



Insurance Charges (Regression)

The goal of this study is to train a model in order to predict insurance charges. The dataset used in this case study is found in <https://www.kaggle.com/datasets/thedevastator/prediction-of-insurance-charges-using-age-gender/data> and has 8 features and 1338 labelled samples. This dataset contains detailed information about insurance customers, including their age, sex, body mass index (BMI), number of children, smoking status and region. Having access to such valuable insights allows analysts to get a better view into customer behaviour and the factors that contribute to their insurance charges.

The dataset contains no missing values and includes several categorical features. Categorical features contain multiple levels, and the data was transformed to corresponding numeric codes, as detailed below:

Gender:

- Female (0)
- Male (1)

Smoker:

- No (0)
- Yes (1)

Region:

- Northwest (0)
- Southeast (1)
- Northeast (2)
- Southwest (3)

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 salary data.

The screenshot displays the Isalos Analytics Platform interface. At the top, a menu bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu, a blue 'IMPORT' button is visible. The main workspace is divided into two panes. The left pane shows a table with 12 rows and 6 columns (Col1 to Col6). The 'User Header' row is highlighted in blue. A context menu is open over the table, offering options: 'Import from Spreadsheet', 'Import from file', 'Export Spread Sheet Data', and 'Clear Spreadsheet'. The right pane shows a table with 21 rows and 8 columns (Col1 to Col8). The 'User Header' row is also highlighted in blue. The bottom of the interface features a status bar with an 'IMPORT' button and a '+' icon.

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.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)
1	0	19	0	27.9	0	1	3	16884.92
2	1	18	1	33.77	1	0	1	1725.552
3	2	28	1	33	3	0	1	4449.462
4	3	33	1	22.705	0	0	0	21984.47
5	4	32	1	28.88	0	0	0	3866.855
6	5	31	0	25.74	0	0	1	3756.621
7	6	46	0	33.44	1	0	1	8240.589
8	7	37	0	27.74	3	0	0	7281.505
9	8	37	1	29.83	2	0	2	6406.410
10	9	60	0	25.84	0	0	0	28923.13
11	10	25	1	26.22	0	0	2	2721.320
12	11	62	0	26.29	0	1	1	27808.72
13	12	23	1	34.4	0	0	3	1826.843
14	13	56	0	39.82	0	0	1	11090.71
15	14	27	1	42.13	0	1	1	39611.75
16	15	19	1	24.6	1	0	3	1837.237
17	16	52	0	30.78	1	0	2	10797.33
18	17	23	1	23.845	0	0	2	2395.171
19	18	56	1	40.3	0	0	3	10602.38
20	19	30	1	35.3	0	1	3	36837.46
21	20	60	0	36.005	0	0	2	13228.84

The data will appear in the output spreadsheet.

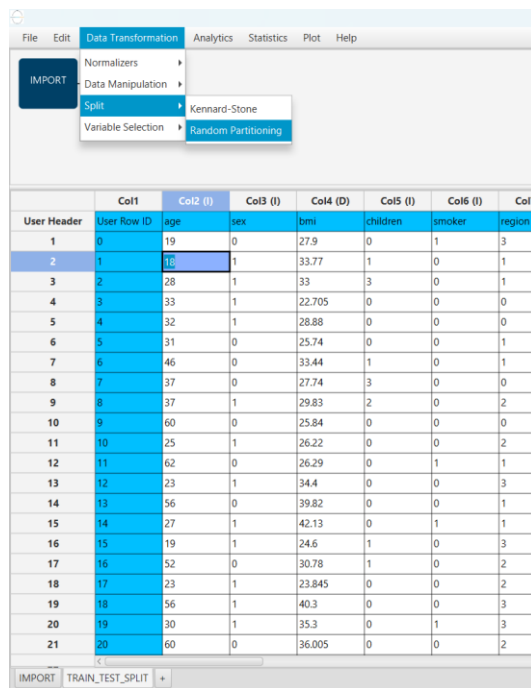
Step 3: Split data

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

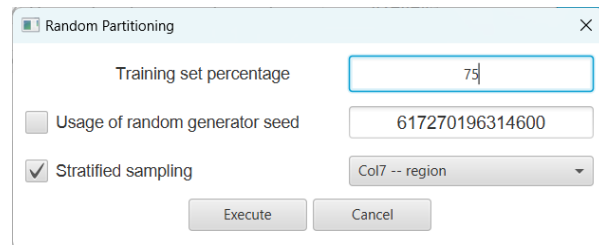
User Header	Col1	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)	Col9
1	0	19	0	27.9	0	1	3	16884.924	
2	1	18	1	33.77	1	0	1	1725.5523	
3	2	28	1	33	3	0	1	4449.462	
4	3	33	1	22.705	0	0	0	21984.47061	
5	4	32	1	28.88	0	0	0	3866.8552	
6	5	31	0	25.74	0	0	1	3756.6216	
7	6	46	0	33.44	1	0	1	8240.5896	
8	7	37	0	27.74	3	0	0	7281.5056	
9	8	37	1	29.83	2	0	2	6406.4107	
10	9	60	0	25.84	0	0	0	28923.13692	
11	10	25	1	26.22	0	0	2	2721.3208	
12	11	62	0	26.29	0	1	1	27808.7251	
13	12	23	1	34.4	0	0	3	1826.843	
14	13	56	0	39.82	0	0	1	11090.7178	
15	14	27	1	42.13	0	1	1	39611.7577	
16	15	19	1	24.6	1	0	3	1837.237	
17	16	52	0	30.78	1	0	2	10797.3362	
18	17	23	1	23.845	0	0	2	2395.17155	
19	18	56	1	40.3	0	0	3	10602.385	
20	19	30	1	35.3	0	1	3	36837.467	
21	20	60	0	36.005	0	0	2	13228.84695	

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



The screenshot shows the 'Data Transformation' menu with 'Random Partitioning' selected. Below the menu is a table with 7 columns: User Header, User Row ID, age, sex, bmi, children, smoker, and region. The table contains 21 rows of data.

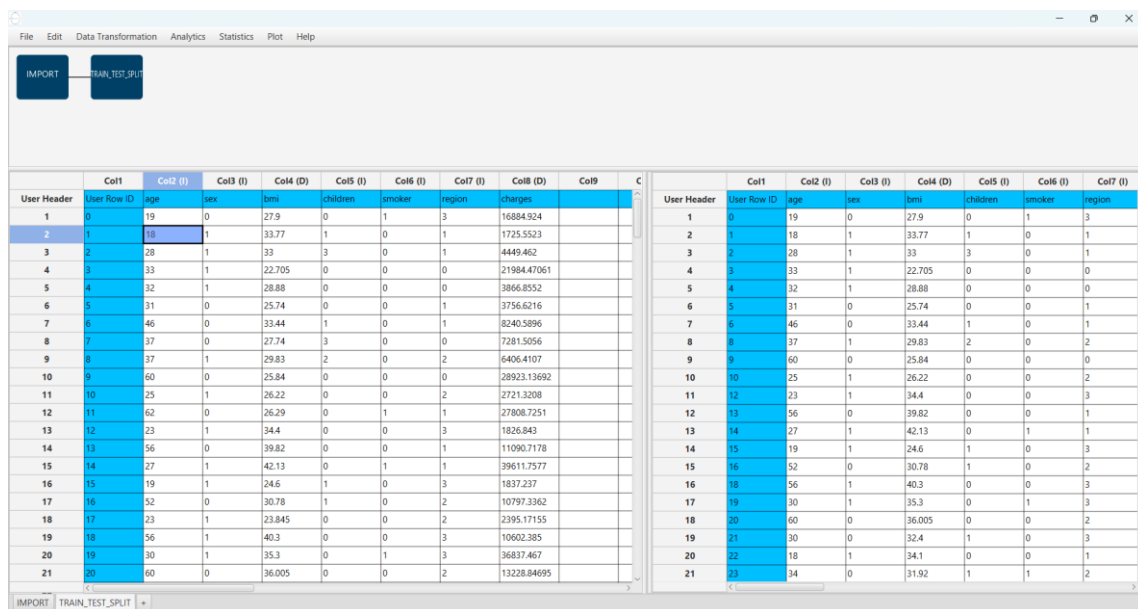
User Header	User Row ID	age	sex	bmi	children	smoker	region
1	0	19	0	27.9	0	1	3
2	1	18	1	33.77	1	0	1
3	2	28	1	33	3	0	1
4	3	33	1	22.705	0	0	0
5	4	32	1	28.88	0	0	0
6	5	31	0	25.74	0	0	1
7	6	46	0	33.44	1	0	1
8	7	37	0	27.74	3	0	0
9	8	37	1	29.83	2	0	2
10	9	60	0	25.84	0	0	0
11	10	25	1	26.22	0	0	2
12	11	62	0	26.29	0	1	1
13	12	23	1	34.4	0	0	3
14	13	56	0	39.82	0	0	1
15	14	27	1	42.13	0	1	1
16	15	19	1	24.6	1	0	3
17	16	52	0	30.78	1	0	2
18	17	23	1	23.845	0	0	2
19	18	56	1	40.3	0	0	3
20	19	30	1	35.3	0	1	3
21	20	60	0	36.005	0	0	2



The 'Random Partitioning' dialog box shows the following settings:

- Training set percentage: 75
- Usage of random generator seed: 617270196314600
- Stratified sampling: ☒ (checked)
- Col7 -- region (dropdown menu)
- Buttons: Execute, Cancel

The results will appear on the output spreadsheet.



The screenshot shows the 'TRAIN_TEST_SPLIT' tab with two output spreadsheets. The left spreadsheet contains the original data with an additional 'charges' column. The right spreadsheet contains the data after random partitioning, with the 'region' column highlighted in blue.

User Header	User Row ID	age	sex	bmi	children	smoker	region	charges
1	0	19	0	27.9	0	1	3	16884.924
2	1	18	1	33.77	1	0	1	1725.5523
3	2	28	1	33	3	0	1	4449.462
4	3	33	1	22.705	0	0	0	21984.47061
5	4	32	1	28.88	0	0	0	3866.8552
6	5	31	0	25.74	0	0	1	3756.6216
7	6	46	0	33.44	1	0	1	8240.5896
8	7	37	0	27.74	3	0	0	7281.5056
9	8	37	1	29.83	2	0	2	6406.4107
10	9	60	0	25.84	0	0	0	28923.13692
11	10	25	1	26.22	0	0	2	2721.3208
12	11	62	0	26.29	0	1	1	27808.7251
13	12	23	1	34.4	0	0	3	1826.843
14	13	56	0	39.82	0	0	1	11090.7178
15	14	27	1	42.13	0	1	1	39611.7577
16	15	19	1	24.6	1	0	3	1837.237
17	16	52	0	30.78	1	0	2	10797.3362
18	17	23	1	23.845	0	0	2	2395.17155
19	18	56	1	40.3	0	0	3	10602.385
20	19	30	1	35.3	0	1	3	36837.467
21	20	60	0	36.005	0	0	2	13228.84695

Step 4: 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"

The screenshot shows the 'Data Transformation' menu with 'Normalizers' selected, and the 'Z Score' option highlighted. The background displays a data table with 21 rows and 8 columns (User Row ID, age, sex, bmi, children, smoker, region, charges). The 'charges' column is highlighted in blue, indicating it is the target for normalization.

User Header	Col1 (I)	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)
1	0	19	0	27.9	0	1	3	16884.924
2	1	18	1	33.77	1	0	1	1725.5523
3	2	28	1	33	3	0	1	4449.462
4	3	33	1	22.705	0	0	0	21984.47051
5	4	32	1	28.88	0	0	0	3866.8552
6	5	31	0	25.74	0	0	1	3756.6216
7	6	46	0	33.44	1	0	1	8240.5896
8	8	37	1	29.83	2	0	2	6406.4107
9	9	60	0	25.84	0	0	0	28923.13692
10	10	25	1	26.22	0	0	2	2721.3208
11	12	23	1	34.4	0	0	3	1826.843
12	13	56	0	39.82	0	0	1	11090.7178
13	14	27	1	42.13	0	1	1	39611.7577
14	15	19	1	24.6	1	0	3	1837.237
15	16	52	0	30.78	1	0	2	10797.3362
16	18	56	1	40.3	0	0	3	10602.385
17	19	30	1	35.3	0	1	3	36837.467
18	20	60	0	36.005	0	0	2	13228.84695
19	21	30	0	32.4	1	0	3	4149.736
20	22	18	1	34.1	0	0	1	1137.011
21	23	34	0	31.92	1	1	2	37701.8768

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

The screenshot shows the 'ZScore Normalizer' dialog box. The 'Excluded Columns' list contains 'Col8 -- charges'. The 'Included Columns' list contains 'Col2 -- age', 'Col3 -- sex', 'Col4 -- bmi', 'Col5 -- children', 'Col6 -- smoker', and 'Col7 -- region'. The 'Execute' button is highlighted.

User Header	Col1 (I)	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)
1	0	19	0	27.9	0	1	3	16884.924
2	1	18	1	33.77	1	0	1	1725.5523
3	2	28	1	33	3	0	1	4449.462
4	3	33	1	22.705	0	0	0	21984.47051
5	4	32	1	28.88	0	0	0	3866.8552
6	5	31	0	25.74	0	0	1	3756.6216
7	6	46	0	33.44	1	0	1	8240.5896
8	8	37	1	29.83	2	0	2	6406.4107
9	9	60	0	25.84	0	0	0	28923.13692
10	10	25	1	26.22	0	0	2	2721.3208
11	12	23	1	34.4	0	0	3	1826.843
12	13	56	0	39.82	0	0	1	11090.7178
13	14	27	1	42.13	0	1	1	39611.7577
14	15	19	1	24.6	1	0	3	1837.237
15	16	52	0	30.78	1	0	2	10797.3362
16	18	56	1	40.3	0	0	3	10602.385
17	19	30	1	35.3	0	1	3	36837.467
18	20	60	0	36.005	0	0	2	13228.84695
19	21	30	0	32.4	1	0	3	4149.736
20	22	18	1	34.1	0	0	1	1137.011
21	23	34	0	31.92	1	1	2	37701.8768

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)
1	0	19	0	27.9	0	1	3	16884.924
2	1	18	1	33.77	1	0	1	1725.5523
3	2	28	1	33	3	0	1	4449.462
4	3	33	1	22.705	0	0	0	21984.47061
5	4	32	1	28.88	0	0	0	3866.8552
6	5	31	0	25.74	0	0	1	3756.6216
7	6	46	0	33.44	1	0	1	8240.5896
8	8	37	1	29.83	2	0	2	6406.4107
9	9	60	0	25.84	0	0	0	28923.13692
10	10	25	1	26.22	0	0	2	2721.3208
11	12	23	1	34.4	0	0	3	1826.843
12	13	56	0	39.82	0	0	1	11090.7178
13	14	27	1	42.13	0	1	1	39611.7577
14	15	19	1	24.6	1	0	3	1837.237
15	16	52	0	30.78	1	0	2	10797.3362
16	18	56	1	40.3	0	0	3	10602.385
17	19	30	1	35.3	0	1	3	36837.467
18	20	60	0	36.005	0	0	2	13228.84695
19	21	30	0	32.4	1	0	3	4149.736
20	22	18	1	34.1	0	0	1	1137.011
21	23	34	0	31.92	1	1	2	37701.8768

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	9	-1.442052194	-1.013537686	-0.448337111	-0.894088423	2.0358255719	1.3698195197	16884.924
2	1	0.451325	61993	49191124	5403234	795726	160479	1725.5523
3	3	-1.513797506	0.9856604221	0.5084179971	-0.072051081	-0.490711973	-0.438594415	4449.462
4	3	3752274	549418	769941	75050219	46850026	58298184	21984.47061
5	4	-0.796344383	0.9856604221	0.3829151975	1.5720236018	-0.490711973	-0.438594415	3866.8552
6	5	0.742793	549418	7306575	2914	46850026	58298184	3756.6216
7	6	-0.437617821	0.9856604221	-1.295073532	-0.894088423	-0.490711973	-1.342801383	8240.5896
8	8	42380545	549418	1963338	5403234	46850026	2324968	28923.13692
9	9	0.509363133	0.9856604221	-0.288606275	-0.894088423	-0.490711973	-1.342801383	2721.3208
10	10	7539002	549418	6323666	5403234	46850026	2324968	10797.3362
11	12	-0.581108446	-1.013537686	-0.800396912	-0.894088423	-0.490711973	-0.438594415	10602.385
12	13	0.83995	61993	9782544	5403234	46850026	58298184	36837.467
13	14	0.4950712388	-1.013537686	0.4546310830	-0.072051081	-0.490711973	-0.438594415	13228.84695
14	15	674268	61993	610241	75050219	46850026	58298184	4149.736
15	16	-0.150636572	0.9856604221	-0.133765159	0.7499662600	-0.490711973	-0.4656125520	1137.011
16	18	1034263	549418	2379103	39319	46850026	6653307	37701.8768
17	19	1.4995056114	-1.013537686	-0.784097848	-0.894088423	-0.490711973	-1.342801383	28923.13692
18	20	887537	61993	0946271	5403234	46850026	2324968	2721.3208
19	21	-1.011580320	0.9856604221	-0.722161401	-0.894088423	-0.490711973	-0.4656125520	10797.3362
20	22	0.645637	549418	5368447	5403234	46850026	6653307	10602.385
21	23	-1.155070944	0.9856604221	0.6111021059	-0.894088423	-0.490711973	-1.3698195197	1826.843
		7247534	549418	438434	5403234	46850026	160479	11090.7178
		1.2125243621	-1.013537686	1.4945114226	-0.894088423	-0.490711973	-0.438594415	39611.7577
		683748	61993	36427	5403234	46850026	58298184	4043742
		-0.868089695	0.9856604221	1.8710198214	-0.894088423	2.0358255719	-0.438594415	1826.843
		4043742	549418	482108	5403234	795726	58298184	11090.7178
		1.442052194	0.9856604221	-0.986206252	-0.072051081	-0.490711973	1.3698195197	1826.843
		0.451325	549418	6516017	75050219	46850026	160479	1137.011
		0.9255431128	-1.013537686	0.0210759571	-0.072051081	-0.490711973	0.4656125520	10797.3362
		470055	61993	6646633	75050219	46850026	6653307	37701.8768

Step 5: 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 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)
34	132	53	0	35.9	2	0	3	11163.568
35	124	20	0	28.785	0	0	2	2457.21115
36	139	22	0	36	0	0	3	2166.732
37	140	34	1	22.42	2	0	2	27375.90478
38	142	34	1	25.3	2	1	1	18972.495
39	143	29	1	29.735	2	0	0	18157.876
40	145	29	0	38.83	3	0	1	5138.2567
41	146	46	1	30.495	3	1	0	40720.55105
42	149	19	1	28.4	1	0	3	1842.519
43	150	35	1	24.13	1	0	0	5125.2157
44	151	48	1	29.7	0	0	1	7789.635
45	157	18	1	25.175	0	1	2	15518.18025
46	161	18	0	36.85	0	1	1	36149.4835
47	163	32	0	29.8	2	0	3	5152.134
48	166	20	0	37	5	0	3	4830.63
49	167	32	0	33.155	3	0	0	6128.79745
50	168	19	0	31.825	1	0	0	2719.27975
51	174	24	0	33.345	0	0	0	2855.43755
52	177	54	1	29.2	1	0	3	10436.096
53	182	22	1	19.95	3	0	2	4005.4225
54	188	41	0	32.2	1	0	3	6775.961
55	189	29	0	32.11	2	0	0	4922.9159

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

	Col1	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6
34	132	53	0	35.9	2	0
35	134	20	0	28.785	0	0
36	139	22	0	36	0	0
37	140	34	1	22.42	2	0
38	142	34	1	25.3	2	1
39	143	29	1	29.735	2	0
40	145	29	0	38.83	3	0
41	146	46	1	30.495	3	1
42	149	19	1	28.4	1	0
43	150	35	1	24.13	1	0
44	151	48	1	29.7	0	0
45	157	18	1	25.175	0	1
46	161	18	0	36.85	0	1
47	163	32	0	29.8	2	0
48	166	20	0	37	5	0
49	167	32	0	33.155	3	0
50	168	19	0	31.825	1	0
51	174	24	0	33.345	0	0
52	177	54	1	29.2	1	0
53	182	22	1	19.95	3	0
54	188	41	0	32.2	1	0
55	189	29	0	32.11	2	0

Existing Model Execution

Model: (from Tab:) NORMALIZE_TR...

Type: Z Score Normalizer Model

Description:

Model Input:

Header	Datatype
age	Double
sex	Double
bmi	Double
children	Double
smoker	Double
region	Double

☐ Transfer Column(s) to Output

Execute Cancel

The results will appear on the output spreadsheet.

	Col1	Col2 (I)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)	Col9	Col10
34	132	53	0	35.9	2	0	3	11163.568		
35	134	20	0	28.785	0	0	2	2457.21115		
36	139	22	0	36	0	0	3	2166.732		
37	140	34	1	22.42	2	0	2	27375.90478		
38	142	34	1	25.3	2	1	1	18972.495		
39	143	29	1	29.735	2	0	0	18157.876		
40	145	29	0	38.83	3	0	1	5138.2567		
41	146	46	1	30.495	3	1	0	40720.55105		
42	149	19	1	28.4	1	0	3	1842.519		
43	150	35	1	24.13	1	0	0	5125.2157		
44	151	48	1	29.7	0	0	1	7789.635		
45	157	18	1	25.175	0	1	2	15518.18025		
46	161	18	0	36.85	0	1	1	36149.4835		
47	163	32	0	29.8	2	0	3	5152.134		
48	166	20	0	37	5	0	3	4830.63		
49	167	32	0	33.155	3	0	0	6128.79745		
50	168	19	0	31.825	1	0	0	2719.27975		
51	174	24	0	33.345	0	0	0	2855.43755		
52	177	54	1	29.2	1	0	3	10436.096		
53	182	22	1	19.95	3	0	2	4005.4225		
54	188	41	0	32.2	1	0	3	6775.961		
55	189	29	0	32.11	2	0	0	4922.9159		

Step 6: 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 from the output of the "NORMALIZE_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

The screenshot shows the 'Data Transformation' menu with options: 'IMPORT', 'TRAIN_TEST_SPLIT', 'NORMALIZE_TRAIN_SET', 'NORMALIZE_TEST_SET', and 'FEATURE_SELECTION_REGRESSION'. Below the menu is a data spreadsheet with columns: User Header, User Row ID, age, sex, bmi, children, smoker, region, and charges. The data is organized into two tables, each with 15 rows of data.

User Header	User Row ID	age	sex	bmi	children	smoker	region	charges
1	0	-1.442052194	-1.013537686	-0.448337111	-0.894088423	2.0358255719	1.3698195197	16884.924
2	1	-1.513797506	0.9856604221	0.5084179971	-0.072051081	-0.490711973	-0.438594415	1725.5523
3	2	-0.796344383	0.9856604221	0.3829151975	1.5720236018	-0.490711973	-0.438594415	4449.462
4	3	-0.437617821	0.9856604221	-1.295073532	-0.894088423	-0.490711973	-1.342801383	21984.47061
5	4	42380545	549418	1963338	5403234	46850026	2324968	3866.8552
6	5	-0.581108446	-1.013537686	-0.800396912	-0.894088423	-0.490711973	-0.438594415	3756.6216
7	6	0.4950712388	-1.013537686	0.4546310830	-0.072051081	-0.490711973	-0.438594415	8240.5896
8	8	-0.150636572	0.9856604221	-0.133765159	0.7499862600	-0.490711973	0.4656125520	6406.4107
9	9	1.4995056114	-1.013537686	-0.784097848	-0.894088423	-0.490711973	-1.342801383	28923.13692
10	10	887537	61993	0946271	5403234	46850026	2324968	2721.3208
11	12	-1.011580320	0.9856604221	-0.722161401	-0.894088423	-0.490711973	0.4656125520	1826.843
12	13	1.2125243621	-1.013537686	1.4945114226	-0.894088423	-0.490711973	-0.438594415	11090.7178
13	14	-0.868089695	0.9856604221	1.8710198214	-0.894088423	2.0358255719	-0.438594415	39611.7577
14	15	-1.442052194	0.9856604221	-0.986206252	-0.072051081	-0.490711973	1.3698195197	1837.237
15	16	0.9255431128	-1.013537686	0.0210759571	-0.072051081	-0.490711973	0.4656125520	10797.3362

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

The screenshot shows the 'Regression Analysis Model' dialog box. The 'Significance Level (α)' is set to 0.05. The 'Select Intercept Column' dropdown is set to 'Col8 -- charges'. The 'Excluded Columns' list is empty, and the 'Included Columns' list contains: Col2 -- age, Col3 -- sex, Col4 -- bmi, Col5 -- children, Col6 -- smoker, and Col7 -- region. The 'Execute' button is highlighted.

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	0	-1.442052194	-1.013537686	-0.448337111	-0.894088423	2.0358255719	1.3698195197	16884.924
2	1	0451325	61993	49191124	5403234	795726	58298184	1725.5523
3	2	-1.513797506	0.9856604221	0.5084179971	-0.072051081	-0.490711973	-0.438594415	4449.462
4	3	3752274	549418	769941	75050219	46850026	58298184	21984.4706
5	4	-0.796344383	0.9856604221	0.3829151975	1.5720236018	-0.490711973	-0.438594415	3866.8552
6	5	0742793	549418	7306575	2914	46850026	58298184	3756.6216
7	6	-0.437617821	0.9856604221	-1.295073532	-0.894088423	-0.490711973	-1.342801383	8240.5896
8	7	42380545	549418	1963338	5403234	46850026	2324968	6406.4107
9	8	-0.509363133	0.9856604221	-0.288606275	-0.894088423	-0.490711973	-1.342801383	28923.1369
10	9	7539002	549418	6323666	5403234	46850026	2324968	1837.237
11	10	-0.581108446	-1.013537686	-0.800396912	-0.894088423	-0.490711973	-0.438594415	1090.7178
12	11	083995	61993	9782544	5403234	46850026	58298184	39611.7577
13	12	0.4950712388	-1.013537686	0.4546310830	-0.072051081	-0.490711973	-0.438594415	1837.237
14	13	674268	61993	610241	75050219	46850026	58298184	10797.3362
15	14	-0.150636572	0.9856604221	-0.133765159	0.749986260	-0.490711973	0.4656125520	10602.385
16	15	1034263	549418	2379103	39319	46850026	6653307	36837.467
17	16	1.4995056114	-1.013537686	-0.784097848	-0.894088423	-0.490711973	-1.342801383	13228.8469
18	17	887537	61993	0946271	5403234	46850026	2324968	
19	18	-1.011580320	0.9856604221	-0.722161401	-0.894088423	-0.490711973	0.4656125520	
20	19	0645637	549418	5368447	-0.894088423	-0.490711973	1.3698195197	
21	20	-1.155070944	0.9856604221	0.6111021059	-0.894088423	-0.490711973	1.3698195197	
22	21	7247534	549418	438344	5403234	46850026	160479	
23	22	1.2125243621	-1.013537686	1.4945114226	-0.894088423	-0.490711973	-0.438594415	
24	23	683748	61993	36427	5403234	46850026	58298184	
25	24	-0.868089695	0.9856604221	1.8710198214	-0.894088423	-0.490711973	-0.438594415	
26	25	4043742	549418	482108	5403234	795726	58298184	
27	26	1.442052194	0.9856604221	-0.886206252	-0.072051081	-0.490711973	1.3698195197	
28	27	0451325	549418	6516017	75050219	46850026	160479	
29	28	0.925431128	-1.013537686	0.0210795971	-0.072051081	-0.490711973	0.4656125520	
30	29	479955	61993	5654622	75050219	46850026	6653307	
31	30	1.2125243621	0.9856604221	1.5727469340	-0.894088423	-0.490711973	1.3698195197	
32	31	683748	549418	77836	5403234	46850026	160479	
33	32	-0.652853758	0.9856604221	0.7577936888	-0.894088423	-0.490711973	1.3698195197	
34	33	4140898	549418	964862	5403234	795726	58298184	
35	34	1.4995056114	-1.013537686	0.8727020973	-0.894088423	-0.490711973	0.4656125520	
36	35	887537	61993	260574	5403234	46850026	6653307	

User Header	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
1	Regression							
2	Statistics							
3	Multiple R	0.8654534384						
4	R Square	0.7490096541						
5	Adjusted R Square	0.7474991806						
6	Standard Error	239772						
7	Observations	1004						
8	Regression	df	SS	MS	F	Significance F		
9	Residual	6	1.0444945778	1.7408242963	495.87738958	3.7626178705		
10	Total	997	302507E11	837513E10	53771	99801E-295		
11		1003	1.3945006212	981787E11				
12	Coefficients	Standard Error	t Stat	P-value	Lower 95.0%	Upper 95.0%		
13	charges	12823.924104	186.99218663	68.57999706	0.0	12456.980690	13190.867517	
14	age	441232	319934	59962	0.0	907983	97448	
15	sex	3711.8504717	189.03870058	19.635442161	7.9196977388	3340.8990892	4082.8178541	
16	bmi	104914	50017	96855	3132E-73	27434	932393	
17	children	-87.68106927	187.90499132	-0.466624482	0.6408705114	-456.4157216	281.05358311	
18	smoker	809772	956933	17946806	475497	711408	494535	
19	region	1945.4451983	188.73848880	10.307623053	9.6343670392	1575.0749353	2315.8154613	
20		399272	766047	623633	65465E-24	883388	715155	
21		481.09377539	187.58999025	2.5446025928	0.0104748572	112.97726418	849.21028661	
22		68264	232955	672005	47518612	210311	15497	
23		9343.3161509	187.69852848	49.778313268	1.4680796252	8974.9866501	9711.6456517	
24		30756	397093	60218	624174E-272	25571	3594	
25		-171.1345266	187.2072773	-0.914144614	0.3608618806	-538.5000272	196.23097405	
26		1098	09581	7656395	3934177	749278	296784	

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

- charges (p-value = 0.0)
- age (p-value = 7.91969773883132E-73)
- bmi (p-value = 9.634367039265465E-24)
- children (p-value = 0.010474857247518612)
- smoker (p-value = 1.4680796252634174E-272)

Step 7: 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".

FileEditData TransformationAnalyticsStatisticsPlotHelp

IMPORT

TRAIN TEST SPLIT

NORMALIZE TRAIN SET

FEATURE SELECTION

NORMALIZE TEST SET

NORMALIZE TRAIN SET

NORMALIZE TEST SET

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	age	sex	bmi	children	smoker	region	charges
1	0	-1.442052194	-1.013537686	-0.448337111	-0.894088423	2.0358255719	1.3698195197	16884.924
		0451325	61993	49191124	5403234	795726	58298184	
2	1	-1.513797506	0.9856604221	0.5084179971	-0.072051081	-0.490711973	-0.438594415	1725.5523
		3752274	549418	769941	75050219	46850026	58298184	
3	2	-0.796344383	0.9856604221	0.3829151975	1.5720236018	-0.490711973	-0.438594415	4449.462
		0742793	549418	7306575	2914	46850026	58298184	
4	3	-0.437617821	0.9856604221	-1.295073532	-0.894088423	-0.490711973	-1.342801383	21984.47061
		42380545	549418	1963338	5403234	46850026	2324968	
5	4	-0.509363133	0.9856604221	-0.288606275	-0.894088423	-0.490711973	-1.342801383	3866.8552
		7539002	549418	6323666	5403234	46850026	2324968	
6	5	-0.581108446	-1.013537686	-0.800396912	-0.894088423	-0.490711973	-0.438594415	3756.6216
		083995	61993	9782544	5403234	46850026	58298184	
7	6	0.4950712388	-1.013537686	0.4546310830	-0.072051081	-0.490711973	-0.438594415	10797.3362
		674268	61993	610241	75050219	46850026	58298184	
8	7	-0.150636572	0.9856604221	-0.133765159	0.749986260	-0.490711973	0.4656125520	6406.4107
		1034263	549418	2379103	39319	46850026	6653307	
9	8	1.4995056114	-1.013537686	-0.784097848	-0.894088423	-0.490711973	-1.342801383	28923.13692
		887537	61993	0946271	5403234	46850026	2324968	
10	9	-1.011580320	0.9856604221	-0.722161401	-0.894088423	-0.490711973	0.4656125520	
		0645637	549418	5368447	5403234	46850026	6653307	
11	10	-1.155070944	0.9856604221	0.6111021059	-0.894088423	-0.490711973	1.3698195197	1826.843
		7247534	549418	438344	5403234	46850026	160479	
12	11	1.2125243621	-1.013537686	1.4945114226	-0.894088423	-0.490711973	-0.438594415	11090.7178
		683748	61993	36427	5403234	46850026	58298184	
13	12	-0.868089695	0.9856604221	1.8710198214	-0.894088423	-0.490711973	-0.438594415	39611.7577
		4043742	549418	482108	5403234	795726	58298184	
14	13	1.442052194	0.9856604221	-0.886206252	-0.072051081	-0.490711973	1.3698195197	1837.237
		0451325	549418	6516017	75050219	46850026	160479	

IMPORTTRAIN TEST SPLITNORMALIZE TRAIN SETNORMALIZE TEST SETFEATURE SELECTION REGRESSIONFEATURE SELECTION TRAIN SET

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

The screenshot shows the 'Select Column(s)' dialog box in the Isalos Analytics Platform. The dialog has two panes: 'Excluded Columns' and 'Included Columns'. In the 'Excluded Columns' pane, 'Col3 -- sex' and 'Col7 -- region' are listed. In the 'Included Columns' pane, 'Col2 -- age', 'Col4 -- bmi', 'Col5 -- children', 'Col6 -- smoker', and 'Col8 -- charges' are listed. The 'Execute' button is highlighted.

The results will appear on the output spreadsheet.

The screenshot shows the 'FEATURE_SELECTION_TEST_SET' tab in the Isalos Analytics Platform. The tab contains a table with columns: User Header, User Row ID, age, sex, bmi, children, smoker, region, and charges. The table shows data for 14 rows. The 'FEATURE_SELECTION_TEST_SET' tab is selected, and the 'Import' button is visible.

Step 8: 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".

