



VTL for Data Science

Gustavo Homem
gustavo.homem@meaningfuldata.eu

*SDMX Global Conference
Rome 29/09/2025*

The VTL Suite



VTL Suite



A VTL-centric VTL experience
~~friendly to business users~~
friendly to every user

Opinionated VTL IDE



bird

Code: bird

Profile

gustavo

TS Transformation schemes (3)

AnaCredit (20)

Derivation of the variables...

Generate artificial cube for...

Generation of cash reser...

Generation of the output c...

Generation of the output c...

Generation of the output c...

Generation of the output c...

Generation of the output c...

Generation of the output c...

Get the perspective

Joint liability datasets

Loans

Preparation of counterpart...

Preparation of instruments

Preparation of subsets of ...

Preparation of subsets of ...

Reporting Requirements

Uniq credit facilities

Input layer to Enriched in...

Provision flows (excluding ...

Advances that are not loa...

Approach taken in the cas...

AnaCredit

Generation of cash reserve with respect to AnaCredit

FILE

EDIT

VALIDATE

RUN

HELP

MODULE

```
1  /* Inner join between the cash reserve and instruments to be reported to Anacredit.
2  /* The protection provider is actually the debtor while the protection id can be any
3
4  INSTRMNTS_CSTMRS_APPRCH_1 :=
5      BIRD_INSTRMNTS_CSTMRS_APPRCH_EIL[rename CNTRPTY_ID to PRCTCN_PRVDR_ID];
6
7  INSTRMNT_PRCTCN_PRVDR_AC:=inner_join(INSTRMNTS_CSTMRS_APPRCH_1 as A,
8      RPRTNG_RQRMNTS.LNS_AC[keep CNTRCT_ID,
9      OBSRVD_AGNT_CD,
10     INSTRMNT_ID] as B);
11
12  FCTRNG_AXLRY_1 :=
13      BIRD_FCTRNG_AXLRY_EIL
14      [calc PRCTCN_ID := INSTRMNT_UNQ_ID || " _CSH_RSRV_PRCTCN"];
15
16  CSH_RSRV_AC_1:=inner_join(INSTRMNT_PRCTCN_PRVDR_AC[keep OBSRVD_AGNT_CD] as A,
17      FCTRNG_AXLRY_1 as B);
18
19  /* Variables needed for PROTECTION- INSTRUMENT datasets. */
20  /* For cash reserve we may consider cash reserve amount to be specific for an instru
21  CSH_RSRV_AC_2:=inner_join(RPRTNG_RQRMNTS.LNS_AC[keep CNTRCT_ID,
22      OBSRVD_AGNT_CD,
23      INSTRMNT_ID] as A,
24      FCTRNG_AXLRY_1 as B);
25
26  PRCTCN_INSTRMNT_CSH_RSRV_CLC:=CSH_RSRV_AC_2[calc identifier PRCTCN_ID:=PRCTCN_ID,
27      THRD_PRTY_PRTY_CLMS :=
28      [rename CSH_RSRV_AMNT to PRCTCN_ALLCTD_V
29
30  PRCTCN_INSTRMNT_CSH_RSRV:=PRCTCN_INSTRMNT_CSH_RSRV_CLC[keep PRCTCN_ALLCTD_VL,
31      THRD_PRTY_PRTY_CLMS,
32      CNTRCT_ID,
33      INSTRMNT_ID,
34      OBSRVD_AGNT_CD];
35
36  PRCTCN_CSH_RSRV:=CSH_RSRV_AC_1[calc identifier PRCTCN_ID:=PRCTCN_ID];
37
38  PRCTCN_CSH_RSRV_IMPL:=PRCTCN_CSH_RSRV[calc TYP PRCTCN := "15",
```

Inputs

Execution

History

Graph

Please Input

SCALARS (0)

DATASETS (22)

BIRD_CMMTMNTS_GVN_EIL

BIRD_CMPSTN_LGL_ENTTY_EIL

BIRD_CNTRPRTS_CDS_EIL

BIRD_CNTRPRTS_EIL

BIRD_CRDT_FCLTS_INSTRMNTS_EIL

BIRD_DPSTS_THT_LBLTS_EIL

BIRD_FCTRNG_AXLRY_EIL

BIRD_INSTRMNTS_BNFCRS_EIL

BIRD_INSTRMNTS_CRDTRS_EIL

BIRD_INSTRMNTS_CSTMRS_APPRCH_EIL

BIRD_INSTRMNTS_PRCTCN_EIL

BIRD_INSTRMNTS_SRVCRS_EIL

BIRD_INSTRMNTS_TRNSFR_EIL

VTL Playground Manual

Navigation

[Introduction](#)
[Overview](#)
[Organisation](#)
[Working with Vtl Playground](#)
[Transformation Schemes and Modules](#)

- [Referencing output of transformation](#)
- [Building transformation schemes from modules and library items: VTL strict](#)
- [Creating new transformation schemes and modules](#)
- [Deleting transformation schemes and modules](#)
- [Options for transformation schemes and modules](#)

[Datasets](#)
[Library Items](#)
[Concurrency](#)
[Versioning](#)
[Auditing](#)
[User Roles and actions](#)

Quick search

Transformation Schemes and Modules

- [Referencing output of transformation](#)
- [Building transformation schemes from modules and library items: VTL strict](#)
- [Creating new transformation schemes and modules](#)
- [Deleting transformation schemes and modules](#)
- [Options for 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

Q 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	Q	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	Q	Q year	Q coefficient	Q 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

Builtin version control



6

House of Pain

Code: HOP

Profile
gustavo

Q Please Input

TS Transformation schemes (1)

Housing Research (3)

Housing data

Inflation

Output

Library items

User Defined Operators (3)

Hierarchical Rulesets (0)

Data Points Rulesets (0)

External Routines (0)

Value Domains (0)

<> Inflation X

Version 1.7

Version 1.8

```
1 1 /***** inflation *****/
2 2
3 3 // start transforming inflation values to coefficients
4 4 // we need to use ln so that we can sum
5 5 inflation_base_01 := Inflation_PT[calc coefficient := 1 + value]
6 6 | | | | | | | | | | [calc coefficient_l := ln(coefficient)];
7 7
8 8 // we have here the basic yearly inflation table;
9 9 inflation_base := inflation_base_01[calc coefficient_lc := sum(coefficient_l over (orde
10 10 | | | | | | | | | | [calc coefficient_c := exp(coefficient_lc)]
11 11 | | | | | | | | | | [drop coefficient_l, coefficient_lc, value];
12 12
13 13 // distribute the inflation per Q
14 14 inflation_q_01 := cross_join(inflation_base, Inflation_divisors_Q);
15 15
16 16 // here we obtain the table with quarterly values
17 17 inflation_q_02 := inflation_q_01[calc coefficient_q := power(coefficient, 1/divisor)]
18 18 | | | | | | | | | | [calc coefficient_lq := ln(coefficient_q)];
19 19
20 20 // note:
21 21 // the last value of coefficient_c must be
22 22 // equal to the last value of coefficient_cq
23 23 inflation_q_03 := inflation_q_02[calc coefficient_lcq := sum(coefficient_lq over (order
24 24 | | | | | | | | | | [calc coefficient_cq := exp(coefficient_lcq)];
25 25
26 26 // from this point on all coefficients are cumulative
27 27 // and we obtain the inverse coefficient as well
28 28 inflation := inflation_q_03[calc coefficient_inv := 1/coefficient_cq]
29 29 | | | | | | | | | | [drop coefficient, coefficient_q, coefficient_c]
30 30 | | | | | | | | | | [drop coefficient_lq, coefficient_lcq, divisor]
31 31 | | | | | | | | | | [rename coefficient_cq to coefficient];
32+
33+output_inflation <- inflation;
```

Inputs

Execution

History

Graph



VERSIONS

1.8
4/4/2025 | Valid since 4/4/2025

1.7
4/4/2025 | Valid since 4/4/2025

1.4
4/4/2025 | Valid since 4/4/2025

COMPARE VERSIONS

Project backup / restore



House of Pain

Code: HOP

Profile
gustavo



Project settings

Adjust project settings and perform management operations

Name

House of Pain



Code

HOP

Name has not changed

Description

Testing datasets related to houses



HO

SAVE

Manage your project



DOWNLOAD



EXPORT



VIEW RESOURCE USE



DELETE

Users and Teams



House of Pain

Code: HOP

Profile
gustavo



Project users

Add / remove or change roles of users in this project

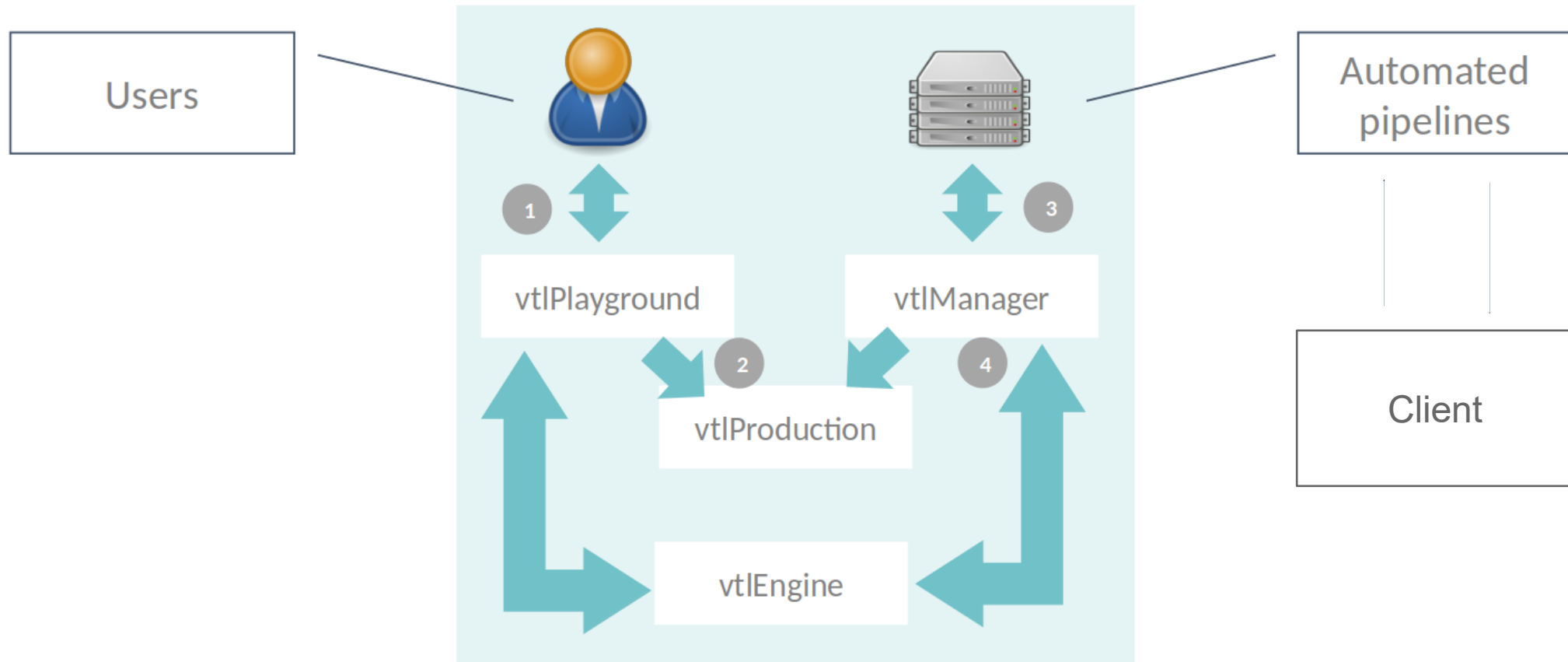


Sort by

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	⋮

Web + REST + queues





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

12553

Minimum

6303

Average

9287

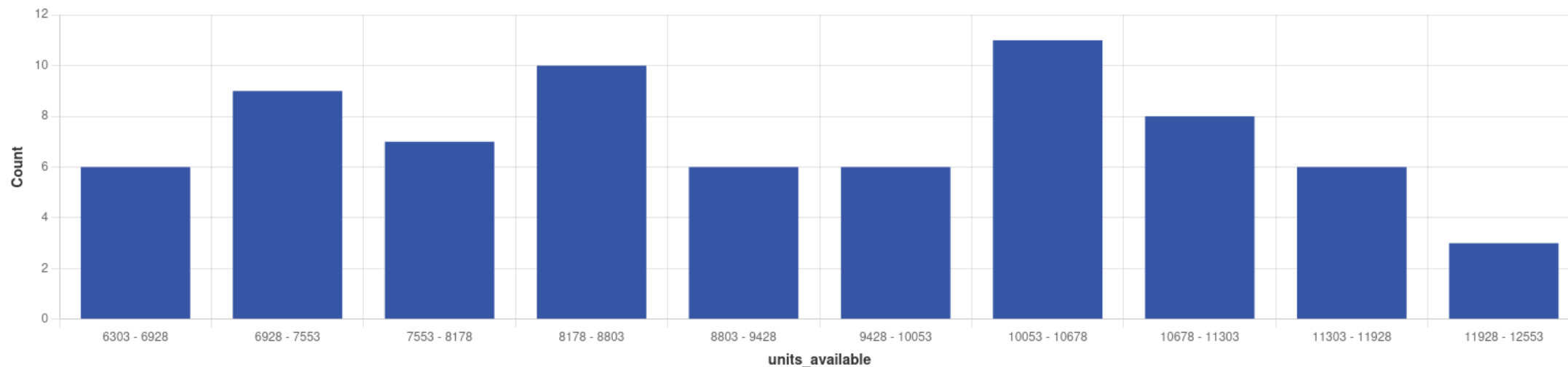
Median

9189

Standard deviation

1696

Count distribution



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



Component to explore

transaction_price_d (Number)

