

Assistente de Analista de dados



Alejandra Caggiano

O Data Analyst Assistant é um assistente de GeneXus Enterprise AI que utiliza inteligência artificial para analisar dados e gerar listas, resumos e descrições úteis. Oferece respostas específicas em prazos curtos para a tomada de decisões.

Assistente de Analista de dados

- Metadados
- Glossário
- Conjunto de dados

Para definir um assistente deste tipo, você deve ter a seguinte informação:

- Os **Metadados**: Isso inclui uma descrição de cada conjunto de dados, suas colunas (incluindo os tipos de dados e valores possíveis) e considerações para o codificador e o interpretador.
- Também deve ser indicado o **Glossário**: Significa uma lista de termos utilizados dentro da empresa e/ou o domínio do usuário que seriam relevantes para que o LLM entenda as perguntas.
- E também **os Conjuntos de Dados**, que é um conjunto de arquivos com formato CSV que seguem as descrições adicionadas como metadados.

Toda esta informação é então utilizada para se comunicar com a API, permitindo que os usuários finais possam interagir com ela através de GeneXus Enterprise AI.

Assistente de Analista de dados

The image displays two overlapping screenshots of the GeneXus Enterprise AI interface. The top screenshot shows the 'Assistant' page for a project named 'Training/Project (Training)'. It features a sidebar with 'PROJECT OPTIONS' (Dashboard, Assistants, RAG Assistants, Playground, Requests, Api Tokens, Members) and 'ORGANIZATION OPTIONS' (Projects, Central dashboard, Members, Api Tokens, Settings). The main content area is titled 'Assistant' and includes a 'Create your own Assistants' section with buttons for 'CREATE CHAT ASSISTANT', 'CREATE API ASSISTANT', 'CREATE CHAT WITH DATA ASSISTANT', and 'CREATE DATA ANALYST ASSISTANT'. Below this is a 'My Assistants' section with a table of existing assistants.

Assistant Name	Type	Active Revision	Last Update
MarketingAssistant	Prompt/Chat	1	02/15/24 02:33 PM UPDATE
WeatherAssistant	Prompt/Chat	1	02/03/24 09:37 AM UPDATE

The bottom screenshot shows the 'Assistant Editor' for a 'New data analyst assistant'. It has a similar sidebar and a main area with two text input fields for 'Metadata (Chat)' and 'Memory (Chat)', both with a placeholder '(Step 2/2) (Format is supported)'. Below the fields is a note: 'Please, select files containing the datasets for the assistant.' and a '+ Add Files' button. At the bottom are 'SAVE' and 'CANCEL' buttons.

Vejamos agora um exemplo de como criar este tipo de assistente.

Em primeiro lugar entramos no Backoffice de GeneXus Enterprise AI.

Selecionamos o projeto sobre o qual vamos trabalhar e na opção Assistants do menu escolhemos Create Data Analyst Assistant..

Devemos então definir os metadados

Se trata de um JSON com informação descritiva sobre a informação contida em um conjunto de dados. É utilizado para fornecer contexto e compreender melhor os dados disponíveis.

Assistente de Analista de dados: Metadados

Nosso exemplo:

```
{
  "dataset_name": {
    "dataframe name": "dataset_name",
    "description": "contains data regarding different types of revenues, their types, dates,
and associated details.",
    "column explanations": {
      "column1": "description column 1. dtype: dtype1",
      "column2": "description column 2. dtype: dtype2",
      ...
      "columnN": "description column N. dtype: dtypeN"
    },
    "considerations": {
      "coder": [
        "Describe you consideration here",
        .....
        "Example of consideration: use co
      ],
      "interpreter": [
        "Example 1",
        "Example N"
      ]
    }
  }
}
```

Product	Price	January	February	March	April
Notebook	850	85	110	99	83
Wireless mou	15	52	27	36	75
Notebook des	180	15	22	10	19
Chair	50	60	72	52	49
Smart TV	450	16	20	24	18
Smart Watch	48	80	77	92	102
Cell phone	490	53	49	60	58
Gaming consc	120	15	10	17	14
Gaming keybc	80	21	30	19	27
Pendrdrive	12	120	132	98	114

```
"mysales": {
  "dataframe name": "mysales ",
  "description": " contains data on technology products, prices and sales in January, Febr
uary, March and April ",
  "column explanations": {
    "product": "contains the product name. dtype: string",
    "price": "contains the product price. dtype: integer",
    "january": "contains the sales in January. dtype: integer",
    "february": "contains the sales in February. dtype: integer",
    "march": "contains the sales in March. dtype: integer",
    "april": "contains the sales in April. dtype: integer"
  },
  "considerations": {
    "coder": [
      "Use Product to determine the name of the Product",
      "Use Price to determine the price of the Product",
      "Use January to determine the sales of the product in January",
      "Use February to determine the sales of the product in February",
      "Use March to determine the sales of the product in March",
      "Use April to determine the sales of the product in April"
    ],
    "interpreter": [
      "Responses should be returned in the language preferred by the end user or acco
the context of use "
    ]
  }
}
```

Esses metadados incluem detalhes como o nome do conjunto de dados, uma descrição de seu conteúdo e a estrutura de suas colunas (incluindo os tipos de dados).

Para então completar esses metadados corretamente, deve-se definir um JSON com a seguinte estrutura:

Onde:

O "dataframe_name" deve conter o nome do conjunto de dados e corresponder ao nome de um dos arquivos .csv que serão carregados.

A "descrição" deve fornecer uma descrição do conjunto de dados, incluindo os tipos de dados de suas colunas e os possíveis valores.

As "explicações de colunas" devem incluir uma descrição de cada coluna do conjunto de dados, especificando o seu objetivo e o tipo de dados que contém.

É importante considerar que "column1", a "columnN": corresponde aos nomes exatos (sem espaços em branco ou caracteres especiais) das colunas dos arquivos .csv a serem carregados.

A quantidade de colunas nos arquivos .csv deve coincidir com a quantidade

de colunas fornecidas nesses metadados. Além disso, é importante evitar a duplicação de nomes de colunas.

Quanto às "considerações", devem detalhar considerações específicas para o codificador e o interpretador.

O "codificador" gera o código necessário para processar os dados. Além disso, utiliza a informação dos metadados para guiar o processo de extração de informação relevante dos dados.

Enquanto o "interpretador", sua função é produzir a resposta final com a informação obtida no processamento de dados realizado pelo codificador. Analisa os resultados para compreender melhor os dados e fornecer respostas relevantes.

Pode incluir considerações que indiquem a compreensão dos termos técnicos e das abreviaturas utilizadas nos dados, bem como especificar que as respostas devem ser devolvidas no idioma preferido do usuário final ou de acordo com o contexto de uso.

No nosso exemplo temos um conjunto de dados bem simples, onde são registrados produtos de tecnologia, e sua respectiva quantidade de unidades vendidas nos meses de Janeiro a Abril.

Carregamos os metadados levando em consideração tudo o que mencionamos anteriormente.

Assistente de Analista de dados: Glossário

```
{
  "glossary": {
    "term1": "Definition of term 1",
    "term2": "Definition of term 2",
    "term3": "Definition of term 3",
    "abbr1": "Abbreviation 1 - Meaning of abbreviation 1",
    "abbr2": "Abbreviation 2 - Meaning of abbreviation 2",
    "abbr3": "Abbreviation 3 - Meaning of abbreviation 3"
  }
}
```

Nosso exemplo:

```
{
  "glossary": {
    "Product": "A technology product",
    "Sales": "Units sold of the product",
    "Best selling product": "The product with the most units sold in the month",
    "January": "Units of the product sold in January",
    "February": "Units of the product sold in February",
    "March": "Units of the product sold in March",
    "April": "Units of the product sold in April"
  }
}
```

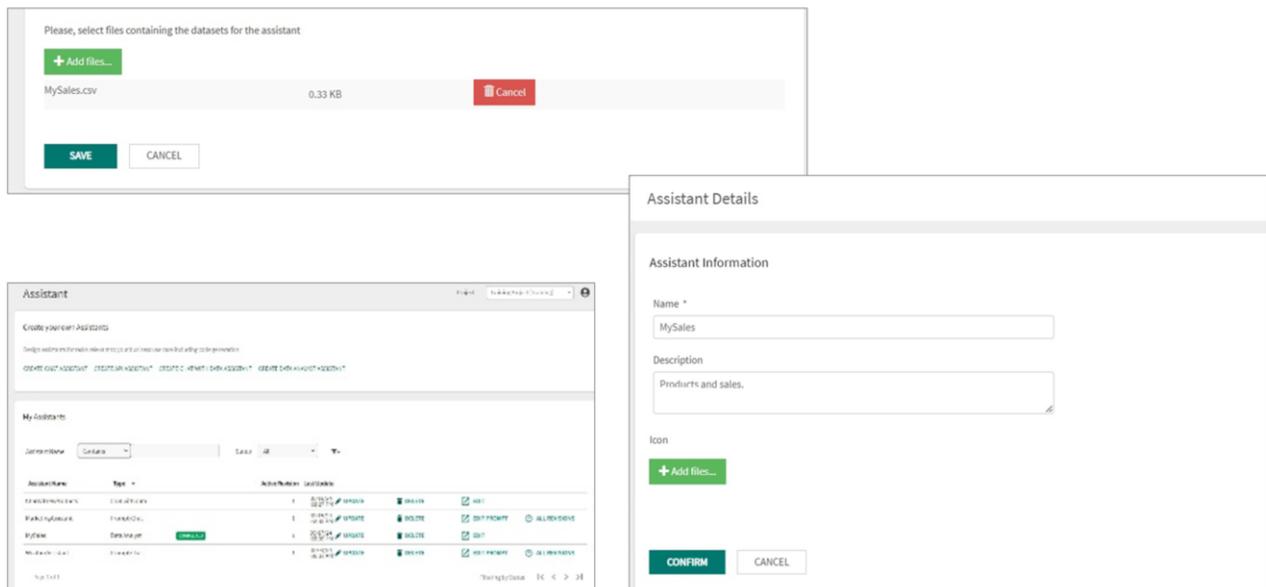
Bem. O passo seguinte é definir o **glossário**.

Trata-se também de um JSON que contém uma lista de termos utilizados na empresa ou domínio do usuário final, juntamente com suas respectivas definições. Esses termos e definições são relevantes para que o LLM compreenda as perguntas e forneça respostas mais precisas.

Por exemplo, pode incluir abreviaturas que são utilizadas frequentemente na empresa e seus significados para uma melhor compreensão em um formato como o seguinte:

Carregamos então, o glossário associado ao nosso exemplo.

Assistente de Analista de dados: Criação



Bem. Chegado a este ponto, a próxima etapa é **adicionar arquivos**.

Para fazer isso, devemos carregar um arquivo do conjunto de dados, selecionando Add files...

É importante garantir que o arquivo CSV siga as descrições e metadados adicionados anteriormente.

Finalmente, para criar o assistente, selecionamos Save.

Aparece então uma janela para inserir um Nome, Descrição e Ícone, se desejado. Pressionamos Confirm.

E voltamos à página onde vemos o status atual de criação do assistente.

Este status está relacionado ao processo de carga dos conjuntos de dados com a informação dos metadados e glossário.

Os status possíveis são FALHO, CONCLUÍDO, ou PROCESSANDO (indicando a porcentagem de progresso).

Asistente de Analista de datos: Atualização

Assistant Details

Assistant: MySales

Assistant Id
ce96d69a-6a89-42f5-9d32-09d851b21d7b

Name *
MySales

Description
Products and sales.

Status
Enabled

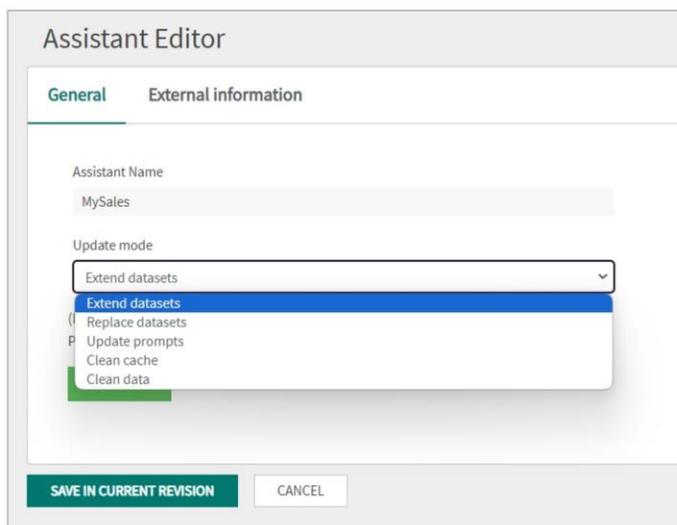
Icon

+ Add files...

CONFIRM CANCEL

Una vez que se complete el estado de carga, podemos seleccionar Update y ver el identificador de versión con la que se guardó, podemos cambiar el nombre, descripción, configurarlo como habilitado o deshabilitado, o agregar un ícono si se desea.

Assistente de Analista de dados: Edição



Assistant Editor

General External information

Assistant Name
MySales

Update mode
Extend datasets

- Extend datasets
- Replace datasets
- Update prompts
- Clean cache
- Clean data

SAVE IN CURRENT REVISION CANCEL

Também é possível editá-lo, para ampliar ou substituir o conjunto de dados, ou para atualizar o prompt.

Ao ampliar o conjunto de dados se amplia a informação disponível sem afetar a estrutura original do conjunto de dados.

Os arquivos selecionados para a integração devem atender às mesmas condições iniciais, como ter a mesma quantidade de colunas e nomes de arquivos.

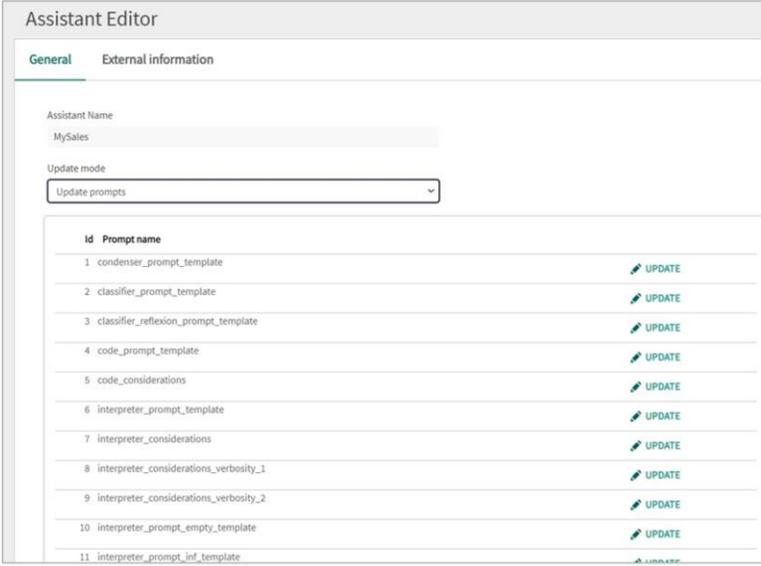
É fundamental evitar que os arquivos utilizados para a extensão contenham as mesmas linhas que já estão carregadas no assistente, pois isso poderia gerar chaves duplicadas e causar problemas na execução das consultas.

Também deve ser considerado que ao ampliar o conjunto de dados não é possível realizar alterações nos metadados ou no glossário. Este recurso concentra-se exclusivamente em ampliar os conjuntos de dados existentes sem alterar outros aspectos do assistente.

Por sua vez, ao substituir o conjunto de dados, é permitido substituir completamente os dados no conjunto de dados existente, bem como modificar os metadados e o glossário associado.

Isso é útil quando é necessário atualizar completamente a informação ou realizar alterações significativas na estrutura e na descrição do conjunto de dados.

Assistente de Analista de dados: Atualizar prompt



The screenshot shows the 'Assistant Editor' interface. It has two tabs: 'General' and 'External information'. Under 'General', there is a text field for 'Assistant Name' containing 'MySales' and a dropdown menu for 'Update mode' set to 'Update prompts'. Below this is a table of prompts.

Id	Prompt name	UPDATE
1	condenser_prompt_template	UPDATE
2	classifier_prompt_template	UPDATE
3	classifier_reflexion_prompt_template	UPDATE
4	code_prompt_template	UPDATE
5	code_considerations	UPDATE
6	interpreter_prompt_template	UPDATE
7	interpreter_considerations	UPDATE
8	interpreter_considerations_verbosity_1	UPDATE
9	interpreter_considerations_verbosity_2	UPDATE
10	interpreter_prompt_empty_template	UPDATE
11	interpreter_prompt_inf_template	UPDATE

Se for selecionado Update prompt, será exibida uma lista dos 28 prompts padrão.

É possível consultar e atualizar cada uma das mensagens de forma individual clicando em Update.

Para fazer isso é necessário ter a role de Administrador, ProjectRole ou OrganizationRole.

É importante levar em consideração que essas atualizações são realizadas a partir da lista de prompts e não no nível geral do assistente.

Ao gerar uma nova versão do assistente e realizar alterações com as opções Estender conjuntos de dados ou Substituir conjuntos de dados, serão ignoradas as alterações realizadas nas indicações em versões anteriores do assistente.

Como sempre, podemos finalmente testar o assistente criado utilizando a opção Playground no menu do Backoffice.

GeneXus[™]
by **Globant**

training.genexus.com