



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.

	Col1	Col2	Col3	Col4	Col5	Col6
User Header	User Row ID					
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

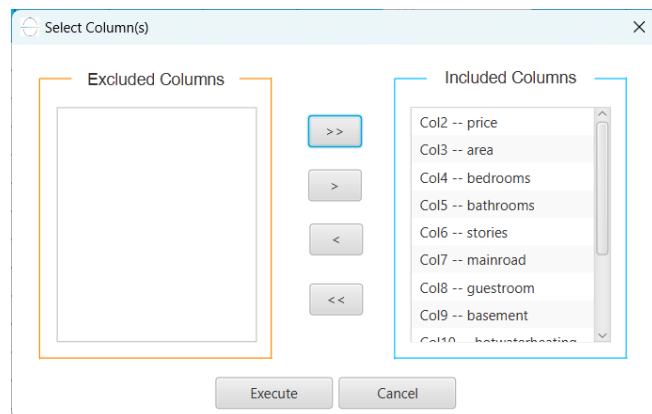
User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1		1.33E7	7420.0	4.0	2.0	3.0	1.0	0.0	0.0
2		1.225E7	8960.0	4.0	4.0	4.0	1.0	0.0	0.0
3		1.225E7	9960.0	3.0	2.0	2.0	1.0	0.0	1.0
4		1.2215E7	7500.0	4.0	2.0	2.0	1.0	0.0	1.0
5		1.141E7	7420.0	4.0	1.0	2.0	1.0	1.0	1.0
6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

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



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 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	13300000.000000000	7420.000000000	4.000000000	2.000000000	3.000000000	1.000000000	0E-7	0E-7	
2	12250000.000000000	8960.000000000	4.000000000	4.000000000	4.000000000	1.000000000	0E-7	0E-7	
3	12250000.000000000	9960.000000000	3.000000000	2.000000000	2.000000000	1.000000000	0E-7	1.000000000	
4	12215000.000000000	7500.000000000	4.000000000	2.000000000	2.000000000	1.000000000	0E-7	1.000000000	
5	11410000.000000000	7420.000000000	4.000000000	1.000000000	2.000000000	1.000000000	1.000000000	1.000000000	
6	10850000.000000000	7500.000000000	3.000000000	3.000000000	1.000000000	1.000000000	0E-7	1.000000000	
7	10150000.000000000	8580.000000000	4.000000000	3.000000000	4.000000000	1.000000000	0E-7	0E-7	
8	10150000.000000000	16200.000000000	5.000000000	3.000000000	2.000000000	1.000000000	0E-7	0E-7	
9	9870000.000000000	8100.000000000	4.000000000	1.000000000	2.000000000	1.000000000	1.000000000	1.000000000	
10	9800000.000000000	5750.000000000	3.000000000	2.000000000	4.000000000	1.000000000	1.000000000	0E-7	
11	9800000.000000000	13200.000000000	3.000000000	1.000000000	2.000000000	1.000000000	0E-7	1.000000000	
12	9681000.000000000	6000.000000000	4.000000000	3.000000000	2.000000000	1.000000000	1.000000000	1.000000000	
13	9310000.000000000	6550.000000000	4.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7	
14	9240000.000000000	3500.000000000	4.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7	
15	9240000.000000000	7800.000000000	3.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7	

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:

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
1		13300000.000000	7420.000000	4.000000	2.000000	3.000000	1.000000	0E-7
2		12250000.000000	8960.000000	4.000000	4.000000	4.000000	1.000000	0E-7
3		12250000.000000	9960.000000	3.000000	2.000000	2.000000	1.000000	0E-7
4		12215000.000000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E-7
5		11410000.000000	7420.000000	4.000000	1.000000	2.000000	1.000000	1.000000
6		10850000.000000	7500.000000	3.000000	3.000000	1.000000	1.000000	0E-7
7		10150000.000000	8580.000000	4.000000	3.000000	4.000000	1.000000	0E-7
8		10150000.000000	16200.000000	0	3.000000	2.000000	1.000000	0E-7
9		9870000.000000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.000000
10		9800000.000000	5750.000000	3.000000	2.000000	4.000000	1.000000	1.000000
11		9800000.000000	13200.000000	3.000000	1.000000	2.000000	1.000000	0E-7
12		9681000.000000	6000.000000	4.000000	3.000000	2.000000	1.000000	1.000000
13		9310000.000000	6550.000000	4.000000	2.000000	2.000000	1.000000	0E-7
14		9240000.000000	3500.000000	4.000000	2.000000	2.000000	1.000000	0E-7
15		9240000.000000	7800.000000	3.000000	2.000000	2.000000	1.000000	0E-7

Random Partitioning

Training Set Percentage: 75

☐ Time-based RNG Seed: 2444166481100

☒ Stratified sampling: Col3 -- area

Execute Cancel

The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
1		13300000.000000	7420.000000	4.000000	2.000000	3.000000	1.000000	0E-7
2		12250000.000000	8960.000000	4.000000	4.000000	4.000000	1.000000	0E-7
3		12250000.000000	9960.000000	3.000000	2.000000	2.000000	1.000000	0E-7
4		12215000.000000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E-7
5		11410000.000000	7420.000000	4.000000	1.000000	2.000000	1.000000	1.000000
6		10850000.000000	7500.000000	3.000000	3.000000	1.000000	1.000000	0E-7
7		10150000.000000	8580.000000	4.000000	3.000000	4.000000	1.000000	0E-7
8		10150000.000000	16200.000000	0	3.000000	2.000000	1.000000	0E-7
9		9870000.000000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.000000
10		9800000.000000	5750.000000	3.000000	2.000000	4.000000	1.000000	0E-7
11		9800000.000000	13200.000000	3.000000	1.000000	2.000000	1.000000	0E-7
12		9681000.000000	6000.000000	4.000000	3.000000	2.000000	1.000000	1.000000
13		9310000.000000	6550.000000	4.000000	2.000000	2.000000	1.000000	0E-7
14		9240000.000000	3500.000000	4.000000	2.000000	2.000000	1.000000	0E-7
15		9240000.000000	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".

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement
1	1	13300000.000000000	7420.000000000	4.000000000	2.000000000	3.000000000	1.000000000	0E-7	0E-7
2	2	12250000.000000000	8960.000000000	4.000000000	4.000000000	4.000000000	1.000000000	0E-7	0E-7
3	3	12250000.000000000	9960.000000000	3.000000000	2.000000000	2.000000000	1.000000000	0E-7	1.000000000
4	4	11410000.000000000	7420.000000000	4.000000000	1.000000000	2.000000000	1.000000000	1.000000000	1.000000000
5	5	10850000.000000000	7500.000000000	3.000000000	3.000000000	1.000000000	1.000000000	0E-7	1.000000000
6	6	10150000.000000000	8580.000000000	4.000000000	3.000000000	4.000000000	1.000000000	0E-7	0E-7
7	7	10150000.000000000	16200.000000000	5.000000000	3.000000000	2.000000000	1.000000000	0E-7	0E-7
8	8	9800000.000000000	13200.000000000	3.000000000	1.000000000	2.000000000	1.000000000	0E-7	1.000000000
9	9	9681000.000000000	6000.000000000	4.000000000	3.000000000	2.000000000	1.000000000	1.000000000	1.000000000
10	10	9310000.000000000	6550.000000000	4.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7
11	11	9240000.000000000	3500.000000000	4.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7
12	12	9240000.000000000	7800.000000000	3.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7
13	13	9100000.000000000	6600.000000000	4.000000000	2.000000000	2.000000000	1.000000000	1.000000000	1.000000000
14	14	8960000.000000000	8500.000000000	3.000000000	2.000000000	4.000000000	1.000000000	0E-7	0E-7
15	15	8890000.000000000	4600.000000000	3.000000000	2.000000000	2.000000000	1.000000000	0E-7	0E-7

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

ZScore Normalizer

Excluded Columns: Col2 -- price

Included Columns: Col3 -- area, Col4 -- bedrooms, Col5 -- bathrooms, Col6 -- stories, Col7 -- mainroad, Col8 -- guestroom, Col9 -- basement, Col10 -- hotwaterheating, Col11 -- airconditioning

Execute Cancel

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 (D)
1	13300000.000000	7420.000000	4.000000	2.000000	3.000000	1.000000	0E-7	0E-7	
2	12250000.000000	8960.000000	4.000000	4.000000	4.000000	1.000000	0E-7	0E-7	
3	12250000.000000	9960.000000	3.000000	2.000000	2.000000	1.000000	0E-7	1.000000	
4	11410000.000000	7420.000000	4.000000	1.000000	2.000000	1.000000	1.000000	1.000000	
5	10850000.000000	7590.000000	3.000000	3.000000	1.000000	1.000000	0E-7	1.000000	
6	10150000.000000	8580.000000	4.000000	3.000000	4.000000	1.000000	0E-7	0E-7	
7	10150000.000000	16200.000000	5.000000	3.000000	2.000000	1.000000	0E-7	0E-7	
8	9800000.000000	13200.000000	3.000000	1.000000	2.000000	1.000000	0E-7	1.000000	
9	9681000.000000	6000.000000	4.000000	3.000000	2.000000	1.000000	1.000000	1.000000	
10	9310000.000000	6550.000000	4.000000	2.000000	2.000000	1.000000	0E-7	0E-7	
11	9240000.000000	3500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	0E-7	
12	9240000.000000	7800.000000	3.000000	2.000000	2.000000	1.000000	0E-7	0E-7	
13	9100000.000000	6600.000000	4.000000	2.000000	2.000000	1.000000	1.000000	1.000000	
14	8960000.000000	8500.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	
15	8890000.000000	4600.000000	3.000000	2.000000	2.000000	1.000000	1.000000	0E-7	

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 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)
1	12215000.000000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	1.000000	0E-7	
2	9870000.000000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.000000	1.000000	0E-7	
3	9800000.000000	5750.000000	3.000000	2.000000	4.000000	1.000000	1.000000	0E-7	0E-7	
4	9100000.000000	6000.000000	4.000000	1.000000	2.000000	1.000000	0E-7	1.000000	0E-7	
5	8855000.000000	6420.000000	3.000000	2.000000	2.000000	1.000000	0E-7	0E-7	0E-7	
6	8575000.000000	8800.000000	3.000000	2.000000	2.000000	1.000000	0E-7	0E-7	0E-7	
7	8120000.000000	6840.000000	5.000000	1.000000	2.000000	1.000000	1.000000	1.000000	0E-7	
8	8080940.000000	7000.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
9	7840000.000000	6360.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
10	7700000.000000	6000.000000	4.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
11	7560000.000000	6000.000000	3.000000	2.000000	3.000000	1.000000	0E-7	0E-7	0E-7	
12	7525000.000000	6000.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
13	7490000.000000	6600.000000	3.000000	1.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
14	7350000.000000	6000.000000	3.000000	1.000000	2.000000	1.000000	0E-7	0E-7	0E-7	
15	7343000.000000	11440.000000	4.000000	1.000000	2.000000	1.000000	0E-7	1.000000	0E-7	

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

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom
1	12215000.000000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E	
2	9870000.000000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.0	
3	9800000.000000	5750.000000	3.000000	2.000000	4.000000	1.000000	1.0	
4	9100000.000000	6000.000000	4.000000	1.000000	2.000000	1.000000	0E	
5	8855000.000000	6420.000000	3.000000	2.000000	2.000000	1.000000	0E	
6	8575000.000000	8800.000000	3.000000	2.000000	2.000000	1.000000	0E	
7	8120000.000000	6840.000000	5.000000	1.000000	2.000000	1.000000	1.0	
8	8080940.000000	7000.000000	3.000000	2.000000	4.000000	1.000000	0E	
9	7840000.000000	6360.000000	3.000000	2.000000	4.000000	1.000000	0E	
10	7700000.000000	6000.000000	4.000000	2.000000	4.000000	1.000000	0E	
11	7560000.000000	6000.000000	3.000000	2.000000	3.000000	1.000000	0E	
12	7525000.000000	6000.000000	3.000000	2.000000	4.000000	1.000000	0E	
13	7490000.000000	6600.000000	3.000000	1.000000	4.000000	1.000000	0E	
14	7350000.000000	6000.000000	3.000000	1.000000	2.000000	1.000000	0E	
15	7343000.000000	11440.000000	4.000000	1.000000	2.000000	1.000000	0E	

Existing Model Execution

Model: (from Tab:)NORMALIZE_TR...

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
- heating -> Double

☐ Transfer Column(s) to Output

Execute Cancel

The results will appear on the output spreadsheet.

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating
1	12215000.000000	7500.000000	4.000000	2.000000	2.000000	1.000000	0E-7	1.000000	0E-7	
2	9870000.000000	8100.000000	4.000000	1.000000	2.000000	1.000000	1.000000	1.000000	0E-7	
3	9800000.000000	5750.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
4	9100000.000000	6000.000000	4.000000	1.000000	2.000000	1.000000	0E-7	1.000000	0E-7	
5	8855000.000000	6420.000000	3.000000	2.000000	2.000000	1.000000	0E-7	0E-7	0E-7	
6	8575000.000000	8800.000000	3.000000	2.000000	2.000000	1.000000	0E-7	0E-7	0E-7	
7	8120000.000000	6840.000000	5.000000	1.000000	2.000000	1.000000	1.000000	1.000000	0E-7	
8	8080940.000000	7000.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
9	7840000.000000	6360.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
10	7700000.000000	6000.000000	4.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
11	7560000.000000	6000.000000	3.000000	2.000000	3.000000	1.000000	0E-7	0E-7	0E-7	
12	7525000.000000	6000.000000	3.000000	2.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
13	7490000.000000	6600.000000	3.000000	1.000000	4.000000	1.000000	0E-7	0E-7	0E-7	
14	7350000.000000	6000.000000	3.000000	1.000000	2.000000	1.000000	0E-7	0E-7	0E-7	
15	7343000.000000	11440.000000	4.000000	1.000000	2.000000	1.000000	0E-7	1.000000	0E-7	

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 table with 15 rows and 11 columns. The columns are: User Header, User Row ID, price, area, bedrooms, bathrooms, stories, mainroad, guestroom, basement, and hotelling. The data is as follows:

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotelling
1	1	13300000.000000	1.0188414	1.4513175	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.24
2	2	12250000.000000	1.7183192	1.4513175	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.24
3	3	12250000.000000	2.1725255	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.24
4	4	11410000.000000	1.0188414	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.24
5	5	10850000.000000	1.0551779	0.0644276	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094	-0.24
6	6	10150000.000000	1.5457208	1.4513175	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.24
7	7	10150000.000000	5.0067730	2.8382074	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.24
8	8	9800000.000000	3.6441540	0.0644276	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.24
9	9	9681000.000000	0.3738684	1.4513175	3.2224865	0.2132186	0.4019153	2.0906616	1.3405094	4.09
10	10	9310000.000000	0.6236819	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.24
11	11	9240000.000000	-0.7616474	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	4.09
12	12	9240000.000000	1.1914398	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.24
13	13	9100000.000000	0.6463922	1.4513175	1.3206150	0.2132186	0.4019153	2.0906616	1.3405094	-0.24
14	14	8960000.000000	1.5093843	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.24
15	15	8890000.000000	-0.2620204	0.0644276	1.3206150	0.2132186	0.4019153	2.0906616	-0.7441611	-0.24

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.

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. 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.

File Edit Data Transformation Analytics Statistics Plot Help																				
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	User Header	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)	Col9	Col10	Col11	Col12	Col13
88		6090000.0000	1.4512458	0.0644276	-0.5812566	1.3606716	1	Regression Statistics												
89		6083000.0000	2.0180953	0.0644276	-0.5812566	-0.9342343	2	Multiple R	0.8247151											
		6020000.0000					3	R Square	0.6801350											
90		0.7826541	0.0644276	1.3206150	-0.9342343		4	Adjusted R Square	0.6704827											
91		5950000.0000	-0.6708061	1.4513175	-0.5812566	0.2132186	5	Standard Error	1079890.2152											
		000					6	Observations	409.0000000											
92		5950000.0000	0.5646351	0.0644276	-0.5812566	-0.9342343	7													
		000					8	Degrees of Freedom		Sum of Squares		Mean Square	F-statistic	Significance F						
93		5950000.0000	0.8371589	0.0644276	-0.5812566	-0.9342343	9	Regression	12.0000000	8032543121	81813453601	500.2700000	70.1749766	0E-7						
94		5950000.0000	0.9329964	0.0644276	-0.5812566	0.2132186	10	Residual	396.0000000	1034249026	1166162819	1166162819								
95		5950000.0000	0.9734208	1.4513175	1.3206150	0.2132186	11	Total	408.0000000	4118.3000000	34.6420000									
		000					12	Coefficients		Standard Error	t-statistic	P-value	Lower 95.0%	Upper 95.0%						
96		5950000.0000	0.6123268	0.0644276	1.3206150	2.5081245	13	price	4794101.4669	53397.134547	88.7819887	0E-7	4689124.1631	4899078.7708						
		000					14	area	560208.79134	61282.501049	9.1399238	0E-7	4397028.41329	6087108.16949						
97		5943000.0000	4.7342492	0.0644276	-0.5812566	-0.9342343	15	bedrooms	75039.579377	63116.105348	1.1889133	0.2351862	-48044.95488	199124.11361						
98		5880000.0000	0.9007478	0.0644276	-0.5812566	-0.9342343	16	bathrooms	517787.31358	61150.704923	8.4873973	0E-7	397566.70274	638007.82443						
99		5880000.0000	0.6009716	0.0644276	1.3206150	1.3606716	17	stories	389080.36150	64395.503818	6.0420424	0E-7	262480.54882	515680.15809						
100		5873000.0000	2.8538350	0.0644276	-0.5812566	1.3606716	18	mainroad	147362.87864	57746.843480	2.5522240	0.0110783	33854.168130	240911.58916						
101		5866000.0000	-0.1711791	0.0644276	-0.5812566	-0.9342343	19	guestroom	150890.32531	58726.451010	2.5893759	0.0105534	35435.731254	268344.91896						
102		5810000.0000	0.2957450	1.4513175	-0.5812566	2.5081245	20	basement	148045.07738	62497.246238	2.3386242	0.0199461	23177.203202	268912.95156						
103		5810000.0000	0.0105034	0.0644276	-0.5812566	1.3606716	21	hotwaterheat	194876.78249	54770.487323	3.5580632	0.0004189	67199.544009	302554.02098						
104		5810000.0000	-0.1711791	0.0644276	-0.5812566	1.3606716	22	aircondition	194146.58127	58157.058448	6.7772197	0E-7	279811.82462	508485.33792						
105		5803000.0000	0.8280748	0.0644276	-0.5812566	-0.9342343	23	parking	238011.54284	59231.196101	4.0183477	0.0000702	121564.63377	354458.45192						
106		5775000.0000	0.3738684	0.0644276	1.3206150	2.5081245	24	prefarea	286881.67317	57734.062958	4.9890193	0.0000010	173378.08889	400385.25946						
107		5740000.0000	-0.2438522	1.4513175	-0.5812566	0.2132186	25	furnishingstat	144243.96208	56103.819104	2.5710542	0.0105032	33947.390497	254544.53363						
		56455000.0000					26													
IMPORT TRAIN_TEST_SPLIT NORMALIZE_TRAIN_SET NORMALIZE_TEST_SET FEATURE_SELECTION_REGRESSION																				

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".

User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement
1	13300000.0000	1.0188414	1.4513175	1.3206150	1.3606716	0.4019153	-0.4771480	-0.74	
2	12250000.0000	1.7183192	1.4513175	5.1243580	2.5081245	0.4019153	-0.4771480	-0.74	
3	12250000.0000	2.1725255	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.340	
4	11410000.0000	1.0188414	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	1.340	
5	10850000.0000	1.0551779	0.0644276	3.2224865	-0.9342343	0.4019153	-0.4771480	1.340	
6	10150000.0000	1.5457208	1.4513175	3.2224865	2.5081245	0.4019153	-0.4771480	-0.74	
7	10150000.0000	5.0067730	2.8382074	3.2224865	0.2132186	0.4019153	-0.4771480	-0.74	
8	9800000.0000	3.6441540	0.0644276	-0.5812566	0.2132186	0.4019153	-0.4771480	1.340	
9	9681000.0000	0.3738684	1.4513175	3.2224865	0.2132186	0.4019153	2.0906616	1.340	
10	9310000.0000	0.6236819	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.74	
11	9240000.0000	-0.7616474	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.74	
12	9240000.0000	1.1914398	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.74	
13	9100000.0000	0.6463922	1.4513175	1.3206150	0.2132186	0.4019153	2.0906616	1.340	
14	8960000.0000	1.6002843	0.0644276	1.3206150	3.6081345	0.4019153	-0.4771480	-0.74	

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)".

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

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1		13300000.0000	1.0188414	1.4513175	1.3206150	1.3606716	0.4019153
2		12250000.0000	1.7183192	1.4513175	5.1243580	2.5081245	0.4019153
3		12250000.0000	2.1725255	0.0644276	1.3206150	0.2132186	0.4019153
4		11410000.0000	1.0188414	1.4513175	-0.5812566	0.2132186	0.4019153
5		10850000.0000	1.0551779	0.0644276	3.2224865	-0.9342343	0.4019153
6		10150000.0000	1.5457208	1.4513175	3.2224865	2.5081245	0.4019153
7		10150000.0000	5.0067730	2.8382074	3.2224865	0.2132186	0.4019153
8		9800000.0000	3.6441540	0.0644276	-0.5812566	0.2132186	0.4019153
9		9681000.0000	0.3738684	1.4513175	3.2224865	0.2132186	0.4019153
10		9310000.0000	0.6236819	1.4513175	1.3206150	0.2132186	0.4019153
11		9240000.0000	-0.7616474	1.4513175	1.3206150	0.2132186	0.4019153
12		9240000.0000	1.1914398	0.0644276	1.3206150	0.2132186	0.4019153
13		9100000.0000	0.6463922	1.4513175	1.3206150	0.2132186	0.4019153
14		8960000.0000	1.6002843	0.0644276	1.3206150	3.6081345	0.4019153

Select Column(s)

Excluded Columns

Col4 -- bedrooms

Included Columns

Col2 -- price

Col3 -- area

Col5 -- bathrooms

Col6 -- stories

Col7 -- mainroad

Col8 -- guestroom

Col9 -- basement

Col10 -- hotwaterheating

>>

>

<

<<

Execute

Cancel

The results will appear on the output spreadsheet.

FileEditData TransformationAnalyticsStatisticsPlotHelp

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

FEATURE_SELECTION_REGRESSION

NORMALIZE_TEST_SET

FEATURE_SELECTION_TRAIN_SET

FEATURE_SELECTION_TEST_SET

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
User Header	User Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement
1		13300000.000000	1.0188414	1.4513175	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441
2		12250000.000000	1.7183192	1.4513175	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441
3		12250000.000000	2.1725255	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	1.34050
4		11410000.000000	1.0188414	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	1.34050
5		10850000.000000	1.0551779	0.0644276	3.2224865	-0.9342343	0.4019153	-0.4771480	1.34050
6		10150000.000000	1.5457208	1.4513175	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441
7		10150000.000000	5.0067730	2.8382074	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441
8		9600000.000000	3.6441540	0.0644276	-0.5812566	0.2132186	0.4019153	-0.4771480	1.34050
9		9681000.000000	0.3738684	1.4513175	3.2224865	0.2132186	0.4019153	2.0906616	1.34050
10		9310000.000000	0.6236819	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441
11		9240000.000000	-0.7616474	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441
12		9240000.000000	1.1914398	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441
13		9100000.000000	0.6463922	1.4513175	1.3206150	0.2132186	0.4019153	2.0906616	1.34050
14		8960000.000000	1.5093043	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom	basement
1		13300000.000000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441
2		12250000.000000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441
3		12250000.000000	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.34050
4		11410000.000000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.34050
5		10850000.000000	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.34050
6		10150000.000000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441
7		10150000.000000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441
8		9600000.000000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.34050
9		9681000.000000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.34050
10		9310000.000000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441
11		9240000.000000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441
12		9240000.000000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441
13		9100000.000000	0.6463922	1.3206150	0.2132186	0.4019153	2.0906616	1.34050
14		8960000.000000	1.5093043	1.3206150	2.5081245	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".

File

Edit

Data Transformation

Analytics

Statistics

Plot

Help

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

NORMALIZE_TEST_SET

FEATURE_SELECTION_REGRESSION

FEATURE_SELECTION_TRAIN_SET

FEATURE_SELECTION_TEST_SET

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1	12215000.000000	1.0551779	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	
2	9870000.000000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	
3	9800000.000000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	2.0906616	
4	9100000.000000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.4771480	
5	8855000.000000	0.5646351	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	
6	8575000.000000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	
7	8120000.000000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.0906616	
8	8080940.000000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	
9	7840000.000000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	
10	7700000.000000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.4771480	
11	7560000.000000	0.3738684	0.0644276	1.3206150	1.3606716	0.4019153	-0.4771480	
12	7525000.000000	0.3738684	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	

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									
17									

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

NORMALIZE_TEST_SET

FEATURE_SELECTION_REGRESSION

FEATURE_SELECTION_TRAIN_SET

FEATURE_SELECTION_TEST_SET

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 'Data Transformation' menu with options: Normalizers, Data Manipulation, Split, Variable Selection, Remove Column(s), Select Column(s), Matrix Transpose, Sort by Column, and Fill Missing Column(s) Values. The 'Select Column(s)' dialog box is open, showing 'Excluded Columns' (Col4 -- bedrooms) and 'Included Columns' (Col2 -- price, Col3 -- area, Col5 -- bathrooms, Col6 -- stories, Col7 -- mainroad, Col8 -- guestroom, Col9 -- basement, Col10 -- hotwaterheating, Col11 -- airconditioning). The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1	12215000.000000	1.0551779	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	
2	9870000.000000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	
3	9800000.000000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	2.0906616	
4	9100000.000000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.4771480	
5	8855000.000000	0.5646351	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	
6	8575000.000000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	
7	8120000.000000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.0906616	
8	8080940.000000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	
9	7840000.000000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	
10	7700000.000000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.4771480	
11	7560000.000000	0.3738684	0.0644276	1.3206150	1.3606716	0.4019153	-0.4771480	
12	7525000.000000	0.3738684	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	

The results will appear on the output spreadsheet.

The screenshot shows the 'Data Transformation' menu with options: Normalizers, Data Manipulation, Split, Variable Selection, Remove Column(s), Select Column(s), Matrix Transpose, Sort by Column, and Fill Missing Column(s) Values. The 'Select Column(s)' dialog box is open, showing 'Excluded Columns' (Col4 -- bedrooms) and 'Included Columns' (Col2 -- price, Col3 -- area, Col5 -- bathrooms, Col6 -- stories, Col7 -- mainroad, Col8 -- guestroom, Col9 -- basement, Col10 -- hotwaterheating, Col11 -- airconditioning). The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1	12215000.000000	1.0551779	1.4513175	1.3206150	0.2132186	0.4019153	-0.4771480	
2	9870000.000000	1.3277017	1.4513175	-0.5812566	0.2132186	0.4019153	2.0906616	
3	9800000.000000	0.2603169	0.0644276	1.3206150	2.5081245	0.4019153	2.0906616	
4	9100000.000000	0.3738684	1.4513175	-0.5812566	0.2132186	0.4019153	-0.4771480	
5	8855000.000000	0.5646351	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	
6	8575000.000000	1.6456462	0.0644276	1.3206150	0.2132186	0.4019153	-0.4771480	
7	8120000.000000	0.7554018	2.8382074	-0.5812566	0.2132186	0.4019153	2.0906616	
8	8080940.000000	0.8280748	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	
9	7840000.000000	0.5373827	0.0644276	1.3206150	2.5081245	0.4019153	-0.4771480	
10	7700000.000000	0.3738684	1.4513175	1.3206150	2.5081245	0.4019153	-0.4771480	
11	7560000.000000	0.3738684	0.0644276	1.3206150	1.3606716	0.4019153	-0.4771480	
12	7525000.000000	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".

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 bar, a workflow diagram is visible with steps: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, and TRAIN_MODEL. Below the workflow, a data table is displayed with columns: User Header, User Row ID, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D). The table contains 12 rows of data. To the right of the main table, there is a smaller table with columns: User Header, User Row ID, Col1, Col2, Col3, Col4, Col5, Col6, Col7, Col8. This table is currently empty.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	1	13300000.0000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	
2	2	12250000.0000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611	
3	3	12250000.0000	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	
4	4	11410000.0000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	
5	5	10850000.0000	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094	
6	6	10150000.0000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611	
7	7	10150000.0000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	
8	8	9800000.0000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	
9	9	9681000.0000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.3405094	
10	10	9310000.0000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	
11	11	9240000.0000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	
12	12	9240000.0000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	

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 screenshot shows the Isalos Analytics Platform interface. The 'Analytics' menu is open, showing options: Regression, Classification, Clustering, Anomaly Detection, and Existing Model Utilization. The 'Regression' option is selected, and a sub-menu is displayed with options: k-Nearest Neighbors (kNN), Fully Connected Neural Network, Radial Basis Function Network, Linear SGD, XGBoost, and Random Forest. Below the menu, a data table is displayed with columns: User Header, User Row ID, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D). The table contains 12 rows of data.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	1	13300000.0000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	
2	2	12250000.0000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	
3	3	12250000.0000	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	
4	4	11410000.0000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	
5	5	10850000.0000	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	
6	6	10150000.0000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	
7	7	10150000.0000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	
8	8	9800000.0000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	
9	9	9681000.0000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	
10	10	9310000.0000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	
11	11	9240000.0000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	
12	12	9240000.0000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	

The screenshot shows the 'kNN Regression Model' configuration dialog. It has a title bar with a close button. The 'Target Column' is set to 'Col2 -- price'. The 'Number of Neighbors' is set to 5. There are 'Execute' and 'Cancel' buttons at the bottom.

The predictions will appear on the output spreadsheet.

FileEditData TransformationAnalyticsStatisticsPlotHelp

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

FEATURE_SELECTION_REGRESSION

FEATURE_SELECTION_TRAIN_SET

FEATURE_SELECTION_TEST_SET

TRAIN_MODEL.fit

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom	basement
1		13300000.000000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611
2		12250000.000000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611
3		12250000.000000	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094
4		11410000.000000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094
5		10850000.000000	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094
6		10150000.000000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611
7		10150000.000000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611
8		9800000.000000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094
9		9681000.000000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.3405094
10		9310000.000000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611
11		9240000.000000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611
12		9240000.000000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)	Col8 (S)
User Header	User Row ID	price	RNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3
1		13300000.000000	13188929.4110539	Entry 1	OE-7	Entry 48	0.0741518	Entry 27
2		12250000.000000	12229889.4121292	Entry 2	OE-7	Entry 30	0.7453612	Entry 14
3		12250000.000000	12205901.1738684	Entry 3	OE-7	Entry 105	0.5148770	Entry 213
4		11410000.000000	11363133.2398893	Entry 4	OE-7	Entry 311	0.3809051	Entry 109
5		10850000.000000	10826633.7549654	Entry 5	OE-7	Entry 24	0.5940605	Entry 93
6		10150000.000000	10145159.9494460	Entry 6	OE-7	Entry 1	0.6919589	Entry 48
7		10150000.000000	10119807.5316795	Entry 7	OE-7	Entry 304	0.7793816	Entry 361
8		9800000.000000	9781644.4849248	Entry 8	OE-7	Entry 49	0.4922288	Entry 65
9		9681000.000000	9664271.3162595	Entry 9	OE-7	Entry 124	0.9579445	Entry 148
10		9310000.000000	9292345.9693808	Entry 10	OE-7	Entry 114	0.4948162	Entry 27
11		9240000.000000	9205858.5426032	Entry 11	OE-7	Entry 208	0.4765667	Entry 242
12		9240000.000000	9198353.0763197	Entry 12	OE-7	Entry 172	0.3732271	Entry 231

IMPORTTRAIN_TEST_SPLITNORMALIZE_TRAIN_SETFEATURE_SELECTION_REGRESSIONFEATURE_SELECTION_TRAIN_SETFEATURE_SELECTION_TEST_SETTRAIN_MODEL.fit+

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 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, and TRAIN_MODEL.fit. Below the diagram, there are two data tables. The left table has columns: User Header, User Row ID, price, area, bathrooms, stories, mainroad, guestroom, basement, hotwatering. The right table has columns: User Header, User Row ID, price, area, bathrooms, stories, mainroad, guestroom, basement, hotwatering.

User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom	basement	hotwatering
1	12215000.000000	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438000	
2	9870000.000000	1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438000	
3	9800000.000000	0.2603169	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438000	
4	9100000.000000	0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438000	
5	8855000.000000	0.5646351	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438000	
6	8575000.000000	1.6456462	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438000	
7	8120000.000000	0.7554018	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438000	
8	8080940.000000	0.8280748	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000	
9	7840000.000000	0.5373827	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000	
10	7700000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000	
11	7560000.000000	0.3738684	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438000	
12	7525000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000	

User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom	basement	hotwatering
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									

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 screenshot shows the hp.ekk interface with the 'Analytics' menu open, highlighting 'Existing Model Utilization'. Below it, a table displays data for 12 rows. To the right, the 'Existing Model Execution' dialog is open, showing the 'kNN Model' configuration. The 'Model Input' section lists various features like 'area', 'bedrooms', 'bathrooms', etc., with their datatypes. The 'Excluded Columns' list includes 'area', 'bathrooms', 'stories', 'mainroad', 'guestroom', 'basement', 'hotwaterheating', and 'airconditioning'. The 'Included Columns' list includes 'price'.

User Header	Col1 (D)	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	12215000.000000	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3400509	-0.2438000
2	9870000.000000	1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616	1.3400509	-0.2438000
3	9800000.000000	0.2603169	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438000
4	9100000.000000	0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3400509	-0.2438000
5	8855000.000000	0.5646351	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438000
6	8575000.000000	1.6456462	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438000
7	8120000.000000	0.7554018	-0.5812566	0.2132186	0.4019153	2.0906616	1.3400509	-0.2438000
8	8080940.000000	0.8280748	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000
9	7840000.000000	0.5373827	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000
10	7700000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000
11	7560000.000000	0.3738684	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438000
12	7525000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000

The predictions will appear on the output spreadsheet.

The screenshot shows the hp.ekk interface with the 'VALIDATE_MODEL(predict)' tab selected. The output spreadsheet displays predictions for the 12 rows. The 'User Header' row includes 'User Row ID', 'price', 'area', 'bathrooms', 'stories', 'mainroad', 'guestroom', 'basement', and 'hotwaterheating'. The 'User Header' row for the output includes 'User Row ID', 'kNN Prediction', 'Closest NN1', 'Distance from NN1', 'Closest NN2', 'Distance from NN2', 'Closest NN3', and 'Distance from NN3'.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	12215000.000000	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3400509	-0.2438000
2	9870000.000000	1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616	1.3400509	-0.2438000
3	9800000.000000	0.2603169	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438000
4	9100000.000000	0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3400509	-0.2438000
5	8855000.000000	0.5646351	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438000
6	8575000.000000	1.6456462	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438000
7	8120000.000000	0.7554018	-0.5812566	0.2132186	0.4019153	2.0906616	1.3400509	-0.2438000
8	8080940.000000	0.8280748	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000
9	7840000.000000	0.5373827	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000
10	7700000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000
11	7560000.000000	0.3738684	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438000
12	7525000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438000

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 hp.ekk interface with a workflow diagram at the top and a data table below. The workflow includes steps: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, FEATURE_SELECTION_REGRESSION, NORMALIZE_TEST_SET, TRAIN_MODEL, and STATISTICS_ACCURACIES. The data table has columns: User Header, Col1, Col2 (D), Col3 (S), Col4 (D), Col5 (S), Col6 (D), Col7 (S), Col8 (D), Col9 (S). The table contains 12 rows of data, including User Row ID, kNN Prediction, and various distances from NN1, NN2, and NN3.

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)
User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4
1	8394729.1211	079	Entry 8	0.6174273	Entry 5	0.7637626	Entry 93	0.7645004	Entry 65
2	6824551.4683	578	Entry 20	0.3507424	Entry 61	0.3732271	Entry 115	0.6929062	Entry 94
3	6439216.7259	801	Entry 120	0.7641550	Entry 31	1.1056799	Entry 56	1.1067204	Entry 78
4	4813236.0342	003	Entry 234	0.3602715	Entry 70	0.3625968	Entry 176	0.3648357	Entry 192
5	9151949.6764	435	Entry 10	0.0090941	Entry 114	0.4921283	Entry 27	0.5114741	Entry 59
6	5759676.3731	597	Entry 144	0.3437411	Entry 47	0.3783940	Entry 132	0.4718113	Entry 53
7	6462229.4133	875	Entry 311	0.1437566	Entry 109	0.1497027	Entry 4	0.3357936	Entry 18
8	8305353.8342	013	Entry 14	0.1049318	Entry 30	0.1399091	Entry 35	0.3405947	Entry 41
9	7800748.8221	759	Entry 59	0.3333598	Entry 60	0.3334772	Entry 48	0.7453560	Entry 1
10	5409578.3681	111	Entry 75	0.3333333	Entry 96	0.6020464	Entry 286	0.6793799	Entry 188
11	6017173.8042	042	Entry 66	0.3333333	Entry 103	0.3379986	Entry 32	0.4714045	Entry 150
12	7563332.4143	000	Entry 35	0E-7	Entry 14	0.3764257	Entry 30	0.3938957	Entry 68

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

The screenshot shows the hp.ekk interface with the Statistics menu open, highlighting the Regression Metrics option. Below the menu is a data table with columns: User Header, Col1, Col2 (D), Col3 (S), Col4 (D), Col5 (S), Col6 (D), Col7 (S), Col8. The table contains 12 rows of data, including User Row ID, kNN Prediction, and various distances from NN1, NN2, and NN3.

The Regression Statistics Metrics dialog is open, showing the following settings:

- Actual Value Column: Col13 -- price
- Prediction Value Column: Col2 -- kNN Prediction
- Buttons: Execute, Cancel

The results 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, 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.

	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)
User Header	User Row ID	NN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4
1		8394729.1211079	Entry 8	0.6174273	Entry 5	0.7637626	Entry 93	0.7645004	Entry 65
2		6824551.4683578	Entry 20	0.3507424	Entry 61	0.3732271	Entry 115	0.6929062	Entry 94
3		6438216.7259801	Entry 120	0.7641550	Entry 31	1.1056799	Entry 56	1.1067204	Entry 78
4		4813236.034003	Entry 234	0.3602715	Entry 70	0.3625968	Entry 176	0.3648357	Entry 192
5		9151949.6764435	Entry 10	0.0090941	Entry 114	0.4921283	Entry 27	0.5114741	Entry 59
6		5759676.3731597	Entry 144	0.3437411	Entry 47	0.3783940	Entry 132	0.4718113	Entry 53
7		6462229.4133875	Entry 311	0.1437566	Entry 109	0.1497027	Entry 4	0.3357936	Entry 18
8		8305353.8342013	Entry 14	0.1049318	Entry 30	0.1399091	Entry 35	0.3405947	Entry 41
9		7808748.8221759	Entry 59	0.3333598	Entry 60	0.3334772	Entry 48	0.7453560	Entry 1
10		5409578.3681111	Entry 75	0.3333333	Entry 96	0.6020464	Entry 286	0.6793799	Entry 188
11		6017173.8042042	Entry 66	0.3333333	Entry 103	0.3379986	Entry 32	0.4714045	Entry 150
12		7563332.4143000	Entry 35	0E-7	Entry 14	0.3764257	Entry 30	0.3938957	Entry 68

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared		
1		14108899052	1187808.8673	853056.98213	0.5852034		
2		94.4163000	244	28			
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							

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".

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, 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.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom	basement
1		13300000.0000000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611
2		12250000.0000000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611
3		12250000.0000000	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094
4		11410000.0000000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094
5		10850000.0000000	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094
6		10150000.0000000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611
7		10150000.0000000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611
8		9800000.0000000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094
9		9681000.0000000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.3405094
10		9310000.0000000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611
11		9240000.0000000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611
12		9240000.0000000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
User Header	User Row ID							
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								

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".

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom
1		1330000.000000	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480
2		1225000.000000	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480
3		1225000.000000	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480
4		1141000.000000	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616
5		1085000.000000	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480
6		1015000.000000	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480
7		1015000.000000	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480
8		980000.000000	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480
9		9681000.000000	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616
10		931000.000000	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480
11		924000.000000	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480
12		924000.000000	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480

Excluded Columns

Col2 -- price

Included Columns

Col3 -- area
Col4 -- bathrooms
Col5 -- stories
Col6 -- mainroad
Col7 -- guestroom
Col8 -- basement
Col9 -- hotwaterheating
Col10 -- airconditioning
Col11 -- ...

Execute

Cancel

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".

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating
1		1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438028
2		1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028
3		2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028
4		1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028
5		1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094	-0.2438028
6		1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028
7		5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028
8		3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028
9		0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.3405094	4.0916470
10		0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028
11		-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	4.0916470
12		1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028
13		0.6463922	1.3206150	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028
14		1.5093843	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028
15		-0.2620204	1.3206150	0.2132186	0.4019153	2.0906616	-0.7441611	-0.2438028
16		-0.3891982	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	4.0916470
17		0.8984767	1.3206150	-0.9342343	0.4019153	2.0906616	1.3405094	-0.2438028

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

hp.ekk

File Edit Data Transformation Analytics Statistics Plot Help

Domain - APD

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

Perform Computations CPU (double precision)

Execute Cancel

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7
User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	
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	0.8984767	1.3206150	-0.9342343	0.4019153	2.0906616	1.3405	

The results will appear on the output spreadsheet.

hp.ekk

File Edit Data Transformation Analytics Statistics Plot Help

EXCLUDE_PRICE

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

Perform Computations CPU (double precision)

Execute Cancel

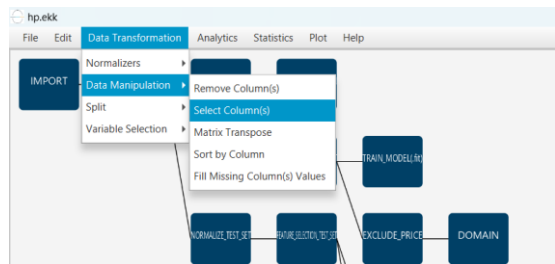
User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	
1	1.0188414	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.4
2	1.7183192	5.1243580	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.4
3	2.1725255	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	-0
4	1.0188414	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	1.4
5	1.0551779	3.2224865	-0.9342343	0.4019153	-0.4771480	1.3405094	-0.2438028	1.4
6	1.5457208	3.2224865	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.4
7	5.0067730	3.2224865	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	-0
8	3.6441540	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	1.4
9	0.3738684	3.2224865	0.2132186	0.4019153	2.0906616	1.3405094	4.0916470	-0
10	0.6236819	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.4
11	-0.7616474	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	4.0916470	-0
12	1.1914398	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	-0
13	0.6463922	1.3206150	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	1.4
14	1.5093843	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.4
15	-0.2620204	1.3206150	0.2132186	0.4019153	2.0906616	-0.7441611	-0.2438028	1.4
16	-0.3891982	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	4.0916470	-0
17	0.8984767	1.3206150	-0.9342343	0.4019153	2.0906616	1.3405094	-0.2438028	1.4

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".

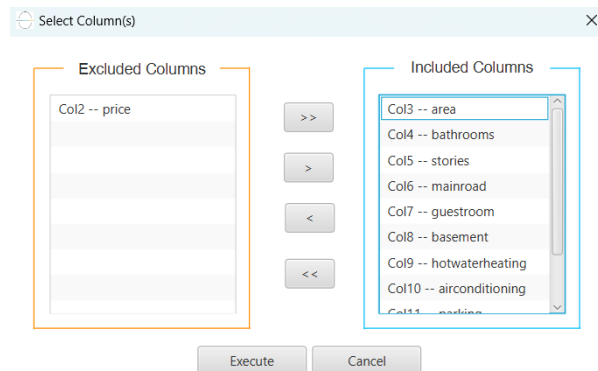
Import data into the input spreadsheet of the "EXCLUDE_PRICE_TEST_SET" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

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".



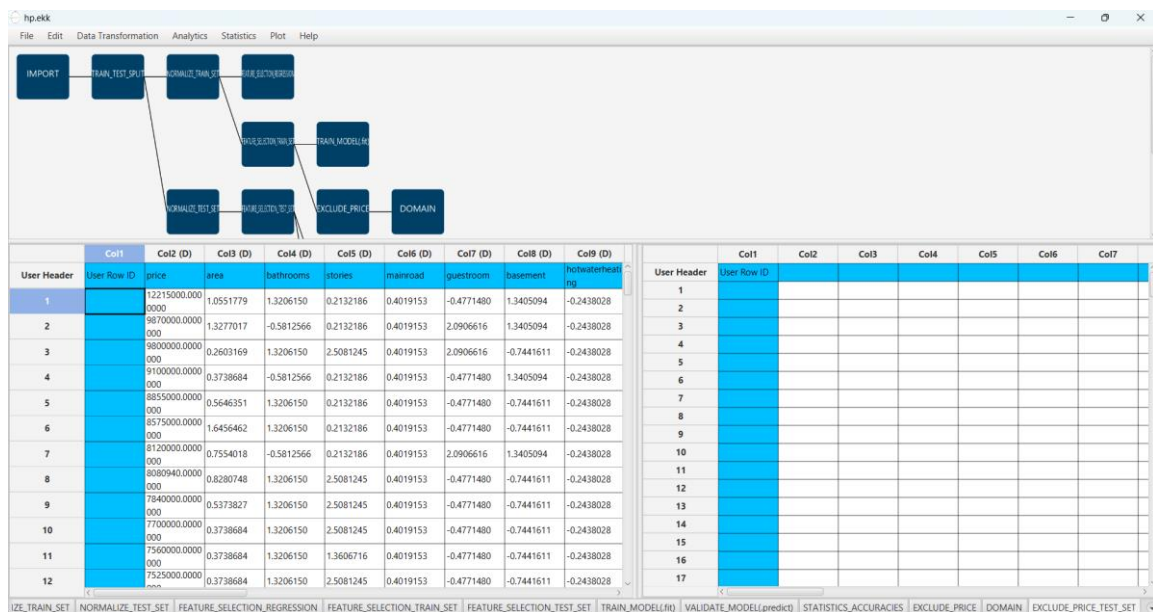
The screenshot shows the 'Data Transformation' menu in the hp.ekk application. The 'Select Column(s)' option is highlighted. Below the menu, a workflow diagram shows data flow from 'IMPORT' through 'NORMALIZE_TEST_SET', 'FEATURE_SELECTION_REGRESSION', and 'EXCLUDE_PRICE' to 'DOMAIN'. A data table is displayed below the workflow.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
User Header	User Row ID	price	area	bathrooms	stories	mainroad	guestroom
1		12215000.000000	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771481
2		9870000.000000	1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616
3		9800000.000000	0.2603169	1.3206150	2.5081245	0.4019153	2.0906616
4		9100000.000000	0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771481
5		8855000.000000	0.5646351	1.3206150	0.2132186	0.4019153	-0.4771481
6		8575000.000000	1.6456462	1.3206150	0.2132186	0.4019153	-0.4771481
7		8120000.000000	0.7554018	-0.5812566	0.2132186	0.4019153	2.0906616
8		8080940.000000	0.8280748	1.3206150	2.5081245	0.4019153	-0.4771481
9		7840000.000000	0.5373827	1.3206150	2.5081245	0.4019153	-0.4771481
10		7700000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771481
11		7560000.000000	0.3738684	1.3206150	1.3606716	0.4019153	-0.4771481
12		7525000.000000	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771481



The 'Select Column(s)' dialog box shows a list of columns. The 'Excluded Columns' list contains 'Col2 -- price'. The 'Included Columns' list contains 'Col3 -- area', 'Col4 -- bathrooms', 'Col5 -- stories', 'Col6 -- mainroad', 'Col7 -- guestroom', 'Col8 -- basement', 'Col9 -- hotwaterheating', and 'Col10 -- airconditioning'. Buttons for '>>', '>', '<', and '<<' are available between the lists. 'Execute' and 'Cancel' buttons are at the bottom.

The results will appear on the output spreadsheet.



The screenshot shows the hp.ekk application with the output spreadsheet. The spreadsheet has columns for 'User Header', 'User Row ID', 'price', 'area', 'bathrooms', 'stories', 'mainroad', 'guestroom', 'basement', 'hotwaterheating', and 'airconditioning'. The data is organized into tabs: 'IMPORT', 'TRAIN_TEST_SPLIT', 'NORMALIZE_TEST_SET', 'FEATURE_SELECTION_REGRESSION', 'EXCLUDE_PRICE', and 'DOMAIN'. The output spreadsheet shows the results of the feature selection process, with columns for 'User Header', 'User Row ID', 'price', 'area', 'bathrooms', 'stories', 'mainroad', 'guestroom', 'basement', 'hotwaterheating', and 'airconditioning'.

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".

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
User Header	User Row ID	area	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning
1		1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	1.41
2		1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	1.41
3		0.2603169	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438028	1.41
4		0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	-0.71
5		0.5646351	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
6		1.6456462	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
7		0.7554018	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	1.41
8		0.8280748	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
9		0.5373827	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
10		0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	-0.71
11		0.3738684	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
12		0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
13		0.6463922	-0.5812566	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
14		0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41
15		2.8447509	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	-0.71
16		1.7364874	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438028	1.41
17		-0.5345442	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	1.41

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

Existing Model Execution

Model: (from Tab:) DOMAIN

Type: APD Model

Description:

Model Input:

Header	Datatype
area	Double
bedrooms	Double
bathrooms	Double
stories	Double
mainroad	Double
guestroom	Double
basement	Double
hotwaterheating	Double
airconditioning	Double

☒ Transfer Column(s) to Output

Excluded Columns:

Included Columns:

- Col2 -- area
- Col3 -- bathrooms
- Col4 -- stories
- Col5 -- mainroad
- Col6 -- guestroom
- Col7 -- basement
- Col8 -- hotwaterheating
- Col9 -- airconditioning

Execute Cancel

The results will appear on the output spreadsheet. There are no unreliable samples in the test set.

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 workflow diagram shows the process: IMPORT → TRAIN_TEST_SPLIT → NORMALIZE_TRAIN_SET → FEATURE_SELECTION_REGRESSION → FEATURE_SELECTION_TRAIN_SET → TRAIN_MODEL(fit). Below the workflow, a data table is shown with columns for User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), and Col9. The table contains 17 rows of data, including User Row ID, area, bathrooms, stories, mainroad, guestroom, basement, and reliability. The bottom of the interface shows a list of data sets: NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(fit), VALIDATE_MODEL(predict), STATISTICS_ACCURACIES, EXCLUDE_PRICE, DOMAIN, EXCLUDE_PRICE_TEST_SET, and RELIABILITY.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
1	1.0551779	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	1.41	
2	1.3277017	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	1.41	
3	0.2603169	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438028	1.41	
4	0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	-0.7	
5	0.5646351	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
6	1.6456462	1.3206150	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
7	0.7554018	-0.5812566	0.2132186	0.4019153	2.0906616	1.3405094	-0.2438028	1.41	
8	0.8280748	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
9	0.5373827	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
10	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	-0.7	
11	0.3738684	1.3206150	1.3606716	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
12	0.3738684	1.3206150	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
13	0.6463922	-0.5812566	2.5081245	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
14	0.3738684	-0.5812566	0.2132186	0.4019153	-0.4771480	-0.7441611	-0.2438028	1.41	
15	2.8447509	-0.5812566	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	-0.7	
16	1.7364874	1.3206150	2.5081245	0.4019153	2.0906616	-0.7441611	-0.2438028	1.41	
17	-0.5345442	1.3206150	0.2132186	0.4019153	-0.4771480	1.3405094	-0.2438028	1.41	

Final Isalos Workflow

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

