



Housing Prices (Regression)

The goal of this study is to train a model in order to predict housing prices. The dataset used in this case study is found in <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset/data> and has 13 features and 545 samples. This dataset contains information on certain factors like house area, bedrooms, furnished, nearness 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. Categorical features contain multiple levels, and the data was transformed to corresponding numeric codes, as detailed below:

mainroad:

- No (0)
- Yes (1)

guestroom:

- No (0)
- Yes (1)

basement:

- No (0)
- Yes (1)

hotwaterheating:

- No (0)
- Yes (1)

airconditioning:

- No (0)
- Yes (1)

prefarea:

- No (0)
- Yes (1)

furnishingstatus:

- unfurnished (0)

- semi - furnished (1)
- furnished (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 housing price data.

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

Step 2: Manipulate data

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

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

The screenshot shows the Isalos Analytics Platform interface. In the top navigation bar, 'Data Transformation' is selected. Under 'Data Manipulation', the 'Select Column(s)' option is highlighted. A secondary window titled 'Select Column(s)' is open, showing a list of columns from Col2 to Col10. The columns 'Col2 -- price', 'Col3 -- area', 'Col4 -- bedrooms', 'Col5 -- bathrooms', 'Col6 -- stories', 'Col7 -- mainroad', 'Col8 -- guestroom', 'Col9 -- basement', and 'Col10 -- hotwaterheating' are listed under 'Included Columns'. The 'Excluded Columns' section is empty. At the bottom of the window are 'Execute' and 'Cancel' buttons.

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".

The screenshot shows the Isalos Analytics Platform interface with two tabs: 'IMPORT' and 'TRAIN_TEST_SPLIT'. The 'TRAIN_TEST_SPLIT' tab is active and displays a large table with columns labeled 'Col1' through 'Col9'. The table has a header row and approximately 21 data rows. The data includes various numerical values such as 13300000.0, 7420.0, 4.0, etc. The 'IMPORT' tab is also visible at the bottom left.

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

The screenshot shows the Isalos Analytics Platform interface. The top navigation bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Data Transformation menu is open, showing sub-options: Import, Normalizers, Data Manipulation, Split, Variable Selection, and Random Partitioning. The Random Partitioning option is highlighted.

The main workspace displays a spreadsheet with columns labeled Col1 through Col8. The first row is the User Header. The data rows show various values for price, area, bedrooms, bathrooms, stories, mainroad, guestroom, and basements. Row 6 is selected, and its entire row is highlighted in blue.

A modal dialog box titled "Random Partitioning" is open. It contains the following settings:

- Training Set Percentage: 75
- Time-based RNG Seed: 2444166481100
- Stratified sampling: checked, with "Col3 -- area" selected

At the bottom of the dialog are Execute and Cancel buttons.

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface with the TRAIN_TEST_SPLIT tab selected in the ribbon. The main workspace displays two separate spreadsheets side-by-side.

The left spreadsheet has columns Col1 through Col9. The data rows correspond to the rows from the original dataset, with the first few rows showing price, area, bedrooms, bathrooms, stories, mainroad, guestroom, and basements.

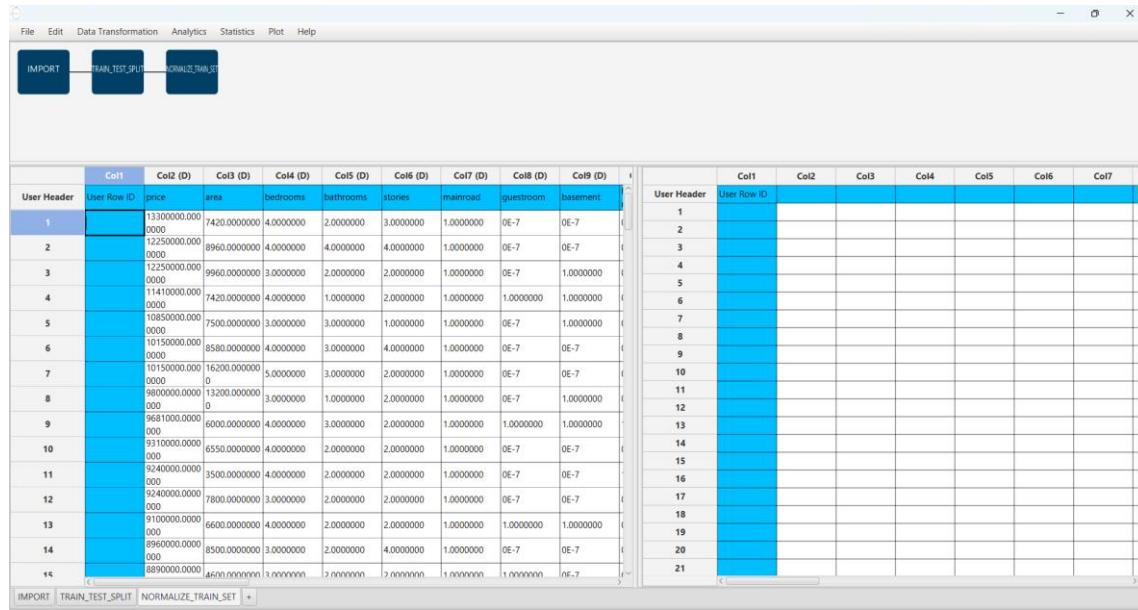
The right spreadsheet also has columns Col1 through Col9. The data rows correspond to the rows from the original dataset, with the first few rows showing price, area, bedrooms, bathrooms, stories, mainroad, guestroom, and basements.

Step 4: Normalize the training set

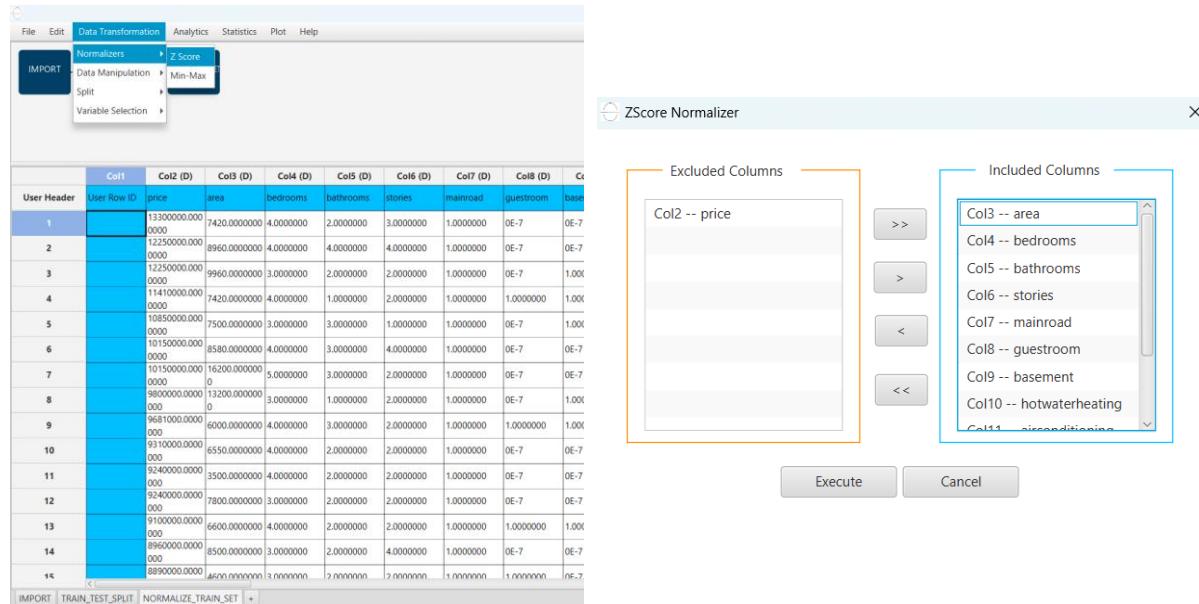
Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TRAIN_SET".

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

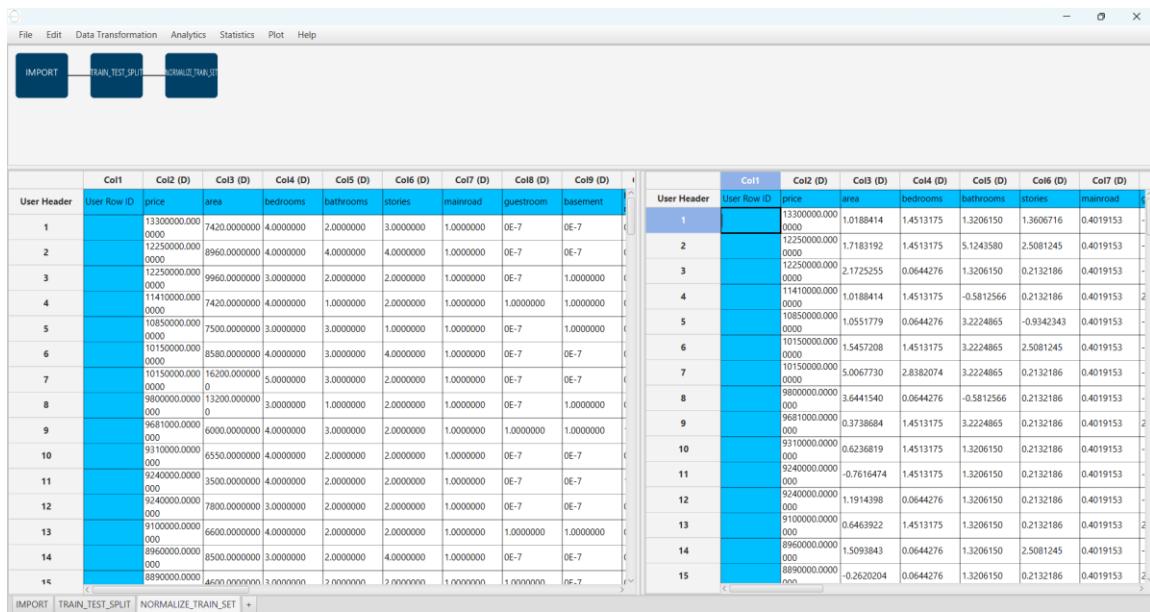
and then choosing "Import from SpreadSheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT: Training Set".



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



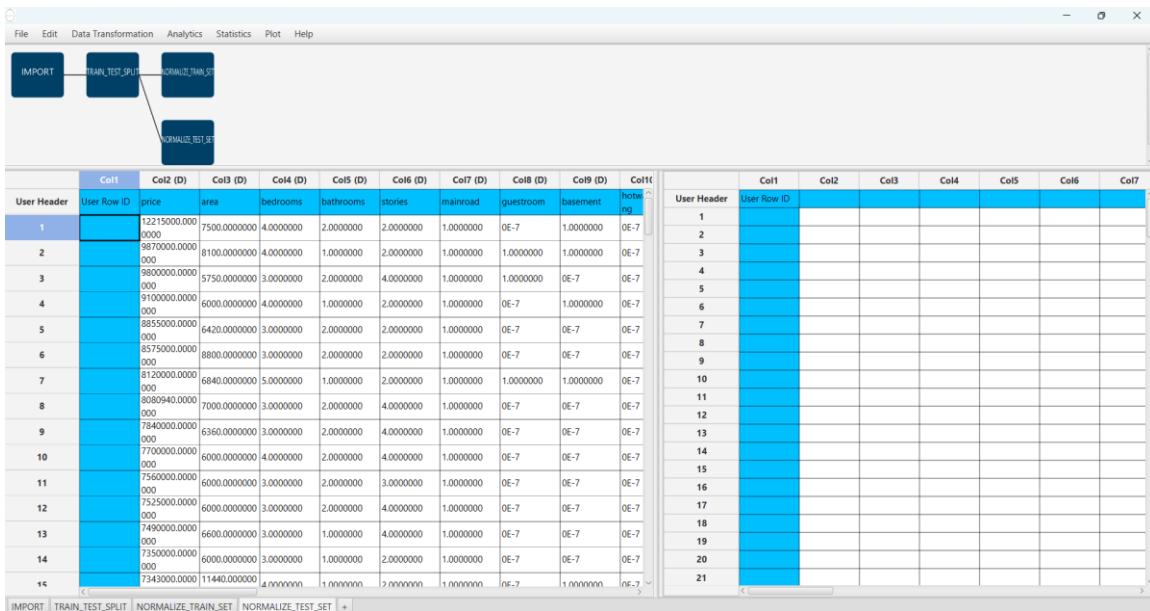
The results will appear on the output spreadsheet.



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".



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

The screenshot shows the Isalos Analytics Platform interface. On the left, there is a flowchart with nodes: 'IMPORT', 'TRAIN_TEST_SPLIT', and 'NORMALIZE_TRAIN_SET'. The 'NORMALIZE_TRAIN_SET' node is highlighted with a red box. On the right, a data spreadsheet is displayed with columns: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D). The data consists of 15 rows of house price information. Below the spreadsheet is a navigation bar with tabs: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, and a '+' button.

Existing Model Execution

Model: (from Tab:)NORMALIZE_TRAIN_SET

Type: Z Score Normalizer Model

Description:

Model Input:

- Header -> Datatype
- area -> Double
- bedrooms -> Double
- bathrooms -> Double
- stories -> Double
- mainroad -> Double
- guestroom -> Double
- basement -> Double
- hotwaterheating -> Double
- parking -> Double

Transfer Column(s) to Output

Execute Cancel

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. On the left, there is a flowchart with nodes: 'IMPORT', 'TRAIN_TEST_SPLIT', 'NORMALIZE_TRAIN_SET', and 'NORMALIZE_TEST_SET'. The 'NORMALIZE_TEST_SET' node is highlighted with a red box. On the right, two data spreadsheets are shown side-by-side. Both spreadsheets have columns: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), Col9 (D), and Col10. The data consists of 15 rows of house price information. Below each spreadsheet is a navigation bar with tabs: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, and a '+' button.

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 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 is a data flow diagram. The flow starts with an IMPORT node, followed by a TRAIN_TEST_SPLIT node, then two parallel paths: one through a NORMALIZE_TRAIN_SET node and another through a FEATURE_SELECTION_REGRESSION node. These paths converge at a UNION node, which then leads to a FINAL_OUTPUT node. Below the flow diagram is a large spreadsheet table with 15 rows of data. The columns are labeled Col1 through Col10. The first row is the User Header. The data includes various numerical values such as price, area, bedrooms, bathrooms, stories, mainroad, guestroom, basement, and highway. To the right of the main window, there's a smaller preview window showing a subset of the data from rows 1 to 21.

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

This screenshot shows the Data Transformation menu open under the main menu. The "Variable Selection" option is selected, revealing three sub-options: Best First, Stepwise, and Regression Analysis. The "Regression Analysis" option is highlighted. To the right of the menu, a "Regression Analysis Model" dialog box is displayed. It contains fields for "Significance Level (α)" set to 0.05, "Select Intercept Column" set to "Col8 -- charges", and a list of "Included Columns" which includes "Col2 -- age", "Col3 -- sex", "Col4 -- bmi", "Col5 -- children", "Col6 -- smoker", and "Col7 -- region". There are also "Excluded Columns" and "Included Columns" buttons. At the bottom of the dialog are "Execute" and "Cancel" buttons. Below the menu and dialog, there's a smaller view of the same data spreadsheet as in the previous screenshot.

The results will appear on the output spreadsheet.

The screenshot shows a feature selection report in the NovaMechanics platform. The top part of the interface has a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, Help, and a search bar. Below the menu is a toolbar with buttons for Import, Train/Test Split, Normalize Train Set, Normalize Test Set, and Feature Selection Regression.

The main area displays two tables. The first table, titled 'User Header', contains statistical information such as R-squared (0.80801550), Adjusted R-squared (0.7979890152), and Observations (409.000000). The second table, titled 'Coefficients', lists features and their corresponding coefficients and p-values. Key features include price (p-value = 0.0), area (p-value = 0.0), bathrooms (p-value = 0.0), stories (p-value = 0.0), mainroad (p-value = 0.011078305445895668), guestroom (p-value = 0.010553358931198279), basement (p-value = 0.01994612978933753), hotwaterheating (p-value = 4.1889240314438027E-4), airconditioning (p-value = 0.0), parking (p-value = 7.015384208333902E-5), prefarea (p-value = 1.0032449728224924E-6), and furnishingstatus (p-value = 0.010503202918605318).

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

- price (p-value = 0.0)
- area (p-value = 0.0)
- bathrooms (p-value = 0.0)
- stories (p-value = 0.0)
- mainroad (p-value = 0.011078305445895668)
- guestroom (p-value = 0.010553358931198279)
- basement (p-value = 0.01994612978933753)
- hotwaterheating (p-value = 4.1889240314438027E-4)
- airconditioning (p-value = 0.0)
- parking (p-value = 7.015384208333902E-5)
- prefarea (p-value = 1.0032449728224924E-6)
- furnishingstatus (p-value = 0.010503202918605318)

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".

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a navigation bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the navigation bar is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and FEATURE_SELECTION_REGRESSION. Arrows indicate data flow from IMPORT to TRAIN_TEST_SPLIT, then to both NORMALIZE_TRAIN_SET and FEATURE_SELECTION_REGRESSION. Both of these nodes then point to NORMALIZE_TEST_SET and FEATURE_SELECTION_TRAIN_SET respectively. Below the workflow is a large spreadsheet view. The first row contains column headers: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), and Col9. The second row contains "User Header". Rows 1 through 14 represent data points, each with a User Row ID and various numerical values across the columns. A vertical scrollbar is visible on the right side of the spreadsheet.

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)".

This screenshot shows the "Select Column(s)" dialog box, which is part of the Data Manipulation process. The dialog has two main sections: "Excluded Columns" (left, orange border) containing "Col4 -- bedrooms", and "Included Columns" (right, blue border) containing "Col2 -- price", "Col3 -- area", "Col5 -- bathrooms", "Col6 -- stories", "Col7 -- mainroad", "Col8 -- guestroom", "Col9 -- basement", and "Col10 -- hotwaterheating". There are four buttons at the bottom: "Execute", "Cancel", and two arrows (">>>" and "<<") for moving columns between the lists.

The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
1	133000000.0000	1.0188414	1.4513175	1.3206150	1.3606716	0.4019153	-0.4771480	-0.740		
2	122500000.0000	1.7183192	1.4513175	5.1243580	2.5081245	0.4019153	-0.4771480	-0.740		
3	122500000.0000	2.1725255	0.6644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.340		
4	114100000.0000	1.0188414	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	1.340		
5	108500000.0000	1.0551779	0.6644276	3.2224865	-0.9342343	0.4019153	-0.4771480	1.340		
6	101500000.0000	1.5457208	1.4513175	3.2224865	2.5081245	0.4019153	-0.4771480	-0.740		
7	101500000.0000	5.0667730	2.8382074	3.2224865	0.2132186	0.4019153	-0.4771480	-0.740		
8	98000000.0000	3.6441540	0.6644276	-0.5812566	0.2132186	0.4019153	-0.4771480	1.340		
9	968100000.0000	0.3738684	1.4513175	3.2224865	0.2132186	0.4019153	2.0906616	1.340		
10	931000000.0000	0.6236819	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.740		
11	924000000.0000	0.7616474	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.740		
12	924000000.0000	1.1914398	0.6644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.740		
13	910000000.0000	0.6463922	1.4513175	1.3206150	0.2132186	0.4019153	2.0906616	1.340		
14	896000000.0000	1.3738684	0.6644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.740		

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
1	133000000.0000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441			
2	122500000.0000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441			
3	122500000.0000	2.1725255	0.6644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.34050		
4	114100000.0000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.34050			
5	108500000.0000	1.0551779	0.6644276	3.2224865	-0.9342343	0.4019153	-0.4771480	1.34050		
6	101500000.0000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441			
7	101500000.0000	5.0667730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441			
8	98000000.0000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.34050			
9	968100000.0000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.34050			
10	931000000.0000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441			
11	924000000.0000	0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441			
12	924000000.0000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441			
13	910000000.0000	0.6463922	1.3206150	0.2132186	0.4019153	2.0906616	1.34050			
14	896000000.0000	1.3738684	0.6644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441		

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".

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1	122150000.0000	1.0551779	1.4513175	1.3206150	0.2132186	0.4019153	-0.477		
2	987000000.0000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.090		
3	987000000.0000	0.2603169	0.6644276	1.3206150	2.5081245	0.4019153	2.090		
4	910000000.0000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.477		
5	885500000.0000	0.5646351	0.6644276	1.3206150	0.2132186	0.4019153	-0.477		
6	857500000.0000	1.6456462	0.6644276	1.3206150	0.2132186	0.4019153	-0.477		
7	812000000.0000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.090		
8	808094000.0000	0.8280748	0.6644276	1.3206150	2.5081245	0.4019153	-0.477		
9	784000000.0000	0.5373827	0.6644276	1.3206150	2.5081245	0.4019153	-0.477		
10	770000000.0000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.477		
11	756000000.0000	0.3738684	0.6644276	1.3206150	1.3606716	0.4019153	-0.477		
12	752500000.0000	0.3738684	0.6644276	1.3206150	2.5081245	0.4019153	-0.477		

User Header	User Row ID	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										
17										

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 Isalos Analytics Platform interface. At the top, the 'Data Transformation' menu is open under the 'IMPORT' section, with 'Select Column(s)' highlighted. A context menu is also open over a table, also showing 'Select Column(s)'. To the right, a 'Select Column(s)' dialog box is displayed, divided into 'Excluded Columns' (containing 'Col4 -- bedrooms') and 'Included Columns' (containing 'Col2 -- price', 'Col3 -- area', 'Col5 -- bathrooms', 'Col6 -- stories', 'Col7 -- mainroad', 'Col8 -- guestroom', 'Col9 -- basement', 'Col10 -- hotwaterheating', and 'Col11 -- airconditioning'). Buttons for 'Execute' and 'Cancel' are at the bottom.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
2		12215000.0000	1.0551779	1.4513175	1.3206150	0.2132186	0.4019153	-0.477
3		9870000.0000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.090
4		9800000.0000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	2.090
5		9100000.0000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.477
6		8855000.0000	0.5646351	0.0644276	1.3206150	0.2132186	0.4019153	-0.477
7		8575000.0000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.477
8		8120000.0000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.090
9		8089940.0000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.477
10		7840000.0000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.477
11		7700000.0000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.477
12		7525000.0000	0.3738684	0.0644276	1.3206150	2.5081245	0.4019153	-0.477

