

Asistente de Analista de datos



Alejandra Caggiano

El Data Analyst Assistant es un asistente de GeneXus Enterprise AI que utiliza inteligencia artificial para analizar datos y generar listas, resúmenes y descripciones útiles. Ofrece respuestas específicas en plazos cortos para la toma de decisiones.

Asistente de Analista de datos

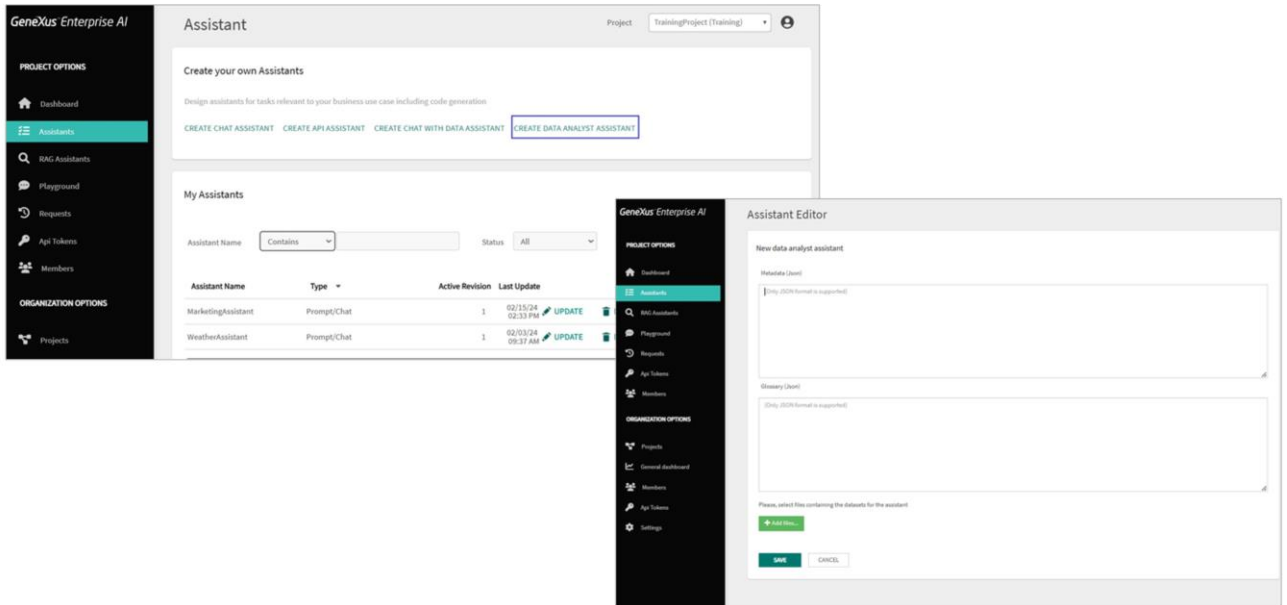
- Metadata
- Glossary
- Datasets

Para definir un asistente de este tipo, se debe disponer de la siguiente información:

- La **Metadata**: Esto incluye una descripción de cada conjunto de datos, sus columnas (incluidos los tipos de datos y valores posibles) y consideraciones para el codificador y el intérprete.
- También se debe indicar el **Glosario**: Significa una lista de términos utilizados dentro de la empresa y/o el dominio del usuario que serían relevantes para que el LLM comprenda las preguntas.
- Y también el **Conjuntos de datos**, que es un conjunto de archivos con formato CSV que siguen las descripciones agregadas como metadata.

Toda esta información es utilizada luego para comunicarse con la API permitiendo a los usuarios finales poder interactuar con ella a través de GeneXus Enterprise AI.

Asistente de Analista de datos



Veamos ahora un ejemplo de cómo crear este tipo de asistente.

En primer lugar ingresamos al Backoffice de GeneXus Enterprise AI.

Seleccionamos el proyecto sobre el cual vamos a trabajar, y en la opción Assistants del menú, elegimos Create Data Analyst Assistant..

Debemos definir entonces la metadata

Se trata de un JSON con información descriptiva sobre la información contenida en un conjunto de datos. Se utiliza para proporcionar contexto y comprender mejor los datos disponibles.

