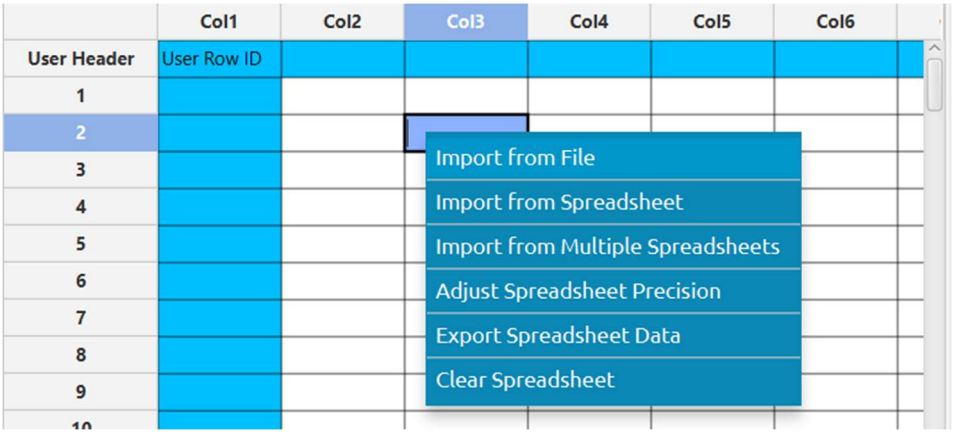
Life Expectancy (WHO)

The data contained in this dataset, which can be found in https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who, were tracked by the World Health Organization (WHO) and were made available to the public for the purpose of health data analysis. The dataset has 21 features and 2938 samples, containing information on immunization, mortality, economic, social and other health related factors. The goal is to predict the life expectancy of individuals from different backgrounds.

Isalos version used: 2.0.6

Step 1: Import data from file

Right click on the input spreadsheet (left) and choose the option “Import from File”. Then navigate through your files to load the one with the life expectancy data.

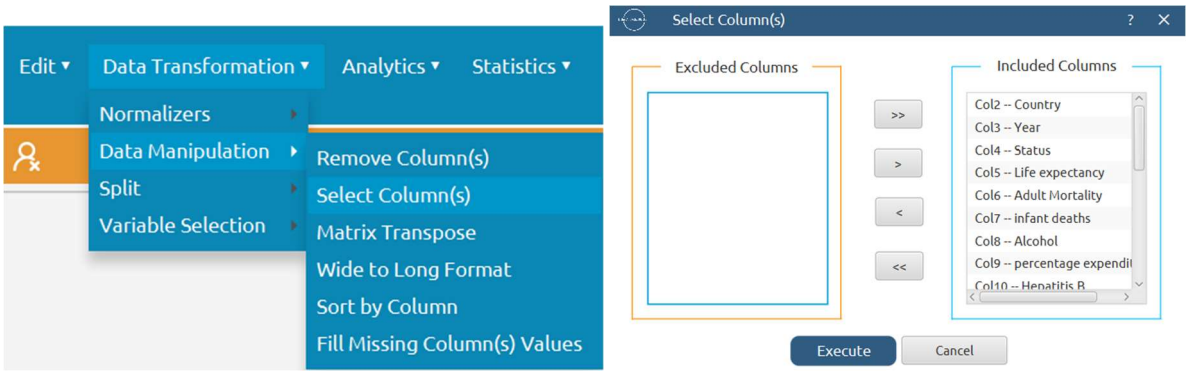


The data will appear on the left spreadsheet.

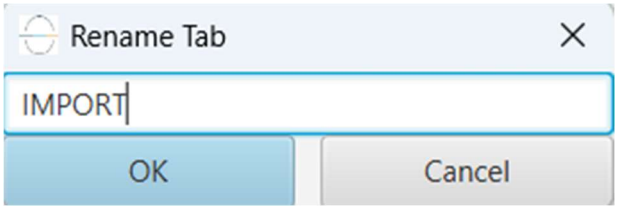
	Col1	Col2 (S)	Col3 (I)	Col4 (S)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (D)	Col9 (D)	Col10 (I)	Col11 (I)	Col12 (D)	Col13 (I)	Col14 (I)	Col15 (D)	Col16 (I)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)
User Header	User Row ID	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition of resources	Schooling
1		Afghanistan	2015	Developing	65	263	62	0.01	71.27962362	65	1154	19.1	83	6	8.16	65	0.1	584.25921	33736494	17.2	17.3	0.479	10.1
2		Afghanistan	2014	Developing	59.9	271	64	0.01	73.52358168	62	492	18.6	86	58	8.18	62	0.1	612.696514	327582	17.5	17.5	0.476	10
3		Afghanistan	2013	Developing	59.9	268	66	0.01	73.21924272	64	430	18.1	89	62	8.13	64	0.1	631.744976	31731688	17.7	17.7	0.47	9.9
4		Afghanistan	2012	Developing	59.5	272	69	0.01	78.1842153	67	2787	17.6	93	67	8.52	67	0.1	669.959	3696958	17.9	18	0.463	9.8
5		Afghanistan	2011	Developing	59.2	275	71	0.01	7.097108703	68	3013	17.2	97	68	7.87	68	0.1	63.537231	2978599	18.2	18.2	0.454	9.5
6		Afghanistan	2010	Developing	58.8	279	74	0.01	79.67936736	66	1989	16.7	102	66	9.2	66	0.1	553.32894	2883167	18.4	18.4	0.448	9.2
7		Afghanistan	2009	Developing	58.6	281	77	0.01	56.76221682	63	2861	16.2	106	63	9.42	63	0.1	445.8932979	284331	18.6	18.7	0.434	8.9
8		Afghanistan	2008	Developing	58.1	287	80	0.03	25.87392536	64	1599	15.7	110	64	8.33	64	0.1	373.3611163	2729431	18.8	18.9	0.433	8.7
9		Afghanistan	2007	Developing	57.5	295	82	0.02	10.91015598	63	1141	15.2	113	63	6.73	63	0.1	369.835796	26616792	19	19.1	0.415	8.4
10		Afghanistan	2006	Developing	57.3	295	84	0.03	17.17151751	64	1990	14.7	116	58	7.43	58	0.1	272.56377	2589345	19.2	19.3	0.405	8.1
11		Afghanistan	2005	Developing	57.3	291	85	0.02	1.388647732	66	1296	14.2	118	58	8.7	58	0.1	25.2941299	257798	19.3	19.5	0.396	7.9
12		Afghanistan	2004	Developing	57	293	87	0.02	15.29606643	67	466	13.8	120	5	8.79	5	0.1	219.1413528	24118979	19.5	19.7	0.381	6.8
13		Afghanistan	2003	Developing	56.7	295	87	0.01	11.08905273	65	798	13.4	122	41	8.82	41	0.1	198.7285436	2364851	19.7	19.9	0.373	6.5
14		Afghanistan	2002	Developing	56.2	3	88	0.01	16.68735091	64	2486	13	122	36	7.76	36	0.1	187.84595	21979923	19.9	2.2	0.341	6.2
15		Afghanistan	2001	Developing	55.3	316	88	0.01	10.5747282	63	8762	12.6	122	35	7.8	33	0.1	117.49698	2966463	2.1	2.4	0.34	5.9

Step 2: Manipulate data

We will select all the columns to be used for predictions. On the menu click on *Data Transformation → Data Manipulation → Select Column(s)* and select all columns.



