

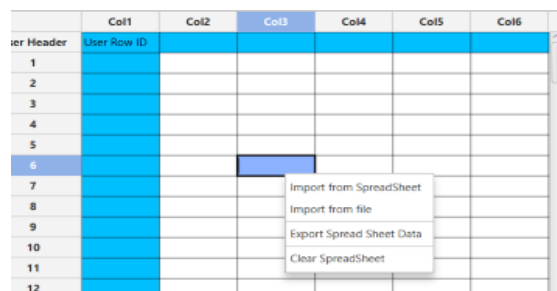


MA score Data (Regression)

The goal of this study is to train a model in order to predict the test score for school districts in Massachusetts. The dataset used in this case study is found in [R: Massachusetts Test Score Data \(vincentarelbundock.github.io\)](https://github.com/vincentarelbundock) and has 16 features and 220 labelled samples. The dataset contains data on test performance, school characteristics and student demographic backgrounds for school districts in Massachusetts. The data analysed here are the overall total score, which is the sum of the scores on the English, Math, and Science portions of the test.

Step 1: Import data from file

Right click on the input spreadsheet and choose the option "Import from file". Then navigate through your files to load the one with the MA score data.

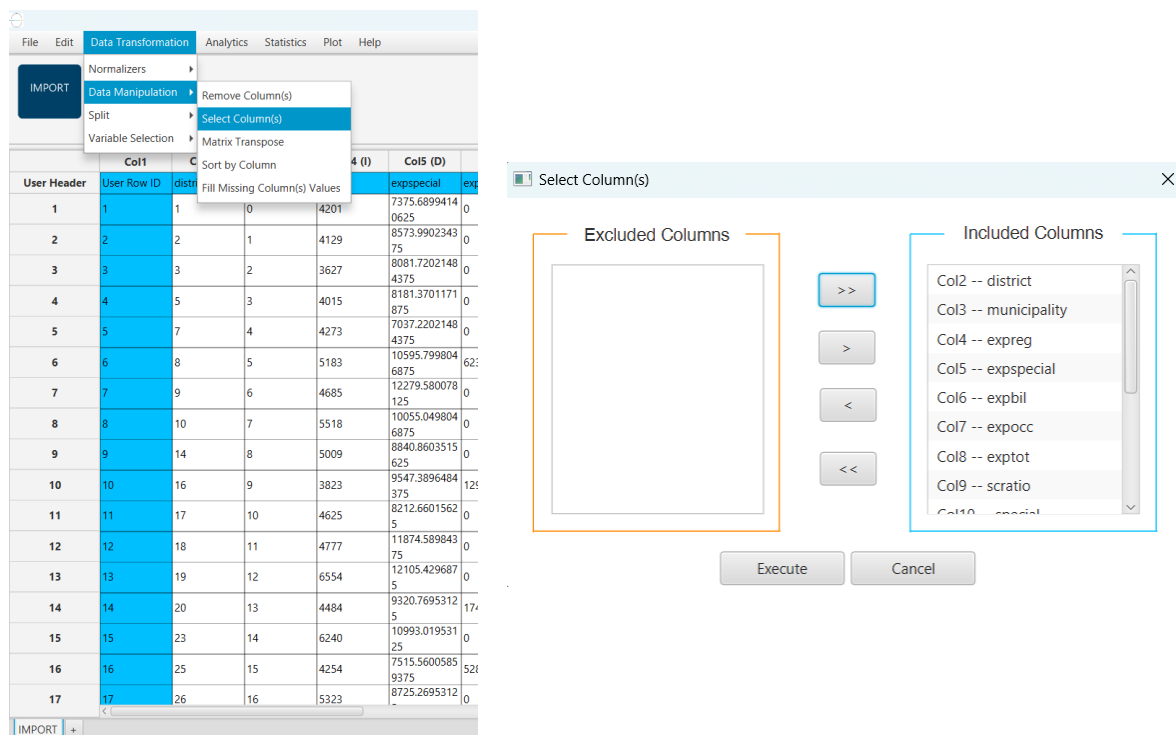


User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11
1	1	1	0	4201	7375.6899414	0	0	4646	16.600000381	14.600000381	11.800000381
2	2	2	1	4129	8573.9902343	0	0	4930	5.6999998092	17.399999618	2.5
3	3	3	2	3627	8081.7202148	0	0	4281	6.5137	530273	14.100000381
4	4	5	3	4015	8181.3701171	0	0	4826	8.600000381	21.100000381	12.100000381
5	5	7	4	4273	7037.2202148	0	0	4824	6.0999999046	16.799999237	17.390000381
6	6	8	5	5183	10595.799804	6235	0	6454	7.6999998092	17.200000762	26.790000381
7	7	9	6	4685	12279.580078	125	0	5537	32568	530277	26513
8	8	10	7	5518	10055.049804	0	0	6405	6.5137	939453	06054
9	9	14	8	5009	8840.8603515	625	0	5649	5.4000000953	11.300000190	3.299000381
10	10	16	9	3823	9547.3896484	12943	11519	4814	7.0999999046	20.399999618	11.190000381
11	11	17	10	4625	8212.6601562	5	0	5210	6.5137	11.100000381	10.690000381
12	12	18	11	4777	11874.589843	75	0	5615	4.69728	530272	26513
13	13	19	12	6554	12105.429687	5	0	7389	7.0999999046	13.199999809	20.700000381
14	14	20	13	4484	9320.7695312	5	1741	5323	6.5137	939453	67432
15	15	23	14	6240	10993.019531	25	0	7234	8.8000001907	14.899999618	18.700000381
16	16	25	15	4254	7515.5600585	9375	5281	1470	34863	530272	2.900000381
17	17	26	16	5323	8725.2695312	0	0	6065	9.399999618	20.0	69727

Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features. In our dataset there are no such columns. Therefore, we will include all columns in the training. We follow these steps to execute this:

- On the menu click on "Data Transformation" → "Data Manipulation" → "Select Column(s)"
- Select all columns.



The data will appear in the output spreadsheet.

Step 3: Fill missing values

Create a new tab by pressing the "+" button on the bottom of the page with the name "FILL_MISSING_VALUES" which we will use for filling the missing values of the dataset

Import data into the input spreadsheet of the "FILL_MISSING_VALUES" tab from the output of the "IMPORT" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (I)
1	1	1.0000000	0E-7	4201.0000000	7375.6899414	0E-7	0E-7	4646.0000000
2	2	2.0000000	1.0000000	4129.0000000	8573.9902344	0E-7	0E-7	4930.0000000
3	3	3.0000000	2.0000000	3627.0000000	8081.7202148	0E-7	0E-7	4281.0000000
4	4	5.0000000	3.0000000	4015.0000000	8181.3701172	0E-7	0E-7	4826.0000000
5	5	7.0000000	4.0000000	4273.0000000	7037.2202148	0E-7	0E-7	4824.0000000
6	6	8.0000000	5.0000000	5183.0000000	10595.7998047	6235.0000000	0E-7	6454.0000000
7	7	9.0000000	6.0000000	4685.0000000	12279.5800781	0E-7	0E-7	5537.0000000
8	8	10.0000000	7.0000000	5518.0000000	10055.0498047	0E-7	0E-7	6405.0000000
9	9	14.0000000	8.0000000	5009.0000000	8840.8603516	0E-7	0E-7	5649.0000000
10	10	16.0000000	9.0000000	3823.0000000	9547.38964840	12943.0000000	11519.0000000	4814.0000000
11	11	17.0000000	10.0000000	4625.0000000	8212.6601563	0E-7	0E-7	5210.0000000
12	12	18.0000000	11.0000000	4777.0000000	11874.5898438	0E-7	0E-7	5615.0000000
13	13	19.0000000	12.0000000	6554.0000000	12105.4296875	0E-7	0E-7	7389.0000000
14	14	20.0000000	13.0000000	4484.0000000	9320.7695313	1741.0000000	0E-7	5323.0000000
15	15	23.0000000	14.0000000	6240.0000000	10993.0195313	0E-7	0E-7	7234.0000000
16	16	25.0000000	15.0000000	4254.0000000	7515.5600586	5281.0000000	1470.0000000	5048.0000000
17	17	26.0000000	16.0000000	5323.0000000	8725.2695313	0E-7	0E-7	6085.0000000
18	18	27.0000000	17.0000000	3079.0000000	7755.4399414	0E-7	0E-7	3930.0000000

Fill the missing values in the dataset by browsing: “Data Transformation” → “Data Manipulation” → “Fill Missing Column Values”. Then include all columns and select the “Mean” as the “Numerical Method” to fill the missing values.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (D)	Col6 (I)
1	1	1.0000000	0E-7	4201.0000000	7375.6899414	0E-7
2	2	2.0000000	1.0000000	4129.0000000	8573.9902344	0E-7
3	3	3.0000000	2.0000000	3627.0000000	8081.7202148	0E-7
4	4	5.0000000	3.0000000	4015.0000000	8181.3701172	0E-7
5	5	7.0000000	4.0000000	4273.0000000	7037.2202148	0E-7
6	6	8.0000000	5.0000000	5183.0000000	10595.7998047	6235.0000000
7	7	9.0000000	6.0000000	4685.0000000	12279.5800781	0E-7
8	8	10.0000000	7.0000000	5518.0000000	10055.0498047	0E-7
9	9	14.0000000	8.0000000	5009.0000000	8840.8603516	0E-7
10	10	16.0000000	9.0000000	3823.0000000	9547.38964840	12943.0000000
11	11	17.0000000	10.0000000	4625.0000000	8212.6601563	0E-7
12	12	18.0000000	11.0000000	4777.0000000	11874.5898438	0E-7
13	13	19.0000000	12.0000000	6554.0000000	12105.4296875	0E-7
14	14	20.0000000	13.0000000	4484.0000000	9320.7695313	1741.0000000
15	15	23.0000000	14.0000000	6240.0000000	10993.0195313	0E-7
16	16	25.0000000	15.0000000	4254.0000000	7515.5600586	5281.0000000
17	17	26.0000000	16.0000000	5323.0000000	8725.2695313	0E-7
18	18	27.0000000	17.0000000	3079.0000000	7755.4399414	0E-7

Fill Missing Column(s) Values

Excluded Columns

Included Columns

- Col2 -- district
- Col3 -- municipality
- Col4 -- expreg
- Col5 -- expspecial
- Col6 -- expbil
- Col7 -- expocc
- Col8 -- expatot
- Col9 -- scratio
- Col10 -- special

Numerical Method: Mean

Span: Integer (0,+∞), Default: -

Categorical Method: None

Execute Cancel

The data will appear in the output spreadsheet.

User Header	User Row ID	district	municipality	expreg	expspecial	expbil	expoc	expit
1	1	1.000000	0E-7	4201.000000	7375.6899414	0E-7	0E-7	4646.0000000
2	2	2.000000	1.000000	4129.000000	8573.9902344	0E-7	0E-7	4930.0000000
3	3	3.000000	2.000000	3627.000000	8081.7202148	0E-7	0E-7	4281.0000000
4	4	5.000000	3.000000	4015.000000	8181.3701172	0E-7	0E-7	4826.0000000
5	5	7.000000	4.000000	4273.000000	7037.2202148	0E-7	0E-7	4824.0000000
6	6	8.000000	5.000000	5183.000000	10595.799804	6235.0000000	0E-7	6454.0000000
7	7	9.000000	6.000000	4685.000000	12279.580078	0E-7	0E-7	5537.0000000
8	8	10.000000	7.000000	5518.000000	10055.049804	0E-7	0E-7	6405.0000000
9	9	14.000000	8.000000	5009.000000	8840.8603516	0E-7	0E-7	5649.0000000
10	10	16.000000	9.000000	3823.000000	9547.389648	12943.0000000	11519.0000000	4814.0000000
11	11	17.000000	10.000000	4625.000000	8212.6601563	0E-7	0E-7	5210.0000000
12	12	18.000000	11.000000	4777.000000	11874.589843	0E-7	0E-7	5615.0000000
13	13	19.000000	12.000000	6554.000000	12105.429687	0E-7	0E-7	7389.0000000
14	14	20.000000	13.000000	4484.000000	9320.7695313	1741.0000000	0E-7	5323.0000000
15	15	23.000000	14.000000	6240.000000	10993.019531	0E-7	0E-7	7234.0000000
16	16	25.000000	15.000000	4254.000000	7515.5600586	5281.0000000	1470.0000000	5048.0000000
17	17	26.000000	16.000000	5323.000000	8725.2695313	0E-7	0E-7	6065.0000000
18	18	27.000000	17.000000	3079.000000	7755.4399414	0E-7	0E-7	3930.0000000
19	19	28.000000	18.000000	4836.000000	10220.429687	0E-7	0E-7	6121.0000000
20	20	30.000000	19.000000	4205.000000	8288.4501953	0E-7	0E-7	4961.0000000
21	21	31.000000	20.000000	4271.000000	8314.6396484	0E-7	0E-7	4901.0000000

Step 4: Split data

Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_TEST_SPLIT" which we will use for splitting to create the train and test set.

