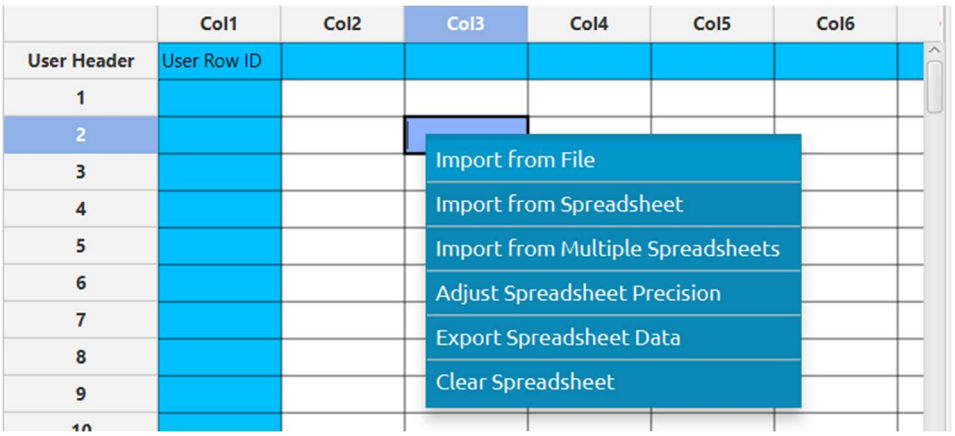
Housing Prices Dataset

The goal of this study is to train a model in order to predict housing prices. The dataset used in this case study, which can be found in https://www.kaggle.com/datasets/yasserh/housing-prices-dataset/data, has 13 features and 545 samples. It contains information on certain factors like house area, number of bedrooms, furnishing status, proximity to main road etc., aiming to predict housing prices in the Northeast states of USA. The dataset contains no missing values and includes several categorical features.

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 housing prices data.

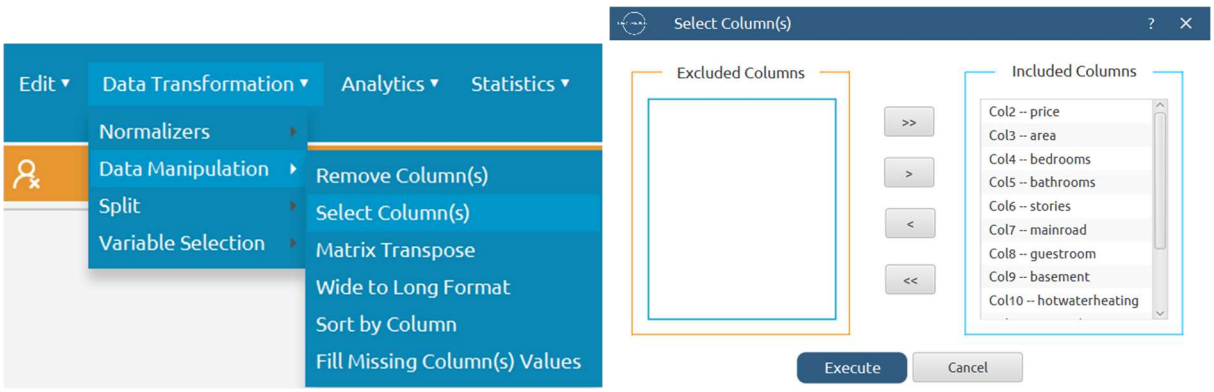


The data will appear on the left spreadsheet.

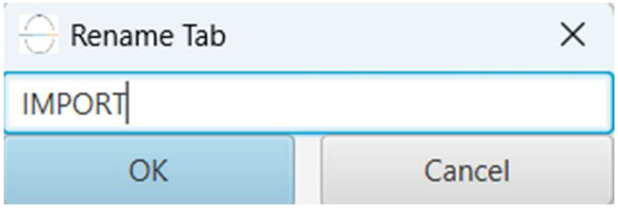
	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (I)	Col13 (S)	Col14 (S)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
1		13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
2		12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
3		12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
4		12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
5		11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
6		10850000	7500	3	3	1	yes	no	yes	no	yes	2	yes	semi-furnished
7		10150000	8580	4	3	4	yes	no	no	no	yes	2	yes	semi-furnished
8		10150000	16200	5	3	2	yes	no	no	no	no	0	no	unfurnished
9		9870000	8100	4	1	2	yes	yes	yes	no	yes	2	yes	furnished
10		9800000	5750	3	2	4	yes	yes	no	no	yes	1	yes	unfurnished
11		9800000	13200	3	1	2	yes	no	yes	no	yes	2	yes	furnished
12		9681000	6000	4	3	2	yes	yes	yes	yes	no	2	no	semi-furnished
13		9310000	6550	4	2	2	yes	no	no	no	yes	1	yes	semi-furnished
14		9240000	3500	4	2	2	yes	no	no	yes	no	2	no	furnished
15		9240000	7800	3	2	2	yes	no	no	no	no	0	yes	semi-furnished

Step 2: Manipulate data

In this dataset there are not any empty values and the categorical features have been encoded, so we can select all the columns to be used. On the menu click on *Data Transformation → Data Manipulation → Select Column(s)* and select all columns.