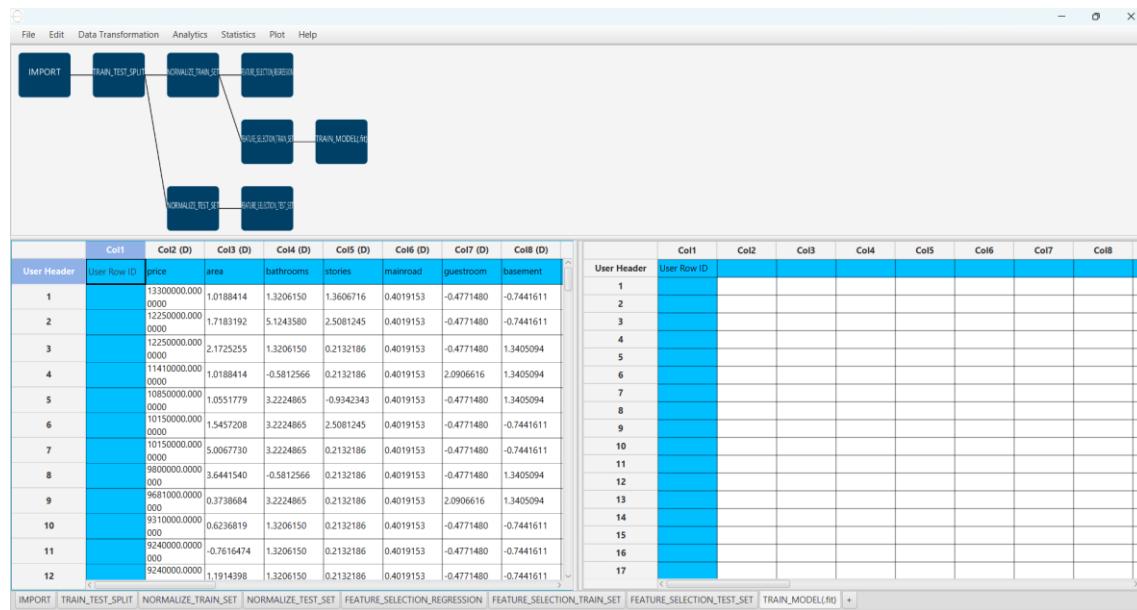
The results will appear on the output spreadsheet.

The screenshot shows the completed data flow in the Isalos Analytics Platform. The 'FEATURE_SELECTION_TRAIN_SET' tab has been selected, and the resulting normalized dataset is displayed in a table below. The columns are labeled Col1 through Col9. The data is identical to the one shown in the previous screenshot, with rows 1 through 12.

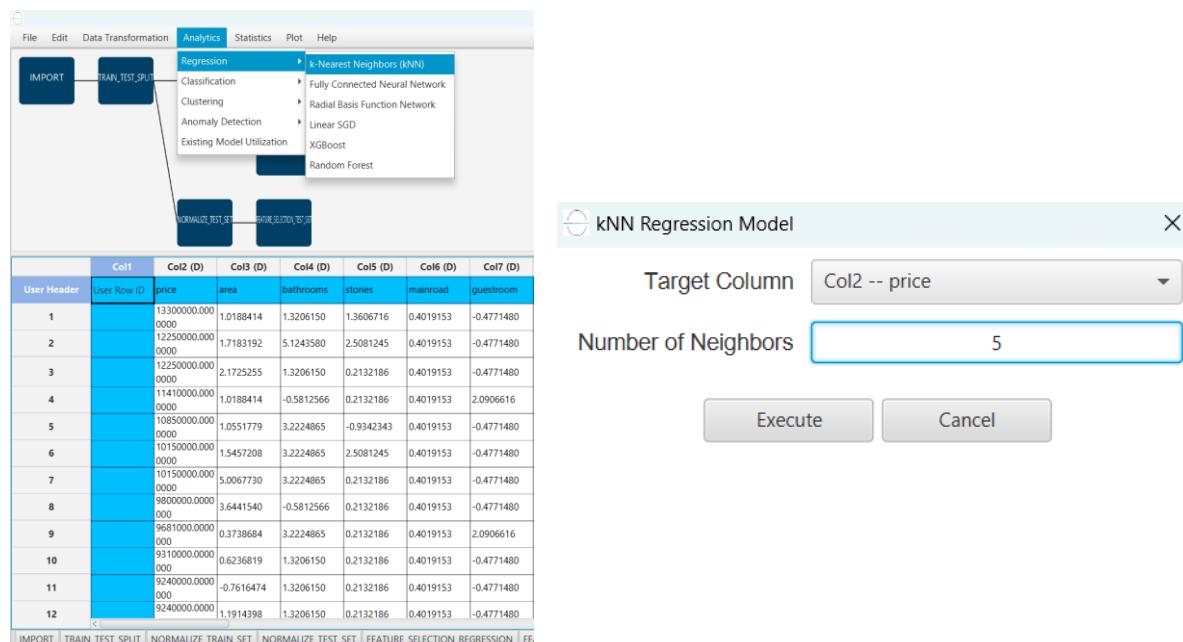
User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
2		12215000.0000	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094
3		9870000.0000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616
4		9800000.0000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	-0.7441611
5		9100000.0000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.4771480
6		8855000.0000	0.5646351	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480
7		8575000.0000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.7441611
8		8120000.0000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.0906616
9		8089940.0000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.7441611
10		7840000.0000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480
11		7700000.0000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.7441611
12		7525000.0000	0.3738684	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480

Step 9: Train the model

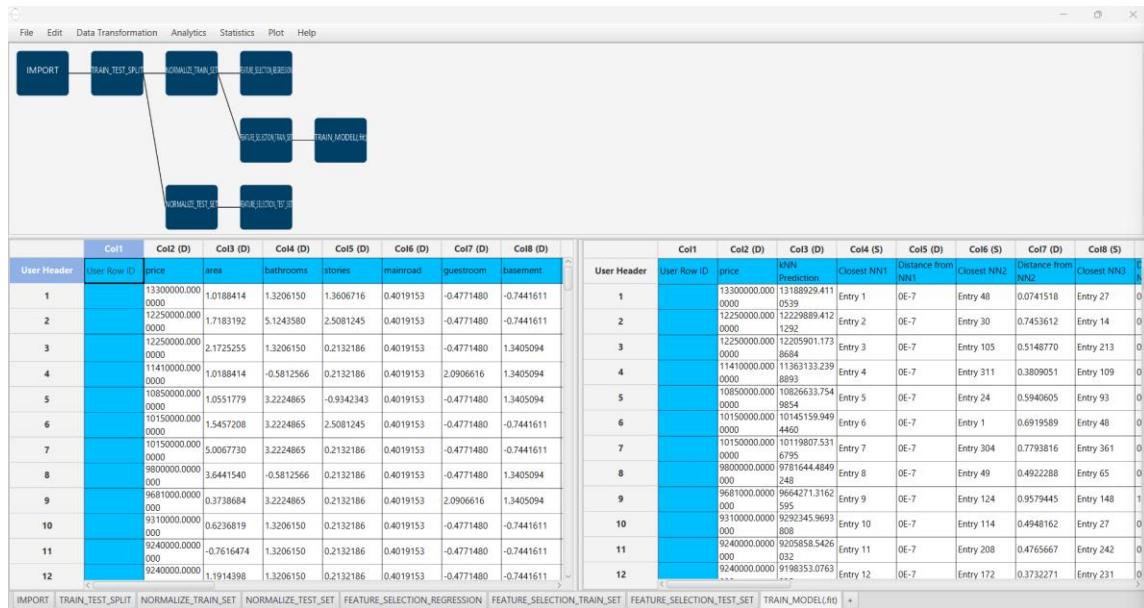
Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_MODEL(.fit)". Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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 "price" and the "Number of Neighbors" to 5.



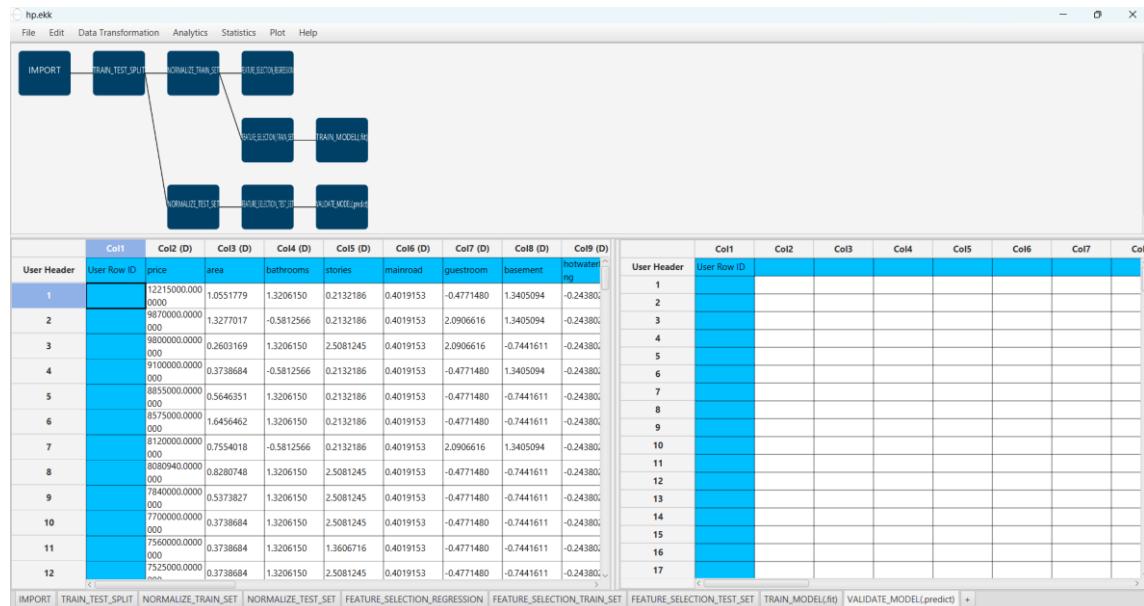
The predictions will appear on the output spreadsheet.



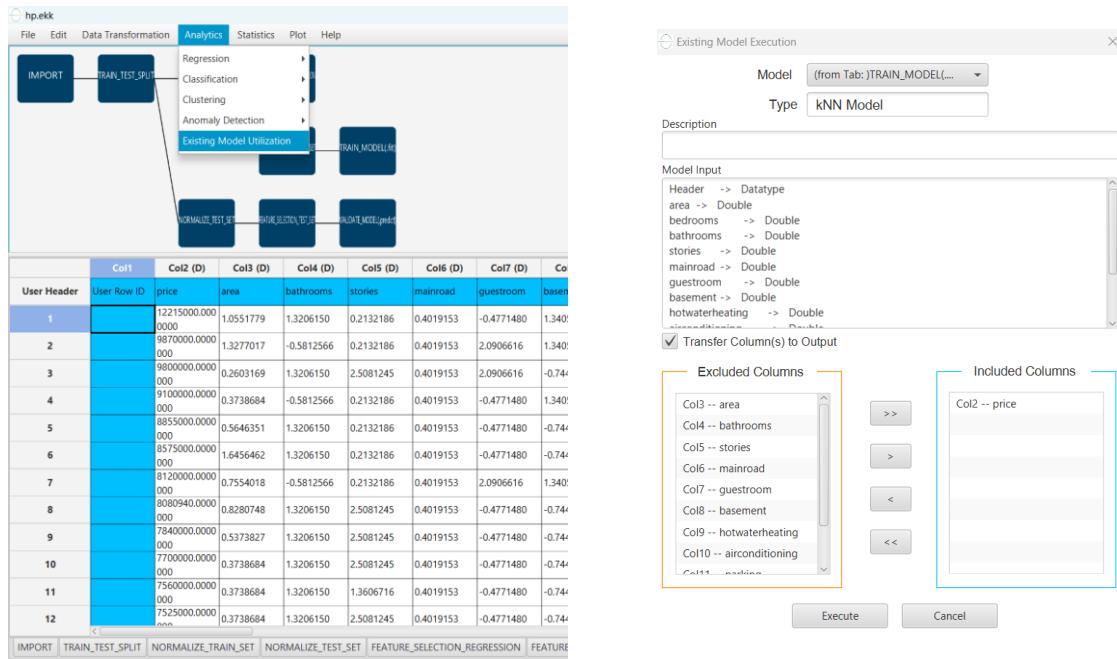
Step 10: Validate the model

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

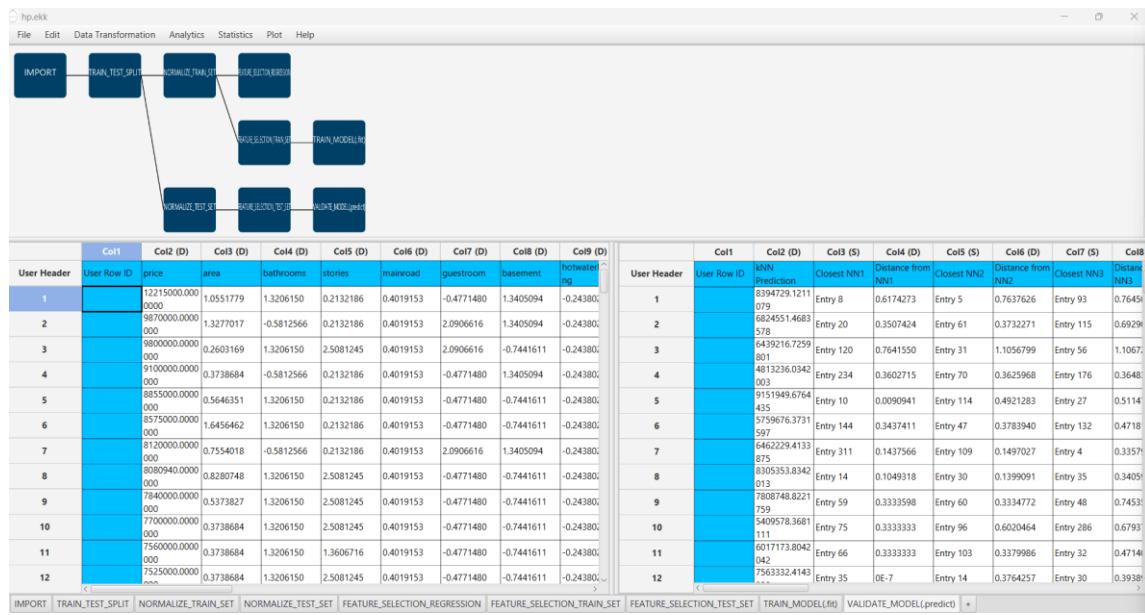
Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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



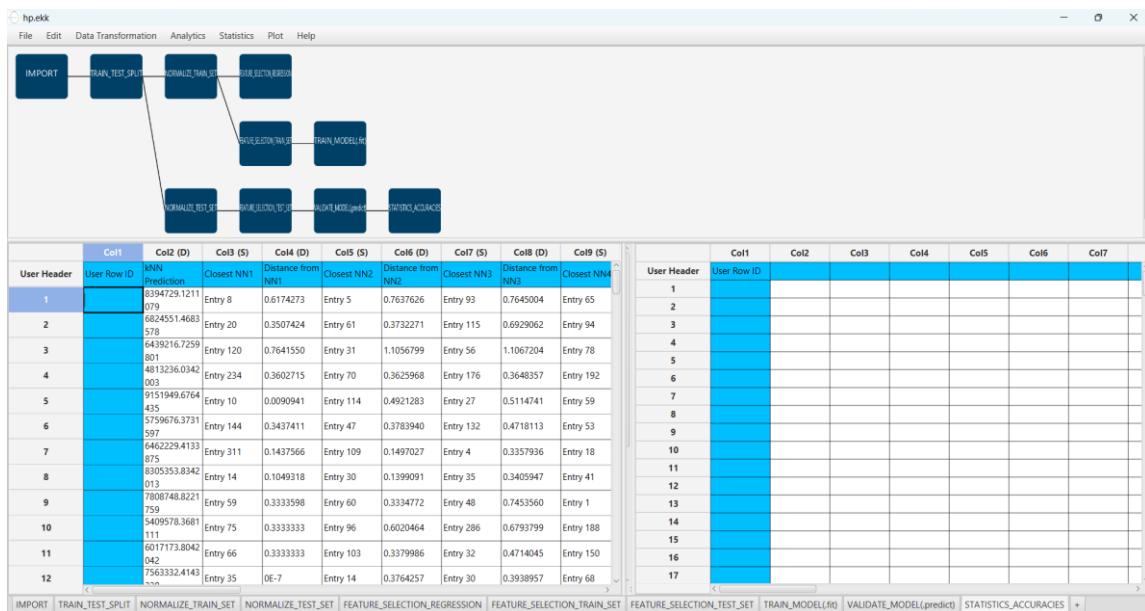
The predictions will appear on the output spreadsheet.



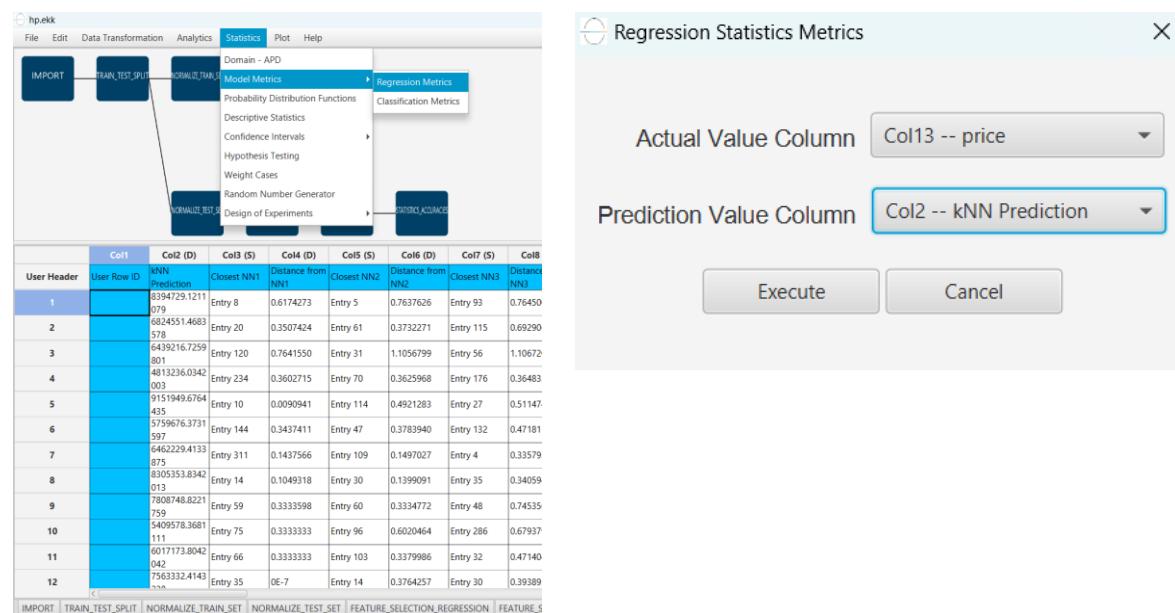
Step 11: Statistics calculation

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

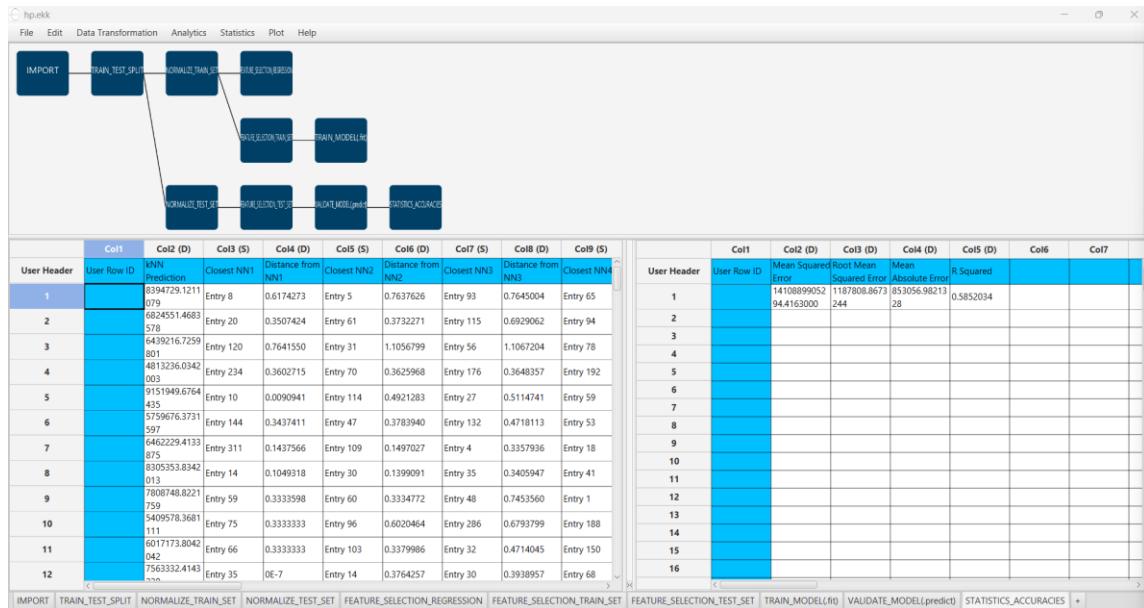
Import data into the input spreadsheet of the "STATISTICS_ACCURACIES" tab from the output of the "VALIDATE_MODEL(.predict)" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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



The results will appear on the output spreadsheet.

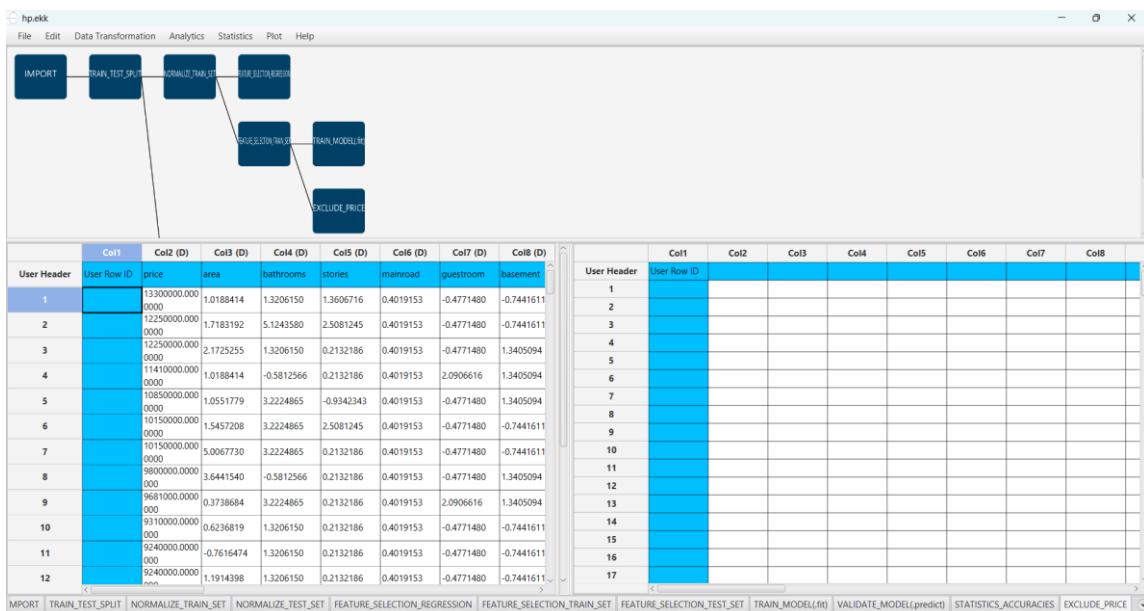


Step 12: Reliability check of each record of the test set

Step 12.a: Create the domain

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

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



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

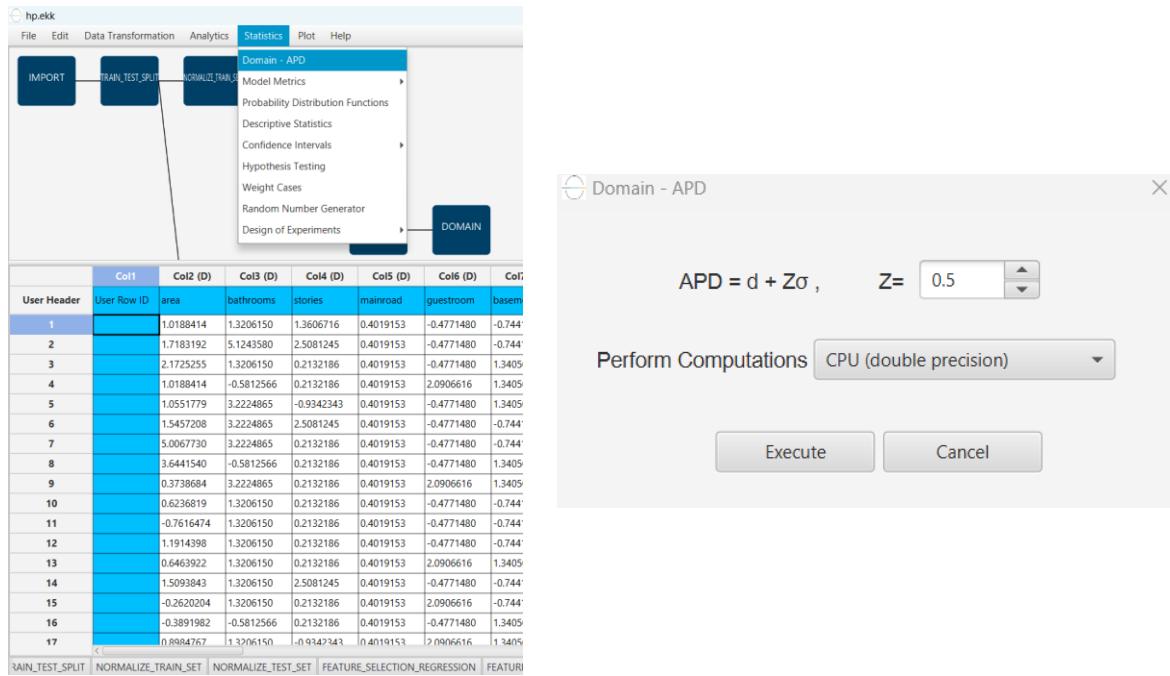
The results will appear on the output spreadsheet.

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

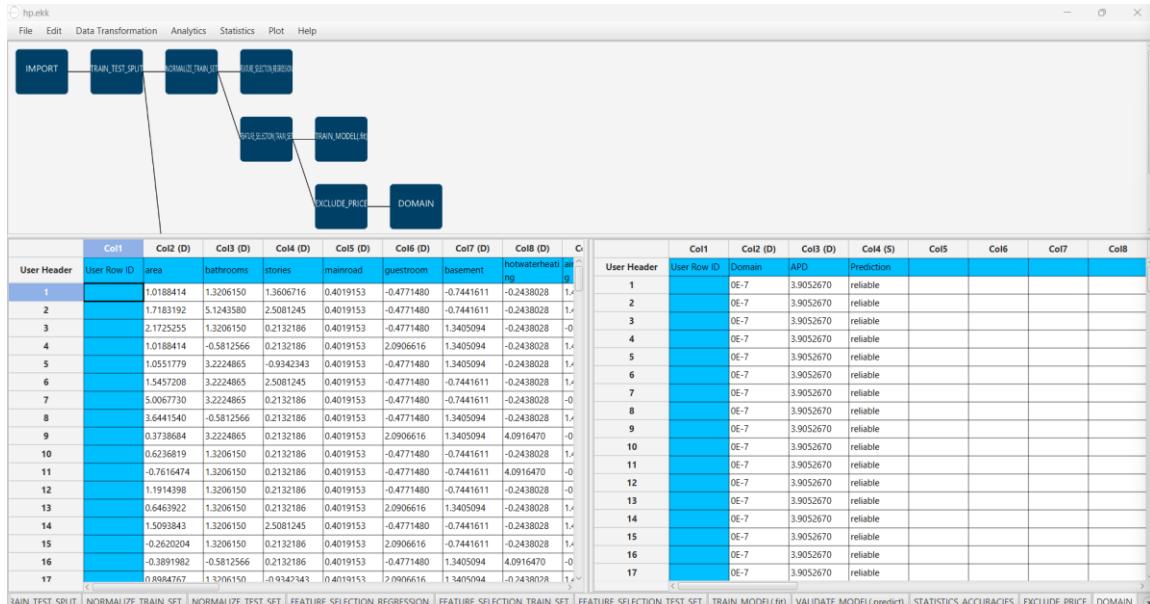
Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE_PRICE" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	Col8 (D)	
1	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438028	1		
2	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1		
3	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1		
4	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	-0.7441611	-0.2438028	1		
5	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	-0.7441611	-0.2438028	1		
6	1.015000000	1.5457208	2.3224865	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1	
7	1.015000000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1	
8	9.800000000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1	
9	9.968100000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	-0.7441611	-0.2438028	1	
10	9.931000000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1	
11	9.924000000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1	
12	9.924000000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1	

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



The results will appear on the output spreadsheet.



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_PRICE_TEST_SET".

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

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

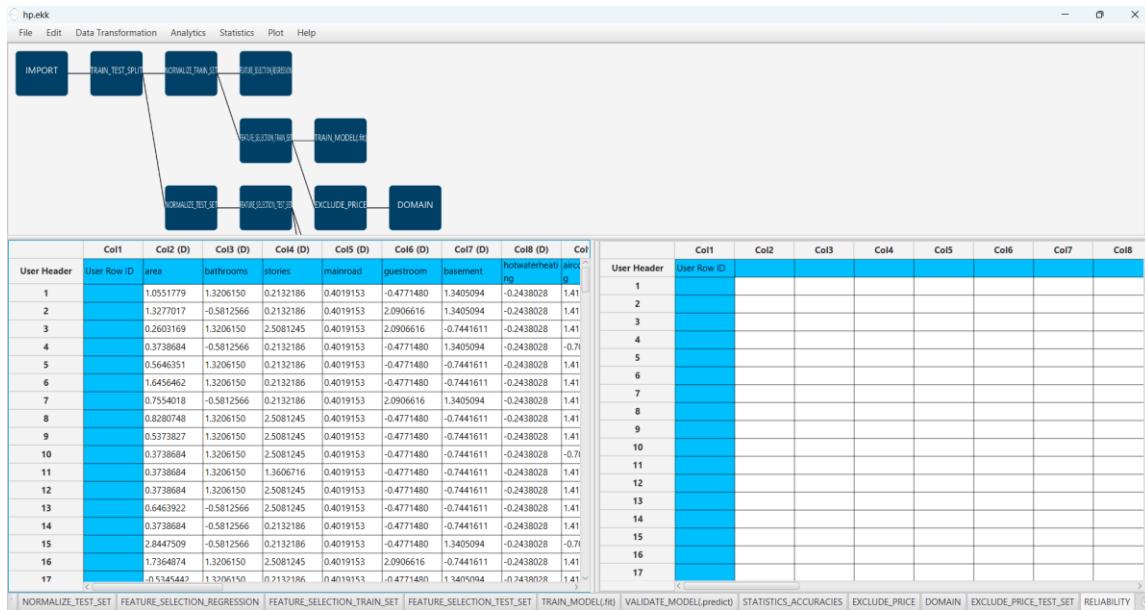
The screenshot shows the Isalos Analytics Platform interface. The main window displays a data transformation workflow with nodes like IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, SELECT_REGRESSION, and TRAIN_MODELING. A context menu is open over the 'Data Manipulation' node, with 'Select Column(s)' selected. A modal dialog titled 'Select Column(s)' lists various columns from the dataset, allowing the user to include or exclude specific ones. The 'Included Columns' list includes Col3 through Col10, while Col2 is explicitly excluded.

The results will appear on the output spreadsheet.

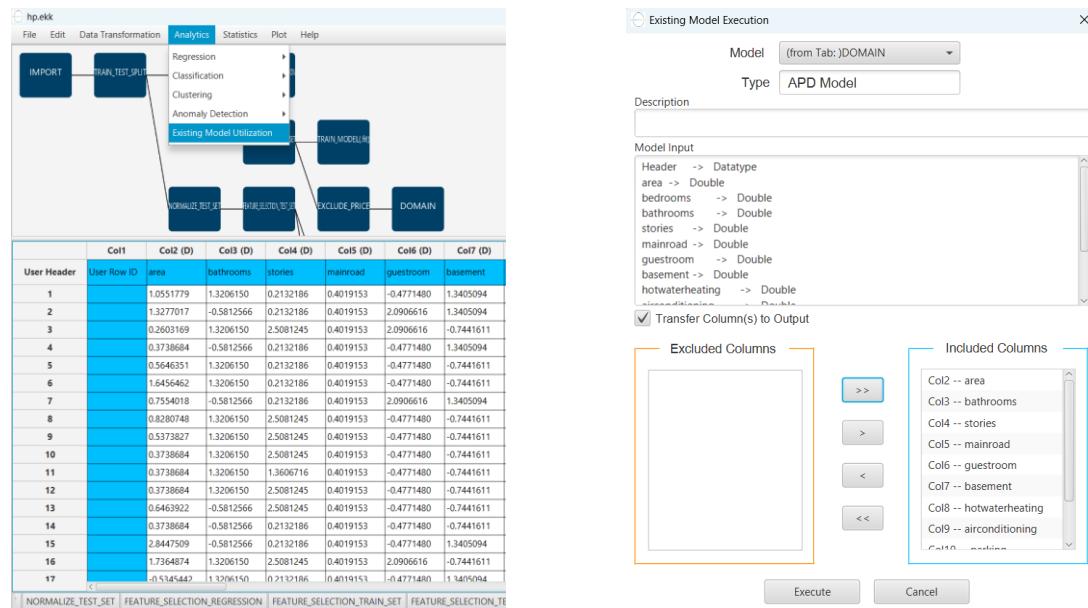
The screenshot shows the output spreadsheet after applying the column selection. The columns are labeled Col1 through Col9. The data rows show the same 12 entries as the input, with the 'price' column removed. The columns are labeled Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), and Col9 (D).

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

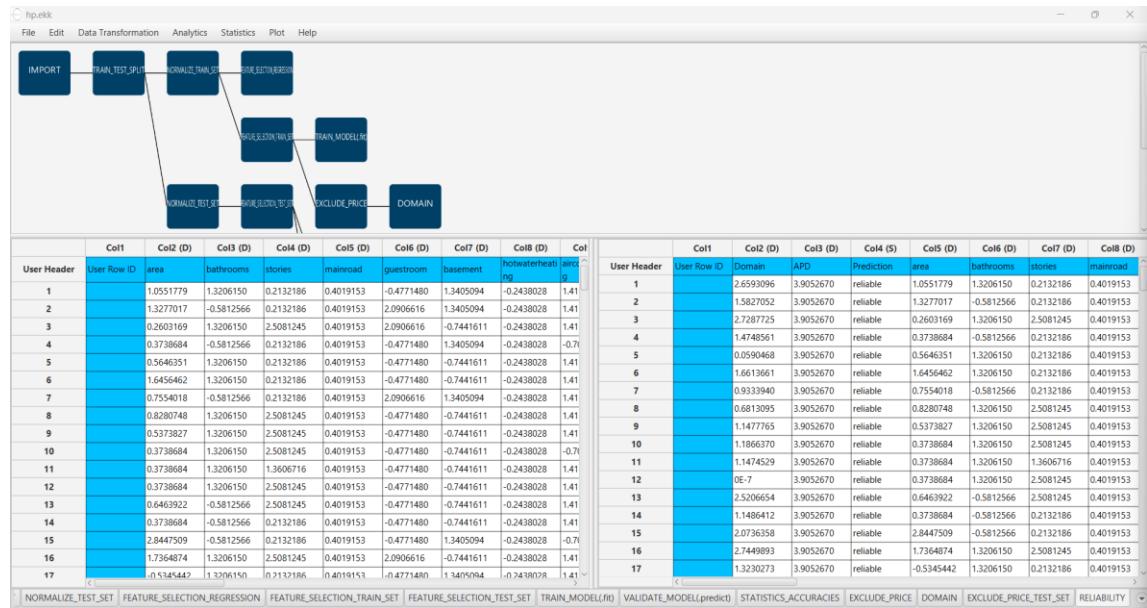
Import data into the input spreadsheet of the "RELIABILITY" tab from the output of the "EXCLUDE_PRICE_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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



The results will appear on the output spreadsheet. 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:

