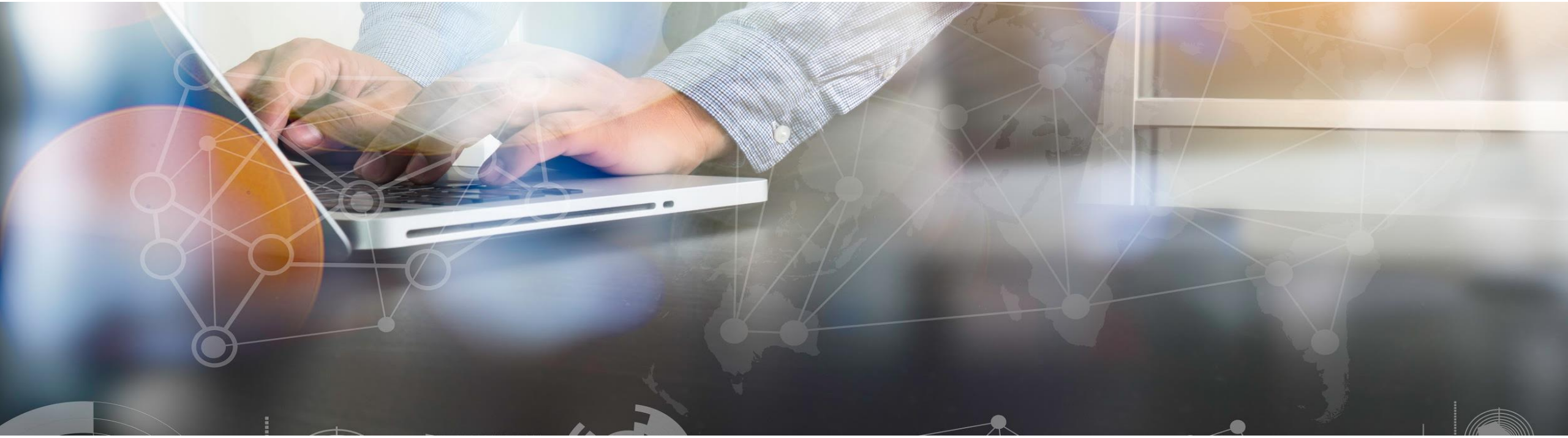
meaningful DATA

Making your data **worthwhile.**

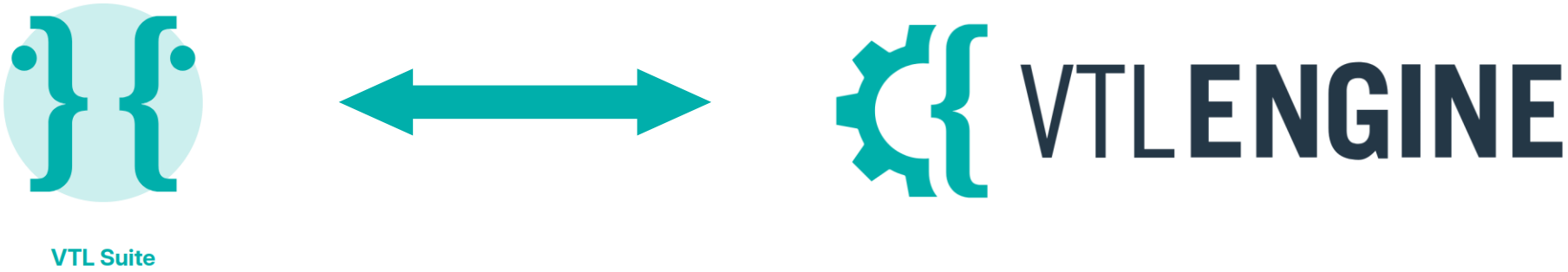


VTL for Data Science

Gustavo Homem
gustavo.homem@meaningfuldata.eu

SDMX Global Conference
Rome 29/09/2025

The VTL Suite

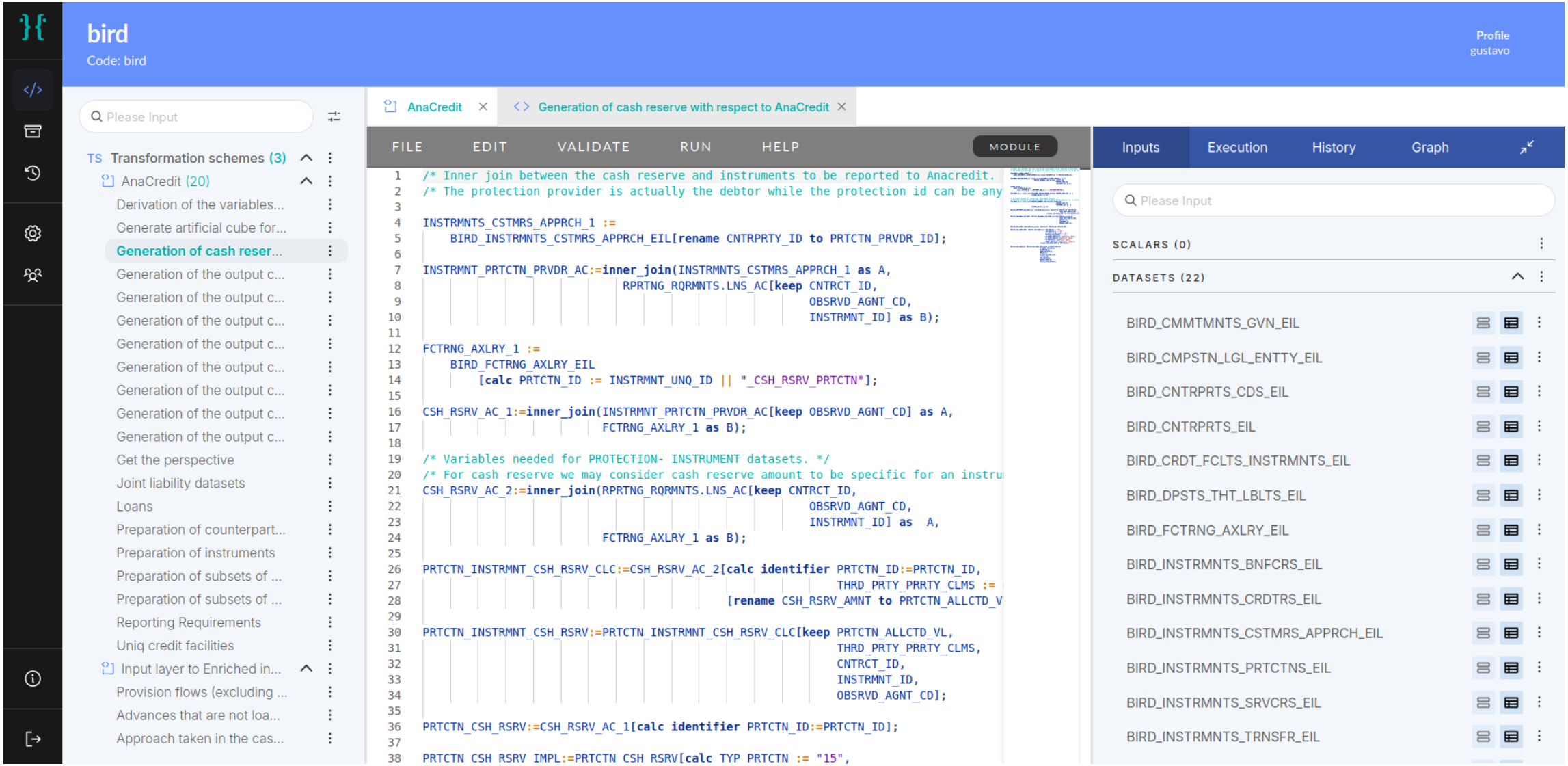


A VTL-centric VTL experience

~~**friendly to business users**~~

friendly to every user

Opinionated VTL IDE



Integrated user manual

Quick search

Go

Transformation Schemes and Modules

Transformation schemes are artefacts defined in the VTL standard that represent the smaller set of transformations to be executed in the same run, so that they can be considered a VTL program.

The individual transformations composing a transformation scheme do not have any specific order or grouping. This circumstance can lead to difficulties in the interpretation and maintenance of VTL code, when there is a high number of individual transformations. The VTL standard does not provide any means to support a meaningful organisation of the transformations.

The vtlSuite is providing an artefact called Modules with the aim of assisting with this organisation. The concept of a module is not part of the standard, but the vtlSuite is able to build a fully VTL compliant version of the transformation scheme from the modules. In the vtlSuite, the transformation scheme is created with an algorithm from the modules belonging to the transformation scheme and the library items used in the modules.

Referencing output of transformation

In VTL, the output of a transformation can be used as input for other transformations. An example of this would be the following VTL script:

```
DSr1 := DS1 +DS2;
DSr2 := DSr1 * 5;
```

The second transformation of the script is taking as an input the output of the first transformation. This is known because both the output of transformation 1 and one of the inputs of transformation 2 share the same name.

The vtlSuite uses a namespace convention for dealing with the modules. For referencing the output of a transformation located in the same module, the reference should use the same name that transformationScheme

As an example, let's consider a transformation scheme with two modules (with codes M1 and M2).

M1:

Builtin debugging

Inputs | Execution | History | Graph

Please Input

INTERMEDIATE RESULTS (6)

- inflation
- inflation_base
- inflation_base_01
- inflation_q_01
- inflation_q_02
- inflation_q_03

PERSISTENT RESULTS (1)

output_inflation

Name	Role	Type	Nulla...
Q	Identifier	String	No
year	Identifier	Integer	No
coefficient	Measure	Number	No
coefficient_inv	Measure	Number	No

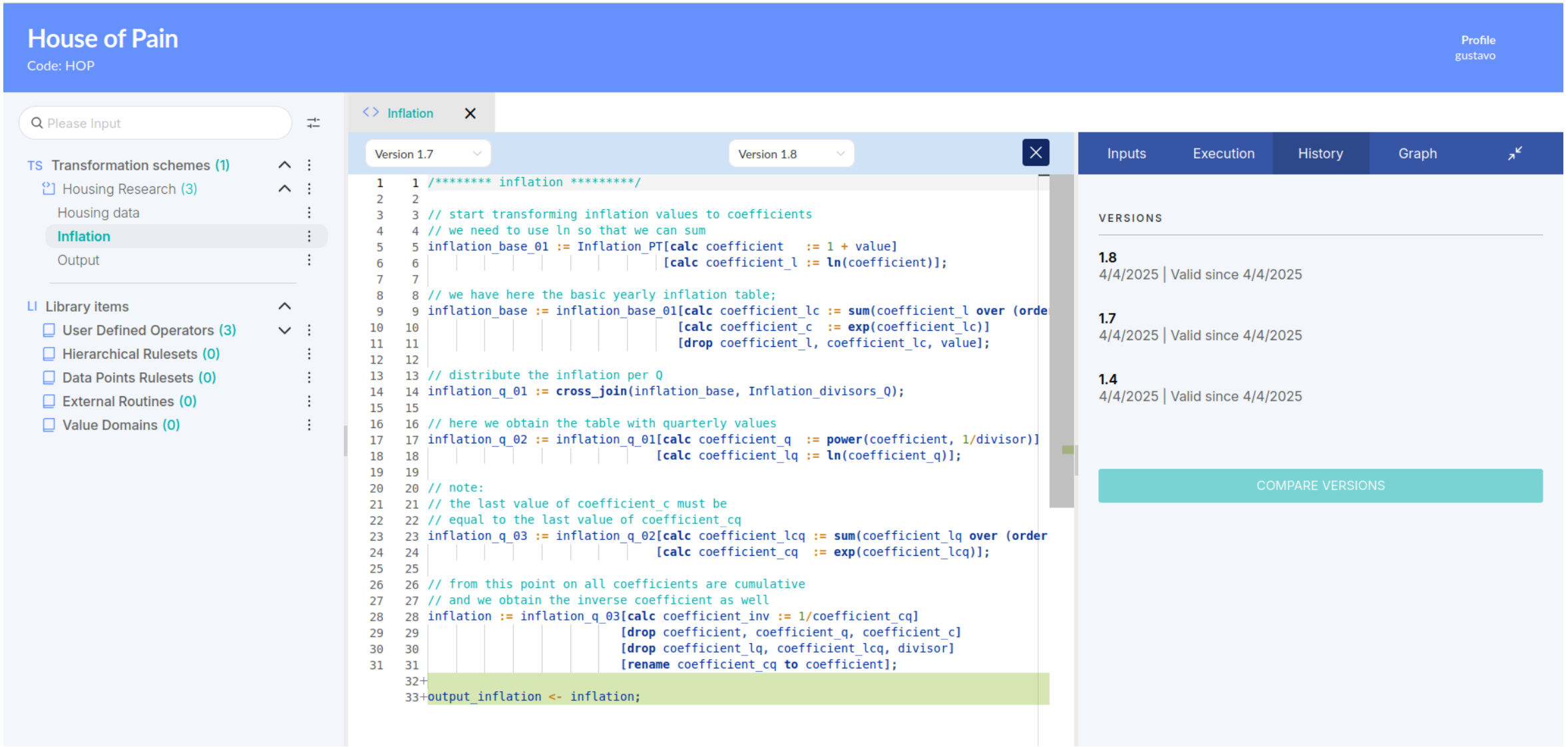
Table of data points from output_inflation (76 rows x 4 columns)

Q	year	coefficient	coefficient_inv
1	2007	1	1
2	2007	1	1
3	2007	1	1
4	2007	1	1
1	2008	1.0064130451055977	0.9936278199723406
2	2008	1.0128672173587219	0.9872962446229862
3	2008	1.0193627805096246	0.9810050152116165
4	2008	1.0259000000000003	0.9747538746466514
1	2009	1.0237387902252366	0.9768116726142478
2	2009	1.0215821333578619	0.978873814788711
3	2009	1.0194300198065052	0.9809403103410735
4	2009	1.0172824400000002	0.9830111684617301
1	2010	1.0208243871770566	0.9796004215429811
2	2010	1.0243786666124042	0.9762015088687624
3	2010	1.0279453212441747	0.9728143893778795

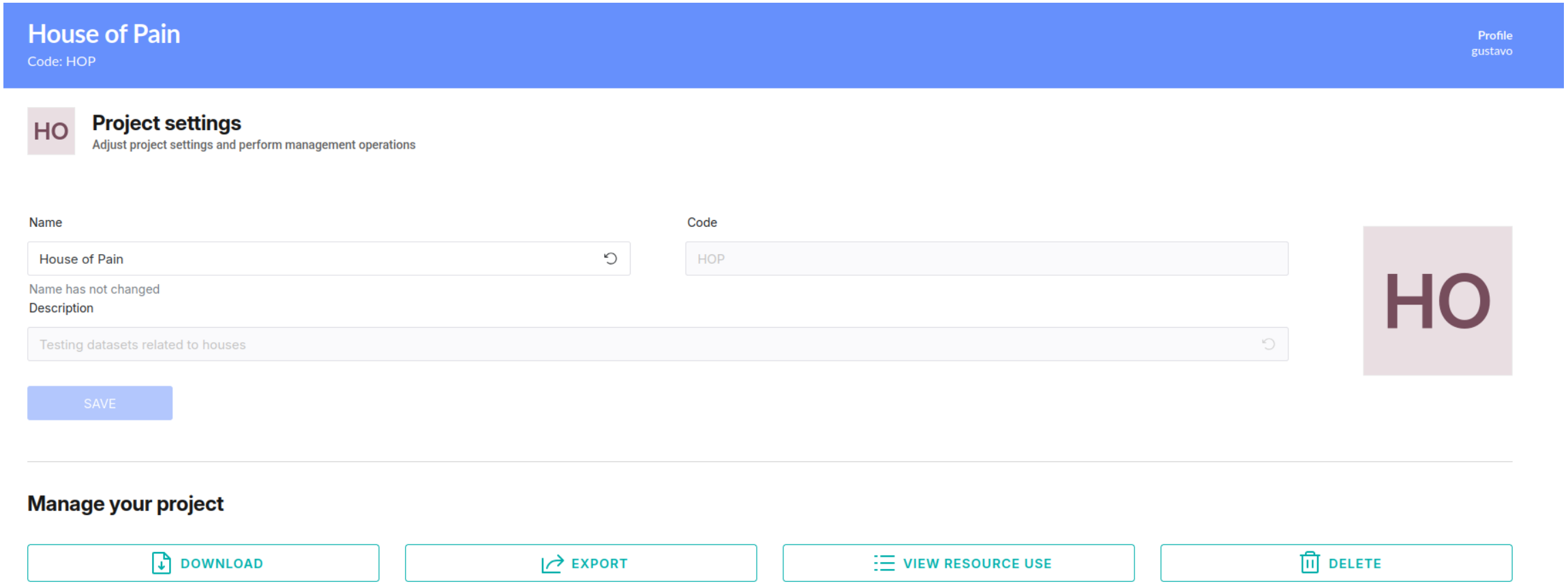
Table | Component explorer | Relationship explorer

Download data points file as CSV or JSON

Builtin version control



Project backup / restore



Users and Teams

House of Pain

Code: HOP

Profile
gustavo

HO

Project users

Add / remove or change roles of users in this project

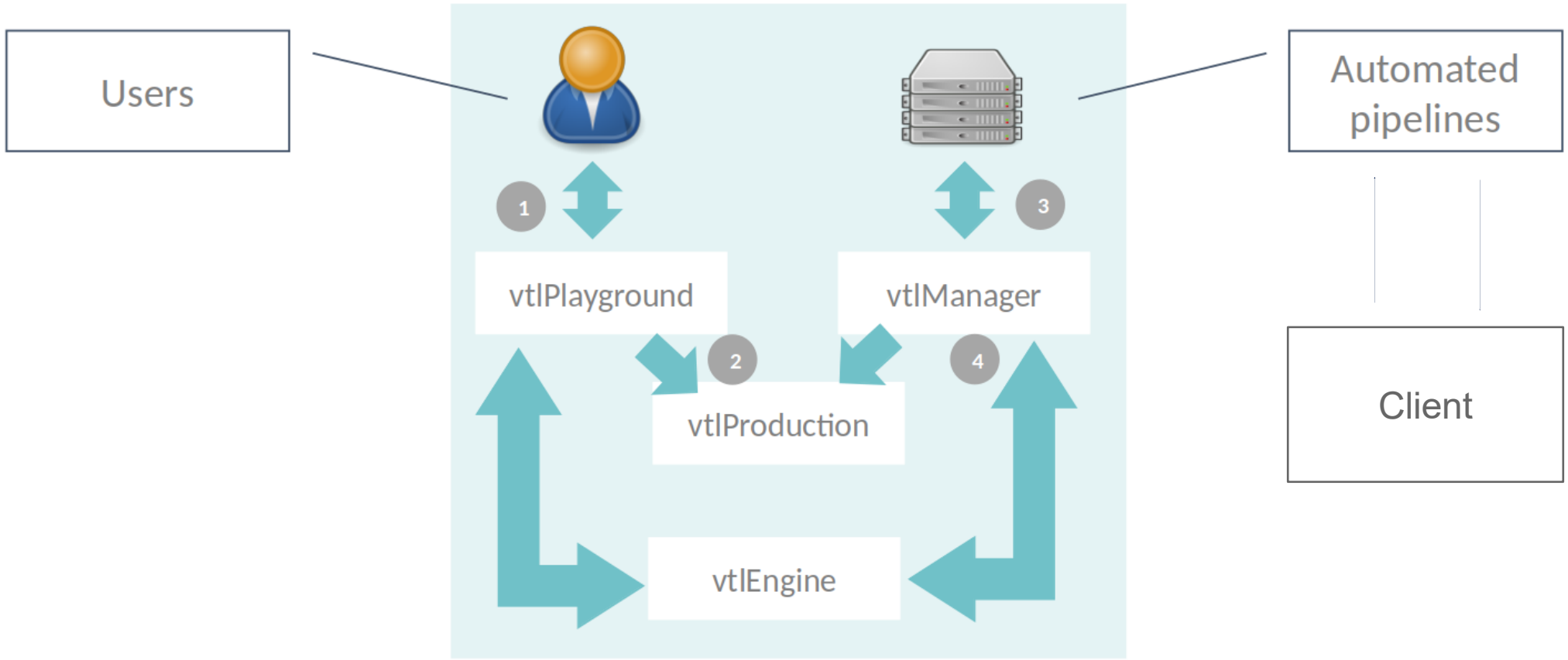
Search by

Sort by
Username (A-Z)

User	First name	Last name	Team relation	Project role
allyson	-	-	Member	Administrator
antonio	-	-	Member	Collaborator
gustavo	-	-	Member	Administrator
javier	-	-	Member	Collaborator
jesus	-	-	Member	Collaborator
pablo	-	-	Member	Collaborator

1

Web + REST + queues



Data Science

KYC

Data Science

Data Science

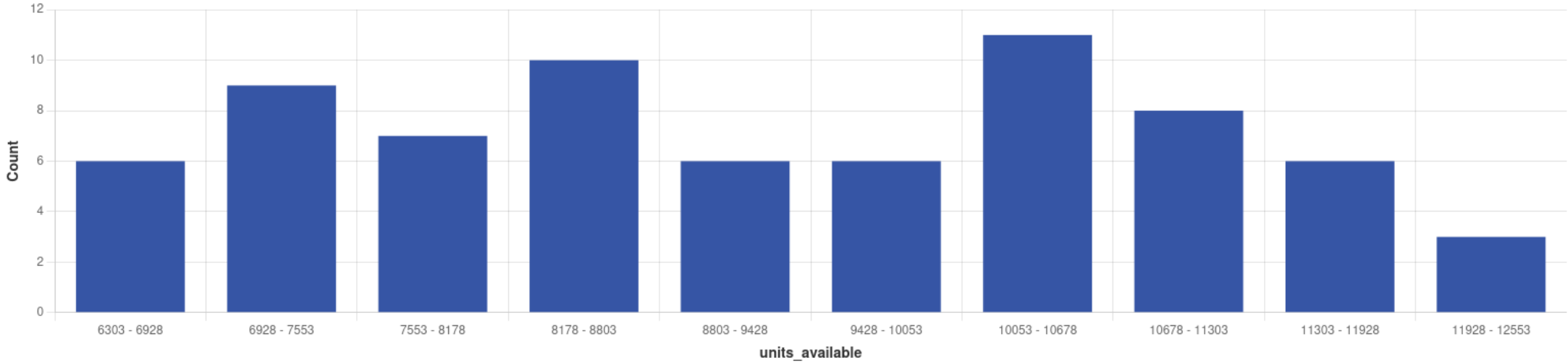
Component explorer of data points from output_generic (73 rows x 81 columns)

Component to explore

units_available (Integer)

Maximum	Minimum	Average	Median	Standard deviation
12553	6303	9287	9189	1696

Count distribution



Data Science

Component explorer of data points from output_generic (73 rows x 81 columns)

Component to explore

transaction_price_d (Number)

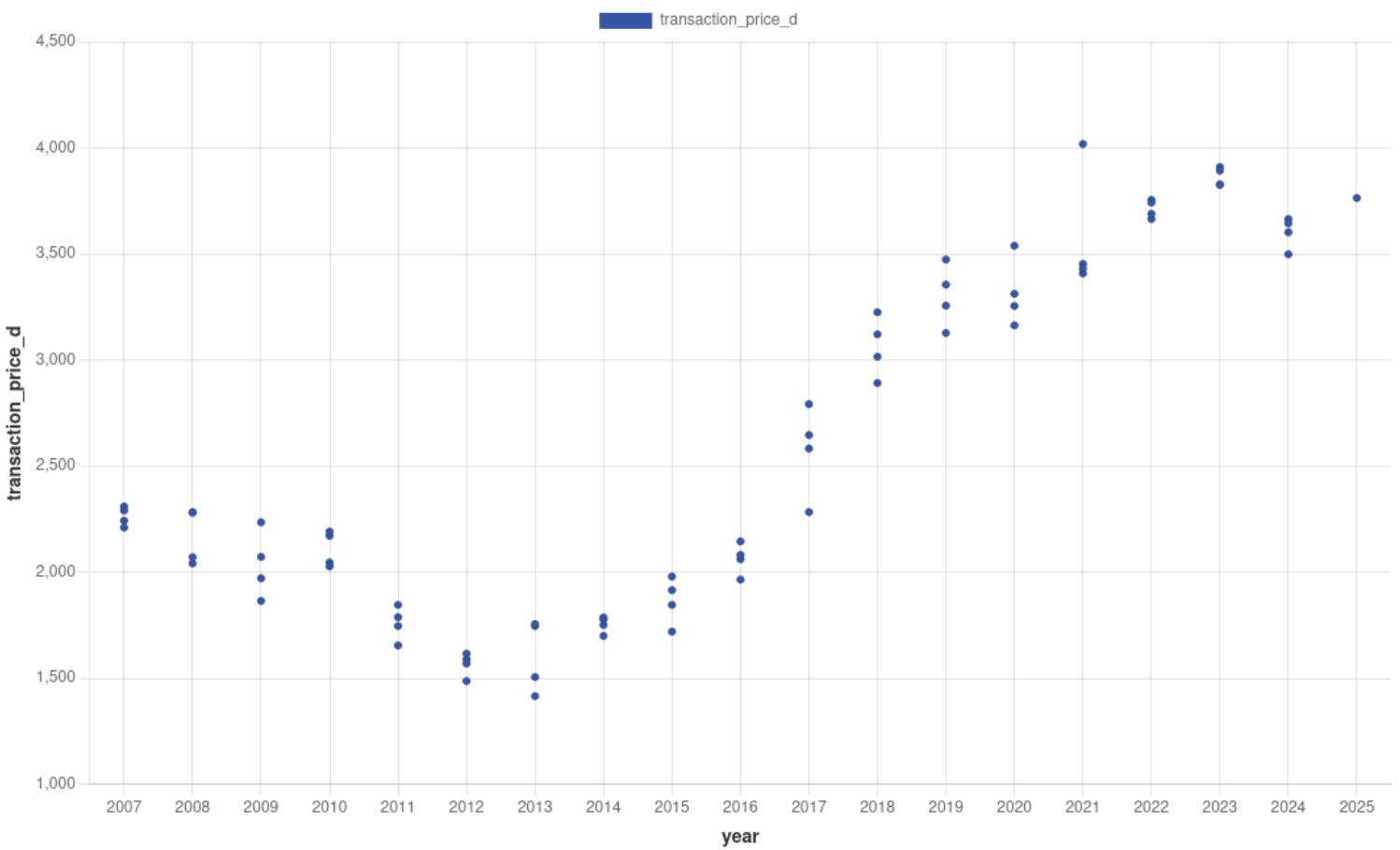
Maximum	Minimum	Average	Median	Standard deviation
4018.53	1416.97	2584.02	2280.65	802.84

Count distribution

Data Science

Data Science

Relationship explorer of data points from output_generic (73 rows x 81 columns)



X axis

year (Number)

Return to default

Y axis

transaction_price_d (Number)

X sorting

Alphabetical By average Y value

Y transformation

None Logarithm Exponential Sum

Average Median Count

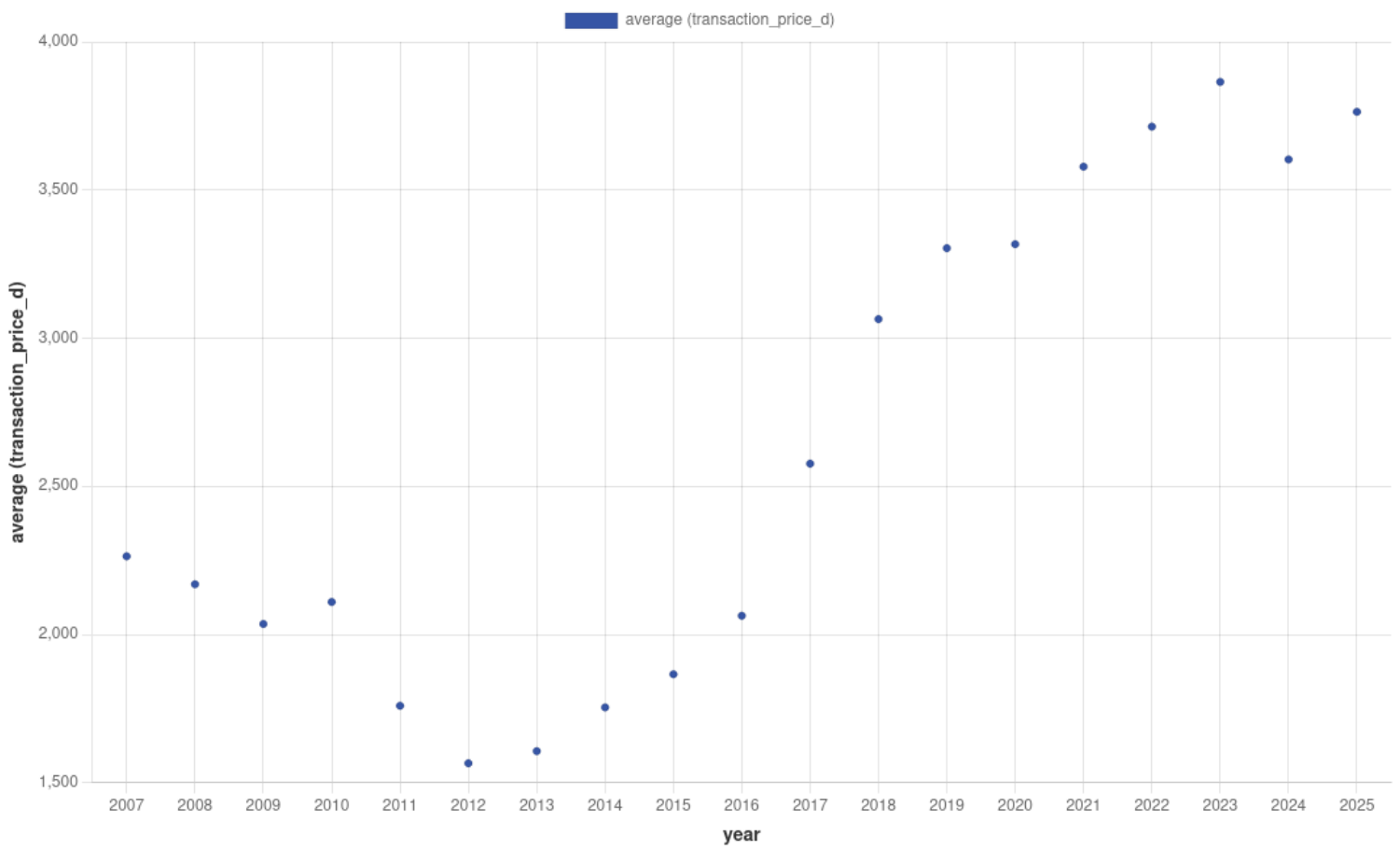
Data range

19 / 19 data points selected

GENERATE

Data Science

Relationship explorer of data points from output_generic (73 rows x 81 columns)



X axis

year (Number)

Return to default

Y axis

transaction_price_d (Number)

X sorting

Alphabetical By average Y value

Y transformation

None Logarithm Exponential Sum

Average Median Count

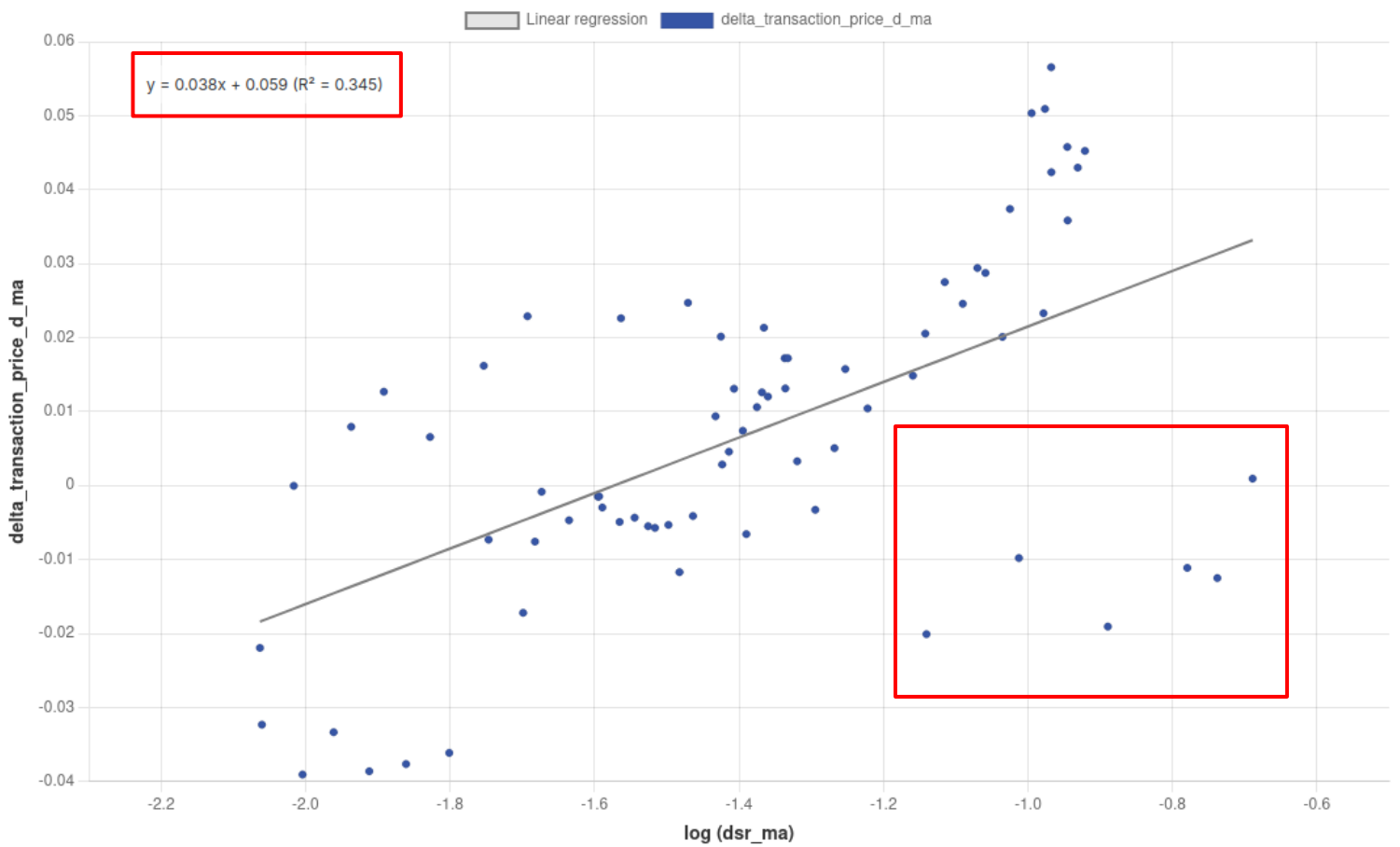
Data range

19 / 19 data points selected

GENERATE

Data Science

Relationship explorer of data points from output_generic_eda (73 rows x 12 columns)



X axis

dsr_ma (Number)

Force categorical

Y axis

delta_transaction_price_d_ma (Number)

X transformation

None | Logarithm | Exponential

Y transformation

None | Logarithm | Exponential

Functions

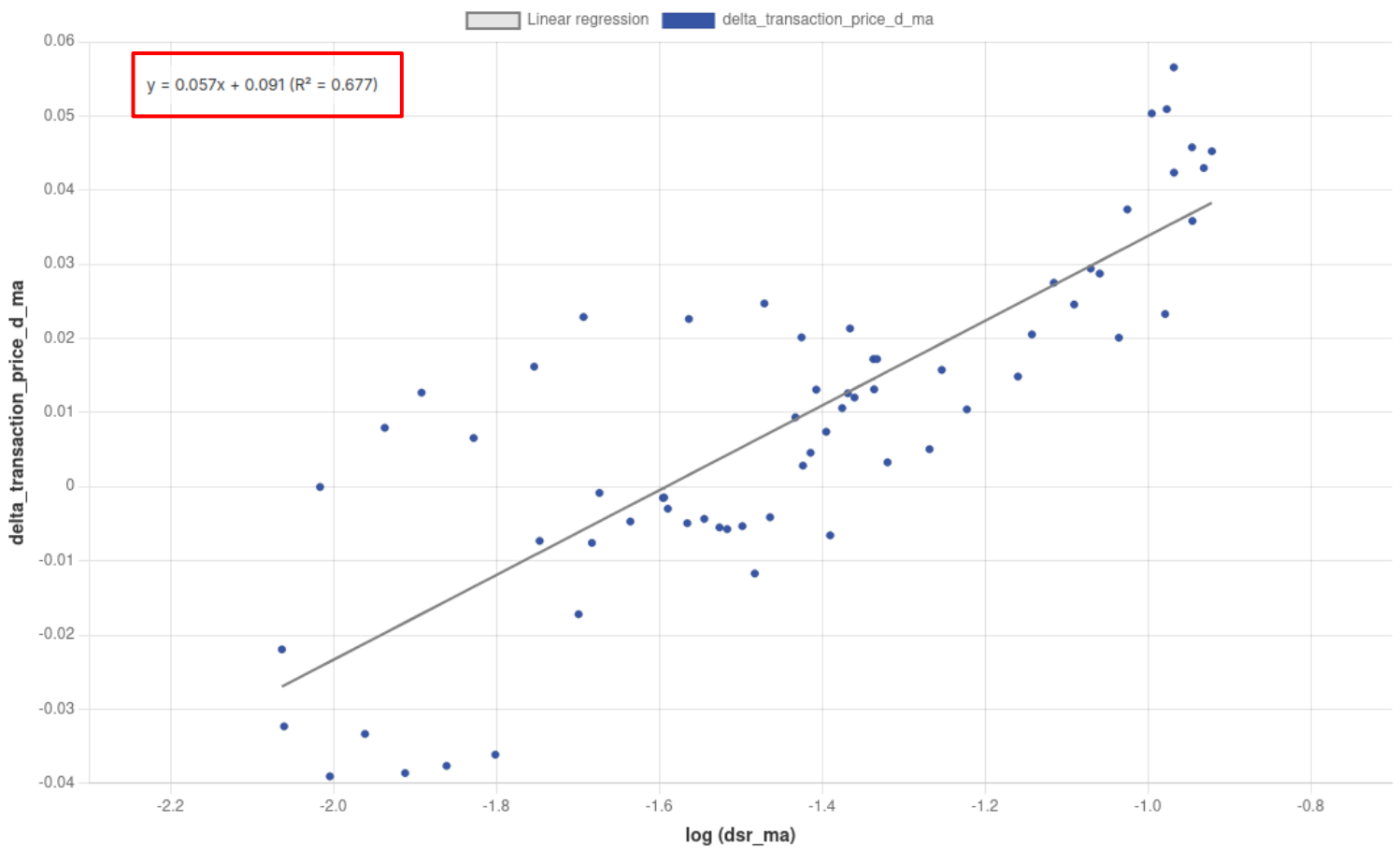
None | Linear regression | Fitted lowess curve

* You are using an exponential or logarithmic transformation: if the result value tends to infinity the data will not be displayed.

GENERATE

Data Science

Relationship explorer of data points from output_generic_eda (65 rows x 12 columns)



X axis

dsr_ma (Number)

Force categorical

Y axis

delta_transaction_price_d_ma (Number)

X transformation

None | Logarithm | Exponential

Y transformation

None | Logarithm | Exponential

Functions

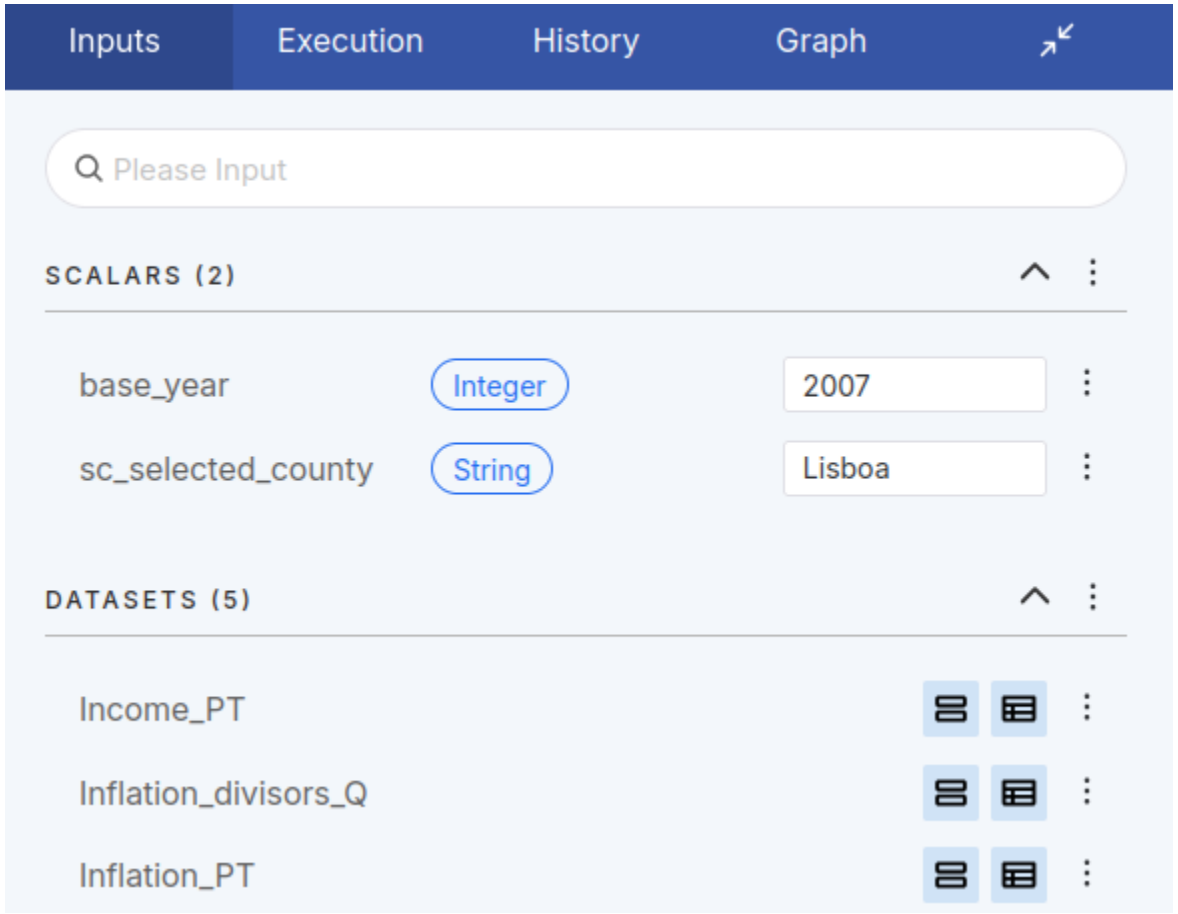
None | Linear regression | Fitted lowess curve

* You are using an exponential or logarithmic transformation: if the result value tends to infinity the data will not be displayed.

GENERATE

Data Science

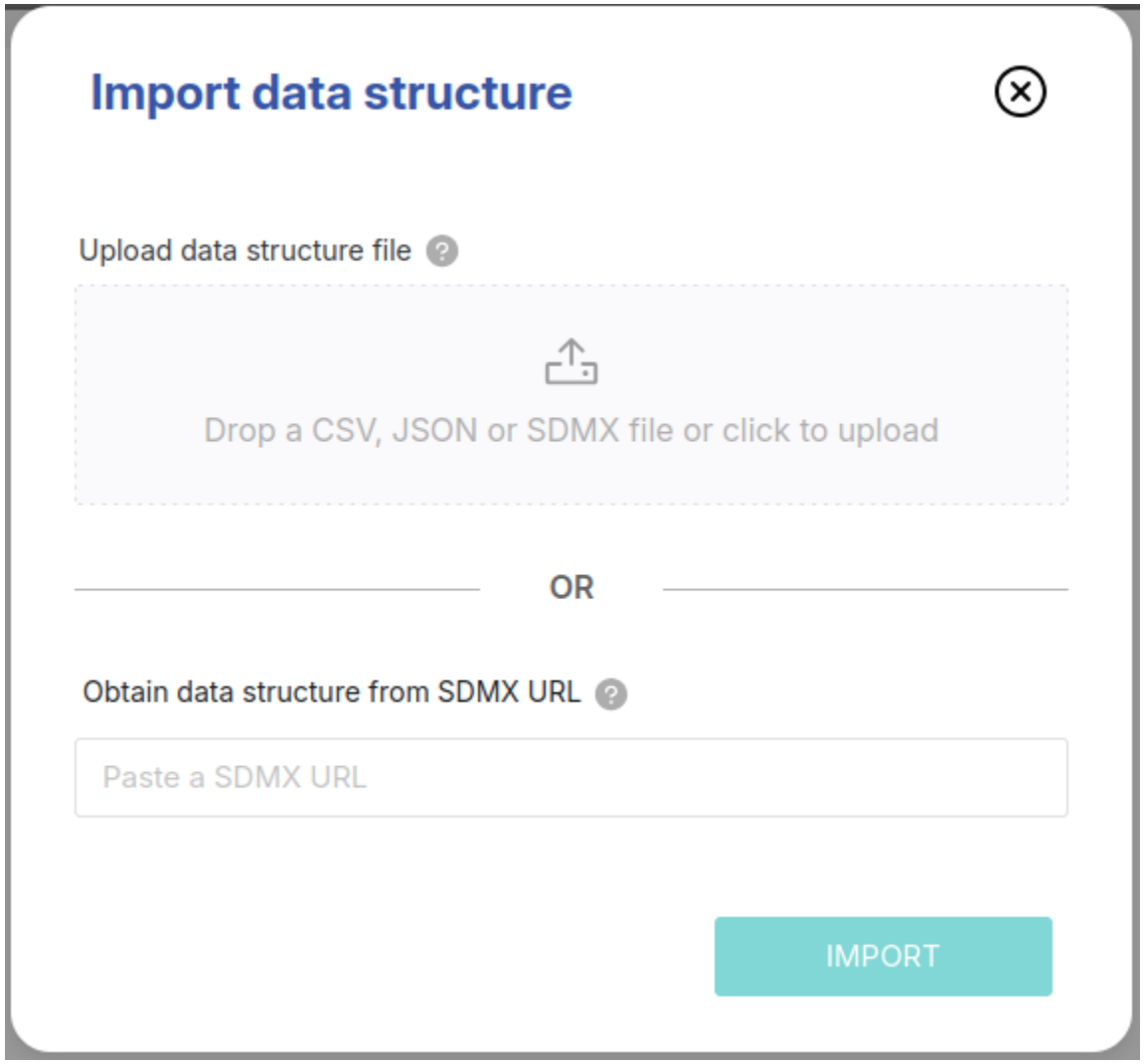
Parameters

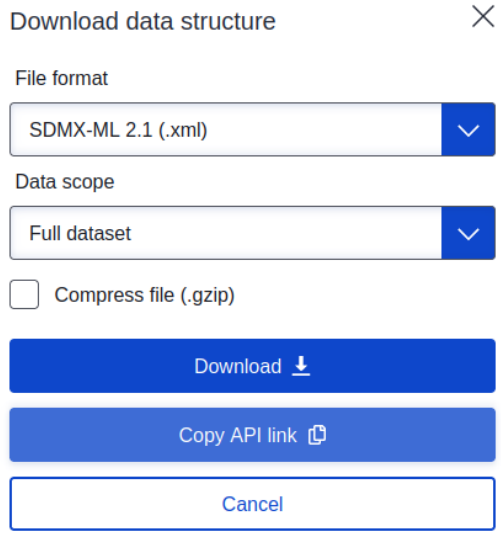


Real GDP series!

(from Eurostat)

Data Science





Data Science

Table of data points from TIPSNA40 (865 rows x 8 columns)

freq	geo	na_item	TIME_PERIOD	unit	OBS_VALUE	CONF_STATUS	OBS_FLAG
A	AT	B1GQ	1995	CLV15_EUR_HAB	30290	-	-
A	AT	B1GQ	1996	CLV15_EUR_HAB	30920	-	-
A	AT	B1GQ	1997	CLV15_EUR_HAB	31540	-	-
A	AT	B1GQ	1998	CLV15_EUR_HAB	32610	-	-
A	AT	B1GQ	1999	CLV15_EUR_HAB	33770	-	-
A	AT	B1GQ	2000	CLV15_EUR_HAB	34760	-	-
A	AT	B1GQ	2001	CLV15_EUR_HAB	35090	-	-
A	AT	B1GQ	2002	CLV15_EUR_HAB	35430	-	-
A	AT	B1GQ	2003	CLV15_EUR_HAB	35680	-	-
A	AT	B1GQ	2004	CLV15_EUR_HAB	36360	-	-
A	AT	B1GQ	2005	CLV15_EUR_HAB	36950	-	-
A	AT	B1GQ	2006	CLV15_EUR_HAB	37970	-	-
A	AT	B1GQ	2007	CLV15_EUR_HAB	39270	-	-
A	AT	B1GQ	2008	CLV15_EUR_HAB	39710	-	-
A	AT	B1GQ	2009	CLV15_EUR_HAB	38200	-	-

Table | Component explorer | Relationship explorer

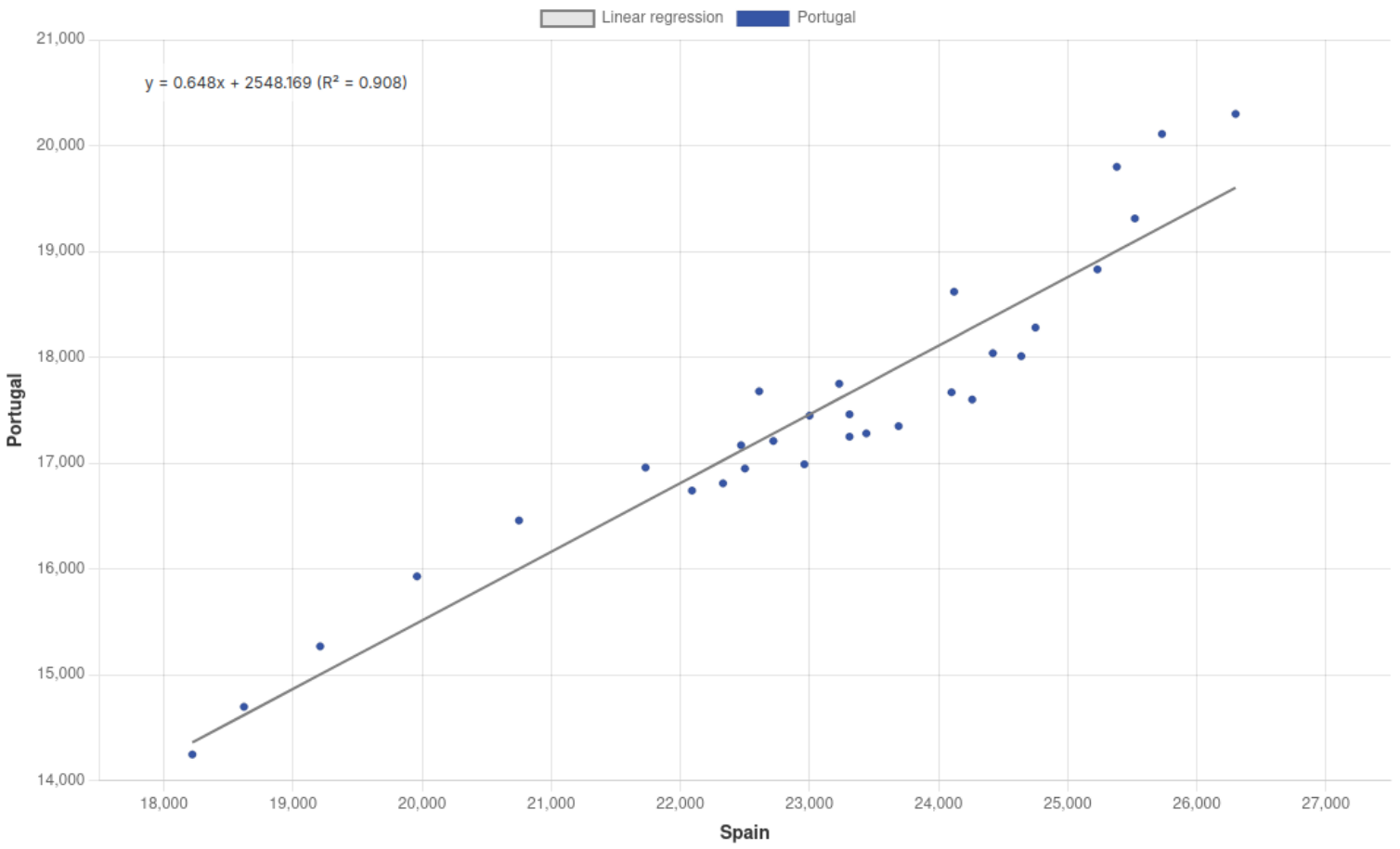
Download data points file as CSV or JSON

Data Science


```
// we are using the real GDP metric from here
// https://ec.europa.eu/eurostat/databrowser/view/tipsna40/default/table?lang=en

// data structure:
// https://ec.europa.eu/eurostat/api/dissemination/sdmx/2.1/dataflow/ESTAT/tipsna40/

// data points
// https://ec.europa.eu/eurostat/api/dissemination/sdmx/3.0/data/dataflow/ESTAT/tips

estat_real_gdp_portugal <- TIPSNA40[sub geo = "PT"]
                                   [keep OBS_VALUE]
                                   [rename OBS_VALUE to Portugal];

estat_real_gdp_italy <- TIPSNA40[sub geo = "IT"]
                                [keep OBS_VALUE]
                                [rename OBS_VALUE to Italy];

estat_real_gdp_spain <- TIPSNA40[sub geo = "ES"]
                                [keep OBS_VALUE]
                                [rename OBS_VALUE to Spain];

estat_real_gdp_combined <- inner_join(
                                estat_real_gdp_portugal,
                                estat_real_gdp_italy,
                                estat_real_gdp_spain);
```

Data Science

Relationship explorer of data points from estat_real_gdp_combined (30 rows x 7 columns)



X axis

Spain (Number)

Force categorical

Y axis

Portugal (Number)

X transformation

None Logarithm Exponential

Y transformation

None Logarithm Exponential

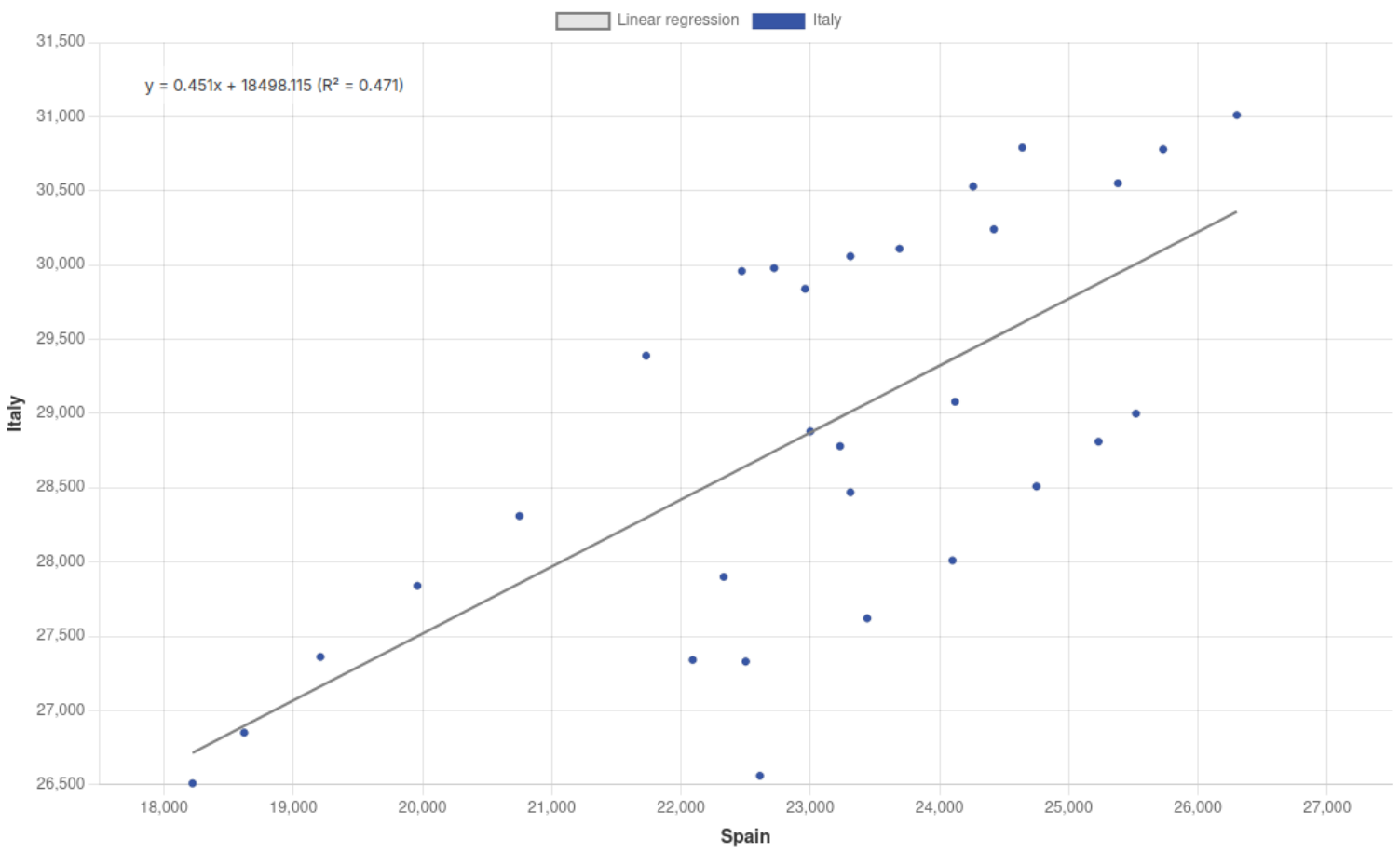
Functions

None Linear regression Fitted lowess curve

GENERATE

Data Science

Relationship explorer of data points from estat_real_gdp_combined (30 rows x 7 columns)



X axis

Spain (Number)

Force categorical

Y axis

Italy (Number)

X transformation

None Logarithm Exponential

Y transformation

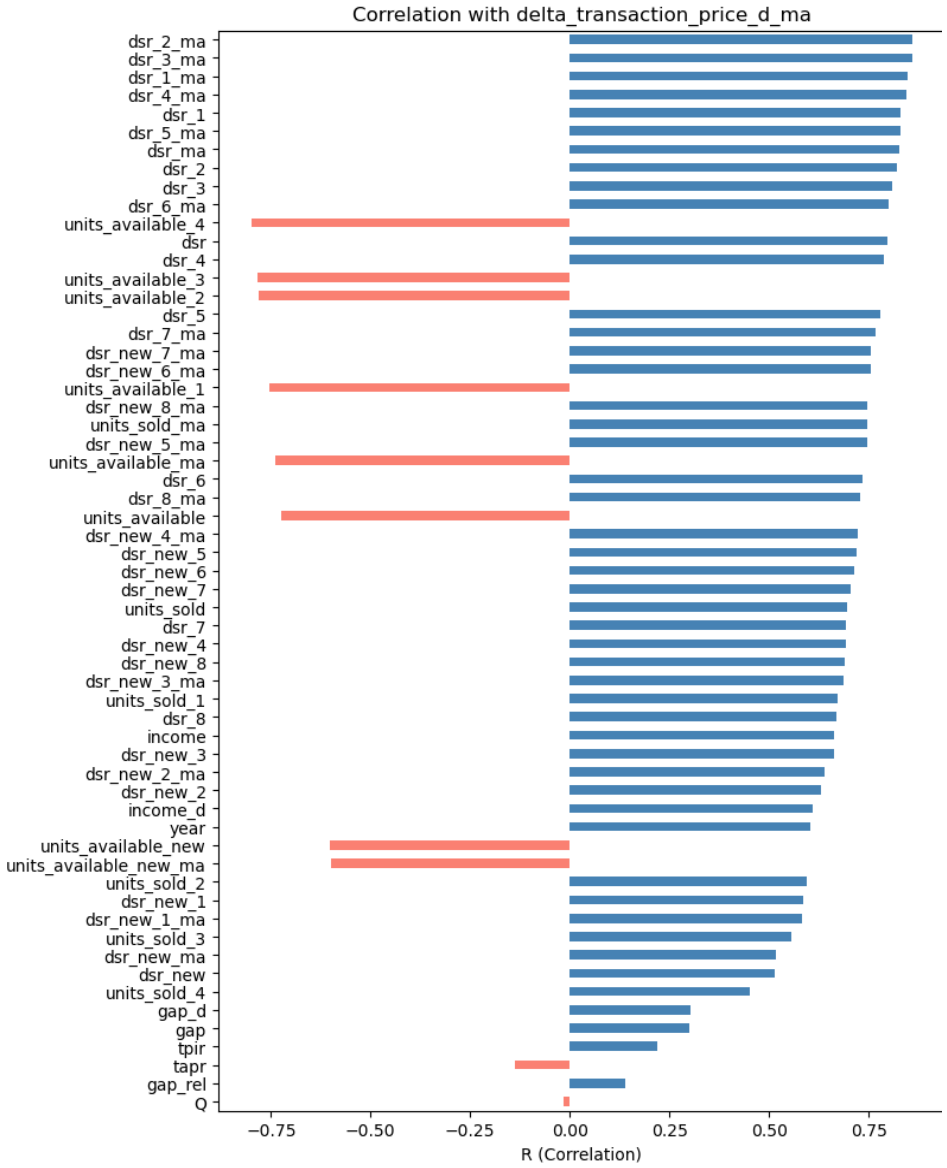
None Logarithm Exponential

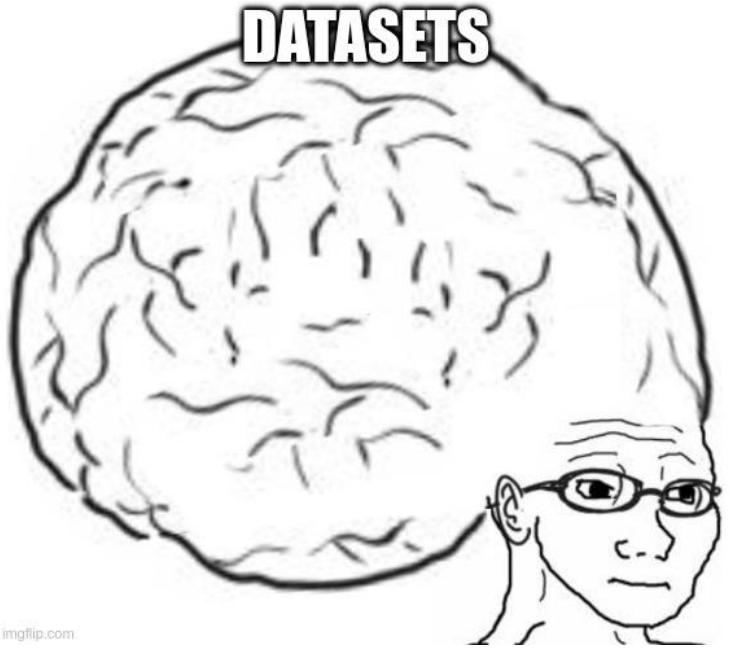
Functions

None Linear regression Fitted lowess curve

GENERATE

What else?





Datasets larger than memory?

Let us know!

See you at our booth!

Gustavo Homem
gustavo.homem@meaningfuldata.eu

www.meaningfuldata.eu