Dataiku stories with Generative AI for alfonso.cervantes

Version	Author	Date
1.0	alfonso.cervantes	2025-10-03 10:58:39

Table of contents

I. Project general description

Project references for the created flow:

- Project name: **Dataiku stories with Generative AI for alfonso.cervantes**
- Project key: **TUT_STORIES_GENAI**
- Status: **Sandbox**
- Current git branch: **master**
- Tags: **starter, academy, downloads, tutorial**
- Created by **isabelle.godio (deleted)** on the 2023-04-21 07:56:10
- Last Modification by **alfonso.cervantes** on the 2025-09-26 16:19:11

II. Graphical Representation of the Flow
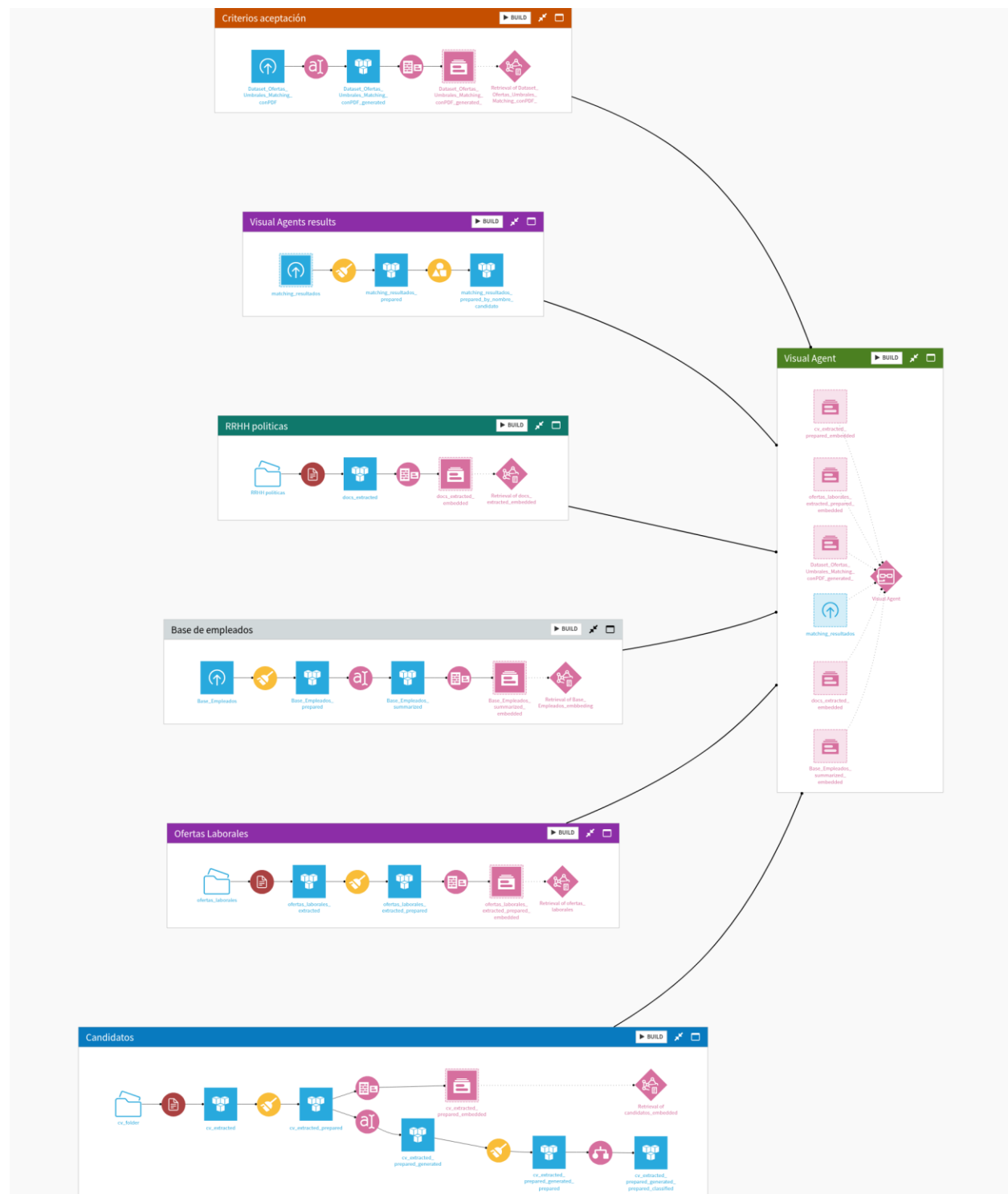
III. Datasets

List all the datasets in the flow.

