

Project_Online_Retailer



Version	Author	Date
1.0	yunting.lu	2023-04-27 09:06:08

Table of contents

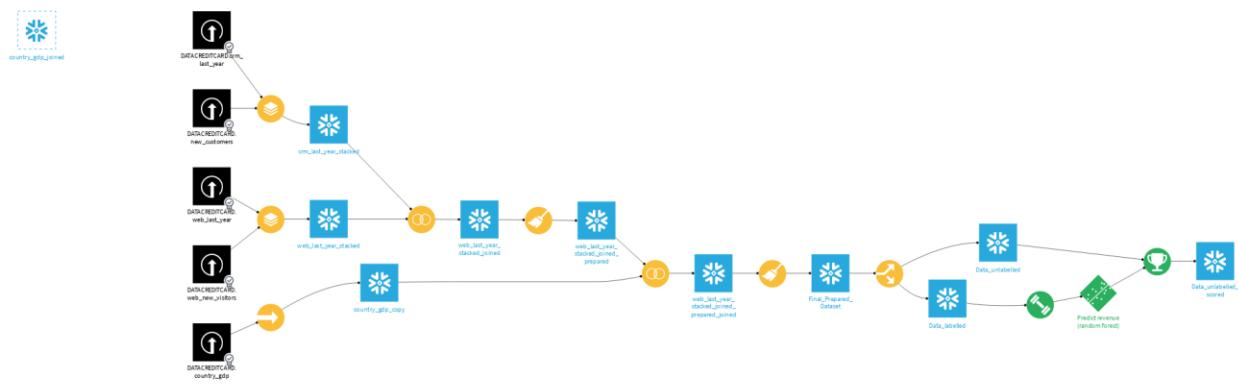
I.	Project general description	4
II.	Graphical Representation of the Flow	5
III.	Datasets	6
1.	{{\\$dataset.name}}	Error! Bookmark not defined.
a)	Status of the dataset {{\\$dataset.name}}	Error! Bookmark not defined.
b)	Related Recipes	23
c)	Schema	23
IV.	Managed folders	Error! Bookmark not defined.
1.	{{\\$folder.name}}	Error! Bookmark not defined.
a)	Status	Error! Bookmark not defined.
b)	Derived datasets	Error! Bookmark not defined.
c)	Related Recipes	Error! Bookmark not defined.
V.	Recipes	25
1.	{{ \\$recipe.name }}	Error! Bookmark not defined.
a)	Status	31
b)	Inputs / Outputs	31
c)	Prepare Steps	30
d)	Join	29
e)	Code	Error! Bookmark not defined.
f)	Windows	Error! Bookmark not defined.
g)	Sorted Columns	Error! Bookmark not defined.
h)	Top rows and sorting direction	Error! Bookmark not defined.
i)	Parameters	Error! Bookmark not defined.
VI.	Saved Models	32
1.	{{ \$model.name }}	Error! Bookmark not defined.
a)	Status	32
b)	Related Recipes	32
VII.	Model Evaluation Stores	Error! Bookmark not defined.
1.	{{ \$mes.name }}	Error! Bookmark not defined.
a)	Related Recipes	Error! Bookmark not defined.
VIII.	Labeling Tasks	Error! Bookmark not defined.
1.	{{ \$labeling_task.name }}	Error! Bookmark not defined.
a)	Status	Error! Bookmark not defined.
b)	Inputs / Outputs	Error! Bookmark not defined.
IX.	Deployment and Monitoring	33
A.	Implementation Details	33
B.	Version Control	33

I. Project general description

Project references for the created flow:

- Project name: **Project_Online_Retailer**
- Project key: **POR**
- Project short description:
*The project *Project_Online_Retailer* was created by stefanos.lambropoulos on Apr 20th 2023*
- Status: **Sandbox**
- Current git branch: **master**
- Created by **stefanos.lambropoulos** on the 2023-04-20 11:55:36
- Last Modification by **yunting.lu** on the 2023-04-20 14:42:13

II. Graphical Representation of the Flow



III. Datasets

List all the datasets in the flow.

1. web_last_year

Summarized information about your dataset web_last_year:

- This dataset has not been created in this project. Its origin project is **data credit card**

a) Status of the dataset web_last_year

Characteristics of the dataset web_last_year:

- Source project name: **data credit card**
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **maxime.capron** on the 2023-04-19 19:31:09
- Last Modification by **maxime.capron** on the 2023-04-19 19:32:00

b) Related Recipes

The dataset web_last_year relates to the following recipes.

