Azure ML Classic Studio

Predicting Automobile prices using Regression Model in Azure ML Classic Studio.

This model (Pipeline) trains a linear regressor to predict a car's price based on technical features such as make, model, horsepower, and size. Because you're trying to answer the question "How much?" this is called a regression problem. However, you can apply the same fundamental steps in this example to tackle any type of machine learning problem whether it be regression, classification, clustering, and so on.

Gallery Link:

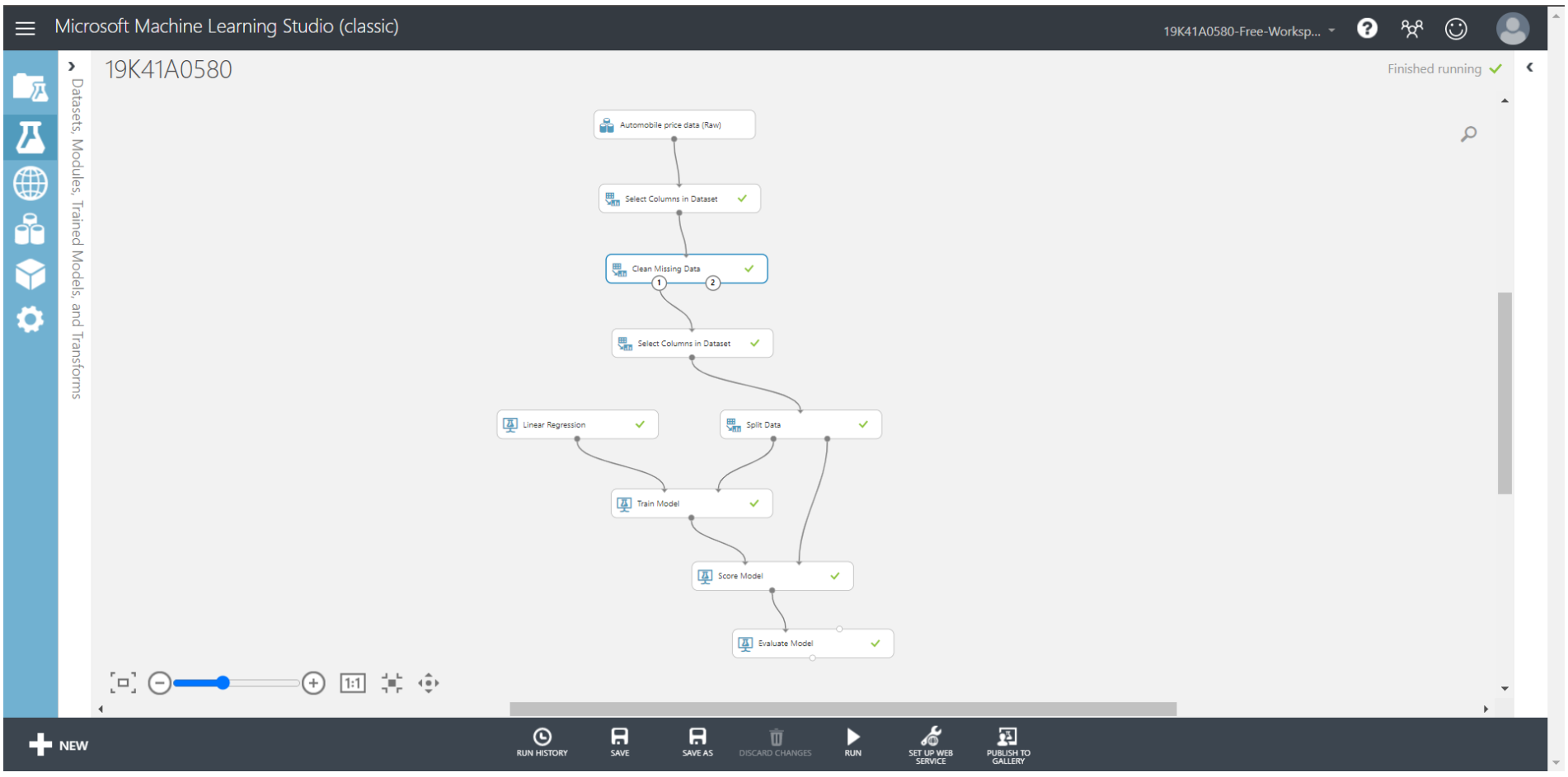
- https://gallery.cortanaintelligence.com/Experiment/19K41A0580

navigate to the link to see the Workflow and you can download the project as well.

Machine Learning Project Workflow

1. **Import Data**
2. **Explore Data (Missing values, outliers)**
3. **Preprocess data (Missing value imputation, outlier treatment, normalization)**
4. **Model Selection**
5. **Model Training**
6. **Model Testing**
7. **Model Deployment**

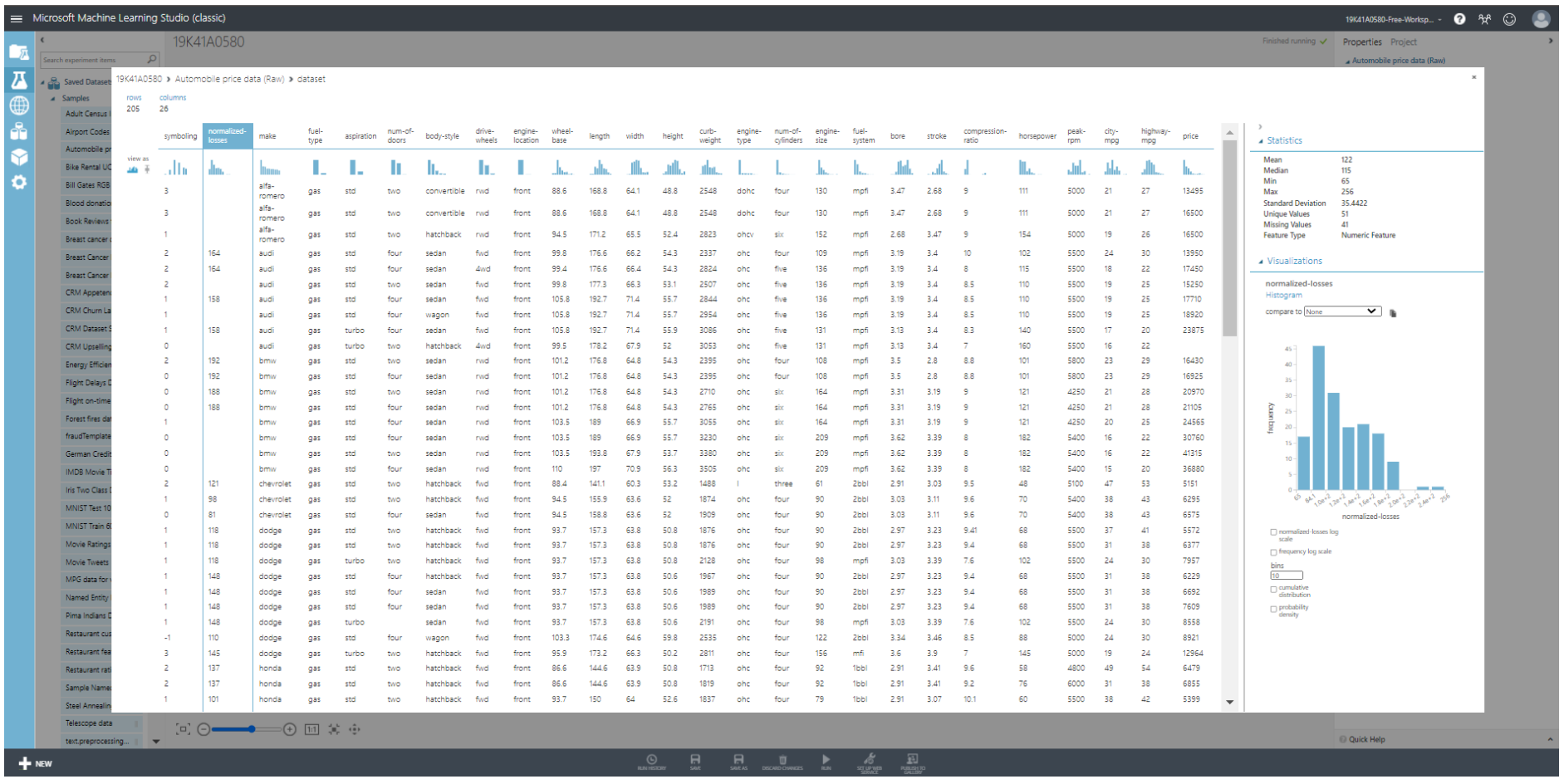
Workflow



Project Workflow

Import Data:

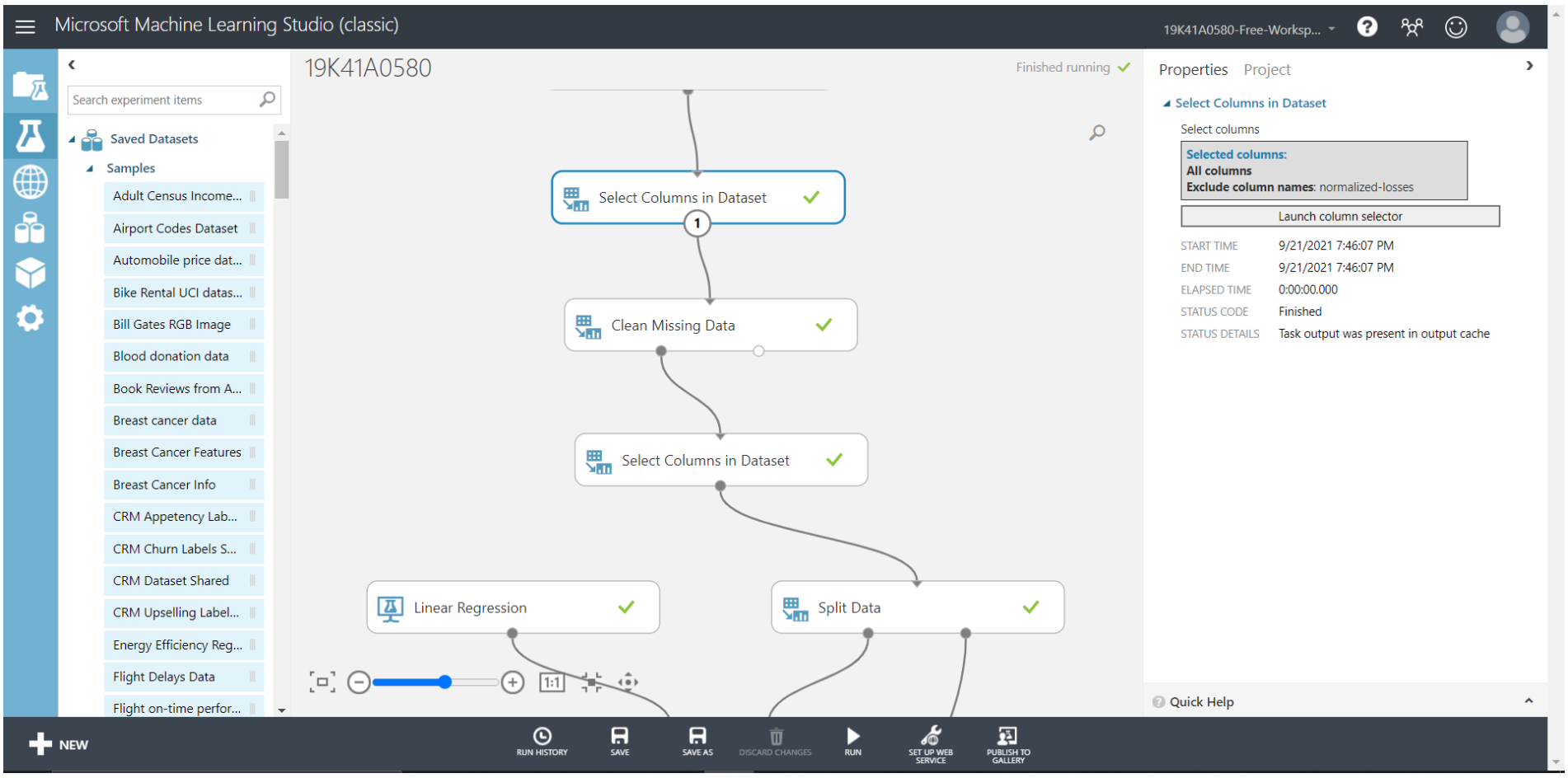
- importing the RAW dataset which is in CSV format.
- the dataset is pre-available in the Azure ML Classic Studio.



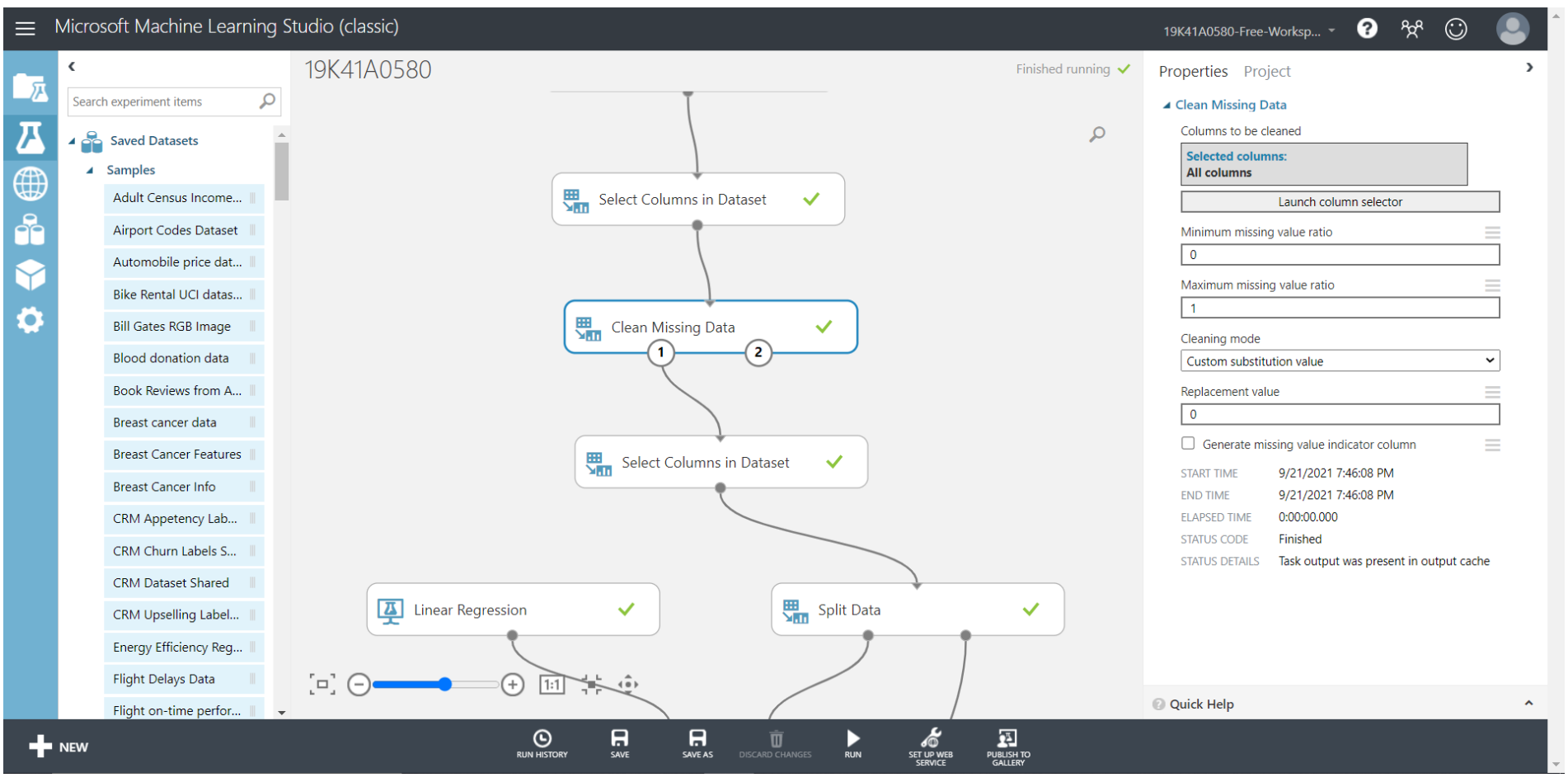
Automobile Price RAW dataset (CSV format)

Explore Data

- this basically includes data visualization to search for any missing values in the Dataset.
- if any missing values are found, then they needs to be cleaned.
- selecting the required columns and clean the data using the Clean Missing Value module (Just Drag n' Drop)



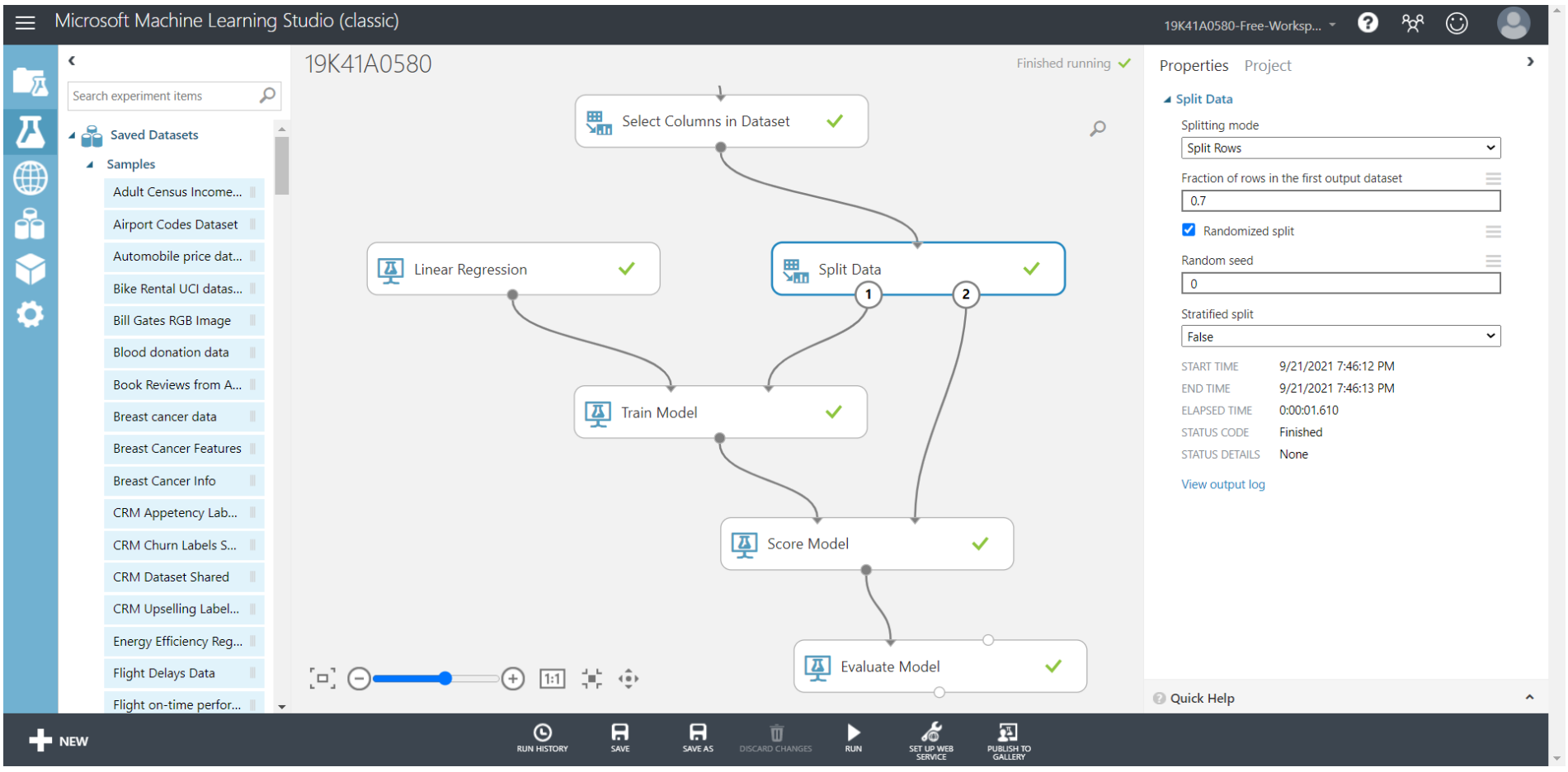
as the normalized loss has 41 missing values in the Dataset, those missing values are to be cleaned.



Data Cleaning

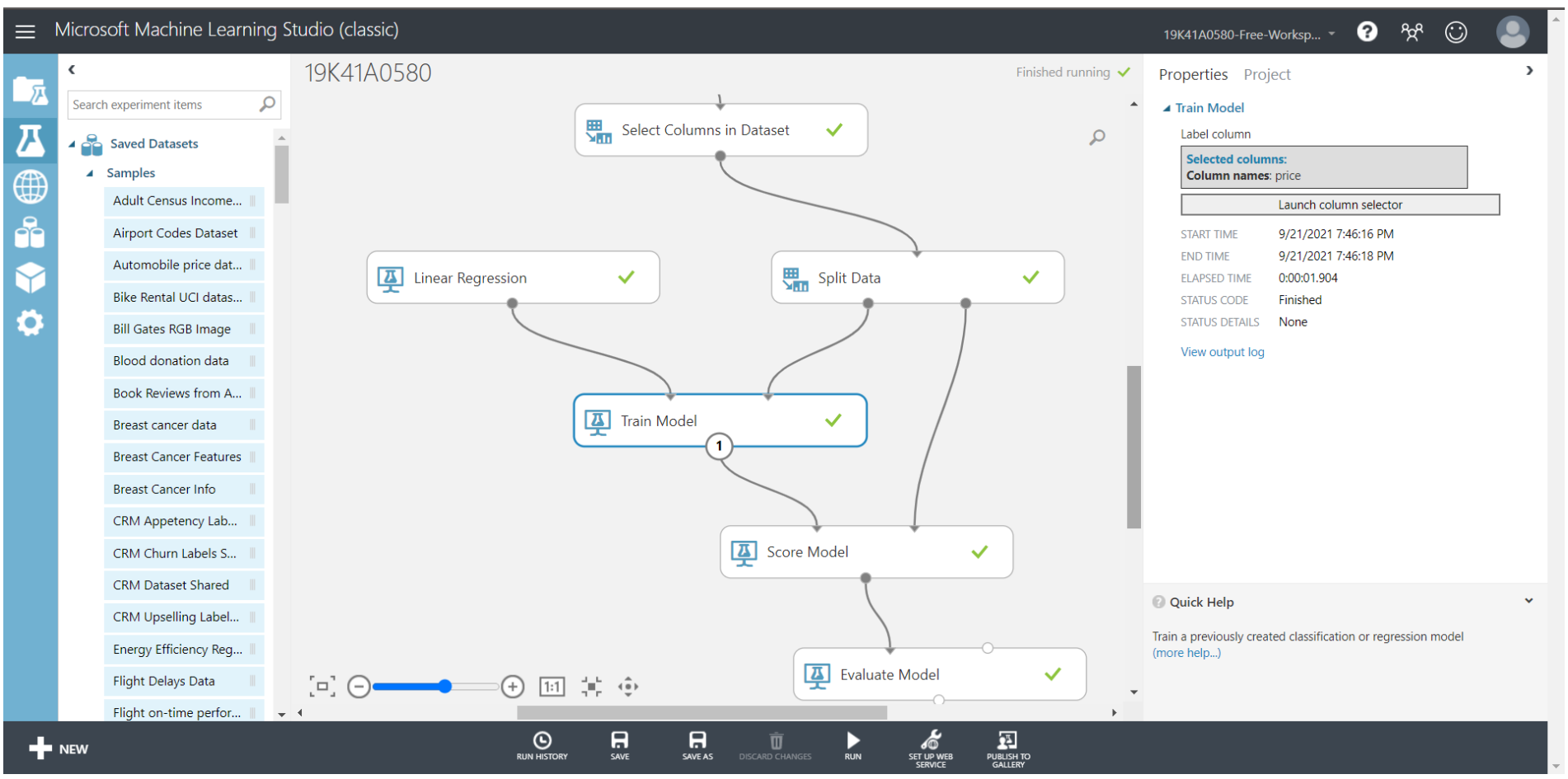
Split Data

- Use the Split Data module to randomly divide the input data so that the training dataset contains 70% of the original data and the testing dataset contains 30% of the original data.



Data Splitting

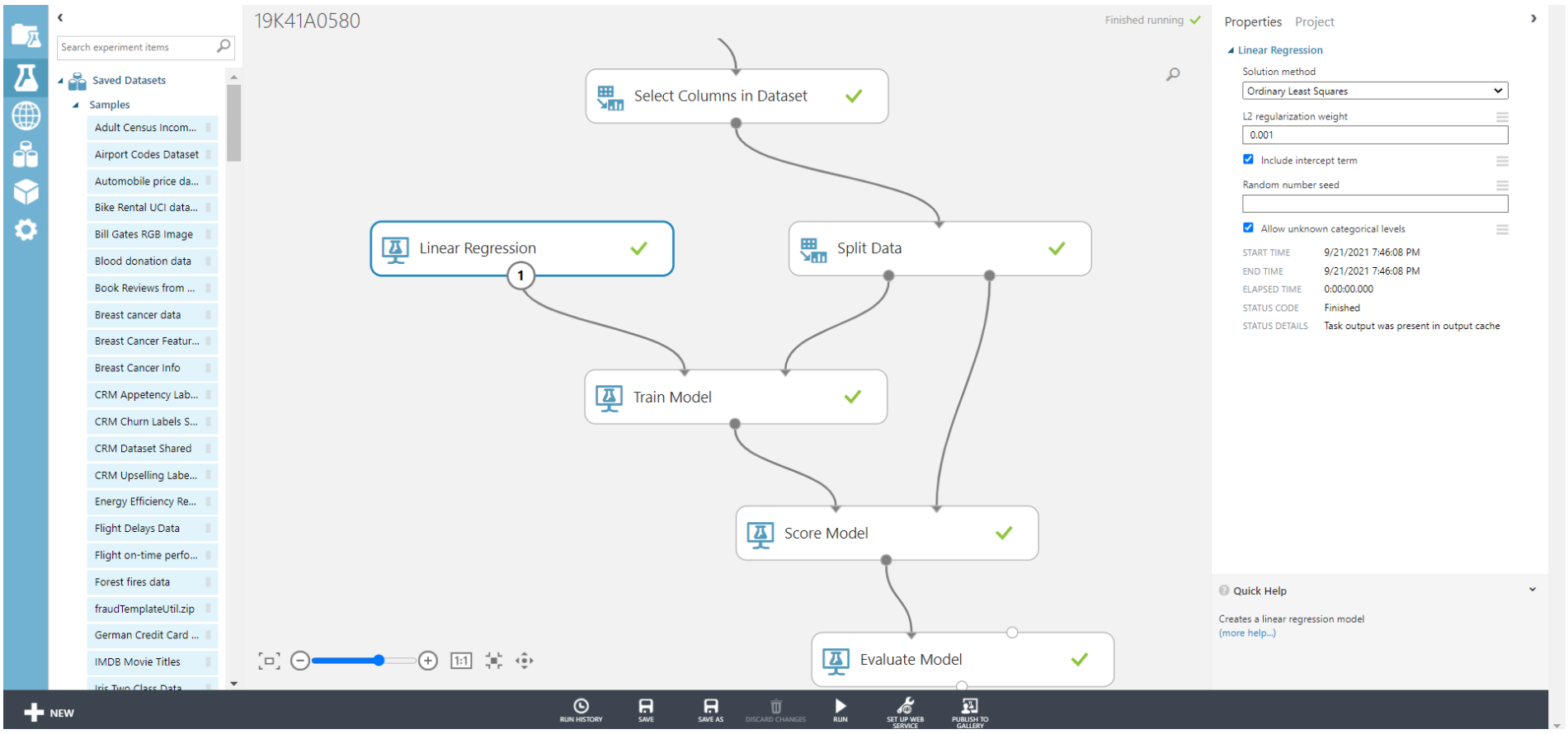
Model Training and Algorithm



Model Training

Using Linear regression to train the model

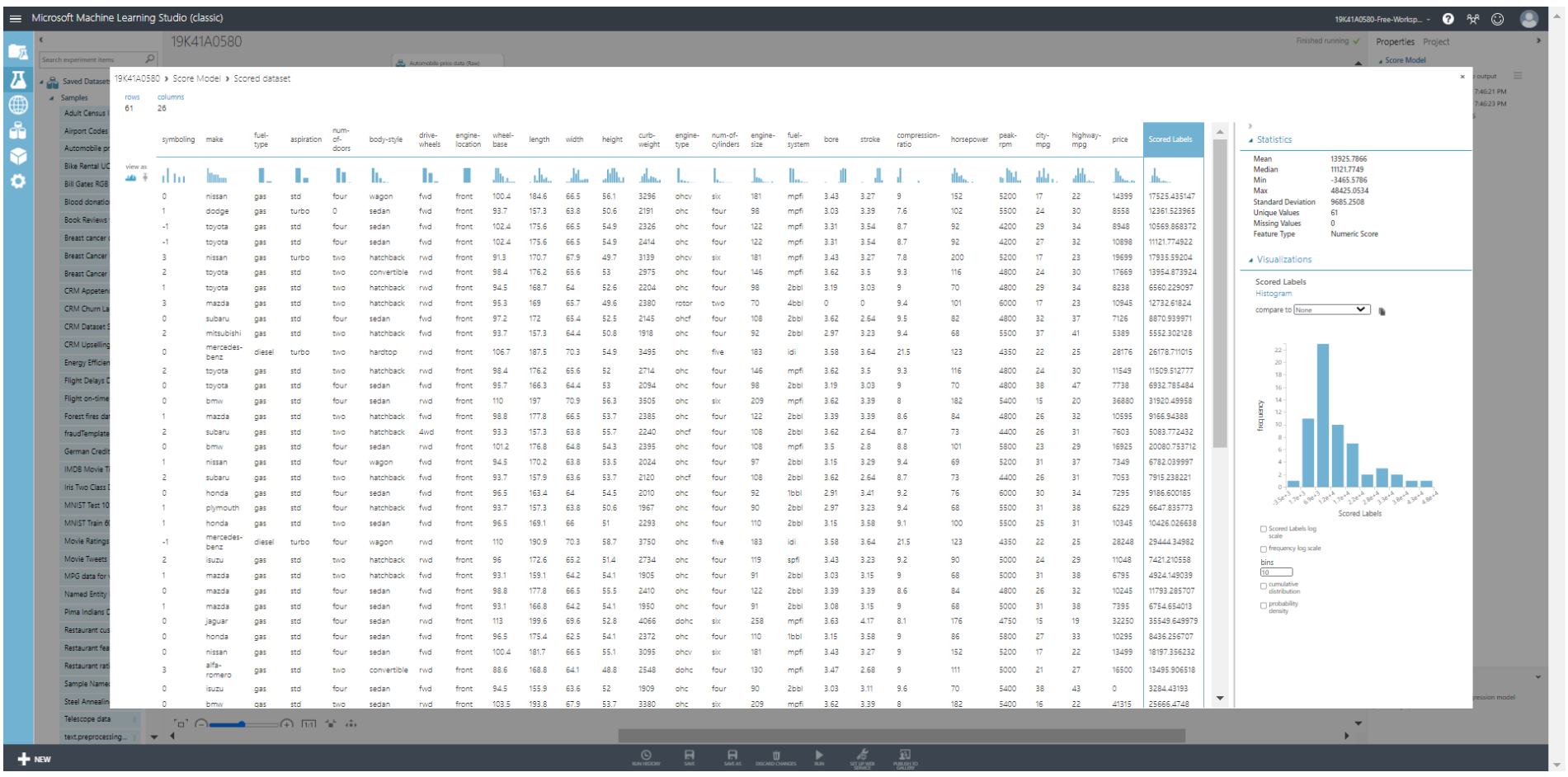
- Since the goal of this sample is to predict automobile prices, and because the label column (price) is continuous data, a regression model can be a good choice. We use Linear Regression for this pipeline.



Linear Regression

Score Model and Evaluate Model

- After the model is trained, we can use the Score Model and Evaluate Model modules to generate predicted results and evaluate the models.



Score Labels

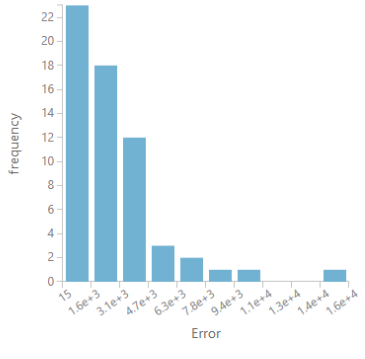
Evaluation Results

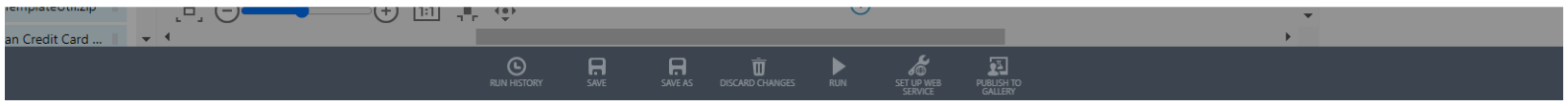
19K41A0580 › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	2688.238855
Root Mean Squared Error	3758.693629
Relative Absolute Error	0.36918
Relative Squared Error	0.158758
Coefficient of Determination	0.841242

Error Histogram





Model Evaluation Results