Maximum

4018.53

Minimum

1416.97

Average

2584.02

Median

2280.65

Standard deviation

802.84

Count distribution

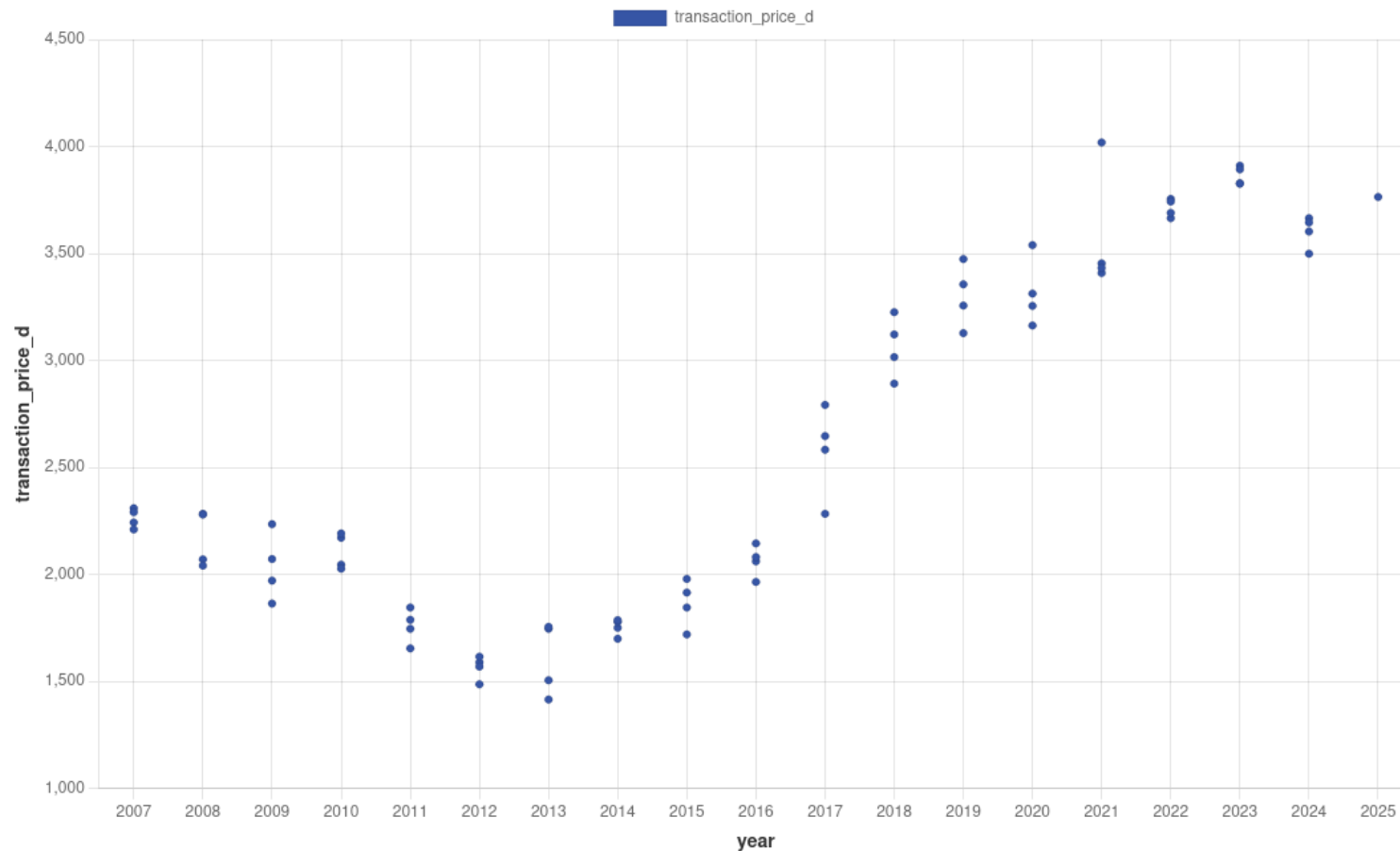


Data Science



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

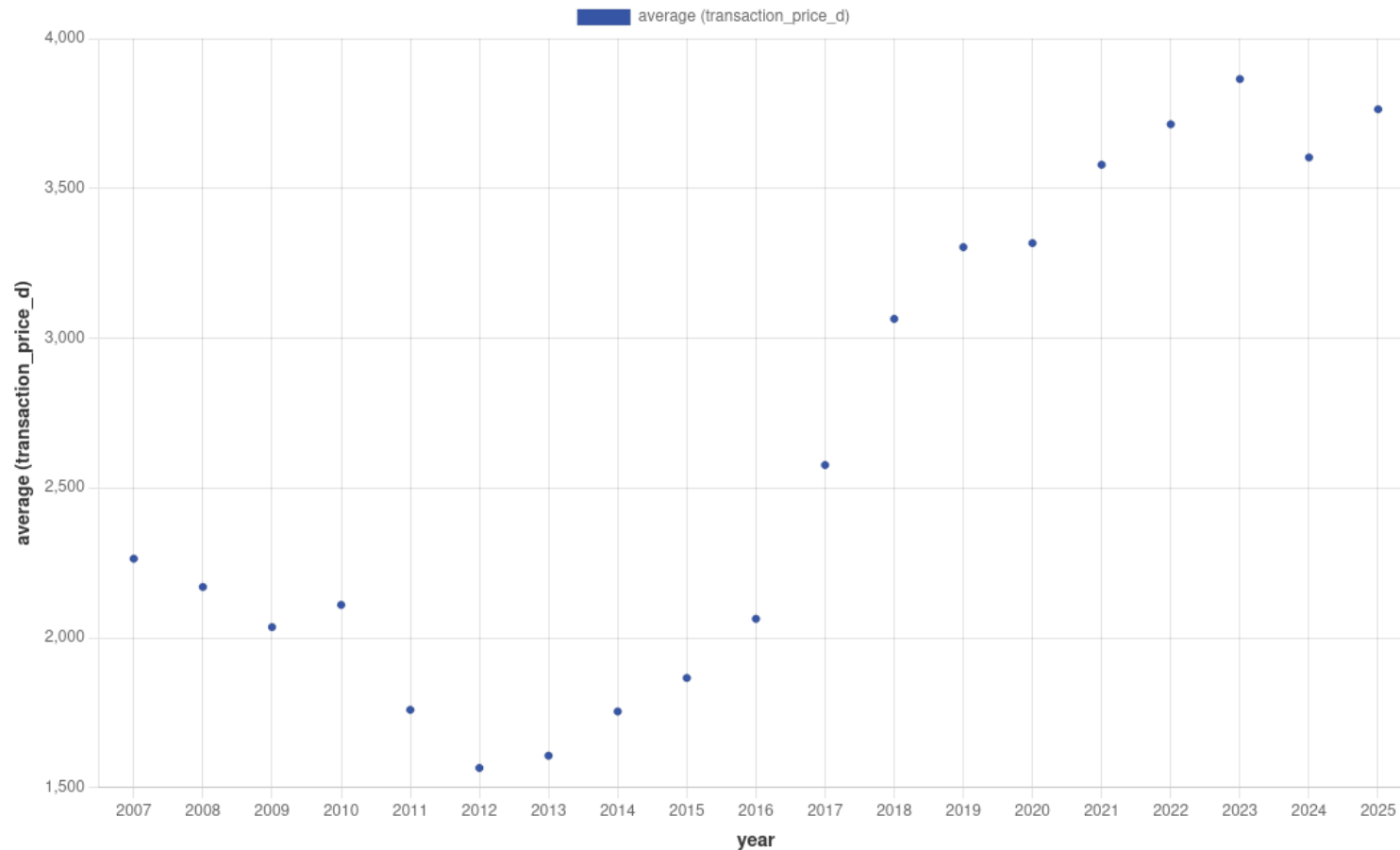
Data range

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

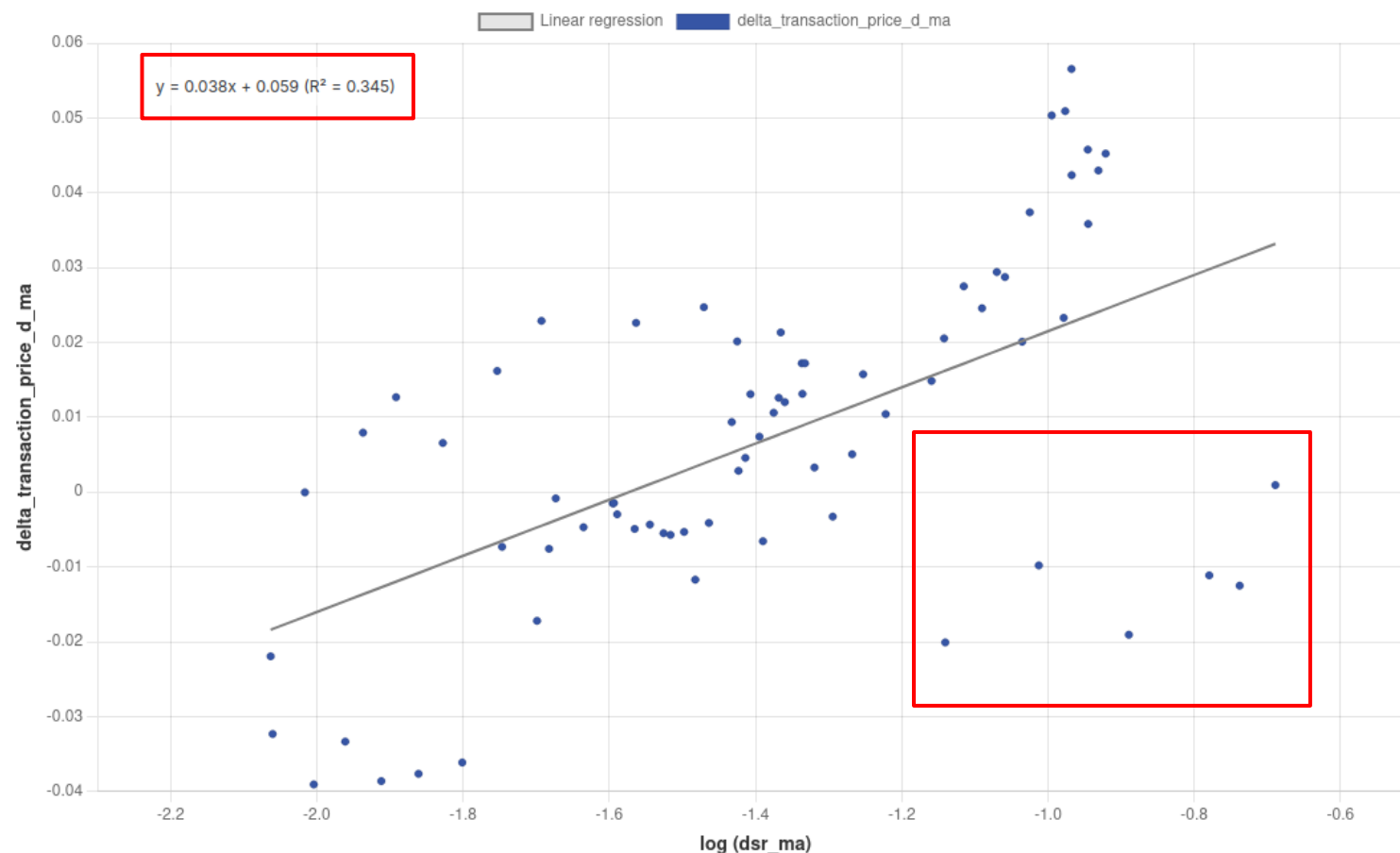
Data range



19 / 19 data points selected ?

GENERATE

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

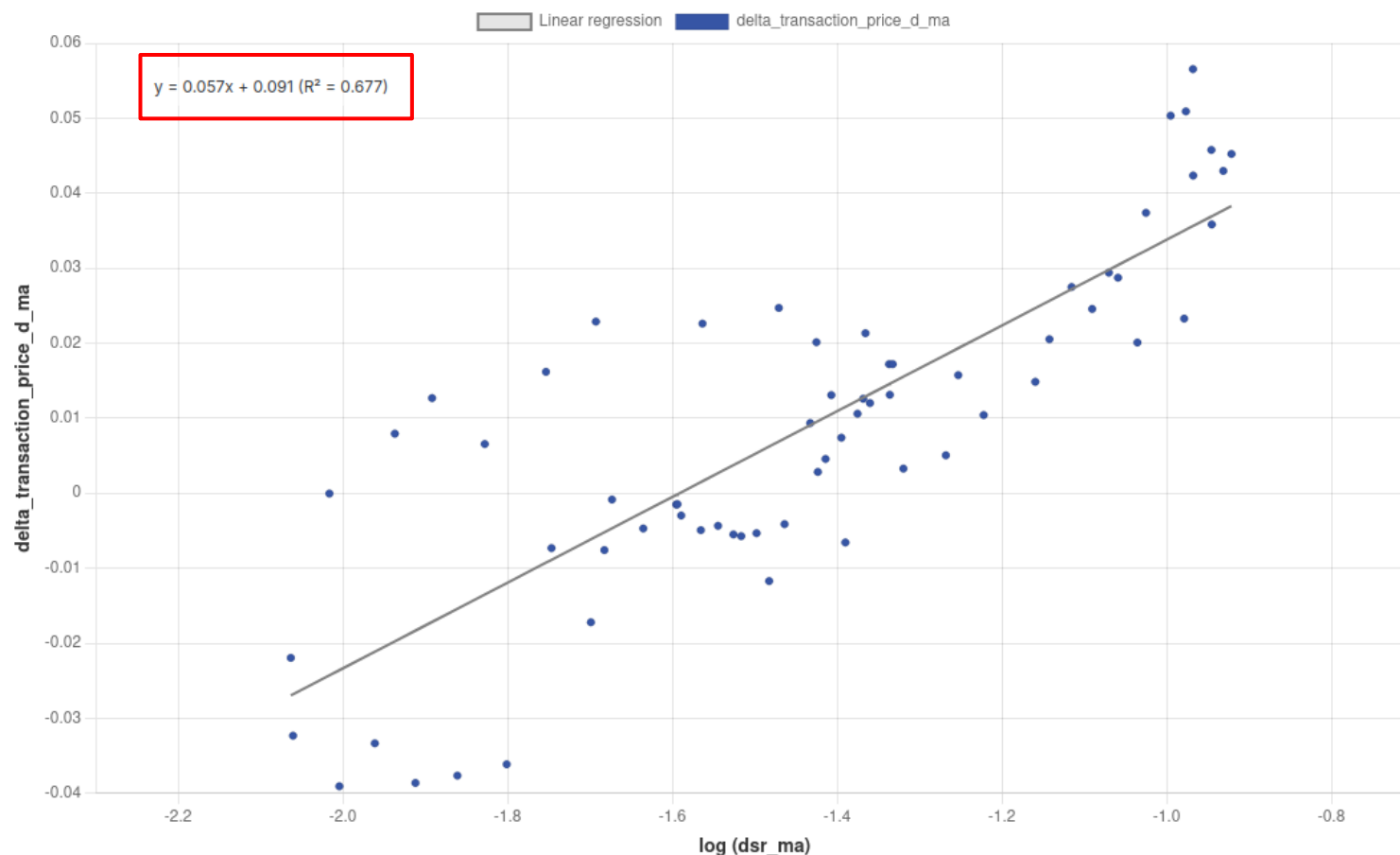
Functions

☐ 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

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

Inputs

Execution

History

Graph

Q Please Input

SCALARS (2)

base_yearInteger2007

sc_selected_countyStringLisboa

DATASETS (5)

Income_PT

Inflation_divisors_Q

Inflation_PT

Real GDP series!

(from Eurostat)

Data Science

Import data structure ✕

Upload data structure file ?



Drop a CSV, JSON or SDMX file or click to upload

OR

Obtain data structure from SDMX URL ?

Paste a SDMX URL

IMPORT

Download data structure ✕

File format

SDMX-ML 2.1 (.xml) ▼

Data scope

Full dataset ▼

☐ Compress file (.gzip)

Download 

Copy API link 

Cancel

Data Science

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

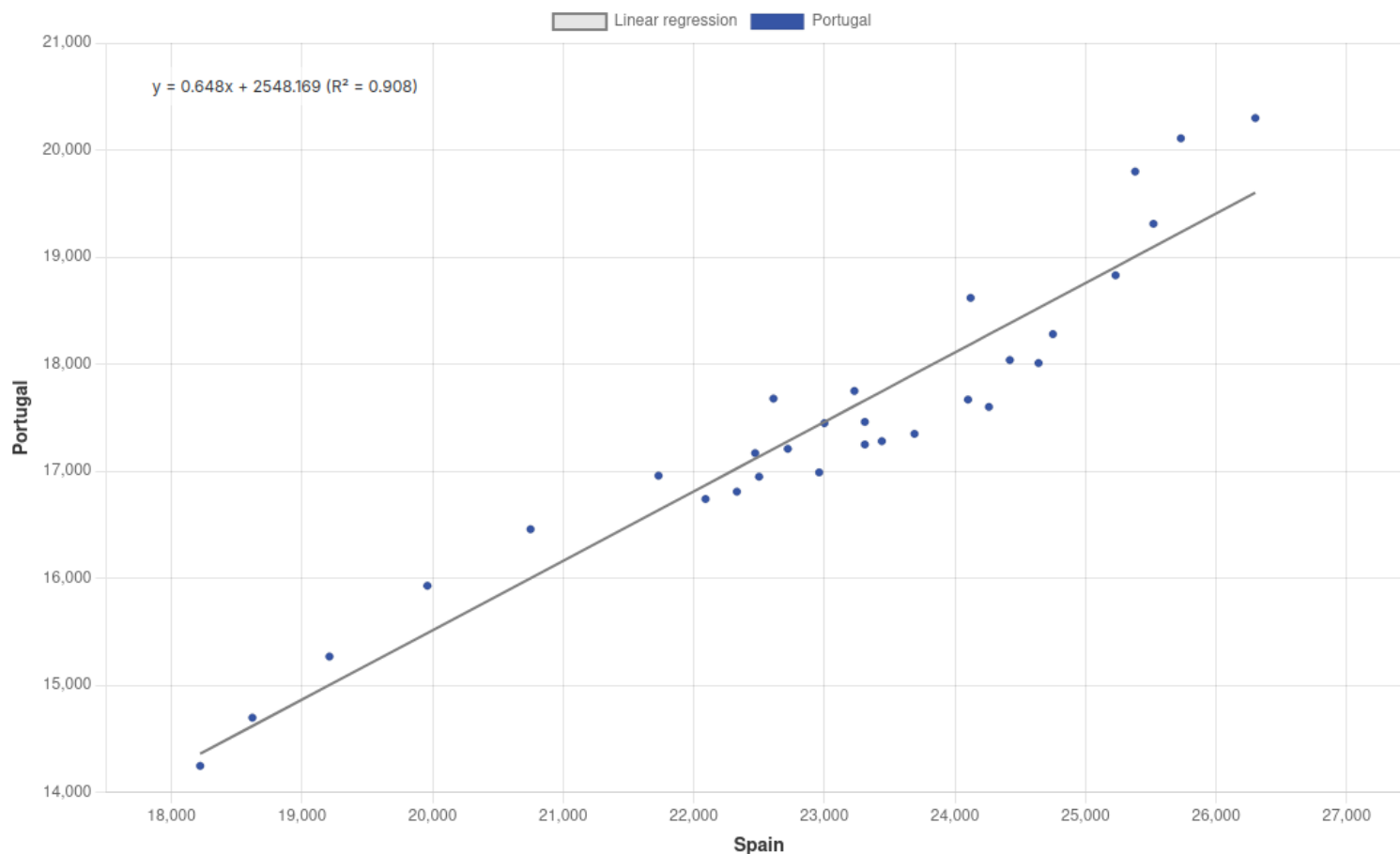


Q freq	↓	Q geo	↓	Q na_item	↓	Q TIME_PERIOD	↓	Q unit	↓	Q OBS_VALUE	↓	Q CONF_STATUS	↓	Q OBS_FLAG	↓
A		AT		B1GQ		1995		CLV15_EUR_HAB		30290		-		-	
A		AT		B1GQ		1996		CLV15_EUR_HAB		30920		-		-	
A		AT		B1GQ		1997		CLV15_EUR_HAB		31540		-		-	
A		AT		B1GQ		1998		CLV15_EUR_HAB		32610		-		-	
A		AT		B1GQ		1999		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		-		-	

Data Science

```
1 // we are using the real GDP metric from here
2 // https://ec.europa.eu/eurostat/databrowser/view/tipsna40/default/table?lang=en
3
4 // data structure:
5 // https://ec.europa.eu/eurostat/api/dissemination/sdmx/2.1/dataflow/ESTAT/tipsna40/
6
7 // data points
8 // https://ec.europa.eu/eurostat/api/dissemination/sdmx/3.0/data/dataflow/ESTAT/tips
9
10 ∨ estat_real_gdp_portugal <- TIPSNA40[sub geo = "PT"]
11 | | | | | | | | | | [keep OBS_VALUE]
12 | | | | | | | | | | [rename OBS_VALUE to Portugal];
13
14 ∨ estat_real_gdp_italy <- TIPSNA40[sub geo = "IT"]
15 | | | | | | | | | | [keep OBS_VALUE]
16 | | | | | | | | | | [rename OBS_VALUE to Italy];
17
18 ∨ estat_real_gdp_spain <- TIPSNA40[sub geo = "ES"]
19 | | | | | | | | | | [keep OBS_VALUE]
20 | | | | | | | | | | [rename OBS_VALUE to Spain];
21
22 ∨ estat_real_gdp_combined <- inner_join(
23 | | | | | | | | | | estat_real_gdp_portugal,
24 | | | | | | | | | | estat_real_gdp_italy,
25 | | | | | | | | | | estat_real_gdp_spain);
```

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

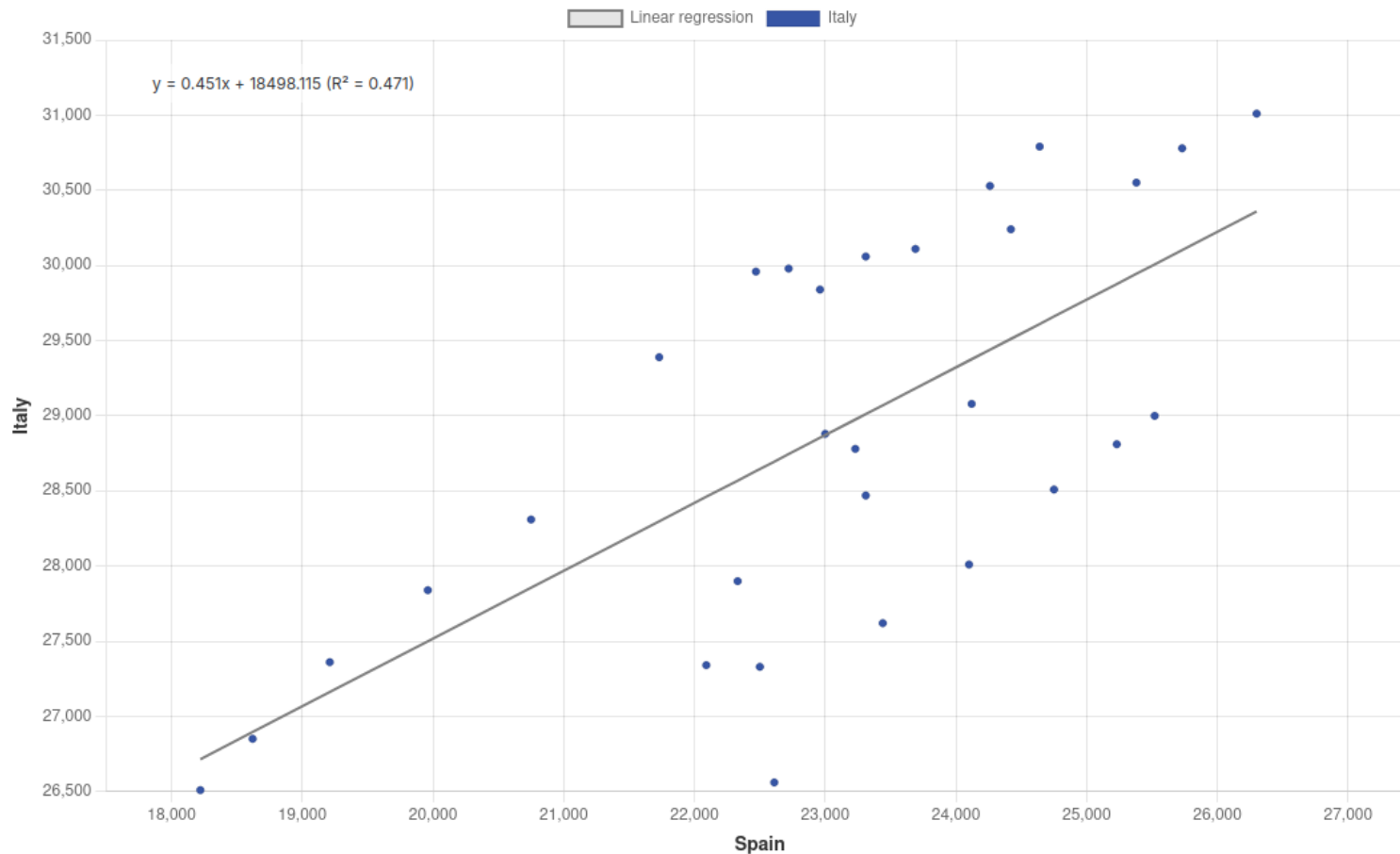
Functions

☐ 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

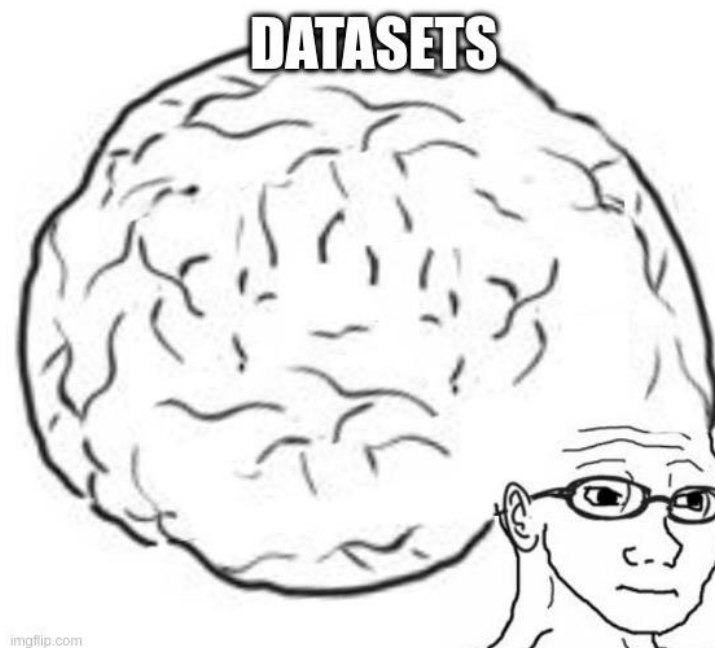
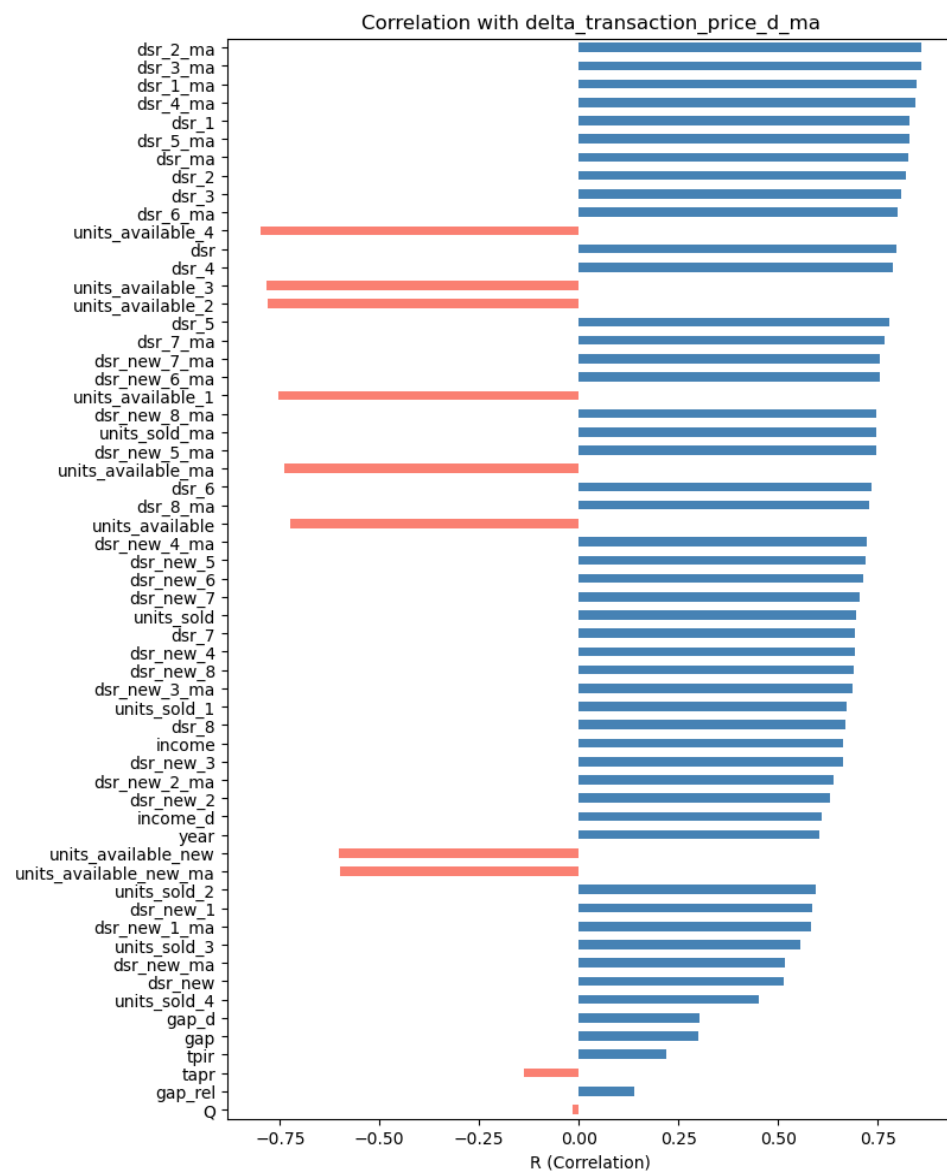
☒ None ☐ Logarithm ☐ Exponential

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