Asistente de Analista de datos: Metadata

Nuestro ejemplo:

```
{
  "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
ntext of use "
    ]
  }
}
```

Esta metadata incluye detalles como ser el nombre del conjunto de datos, una descripción de su contenido y la estructura de sus columnas (incluidos los tipos de datos).

Para completar entonces esta metadata correctamente, se debe definir un JSON con la siguiente estructura:

Dónde:

El "dataframe_name" debe contener el nombre del conjunto de datos y corresponder al nombre de uno de los archivos .csv que se cargarán.

La "descripción" debe proporcionar una descripción del conjunto de datos, incluidos los tipos de datos de sus columnas y los posibles valores.

Las "explicaciones de columnas" deben incluir una descripción de cada columna del conjunto de datos, especificando su propósito y el tipo de datos que contiene.

Es importante considerar que "column1", a "columnN": corresponde a los nombres exactos (sin espacios en blanco ni caracteres especiales) de las columnas de los archivos .csv a cargar.

La cantidad de columnas en los archivos .csv debe coincidir con la cantidad de

columnas proporcionadas en esta metadata. Además, es importante evitar la duplicación de nombres de columnas.

En cuanto a las "consideraciones", deben detallar consideraciones específicas para el codificador y el intérprete.

El "codificador" genera el código necesario para procesar los datos. Además, utiliza la información de la metadata para guiar el proceso de extracción de información relevante de los datos.

En cuanto a "intérprete, su función es producir la respuesta final con la información obtenida del procesamiento de datos realizado por el codificador. Analiza los resultados para comprender mejor los datos y proporcionar respuestas relevantes.

Puede incluir consideraciones que indiquen la comprensión de los términos técnicos y abreviaturas utilizadas en los datos, así como especificar que las respuestas deben devolverse en el idioma preferido por el usuario final o según el contexto de uso.

En nuestro ejemplo tenemos un conjunto de datos muy sencillo, donde se registran productos de tecnología, y su respectiva cantidad de unidades vendidas en los meses de Enero a Abril.

Cargamos la metadata teniendo en cuenta todo lo que mencionamos anteriormente.

Asistente de Analista de datos: Glosario

```
{
  "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"
  }
}
```

Nuestro ejemplo:

```
{
  "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"
  }
}
```

Bien. El paso siguiente es definir el **glosario**.

Se trata también de un JSON que contiene una lista de términos utilizados en la empresa o dominio del usuario final, junto con sus respectivas definiciones. Estos términos y definiciones son relevantes para que el LLM comprenda las preguntas y proporcione respuestas más precisas.

Por ejemplo, puede incluir abreviaturas que se utilizan frecuentemente en la empresa y sus significados para una mejor comprensión en un formato como el siguiente:

Cargamos entonces el glosario asociado a nuestro ejemplo.

Asistente de Analista de datos: Creación

Please, select files containing the datasets for the assistant

+ Add files...

MySales.csv 0.33 KB Cancel

SAVE CANCEL

Assistant

Create Assistant

My Assistants

Assistant Name	Type	Status	Created	Updated	Deleted	Export	Import	Refresh
TRANSACCIONES	TRANSACCIONES	FALLADO	2023-10-27 10:00					
TRANSACCIONES	TRANSACCIONES	COMPLETADO	2023-10-27 10:00					
TRANSACCIONES	TRANSACCIONES	PROCESANDO	2023-10-27 10:00					

Assistant Details

Assistant Information

Name *

MySales

Description

Products and sales.

Icon

+ Add files...

CONFIRM CANCEL

Bien. Llegado a este punto, el paso siguiente es agregar archivos

Para eso, debemos cargar un archivo del conjunto de datos, seleccionando Add files...

Es importante asegurarse de que el archivo CSV siga si las descripciones y metadata agregados anteriormente.

Cargamos entonces nuestro archivo.

Finalmente, para crear el asistente, seleccionamos Save.

Aparece entonces una ventana para ingresar un Nombre, Descripción e Icono si se desea-

Presionamos Confirm.

Y volvemos a la página donde vemos el estado actual de creación del asistente.

Este estado está relacionado con el proceso de carga de los conjuntos de datos con la información de la metadata y el glosario.

Los estados posibles son FALLADO, COMPLETADO, o PROCESANDO (indicando el porcentaje de progreso).

Asistente de Analista de datos: Actualización

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.

Asistente de Analista de datos: Edición

The image shows a dialog box titled "Assistant Editor" with two tabs: "General" and "External information". The "General" tab is active. It contains the following fields and controls:

- Assistant Name:** A text input field containing "MySales".
- Update mode:** A dropdown menu with a downward arrow. The menu is open, showing the following options:
 - Extend datasets (highlighted in blue)
 - Replace datasets
 - Update prompts
 - Clean cache
 - Clean data
- Buttons:** At the bottom, there are two buttons: "SAVE IN CURRENT REVISION" (in a green box) and "CANCEL" (in a white box).

También es posible editarlo, para ampliar o reemplazar el conjunto de datos, o para actualizar el prompt. Al ampliar el conjunto de datos se amplía la información disponible sin afectar la estructura original del conjunto de datos.

Los archivos seleccionados para la integración deben cumplir las mismas condiciones iniciales, como tener la misma cantidad de columnas y nombres de archivos.

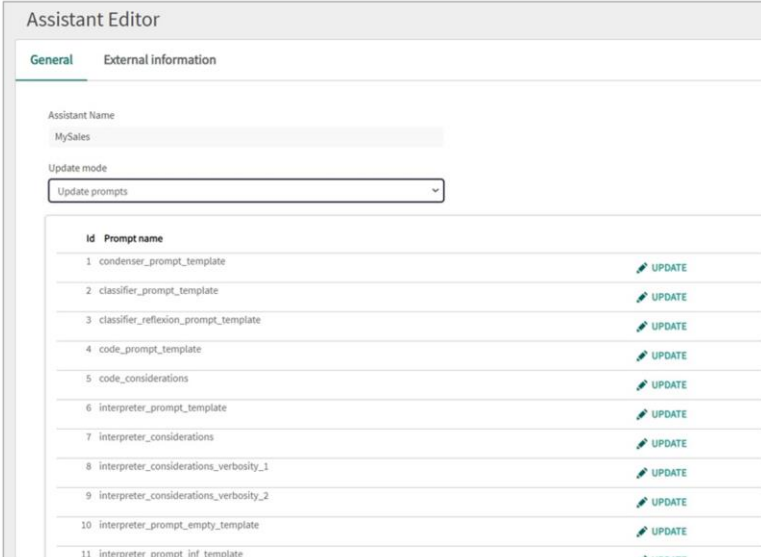
Es fundamental evitar que los archivos utilizados para la extensión contengan las mismas filas que ya están cargadas en el asistente, ya que esto podría generar claves duplicadas y causar problemas en la ejecución de las consultas.

También hay que considerar que al ampliar el conjunto de datos no es posible realizar cambios en la metadata ni en el glosario. Esta función se centra exclusivamente en ampliar los conjuntos de datos existentes sin alterar otros aspectos del asistente.

. Por su parte, al reemplazar el conjunto de datos, se permite reemplazar completamente los datos en el conjunto de datos existente, así como también modificar la metadata y el glosario asociado.

Esto es útil cuando se necesita actualizar completamente la información o realizar cambios significativos en la estructura y descripción del conjunto de datos.

Asistente de Analista de datos: Actualizar 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

Si se selecciona Update prompt, se mostrará una lista de los 28 prompts predeterminados.

Es posible consultar y actualizar cada uno de los mensajes en forma individual haciendo clic en Update.

Para esto se requiere tener el rol de Administrador, ProjectRole u OrganizationRole.

Es importante tener en cuenta que estas actualizaciones se realizan desde la lista de prompts y no a nivel general del asistente.

Al generar una nueva versión del asistente y realizar cambios con las opciones Extender conjuntos de datos o Reemplazar conjuntos de datos, se ignorarán los cambios realizados en las indicaciones en versiones anteriores del asistente.

Como siempre, podemos finalmente probar el asistente creado utilizando la opción Playground en el menú del Backoffice.

GeneXus[™]
by **Globant**

training.genexus.com