



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						

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement
2		1.33E7	7420.0	4.0	2.0	3.0	1.0	0.0	0.0
3		1.225E7	8960.0	4.0	4.0	4.0	1.0	0.0	0.0
4		1.225E7	9960.0	3.0	2.0	2.0	1.0	0.0	1.0
5		1.2215E7	7500.0	4.0	2.0	2.0	1.0	0.0	1.0
6		1.141E7	7420.0	4.0	1.0	2.0	1.0	1.0	1.0
7		1.085E7	7500.0	3.0	3.0	1.0	1.0	0.0	1.0
8		1.015E7	8580.0	4.0	3.0	4.0	1.0	0.0	0.0
9		1.015E7	16200.0	5.0	3.0	2.0	1.0	0.0	0.0
10		9870000.0	8100.0	4.0	1.0	2.0	1.0	1.0	1.0
11		9800000.0	5750.0	3.0	2.0	4.0	1.0	1.0	0.0
12		9800000.0	13200.0	3.0	1.0	2.0	1.0	0.0	1.0
13		9681000.0	6000.0	4.0	3.0	2.0	1.0	1.0	1.0
14		9310000.0	6550.0	4.0	2.0	2.0	1.0	0.0	0.0
15		9240000.0	3500.0	4.0	2.0	2.0	1.0	0.0	0.0
16		9240000.0	7800.0	3.0	2.0	2.0	1.0	0.0	0.0
17		9100000.0	6000.0	4.0	1.0	2.0	1.0	0.0	1.0
18		9100000.0	6600.0	4.0	2.0	2.0	1.0	1.0	1.0
19		8960000.0	8500.0	3.0	2.0	4.0	1.0	0.0	0.0
20		8890000.0	4600.0	3.0	2.0	2.0	1.0	1.0	0.0
21		8855000.0	6420.0	3.0	2.0	2.0	1.0	0.0	0.0
		8750000.0	4320.0	1.0	2.0	1.0	0.0	0.0	1.0

Step 2: Manipulate data

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

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

The data will appear in the output spreadsheet.

Step 3: 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".

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:

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
1	133000000.0000	7420.000000	4.000000	2.000000	3.000000	1.000000	0E-7	
2	122500000.0000	8960.000000	4.000000	4.000000	4.000000	1.000000	0E-7	
3	122500000.0000	9960.000000	3.000000	2.000000	2.000000	1.000000	0E-7	
4	122150000.0000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	
5	114100000.0000	7420.000000	4.000000	1.000000	2.000000	1.000000	1.000000	
6	108500000.0000	7500.000000	3.000000	3.000000	1.000000	1.000000	0E-7	
7	101500000.0000	8580.000000	4.000000	3.000000	4.000000	1.000000	0E-7	
8	101500000.0000	16200.000000	5.000000	3.000000	2.000000	1.000000	0E-7	
9	98700000.0000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.000000	
10	98000000.0000	5750.000000	3.000000	2.000000	4.000000	1.000000	1.000000	
11	98000000.0000	13200.000000	3.000000	1.000000	2.000000	1.000000	0E-7	
12	96810000.0000	6000.000000	4.000000	3.000000	2.000000	1.000000	1.000000	
13	93100000.0000	6550.000000	4.000000	2.000000	2.000000	1.000000	0E-7	
14	92400000.0000	3500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	
15	92400000.0000	7800.000000	3.000000	2.000000	2.000000	1.000000	0E-7	

The results will appear on the output spreadsheet.

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
1	133000000.0000	7420.000000	4.000000	2.000000	3.000000	1.000000	0E-7	
2	122500000.0000	8960.000000	4.000000	4.000000	4.000000	1.000000	0E-7	
3	122500000.0000	9960.000000	3.000000	2.000000	2.000000	1.000000	0E-7	
4	122150000.0000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	
5	114100000.0000	7420.000000	4.000000	1.000000	2.000000	1.000000	1.000000	
6	108500000.0000	7500.000000	3.000000	3.000000	1.000000	1.000000	0E-7	
7	101500000.0000	8580.000000	4.000000	3.000000	4.000000	1.000000	0E-7	
8	101500000.0000	16200.000000	5.000000	3.000000	2.000000	1.000000	0E-7	
9	98700000.0000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.000000	
10	98000000.0000	5750.000000	3.000000	2.000000	4.000000	1.000000	0E-7	
11	98000000.0000	13200.000000	3.000000	1.000000	2.000000	1.000000	1.000000	
12	96810000.0000	6000.000000	4.000000	3.000000	2.000000	1.000000	1.000000	
13	93100000.0000	6550.000000	4.000000	2.000000	2.000000	1.000000	0E-7	
14	92400000.0000	3500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	
15	92400000.0000	7800.000000	3.000000	2.000000	2.000000	1.000000	0E-7	

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 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 toolbar with buttons for IMPORT, TRAIN_TEST_SPLIT, and NORMALIZE_TRAIN_SET.

The main area consists of two side-by-side spreadsheets. The left spreadsheet has columns labeled Col1 through Col9. The right spreadsheet has columns labeled Col1 through Col7. Both spreadsheets have a header row labeled "User Header". The data rows are numbered 1 through 21. The bottom of the interface shows tabs for IMPORT, TRAIN_TEST_SPLIT, and NORMALIZE_TRAIN_SET, with the NORMALIZE_TRAIN_SET tab currently selected.

Normalize the data using Z-score:

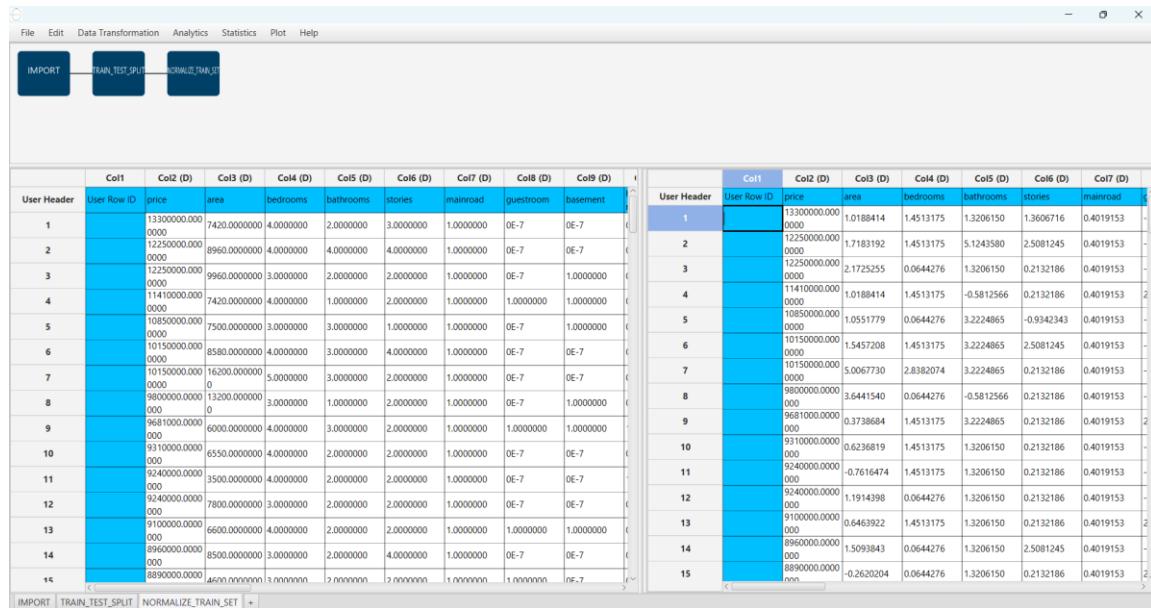
"Data Transformation" → "Normalizers" → "Z-Score"

Then select all columns except "price" and click "Execute".

The screenshot shows the Data Transformation interface. The "Data Transformation" tab is selected in the top navigation bar. A context menu is open under the "Normalizers" option, with "Z Score" selected. The main area shows a table of data with columns from Col1 to Col9. To the right, there's a configuration dialog for the "ZScore Normalizer".

The configuration dialog has two sections: "Excluded Columns" and "Included Columns". The "Excluded Columns" section contains "Col2 -- price". The "Included Columns" section contains "Col3 -- area", "Col4 -- bedrooms", "Col5 -- bathrooms", "Col6 -- stories", "Col7 -- mainroad", "Col8 -- guestroom", "Col9 -- basement", "Col10 -- hotwaterheating", and "Col11 -- airconditioning". There are buttons for "Execute" and "Cancel" at the bottom.

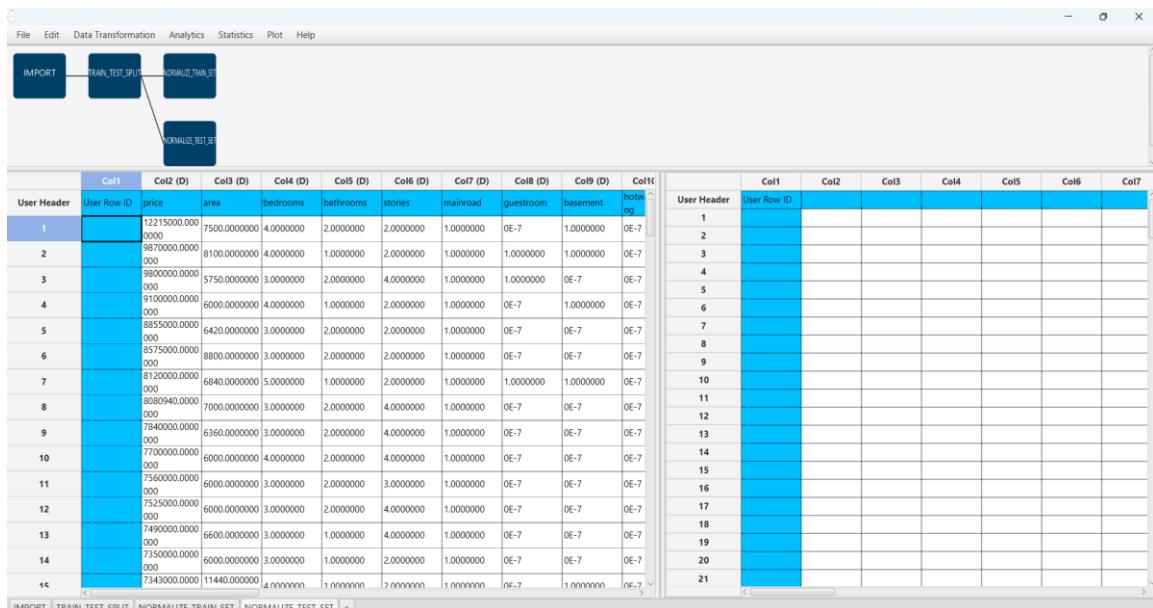
The results will appear on the output spreadsheet.



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:

"Analytics" → "Existing Model Utilization" → "Model (from Tab:) NORMALIZE_TRAIN_SET".

The screenshot shows the Isalos Analytics Platform interface. The top navigation bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Analytics menu is open, with Existing Model Utilization highlighted. A floating window titled "Existing Model Execution" is displayed, showing a dropdown "Model" set to "(from Tab:) NORMALIZE_TRAIN_SET" and a "Type" set to "Z Score Normalizer Model". Below this, a "Description" field is empty. A "Model Input" section lists column mappings: Header -> Datatype, area -> Double, bedrooms -> Double, bathrooms -> Double, stories -> Double, mainroad -> Double, guestroom -> Double, basement -> Double, hotwaterheating -> Double, and saniadition -> Double. A checkbox "Transfer Column(s) to Output" is checked. At the bottom are "Execute" and "Cancel" buttons.

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface with a data flow diagram. The flow starts with an "IMPORT" node, followed by a "TRAIN_TEST_SPLIT" node, then a "NORMALIZE_TRAIN_SET" node, and finally a "NORMALIZE_TEST_SET" node. The "NORMALIZE_TRAIN_SET" node has a line pointing to a new tab labeled "NORMALIZE_TRAIN_SET" in the bottom tabs. This tab contains a normalized dataset with columns Col1 through Col11. To the right, another tab labeled "NORMALIZE_TEST_SET" also contains a normalized dataset with columns Col1 through Col11. Both datasets have identical data rows.

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 with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and FEATURE_SELECTION_REGRESSION. The TRAIN_TEST_SPLIT node has two outputs: one to NORMALIZE_TRAIN_SET and one to FEATURE_SELECTION_REGRESSION. The NORMALIZE_TRAIN_SET node has an output to FEATURE_SELECTION_REGRESSION. The FEATURE_SELECTION_REGRESSION node has an output to a spreadsheet view.

The spreadsheet view contains two tabs: "User Header" and "User Row ID". The "User Header" tab shows columns Col1 through Col10. The "User Row ID" tab shows rows 1 through 21, each with values for price, area, bedrooms, bathrooms, stories, mainroad, guestroom, basement, hoarding, and Col11. The bottom of the screen shows a toolbar with buttons for IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, and FEATURE_SELECTION_REGRESSION.

Choose the most important features using the Regression Analysis:

"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.

The screenshot shows the Isalos Analytics Platform interface with the "Data Transformation" tab selected. A context menu is open over the "Variable Selection" node, with options: Best First, Stepwise, and Regression Analysis. The "Regression Analysis" option is highlighted.

To the right, a "Regression Analysis Model" dialog box is open. It has fields for "Significance Level (α)" set to 0.05, "Select Intercept Column" set to "Col8 -- charges", and "Included Columns" containing "Col2 -- age", "Col3 -- sex", "Col4 -- bmi", "Col5 -- children", "Col6 -- smoker", and "Col7 -- region". There are also "Excluded Columns" and "Included Columns" buttons with arrows for managing column selection.

The bottom of the screen shows a toolbar with buttons for IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, and FEATURE_SELECTION_REGRESSION.

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' and 'User Row ID', contains statistical information such as R-squared, F-statistic, and p-values for different models. The second table, titled 'Coefficients', lists the coefficient values for each feature (price, area, bathrooms, stories, mainroad, guestroom, basement, hotwaterheating, airconditioning, parking, prefarea, furnishingstatus) along with their standard error, t-statistic, and p-value.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	
88	1	6090000000000000	1.4512458	0.0644276	-0.5812566	1.3606716		
89	2	5083000000000000	2.0180953	0.0644276	-0.5812566	-0.9342343		
90	3	6020000000000000	0.7826541	0.0644276	1.3206150	-0.9342343		
91	4	5950000000000000	-0.6708061	1.4513175	-0.5812566	0.2132186		
92	5	5950000000000000	0.5646351	0.0644276	-0.5812566	-0.9342343		
93	6	5950000000000000	0.8371589	0.0644276	-0.5812566	-0.9342343		
94	7	5950000000000000	0.9329964	0.0644276	-0.5812566	0.2132186		
95	8	5950000000000000	0.9734208	1.4513175	1.3206150	0.2132186		
96	9	5950000000000000	0.6123268	0.0644276	1.3206150	2.5081245		
97	10	5943000000000000	4.7342492	0.0644276	-0.5812566	-0.9342343		
98	11	5880000000000000	0.9007478	0.0644276	-0.5812566	-0.9342343		
99	12	5880000000000000	0.6009716	0.0644276	1.3206150	1.3606716		
100	13	5873000000000000	2.8538350	0.0644276	-0.5812566	1.3606716		
101	14	5866000000000000	-0.1711791	0.0644276	-0.5812566	-0.9342343		
102	15	5810000000000000	0.2957450	1.4513175	-0.5812566	2.5081245		
103	16	5810000000000000	0.0105034	0.0644276	-0.5812566	1.3606716		
104	17	5810000000000000	-0.1711791	0.0644276	-0.5812566	1.3606716		
105	18	5803000000000000	0.8280748	0.0644276	-0.5812566	-0.9342343		
106	19	5775000000000000	0.3738684	0.0644276	1.3206150	2.5081245		
107	20	5740000000000000	5740000000000000	-0.2438522	1.4513175	-0.5812566	0.2132186	

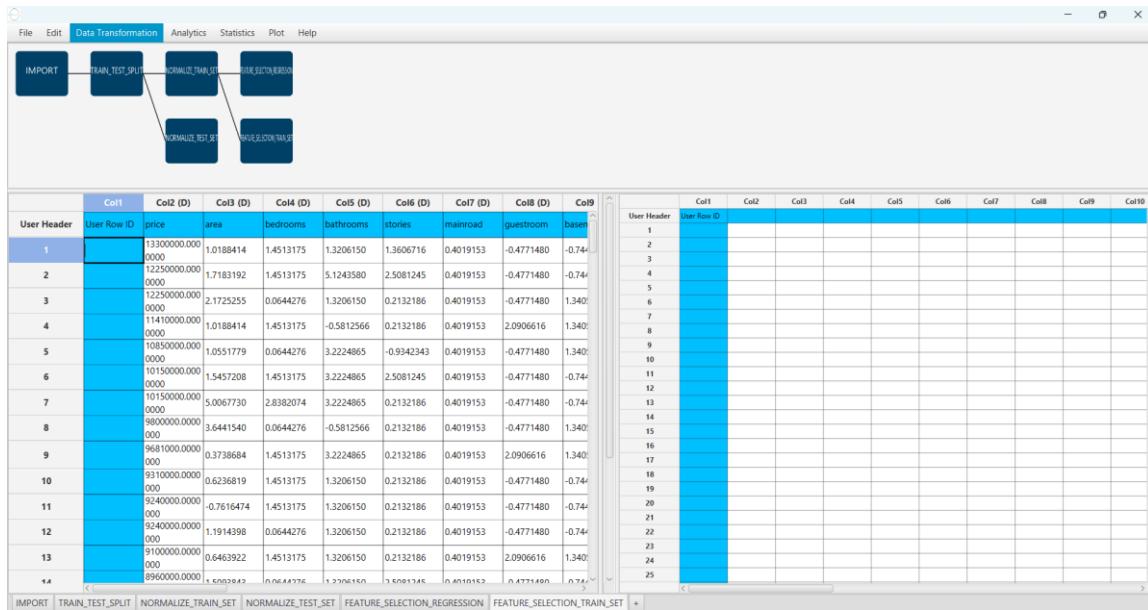
The significant features according to the p-value are the following:

- 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".

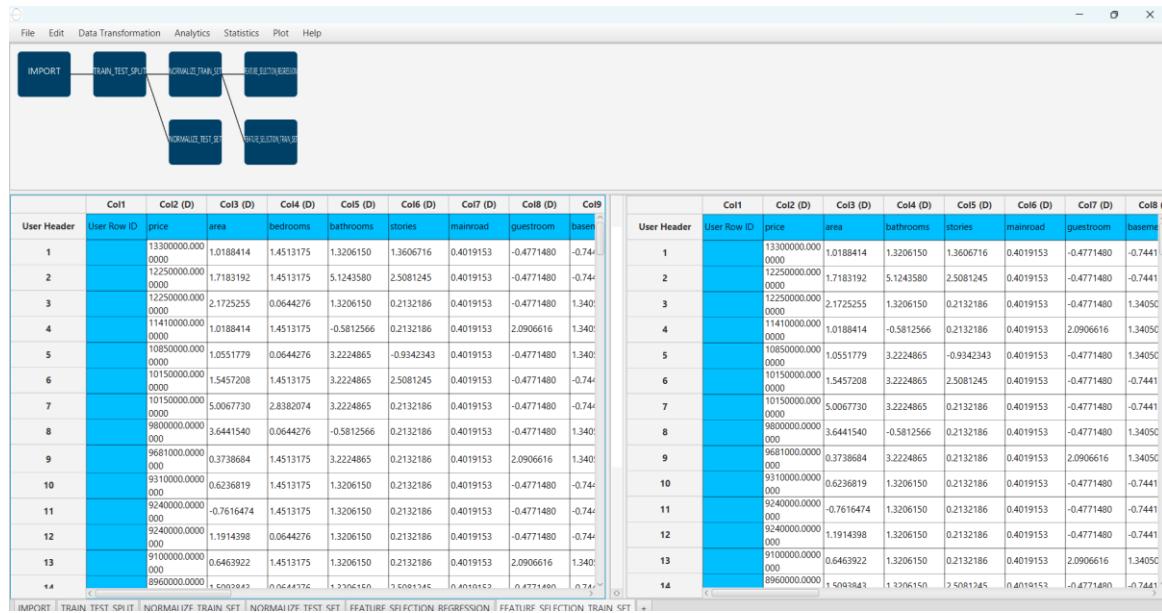


Manipulate the data by choosing the columns that correspond to the significant features (from the previous step):

"Data Transformation" → "Data Manipulation" → "Select Column(s)".

The screenshot shows the Isalos Analytics Platform interface with a "Select Column(s)" dialog box overlaid. The dialog box has two main sections: "Excluded Columns" (left, orange border) and "Included Columns" (right, blue border). In the "Excluded Columns" section, there is a list box containing "Col4 -- bedrooms". In the "Included Columns" section, there is a list box containing "Col2 -- price", "Col3 -- area", "Col5 -- bathrooms", "Col6 -- stories", "Col7 -- mainroad", "Col8 -- guestroom", "Col9 -- basement", and "Col10 -- hotwaterheating". At the bottom of the dialog box are "Execute" and "Cancel" buttons. To the left of the dialog box, the main interface shows the "Data Transformation" menu open, with "Data Manipulation" selected. Under "Data Manipulation", the "Select Column(s)" option is highlighted. The main data spreadsheet below the menu shows the same structure as the first screenshot, with columns Col1 through Col10 and rows 1 through 25.

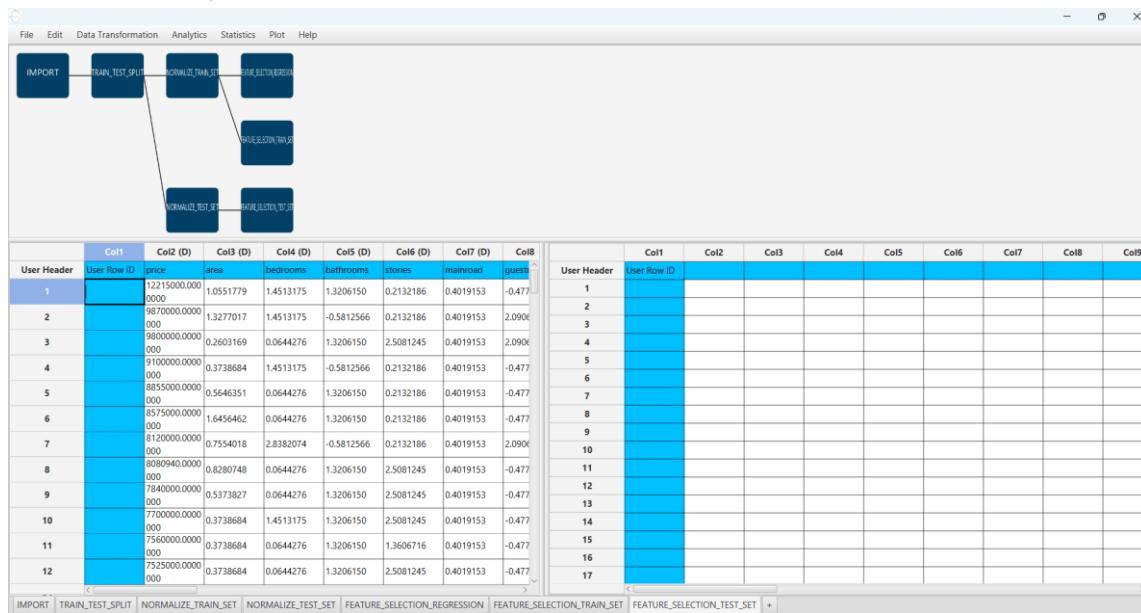
The results will appear on the output spreadsheet.



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".



Manipulate the data by choosing the columns that correspond to the significant features (from the step 6):

"Data Transformation" → "Data Manipulation" → "Select Column(s)".

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1		12215000.0000	1.0551779	1.4513175	1.3206150	0.2132186	0.4019153	-0.477	
2		9870000.0000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906	
3		9800000.0000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	2.0906	
4		9100000.0000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.477	
5		8850000.0000	0.5646351	0.0644276	1.3206150	2.5081245	0.4019153	-0.477	
6		8575000.0000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.477	
7		8120000.0000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.0906	
8		8080940.0000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.477	
9		7840000.0000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.477	
10		7700000.0000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.477	
11		7560000.0000	0.3738684	0.0644276	1.3206150	1.3606716	0.4019153	-0.477	
12		7525000.0000	0.3738684	0.0644276	1.3206150	2.5081245	0.4019153	-0.477	

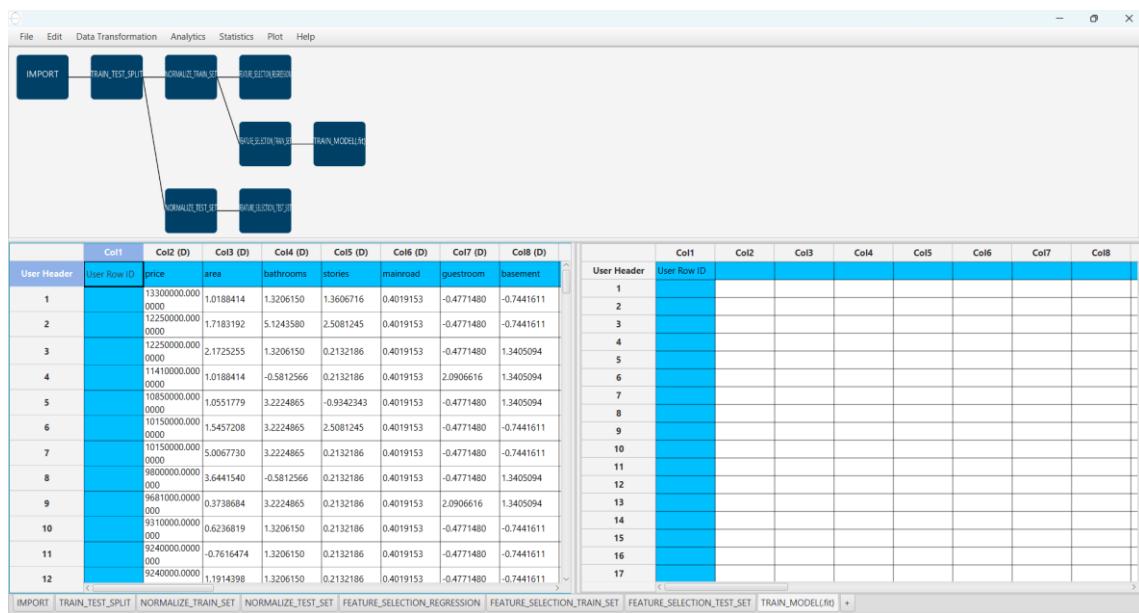
The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8	
1		12215000.0000	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.243802	
2		9870000.0000	1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.243802	
3		9800000.0000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	-0.7441611	-0.243802	
4		9100000.0000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.243802
5		8850000.0000	0.5646351	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.243802
6		8575000.0000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.243802
7		8120000.0000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.243802
8		8080940.0000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	1.3405094	-0.243802
9		7840000.0000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	1.3405094	-0.243802
10		7700000.0000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.4771480	1.3405094	-0.243802
11		7560000.0000	0.3738684	0.0644276	1.3206150	1.3606716	0.4019153	-0.4771480	1.3405094	-0.243802
12		7525000.0000	0.3738684	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	1.3405094	-0.243802

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)".

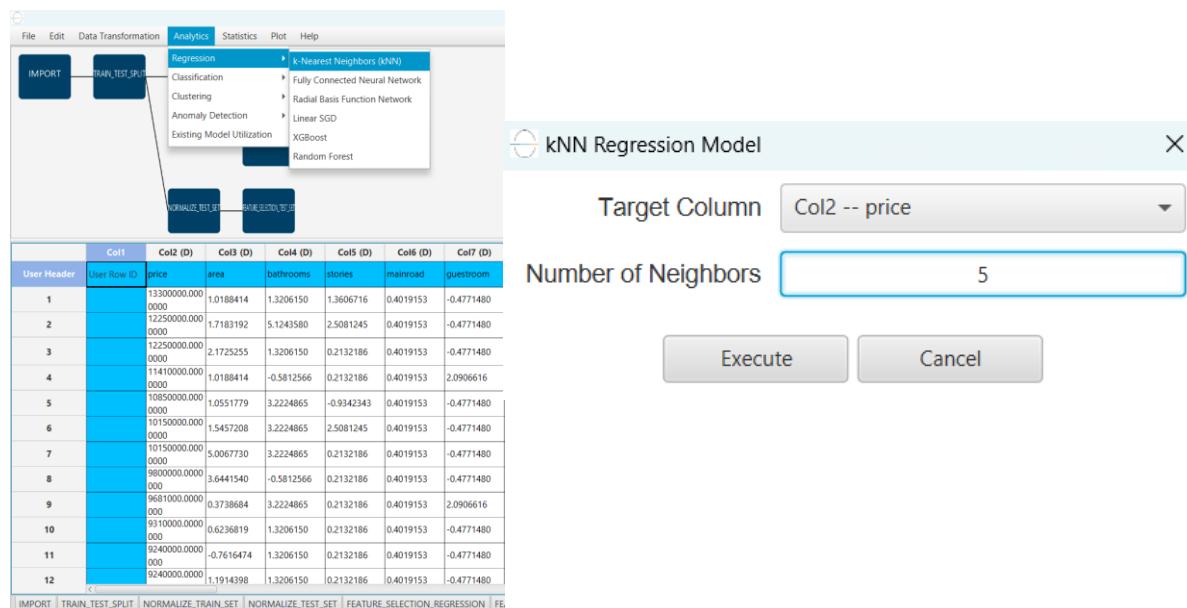
Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



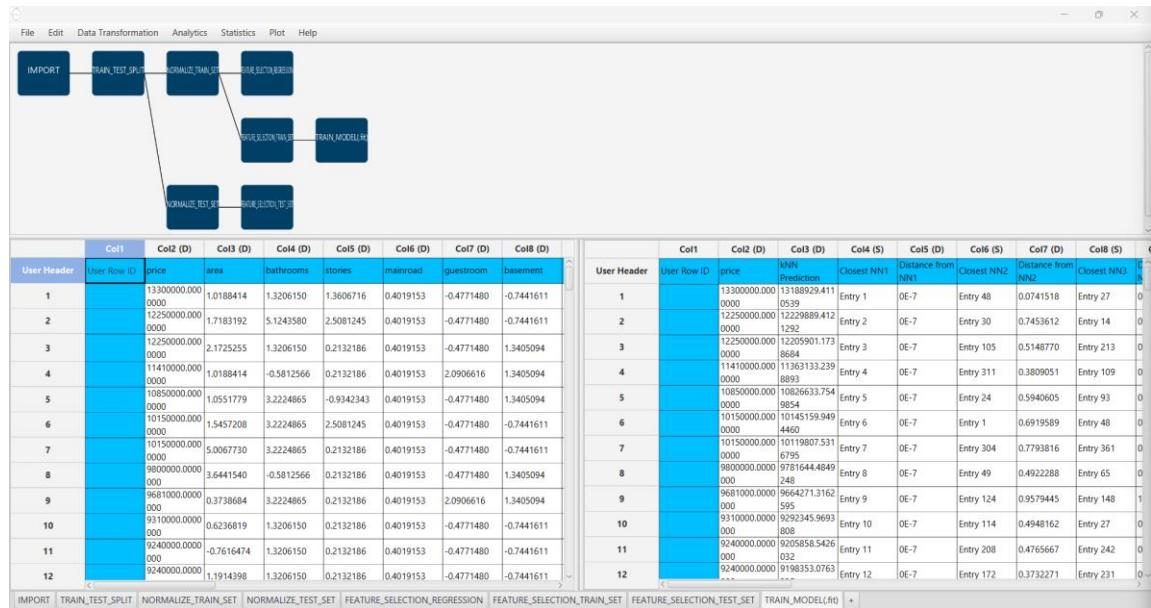
Use the k Nearest Neighbors (kNN) method to train and fit the model:

"Analytics" → "Regression" → "k Nearest Neighbors (kNN)"

and set the "Target Column" as the column corresponding to "price" and the "Number of Neighbors" to 5.



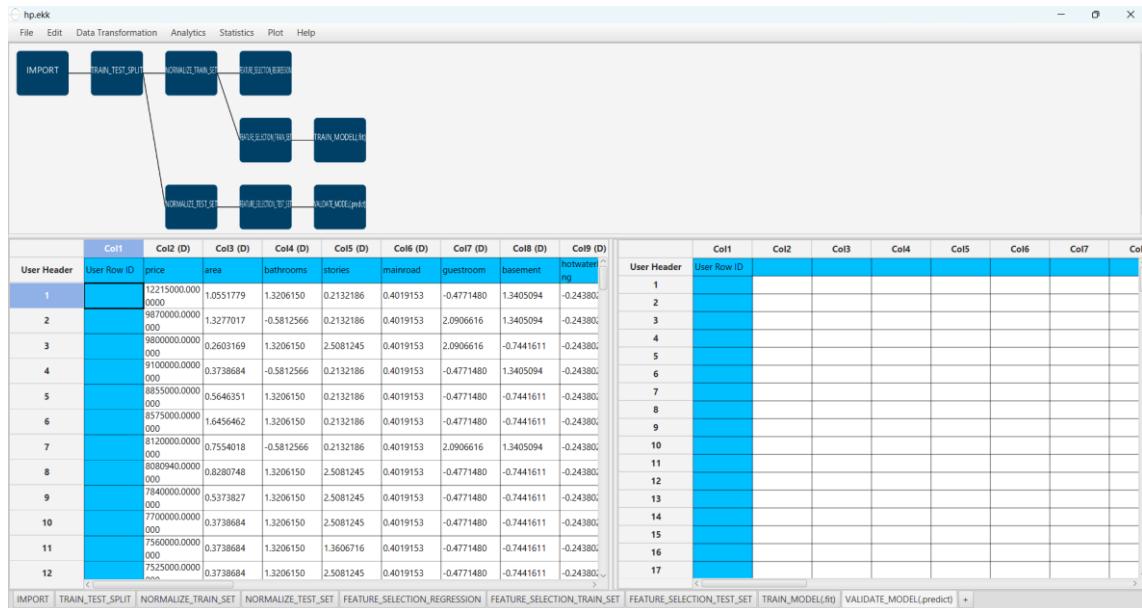
The predictions will appear on the output spreadsheet.



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)".

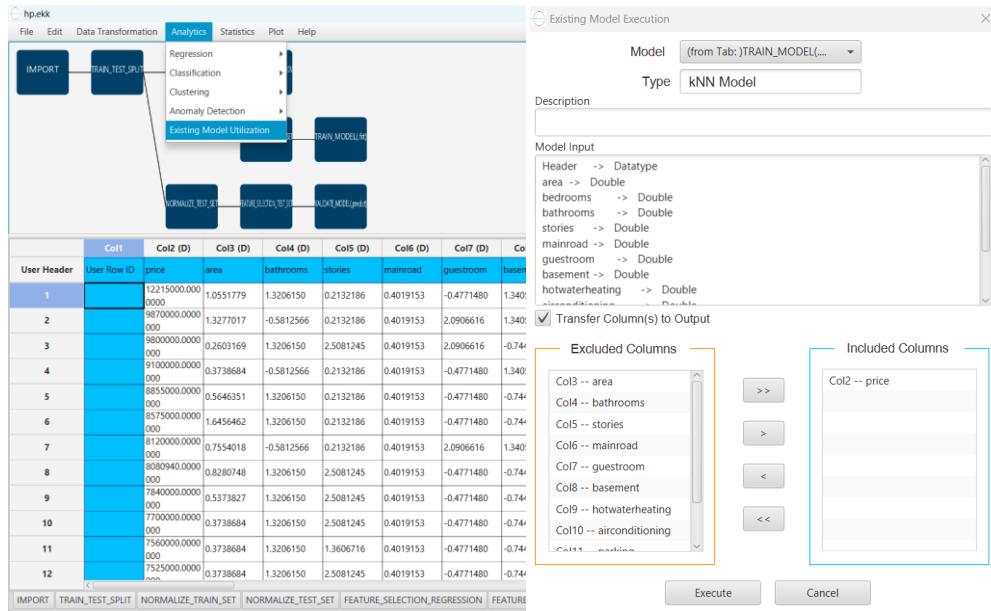
Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



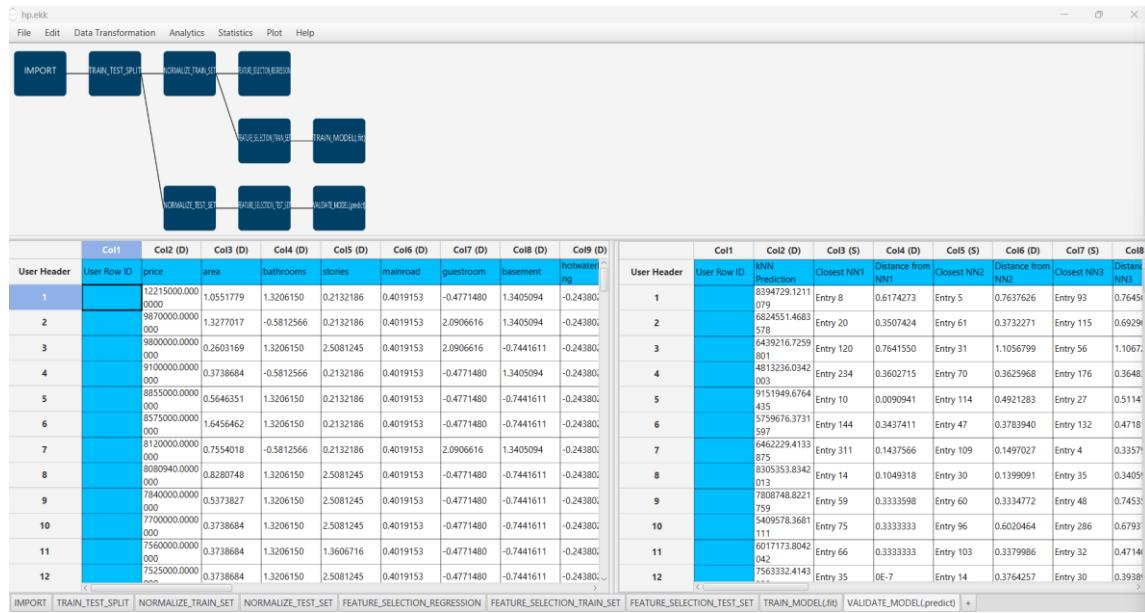
To validate the model:

"Analytics" → "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN_MODEL(.fit)".

and transfer the "price" column in the output.



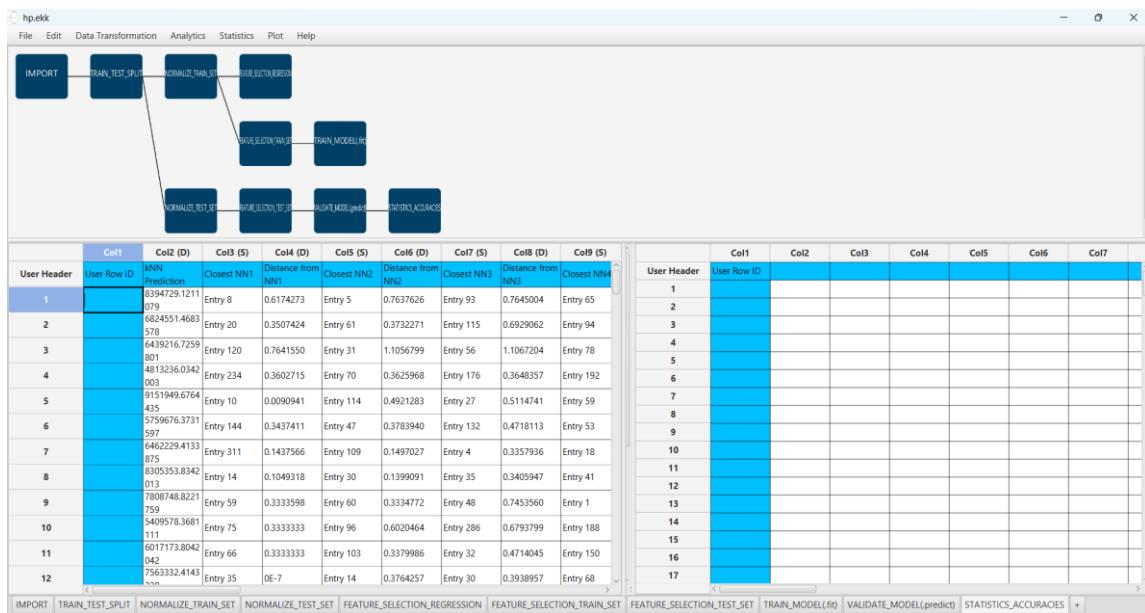
The predictions will appear on the output spreadsheet.



Step 11: Statistics calculation

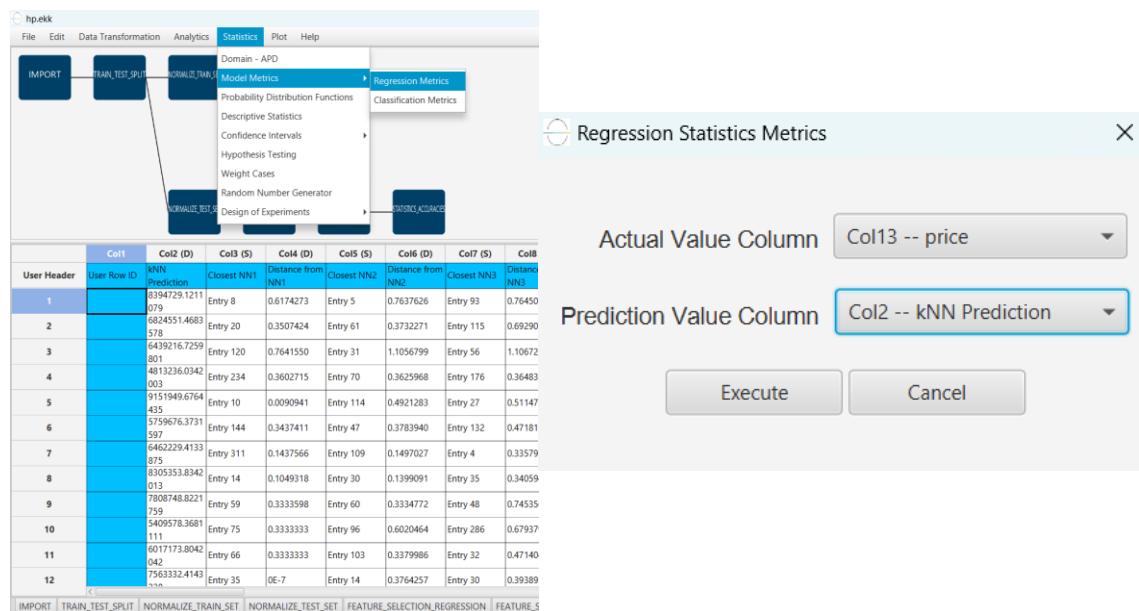
Create a new tab by pressing the "+" button on the bottom of the page with the name "STATISTICS_ACCURACIES".

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".

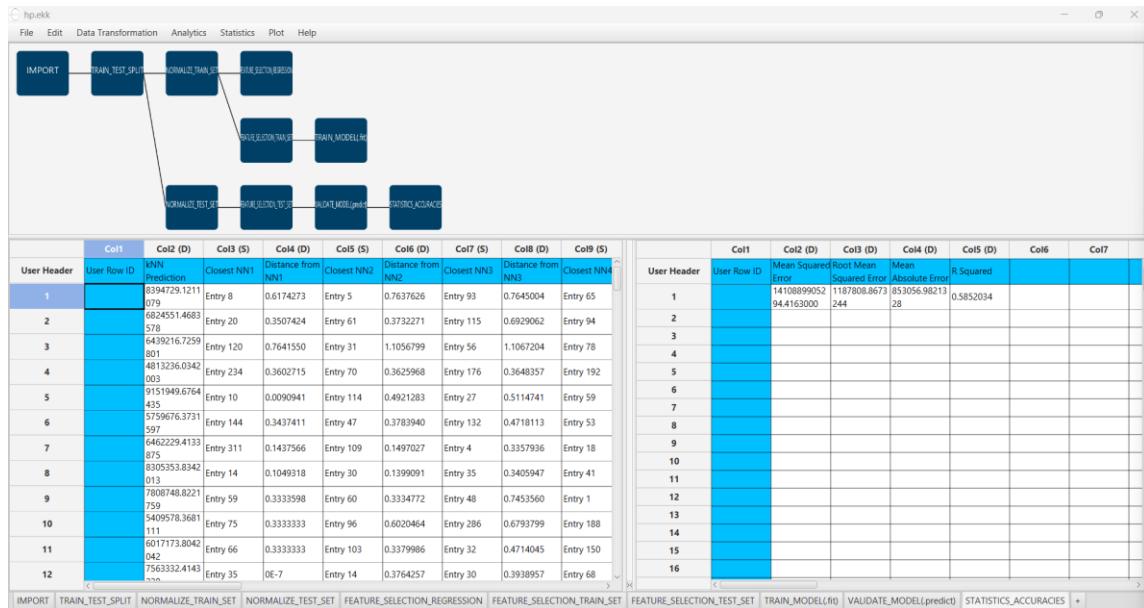


Calculate the statistical metrics for the regression:

"Statistics" → "Model Metrics" → "Regression Metrics".



The results will appear on the output spreadsheet.

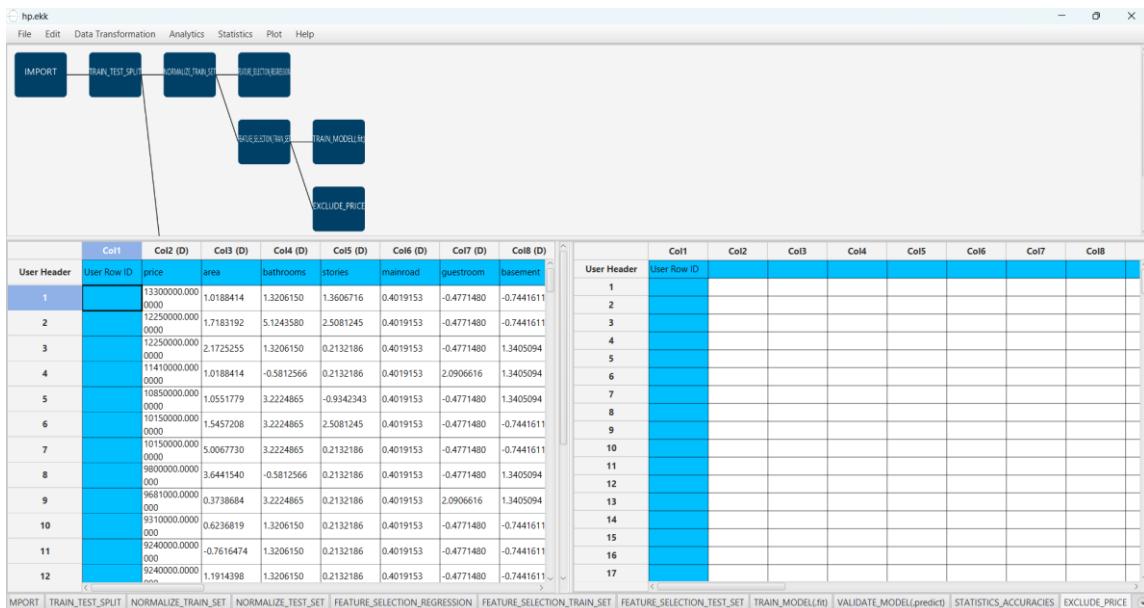


Step 12: Reliability check of each record of the test set

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".

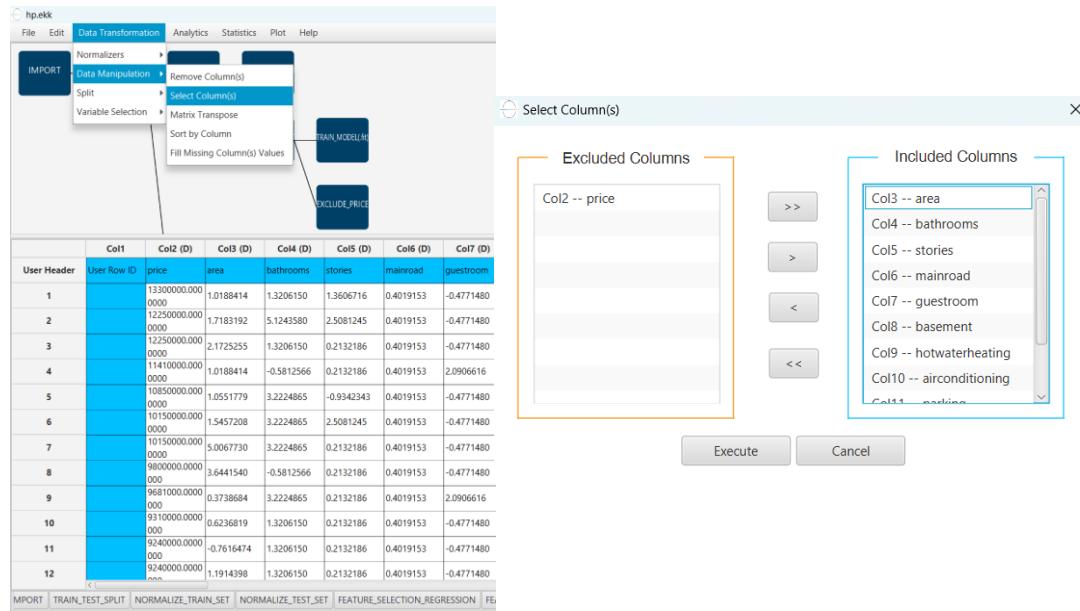
Import data into the input spreadsheet of the "EXCLUDE_PRICE" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Manipulate the data to exclude the column that corresponds to the "price"

"Data Transformation" → "Data Manipulation" → "Select Columns"

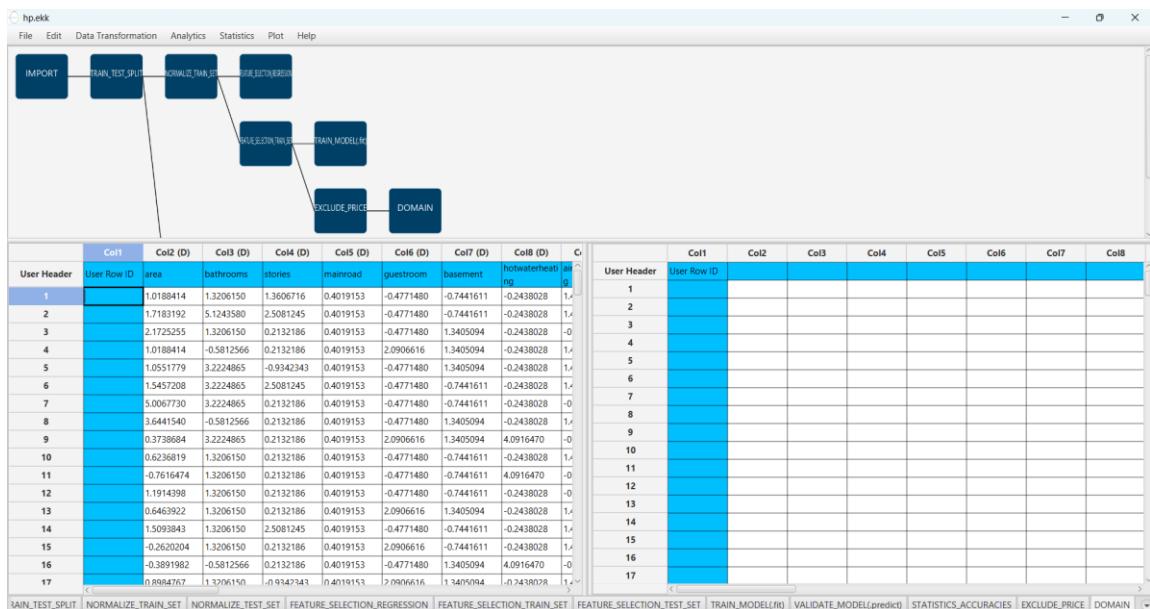
Then select all the columns except the "price".



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".



Create the domain:

"Statistics" → "Domain APD"

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7
User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	Incheateheatinsg
1	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.744	
2	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.744	
3	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405	
4	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405	
5	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405	
6	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.744	
7	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.744	
8	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405	
9	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.3405	
10	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.744	
11	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.744	
12	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.744	
13	0.6463922	1.3206150	0.2132186	0.4019153	2.0906616	1.3405	
14	1.5093843	1.3206150	2.5081245	0.4019153	-0.4771480	-0.744	
15	-0.2620204	1.3206150	0.2132186	0.4019153	2.0906616	-0.744	
16	0.3891982	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405	
17	0.8944767	1.3206150	-0.9342343	0.4019153	2.0906616	1.3405	

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)
User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	Incheateheatinsg	APD
1	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
2	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
3	2.1725255	1.3206150	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	0E-7
4	1.0188414	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	0E-7
5	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094	-0.2438028	0E-7
6	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
7	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
8	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	0.0916470	0E-7
9	0.3738684	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
10	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
11	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
12	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
13	0.6463922	1.3206150	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	0E-7
14	1.5093843	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	0E-7
15	-0.2620204	1.3206150	0.2132186	0.4019153	2.0906616	-0.7441611	-0.2438028	0E-7
16	0.3891982	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	0.0916470	0E-7
17	0.8944767	1.3206150	-0.9342343	0.4019153	2.0906616	1.3405094	-0.2438028	0E-7

Step 12.b: Check the test set reliability

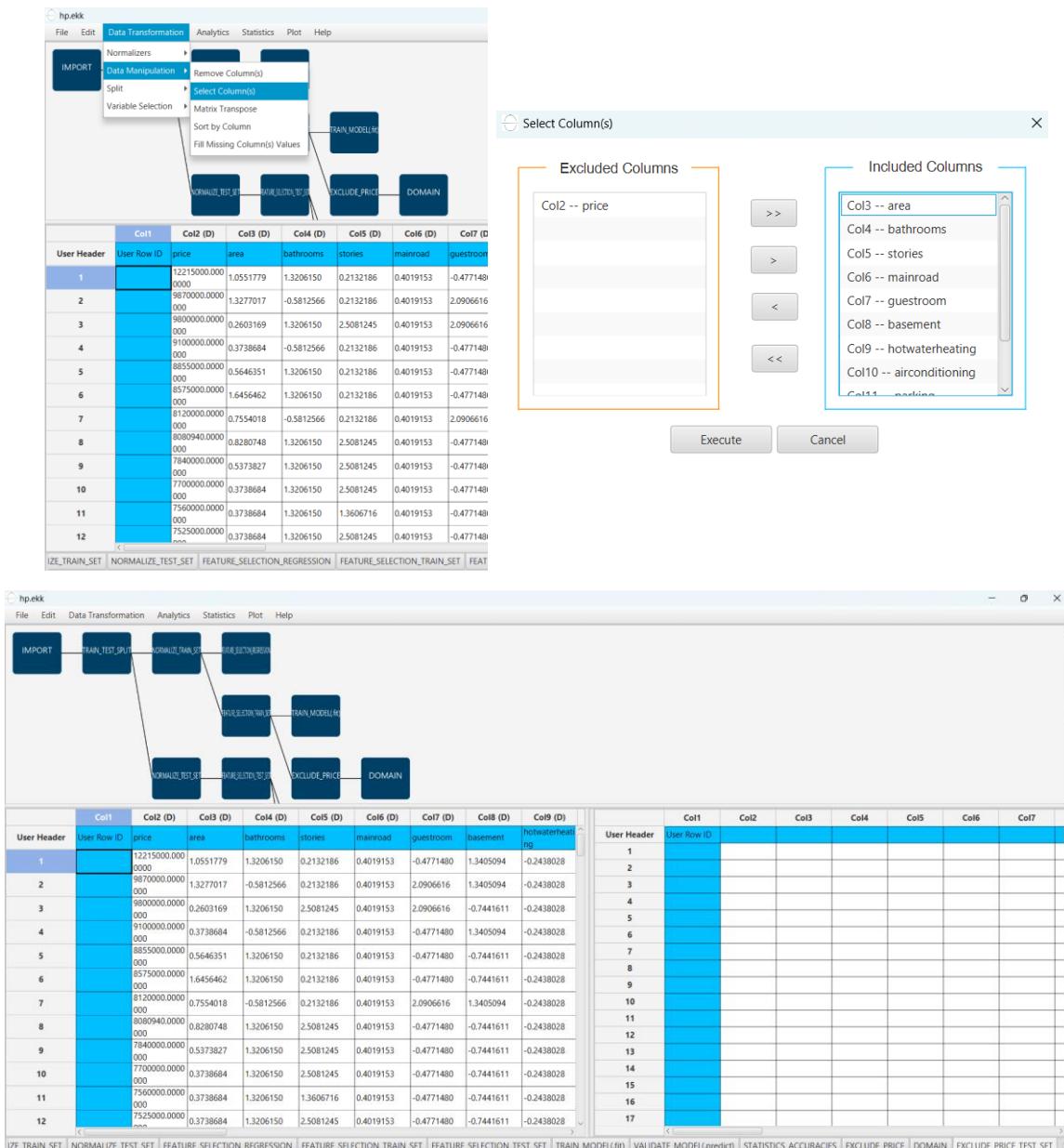
Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_PRICE_TEST_SET".

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"

"Data Transformation" → "Data Manipulation" → "Select Columns".

Then select all the columns except "price".



The results will appear on the output spreadsheet.

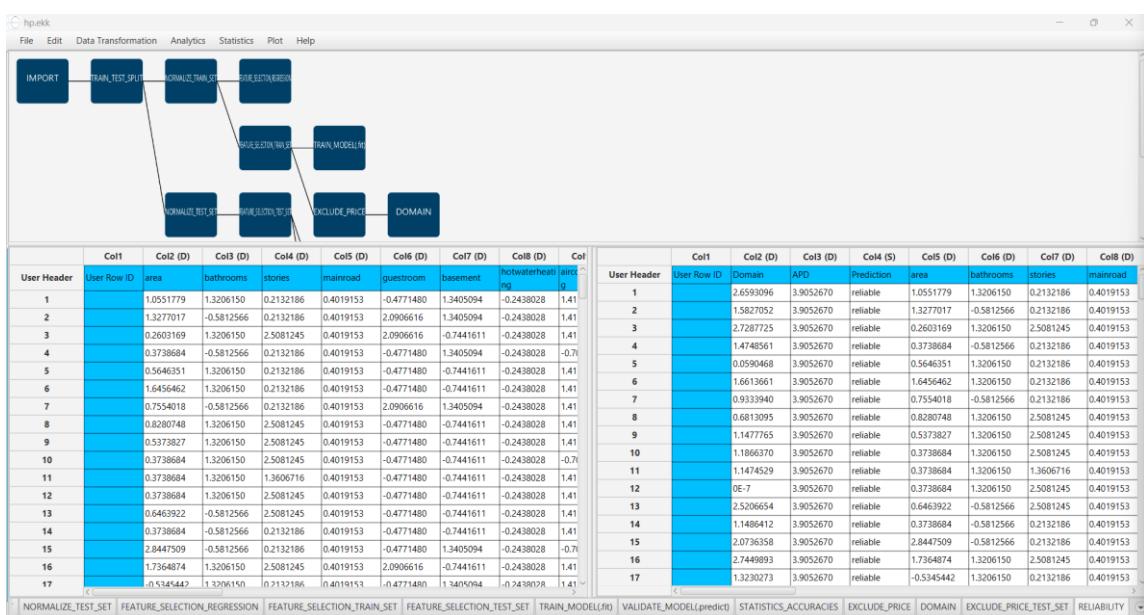
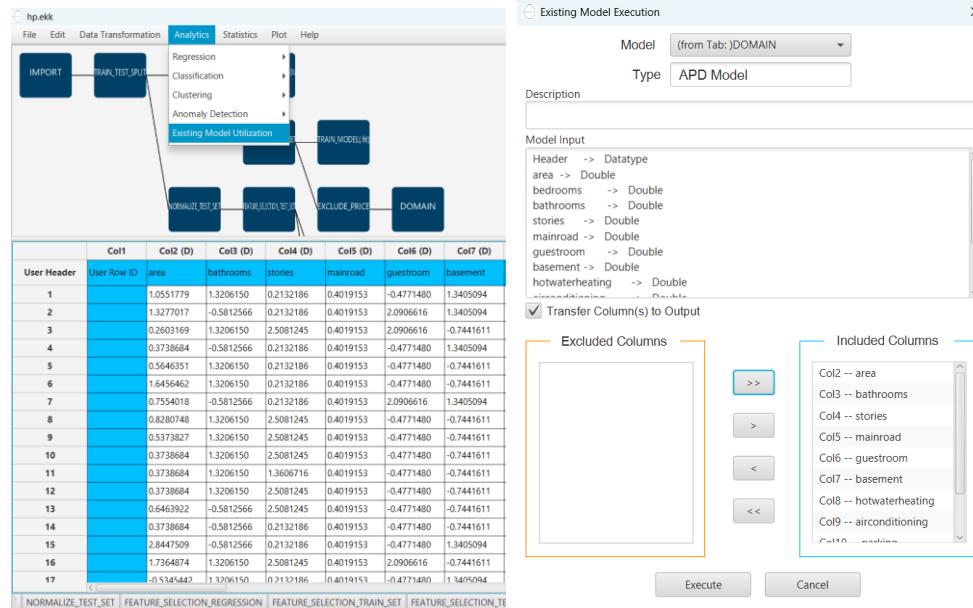
Create a new tab by pressing the "+" button on the bottom of the page with the name "RELIABILITY".

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".

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:

