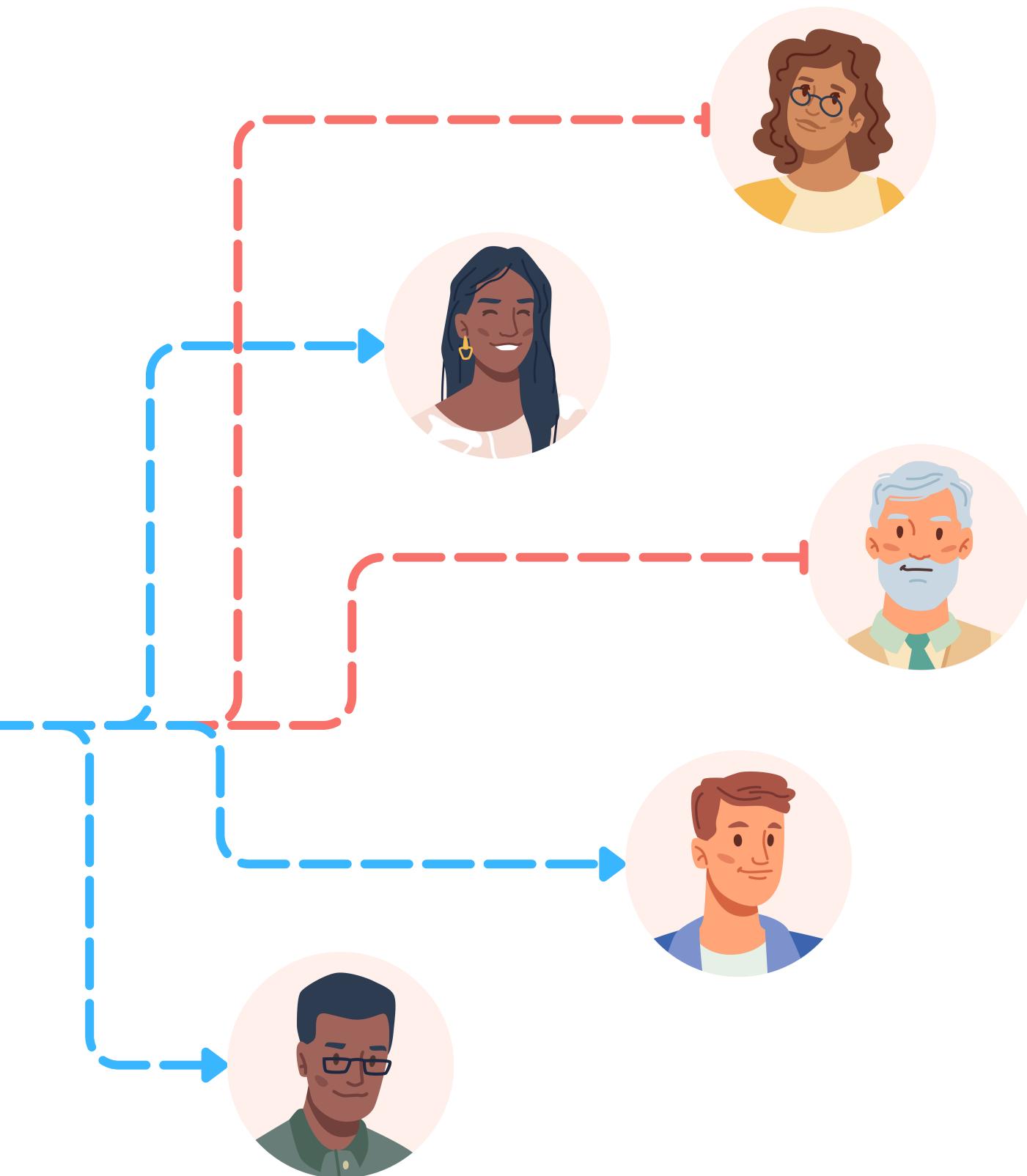
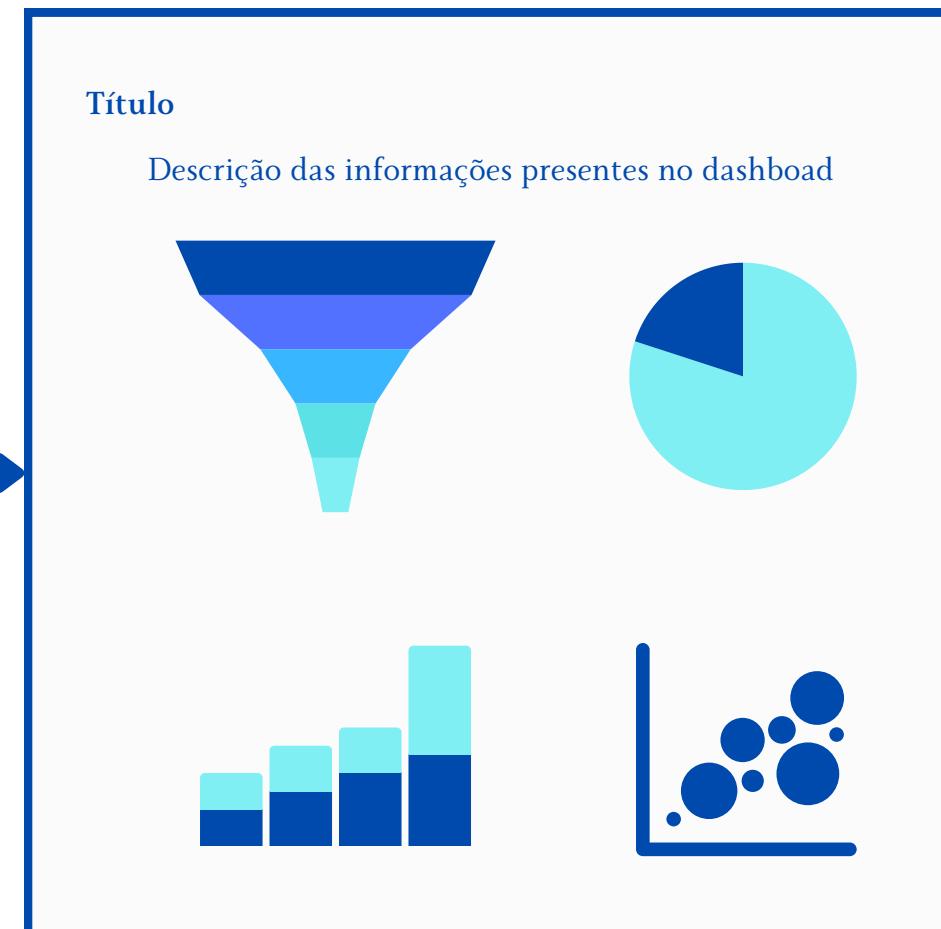
E os painéis podem ser exportados como **relatórios/figuras**.

Os dados das **tabelas e figuras** também podem ser **exportados individualmente**.

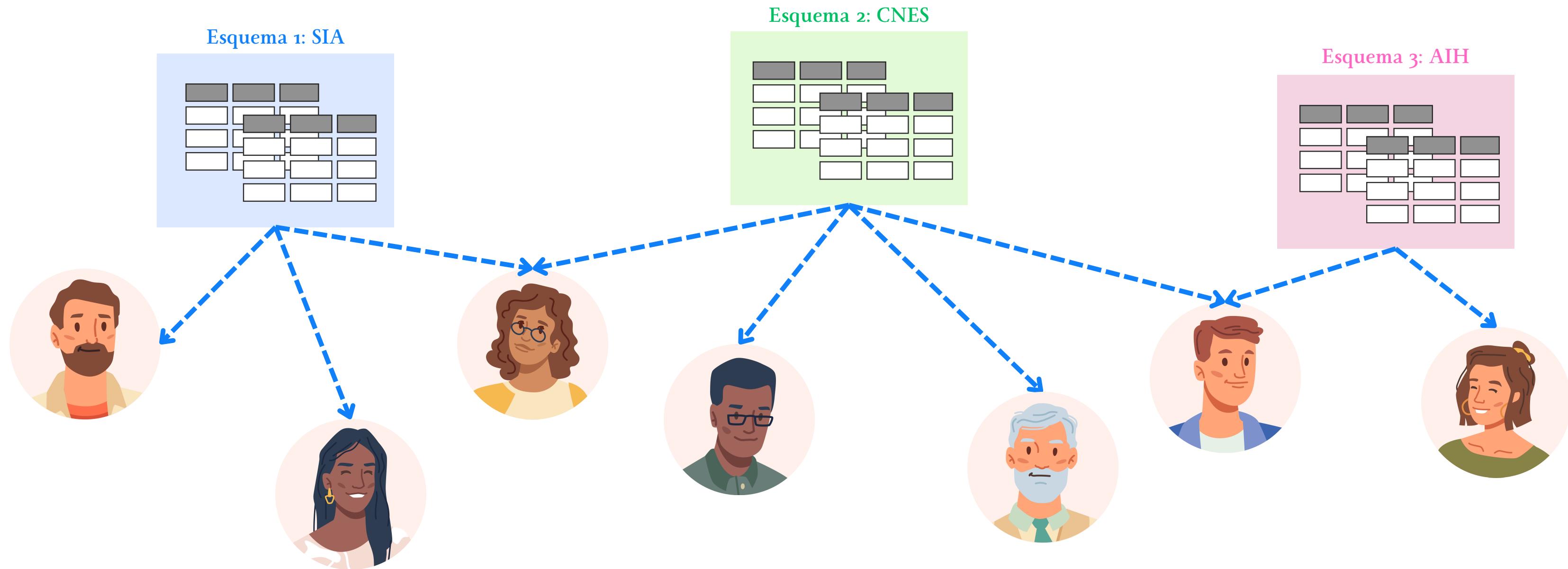


Sempre que as tabelas são atualizadas em sua fonte original, os dados no Apache Superset também recebem atualização.

Os usuários têm a capacidade de **criar seus próprios relatórios interativos e aplicar restrições de acesso.**

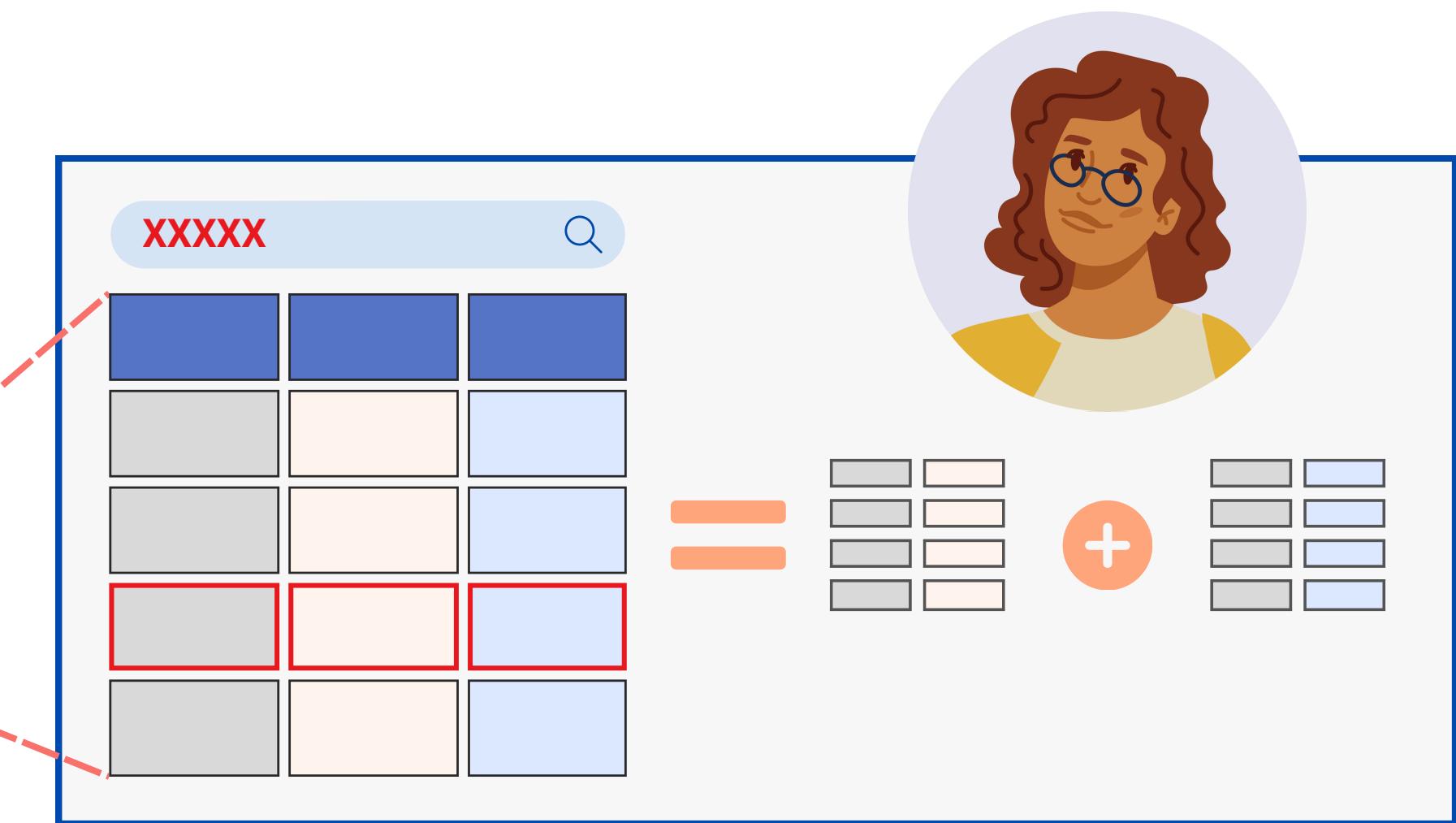


Inclusive, por meio das **restrições de acesso**, é possível **definir níveis de permissões** para diferentes usuários, controlando o acesso a informações sensíveis, desde *schemas* de dados e até a criação de dashboards.

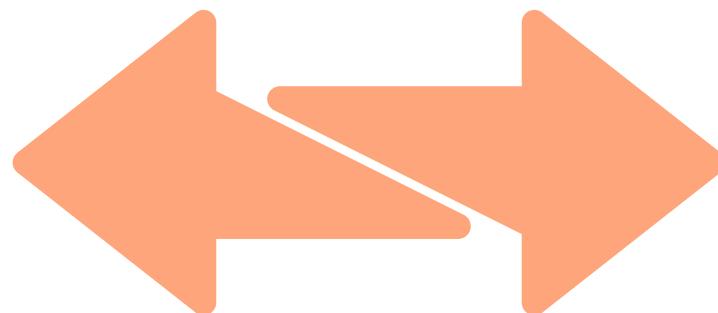
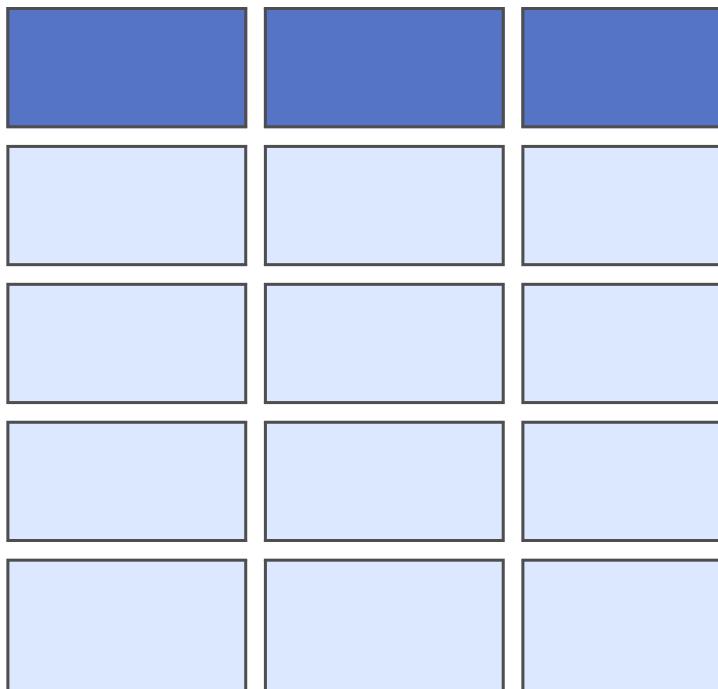




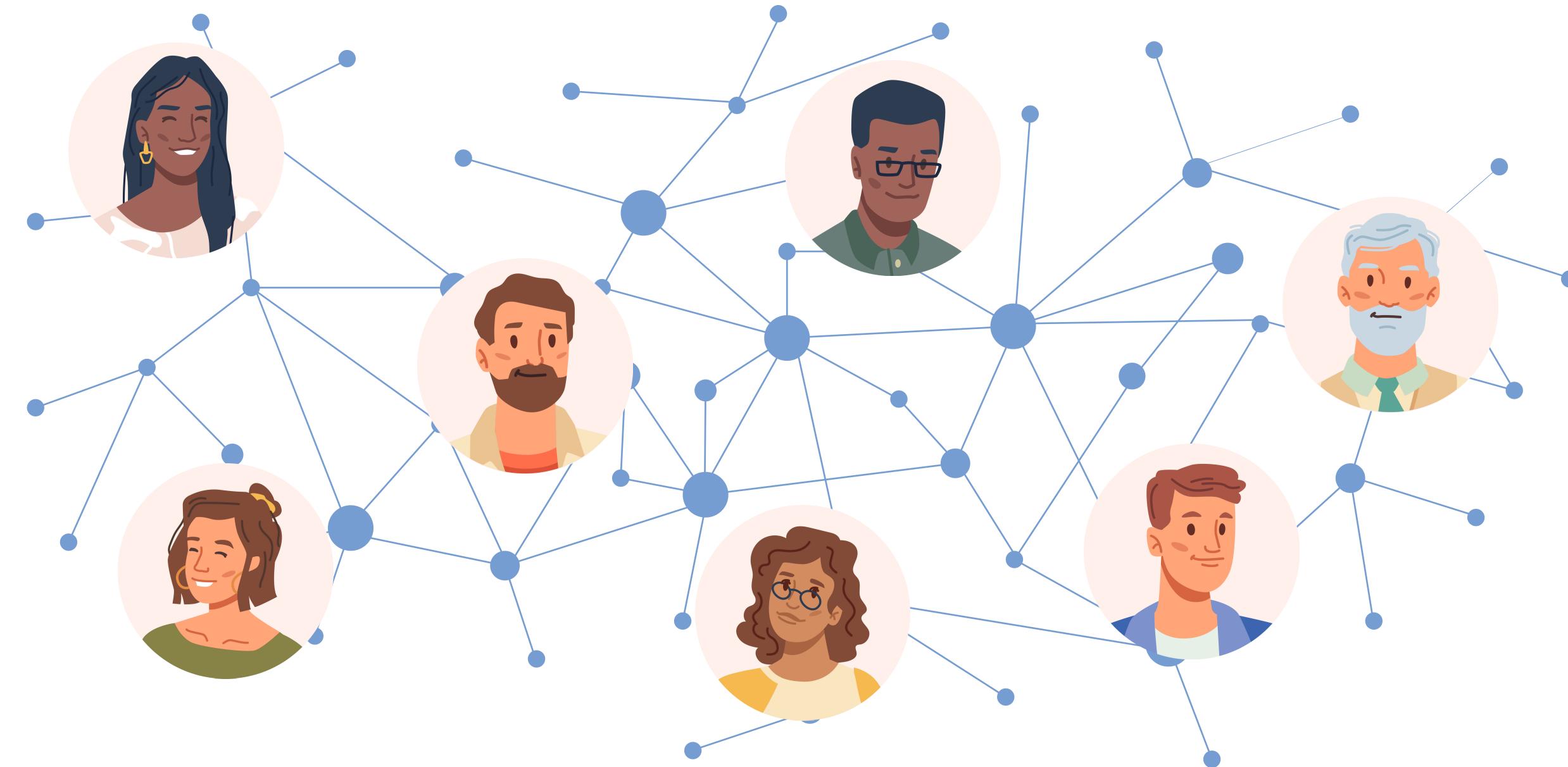
Além de ser um meio para **consultar os dados disponíveis** (consultas/filtros).



O Apache Superset também permite
reproduzir um mesmo processo para outras
fontes de dados.



Em resumo, o Apache Superset capacita os usuários a acelerar a obtenção, consulta, processamento, visualização e compartilhamento de informações da fonte de dados selecionada.

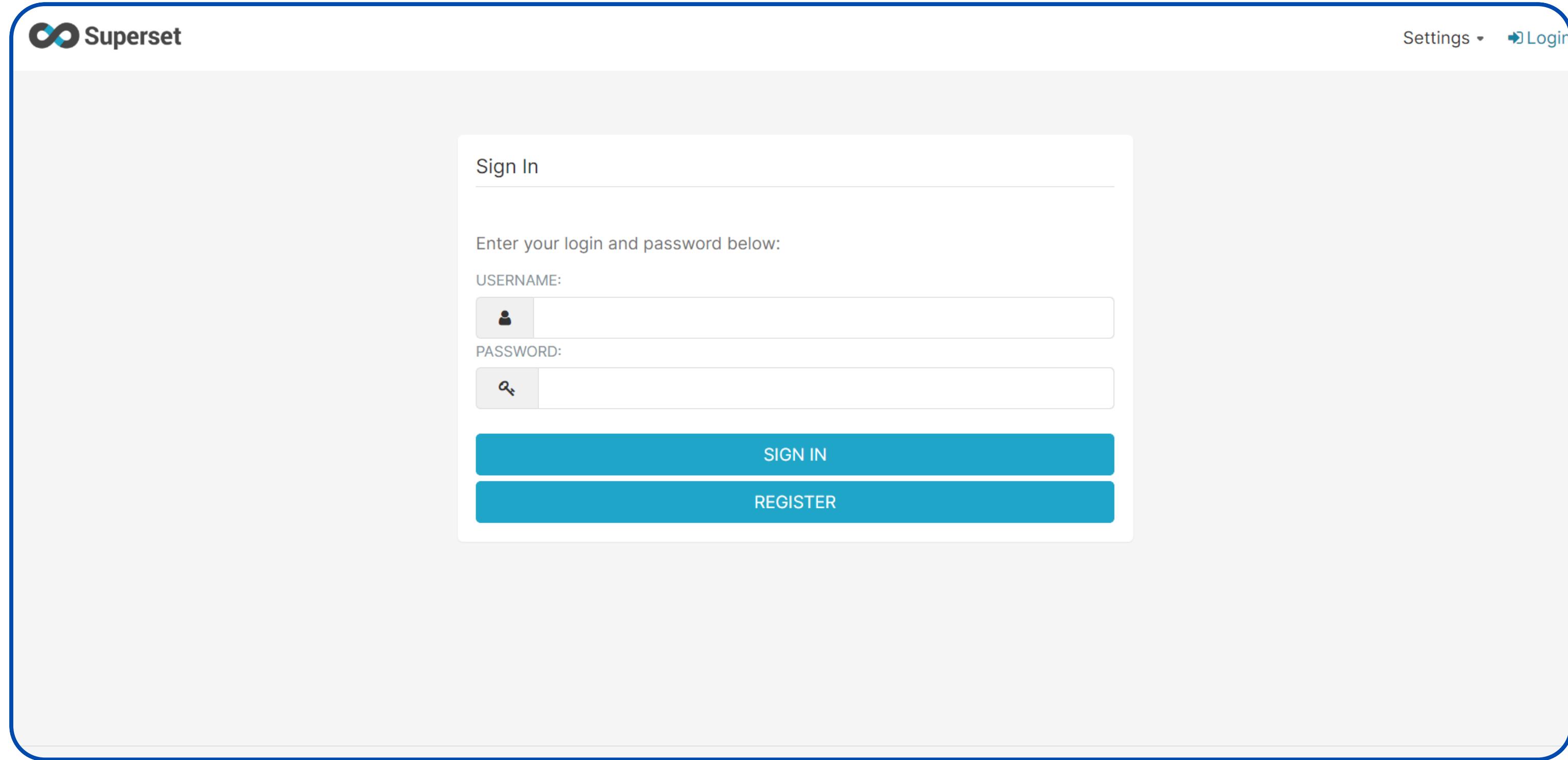




Maxime Beauchemin

- Projeto de **hackathon no Airbnb (2015)** cujo objetivo era a criação de uma ferramenta para a visualização de dados em código aberto.
- Cresceu rapidamente e **superou o Tableau** como principal solução de visualização de dados do Airbnb.
- Em 2016, Superset tornou-se um **projeto de código aberto completo**, incubado pela Apache Software Foundation.
- Empresas como Airbnb, Lyft e Twitter (X) são usuários corporativos do Superset.

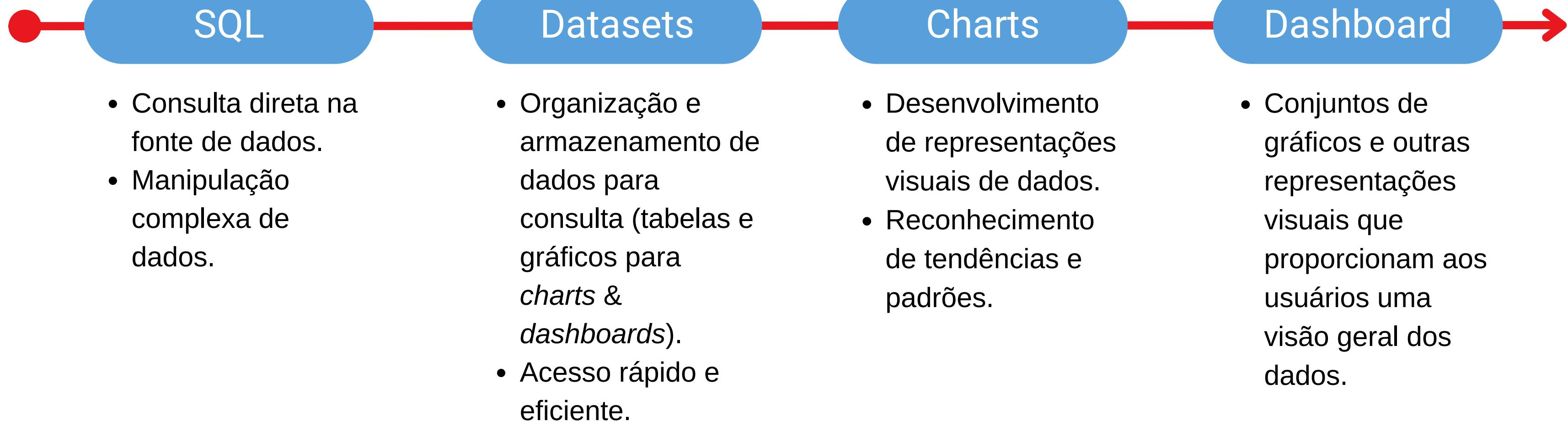
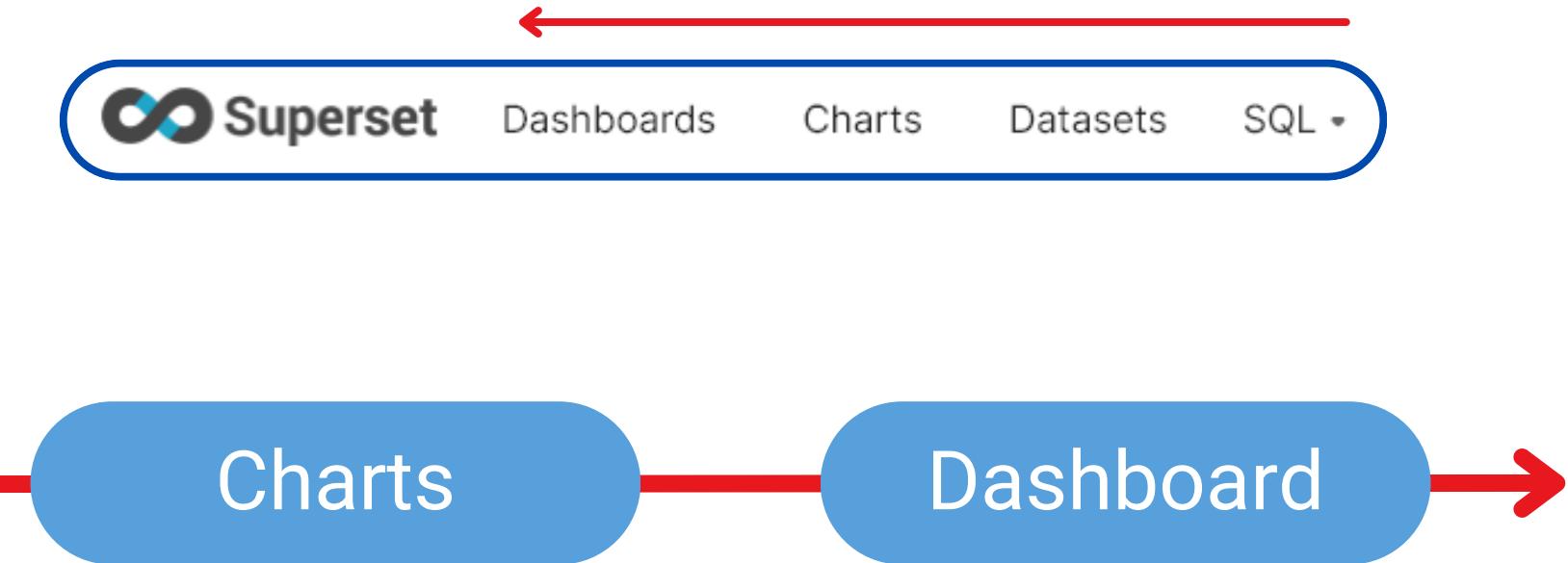
Conhecendo a ferramenta

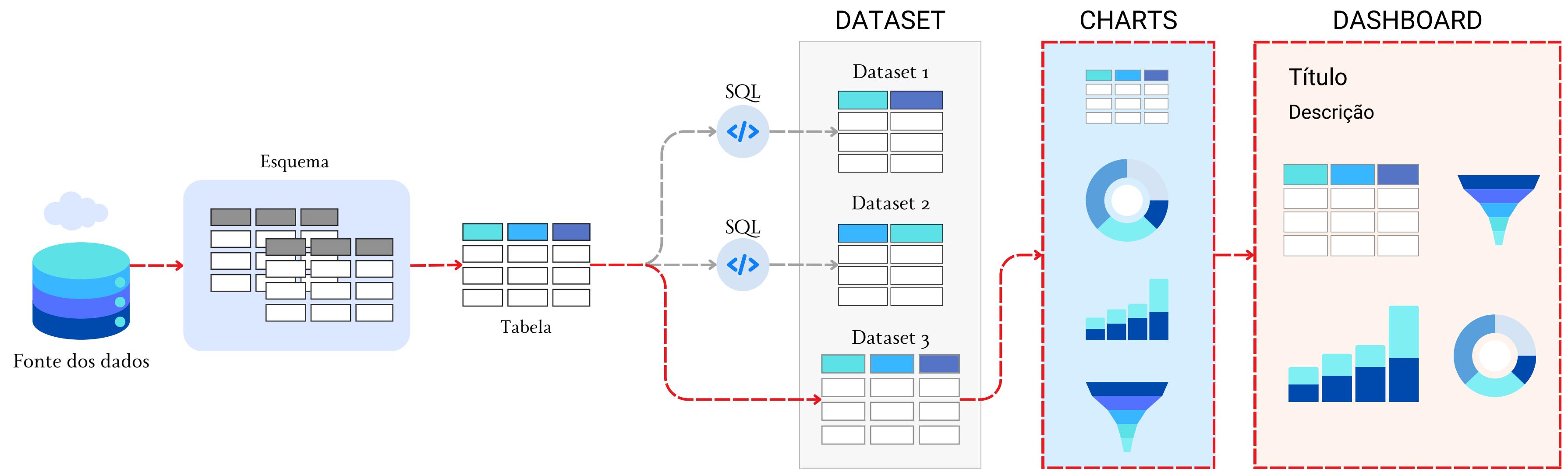


The screenshot displays the Apache Superset web application interface. At the top, there is a navigation bar with the Superset logo, 'Superset' text, and 'Home' link. To the right of the navigation bar are 'Dashboards', 'Charts', 'Datasets', and 'SQL' buttons, with 'Dashboards' being highlighted by a red box. Further to the right are 'INVITE USERS' and 'Settings' buttons. Below the navigation bar, the main content area is titled 'Dashboards'. It features a search bar, filter dropdowns for 'OWNER', 'CREATED BY', 'STATUS', 'FAVORITE', and 'CERTIFIED', and a sort dropdown set to 'Recently modified'. The main area contains eight dashboard cards, each with a preview image, title, modification timestamp, and status (published or draft). A vertical 'Documentation' sidebar is visible on the right side of the page.

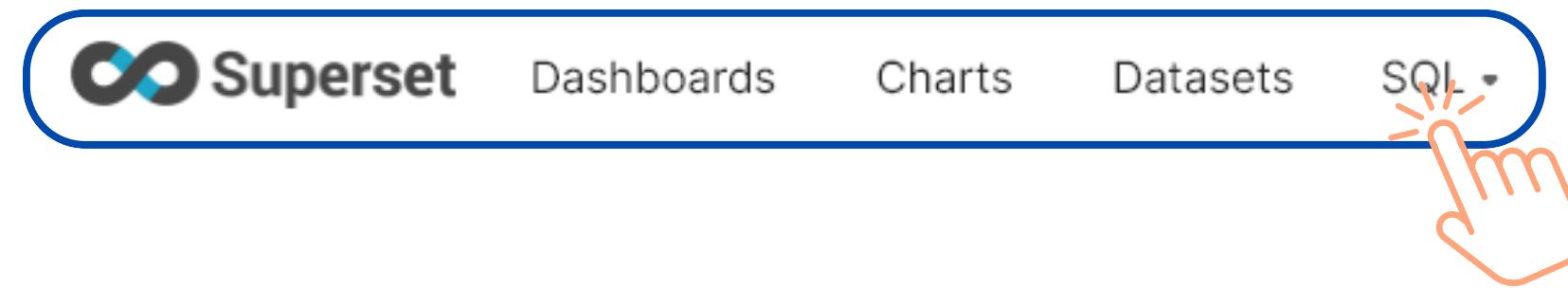
- SaaS Business Growth analysis** (published) - Modified 2 minutes ago
- Marketing Ad Spend [2023]** (published) - Modified a day ago
- Gaming user engagement** (draft) - Modified 4 days ago
- Citi Bike Rides** (published) - Modified 26 days ago
- Sales Dashboard** (published) - Modified 2 months ago
- Slack Dashboard** (published) - Modified 2 months ago
- Video Game Sales** (published) - Modified 2 months ago
- Political Ad Spend** (published) - Modified 3 months ago

No Apache Superset, você verá um **menu** contendo as seguintes ferramentas:





SQL



SQL é uma **linguagem de programação** globalmente utilizada para comunicação e manipulação de bancos de dados.

As linguagens de programação atuam como pontes de **comunicação entre humanos e computadores**, possuindo suas próprias regras e vocabulário.

Elas permitem que os desenvolvedores criem **instruções para o computador executar tarefas**, desde cálculos simples até operações complexas.





```
SELECT coluna1, coluna2 FROM schema.tabela1
```



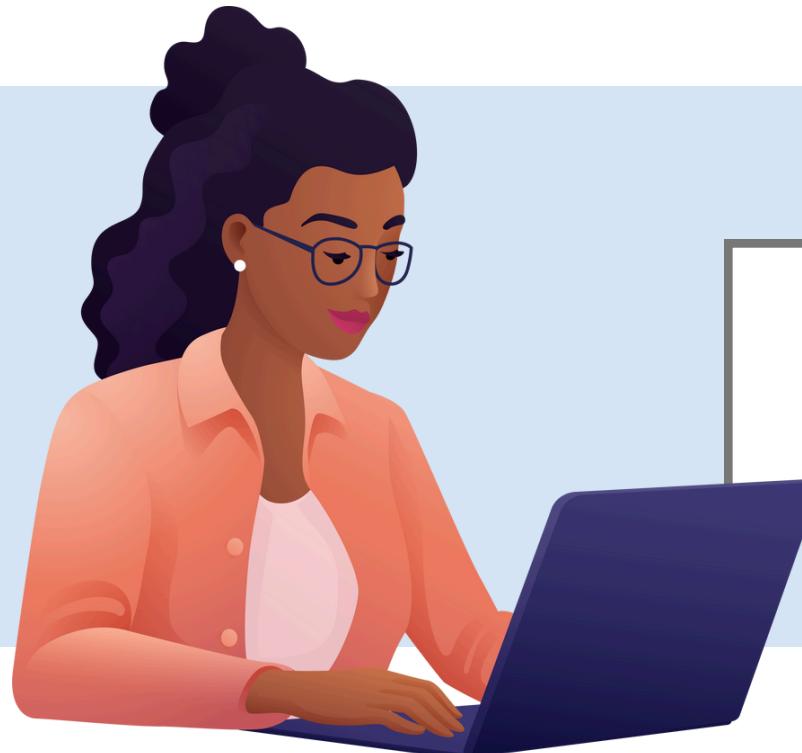
SELECT coluna1, coluna2 **FROM** schema.tabela1

Selezione

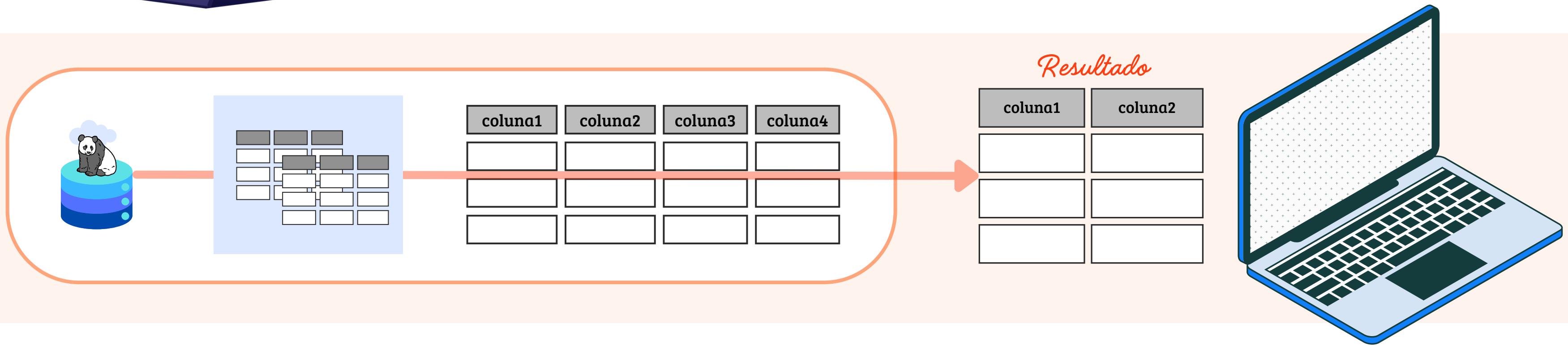
coluna1 e coluna 2

da

tabela 1



SELECT coluna1, coluna2 **FROM** schema.tabela1



The screenshot shows the Apache Superset web application. A black rounded rectangle highlights the central workspace. Six numbered callouts point to specific features:

- [1] A red circle with the number 1 points to the top navigation bar.
- [2] A red circle with the number 2 points to the "SCHEMA" dropdown menu.
- [3] A red circle with the number 3 points to the "SEE TABLE SCHEMA" dropdown menu.
- [4] A red circle with the number 4 points to the SQL query editor area.
- [5] A red circle with the number 5 points to the "RUN" button.
- [6] A red circle with the number 6 points to the results table.

The SQL query in the editor is:

```
1 SELECT
2 *
3 FROM public."FCC 2018 Survey"
4 LIMIT 100
```

The results table displays the following data:

ID	is_software_dev	is_first_dev_job	months_job_search	job_pref
64e4170f7f25253d56019818b7e9fdc0	0	NULL	NULL	freelance
bc030166e4c7f8cb045c28ee98d5a6cd	0	NULL	NULL	work for a medium-sized company
d23d65229c00192bb5a48e73c558a0bb	0	NULL	NULL	work for a multinational corporation
c601014027aa378a84a8e4047cab50e3	0	NULL	NULL	freelance
3a461a8244498a13964071d93ab977ea	1	1	6	NULL

[1] Guia “SQL”; [2] Selecionar um esquema; [3] Selecionar tabela; [4] Query (código SQL) a ser executado; [5] Executar o código; [6] Resultados.

The screenshot shows the Apache Superset web application. In the top navigation bar, the 'SQL' tab is selected. On the left, there's a sidebar for 'DATABASE' (set to 'postgresql examples') and 'SCHEMA' (set to 'public'). Below that is a 'SEE TABLE SCHEMA' section showing tables 'FCC 2018 Survey' and 'bart_lines'. The main area contains a query editor with the following SQL code:

```
1 SELECT
2 *
3 FROM public."FCC 2018 Survey"
4 LIMIT 100
```

Below the query editor are buttons for 'RUN', 'LIMIT: 1000', and a timer showing '00:00:00.22'. To the right of the run button are 'SAVE', 'COPY LINK', and a three-dot menu. A red circle with the number '7' highlights the 'SAVE' button. The results panel below shows the query again and a table of data with 100 rows. The table has columns: ID, is_software_dev, is_first_dev_job, months_job_search, and job_pref. The first few rows are:| ID | is_software_dev | is_first_dev_job | months_job_search | job_pref |
| --- | --- | --- | --- | --- |
| 64e4170f7f25253d56019818b7e9fdc0 | 0 | NULL | NULL | freelance |
| bc030166e4c7f8cb045c28ee98d5a6cd | 0 | NULL | NULL | work for a medium-sized company |
| d23d65229c00192bb5a48e73c558a0bb | 0 | NULL | NULL | work for a multinational corporation |
| c601014027aa378a84a8e4047cab50e3 | 0 | NULL | NULL | freelance |
| 3a461a8244498a13964071d93ab977ea | 1 | 1 | 6 | NULL |

[7] Salvar como um Dataset; [8] Criar gráfico.

Mas eu preciso aprender SQL para usar o Superset?

NÃO!

A plataforma oferece uma **interface amigável** que permite a visualização e manipulação de dados originados a partir de datasets originados das tabelas existentes nos *schemas* de banco de dados.

```
SELECT *  
FROM tabela1  
JOIN tabela2  
ON tabela1.Código = tabela2.Código
```

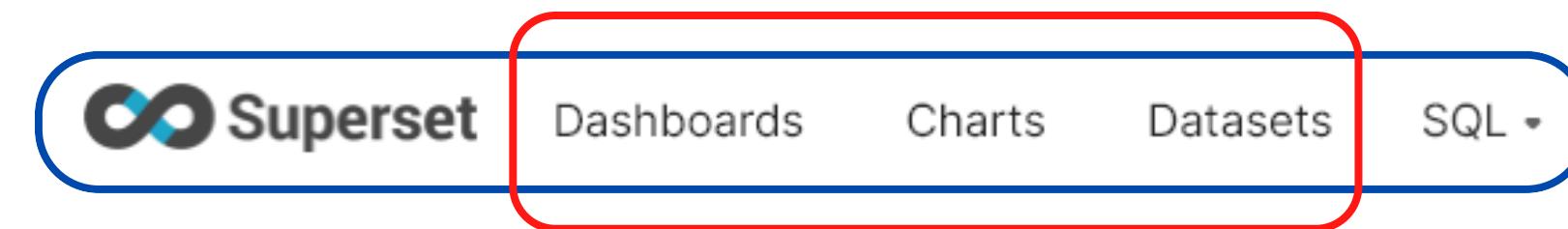
The diagram illustrates a JOIN operation between two tables.
TABELA 1 has columns: Grupo (blue), Código (pink), and Estado (blue). Data rows: (1, A, PR), (1, B, ES), (2, C, PA).
TABELA 2 has columns: Código (pink), CID10 (green), and CNES (green). Data rows: (A, C44, 5897), (B, C18, 6987), (C, C20, 7525).
A red dashed arrow points from the Código column of TABELA 1 to the Código column of TABELA 2, indicating the join condition.

TABELA 1	Código	Estado	TABELA 2	Código	CID10	CNES
1	A	PR		A	C44	5897
1	B	ES		B	C18	6987
2	C	PA		C	C20	7525

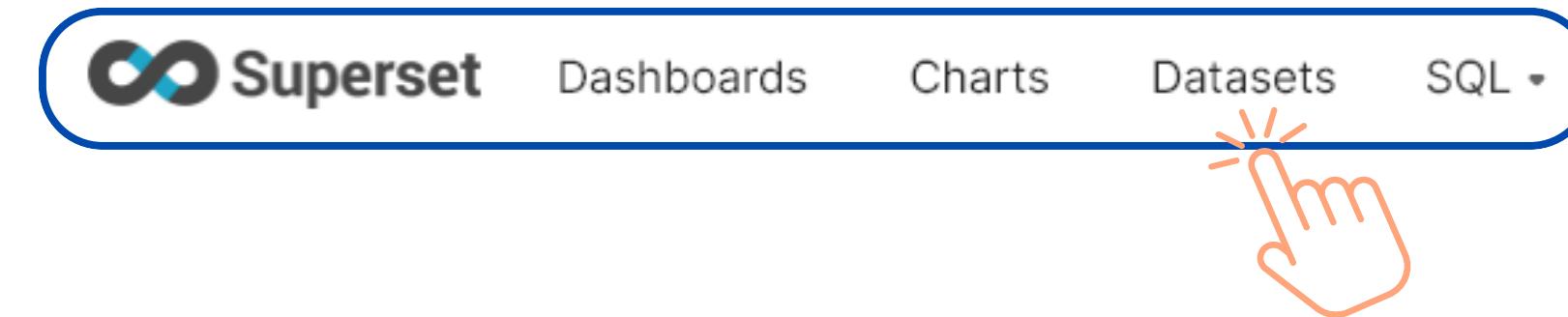
TABELA 1 + TABELA 2

Grupo	Código	Estado	CID10	CNES
1	A	PR	C44	5897
1	B	ES	C18	6987
2	C	PA	C20	7525

OUTROS



DATASETS

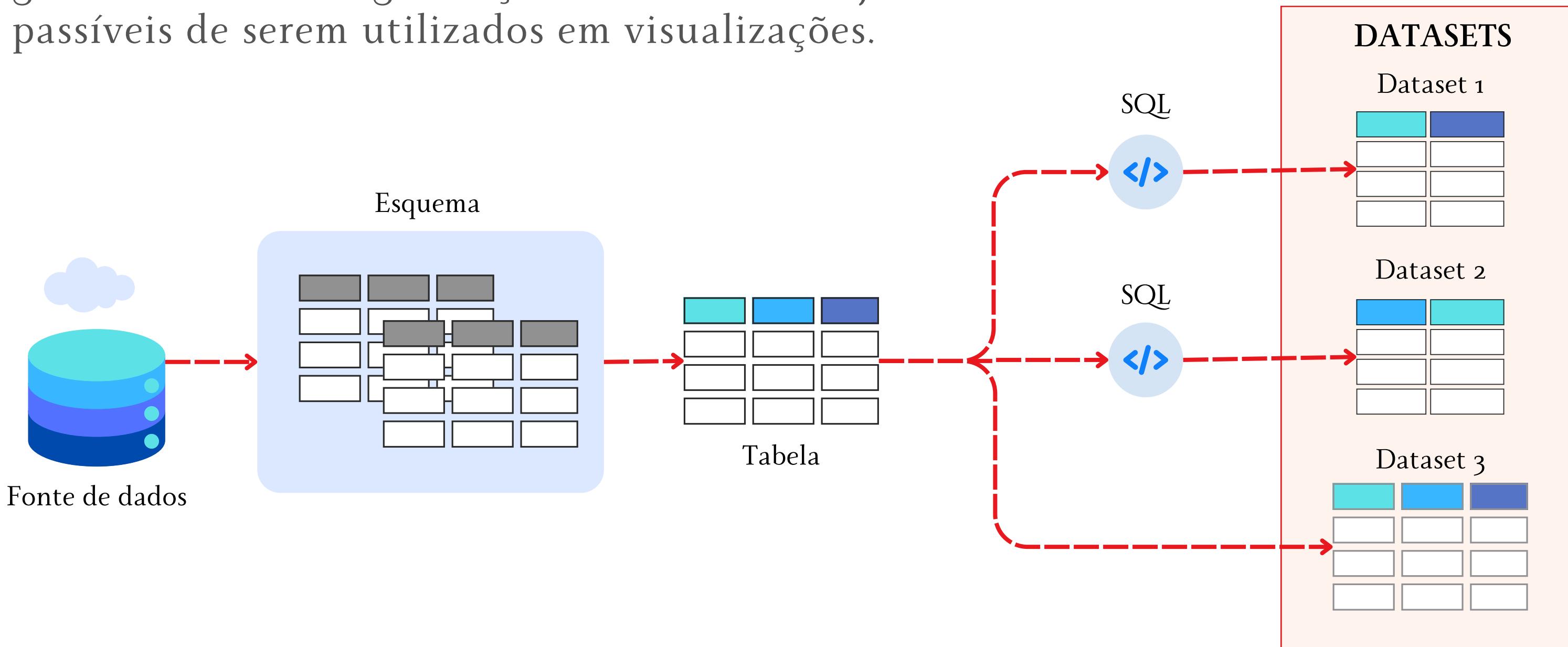


The screenshot shows the Apache Superset interface for managing datasets. A red circle labeled '1' points to the 'Datasets' tab in the top navigation bar. Another red circle labeled '2' points to the '+ DATASET' button in the top right corner. A third red circle labeled '3' points to the download icon (a downward arrow) in the same area. A fourth red circle labeled '4' points to the first dataset entry in the list below.

Name	Type	Database	Schema	Owners	Last modified	Actions
messages	Physical	examples	public		3 months ago	
FCC 2018 Survey	Physical	examples	public		3 months ago	
channels	Physical	examples	public		3 months ago	
unicode_test	Physical	examples	public		3 months ago	
threads	Physical	examples	public		3 months ago	
users_channels	Physical	examples	public		3 months ago	
users	Physical	examples	public		3 months ago	
video_game_sales	Physical	examples	public		3 months ago	
exported_stats	Physical	examples	public		3 months ago	
messages_channels	Virtual	examples	public		3 months ago	

[1] Guia “Datasets”; [2] Adicionar dataset de um esquema; [3] Importar dataset de um CSV;
[4] Lista com todos os datasets e detalhes;

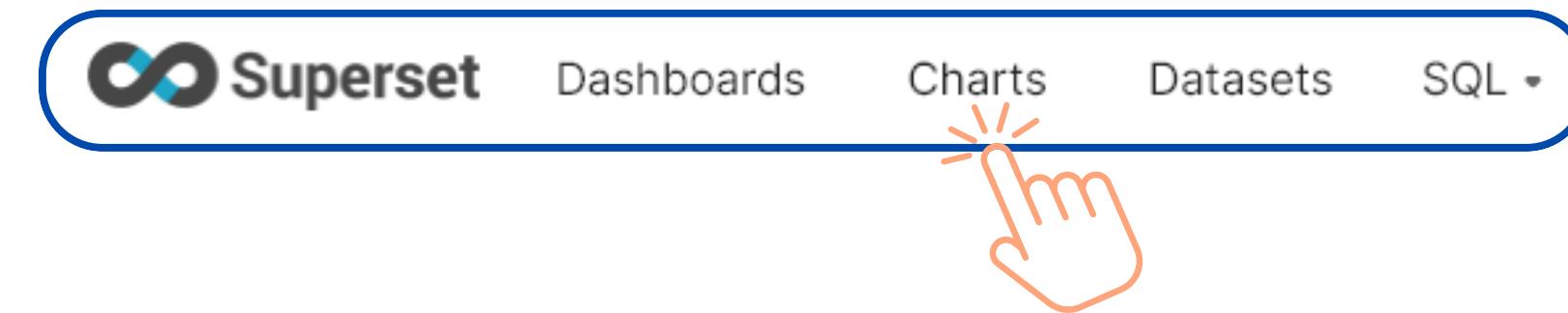
A seção "datasets" no Apache Superset é um **hub central** para o gerenciamento e organização de todos os conjuntos de dados passíveis de serem utilizados em visualizações.



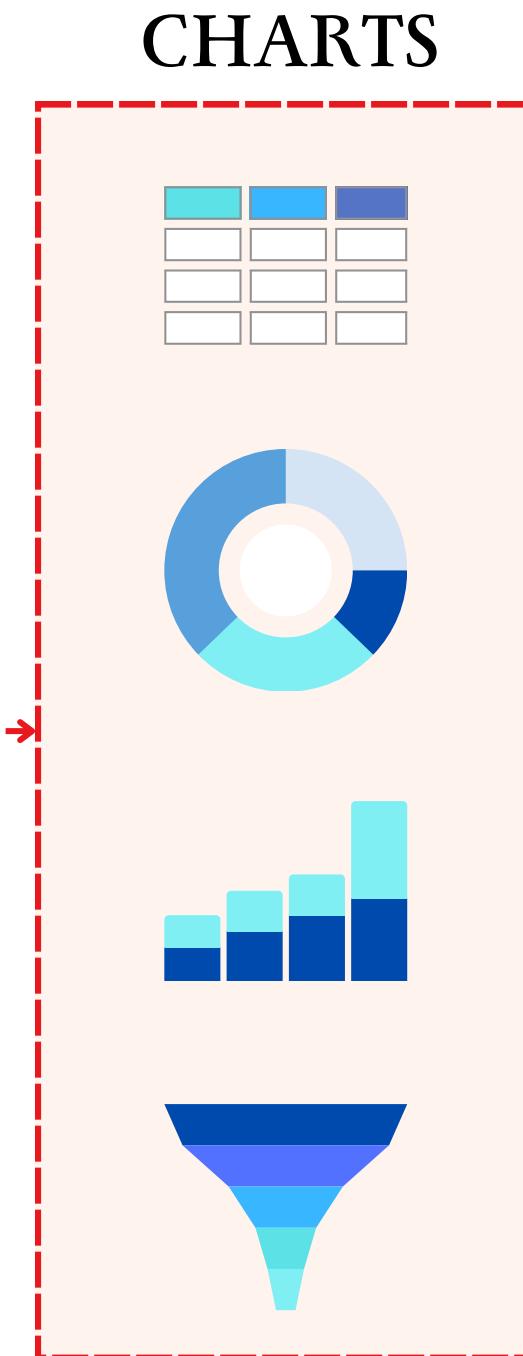
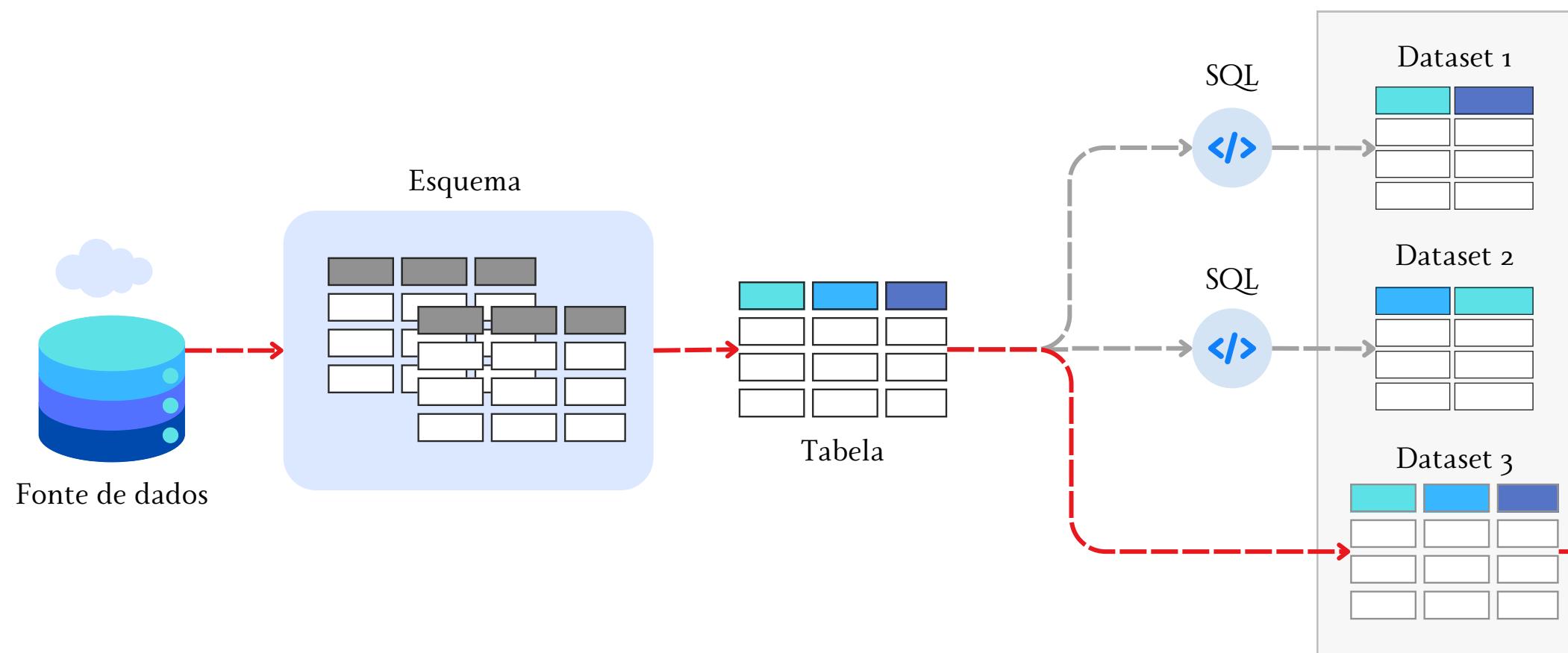
Além de adicionar conjuntos de dados, a seção "datasets" permite **editar conjuntos existentes**, alterando fonte, campos, tipos de dados e metadados para preparar dados para análise e otimizar visualizações.

The screenshot shows the Apache Superset interface for editing datasets. A modal window titled 'Edit Dataset FCC 2018 Survey' is open. At the top, a warning message reads: 'Be careful. Changing these settings will affect all charts using this dataset, including charts owned by other people.' Below this, tabs for 'SOURCE', 'METRICS', 'COLUMNS', 'CALCULATED COLUMNS', and 'SETTINGS' are visible, with 'SOURCE' being the active tab. Under 'SOURCE', there is a note: 'Click the lock to make changes.' followed by radio buttons for 'Physical (table or view)' (selected) and 'Virtual (SQL)'. The 'Physical' section includes fields for 'DATABASE' (set to 'postgresql examples'), 'SCHEMA' (set to 'public'), and 'TABLE' (set to 'FCC 2018 Survey'). At the bottom right of the modal are 'CANCEL' and 'SAVE' buttons. In the background, the main 'Datasets' page lists various datasets like 'messages', 'FCC 2018 Survey', 'channels', etc., with columns for NAME, TYPE, MODIFIED BY, and last modified time. The 'FCC 2018 Survey' entry is highlighted.

CHARTS



Na área "Charts" do Apache Superset
são criadas **visualizações de dados** a
partir dos datasets.



Create a new chart

1 Choose a dataset
tabela_com_dados_para_dataviz

2 Choose chart type

All charts

Recommended tags: Popular

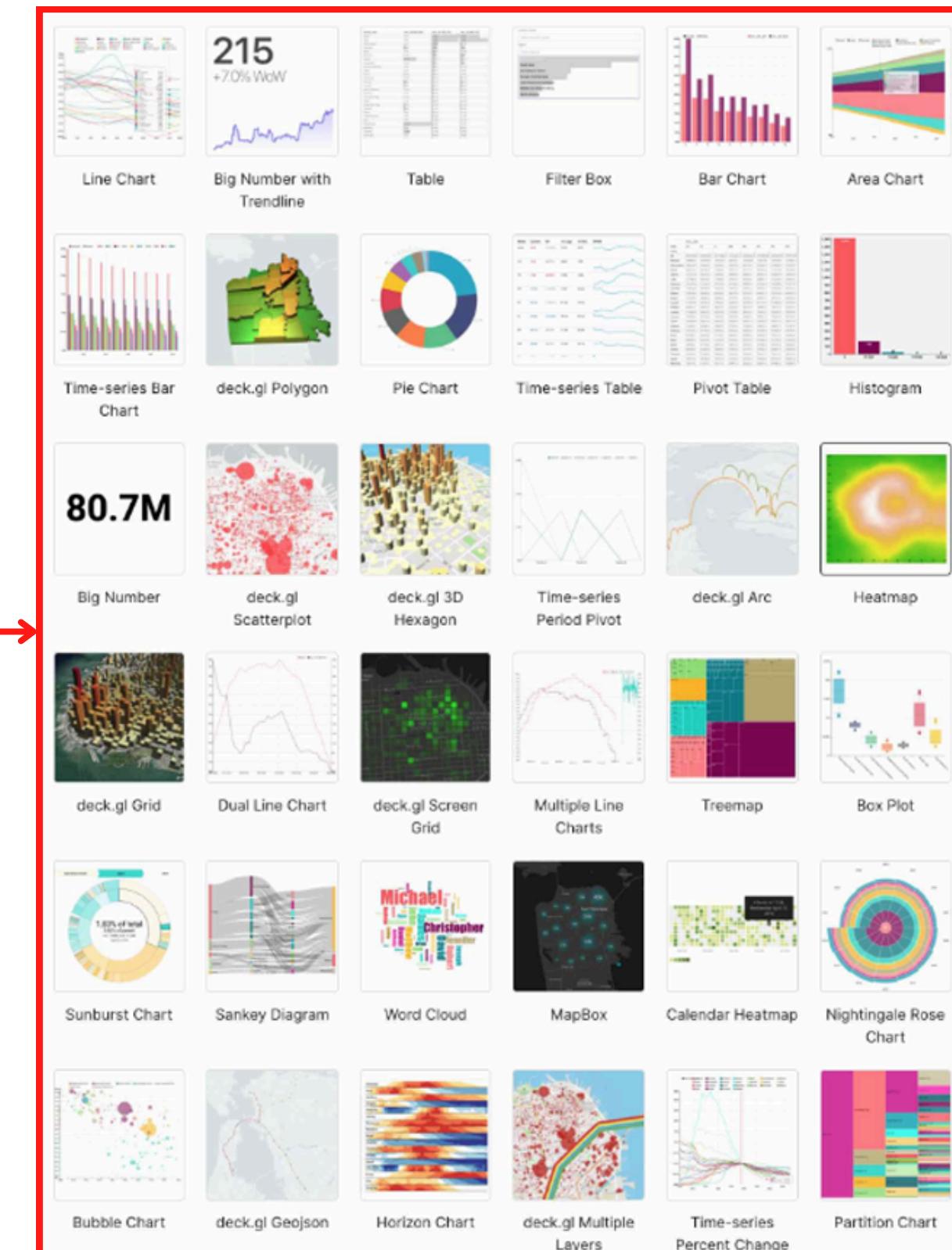
Category: Correlation, Distribution, Evolution, Flow, KPI, Map, Part of a Whole, Ranking, Table, Tools, Other

Tags: 2D, 3D, Additive, Aesthetic

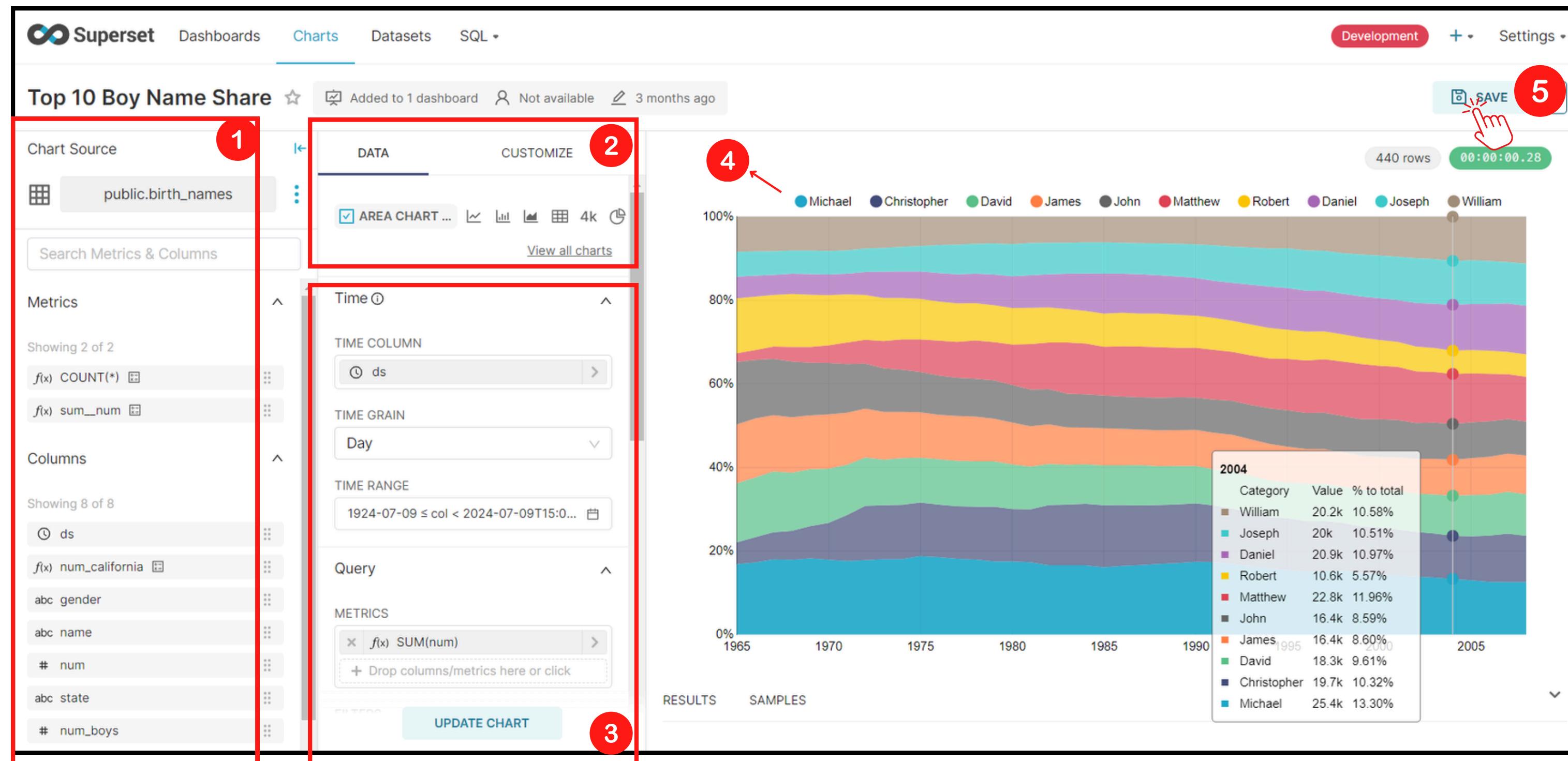
Search all charts: 215, 80.7M, Big Number with Trendline, Big Number, Table, Pivot Table, Time-series Line Chart, Time-series Area Chart, Time-series Bar Chart, Time-series Scatter Plot, Pie Chart, Bar Chart, World Map

Please select both a Dataset and a Chart type to proceed

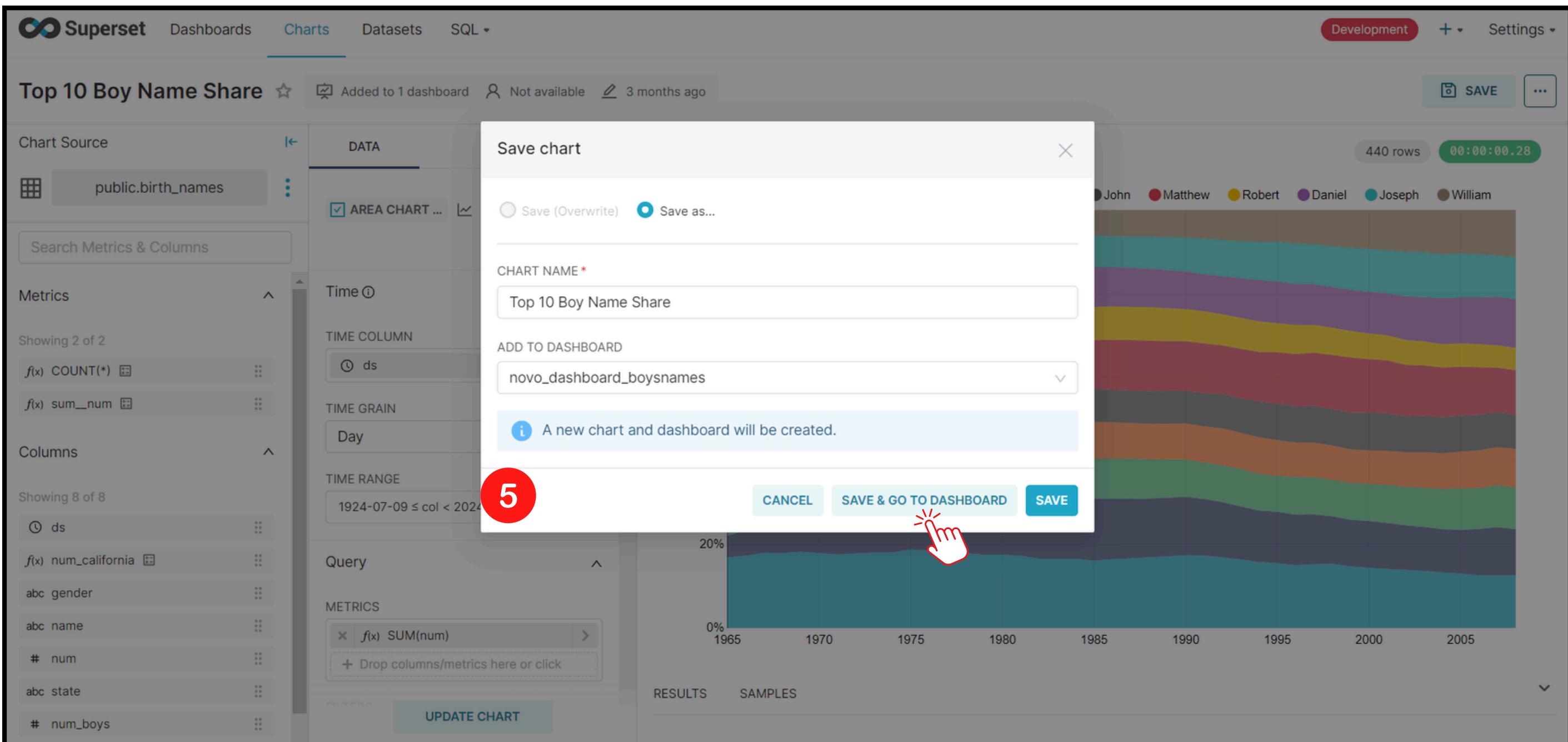
CREATE NEW CHART



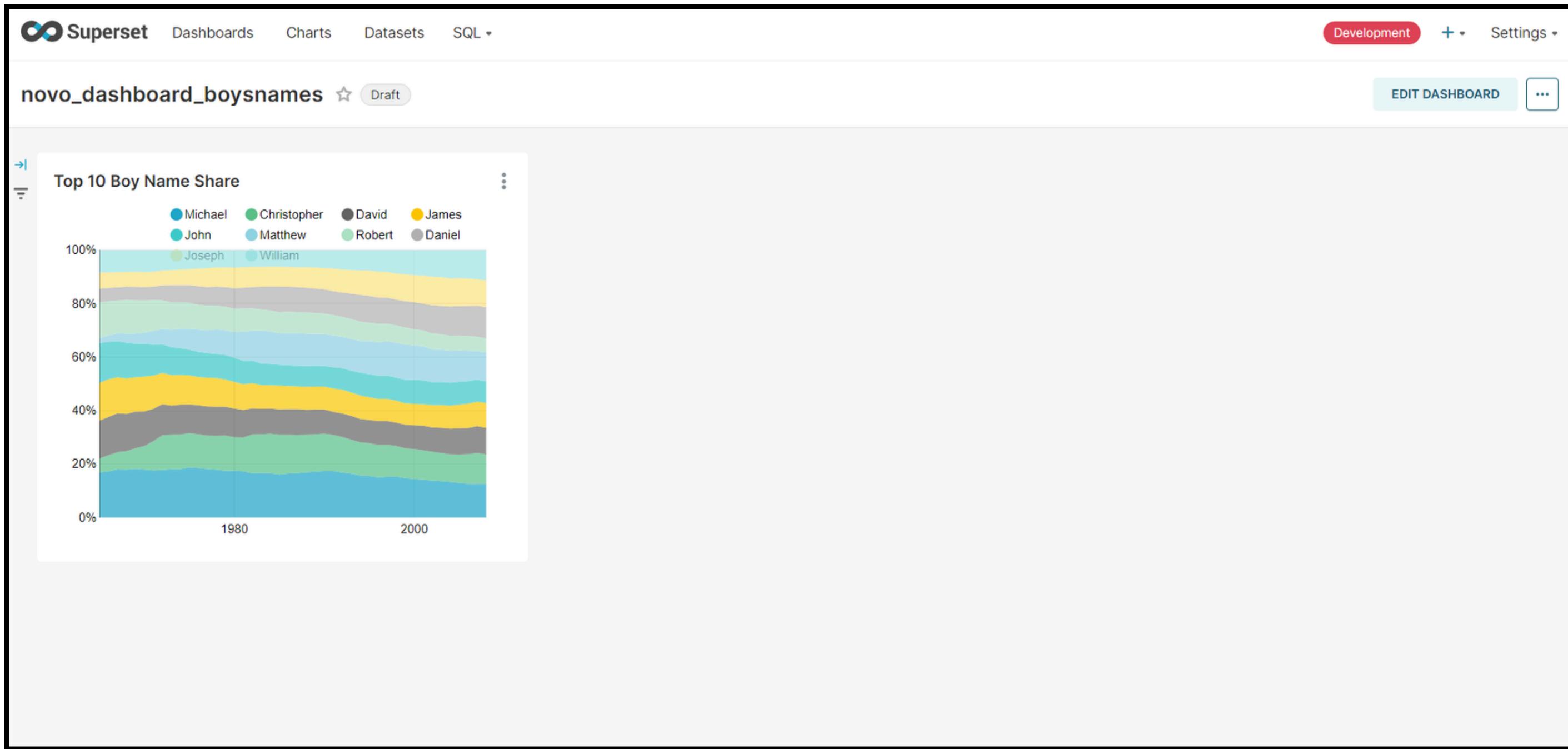
[1] Fonte de dados (dataset selecionado); [2] Seleciona o tipo de visualização; [3] Criar o chart.



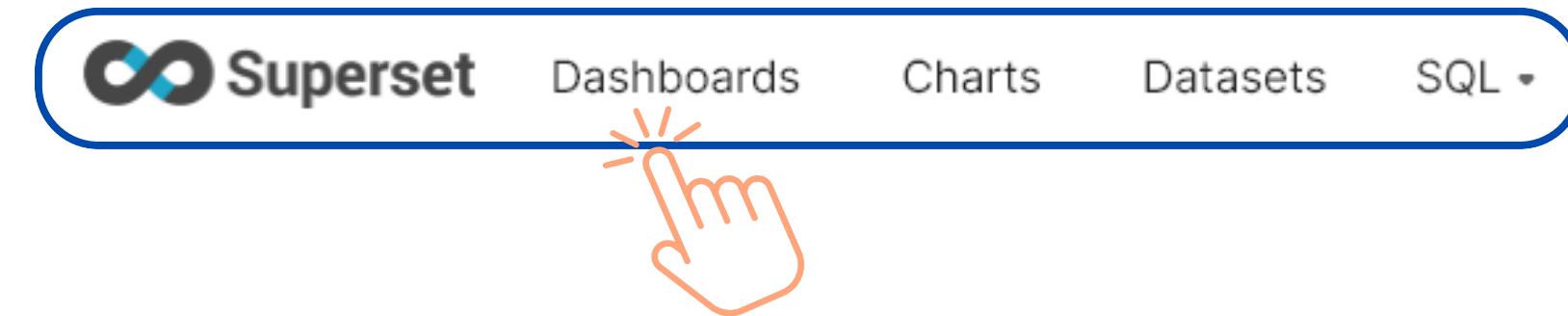
[1] Fonte de dados (dataset selecionado); [2] Tipo de visualização escolhido; [3] Colunas/informações para a composição do gráfico; [4] Resultado; [5] Salvar o chart em um dashboard existente ou novo.



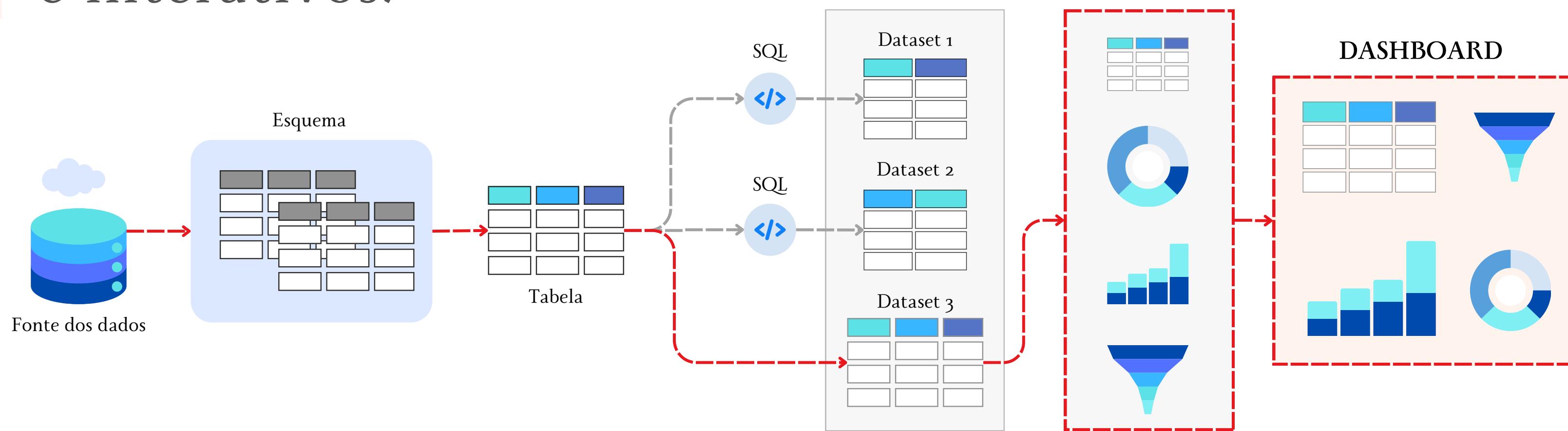
[1] Fonte de dados (dataset selecionado); [2] Tipo de visualização escolhido; [3] Colunas/informações para a composição do gráfico; [4] Resultado; [5] Salvar o chart em um dashboard existente ou novo.



DASHBOARDS

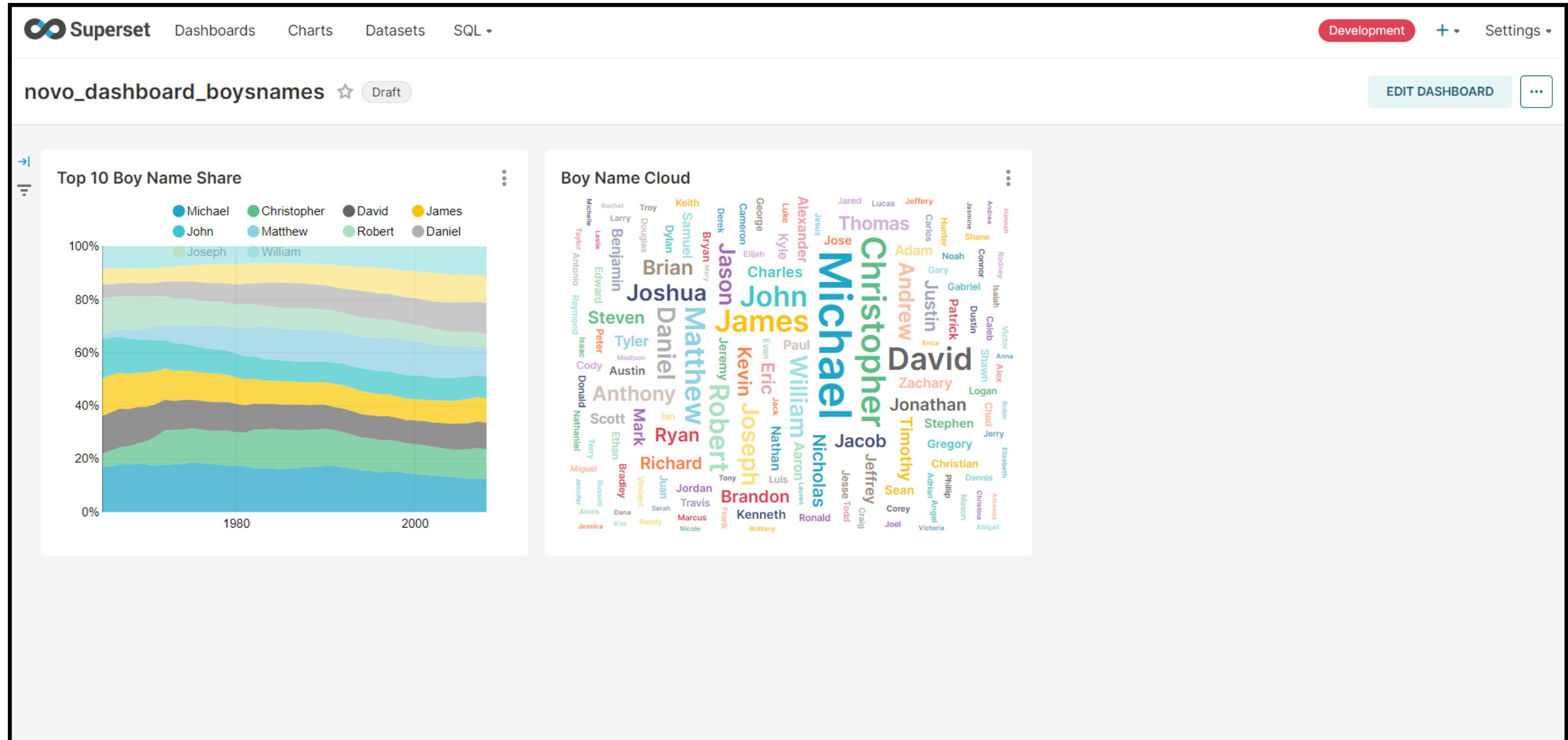


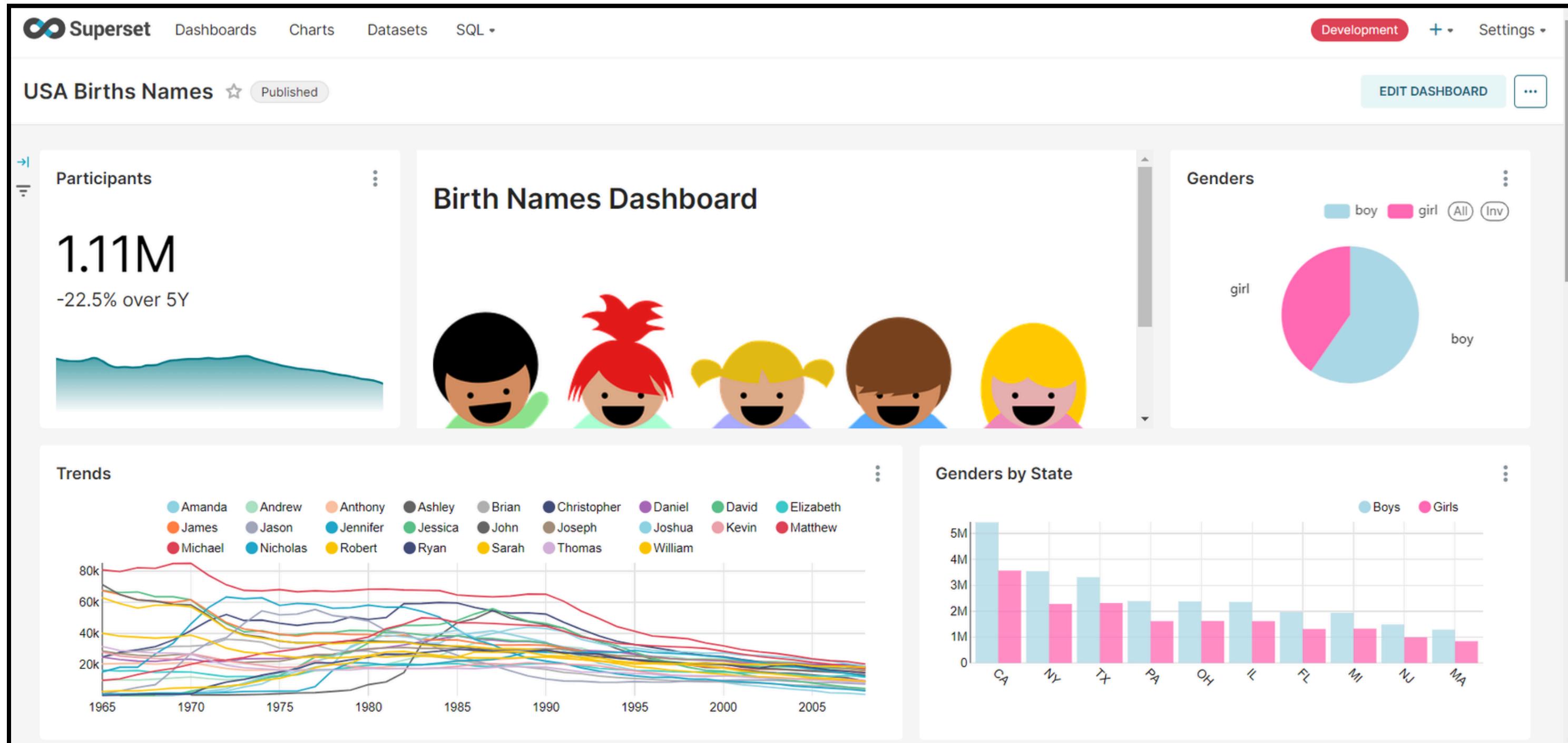
Na área "Dashboards" do Apache Superset unimos as visualizações para criar os relatórios dinâmicos e interativos.



The screenshot shows the Apache Superset dashboard editor interface. On the left, there is a chart titled "Top 10 Boy Name Share" which is an area chart showing the share of boy names over time from 1980 to 2000. To the right of the chart is a card for a "Boy Name Cloud". The card displays the following information: Viz type (Word Cloud), Dataset (public.birth_names), and Last modified (3 months ago). A red circle with the number "1" is overlaid on the top-left corner of the chart area. A red arrow points from the bottom-right corner of the "Boy Name Cloud" card towards the "+ CREATE NEW CHART" button in the sidebar.

[1] Depois de clicar em "editar", você poderá adicionar facilmente outros gráficos ao seu painel interativo, simplesmente clicando e arrastando a caixa com o gráfico selecionado.





Deseja obter mais informações
sobre o Apache Superset?

[Documentação
oficial do SUPERSET](#)

[Conheça o
SUPERSET](#)

[PRESET
Do mesmo criador
do SUPERSET](#)

obrigada

marilia.melo.favalesso@gmail.com

<https://www.linkedin.com/in/mariliafavalesso/>