

# Project\_Online\_Retailer

---



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

## Table of contents

<b>I. Project general description</b>	<b>5</b>
<b>II. Graphical Representation of the Flow</b>	<b>6</b>
<b>III. Datasets</b>	<b>7</b>
1. web_last_year	7
a) Status of the dataset web_last_year	7
b) Related Recipes	7
c) Schema	7
2. web_new_visitors	8
a) Status of the dataset web_new_visitors	8
b) Related Recipes	8
c) Schema	8
3. web_last_year_stacked	9
a) Status of the dataset web_last_year_stacked	9
b) Related Recipes	9
c) Schema	9
4. crm_last_year	10
a) Status of the dataset crm_last_year	10
b) Related Recipes	10
c) Schema	10
5. new_customers	11
a) Status of the dataset new_customers	11
b) Related Recipes	11
c) Schema	11
6. crm_last_year_stacked	12
a) Status of the dataset crm_last_year_stacked	12
b) Related Recipes	12
c) Schema	12
7. web_last_year_stacked_joined	13
a) Status of the dataset web_last_year_stacked_joined	13
b) Related Recipes	13
c) Schema	13
8. web_last_year_stacked_joined_prepared	14
a) Status of the dataset web_last_year_stacked_joined_prepared	14
b) Related Recipes	14
c) Schema	15
9. country_gdp	15
a) Status of the dataset country_gdp	15
b) Related Recipes	16
c) Schema	16
10. country_gdp_copy	16

a)	Status of the dataset country_gdp_copy	16
b)	Related Recipes	17
c)	Schema	17
11.	web_last_year_stacked_joined_prepared_joined	17
a)	Status of the dataset web_last_year_stacked_joined_prepared_joined	17
b)	Related Recipes	18
c)	Schema	18
12.	Final_Prepared_Dataset	19
a)	Status of the dataset Final_Prepared_Dataset	19
b)	Related Recipes	19
c)	Schema	19
13.	Data_labelled	20
a)	Status of the dataset Data_labelled	20
b)	Related Recipes	20
c)	Schema	21
14.	Data_unlabelled	21
a)	Status of the dataset Data_unlabelled	21
b)	Related Recipes	22
c)	Schema	22
15.	Data_unlabelled_scored	22
a)	Status of the dataset Data_unlabelled_scored	23
b)	Related Recipes	23
c)	Schema	23
16.	country_gdp_joined	24
a)	Status of the dataset country_gdp_joined	24
b)	Related Recipes	24
c)	Schema	24
<b>IV.</b>	<b>Recipes</b>	<b>26</b>
1.	compute_web_last_year_stacked	26
a)	Status	26
b)	Inputs / Outputs	26
2.	compute_crm_last_year_stacked	26
a)	Status	26
b)	Inputs / Outputs	27
3.	compute_web_last_year_stacked_joined	27
a)	Status	27
b)	Inputs / Outputs	27
c)	Join	28
4.	compute_web_last_year_stacked_joined_prepared	28
a)	Status	28
b)	Inputs / Outputs	28
c)	Prepare Steps	28
5.	compute_country_gdp_copy	29

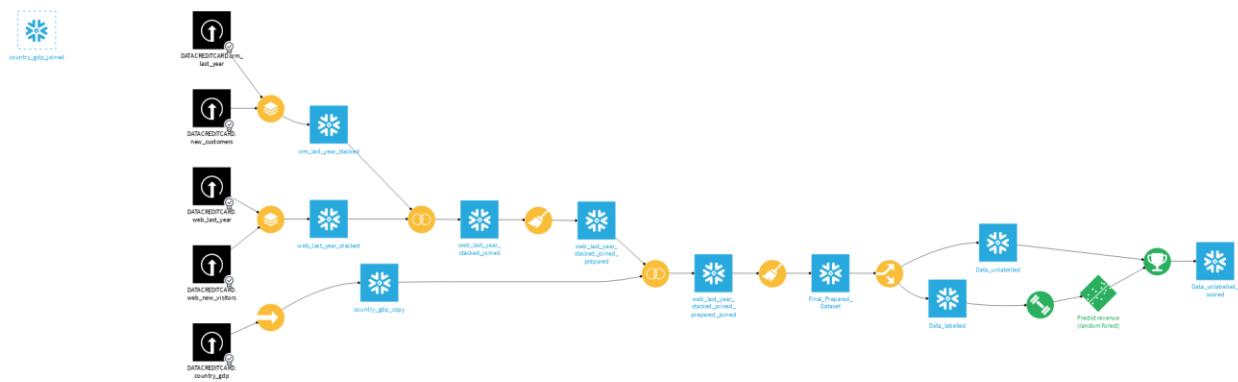
a)	Status	29
b)	Inputs / Outputs	29
6.	compute_web_last_year_stacked_joined_prepared_joined	29
a)	Status	29
b)	Inputs / Outputs	30
c)	Join	30
7.	compute_web_last_year_stacked_joined_prepared_joined_prepared	30
a)	Status	30
b)	Inputs / Outputs	30
c)	Prepare Steps	31
8.	split_Final_Prepared_Dataset	31
a)	Status	31
b)	Inputs / Outputs	31
9.	train_Predict_revenue_random_forest_	31
a)	Status	31
b)	Inputs / Outputs	32
10.	score_Data_unlabelled	32
a)	Status	32
b)	Inputs / Outputs	32
<b>V.</b>	<b>Saved Models</b>	<b>33</b>
1.	Predict revenue (random forest)	33
a)	Status	33
b)	Related Recipes	33
<b>VI.</b>	<b>Deployment and Monitoring</b>	<b>34</b>
A.	Implementation Details	34
B.	Version Control	34

## 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	

<b>ip</b>	string	
<b>pages_visited</b>	string	
<b>campain</b>	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
<b>customer_id</b>	string	
<b>ip</b>	string	

<b>pages_visited</b>	string	
<b>campain</b>	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
<b>customer_id</b>	string	

<b>ip</b>	string	
<b>pages_visited</b>	string	
<b>campain</b>	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
<b>customer_id</b>	string	
<b>birth</b>	string	

<b>price_first_item_purchased</b>	string	
<b>gender</b>	string	
<b>revenue</b>	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
<b>customer_id</b>	string	
<b>birth</b>	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	

<b>birth</b>	string	
<b>price_first_item_purchased</b>	string	
<b>gender</b>	string	
<b>revenue</b>	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
<b>Country</b>	string	
<b>GDP_cap</b>	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	

<b>pages_visited</b>	double	
<b>campain</b>	boolean	
<b>age</b>	float	
<b>price_first_item_purchased</b>	double	
<b>gender</b>	string	
<b>revenue</b>	float	
<b>first_purchased/revenue</b>	double	
<b>GDP_cap</b>	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	

<b>age</b>	bigint	
<b>price_first_item_purchased</b>	double	
<b>gender</b>	string	
<b>revenue</b>	bigint	
<b>first_purchased/revenue</b>	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