1. cv_extracted

Summarized information about your dataset cv_extracted:

a) Status of the dataset cv_extracted

Characteristics of the dataset cv_extracted:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **7.3 KB**
- Record Count: **5**
- Created by **alfonso.cervantes** on the 2025-09-25 06:50:57
- Last Modification by **alfonso.cervantes** on the 2025-09-25 06:51:57

b) Related Recipes

The dataset cv_extracted relates to the following recipes.

Parent Recipe

List of recipes with cv_extracted as an output:
- compute_cv_extracted *(Text extraction)*

Successor Recipes

List of recipes with cv_extracted as an input:
- compute_cv_extracted_prepared *(Prepare)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
text	string	
error_message	string	

2. cv_extracted_prepared

Summarized information about your dataset cv_extracted_prepared:

a) Status of the dataset cv_extracted_prepared

Characteristics of the dataset cv_extracted_prepared:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **8.1 KB**
- Record Count: **5**
- Created by **alfonso.cervantes** on the 2025-09-25 17:29:28
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:29:49

b) Related Recipes

The dataset cv_extracted_prepared relates to the following recipes.

Parent Recipe

List of recipes with cv_extracted_prepared as an output:
- compute_cv_extracted_prepared *(Prepare)*

Successor Recipes

List of recipes with cv_extracted_prepared as an input:
- compute_cv_extracted_prepared_embedded *(embed dataset)*
- compute_cv_extracted_prepared_generated *(prompt)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
text	string	
error_message	string	
metadata	string	

3.   cv_extracted_prepared_generated

Summarized information about your dataset cv_extracted_prepared_generated:

a)   Status of the dataset cv_extracted_prepared_generated

Characteristics of the dataset cv_extracted_prepared_generated:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **20.6 KB**
- Record Count: **5**
- Created by **alfonso.cervantes** on the 2025-09-26 05:09:15
- Last Modification by **alfonso.cervantes** on the 2025-09-26 05:09:59

b)   Related Recipes

The dataset cv_extracted_prepared_generated relates to the following recipes.

Parent Recipe

List of recipes with cv_extracted_prepared_generated as an output:
- compute_cv_extracted_prepared_generated *(prompt)*

Successor Recipes

List of recipes with cv_extracted_prepared_generated as an input:
- compute_cv_extracted_prepared_generated_prepared *(Prepare)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
text	string	
error_message	string	
metadata	string	
llm_output	string	
llm_validation_status	string	
llm_raw_response	string	
llm_error_message	string	
llm_raw_query	string	

4.   cv_extracted_prepared_generated_prepared

Summarized information about your dataset cv_extracted_prepared_generated_prepared:

a)   Status of the dataset cv_extracted_prepared_generated_prepared

Characteristics of the dataset cv_extracted_prepared_generated_prepared:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **6.7 KB**
- Record Count: **5**
- Created by **alfonso.cervantes** on the 2025-09-26 05:11:45
- Last Modification by **alfonso.cervantes** on the 2025-09-26 06:40:58

b)   Related Recipes

The dataset cv_extracted_prepared_generated_prepared relates to the following recipes.

Parent Recipe

List of recipes with cv_extracted_prepared_generated_prepared as an output:
- compute_cv_extracted_prepared_generated_prepared *(Prepare)*

Successor Recipes

List of recipes with cv_extracted_prepared_generated_prepared as an input:
- compute_cv_extracted_prepared_generated_prepared_classified *(classification with model-defined classes)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
nombre	string	
telefono	string	
linked	string	
github	string	
email	string	
resumen	string	
skills	string	
proyectos	string	
certificaciones	string	

5. cv_extracted_prepared_generated_prepared_classified

Summarized information about your dataset
cv_extracted_prepared_generated_prepared_classified:

a) Status of the dataset cv_extracted_prepared_generated_prepared_classified

Characteristics of the dataset cv_extracted_prepared_generated_prepared_classified:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **6.8 KB**
- Record Count: **5**
- Created by **alfonso.cervantes** on the 2025-09-26 06:42:49
- Last Modification by **alfonso.cervantes** on the 2025-09-26 06:42:49

b)   Related Recipes

The dataset cv_extracted_prepared_generated_prepared_classified relates to the following recipes.

