

[Siyu] Advanced Designer



Version

Author

Date

1.0

siyu.chen

2024-04-29 11:09:25

Table of contents

I.	Project general description	4
II.	Graphical Representation of the Flow	5
III.	Datasets	6
1.	{{ \$dataset.name }}	错误!未定义书签。
a)	Status of the dataset {{ \$dataset.name }}	错误!未定义书签。
b)	Related Recipes	28
c)	Schema	28
IV.	Managed folders	错误!未定义书签。
1.	{{ \$folder.name }}	错误!未定义书签。
a)	Status	错误!未定义书签。
b)	Derived datasets	错误!未定义书签。
c)	Related Recipes	错误!未定义书签。
V.	Recipes	31
1.	{{ \$recipe.name }}	错误!未定义书签。
a)	Status	38
b)	Inputs / Outputs	38
c)	Prepare Steps	33
d)	Join	35
e)	Data Relationships	错误!未定义书签。
f)	Feature Transformations	错误!未定义书签。
g)	Code	错误!未定义书签。
h)	Windows	37
i)	Sorted Columns	错误!未定义书签。
j)	Top rows and sorting direction	36
k)	Parameters	错误!未定义书签。
VI.	Saved Models	39
1.	{{ \$model.name }}	错误!未定义书签。
a)	Status	39
b)	Related Recipes	39
VII.	Model Evaluation Stores	错误!未定义书签。
1.	{{ \$mes.name }}	错误!未定义书签。
a)	Related Recipes	错误!未定义书签。
VIII.	Labeling Tasks	错误!未定义书签。
1.	{{ \$labeling_task.name }}	错误!未定义书签。
a)	Status	错误!未定义书签。
b)	Inputs / Outputs	错误!未定义书签。

IX.	Deployment and Monitoring	40
A.	Implementation Details	40
B.	Version Control	40

I. Project general description

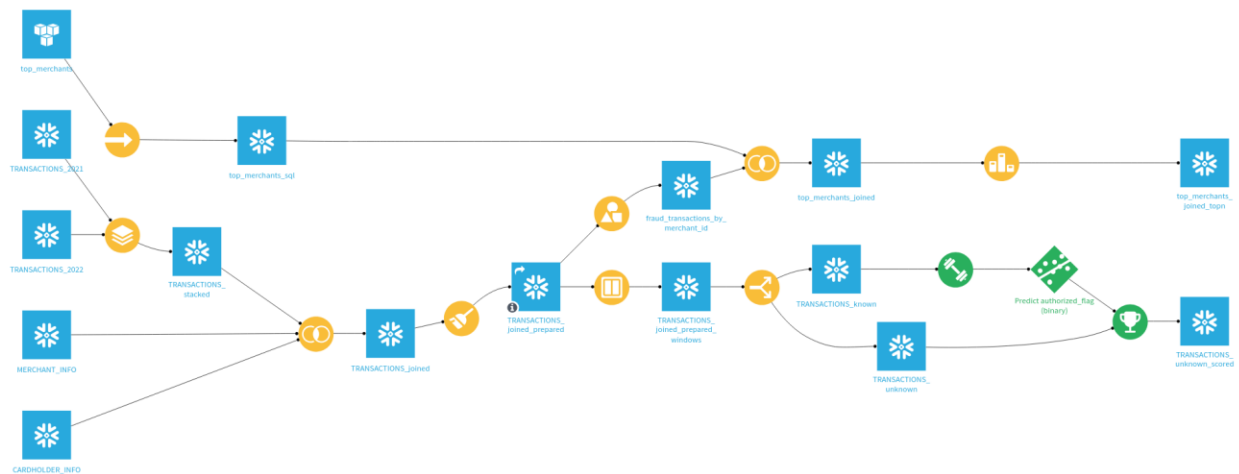
Project references for the created flow:

- Project name: **[Siyu] Advanced Designer**
- Project key: **SIYUADVANCEDDESIGNER**
- Project short description:

*The project *Sylvain's full walkthrough* was created by sylvain.franco on Dec 08th 2022
The project *STARTERADVANCEDDESIGNER* was copied from
SYLVAINSFULLWALKTHROUGH by sylvain.franco on Dec 20th 2022
The project *SIYUADVANCEDDESIGNER* was copied from
STARTERADVANCEDDESIGNER by siyu.chen on Apr 25th 2024*

- Status:
- Current git branch: **master**
- Tags: **duplicated, imported**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:08:27
- Last Modification by **api:XJHUEYK71G9Z8M5F (deleted)** on the 2024-04-26 11:10:42

II. Graphical Representation of the Flow



III. Datasets

List all the datasets in the flow.

1. TRANSACTIONS_2021

Summarized information about your dataset TRANSACTIONS_2021:

- Tags: **origin:sql_import, managed-snowflake**

a) Status of the dataset TRANSACTIONS_2021

Characteristics of the dataset TRANSACTIONS_2021:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:10:17
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 17:11:16

b) Related Recipes

The dataset TRANSACTIONS_2021 relates to the following recipes.

Successor Recipes

List of recipes with TRANSACTIONS_2021 as an input:

- `compute_TRANSACTIONS_stacked` ([Stack](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	string	
authorized_flag	string	

purchase_date	string	
card_id	string	
merchant_id	string	
merchant_category_id	string	
item_category	string	
purchase_amount	string	
signature_provided	string	

2. TRANSACTIONS_2022

Summarized information about your dataset TRANSACTIONS_2022:

- Tags: **origin:sql_import, managed-snowflake**

a) Status of the dataset TRANSACTIONS_2022

Characteristics of the dataset TRANSACTIONS_2022:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:10:17
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 17:11:26

b) Related Recipes

The dataset TRANSACTIONS_2022 relates to the following recipes.

Successor Recipes

List of recipes with TRANSACTIONS_2022 as an input:

- compute_TRANSACTIONS_stacked ([Stack](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	string	
authorized_flag	string	
purchase_date	string	
card_id	string	
merchant_id	string	
merchant_category_id	string	
item_category	string	
purchase_amount	string	
signature_provided	string	

3. TRANSACTIONS_stacked

Summarized information about your dataset TRANSACTIONS_stacked:

a) Status of the dataset TRANSACTIONS_stacked

Characteristics of the dataset TRANSACTIONS_stacked:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:11:29
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:11:29

b) Related Recipes

The dataset TRANSACTIONS_stacked relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_stacked as an output:

- compute_TRANSACTIONS_stacked ([Stack](#))

Successor Recipes

List of recipes with TRANSACTIONS_stacked as an input:

- compute_TRANSACTIONS_stacked_joined ([Join](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	string	
authorized_flag	string	
purchase_date	string	
card_id	string	
merchant_id	string	
merchant_category_id	string	
item_category	string	
purchase_amount	string	
signature_provided	string	

4. MERCHANT_INFO

Summarized information about your dataset MERCHANT_INFO:

- Tags: **origin:sql_import, managed-snowflake**

a) Status of the dataset MERCHANT_INFO

Characteristics of the dataset MERCHANT_INFO:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:10:17
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:12:54

b) Related Recipes

The dataset MERCHANT_INFO relates to the following recipes.

Successor Recipes

List of recipes with MERCHANT_INFO as an input:

- compute_TRANSACTIONS_stacked_joined ([Join](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
merchant_id	string	
merchant_category_id	string	
subsector_description	string	
latitude	string	
longitude	string	

5. CARDHOLDER_INFO

Summarized information about your dataset CARDHOLDER_INFO:

- Tags: **origin:sql_import, managed-snowflake**

a) Status of the dataset CARDHOLDER_INFO

Characteristics of the dataset CARDHOLDER_INFO:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:10:17
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 17:11:39

b) Related Recipes

The dataset CARDHOLDER_INFO relates to the following recipes.

Successor Recipes

List of recipes with CARDHOLDER_INFO as an input:

- compute_TRANSACTIONS_stacked_joined ([Join](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
internal_card_mapping	string	
first_active_month	string	
reward_program	string	
latitude	string	
longitude	string	
fico_score	string	
age	string	

6. TRANSACTIONS_joined

Summarized information about your dataset TRANSACTIONS_joined:

a) Status of the dataset TRANSACTIONS_joined

Characteristics of the dataset TRANSACTIONS_joined:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 16:39:20
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 16:39:22

b) Related Recipes

The dataset TRANSACTIONS_joined relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_joined as an output:

- compute_TRANSACTIONS_stacked_joined ([Join](#))

Successor Recipes

List of recipes with TRANSACTIONS_joined as an input:

- compute_TRANSACTIONS_joined_prepared ([Prepare](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	string	
authorized_flag	string	
purchase_date	string	
card_id	string	
merchant_id	string	
merchant_category_id	string	
item_category	string	
purchase_amount	string	
signature_provided	string	
merchant_subsector_description	string	
merchant_latitude	string	
merchant_longitude	string	
card_first_active_month	string	
card_reward_program	string	
card_latitude	string	

card_longitude	string	
card_fico_score	string	
card_age	string	

7. TRANSACTIONS_joined_prepared

Summarized information about your dataset TRANSACTIONS_joined_prepared:

- Short description:
This dataset contains enriched features on credit card transactions.
- Tags: **credit-card-fraud**

a) Status of the dataset TRANSACTIONS_joined_prepared

Characteristics of the dataset TRANSACTIONS_joined_prepared:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 16:39:43
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:57:40

b) Related Recipes

The dataset TRANSACTIONS_joined_prepared relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_joined_prepared as an output:

- compute_TRANSACTIONS_joined_prepared ([Prepare](#))

Successor Recipes

List of recipes with TRANSACTIONS_joined_prepared as an input:

- compute_TRANSACTIONS_joined_prepared_windows ([Window](#))
- compute_Tfraud_transactions_by_merchant_id ([Group](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	bigint	
authorized_flag	bigint	
authorized_flag_description	string	
purchase_date	string	
purchase_date_parsed	date	
purchase_date_parsed_month	bigint	
purchase_date_parsed_day	bigint	
purchase_date_dow	bigint	
purchase_weekend	bigint	
card_id	string	
merchant_id	string	
merchant_category_id	bigint	
item_category	string	
purchase_amount	double	
signature_provided	string	
risk	string	
merchant_subsector_description	string	
spending_type	string	
merchant_latitude	double	
merchant_longitude	double	
card_first_active_month	string	
card_first_active_month_parsed	date	
days_active	bigint	
card_reward_program	string	

card_latitude	double	
card_longitude	double	
card_fico_score	bigint	
card_age	bigint	
merchant_location	string	
merchant_state	string	
merchant_state_enName	string	
card_location	string	
distance_card_merchant	double	
card_state	string	
card_state_enName	string	

8. fraud_transactions_by_merchant_id

Summarized information about your dataset fraud_transactions_by_merchant_id:

a) Status of the dataset fraud_transactions_by_merchant_id

Characteristics of the dataset fraud_transactions_by_merchant_id:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **0**
- Created by **siyu.chen** on the 2024-04-26 08:54:13
- Last Modification by **siyu.chen** on the 2024-04-26 09:32:09

b) Related Recipes

The dataset fraud_transactions_by_merchant_id relates to the following recipes.

Parent Recipe

List of recipes with fraud_transactions_by_merchant_id as an output:

- compute_Tfraud_transactions_by_merchant_id (*Group*)

Successor Recipes

List of recipes with fraud_transactions_by_merchant_id as an input:

- compute_top_merchants_joined (*Join*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
merchant_id	string	
merchant_category	bigint	
sum_fraud_transactions	double	
merchant_subsector_description	string	
merchant_geopoint	string	
merchant_state	string	

9. top_merchants

Summarized information about your dataset top_merchants:

a) Status of the dataset top_merchants

Characteristics of the dataset top_merchants:

- Type: **Amazon S3**
- Connection: **sample-data**
- Total Size: **Not computed**
- Record Count: **100**
- Created by **siyu.chen** on the 2024-04-26 09:14:11
- Last Modification by **siyu.chen** on the 2024-04-26 09:14:11

b) Related Recipes

The dataset top_merchants relates to the following recipes.

Successor Recipes

List of recipes with top_merchants as an input:

- compute_top_merchants_sql ([Sync](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
merchant_id	string	
merchant_name	string	
merchant_category_id	bigint	
merchant_turnover	double	

10. top_merchants_sql

Summarized information about your dataset top_merchants_sql:

a) Status of the dataset top_merchants_sql

Characteristics of the dataset top_merchants_sql:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **siyu.chen** on the 2024-04-26 09:37:38
- Last Modification by **siyu.chen** on the 2024-04-26 09:37:38

b) Related Recipes

The dataset top_merchants_sql relates to the following recipes.

Parent Recipe

List of recipes with top_merchants_sql as an output:

- compute_top_merchants_sql ([Sync](#))

Successor Recipes

List of recipes with top_merchants_sql as an input:

- compute_top_merchants_joined ([Join](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
merchant_id	string	
merchant_name	string	
merchant_category_id	bigint	
merchant_turnover	double	

11. top_merchants_joined

Summarized information about your dataset top_merchants_joined:

a) Status of the dataset top_merchants_joined

Characteristics of the dataset top_merchants_joined:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **100**
- Created by **siyu.chen** on the 2024-04-26 09:21:34
- Last Modification by **siyu.chen** on the 2024-04-26 09:33:35

b) Related Recipes

The dataset top_merchants_joined relates to the following recipes.

Parent Recipe

List of recipes with top_merchants_joined as an output:

- compute_top_merchants_joined ([Join](#))

Successor Recipes

List of recipes with `top_merchants_joined` as an input:

- `compute_top_merchants_joined_topn` (*Top N*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
<code>merchant_id</code>	string	
<code>merchant_name</code>	string	
<code>merchant_category_id</code>	bigint	
<code>merchant_turnover</code>	double	
<code>merchant_category</code>	bigint	
<code>sum_fraud_transcations</code>	double	
<code>merchant_subsector_description</code>	string	
<code>merchant_geopoint</code>	string	
<code>merchant_state</code>	string	
<code>percent_fraud</code>	double	

12. `top_merchants_joined_topn`

Summarized information about your dataset `top_merchants_joined_topn`:

a) Status of the dataset `top_merchants_joined_topn`

Characteristics of the dataset `top_merchants_joined_topn`:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **10**
- Created by **siyu.chen** on the 2024-04-26 09:35:37
- Last Modification by **siyu.chen** on the 2024-04-26 09:36:12

b) Related Recipes

The dataset `top_merchants_joined_topn` relates to the following recipes.

Parent Recipe

List of recipes with `top_merchants_joined_topn` as an output:

- `compute_top_merchants_joined_topn` (*Top N*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
<code>merchant_id</code>	string	
<code>merchant_name</code>	string	
<code>merchant_category_id</code>	bigint	
<code>merchant_turnover</code>	double	
<code>merchant_category</code>	bigint	
<code>sum_fraud_transactions</code>	double	
<code>merchant_subsector_description</code>	string	
<code>merchant_geopoint</code>	string	
<code>merchant_state</code>	string	
<code>percent_fraud</code>	double	

13. TRANSACTIONS_joined_prepared_windows

Summarized information about your dataset `TRANSACTIONS_joined_prepared_windows`:

a) Status of the dataset `TRANSACTIONS_joined_prepared_windows`

Characteristics of the dataset `TRANSACTIONS_joined_prepared_windows`:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**

