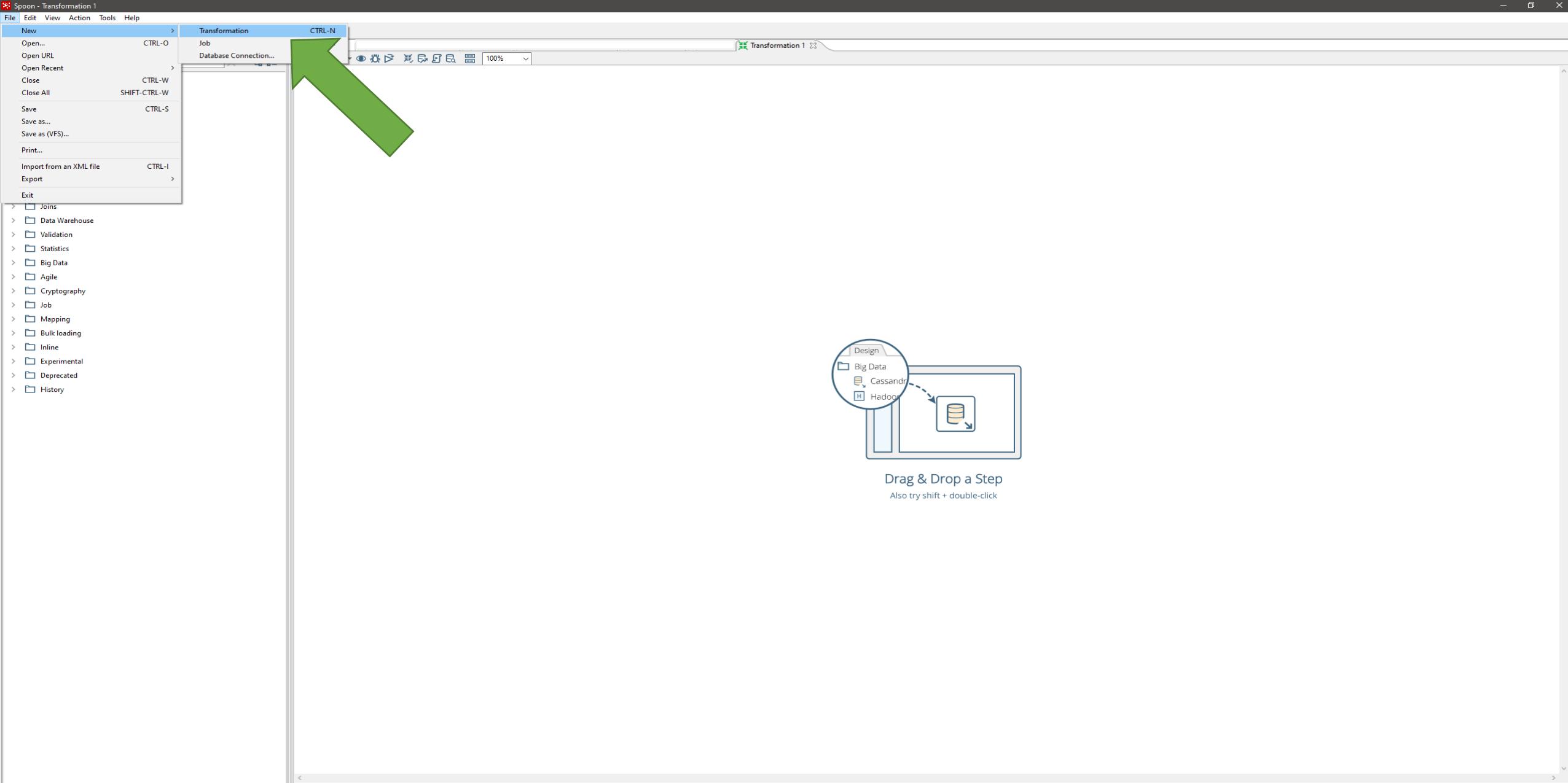
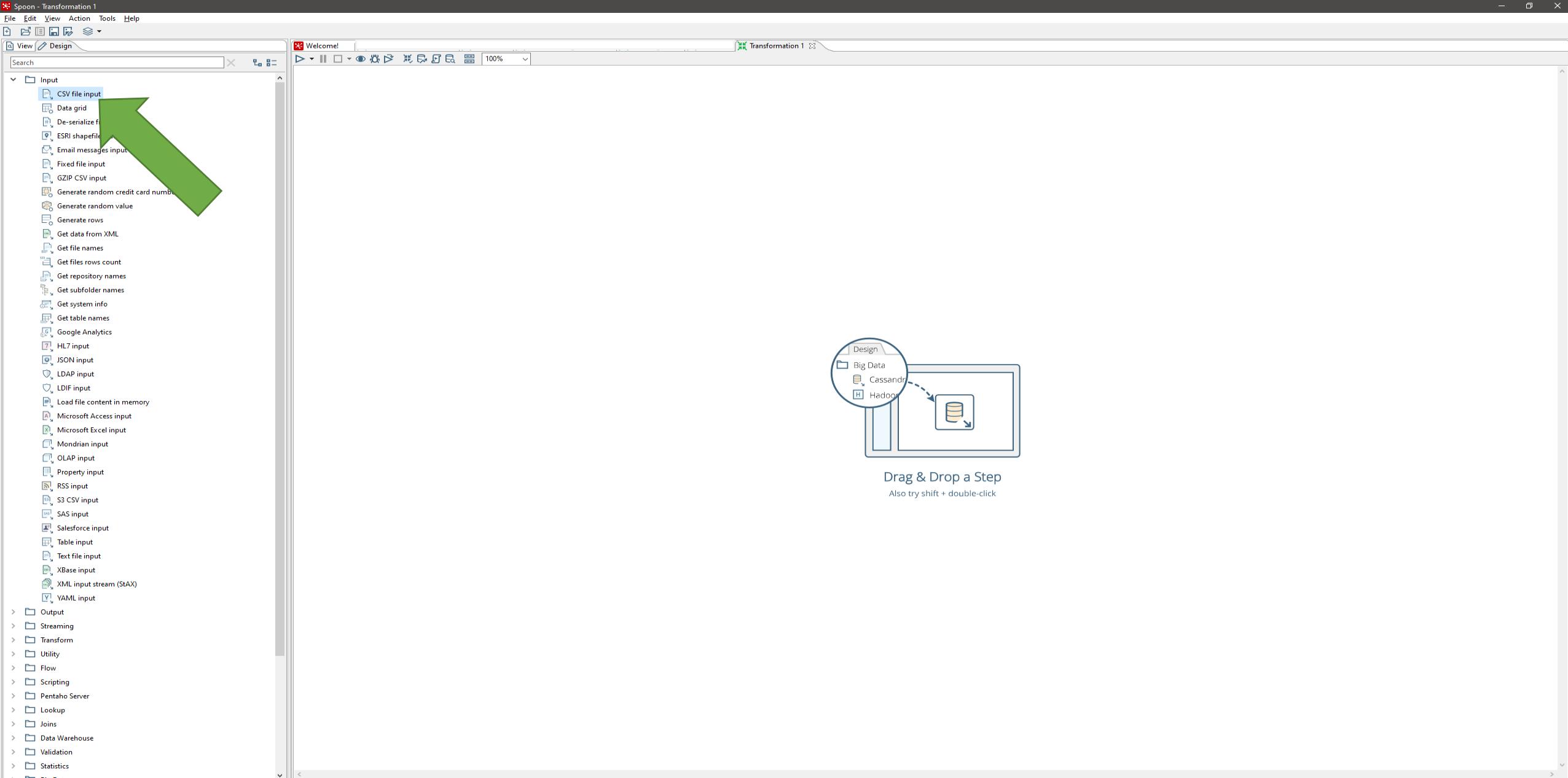


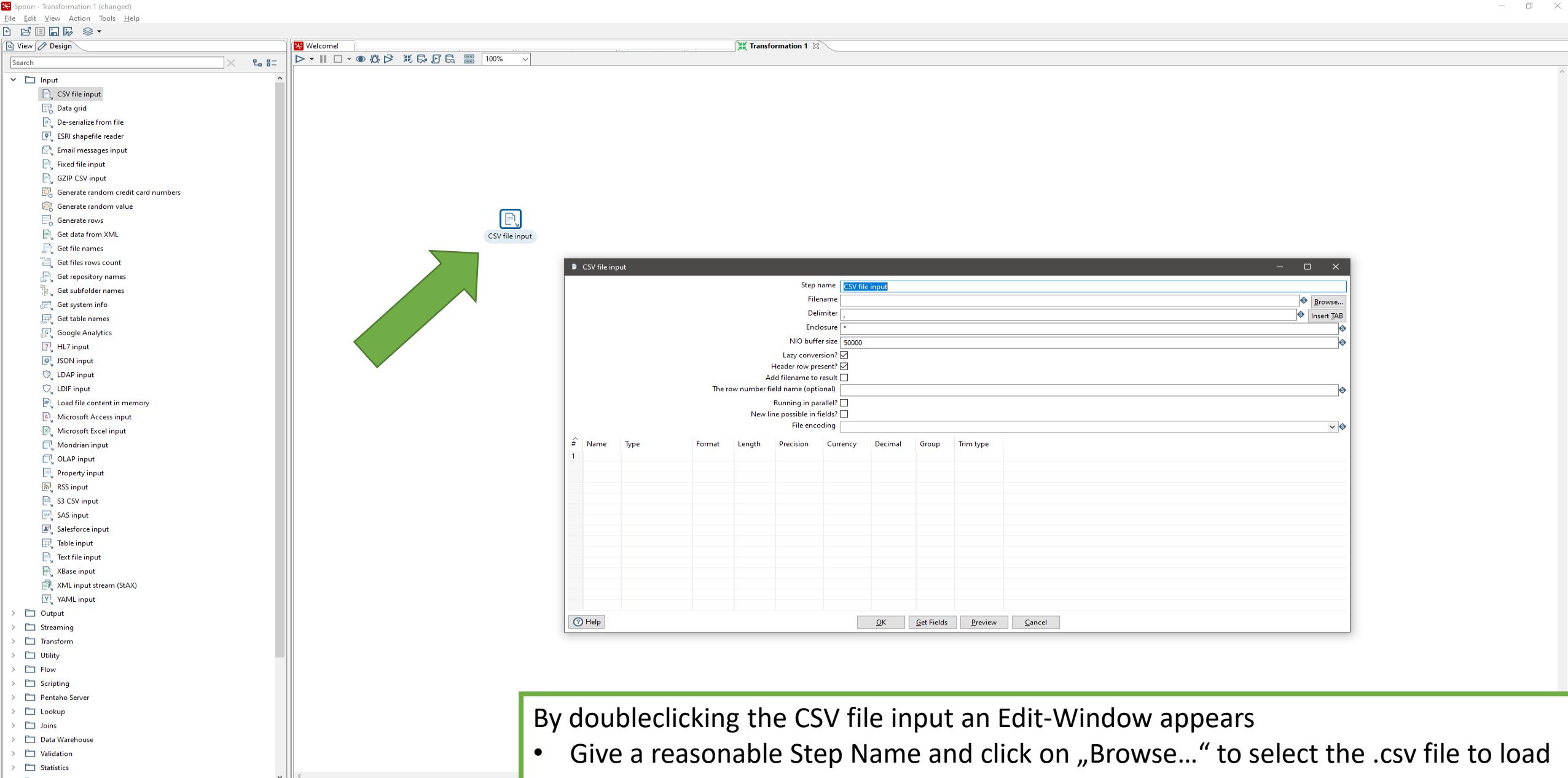
Tutorial ETL - Processing



In order to start a new transformation, click on File > New > Transformation



- To import the operational data, click on Input and drag the CSV file input into the main window



By doubleclicking the CSV file input an Edit-Window appears

- Give a reasonable Step Name and click on „Browse...“ to select the .csv file to load
- Click on „Get Fields“ to load the data
- Press OK

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

Input

- CSV file input
- Data grid
- De-serialize from file
- ESRI shapefile reader
- Email messages input
- Fixed file input
- GZIP CSV input
- Generate random credit card numbers
- Generate random value
- Generate rows
- Get data from XML
- Get file names
- Get files rows count
- Get repository names
- Get subfolder names
- Get system info
- Get table names
- Google Analytics
- HL7 input
- JSON input
- LDAP input
- LDIF input
- Load file content in memory
- Microsoft Access input
- Microsoft Excel input
- Mondrian input
- OLAP input
- Property input
- RSS input
- S3 CSV input
- SAS input
- Salesforce input
- Table input
- Text file input
- XBase input
- XML input stream (StAX)
- YAML input

Output

Streaming

Transform

Utility

Flow

Scripting

Pentaho Server

Lookup

Joins

Data Warehouse

Validation

Statistics

Dim Data

Welcome! Transformation 1

CSV file input

Step name: Exports_2014_2015_Input

Filename: C:\Users\...\PC_Export_2014_2015.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional)

Running in parallel?

New line possible in fields?

File encoding: UTF-8

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	pc_code	Integer	#	15	0	€	,	.	none
2	pc_description	String		17		€	,	.	none
3	unit	String		2		€	,	.	none
4	country_code	Integer	#	15	0	€	,	.	none
5	country_name	String		15		€	,	.	none
6	quantity	String		2		€	,	.	none
7	value	Number	#,##	15	0	€	,	.	none

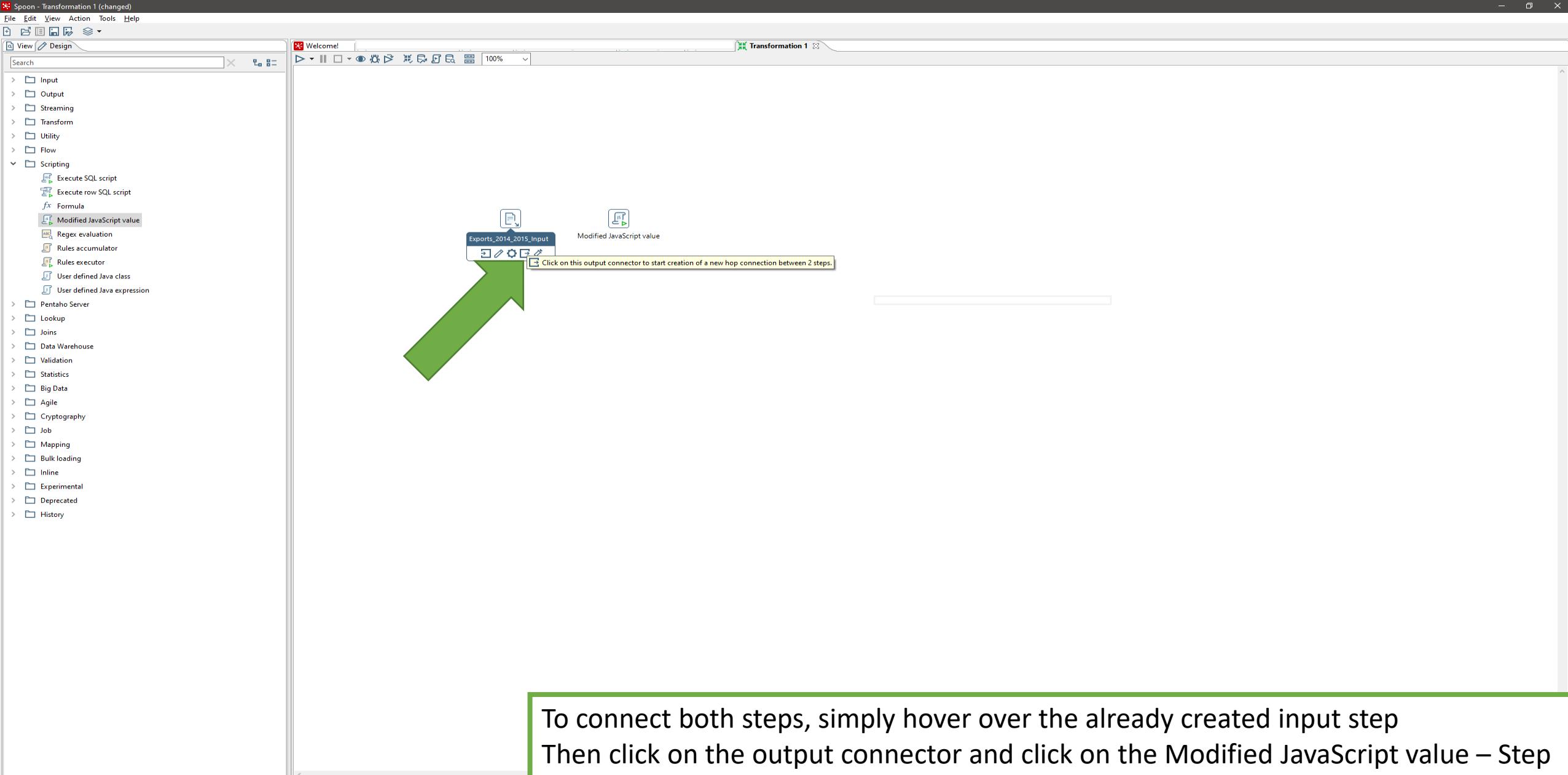
Help

Loading the Fields should get you a screen like this
 Edit the datatype of the corresponding column to match the following:

pc_code	Product ID	A1	String
pc_description	Product Description	Tea	String
unit	Unit of quantity	Kgs	String
country_code	Country ID	1213	Integer
country_name	Country Description	Kenya	String
quantity	Count of Goods	2277547	Integer
value	Value in Mio. USD	4,276366	Number



The first step is to add a random date to each transaction
To achieve this, we integrate a step named „Modified JavaScript value“
This step can be found under Scripting > Modified JavaScript value
Drag it into the main screen



Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

> Input
> Output
> Streaming
> Transform
> Utility
> Flow
Scripting
 Execute SQL script
 Execute row SQL script
 Formula
 Modified JavaScript value
 Regex evaluation
 Rules accumulator
 Rules executor
 User defined Java class
 User defined Java expression
> Pentaho Server
> Lookup
> Joins
> Data Warehouse
> Validation
> Statistics
> Big Data
> Agile
> Cryptography
> Job
> Mapping
> Bulk loading
> Inline
> Experimental
> Deprecated
> History

Welcome!

100%

Exports_2014_2015_Input → Modified JavaScript value

Java script functions :

- > Transform Scripts
- > Transform Constants
- > Transform Functions
- > Input fields
 - pc_code
 - pc_description
 - unit
 - country_code
 - country_name
 - quantity
 - value
- > Output fields
 - Please use the 'Replace value 'Fieldname'' or 'Rename to' field

Step name : Modified JavaScript value

Java script :

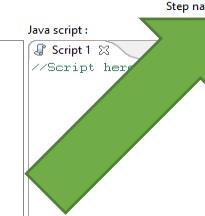
```
//Script here
```

Linien: 0 Compatibility mode? Optimization level: 9

Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1						

Help OK Cancel Get variables Test script



Again give the step a reasonable name
We will now proceed with scripting the random date assignment

Spoon - Tuttrans (changed)
File Edit View Action Tools Help

Modified JavaScript value

Step name Random_Date_Assignment_Export_2014_2015

Java script functions :

Input fields

- pc_code
- pc_description
- unit
- country_code
- country_name
- quantity
- value

Output fields

Please use the 'Replace value 'Fieldname' or 'Rename To' field.

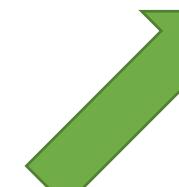
```
// Script for random assignation of month and year  
// d1 creates present date, sets its Day to the 1st, then randomizes its month  
// if month is less than April, assign 2015; if month is more April and higher assign 2014  
// reason is CSV-entries are aggregated from range April to March  
// d3 finally creates a string representing the adequate year + randomized month: "YYYYMM" <- Month_ID for DIM_Zeit  
  
var d1 = new Date();  
d1.setDate(1);  
d1.setMonth((Math.random()*12));  
var d2 = date2str(d1,"MM");  
if (str2num(d2) >= 4) {  
    d1.setFullYear(2014);  
} else {  
    d1.setFullYear(2015);  
}  
var d3 = date2str(d1,"yyyy") + d2;
```

Position: 17, 32
Compatibility mode? Optimization level 9

Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	d1		Date			N
2	d2		String			N
3	d3		String			N

OK Cancel Get variables Test script



The procedure of the script is described in the comment section above the code
After typing the script a click on „Get variables“ should give you an output like this
Change the datatype of d3 to Integer and click on „Test script“ and press OK

Spoon - Tuttrans (changed)
 File Edit View Action Tools Help

View Design

Search

Input
 Output
 Streaming
Transform
 Add XML
 Add a checksum
 Add constants
 Add sequence
 Add value fields changing sequence
 Calculator
 Closure generator
 Concat fields
 Get ID from slave server
 Number range
 Replace in string
 Row denormaliser
 Row flattener
 Row normaliser
 Select values
 Set field value
 Set field value to a constant
 Sort rows
 Split field to rows
 Split fields
 String operations
 Strings cut
 Unique rows
 Unique rows (HashSet)
 Value mapper
 XSL transformation
 Utility
 Flow
 Scripting
 Pentaho Server
 Lookup
 Joins
 Data Warehouse
 Validation
 Statistics
 Big Data
 Agile
 Cryptography
 Job
 Mapping
 Bulk loading
 Inline
 Experimental
 Deprecated
 History

Modified JavaScript value

Step name Random_Date_Assignment_Export_2014_2015

Java script functions:

- > Transform Scripts
- > Transform Constants
- > Transform Functions
- > Input fields
 - pc_code
 - pc_description
 - unit
 - country_code
 - country_name
 - quantity
 - value
- > Output fields

Please use the 'Replace value 'Fieldname' or 'Rename To' field.

Java script:

```

// Script for random assignation of month and year
// d1 creates present date, sets its Day to the 1st, then randomizes its month
// if month is less than April, assign 2015; if month is more April and higher assign 2014
// reason is CSV-entries are aggregated from range April to March
// d3 finally creates a string representing the adequate year + randomized month: "YYYYMM" <- Month_ID for DIM_Zeit

var d1 = new Date();
d1.setDate(1);
d1.setMonth((Math.random()*12));
var d2 = date2str(d1,"MM");
if (str2num(d2) >= 4) {
    d1.setFullYear(2014);
} else {
    d1.setFullYear(2015);
}
var d3 = date2str(d1,"yyyy") + d2;

```

Examine preview data

Rows of step: Random_Date_Assignment_Export_2014_2015 (10 rows)

#	pc_code	pc_description	unit	country_code	country_name	quantity	value	d1	d2	d3
1	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2014/08/01 22:31:06.132	08	201408
2	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2014/09/01 22:31:06.133	09	201409
3	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2015/03/01 22:31:06.133	03	201503
4	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2014/05/01 22:31:06.133	05	201405
5	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2015/01/01 22:31:06.133	01	201501
6	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2014/04/01 22:31:06.133	05	201405
7	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2014/01/01 22:31:06.133	04	201404
8	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2015/05/01 22:31:06.133	01	201501
9	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2014/11/01 22:31:06.133	11	201411
10	test value test value	test value test value	test value test value	0	test value test value	test value test value	0	2015/01/01 22:31:06.133	01	201501

Close **Show Log**



The Script test should prompt something like this
 Variable d3 is everything we are interested in for now
 Press on the Close button

Spoon - Tuttrans (changed)
File Edit View Action Tools Help

Modified JavaScript value

Step name Random_Date_Assignment_Export_2014_2015

Java script functions :

Input fields

- pc_code
- pc_description
- unit
- country_code
- country_name
- quantity
- value

Output fields

Please use the 'Replace value 'Fieldname' or 'Rename To' field.

```
// Script for random assignation of month and year  
// d1 creates present date, sets its Day to the 1st, then randomizes its month  
// if month is less than April, assign 2015; if month is more April and higher assign 2014  
// reason is CSV-entries are aggregated from range April to March  
// d3 finally creates a string representing the adequate year + randomized month: "YYYYMM" <- Month_ID for DIM_Zeit  
  
var d1 = new Date();  
d1.setDate(1);  
d1.setMonth((Math.random()*12));  
var d2 = date2str(d1, "MM");  
if (str2num(d2) >= 4) {  
    d1.setFullYear(2014);  
} else {  
    d1.setFullYear(2015);  
}  
var d3 = date2str(d1, "yyyy") + d2;
```

Position: 17, 32
Compatibility mode? Optimization level 9

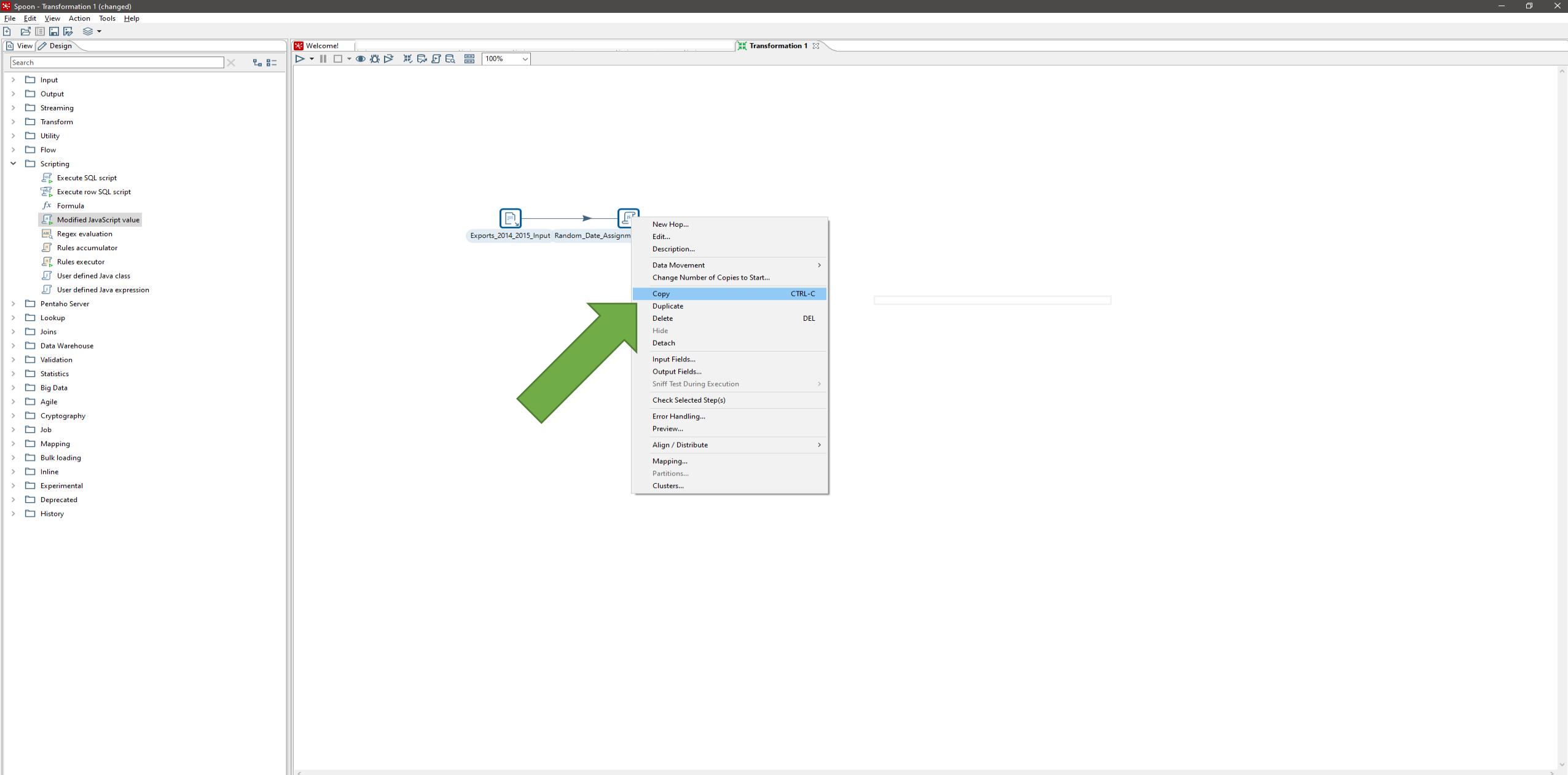
Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	d1		Date			N
2	d2		String			N
3	d3		String			N

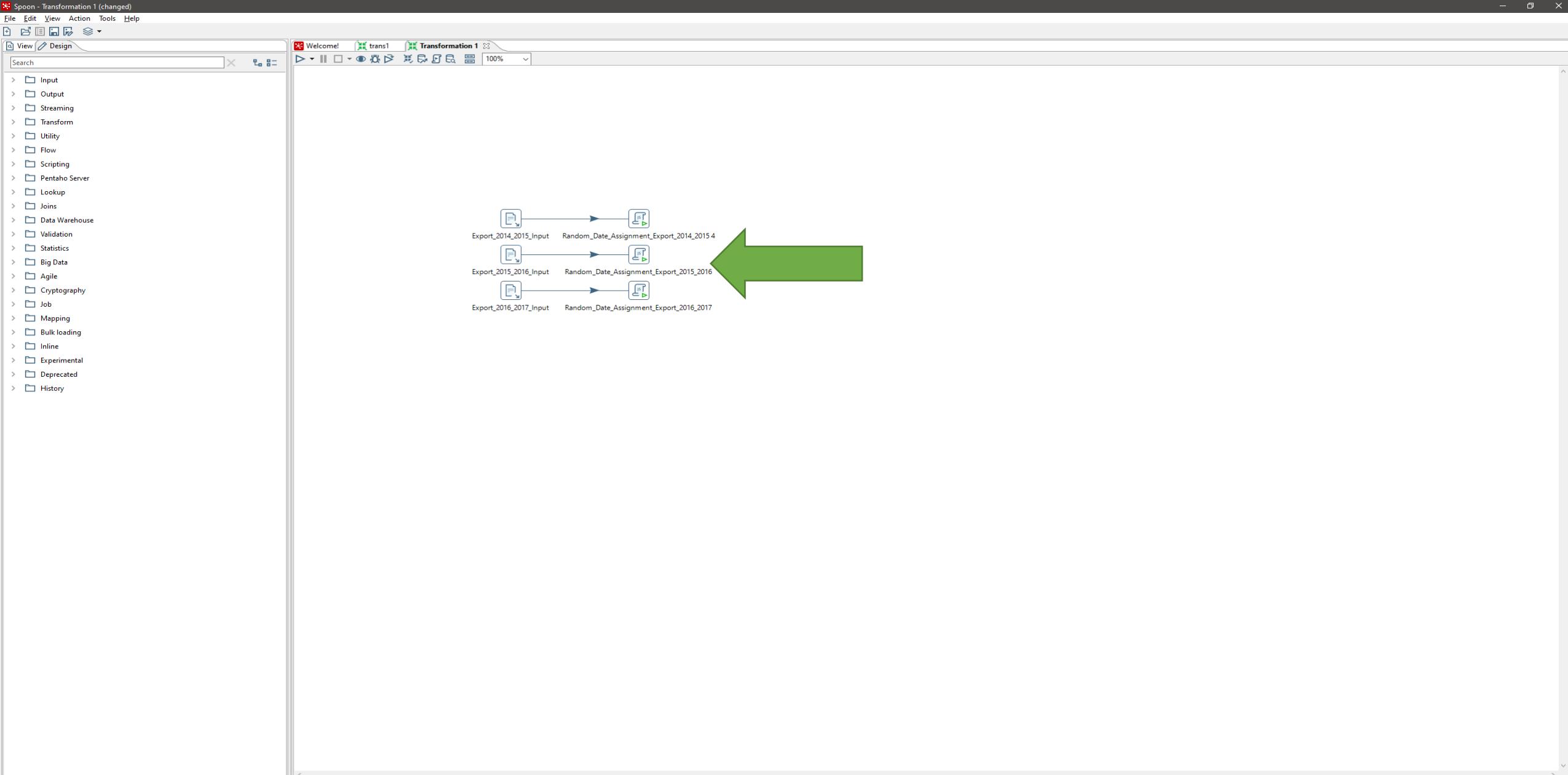
OK Cancel Get variables Test script



Delete both variables d1 and d2 since we are not interested in it by now
Press OK



Now copy both steps and paste them to create inputs for the other two export files



Rearrange the steps and rename them according to the picture above
Then double-click on „Random_Date_Assignment_Export_2015_2016“

Spoon - Tuttrans (changed)
 File Edit View Action Tools Help

View Design
 Search

> Input
 Output
 Streaming
 > Transform
 Add XML
 Add a checksum
 Add constants
 Add sequence
 Add value fields changing sequence
 Calculator
 Closure generator
 Concat fields
 Get ID from slave server
 Number range
 Replace in string
 Row denormaliser
 Row flattener
 Row normaliser
 Select values
 Set field value
 Set field value to a constant
 Sort rows
 Split field to rows
 Split fields
 String operations
 Strings cut
 Unique rows
 Unique rows (HashSet)
 Value mapper
 XSL transformation

> Utility
 Flow
 Scripting
 Pentaho Server
 Lookup
 Joins
 Data Warehouse
 Validation
 Statistics
 Big Data
 Agile
 Cryptography
 Job
 Mapping
 Bulk loading
 Inline
 Experimental
 Deprecated
 History

Welcome! trans1 Tuttrans
 100%
 Export_2014_2015_Input Random_Date_Assignment_Export_2014
 Export_2015_2016_Input Random_Date_Assignment_Export_2015
 Export_2016_2017_Input Random_Date_Assignment_Export_2016

Modified JavaScript value
 Step name: Random_Date_Assignment_Export_2015_2016
 Java script functions:
 > Transform Scripts
 > Transform Constants
 > Transform Functions
 > Input fields
 pc_code
 pc_description
 unit
 country_code
 country_name
 quantity
 value
 > Output fields
 Please use the 'Replace value 'Fieldname''

```

    Script 1
    // Script for random assignation of month and year
    // d1 creates present date, sets its Day to the 1st, then randomizes its month
    // if month is less than April, assign 2015; if month is more April and higher assign 2014
    // reason is CSV-entries are aggregated from range April to March
    // d3 finally creates a string representing the adequate year + randomized month: "YYYYMM" <- Month_ID for DIM_Zeit

    var d1 = new Date();
    d1.setDate(1);
    d1.setMonth(Math.random()*12);
    var d2 = date2str(d1, "MM");
    if (str2num(d2) >= 4) {
        d1.setFullYear(2014);
    } else {
        d1.setFullYear(2015);
    }

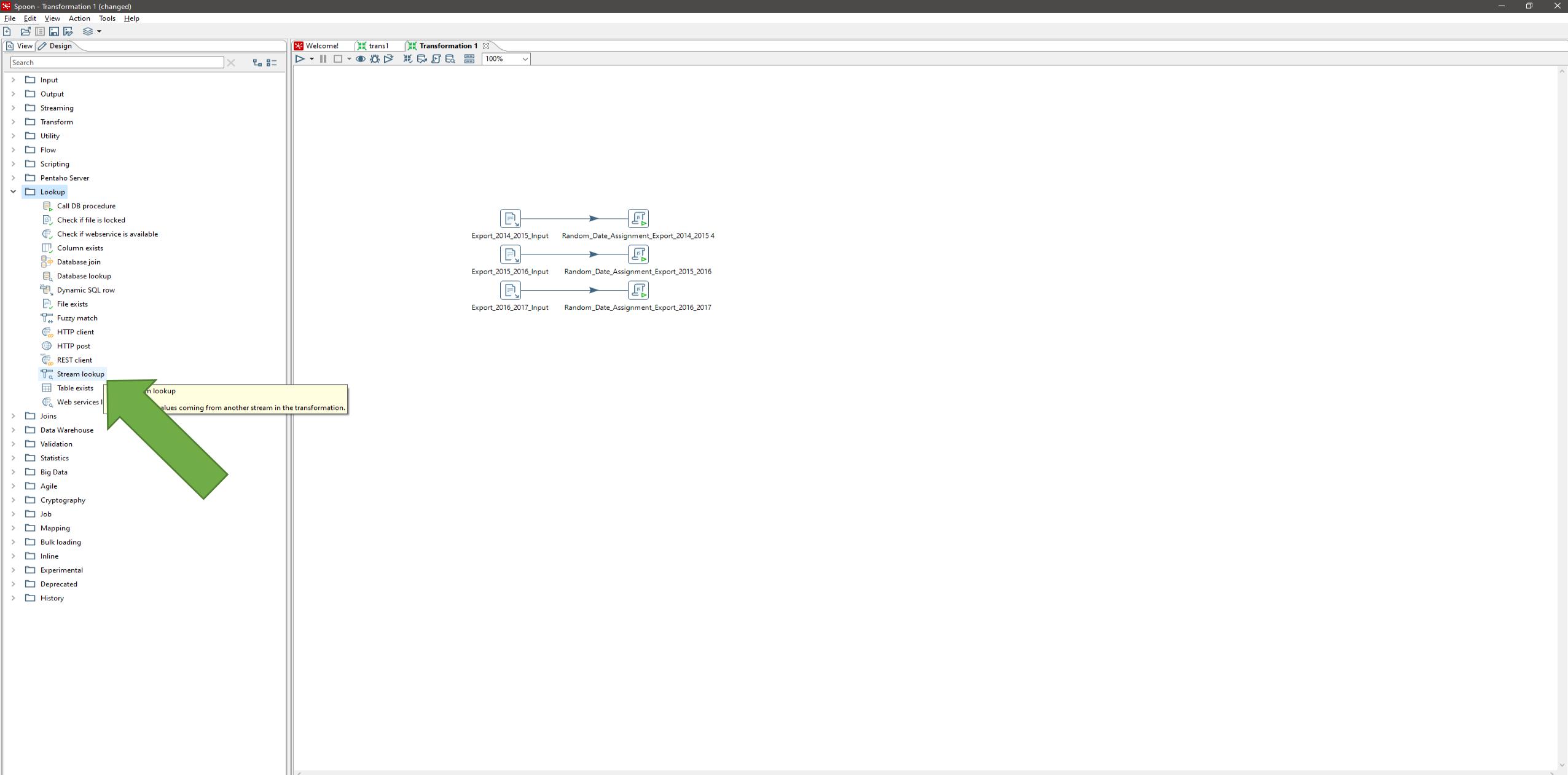
    var d3 = date2str(d1, "yyyy") + d2;
    
```

Execution Results
 Logging Execution History Step Metrics Performance Graph Metrics Preview data
 Stepname Copynr Read Written Input Output
 1 DIM_Product 0 0 168 168
 2 Export_2015_2016_Input 0 0 20771 20772
 3 Export_2014_2015_Input 0 0 20830 20831
 4 Export_2016_2017_Input 0 0 20830 20831
 5 Random_Date_Assignment_Export_2014_2015 0 20830 20830 0
 6 Random_Date_Assignment_Export_2015_2016 0 20771 20771 0
 7 Random_Date_Assignment_Export_2016_2017 0 20830 20830 0
 8 DIM_Region 0 0 233 233
 9 DIM_Time 0 0 1460 1460
 10 Stream_Lookup_DIM_Time_Exports 0 63891 62431 0
 11 Stream_Lookup_DIM_Product_Exports 0 62599 62431 0
 12 Stream_Lookup_DIM_Region_Exports 0 62664 62431 0
 13 Cleansing_Exports 0 62431 62431 0
 14 Output_Exports 0 60370 60369 0 60369

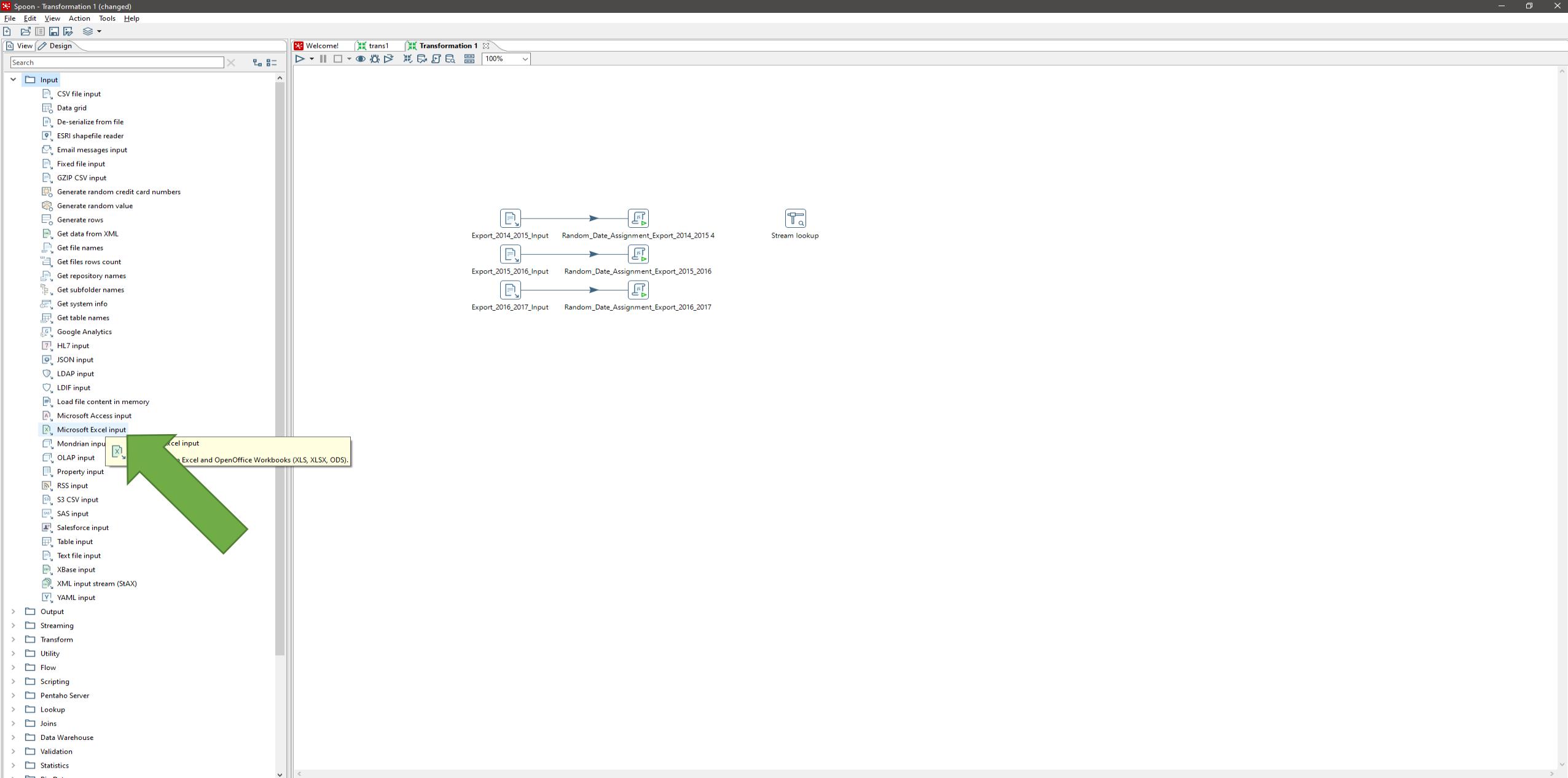
Fields
 # Fieldname Rename to Type Length Precision Replace value 'Fieldname' or 'Rename to'
 1 d3 Integer N

Position: 17, 32
 Compatibility mode? Optimization level: 9
 OK Cancel Get variables Test script

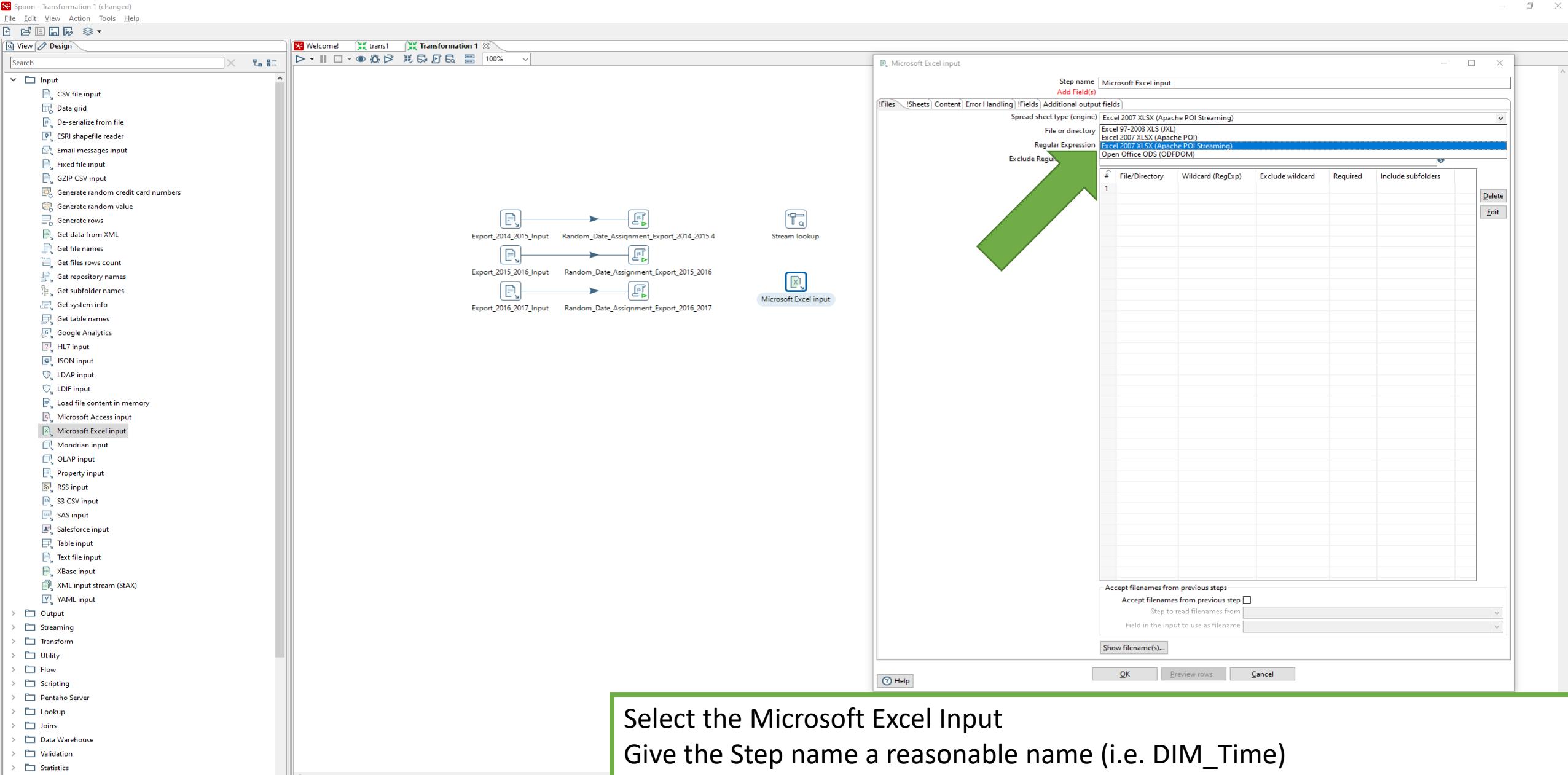
Adjust d1.setFullYear as described above and press OK
 Do the same for „Random_Date_Assignment_Export_2016_2017“



We will now create the first Lookup for our inputs to enrich dimensions
Find Stream Lookup under Lookup and drag it onto the main screen



Since our Time dimension DIM_Time is an Excel file, we additionally need a MS Excel Input Step
Drag the MS Input Step onto the main screen



Select the Microsoft Excel Input
Give the Step name a reasonable name (i.e. DIM_Time)
Select Spread sheet type and choose Excel 2007 XLSX (Apache POI Streaming)
Click on „Browse...“ , search and choose DIM_Time.xlsx
Finally press „Add“

Spoon - Transformation 1 (changed)
 File Edit View Action Tools Help

View Design
 Search

Input

- CSV file input
- Data grid
- De-serialize from file
- ESRI shapefile reader
- Email messages input
- Fixed file input
- GZIP CSV input
- Generate random credit card numbers
- Generate random value
- Generate rows
- Get data from XML
- Get file names
- Get files rows count
- Get repository names
- Get subfolder names
- Get system info
- Get table names
- Google Analytics
- HL7 input
- JSON input
- LDAP input
- LDIF input
- Load file content in memory
- Microsoft Access input
- Microsoft Excel input**
- Mondrian input
- OLAP input
- Property input
- RSS input
- S3 CSV input
- SAS input
- Salesforce input
- Table input
- Text file input
- XBase input
- XML input stream (StAX)
- YAML input

Output

Streaming

Transform

Utility

Flow

Scripting

Pentaho Server

Lookup

Joins

Data Warehouse

Validation

Statistics

DIM Data

Welcome! trans1 Transformation 1

Transformation 1

Stream lookup

Microsoft Excel input

Step name: Microsoft Excel input
Add Field(s)

Files | Sheets | Content | Error Handling | Fields | Additional output fields

Spread sheet type (engine): Excel 2007 XLSX (Apache POI Streaming)

File or directory:

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard
1	C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\DIM_Time.xlsx	

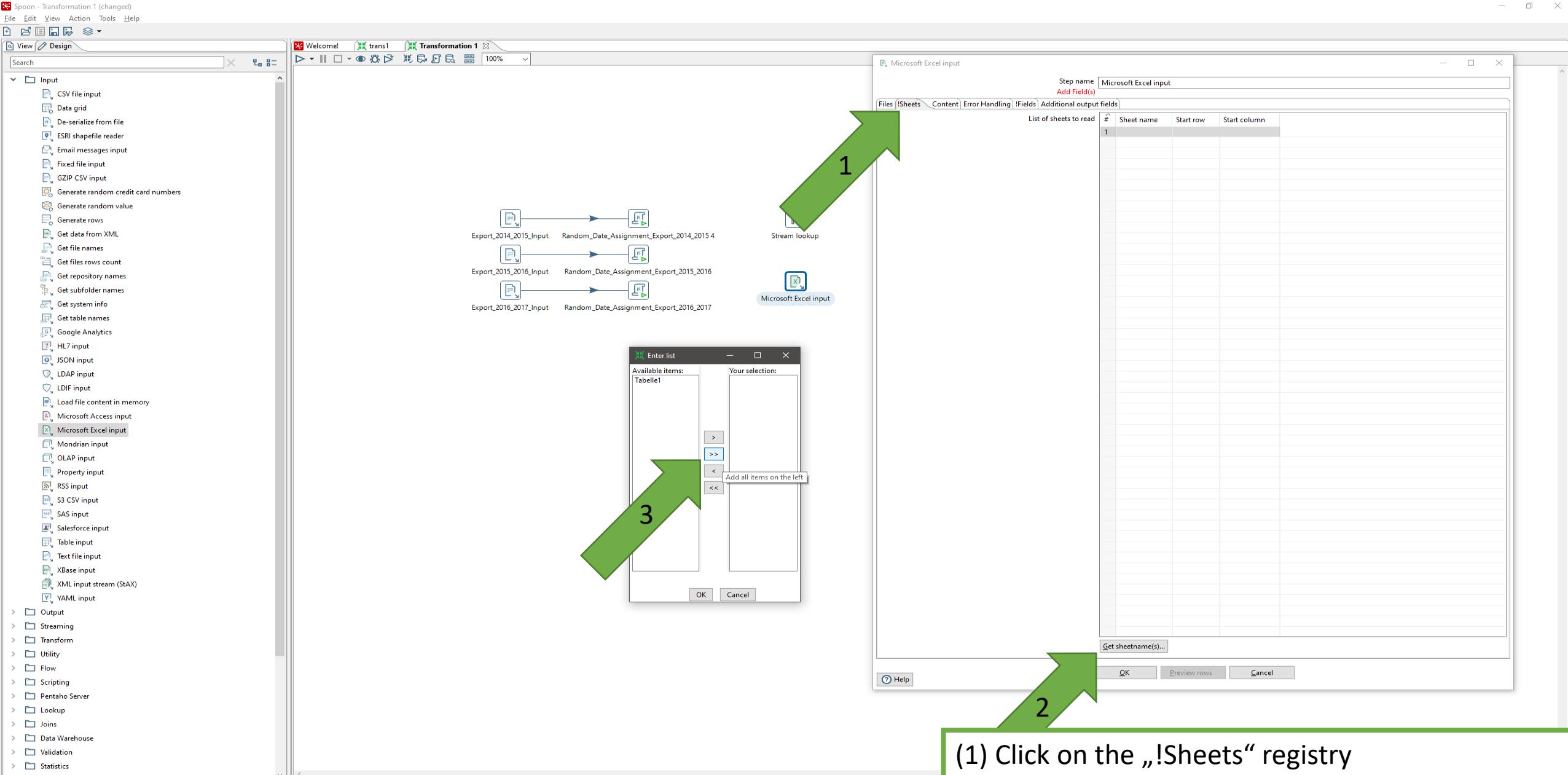
Delete

Edit

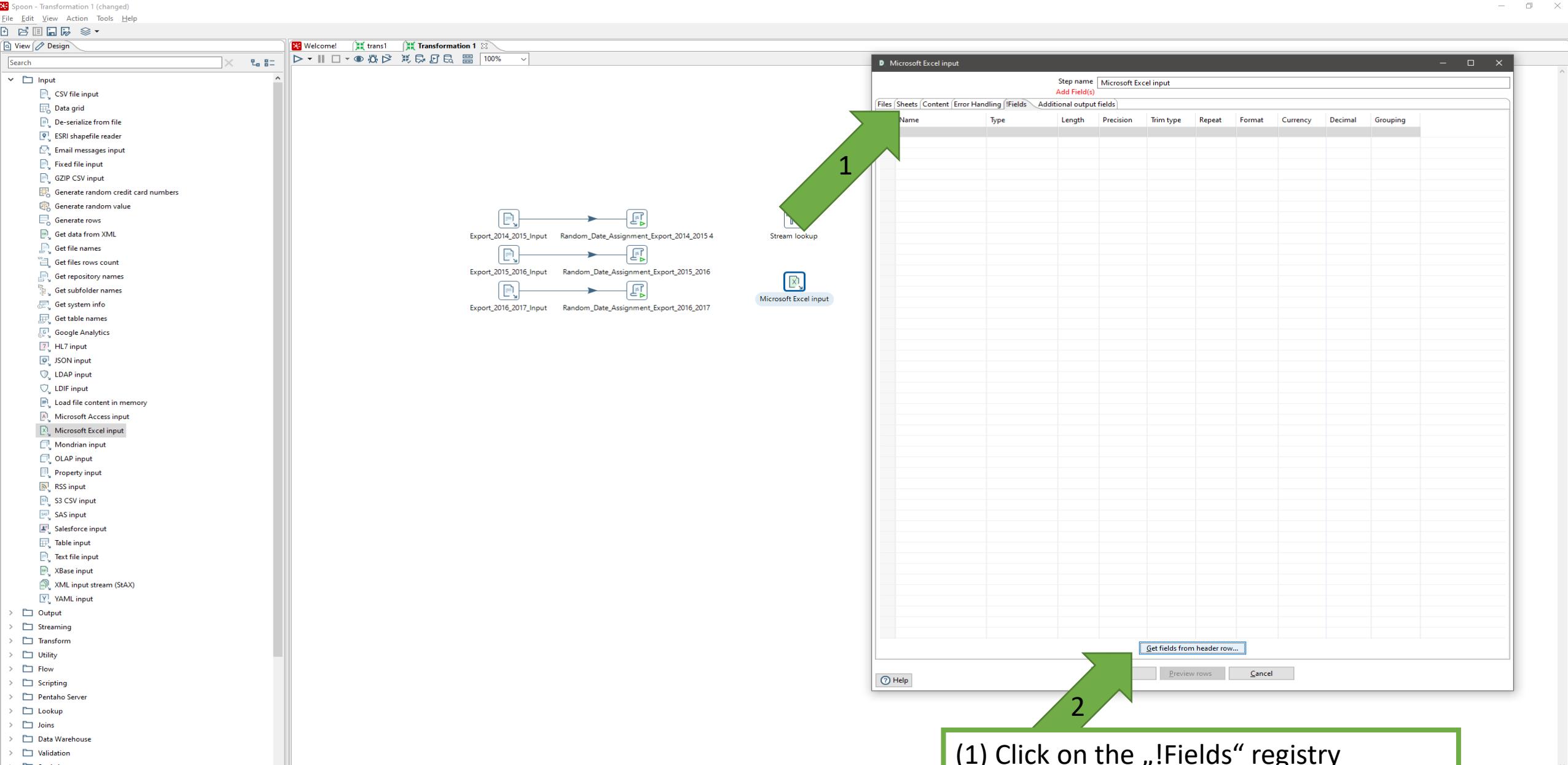
Accept filenames from previous steps
 Step to read filenames from:
 Field in the input to use as filename:
 Show filename(s)...

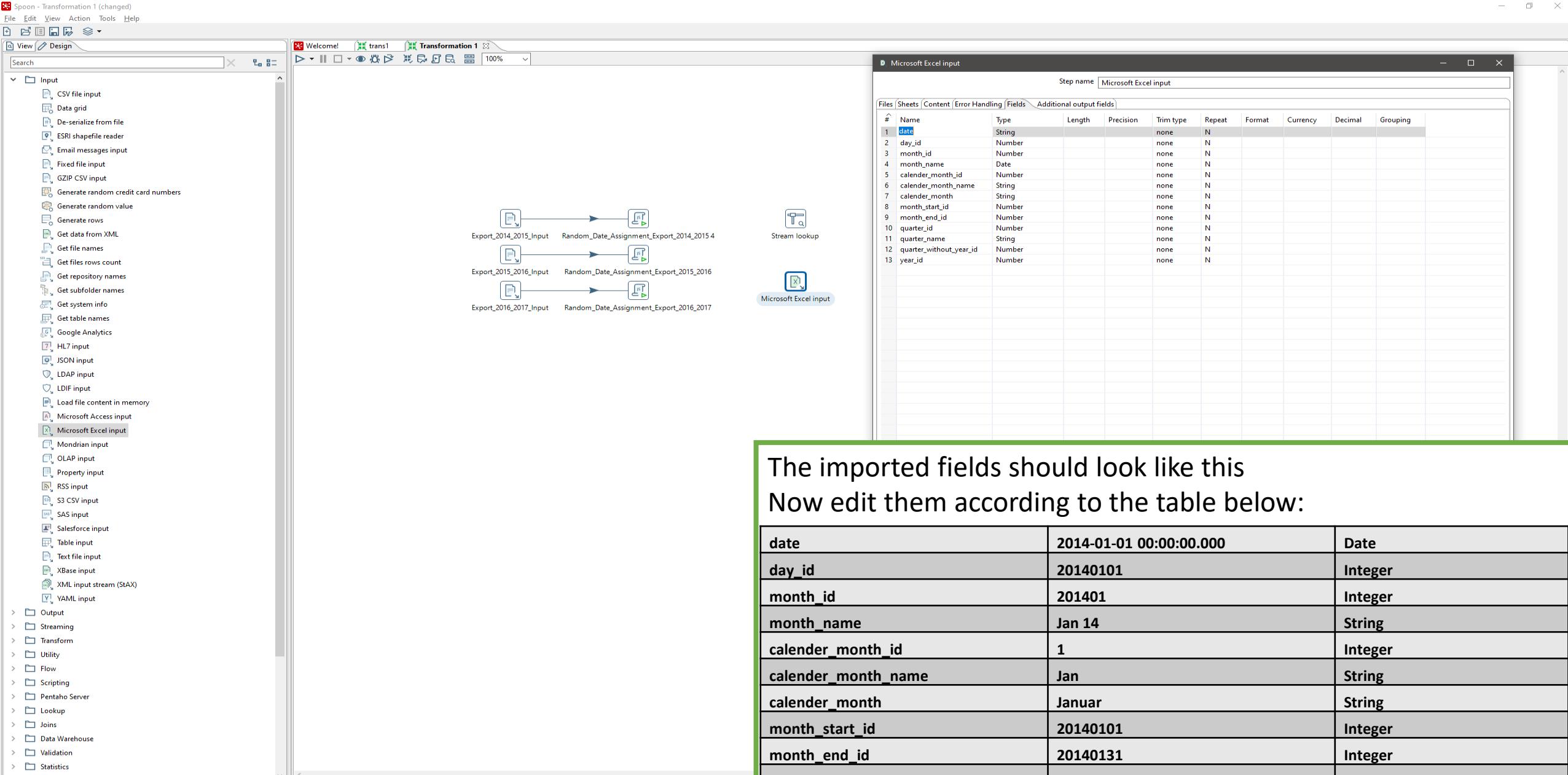
OK Preview rows Cancel

DIM_Time.xlsx should be added to the selected files
 Screen should look like stated above



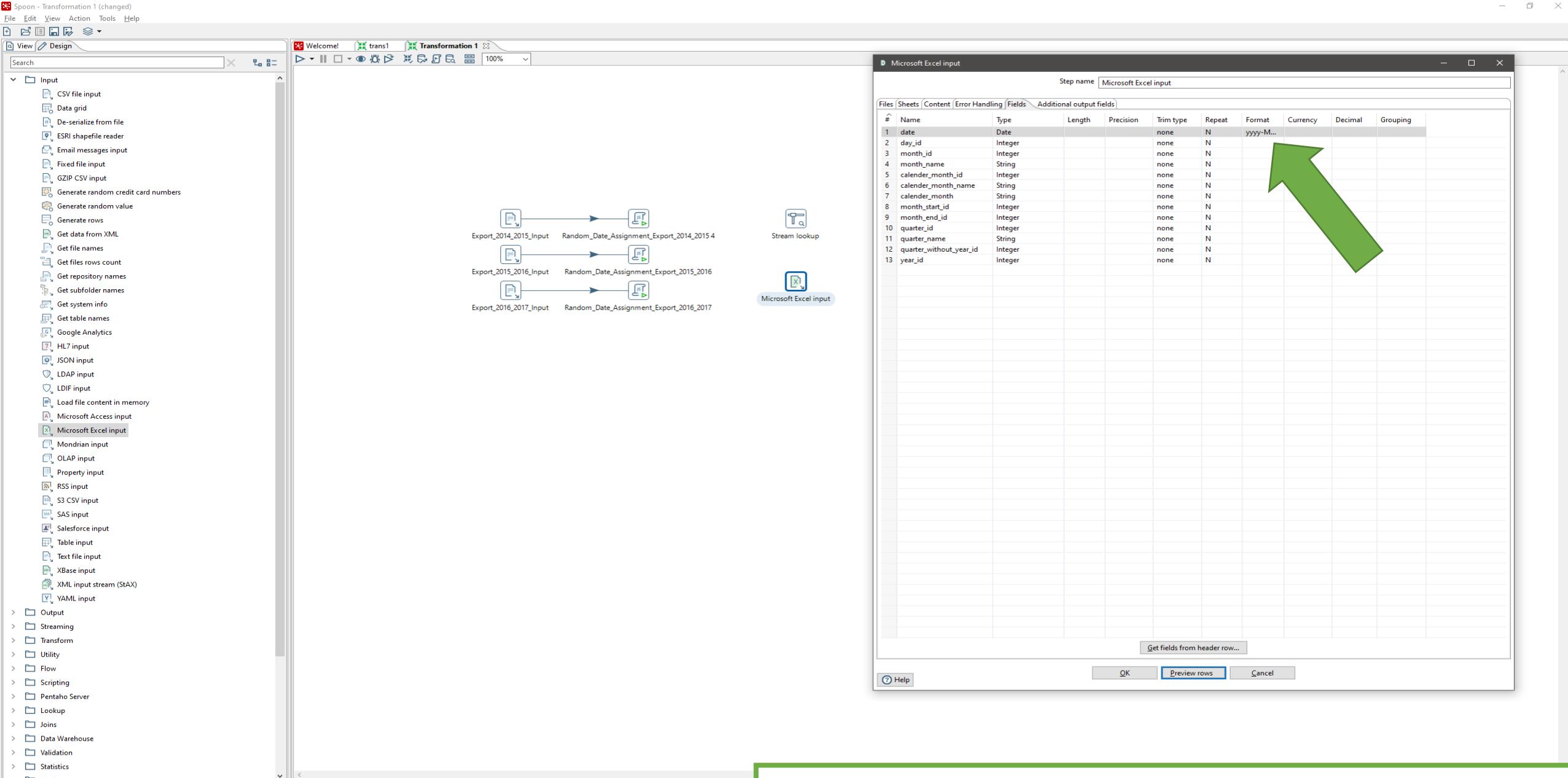
- (1) Click on the „!Sheets“ registry
- (2) Select „Get sheetnames...“
- (3) Press on „>>“ to add all items to your selection
Press OK



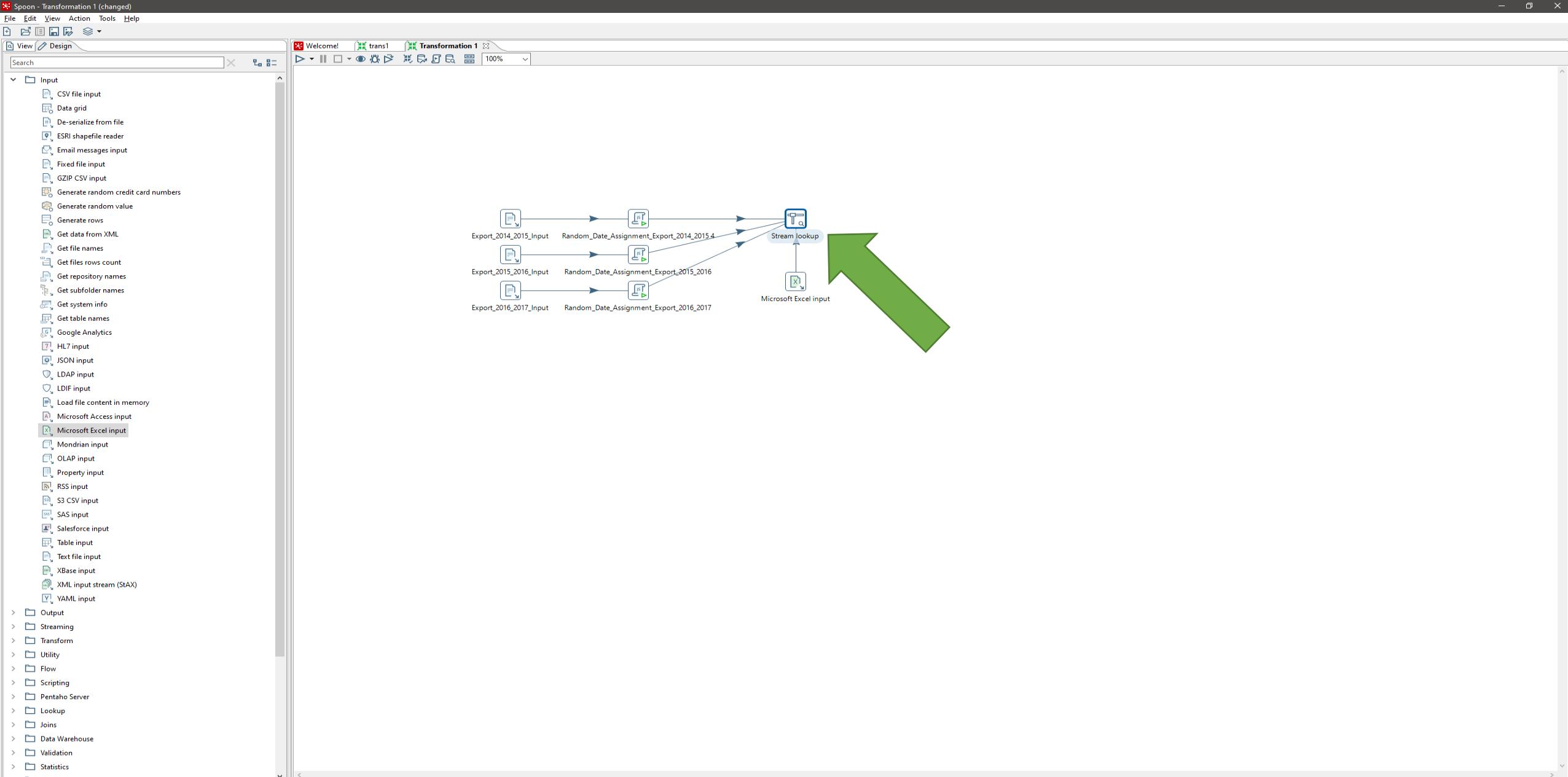


The imported fields should look like this
Now edit them according to the table below:

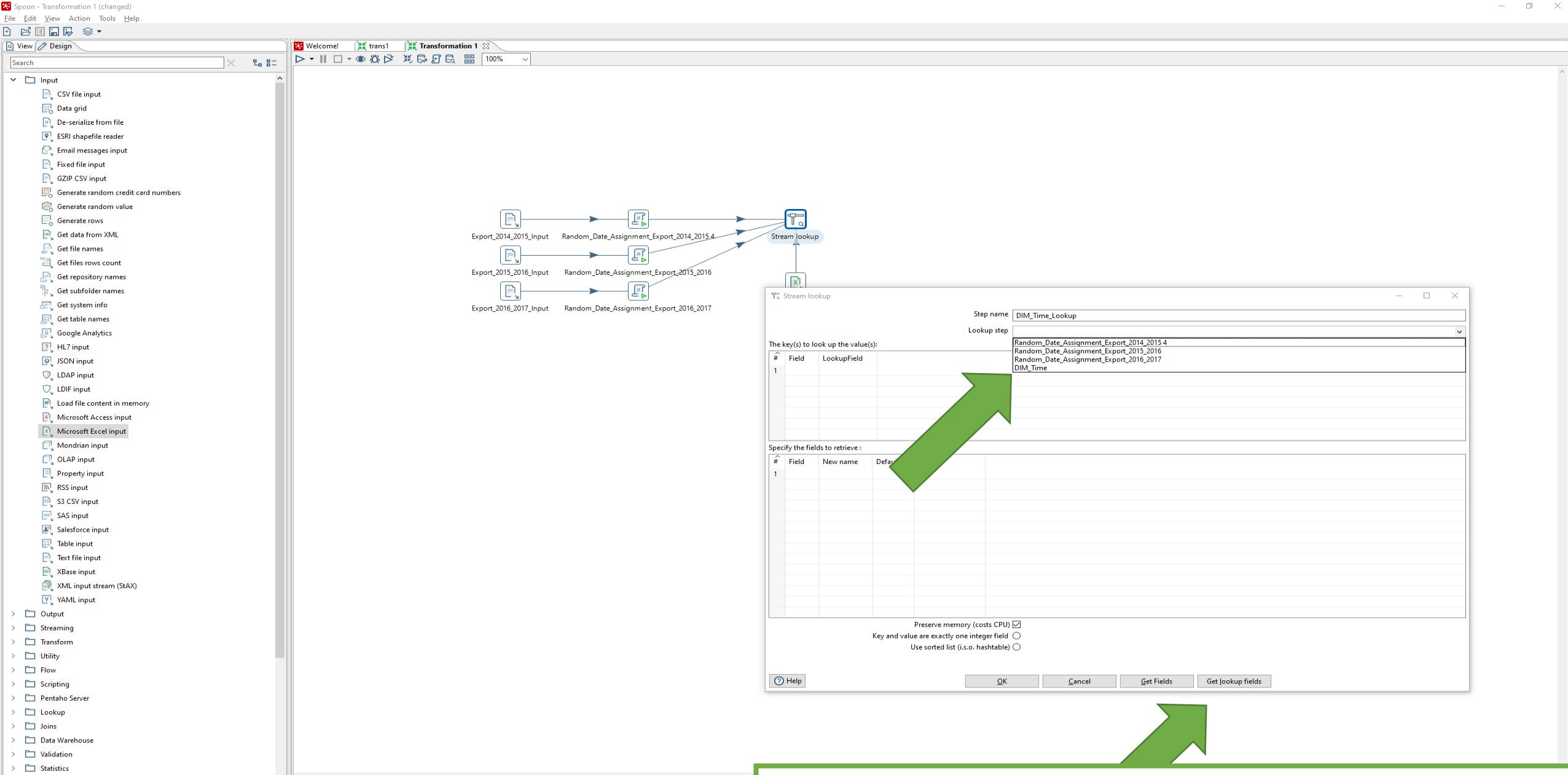
date	2014-01-01 00:00:00.000	Date
day_id	20140101	Integer
month_id	201401	Integer
month_name	Jan 14	String
calender_month_id	1	Integer
calender_month_name	Jan	String
calender_month	Januar	String
month_start_id	20140101	Integer
month_end_id	20140131	Integer
quarter_id	20141	Integer
quarter_name	Q1 14	String
quarter_without_year_ID	1	Integer
year_id	2014	Integer



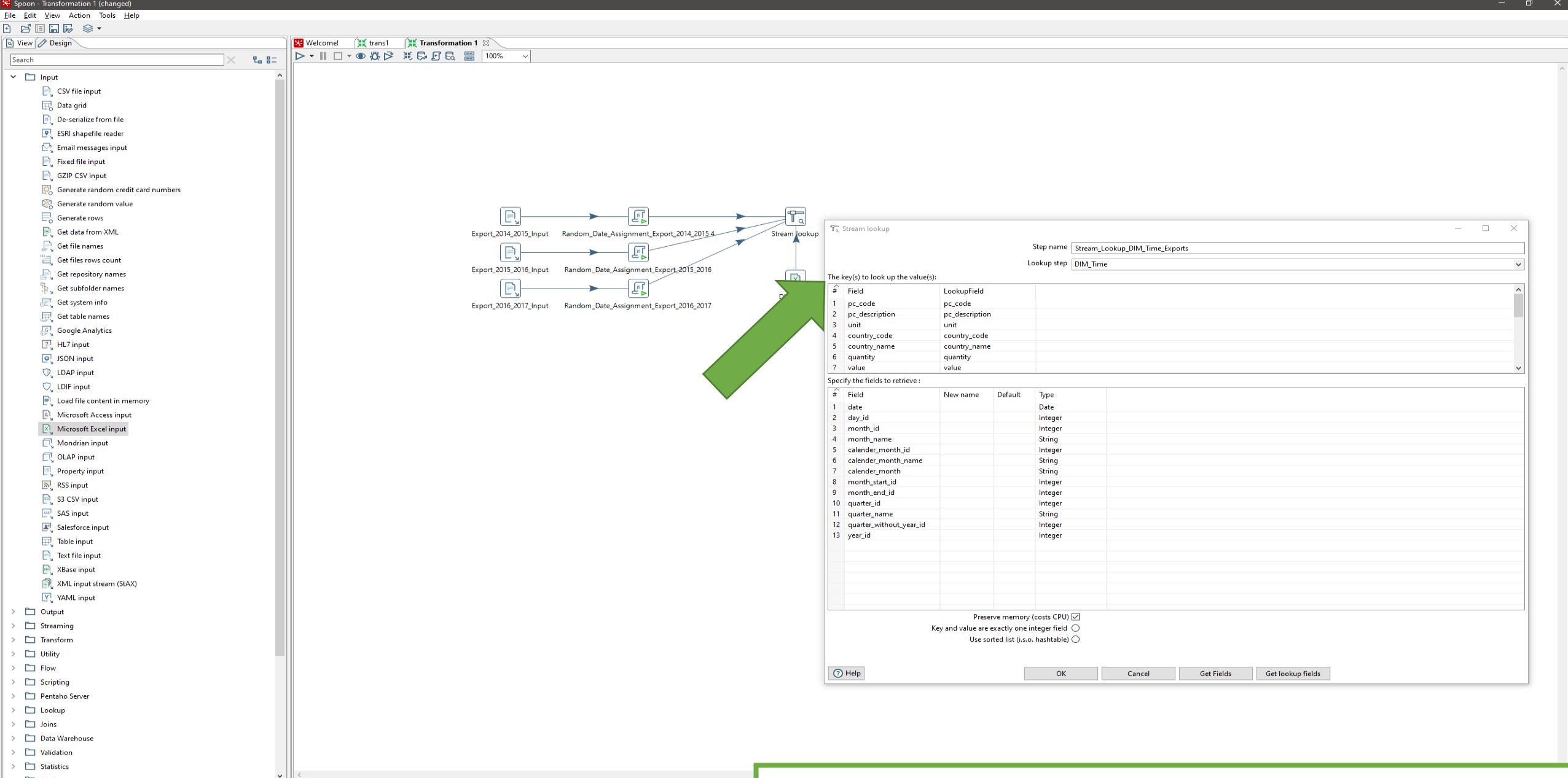
Do not forget to adjust the format for the date to:
 „yyyy-MM-dd HH:mm:ss.SSS“
 Press OK



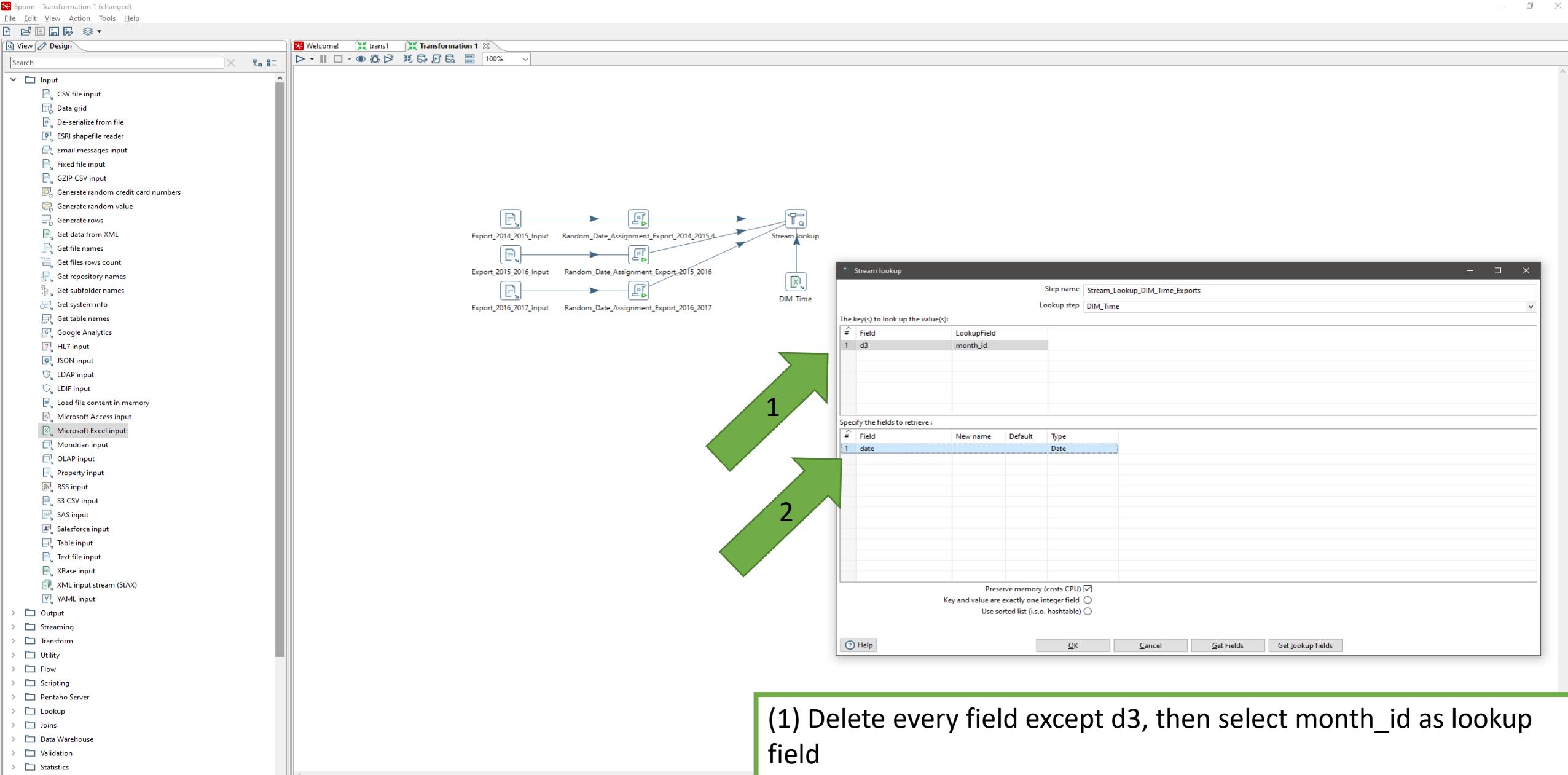
Now connect the steps as shown in the picture above
Then double-click „Stream lookup“ to edit it



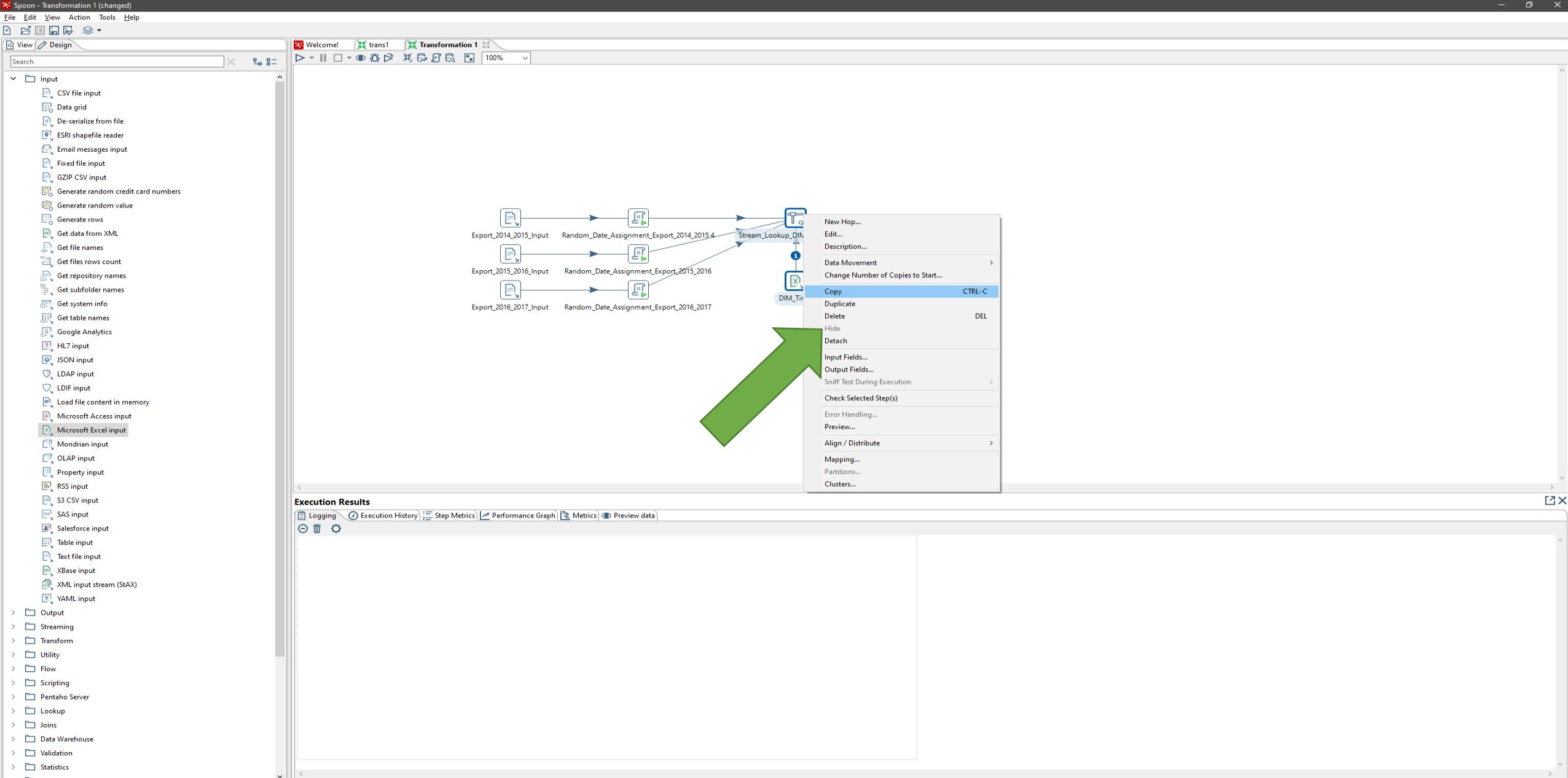
Name the step name properly
 Select the Lookup step which is „DIM_Time“ in our case
 Select „Get Fields“ and „Get lookup fields“



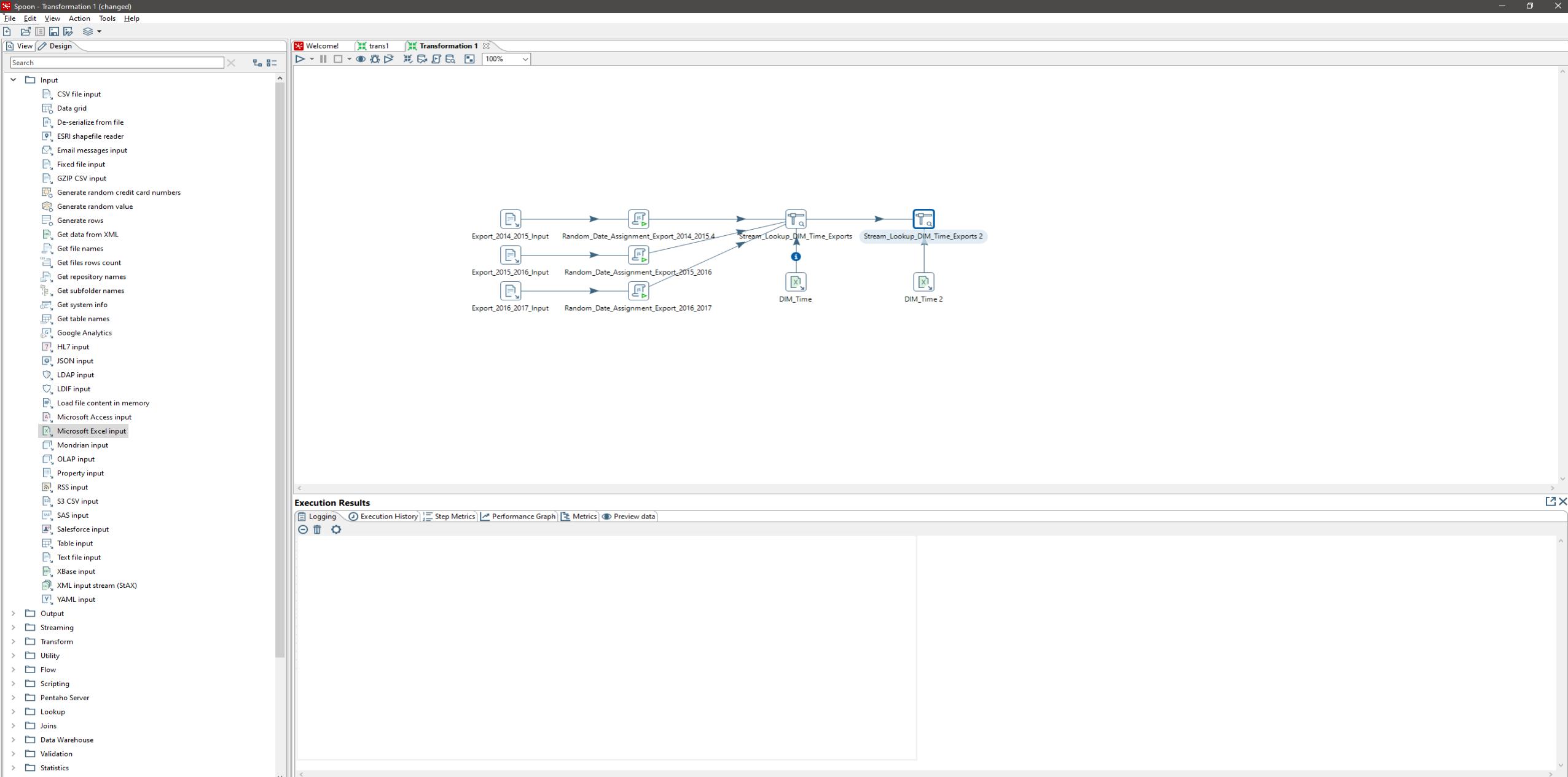
The screen should now look like this
 Since we look up the columns of DIM_Time, we already know that d3 is the primary key to connect with DIM_Time



