1. Drag to canvas Automobile price data

Output port - Visualize

Prepare the data - normalized-losses missing data

2. Select columns

Launch column selector

With Rules - All columns - Exclude - column names - normalized-losses

3. Drag to Canvas Clean missing data

Remove entire row

Define features - what do we believe impacts the price (make, body-style, wheel-base, engine-size, horsepower, peak-rpm, highway-mpg, price)

4. Select columns

Launch column selector

With Rules - No Columns - Include - column names

Done with feature engineering. Time to be a data scientist.

5. Split data

Fraction data in first output .75 - training

Run

6. Linear Regression

7. Train the model

Launch column selector - By name - Price

Run

8. Score Model

Visualize, see the prices

9. Evaluate

Run

Mean Absolute Error (MAE): The average of absolute errors (an error is the difference between the predicted value and the actual value)

Root Mean Squared Error (RMSE): The square root of the average of squared errors of predictions made on the test dataset.

Relative Absolute Error: The average of absolute errors relative to the absolute difference between actual values and the average of all actual values.

Relative Squared Error: The average of squared errors relative to the squared difference between the actual values and the average of all actual values.

Coefficient of Determination: Also known as the R squared value, this is a statistical metric indicating how well a model fits the data.

For each of the error statistics, smaller is better. A smaller value indicates that the predictions more closely match the actual values. For Coefficient of Determination, the closer its value is to one (1.0), the better the predictions