- Record Count: **294,135**
- Created by **siyu.chen** on the 2024-04-26 08:28:18
- Last Modification by **siyu.chen** on the 2024-04-26 08:30:41

b) Related Recipes

The dataset TRANSACTIONS_joined_prepared_windows relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_joined_prepared_windows as an output:

- compute_TRANSACTIONS_joined_prepared_windows ([Window](#))

Successor Recipes

List of recipes with TRANSACTIONS_joined_prepared_windows as an input:

- split_TRANSACTIONS_joined_prepared ([Split](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	bigint	
authorized_flag	bigint	
authorized_flag_description	string	
purchase_date	string	
purchase_date_parsed	date	
purchase_date_parsed_month	bigint	
purchase_date_parsed_day	bigint	
purchase_date_dow	bigint	
purchase_weekend	bigint	
card_id	string	
merchant_id	string	

merchant_category_id	bigint	
item_category	string	
purchase_amount	double	
signature_provided	string	
risk	string	
merchant_subsector_description	string	
spending_type	string	
merchant_latitude	double	
merchant_longitude	double	
card_first_active_month	string	
card_first_active_month_parsed	date	
days_active	bigint	
card_reward_program	string	
card_latitude	double	
card_longitude	double	
card_fico_score	bigint	
card_age	bigint	
merchant_location	string	
merchant_state	string	
merchant_state_enName	string	
card_location	string	
distance_card_merchant	double	
card_state	string	
card_state_enName	string	
purchase_amount_min	double	
purchase_amount_max	double	
purchase_amount_avg	double	

purchase_amount_sum	double	
w2_purchase_amount_min	double	
w2_purchase_amount_max	double	
w2_purchase_amount_avg	double	
w2_purchase_amount_sum	double	

14. TRANSACTIONS_known

Summarized information about your dataset TRANSACTIONS_known:

a) Status of the dataset TRANSACTIONS_known

Characteristics of the dataset TRANSACTIONS_known:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **0**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:12:18
- Last Modification by **siyu.chen** on the 2024-04-26 08:33:12

b) Related Recipes

The dataset TRANSACTIONS_known relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_known as an output:

- split_TRANSACTIONS_joined_prepared ([Split](#))

Successor Recipes

List of recipes with TRANSACTIONS_known as an input:

- train_Predict_authorized_flag__binary_ ([Prediction Training](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	bigint	
authorized_flag	bigint	
authorized_flag_description	string	
purchase_date	string	
purchase_date_parsed	date	
purchase_date_parsed_month	bigint	
purchase_date_parsed_day	bigint	
purchase_date_dow	bigint	
purchase_weekend	bigint	
card_id	string	
merchant_id	string	
merchant_category_id	bigint	
item_category	string	
purchase_amount	double	
signature_provided	string	
risk	string	
merchant_subsector_description	string	
spending_type	string	
merchant_latitude	double	
merchant_longitude	double	
card_first_active_month	string	
card_first_active_month_parsed	date	
days_active	bigint	
card_reward_program	string	

card_latitude	double	
card_longitude	double	
card_fico_score	bigint	
card_age	bigint	
merchant_location	string	
merchant_state	string	
merchant_state_enName	string	
card_location	string	
distance_card_merchant	double	
card_state	string	
card_state_enName	string	
purchase_amount_min	double	
purchase_amount_max	double	
purchase_amount_avg	double	
purchase_amount_sum	double	
w2_purchase_amount_min	double	
w2_purchase_amount_max	double	
w2_purchase_amount_avg	double	
w2_purchase_amount_sum	double	

15. TRANSACTIONS_unknown

Summarized information about your dataset TRANSACTIONS_unknown:

a) Status of the dataset TRANSACTIONS_unknown

Characteristics of the dataset TRANSACTIONS_unknown:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**

- Record Count: **Not computed**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:12:29
- Last Modification by **siyu.chen** on the 2024-04-26 08:33:12

b) Related Recipes

The dataset TRANSACTIONS_unknown relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_unknown as an output:

- split_TRANSACTIONS_joined_prepared ([Split](#))

Successor Recipes

List of recipes with TRANSACTIONS_unknown as an input:

- score_TRANSACTIONS_unknown ([Prediction Scoring](#))

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	bigint	
authorized_flag	bigint	
authorized_flag_description	string	
purchase_date	string	
purchase_date_parsed	date	
purchase_date_parsed_month	bigint	
purchase_date_parsed_day	bigint	
purchase_date_dow	bigint	
purchase_weekend	bigint	
card_id	string	
merchant_id	string	

merchant_category_id	bigint	
item_category	string	
purchase_amount	double	
signature_provided	string	
risk	string	
merchant_subsector_description	string	
spending_type	string	
merchant_latitude	double	
merchant_longitude	double	
card_first_active_month	string	
card_first_active_month_parsed	date	
days_active	bigint	
card_reward_program	string	
card_latitude	double	
card_longitude	double	
card_fico_score	bigint	
card_age	bigint	
merchant_location	string	
merchant_state	string	
merchant_state_enName	string	
card_location	string	
distance_card_merchant	double	
card_state	string	
card_state_enName	string	
purchase_amount_min	double	
purchase_amount_max	double	
purchase_amount_avg	double	

purchase_amount_sum	double	
w2_purchase_amount_min	double	
w2_purchase_amount_max	double	
w2_purchase_amount_avg	double	
w2_purchase_amount_sum	double	

16. TRANSACTIONS_unknown_scored

Summarized information about your dataset TRANSACTIONS_unknown_scored:

a) Status of the dataset TRANSACTIONS_unknown_scored

Characteristics of the dataset TRANSACTIONS_unknown_scored:

- Type: **Snowflake**
- Connection: **managed-snowflake**
- Total Size: **Not computed**
- Record Count: **65,389**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:29:06
- Last Modification by **siyu.chen** on the 2024-04-26 08:40:48

b) Related Recipes

The dataset TRANSACTIONS_unknown_scored relates to the following recipes.

Parent Recipe

List of recipes with TRANSACTIONS_unknown_scored as an output:

- score_TRANSACTIONS_unknown (*Prediction Scoring*)

c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
transaction_id	bigint	
authorized_flag	bigint	

authorized_flag_description	string	
purchase_date	string	
purchase_date_parsed	date	
purchase_date_parsed_month	bigint	
purchase_date_parsed_day	bigint	
purchase_date_dow	bigint	
purchase_weekend	bigint	
card_id	string	
merchant_id	string	
merchant_category_id	bigint	
item_category	string	
purchase_amount	double	
signature_provided	bigint	
risk	string	
merchant_subsector_description	string	
spending_type	string	
merchant_latitude	double	
merchant_longitude	double	
card_first_active_month	string	
card_first_active_month_parsed	date	
days_active	bigint	
card_reward_program	string	
card_latitude	double	
card_longitude	double	
card_fico_score	bigint	
card_age	bigint	
merchant_location	string	

merchant_state	string	
merchant_state_enName	string	
card_location	string	
distance_card_merchant	double	
card_state	string	
card_state_enName	string	
purchase_amount_min	double	
purchase_amount_max	double	
purchase_amount_avg	double	
purchase_amount_sum	double	
w2_purchase_amount_min	double	
w2_purchase_amount_max	double	
w2_purchase_amount_avg	double	
w2_purchase_amount_sum	double	
proba_0	double	
proba_1	double	
prediction	string	

IV. Recipes

List all the recipes in the flow.

1. **compute_TRANSACTIONS_stacked**

Summarized information about the recipe compute_TRANSACTIONS_stacked.

a) **Status**

Characteristics of the recipe compute_TRANSACTIONS_stacked:

- Type: **Stack**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:11:29
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:11:32

b) **Inputs / Outputs**

List of the inputs of the recipe:

- Dataset TRANSACTIONS_2021 ([Snowflake](#))
- Dataset TRANSACTIONS_2022 ([Snowflake](#))

List of the outputs of the recipe:

- Dataset TRANSACTIONS_stacked ([Snowflake](#))

Stacking mode: Union of input schemas

2. **compute_TRANSACTIONS_stacked_joined**

Summarized information about the recipe compute_TRANSACTIONS_stacked_joined.

a) **Status**

Characteristics of the recipe compute_TRANSACTIONS_stacked_joined:

- Type: **Join**

- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 15:12:00
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 16:39:25

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_stacked (*Snowflake*)
- Dataset MERCHANT_INFO (*Snowflake*)
- Dataset CARDHOLDER_INFO (*Snowflake*)

List of the outputs of the recipe:

- Dataset TRANSACTIONS_joined (*Snowflake*)

c) Join

Left TRANSACTIONS_stacked - MERCHANT_INFO And

Left input	Condition	Right input
merchant_id	=	merchant_id

merchant_category_id	=	merchant_category_id
----------------------	---	----------------------

Left TRANSACTIONS_stacked - CARDHOLDER_INFO And

Left input	Condition	Right input
card_id	=	internal_card_mapping

3. compute_TRANSACTIONS_joined_prepared

Summarized information about the recipe compute_TRANSACTIONS_joined_prepared.

a) Status

Characteristics of the recipe compute_TRANSACTIONS_joined_prepared:

- Type: **Prepare**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-08 16:39:43
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:57:43

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_joined (*Snowflake*)

List of the outputs of the recipe:

- Dataset TRANSACTIONS_joined_prepared (*Snowflake*)

c) Prepare Steps

Type	Comment
FilterOnBadType	
FindReplace	
GeoPointCreator	
GeoPointCreator	
CityLevelReverseGeocoder	
CityLevelReverseGeocoder	
GeoDistanceProcessor	
SwitchCase	
VisualIfRule	
DateParser	
DateComponentsExtractor	
CreateColumnWithGREL	
DateParser	
DateDifference	

4. compute_Tfraud_transactions_by_merchant_id

Summarized information about the recipe compute_Tfraud_transactions_by_merchant_id.

a) Status

Characteristics of the recipe compute_Tfraud_transactions_by_merchant_id:

- Type: **Group**
- Created by **siyu.chen** on the 2024-04-26 08:54:13
- Last Modification by **siyu.chen** on the 2024-04-26 09:49:27

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_joined_prepared ([Snowflake](#))

List of the outputs of the recipe:

- Dataset fraud_transactions_by_merchant_id ([Snowflake](#))

Group keys: merchant_id

5. compute_top_merchants_sql

Summarized information about the recipe compute_top_merchants_sql.

a) Status

Characteristics of the recipe compute_top_merchants_sql:

- Type: **Sync**
- Created by **siyu.chen** on the 2024-04-26 09:37:38
- Last Modification by **siyu.chen** on the 2024-04-26 09:37:38

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset top_merchants ([Amazon S3](#))

List of the outputs of the recipe:

- Dataset top_merchants_sql ([Snowflake](#))

6. compute_top_merchants_joined

Summarized information about the recipe compute_top_merchants_joined.

a) Status

Characteristics of the recipe compute_top_merchants_joined:

- Type: **Join**
- Created by **siyu.chen** on the 2024-04-26 09:21:34
- Last Modification by **siyu.chen** on the 2024-04-26 09:38:59

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset fraud_transactions_by_merchant_id (*Snowflake*)
- Dataset top_merchants_sql (*Snowflake*)

List of the outputs of the recipe:

- Dataset top_merchants_joined (*Snowflake*)

c) Join

Left top_merchants_sql - fraud_transactions_by_merchant_id And

Left input	Condition	Right input
merchant_id	=	merchant_id

7. compute_top_merchants_joined_topn

Summarized information about the recipe compute_top_merchants_joined_topn.

a) Status

Characteristics of the recipe compute_top_merchants_joined_topn:

- Type: **Top N**
- Created by **siyu.chen** on the 2024-04-26 09:35:37
- Last Modification by **siyu.chen** on the 2024-04-26 09:36:13

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset top_merchants_joined ([Snowflake](#))

List of the outputs of the recipe:

- Dataset top_merchants_joined_topn ([Snowflake](#))

c) Top rows and sorting direction

Top rows: **10**

Bottom rows: **0**

Column name	Sorting direction
percent_fraud	Descending

8. compute_TRANSACTIONS_joined_prepared_windows

Summarized information about the recipe
compute_TRANSACTIONS_joined_prepared_windows.

a) Status

Characteristics of the recipe compute_TRANSACTIONS_joined_prepared_windows:

- Type: **Window**
- Created by **siyu.chen** on the 2024-04-26 08:28:18
- Last Modification by **siyu.chen** on the 2024-04-26 08:30:52

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_joined_prepared ([Snowflake](#))

List of the outputs of the recipe:

- Dataset TRANSACTIONS_joined_prepared_windows ([Snowflake](#))

c) Windows

Window name	Partitioning columns
	card_id
	merchant_id

9. split_TRANSACTIONS_joined_prepared

Summarized information about the recipe split_TRANSACTIONS_joined_prepared.

a) Status

Characteristics of the recipe split_TRANSACTIONS_joined_prepared:

- Type: **Split**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:12:30
- Last Modification by **siyu.chen** on the 2024-04-26 08:33:17

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_joined_prepared_windows ([Snowflake](#))

List of the outputs of the recipe:

- Dataset TRANSACTIONS_known ([Snowflake](#))
- Dataset TRANSACTIONS_unknown ([Snowflake](#))

Split mode: Specific filter for each output dataset

10. train_Predict_authorized_flag__binary__

Summarized information about the recipe train_Predict_authorized_flag__binary__.

a) Status

Characteristics of the recipe train_Predict_authorized_flag__binary__:

- Type: **Prediction Training**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:27:51

- Last Modification by **siyu.chen** on the 2024-04-26 08:52:22

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_known ([Snowflake](#))

List of the outputs of the recipe:

- Saved Model Predict authorized_flag (binary) ([LightGBM \(session 1 retrain\) - v1](#))

11. score_TRANSACTIONS_unknown

Summarized information about the recipe score_TRANSACTIONS_unknown.

a) Status

Characteristics of the recipe score_TRANSACTIONS_unknown:

- Type: **Prediction Scoring**
- Created by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:29:06
- Last Modification by **sylvain.franco@dataiku.com (deleted)** on the 2022-12-20 09:29:12

b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset TRANSACTIONS_unknown ([Snowflake](#))
- Saved Model Predict authorized_flag (binary) ([LightGBM \(session 1 retrain\) - v1](#))

List of the outputs of the recipe:

- Dataset TRANSACTIONS_unknown_scored ([Snowflake](#))

V. Saved Models

List all the saved models in the flow.

1. Predict authorized_flag (binary)

Summarized information about the saved model Predict authorized_flag (binary):.

a) Status

Model id: **3yL0JQWo**

Active version: **LightGBM (session 1 retrain) - v1**

b) Related Recipes

List of the recipes related to the saved model.

Parent Recipe

Recipe with Predict authorized_flag (binary) as an output:

- train_Predict_authorized_flag__binary_ ([Prediction Training](#))

Successor Recipes

List of recipes with Predict authorized_flag (binary) as an input:

- score_TRANSACTIONS_unknown ([Prediction Scoring](#))

VI. Deployment and Monitoring

A. Implementation Details

- The flow can be found here: <https://dss-ffe10b81-f9822877-dku.ap-southeast-1.app.dataiku.io/projects/SIYUADVANCEDDESIGNER/>
- The name of the generated file is: Dataiku Flow Documentation - SIYUADVANCEDDESIGNER.docx

B. Version Control

- The flow was executed with the following version of DSS: 12.6.0