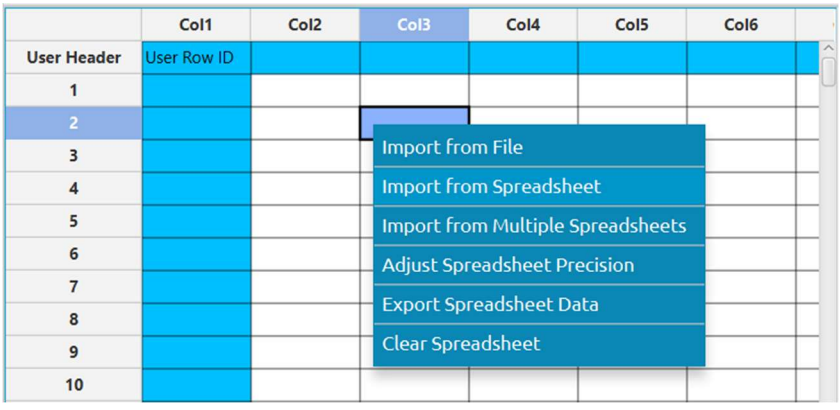
All the data will appear in the output (right) spreadsheet. This tab can be renamed "IMPORT" by right-clicking on it and choosing the "Rename" option.



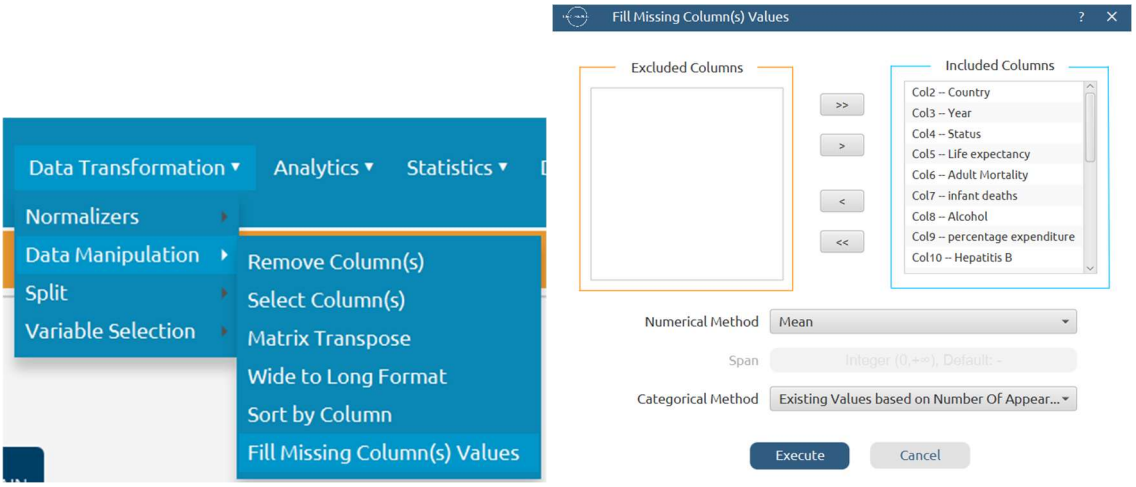
Step 3: Fill missing data

The dataset contains some missing values, so we will fill them appropriately depending on whether the corresponding factor is numerical of categorical. Create a new tab by pressing the "+" button on the bottom of the page with the name "FILL_MISSING" which we will use for filling the missing data.

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



Fill the missing values by choosing: *Data Transformation → Data Manipulation → Fill Missing Column(s) Values*



The results will appear on the output (right) spreadsheet.

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 the train and test set.

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

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

The results will be two separate spreadsheets, "TRAIN_TEST_SPLIT: Training Set" and "TRAIN_TEST_SPLIT: Test Set", which will be available to import into the next tabs.

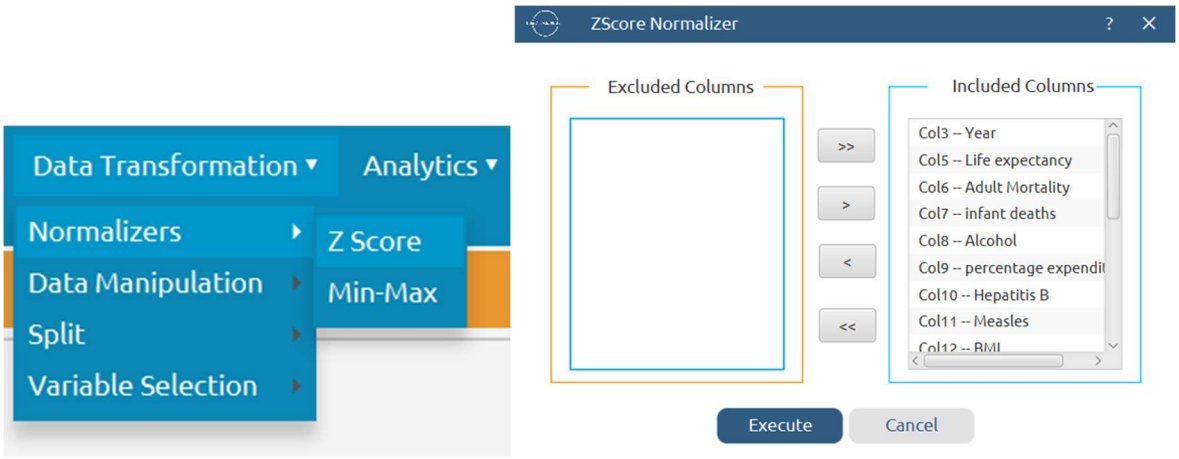
Step 5: Normalize the training set

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

Import 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".

	Col1	Col2 (S)	Col3 (I)	Col4 (S)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (D)	Col9 (D)	Col10 (I)	Col11 (I)	Col12 (D)	Col13 (I)	Col14 (I)	Col15 (D)	Col16 (I)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)
User Header	User Row ID	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition ...	Schooling
1		Afghanistan	2014	Developing	59.9	271	64	0.01	73.5235817	62	492	18.6	86	58	8.18	62	0.1	612.696514	327582	17.5	17.5	0.476	10
2		Afghanistan	2012	Developing	59.5	272	69	0.01	78.1842153	67	2787	17.6	93	67	8.52	67	0.1	669.959	3696958	17.9	18	0.463	9.8
3		Afghanistan	2010	Developing	58.8	279	74	0.01	79.6793674	66	1989	16.7	102	66	9.2	66	0.1	553.32894	2883167	18.4	18.4	0.448	9.2
4		Afghanistan	2009	Developing	58.6	281	77	0.01	56.7622168	63	2861	16.2	106	63	9.42	63	0.1	445.8932979	284331	18.6	18.7	0.434	8.9
5		Afghanistan	2008	Developing	58.1	287	80	0.03	25.8739254	64	1599	15.7	110	64	8.33	64	0.1	373.3611163	2729431	18.8	18.9	0.433	8.7
6		Afghanistan	2007	Developing	57.5	295	82	0.02	10.9101560	63	1141	15.2	113	63	6.73	63	0.1	369.835796	26616792	19	19.1	0.415	8.4
7		Afghanistan	2006	Developing	57.3	295	84	0.03	17.1715175	64	1990	14.7	116	58	7.43	58	0.1	272.56377	2589345	19.2	19.3	0.405	8.1
8		Afghanistan	2004	Developing	57	293	87	0.02	15.2960664	67	466	13.8	120	5	8.79	5	0.1	219.1413528	24118979	19.5	19.7	0.381	6.8
9		Afghanistan	2003	Developing	56.7	295	87	0.01	11.0890527	65	798	13.4	122	41	8.82	41	0.1	198.7285436	2364851	19.7	19.9	0.373	6.5
10		Afghanistan	2002	Developing	56.2	3	88	0.01	16.8873509	64	2486	13	122	36	7.76	36	0.1	187.84595	21979923	19.9	2.2	0.341	6.2
11		Albania	2015	Developing	77.8	74	0	4.6	364.9752287	99	0	58	0	99	6	99	0.1	3954.22783	28873	1.2	1.3	0.762	14.2
12		Albania	2014	Developing	77.5	8	0	4.51	428.7490668	98	0	57.2	1	98	5.88	98	0.1	4575.763787	288914	1.2	1.3	0.761	14.2
13		Albania	2013	Developing	77.2	84	0	4.76	430.8769785	99	0	56.5	1	99	5.66	99	0.1	4414.72314	289592	1.3	1.4	0.759	14.2
14		Albania	2012	Developing	76.9	86	0	5.14	412.4433563	99	9	55.8	1	99	5.59	99	0.1	4247.61438	2941	1.3	1.4	0.752	14.2
15		Albania	2011	Developing	76.6	88	0	5.37	437.0621	99	28	55.1	1	99	5.71	99	0.1	4437.17868	295195	1.4	1.5	0.738	13.3

Normalize the data using Z-score: *Data Transformation → Normalizers → Z Score* and select all columns except the "Life expectancy" target column.



The results will appear on the output spreadsheet.

	Col1	Col2 (S)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)
User Header	User Row ID	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition of resources	Schooling
1		Afghanistan	1.4009935	Developing	-0.9821559	0.8465961	0.2736577	-1.1773632	-0.3310132	-0.8047237	-0.1586791	-1.0075729	0.2639508	-1.0371383	0.9344303	-0.8641246	-0.3119128	-0.5262205	-0.2125619	2.8937125	2.8380444	-0.7404049	-0.6011758
2		Afghanistan	0.9664608	Developing	-1.0241775	0.8545961	0.3143709	-1.1773632	-0.3286563	-0.5892600	0.0421280	-1.0582113	0.3059976	-0.6559299	1.0752788	-0.6524678	-0.3119128	-0.5217364	-0.1564428	2.9850127	2.9502178	-0.8041869	-0.6630631
3		Afghanistan	0.5319280	Developing	-1.0977154	0.9105962	0.3550841	-1.1773632	-0.3279002	-0.6323527	-0.0276951	-1.1037858	0.3600578	-0.6982864	1.3569758	-0.6947992	-0.3119128	-0.5308694	-0.1699970	3.0991380	3.0399566	-0.8777615	-0.8487250
4		Afghanistan	0.3146616	Developing	-1.1187262	0.9265962	0.3795120	-1.1773632	-0.3394897	-0.7616309	0.0486028	-1.1291050	0.3840846	-0.8253558	1.4481131	-0.8217933	-0.3119128	-0.5392825	-0.2132822	3.1447881	3.1072606	-0.9464698	-0.9415559
5		Afghanistan	0.0973953	Developing	-1.1712533	0.9745962	0.4039400	-1.1721731	-0.3551103	-0.7185382	-0.0618192	-1.1544242	0.4081113	-0.7829994	0.9965694	-0.7794619	-0.3119128	-0.5449623	-0.1725576	3.1904382	3.1521300	-0.9513761	-1.0034432
6		Afghanistan	-0.1198711	Developing	-1.2342858	1.0385963	0.4202252	-1.1747681	-0.3626776	-0.7616309	-0.1018931	-1.1797434	0.4261314	-0.8253558	0.3337529	-0.8217933	-0.3119128	-0.5452384	0.2253009	3.2360883	3.1969993	-1.0396897	-1.0962742
7		Afghanistan	-0.3371375	Developing	-1.2552966	1.0385963	0.4365105	-1.1721731	-0.3595112	-0.7185382	-0.0276076	-1.2050626	0.4441514	-1.0371383	0.6237351	-1.0334501	-0.3119128	-0.5528556	-0.1748908	3.2817384	3.2418687	-1.0887527	-1.1891051
8		Afghanistan	-0.7716702	Developing	-1.2868128	1.0225963	0.4609384	-1.1747681	-0.3604596	-0.5892600	-0.1609541	-1.2506371	0.4681782	-3.2820324	1.1871291	-3.2770128	-0.3119128	-0.5570390	0.1836983	3.3502136	3.3316074	-1.2065041	-1.5913725
9		Afghanistan	-0.9889366	Developing	-1.3183291	1.0385963	0.4609384	-1.1773632	-0.3625872	-0.6754455	-0.1319048	-1.2708924	0.4801915	-1.7571987	1.1995569	-1.7530834	-0.3119128	-0.5586375	-0.1786299	3.3958637	3.3764768	-1.2457546	-1.6842035
10		Afghanistan	-1.2062030	Developing	-1.3708561	-1.2974060	0.4690811	-1.1773632	-0.3596549	-0.7185382	0.0157912	-1.2911478	0.4801915	-1.9689811	0.7604410	-1.9647403	-0.3119128	-0.5594897	0.1480710	3.4415138	-0.5944619	-1.4027564	-1.7770344
11		Albania	1.6182599	Developing	0.8983129	-0.7294054	-0.2474712	0.0137605	-0.1836225	0.7897077	-0.2017280	0.9875788	-0.2526242	0.6994779	0.0313429	0.7021361	-0.3119128	-0.2645519	-0.2175370	-0.8267715	-0.7963741	0.6627989	0.6984574
12		Albania	1.4009935	Developing	0.8667966	-1.2574059	-0.2474712	-0.0095949	-0.1513713	0.7466150	-0.2017280	0.9470681	-0.2466176	0.6571214	-0.0183684	0.6598047	-0.3119128	-0.2158807	-0.2132059	-0.8267715	-0.7963741	0.6578926	0.6984574
13		Albania	1.1837272	Developing	0.8352804	-0.6494053	-0.2474712	0.0552811	-0.1502952	0.7897077	-0.2017280	0.9116213	-0.2466176	0.6994779	-0.1095057	0.7021361	-0.3119128	-0.2284915	-0.2131946	-0.8039464	-0.7739394	0.6480800	0.6984574
14		Albania	0.9664608	Developing	0.8037641	-0.6334053	-0.2474712	0.1538927	-0.1596173	0.7897077	-0.2009405	0.8761744	-0.2466176	0.6994779	-0.1385039	0.7021361	-0.3119128	-0.2415774	-0.2179689	-0.8039464	-0.7739394	0.6137358	0.6984574
15		Albania	0.7491944	Developing	0.7722479	-0.6174053	-0.2474712	0.2135786	-0.1471673	0.7897077	-0.1992780	0.8407275	-0.2466176	0.6994779	-0.0887926	0.7021361	-0.3119128	-0.2267330	-0.2131013	-0.7811214	-0.7515047	0.5450475	0.4199646

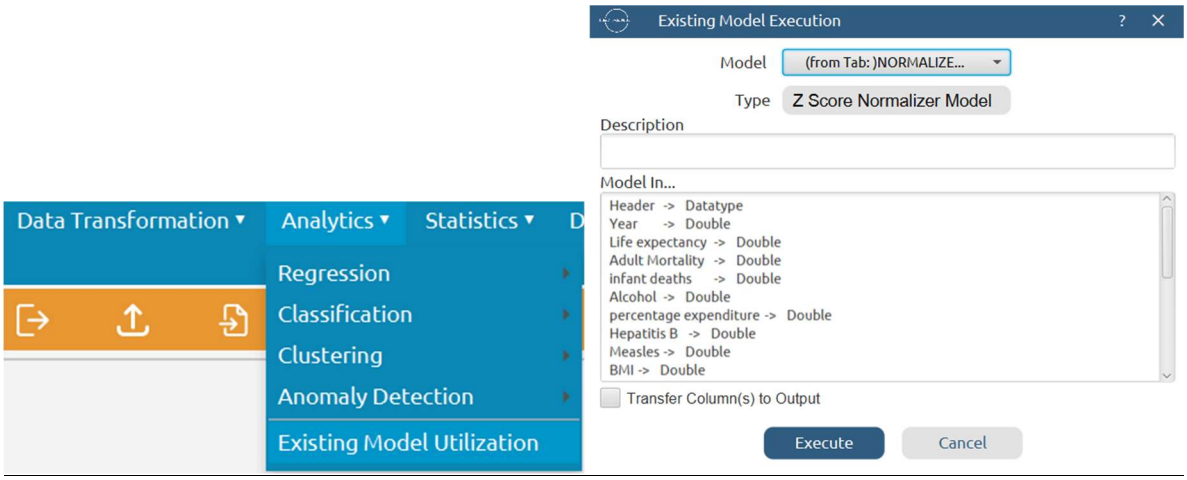
Step 6: Normalize the test set

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

Import 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".

	Col1	Col2 (S)	Col3 (I)	Col4 (S)	Col5 (D)	Col6 (I)	Col7 (I)	Col8 (D)	Col9 (D)	Col10 (I)	Col11 (I)	Col12 (D)	Col13 (I)	Col14 (I)	Col15 (D)	Col16 (I)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)
User Header	User Row ID	Country	Year	Status	Life expectancy	Adult Mortality	Infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition ...	Schooling
1		Afghanistan	2015	Developing	65	263	62	0.01	71.2796236	65	1154	19.1	83	6	8.16	65	0.1	584.25921	33736494	17.2	17.3	0.479	10.1
2		Afghanistan	2013	Developing	59.9	268	66	0.01	73.2192427	64	430	18.1	89	62	8.13	64	0.1	631.744976	31731688	17.7	17.7	0.47	9.9
3		Afghanistan	2011	Developing	59.2	275	71	0.01	7.0971087	68	3013	17.2	97	68	7.87	68	0.1	63.537231	2978599	18.2	18.2	0.454	9.5
4		Afghanistan	2005	Developing	57.3	291	85	0.02	1.3886477	66	1296	14.2	118	58	8.7	58	0.1	25.2941299	257798	19.3	19.5	0.396	7.9
5		Afghanistan	2001	Developing	55.3	316	88	0.01	10.5747282	63	8762	12.6	122	35	7.8	33	0.1	117.49698	2966463	2.1	2.4	0.34	5.9
6		Afghanistan	2000	Developing	54.8	321	88	0.01	10.42496	62	6532	12.2	122	24	8.2	24	0.1	114.56	293756	2.3	2.5	0.338	5.5
7		Algeria	2013	Developing	75.3	112	21	0.53	544.4507432	95	25	57.2	24	95	7.12	95	0.1	5471.866766	38338562	5.9	5.8	0.737	14.4
8		Algeria	2012	Developing	75.1	113	21	0.66	555.9260834	95	18	56.1	24	95	6.14	95	0.1	5564.82566	37565847	5.9	5.8	0.732	14.4
9		Algeria	2007	Developing	73.8	129	20	0.44	320.3239241	9	0	5.8	23	95	3.82	95	0.1	3935.183343	34376	6	5.9	0.69	12.3
10		Algeria	2003	Developing	71.7	146	20	0.34	25.0185226	81	15374	47	23	87	3.6	87	0.1	294.33556	3243514	6.3	6.1	0.663	11.5
11		Algeria	2000	Developing	71.3	145	21	0.25	154.4559436	81	0	44.4	25	86	3.49	86	0.1	1757.17797	3118366	6.5	6.4	0.636	10.7
12		Angola	2012	Developing	56	358	72	8.24	256.1225243	75	4458	21.5	110	75	3.3	75	2.6	4598.249988	259615	8.8	8.6	0.508	10.3
13		Angola	2011	Developing	51	361	75	8.06	239.8913921	72	1449	21	115	73	3.38	71	2.5	4299.12889	24218565	8.9	8.8	0.495	9.4
14		Angola	2002	Developing	46.5	391	96	2.82	24.037942	81	11945	16.3	157	37	3.63	41	2.3	711.181716	17572649	1.5	1.5	0.401	5.5
15		Antigua and Barbuda	2015	Developing	76.4	13	0	4.6029	0	99	0	47.7	0	66	5.9382	99	0.2	13566.9541	12753375.120 1	3.3	3.3	0.784	13.9

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



The results will appear on the output spreadsheet.

	Col1	Col2 (S)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)
User Header	User Row ID	Country	Year	Status	Life expectancy	Adult Mortality	Infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition ...	Schooling
1		Afghanistan	1.6182599	Developing	-0.4463799	0.7825960	0.2573724	-1.1773632	-0.3321480	-0.6754455	-0.1007557	-0.9822537	0.2459308	-3.2396759	0.9261451	-0.7371305	-0.3119128	-0.5284473	0.3438839	2.8252373	2.7931751	-0.7256860	-0.5702321
2		Afghanistan	1.1837272	Developing	-0.9821559	0.8225961	0.2899430	-1.1773632	-0.3311671	-0.7185382	-0.1641040	-1.0328921	0.2819709	-0.8677123	0.9137173	-0.7794619	-0.3119128	-0.5247288	0.3104926	2.9393626	2.8829138	-0.7698428	-0.6321194
3		Afghanistan	0.7491944	Developing	-1.0556938	0.8785961	0.3306562	-1.1773632	-0.3646059	-0.5461672	0.0619025	-1.0784666	0.3300244	-0.6135734	0.8060096	-0.6101364	-0.3119128	-0.5692240	-0.1684076	3.0534879	2.9950872	-0.8483437	-0.7558940
4		Afghanistan	-0.5544039	Developing	-1.2552966	1.0065962	0.4446532	-1.1747681	-0.3674928	-0.6323527	-0.0883310	-1.2303817	0.4561648	-1.0371383	1.1498457	-1.0334501	-0.3119128	-0.5722188	-0.2137242	3.3045635	3.2867381	-1.1329095	-1.2509924
5		Afghanistan	-1.4234694	Developing	-1.4654048	1.2065964	0.4690811	-1.1773632	-0.3628473	-0.7616309	0.5649266	-1.3114031	0.4801915	-2.0113376	0.7770114	-2.0917344	-0.3119128	-0.5649985	-0.1686097	-0.6213460	-0.5495926	-1.4076627	-1.8698654
6		Afghanistan	-1.6407357	Developing	-1.5179319	1.2465965	0.4690811	-1.1773632	-0.3629230	-0.8047237	0.3698068	-1.3316585	0.4801915	-2.4772591	0.9427155	-2.4727167	-0.3119128	-0.5652285	-0.2131252	-0.5756959	-0.5271579	-1.4174753	-1.9936400
7		Algeria	1.1837272	Developing	0.6356776	-0.4254051	-0.0764758	-1.0424210	-0.0928595	0.6173367	-0.1995405	0.9470681	-0.1084638	0.5300519	0.4953144	0.5328106	-0.3119128	-0.1457087	0.4205341	0.2460061	0.2131866	0.5401412	0.7603447
8		Algeria	0.9664608	Developing	0.6146667	-0.4174051	-0.0764758	-1.0086855	-0.0870563	0.6173367	-0.2001530	0.8913659	-0.1084638	0.5300519	0.0893393	0.5328106	-0.3119128	-0.1384293	0.4076641	0.2460061	0.2131866	0.5156097	0.7603447
9		Algeria	-0.1198711	Developing	0.4780964	-0.2894050	-0.0846184	-1.0657764	-0.2062032	-3.0886389	-0.2017280	-1.6557440	-0.1144704	0.5300519	-0.8717446	0.5328106	-0.3119128	-0.2660433	-0.2174454	0.2688312	0.2356213	0.3095448	0.1105281
10		Algeria	-0.9889366	Developing	0.2574827	-0.1534049	-0.0846184	-1.0917268	-0.3555429	0.0140384	1.1434611	0.4305568	-0.1144704	0.1912000	-0.9628819	0.1941596	-0.3119128	-0.3511507	-0.1639952	0.3373063	0.2804907	0.1770745	-0.1370211
11		Algeria	-1.6407357	Developing	0.2154611	-0.1614049	-0.0764758	-1.1150822	-0.2900847	0.0140384	-0.2017280	0.2988970	-0.1024571	0.1488435	-1.0084505	0.1518283	-0.3119128	-0.4365985	-0.1660797	0.3829564	0.3477947	0.0446042	-0.3845702
12		Angola	0.9664608	Developing	-1.3918669	1.5425968	0.3387988	0.9583553	-0.2386707	-0.2445161	0.1883366	-0.8607216	0.4081113	-0.3170779	-1.0871600	-0.3138168	0.1645606	-0.2141199	-0.2136939	0.9079327	0.8413577	-0.5834031	-0.5083448
13		Angola	0.7491944	Developing	-1.9171375	1.5665968	0.3632268	0.9116445	-0.2468790	-0.3737963	-0.0749439	-0.8860408	0.4381447	-0.4017909	-1.0540191	-0.4831423	0.1455016	-0.2375434	0.1853570	0.9307578	0.8862271	-0.6471851	-0.7868377
14		Angola	-1.2062030	Developing	-2.3898811	1.8065970	0.5342222	-0.4481567	-0.3560387	0.0140384	0.8434316	-1.1240412	0.6904256	-1.9266247	-0.9504541	-1.7530834	0.1073838	-0.5185083	0.0746652	-0.7582963	-0.7515047	-1.1083780	-1.9936400
15		Antigua and Barbuda	1.6182599	Developing	0.7512371	-1.2174059	-0.2474712	0.0145130	-0.3681950	0.7897077	-0.2017280	0.4660036	-0.2526242	0.1488435	0.0057416	0.7021361	-0.2928538	0.4882011	-0.0056027	-0.3474453	-0.3476804	0.7707377	0.6056265

Step 7: 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 "NORMALIZE_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

Use the k-Nearest Neighbors (kNN) method to train and fit the model: *Analytics → Regression → k-Nearest Neighbors (kNN)*

The predictions will appear on the output spreadsheet.

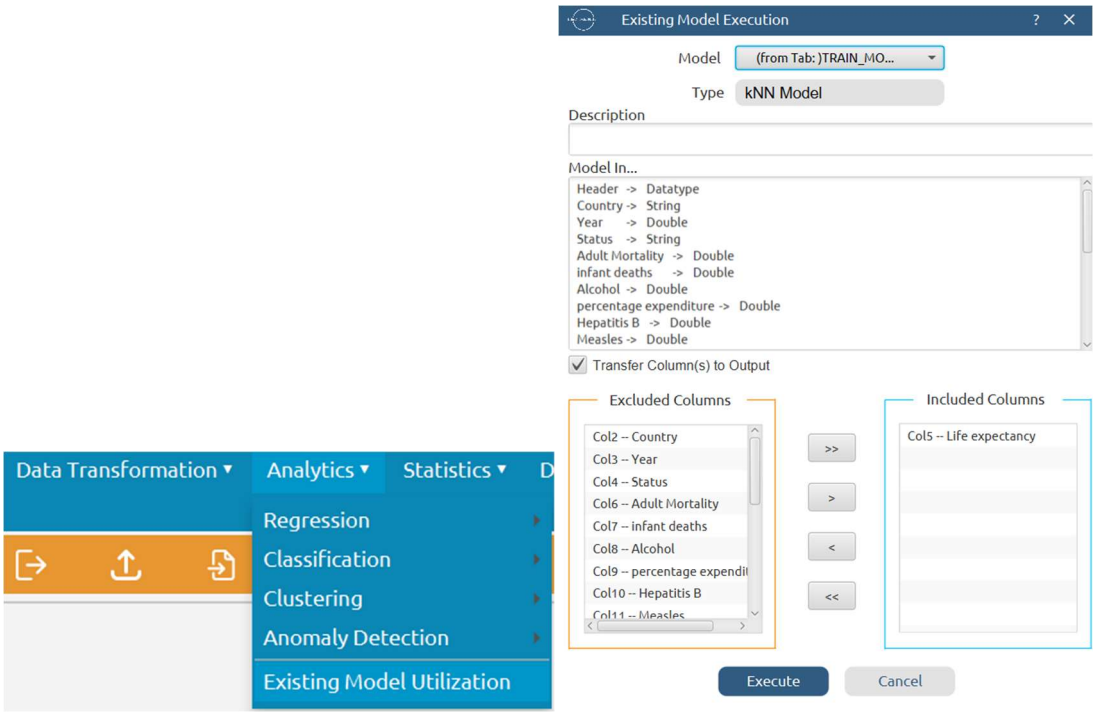
	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)	Col8 (S)	Col9 (D)	Col10 (S)	Col11 (D)	Col12 (S)	Col13 (D)
User Header	User Row ID	Life expectancy	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5
1		-0.9821559	-0.9835758	Entry 1	0.0	Entry 2	0.1826497	Entry 3	0.3004219	Entry 4	0.3588490	Entry 5	0.4211178
2		-1.0241775	-1.0253480	Entry 2	0.0	Entry 3	0.1468715	Entry 1	0.1826497	Entry 4	0.2291854	Entry 5	0.2847582
3		-1.0977154	-1.0984239	Entry 3	0.0	Entry 4	0.0903680	Entry 2	0.1468715	Entry 5	0.1527980	Entry 6	0.2629090
4		-1.1187262	-1.1191531	Entry 4	0.0	Entry 3	0.0903680	Entry 5	0.0961725	Entry 6	0.2130117	Entry 2	0.2291854
5		-1.1712533	-1.1712245	Entry 5	0.0	Entry 4	0.0961725	Entry 6	0.1218845	Entry 3	0.1527980	Entry 7	0.1755262
6		-1.2342858	-1.2329360	Entry 6	0.0	Entry 7	0.1118158	Entry 5	0.1218845	Entry 4	0.2130117	Entry 3	0.2629090
7		-1.2552966	-1.2536430	Entry 7	0.0	Entry 6	0.1118158	Entry 5	0.1755262	Entry 4	0.2518131	Entry 3	0.3227616
8		-1.2868128	-1.2868686	Entry 8	0.0	Entry 9	0.5355506	Entry 7	0.8000838	Entry 10	0.8788337	Entry 6	0.8915512
9		-1.3183291	-1.3176119	Entry 9	0.0	Entry 7	0.3431594	Entry 6	0.4521938	Entry 5	0.4954257	Entry 8	0.5355506
10		-1.3708561	-1.3704046	Entry 10	0.0	Entry 9	0.7515820	Entry 7	0.8452388	Entry 8	0.8788337	Entry 6	0.8986636
11		0.8983129	0.8967173	Entry 11	0.0	Entry 12	0.1154304	Entry 13	0.1371760	Entry 14	0.2063361	Entry 15	0.2782994
12		0.8667966	0.8660983	Entry 12	0.0	Entry 11	0.1154304	Entry 13	0.1277026	Entry 14	0.1780298	Entry 15	0.2412688
13		0.8352804	0.8351242	Entry 13	0.0	Entry 14	0.0710297	Entry 12	0.1277026	Entry 11	0.1371760	Entry 15	0.1472139
14		0.8037641	0.8037385	Entry 14	0.0	Entry 13	0.0710297	Entry 15	0.0830090	Entry 16	0.1657325	Entry 12	0.1780298
15		0.7722479	0.7722431	Entry 15	0.0	Entry 14	0.0830090	Entry 16	0.0912021	Entry 13	0.1472139	Entry 17	0.1492203

Step 8: 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 "NORMALIZE_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

To validate the model: *Analytics → Existing Model Utilization→ Model (from Tab:) TRAIN_MODEL(.fit)*. Choose the column "Life expectancy" to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

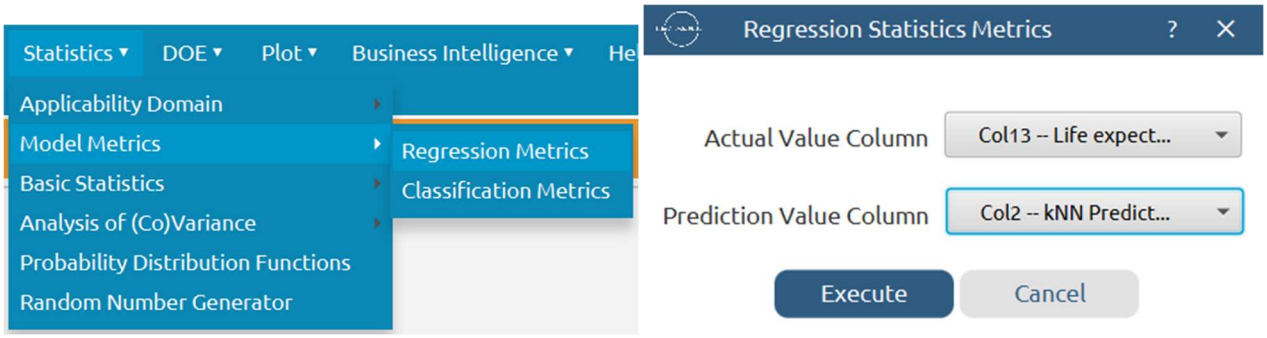
	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)	Col11 (S)	Col12 (D)	Col13 (D)
User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5	Life expectancy
1		-1.0710328	Entry 1	0.5484513	Entry 2	0.6692436	Entry 3	0.7169537	Entry 4	0.7295122	Entry 5	0.7746331	-0.4463799
2		-1.0431921	Entry 1	0.0885883	Entry 2	0.1024902	Entry 3	0.2249559	Entry 4	0.2914365	Entry 5	0.3494341	-0.9821559
3		-1.0715188	Entry 2	0.0824046	Entry 3	0.1112022	Entry 4	0.1903402	Entry 5	0.2227860	Entry 1	0.2484389	-1.0556938
4		-1.2272089	Entry 7	0.1039535	Entry 6	0.1988734	Entry 5	0.2299472	Entry 4	0.2924381	Entry 9	0.2939299	-1.2552966
5		-1.2980932	Entry 10	0.7815906	Entry 9	0.9059494	Entry 7	0.9954774	Entry 8	1.0002534	Entry 6	1.0437404	-1.4654048
6		-1.3898225	Entry 10	0.8079307	Entry 9	0.9386831	Entry 8	0.9565043	Entry 7	1.0767070	Entry 799	1.1128704	-1.5179319
7		0.6068112	Entry 28	0.1660018	Entry 29	0.1751787	Entry 30	0.2402264	Entry 31	0.3018280	Entry 27	0.3048206	0.6356776
8		0.5807076	Entry 29	0.0873281	Entry 30	0.1557503	Entry 28	0.2154119	Entry 31	0.2233825	Entry 32	0.3202332	0.6146667
9		0.3392586	Entry 33	0.5157859	Entry 34	0.9229403	Entry 35	0.9287926	Entry 36	0.9452042	Entry 37	0.9624389	0.4780964
10		0.3034417	Entry 36	0.0868685	Entry 37	0.1575971	Entry 34	0.1579124	Entry 35	0.2054402	Entry 32	0.3777740	0.2574827
11		0.2791587	Entry 37	0.0881479	Entry 36	0.1413115	Entry 35	0.3320029	Entry 34	0.3509622	Entry 32	0.5722366	0.2154611
12		-1.9973706	Entry 40	0.1433306	Entry 39	0.2354864	Entry 41	0.2787044	Entry 43	0.3332800	Entry 44	0.3829600	-1.3918669
13		-2.0286322	Entry 40	0.2104317	Entry 43	0.2425495	Entry 41	0.2553180	Entry 39	0.2659054	Entry 44	0.3048817	-1.9171375
14		-2.3423489	Entry 46	0.3912395	Entry 49	0.4998124	Entry 48	0.5271789	Entry 45	0.5415828	Entry 47	0.5439920	-2.3898811
15		0.6847632	Entry 51	0.3503198	Entry 52	0.3681257	Entry 53	0.4018255	Entry 54	0.4526655	Entry 55	0.4751613	0.7512371

Step 9: 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".

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Regression Metrics*



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared
1		0.0478849	0.2188261	0.1283504	0.9516540

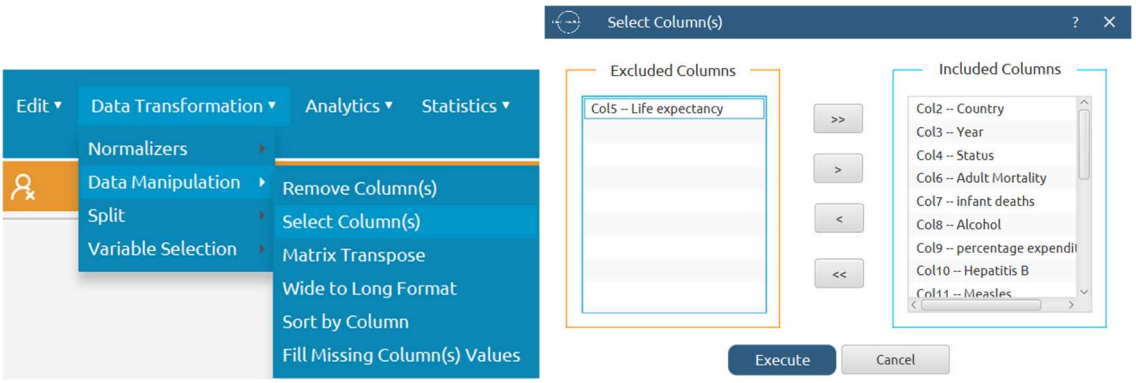
Step 10: Reliability check for each record of the test set

Step 10.a: Create the domain

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

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

Manipulate the data to exclude the target column "Life expectancy": *Data Transformation → Data Manipulation → Select Column(s)*

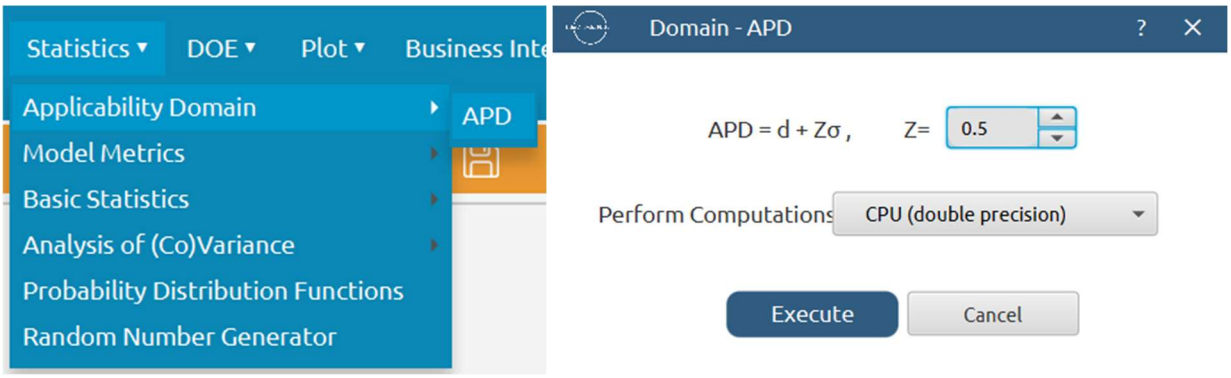


The results will appear on the output spreadsheet.

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_LIFE" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

Create the domain: *Statistics → Applicability Domain → APD*



The results will appear on the output spreadsheet.

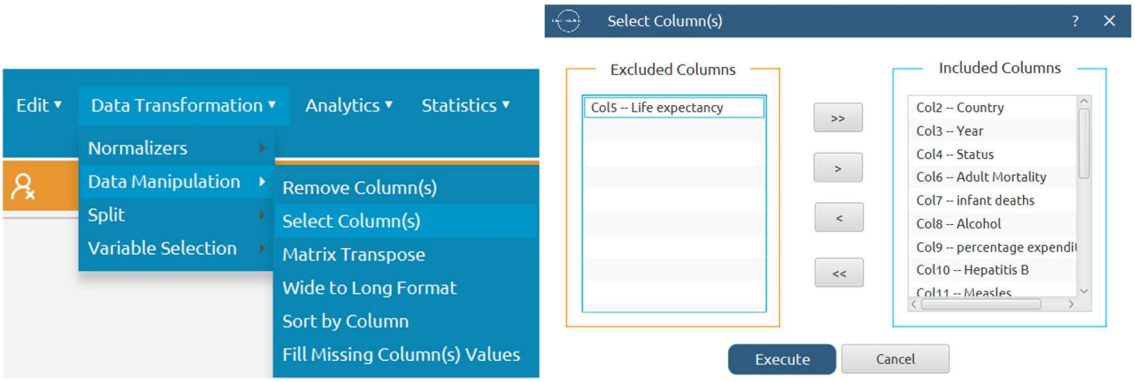
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	4.4021455	reliable
2		0.0	4.4021455	reliable
3		0.0	4.4021455	reliable
4		0.0	4.4021455	reliable
5		0.0	4.4021455	reliable
6		0.0	4.4021455	reliable
7		0.0	4.4021455	reliable
8		0.0	4.4021455	reliable
9		0.0	4.4021455	reliable
10		0.0	4.4021455	reliable
11		0.0	4.4021455	reliable
12		0.0	4.4021455	reliable
13		0.0	4.4021455	reliable
14		0.0	4.4021455	reliable
15		0.0	4.4021455	reliable

Step 10.b: Check the test set reliability

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

Import data into the input spreadsheet of the "EXCLUDE_LIFE_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".

Manipulate the data to exclude the target column "Life expectancy": *Data Transformation → Data Manipulation → Select Column(s)*

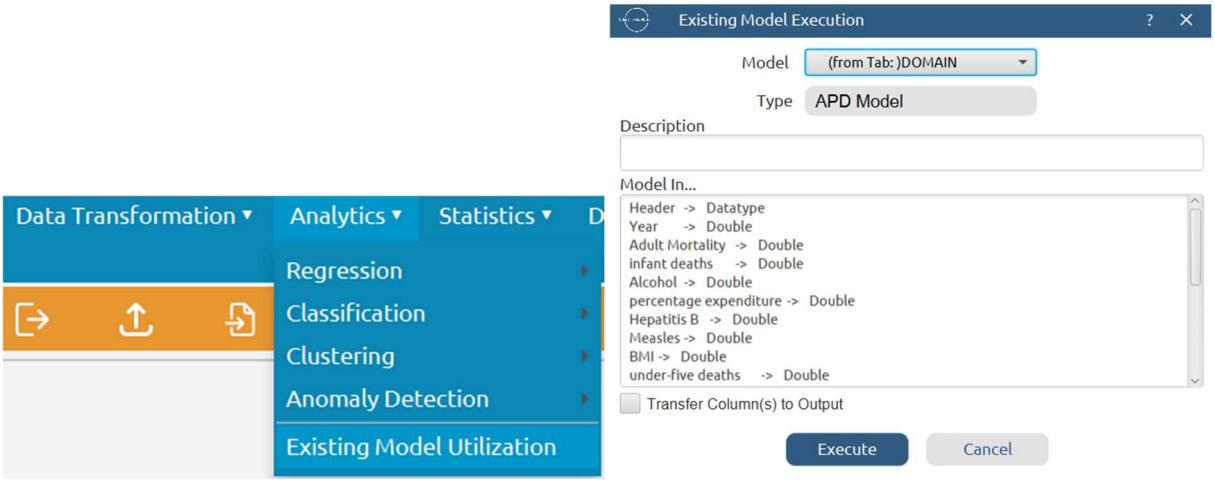


The results will appear on the output spreadsheet.

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

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

Check the Reliability: *Analytics → Existing Model Utilization → Model (from Tab:) DOMAIN*



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		2.2929200	4.4021455	reliable
2		0.6101480	4.4021455	reliable
3		0.3851409	4.4021455	reliable
4		0.5892839	4.4021455	reliable
5		2.1604021	4.4021455	reliable
6		2.5673226	4.4021455	reliable
7		0.8929959	4.4021455	reliable
8		0.4396590	4.4021455	reliable
9		2.1135013	4.4021455	reliable
10		0.8854990	4.4021455	reliable
11		0.6239087	4.4021455	reliable
12		0.8428629	4.4021455	reliable
13		1.0295712	4.4021455	reliable
14		1.4288277	4.4021455	reliable
15		1.0239509	4.4021455	reliable

Final Isalos Workflow

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