Parent Recipe

List of recipes with cv_extracted_prepared_generated_prepared_classified as an output:
- compute_cv_extracted_prepared_generated_prepared_classified *(classification with model-defined classes)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
nombre	string	
telefono	string	
linked	string	
github	string	
email	string	
resumen	string	
skills	string	
proyectos	string	
certificaciones	string	
prediction	string	

prediction_score	double	
llm_raw_response	string	
llm_error_message	string	

6.   Base_Empleados

Summarized information about your dataset Base_Empleados:

a)   Status of the dataset Base_Empleados

Characteristics of the dataset Base_Empleados:
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **alfonso.cervantes** on the 2025-09-25 14:15:53
- Last Modification by **alfonso.cervantes** on the 2025-09-25 14:15:53

b)   Related Recipes

The dataset Base_Empleados relates to the following recipes.

Successor Recipes

List of recipes with Base_Empleados as an input:
- compute_Base_Empleados_prepared *(Prepare)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
Nombre	string	
Email	string	
Teléfono	string	
Rol	string	

Salario (USD)	string	
Jefe Inmediato	string	
Departamento	string	
Ubicación	string	
Ciudad	string	
País	string	

7. Base_Empleados_prepared

Summarized information about your dataset Base_Empleados_prepared:

a) Status of the dataset Base_Empleados_prepared

Characteristics of the dataset Base_Empleados_prepared:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **5.6 KB**
- Record Count: **50**
- Created by **alfonso.cervantes** on the 2025-09-25 14:16:31
- Last Modification by **alfonso.cervantes** on the 2025-09-25 14:16:31

b) Related Recipes

The dataset Base_Empleados_prepared relates to the following recipes.

Parent Recipe

List of recipes with Base_Empleados_prepared as an output:
- compute_Base_Empleados_prepared *(Prepare)*

Successor Recipes

List of recipes with Base_Empleados_prepared as an input:
- compute_Base_Empleados_summarized *(prompt)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Nombre	string	
Email	string	
Teléfono	double	
Rol	string	
Salario (USD)	bigint	
Jefe Inmediato	string	
Departamento	string	
Ubicación	string	
Ciudad	string	
País	string	

8.   Base_Empleados_summarized

Summarized information about your dataset Base_Empleados_summarized:

a)   Status of the dataset Base_Empleados_summarized

Characteristics of the dataset Base_Empleados_summarized:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **23.4 KB**
- Record Count: **50**
- Created by **alfonso.cervantes** on the 2025-09-25 14:19:45
- Last Modification by **alfonso.cervantes** on the 2025-09-25 14:22:44

b)   Related Recipes

The dataset Base_Empleados_summarized relates to the following recipes.

Parent Recipe

List of recipes with Base_Empleados_summarized as an output:
- compute_Base_Empleados_summarized *(prompt)*

Successor Recipes

List of recipes with Base_Empleados_summarized as an input:
- compute_Base_Empleados_summarized_embedded *(embed dataset)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
Nombre	string	
Email	string	
Teléfono	double	
Rol	string	
Salario (USD)	bigint	
Jefe Inmediato	string	
Departamento	string	
Ubicación	string	
Ciudad	string	
País	string	
llm_output	string	
llm_validation_status	string	
llm_raw_response	string	
llm_error_message	string	
llm_raw_query	string	

9.   docs_extracted

Summarized information about your dataset docs_extracted:

a)   Status of the dataset docs_extracted

Characteristics of the dataset docs_extracted:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **7.4 KB**
- Record Count: **5**
- Created by **alfonso.cervantes** on the 2025-09-25 05:06:42
- Last Modification by **alfonso.cervantes** on the 2025-09-25 05:08:00

b)   Related Recipes

The dataset docs_extracted relates to the following recipes.

Parent Recipe

List of recipes with docs_extracted as an output:
- compute_docs_extracted *(Text extraction)*

Successor Recipes

List of recipes with docs_extracted as an input:
- compute_docs_extracted_embedded *(embed dataset)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
text	string	
chunk_id	bigint	
metadata	string	

error_message	string	

10. ofertas_laborales_extracted

Summarized information about your dataset ofertas_laborales_extracted:

a)   Status of the dataset ofertas_laborales_extracted

Characteristics of the dataset ofertas_laborales_extracted:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **9.8 KB**
- Record Count: **4**
- Created by **alfonso.cervantes** on the 2025-09-25 16:58:11
- Last Modification by **alfonso.cervantes** on the 2025-09-25 16:59:52

b)   Related Recipes

The dataset ofertas_laborales_extracted relates to the following recipes.

Parent Recipe

List of recipes with ofertas_laborales_extracted as an output:
- compute_ofertas_laborales_extracted *(Text extraction)*

Successor Recipes

List of recipes with ofertas_laborales_extracted as an input:
- compute_ofertas_laborales_extracted_prepared *(Prepare)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
text	string	

error_message	string	

11.   ofertas_laborales_extracted_prepared

Summarized information about your dataset ofertas_laborales_extracted_prepared:

a)   Status of the dataset ofertas_laborales_extracted_prepared

Characteristics of the dataset ofertas_laborales_extracted_prepared:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **10.4 KB**
- Record Count: **4**
- Created by **alfonso.cervantes** on the 2025-09-25 17:13:43
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:14:07

b)   Related Recipes

The dataset ofertas_laborales_extracted_prepared relates to the following recipes.

Parent Recipe

List of recipes with ofertas_laborales_extracted_prepared as an output:
- compute_ofertas_laborales_extracted_prepared *(Prepare)*

Successor Recipes

List of recipes with ofertas_laborales_extracted_prepared as an input:
- compute_ofertas_laborales_extracted_prepared_embedded *(embed dataset)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
file	string	
text	string	

error_message	string	
metadata	string	

12. Dataset_Ofertas_Umbrales_Matching_conPDF

Summarized information about your dataset Dataset_Ofertas_Umbrales_Matching_conPDF:

a) Status of the dataset Dataset_Ofertas_Umbrales_Matching_conPDF

Characteristics of the dataset Dataset_Ofertas_Umbrales_Matching_conPDF:
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **20**
- Created by **alfonso.cervantes** on the 2025-09-25 16:57:33
- Last Modification by **alfonso.cervantes** on the 2025-09-25 16:57:34

b) Related Recipes

The dataset Dataset_Ofertas_Umbrales_Matching_conPDF relates to the following recipes.

Successor Recipes

List of recipes with Dataset_Ofertas_Umbrales_Matching_conPDF as an input:
- compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated *(prompt)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Oferta	string	
Categoria	string	
Definicion	string	
MinMatchPct	string	
Archivo_PDF	string	

13. Dataset_Ofertas_Umbrales_Matching_conPDF_generated

Summarized information about your dataset
Dataset_Ofertas_Umbrales_Matching_conPDF_generated:

a) Status of the dataset Dataset_Ofertas_Umbrales_Matching_conPDF_generated

Characteristics of the dataset Dataset_Ofertas_Umbrales_Matching_conPDF_generated:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **15.6 KB**
- Record Count: **20**
- Created by **alfonso.cervantes** on the 2025-09-25 17:42:32
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:44:08

b) Related Recipes

The dataset Dataset_Ofertas_Umbrales_Matching_conPDF_generated relates to the following recipes.

Parent Recipe

List of recipes with Dataset_Ofertas_Umbrales_Matching_conPDF_generated as an output:
- compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated *(prompt)*

Successor Recipes

List of recipes with Dataset_Ofertas_Umbrales_Matching_conPDF_generated as an input:
- compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated_embedded *(embed dataset)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Oferta	string	
Categoria	string	

Definicion	string	
MinMatchPct	string	
Archivo_PDF	string	
llm_output	string	
llm_validation_status	string	
llm_raw_response	string	
llm_error_message	string	
llm_raw_query	string	

14. matching_resultados

Summarized information about your dataset matching_resultados:

a) Status of the dataset matching_resultados

Characteristics of the dataset matching_resultados:
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **19**
- Created by **alfonso.cervantes** on the 2025-09-26 01:52:59
- Last Modification by **alfonso.cervantes** on the 2025-09-26 03:22:35

b) Related Recipes

The dataset matching_resultados relates to the following recipes.

Successor Recipes

List of recipes with matching_resultados as an input:
- compute_matching_resultados_prepared *(Prepare)*

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
nombre_candidato	string	
oferta_laboral	string	
puntaje_matching	int	
pasa_o_no	string	

15.   matching_resultados_prepared

Summarized information about your dataset matching_resultados_prepared:

a)   Status of the dataset matching_resultados_prepared

Characteristics of the dataset matching_resultados_prepared:
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **1.6 KB**
- Record Count: **16**
- Created by **alfonso.cervantes** on the 2025-09-26 03:24:29
- Last Modification by **alfonso.cervantes** on the 2025-09-26 03:28:10

b)   Related Recipes

The dataset matching_resultados_prepared relates to the following recipes.

Parent Recipe

List of recipes with matching_resultados_prepared as an output:
- compute_matching_resultados_prepared *(Prepare)*

Successor Recipes

List of recipes with matching_resultados_prepared as an input:
- compute_matching_resultados_prepared_by_nombre_candidato *(Group)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
nombre_candidato	string	
oferta_laboral	string	
puntaje_matching	int	
pasa_o_no	string	

16.   matching_resultados_prepared_by_nombre_candidato

Summarized information about your dataset
matching_resultados_prepared_by_nombre_candidato:

a)   Status of the dataset matching_resultados_prepared_by_nombre_candidato

Characteristics of the dataset matching_resultados_prepared_by_nombre_candidato:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **2 KB**
- Record Count: **7**
- Created by **alfonso.cervantes** on the 2025-09-26 03:32:17
- Last Modification by **alfonso.cervantes** on the 2025-09-26 03:32:32

b)   Related Recipes

The dataset matching_resultados_prepared_by_nombre_candidato relates to the following recipes.

Parent Recipe

List of recipes with matching_resultados_prepared_by_nombre_candidato as an output:
- compute_matching_resultados_prepared_by_nombre_candidato *(Group)*

c)   Schema

Schema details of the dataset.

Column Name	Column Type	Description
nombre_candidato	string	
oferta_laboral_first	string	
puntaje_matching_max	int	
count	bigint	

IV. Managed folders

List all the managed folders in the flow.

1. cv_folder

Summarized information about the managed folder cv_folder:

a) Status

Characteristics of the managed folder cv_folder:
- full_id: **TUT_STORIES_GENAI.IZj2rTsT**
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Created by **alfonso.cervantes** on the 2025-09-25 06:50:31
- Last Modification by **alfonso.cervantes** on the 2025-09-25 06:50:31

b) Related Recipes

List of the recipes related to the managed folder.

Successor Recipes

List of recipes with cv_folder as an input:
- compute_cv_extracted *(Text extraction)*

2. RRHH politicas

Summarized information about the managed folder RRHH politicas:

a) Status

Characteristics of the managed folder RRHH politicas:
- full_id: **TUT_STORIES_GENAI.Q1YtLaWF**
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**

- Created by **alfonso.cervantes** on the 2025-09-25 05:02:36
- Last Modification by **alfonso.cervantes** on the 2025-09-25 13:59:13

b)   Related Recipes

List of the recipes related to the managed folder.

Successor Recipes

List of recipes with RRHH politicas as an input:
- compute_docs_extracted *(Text extraction)*

3.   ofertas_laborales

Summarized information about the managed folder ofertas_laborales:

a)   Status

Characteristics of the managed folder ofertas_laborales:
- full_id: **TUT_STORIES_GENAI.Jme0ehop**
- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Created by **alfonso.cervantes** on the 2025-09-25 16:24:06
- Last Modification by **alfonso.cervantes** on the 2025-09-25 16:24:06

b)   Related Recipes

List of the recipes related to the managed folder.

Successor Recipes

List of recipes with ofertas_laborales as an input:
- compute_ofertas_laborales_extracted *(Text extraction)*

V. Recipes

List all the recipes in the flow.

1.   compute_cv_extracted

Summarized information about the recipe compute_cv_extracted.

a)   Status

Characteristics of the recipe compute_cv_extracted:
- Type: **Text extraction**
- Created by **alfonso.cervantes** on the 2025-09-25 06:50:59
- Last Modification by **alfonso.cervantes** on the 2025-09-25 06:51:07

b)   Inputs / Outputs

List of the inputs of the recipe:
- Folder cv_folder *(Amazon S3)*

List of the outputs of the recipe:
- Dataset cv_extracted *(Amazon S3)*

2.   compute_cv_extracted_prepared

Summarized information about the recipe compute_cv_extracted_prepared.

a)   Status

Characteristics of the recipe compute_cv_extracted_prepared:
- Type: **Prepare**
- Created by **alfonso.cervantes** on the 2025-09-25 17:29:28

- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:29:50

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset cv_extracted *(Amazon S3)*

List of the outputs of the recipe:
- Dataset cv_extracted_prepared *(Amazon S3)*

c)   Prepare Steps

Type	Comment
CreateColumnWithGREL	

3.   compute_cv_extracted_prepared_generated

Summarized information about the recipe compute_cv_extracted_prepared_generated.

a)   Status

Characteristics of the recipe compute_cv_extracted_prepared_generated:
- Type: **prompt**
- Created by **alfonso.cervantes** on the 2025-09-26 05:09:15
- Last Modification by **alfonso.cervantes** on the 2025-09-26 05:22:50

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset cv_extracted_prepared *(Amazon S3)*

List of the outputs of the recipe:
- Dataset cv_extracted_prepared_generated *(Amazon S3)*

4. compute_cv_extracted_prepared_generated_prepared

Summarized information about the recipe
compute_cv_extracted_prepared_generated_prepared.

a)   Status

Characteristics of the recipe compute_cv_extracted_prepared_generated_prepared:
- Type: **Prepare**
- Created by **alfonso.cervantes** on the 2025-09-26 05:11:45
- Last Modification by **alfonso.cervantes** on the 2025-09-26 06:40:59

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset cv_extracted_prepared_generated *(Amazon S3)*

List of the outputs of the recipe:
- Dataset cv_extracted_prepared_generated_prepared *(Amazon S3)*

c)   Prepare Steps

Type	Comment
ColumnSplitter	
ColumnRenamer	
ColumnsSelector	

5. compute_cv_extracted_prepared_generated_prepared_classified

Summarized information about the recipe
compute_cv_extracted_prepared_generated_prepared_classified.

a)   Status

Characteristics of the recipe compute_cv_extracted_prepared_generated_prepared_classified:
- Type: **classification with model-defined classes**

- Created by **alfonso.cervantes** on the 2025-09-26 06:42:49
- Last Modification by **alfonso.cervantes** on the 2025-09-26 06:43:15

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset cv_extracted_prepared_generated_prepared *(Amazon S3)*

List of the outputs of the recipe:
- Dataset cv_extracted_prepared_generated_prepared_classified *(Amazon S3)*

6.   compute_cv_extracted_prepared_embedded

Summarized information about the recipe compute_cv_extracted_prepared_embedded.

a)   Status

Characteristics of the recipe compute_cv_extracted_prepared_embedded:
- Type: **embed dataset**
- Created by **alfonso.cervantes** on the 2025-09-25 17:31:07
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:31:16

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset cv_extracted_prepared *(Amazon S3)*

7.   compute_Base_Empleados_prepared

Summarized information about the recipe compute_Base_Empleados_prepared.

a) Status

Characteristics of the recipe compute_Base_Empleados_prepared:
- Type: **Prepare**
- Created by **alfonso.cervantes** on the 2025-09-25 14:16:31
- Last Modification by **alfonso.cervantes** on the 2025-09-25 16:22:12

b) Inputs / Outputs

List of the inputs of the recipe:
- Dataset Base_Empleados *(Uploaded Files)*

List of the outputs of the recipe:
- Dataset Base_Empleados_prepared *(Amazon S3)*

c) Prepare Steps

Type	Comment
StringTransformer	

8. compute_Base_Empleados_summarized

Summarized information about the recipe compute_Base_Empleados_summarized.

a) Status

Characteristics of the recipe compute_Base_Empleados_summarized:
- Type: **prompt**
- Created by **alfonso.cervantes** on the 2025-09-25 14:19:45
- Last Modification by **alfonso.cervantes** on the 2025-09-25 14:22:47

b) Inputs / Outputs

List of the inputs of the recipe:
- Dataset Base_Empleados_prepared *(Amazon S3)*

List of the outputs of the recipe:
- Dataset Base_Empleados_summarized *(Amazon S3)*

9.   compute_Base_Empleados_summarized_embedded

Summarized information about the recipe compute_Base_Empleados_summarized_embedded.

a)   Status

Characteristics of the recipe compute_Base_Empleados_summarized_embedded:
- Type: **embed dataset**
- Created by **alfonso.cervantes** on the 2025-09-26 04:32:41
- Last Modification by **alfonso.cervantes** on the 2025-09-26 04:33:33

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset Base_Empleados_summarized *(Amazon S3)*

10.   compute_docs_extracted

Summarized information about the recipe compute_docs_extracted.

a)   Status

Characteristics of the recipe compute_docs_extracted:
- Type: **Text extraction**
- Created by **alfonso.cervantes** on the 2025-09-25 05:06:44
- Last Modification by **alfonso.cervantes** on the 2025-09-25 05:06:58

b)   Inputs / Outputs

List of the inputs of the recipe:
- Folder RRHH politicas *(Amazon S3)*

List of the outputs of the recipe:
- Dataset docs_extracted *(Amazon S3)*

11.   compute_docs_extracted_embedded

Summarized information about the recipe compute_docs_extracted_embedded.

a)   Status

Characteristics of the recipe compute_docs_extracted_embedded:
- Type: **embed dataset**
- Created by **alfonso.cervantes** on the 2025-09-25 17:26:59
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:27:27

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset docs_extracted *(Amazon S3)*

12.   compute_ofertas_laborales_extracted

Summarized information about the recipe compute_ofertas_laborales_extracted.

a)   Status

Characteristics of the recipe compute_ofertas_laborales_extracted:
- Type: **Text extraction**

- Created by **alfonso.cervantes** on the 2025-09-25 16:58:13
- Last Modification by **alfonso.cervantes** on the 2025-09-25 16:58:26

b)   Inputs / Outputs

List of the inputs of the recipe:
- Folder ofertas_laborales *(Amazon S3)*

List of the outputs of the recipe:
- Dataset ofertas_laborales_extracted *(Amazon S3)*

13.   compute_ofertas_laborales_extracted_prepared

Summarized information about the recipe compute_ofertas_laborales_extracted_prepared.

a)   Status

Characteristics of the recipe compute_ofertas_laborales_extracted_prepared:
- Type: **Prepare**
- Created by **alfonso.cervantes** on the 2025-09-25 17:13:43
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:14:07

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset ofertas_laborales_extracted *(Amazon S3)*

List of the outputs of the recipe:
- Dataset ofertas_laborales_extracted_prepared *(Amazon S3)*

c)   Prepare Steps

Type	Comment
CreateColumnWithGREL	

14.   compute_ofertas_laborales_extracted_prepared_embedded

Summarized information about the recipe
compute_ofertas_laborales_extracted_prepared_embedded.

a)   Status

Characteristics of the recipe compute_ofertas_laborales_extracted_prepared_embedded:
- Type: **embed dataset**
- Created by **alfonso.cervantes** on the 2025-09-25 17:18:28
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:19:10

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset ofertas_laborales_extracted_prepared *(Amazon S3)*

15.   compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated

Summarized information about the recipe
compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated.

a)   Status

Characteristics of the recipe
compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated:
- Type: **prompt**
- Created by **alfonso.cervantes** on the 2025-09-25 17:42:32
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:44:49

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset Dataset_Ofertas_Umbrales_Matching_conPDF *(Uploaded Files)*

List of the outputs of the recipe:
- Dataset Dataset_Ofertas_Umbrales_Matching_conPDF_generated *(Amazon S3)*

16.   compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated_embedded

Summarized information about the recipe compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated_embedded.

a)   Status

Characteristics of the recipe compute_Dataset_Ofertas_Umbrales_Matching_conPDF_generated_embedded:
- Type: **embed dataset**
- Created by **alfonso.cervantes** on the 2025-09-25 17:46:14
- Last Modification by **alfonso.cervantes** on the 2025-09-25 17:46:49

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset Dataset_Ofertas_Umbrales_Matching_conPDF_generated *(Amazon S3)*

17.   compute_matching_resultados_prepared

Summarized information about the recipe compute_matching_resultados_prepared.

a)   Status

Characteristics of the recipe compute_matching_resultados_prepared:
- Type: **Prepare**
- Created by **alfonso.cervantes** on the 2025-09-26 03:24:29
- Last Modification by **alfonso.cervantes** on the 2025-09-26 03:31:49

b)   Inputs / Outputs

List of the inputs of the recipe:
- Dataset matching_resultados *(Uploaded Files)*

List of the outputs of the recipe:
- Dataset matching_resultados_prepared *(Amazon S3)*

c)   Prepare Steps

Type	Comment
RemoveRowsOnEmpty	

18.   compute_matching_resultados_prepared_by_nombre_candidato

Summarized information about the recipe
compute_matching_resultados_prepared_by_nombre_candidato.

a)   Status

Characteristics of the recipe compute_matching_resultados_prepared_by_nombre_candidato:
- Type: **Group**
- Created by **alfonso.cervantes** on the 2025-09-26 03:32:17
- Last Modification by **alfonso.cervantes** on the 2025-09-26 03:32:32

b)   Inputs / Outputs

List of the inputs of the recipe:

- Dataset matching_resultados_prepared *(Amazon S3)*

List of the outputs of the recipe:
- Dataset matching_resultados_prepared_by_nombre_candidato *(Amazon S3)*

Group keys: nombre_candidato

VI. Saved Models

List all the saved models in the flow.

1.   Retrieval of ofertas_laborales

Summarized information about the saved model Retrieval of ofertas_laborales:.

a)   Status

Model id: **VDYXmKwC**
Active version:

b)   Related Recipes

List of the recipes related to the saved model.

2.   Retrieval of Base_Empleados_embbeding

Summarized information about the saved model Retrieval of Base_Empleados_embbeding:.

a)   Status

Model id: **VwkvR0ZW**
Active version:

b)   Related Recipes

List of the recipes related to the saved model.

3.   Visual Agent

Summarized information about the saved model Visual Agent:.

a)   Status

Model id: **A58gpiyb**

Active version:

b)   Related Recipes

List of the recipes related to the saved model.

4.   Retrieval of candidatos_embedded

Summarized information about the saved model Retrieval of candidatos_embedded:.

a)   Status

Model id: **EqaYHW6i**
Active version:

b)   Related Recipes

List of the recipes related to the saved model.

5.   Retrieval of Dataset_Ofertas_Umbrales_Matching_conPDF_generated_embedded

Summarized information about the saved model Retrieval of
Dataset_Ofertas_Umbrales_Matching_conPDF_generated_embedded:.

a)   Status

Model id: **pClK9GW2**
Active version:

b)   Related Recipes

List of the recipes related to the saved model.

6.   Retrieval of docs_extracted_embedded

Summarized information about the saved model Retrieval of docs_extracted_embedded:.

a) Status

Model id: **2qGhwwoS**
Active version:

b) Related Recipes

List of the recipes related to the saved model.

VII. Deployment and Monitoring

A. Implementation Details

- The flow can be found here: https://dss-39049e88-af3b9684-dku.us-east-1.app.dataiku.io/projects/TUT_STORIES_GENAI/
- The name of the generated file is: Dataiku Flow Documentation - TUT_STORIES_GENAI.docx

B. Version Control

- The flow was executed with the following version of DSS: 14.1.0