



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:

Automobile Price prediction [19K41A0575]

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

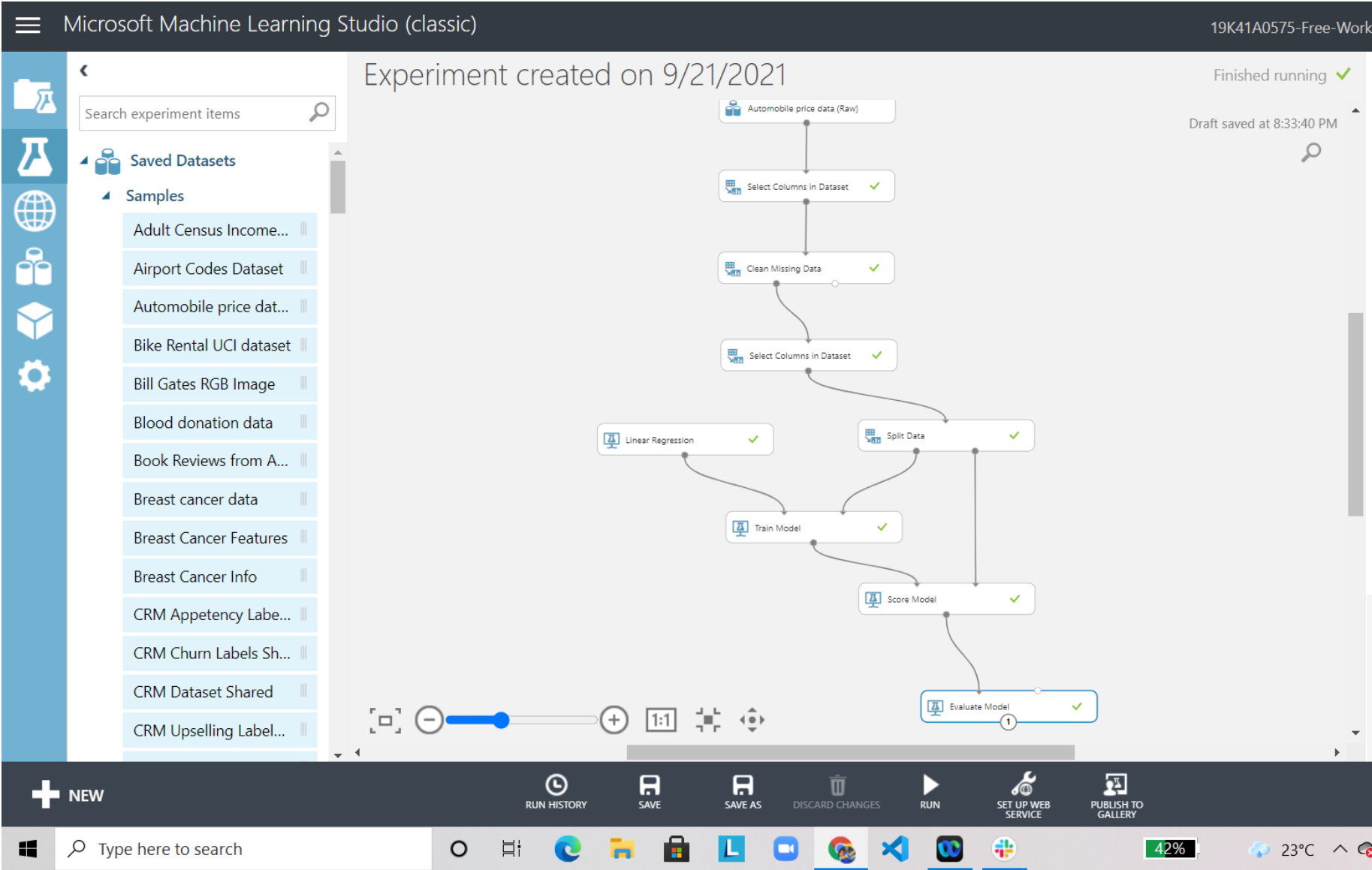
 <https://gallery.cortanaintelligence.com/Experiment/Automobile-price-prediction-412>

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

Machine Learning Project Workflow

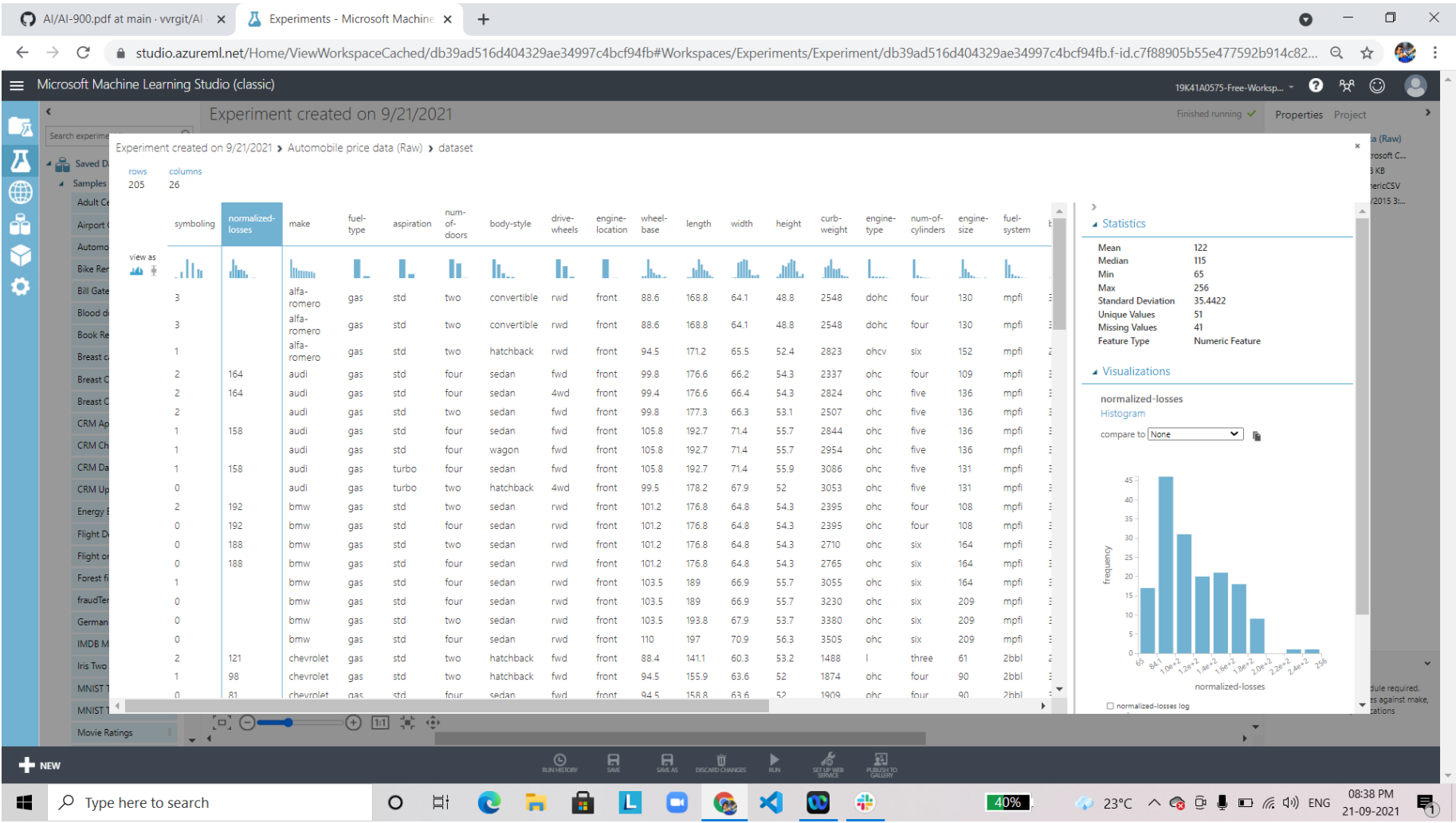
1. Import Data
2. Explore Data (Missing values, outliers)
3. Preprocess data (Missing value imputation, outlier treatment, normalization)
4. Model Selection
5. Model Training
6. Model Testing
7. Model Deployment

Workflow



Project Workflow

Import Data:



Explore Data

AI/AI-900.pdf at main · vvrigit/AI · xExperiments - Microsoft Machine · x+

studio.azureml.net/Home/ViewWorkspaceCached/db39ad516d404329ae34997c4bcf94fb#Workspaces/Experiments/Experiment/db39ad516d404329ae34997c4bcf94fb.f-id.c7f88905b55e477592b914c82f1bd...

Microsoft Machine Learning Studio (classic)19K41A0575-Free-Worksp...?

Experiment created on 9/21/2021

Finished running ✓Draft saved at 8:45:26 PM

Search experiment items

Saved Datasets

Samples

Adult Census Income...

Airport Codes Dataset

Automobile price dat...

Bike Rental UCI dataset

Bill Gates RGB Image

Blood donation data

Book Reviews from A...

Breast cancer data

Breast Cancer Features

Breast Cancer Info

CRM Appetency Labe...

CRM Churn Labels Sh...

CRM Dataset Shared

CRM Upselling Label...

Automobile price data (Raw)

Select Columns in Dataset ✓1

Clean Missing Data ✓

Select Columns in Dataset ✓

Linear Regression ✓

Split Data ✓

Train Model ✓

Score Model ✓

Evaluate Model ✓

Select Columns

Selected columns:

All columns

Exclude column names:

normalized-losses

Launch column selector

START TIME9/21/2021 ...

END TIME9/21/2021 ...

ELAPSED TIME0:00:00.000

STATUS CODEFinished

STATUS DETAILSTask output was present in output cache

Quick Help

Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns.
(more help...)

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

Type here to search

37%23°C08:48 PM21-09-2021

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

AI/AI-900.pdf at main · vvrigit/AI · xExperiments - Microsoft Machine · x+

studio.azureml.net/Home/ViewWorkspaceCached/db39ad516d404329ae34997c4bcf94fb#Workspaces/Experiments/Experiment/db39ad516d404329ae34997c4bcf94fb.f-id.c7f88905b55e477592b914c82f1bd...

Microsoft Machine Learning Studio (classic)19K41A0575-Free-Worksp...?

Experiment created on 9/21/2021

Finished running ✓Draft saved at 8:49:23 PM

Search experiment items

Saved Datasets

Samples

Adult Census Income...

Airport Codes Dataset

Automobile price dat...

Bike Rental UCI dataset

Bill Gates RGB Image

Blood donation data

Book Reviews from A...

Breast cancer data

Breast Cancer Features

Breast Cancer Info

CRM Appetency Labe...

CRM Churn Labels Sh...

CRM Dataset Shared

CRM Upselling Label...

Automobile price data (Raw)

Select Columns in Dataset ✓

Clean Missing Data ✓12

Select Columns in Dataset ✓

Linear Regression ✓

Split Data ✓

Train Model ✓

Score Model ✓

Evaluate Model ✓

Clean Missing Data

Columns to be cleaned

Selected columns:

All columns

Launch column selector

Minimum missing value ratio0

Maximum missing value ratio1

Cleaning modeRemove entire row

START TIME9/21/2021 8:41:59 PM

END TIME9/21/2021 8:41:59 PM

ELAPSED TIME0:00:00.000

STATUS CODEFinished

STATUS DETAILSTask output was present in output cache

Quick Help

Specifies how to handle the values missing from a dataset
(more help...)

NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

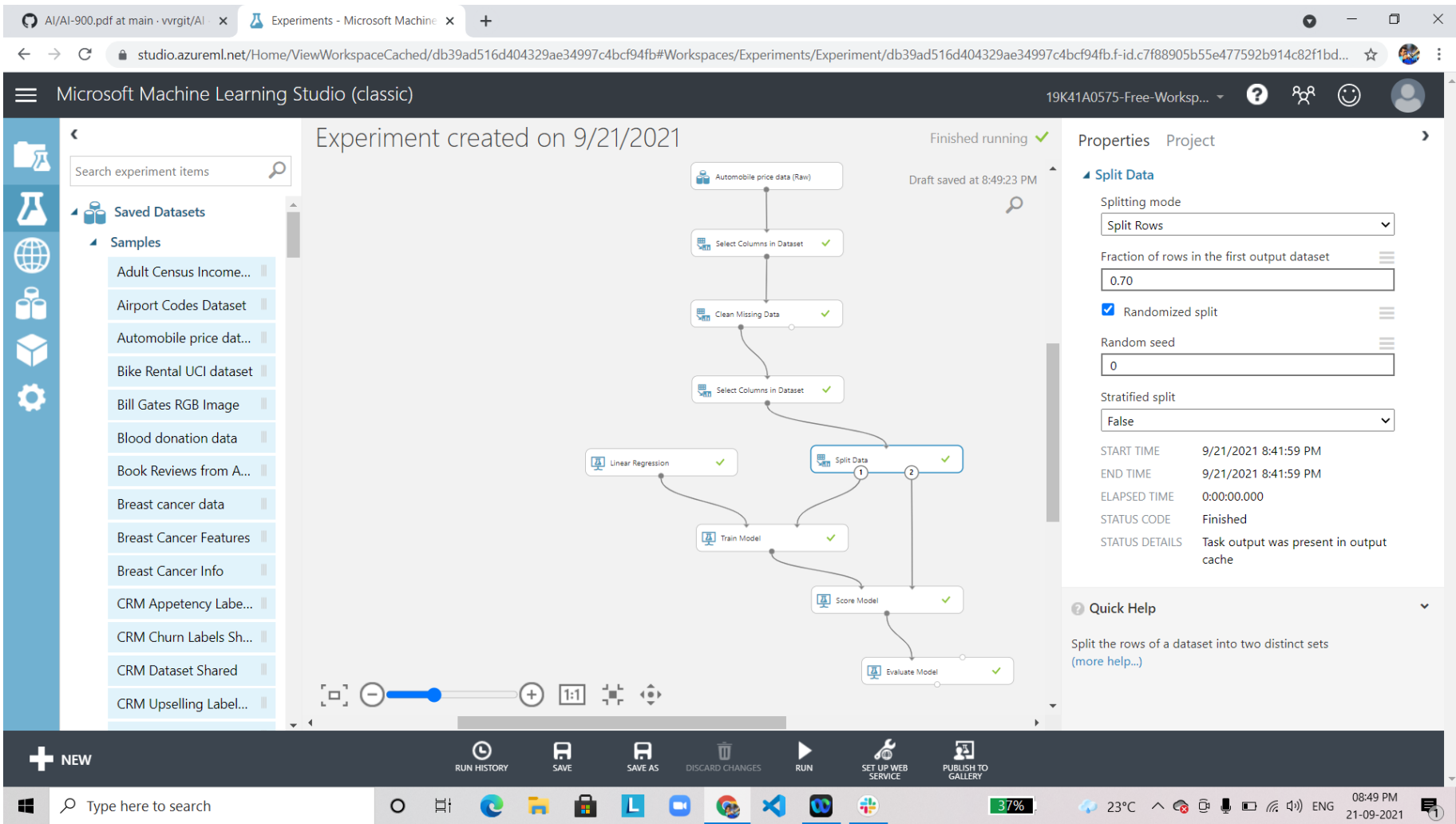
Type here to search

37%23°C08:49 PM21-09-2021

Data Cleaning

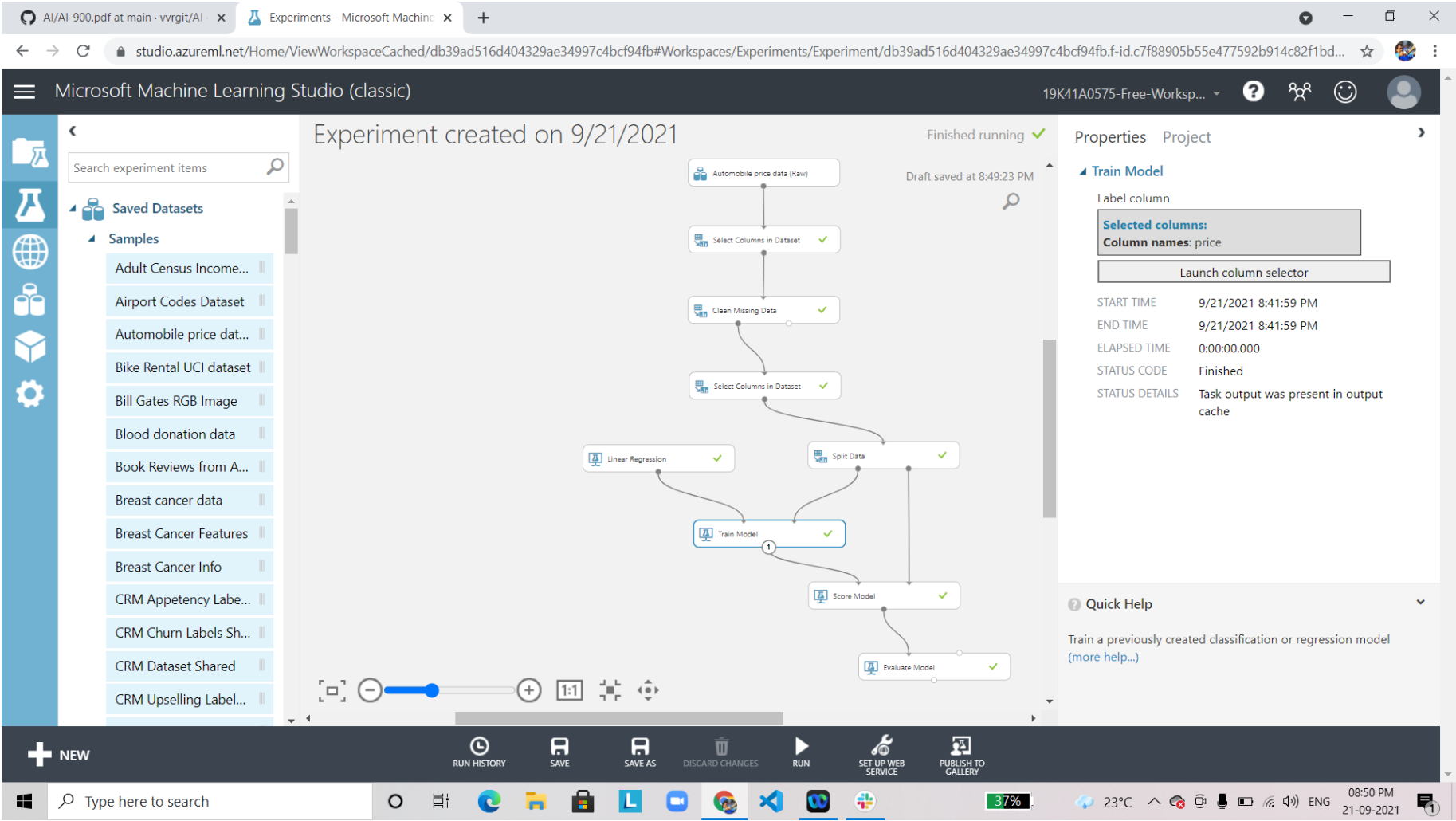
Split Data

- 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.



Data Splitting

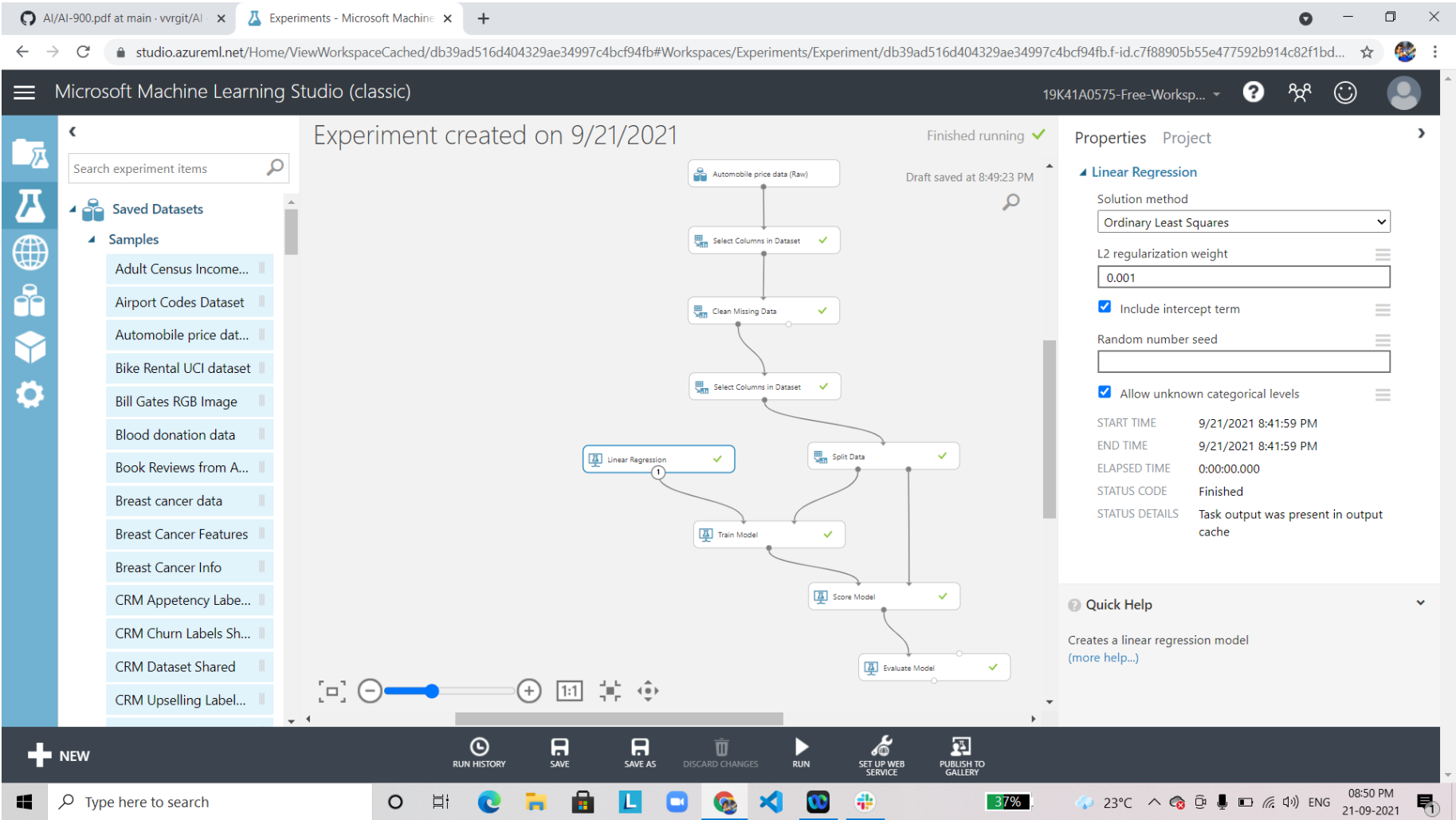
Model Training and Algorithm



Model Training

using Linear regression to train the model

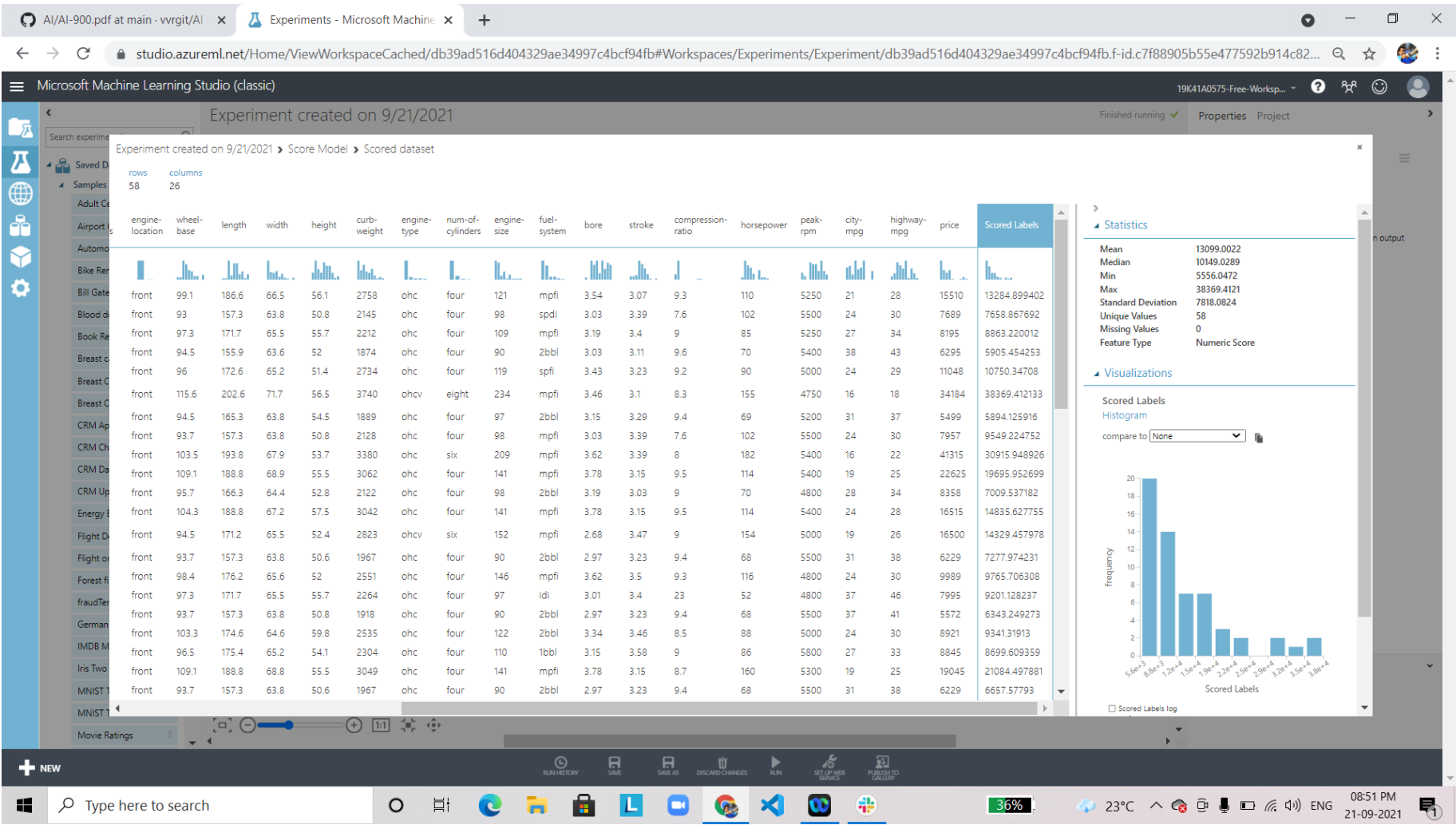
- Since the goal of this sample is to predict automobile prices, and because the label column (price) is continuous data, a regression model can be a good choice. We use Linear Regression for this pipeline.



Linear Regression

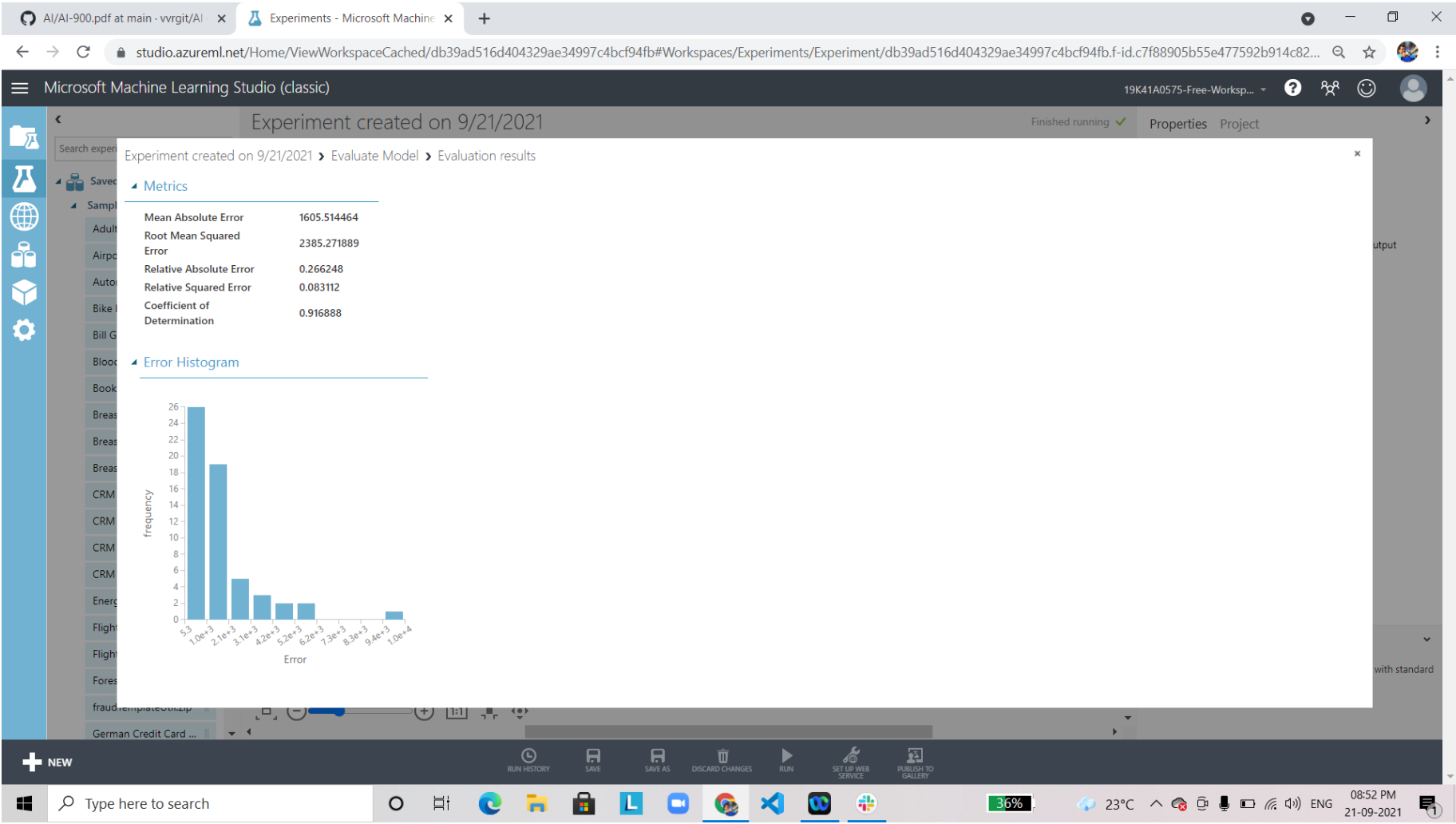
Score Model and Evaluate Model

- 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



Model Evaluation Results