# **Data Analyst Assistant**



The Data Analyst Assistant is a GeneXus Enterprise AI assistant that uses artificial intelligence to analyze data and generate useful lists, summaries, and descriptions. It provides specific answers in short timeframes to support decision-making.



To define an assistant of this type, the following information must be available:

• **Metadata**: This includes a description of each data set, its columns (including possible data types and values), and considerations for the encoder and interpreter.

• The **Glossary** should be indicated: It means a list of terms used in the company and/or the user's domain that would be relevant for the LLM to understand the questions.

• And also the **Dataset**, which is a set of CSV-formatted files that follow the descriptions added as metadata.

All this information is then used to communicate with the API, enabling end users to interact with it through GeneXus Enterprise AI.

#### **Data Analyst Assistant**

in a prise ru	Assistant		
PROJECT OPTIONS	Create your own Assistants		
A Dashboard	Design assistants for tasks relevant to your business use case including code generation		
🖅 Assistants	CREATE CHAT ASSISTANT CREATE API ASSISTANT CREATE CHAT WITH DATA ASSISTANT CREATE C	A ANALYST ASSISTANT	
Q RAG Assistants			
Playground	My Assistants		
D Requests		GeneXus Enterprise Al Assistant Editor	
🔎 Api Tokens	Assistant Name Contains * Status All	work (expose)     New data analyst assistant	
221 Members		Construent Metalolo (June)	
ORGANIZATION OPTIONS	Assistant Name type Active Hervision Last Opate MarketingAssistant Prompt/Chat 1 0215244	PDATE I Q. BAS Annihile	
Y Projects	WeatherAssistant Prompt/Chat 1 00:37.84 09:37.84	IPDATE	
		· D Reports	
		22 Monstern Granary (Josef	
		(Dely 2009 Surgershiel)	
		🐨 . Projects	
		er General dashbeard 295 Manuara	
		Planes, select Alexanders and a selected and a sele	
		Selege     AMMen	
		SANE CHINCE,	

Let's see an example of how to create this type of assistant.

First, we go to the GeneXus Enterprise AI Backoffice

We select the project we will work on, and in the Assistants option of the menu, we choose Create Data Analyst Assistant.

Next, we must define the metadata.

This is a JSON with descriptive information about the information contained in a dataset. It is used to provide context and better understand the available data.

GeneXus by Globant

Data	Anal	vst	Assi	stant:	Metadata

							7	'mvsales,": {
dataset_name": {								"dataframe name": "mysales ",
"dataframe name": "dataset_name",							uar	"description": " contains data on technology products, prices and sales in January, Febr y, March and April ",
"description": "contains data regarding o associated details.",	different types of i	revenues, their t	ypes, dates,				5000	"column explanations": {
"column explanations":								"product": "contains the product name. dtyne: string",
								"price": "contains the product price. dtype: integer",
"colum1": "description column 1. dtyp	pe: dtype1",							"ianuarv": "contains the sales in January. dtype; integer",
"colum2": "description column 2. dtyp	pe: dtype2",							"february": "contains the sales in February. dtyne; integer",
								"march"; "contains the sales in March. dtvoe; integer",
"columN": "description column N. dtv	vpe: dtypeN"							"april": "contains the sales in April. dtvpe: integer"
column : description column it of	perotypen							1
1.								"considerations": {
"considerations": {								"coder": [
"coder": [								"Lise Product to determine the name of the Product"
"Describe you consideration bero"								"Lise Price to determine the price of the Product"
beschoe you consideration here ;								"Use Investo determine the price of the modulet in January"
	Product	Price	January	February	March	April		"Use Salidary to determine the sales of the product in Salidary ,
"Example of consideration: use co	Notebook	850	85	110	)	99	83	Use Peordary to determine the sales of the product in Peordary ,
k	Wireless mou	15	52	27	'	36	75	"Use March to determine the sales of the product in March",
	Notebook des	180	15	22	2	10	19	"Use April to determine the sales of the product in April"
"interpreter": [	Chair	50	60	72	2	52	49	).
"Example 1",	Smart TV	450	16	20	)	24	18	"interpreter": [
"Example N"	Smart Watch	48	80	77		92	102	" Responses should be returned in the language preferred by the end user or acco
source of	Cell phone	490	53	49		60	58	the context of use "
1	Gaming conso	120	15	10		1/	14	
}	Pendrive	12	120	122	,	98	114	
	renuive	12	120	152		50	114	

This **metadata** includes details such as the name of the dataset, a description of its contents and the structure of its columns (including data types).

To complete this metadata correctly, a JSON must be defined with the following structure:

Where:

The "dataframe\_name" must contain the name of the dataset and correspond to the name of one of the .csv files to be loaded.

The "description" must provide a description of the dataset, including the data types of its columns and possible values.

The "column explanations" must include a description of each column in the dataset, specifying its purpose and the type of data it contains.

It is important to consider that "colum1" to "columN": correspond to the exact names (without blanks or special characters) of the columns in the .csv files to be loaded.

The number of columns in the .csv files must match the number of columns provided in this metadata. In addition, it is important to avoid duplication of column names.

**GeneXus by Globant** 

As for the "considerations", they should describe specific considerations for the encoder and the interpreter.

The "encoder" generates the code necessary to process the data. In addition, it uses the metadata information to guide the process of extracting relevant information from the data.

As for the "interpreter", its role is to generate the final response with the information obtained from the data processing performed by the encoder. It analyzes the results to better understand the data and provide relevant answers.

It may include considerations for understanding the technical terms and abbreviations used in the data, as well as specifying that the answers must be provided in the language selected by the end user or according to the context.

In this example, we have a very simple dataset, where technology products are recorded, including the respective number of units sold in January to April.

We load the metadata taking into account everything we mentioned above.

### **Data Analyst Assistant: Glossary**

	"dossan".
	Biossary . (
	"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"
1	}
}	

#### Our example:

{		
	"g	lossary": {
		"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"
	}	
}		

Good. The next step is to define the **glossary**.

This is also a JSON containing a list of terms used in the end user's company or domain, along with their corresponding definitions. These terms and definitions are relevant for the LLM to understand the questions and provide more accurate answers.

For example, it may include abbreviations that are frequently used in the company and their meanings for better understanding in a format like the following:

We load the glossary associated with our example.

#### **Data Analyst Assistant: Creation**

	is for the assistant		
+ Add files			
MySales.csv	0.33 KB	Cancel	
SAVE CANCEL			
			Assistant Details
			Assistant Information
Assistant		Print [hikkgrainflooring] • 0	Name *
Create your even Assistants			MySales
Design and/rame for radio mission must be an instrumentation for a first case and	enco		
Dedge and/rankillerinde index mays into reasons care but aday to be pre- dependent care account of statuture accounts of a statute of the	VIZIDAL ORDIT DEPARADEL VIZIDAL		Description
Sedge with a set for with when a stary of a linear work of a fingle set of a opport out according constant and according constant of the final set of a	ATTENT (REAL PRAYER), ATTENT		Description Products and sales.
longen mitter en formale minister en segure son de la del de la segure de la segu	AZZERAL, GERELDERANDEL AZERAL,		Description Products and sales.
lenge nadra minima nina mina mina mina ka nina wa nin ha king tale pin oromti olar nagrana " attattan nagetha", attatta nina ka Ny Andrawiti	ATTENT, ATTENT PARAMATIN, ATTENT,		Description Products and sales.
tanga mataran darawa kasa mayo kata na kasa kasa kasa kasa kasa kasa kasa	anner annersensensensensensensensensensensensensens		Description Products and sales.
balg adara taraka dala mayuka kua kua kua bala dagi ta bala aperti dari kuataka Mg Andra ta Jamanika	ren a		Description Products and sales.  Icon Add files
haliga salawa hanaka aka arang untu manua sa sakuha aka pa sa ya agunta casi kasanan a satatu kaganar aka paka sakuha saku Ng Anstanan Jagranahan (Jana aka aka aka aka aka aka aka aka aka	алан аралын аралын таралы Баш аралын таралын Алан Алан Алан аралын Алан Алан Алан аралын	2 m	Description Products and sales. Kon Add files
tingko selava stransko skola zrazu u star sena se se obskola dog stra se osoba dog stra se osoba dog stra se os defetti kala Adattari i stata star se osoba dog stra se osoba dog stra se osoba dog stra se osoba dog stra se os desentare taga e osoba dog se osoba dog	5000 5200-7 (0000 540-0400 + 5200-7 5400	⊠ел: Хатиснт ⊙ацириона	Description Products and sales. Kon Add files
todge undersentervelse einer enzy ist an versue an ande after stegens der efficielt daef Australier instantisk degtiver instantisk einer einer High dadsamte: Der versiteren Sande in versue Sandersteren Sande instantisk einer stegens Fachaltigkanse, Frankrikke, Frankrikkense, Frankrikke, Frankrikkense, Frankrikke,	5000	문 en: 전 Shirteart © Autoriest 전 Shi	Description Products and sales. Icon Add files
ningi ushum homan kwa engu utu munu anaka du bu su pan agama kwantan susarana susara kwantan suka du bu Keji dadatata Manakatwa kwantan susara kwanta Manakatwa kwa kwantan Kabunga kwantan Kabunga kwantan	1         2000 *         1000 *         1000 *           1         2000 *         2000 *         1000 *           1         2000 *         1000 *         1000 *           1         2000 *         1000 *         1000 *           1         2000 *         1000 *         1000 *           1         2000 *         1000 *         1000 *           1         2000 *         1000 *         1000 *	문 ar: 또 Driver 은 Autoriou 문 Driver 은 Autoriou	Description Products and sales.  Icon  Add files  Calcon

OK. At this point, the next step is to add files.

For that, we must load a file from the dataset by selecting Add files...

It is important to make sure that the CSV file follows the descriptions and metadata added earlier.

We load our file.

Finally, to create the assistant, we select Save.

A window is displayed to enter a Name, Description and Icon if desired.

We click on Confirm... ...and go back to the page showing the current status of the assistant.

This status is related to the loading process of the datasets with the metadata and glossary information.

The possible statuses are FAILED, COMPLETED, or PROCESSING (indicating the progress percentage).

GeneXus by Globant

# **Data Analyst Assistant: Update**

Assistant Details
Assistant: MySales
Assistant Id
ce96d69a-6a89-4215-9d32-09d851b21d7b
Name *
Description
Products and sales.
Status Enabled
KON
+ Add files
CONFIRM

Once the loading status is completed, we can select Update and see the version identifier with which it was saved, change the name and description, set it as enabled or disabled, or add an icon.

#### **Data Analyst Assistant: Edition**

General	External information	
Assistant	Name	
MySale	S	
Update n	node	
Extend	datasets	~
Extend	datasets	
P Updat	e prompts	
Clean	data	
-		

It is also possible to edit it, to extend or replace the dataset, or to update the prompt.

Expanding the dataset expands the information available without affecting the original structure of the dataset.

The files selected for integration must meet the same initial conditions, such as having the same number of columns and file names.

The files used for the extension must not contain the same rows that are already loaded in the assistant, as this could generate duplicate keys and cause problems when running the queries.

It should also be noted that when increasing the dataset it is not possible to make changes to the metadata or the glossary. This function focuses only on increasing existing datasets without affecting other aspects of the assistant.

Replacing the dataset, on the other hand, allows you to completely replace the data in the existing dataset, as well as modify the metadata and associated glossary.

This is useful when you need to update the information completely or make significant changes to the structure and description of the dataset.

# Data Analyst Assistant: Update prompt

Assista	int Editor	
General	External information	
MySale	es	
Update	mode	
Copua	e prompts	
	Id Prompt name	
	1 condenser_prompt_template	P UPDATE
	2 classifier_prompt_template	P UPDATE
	3 classifier_reflexion_prompt_template	🖋 UPDATE
	4 code_prompt_template	📌 UPDATE
	5 code_considerations	€ UPDATE
	6 interpreter_prompt_template	& UPDATE
	7 interpreter_considerations	JPDATE
	8 interpreter_considerations_verbosity_1	✓ UPDATE
	9 interpreter_considerations_verbosity_2	♥ UPDATE
	10 interpreter_prompt_empty_template	
	11. Internetter moment lof template	P OF DATE

If Update prompt is selected, a list of the 28 default prompts will be displayed.

It is possible to query and update each of the prompts individually by clicking on Update.

This requires the Administrator, ProjectRole or OrganizationRole role.

It is important to note that these updates are made from the list of prompts and not at the general level of the assistant.

When generating a new version of the assistant and making changes with the Extend Datasets or Replace Datasets options, changes made to the prompts in previous versions of the assistant will be ignored.

As usual, we can finally test the created assistant using the Playground option in the Backoffice menu.



training.genexus.com