Import data into the input spreadsheet of the "TRAIN_TEST_SPLIT" tab from the output of the "FILL_MISSING_VALUES" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	User Row ID	district	municipality	expreg	expspecial	expbil	expoc	expit
1	1	1.000000	0E-7	4201.000000	7375.6899414	0E-7	0E-7	4646.0000000
2	2	2.000000	1.000000	4129.000000	8573.9902344	0E-7	0E-7	4930.0000000
3	3	3.000000	2.000000	3627.000000	8081.7202148	0E-7	0E-7	4281.0000000
4	4	5.000000	3.000000	4015.000000	8181.3701172	0E-7	0E-7	4826.0000000
5	5	7.000000	4.000000	4273.000000	7037.2202148	0E-7	0E-7	4824.0000000
6	6	8.000000	5.000000	5183.000000	10595.799804	6235.0000000	0E-7	6454.0000000
7	7	9.000000	6.000000	4685.000000	12279.580078	0E-7	0E-7	5537.0000000
8	8	10.000000	7.000000	5518.000000	10055.049804	0E-7	0E-7	6405.0000000
9	9	14.000000	8.000000	5009.000000	8840.8603516	0E-7	0E-7	5649.0000000
10	10	16.000000	9.000000	3823.000000	9547.389648	12943.0000000	11519.0000000	4814.0000000
11	11	17.000000	10.000000	4625.000000	8212.6601563	0E-7	0E-7	5210.0000000
12	12	18.000000	11.000000	4777.000000	11874.589843	0E-7	0E-7	5615.0000000
13	13	19.000000	12.000000	6554.000000	12105.429687	0E-7	0E-7	7389.0000000
14	14	20.000000	13.000000	4484.000000	9320.7695313	1741.0000000	0E-7	5323.0000000
15	15	23.000000	14.000000	6240.000000	10993.019531	0E-7	0E-7	7234.0000000
16	16	25.000000	15.000000	4254.000000	7515.5600586	5281.0000000	1470.0000000	5048.0000000
17	17	26.000000	16.000000	5323.000000	8725.2695313	0E-7	0E-7	6065.0000000
18	18	27.000000	17.000000	3079.000000	7755.4399414	0E-7	0E-7	3930.0000000
19	19	28.000000	18.000000	4836.000000	10220.429687	0E-7	0E-7	6121.0000000
20	20	30.000000	19.000000	4205.000000	8288.4501953	0E-7	0E-7	4961.0000000
21	21	31.000000	20.000000	4271.000000	8314.6396484	0E-7	0E-7	4901.0000000

Split the dataset by browsing: "Data Transformation" → "Split" → "Random Partitioning". Then choose the "Training set percentage" and the column for the sampling as shown below:

User Header	User Row ID	District	municipality	exposmg	expospecial	exposbil	exposoc	exposit	scat
1		1.0000000	0E-7	4201.0000000	7375.6899414	0E-7	0E-7	4646.0000000	16.60
2		2.0000000	1.0000000	4129.0000000	8573.9902344	0E-7	0E-7	4930.0000000	5.699
3		3.0000000	2.0000000	3627.0000000	8081.7202148	0E-7	0E-7	4281.0000000	7.500
4		5.0000000	3.0000000	4015.0000000	8181.3701172	0E-7	0E-7	4826.0000000	8.600
5		7.0000000	4.0000000	4273.0000000	7037.2202148	0E-7	0E-7	4824.0000000	6.099
6		8.0000000	5.0000000	5183.0000000	10595.7998047	6235.0000000	0E-7	6454.0000000	7.699
7		9.0000000	6.0000000	4685.0000000	12279.5800781	0E-7	0E-7	5537.0000000	5.400
8		10.0000000	7.0000000	5518.0000000	10055.0498047	0E-7	0E-7	6405.0000000	7.099
9		14.0000000	8.0000000	5009.0000000	8840.8603516	0E-7	0E-7	5649.0000000	10.60
10		16.0000000	9.0000000	3823.0000000	9547.3896484	12943.0000000	11519.0000000	4814.0000000	6.699
11		17.0000000	10.0000000	4625.0000000	8212.6601563	0E-7	0E-7	5210.0000000	12.50
12		18.0000000	11.0000000	4777.0000000	11874.5898438	0E-7	0E-7	5615.0000000	7.599
13		19.0000000	12.0000000	6554.0000000	12105.4296875	0E-7	0E-7	7389.0000000	4.199
14		20.0000000	13.0000000	4484.0000000	9320.7695313	1741.0000000	0E-7	5323.0000000	8.800
15		23.0000000	14.0000000	6240.0000000	10993.0195313	0E-7	0E-7	7234.0000000	4.800
16		25.0000000	15.0000000	4254.0000000	7515.5600586	5281.0000000	1470.0000000	5048.0000000	9.399
17		26.0000000	16.0000000	5323.0000000	8725.2695313	0E-7	0E-7	6065.0000000	15.00

Random Partitioning

Training set percentage

75

☐ Usage of random generator seed

79827627322600

☒ Stratified sampling

Col16 -- salary

Execute

Cancel

The results will appear on the output spreadsheet.

User Header	User Row ID	District	municipality	exposmg	expospecial	exposbil	exposoc	exposit	scat
1		1.0000000	0E-7	4201.0000000	7375.6899414	0E-7	0E-7	4646.0000000	16.60
2		2.0000000	1.0000000	4129.0000000	8573.9902344	0E-7	0E-7	4930.0000000	5.699
3		3.0000000	2.0000000	3627.0000000	8081.7202148	0E-7	0E-7	4281.0000000	7.500
4		5.0000000	3.0000000	4015.0000000	8181.3701172	0E-7	0E-7	4826.0000000	8.600
5		7.0000000	4.0000000	4273.0000000	7037.2202148	0E-7	0E-7	4824.0000000	6.099
6		8.0000000	5.0000000	5183.0000000	10595.7998047	6235.0000000	0E-7	6454.0000000	7.699
7		9.0000000	6.0000000	4685.0000000	12279.5800781	0E-7	0E-7	5537.0000000	5.400
8		10.0000000	7.0000000	5518.0000000	10055.0498047	0E-7	0E-7	6405.0000000	7.099
9		14.0000000	8.0000000	5009.0000000	8840.8603516	0E-7	0E-7	5649.0000000	10.60
10		16.0000000	9.0000000	3823.0000000	9547.3896484	12943.0000000	11519.0000000	4814.0000000	6.699
11		17.0000000	10.0000000	4625.0000000	8212.6601563	0E-7	0E-7	5210.0000000	12.50
12		18.0000000	11.0000000	4777.0000000	11874.5898438	0E-7	0E-7	5615.0000000	7.599
13		19.0000000	12.0000000	6554.0000000	12105.4296875	0E-7	0E-7	7389.0000000	4.199
14		20.0000000	13.0000000	4484.0000000	9320.7695313	1741.0000000	0E-7	5323.0000000	8.800
15		23.0000000	14.0000000	6240.0000000	10993.0195313	0E-7	0E-7	7234.0000000	4.800
16		25.0000000	15.0000000	4254.0000000	7515.5600586	5281.0000000	1470.0000000	5048.0000000	9.399
17		26.0000000	16.0000000	5323.0000000	8725.2695313	0E-7	0E-7	6065.0000000	15.00
18		27.0000000	17.0000000	3079.0000000	7755.4399414	0E-7	0E-7	3930.0000000	13.80
19		28.0000000	18.0000000	4836.0000000	10220.4296875	0E-7	0E-7	6121.0000000	6.000
20		30.0000000	19.0000000	4205.0000000	8288.4501953	0E-7	0E-7	4961.0000000	9.800

Step 5: Normalize the training set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TRAIN_SET".

Import data into the input spreadsheet of the "NORMALIZE_TRAIN_SET" tab the train set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT : Training Set"

ma_score_data.xlsx

FileEditData TransformationAnalyticsStatisticsPlotHelp

IMPORT

FILE MISSING VALUES

TRAIN TEST SPLIT

NORMALIZE TRAIN SET

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (D)
User Header	User Row ID	district	municipality	expreg	expspecial	exbl	expocc	exp tot	scrio
1	1	1.0000000	0E-7	4201.0000000	7375.6899414	0E-7	0E-7	4646.0000000	16.6000
2	2	2.0000000	1.0000000	4129.0000000	8573.9902344	0E-7	0E-7	4930.0000000	5.69999
3	3	3.0000000	2.0000000	3627.0000000	8081.7202148	0E-7	0E-7	4281.0000000	7.50000
4	4	5.0000000	3.0000000	4015.0000000	8181.3701172	0E-7	0E-7	4826.0000000	8.60000
5	5	7.0000000	4.0000000	4273.0000000	7037.2202148	0E-7	0E-7	4824.0000000	6.09999
6	6	8.0000000	5.0000000	5183.0000000	10595.7998047	6235.0000000	0E-7	6454.0000000	7.69999
7	7	9.0000000	6.0000000	4685.0000000	12279.5800787	0E-7	0E-7	5537.0000000	5.40000
8	8	10.0000000	7.0000000	5518.0000000	10055.0498047	0E-7	0E-7	6405.0000000	7.09999
9	9	14.0000000	8.0000000	5009.0000000	8840.8603516	0E-7	0E-7	5649.0000000	10.6000
10	10	16.0000000	9.0000000	3823.0000000	9547.3896484	12943.0000000	11519.0000000	4814.0000000	6.69999
11	11	17.0000000	10.0000000	4625.0000000	8212.6601563	0E-7	0E-7	5210.0000000	12.5000
12	12	18.0000000	11.0000000	4777.0000000	11874.5898439	0E-7	0E-7	5615.0000000	7.59999
13	13	19.0000000	12.0000000	6554.0000000	12105.4296875	0E-7	0E-7	7389.0000000	4.19999
14	14	20.0000000	13.0000000	4484.0000000	9320.7695313	1741.0000000	0E-7	5323.0000000	8.80000
15	15	23.0000000	14.0000000	6240.0000000	10993.0195313	0E-7	0E-7	7234.0000000	4.80000
16	16	26.0000000	16.0000000	5323.0000000	8725.2695313	0E-7	0E-7	6065.0000000	15.0000
17	17	27.0000000	17.0000000	3079.0000000	7755.4399414	0E-7	0E-7	3930.0000000	13.8000
18	18	28.0000000	18.0000000	4836.0000000	10220.4296875	0E-7	0E-7	6121.0000000	6.00000
19	19	30.0000000	19.0000000	4205.0000000	8288.4501953	0E-7	0E-7	4961.0000000	9.80000
20	20	31.0000000	20.0000000	4271.0000000	8314.6396484	0E-7	0E-7	4901.0000000	10.1000

IMPORTFILE MISSING VALUESTRAIN TEST SPLITNORMALIZE TRAIN SET

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
User Header	User Row ID							
1	1							
2	2							
3	3							
4	4							
5	5							
6	6							
7	7							
8	8							
9	9							
10	10							
11	11							
12	12							
13	13							
14	14							
15	15							
16	16							
17	17							
18	18							
19	19							
20	20							
21	21							
22	22							
23	23							
24	24							

Normalize the data using Z-score by browsing: "Data Transformation" → "Normalizers" → "Z-Score". Then select all columns except "score4" and click "Execute".