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a menu bar with 'File', 'Edit', 'Data Transformation', 'Analytics', 'Statistics', 'Plot', and 'Help'. Below the menu, a workflow diagram shows the process: 'IMPORT' → 'TRAIN_TEST_SPLIT' → 'NORMALIZE_TRAIN_SET' → 'NORMALIZE_TEST_SET' → 'FEATURE_SELECTION_REGRESSION' → 'FEATURE_SELECTION_TRAIN_SET' → 'FEATURE_SELECTION_TEST_SET'. The main area displays a large data table with 14 rows and 9 columns. The columns are labeled 'User Header', 'User Row ID', 'age', 'sex', 'bmi', 'children', 'smoker', 'region', and 'charges'. The data is organized into a grid with alternating blue and white rows. The 'User Header' column contains the text 'User Header' and 'User Row ID'. The 'User Row ID' column contains numerical values from 1 to 14. The 'age' column contains values from 24 to 47. The 'sex' column contains values 'f' and 'm'. The 'bmi' column contains values from 1.109264192 to 1.109264192. The 'children' column contains values from 0.0066908845 to 0.0066908845. The 'smoker' column contains values from 0.0066908845 to 0.0066908845. The 'region' column contains values from 0.0066908845 to 0.0066908845. The 'charges' column contains values from 0.0066908845 to 0.0066908845.

User Header	User Row ID	age	sex	bmi	children	smoker	region	charges
1	7	-0.150636572	-1.013537686	-0.474415615	1.5720236018	-0.490711973	-1.342801383	7281.5056
2	11	1.6429962361	-1.013537686	-0.710752056	-0.894088423	2.0358255719	-0.438594415	2324968
3	17	-1.155070944	0.9856604221	-1.109264192	-0.894088423	-0.490711973	0.4656125520	27808.7251
4	24	-0.150636572	0.9856604221	-0.472963280	0.7499862600	-0.490711973	-1.342801383	2395.17155
5	25	1.4277602991	-1.013537686	-0.477675428	1.5720236018	-0.490711973	-0.438594415	6203.90175
6	29	-0.581108446	0.9856604221	0.9207843387	0.7499862600	2.0358255719	1.3698195197	14001.1338
7	30	0.83995	549418	327561	0.9319	795726	160479	38711.0
8	33	-1.226816257	0.9856604221	0.8066908845	-0.894088423	2.0358255719	1.3698195197	35585.576
9	37	0.548482	549418	473678	5403234	795726	160479	13770.0979
10	38	1.7147415484	0.9856604221	-0.381510945	-0.894088423	-0.490711973	-1.342801383	2302.3
11	41	790383	549418	4690406	5403234	46850026	58298184	39774.2763
12	42	-0.939835007	0.9856604221	-1.605570718	-0.894088423	-0.490711973	1.3698195197	6079.6715
13	44	7344689	549418	2294277	5403234	46850026	160479	6272.4772
14	47	-0.294127196	0.9856604221	0.9810908788	-0.072051081	2.0358255719	0.4656125520	6653307

Manipulate the data by choosing the columns that correspond to the significant features (from the step 6) by browsing: "Data Transformation" → "Data Manipulation" → "Select Column(s)".

The screenshot shows the 'Select Column(s)' dialog box. It has two main sections: 'Excluded Columns' and 'Included Columns'. The 'Excluded Columns' section contains a list of columns: 'Col3 -- sex' and 'Col7 -- region'. The 'Included Columns' section contains a list of columns: 'Col2 -- age', 'Col4 -- bmi', 'Col5 -- children', 'Col6 -- smoker', and 'Col8 -- charges'. There are buttons for '>>', '>', '<', and '<<' to move columns between the two sections. At the bottom, there are 'Execute' and 'Cancel' buttons.

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)	Col9
1	7	-0.150636572	-1.013537686	-0.474415615	1.5720236018	-0.490711973	-1.342801383	7281.5056	
2	11	1.6429962361	-1.013537686	-0.710752056	-0.894088423	2.0358255719	-0.438594415	27808.7251	
3	17	-1.155070944	0.9856604221	-1.109264192	-0.894088423	-0.490711973	0.4656125520	2395.17155	
4	24	0.150636572	0.9856604221	-0.427963280	0.7499862600	-0.490711973	-1.342801383	6203.90175	
5	25	1.4277602991	-1.013537686	-0.477675428	1.5720236018	-0.490711973	-0.438594415	14001.1338	
6	29	-0.581108446	0.9856604221	0.9507843387	0.7499862600	2.0358255719	1.3698195197	38711.0	
7	30	1.226816257	0.9856604221	0.8066908845	-0.894088423	2.0358255719	1.3698195197	35585.576	
8	33	1.7147415484	0.9856604221	-0.381510945	-0.894088423	-0.490711973	-1.342801383	13770.0979	
9	37	-0.939835007	0.9856604221	-1.605570718	-0.894088423	-0.490711973	1.3698195197	2302.3	
10	38	-0.294127196	0.9856604221	0.9810908788	-0.072051081	2.0358255719	0.4656125520	39774.2763	
11	41	-0.581108446	0.9856604221	0.9745712528	0.7499862600	-0.490711973	-0.438594415	4949.7587	
12	44	0.1363446772	0.9856604221	-1.445839882	-0.072051081	-0.490711973	-0.438594415	6272.4772	
13	46	-0.078891259	0.9856604221	1.0430273253	-0.072051081	-0.490711973	0.4656125520	6079.6715	
14	48	-0.796344383	-1.013537686	0.6714086460	-0.894088423	-0.490711973	-1.342801383	3556.9223	

Step 9: 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	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8
1	0	-1.442052194	-0.440337111	-0.894088423	2.0358255719	16884.924		
2	1	0.451325	49191124	5403234	795726			
3	2	-1.513797506	0.5084179971	-0.072051081	-0.490711973	1725.5523		
4	3	3752274	769941	75050219	46850026	4449.462		
5	4	-0.796344383	0.38329151975	1.5720236018	-0.490711973			
6	5	0.7427293	7306575	2914	46850026			
7	6	-0.437617821	-1.295073532	-0.894088423	-0.490711973	21984.47061		
8	7	42380545	1963338	5403234	46850026			
9	8	-0.509363133	-0.288606275	-0.894088423	-0.490711973	3866.8552		
10	9	7339002	6323666	5403234	46850026			
11	10	-0.581108446	-0.800396912	-0.894088423	-0.490711973	3756.6216		
12	11	0.83995	9782544	5403234	46850026			
13	12	0.4950712388	0.4546310830	-0.072051081	-0.490711973	8240.5896		
14	13	674268	610241	75050219	46850026			
15	14	-0.150636572	-0.133765159	0.7499862600	-0.490711973	6406.4107		
16	15	1034263	2379103	39319	46850026			
17	16	1.4995056114	-0.784097848	-0.894088423	-0.490711973	28923.13692		
18	17	887537	0946271	5403234	46850026			
19	18	-1.011580320	-0.722161401	-0.894088423	-0.490711973	2721.3208		
20	19	0645637	536847	5403234	46850026			
21	20	-1.155070944	0.6111021059	-0.894088423	-0.490711973	1826.843		
22	21	7247534	438434	5403234	46850026			

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 "charges" and the "Number of Neighbors" to 3.

User Header	User Row ID	age	bmr	children	smoker	charges	Col8
1	1	1.442052194	-0.448337111	-0.894088423	2.0356255719	16884.924	
2	2	0.451325	49191124	5403234	795726	1725.5523	
3	3	1.513797506	0.5084179971	-0.072051081	-0.490711973	4449.462	
4	4	3752274	769941	75050219	46850026	21984.47061	
5	5	-0.796344383	0.3829151975	1.5720236018	-0.490711973	3866.8552	
6	6	0.742793	7306575	2914	46850026	3756.6216	
7	7	-0.437617821	-1.295073532	-0.894088423	-0.490711973	8240.5896	
8	8	42380545	1963338	5403234	46850026	6406.4107	
9	9	-0.509363133	-0.288606275	-0.894088423	-0.490711973	28923.13692	
10	10	7539002	6323666	5403234	46850026	2721.3208	
11	11	-0.581108446	-0.800396912	-0.894088423	-0.490711973	1826.843	

kNN Regression Model

Target Column: Col6 -- charges

Number of Neighbors: 3

Execute Cancel

The predictions will appear on the output spreadsheet.

User Header	User Row ID	age	bmr	children	smoker	charges	Col8	Col9
1	1	1.442052194	-0.448337111	-0.894088423	2.0356255719	16884.924		
2	2	0.451325	49191124	5403234	795726	1725.5523		
3	3	1.513797506	0.5084179971	-0.072051081	-0.490711973	4449.462		
4	4	-0.796344383	0.3829151975	1.5720236018	-0.490711973	21984.47061		
5	5	0.742793	7306575	2914	46850026	3866.8552		
6	6	-0.437617821	-1.295073532	-0.894088423	-0.490711973	3756.6216		
7	7	42380545	1963338	5403234	46850026	8240.5896		
8	8	-0.509363133	-0.288606275	-0.894088423	-0.490711973	6406.4107		
9	9	7539002	6323666	5403234	46850026	28923.13692		
10	10	-0.581108446	-0.800396912	-0.894088423	-0.490711973	2721.3208		
11	11	0.4950712388	0.4546310830	-0.072051081	-0.490711973	1826.843		

User Header	User Row ID	charges	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3
1	1	16884.924	16824.258955	0	0.0	296	0.0053306833	126	
2	2	1725.5523	39548	0	0.0	710	46785042	210	
3	3	4449.462	4470.4764133	2	0.0	906	0.0384718859	670	
4	4	21984.47061	16677	3	0.0	971	0.02997555	981	
5	5	3866.8552	21103.694406	3	0.0	570	0.0316496631	1277	
6	6	3756.6216	3881.5543337	4	0.0	101	0.0225304552	961	
7	7	8240.5896	267377	5	0.0	347	4456439		
8	8	6406.4107	3743.7734382	5	0.0	101	0.0230072864		
9	9	28923.13692	124195	6	0.0	101	21612056		
10	10	2721.3208	8293.3674571	556	0.0	6	0.0		
11	11	1826.843	6395.0045758	8	0.0	879	0.0088781275		

Step 10: 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".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(fit), and VALIDATE_MODEL(predict). The main area displays two data tables. The left table has columns: User Header, User Row ID, age, bmi, children, smoker, charges, and empty columns Col7, Col8, Col9. The right table has columns: User Header, User Row ID, and empty columns Col2 through Col7. Both tables contain 11 rows of data.

User Header	User Row ID	age	bmi	children	smoker	charges	Col7	Col8	Col9
1	7	-0.150636572	-0.474415615	1.5720236018	-0.490711973	7281.5056			
2	11	1.6429962361	-0.710752056	-0.894088423	2.0358255719	27808.7251			
3	17	-1.155070944	-1.109264192	-0.894088423	-0.490711973	2395.17155			
4	24	-0.150636572	-0.474415615	1.5720236018	-0.490711973	6203.90175			
5	25	1.4277602991	-0.477675428	1.5720236018	-0.490711973	14001.1338			
6	29	-0.581108446	0.9207843387	0.7499862600	2.0358255719	38711.0			
7	30	-1.226816257	0.8066908845	-0.894088423	2.0358255719	35585.576			
8	33	1.7147415484	-0.381510945	-0.894088423	-0.490711973	13770.0979			
9	37	-0.939835007	-1.605570718	-0.894088423	-0.490711973	2302.3			
10	38	-0.294127196	0.9810908788	-0.072051081	2.0358255719	39774.2763			
11	41	-0.581108446	0.9745712528	0.7499862600	-0.490711973	4949.7587			

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

The screenshot shows the Isalos Analytics Platform interface. The 'Analytics' menu is open, and 'Existing Model Utilization' is selected. The workflow diagram shows the process from IMPORT to VALIDATE_MODEL(predict). The main area displays two data tables. The left table has columns: User Header, User Row ID, age, bmi, children, smoker, charges, and empty columns Col7, Col8, Col9. The right table has columns: User Header, User Row ID, and empty columns Col2 through Col7. Both tables contain 11 rows of data.

User Header	User Row ID	age	bmi	children	smoker	charges	Col7	Col8	Col9
1	7	-0.150636572	-0.474415615	1.5720236018	-0.490711973	7281.5056			
2	11	1.6429962361	-0.710752056	-0.894088423	2.0358255719	27808.7251			
3	17	-1.155070944	-1.109264192	-0.894088423	-0.490711973	2395.17155			
4	24	-0.150636572	-0.474415615	1.5720236018	-0.490711973	6203.90175			
5	25	1.4277602991	-0.477675428	1.5720236018	-0.490711973	14001.1338			
6	29	-0.581108446	0.9207843387	0.7499862600	2.0358255719	38711.0			
7	30	-1.226816257	0.8066908845	-0.894088423	2.0358255719	35585.576			
8	33	1.7147415484	-0.381510945	-0.894088423	-0.490711973	13770.0979			
9	37	-0.939835007	-1.605570718	-0.894088423	-0.490711973	2302.3			
10	38	-0.294127196	0.9810908788	-0.072051081	2.0358255719	39774.2763			
11	41	-0.581108446	0.9745712528	0.7499862600	-0.490711973	4949.7587			

The 'Existing Model Execution' dialog box is open, showing the 'Model' dropdown set to '(from Tab:) TRAIN_MODEL (.fit)' and the 'Type' set to 'kNN Model'. The 'Model Input' section lists the input variables: age, sex, bmi, children, smoker, and region, all with a datatype of 'Double'. The 'Transfer Column(s) to Output' section shows the 'charges' column being transferred to the output.

The predictions will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a workflow diagram with steps: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(fit), and VALIDATE_MODEL(predict). Below the diagram, there are two data tables.

Table 1 (Left):

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	7	-0.150636572	-0.474415615	1.5720236018	-0.490711973	7281.5056			
2	11	1.6429962361	-0.710752056	-0.894088423	2.0358255719	27808.7251			
3	17	-1.155070944	-1.109264192	-0.894088423	-0.490711973	2395.17155			
4	24	7247534	5229658	5403234	46850026	6203.90175			
5	25	-0.150636572	-0.473763280	0.7499862600	-0.490711973	14001.1338			
6	29	1.4277602991	-0.477675428	1.5720236018	-0.490711973	46850026			
7	30	0.83995	327561	39319	795726	38711.0			
8	33	-1.226816257	0.806690845	-0.894088423	2.0358255719	35585.576			
9	37	0.548482	473678	5403234	795726	13770.0979			
10	38	1.7147415484	-0.381510945	-0.894088423	-0.490711973	46850026			
11	41	790383	4690406	5403234	46850026	2302.3			

Table 2 (Right):

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (I)	Col6 (D)	Col7 (I)	Col8
1	7	6605.4818823	300	0.0223320117	1063	0.0316496631	785	0.042
2	11	28529.490195	1321	0.0108958837	419	0.0285865009	244	0.044
3	17	2513.9648511	1114	0.0178907721	693	0.0233320117	816	0.024
4	24	6417.9442891	176	0.0223320117	879	0.0396825396	1173	0.035
5	25	14088.949734	912	0.0275760021	255	0.1075900337	734	0.111
6	29	38453.120961	609	0.0458380400	1207	0.1337977986	373	0.142
7	30	35200.740041	1291	0.0678820087	263	0.0747142001	618	0.095
8	33	13274.296965	918	0.0435788603	789	0.0484637574	341	0.055
9	37	504372	1054	0.0282400574	1137	0.0384718859	388	0.046
10	38	38357.782707	1118	0.0500297738	947	0.0794112898	1249	0.104
11	41	4511.4982658	934	0.0262970991	1168	0.0441890914	1217	0.046

Step 11: 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".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a workflow diagram with steps: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(fit), VALIDATE_MODEL(predict), and STATISTICS_ACCURACIES. Below the diagram, there are two data tables.

Table 1 (Left):

User Header	Col1	Col2 (D)	Col3 (I)	Col4 (D)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (D)	Col9
1	7	6605.4818823	300	0.0223320117	1063	0.0316496631	785	0.0434915766	7281.5056
2	11	28529.490195	1321	0.0108958837	419	0.0285865009	244	0.0446583287	27808.7251
3	17	2513.9648511	1114	0.0178907721	693	0.0233320117	816	0.0240230227	2395.17155
4	24	6417.9442891	176	0.0223320117	879	0.0396825396	1173	0.0397056449	6203.90175
5	25	14088.949734	912	0.0275760021	255	0.1075900337	734	0.1075900337	14001.1338
6	29	38453.120961	609	0.0458380400	1207	0.1337977986	373	0.1420626675	46850026
7	30	35200.740041	1291	0.0678820087	263	0.0747142001	618	0.0934927925	38711.0
8	33	13274.296965	918	0.0435788603	789	0.0484637574	341	0.0508825641	35585.576
9	37	504372	1054	0.0282400574	1137	0.0384718859	388	0.0384718859	13770.0979
10	38	38357.782707	1118	0.0500297738	947	0.0794112898	1249	0.1040055505	46850026
11	41	4511.4982658	934	0.0262970991	1168	0.0441890914	1217	0.0441890914	4949.7587

Table 2 (Right):

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								

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

The screenshot shows the 'Statistics' menu with options: Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Model Metrics' option is selected, leading to a 'Regression Statistics Metrics' dialog box.

The dialog box has two dropdown menus: 'Actual Value Column' set to 'Col9 -- charges' and 'Prediction Value Column' set to 'Col2 -- kNN Prediction'. There are 'Execute' and 'Cancel' buttons.

Below the dialog box is a table showing the results of the regression analysis. The table has columns for User Header, User Row ID, kNN Prediction, Closest NN1, Distance from NN1, Closest NN2, Distance from NN2, Closest NN3, and Distance from NN3.

User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3
1	7	6605.4818823	300	0.0223320117	1063	0.0316496631	785	0.0434915
2	11	28529.490195	1321	0.0108958837	419	0.0285865009	244	0.0446583
3	17	2513.9648511	1114	0.0178907721	693	0.0223320117	816	0.0240230
4	24	6417.9442891	176	0.0223320117	879	0.0396825396	1173	0.0397056
5	25	14088.949734	912	0.0275760021	255	0.1075900337	734	0.1178369
6	29	38453.120961	609	0.0458380400	1207	0.1337977986	373	0.1420626
7	30	35200.740041	1291	0.0678820087	263	0.0747142001	618	0.0934927
8	33	13274.296965	918	0.0435788603	789	0.0484637574	341	0.0508825
9	37	3252.4922644	1054	0.0282400574	1137	0.0384718859	388	0.0486951
10	38	38357.782707	1118	0.0500297738	947	0.0794112898	1249	0.1046055
11	41	4511.4982658	934	0.0262970991	1168	0.0441890914	1217	0.0469642

The results will appear on the output spreadsheet.

The screenshot shows the 'Statistics' menu with options: Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Model Metrics' option is selected, leading to a 'Regression Statistics Metrics' dialog box.

The dialog box has two dropdown menus: 'Actual Value Column' set to 'Col9 -- charges' and 'Prediction Value Column' set to 'Col2 -- kNN Prediction'. There are 'Execute' and 'Cancel' buttons.

Below the dialog box is a table showing the results of the regression analysis. The table has columns for User Header, User Row ID, kNN Prediction, Closest NN1, Distance from NN1, Closest NN2, Distance from NN2, Closest NN3, and Distance from NN3.

User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3
1	7	6605.4818823	300	0.0223320117	1063	0.0316496631	785	0.0434915
2	11	28529.490195	1321	0.0108958837	419	0.0285865009	244	0.0446583
3	17	2513.9648511	1114	0.0178907721	693	0.0223320117	816	0.0240230
4	24	6417.9442891	176	0.0223320117	879	0.0396825396	1173	0.0397056
5	25	14088.949734	912	0.0275760021	255	0.1075900337	734	0.1178369
6	29	38453.120961	609	0.0458380400	1207	0.1337977986	373	0.1420626
7	30	35200.740041	1291	0.0678820087	263	0.0747142001	618	0.0934927
8	33	13274.296965	918	0.0435788603	789	0.0484637574	341	0.0508825
9	37	3252.4922644	1054	0.0282400574	1137	0.0384718859	388	0.0486951
10	38	38357.782707	1118	0.0500297738	947	0.0794112898	1249	0.1046055
11	41	4511.4982658	934	0.0262970991	1168	0.0441890914	1217	0.0469642

Step 12: Reliability check of each record of the test set

Step 12.a: Create the domain

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

Import data into the input spreadsheet of the "EXCLUDE_CHARGES" 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	Col9
1	0	-1.442052194	-0.448337111	-0.894088423	2.0358255719	16884.924			
2	1	-1.513797506	0.5084179971	-0.072051081	-0.490711973	1725.5523			
3	2	-0.796344383	0.3829151975	1.5720236018	-0.490711973	4449.462			
4	3	-0.437617821	-1.295073532	-0.894088423	-0.490711973	21984.47061			
5	4	-0.509363133	-0.288606275	-0.894088423	-0.490711973	3866.8552			
6	5	-0.581108446	-0.800396912	-0.894088423	-0.490711973	3756.6216			
7	6	0.4950712388	0.4546310830	-0.072051081	-0.490711973	8240.5896			
8	7	0.674268	0.133765159	0.7499862600	-0.490711973	6406.4107			
9	8	1.4995056114	-0.784097848	-0.894088423	-0.490711973	28923.13692			
10	9	0.645637	0.536847	0.4303234	46850026	2721.3208			
11	10	-1.155070944	0.6111021059	-0.894088423	-0.490711973	1826.843			
12	11	0.7247534	0.438434	0.4303234	46850026				

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

File

Edit

Data Transformation

Analytics

Statistics

Plot

Help

IMPORT

Normalizers

Data Manipulation

Split

Variable Selection

Remove Column(s)

Select Column(s)

Matrix Transpose

Sort by Column

Fill Missing Column(s) Values

TRAIN_MODEL_LR

VALIDATE_MODEL(predict)

STATISTICS_ACCURACIES

EXCLUDE_CHARGES

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	0	-1.442052194	-0.448337111	-0.894088423	2.0358255719	16884.924			
2	1	-1.513797506	0.5084179971	-0.072051081	-0.490711973	1725.5523			
3	2	-0.796344383	0.3829151975	1.5720236018	-0.490711973	4449.462			
4	3	-0.437617821	-1.295073532	-0.894088423	-0.490711973	21984.47061			
5	4	-0.509363133	-0.288606275	-0.894088423	-0.490711973	3866.8552			
6	5	-0.581108446	-0.800396912	-0.894088423	-0.490711973	3756.6216			
7	6	0.4950712388	0.4546310830	-0.072051081	-0.490711973	8240.5896			
8	7	0.674268	0.133765159	0.7499862600	-0.490711973	6406.4107			
9	8	1.4995056114	-0.784097848	-0.894088423	-0.490711973	28923.13692			
10	9	0.645637	0.536847	0.4303234	46850026	2721.3208			
11	10	-1.155070944	0.6111021059	-0.894088423	-0.490711973	1826.843			
12	11	0.7247534	0.438434	0.4303234	46850026				

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

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a workflow diagram with nodes: IMPORT, TRAIN_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(fit), VALIDATE_MODEL(predict), STATISTICS_ACCURACIES, EXCLUDE_CHARGES, and DOMAIN. Below the diagram is a data table with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6, Col7, Col8, Col9. The table contains 11 rows of data. The DOMAIN node is highlighted in blue.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8	Col9
1	0	-1.442052194	-0.448337111	-0.894088423	2.0358255719				
2	1	-1.513797506	0.5084179971	-0.072051081	-0.490711973				
3	2	0.742793	7306575	2914	46850026				
4	3	-0.437617821	-1.295073532	884215403234	-0.490711973				
5	4	-0.509363133	-0.288606275	-0.894088423	-0.490711973				
6	5	-0.581108446	-0.800396912	-0.894088423	-0.490711973				
7	6	0.4950712388	0.4546310830	-0.072051081	-0.490711973				
8	7	-0.150636572	-0.133765159	0.749962600	-0.490711973				
9	8	1.4995056114	-0.784097848	-0.894088423	-0.490711973				
10	9	0.887537	0.946271	5403234	46850026				
11	10	-1.011580320	-0.722161401	-0.894088423	-0.490711973				
12	11	-1.155070944	0.6111021059	-0.894088423	-0.490711973				
13	12	7247534	438434	5403234	46850026				

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

The screenshot shows the Isalos Analytics Platform interface. The 'Statistics' menu is open, showing options: Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Domain - APD' option is selected. Below the menu is a data table with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6, Col7, Col8, Col9. The table contains 11 rows of data. The 'Domain - APD' dialog box is open, showing the formula $APD = d + Z\sigma$, the value of Z as 0.5, and the option to perform computations using CPU (double precision). The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8	Col9
1	0	-1.442052194	-0.448337111	-0.894088423	2.0358255719				
2	1	-1.513797506	0.5084179971	-0.072051081	-0.490711973				
3	2	0.742793	7306575	2914	46850026				
4	3	-0.437617821	-1.295073532	884215403234	-0.490711973				
5	4	-0.509363133	-0.288606275	-0.894088423	-0.490711973				
6	5	-0.581108446	-0.800396912	-0.894088423	-0.490711973				
7	6	0.4950712388	0.4546310830	-0.072051081	-0.490711973				
8	7	-0.150636572	-0.133765159	0.749962600	-0.490711973				
9	8	1.4995056114	-0.784097848	-0.894088423	-0.490711973				
10	9	0.887537	0.946271	5403234	46850026				
11	10	-1.011580320	-0.722161401	-0.894088423	-0.490711973				
12	11	-1.155070944	0.6111021059	-0.894088423	-0.490711973				
13	12	7247534	438434	5403234	46850026				

Domain - APD

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

Perform Computations CPU (double precision)

Execute Cancel

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(REG), VALIDATE_MODEL(predict), STATISTICS_ACCURACIES, EXCLUDE_CHARGES, and DOMAIN. The main area displays two data tables.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8	Col9
1	0	-1.442052194	-0.44837111	-0.894088423	2.0358255719				
2	1	-1.513797506	0.5084179971	-0.072051081	-0.490711973				
3	2	-0.796344383	0.3829151975	1.5720236018	-0.490711973				
4	3	-0.437617821	-1.295073532	-0.894088423	-0.490711973				
5	4	42380545	1963338	5403234	46850026				
6	5	-0.581108446	-0.800396912	-0.894088423	-0.490711973				
7	6	0.4950712388	0.4546310830	-0.072051081	-0.490711973				
8	8	-0.150636572	-0.133765159	0.7499862600	-0.490711973				
9	9	1.4995056114	-0.784097848	-0.894088423	-0.490711973				
10	10	-1.011580320	-0.722161401	-0.894088423	-0.490711973				
11	12	0.645637	5368447	5403234	46850026				

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5	Col6	Col7	Col8	Col9
1	0	0.0	APD	2.0546260207	reliable				
2	1	0.0		2.0546260207	reliable				
3	2	0.0		2.0546260207	reliable				
4	3	0.0		2.0546260207	reliable				
5	4	0.0		2.0546260207	reliable				
6	5	0.0		2.0546260207	reliable				
7	6	0.0		2.0546260207	reliable				
8	8	0.0		2.0546260207	reliable				
9	9	0.0		2.0546260207	reliable				
10	10	0.0		2.0546260207	reliable				
11	12	0.0		2.0546260207	reliable				

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_CHARGES_TEST_SET".

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

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(REG), VALIDATE_MODEL(predict), STATISTICS_ACCURACIES, EXCLUDE_CHARGES, and DOMAIN. The main area displays two data tables.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	7	-0.150636572	-0.474415615	1.5720236018	-0.490711973	7281.5056			
2	11	1.6423962361	-0.710752056	-0.894088423	2.0358255719	27808.7251			
3	17	-1.155070944	-1.109264192	-0.894088423	-0.490711973	2395.17155			
4	24	-0.150636572	-0.427963280	0.7499862600	-0.490711973	6203.90175			
5	25	1.4277602991	-0.477675428	1.5720236018	-0.490711973	14001.1338			
6	29	-0.581108446	0.9207843387	0.7499862600	-0.490711973	38711.0			
7	30	1.228816257	0.806690845	-0.894088423	2.0358255719	35585.576			
8	33	1.7147415484	-0.381510945	-0.894088423	-0.490711973	13770.0979			
9	37	-0.939635007	-1.605570718	-0.894088423	-0.490711973	2302.3			
10	38	-0.294127196	0.9810908788	-0.072051081	2.0358255719	39774.2763			
11	41	-0.581108446	0.9745712528	0.7499862600	-0.490711973	4949.7587			

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									

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

The screenshot shows the 'Data Transformation' menu with options: Normalizers, Data Manipulation, Variable Selection, and Matrix Transpose. The 'Data Manipulation' submenu is open, showing 'Remove Column(s)', 'Select Column(s)', 'Split', 'Sort by Column', and 'Fill Missing Column(s) Values'. The 'Select Column(s)' option is highlighted. Below the menu, a data table is visible with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7, and Col8. The table contains 11 rows of data.

The 'Select Column(s)' dialog box shows two lists: 'Excluded Columns' and 'Included Columns'. The 'Excluded Columns' list contains 'Col6 -- charges'. The 'Included Columns' list contains 'Col2 -- age', 'Col3 -- bmi', 'Col4 -- children', and 'Col5 -- smoker'. Navigation buttons (>>, >, <, <<) are between the lists. 'Execute' and 'Cancel' buttons are at the bottom.

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

The screenshot shows the 'Data Transformation' menu with options: Normalizers, Data Manipulation, Variable Selection, and Matrix Transpose. The 'Data Manipulation' submenu is open, showing 'Remove Column(s)', 'Select Column(s)', 'Split', 'Sort by Column', and 'Fill Missing Column(s) Values'. The 'Select Column(s)' option is highlighted. Below the menu, a data table is visible with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6, Col7, and Col8. The table contains 11 rows of data.

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

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
User Header	User Row ID	age	bmi	children	smoker		
1	7	-0.150636572	-0.474415615	1.5720236018	-0.490711973		
		1034263	30571447	2914	46850026		
2	11	1.6429962361	-0.710752056	-0.894088423	2.0358255719		
		489434	1183058	5403234	795726		
3	17	-1.155070944	-1.109264192	-0.894088423	-0.490711973		
		7247534	5229858	5403234	46850026		
4	24	-0.150636572	-0.474415615	1.5720236018	-0.490711973		
		1034263	30571447	2914	46850026		
5	25	1.6429962361	-0.710752056	-0.894088423	2.0358255719		
		489434	1183058	5403234	795726		
6	29	-0.581108446	0.9207843387	0.7499862600	2.0358255719		
		083995	327561	39319	795726		
7	30	-1.226816257	0.8066908845	-0.894088423	2.0358255719		
		0548482	473678	5403234	795726		
8	33	1.7147415484	-0.381510945	-0.894088423	-0.490711973		
		790383	4690406	5403234	46850026		
9	37	-0.939835007	-1.605570718	-0.894088423	-0.490711973		
		7344689	2294277	5403234	46850026		
10	38	-0.294127196	0.9810908788	-0.072051081	2.0358255719		
		7636159	021767	75050219	795726		
11	41	-0.581108446	0.9207843387	0.7499862600	-0.490711973		
		083995	487261	39319	46850026		

The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8
User Header	User Row ID	age	bmi	children	smoker			
1	7	-0.150636572	-0.474415615	1.5720236018	-0.490711973			
		1034263	30571447	2914	46850026			
2	11	1.6429962361	-0.710752056	-0.894088423	2.0358255719			
		489434	1183058	5403234	795726			
3	17	-1.155070944	-1.109264192	-0.894088423	-0.490711973			
		7247534	5229858	5403234	46850026			
4	24	-0.150636572	-0.474415615	1.5720236018	-0.490711973			
		1034263	30571447	2914	46850026			
5	25	1.6429962361	-0.710752056	-0.894088423	2.0358255719			
		489434	1183058	5403234	795726			
6	29	-0.581108446	0.9207843387	0.7499862600	2.0358255719			
		083995	327561	39319	795726			
7	30	-1.226816257	0.8066908845	-0.894088423	2.0358255719			
		0548482	473678	5403234	795726			
8	33	1.7147415484	-0.381510945	-0.894088423	-0.490711973			
		790383	4690406	5403234	46850026			
9	37	-0.939835007	-1.605570718	-0.894088423	-0.490711973			
		7344689	2294277	5403234	46850026			
10	38	-0.294127196	0.9810908788	-0.072051081	2.0358255719			
		7636159	021767	75050219	795726			
11	41	-0.581108446	0.9207843387	0.7499862600	-0.490711973			
		083995	487261	39319	46850026			

	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
User Header	User Row ID	Domain	APD	Prediction	age	bmi	children	smoker	
1	7	0.0781435902	2.0546260207	reliable	-0.150636572	-0.474415615	1.5720236018	-0.490711973	
		323031	790885		1034263	30571447	2914	46850026	
2	11	0.0660112127	2.0546260207	reliable	1.6429962361	-0.710752056	-0.894088423	2.0358255719	
		786896	790885		489434	1183058	5403234	795726	
3	17	0.0781435902	2.0546260207	reliable	-1.155070944	-1.109264192	-0.894088423	-0.490711973	
		3230318	790885		7247534	5229858	5403234	46850026	
4	24	0.0781435902	2.0546260207	reliable	-0.150636572	-0.474415615	1.5720236018	-0.490711973	
		3230308	790885		1034263	30571447	2914	46850026	
5	25	0.1670654150	2.0546260207	reliable	1.6429962361	-0.710752056	-0.894088423	2.0358255719	
		5717643	790885		489434	1183058	5403234	795726	
6	29	0.2547955669	2.0546260207	reliable	-0.581108446	0.9207843387	0.7499862600	2.0358255719	
		924162	790885		083995	327561	39319	795726	
7	30	0.2436058801	2.0546260207	reliable	-1.226816257	0.8066908845	-0.894088423	2.0358255719	
		8362608	790885		0548482	473678	5403234	795726	
8	33	0.1446063877	2.0546260207	reliable	1.7147415484	-0.381510945	-0.894088423	-0.490711973	
		5596893	790885		790383	4690406	5403234	46850026	
9	37	0.1306630992	2.0546260207	reliable	-0.939835007	-1.605570718	-0.894088423	-0.490711973	
		9817455	790885		7344689	2294277	5403234	46850026	
10	38	0.2075451295	2.0546260207	reliable	-0.294127196	0.9810908788	-0.072051081	2.0358255719	
		6569346	790885		7636159	021767	75050219	795726	
11	41	0.1148198163	2.0546260207	reliable	-0.581108446	0.9207843387	0.7499862600	-0.490711973	
		4607938	790885		083995	487261	39319	46850026	

There are no unreliable samples in the test set.

Final Isalos Workflow

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