Successor Recipes

List of recipes with web_last_year as an input:

- **compute_web_last_year_stacked** (*Stack*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	

ip	string	
pages_visited	string	
campain	string	

2. web_new_visitors

Summarized information about your dataset web_new_visitors:

- This dataset has not been created in this project. Its origin project is **data credit card**

a) Status of the dataset web_new_visitors

Characteristics of the dataset web_new_visitors:

- Source project name: **data credit card**
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **4,600**
- Created by **maxime.capron** on the 2023-04-19 19:31:06
- Last Modification by **maxime.capron** on the 2023-04-19 19:31:51

b) Related Recipes

The dataset web_new_visitors relates to the following recipes.

Successor Recipes

List of recipes with web_new_visitors as an input:

- compute_web_last_year_stacked (*Stack*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	

pages_visited	string	
campain	string	

3. web_last_year_stacked

Summarized information about your dataset web_last_year_stacked:

a) Status of the dataset web_last_year_stacked

Characteristics of the dataset web_last_year_stacked:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:03:37
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:03:37

b) Related Recipes

The dataset web_last_year_stacked relates to the following recipes.

Parent Recipe

List of recipes with web_last_year_stacked as an output:

- compute_web_last_year_stacked (*Stack*)

Successor Recipes

List of recipes with web_last_year_stacked as an input:

- compute_web_last_year_stacked_joined (*Join*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	

ip	string	
pages_visited	string	
campain	string	

4. crm_last_year

Summarized information about your dataset crm_last_year:

- This dataset has not been created in this project. Its origin project is **data credit card**

a) Status of the dataset crm_last_year

Characteristics of the dataset crm_last_year:

- Source project name: **data credit card**
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **maxime.capron** on the 2023-04-19 19:31:17
- Last Modification by **maxime.capron** on the 2023-04-19 19:31:55

b) Related Recipes

The dataset crm_last_year relates to the following recipes.

Successor Recipes

List of recipes with crm_last_year as an input:

- compute_crm_last_year_stacked ([Stack](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
birth	string	

price_first_item_purchased	string	
gender	string	
revenue	string	

5. new_customers

Summarized information about your dataset new_customers:

- This dataset has not been created in this project. Its origin project is **data credit card**

a) Status of the dataset new_customers

Characteristics of the dataset new_customers:

- Source project name: **data credit card**
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **4,600**
- Created by **maxime.capron** on the 2023-04-19 19:31:12
- Last Modification by **maxime.capron** on the 2023-04-19 19:31:48

b) Related Recipes

The dataset new_customers relates to the following recipes.

Successor Recipes

List of recipes with new_customers as an input:

- compute_crm_last_year_stacked (*Stack*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
birth	string	

price_first_item_purchased	string	
gender	string	

6. crm_last_year_stacked

Summarized information about your dataset crm_last_year_stacked:

a) Status of the dataset crm_last_year_stacked

Characteristics of the dataset crm_last_year_stacked:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:02:21
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:02:21

b) Related Recipes

The dataset crm_last_year_stacked relates to the following recipes.

Parent Recipe

List of recipes with crm_last_year_stacked as an output:

- compute_crm_last_year_stacked (*Stack*)

Successor Recipes

List of recipes with crm_last_year_stacked as an input:

- compute_web_last_year_stacked_joined (*Join*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	

birth	string	
price_first_item_purchased	string	
gender	string	
revenue	string	

7. web_last_year_stacked_joined

Summarized information about your dataset `web_last_year_stacked_joined`:

a) Status of the dataset `web_last_year_stacked_joined`

Characteristics of the dataset `web_last_year_stacked_joined`:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **34,237**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:05:53
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:05:53

b) Related Recipes

The dataset `web_last_year_stacked_joined` relates to the following recipes.

Parent Recipe

List of recipes with `web_last_year_stacked_joined` as an output:

- `compute_web_last_year_stacked_joined` ([Join](#))

Successor Recipes

List of recipes with `web_last_year_stacked_joined` as an input:

- `compute_web_last_year_stacked_joined_prepared` ([Prepare](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	
pages_visited	string	
campain	string	
birth	string	
price_first_item_purchased	string	
gender	string	
revenue	string	

8. web_last_year_stacked_joined_prepared

Summarized information about your dataset web_last_year_stacked_joined_prepared:

a) Status of the dataset web_last_year_stacked_joined_prepared

Characteristics of the dataset web_last_year_stacked_joined_prepared:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **34,205**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:07:11
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:49:08

b) Related Recipes

The dataset web_last_year_stacked_joined_prepared relates to the following recipes.

Parent Recipe

List of recipes with web_last_year_stacked_joined_prepared as an output:

- compute_web_last_year_stacked_joined_prepared ([Prepare](#))

Successor Recipes

List of recipes with web_last_year_stacked_joined_prepared as an input:

- compute_web_last_year_stacked_joined_prepared_joined ([Join](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	
geo_geopoint	string	
geo_country	string	
pages_visited	double	
campain	boolean	
age	double	
price_first_item_purchased	double	
gender	string	
revenue	float	
first_purchased/revenue	double	

9. country_gdp

Summarized information about your dataset country_gdp:

- This dataset has not been created in this project. Its origin project is **data credit card**

a) Status of the dataset country_gdp

Characteristics of the dataset country_gdp:

- Source project name: **data credit card**
- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**

- Record Count: **188**
- Created by **maxime.capron** on the 2023-04-19 19:31:19
- Last Modification by **maxime.capron** on the 2023-04-19 19:32:04

b) Related Recipes

The dataset country_gdp relates to the following recipes.

Successor Recipes

List of recipes with country_gdp as an input:

- compute_country_gdp_copy (*Sync*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Country	string	
GDP_cap	string	

10. country_gdp_copy

Summarized information about your dataset country_gdp_copy:

a) Status of the dataset country_gdp_copy

Characteristics of the dataset country_gdp_copy:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:45:16
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:45:16

b) Related Recipes

The dataset country_gdp_copy relates to the following recipes.

Parent Recipe

List of recipes with country_gdp_copy as an output:

- compute_country_gdp_copy (*Sync*)

Successor Recipes

List of recipes with country_gdp_copy as an input:

- compute_web_last_year_stacked_joined_prepared_joined (*Join*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Country	string	
GDP_cap	string	

11. web_last_year_stacked_joined_prepared_joined

Summarized information about your dataset web_last_year_stacked_joined_prepared_joined:

a) Status of the dataset web_last_year_stacked_joined_prepared_joined

Characteristics of the dataset web_last_year_stacked_joined_prepared_joined:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **34,205**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:37:19
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:50:00

b) Related Recipes

The dataset `web_last_year_stacked_joined_prepared_joined` relates to the following recipes.

Parent Recipe

List of recipes with `web_last_year_stacked_joined_prepared_joined` as an output:

- `compute_web_last_year_stacked_joined_prepared_joined` ([Join](#))

Successor Recipes

List of recipes with `web_last_year_stacked_joined_prepared_joined` as an input:

- `compute_web_last_year_stacked_joined_prepared_joined_prepared` ([Prepare](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
<code>customer_id</code>	string	
<code>ip</code>	string	
<code>geo_geopoint</code>	string	
<code>geo_country</code>	string	
<code>pages_visited</code>	double	
<code>campain</code>	boolean	
<code>age</code>	double	
<code>price_first_item_purchased</code>	double	
<code>gender</code>	string	
<code>revenue</code>	float	
<code>first_purchased/revenue</code>	double	
<code>Country</code>	string	
<code>GDP_cap</code>	string	

12. Final_Prepared_Dataset

Summarized information about your dataset Final_Prepared_Dataset:

a) Status of the dataset Final_Prepared_Dataset

Characteristics of the dataset Final_Prepared_Dataset:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **34,205**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:39:39
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:52:44

b) Related Recipes

The dataset Final_Prepared_Dataset relates to the following recipes.

Parent Recipe

List of recipes with Final_Prepared_Dataset as an output:

- compute_web_last_year_stacked_joined_prepared_joined_prepared (*Prepare*)

Successor Recipes

List of recipes with Final_Prepared_Dataset as an input:

- split_Final_Prepared_Dataset (*Split*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	
geo_geopoint	string	
geo_country	string	

pages_visited	double	
campain	boolean	
age	float	
price_first_item_purchased	double	
gender	string	
revenue	float	
first_purchased/revenue	double	
GDP_cap	bigint	

13. Data_labelled

Summarized information about your dataset Data_labelled:

a) Status of the dataset Data_labelled

Characteristics of the dataset Data_labelled:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **29,271**
- Created by **stefanos.lambropoulos** on the 2023-04-20 13:09:25
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:54:15

b) Related Recipes

The dataset Data_labelled relates to the following recipes.

Parent Recipe

List of recipes with Data_labelled as an output:

- [split_Final_Prepared_Dataset \(*Split*\)](#)

Successor Recipes

List of recipes with Data_labelled as an input:

- train_Predict_revenue_random_forest_ (*Prediction Training*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	
geo_geopoint	string	
geo_country	string	
pages_visited	double	
campain	boolean	
age	float	
price_first_item_purchased	double	
gender	string	
revenue	float	
first_purchased/revenue	double	
GDP_cap	bigint	

14. Data_unlabelled

Summarized information about your dataset Data_unlabelled:

a) Status of the dataset Data_unlabelled

Characteristics of the dataset Data_unlabelled:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **4,934**
- Created by **stefanos.lambropoulos** on the 2023-04-20 13:09:35
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:54:15

b) Related Recipes

The dataset Data_unlabelled relates to the following recipes.

Parent Recipe

List of recipes with Data_unlabelled as an output:

- split_Final_Prepared_Dataset (*Split*)

Successor Recipes

List of recipes with Data_unlabelled as an input:

- score_Data_unlabelled (*Prediction Scoring*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	
geo_geopoint	string	
geo_country	string	
pages_visited	double	
campain	boolean	
age	float	
price_first_item_purchased	double	
gender	string	
revenue	float	
first_purchased/revenue	double	
GDP_cap	bigint	

15. Data_unlabelled_scored

Summarized information about your dataset Data_unlabelled_scored:

a) Status of the dataset Data_unlabelled_scored

Characteristics of the dataset Data_unlabelled_scored:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **4,934**
- Created by **yunting.lu** on the 2023-04-20 14:38:21
- Last Modification by **yunting.lu** on the 2023-04-20 14:38:21

b) Related Recipes

The dataset Data_unlabelled_scored relates to the following recipes.

Parent Recipe

List of recipes with Data_unlabelled_scored as an output:

- score_Data_unlabelled (*Prediction Scoring*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
customer_id	string	
ip	string	
geo_geopoint	string	
geo_country	string	
pages_visited	double	
campain	boolean	
age	double	
price_first_item_purchased	double	
gender	string	
revenue	double	

<code>first_purchased/revenue</code>	double	
<code>GDP_cap</code>	bigint	
<code>prediction</code>	float	

16. `country_gdp_joined`

Summarized information about your dataset `country_gdp_joined`:

a) Status of the dataset `country_gdp_joined`

Characteristics of the dataset `country_gdp_joined`:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:35:21
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:35:21

b) Related Recipes

The dataset `country_gdp_joined` relates to the following recipes.

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
<code>Country</code>	string	
<code>GDP_cap</code>	string	
<code>customer_id</code>	string	
<code>ip</code>	string	
<code>geo_country</code>	string	
<code>pages_visited</code>	double	
<code>campain</code>	boolean	

age	bigint	
price_first_item_purchased	double	
gender	string	
revenue	bigint	
first_purchased/revenue	double	

IV. Recipes

List all the recipes in the flow.

1. compute_web_last_year_stacked

Summarized information about the recipe compute_web_last_year_stacked.

a) Status

Characteristics of the recipe compute_web_last_year_stacked:

- Type: **Stack**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:03:37
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:03:40

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset web_last_year (*Uploaded Files*)
- Dataset web_new_visitors (*Uploaded Files*)

List of the outputs of the recipe:

- Dataset web_last_year_stacked (*Snowflake*)

Stacking mode: Union of input schemas

2. compute_crm_last_year_stacked

Summarized information about the recipe compute_crm_last_year_stacked.

a) Status

Characteristics of the recipe compute_crm_last_year_stacked:

- Type: **Stack**

- Created by **stefanos.lambropoulos** on the 2023-04-20 12:02:21
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:02:28

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset `crm_last_year` (*Uploaded Files*)
- Dataset `new_customers` (*Uploaded Files*)

List of the outputs of the recipe:

- Dataset `crm_last_year_stacked` (*Snowflake*)

Stacking mode: Union of input schemas

3. `compute_web_last_year_stacked_joined`

Summarized information about the recipe `compute_web_last_year_stacked_joined`.

a) Status

Characteristics of the recipe `compute_web_last_year_stacked_joined`:

- Type: **Join**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:05:53
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:06:24

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset `web_last_year_stacked` (*Snowflake*)
- Dataset `crm_last_year_stacked` (*Snowflake*)

List of the outputs of the recipe:

- Dataset `web_last_year_stacked_joined` (*Snowflake*)

c) Join

Left web_last_year_stacked - crm_last_year_stacked And

Left input	Condition	Right input
customer_id	=	customer_id

4. compute_web_last_year_stacked_joined_prepared

Summarized information about the recipe compute_web_last_year_stacked_joined_prepared.

a) Status

Characteristics of the recipe compute_web_last_year_stacked_joined_prepared:

- Type: **Prepare**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:07:11
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:49:10

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset **web_last_year_stacked_joined** (*Snowflake*)

List of the outputs of the recipe:

- Dataset **web_last_year_stacked_joined_prepared** (*Snowflake*)

c) Prepare Steps

Type	Comment
DateParser	
DateDifference	
ColumnsSelector	
FilterOnNumericalRange	
GeoIPResolver	
CreateColumnWithGREL	
RoundProcessor	
FilterOnBadType	

RemoveRowsOnEmpty	
FillEmptyWithValue	

5. compute_country_gdp_copy

Summarized information about the recipe compute_country_gdp_copy.

a) Status

Characteristics of the recipe compute_country_gdp_copy:

- Type: **Sync**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:45:16
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 12:45:22

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset country_gdp (*Uploaded Files*)

List of the outputs of the recipe:

- Dataset country_gdp_copy (*Snowflake*)

6. compute_web_last_year_stacked_joined_prepared_joined

Summarized information about the recipe compute_web_last_year_stacked_joined_prepared_joined.

a) Status

Characteristics of the recipe compute_web_last_year_stacked_joined_prepared_joined:

- Type: **Join**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:37:19
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:50:02

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset web_last_year_stacked_joined_prepared ([Snowflake](#))
- Dataset country_gdp_copy ([Snowflake](#))

List of the outputs of the recipe:

- Dataset web_last_year_stacked_joined_prepared_joined ([Snowflake](#))

c) Join

Left web_last_year_stacked_joined_prepared - country_gdp_copy And

Left input	Condition	Right input
geo_country	=	Country

7. compute_web_last_year_stacked_joined_prepared_joined_prepared

Summarized information about the recipe

compute_web_last_year_stacked_joined_prepared_joined_prepared.

a) Status

Characteristics of the recipe

compute_web_last_year_stacked_joined_prepared_joined_prepared:

- Type: **Prepare**
- Created by **stefanos.lambropoulos** on the 2023-04-20 12:39:39
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:53:42

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset web_last_year_stacked_joined_prepared_joined ([Snowflake](#))

List of the outputs of the recipe:

- Dataset Final_Prepared_Dataset ([Snowflake](#))

c) Prepare Steps

Type	Comment
ColumnsSelector	
RoundProcessor	

8. split_Final_Prepared_Dataset

Summarized information about the recipe split_Final_Prepared_Dataset.

a) Status

Characteristics of the recipe split_Final_Prepared_Dataset:

- Type: **Split**
- Created by **stefanos.lambropoulos** on the 2023-04-20 13:09:40
- Last Modification by **stefanos.lambropoulos** on the 2023-04-20 13:54:18

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset Final_Prepared_Dataset (*Snowflake*)

List of the outputs of the recipe:

- Dataset Data_labelled (*Snowflake*)
- Dataset Data_unlabelled (*Snowflake*)

Split mode: Specific filter for each output dataset

9. train_Predict_revenue_random_forest_

Summarized information about the recipe train_Predict_revenue_random_forest_.

a) Status

Characteristics of the recipe train_Predict_revenue_random_forest_:

- Type: **Prediction Training**
- Created by **yunting.lu** on the 2023-04-20 14:37:47

- Last Modification by **yunting.lu** on the 2023-04-20 14:37:47

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset Data_labelled (*Snowflake*)

List of the outputs of the recipe:

- Saved Model Predict revenue (random forest) (*Random forest (s1) - v1*)

10. score_Data_unlabelled

Summarized information about the recipe score_Data_unlabelled.

a) Status

Characteristics of the recipe score_Data_unlabelled:

- Type: **Prediction Scoring**
- Created by **yunting.lu** on the 2023-04-20 14:38:21
- Last Modification by **yunting.lu** on the 2023-04-20 14:38:30

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset Data_unlabelled (*Snowflake*)
- Saved Model Predict revenue (random forest) (*Random forest (s1) - v1*)

List of the outputs of the recipe:

- Dataset Data_unlabelled_scored (*Snowflake*)

V. Saved Models

List all the saved models in the flow.

1. Predict revenue (random forest)

Summarized information about the saved model Predict revenue (random forest):.

a) Status

Model id: **OQ3YQ0Ly**

Active version: **Random forest (s1) - v1**

b) Related Recipes

List of the recipes related to the saved model.

Parent Recipe

Recipe with Predict revenue (random forest) as an output:

- train_Predict_revenue_random_forest_ (*Prediction Training*)

Successor Recipes

List of recipes with Predict revenue (random forest) as an input:

- score_Data_unlabelled (*Prediction Scoring*)

VI. Deployment and Monitoring

A. Implementation Details

- The flow can be found here: <https://dss-44270426-726225ce-dku.eu-west-3.app.dataiku.io/projects/POR/>
- The name of the generated file is: Dataiku Flow Documentation - POR.docx

B. Version Control

- The flow was executed with the following version of DSS: 11.4.1