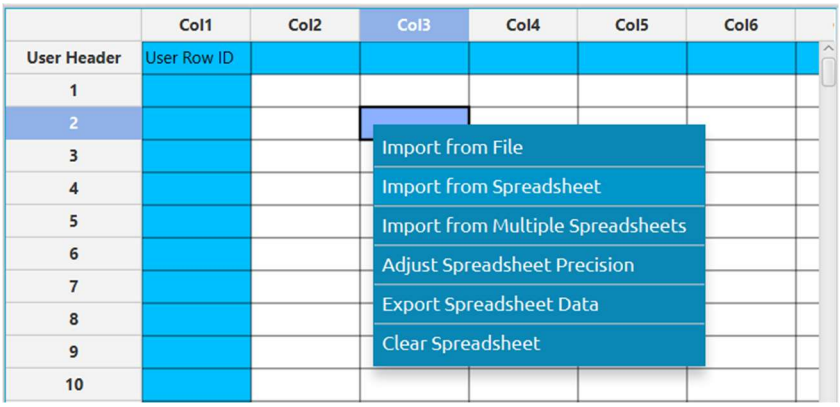
All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



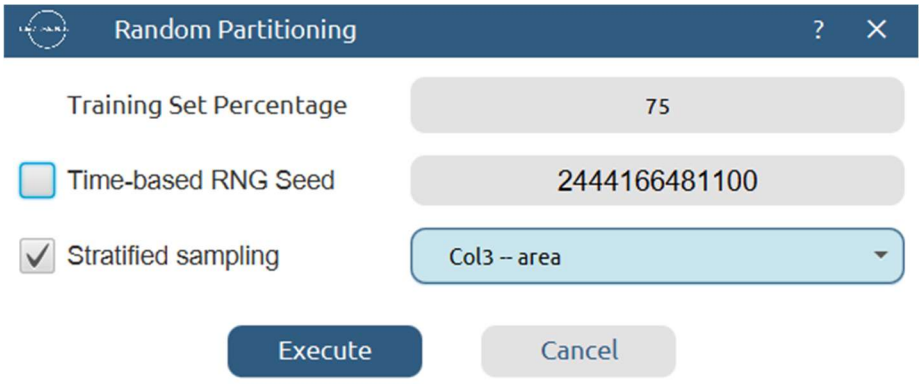
Step 3: Split data

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 "IMPORT" 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.

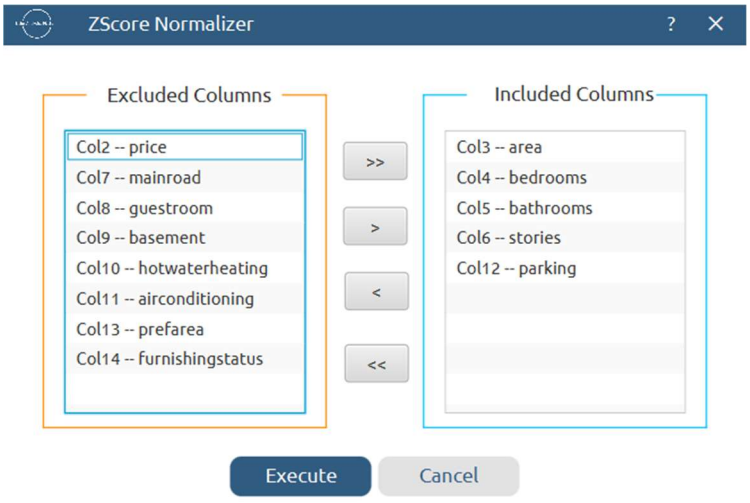
Step 4: Normalize the training set

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

Import 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 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (I)	Col13 (S)	Col14 (S)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
1		13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
2		12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
3		12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
4		11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
5		10850000	7500	3	3	1	yes	no	yes	no	yes	2	yes	semi-furnished
6		10150000	8580	4	3	4	yes	no	no	no	yes	2	yes	semi-furnished
7		10150000	16200	5	3	2	yes	no	no	no	no	0	no	unfurnished
8		9800000	13200	3	1	2	yes	no	yes	no	yes	2	yes	furnished
9		9681000	6000	4	3	2	yes	yes	yes	yes	no	2	no	semi-furnished
10		9310000	6550	4	2	2	yes	no	no	no	yes	1	yes	semi-furnished
11		9240000	3500	4	2	2	yes	no	no	yes	no	2	no	furnished
12		9240000	7800	3	2	2	yes	no	no	no	no	0	yes	semi-furnished
13		9100000	6600	4	2	2	yes	yes	yes	no	yes	1	yes	unfurnished
14		8960000	8500	3	2	4	yes	no	no	no	yes	2	no	furnished
15		8890000	4600	3	2	2	yes	yes	no	no	yes	2	no	furnished

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



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (D)	Col13 (S)	Col14 (S)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
1		1.33E+7	1.0188414	1.4513175	1.3206150	1.3606716	yes	no	no	no	yes	1.5725117	yes	furnished
2		1.225E+7	1.7183192	1.4513175	5.1243580	2.5081245	yes	no	no	no	yes	2.7591487	no	furnished
3		1.225E+7	2.1725255	0.0644276	1.3206150	0.2132186	yes	no	yes	no	no	1.5725117	yes	semi-furnished
4		1.141E+7	1.0188414	1.4513175	-0.5812566	0.2132186	yes	yes	yes	no	yes	1.5725117	no	furnished
5		1.085E+7	1.0551779	0.0644276	3.2224865	-0.9342343	yes	no	yes	no	yes	1.5725117	yes	semi-furnished
6		1.015E+7	1.5457208	1.4513175	3.2224865	2.5081245	yes	no	no	no	yes	1.5725117	yes	semi-furnished
7		1.015E+7	5.0067730	2.8382074	3.2224865	0.2132186	yes	no	no	no	no	-0.8007624	no	unfurnished
8		9800000.0	3.6441540	0.0644276	-0.5812566	0.2132186	yes	no	yes	no	yes	1.5725117	yes	furnished
9		9681000.0	0.3738684	1.4513175	3.2224865	0.2132186	yes	yes	yes	yes	no	1.5725117	no	semi-furnished
10		9310000.0	0.6236819	1.4513175	1.3206150	0.2132186	yes	no	no	no	yes	0.3858746	yes	semi-furnished
11		9240000.0	-0.7616474	1.4513175	1.3206150	0.2132186	yes	no	no	yes	no	1.5725117	no	furnished
12		9240000.0	1.1914398	0.0644276	1.3206150	0.2132186	yes	no	no	no	no	-0.8007624	yes	semi-furnished
13		9100000.0	0.6463922	1.4513175	1.3206150	0.2132186	yes	yes	yes	no	yes	0.3858746	yes	unfurnished
14		8960000.0	1.5093843	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	1.5725117	no	furnished
15		8890000.0	-0.2620204	0.0644276	1.3206150	0.2132186	yes	yes	no	no	yes	1.5725117	no	furnished

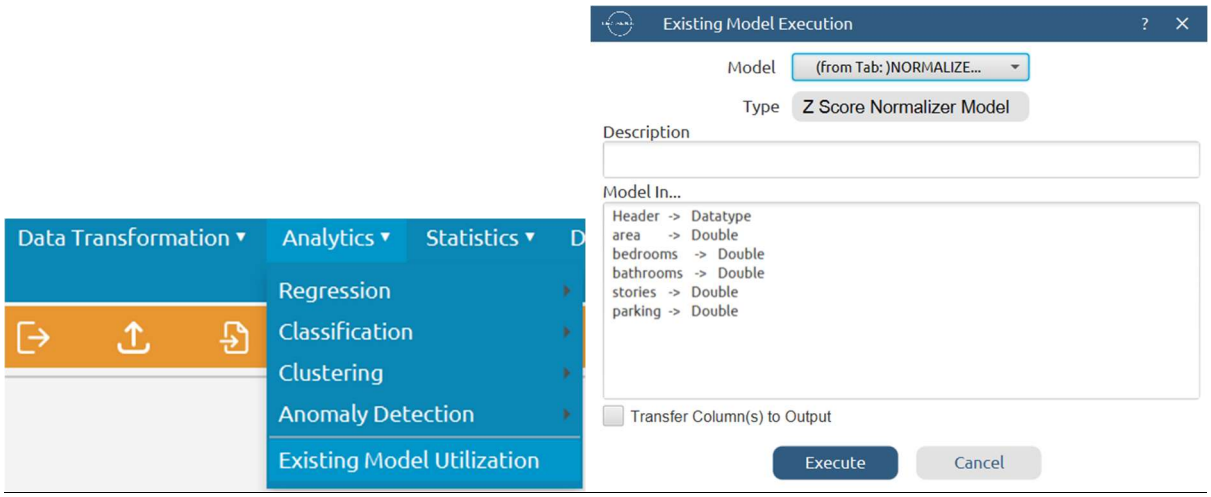
Step 5: Normalize the test set

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

Import 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 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (I)	Col13 (S)	Col14 (S)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
1		12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
2		9870000	8100	4	1	2	yes	yes	yes	no	yes	2	yes	furnished
3		9800000	5750	3	2	4	yes	yes	no	no	yes	1	yes	unfurnished
4		9100000	6000	4	1	2	yes	no	yes	no	no	2	no	semi-furnished
5		8855000	6420	3	2	2	yes	no	no	no	yes	1	yes	semi-furnished
6		8575000	8800	3	2	2	yes	no	no	no	yes	2	no	furnished
7		8120000	6840	5	1	2	yes	yes	yes	no	yes	1	no	furnished
8		8080940	7000	3	2	4	yes	no	no	no	yes	2	no	furnished
9		7840000	6360	3	2	4	yes	no	no	no	yes	0	yes	furnished
10		7700000	6000	4	2	4	yes	no	no	no	no	2	no	semi-furnished
11		7560000	6000	3	2	3	yes	no	no	no	yes	0	no	semi-furnished
12		7525000	6000	3	2	4	yes	no	no	no	yes	1	no	furnished
13		7490000	6600	3	1	4	yes	no	no	no	yes	3	yes	furnished
14		7350000	6000	3	1	2	yes	no	no	no	yes	1	no	unfurnished
15		7343000	11440	4	1	2	yes	no	yes	no	no	1	yes	semi-furnished

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.

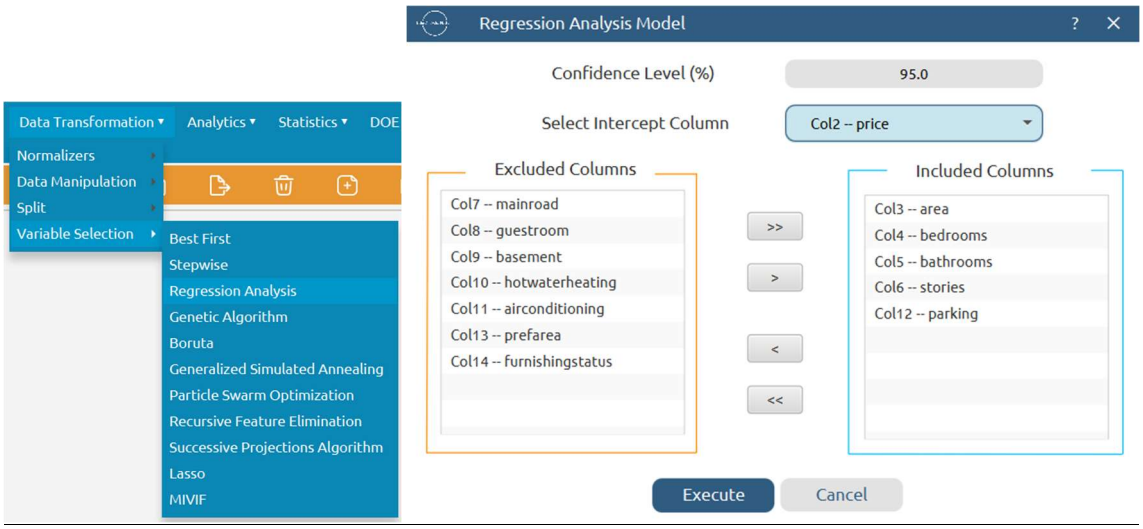
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (D)	Col13 (S)	Col14 (S)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
1		1.2215E+7	1.0551779	1.4513175	1.3206150	0.2132186	yes	no	yes	no	yes	2.7591487	yes	furnished
2		9870000.0	1.3277017	1.4513175	-0.5812566	0.2132186	yes	yes	yes	no	yes	1.5725117	yes	furnished
3		9800000.0	0.2603169	0.0644276	1.3206150	2.5081245	yes	yes	no	no	yes	0.3858746	yes	unfurnished
4		9100000.0	0.3738684	1.4513175	-0.5812566	0.2132186	yes	no	yes	no	no	1.5725117	no	semi-furnished
5		8855000.0	0.5646351	0.0644276	1.3206150	0.2132186	yes	no	no	no	yes	0.3858746	yes	semi-furnished
6		8575000.0	1.6456462	0.0644276	1.3206150	0.2132186	yes	no	no	no	yes	1.5725117	no	furnished
7		8120000.0	0.7554018	2.8382074	-0.5812566	0.2132186	yes	yes	yes	no	yes	0.3858746	no	furnished
8		8080940.0	0.8280748	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	1.5725117	no	furnished
9		7840000.0	0.5373827	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	-0.8007624	yes	furnished
10		7700000.0	0.3738684	1.4513175	1.3206150	2.5081245	yes	no	no	no	no	1.5725117	no	semi-furnished
11		7560000.0	0.3738684	0.0644276	1.3206150	1.3606716	yes	no	no	no	yes	-0.8007624	no	semi-furnished
12		7525000.0	0.3738684	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	0.3858746	no	furnished
13		7490000.0	0.6463922	0.0644276	-0.5812566	2.5081245	yes	no	no	no	yes	2.7591487	yes	furnished
14		7350000.0	0.3738684	0.0644276	-0.5812566	0.2132186	yes	no	no	no	yes	0.3858746	no	unfurnished
15		7343000.0	2.8447509	1.4513175	-0.5812566	0.2132186	yes	no	yes	no	no	0.3858746	yes	semi-furnished

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".

Then do regression analysis with the "price" column as the intercept: *Data Transformation → Variable Selection → Regression Analysis*



The results will appear on the right spreadsheet.

	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)
User Header	User Row ID						
1		Regression Statistics					
2		Multiple R	0.7518375				
3		R Square	0.5652597				
4		Adjusted R Square	0.5598659				
5		Standard Error	1248015.1564352				
6		Observations	409				
7							
8			Degrees of Freedom	Sum of Squares	Mean Square	F-statistic	Significance F
9		Regression	5	816136572715244.4	163227314543048.88	104.7980294	0E-7
10		Residual	403	627689357768876.5	1557541830692.0012		
11		Total	408	1443825930484121			

	Coefficients	Standard Error	t-statistic	P-value	Lower 95.0%	Upper 95.0%
price	4794101.4669927	61710.3778565	77.6871190	0E-7	4672787.0141360	4915415.9198493
area	725338.6813522	66928.8696232	10.8374560	0E-7	593765.3628815	856911.9998229
bedrooms	145289.2157324	71335.4086560	2.0367055	0.0423327	5053.2234892	285525.2079755
bathrooms	600243.8095350	69372.1642081	8.6525167	0E-7	463867.2965931	736620.3224769
stories	454595.2568824	69136.3370380	6.5753448	0E-7	318682.3490115	590508.1647533
parking	289270.9541337	66944.1353581	4.3210798	0.0000196	157667.6252444	420874.2830230

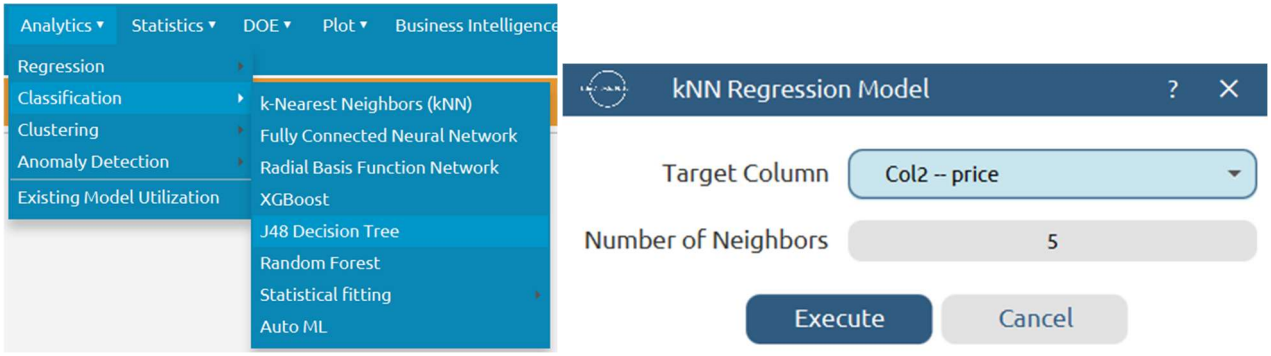
According to the p-value, all numeric features are significant.

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.

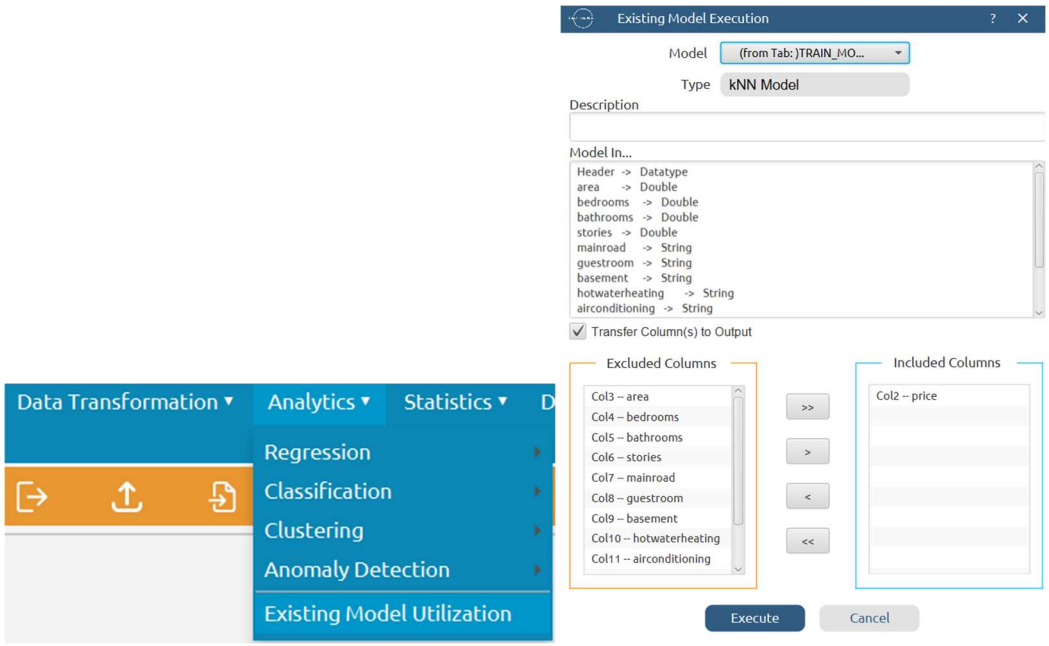
	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)	Col8 (S)	Col9 (D)	Col10 (S)	Col11 (D)	Col12 (S)	Col13 (D)
User Header	User Row ID	price	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5
1		1.33E+7	13190135.4258625	Entry 1	0.0	Entry 48	0.0741518	Entry 27	0.5037797	Entry 60	0.7149031	Entry 59	0.7154286
2		1.225E+7	12230912.4929205	Entry 2	0.0	Entry 30	0.7453612	Entry 14	0.7868234	Entry 35	0.9652797	Entry 47	1.1702728
3		1.225E+7	12207506.3279932	Entry 3	0.0	Entry 105	0.5148770	Entry 213	0.5671877	Entry 112	0.8205503	Entry 133	0.8610929
4		1.141E+7	11367706.3567307	Entry 4	0.0	Entry 109	0.3838179	Entry 311	0.4556191	Entry 18	0.5354106	Entry 19	0.5698686
5		1.085E+7	10827086.7446985	Entry 5	0.0	Entry 24	0.6445214	Entry 93	0.6675118	Entry 85	0.7991377	Entry 141	0.9478821
6		1.015E+7	10144686.7366917	Entry 6	0.0	Entry 10	0.8287538	Entry 114	0.9871132	Entry 1	1.1085157	Entry 41	1.1095874
7		1.015E+7	10124847.1274158	Entry 7	0.0	Entry 304	0.9259782	Entry 368	1.0381417	Entry 324	1.0703802	Entry 159	1.1159519
8		9800000.0	9782476.2765054	Entry 8	0.0	Entry 49	0.4922288	Entry 65	0.6074602	Entry 33	0.8129288	Entry 92	0.8834637
9		9681000.0	9665003.9798794	Entry 9	0.0	Entry 148	1.0375021	Entry 87	1.1063784	Entry 171	1.1349125	Entry 160	1.2312254
10		9310000.0	9301589.1774574	Entry 10	0.0	Entry 114	0.4948162	Entry 6	0.8287538	Entry 24	1.0026940	Entry 27	1.0068008
11		9240000.0	9216503.8808104	Entry 11	0.0	Entry 208	0.5381597	Entry 252	0.7639039	Entry 242	1.0308165	Entry 67	1.0834860
12		9240000.0	9200369.5483224	Entry 12	0.0	Entry 231	0.4480566	Entry 172	0.4492198	Entry 203	0.4791330	Entry 173	0.5305526
13		9100000.0	9087074.6591109	Entry 13	0.0	Entry 113	0.5433330	Entry 20	1.0541009	Entry 131	1.1008984	Entry 80	1.1334559
14		8960000.0	8953988.8325067	Entry 14	0.0	Entry 30	0.2524350	Entry 35	0.4518809	Entry 68	0.7455611	Entry 2	0.7868234
15		8890000.0	8879184.1653186	Entry 15	0.0	Entry 44	0.7517627	Entry 62	0.7517627	Entry 91	0.7886821	Entry 40	0.7922418

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 "price" to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

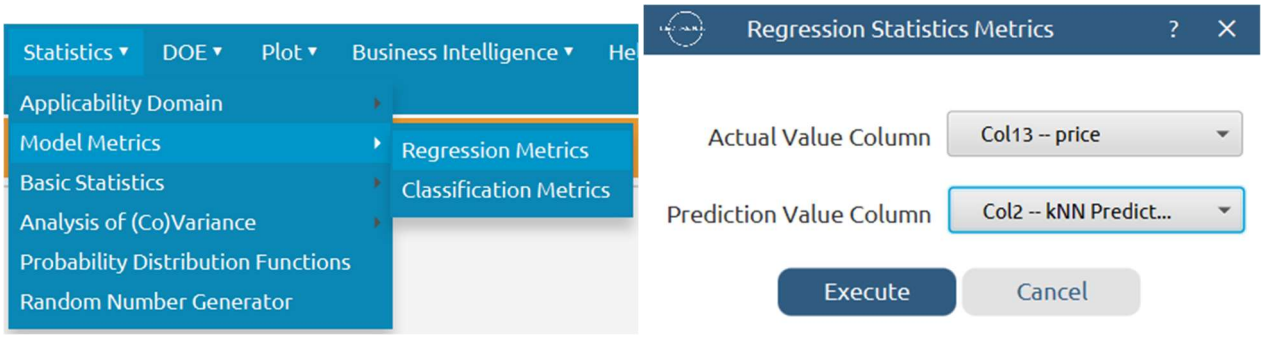
	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)	Col11 (S)	Col12 (D)	Col13 (D)
User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5	price
1		8047916.0234091	Entry 8	0.6661205	Entry 65	0.8136942	Entry 49	0.8917726	Entry 20	1.0562297	Entry 33	1.0853698	1.2215E+7
2		7570782.1194475	Entry 20	0.3507424	Entry 61	0.4492198	Entry 115	0.7366268	Entry 4	1.0011308	Entry 109	1.0805931	9870000.0
3		7080697.9444927	Entry 31	1.1056799	Entry 56	1.1067204	Entry 78	1.1071395	Entry 120	1.1549601	Entry 13	1.2290159	9800000.0
4		4949627.2202772	Entry 234	0.4385152	Entry 70	0.4404276	Entry 176	0.4422726	Entry 192	0.4532460	Entry 147	0.6758817	9100000.0
5		7989873.6113757	Entry 10	0.2501654	Entry 114	0.5519875	Entry 6	0.8671784	Entry 226	1.0080723	Entry 24	1.0327836	8855000.0
6		5742293.8141349	Entry 144	0.4250388	Entry 47	0.4535218	Entry 132	0.4718113	Entry 53	0.6096716	Entry 199	0.6146352	8575000.0
7		7378479.7212269	Entry 109	0.2913947	Entry 4	0.4186375	Entry 311	0.5202557	Entry 18	0.6068574	Entry 19	0.6217316	8120000.0
8		8123951.0110661	Entry 14	0.1049318	Entry 30	0.2864866	Entry 35	0.4224982	Entry 68	0.7504678	Entry 144	0.7872847	8080940.0
9		7772407.0363432	Entry 59	0.3333598	Entry 60	0.3334772	Entry 48	0.7861651	Entry 1	0.7896544	Entry 27	0.7929531	7840000.0
10		4910518.5703297	Entry 75	0.3333333	Entry 286	0.7239178	Entry 111	0.7869428	Entry 233	0.7975559	Entry 356	0.8095266	7700000.0
11		5381658.5330028	Entry 66	0.3333333	Entry 103	0.3379986	Entry 150	0.4714186	Entry 169	0.4749668	Entry 226	0.4814704	7560000.0
12		7708746.7216447	Entry 35	0.25	Entry 14	0.3764257	Entry 30	0.4665339	Entry 68	0.4969872	Entry 47	0.7121982	7525000.0
13		8789858.9085659	Entry 48	0.6293768	Entry 1	0.6317625	Entry 27	1.0377752	Entry 60	1.1055637	Entry 59	1.1056133	7490000.0
14		3700800.7923174	Entry 219	0.3351634	Entry 344	0.3494593	Entry 195	0.4167443	Entry 314	0.4518809	Entry 335	0.4768866	7350000.0
15		6879389.3960511	Entry 3	0.5435451	Entry 105	0.6174086	Entry 133	0.6533702	Entry 112	0.6693054	Entry 213	0.6999597	7343000.0

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		1376770395343.2412	1173358.5962285	843696.6539372	0.5942399

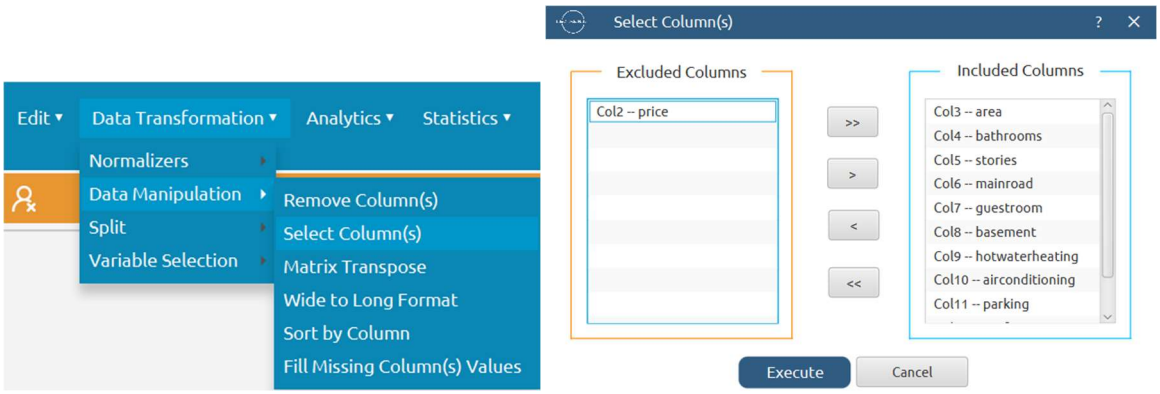
Step 10: Reliability check for each record of the test set

Step 10.a: Create the domain

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

Import data into the input spreadsheet of the "EXCLUDE_PRICE" 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 "price": *Data Transformation → Data Manipulation → Select Column(s)*

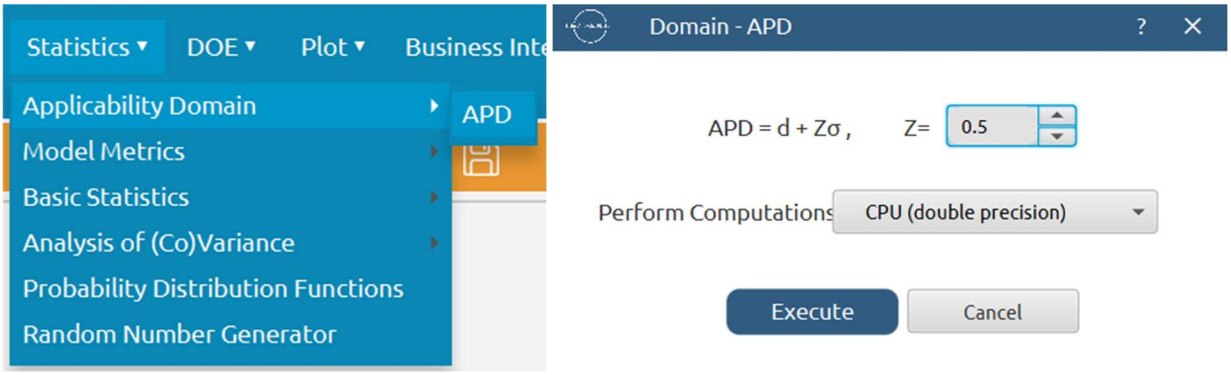


The results will appear on the output spreadsheet.

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_PRICE" 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.

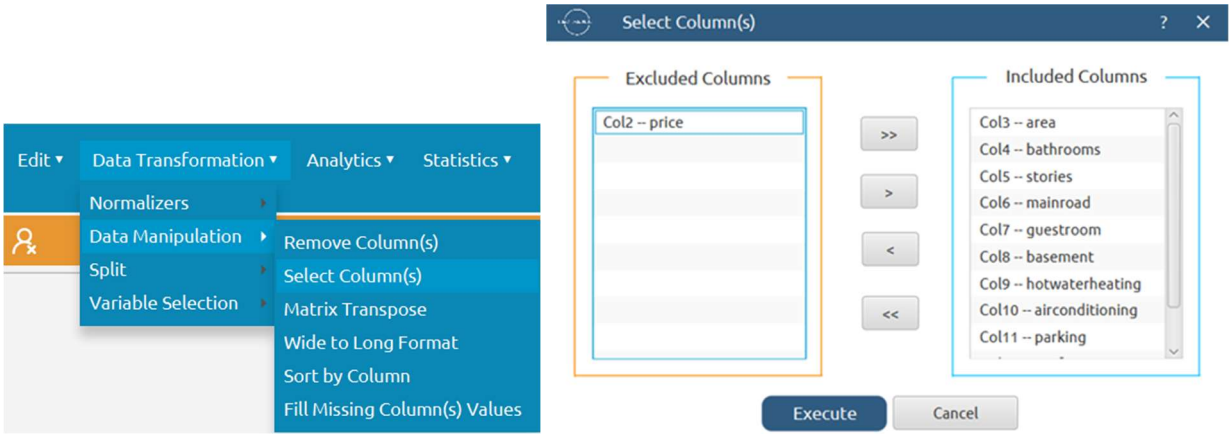
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	2.3433555	reliable
2		0.0	2.3433555	reliable
3		0.0	2.3433555	reliable
4		0.0	2.3433555	reliable
5		0.0	2.3433555	reliable
6		0.0	2.3433555	reliable
7		0.0	2.3433555	reliable
8		0.0	2.3433555	reliable
9		0.0	2.3433555	reliable
10		0.0	2.3433555	reliable
11		0.0	2.3433555	reliable
12		0.0	2.3433555	reliable
13		0.0	2.3433555	reliable
14		0.0	2.3433555	reliable
15		0.0	2.3433555	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_PRICE_TEST_SET".

Import data into the input spreadsheet of the "EXCLUDE_PRICE_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 "price": *Data Transformation → Data Manipulation → Select Column(s)*

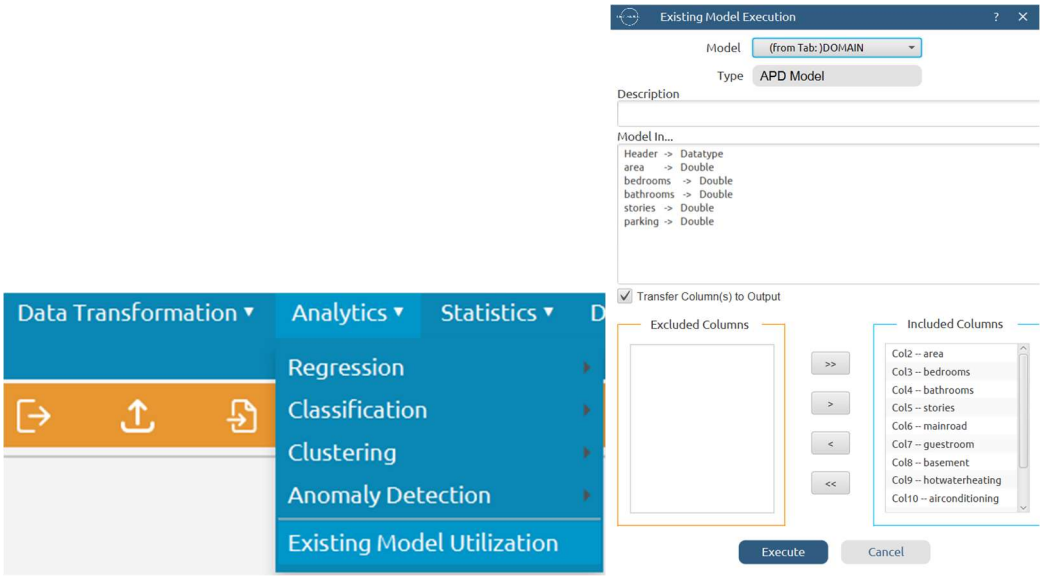


The results will appear on the output spreadsheet.

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_PRICE_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. There are no unreliable samples in the test set.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (S)	Col13 (S)	Col14 (D)	Col15 (S)	Col16 (S)
User Header	User Row ID	Domain	APD	Prediction	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
1		1.2642139	2.3433555	reliable	1.0551779	1.4513175	1.3206150	0.2132186	yes	no	yes	no	yes	2.7591487	yes	furnished
2		0.2271032	2.3433555	reliable	1.3277017	1.4513175	-0.5812566	0.2132186	yes	yes	yes	no	yes	1.5725117	yes	furnished
3		0.1135516	2.3433555	reliable	0.2603169	0.0644276	1.3206150	2.5081245	yes	yes	no	no	yes	0.3858746	yes	unfurnished
4		0.6449730	2.3433555	reliable	0.3738684	1.4513175	-0.5812566	0.2132186	yes	no	yes	no	no	1.5725117	no	semi-furnished
5		0.1317198	2.3433555	reliable	0.5646351	0.0644276	1.3206150	0.2132186	yes	no	no	no	yes	0.3858746	yes	semi-furnished
6		0.5268793	2.3433555	reliable	1.6456462	0.0644276	1.3206150	0.2132186	yes	no	no	no	yes	1.5725117	no	furnished
7		1.3549532	2.3433555	reliable	0.7554018	2.8382074	-0.5812566	0.2132186	yes	yes	yes	no	yes	0.3858746	no	furnished
8		0.2157480	2.3433555	reliable	0.8280748	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	1.5725117	no	furnished
9		0.1635143	2.3433555	reliable	0.5373827	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	-0.8007624	yes	furnished
10		1.1590449	2.3433555	reliable	0.3738684	1.4513175	1.3206150	2.5081245	yes	no	no	no	no	1.5725117	no	semi-furnished
11		0.0	2.3433555	reliable	0.3738684	0.0644276	1.3206150	1.3606716	yes	no	no	no	yes	-0.8007624	no	semi-furnished
12		0.0	2.3433555	reliable	0.3738684	0.0644276	1.3206150	2.5081245	yes	no	no	no	yes	0.3858746	no	furnished
13		1.2004649	2.3433555	reliable	0.6463922	0.0644276	-0.5812566	2.5081245	yes	no	no	no	yes	2.7591487	yes	furnished
14		0.0	2.3433555	reliable	0.3738684	0.0644276	-0.5812566	0.2132186	yes	no	no	no	yes	0.3858746	no	unfurnished
15		1.8040508	2.3433555	reliable	2.8447509	1.4513175	-0.5812566	0.2132186	yes	no	yes	no	no	0.3858746	yes	semi-furnished

Final Isalos Workflow

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