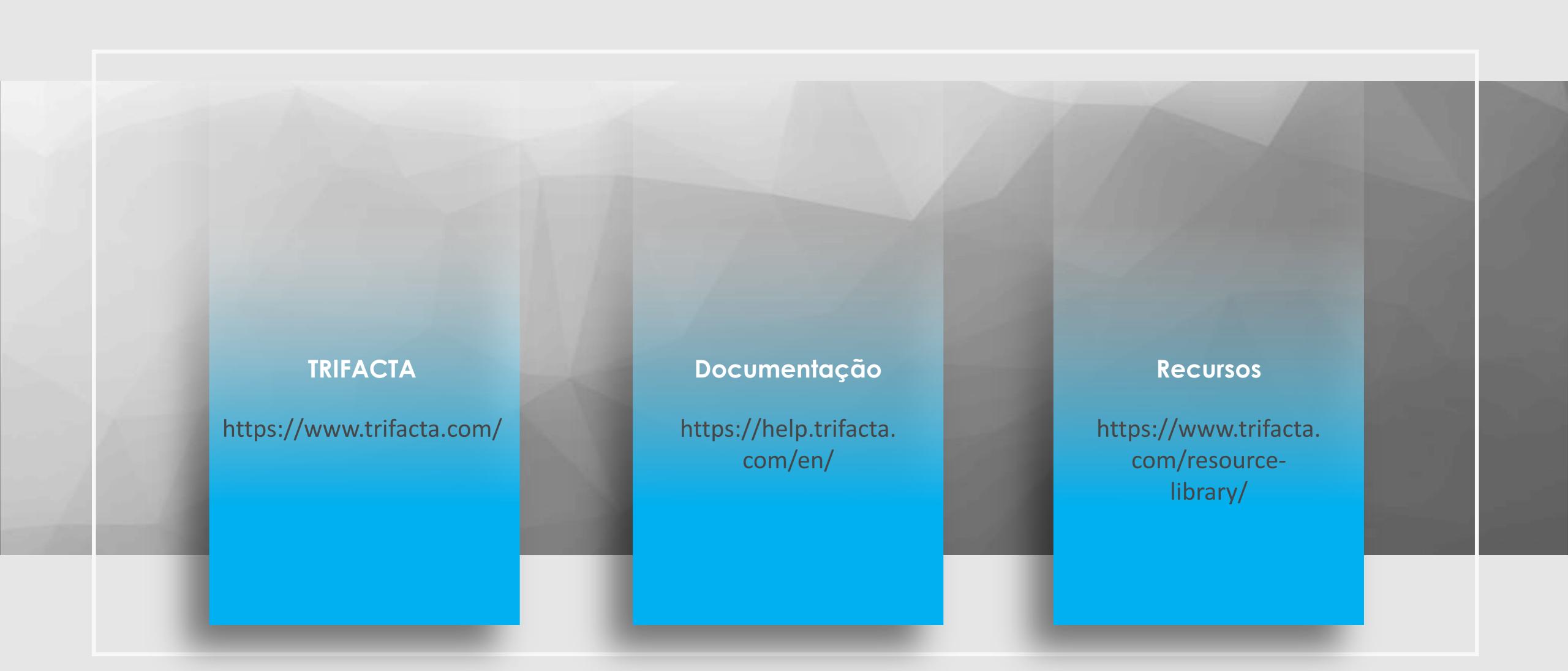
TRIFACTA **Data Preparation** Grimaldo Oliveira



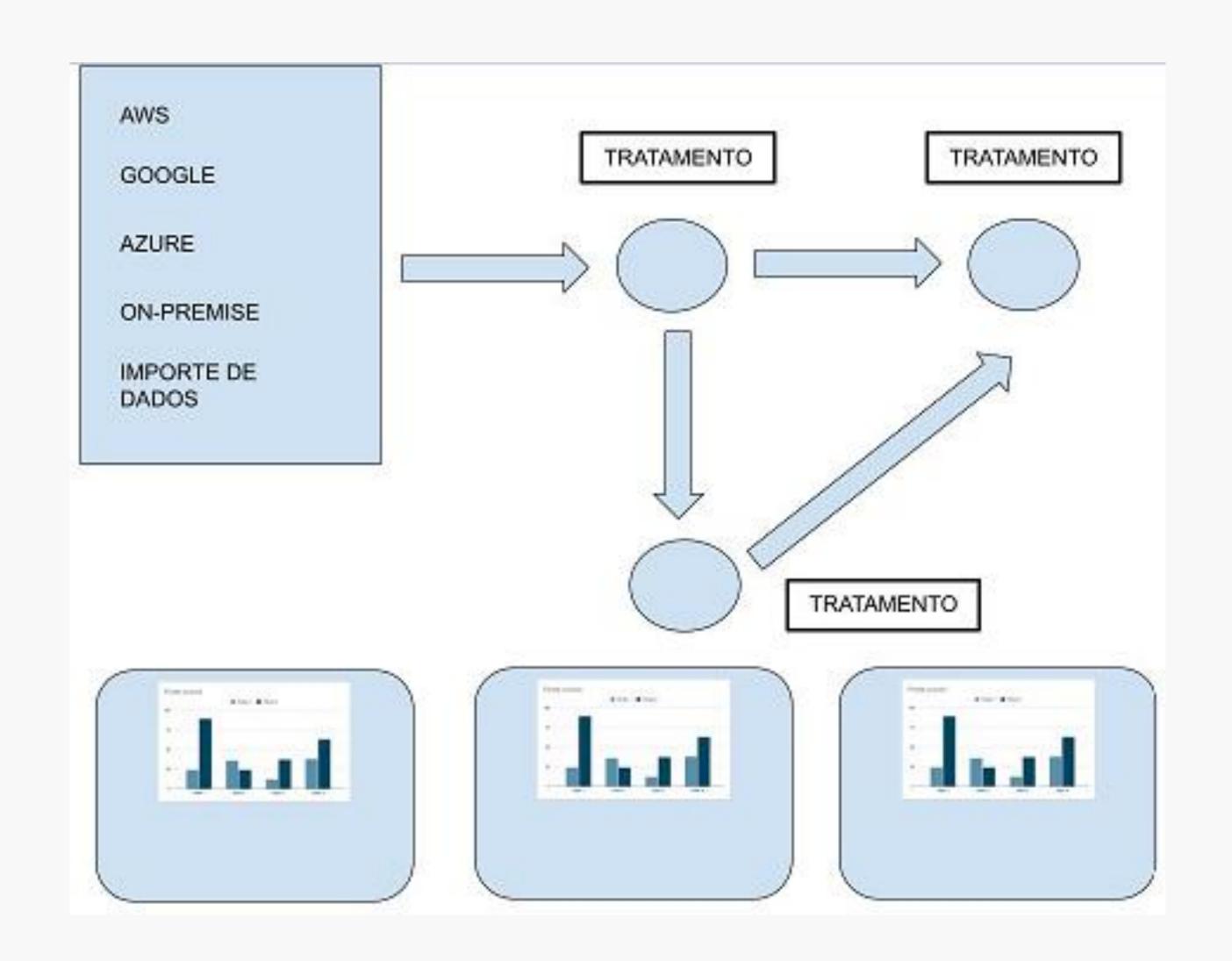
- 1. É uma plataforma de preparação de dados aberta que pode se conectar a diversas fontes de dados, permitindo a integração em todas as principais plataformas de dados em nuvem.
- 2. Você pode escolher entre ETL ou ELT, ou uma combinação ideal dos dois com base no desempenho.
- 3. Suporte para todas as principais nuvens, Google, AWS, Azure e on-premise.
- 4. Possui uma interface intuitiva e simples utilizando objetos.
- 5. utiliza perfil de dados, ajudando na identificação de outliers, tratamento de dados, criação de novos campos, dentre outras tarefas.

Sites Importantes



Arquitetura

Como funciona a estrutura do TRIFACTA.



TRIFACTA

Uma plataforma que permite o tratamento e ajustes de dados, para carga em arquivos ou bases de dados.

Trifacta

Uma ferramenta para tratamento de dados, que permite de forma visual, analisar campo a campo possíveis problemas que estão carregados nas bases de dados.

Tem por característica:

Tratamento de Dados: Permite que você leia bases de dados por meio de conectores, trazendo todas as informações com dados exploratórias, métricas quantitativas.

Ajuste de dados: Permite a criação de novos campos, de acordo com as regras que você determinar, como eliminação de dados nulos, inconsistências, separação de dados, etc.

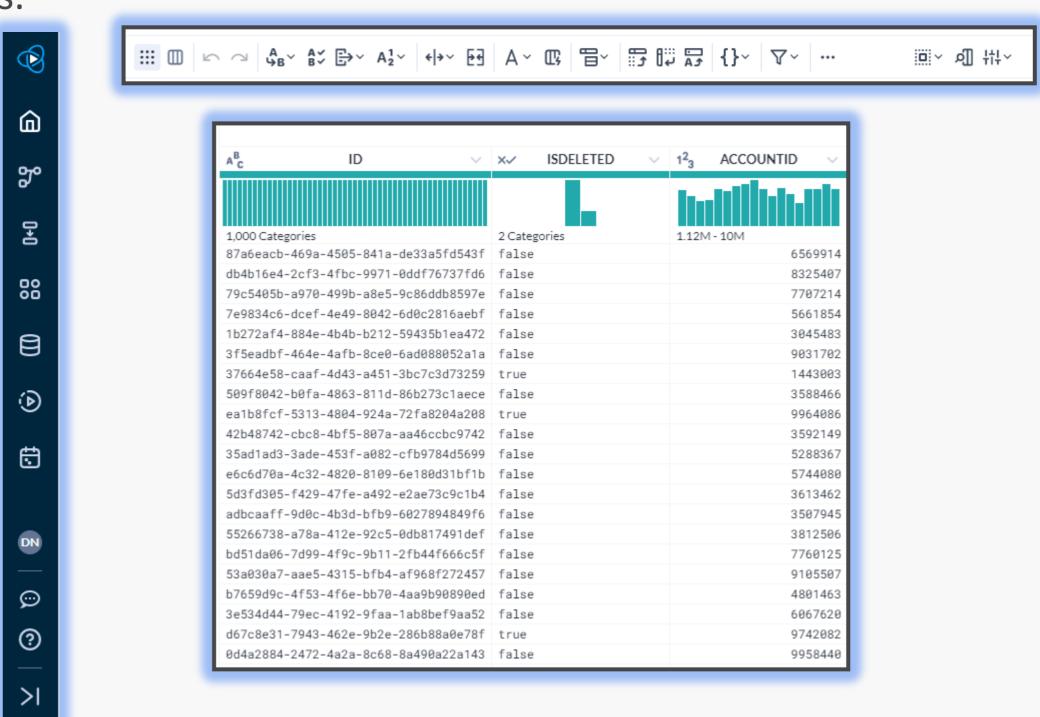
Flow: Cada mudança executada com o Trifacta, é registrada por meio de um flow, que permite ao analista encadear suas ações de tratamento, ao final você pode executar o fluxo de dados e gerar uma nova saída.

TRIFACTA

É local de trabalho dentro do TRIFACTA.

INTERFACE

Um espaço de trabalho, um ambiente para acessar todos os componentes da ferramenta. O espaço de trabalho organiza objetos, e fornece acesso a dados e recursos computacionais, visuais e de leitura dos dados.



Conta com uma gama muito rica de botões para diversas tarefas que aprenderemos no treinamento.

TRIFACTA

É local de trabalho dentro do TRIFACTA.

Detalhando

Vamos conhecer alguns componentes que fazem parte da área de trabalho do Trifacta, utilizaremos durante nosso treinamento.















- Home Retorna a página inicial do Trifacta.
- Flow Apresenta os flow dos trabalhos criados.
- Plan Onde criamos os planos de trabalho com os flows.
- Library Onde estão os arquivos que subimos para o Trifacta.
- Connections Onde podemos conectar a diversas fontes de dados.
- Job History Histórico dos Jobs executados.
- Schedules A programação dos Jobs ou Planos.

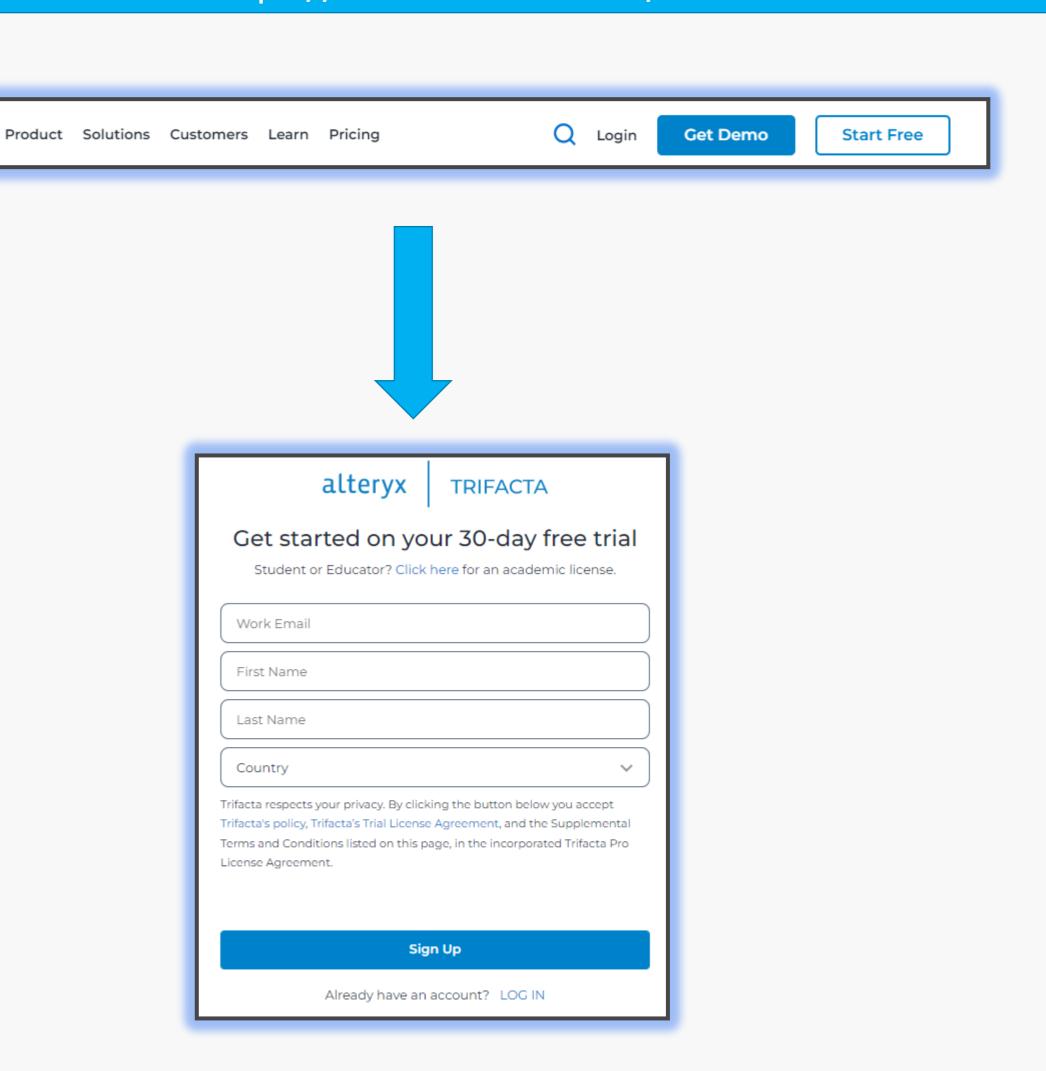
Criação do usuário

Vamos nos preparar criar um usuário para acesso na nuvem da Trifacta.

Criação do usuário

TRIFACTA

https://www.trifacta.com/

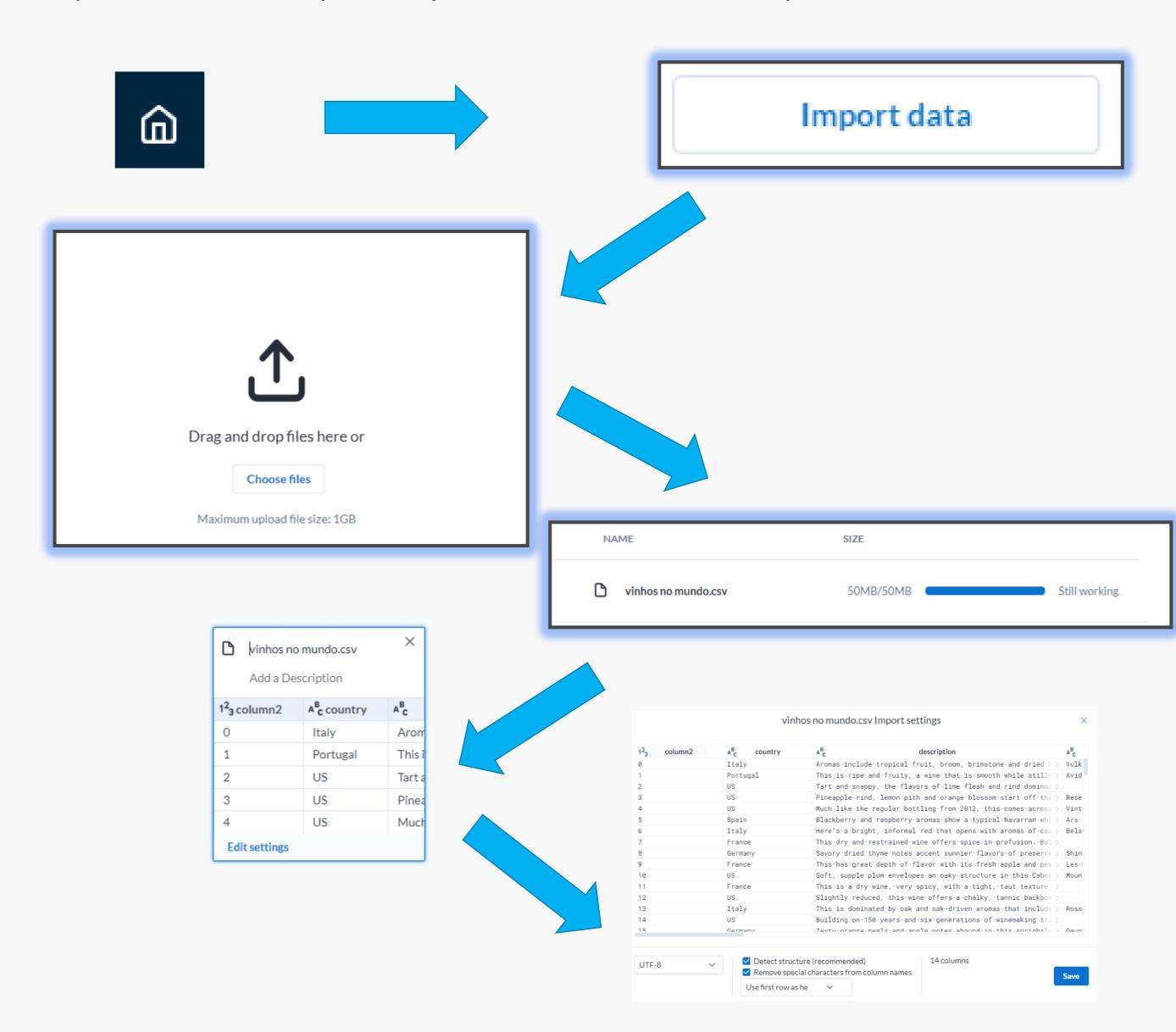


Começando a trabalhar

Vamos carregar os primeiros dados.

Carregando dados

Os primeiros dados para importarmos os dados. Arquivo: vinhos no mundo.csv



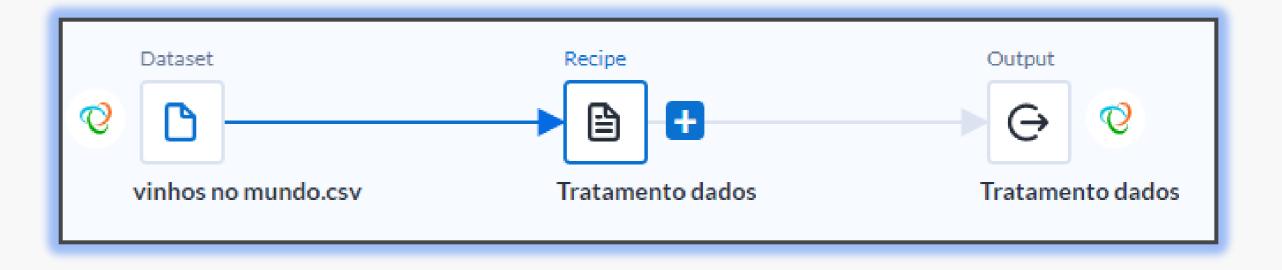
Começando a trabalhar

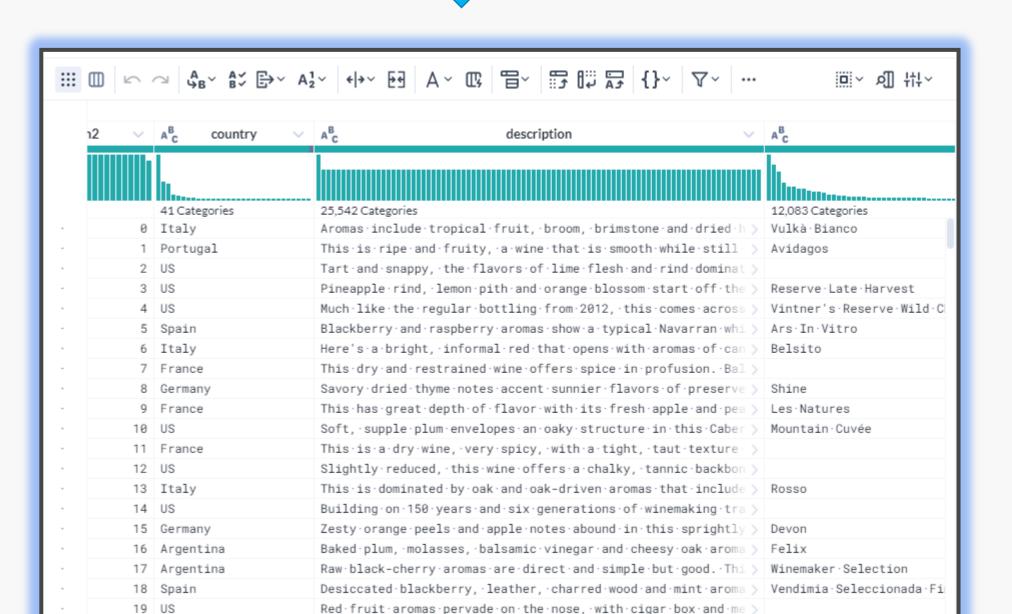
Vamos carregar os primeiros dados.

Carregando dados

20 US

Os primeiros dados serão carregados para fazermos o tratamento dos dados.



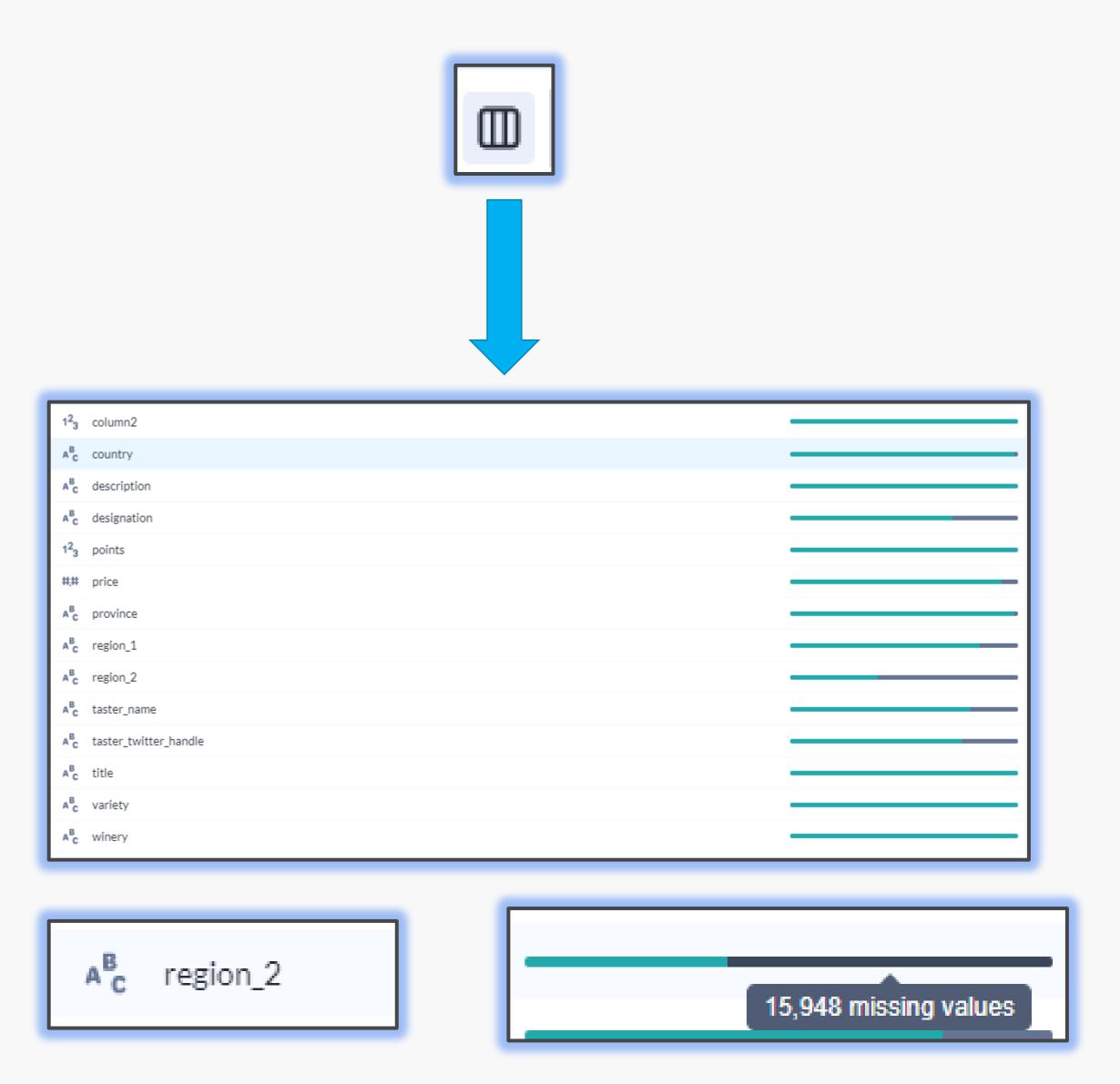


Ripe aromas of dark berries mingle with ample notes of blac > Vin de Maison

Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Função Columns

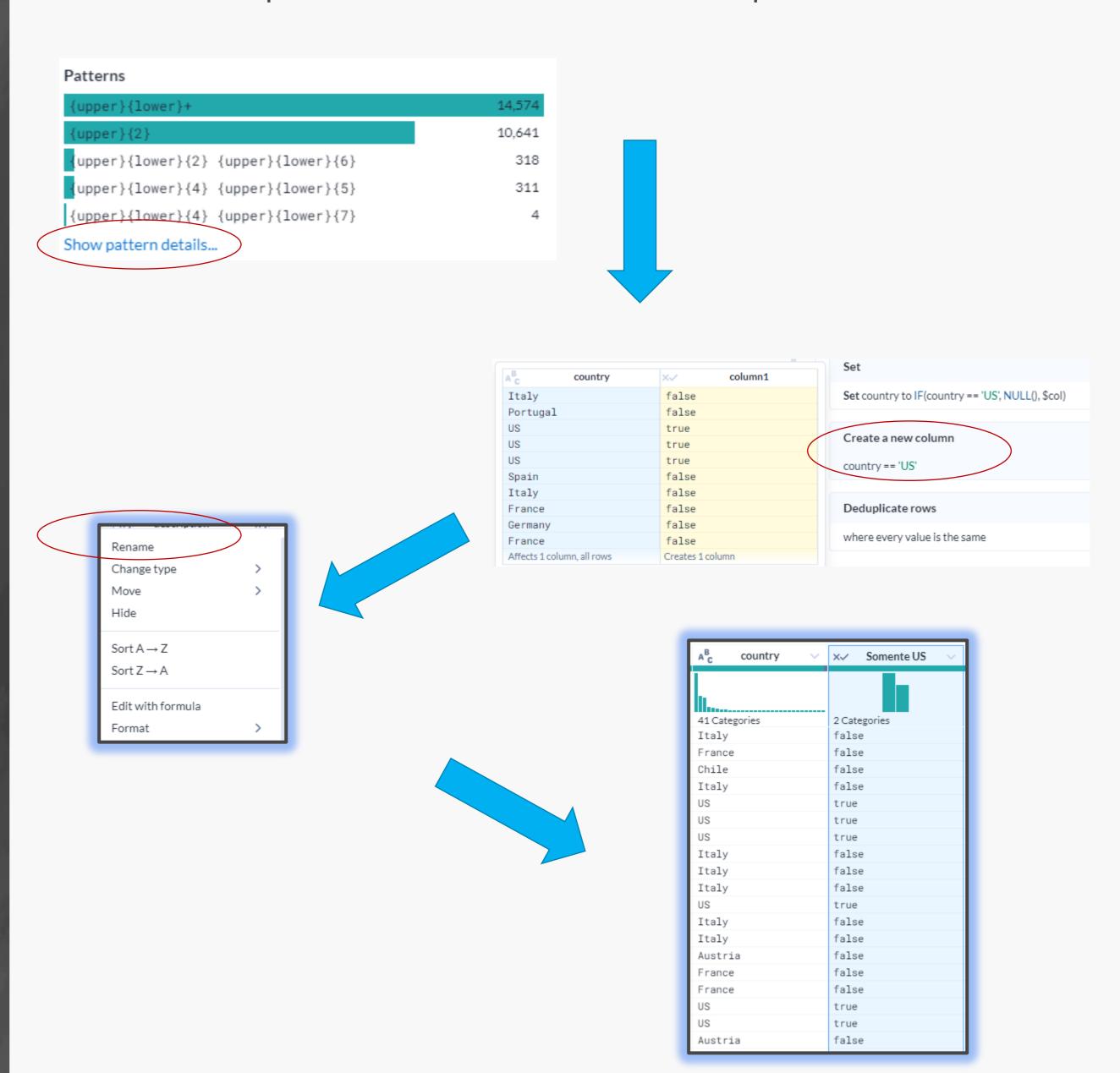
Verificamos o total de registros válidos e sem dados (missings).



Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Função Columns

Selecionando apenas US e criando um novo campo com TRUE e FALSE.

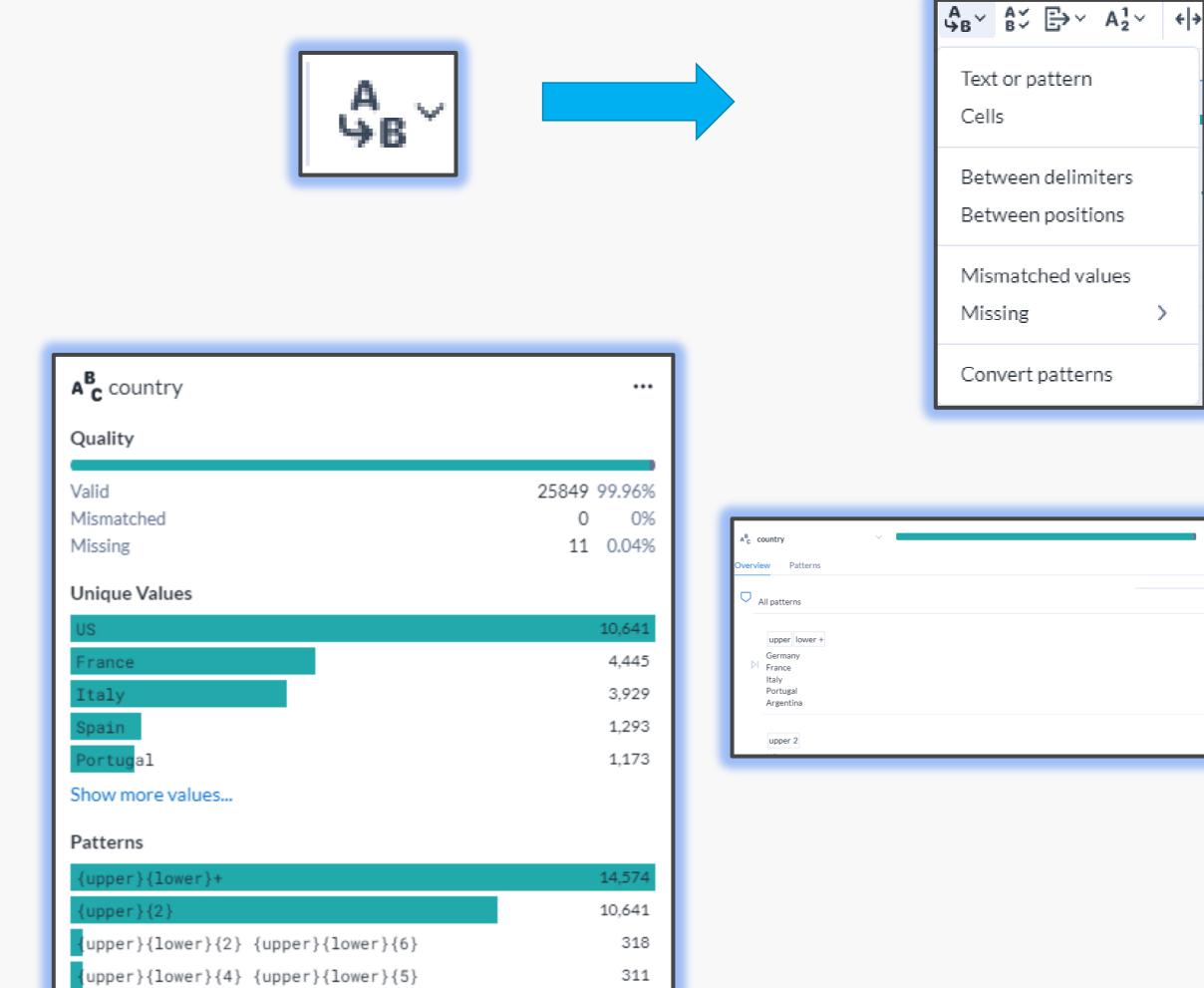


Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Função Substituir

{upper}{lower}{4} {upper}{lower}{7}

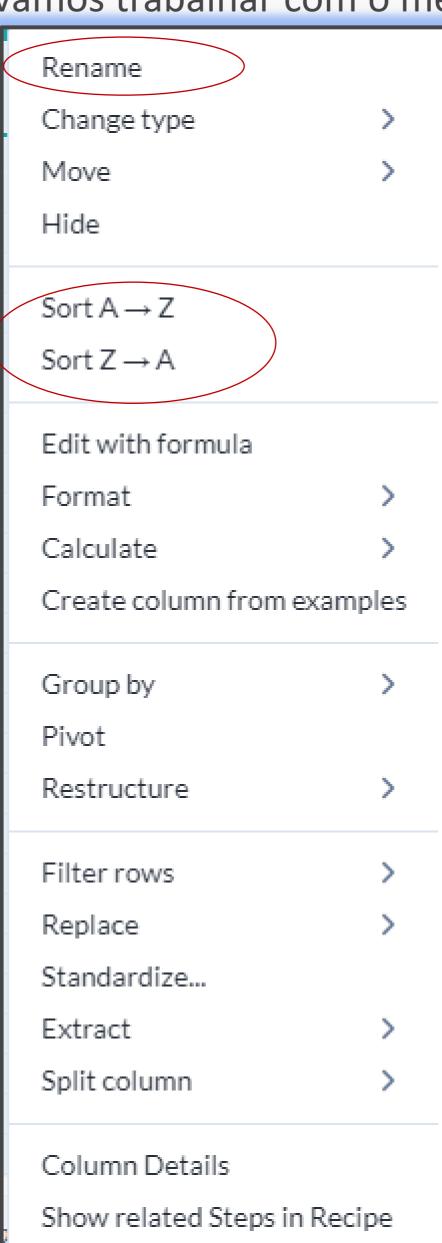
Ajusta dados dentro do TRIFACTA, trocando o conteúdo.

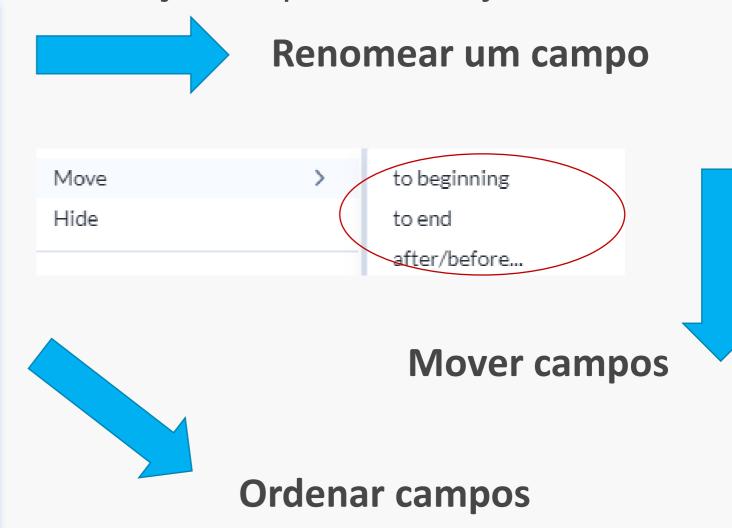


311

Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

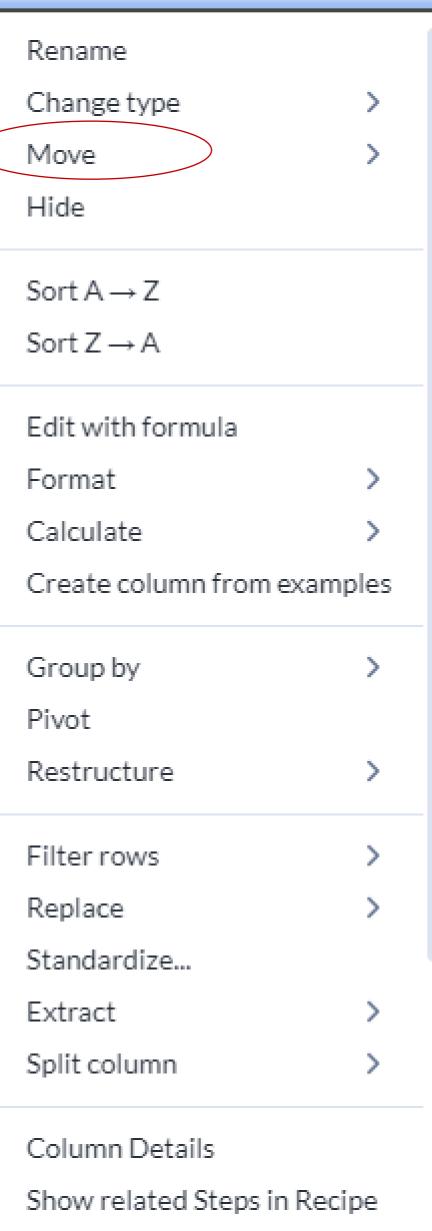
Funções Especiais de ajuste

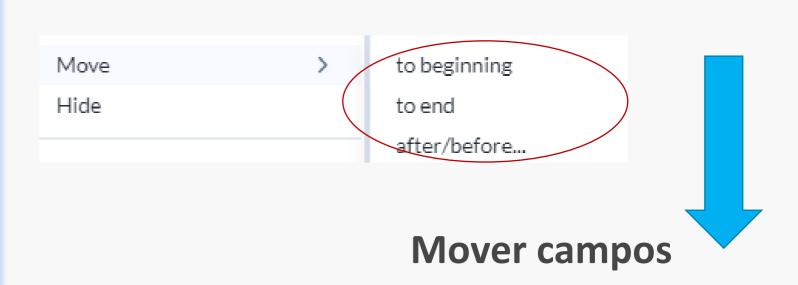




Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

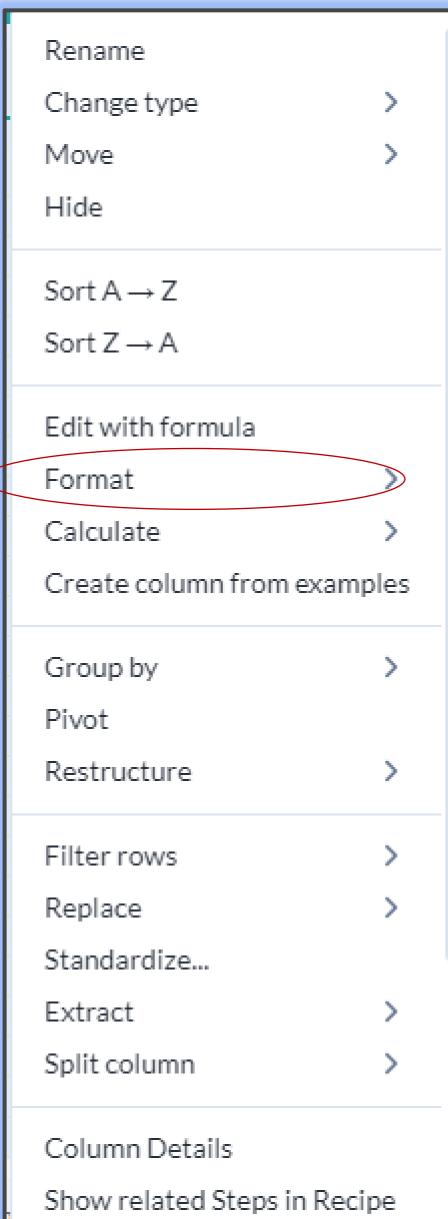
Funções Especiais de ajuste

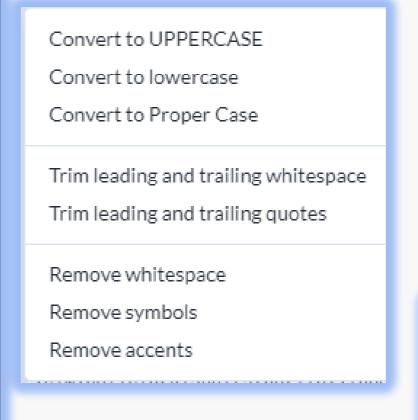




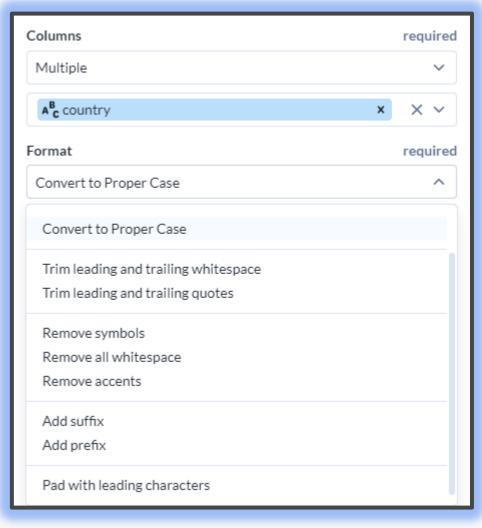
Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Funções Especiais de ajuste



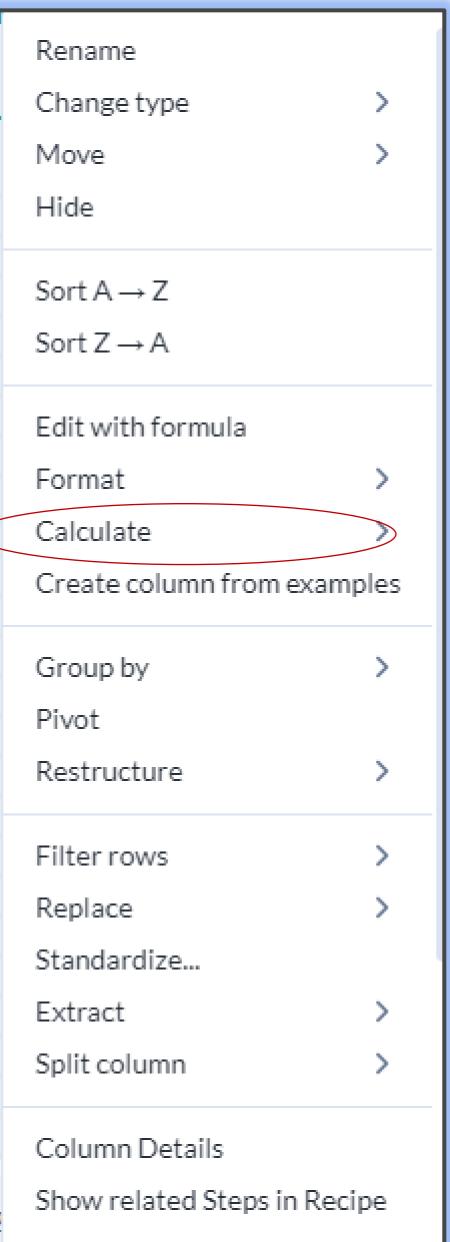


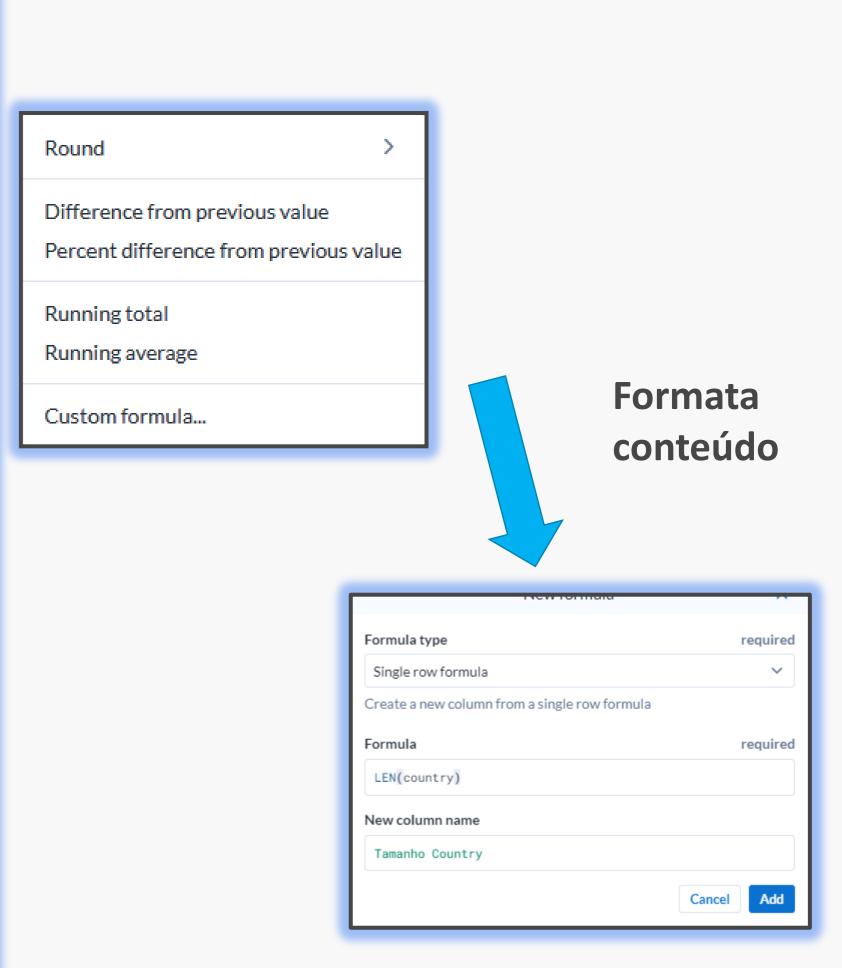




Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Funções Especiais de ajuste

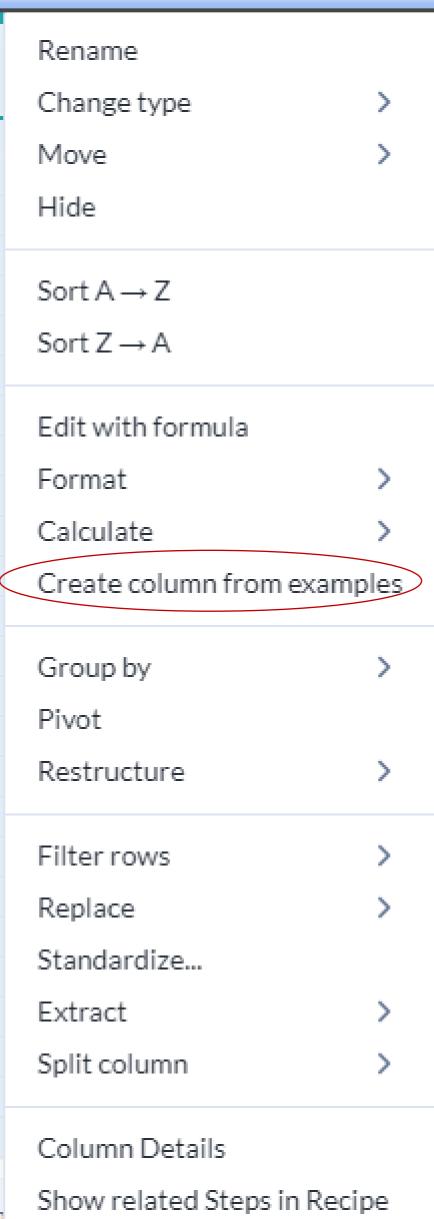




Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Funções Especiais de ajuste

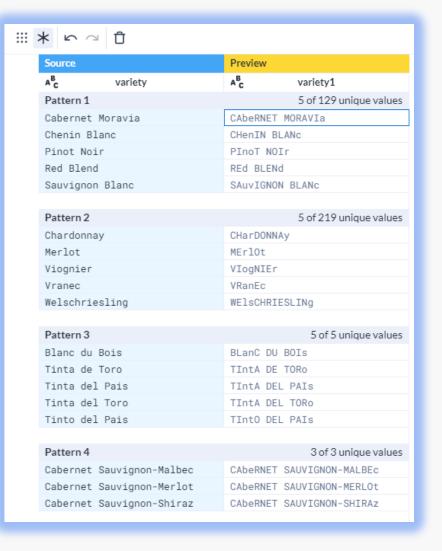
Vamos trabalhar com o menu de funções especiais de ajuste.







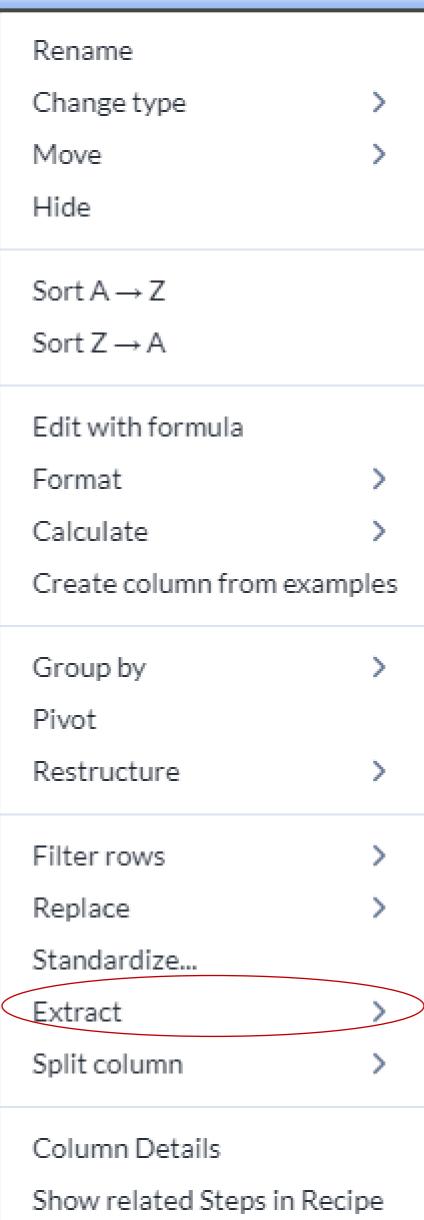
Formata conteúdo por exemplo

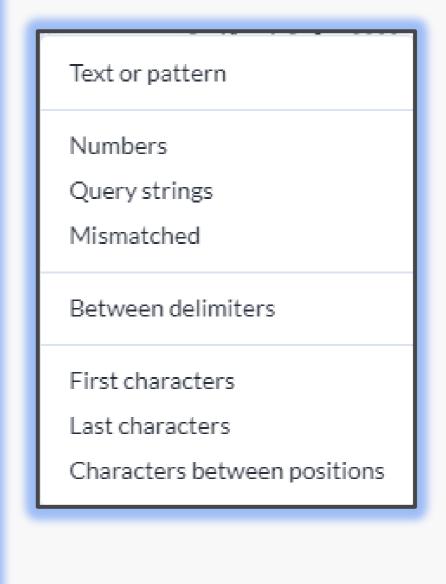


Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

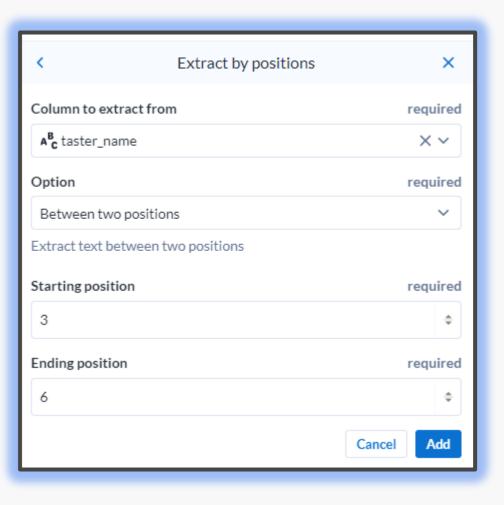
Funções Especiais de ajuste

Vamos trabalhar com o menu de funções especiais de ajuste.





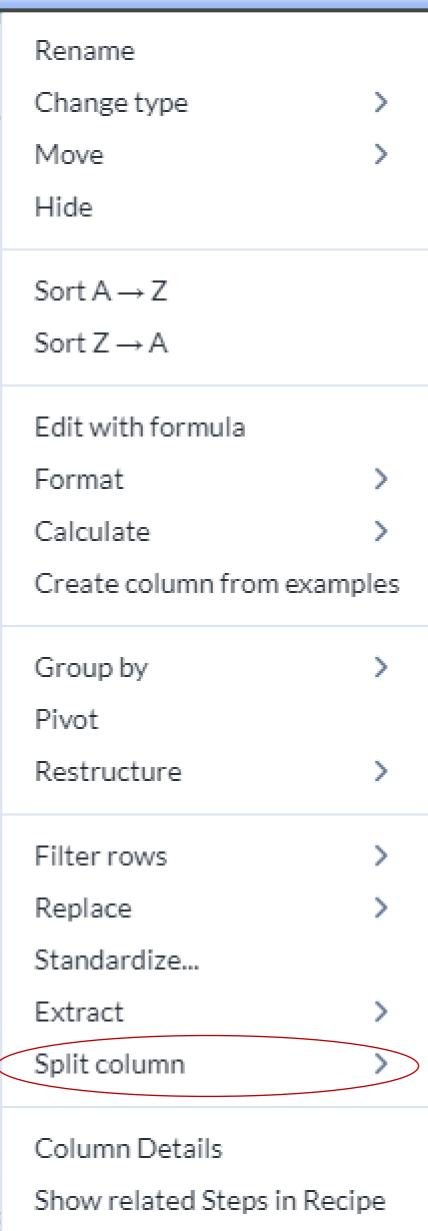
Formata conteúdo por exemplo

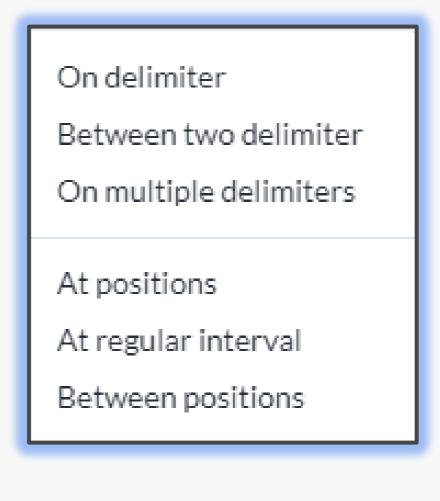


Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

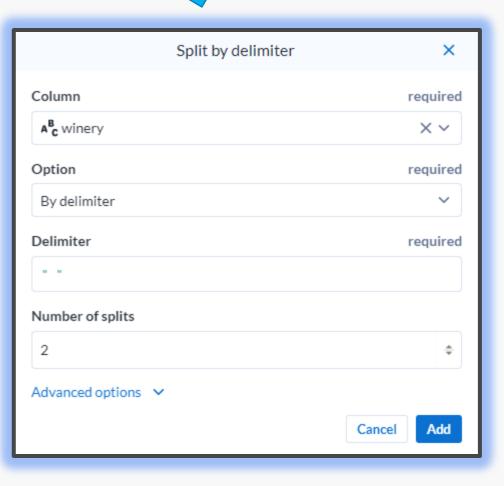
Funções Especiais de ajuste

Vamos trabalhar com o menu de funções especiais de ajuste.





Formata conteúdo



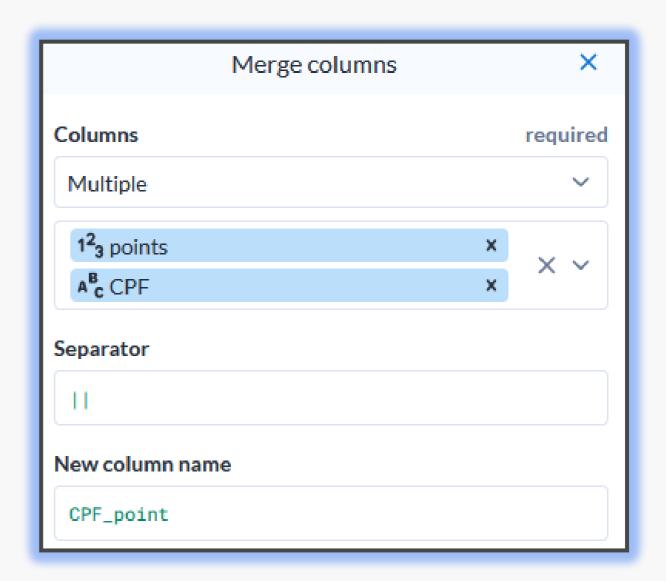
Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Função Merge

Consegue criar novas colunas unindo campos.







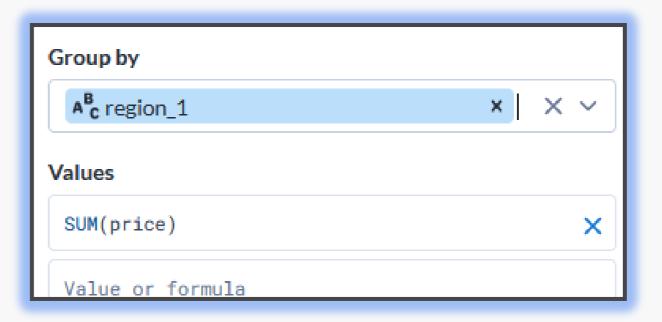
Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Função Agrupamento

Criando agrupamentos com dados numéricos.



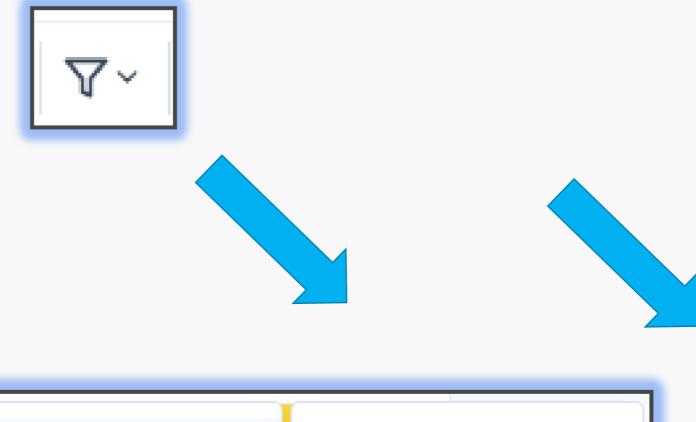


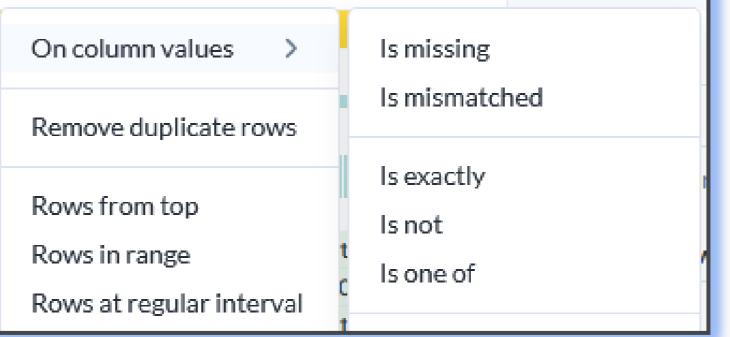


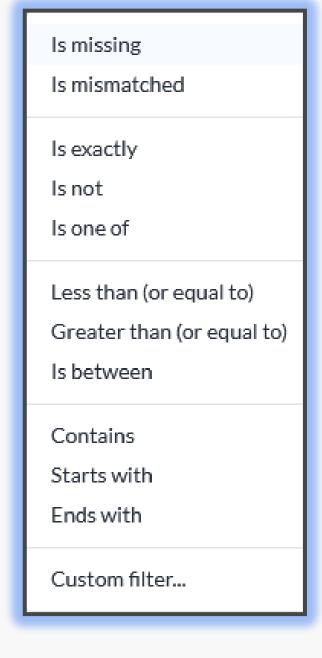
Vamos trabalhar com o tratamento de dados com as funções do TRIFACTA.

Função Filtro

Criação de filtros nos dados.







Funções de controle

São funções de manuseio do TRIFACTA.

Funções de Controle

Como podemos buscar um campo, filtro por tipo, etc.





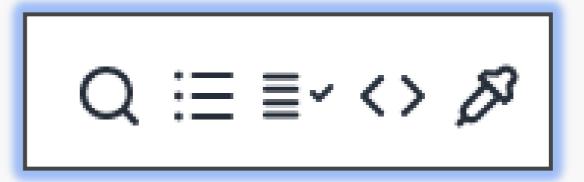
Columns Rows						
Filter by name (Ctrl+Shift+f)						
Filter by data type						
Decimal						
Integer						
String						
Clear all filters						

Funções especiais

São funções de manuseio do TRIFACTA.

Funções especiais

Criadas para fazerem testes e envio de parâmetros.







- 13 Sort rows by price as Ascending
- 14 Create rollingaverage_points from ROLLINGAVERAGE(points, -1, 0)
- 15 Create duplica_average from rollingaverage_points
- 16 Lock duplica_average type to String
- 17 Set duplica_average to SUBSTRING(duplica_average, 3, 11)

Not Missing

points must not contain missing values

Not Missing

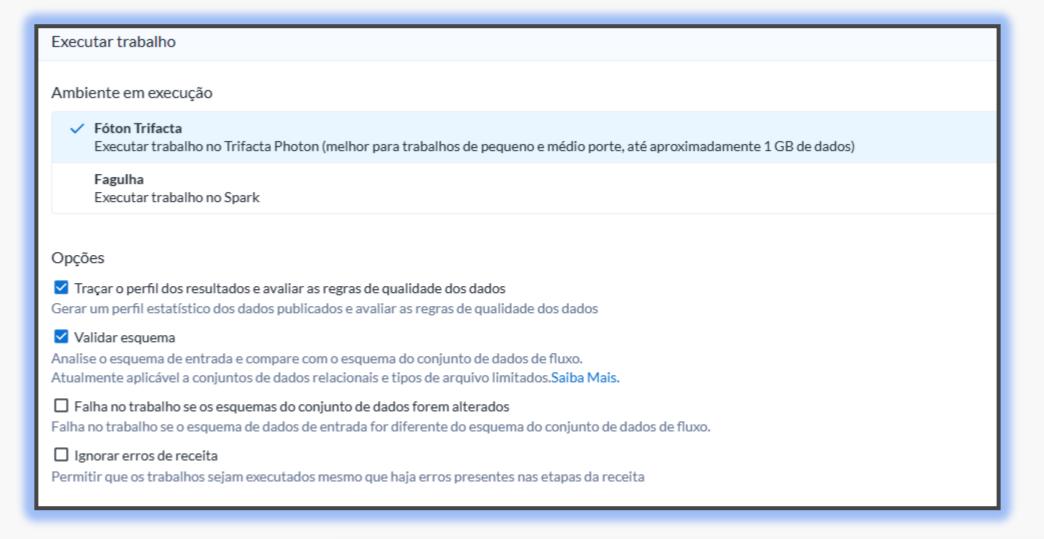
CPF must not contain missing values

Geração arquivo de saída

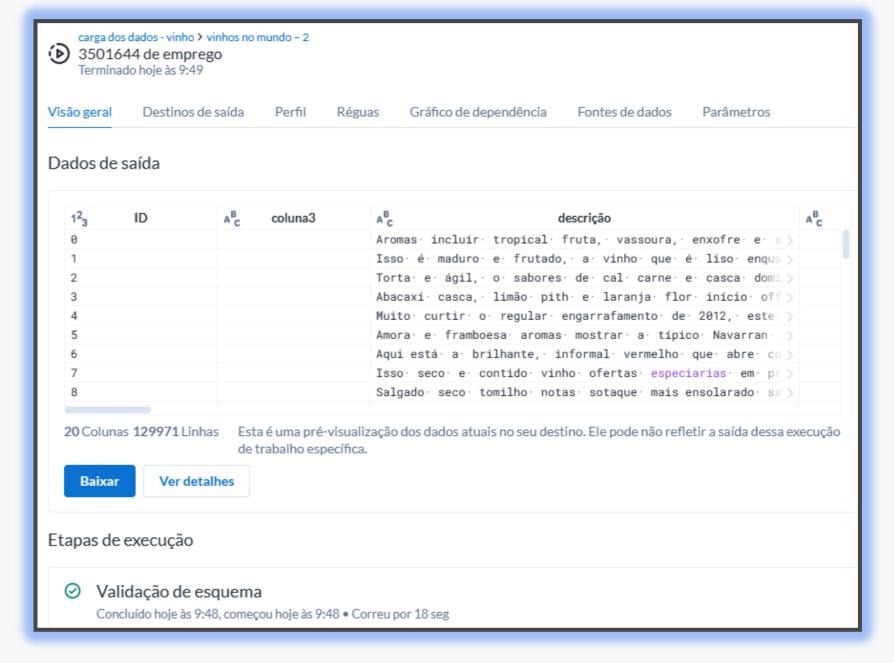
Vamos executar nosso fluxo de dados e gerar a saída.

Gerando o arquivo final

Preparando a saída dos dados com os tratamentos realizados.





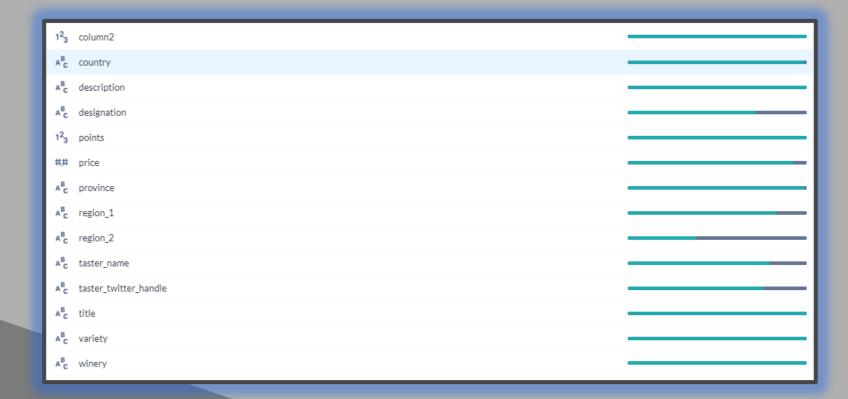


PRÁTICA

ENVIE AO PROFESSOR

PREPARAR

- O aluno deverá utilizar um arquivo de dados do site <u>www.kaggle.com</u>
- O aluno deverá criar um pipeline com 10 actions diferentes.
- O aluno deve enviar ao e-mail <u>grimaldo lopes@hotmail.com</u> um vídeo exibindo o conteúdo de Recipe, quais foram os tratamentos e ajustes nos dados.
- Deverá também mostrar em Excel o arquivo final gravado.



A ^B C	ID	~	×~	ISDELETED	~	1 ² 3	ACCOUNTID	~
							adia.	ī
1,000 Cate	pories		2 Categ	rories		1.12N	1 - 10M	
	-469a-4505-841a-de33	a5fd543f	false	,			65	69914
db4b16e4	-2cf3-4fbc-9971-0ddf	76737fd6	false				83	25407
79c5405b	-a970-499b-a8e5-9c86	ddb8597e	false				77	07214
7e9834c6	-dcef-4e49-8042-6d0c	2816aebf	false				56	61854
1b272af4	-884e-4b4b-b212-5943	5b1ea472	false				30	45483
3f5eadbf	-464e-4afb-8ce0-6ad0	88052a1a	false				90	31702
37664e58	-caaf-4d43-a451-3bc7	c3d73259	true				14	43003
509f8042	-b0fa-4863-811d-86b2	73c1aece	false				35	88466
ea1b8fcf	-5313-4804-924a-72fa	8204a208	true				99	64086
42b48742	-cbc8-4bf5-807a-aa46	ccbc9742	false				35	92149
35ad1ad3	-3ade-453f-a082-cfb9	784d5699	false				52	88367
e6c6d70a	-4c32-4820-8109-6e18	0d31bf1b	false				57	44080
5d3fd305	-f429-47fe-a492-e2ae	73c9c1b4	false				36	13462
adbcaaff	-9d0c-4b3d-bfb9-6027	894849f6	false				35	07945
55266738	-a78a-412e-92c5-0db8	17491def	false				38	12506
bd51da06	-7d99-4f9c-9b11-2fb4	4f666c5f	false				77	60125
53a030a7	-aae5-4315-bfb4-af96	8f272457	false				91	05507
b7659d9c	-4f53-4f6e-bb70-4aa9	b90890ed	false				48	01463
3e534d44	-79ec-4192-9faa-1ab8	bef9aa52	false				60	67620
d67c8e31	-7943-462e-9b2e-286b	88a0e78f	true				97	42082
0d4a2884	-2472-4a2a-8c68-8a49	0a22a143	false				99	58440