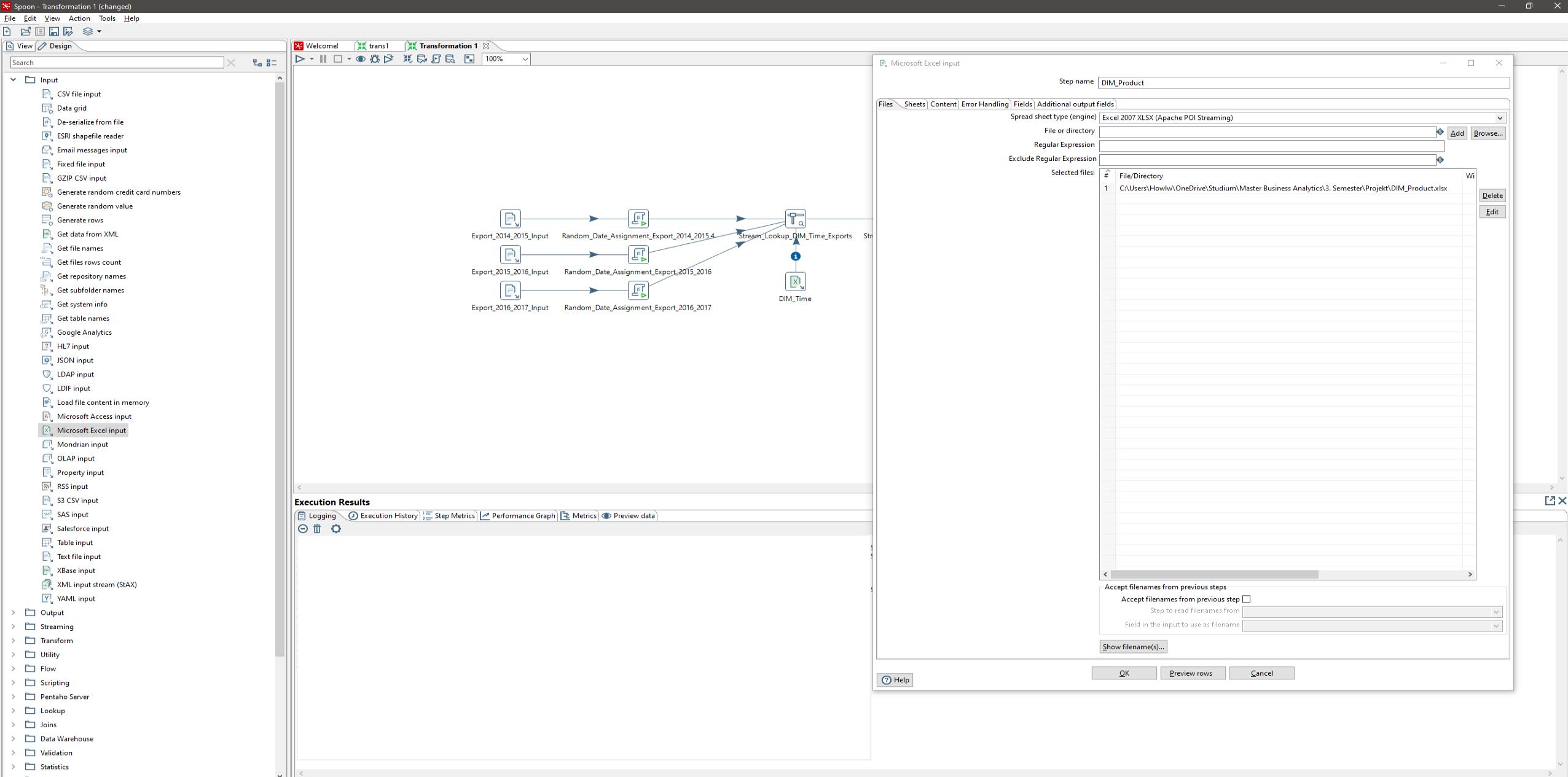
- (1) Delete every field except d3, then select month_id as lookup field
- (2) Since we need dates to visualize timelines in Tableau, this field is the only one we are interested in: Delete every field except date in the fields to retrieve



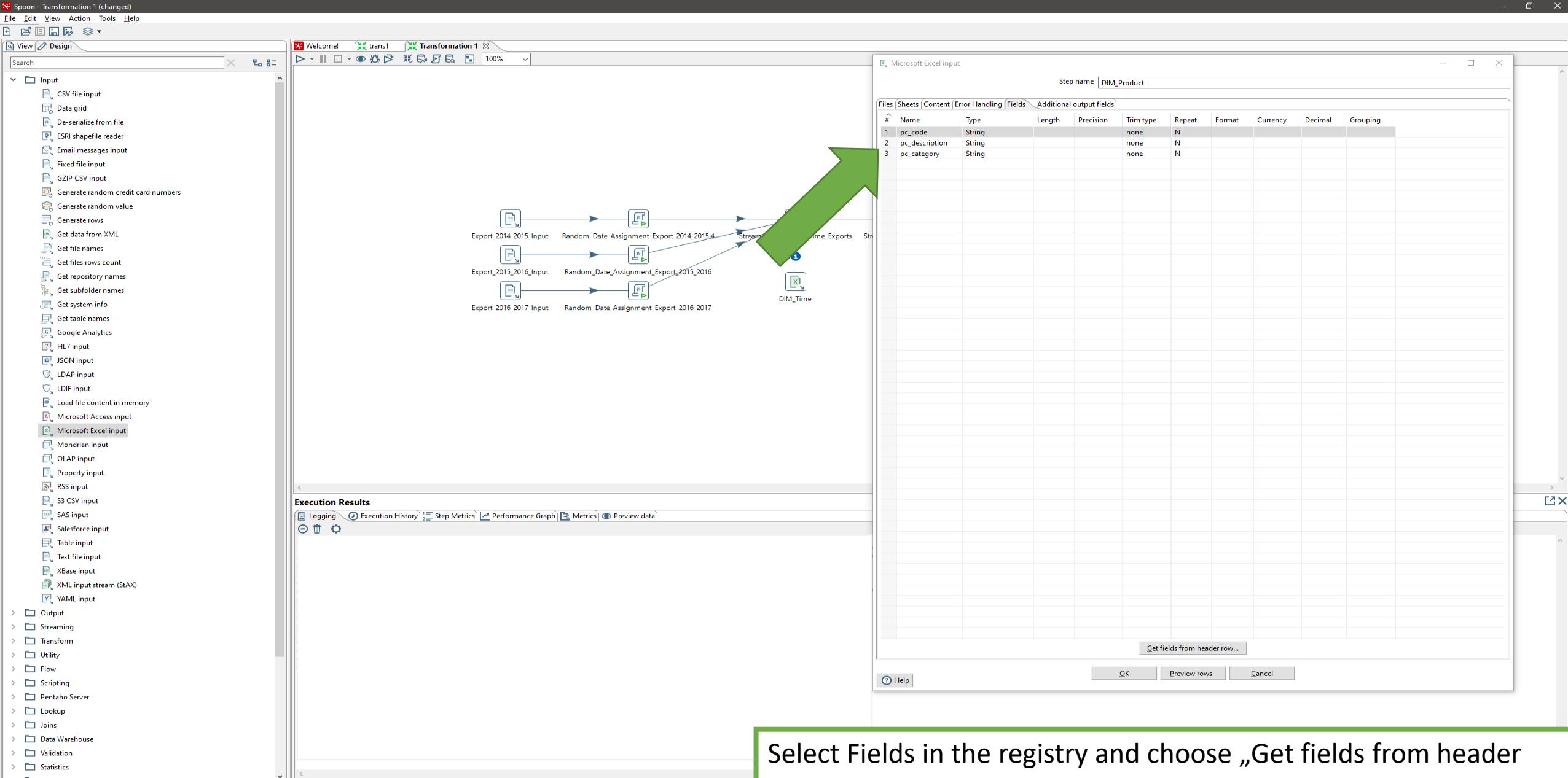
Copy and Paste both the lookup step and the Excel Input Step right next to the first ones



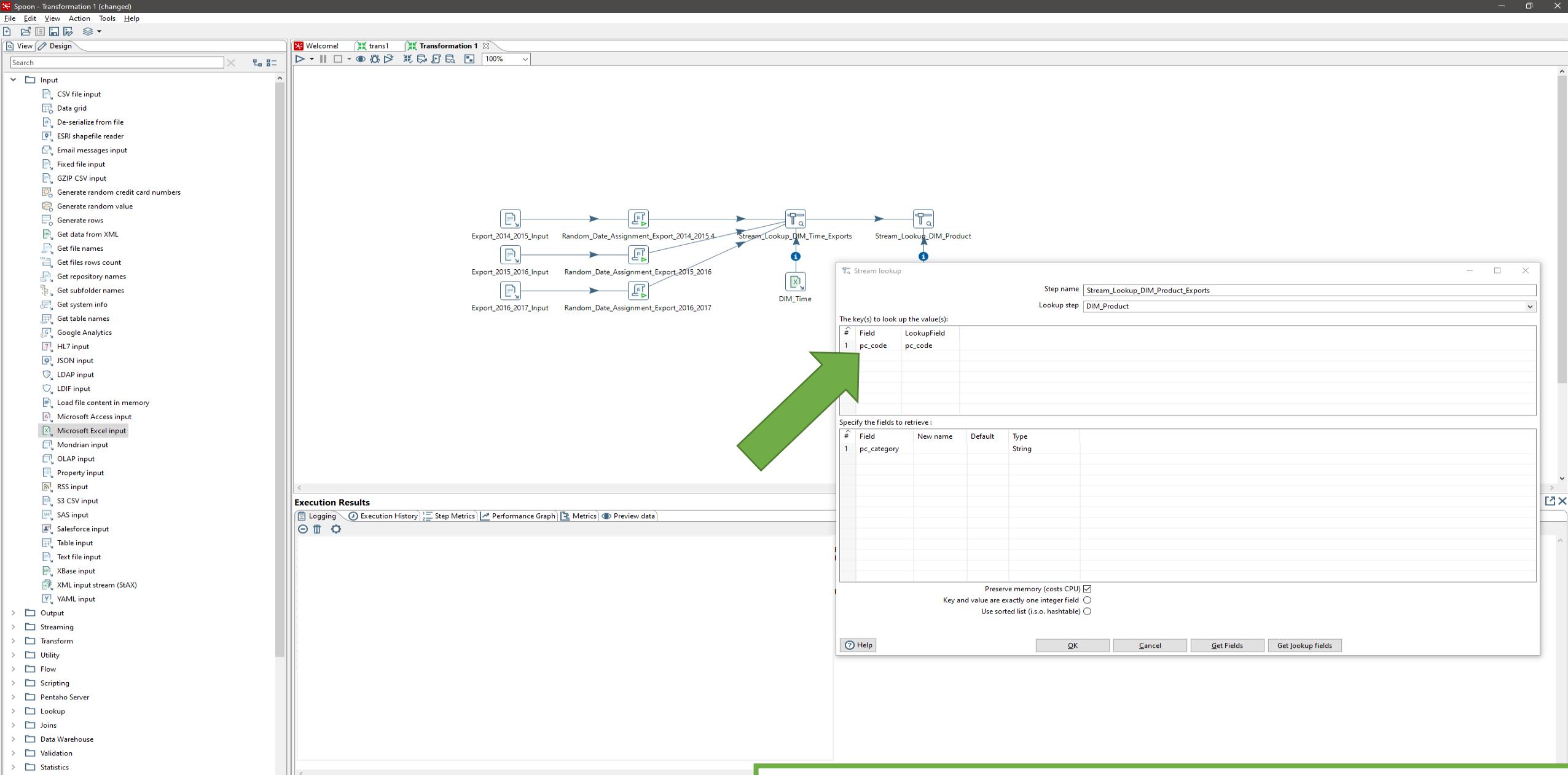
Connect the lookup steps and double-click the newly created Excel Input Step to set up DIM_Product



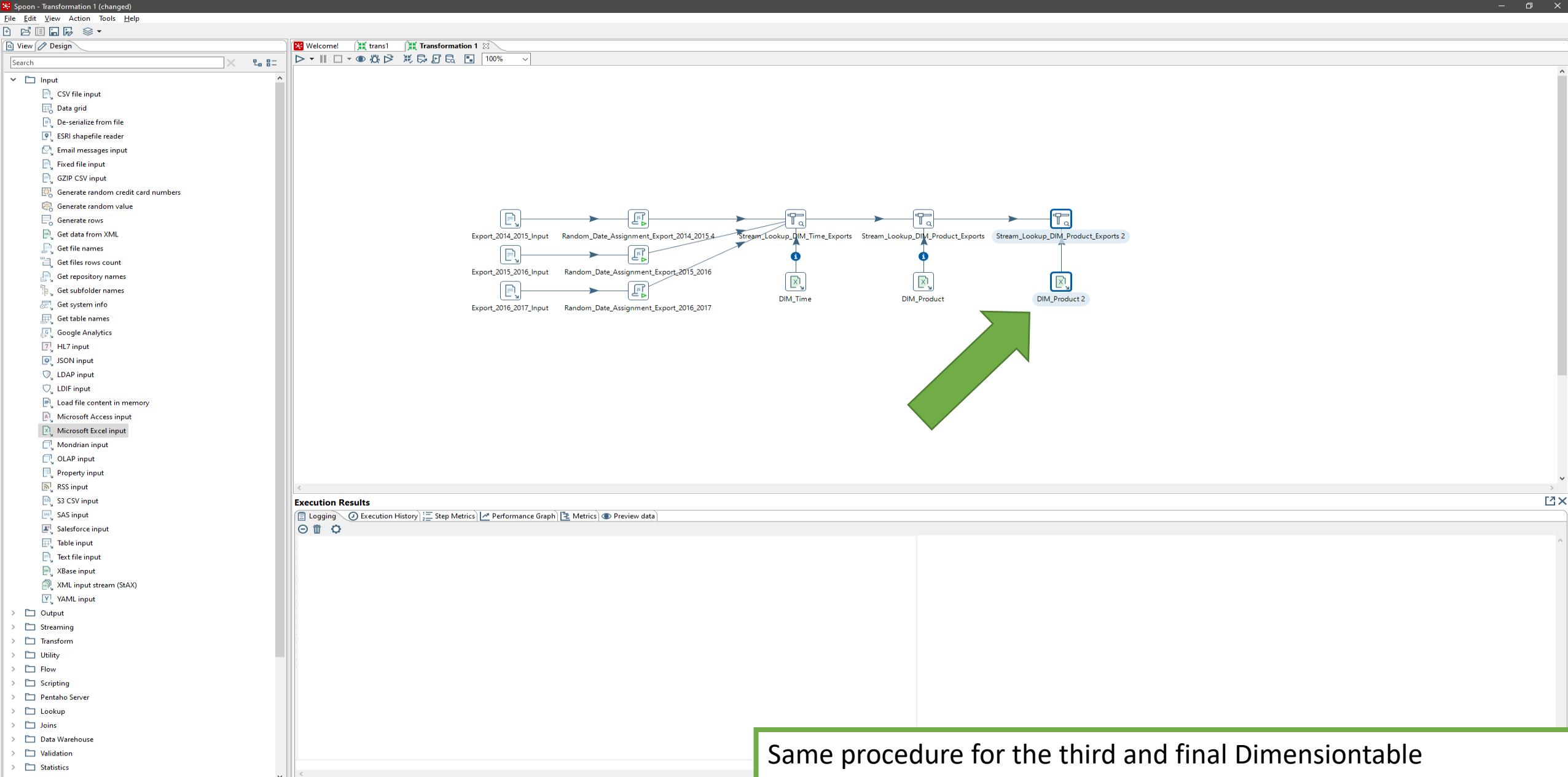
Rename it properly
Select „Browse...“ to select DIM_Product.xlsx and add it



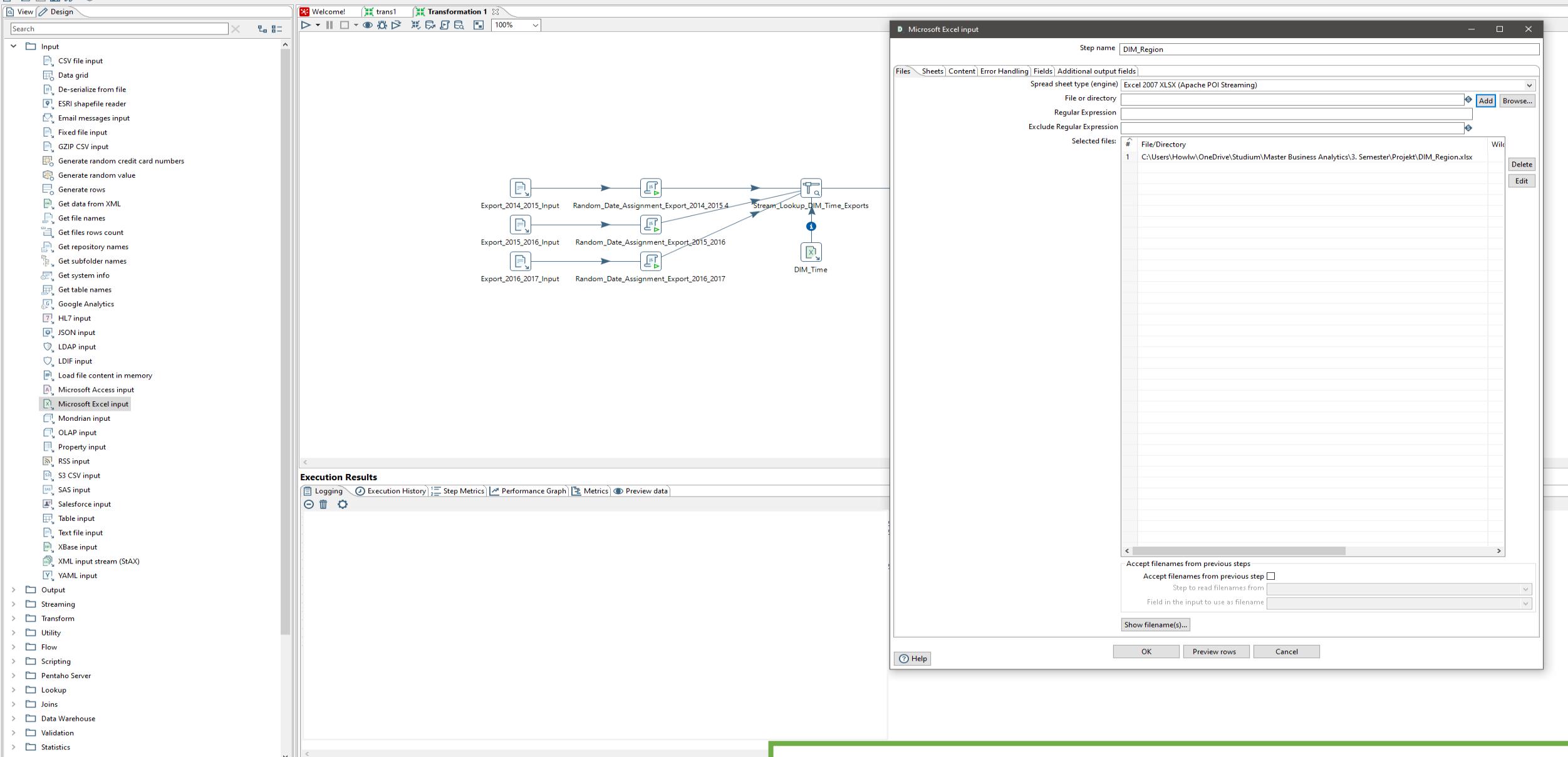
Select Fields in the registry and choose „Get fields from header row...“
The columns of DIM_Product.xlsx should now be loaded
Press OK



Double-click on the Lookup Step for DIM_Product and change it according to the picture above
 Press OK



Same procedure for the third and final Dimensiontable integration
 Again copy and paste both Stream Lookup and Microsoft Excel Input, connect the steps and then edit them



Name the Step properly
Select the DIM_Region.xlsx file via „Browse...“ and add it
Click on Fields

View Design

Search

Input

- CSV file input
- Data grid
- De-serialize from file
- ESRI shapefile reader
- Email messages input
- Fixed file input
- GZIP CSV input
- Generate random credit card numbers
- Generate random value
- Generate rows
- Get data from XML
- Get file names
- Get files rows count
- Get repository names
- Get subfolder names
- Get system info
- Get table names
- Google Analytics
- HL7 input
- JSON input
- LDAP input
- LDIF input
- Load file content in memory
- Microsoft Access input
- Microsoft Excel input
- Mondrian input
- OLAP input
- Property input
- RSS input
- S3 CSV input
- SAS input
- Salesforce input
- Table input
- Text file input
- XBase input
- XML input stream (StAX)
- YAML input

Output

Streaming

Transform

Utility

Flow

Scripting

Pentaho Server

Lookup

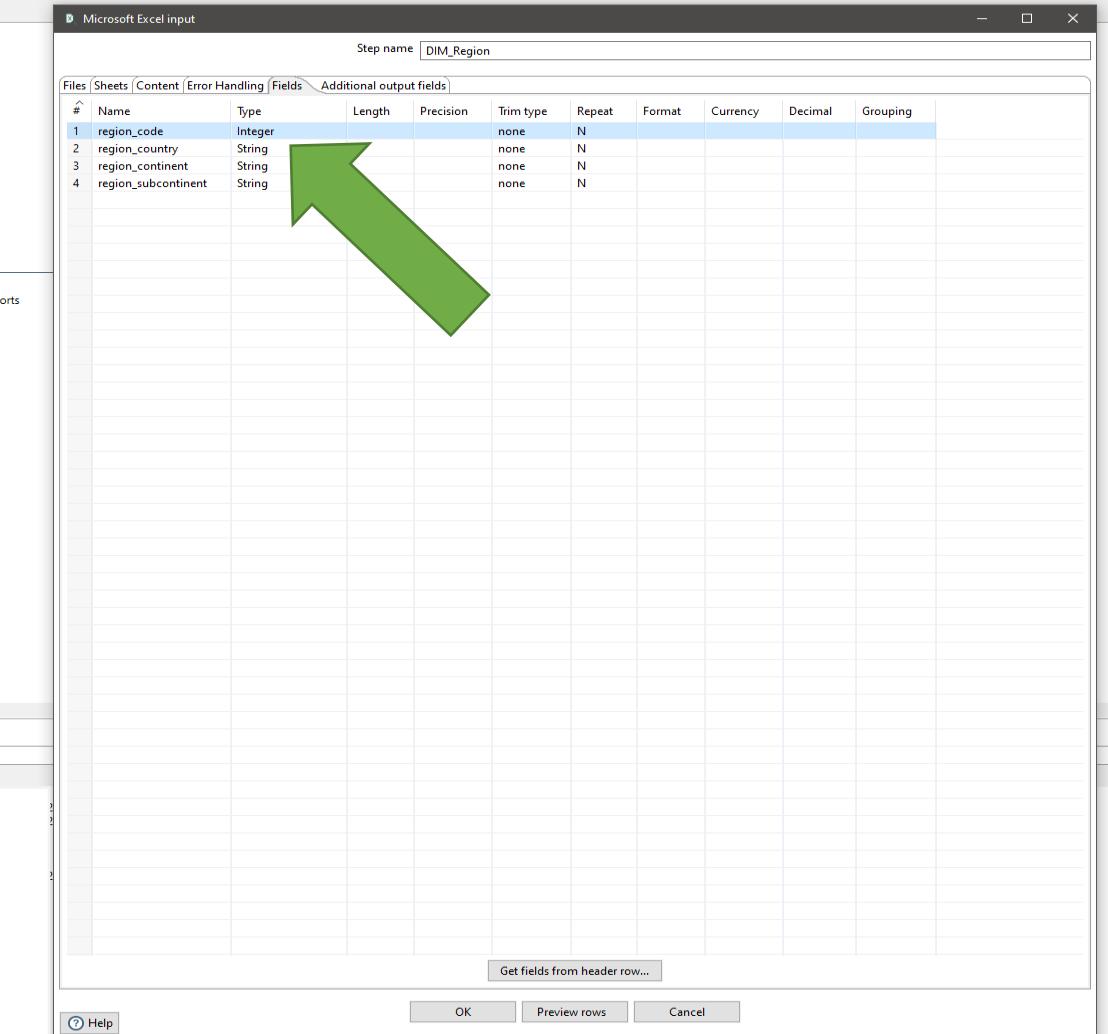
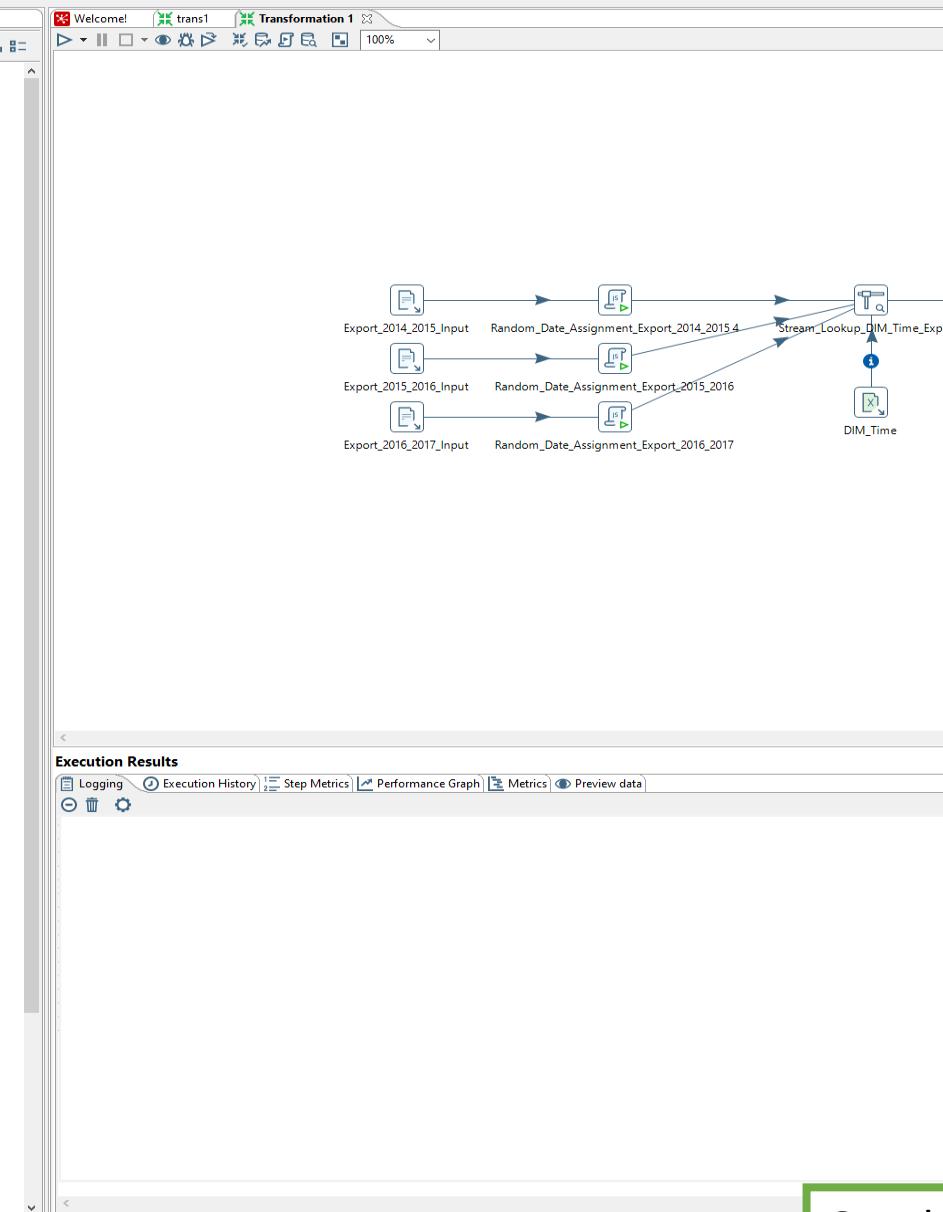
Join

Data Warehouse

Validation

Statistics

Bin Data



Get the fields from the header row
Do not forget to change the datatype of region_code to integer
since this was defined in the operational data

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

- > Input
- > Output
- > Streaming
- Transform**
 - Add XML
 - Add a checksum
 - Add constants
 - Add sequence
 - Add value fields changing sequence
 - Calculator
 - Closure generator
 - Concat fields
 - Get ID from slave server
 - Number range
 - Replace in string
 - Row denormaliser
 - Row flattener
 - Row normaliser
 - Select values
 - Set field value
 - Set field value to a constant
 - Sort rows
 - Split field to rows
 - Split fields
 - String operations
 - Strings cut
 - Unique rows
 - Unique rows (HashSet)
 - Value mapper
 - XSL transformation
- > Utility
- > Flow
- > Scripting
- > Pentaho Server
- > Lookup
- > Joins
- > Data Warehouse
- > Validation
- > Statistics
- > Big Data
- > Agile
- > Cryptography
- > Job
- > Mapping
- > Bulk loading
- > Inline
- > Experimental
- > Deprecated
- > History

Welcome! trans1 Transformation 1

100%

Stream lookup

Step name: Stream_Lookup_DIM_Region_Exports

Lookup step: DIM_Region

The key(s) to look up the value(s):

#	Field	LookupField
1	country_code	region_code

Specify the fields to retrieve:

#	Field	New name	Default	Type
1	region_country			String
2	region_continent			String
3	region_subcontinent			String

Preserve memory (costs CPU)

Key and value are exactly one integer field

Use sorted list (i.s.o. hashtable)

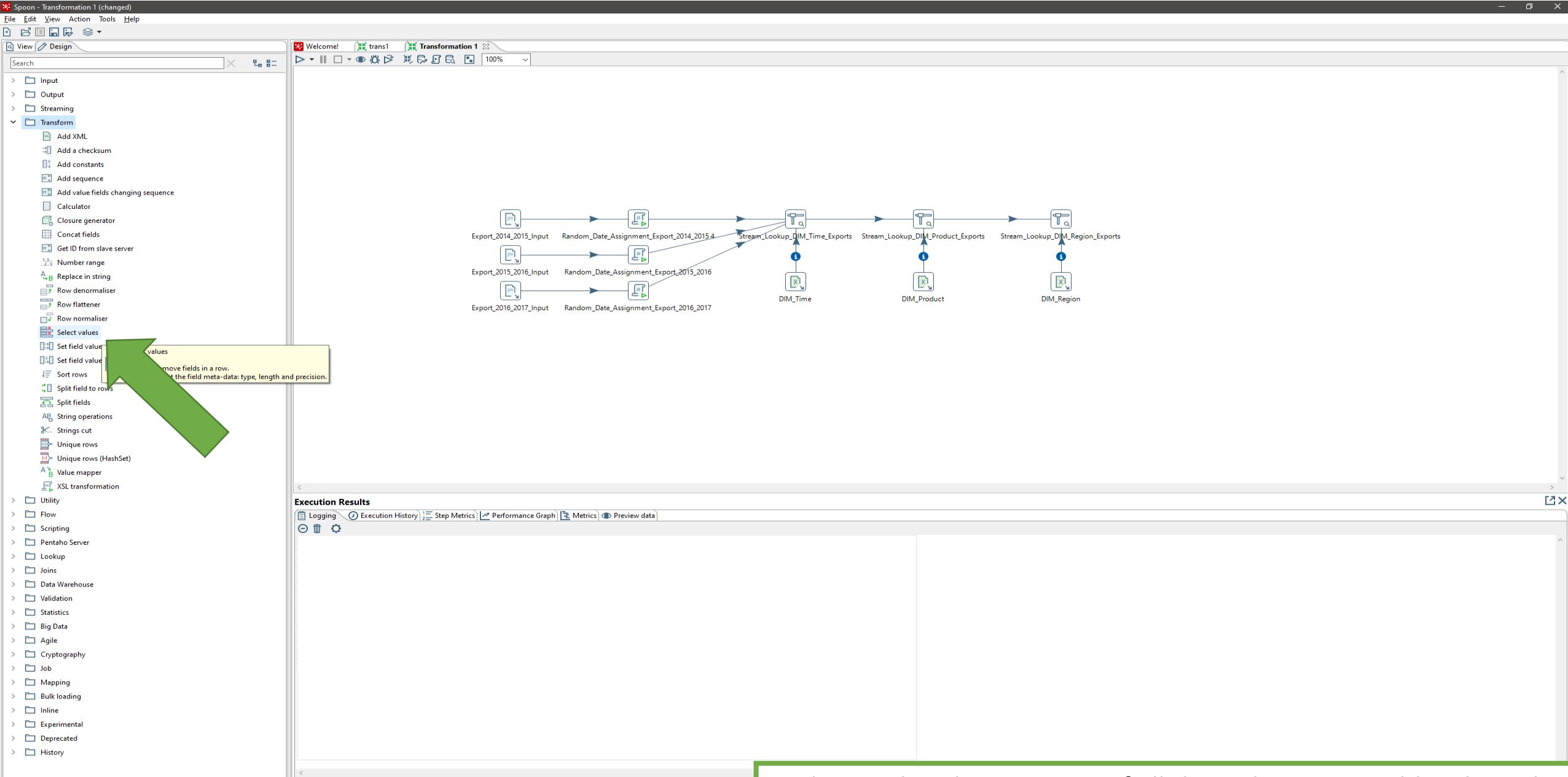
OK Cancel Get Fields Get lookup fields

Execution Results

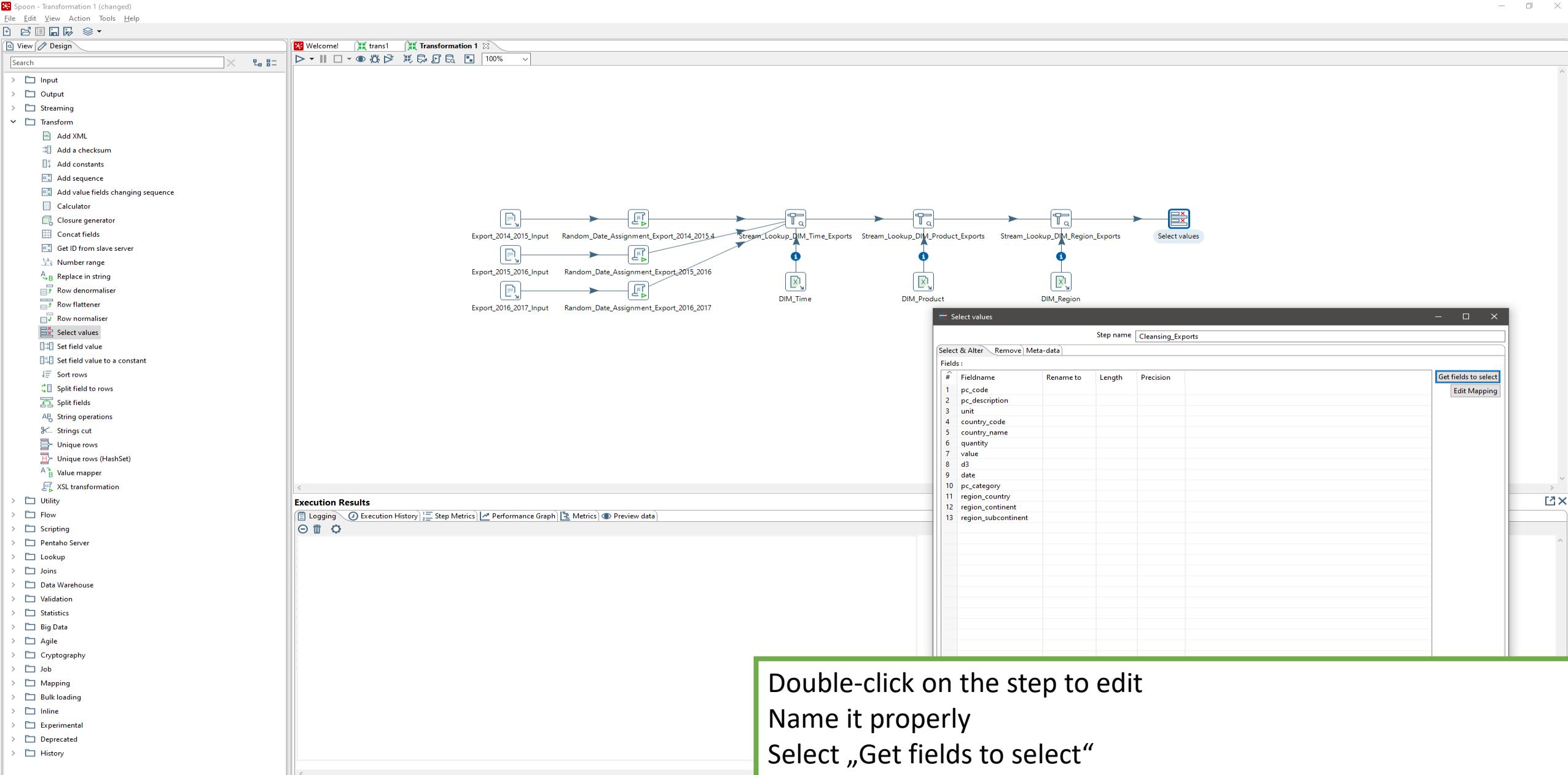
Logging Execution History Step Metrics Performance Graph Metrics Preview data

PC_Export_2015_2016.csv

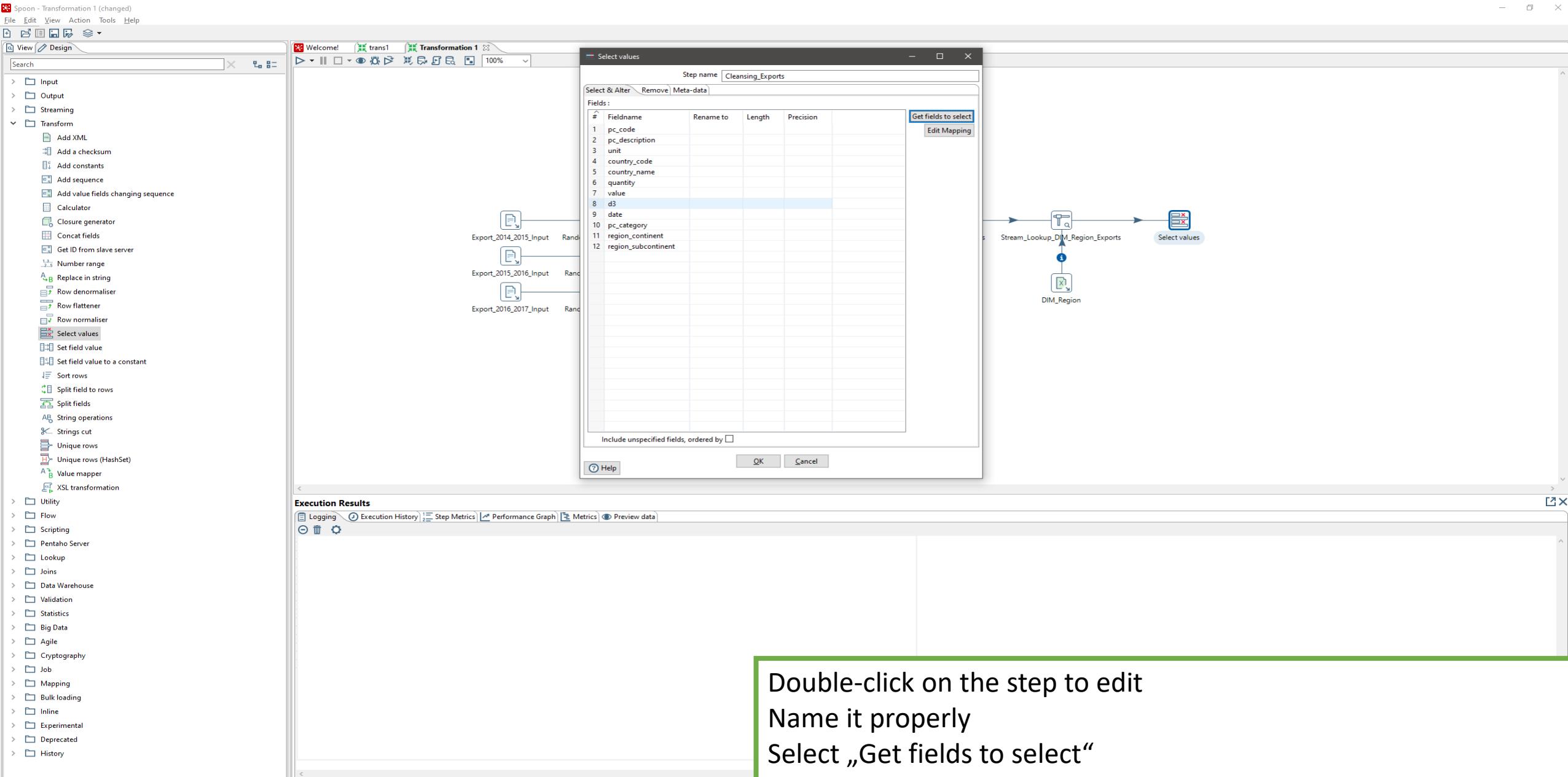
Edit the Lookup Step as stated on the picture above
Press OK



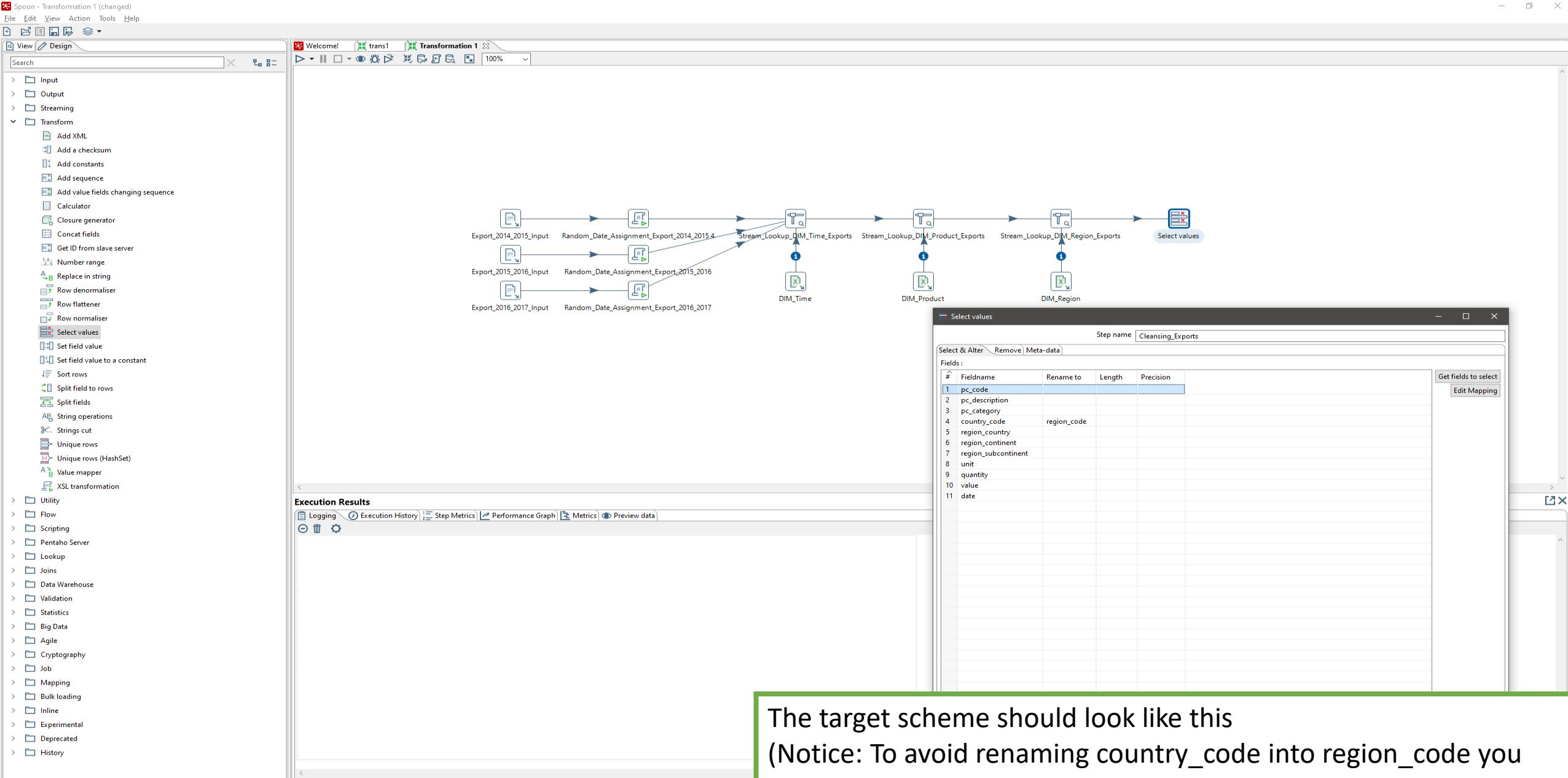
With completed integration of all three dimension tables the only thing left is cleansing the data and finally outputting it
 Drag the „Select values“ step onto the main screen



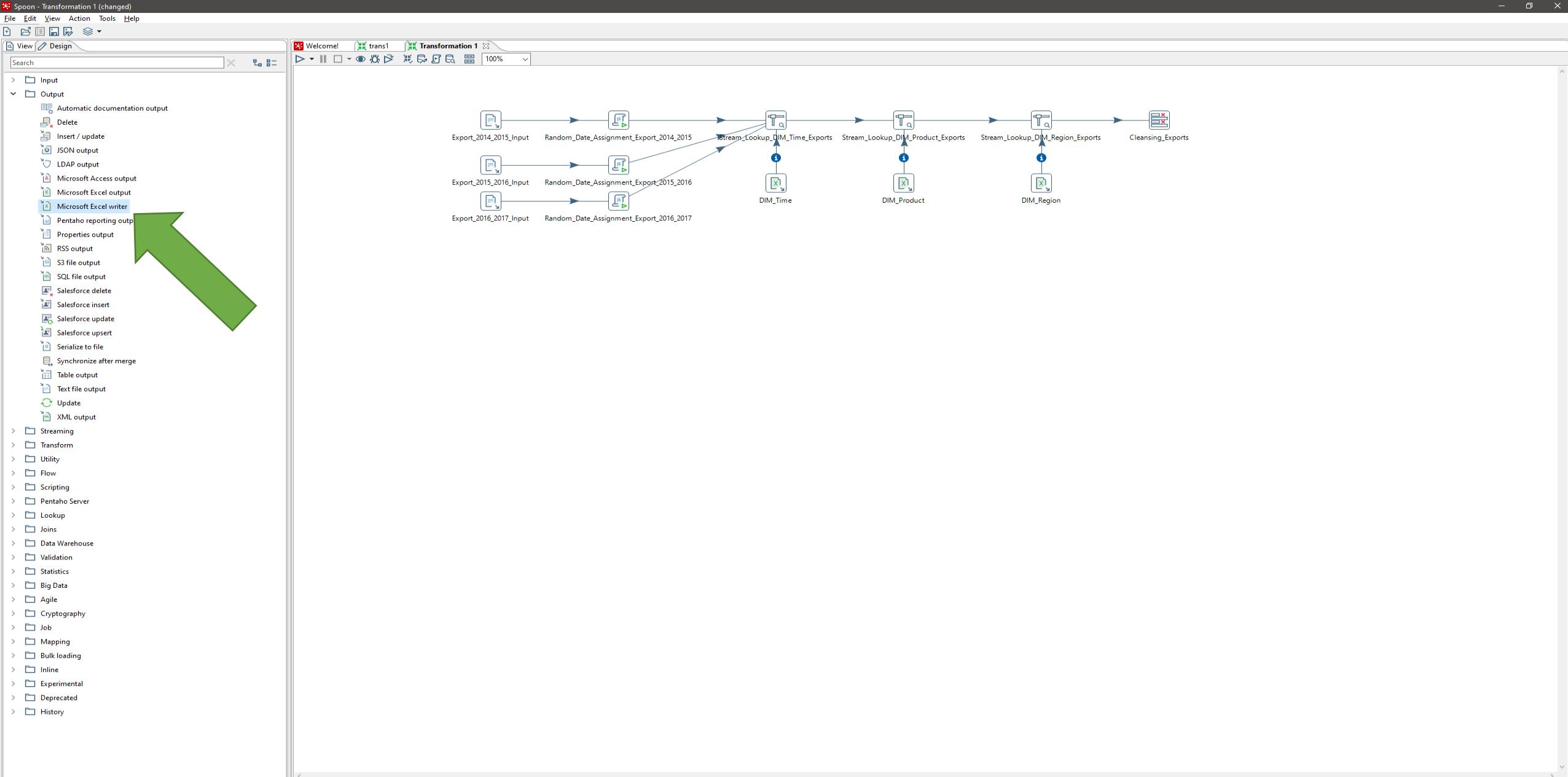
Double-click on the step to edit
Name it properly
Select „Get fields to select“
To bring it into order select the corresponding number next to the filename and press STRG+Arrow Up to move rows upwards and STRG+Arrow Down to move rows downwards



Double-click on the step to edit
 Name it properly
 Select „Get fields to select“
 To bring it into order select the corresponding number next to the filename and press STRG+Arrow Up to move rows upwards and STRG+Arrow Down to move rows downwards



The target schema should look like this
 (Notice: To avoid renaming country_code into region_code you can retrieve the region_code field from DIM_region and finally delete country_code in this step)
 Press OK



Finally for the Output Step drag „Microsoft Excel Writer“ onto the main screen and double-click it

Spoon - Transformation 1 (changed)
File Edit View Action Tools Help

View Design

Search

Input

Output

- Automatic documentation output
- Delete
- Insert / update
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Microsoft Excel writer
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL file output
- Salesforce delete
- Salesforce insert
- Salesforce update
- Salesforce upsert
- Serialize to file
- Synchronize after merge
- Table output
- Text file output
- Update
- XML output

Streaming

Transform

Utility

Flow

Scripting

Pentaho Server

Lookup

Joins

Data Warehouse

Validation

Statistics

Big Data

Agile

Cryptography

Job

Mapping

Bulk loading

Inline

Experimental

Deprecated

History

Welcome! trans1 Transformation 1

Step name: Exports_Output

File & Sheet Content

File

- Name: Exports
- Extension: xlsx [Excel 2007 and above]
- Stream XSLX data
- split every ... data rows: 0
- Include stepnr in filename?
- Include date in filename?
- Include time in filename?
- Specify Date time format
- Date time format:
- Show filename(s)...
- If output file exists: replace with new output file
- Wait for first row before creating file
- Add filenames to result: checked

Sheet

- Sheet name (max. 31 characters): Sheet1
- Make this the active sheet: checked
- If sheet exists in output file: replace with new sheet
- Protect sheet? (XLS format only)
- Protected by user
- Password

Template

- Use template when creating new files: unchecked
- Template file: template.xls
- Use template when creating new sheets: unchecked
- Template sheet:
- Hide Template Sheet:

OK Cancel ? Help

Stream_Lookup_DIM_Region_Exports → Cleansing_Exports → Microsoft Excel writer

DIM_Region

```
graph LR; Stream[Stream_Lookup_DIM_Region_Exports] --> Cleansing[Cleansing_Exports]; Cleansing --> Excel[Microsoft Excel writer]
```

Name it properly
Click „Browse...“ to choose the destination for saving the file

The screenshot shows the Spoon - Transformation 1 interface. A green arrow points from the left towards the 'Content' tab of the 'Microsoft Excel writer' configuration dialog. The dialog is titled 'trans1 Transformation 1' and contains settings for writing to an Excel file named 'Output_Exports'. It includes sections for 'File & Sheet' and 'Content'. The 'Content' tab is active, showing options like 'Start writing at cell A1', 'When writing rows: overwrite existing cells', and a 'Fields' table.

File & Sheet

Step name: Output_Exports

Start writing at cell: A1
When writing rows: overwrite existing cells
Write Header:
Write Footer:
Auto size columns:
Force formula recalculation:
Leave styles of existing cells unchanged:

When writing to existing sheet

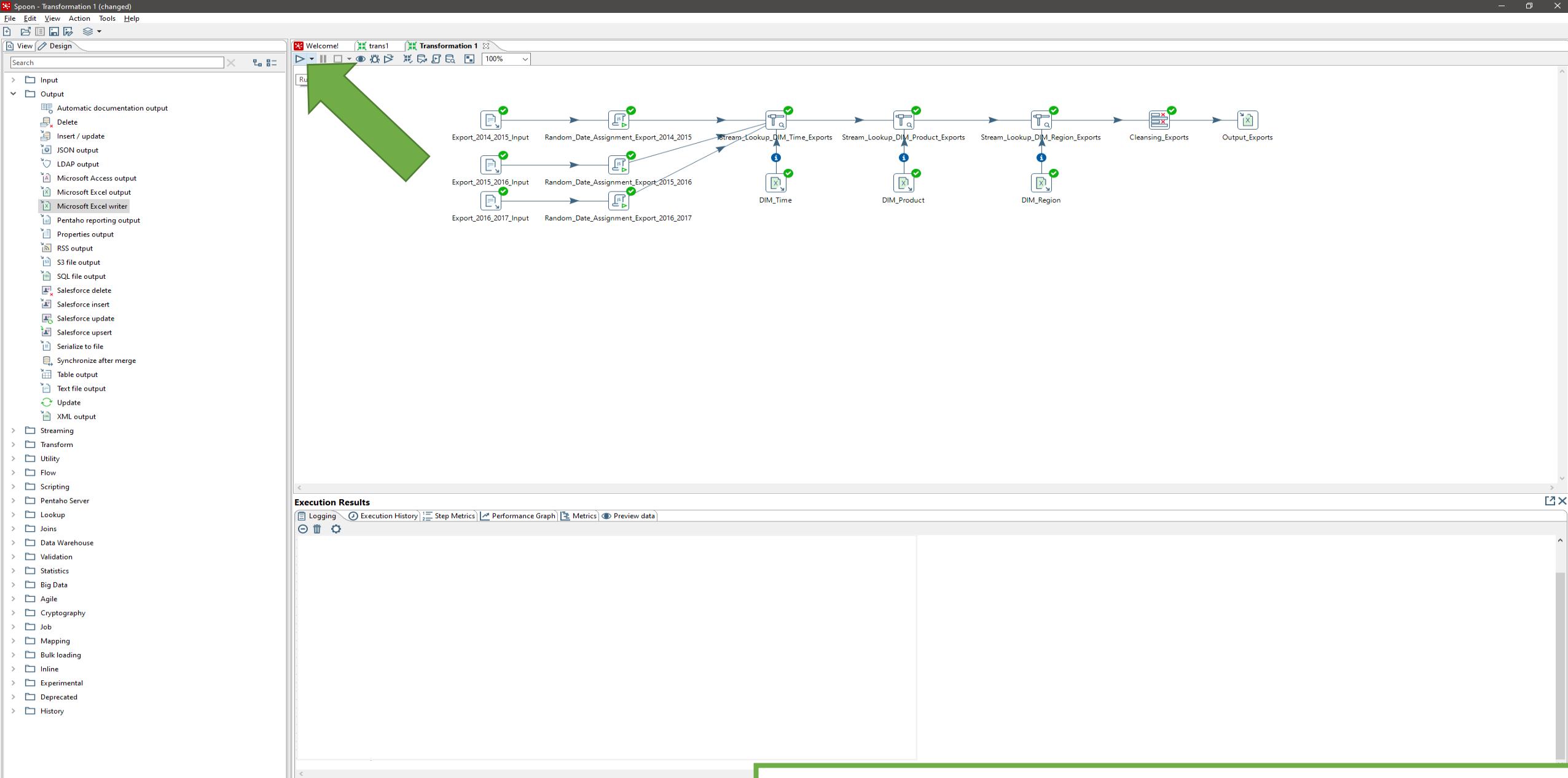
Start writing at end of sheet (appending lines):
Offset by ... rows: 0
Begin by writing ... empty lines: 0
Omit header:

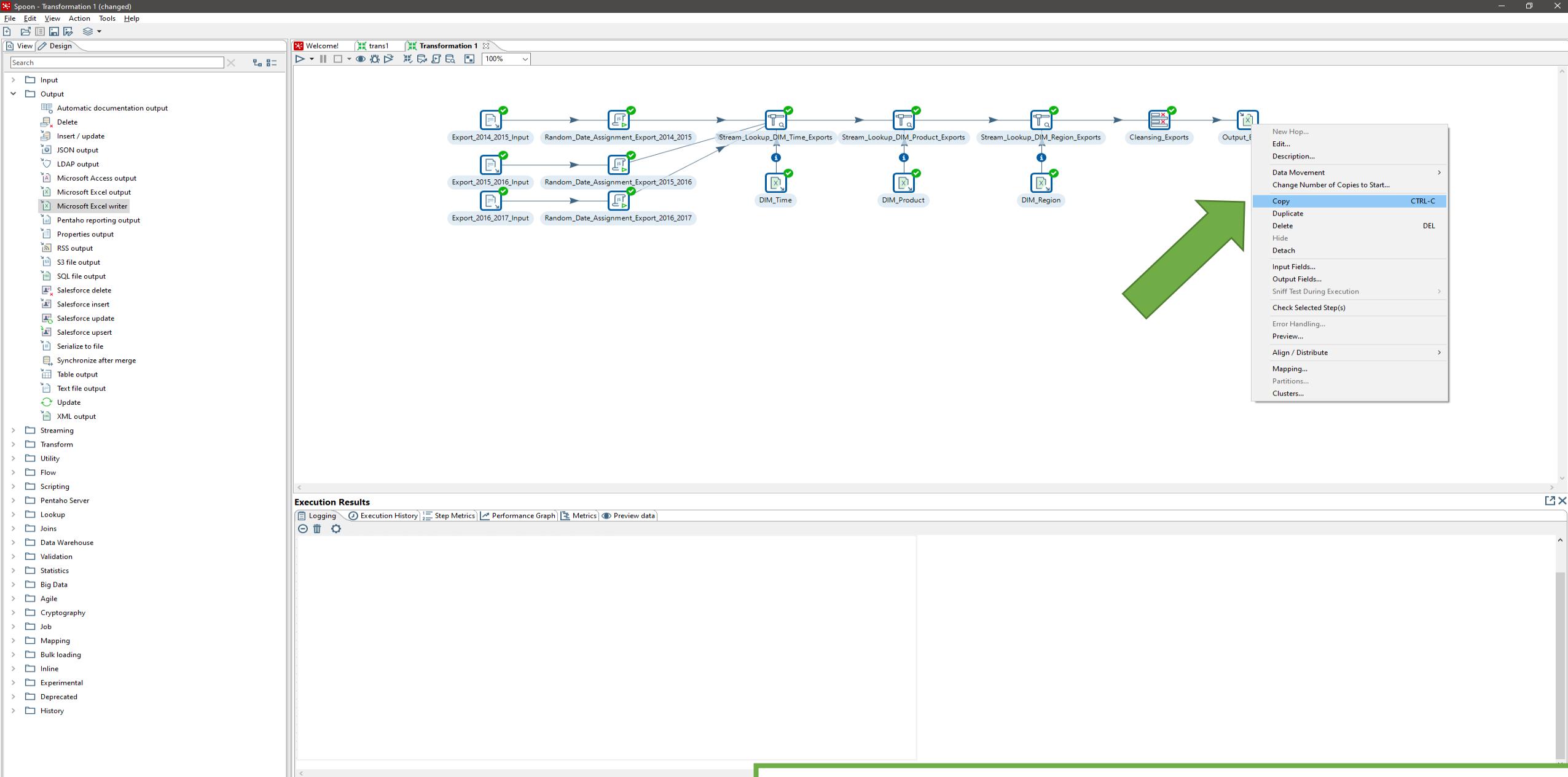
Fields

#	Name	Type	Format	Style from cell	Field title	Header/Footer style from cell	field contains formula	Hyperlink	Cell comment (XLSX)	Cell c
1	pc_code	String			pc_code		N			
2	pc_description	String			pc_description		N			
3	pc_category	String			pc_category		N			
4	region_code	Integer			region_code		N			
5	region_country	String			region_country		N			
6	region_continent	String			region_continent		N			
7	region_subcontinent	String			region_subcontinent		N			
8	unit	String			unit		N			
9	quantity	String			quantity		N			
10	value	Number			value		N			
11	date	Date	mmmm-yy		date		N			

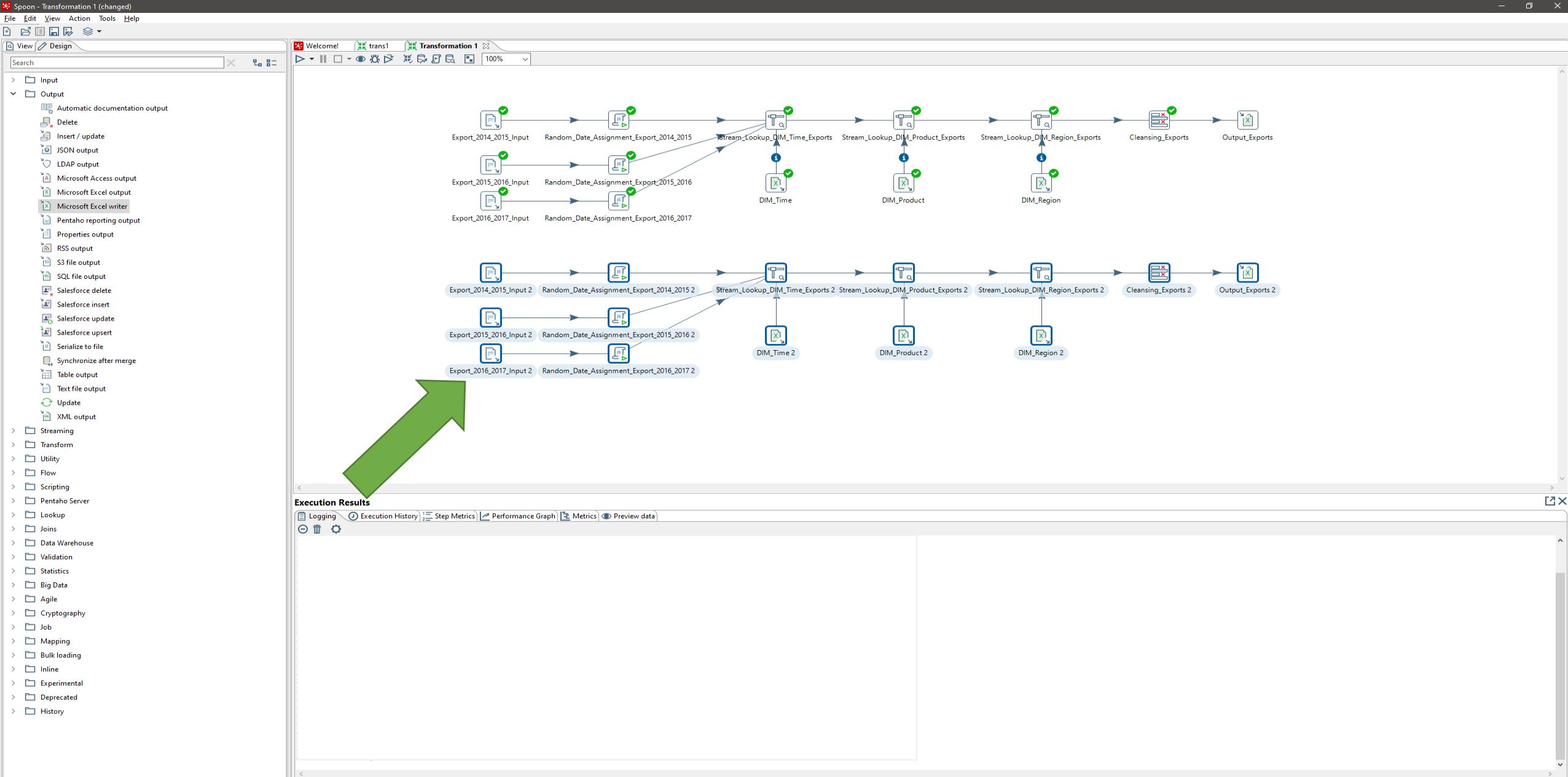
Get Fields Minimal width OK Cancel Help

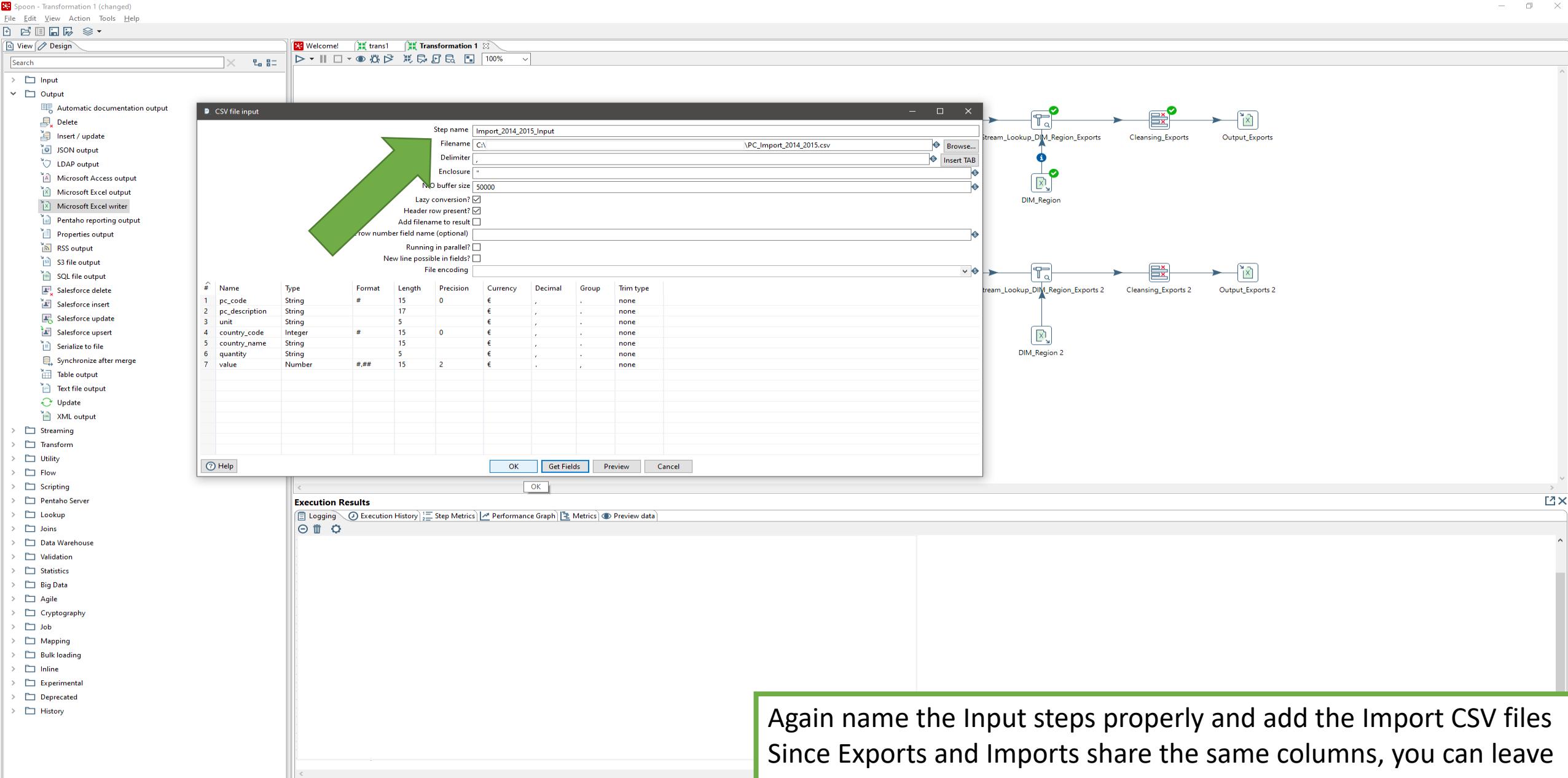
Switch to Content, click „Get Fields“ and edit the Fields accordingly
Press OK



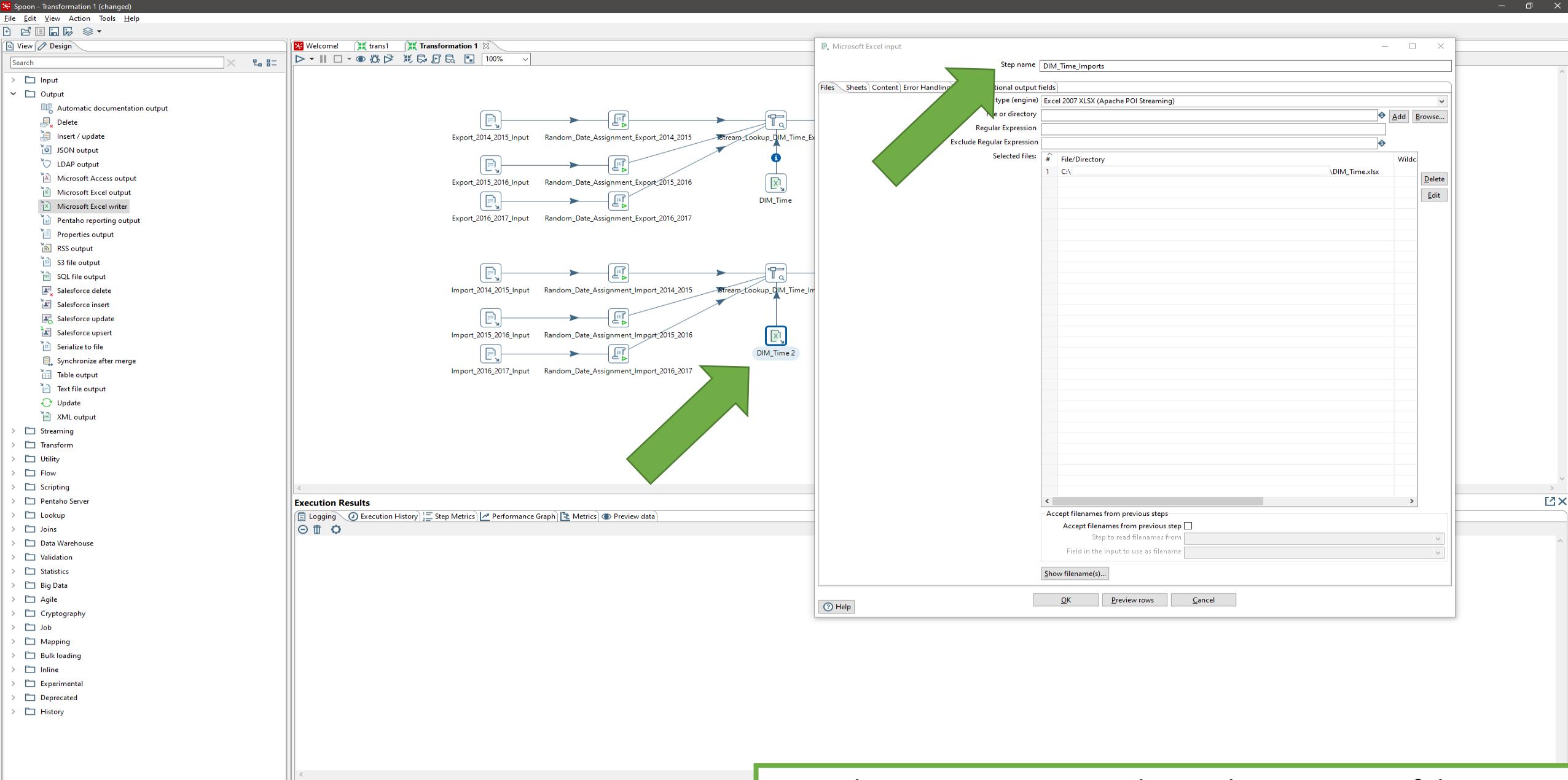


Now that the Exports are done, select the whole process and copy it
We are now building the Import process

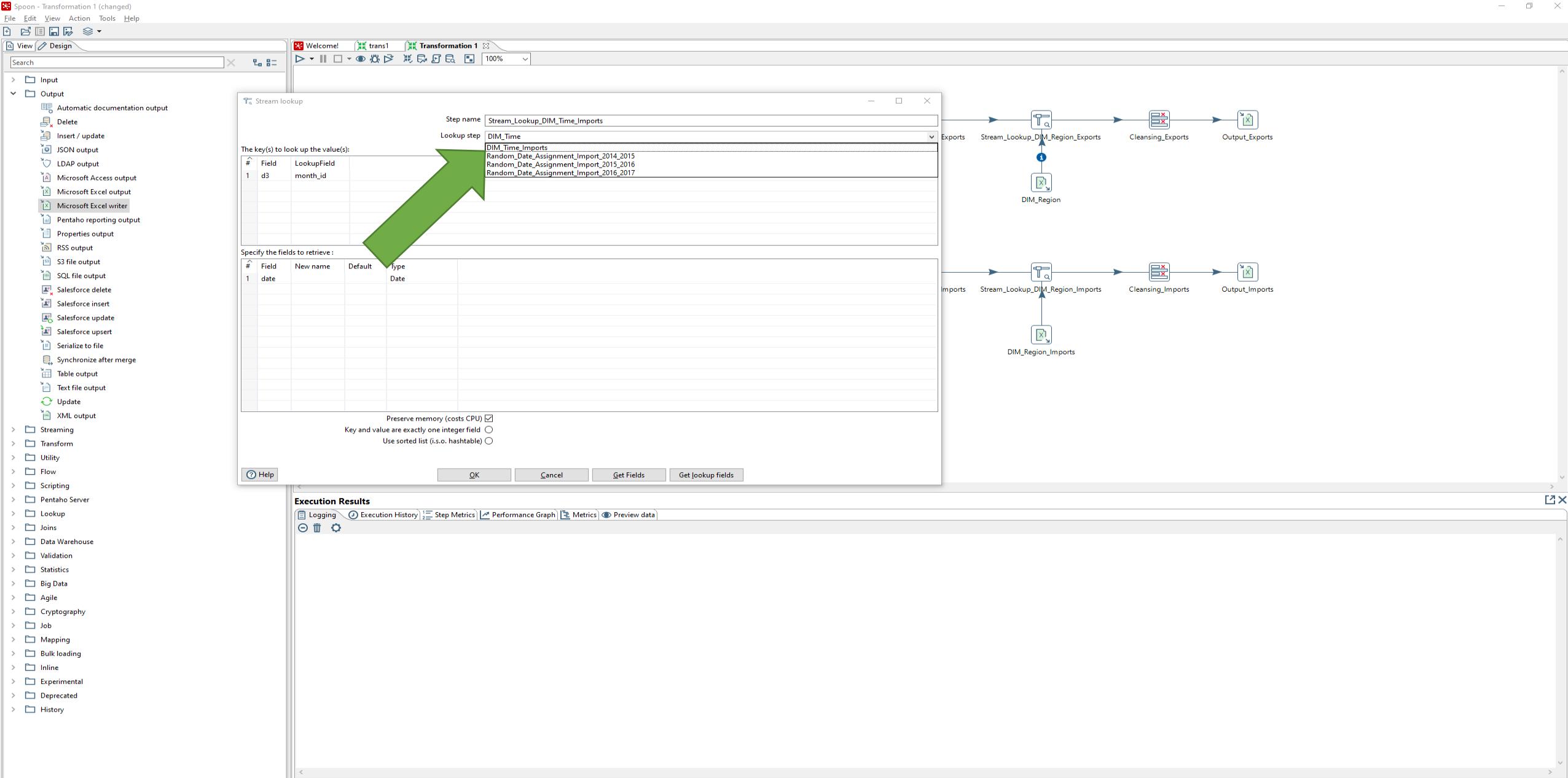




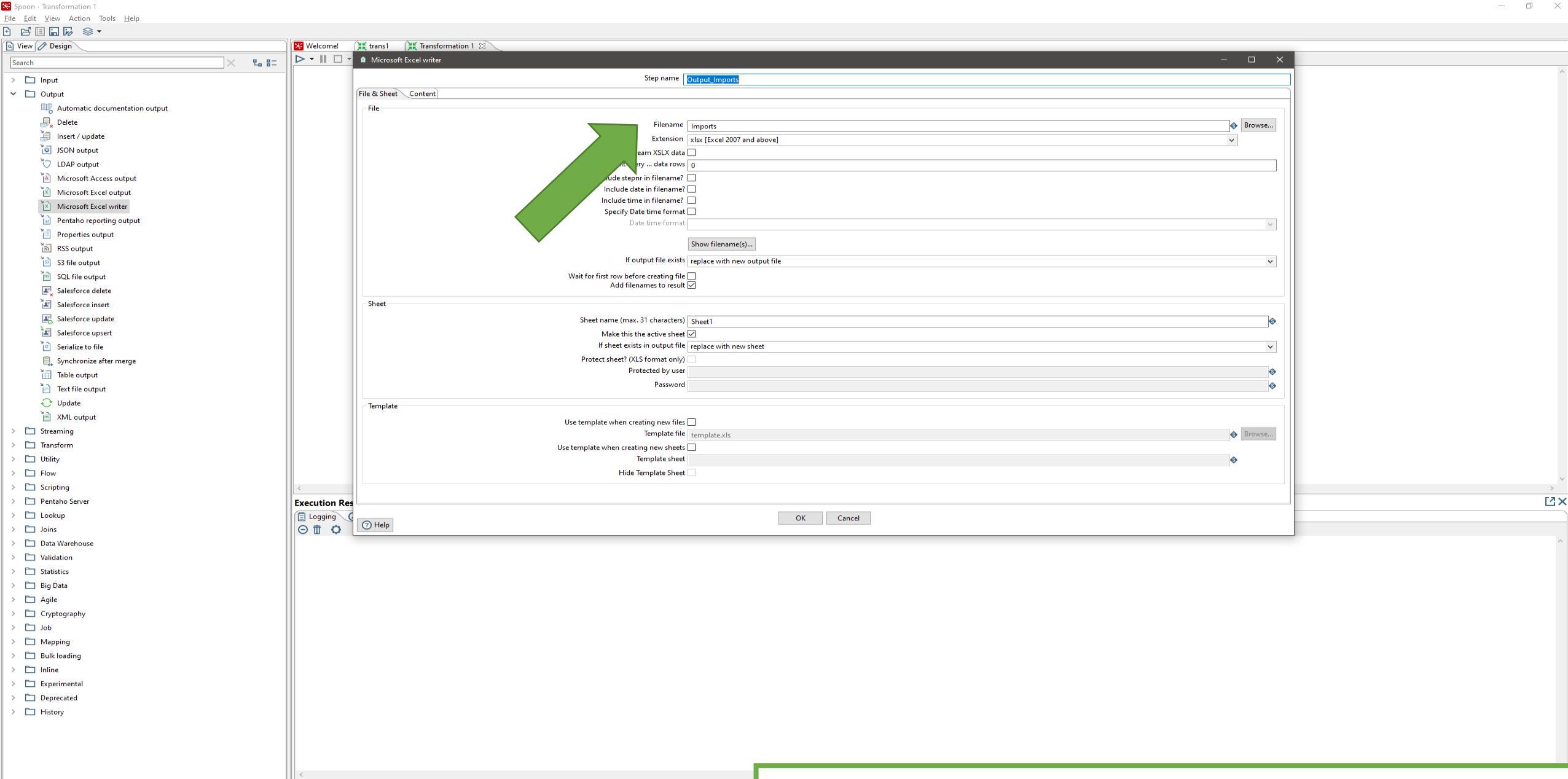
Again name the Input steps properly and add the Import CSV files
 Since Exports and Imports share the same columns, you can leave
 the fields
 Press OK
 Repeat that process for the other two Input steps



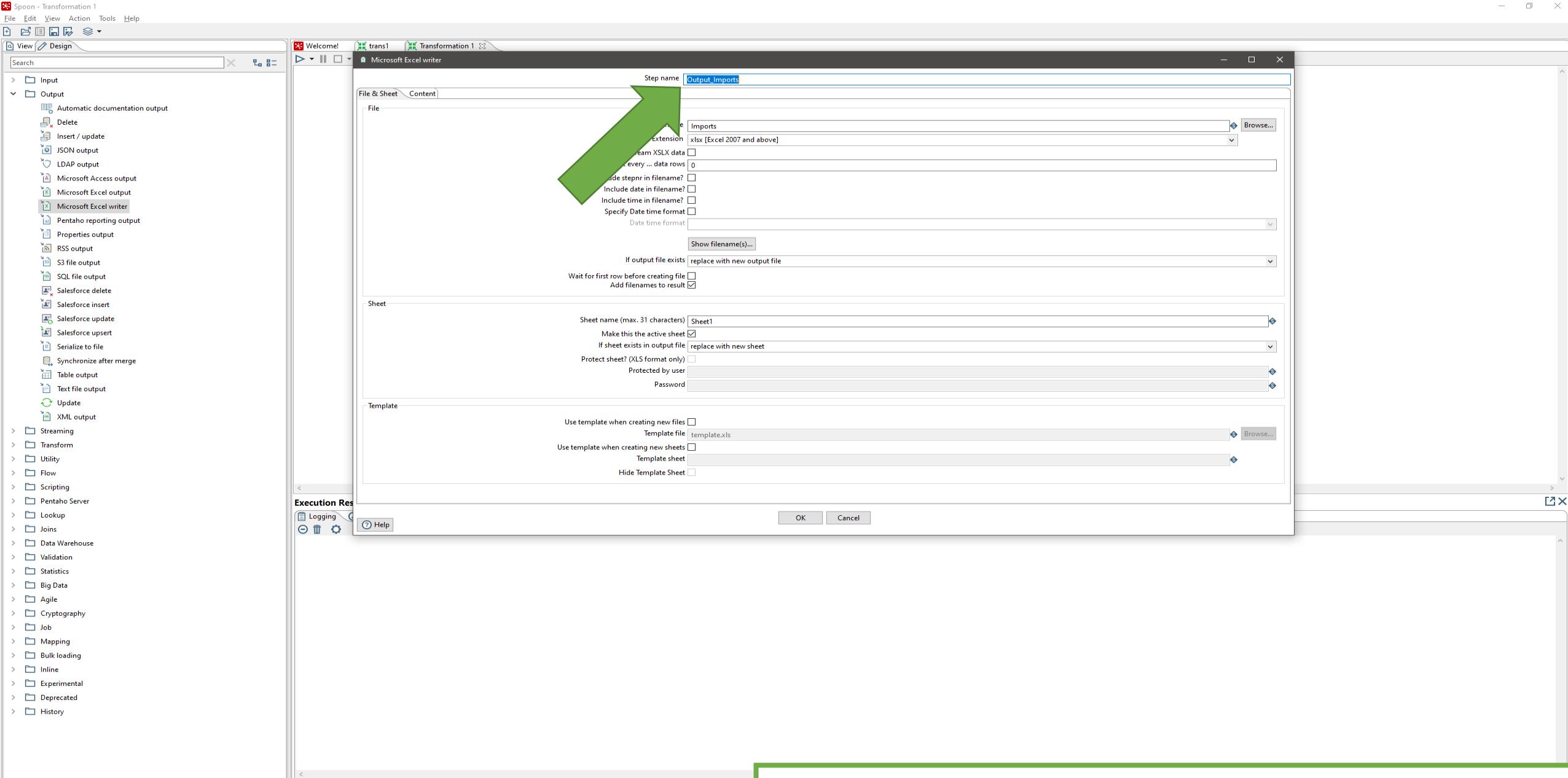
Next edit DIM_Time, DIM_Product and DIM_Region of the Imports process and set a new Step name (i.e. DIM_Time_Import)
Press OK



After editing the DIM files, edit the stream lookups by setting the Lookup step to the renamed DIM file



Next edit the Output step by naming it properly and selecting a fitting filename for the imports
Select content



Next edit the Output step by naming it properly and selecting a fitting filename for the imports
Select content

Spoon - Transformation 1

File Edit View Action Tools Help

View Design

Search

> Input

> Output

- Automatic documentation output
- Delete
- Insert / update
- JSON output
- LDAP output
- Microsoft Access output
- Microsoft Excel output
- Microsoft Excel writer
- Pentaho reporting output
- Properties output
- RSS output
- S3 file output
- SQL file output
- Salesforce delete
- Salesforce insert
- Salesforce update
- Salesforce upsert
- Serialize to file
- Synchronize after merge
- Table output
- Text file output
- Update
- XML output

> Streaming

> Transform

> Utility

> Flow

> Scripting

> Pentaho Server

> Lookup

> Joins

> Data Warehouse

> Validation

> Statistics

> Big Data

> Agile

> Cryptography

> Job

> Mapping

> Bulk loading

> Inline

> Experimental

> Deprecated

> History

Welcome! trans1 Transformation 1

Step name: Output_imports

File & Sheet Content

Content options

- Start cell: A1
- Writing rows: overwrite existing cells
- Write Header
- Write Footer
- Auto size columns
- Force formula recalculation
- Leave values of existing cells unchanged

When writing to existing sheet

- Start writing at end of sheet (appending lines)
- Offset by ... rows: 0
- Begin by writing ... empty lines: 0
- Omit header

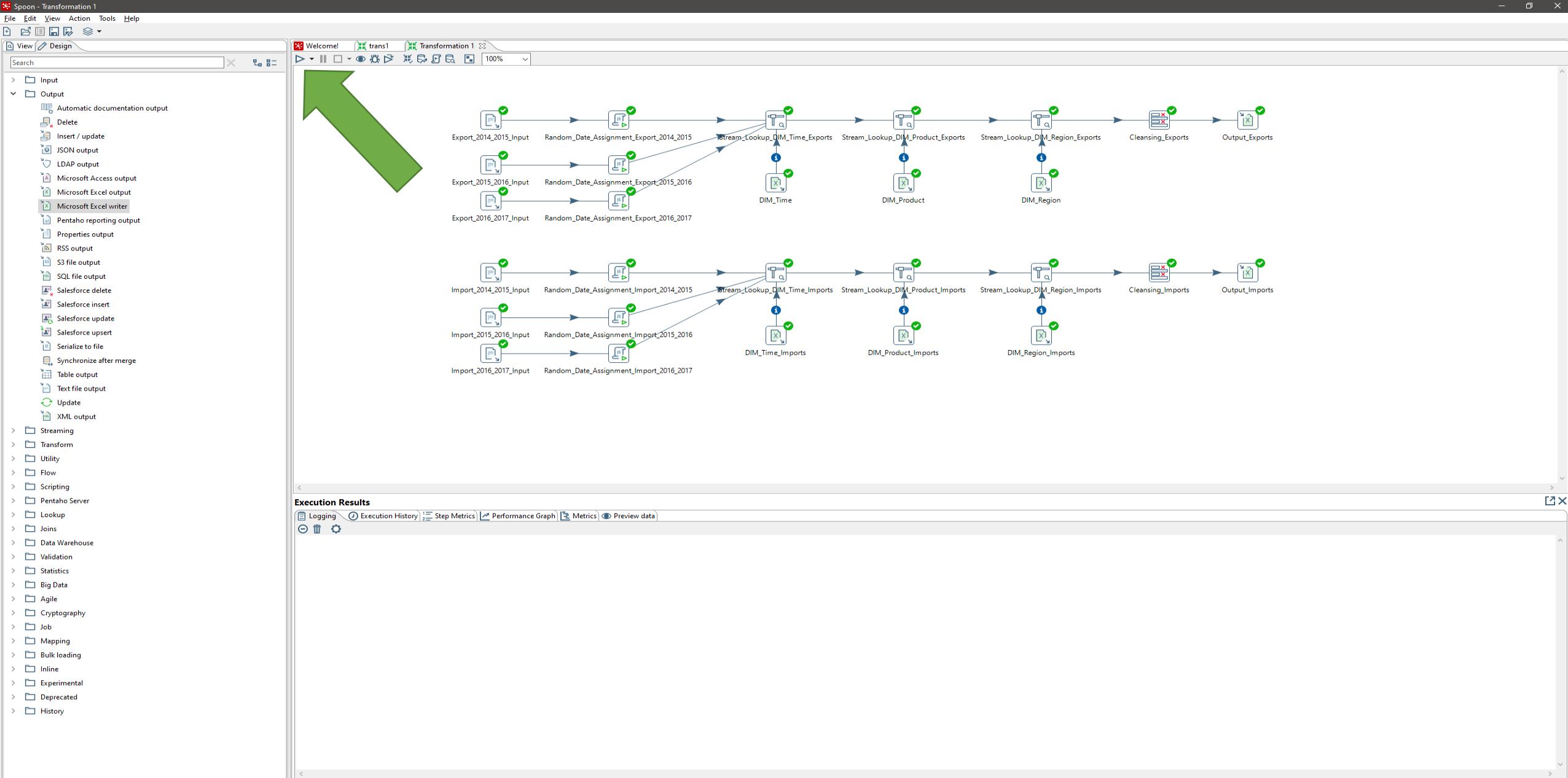
Fields

#	Name	Type	Format	Style from cell	Field title	Header/Footer style from cell	field contains formula	Hyperlink	Cell comment (XLSX)	Cell c
1	pc_code	String			pc_code		N			
2	pc_description	String			pc_description		N			
3	pc_category	String			pc_category		N			
4	region_code	Integer			region_code		N			
5	region_country	String			region_country		N			
6	region_continent	String			region_continent		N			
7	region_subcontinent	String			region_subcontinent		N			
8	unit	String			unit		N			
9	quantity	String			quantity		N			
10	value	Number			value		N			
11	date	Date	mmmm-yy		date		N			

Get Fields Minimal width OK Cancel

A large green arrow points from the text "Edit the content according to the picture above" to the "Content options" section of the dialog.

Edit the content according to the picture above
Press OK



The whole process should look like this now
Press Run to finally output both Exports and Imports