



Vehicle dataset

The "Vehicle Dataset", which can be found in <https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardexho>, contains information about used cars collected from online platforms, aiming to support machine learning applications such as price estimation. It includes key features such as brand, model, year of manufacture, fuel type, transmission, mileage, and price. This dataset is often used to train regression models that predict vehicle prices based on multiple factors, and serves as a practical example for data preprocessing, feature engineering, and evaluation in supervised learning tasks.

Isalos version used: 2.0.6

Step 1: Import data from file

Right click on the input spreadsheet (left) and choose the option "Import from File". Then navigate through your files to load the one with the vehicle data.

The screenshot shows a web-based spreadsheet interface. The spreadsheet has 7 columns labeled Col1 through Col6. The first row is a header row with the following content: "User Header", "User Row ID", and then empty cells for Col2 through Col6. The rows are numbered 1 through 10 on the left. A right-click context menu is open over the cell at row 2, column 3 (Col3). The menu contains the following options: "Import from File", "Import from Spreadsheet", "Import from Multiple Spreadsheets", "Adjust Spreadsheet Precision", "Export Spreadsheet Data", and "Clear Spreadsheet". The "Import from File" option is highlighted.

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

The data will appear on the left spreadsheet.

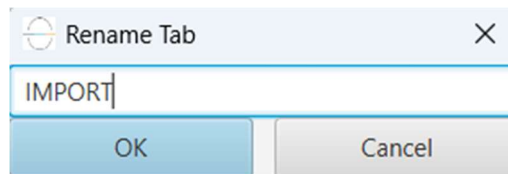
	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID	year	selling_price	km_driven	fuel	seller_type	transmission	owner
1	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual	First Owner
2	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual	First Owner
3	Hyundai Verna 1.6 SX	2012	600000	100000	Diesel	Individual	Manual	First Owner
4	Datsun RediGO T Option	2017	250000	46000	Petrol	Individual	Manual	First Owner
5	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
6	Maruti Alto LX BSIII	2007	140000	125000	Petrol	Individual	Manual	First Owner
7	Hyundai Xcent 1.2 Kappa S	2016	550000	25000	Petrol	Individual	Manual	First Owner
8	Tata Indigo Grand Petrol	2014	240000	60000	Petrol	Individual	Manual	Second Owner
9	Hyundai Creta 1.6 VTVT S	2015	850000	25000	Petrol	Individual	Manual	First Owner
10	Maruti Celerio Green VXI	2017	365000	78000	CNG	Individual	Manual	First Owner
11	Chevrolet Sail 1.2 Base	2015	260000	35000	Petrol	Individual	Manual	First Owner
12	Tata Indigo Grand Petrol	2014	250000	100000	Petrol	Individual	Manual	First Owner
13	Toyota Corolla Altis 1.8 VL CVT	2018	1650000	25000	Petrol	Dealer	Automatic	First Owner
14	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual	First Owner
15	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual	First Owner

Step 2: Manipulate data

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

The screenshot displays the 'Select Column(s)' dialog box in the Isalos Analytics Platform. The dialog has two main sections: 'Excluded Columns' and 'Included Columns'. The 'Included Columns' section lists all columns from the dataset: Col2 -- year, Col3 -- selling_price, Col4 -- km_driven, Col5 -- fuel, Col6 -- seller_type, Col7 -- transmission, and Col8 -- owner. The 'Excluded Columns' section is currently empty. Navigation buttons (>>, >, <, <<) are located between the two sections. At the bottom of the dialog are 'Execute' and 'Cancel' buttons. To the left of the dialog, the 'Data Manipulation' menu is open, showing options like 'Remove Column(s)', 'Select Column(s)', 'Matrix Transpose', 'Wide to Long Format', 'Sort by Column', and 'Fill Missing Column(s) Values'.

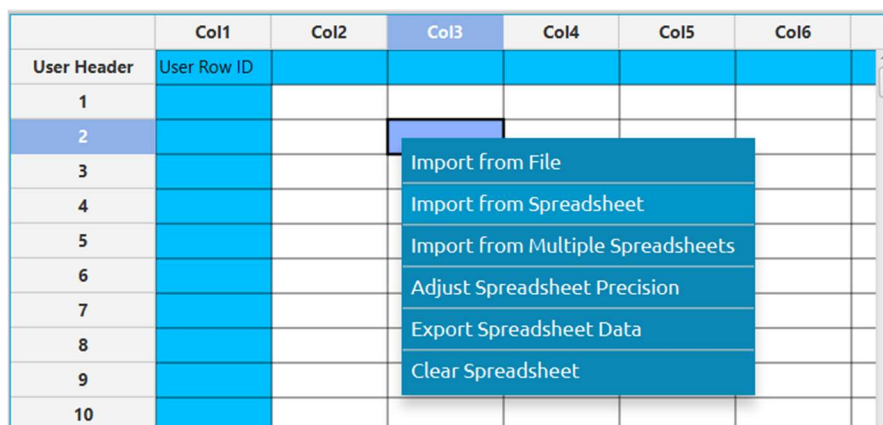
All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



Step 3: Split data

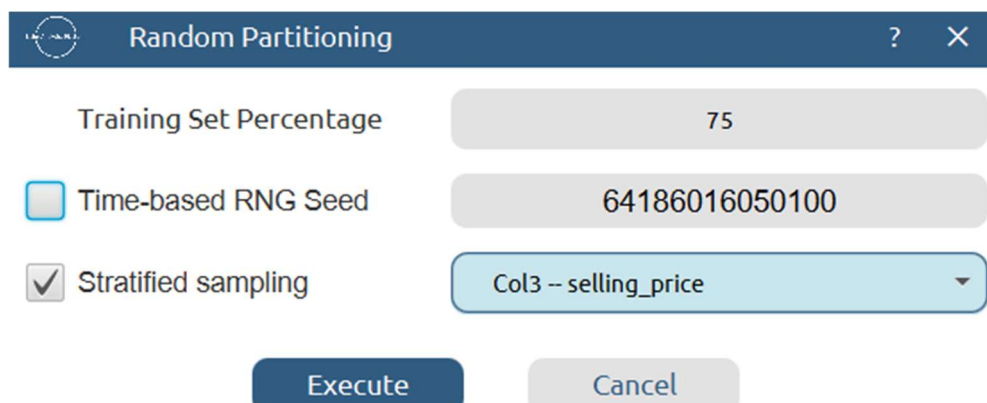
Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN_TEST_SPLIT” which we will use for splitting the train and test set.

Import data into the input spreadsheet of the “TRAIN_TEST_SPLIT” tab from the output of the “IMPORT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

A screenshot of a spreadsheet interface. The spreadsheet has columns labeled Col1 through Col6 and rows labeled 1 through 10. A context menu is open over the spreadsheet, showing options: "Import from File", "Import from Spreadsheet", "Import from Multiple Spreadsheets", "Adjust Spreadsheet Precision", "Export Spreadsheet Data", and "Clear Spreadsheet".

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

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:

A dialog box titled "Random Partitioning" with a help icon (?) and a close button (X) in the top right corner. It contains three settings: "Training Set Percentage" set to 75, "Time-based RNG Seed" set to 64186016050100, and "Stratified sampling" checked with a dropdown menu showing "Col3 -- selling_price". At the bottom are "Execute" and "Cancel" buttons.

Random Partitioning		?	X
Training Set Percentage	75		
<input type="checkbox"/> Time-based RNG Seed	64186016050100		
<input checked="" type="checkbox"/> Stratified sampling	Col3 -- selling_price		
Execute		Cancel	

The results will be two separate spreadsheets, “TRAIN_TEST_SPLIT: Training Set” and “TRAIN_TEST_SPLIT: Test Set”, which will be available to import into the next tabs.

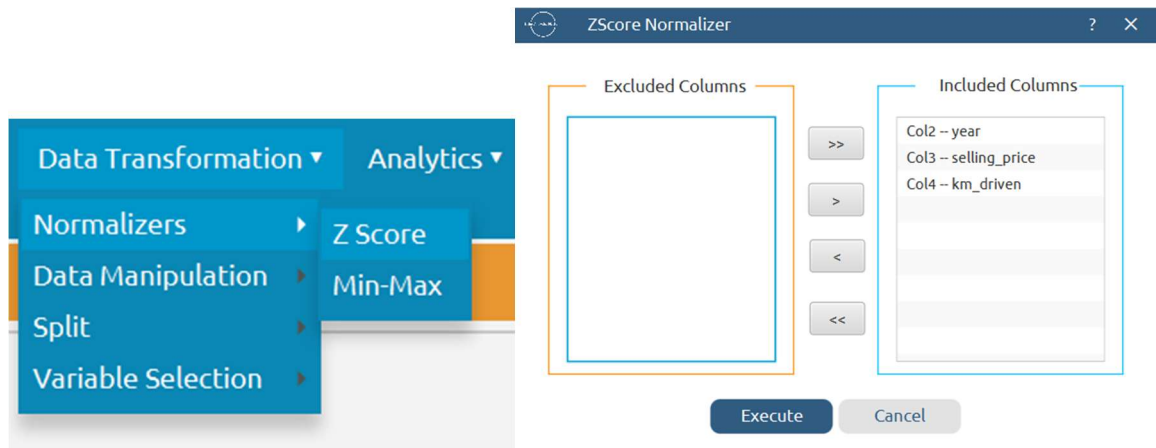
Step 4: Normalize the training set

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

Import into the input spreadsheet of the “NORMALIZE_TRAIN_SET” tab the train set from the output of the “TRAIN_TEST_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN_TEST_SPLIT: Training Set”.

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID	year	selling_price	km_driven	fuel	seller_type	transmission	owner
1	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual	First Owner
2	Datsun RediGO T Option	2017	250000	46000	Petrol	Individual	Manual	First Owner
3	Maruti Alto LX BSIII	2007	140000	125000	Petrol	Individual	Manual	First Owner
4	Hyundai Xcent 1.2 Kappa S	2016	550000	25000	Petrol	Individual	Manual	First Owner
5	Hyundai Creta 1.6 VTVT S	2015	850000	25000	Petrol	Individual	Manual	First Owner
6	Maruti Celerio Green VXI	2017	365000	78000	CNG	Individual	Manual	First Owner
7	Chevrolet Sail 1.2 Base	2015	260000	35000	Petrol	Individual	Manual	First Owner
8	Toyota Corolla Altis 1.8 VL CVT	2018	1650000	25000	Petrol	Dealer	Automatic	First Owner
9	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual	First Owner
10	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual	First Owner
11	Hyundai Verna 1.6 SX	2012	600000	100000	Diesel	Individual	Manual	First Owner
12	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
13	Maruti Alto LX BSIII	2007	140000	125000	Petrol	Individual	Manual	First Owner
14	Tata Indigo Grand Petrol	2014	240000	60000	Petrol	Individual	Manual	Second Owner
15	Hyundai Creta 1.6 VTVT S	2015	850000	25000	Petrol	Individual	Manual	First Owner

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



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID	year	selling_price	km_driven	fuel	seller_type	transmission	owner
1	Maruti 800 AC	-1.4326698	-0.7621981	0.0836910	Petrol	Individual	Manual	First Owner
2	Datsun RediGO T Option	0.9274396	-0.4366123	-0.4220799	Petrol	Individual	Manual	First Owner
3	Maruti Alto LX BSIII	-1.4326698	-0.6251094	1.2427495	Petrol	Individual	Manual	First Owner
4	Hyundai Xcent 1.2 Kappa S	0.6914287	0.0774705	-0.8646295	Petrol	Individual	Manual	First Owner
5	Hyundai Creta 1.6 VTVT S	0.4554177	0.5915534	-0.8646295	Petrol	Individual	Manual	First Owner
6	Maruti Celerio Green VXI	0.9274396	-0.2395472	0.2522813	CNG	Individual	Manual	First Owner
7	Chevrolet Sail 1.2 Base	0.4554177	-0.4194762	-0.6538916	Petrol	Individual	Manual	First Owner
8	Toyota Corolla Altis 1.8 VL CVT	1.1634505	1.9624410	-0.8646295	Petrol	Dealer	Automatic	First Owner
9	Maruti 800 AC	-1.4326698	-0.7621981	0.0836910	Petrol	Individual	Manual	First Owner
10	Maruti Wagon R LXI Minor	-1.4326698	-0.6336774	-0.3377848	Petrol	Individual	Manual	First Owner
11	Hyundai Verna 1.6 SX	-0.2526151	0.1631510	0.7159047	Diesel	Individual	Manual	First Owner
12	Honda Amaze VX i-DTEC	0.2194068	-0.0938904	1.5799301	Diesel	Individual	Manual	Second Owner
13	Maruti Alto LX BSIII	-1.4326698	-0.6251094	1.2427495	Petrol	Individual	Manual	First Owner
14	Tata Indigo Grand Petrol	0.2194068	-0.4537484	-0.1270469	Petrol	Individual	Manual	Second Owner
15	Hyundai Creta 1.6 VTVT S	0.4554177	0.5915534	-0.8646295	Petrol	Individual	Manual	First Owner

Step 5: Normalize the test set

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

Import into the input spreadsheet of the “NORMALIZE_TEST_SET” tab the test set from the output of the “TRAIN_TEST_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN_TEST_SPLIT: Test Set”.

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID	year	selling_price	km_driven	fuel	seller_type	transmission	owner
1	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual	First Owner
2	Hyundai Verna 1.6 SX	2012	600000	100000	Diesel	Individual	Manual	First Owner
3	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
4	Tata Indigo Grand Petrol	2014	240000	60000	Petrol	Individual	Manual	Second Owner
5	Tata Indigo Grand Petrol	2014	250000	100000	Petrol	Individual	Manual	First Owner
6	Datsun RediGO T Option	2017	250000	46000	Petrol	Individual	Manual	First Owner
7	Hyundai Xcent 1.2 Kappa S	2016	550000	25000	Petrol	Individual	Manual	First Owner
8	Maruti Ciaz VXI Plus	2015	585000	24000	Petrol	Dealer	Manual	First Owner
9	Honda City V MT	2018	930000	14500	Petrol	Dealer	Manual	First Owner
10	Skoda Superb Ambition 2.0 TDI CR AT	2011	450000	130400	Diesel	Dealer	Automatic	Second Owner
11	Toyota Innova 2.5 G (Diesel) 7 Seater	2015	1300000	80000	Diesel	Individual	Manual	First Owner
12	Hyundai i10 Magna 1.1L	2014	229999	60000	Petrol	Individual	Manual	Fourth & Above Owner
13	Maruti Wagon R VXI BS IV with ABS	2014	290000	64000	Petrol	Individual	Manual	Second Owner
14	Maruti Wagon R VXI BS IV	2012	275000	60000	Petrol	Individual	Manual	Second Owner
15	Hyundai i20 Sportz 1.4 CRDi	2018	750000	29000	Diesel	Dealer	Manual	First Owner

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

Data Transformation ▾

Analytics ▾

Statistics ▾

Regression

Classification

Clustering

Anomaly Detection

Existing Model Utilization

Existing Model Execution
?
×

Model (from Tab:)NORMALIZE...

Type Z Score Normalizer Model

Description

Model In...

```
Header -> Datatype
year -> Double
selling_price -> Double
km_driven -> Double
```

☐ Transfer Column(s) to Output

Execute
Cancel

The results will appear on the output spreadsheet.

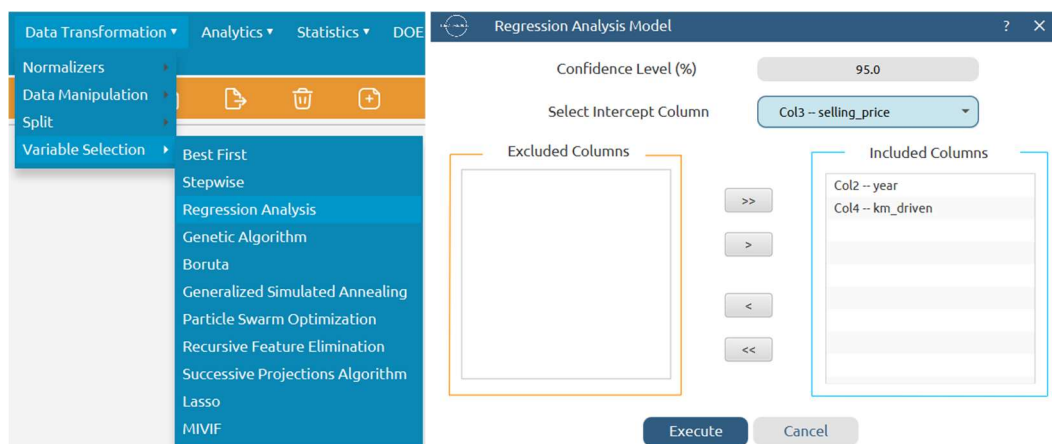
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID	year	selling_price	km_driven	fuel	seller_type	transmission	owner
1	Maruti Wagon R LXI Minor	-1.4326698	-0.6336774	-0.3377848	Petrol	Individual	Manual	First Owner
2	Hyundai Verna 1.6 SX	-0.2526151	0.1631510	0.7159047	Diesel	Individual	Manual	First Owner
3	Honda Amaze VX i-DTEC	0.2194068	-0.0938904	1.5799301	Diesel	Individual	Manual	Second Owner
4	Tata Indigo Grand Petrol	0.2194068	-0.4537484	-0.1270469	Petrol	Individual	Manual	Second Owner
5	Tata Indigo Grand Petrol	0.2194068	-0.4366123	0.7159047	Petrol	Individual	Manual	First Owner
6	Datsun RediGO T Option	0.9274396	-0.4366123	-0.4220799	Petrol	Individual	Manual	First Owner
7	Hyundai Xcent 1.2 Kappa S	0.6914287	0.0774705	-0.8646295	Petrol	Individual	Manual	First Owner
8	Maruti Ciaz VXI Plus	0.4554177	0.1374469	-0.8857033	Petrol	Dealer	Manual	First Owner
9	Honda City V MT	1.1634505	0.7286421	-1.0859043	Petrol	Dealer	Manual	First Owner
10	Skoda Superb Ambition 2.0 TDI CR AT	-0.4886260	-0.0938904	1.3565479	Diesel	Dealer	Automatic	Second Owner
11	Toyota Innova 2.5 G (Diesel) 7 Seater	0.4554177	1.3626777	0.2944289	Diesel	Individual	Manual	First Owner
12	Hyundai i10 Magna 1.1L	0.2194068	-0.4708862	-0.1270469	Petrol	Individual	Manual	Fourth & Above Owner
13	Maruti Wagon R VXI BS IV with ABS	0.2194068	-0.3680679	-0.0427517	Petrol	Individual	Manual	Second Owner
14	Maruti Wagon R VXI BS IV	-0.2526151	-0.3937721	-0.1270469	Petrol	Individual	Manual	Second Owner
15	Hyundai i20 Sportz 1.4 CRDi	1.1634505	0.4201924	-0.7803344	Diesel	Dealer	Manual	First Owner

Step 6: Regression analysis

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”.

Perform regression analysis by choosing *Data Transformation* → *Variable Selection* → *Regression Analysis* and selecting “selling_price” column as the intercept column.



The results will appear on the output spreadsheet.

	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID							
1		Regression Statistics						
2		Multiple R	0.4036433					
3		R Square	0.1629280					
4		Adjusted R Square	0.1624131					
5		Standard Error	0.9151977					
6		Observations	3255					
7								
8			Degrees of Freedom	Sum of Squares	Mean Square	F-statistic	Significance F	
9		Regression	2	530.1675553	265.0837776	316.4851225	0E-7	
10		Residual	3252	2723.8324447	0.8375869			
11		Total	3254	3254.0000000				
12			Coefficients	Standard Error	t-statistic	P-value	Lower 95.0%	Upper 95.0%
13		selling_price	0E-7	0.0160413	0E-7	1.0000000	-0.0314521	0.0314521
14		year	0.3939646	0.0176711	22.2943406	0E-7	0.3593171	0.4286122
15		km_driven	-0.0219195	0.0176711	-1.2404184	0.2149102	-0.0565670	0.0127280

Step 7: Train the model

Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN_MODEL(.fit)”.

Import data into the input spreadsheet of the “TRAIN_MODEL(.fit)” tab from the output of the “NORMALIZE_TRAIN_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the k-Nearest Neighbours (kNN) method to train and fit the model: *Analytics* → *Regression* → *k-Nearest Neighbors (kNN)*



The predictions will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)	Col8 (S)	Col9 (D)	Col10 (S)	Col11 (D)	Col12 (S)	Col13 (D)
User Header	User Row ID	selling_price	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5
1	Maruti 800 AC	-0.7621981	-0.7411973	Maruti Alto LXi	0.0	Maruti 800 AC	0.0	Maruti 800 AC	0.0	Maruti 800 AC	0.0	Maruti SX4 Zxi with Leather BSIII	0.0103893
2	Datsun RediGO T Option	-0.4366123	-0.1853406	Maruti Ciaz 1.4 Zeta	0.0	Datsun RediGO T Option	0.0	Datsun RediGO T Option	0.0	Honda BRV i-VTEC V MT	0.0030994	Maruti Alto 800 LXI	0.0049591
3	Maruti Alto LX BSIII	-0.6251094	-0.6292013	Maruti Alto LX BSIII	0.0	Maruti Alto LX BSIII	0.0	Maruti Wagon R LX BSIII	0.0185966	Maruti Wagon R LX Minor	0.0309944	Maruti Alto LX	0.0362483
4	Hyundai Xcent 1.2 Kappa S	0.0774705	-0.1727165	Renault KWID RXT Optional	0.0	Maruti Alto K10 VXI	0.0	Hyundai Grand i10 Sportz	0.0	Hyundai i20 Magna 1.2	0.0	Hyundai Xcent 1.2 Kappa S	0.0
5	Hyundai Creta 1.6 VTVT S	0.5915534	0.0466256	Maruti Wagon R VXI BS IV	0.0	Datsun GO T Option BSIV	0.0	Hyundai Verna 1.4 VTVT	0.0	Hyundai Creta 1.6 VTVT S	0.0	Hyundai Creta 1.6 VTVT S	0.0
6	Maruti Celerio Green VXI	-0.2395472	-0.2405449	Maruti Celerio Green VXI	0.0	Maruti Celerio Green VXI	0.0	Maruti Alto K10 LXI CNG Optional	0.0498051	Maruti Eeco 5 STR With AC Plus HTR CNG	0.0543079	Maruti Alto 800 CNG LXI Optional	0.0883933
7	Chevrolet Sail 1.2 Base	-0.4194762	-0.3913026	Maruti Celerio VXI	0.0	Chevrolet Sail 1.2 Base	0.0	Chevrolet Sail 1.2 Base	0.0	Maruti Swift VXI Optional	0.0061989	Maruti Swift 1.2 DLX	0.0061989
8	Toyota Corolla Altis 1.8 VL CVT	1.9624410	1.4568694	Toyota Corolla Altis 1.8 VL CVT	0.0	Toyota Corolla Altis 1.8 VL CVT	0.0	Honda Amaze V CVT Petrol BSIV	0.0012398	Honda Amaze V CVT Petrol BSIV	0.0012398	Honda City i-VTEC CVT ZX	0.0037193
9	Maruti 800 AC	-0.7621981	-0.7411973	Maruti Alto LXi	0.0	Maruti 800 AC	0.0	Maruti 800 AC	0.0	Maruti 800 AC	0.0	Maruti SX4 Zxi with Leather BSIII	0.0103893
10	Maruti Wagon R LXI Minor	-0.6336774	-0.6755335	Maruti Wagon R LXI Minor	0.0	Maruti 800 AC BSIII	0.0024795	Maruti Esteem Lxi - BSIII	0.0049591	Maruti 800 AC	0.0247955	Maruti 800 AC	0.0247955
11	Hyundai Verna 1.6 SX	0.1631510	-0.0236153	Toyota Innova 2.5 GX (Diesel) 8 Seater	0.0	Renault Duster 110PS Diesel RxZ	0.0	Hyundai Verna 1.6 SX	0.0	Volkswagen Vento Diesel Comfortline	0.0037193	Volkswagen Vento Diesel Comfortline	0.0037193
12	Honda Amaze VX i-DTEC	-0.0938904	-0.0023863	Honda Amaze VX i-DTEC	0.0	Toyota Innova 2.5 G (Diesel) 7 Seater	0.0049591	Mahindra Scorpio 1.99 S10	0.0074386	Toyota Etios VD	0.0198364	Ford Figo Diesel Titanium	0.0198364
13	Maruti Alto LX BSIII	-0.6251094	-0.6292013	Maruti Alto LX BSIII	0.0	Maruti Alto LX BSIII	0.0	Maruti Wagon R LX BSIII	0.0185966	Maruti Wagon R LX Minor	0.0309944	Maruti Alto LX	0.0362483
14	Tata Indigo Grand Petrol	-0.4537484	-0.4403452	Tata Indigo Grand Petrol	0.0	Maruti Swift Dzire LXI Option	0.0	Hyundai Xcent 1.2 Kappa Base	0.0	Tata Indigo Grand Petrol	0.0	Hyundai Verna 1.6 SX VTVT (O)	0.0123977
15	Hyundai Creta 1.6 VTVT S	0.5915534	0.0466256	Maruti Wagon R VXI BS IV	0.0	Datsun GO T Option BSIV	0.0	Hyundai Verna 1.4 VTVT	0.0	Hyundai Creta 1.6 VTVT S	0.0	Hyundai Creta 1.6 VTVT S	0.0

Step 8: Validate the model

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

Import data into the input spreadsheet of the “VALIDATE_MODEL(.predict)” tab from the output of the “NORMALIZE_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) TRAIN_MODEL(.fit)*. Choose the column “selling_price” to be transferred to the output spreadsheet.

The predictions will appear on the output spreadsheet.

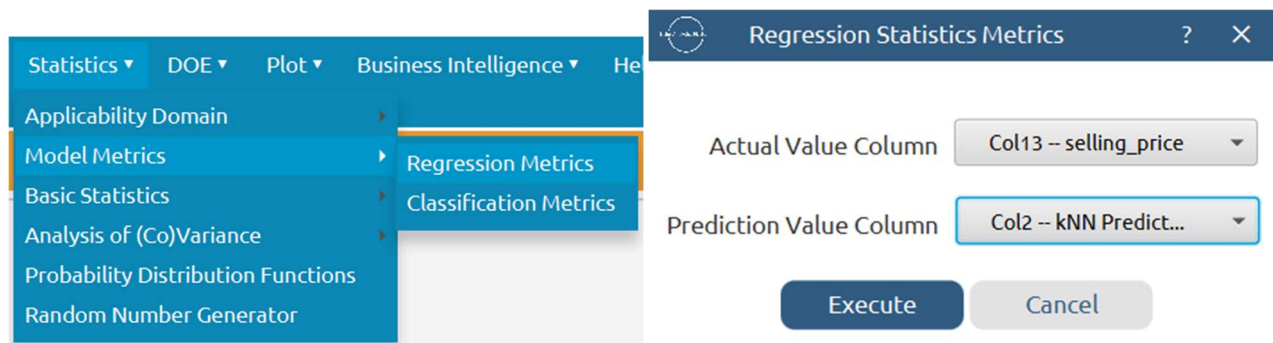
	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)	Col11 (S)	Col12 (D)	Col13 (D)
User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5	selling_price
1	Maruti Wagon R LXI Minor	-0.6755335	Maruti Wagon R LXI Minor	0.0	Maruti 800 AC BSIII	0.0024795	Maruti Esteem Lxi - BSIII	0.0049591	Maruti 800 AC	0.0247955	Maruti 800 AC	0.0247955	-0.6336774
2	Hyundai Verna 1.6 SX	-0.0236153	Toyota Innova 2.5 GX (Diesel) 8 Seater	0.0	Renault Duster 110PS Diesel RxZ	0.0	Hyundai Verna 1.6 SX	0.0	Volkswagen Vento Diesel Comfortline	0.0037193	Volkswagen Vento Diesel Comfortline	0.0037193	0.1631510
3	Honda Amaze VX i-DTEC	-0.0023863	Honda Amaze VX i-DTEC	0.0	Toyota Innova 2.5 G (Diesel) 7 Seater	0.0049591	Mahindra Scorpio 1.99 S10	0.0074386	Toyota Etios VD	0.0198364	Ford Figo Diesel Titanium	0.0198364	-0.0938904
4	Tata Indigo Grand Petrol	-0.4403452	Tata Indigo Grand Petrol	0.0	Maruti Swift Dzire LXI Option	0.0	Hyundai Xcent 1.2 Kappa Base	0.0	Tata Indigo Grand Petrol	0.0	Hyundai Verna 1.6 SX VTVT (O)	0.0123977	-0.4537484
5	Tata Indigo Grand Petrol	-0.4382570	Tata Indigo Grand Petrol	0.0	Tata Indigo Grand Petrol	0.0	Maruti Alto 800 VXI	0.0123977	Hyundai Grand i10 Sportz	0.0247955	Hyundai EON Magna Plus	0.0247955	-0.4366123
6	Datsun RediGO T Option	-0.1853406	Maruti Ciaz 1.4 Zeta	0.0	Datsun RediGO T Option	0.0	Datsun RediGO T Option	0.0	Honda BRV i-VTEC V MT	0.0030994	Maruti Alto 800 LXI	0.0049591	-0.4366123
7	Hyundai Xcent 1.2 Kappa S	-0.1727165	Renault KWID RXT Optional	0.0	Maruti Alto K10 VXI	0.0	Hyundai Grand i10 Sportz	0.0	Hyundai i20 Magna 1.2	0.0	Hyundai Xcent 1.2 Kappa S	0.0	0.0774705
8	Maruti Ciaz VXI Plus	-0.1140918	Toyota Etios Cross 1.2L G	0.0001996	Maruti Wagon R VXI BS IV	0.0004959	Maruti Swift 1.3 VXI ABS	0.0007439	Maruti Swift 1.3 VXI ABS	0.0007439	Maruti Swift 1.3 VXI ABS	0.0007439	0.1374469
9	Honda City V MT	0.4267676	Honda City i-VTEC VX	0.0004029	Maruti Wagon R VXI BS IV	0.0006199	Ford Aspire Titanium Plus BSIV	0.0006199	Honda Amaze VX i-VTEC	0.0012398	Maruti Alto 800 LXI	0.0025837	0.7286421
10	Skoda Superb Ambition 2.0 TDI CR AT	0.9809034	Hyundai Verna SX CRDi AT	0.0807322	BMW 3 Series 320d Sport Line	0.0984084	Skoda Octavia Elegance 2.0 TDI AT	0.1236383	Mercedes-Benz New C-Class C 220 CDI Avantgarde	0.1249740	Audi A8 L 3.0 TDI quattro	0.1256703	-0.0938904
11	Toyota Innova 2.5 G (Diesel) 7 Seater	0.1199937	Maruti Swift Dzire VDI	0.0	Maruti Swift Dzire ZDI	0.0	Maruti Ciaz ZDI SHVS	0.0037193	Maruti Swift VDI BSIV	0.0064220	Hyundai i20 Asta 1.4 CRDi	0.0066784	1.3626777
12	Hyundai i10 Magna 1.1L	-0.5609815	Maruti Wagon R LX Minor	0.0378050	Tata Nano STD	0.0561995	Maruti Alto K10 LXI	0.0825158	Maruti Omni E MPI STD BS IV	0.1076578	Hyundai i20 1.2 Sportz	0.1168939	-0.4708862
13	Maruti Wagon R VXI BS IV with ABS	-0.4589276	Tata Indigo Grand Petrol	0.0049591	Maruti Swift Dzire LXI Option	0.0049591	Hyundai Xcent 1.2 Kappa Base	0.0049591	Tata Indigo Grand Petrol	0.0049591	Maruti 800 AC	0.0136375	-0.3680679
14	Maruti Wagon R VXI BS IV	-0.5158247	Hyundai i10 Magna	0.0	Maruti Alto LXI	0.0	Hyundai EON Sportz	0.0052492	Hyundai Santro Xing GL Plus	0.0123977	Honda City S	0.0123977	-0.3937721
15	Hyundai i20 Sportz 1.4 CRDi	0.6683698	Ford EcoSport 1.5 TDCi Titanium BSIV	0.0017109	Hyundai Verna CRDi 1.6 SX Option	0.0031862	Hyundai Verna CRDi 1.6 SX Option	0.0031862	Maruti Ignis 1.3 Delta	0.0032854	Maruti Ignis 1.3 Delta	0.0032854	0.4201924

Step 9: Statistics calculation

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

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

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Regression Metrics*



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared
1		0.2349704	0.4847374	0.2610929	0.7516431

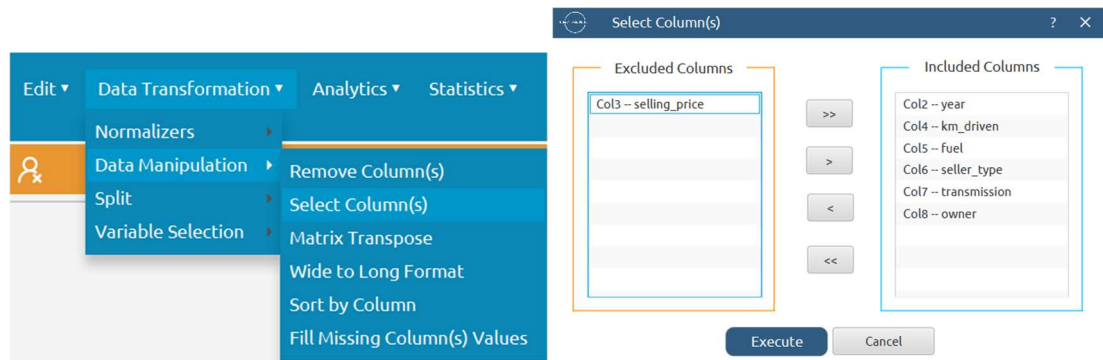
Step 10: Reliability check for each record of the test set

Step 11.a: Create the domain

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

Import data into the input spreadsheet of the “EXCLUDE_PRICE” tab from the output of the “NORMALIZE_TRAIN_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “selling_price”: *Data Transformation → Data Manipulation → Select Column(s)*

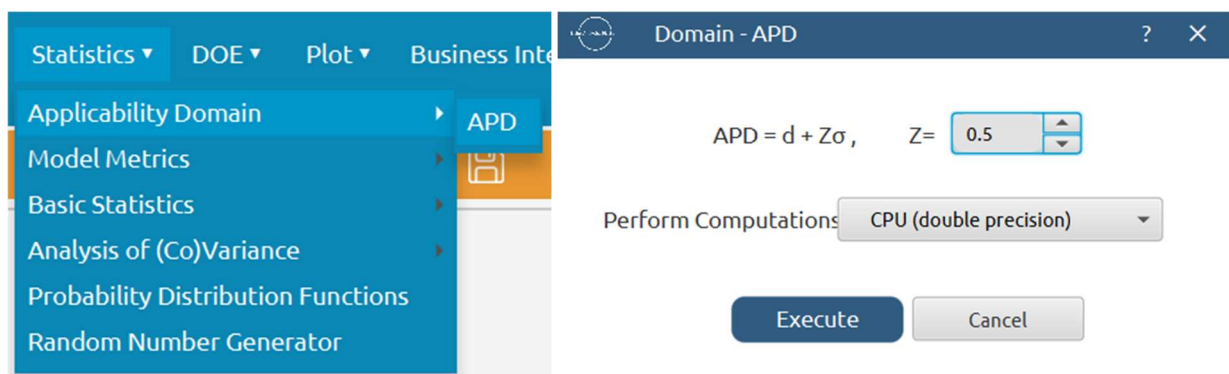


The results will appear on the output spreadsheet.

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

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

Create the domain: Statistics → Applicability Domain → APD



The results will appear on the output spreadsheet.

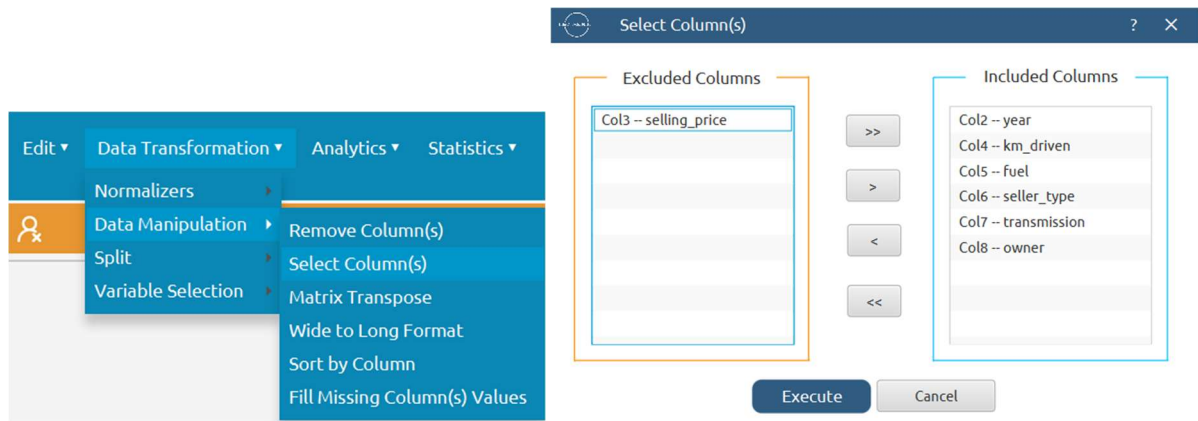
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1	Maruti 800 AC	0.0	1.1381761	reliable
2	Datsun RediGO T Option	0.0	1.1381761	reliable
3	Maruti Alto LX BSIII	0.0	1.1381761	reliable
4	Hyundai Xcent 1.2 Kappa S	0.0	1.1381761	reliable
5	Hyundai Creta 1.6 VTVT S	0.0	1.1381761	reliable
6	Maruti Celerio Green VXI	0.0	1.1381761	reliable
7	Chevrolet Sail 1.2 Base	0.0	1.1381761	reliable
8	Toyota Corolla Altis 1.8 VL CVT	0.0	1.1381761	reliable
9	Maruti 800 AC	0.0	1.1381761	reliable
10	Maruti Wagon R LXI Minor	0.0	1.1381761	reliable
11	Hyundai Verna 1.6 SX	0.0	1.1381761	reliable
12	Honda Amaze VX i-DTEC	0.0	1.1381761	reliable
13	Maruti Alto LX BSIII	0.0	1.1381761	reliable
14	Tata Indigo Grand Petrol	0.0	1.1381761	reliable
15	Hyundai Creta 1.6 VTVT S	0.0	1.1381761	reliable

Step 11.b: Check the test set reliability

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

Import data into the input spreadsheet of the “EXCLUDE_PRICE_TEST_SET” tab from the output of the “NORMALIZE_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “selling_price”: *Data Transformation → Data Manipulation → Select Column(s)*

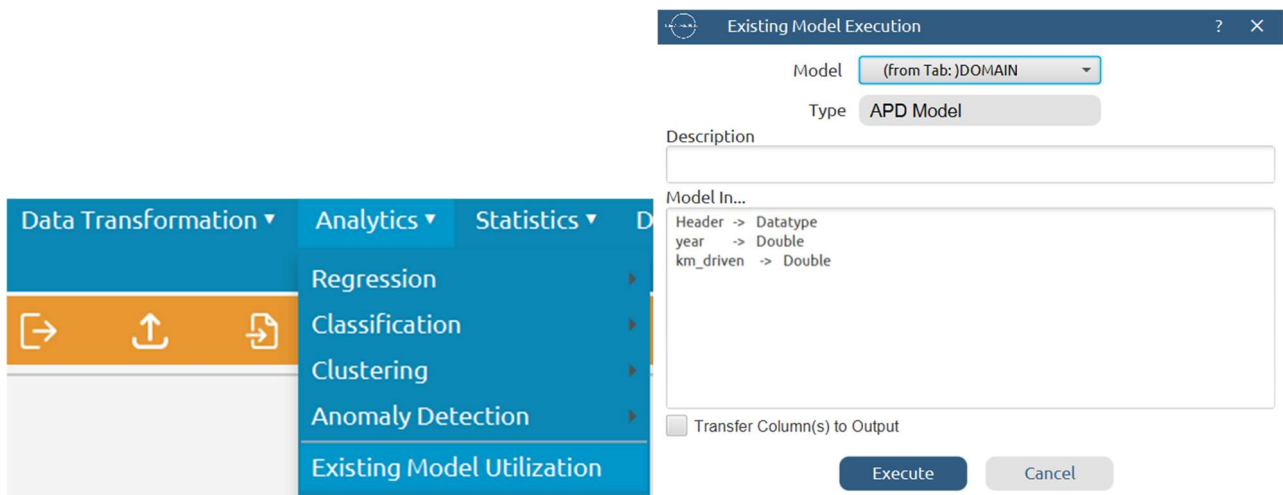


The results will appear on the output spreadsheet.

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

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

Check the Reliability: *Analytics → Existing Model Utilization → Model (from Tab:) DOMAIN*



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1	Maruti Wagon R LXI Minor	0.0	1.1381761	reliable
2	Hyundai Verna 1.6 SX	0.0	1.1381761	reliable
3	Honda Amaze VX i-DTEC	0.0	1.1381761	reliable
4	Tata Indigo Grand Petrol	0.0	1.1381761	reliable
5	Tata Indigo Grand Petrol	0.0	1.1381761	reliable
6	Datsun RediGO T Option	0.0	1.1381761	reliable
7	Hyundai Xcent 1.2 Kappa S	0.0	1.1381761	reliable
8	Maruti Ciaz VXI Plus	0.0	1.1381761	reliable
9	Honda City V MT	0.0038144	1.1381761	reliable
10	Skoda Superb Ambition 2.0 TDI CR AT	0.0084295	1.1381761	reliable
11	Toyota Innova 2.5 G (Diesel) 7 Seater	0.0	1.1381761	reliable
12	Hyundai i10 Magna 1.1L	0.0	1.1381761	reliable
13	Maruti Wagon R VXI BS IV with ABS	0.0	1.1381761	reliable
14	Maruti Wagon R VXI BS IV	0.0	1.1381761	reliable
15	Hyundai i20 Sportz 1.4 CRDi	0.0126443	1.1381761	reliable

Final Isalos Workflow

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

