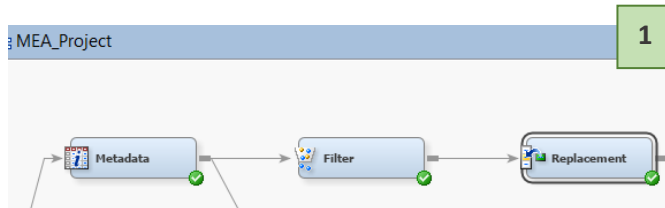
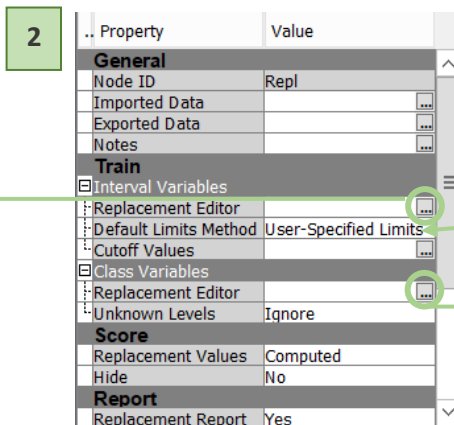


6 – Replacement (to replace variables' values)



Go to the Modify tab and drag and drop the Replacement node. Connect it to the previous node.



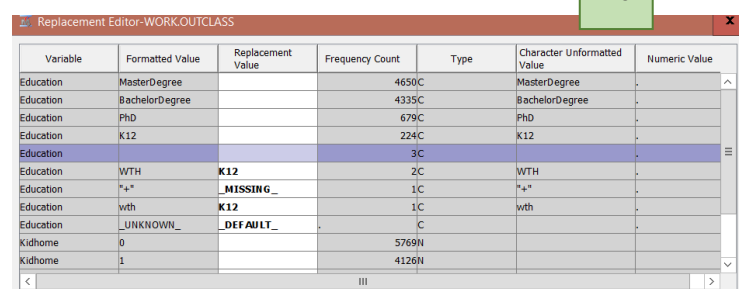
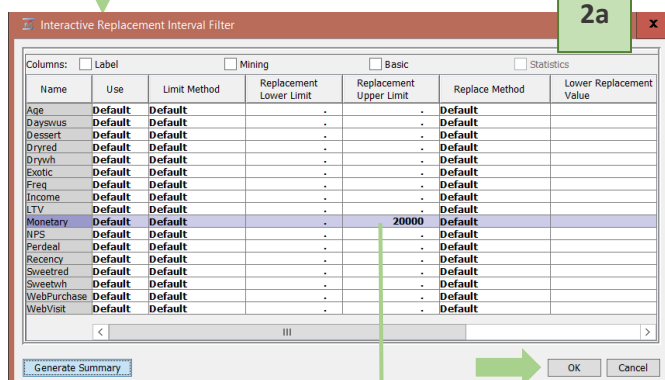
In the properties panel you have to make sure that, for the Interval Variables, the Default Limits Method that is defined is one of the following:

User Specified Method: if you pretend to apply replacement rules for interval variables;

None: if you pretend to do nothing.

If you choose to replace values for interval variables, you have to click on the three dots option to open the pop up window. This window allow you to define lower or/and upper limits for variables values

For class variables, click on the three dots to open the pop up window.



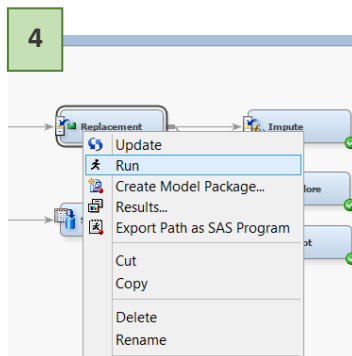
For instance, we observed that our dataset has few customers with extreme values for Monetary. Instead of cutting off the observations on the filter node we can define an upper limit that will replace the original values.

Replacement node

Ex: Customer A : Monetary=42000 → Customer A : Monetary=20000

You can replace variables values using one of the other existing categories or attributing the value `_MISSING_`, transforming those “invalid” values in missing values.

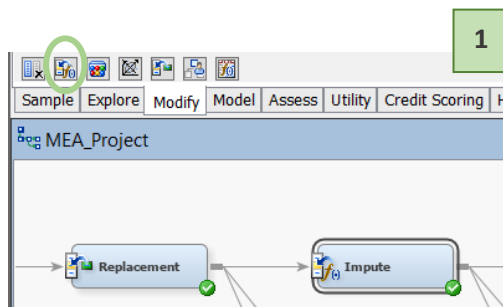
The `_UNKNOWN_` value is defined for other possible values that might appear in the dataset, in case if the company gains a new customer with other value for Education, different from the existing ones, allowing the creation of a new class. This class can assume its own value by `_DEFAULT_` or can be defined a new one.



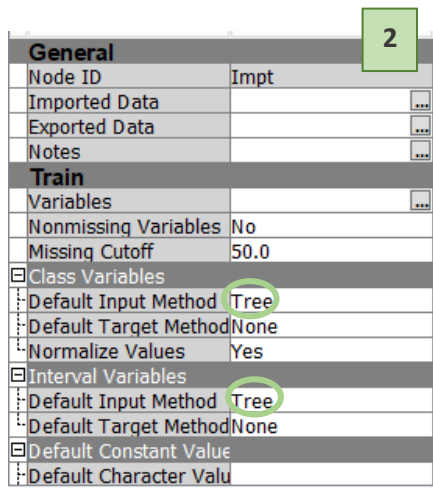
After all the modifications, you have to run the node to update your changes.

For variables that suffer value replacements, SAS creates a new variable with the prefix REP before the variable original name and rejects the original one. In our example, the new variable is REP_Education.

7- Imputation (to “treat” missing values)



Go to the Modify tab and drag and drop the Impute node. Connect it to the previous node.



In the properties panel you have to select the **default input method** the program will use to define values for observations that present missing values. If you have missing values for both class and interval variables, you need to define the input method for both of them.

Count: missing values are replaced with the modal value;

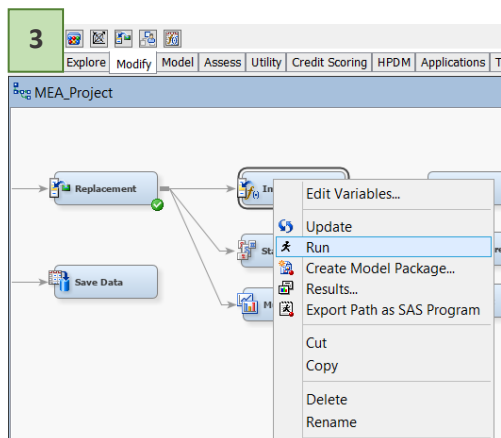
Default Constant: all replacements are equal to a fixed constant value;

Distribution: replacement values are calculated based on the random percentiles of the variable's distribution. In this case, the assignment of values is based on the probability distribution of the non-missing observations;

Tree: replacement values are estimated by analyzing each input as a target, and the remaining input and rejected variables are used as predictors;

Tree Surrogate: the same as Tree Imputation, except with the addition of surrogate splitting rules. A surrogate rule is a back-up to the main splitting rule;

None: replacement values are not imputed, and are left as missing.

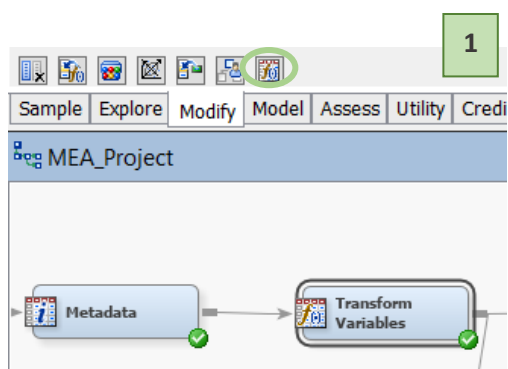


After defining properties, run the node.

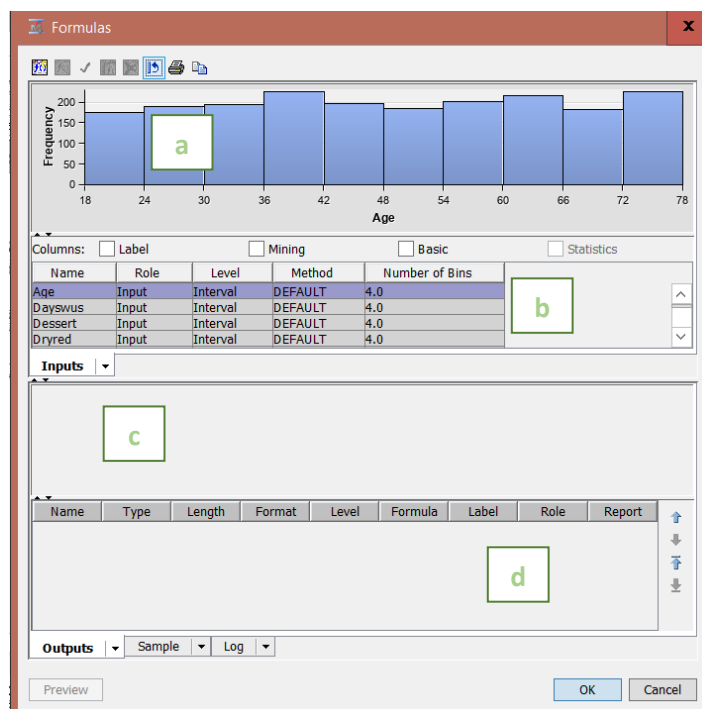
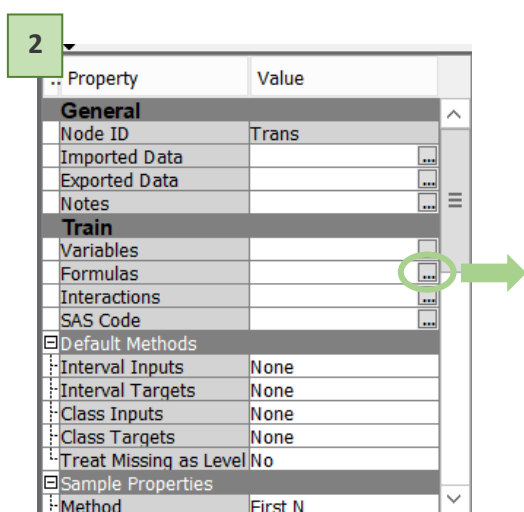
In the same way as in the step before, new variables with the prefix IMP are created for the variables that suffer imputation.

Imputation Summary		
Variable Name	Impute Method	Imputed Variable
Drywh	TREE	IMP Drywh
Freq	TREE	IMP Freq
Income	TREE	IMP Income
REP Education	TREE	IMP REP Education
WebVisit	TREE	IMP WebVisit

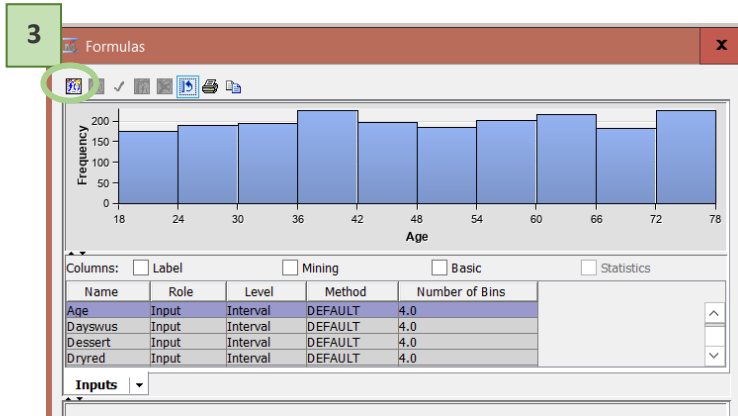
8 – Transform Variables (to transform and create new variables)



Go to the Modify tab and drag and drop the Transform Variables node. Connect it to the previous node.



- Previous variables histogram;
- List of previous variables information;
- New variables information area;
- New variables histogram area.



4

Property	Value
Name	Freq_corr
Type	Numeric
Length	8
Format	Interval
Level	Interval
Label	
Role	Input

Formula:
TRANS_0 =

Build... OK Cancel

5

Expression Text:
SQRT(IMP_Freq ** 2)

Functions:
SQRT(argument)

Variables List:

OK Cancel

Insert formula that describes the new variable

List of existing variables

List of available functions

6

Property	Value
Name	Freq_corr
Type	Numeric
Length	8
Format	Interval
Level	Interval
Label	
Role	Input

Formula:
Freq_corr =
SQRT(IMP_Freq ** 2)

Build... OK Cancel

7

Formulas

Frequency

Age

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Method	Number of Bins
Age	Input	Interval	DEFAULT	4.0
Dayswus	Input	Interval	DEFAULT	4.0
Dessert	Input	Interval	DEFAULT	4.0
Dryred	Input	Interval	DEFAULT	4.0

Inputs

Frequency

Freq_corr

New variable histogram

Name	Type	Length	Format	Level	Formula	Label	Role	Report
Freq_corr	Numeric	8		Interval	SQRT(IM...		Input	No

New variable information

Outputs:

To see new variable histogram

9 – OPTIONAL: SAS Code (uses programming language to create new variable)

1

Sample Explore Modify Model Assess Utility Credit Scoring

MEA_Project

Transform Variables SAS Code

2

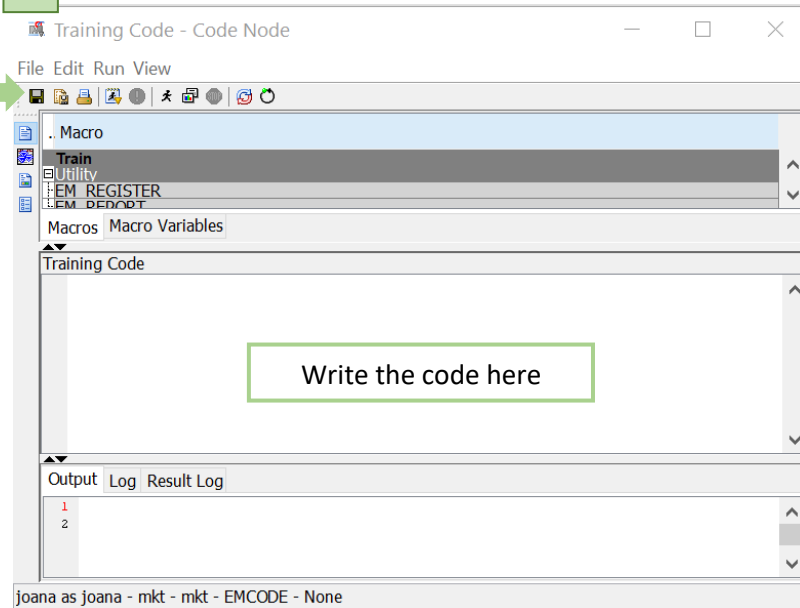
General

Node ID	EMCODE
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Code Editor	
Tool Type	Utility
Data Needed	No
Rerun	No
Use Priors	Yes
Score	
Advisor Type	Basic
Publish Code	Publish
Code Format	DATA step

Click on code editor ...

3

After writing the
code, save it



HOW TO WRITE CODE?

4

Data &EM_EXPORT_TRAIN.;
Set &EM_IMPORT_DATA.;

Start your code always with this

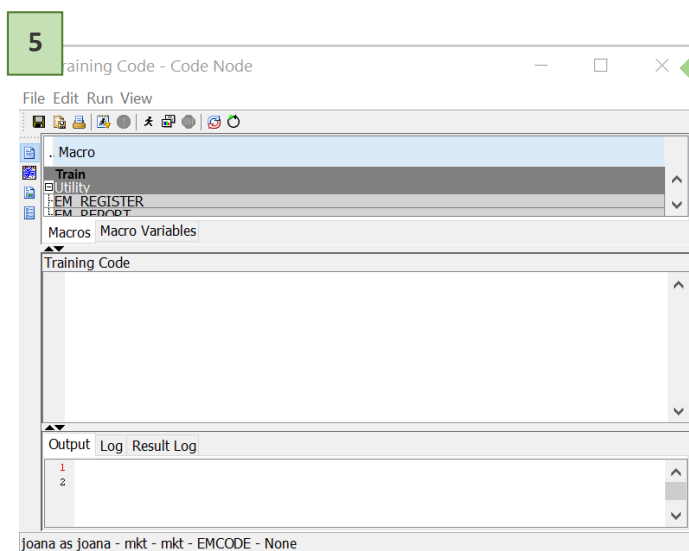
Then, start creating the
variables – SAS language

```
if (Kidhome+Teenhome>0) then TotalChildrenBinary=1;
else TotalChildrenBinary=0;
```

```
if (upcase(Education) in ("GRADUATION","MASTER","PHD")) then HigherEducation=1;
else HigherEducation=0;
```

```
if (upcase(Marital_Status) in ("MARRIED","TOGETHER")) then Partner=1;
else Partner=0;
```

Run; End with "Run;"



After saving the code,
close the window and
run the node