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (D)
1	1.0000000	0E-7	4201.0000000	7375.6899414	0E-7
2	2.0000000	1.0000000	4129.0000000	8573.9902344	0E-7
3	3.0000000	2.0000000	3627.0000000	8081.7202148	0E-7
4	5.0000000	3.0000000	4015.0000000	8181.3701172	0E-7
5	7.0000000	4.0000000	4273.0000000	7037.2202148	0E-7
6	8.0000000	5.0000000	5183.0000000	10595.799804	0E-7
7	9.0000000	6.0000000	4685.0000000	12279.580078	0E-7
8	10.0000000	7.0000000	5518.0000000	10055.049804	0E-7
9	14.0000000	8.0000000	5009.0000000	8840.8603516	0E-7
10	16.0000000	9.0000000	3823.0000000	9547.3896484	0E-7
11	17.0000000	10.0000000	4625.0000000	8212.6601563	0E-7
12	18.0000000	11.0000000	4777.0000000	11874.589843	0E-7
13	19.0000000	12.0000000	6554.0000000	12105.429687	0E-7
14	20.0000000	13.0000000	4484.0000000	9320.7695313	0E-7
15	23.0000000	14.0000000	6240.0000000	10993.019531	0E-7
16	26.0000000	16.0000000	5323.0000000	8725.2695313	0E-7
17	27.0000000	17.0000000	3079.0000000	7755.4399414	0E-7
18	28.0000000	18.0000000	4836.0000000	10220.429687	0E-7
19	30.0000000	19.0000000	4205.0000000	8288.4501953	0E-7
20	31.0000000	20.0000000	4271.0000000	8314.6396484	0E-7

Excluded Columns

Col14 -- score4

Included Columns

Col8 -- exptot
Col9 -- scratio
Col10 -- special
Col11 -- lunch
Col12 -- stratio
Col13 -- income
Col15 -- score8
Col16 -- salary
Col17 -- english

Execute

Cancel

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (D)
1	1.000000	0E-7	4201.000000	7375.6899414	0E-7	0E-7	4646.000000	16.6000	
2	2.000000	1.000000	4129.000000	8573.9902344	0E-7	0E-7	4930.000000	5.69999	
3	3.000000	2.000000	3627.000000	8081.7202148	0E-7	0E-7	4281.000000	7.50000	
4	5.000000	3.000000	4015.000000	8181.3701172	0E-7	0E-7	4826.000000	8.60000	
5	7.000000	4.000000	4273.000000	7037.2202148	0E-7	0E-7	4824.000000	6.09999	
6	8.000000	5.000000	5183.000000	10595.799807	6235.000000	0E-7	6454.000000	7.69999	
7	9.000000	6.000000	4685.000000	12279.580078	1	0E-7	5537.000000	5.40000	
8	10.000000	7.000000	5518.000000	10055.049804	7	0E-7	6405.000000	7.09999	
9	14.000000	8.000000	5009.000000	8840.8603516	1	0E-7	5649.000000	10.6000	
10	16.000000	9.000000	3823.000000	9547.389648	0	11519.000000	4814.000000	6.69999	
11	17.000000	10.000000	4625.000000	8212.6601563	0E-7	0E-7	5210.000000	12.5000	
12	18.000000	11.000000	4777.000000	11874.589843	8	0E-7	5615.000000	7.59999	
13	19.000000	12.000000	6554.000000	12105.429687	5	0E-7	7389.000000	4.19999	
14	20.000000	13.000000	4484.000000	9320.7695313	1741.000000	0E-7	5323.000000	8.80000	
15	23.000000	14.000000	6240.000000	10993.019531	0	0E-7	7234.000000	4.80000	
16	26.000000	16.000000	5323.000000	8725.2695313	0E-7	0E-7	6065.000000	15.0000	
17	27.000000	17.000000	3079.000000	7755.4399414	0E-7	0E-7	3930.000000	13.8000	
18	28.000000	18.000000	4838.000000	10220.429687	5	0E-7	6121.000000	6.00000	
19	30.000000	19.000000	4205.000000	8288.4501953	0E-7	0E-7	4961.000000	9.80000	
20	31.000000	20.000000	4271.000000	8314.6396484	0E-7	0E-7	4901.000000	10.1000	

Step 6: Normalize the test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TEST_SET".

Import data into the input spreadsheet of the "NORMALIZE_TEST_SET" tab the test set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT: Test Set".

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (D)
1	25.000000	15.000000	4254.000000	7515.560586	5281.000000	1470.000000	5048.000000	9.399999	
2	51.000000	35.000000	4832.000000	11449.450195	0E-7	0E-7	5912.000000	8.107100	
3	61.000000	40.000000	4521.000000	15740.580078	7702.000000	4805.000000	5357.000000	8.800002	
4	63.000000	41.000000	4447.000000	5810.6396484	0E-7	0E-7	5290.000000	5.300002	
5	65.000000	43.000000	5548.000000	10008.910156	0E-7	0E-7	6174.000000	6.900001	
6	67.000000	44.000000	5608.000000	12543.660156	0E-7	0E-7	6595.000000	7.300002	
7	74.000000	49.000000	3687.000000	9624.1503906	0E-7	0E-7	4677.000000	7.800002	
8	77.000000	50.000000	3979.000000	6555.8798828	0E-7	0E-7	4552.000000	8.500000	
9	87.000000	55.000000	4122.000000	7546.5400391	0E-7	0E-7	4801.000000	9.399996	
10	88.000000	58.000000	3858.000000	7472.0297852	0E-7	0E-7	4515.000000	8.300002	
11	96.000000	64.000000	4682.000000	6855.020195	0E-7	0E-7	5133.000000	12.000000	
12	101.000000	68.000000	3693.000000	7066.2998047	0E-7	1054.000000	4127.000000	5.199998	
13	103.000000	70.000000	3867.000000	6279.479805	0E-7	0E-7	4382.000000	7.800002	
14	112.000000	75.000000	3400.000000	6819.9301758	0E-7	0E-7	4310.000000	6.800002	
15	114.000000	76.000000	3679.000000	7753.2900391	0E-7	0E-7	4689.000000	4.699998	
16	131.000000	83.000000	5004.000000	9307.5595703	0E-7	0E-7	5772.000000	10.500000	
17	135.000000	85.000000	2905.000000	8572.2402344	0E-7	0E-7	3465.000000	13.399996	
18	137.000000	87.000000	6049.000000	8155.1801758	5322.000000	5457.000000	6595.000000	8.000000	
19	139.000000	89.000000	4328.000000	7801.4301758	0E-7	0E-7	4880.000000	8.107100	
20	158.000000	104.000000	5152.000000	10494.799804	0E-7	0E-7	5812.000000	8.600004	
21	159.000000	105.000000	4961.000000	8982.8896484	0E-7	0E-7	5781.000000	7.800002	

Normalize the test set using the existing normalizer of the training set by browsing: "Analytics" → "Existing Model Utilization" → "Model (from Tab:) NORMALIZE_TRAIN_SET".

User Header	User Row ID	district	municipality	expreg	expspecial	expbil	expocc	expotot	scratio
1	25.0000000	15.0000000	4254.0000000	7515.5600586	5281.0000000	1470.0000000	0.0726519	0.1363825	
2	51.0000000	35.0000000	4832.0000000	11449.4501953	0E-7	0E-7	-0.1538401	-0.3970386	
3	61.0000000	40.0000000	4521.0000000	15740.5800781	7702.0000000	4805.0000000	0.1764840	1.3465590	
4	63.0000000	41.0000000	4447.0000000	8810.6396484	0E-7	0E-7	-0.1538401	-0.3970386	
5	65.0000000	43.0000000	5548.0000000	10008.9101563	0E-7	0E-7	-0.1538401	-0.3970386	
6	67.0000000	44.0000000	5608.0000000	12543.6601563	0E-7	0E-7	-0.1538401	-0.3970386	
7	74.0000000	49.0000000	3687.0000000	9624.1503906	0E-7	0E-7	-0.1538401	-0.3970386	
8	77.0000000	50.0000000	3979.0000000	6555.8798828	0E-7	0E-7	-0.1538401	-0.3970386	
9	87.0000000	55.0000000	4122.0000000	7546.5400391	0E-7	0E-7	-0.1538401	-0.3970386	
10	88.0000000	58.0000000	3858.0000000	7472.0297852	0E-7	0E-7	-0.1538401	-0.3970386	
11	96.0000000	64.0000000	4682.0000000	6855.0200195	0E-7	0E-7	-0.1538401	-0.3970386	
12	101.0000000	68.0000000	3693.0000000	7066.2998047	0E-7	1054.0000000	-0.1538401	-0.3970386	
13	103.0000000	70.0000000	3867.0000000	6279.4799805	0E-7	0E-7	-0.1538401	-0.3970386	
14	112.0000000	75.0000000	3400.0000000	6819.9301758	0E-7	0E-7	-0.1538401	-0.3970386	
15	114.0000000	76.0000000	3679.0000000	7753.2900391	0E-7	0E-7	-0.1538401	-0.3970386	
16	131.0000000	83.0000000	5004.0000000	9307.5595703	0E-7	0E-7	-0.1538401	-0.3970386	
17	135.0000000	85.0000000	2905.0000000	6572.2402344	0E-7	0E-7	-0.1538401	-0.3970386	
18	137.0000000	87.0000000	6049.0000000	8155.1801758	5322.0000000	5457.0000000	-0.1538401	-0.3970386	
19	139.0000000	89.0000000	4328.0000000	7801.4301758	0E-7	0E-7	-0.1538401	-0.3970386	
20	158.0000000	104.0000000	5152.0000000	10494.7998047	0E-7	0E-7	-0.1538401	-0.3970386	
21	159.0000000	105.0000000	4961.0000000	8982.8896484	0E-7	0E-7	-0.1538401	-0.3970386	

Model: (from Tab:)NORMALIZE_TRA...

Type: Z Score Normalizer Model

Description:

Model Input:

- Header -> Datatype
- district -> Double
- municipality -> Double
- expreg -> Double
- expspecial -> Double
- expbil -> Double
- expocc -> Double
- expotot -> Double
- scratio -> Double

☒ Transfer Column(s) to Output

Excluded Columns:

Included Columns:

- Col2 -- district
- Col3 -- municipality
- Col4 -- expreg
- Col5 -- expspecial
- Col6 -- expbil
- Col7 -- expocc
- Col8 -- expotot
- Col9 -- scratio
- Col10 -- expspecial

Execute Cancel

The results will appear on the output spreadsheet.

User Header	User Row ID	district	municipality	expreg	expspecial	expbil	expocc	expotot	scratio
1	25.0000000	15.0000000	4254.0000000	7515.5600586	5281.0000000	1470.0000000	0.0726519	0.1363825	
2	51.0000000	35.0000000	4832.0000000	11449.4501953	0E-7	0E-7	-0.1538401	-0.3970386	
3	61.0000000	40.0000000	4521.0000000	15740.5800781	7702.0000000	4805.0000000	0.1764840	1.3465590	
4	63.0000000	41.0000000	4447.0000000	8810.6396484	0E-7	0E-7	-0.1538401	-0.3970386	
5	65.0000000	43.0000000	5548.0000000	10008.9101563	0E-7	0E-7	-0.1538401	-0.3970386	
6	67.0000000	44.0000000	5608.0000000	12543.6601563	0E-7	0E-7	-0.1538401	-0.3970386	
7	74.0000000	49.0000000	3687.0000000	9624.1503906	0E-7	0E-7	-0.1538401	-0.3970386	
8	77.0000000	50.0000000	3979.0000000	6555.8798828	0E-7	0E-7	-0.1538401	-0.3970386	
9	87.0000000	55.0000000	4122.0000000	7546.5400391	0E-7	0E-7	-0.1538401	-0.3970386	
10	88.0000000	58.0000000	3858.0000000	7472.0297852	0E-7	0E-7	-0.1538401	-0.3970386	
11	96.0000000	64.0000000	4682.0000000	6855.0200195	0E-7	0E-7	-0.1538401	-0.3970386	
12	101.0000000	68.0000000	3693.0000000	7066.2998047	0E-7	1054.0000000	-0.1538401	-0.3970386	
13	103.0000000	70.0000000	3867.0000000	6279.4799805	0E-7	0E-7	-0.1538401	-0.3970386	
14	112.0000000	75.0000000	3400.0000000	6819.9301758	0E-7	0E-7	-0.1538401	-0.3970386	
15	114.0000000	76.0000000	3679.0000000	7753.2900391	0E-7	0E-7	-0.1538401	-0.3970386	
16	131.0000000	83.0000000	5004.0000000	9307.5595703	0E-7	0E-7	-0.1538401	-0.3970386	
17	135.0000000	85.0000000	2905.0000000	6572.2402344	0E-7	0E-7	-0.1538401	-0.3970386	
18	137.0000000	87.0000000	6049.0000000	8155.1801758	5322.0000000	5457.0000000	-0.1538401	-0.3970386	
19	139.0000000	89.0000000	4328.0000000	7801.4301758	0E-7	0E-7	-0.1538401	-0.3970386	
20	158.0000000	104.0000000	5152.0000000	10494.7998047	0E-7	0E-7	-0.1538401	-0.3970386	
21	159.0000000	105.0000000	4961.0000000	8982.8896484	0E-7	0E-7	-0.1538401	-0.3970386	

Step 7: Feature selection

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION_REGRESSION".

Import data into the input spreadsheet of the "FEATURE_SELECTION_REGRESSION" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)
User Row ID	district	municipality	expreg	expspecial	expbil	
1	-1.5453748	-1.5988798	-0.4917361	-0.7569975	-0.1538401	
2	-1.5361207	-1.5839733	-0.5741035	-0.0894423	-0.1538401	
3	-1.5268666	-1.5690668	-1.1483872	-0.3636786	-0.1538401	
4	-1.5083585	-1.5541603	-0.7045185	-0.3081651	-0.1538401	
5	-1.4898503	-1.5392538	-0.4093687	-0.9455540	-0.1538401	
6	-1.4805962	-1.5243473	0.6316636	1.0368776	0.1135672	
7	-1.4713422	-1.5094408	0.0619558	1.9748864	-0.1538401	
8	-1.4620881	-1.4945343	1.0149008	0.7356338	-0.1538401	
9	-1.4250718	-1.4796278	0.4326091	0.0592270	-0.1538401	
10	-1.4065636	-1.4647213	-0.9241649	0.4528239	0.4012606	
11	-1.3973096	-1.4498148	-0.0066837	-0.2907339	-0.1538401	
12	-1.3880555	-1.4349083	0.1672030	1.7492724	-0.1538401	
13	-1.3788014	-1.4200018	2.2000760	1.8778699	-0.1538401	
14	-1.3695473	-1.4050953	-0.1679865	0.3265773	-0.0791720	
15	-1.3417851	-1.3901888	1.8408627	1.2581628	-0.1538401	
16	-1.3140229	-1.3603758	0.7918224	-0.0051669	-0.1538401	
17	-1.3047688	-1.3454693	-1.7752946	-0.5454445	-0.1538401	
18	-1.2955147	-1.3305628	0.2346985	0.8277645	-0.1538401	
19	-1.2770066	-1.3156564	-0.4871601	-0.2485124	-0.1538401	

Choose the most important features using the Regression Analysis by browsing: "Data Transformation" → "Variable Selection" → "Regression Analysis". Then choose the "score4" column as the intercept column, the Significance level (α) as 0.05 and include all columns.

Regression Analysis Model

Significance Level (α)

Select Intercept Column

Excluded Columns

Included Columns

- Col2 -- district
- Col3 -- municipality
- Col4 -- expreg
- Col5 -- expspecial
- Col6 -- expbil
- Col7 -- expocc
- Col8 -- exptot
- Col9 -- scratio
- Col10 -- special

Execute Cancel

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
1	-1.5453748	-1.5988798	-0.4917361	-0.7569975	-0.1538401
2	-1.5361207	-1.5839733	-0.5741035	-0.0894423	-0.1538401
3	-1.5268666	-1.5690668	-1.1483872	-0.3636786	-0.1538401
4	-1.5083585	-1.5541603	-0.7045185	-0.3081651	-0.1538401
5	-1.4898503	-1.5392538	-0.4093687	-0.9455540	-0.1538401
6	-1.4805962	-1.5243473	0.6316636	1.0368776	0.1135672
7	-1.4713422	-1.5094408	0.0619558	1.9748864	-0.1538401
8	-1.4620881	-1.4945343	1.0149008	0.7356338	-0.1538401
9	-1.4250718	-1.4796278	0.4326091	0.0592270	-0.1538401
10	-1.4065636	-1.4647213	-0.9241649	0.4538239	0.4012606
11	-1.3973096	-1.4498148	-0.0066837	-0.2907339	-0.1538401
12	-1.3880555	-1.4349083	0.1672030	1.7492724	-0.1538401
13	-1.3788014	-1.4200018	2.2000760	1.8778699	-0.1538401
14	-1.3695473	-1.4050953	-0.1679865	0.3265773	-0.0791720
15	-1.3417851	-1.3901888	1.8408627	1.2581628	-0.1538401
16	-1.3140229	-1.3603758	0.7918224	-0.0051669	-0.1538401
17	-1.3047688	-1.3454693	-1.7752946	-0.5454445	-0.1538401
18	-1.2955147	-1.3305628	0.2346985	0.8277645	-0.1538401
19	-1.2770066	-1.3156564	-0.4871601	-0.2485124	-0.1538401
20	-1.2677525	-1.3007499	-0.4116567	-0.2339227	-0.1538401
21	-1.2307362	-1.2858434	0.5893359	2.0516135	0.2259769
22	-1.2214821	-1.2709369	-0.1359547	-0.2527243	-0.1538401
23	-1.2122280	-1.2560304	-1.2559224	-0.0110941	-0.1538401
24	-1.2029739	-1.2411239	-1.3337139	-0.9064522	-0.1538401

User Header	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)	Col9	Col10	Col11
1	Regression										
2	Statistics										
3	Multiple R	0.8420505									
4	R Square	0.7096480									
5	Adjusted R Square	0.6797586									
6	Standard Error	8.1538234									
7	Observations	165.0000000									
8	Degrees of Freedom		Sum of Squares	Mean Square	F-statistic	Significance F					
9	Regression	15.0000000	34141.553319	2276.1000000	24.2075704	0E-7					
10	Residual	149.0000000	9906.2406195	66.4842564							
11	Total	164.0000000	34047.793939								
12	Coefficients	Standard Error	t-statistic	P-value	Lower 95.0%	Upper 95.0%					
13	score4	708.5930394	0.8347743	1117.8601114	0E-7	708.3366170	710.8462618				
14	district	-4.5519656	11.2429530	-0.4048728	0.6901517	-26.7681889	17.6642577				
15	municipality	4.6548457	11.2572081	0.4134991	0.6798355	-17.5895459	26.8992373				
16	engprg	3.3917944	3.4812374	0.9743071	0.3314832	-3.4871766	10.2707654				
17	engspecial	1.7432448	1.3444356	1.2966370	0.1967606	-0.9133775	4.3998072				
18	engbil	-0.3472178	0.6513654	-0.5330615	0.5947852	-1.6343244	0.9388887				
19	expacc	0.5710626	0.8133367	0.7021233	0.4836976	-1.0361014	2.1782266				
20	expstot	-4.4255869	4.1361863	-1.0699704	0.2863629	-12.5987557	3.7475618				
21	score8	0.3367487	0.8991702	0.3999810	0.4194720	-0.8050367	1.8225342				
22	special	-0.5935907	0.8349517	-0.7081968	0.4799295	-2.2119277	1.0447463				
23	lunch	-3.0807597	1.9400709	-2.2651463	0.0249454	-5.7622778	-0.3932416				
24	pratio	-0.7476147	0.9058714	-0.8252889	0.4105213	-2.5376286	1.0423992				
25	income	2.4704126	0.9697350	2.5475130	0.0118628	0.5542034	4.3866217				
26	score8	5.8865253	1.1662568	5.1331106	9E-7	3.6819864	8.2910642				
27	salary	0.8044463	0.8344528	1.1406915	0.2558285	-0.6886633	2.5695759				
28	english	-1.8826685	0.9715362	-1.9378263	0.0545355	-3.8024368	0.0370999				

The significant features according to the p-value are the following:

- score4 (p-value = 0.0)
- lunch (p-value = 0.024945423143072513)
- score8 (p-value = 8.766402775289995E-7)
- income (p-value = 0.011862821687884247)
- english (p-value = 0.05453546054604336)

Step 8: Feature selection: train set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION_TRAIN_SET".

Import data into the input spreadsheet of the "FEATURE_SELECTION_TRAIN_SET" tab from the output of the "NORMALIZE_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1		-1.5453748	-1.5988798	-0.4917361	-0.7569975	-0.1538401	-0.3970386	-0.7621131	
2		-1.5361207	-1.5839733	-0.5741035	-0.0894423	-0.1538401	-0.3970386	-0.4750734	
3		-1.5268666	-1.5690668	-1.1483872	-0.3636786	-0.1538401	-0.3970386	-1.1310198	
4		-1.5083585	-1.5541603	-0.7045185	-0.3081651	-0.1538401	-0.3970386	-0.5801865	
5		-1.4898503	-1.5392538	-0.4093687	-0.9455540	-0.1538401	-0.3970386	-0.5822079	
6		-1.4805962	-1.5243473	0.6316636	1.0368776	0.1135672	-0.3970386	1.0652383	
7		-1.4713422	-1.5094408	0.0619558	1.9748864	-0.1538401	-0.3970386	0.1384235	
8		-1.4620881	-1.4945343	1.0149008	0.7356338	-0.1538401	-0.3970386	1.0157138	
9		-1.4250718	-1.4796278	0.4326091	0.0592270	-0.1538401	-0.3970386	0.2516222	
10		-1.4065636	-1.4647213	-0.9241649	0.4528239	0.4012606	3.7828784	-0.5923150	
11		-1.3973096	-1.4498148	-0.0066837	-0.2907339	-0.1538401	-0.3970386	-0.1920765	
12		-1.3880555	-1.4349083	0.1672030	1.7492724	-0.1538401	-0.3970386	0.2172583	
13		-1.3788014	-1.4200018	2.2000760	1.8778699	-0.1538401	-0.3970386	2.0102458	
14		-1.3695473	-1.4050953	-0.1679865	0.3265773	-0.0791720	-0.3970386	-0.0778670	
15		-1.3417851	-1.3901888	1.8408627	1.2581628	-0.1538401	-0.3970386	1.8535868	
16		-1.3140229	-1.3603758	0.7918224	-0.0051669	-0.1538401	-0.3970386	0.6720747	
17		-1.3047688	-1.3454693	-1.7752946	-0.5454445	-0.1538401	-0.3970386	-1.4857766	
18		-1.2955147	-1.3305628	0.2346985	0.8277645	-0.1538401	-0.3970386	0.7286741	
19		-1.2770066	-1.3156564	-0.4871601	-0.2485124	-0.1538401	-0.3970386	-0.4437416	
20		-1.2677525	-1.3007499	-0.4116567	-0.2339227	-0.1538401	-0.3970386	-0.5043838	

Manipulate the data by choosing the columns that correspond to the significant features (from the previous step): "Data Transformation" → "Data Manipulation" → "Select Column(s)".

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1		-1.5453748	-1.5988798	-0.4917361	-0.7569975	-0.1538401	-0.3970386	
2		-1.5361207	-1.5839733	-0.5741035	-0.0894423	-0.1538401	-0.3970386	
3		-1.5268666	-1.5690668	-1.1483872	-0.3636786	-0.1538401	-0.3970386	
4		-1.5083585	-1.5541603	-0.7045185	-0.3081651	-0.1538401	-0.3970386	
5		-1.4898503	-1.5392538	-0.4093687	-0.9455540	-0.1538401	-0.3970386	
6		-1.4805962	-1.5243473	0.6316636	1.0368776	0.1135672	-0.3970386	
7		-1.4713422	-1.5094408	0.0619558	1.9748864	-0.1538401	-0.3970386	
8		-1.4620881	-1.4945343	1.0149008	0.7356338	-0.1538401	-0.3970386	
9		-1.4250718	-1.4796278	0.4326091	0.0592270	-0.1538401	-0.3970386	
10		-1.4065636	-1.4647213	-0.9241649	0.4528239	0.4012606	3.7828784	
11		-1.3973096	-1.4498148	-0.0066837	-0.2907339	-0.1538401	-0.3970386	
12		-1.3880555	-1.4349083	0.1672030	1.7492724	-0.1538401	-0.3970386	
13		-1.3788014	-1.4200018	2.2000760	1.8778699	-0.1538401	-0.3970386	
14		-1.3695473	-1.4050953	-0.1679865	0.3265773	-0.0791720	-0.3970386	
15		-1.3417851	-1.3901888	1.8408627	1.2581628	-0.1538401	-0.3970386	
16		-1.3140229	-1.3603758	0.7918224	-0.0051669	-0.1538401	-0.3970386	
17		-1.3047688	-1.3454693	-1.7752946	-0.5454445	-0.1538401	-0.3970386	
18		-1.2955147	-1.3305628	0.2346985	0.8277645	-0.1538401	-0.3970386	
19		-1.2770066	-1.3156564	-0.4871601	-0.2485124	-0.1538401	-0.3970386	
20		-1.2677525	-1.3007499	-0.4116567	-0.2339227	-0.1538401	-0.3970386	

Excluded Columns

Col4 -- expreg
Col5 -- expspecial
Col6 -- expbit
Col7 -- expocc
Col8 -- expot
Col9 -- scratio
Col10 -- special
Col12 -- stratio

Included Columns

Col11 -- lunch
Col13 -- income
Col14 -- score4
Col15 -- score8
Col17 -- english

Execute Cancel

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	-1.5453748	-1.5988798	-0.4917361	-0.7569975	-0.1538401	-0.3970386	-0.7621131	
2	-1.5361207	-1.5839733	-0.5741035	-0.0894423	-0.1538401	-0.3970386	-0.4750734	
3	-1.5268666	-1.5690668	-1.1483872	-0.3636786	-0.1538401	-0.3970386	-1.1310198	
4	-1.5083585	-1.5541603	-0.7045185	-0.3081651	-0.1538401	-0.3970386	-0.5801865	
5	-1.4898503	-1.5392538	-0.4093687	-0.9455540	-0.1538401	-0.3970386	-0.5822079	
6	-1.4805962	-1.5248473	0.6316636	1.0368776	0.1135672	-0.3970386	1.0652383	
7	-1.4713422	-1.5094408	0.0619558	1.9748864	-0.1538401	-0.3970386	1.1384235	
8	-1.4620881	-1.4945343	1.0149008	0.7356338	-0.1538401	-0.3970386	1.0157138	
9	-1.4250718	-1.4796278	0.4326091	0.0592270	-0.1538401	-0.3970386	0.2516222	
10	-1.4065636	-1.4647213	-0.9241649	0.4528239	0.4012606	3.7828784	-0.5923150	
11	-1.3973096	-1.4498148	-0.0066837	-0.2907339	-0.1538401	-0.3970386	-0.1920765	
12	-1.3880555	-1.4349083	0.1672030	1.7492724	-0.1538401	-0.3970386	0.2172583	
13	-1.3788014	-1.4200018	2.2000760	1.8778699	-0.1538401	-0.3970386	2.0102458	
14	-1.3695473	-1.4050953	-0.1679865	0.3265773	-0.0791720	-0.3970386	-0.0778670	
15	-1.3417851	-1.3901888	1.8408627	1.2581628	-0.1538401	-0.3970386	1.8535868	
16	-1.3140229	-1.3603758	0.7918224	-0.0051669	-0.1538401	-0.3970386	0.6720747	
17	-1.3047688	-1.3454693	-1.7752946	-0.5454445	-0.1538401	-0.3970386	-1.4857766	
18	-1.2955147	-1.3305628	0.2340985	0.8277645	-0.1538401	-0.3970386	0.7286741	
19	-1.2770066	-1.3156564	-0.4871601	-0.2485124	-0.1538401	-0.3970386	-0.4437416	
20	-1.2677525	-1.3007499	-0.4116567	-0.2339227	-0.1538401	-0.3970386	-0.5043838	

Step 9: Feature selection: test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION_TEST_SET".

Import data into the input spreadsheet of the "FEATURE_SELECTION_TEST_SET" tab from the output of the "NORMALIZE_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	-1.3232769	-1.3752823	-0.4311045	-0.6790780	0.0726519	0.1363825	0.3558104
2	-1.0826709	-1.0771524	0.2301226	1.5124335	-0.1538401	-0.3970386	0.5174372
3	-0.9901302	-1.0026199	-0.1256588	3.9029579	0.1764840	1.3465590	-0.0435031
4	-0.9716220	-0.9877134	-0.2103142	0.0423915	-0.1538401	-0.3970386	-0.1112202
5	-0.9531139	-0.9579005	1.0492205	0.7099301	-0.1538401	-0.3970386	0.7822414
6	-0.9346057	-0.9429940	1.1178600	2.1220015	-0.1538401	-0.3970386	1.2077474
7	-0.8698272	-0.8684615	-1.0797478	0.4955862	-0.1538401	-0.3970386	-0.7307813
8	-0.8420649	-0.8535550	-0.7457022	-1.2137015	-0.1538401	-0.3970386	-0.8571192
9	-0.7495242	-0.7790225	-0.5821114	-0.6618195	-0.1538401	-0.3970386	-0.6054541
10	-0.7402701	-0.7343030	-0.8841252	-0.7033281	-0.1538401	-0.3970386	-0.8945152
11	-0.6662375	-0.6448640	0.0585238	-1.0470550	-0.1538401	-0.3970386	-0.2699006
12	-0.6199671	-0.5852381	-1.0728838	-0.9293542	-0.1538401	-0.0145720	-1.2866680
13	-0.6014589	-0.5554251	-0.8738293	-1.3676798	-0.1538401	-0.3970386	-1.0289387
14	-0.5181722	-0.4808926	-1.4080733	-1.0666030	-0.1538401	-0.3970386	-1.1017094
15	-0.4996641	-0.4659861	-1.0888997	-0.5466422	-0.1538401	-0.3970386	-0.7186529
16	-0.3423448	-0.3616406	0.4268891	0.3192182	-0.1538401	-0.3970386	0.3759387
17	-0.3053285	-0.3318276	-1.9743491	-1.2045874	-0.1538401	-0.3970386	-1.9557536
18	-0.2868203	-0.3020146	1.6223603	-0.3227552	0.0744103	1.5831512	1.2077474
19	-0.2683122	-0.2722016	-0.3464492	-0.5198240	-0.1538401	-0.3970386	-0.5256085
20	-0.0924847	-0.0486042	0.5961999	0.9806120	-0.1538401	-0.3970386	0.4163668

Manipulate the data by choosing the columns that correspond to the significant features (from the step 7): "Data Transformation" → "Data Manipulation" → "Select Column(s)".

The screenshot shows the 'ma_score_data.ekk' application window. The 'Data Transformation' menu is open, showing options like 'Remove Column(s)', 'Select Column(s)', 'Matrix Transpose', 'Sort by Column', and 'Fill Missing Column(s) Values'. The 'Select Column(s)' dialog box is also open, showing a list of columns to be selected. The dialog has two panes: 'Excluded Columns' and 'Included Columns'. The 'Excluded Columns' pane contains a list of columns: Col4 -- expreg, Col5 -- expspecial, Col6 -- expbil, Col7 -- expocc, Col8 -- exptot, Col9 -- scratio, Col10 -- special, and Col12 -- stratio. The 'Included Columns' pane contains a list of columns: Col11 -- lunch, Col13 -- income, Col14 -- score4, Col15 -- score8, and Col17 -- english. The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	-1.3232769	-1.3752823	-0.4311045	-0.6790780	0.0726519	0.136	0.049794	0.498794	0.498794
2	-1.0826709	-1.0771524	0.2301226	1.5124335	-0.1538401	-0.39	0.049794	0.498794	0.498794
3	-0.9901302	-1.0026199	-0.1256588	3.9029579	0.1764840	1.346	0.049794	0.498794	0.498794
4	-0.9716220	-0.9877134	-0.2103142	0.0423915	-0.1538401	-0.39	0.049794	0.498794	0.498794
5	-0.9531139	-0.9579005	1.0492205	0.7099301	-0.1538401	-0.39	0.049794	0.498794	0.498794
6	-0.9346057	-0.9429940	1.1178600	2.1220015	-0.1538401	-0.39	0.049794	0.498794	0.498794
7	-0.8698272	-0.8684615	-1.0797478	0.4955862	-0.1538401	-0.39	0.049794	0.498794	0.498794
8	-0.8420649	-0.8535550	-0.7457022	-1.2137015	-0.1538401	-0.39	0.049794	0.498794	0.498794
9	-0.7495242	-0.7790225	-0.5821114	-0.6618195	-0.1538401	-0.39	0.049794	0.498794	0.498794
10	-0.7402701	-0.7343030	-0.8841252	-0.7033281	-0.1538401	-0.39	0.049794	0.498794	0.498794
11	-0.6662375	-0.6448640	0.0585238	-1.0470550	-0.1538401	-0.39	0.049794	0.498794	0.498794
12	-0.6199671	-0.5852381	-1.0728838	-0.9293542	-0.1538401	-0.01	0.049794	0.498794	0.498794
13	-0.6014589	-0.5554251	-0.8738293	-1.3676798	-0.1538401	-0.39	0.049794	0.498794	0.498794
14	-0.5181722	-0.4808926	-1.4080733	-1.0666030	-0.1538401	-0.39	0.049794	0.498794	0.498794
15	-0.4996641	-0.4659861	-1.0888997	-0.5466422	-0.1538401	-0.39	0.049794	0.498794	0.498794
16	-0.3423448	-0.3616406	0.4268891	0.3192182	-0.1538401	-0.39	0.049794	0.498794	0.498794
17	-0.3053285	-0.3318276	-1.9743491	-1.2045874	-0.1538401	-0.39	0.049794	0.498794	0.498794
18	-0.2868203	-0.3020146	1.6223603	-0.3227552	0.0744103	1.583	0.049794	0.498794	0.498794
19	-0.2683122	-0.2722016	-0.3464492	-0.5198240	-0.1538401	-0.39	0.049794	0.498794	0.498794
20	-0.0924847	-0.0486042	0.5961999	0.9806120	-0.1538401	-0.39	0.049794	0.498794	0.498794

The results will appear on the output spreadsheet.

The screenshot shows the 'ma_score_data.ekk' application window with the 'TRAIN_MODEL(.fit)' tab selected. The output spreadsheet is displayed, showing the results of the model training. The spreadsheet has columns for 'User Header', 'Col1', 'Col2 (D)', 'Col3 (D)', 'Col4 (D)', 'Col5 (D)', 'Col6 (D)', 'Col7 (D)', 'Col8 (D)', and 'Col9 (D)'. The data is organized into two main sections: 'User Header' and 'Col1' through 'Col9 (D)'. The 'User Header' section contains the 'User Row ID' and 'district' columns. The 'Col1' through 'Col9 (D)' section contains the model coefficients and other metrics.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	-1.3232769	-1.3752823	-0.4311045	-0.6790780	0.0726519	0.136	0.049794	0.498794	0.498794
2	-1.0826709	-1.0771524	0.2301226	1.5124335	-0.1538401	-0.39	0.049794	0.498794	0.498794
3	-0.9901302	-1.0026199	-0.1256588	3.9029579	0.1764840	1.346	0.049794	0.498794	0.498794
4	-0.9716220	-0.9877134	-0.2103142	0.0423915	-0.1538401	-0.39	0.049794	0.498794	0.498794
5	-0.9531139	-0.9579005	1.0492205	0.7099301	-0.1538401	-0.39	0.049794	0.498794	0.498794
6	-0.9346057	-0.9429940	1.1178600	2.1220015	-0.1538401	-0.39	0.049794	0.498794	0.498794
7	-0.8698272	-0.8684615	-1.0797478	0.4955862	-0.1538401	-0.39	0.049794	0.498794	0.498794
8	-0.8420649	-0.8535550	-0.7457022	-1.2137015	-0.1538401	-0.39	0.049794	0.498794	0.498794
9	-0.7495242	-0.7790225	-0.5821114	-0.6618195	-0.1538401	-0.39	0.049794	0.498794	0.498794
10	-0.7402701	-0.7343030	-0.8841252	-0.7033281	-0.1538401	-0.39	0.049794	0.498794	0.498794
11	-0.6662375	-0.6448640	0.0585238	-1.0470550	-0.1538401	-0.39	0.049794	0.498794	0.498794
12	-0.6199671	-0.5852381	-1.0728838	-0.9293542	-0.1538401	-0.01	0.049794	0.498794	0.498794
13	-0.6014589	-0.5554251	-0.8738293	-1.3676798	-0.1538401	-0.39	0.049794	0.498794	0.498794
14	-0.5181722	-0.4808926	-1.4080733	-1.0666030	-0.1538401	-0.39	0.049794	0.498794	0.498794
15	-0.4996641	-0.4659861	-1.0888997	-0.5466422	-0.1538401	-0.39	0.049794	0.498794	0.498794
16	-0.3423448	-0.3616406	0.4268891	0.3192182	-0.1538401	-0.39	0.049794	0.498794	0.498794
17	-0.3053285	-0.3318276	-1.9743491	-1.2045874	-0.1538401	-0.39	0.049794	0.498794	0.498794
18	-0.2868203	-0.3020146	1.6223603	-0.3227552	0.0744103	1.583	0.049794	0.498794	0.498794
19	-0.2683122	-0.2722016	-0.3464492	-0.5198240	-0.1538401	-0.39	0.049794	0.498794	0.498794
20	-0.0924847	-0.0486042	0.5961999	0.9806120	-0.1538401	-0.39	0.049794	0.498794	0.498794

Step 10: Train the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_MODEL(.fit)".

Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	User Row ID	lunch	income	score4	score5	english	Col7
1		-0.2610802	-0.3923772	714.0000000	-0.3884531	-0.4380666	
2		-0.8831728	1.3481476	731.0000000	0.0230312	0.1017883	
3		-0.1072294	-0.8248734	704.0000000	-0.2774077	-0.4380666	
4		-0.2410127	-0.4419321	704.0000000	-0.3884531	-0.2983138	
5		0.1135131	-0.5691478	701.0000000	0.0557285	-0.4380666	
6		0.7422949	-1.3603626	714.0000000	0.0230312	1.2608885	
7		-0.8296595	1.4470726	725.0000000	1.6658866	-0.4380666	
8		-0.3012153	0.5450987	717.0000000	0.9440916	0.7328349	
9		-0.4751336	0.6307104	702.0000000	0.3888646	-0.4380666	
10		0.3342557	-0.6529104	701.0000000	-0.5550212	-0.2755024	
11		-0.3346611	-0.1850970	713.0000000	0.2778193	-0.4380666	
12		-0.2543911	-0.4299132	707.0000000	-0.8326347	-0.4380666	
13		0.5483090	-0.7239145	703.0000000	-0.9992028	-0.4380666	
14		0.2004724	-0.2080254	704.0000000	0.0002058	-0.0033841	
15		-0.8564162	1.1258900	721.0000000	1.5548413	-0.4380666	
16		-0.6156061	1.5332390	728.0000000	1.4993186	0.5037455	
17		-0.6758086	-0.7141145	710.0000000	-0.0553169	-0.4380666	
18		-0.5487145	0.1140817	731.0000000	0.0230312	-0.4380666	
19		-0.0269594	-0.0120245	713.0000000	0.7220008	-0.4380666	
20		-0.6156061	-0.3894187	710.0000000	-0.1108396	-0.4380666	

Use the k Nearest Neighbors (kNN) method to train and fit the model by browsing: "Analytics" → "Regression" → "k Nearest Neighbors (kNN)" and set the "Target Column" as the column corresponding to "score4" and the "Number of Neighbors" to 5.

kNN Regression Model

Target Column: Col3 -- score4

Number of Neighbors: 5

Execute Cancel

The predictions will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	User Row ID	lunch	income	score4	score8	english			
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)
1	User Row ID	score4	knn Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2
2				Entry 1	0E-7	Entry 4	0.0236435
3				Entry 2	0E-7	Entry 23	0.0959195
4				Entry 3	0E-7	Entry 86	0.0225432
5				Entry 4	0E-7	Entry 1	0.0236435
6				Entry 5	0E-7	Entry 123	0.0128778
7				Entry 6	0E-7	Entry 106	0.1952978
8				Entry 7	0E-7	Entry 15	0.0581360
9				Entry 8	0E-7	Entry 118	0.1152796
10				Entry 9	0E-7	Entry 88	0.0652998
11				Entry 10	0E-7	Entry 22	0.0404206
12				Entry 11	0E-7	Entry 103	0.0371878
13				Entry 12	0E-7	Entry 95	0.0690548
14				Entry 13	0E-7	Entry 108	0.0429316
15				Entry 14	0E-7	Entry 93	0.0532959
16				Entry 15	0E-7	Entry 110	0.0404825
17				Entry 16	0E-7	Entry 99	0.1197324
18				Entry 17	0E-7	Entry 47	0.0278016

Step 11: Validate the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "VALIDATE_MODEL(.predict)".

Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	User Row ID	lunch	income	score4	score8	english			
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	User Row ID						
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

To validate the model browse: "Analytics" → "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN_MODEL (.fit)". and transfer all columns in the output.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7
1		-0.5085794	-0.4866795	708.0000000	-0.5550212	-0.4380666	
2		-1.0236454	3.3072318	739.0000000	2.7208179	-0.4380666	
3		1.5851302	-0.9201003	684.0000000	-1.7209978	0.6010545	
4		0.1603374	-0.9914742	715.0000000	0.0002058	-0.4380666	
5		-0.7962137	2.3418350	724.0000000	1.1106597	-0.4380666	
6		-0.8497270	2.4322543	731.0000000	2.4987271	-0.2446588	
7		-0.4483769	-0.1802894	708.0000000	0.0230312	-0.4380666	
8		-0.5219578	-0.7102314	710.0000000	-0.1108396	-0.4380666	
9		-0.6356736	-0.2707087	729.0000000	0.4999100	-0.4380666	
10		-0.8162812	0.0952212	720.0000000	0.6664781	-0.4380666	
11		0.1737156	-0.2532375	710.0000000	-0.3329304	-0.3313588	
12		-0.8029028	-0.0504851	729.0000000	0.3333419	-0.4380666	
13		0.5750657	-0.9789005	688.0000000	-1.2768162	-0.4380666	
14		-0.2276344	-0.4726266	707.0000000	0.7220008	-0.4380666	
15		1.0499966	-0.8890360	693.0000000	-1.1102481	-0.4380666	
16		-0.8430378	1.3359438	719.0000000	1.3882732	-0.4380666	
17		0.1737156	-0.7451787	706.0000000	0.0230312	-0.4380666	
18		3.6119477	-1.3707174	658.0000000	-3.1645879	10.1735245	
19		-0.9634429	0.7388807	716.0000000	1.2217051	-0.4380666	
20		-0.7025653	0.1958104	730.0000000	0.4999100	-0.4380666	

Existing Model Execution

Model: (from Tab): TRAIN_MODEL(...)

Type: kNN Model

Description:

Model Input:

- Header -> Datatype
- district -> Double
- municipality -> Double
- expreg -> Double
- expspecial -> Double
- expbil -> Double
- expocc -> Double
- exp tot -> Double
- scratio -> Double

☒ Transfer Column(s) to Output

Excluded Columns:

Included Columns:

- Col2 -- lunch
- Col3 -- income
- Col4 -- score4
- Col5 -- score8
- Col6 -- english

Execute Cancel

The predictions will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)
1		708.4565288	Entry 1	0.0593834	Entry 42	0.0646836	Entry 4	0.0648433
2		732.1723174	Entry 148	0.1383970	Entry 133	0.1968367	Entry 146	0.2108242
3		693.9074305	Entry 115	0.0918920	Entry 164	0.0996810	Entry 129	0.1079874
4		705.4957649	Entry 86	0.0645061	Entry 149	0.0706385	Entry 5	0.0728075
5		724.2152123	Entry 87	0.0865315	Entry 85	0.1276155	Entry 161	0.1314347
6		735.2400904	Entry 148	0.0551005	Entry 84	0.1160834	Entry 161	0.1348037
7		714.0929672	Entry 155	0.0153952	Entry 103	0.0244684	Entry 40	0.0259753
8		701.7804416	Entry 47	0.0246733	Entry 17	0.0319624	Entry 116	0.0438552
9		715.5261789	Entry 61	0.0225936	Entry 140	0.0415972	Entry 68	0.0641771
10		713.0335428	Entry 65	0.0439525	Entry 98	0.0521968	Entry 124	0.0714164
11		711.0337694	Entry 60	0.0548319	Entry 80	0.0572327	Entry 156	0.0674449
12		718.3546106	Entry 68	0.0398645	Entry 61	0.0402162	Entry 98	0.0562203
13		707.0339458	Entry 13	0.0659099	Entry 53	0.0793055	Entry 108	0.0858458
14		719.0140180	Entry 140	0.0821372	Entry 111	0.0839176	Entry 19	0.0875075
15		697.1315526	Entry 144	0.0789966	Entry 147	0.1002759	Entry 13	0.1048347
16		725.0091364	Entry 15	0.0464240	Entry 7	0.0530288	Entry 46	0.0556326
17		707.9350285	Entry 5	0.0326252	Entry 123	0.0425212	Entry 149	0.0516342

Step 12: Statistics calculation

Create a new tab by pressing the "+" button on the bottom of the page with the name "STATISTICS_ACCURACIES".

Import data into the input spreadsheet of the "STATISTICS_ACCURACIES" tab from the output of the "VALIDATE_MODEL(.predict)" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)
1	708.4565288	Entry 1	0.0593834	Entry 42	0.0646836	Entry 4	0.0648433	Entry 12	0.0711551	
2	732.1723174	Entry 148	0.1383970	Entry 133	0.1968367	Entry 146	0.2108242	Entry 84	0.2426462	
3	693.9074305	Entry 115	0.0918920	Entry 164	0.0996810	Entry 129	0.1079874	Entry 135	0.1352939	
4	705.4957649	Entry 86	0.0645061	Entry 149	0.0706385	Entry 5	0.0728075	Entry 3	0.0777121	
5	724.2152123	Entry 87	0.0865315	Entry 85	0.1276155	Entry 161	0.1314347	Entry 133	0.1385466	
6	735.2400904	Entry 148	0.0551005	Entry 84	0.1160834	Entry 161	0.1348037	Entry 133	0.1438306	
7	714.0929672	Entry 155	0.0153952	Entry 103	0.0244684	Entry 40	0.0259753	Entry 62	0.0280150	
8	701.7804416	Entry 47	0.0246733	Entry 17	0.0319624	Entry 116	0.0438552	Entry 76	0.0554472	
9	715.5261789	Entry 61	0.0225936	Entry 140	0.0415972	Entry 68	0.0641771	Entry 98	0.0682909	
10	713.0335428	Entry 65	0.0439525	Entry 98	0.0521968	Entry 124	0.0714164	Entry 94	0.0754928	
11	711.0337694	Entry 60	0.0548319	Entry 80	0.0572327	Entry 156	0.0674449	Entry 137	0.0714676	
12	718.3546106	Entry 68	0.0398645	Entry 61	0.0402162	Entry 98	0.0562203	Entry 56	0.0578690	
13	707.0339458	Entry 13	0.0659099	Entry 53	0.0793055	Entry 108	0.0858458	Entry 69	0.1033733	
14	719.0140180	Entry 140	0.0821372	Entry 111	0.0839176	Entry 19	0.0875075	Entry 163	0.0942115	
15	697.1315526	Entry 144	0.0789966	Entry 147	0.1002759	Entry 13	0.1048347	Entry 53	0.1052986	
16	725.0091364	Entry 15	0.0464240	Entry 7	0.0530288	Entry 46	0.0556326	Entry 121	0.0568986	
17	707.9350285	Entry 5	0.0326252	Entry 123	0.0425212	Entry 149	0.0516342	Entry 48	0.0519424	
18	672.8502822	Entry 78	0.6559415	Entry 21	0.9414410	Entry 119	0.9680037	Entry 38	1.0071733	
19	728.6301900	Entry 107	0.0492053	Entry 110	0.0564059	Entry 154	0.0668481	Entry 121	0.0673432	

Calculate the statistical metrics for the regression by browsing: "Statistics" → "Model Metrics" → "Regression Metrics".

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)
1	708.4565288	Entry 1	0.0593834	Entry 42	0.0646836	Entry 4	0.0648433	
2	732.1723174	Entry 148	0.1383970	Entry 133	0.1968367	Entry 146	0.2108242	
3	693.9074305	Entry 115	0.0918920	Entry 164	0.0996810	Entry 129	0.1079874	
4	705.4957649	Entry 86	0.0645061	Entry 149	0.0706385	Entry 5	0.0728075	
5	724.2152123	Entry 87	0.0865315	Entry 85	0.1276155	Entry 161	0.1314347	
6	735.2400904	Entry 148	0.0551005	Entry 84	0.1160834	Entry 161	0.1348037	
7	714.0929672	Entry 155	0.0153952	Entry 103	0.0244684	Entry 40	0.0259753	
8	701.7804416	Entry 47	0.0246733	Entry 17	0.0319624	Entry 116	0.0438552	
9	715.5261789	Entry 61	0.0225936	Entry 140	0.0415972	Entry 68	0.0641771	
10	713.0335428	Entry 65	0.0439525	Entry 98	0.0521968	Entry 124	0.0714164	
11	711.0337694	Entry 60	0.0548319	Entry 80	0.0572327	Entry 156	0.0674449	
12	718.3546106	Entry 68	0.0398645	Entry 61	0.0402162	Entry 98	0.0562203	
13	707.0339458	Entry 13	0.0659099	Entry 53	0.0793055	Entry 108	0.0858458	
14	719.0140180	Entry 140	0.0821372	Entry 111	0.0839176	Entry 19	0.0875075	
15	697.1315526	Entry 144	0.0789966	Entry 147	0.1002759	Entry 13	0.1048347	
16	725.0091364	Entry 15	0.0464240	Entry 7	0.0530288	Entry 46	0.0556326	
17	707.9350285	Entry 5	0.0326252	Entry 123	0.0425212	Entry 149	0.0516342	
18	672.8502822	Entry 78	0.6559415	Entry 21	0.9414410	Entry 119	0.9680037	
19	728.6301900	Entry 107	0.0492053	Entry 110	0.0564059	Entry 154	0.0668481	

Regression Statistics Metrics

Actual Value Column
Col15 -- score4

Prediction Value Column
Col2 -- kNN Prediction

Execute
Cancel

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)
1	708.4565288	Entry 1	0.0593834	Entry 42	0.0646836	Entry 4	0.0648433	Entry 12	0.0711551	
2	732.1723174	Entry 148	0.1383970	Entry 133	0.1968267	Entry 146	0.2108242	Entry 84	0.2426462	
3	693.9074305	Entry 115	0.0918920	Entry 164	0.0996810	Entry 129	0.1079874	Entry 135	0.1352939	
4	705.4957649	Entry 86	0.0645061	Entry 149	0.0706385	Entry 5	0.0728075	Entry 3	0.0777121	
5	724.2152123	Entry 87	0.0865315	Entry 85	0.1276155	Entry 161	0.1314347	Entry 133	0.1385466	
6	735.2400904	Entry 148	0.0551005	Entry 84	0.1160834	Entry 161	0.1348037	Entry 133	0.1438306	
7	714.0929672	Entry 155	0.0153952	Entry 103	0.0244684	Entry 40	0.0259753	Entry 62	0.0280150	
8	701.7804416	Entry 47	0.0246733	Entry 17	0.0319624	Entry 116	0.0438552	Entry 76	0.0554472	
9	715.5261789	Entry 61	0.0225936	Entry 140	0.0415972	Entry 68	0.0641771	Entry 98	0.0682909	
10	713.035428	Entry 65	0.0435525	Entry 98	0.0521968	Entry 124	0.0714164	Entry 94	0.0754928	
11	711.037694	Entry 60	0.0548319	Entry 80	0.0572327	Entry 156	0.0674449	Entry 137	0.0714676	
12	718.3546106	Entry 68	0.0398645	Entry 61	0.0402162	Entry 98	0.0562203	Entry 56	0.0578699	
13	707.0339458	Entry 13	0.0659099	Entry 53	0.0793055	Entry 108	0.0858458	Entry 69	0.1033733	
14	719.0140180	Entry 140	0.0821372	Entry 111	0.0839176	Entry 19	0.0875075	Entry 163	0.0942115	
15	697.1315526	Entry 144	0.0789966	Entry 147	0.1002759	Entry 13	0.1048347	Entry 53	0.1052986	
16	725.091364	Entry 15	0.0464240	Entry 7	0.0530288	Entry 46	0.0556326	Entry 121	0.0568986	
17	707.9350285	Entry 5	0.0326252	Entry 123	0.0425212	Entry 149	0.0516342	Entry 48	0.0519424	
18	672.802822	Entry 78	0.6559415	Entry 21	0.9414410	Entry 119	0.9680037	Entry 38	1.0071733	
19	728.6301900	Entry 107	0.0492053	Entry 110	0.0564059	Entry 154	0.0668481	Entry 121	0.0673432	

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6
1		81.4273095	9.0237082	7.3824663	0.7228803	
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						

Step 13: Reliability check of each record of the test set

Step 13.a: Create the domain

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_SCORE4".

Import data into the input spreadsheet of the "EXCLUDE_SCORE4" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8
1	-0.2610802	-0.3923772	714.0000000	-0.3884531	-0.4380666			
2	-0.8831728	1.3481476	731.0000000	0.0230312	0.1017883			
3	-0.1072294	-0.8248734	704.0000000	-0.2774077	-0.4380666			
4	-0.2410127	-0.4419321	704.0000000	-0.3884531	-0.2983138			
5	0.1135131	-0.5691478	701.0000000	0.0557285	-0.4380666			
6	0.7422949	-1.3603626	714.0000000	0.0230312	1.2608885			
7	-0.8296595	1.4470726	725.0000000	1.6658866	-0.4380666			
8	-0.3012153	0.5450987	717.0000000	0.9440916	0.7328349			
9	-0.4751336	0.6307104	702.0000000	0.3888646	-0.4380666			
10	0.3342557	-0.6529104	701.0000000	-0.5550212	-0.2755024			
11	-0.3346611	-0.1850970	713.0000000	0.2778193	-0.4380666			
12	-0.2543911	-0.4299132	707.0000000	-0.8326347	-0.4380666			
13	0.5483090	-0.7239145	703.0000000	-0.9992028	-0.4380666			
14	0.2004724	-0.2080254	704.0000000	0.0002058	-0.0033841			
15	-0.8564162	1.1258900	721.0000000	1.5548413	-0.4380666			
16	-0.6156061	1.5332390	728.0000000	1.4993186	0.5037455			
17	-0.6758086	-0.7141145	710.0000000	-0.0553169	-0.4380666			
18	-0.5487145	0.1140817	731.0000000	0.0230312	-0.4380666			
19	-0.0269594	-0.0120245	713.0000000	0.7220008	-0.4380666			
20	-0.6156061	-0.3894187	710.0000000	-0.1108396	-0.4380666			

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								

Manipulate the data to exclude the column that corresponds to the "score4" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except the "score4".

The screenshot shows the 'ma_score_data.ekk' application window. The 'Data Transformation' menu is open, and the 'Select Column(s)' option is highlighted. A dialog box titled 'Select Column(s)' is displayed, showing a list of columns. The 'Excluded Columns' list contains 'Col4 -- score4'. The 'Included Columns' list contains 'Col2 -- lunch', 'Col3 -- income', 'Col5 -- score8', and 'Col6 -- english'. The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8
1		-0.2610802	-0.3923772	714.0000000	-0.3884531	-0.4380666		
2		-0.8831728	1.3481476	731.0000000	0.0230312	0.1017883		
3		-0.1072294	-0.8248734	704.0000000	-0.2774077	-0.4380666		
4		-0.2410127	-0.4419321	704.0000000	-0.3884531	-0.2983138		
5		0.1135131	-0.5691478	701.0000000	0.0557285	-0.4380666		
6		0.7422949	-1.3603626	714.0000000	0.0230312	1.2608885		
7		-0.8296595	1.4470726	725.0000000	1.6658866	-0.4380666		
8		-0.3012153	0.5450987	717.0000000	0.9440916	0.7328349		
9		-0.4751336	0.6307104	702.0000000	0.3888646	-0.4380666		
10		0.3342557	-0.6529104	701.0000000	-0.5550212	-0.2755024		
11		-0.3346611	-0.1850970	713.0000000	0.2778193	-0.4380666		
12		-0.2543911	-0.4299132	707.0000000	-0.8326347	-0.4380666		
13		0.5483090	-0.7239145	703.0000000	-0.9992028	-0.4380666		
14		0.2004724	-0.2080254	704.0000000	0.0002058	-0.0033841		
15		-0.8564162	1.1258900	721.0000000	1.5548413	-0.4380666		
16		-0.6156061	1.5332390	728.0000000	1.4993186	0.5037455		
17		-0.6758086	-0.7141145	710.0000000	-0.0553169	-0.4380666		
18		-0.5487145	0.1140817	731.0000000	0.0230312	-0.4380666		
19		-0.0269594	-0.0120245	713.0000000	0.7220008	-0.4380666		
20		-0.6156061	-0.3894187	710.0000000	-0.1108396	-0.4380666		

The results will appear on the output spreadsheet.

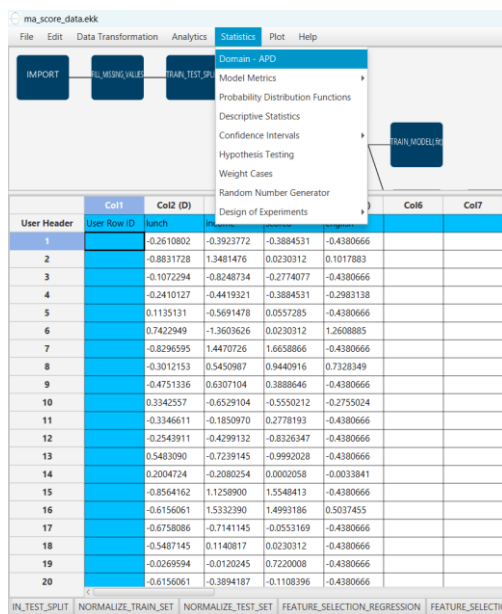
Create a new tab by pressing the "+" button on the bottom of the page with the name "DOMAIN".

Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE_SCORE4" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

The screenshot shows the 'ma_score_data.ekk' application window. The 'Statistics' menu is open, and the 'Domain APD' option is highlighted. The 'Domain APD' tab is selected, showing a table with columns 'User Header', 'User Row ID', 'Col1', 'Col2', 'Col3', 'Col4', 'Col5', 'Col6', 'Col7', and 'Col8'. The table contains data for 20 rows, with the first row having values: User Row ID: 1, Col1: -0.2610802, Col2: -0.3923772, Col3: -0.3884531, Col4: -0.4380666, Col5: 0.0230312, Col6: 0.1017883, Col7: -0.4380666, Col8: -0.2983138.

User Header	User Row ID	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1		-0.2610802	-0.3923772	-0.3884531	-0.4380666	0.0230312	0.1017883	-0.4380666	-0.2983138
2		-0.8831728	1.3481476	0.0230312	0.1017883	-0.4380666	-0.2983138	-0.4380666	-0.2983138
3		-0.1072294	-0.8248734	-0.2774077	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
4		-0.2410127	-0.4419321	-0.3884531	-0.2983138	-0.4380666	-0.2983138	-0.4380666	-0.2983138
5		0.1135131	-0.5691478	0.0557285	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
6		0.7422949	-1.3603626	0.0230312	1.2608885	-0.4380666	-0.2983138	-0.4380666	-0.2983138
7		-0.8296595	1.4470726	1.6658866	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
8		-0.3012153	0.5450987	0.9440916	0.7328349	-0.4380666	-0.2983138	-0.4380666	-0.2983138
9		-0.4751336	0.6307104	0.3888646	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
10		0.3342557	-0.6529104	-0.5550212	-0.2755024	-0.4380666	-0.2983138	-0.4380666	-0.2983138
11		-0.3346611	-0.1850970	0.2778193	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
12		-0.2543911	-0.4299132	-0.8326347	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
13		0.5483090	-0.7239145	-0.9992028	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
14		0.2004724	-0.2080254	0.0002058	-0.0033841	-0.4380666	-0.2983138	-0.4380666	-0.2983138
15		-0.8564162	1.1258900	1.5548413	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
16		-0.6156061	1.5332390	1.4993186	0.5037455	-0.4380666	-0.2983138	-0.4380666	-0.2983138
17		-0.6758086	-0.7141145	-0.0553169	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
18		-0.5487145	0.1140817	0.0230312	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
19		-0.0269594	-0.0120245	0.7220008	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666
20		-0.6156061	-0.3894187	-0.1108396	-0.4380666	-0.2983138	-0.4380666	-0.2983138	-0.4380666

Create the domain by browsing: "Statistics" → "Domain APD".



User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1	-0.2610802	-0.3923772	-0.3884531	-0.4380666			
2	-0.8831728	1.3481476	0.0230312	0.1017883			
3	-0.1072294	-0.8248734	-0.2774077	-0.4380666			
4	-0.2410127	-0.4419321	-0.3884531	-0.2983138			
5	0.1135131	-0.5691478	0.0557285	-0.4380666			
6	0.7422949	-1.3603626	0.0230312	1.2608885			
7	-0.8296595	1.4470726	1.6658866	-0.4380666			
8	-0.3012153	0.5450987	0.9440916	0.7328349			
9	-0.4751336	0.6307104	0.3888646	-0.4380666			
10	0.3342557	-0.6529104	-0.5550212	-0.2755024			
11	-0.3346611	-0.1850970	0.2778193	-0.4380666			
12	-0.2543911	-0.4299132	-0.8326347	-0.4380666			
13	0.5483090	-0.7239145	-0.9992028	-0.4380666			
14	0.2004724	-0.2080254	0.0002058	-0.0033841			
15	-0.8564162	1.1258900	1.5548413	-0.4380666			
16	-0.6156061	1.5332390	1.4993186	0.5037455			
17	-0.6758086	-0.7141145	-0.0553169	-0.4380666			
18	-0.5487145	0.1140817	0.0230312	-0.4380666			
19	-0.0269594	-0.0120245	0.7220008	-0.4380666			
20	-0.6156061	-0.3894187	-0.1108396	-0.4380666			

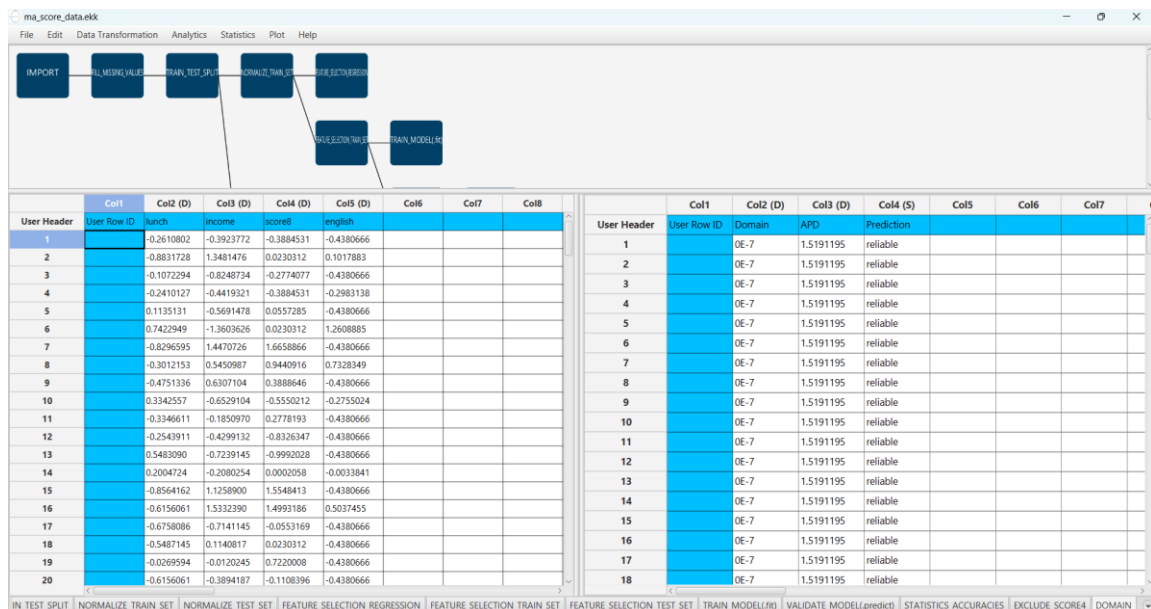
Domain - APD

APD = $d + Z\sigma$, $Z = 0.5$

Perform Computations CPU (double precision)

Execute Cancel

The results will appear on the output spreadsheet.



User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8
1	-0.2610802	-0.3923772	-0.3884531	-0.4380666				
2	-0.8831728	1.3481476	0.0230312	0.1017883				
3	-0.1072294	-0.8248734	-0.2774077	-0.4380666				
4	-0.2410127	-0.4419321	-0.3884531	-0.2983138				
5	0.1135131	-0.5691478	0.0557285	-0.4380666				
6	0.7422949	-1.3603626	0.0230312	1.2608885				
7	-0.8296595	1.4470726	1.6658866	-0.4380666				
8	-0.3012153	0.5450987	0.9440916	0.7328349				
9	-0.4751336	0.6307104	0.3888646	-0.4380666				
10	0.3342557	-0.6529104	-0.5550212	-0.2755024				
11	-0.3346611	-0.1850970	0.2778193	-0.4380666				
12	-0.2543911	-0.4299132	-0.8326347	-0.4380666				
13	0.5483090	-0.7239145	-0.9992028	-0.4380666				
14	0.2004724	-0.2080254	0.0002058	-0.0033841				
15	-0.8564162	1.1258900	1.5548413	-0.4380666				
16	-0.6156061	1.5332390	1.4993186	0.5037455				
17	-0.6758086	-0.7141145	-0.0553169	-0.4380666				
18	-0.5487145	0.1140817	0.0230312	-0.4380666				
19	-0.0269594	-0.0120245	0.7220008	-0.4380666				
20	-0.6156061	-0.3894187	-0.1108396	-0.4380666				

Step 12.b: Check the test set reliability

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_SCORE4_TEST_SET".

Import data into the input spreadsheet of the "EXCLUDE_SCORE4_TEST_SET" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

Filter the data to exclude the column that corresponds to the "score4" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except "score4".

The results will appear on the output spreadsheet.

Import data into the input spreadsheet of the "RELIABILITY" tab from the output of the "EXCLUDE_SCORE4_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

The screenshot shows the 'ma_score_data.ekk' application window. The top menu bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu is a workflow diagram with nodes: IMPORT, ALL MISSING VALUES, TRAIN TEST SPLIT, NORMALIZE TRAIN SET, FEATURE SELECTION REGRESSION, FEATURE SELECTION TRAIN SET, FEATURE SELECTION TEST SET, TRAIN MODEL LR, VALIDATE MODEL predict, STATISTICS ACCURACIES, EXCLUDE SCORE4, DOMAIN, EXCLUDE SCORE4 TEST SET, and RELIABILITY. The bottom part of the window displays a data table with 13 rows and 8 columns.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1		-0.5085794	-0.4866795	-0.5550212	-0.4380666		
2		-1.0236454	3.3072318	2.7208179	-0.4380666		
3		1.5851302	-0.9201003	-1.7209978	0.6010545		
4		0.1603374	-0.9914742	0.0002058	-0.4380666		
5		-0.7962137	2.3418350	1.1106597	-0.4380666		
6		-0.8497270	2.4322543	2.4987271	-0.2446588		
7		-0.4483769	-0.1802894	0.0230312	-0.4380666		
8		-0.5219578	-0.7102314	-0.1108396	-0.4380666		
9		-0.6356736	-0.2707087	0.4999100	-0.4380666		
10		-0.8162812	0.0952212	0.6664781	-0.4380666		
11		0.1737156	-0.2533275	-0.3329304	-0.3313588		
12		-0.8029028	-0.0504851	0.3333419	-0.4380666		
13		0.5750657	-0.9789005	-1.2768162	-0.4380666		
14		-0.3376344	-0.4726366	0.7230008	-0.4380666		

Check the Reliability by browsing: "Analytics" → "Existing Model Utilization". Then select as Model "(from Tab:) DOMAIN".

The screenshot shows the 'ma_score_data.ekk' application window with the 'Analytics' menu open. The 'Existing Model Utilization' option is highlighted. The bottom part of the window displays the same data table as in the previous screenshot.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1		-0.5085794	-0.4866795	-0.5550212	-0.4380666		
2		-1.0236454	3.3072318	2.7208179	-0.4380666		
3		1.5851302	-0.9201003	-1.7209978	0.6010545		
4		0.1603374	-0.9914742	0.0002058	-0.4380666		
5		-0.7962137	2.3418350	1.1106597	-0.4380666		
6		-0.8497270	2.4322543	2.4987271	-0.2446588		
7		-0.4483769	-0.1802894	0.0230312	-0.4380666		
8		-0.5219578	-0.7102314	-0.1108396	-0.4380666		
9		-0.6356736	-0.2707087	0.4999100	-0.4380666		
10		-0.8162812	0.0952212	0.6664781	-0.4380666		
11		0.1737156	-0.2533275	-0.3329304	-0.3313588		
12		-0.8029028	-0.0504851	0.3333419	-0.4380666		
13		0.5750657	-0.9789005	-1.2768162	-0.4380666		
14		-0.3376344	-0.4726366	0.7230008	-0.4380666		

The 'Existing Model Execution' dialog box is shown. It has a 'Model' dropdown set to '(from Tab:) DOMAIN' and a 'Type' dropdown set to 'APD Model'. The 'Description' field is empty. The 'Model Input' section lists various features with their datatypes: Header (Datatype), district (Double), municipality (Double), expreg (Double), expspecial (Double), expbil (Double), expocc (Double), expplot (Double), and scratio (Double). The 'Transfer Column(s) to Output' checkbox is checked. Below this, there are two lists: 'Excluded Columns' (empty) and 'Included Columns' (containing Col2 -- lunch, Col3 -- income, Col4 -- score8, and Col5 -- english). The 'Execute' and 'Cancel' buttons are at the bottom.

The results will appear on the output spreadsheet.

mla_score_data.xlsx

FileEditData TransformationAnalyticsStatisticsPlotHelp

```
graph LR; IMPORT --> TRAINING_VALUES; TRAINING_VALUES --> TRAIN_TEST_SPLIT; TRAIN_TEST_SPLIT --> NORMALIZE_TRAIN_SET; NORMALIZE_TRAIN_SET --> FEATURE_SELECTION_REGRESSION; NORMALIZE_TRAIN_SET --> NORMALIZE_TEST_SET; NORMALIZE_TEST_SET --> FEATURE_SELECTION_TEST_SET; FEATURE_SELECTION_REGRESSION --> EXCLUDE_SCORE4; EXCLUDE_SCORE4 --> DOMAIN; EXCLUDE_SCORE4 --> EXCLUDE_SCORE4_TEST_SET; EXCLUDE_SCORE4_TEST_SET --> RELIABILITY; FEATURE_SELECTION_TEST_SET --> EXCLUDE_SCORE4; FEATURE_SELECTION_TEST_SET --> EXCLUDE_SCORE4_TEST_SET; FEATURE_SELECTION_TEST_SET --> RELIABILITY;
```

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1		-0.5085794	-0.4866795	-0.5550212	-0.4380666		
2		-1.0236454	3.3072318	2.7208179	-0.4380666		
3		1.5851302	-0.9201003	-1.7209978	0.6010545		
4		0.1603374	-0.9914742	0.0002058	-0.4380666		
5		-0.7962137	2.3418350	1.1106597	-0.4380666		
6		-0.8497270	2.4322543	2.4987271	-0.2446588		
7		-0.4483769	-0.1802894	0.0230312	-0.4380666		
8		-0.5219578	-0.7102314	-0.1108396	-0.4380666		
9		-0.6356736	-0.2707087	0.4999100	-0.4380666		
10		-0.8162812	0.0952212	0.6664781	-0.4380666		
11		0.1737156	-0.2533275	-0.3329304	-0.3313588		
12		-0.8029028	-0.0504851	0.3333419	-0.4380666		
13		0.5750657	-0.9789005	-1.2768162	-0.4380666		
14		-0.3376344	-0.4747366	0.7320008	-0.4380666		

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1		0.3128797	1.5191195	reliable	-0.5085794	-0.4866795	-0.5550212	-0.4380666
2		0.8034727	1.5191195	reliable	-1.0236454	3.3072318	2.7208179	-0.4380666
3		0.5117845	1.5191195	reliable	1.5851302	-0.9201003	-1.7209978	0.6010545
4		0.3513367	1.5191195	reliable	0.1603374	-0.9914742	0.0002058	-0.4380666
5		0.4851862	1.5191195	reliable	-0.7962137	2.3418350	1.1106597	-0.4380666
6		0.3230661	1.5191195	reliable	-0.8497270	2.4322543	2.4987271	-0.2446588
7		0.0886498	1.5191195	reliable	-0.4483769	-0.1802894	0.0230312	-0.4380666
8		0.1389370	1.5191195	reliable	-0.5219578	-0.7102314	-0.1108396	-0.4380666
9		0.1267051	1.5191195	reliable	-0.6356736	-0.2707087	0.4999100	-0.4380666
10		0.2557666	1.5191195	reliable	-0.8162812	0.0952212	0.6664781	-0.4380666
11		0.3071865	1.5191195	reliable	0.1737156	-0.2533275	-0.3329304	-0.3313588
12		0.2196993	1.5191195	reliable	-0.8029028	-0.0504851	0.3333419	-0.4380666

ORMALIZE_TEST_SETFEATURE_SELECTION_REGRESSIONFEATURE_SELECTION_TRAIN_SETFEATURE_SELECTION_TEST_SETTRAIN_MODEL(R)VALIDATE_MODEL(predict)STATISTICS_ACCURACIESEXCLUDE_SCORE4DOMAINEXCLUDE_SCORE4_TEST_SETRELIABILITY

There are 3 unreliable samples in the test set.

Final Isalos Workflow

Following the above-described steps, the final workflow on Isalos will look like this:

