



## Azure ML Classic Studio

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

This model (Pipeline) trains a linear regression 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.

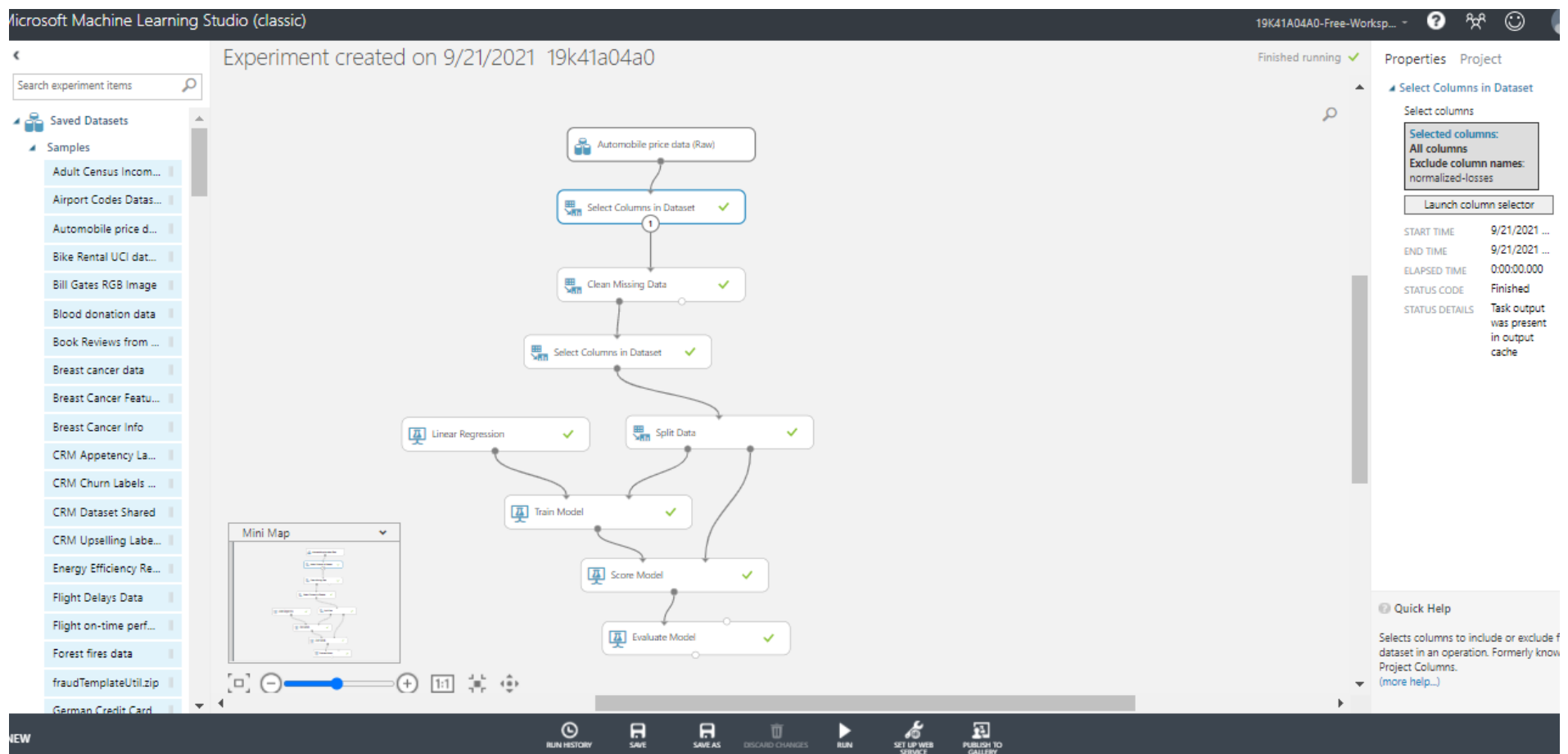
Automobile price prediction 19K41AO4AO

Automobile price prediction using pre-available dataset and training the model using Linear Regression. Tags: Linear Regression, Automobile,

### 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

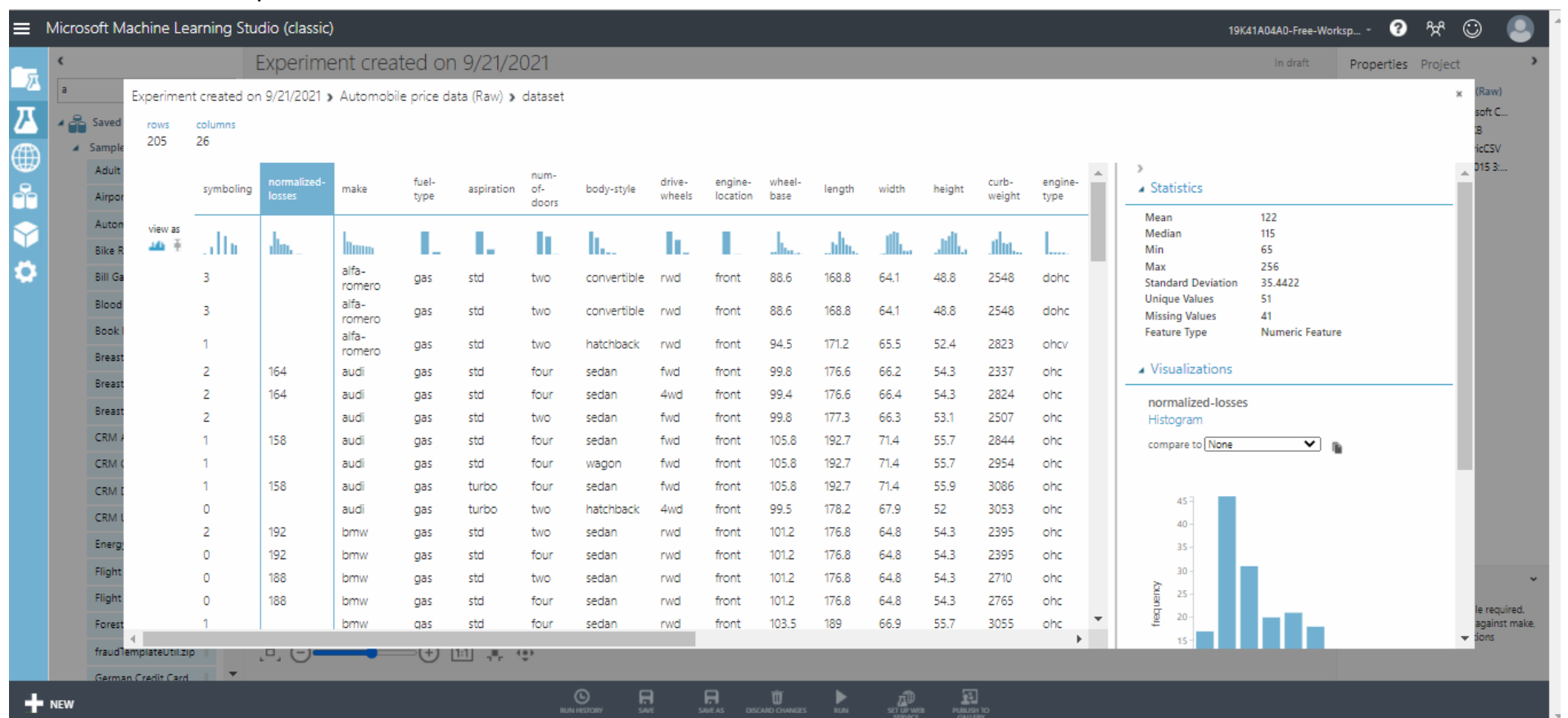
## Workflow



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 )

The screenshot displays the Microsoft Machine Learning Studio interface. The central workspace shows a workflow diagram with the following steps: 'Automobile price data (new)' (dataset), 'Select Columns in Dataset' (1), 'Clean Missing Data', 'Select Columns in Dataset' (2), 'Split Data', 'Linear Regression', 'Train Model', and 'Score Model'. All steps are marked with a green checkmark, indicating they have completed successfully. The 'Clean Missing Data' module is highlighted with a blue border. On the right side, the 'Properties' pane is open for the 'Select Columns in Dataset' module, showing 'Selected columns: All columns' and 'Exclude column names: normalized-losses'. The left sidebar lists various saved datasets, including 'Adult Census Income...', 'Airport Codes Dataset', 'Automobile price da...', 'Bike Rental UCI data...', 'Bill Gates RGB Image', 'Blood donation data', 'Book Reviews from ...', 'Breast cancer data', 'Breast Cancer Featur...', 'Breast Cancer Info', 'CRM Appetency Lab...', 'CRM Churn Labels S...', 'CRM Dataset Shared', 'CRM Upselling Labe...', 'Energy Efficiency Re...', 'Flight Delays Data', and 'Flight on-time perfo...'. The bottom status bar shows the system clock as 9:12 PM on 9/21/2021.

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

This screenshot shows the same workflow as the previous image, but with the 'Clean Missing Data' module selected. The 'Properties' pane on the right is now configured for this module. It shows 'Columns to be cleaned' as 'Selected columns: All columns'. The 'Minimum missing value' is set to 0, and the 'Maximum missing value' is set to 1. The 'Cleaning mode' is set to 'Custom substitution val'. The 'Replacement value' is set to 0. The 'Generate missing ...' checkbox is unchecked. The 'START TIME' is 9/21/2021, and the 'END TIME' is 9/21/2021. The 'ELAPSED TIME' is 0:00:00. The bottom status bar shows the system clock as 9:11 PM on 9/21/2021.

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.

The screenshot displays the Microsoft Machine Learning Studio (classic) interface. The main workspace shows a workflow diagram with the following modules: 'Select Columns in Dataset', 'Clean Missing Data', 'Select Columns in Dataset', 'Split Data', 'Linear Regression', 'Train Model', and 'Score Model'. The 'Split Data' module is highlighted, and its properties are visible on the right. The properties include: Splitting mode: Split Rows, Fraction of rows in the...: 0.7, Randomized split: checked, Random seed: 0, and Stratified split: False. The 'Train Model' module is also highlighted, and its properties are visible on the right. The properties include: Label column: Selected columns: Column names: price, and Launch column selector. The 'Score Model' module is also highlighted, and its properties are visible on the right. The properties include: Task output was present in output cache. The workflow is titled 'Experiment created on 9/21/2021 19k41a04a0'. The status bar at the bottom shows 'Finished running'.

Data Splitting

## Model Training and Algorithm

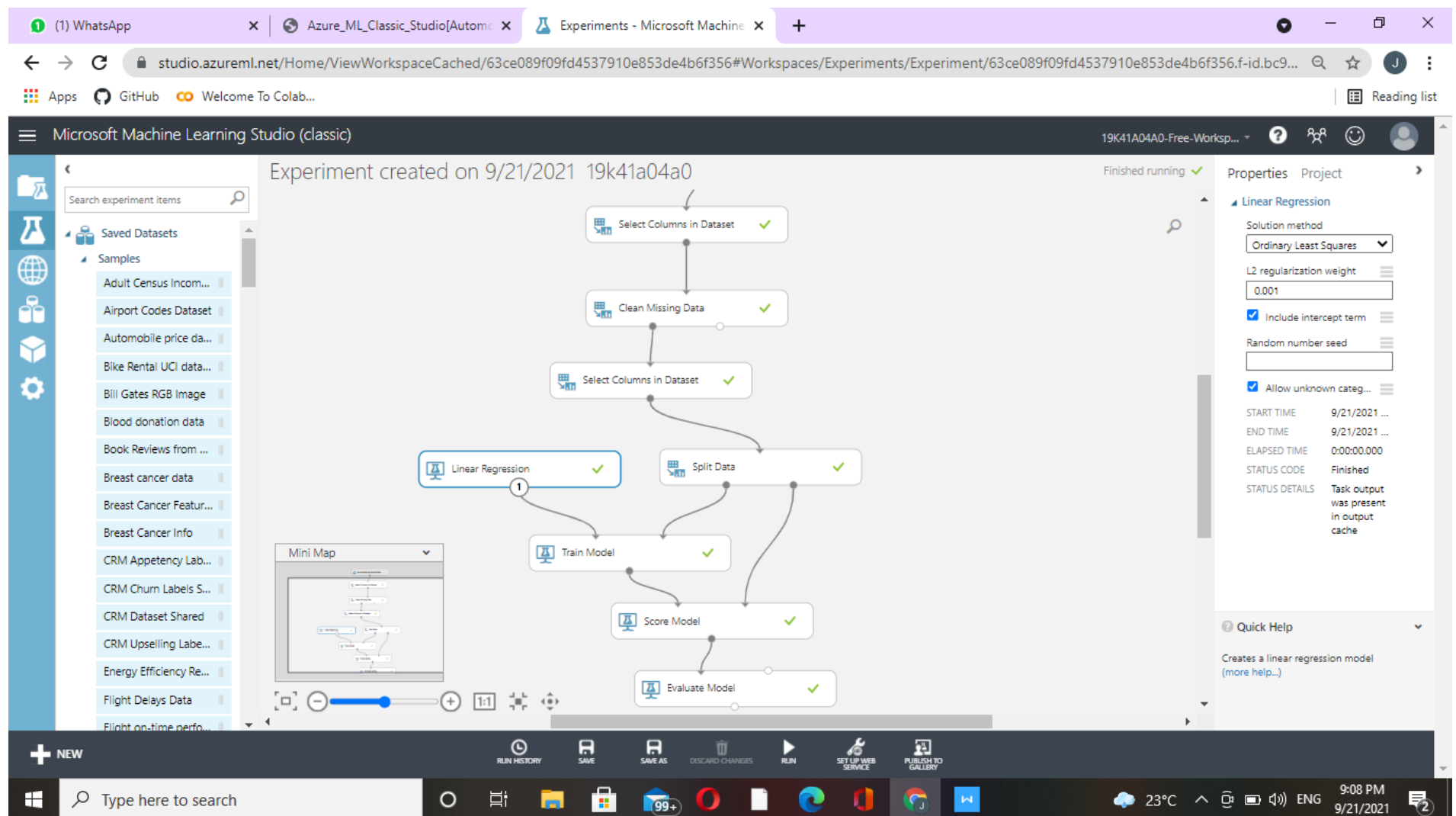
The screenshot displays the Microsoft Machine Learning Studio (classic) interface. The main workspace shows a workflow diagram with the following modules: 'Select Columns in Dataset', 'Clean Missing Data', 'Select Columns in Dataset', 'Split Data', 'Linear Regression', 'Train Model', and 'Score Model'. The 'Train Model' module is highlighted, and its properties are visible on the right. The properties include: Label column: Selected columns: Column names: price, and Launch column selector. The 'Score Model' module is also highlighted, and its properties are visible on the right. The properties include: Task output was present in output cache. The workflow is titled 'Experiment created on 9/21/2021 19k41a04a0'. The status bar at the bottom shows 'Finished running'.

Training model



## using Linear regression to train the model

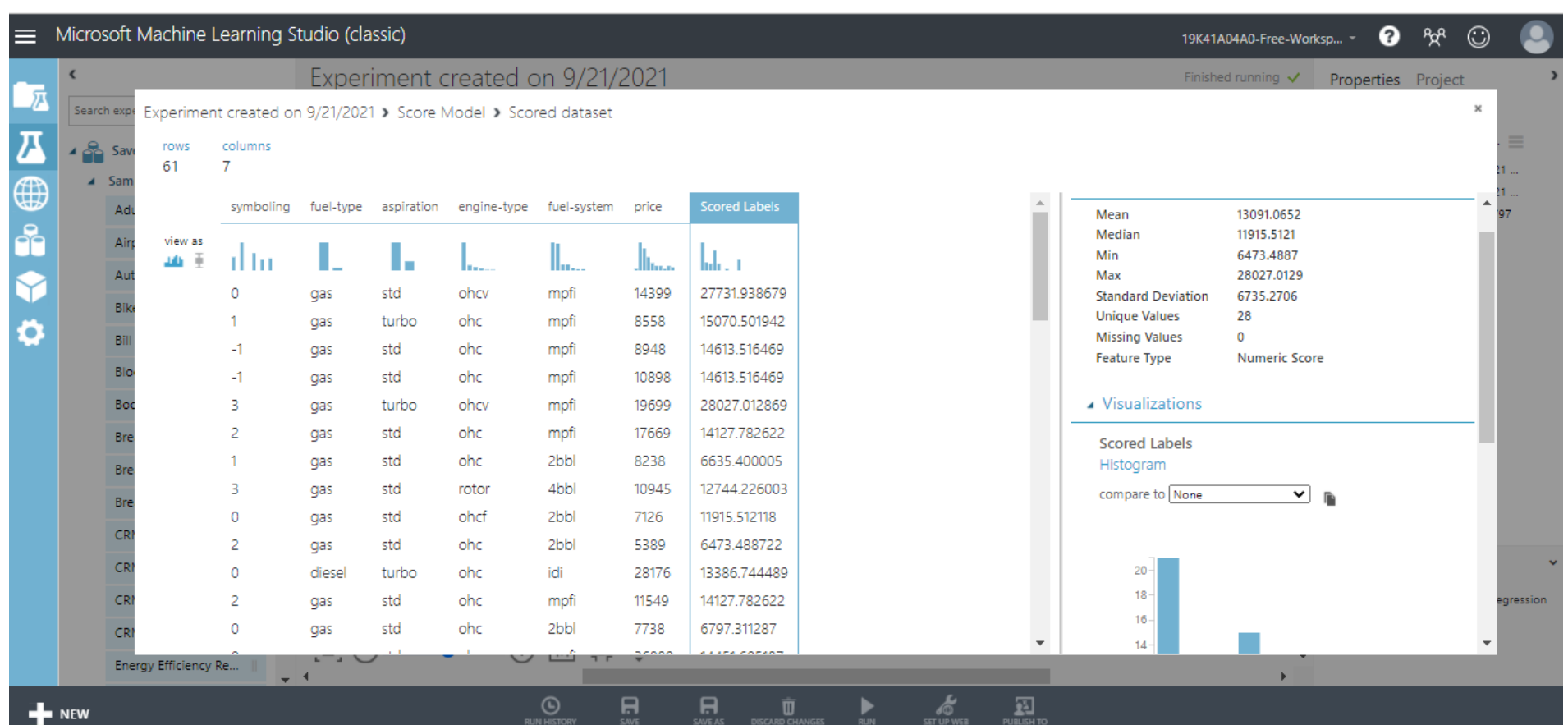
- Since the goal of this sample is to predict automobile prices, and because the label column (price) is continuous data, a regression model can be a good choice. We use Linear Regression for this pipeline.



### Linear Regression

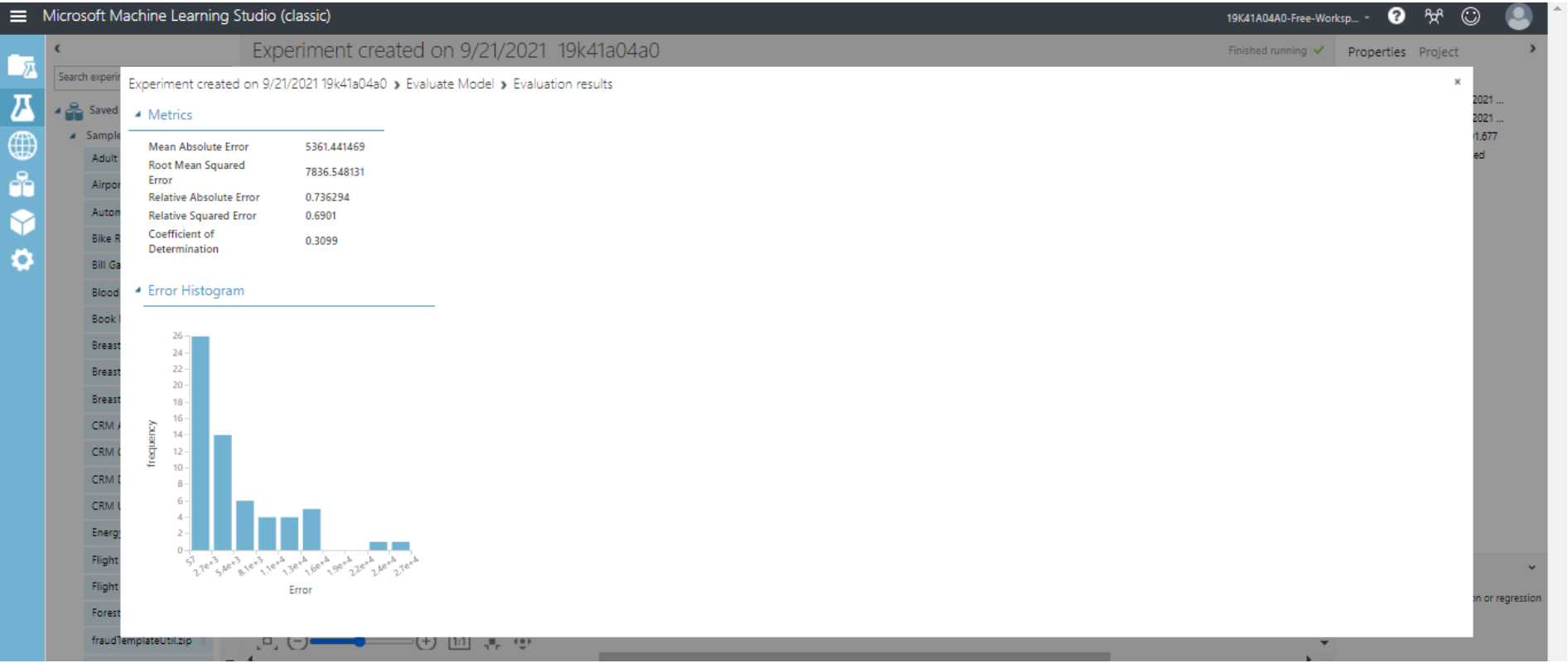
## Score Model and Evaluate Model

- 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



Model Evaluation Results