Azure Microsoft Machine Learning Studio

Automobile price prediction project

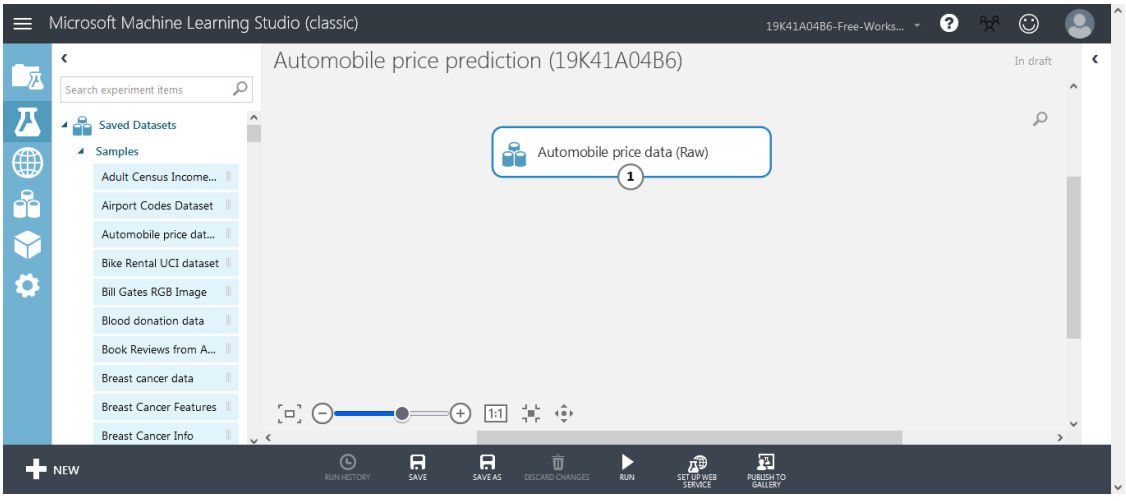
- In this project, we try to predict the automobile price through linear regression model in Azure ML studio. We build the model by drag and drop method of various inbuilt functions, and automobile price prediction data set which are already available in the studio.

Workflow:

1. Data import
2. Check the data for missing values
3. Preprocess the data (By excluding and cleaning missing values)
4. Split the data
5. Training the model
6. Testing the model
7. Evaluation of the model

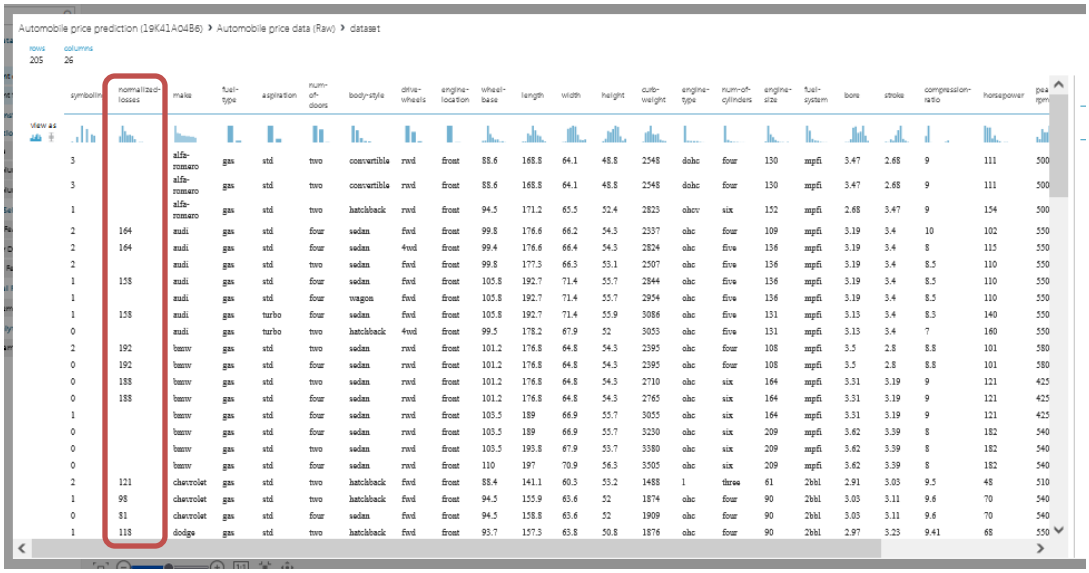
1. Data import

- We import the data from the pre-existing raw data set module of automobile price prediction in azure ML studio.



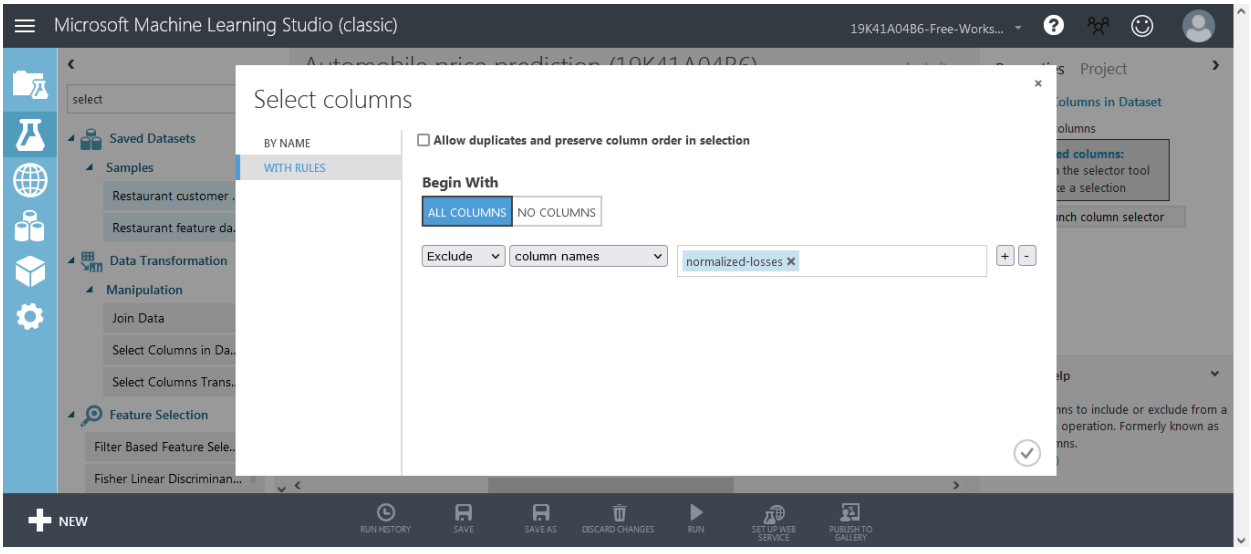
2. Check the data for missing values

- Go through the data set, and check for missing values in the columns.
- Here, we see normalized losses column is missing many values.

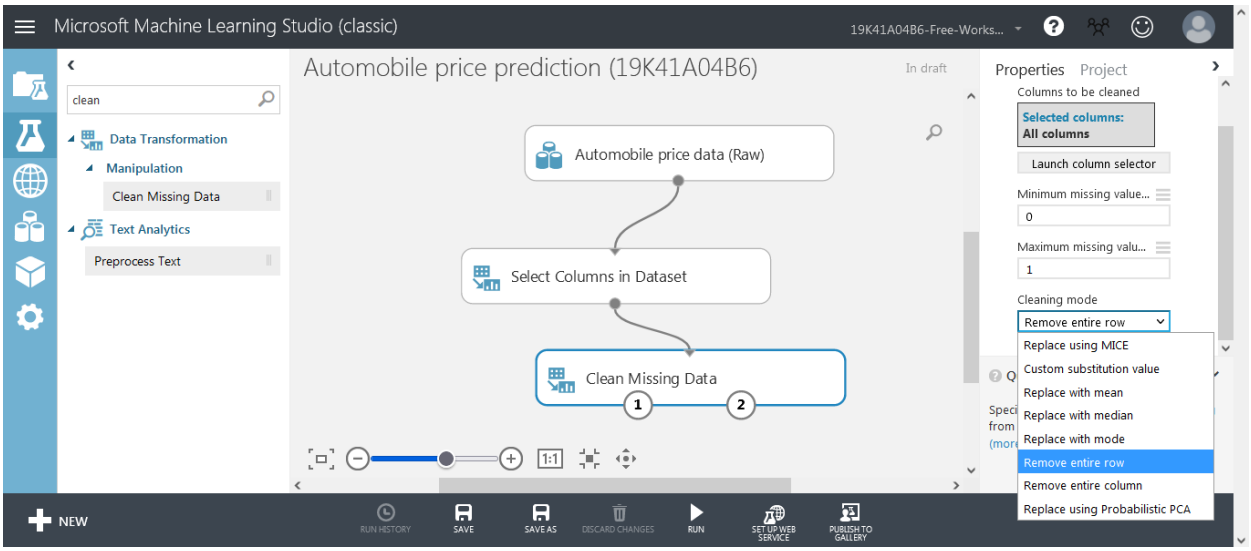


3. Preprocess the data (By excluding and cleaning missing values)

- There are many missing values in the normalized losses column, so we exclude it.



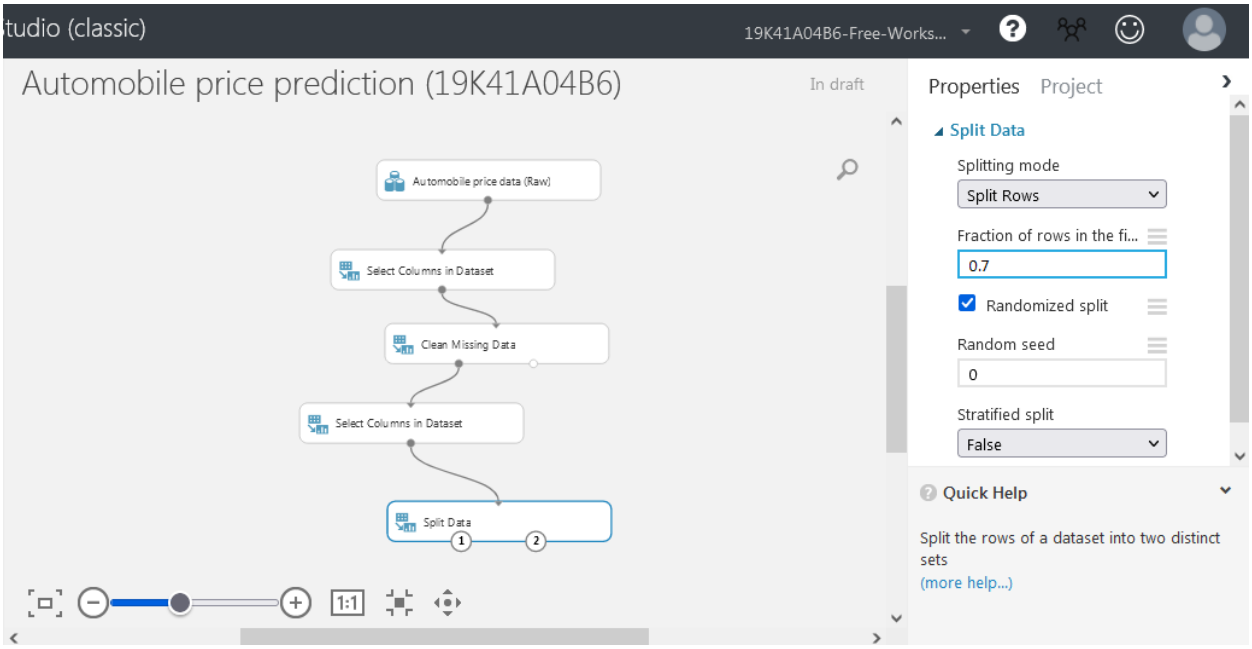
- Now, we process the data of remaining columns by deleting/cleaning the entire row wherever we see a missing value.



- After cleaning the data of all missing values, we select all the columns in data set for splitting process.

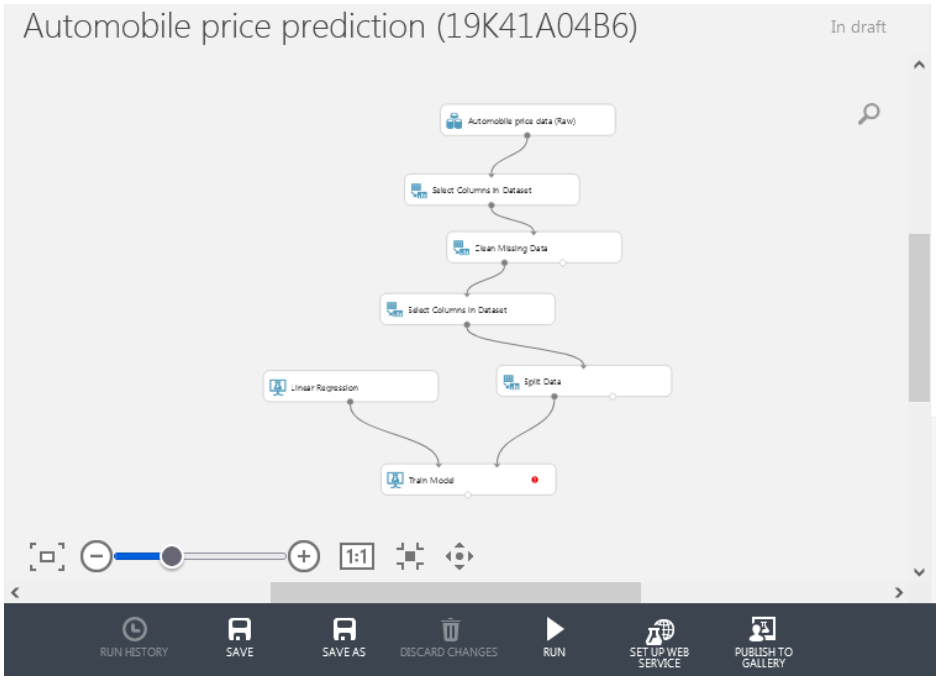
4. Split the data

- The processed data is now split into two parts, that is: Training and testing.
- This split is done with a fraction of 0.7; that is, 70% of data is used for training and 30% is used for testing the data.



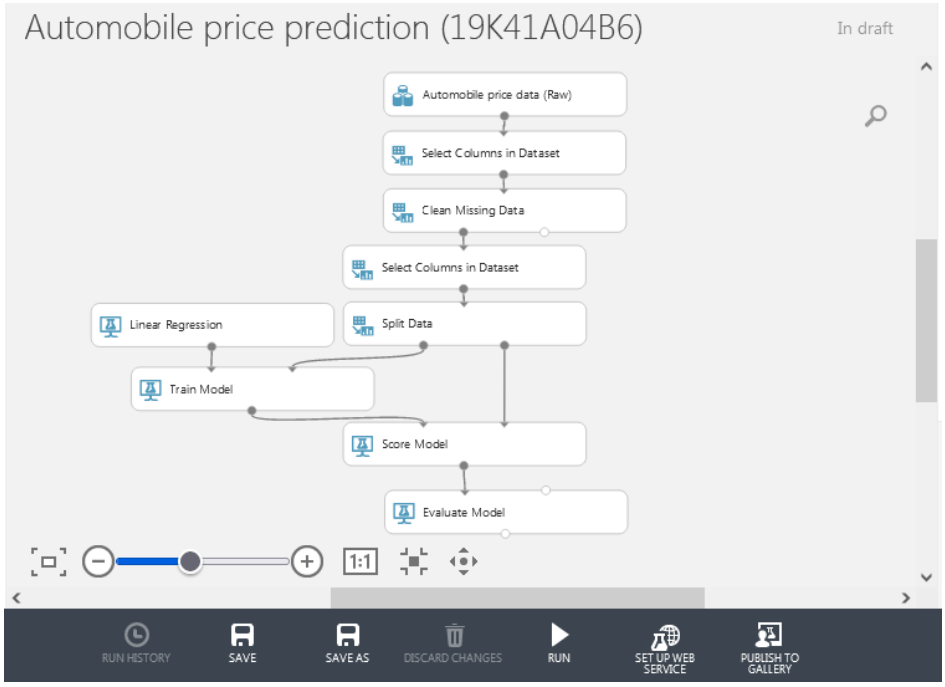
5. Training the model

- For training the model, we use linear regression analysis model.

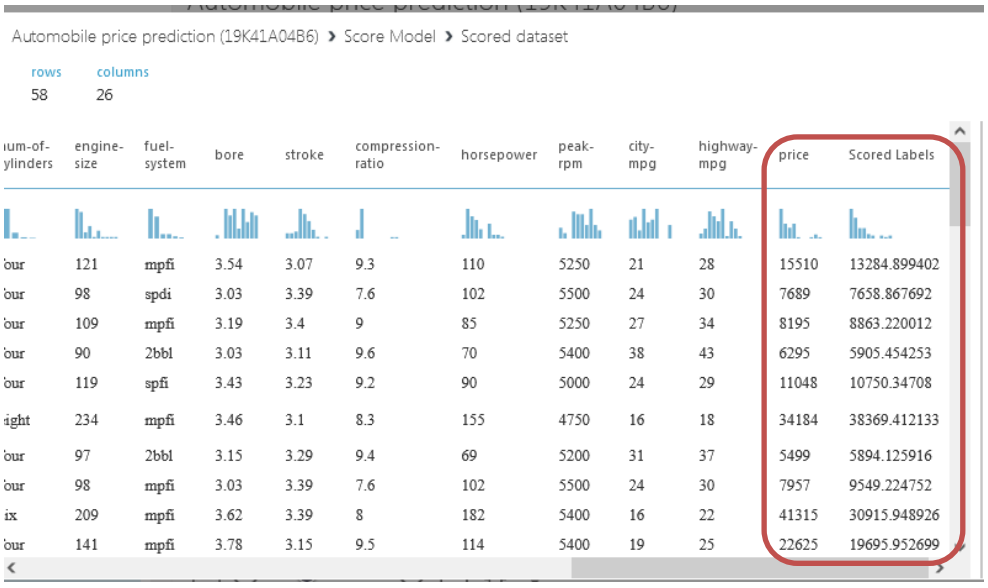


6. Testing the model

- Score model and evaluate model are added for testing our model and evaluating its performance



- Next, we run our model and see the predicted results of our model.
- Here, we can see the price and scored labels, our model's predictions were almost accurate.



Automobile price prediction (19K41A04B6) › Score Model › Scored dataset

rows 58 columns 26

ium-of-ylinders	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price	Scored Labels
our	121	mpfi	3.54	3.07	9.3	110	5250	21	28	15510	13284.899402
our	98	spdi	3.03	3.39	7.6	102	5500	24	30	7689	7658.867692
our	109	mpfi	3.19	3.4	9	85	5250	27	34	8195	8863.220012
our	90	2bbl	3.03	3.11	9.6	70	5400	38	43	6295	5905.454253
our	119	spfi	3.43	3.23	9.2	90	5000	24	29	11048	10750.34708
ight	234	mpfi	3.46	3.1	8.3	155	4750	16	18	34184	38369.412133
our	97	2bbl	3.15	3.29	9.4	69	5200	31	37	5499	5894.125916
our	98	mpfi	3.03	3.39	7.6	102	5500	24	30	7957	9549.224752
ix	209	mpfi	3.62	3.39	8	182	5400	16	22	41315	30915.948926
our	141	mpfi	3.78	3.15	9.5	114	5400	19	25	22625	19695.952699

7. Evaluation of the model

- The evaluation results of our model are as follows: