Microsoft

Introduction to Visual interface for Azure Machine Learning

Contents

Overview

Summary

This lab is intended to serve as an introduction to creating a predictive model with Visual Interface for Azure Machine Learning Studio.

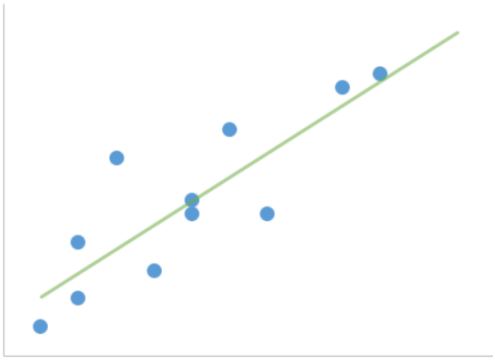
The lab will walk through accessing the Visual Interface environment, exploring and visualizing data in Azure Machine Learning, and creating a simple predictive model.

Business Case

Consumers often evaluate similar products by specific metrics of interest to them. In the auto industry, Miles per Gallon (MPG) always comes up as an important metric for consumers. How do manufacturers know what an acceptable MPG will be for the vehicle they are producing? Using advanced analytics, auto manufacturers can use vehicle attributes and MPG for similar automobiles in the market to predict what an acceptable MPG will be for their car coming off the assembly line.

For this lab, you will be working with a dataset that includes various information about automobiles from the 1970s and early 1980s. The dataset includes attributes like miles per gallon (MPG), horsepower, acceleration, weight, etc.

The lab will use a linear regression algorithm to predict an acceptable MPG for an automobile. Linear regression is used to predict a single, numeric value based on one or many independent variables. It does this by fitting a representative line, or function, to a collection of input variables. This line/function can then be used to predict future values based on new input data.



Learning Objectives

Upon completing this lab, you will have hands-on experience with the following functions and concepts related to Azure Machine Learning Studio:

- Creating and logging in to a free Azure Machine Learning Workspace and launching Azure Machine Learning Studio

Overview

- Creating, modifying, and saving an experiment with ML Studio.
    - Running an experiment
    - Using sample datasets an experiment
    - Browsing modules to use in an experiment
    - Using the search function to find modules to use in an experiment
    - Connecting modules in an experiment
    - Modifying and configuring properties of modules in an experiment
- Visualizing and exploring data in ML Studio
    - Exploring summary statistics about datasets and features
    - Visualizing features with scatterplots, boxplots, and histograms
    - Visualizing relationships between features with scatterplots and boxplots
    - Visualizing predictive model results
- Creating a basic predictive model in ML Studio
    - Splitting data into a training dataset and a test dataset
    - Training a Linear Regression model
    - Testing a trained model
    - Evaluating model performance

Lab Requirements/Prerequisites

- Access to an environment in BP with Azure Machine Learning services enabled.

Access Workspace and Create Experiment

Create/Access an Azure Machine Learning Workspace

To get started, you will need to create and log in to a Azure Machine Learning workspace. A workspace is like an all-inclusive development environment with the tools to create, manage, and publish machine learning models.

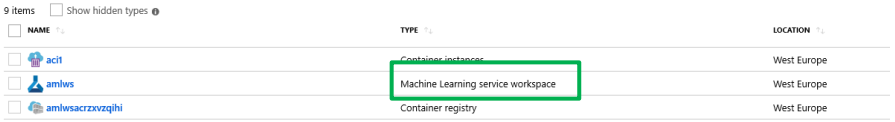
1. Open **Internet Explorer** using the shortcut on the desktop taskbar.

2. Go to the **Azure Portal** website by typing **https://portal.azure.com/** in the address bar.

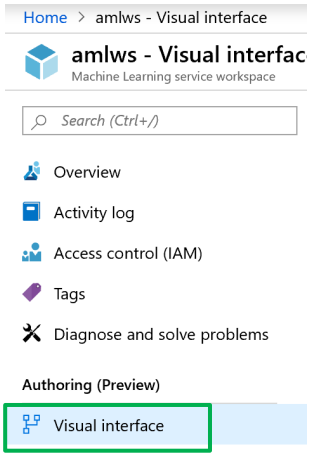
3. **Sign In** using your AD credentials.

4. Enter the **email address** and **password** associated with your organization AD credentials, and click the **Sign In** button.

5. Select the Machine Learning service workspace

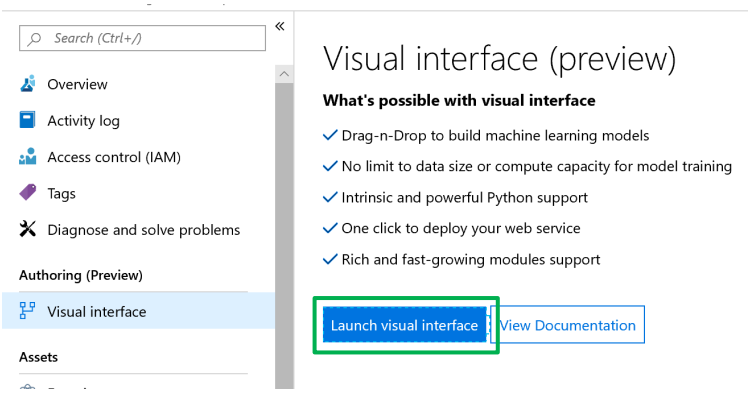


6. On the right side menu, under **Authoring (Preview)**, click on the link to **Visual Interface**.



7. Click on **Launch Visual Interface**. This will open Azure Machine Learning Visual Interface in a new tab of the browser.

Access Workspace and Create Experiment

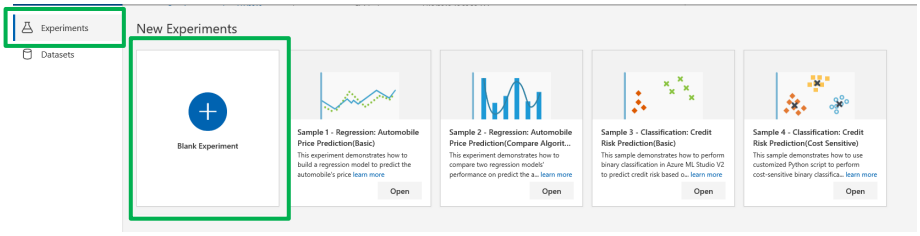


Create a Blank Experiment

Next, we will create our first experiment. An experiment is a collection of data, tasks, and machine learning algorithms that make up a model.

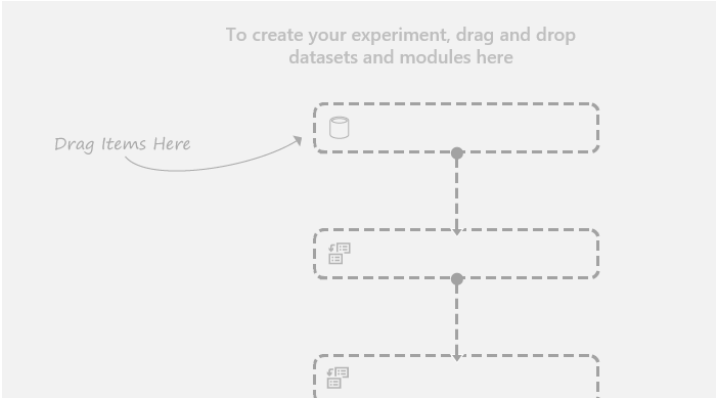
1. Click the **NEW** button in the bottom left corner of the page.

2. Make sure **EXPERIMENTS** is highlighted in the **NEW** dialogue window, and click the **Blank Experiment** pane.



3. You are now in the **ML Studio**. Notice:

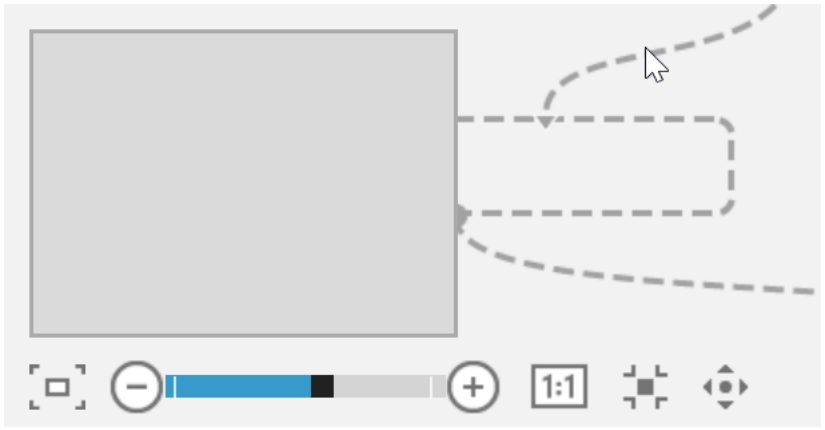
   a. The *Canvas* in the center of the screen. This is where you will drag and drop modules and string them together to create a data flow for your experiment.



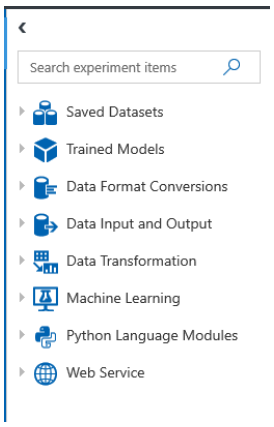
Access Workspace and Create Experiment

b. The navigation icons on the far left of the site allowing you to browse back to different *Experiments*, *Webservices* or *Datasets*

c. There is also a navigation map to traverse/zoom large experiments easily.

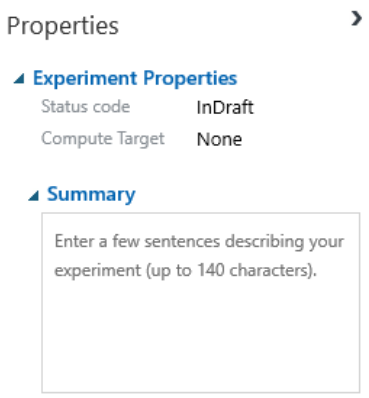


d. The *Modules* pane down the left side of the Canvas. Modules are the individual components that make up your Experiment.

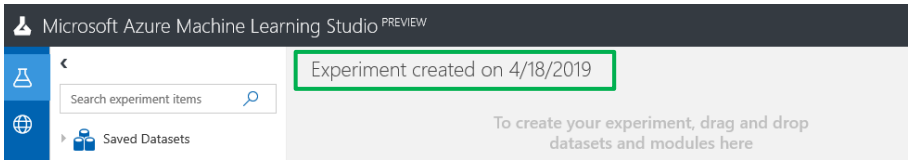


e. The *Properties* pane down the right side of the Canvas. This is where you will configure the properties of the different Modules used in your Experiment.

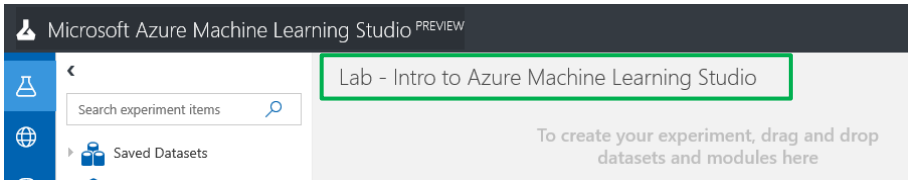
Access Workspace and Create Experiment



4. At the top of the **Canvas**, highlight and delete the text that reads **Experiment created on…**, and replace it with **Lab - Intro to Azure Machine Learning Studio**.



To



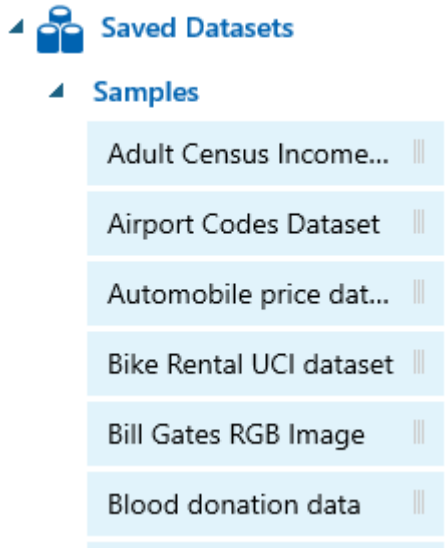
Explore Available Datasets

Input Sample Data

Azure Machine Learning offers several ways to connect to and import data. For this lab, we will work with one of the sample datasets included with Azure Machine Learning.

1. On the **Modules** panel, click **Saved Datasets** and then **Samples**.

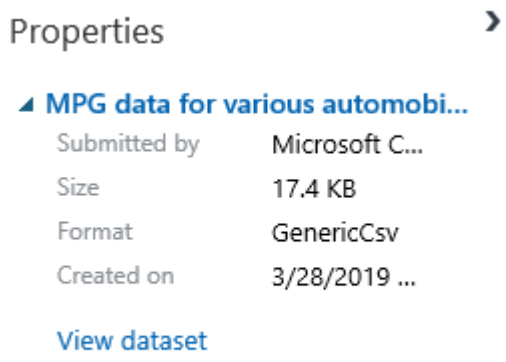
   This expands all of the sample datasets included in ML Studio.



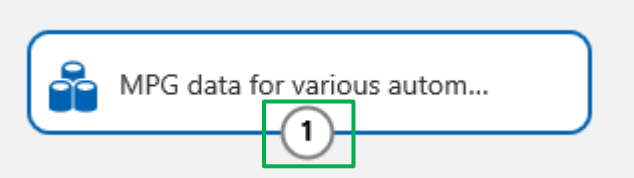
2. Scroll until you find **MPG data for various automobiles**.

3. Over on the **MPG** dataset and notice the description also shows up.

4. Click and drag the **MPG** dataset onto the **Canvas**. Notice the Properties pane is now reflecting information about the dataset.



Notice at the bottom of the *MPG dataset* module on the Canvas, there is a small circle called a port. Ports on the top of modules are called ***input ports***, and ports on the bottom of modules are ***output ports***. These ports are used to connect modules to one another and to provide a menu of additional options for the module.

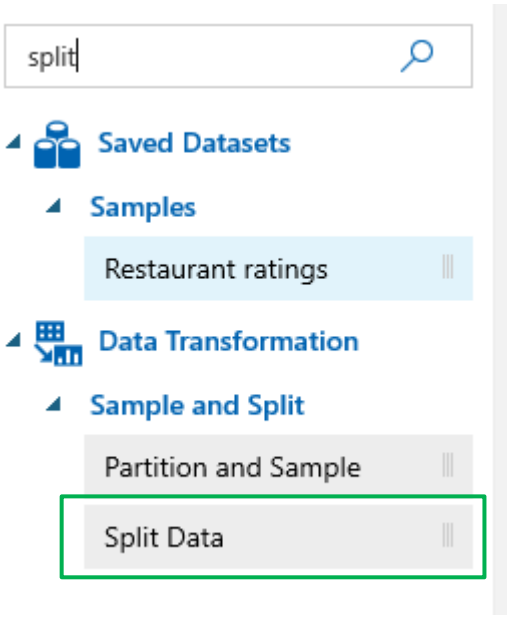


Create a Simple Predictive Model

Split Input Data into Train and Test Data Sets

We are ready to create a predictive model. The first thing we will do is split the original dataset into 2 datasets: one dataset will be used for training a model, and one will be used for testing our model (as it is typically better to test our models with different data than what we trained it with).

1. In the search box at the top of the **Modules** pane, type the word **split** and hit **Enter**.



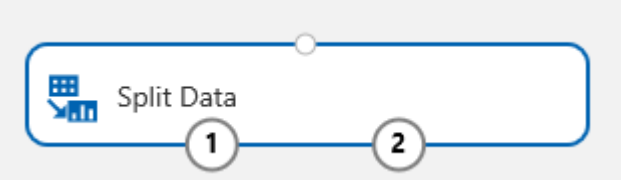
Notice the list of modules has been filtered to show only those relevant to the search term.

2. Click and drag the **Split Data** module onto the **Canvas** anywhere under the **MPG** dataset**.**

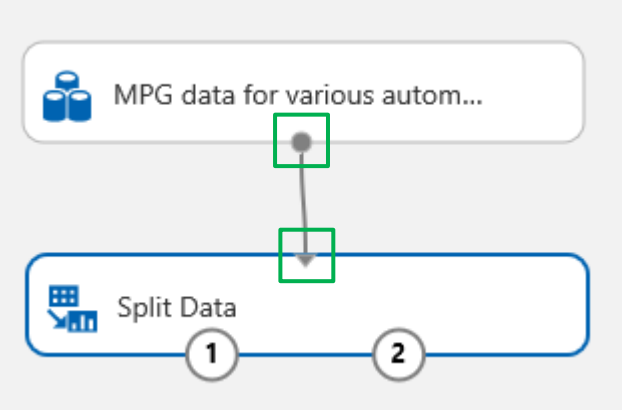
Notice the Split Data module has 1 input port and 2 output ports. The Properties pane displays properties that can be modified for this module. There is also a description of the

Create a Simple Predictive Model

module at the bottom of the Properties pane with a *(more help…)* link. A page will open with more details about the module and its configurable properties when this link is clicked.



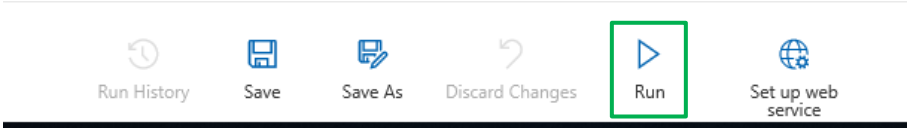
3. Click and drag the **output port** from the **MPG** dataset module to the **input port** of the **Split Data** module**.**



4. In the **Properties** pane of **Split Data** module, type **0.75** in the **Fraction of rows in the first output dataset** textbox.

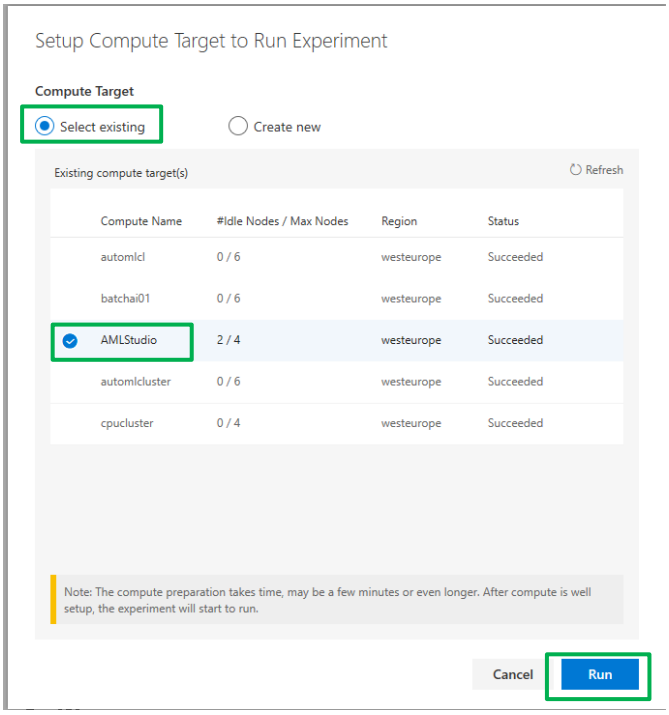
   This configures the module to split 75% of the input rows to the left output port, and 25% of the input rows to the right output port.

5. Click **RUN** at the bottom of the **Canvas**.

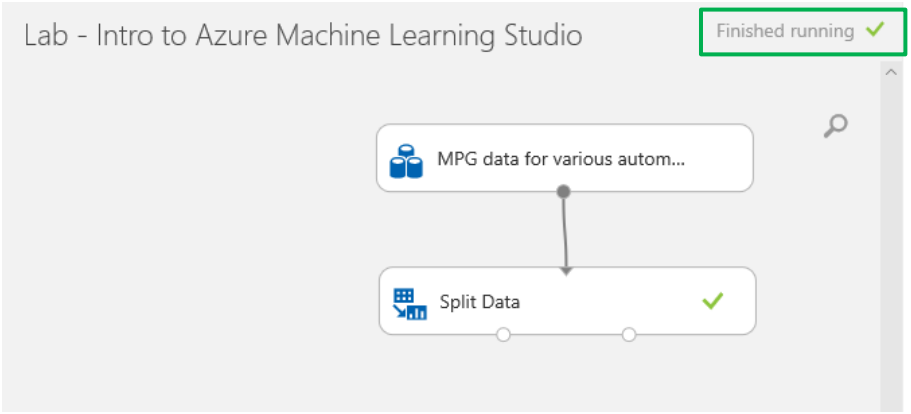


6. In the new window **Setup Compute Target to Run Experiment**, click **Select existing** and choose one of the available clusters and click **Run**. (*Note: that your instructor should indicate which cluster to be used*)

Create a Simple Predictive Model



The experiment will now execute each module in order starting from the first module in the workflow. When the experiment is done executing, the words *Finished running* will display in the top right corner of the Canvas. Notice the Split module has a *green check mark* indicating it completed successfully.



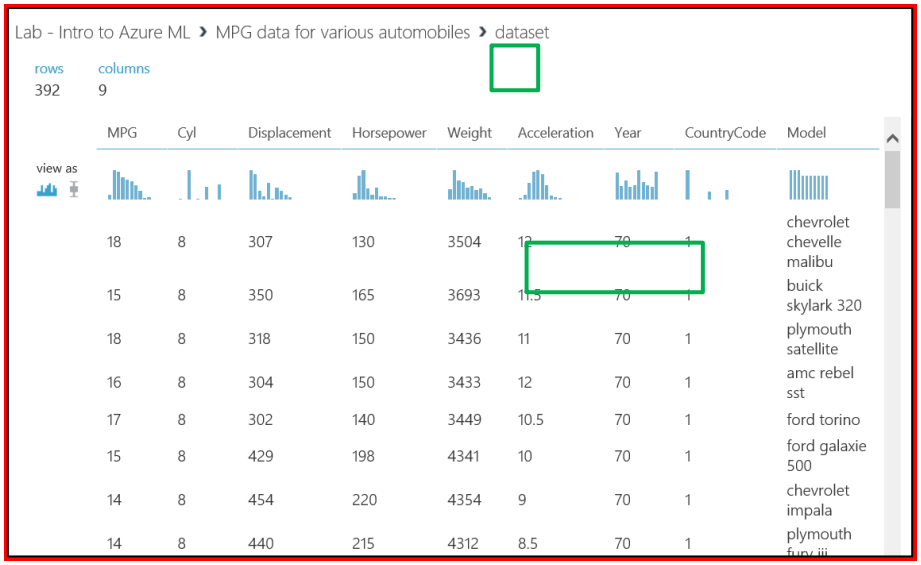
Explore the Input Data

A common task in any advanced analytics workflow is to analyze and profile the data you are working with. The following set of steps highlights some of the ways we can explore and visualize the data we just imported.

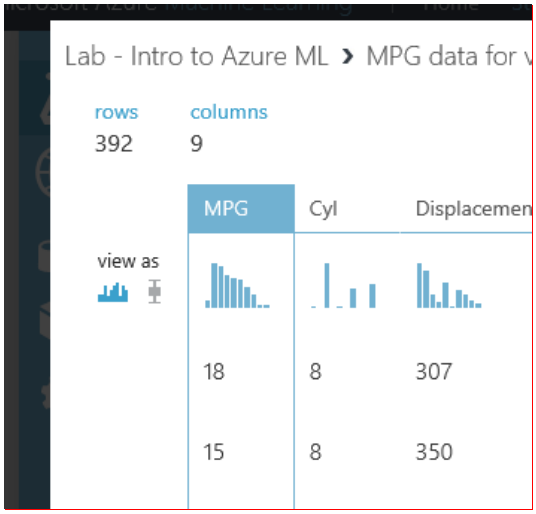
Note: you will need to run the experiment to visualize the data.

1. Click the **output port** on the **MPG** dataset, and select **Visualize** from the menu that is displayed.

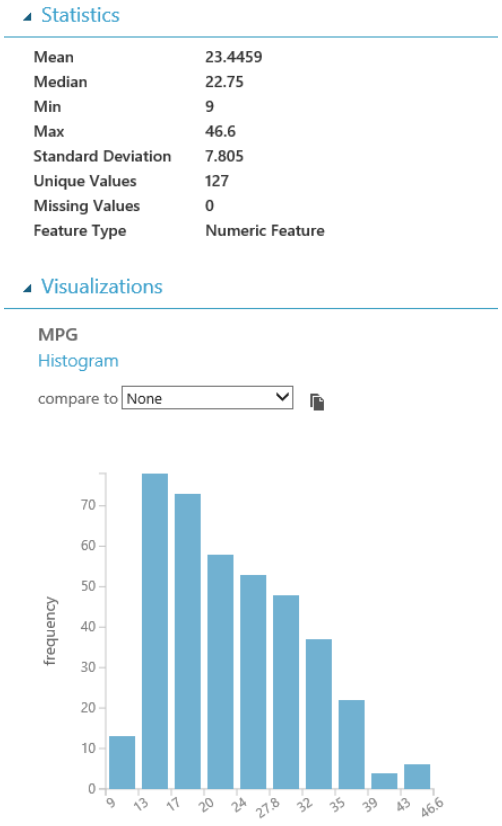
The resulting dialogue box provides the number of rows and columns in the dataset as well as the first 100 rows and first 100 columns in the dataset with a histogram for each column.



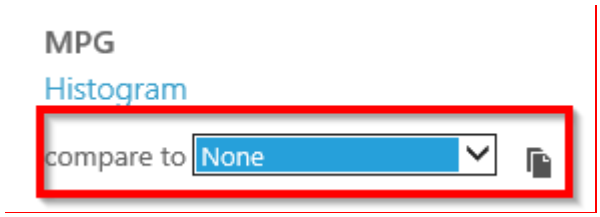
2. Click anywhere in the first column, **MPG**, to highlight the column.



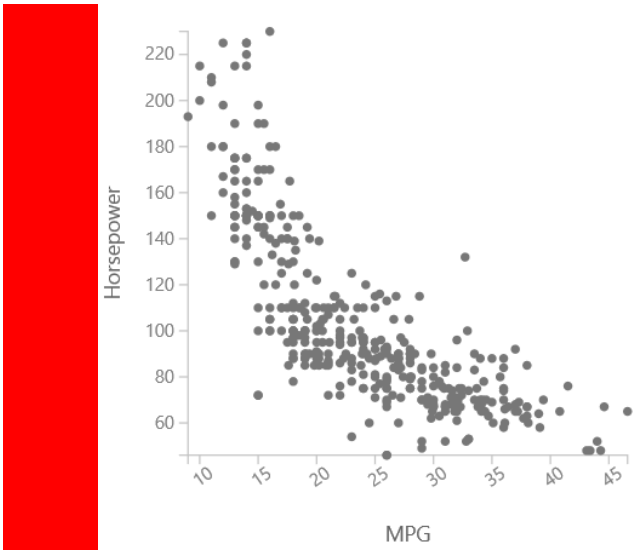
Notice on the right side of the dialogue box, there is now information in the *Statistics* pane and *Visualizations* pane about MPG (you might need to use the horizontal scroll bar in the dialogue box to scroll all the way to the right if Statistics and Visualizations are not visible).



3. In the **Visualizations** pane, change the **compare to** dropdown box from **None** to **Horsepower**.

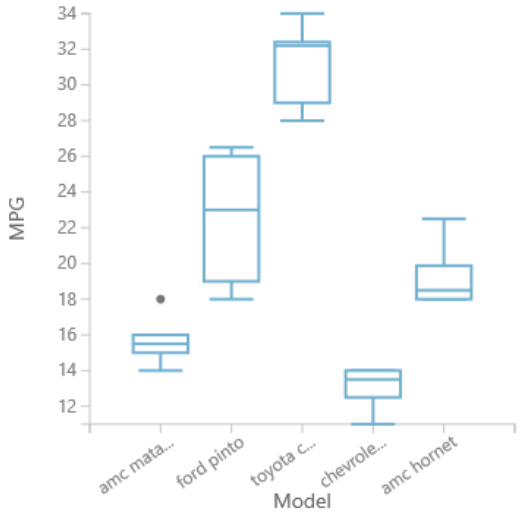


Notice the histogram changed to a *ScatterPlot* comparing MPG to Horsepower.

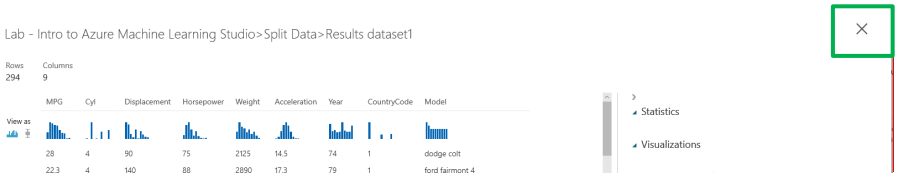


4. Next, change the **compare to** dropdown option from **Horsepower** to **Model.**

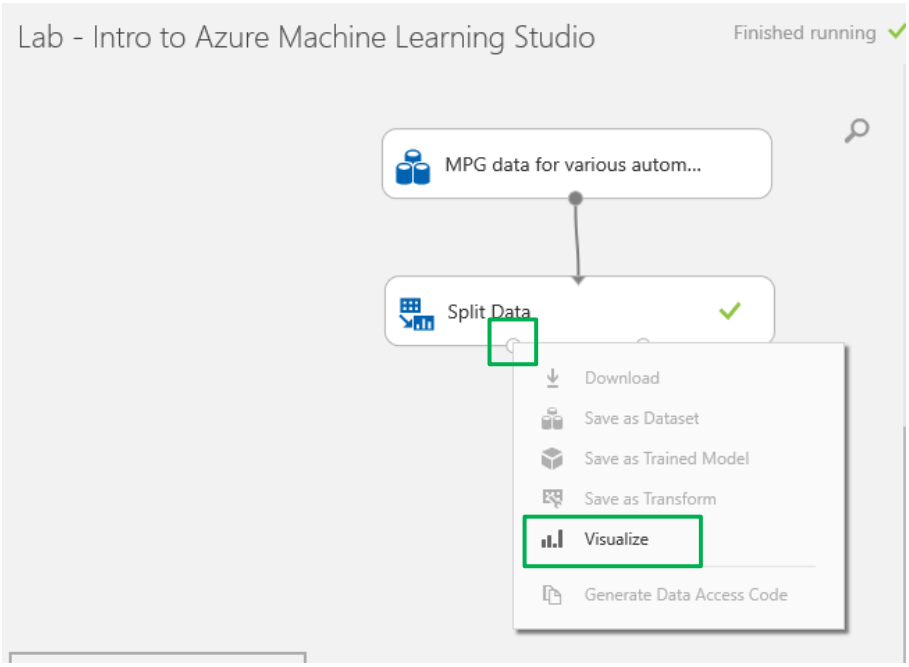
   Notice the resulting chart is now a *MultiboxPlot* with an MPG boxplot displayed for each of the values in the Model column.



5. Click the **X** in the top right corner to close the **Visualize** dialogue box.

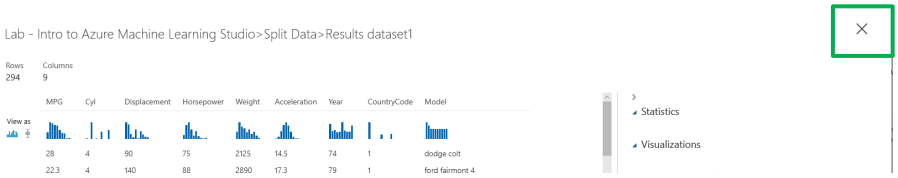


6. Click the **left output port** on the **Split Data** module, and select **Visualize** from the menu that is displayed.



Notice only 294 of the original 392 rows (75%) have been routed to the left output port. The remaining 98 rows (25%) have been routed to the right output port.

7. Click the **X** in the top right corner to close the **Visualize** dialogue box.



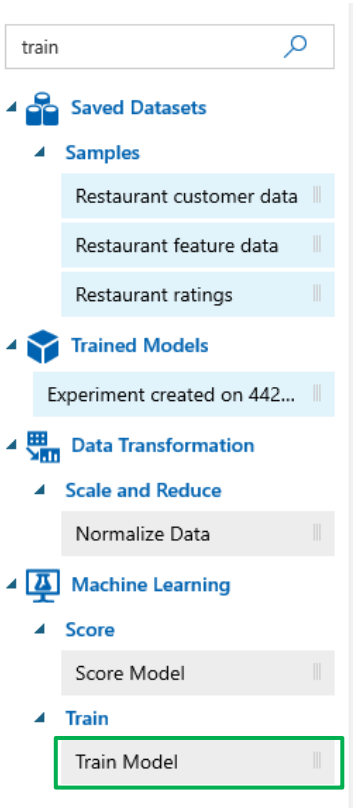
Train Predictive Model

Train a Predictive Model

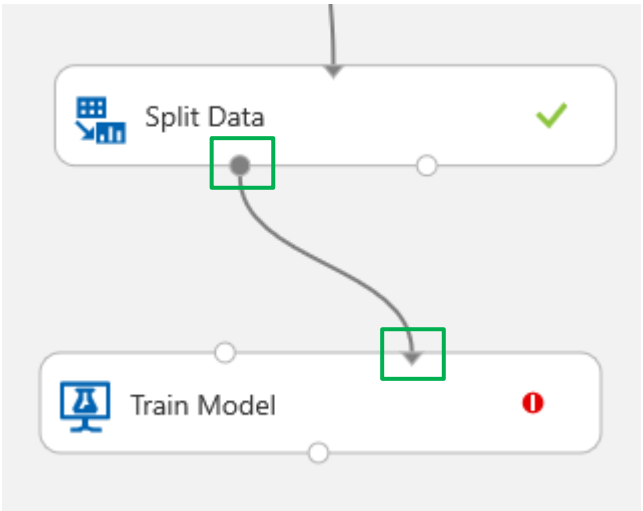
Next, you will use a common Linear Regression algorithm to train a model that will predict an automobile's MPG.

1. Type **train** in the search box at the top of the **Modules** pane.

2. Find the **Train Model** module, and click and drag it onto the **Canvas** below the **Split Data** module.



3. Connect the **left output port** from the **Split Data** module to the **right input port** on the **Train Model** module

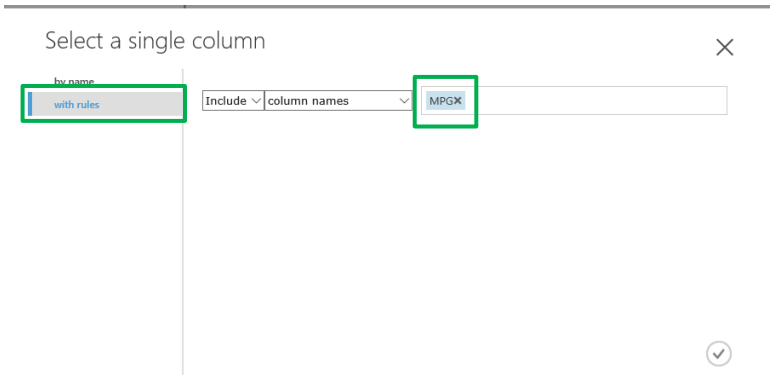


Train Predictive Model

4. Click on the **Train Model** and in the **Properties pane**, click the **Launch column selector** button**.**

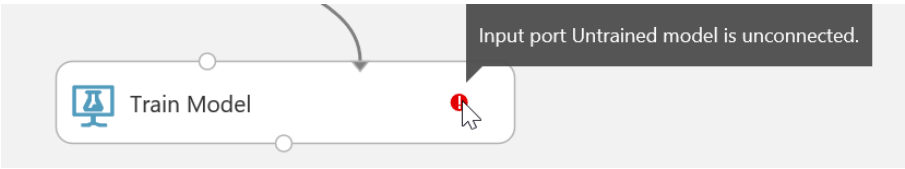
   This launches the *Select Column* dialogue box. Here, we will select the column we want the model to predict.

5. Click the text box with the **red circle** in it, and select **MPG** from the list of columns.



6. Click the  button to save the selection and close the dialogue box.

7. Hover your mouse cursor over the red error icon in the Train Model module. Notice the error message now reads:

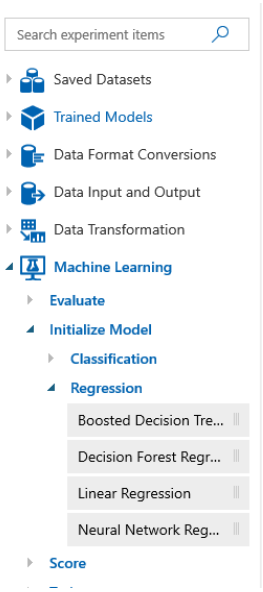


   We now need to select an appropriate machine learning algorithm to forecast MPG.

8. Clear the search box in the **Modules** pane and hit **enter**.

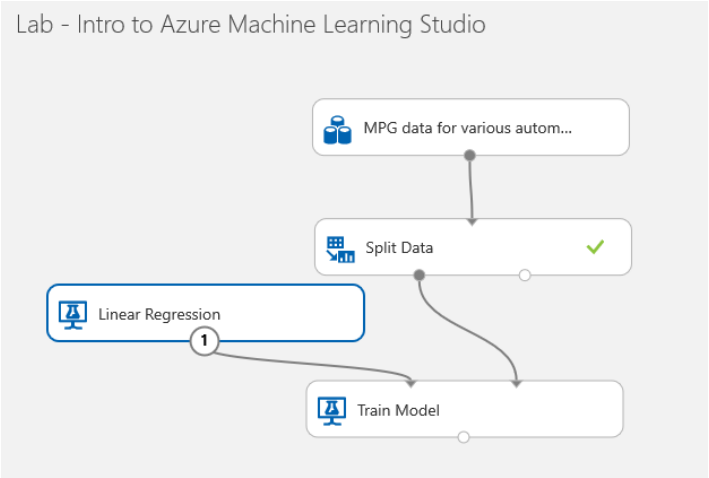
9. In the **Modules** pane, find and click to expand **Machine Learning,** and then click **Initialize Model**, and then click **Regression**.

Train Predictive Model



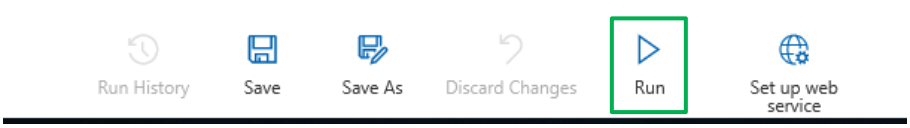
10. Click and drag the **Linear Regression** module onto the **Canvas** just above and to the left of the **Train Model** module.

11. Connect the **output port** of the **Linear Regression** module to the **left input port** of the **Train Model** module.



You might notice there are several parameters that can be modified in the Properties pane for the Linear Regression module. For this lab, we will use the defaults.

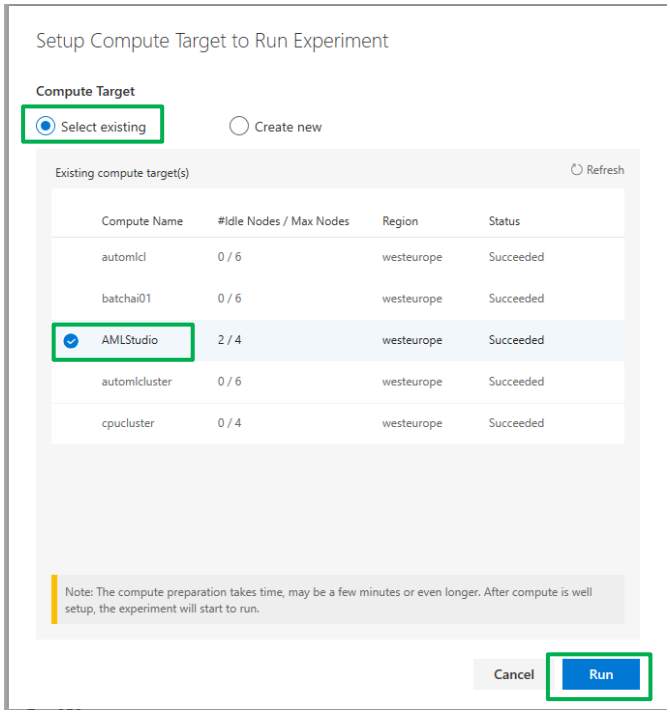
8. Click **RUN** at the bottom of the **Canvas**.



9. In the new window **Setup Compute Target to Run Experiment**, click **Select existing** and choose one of the

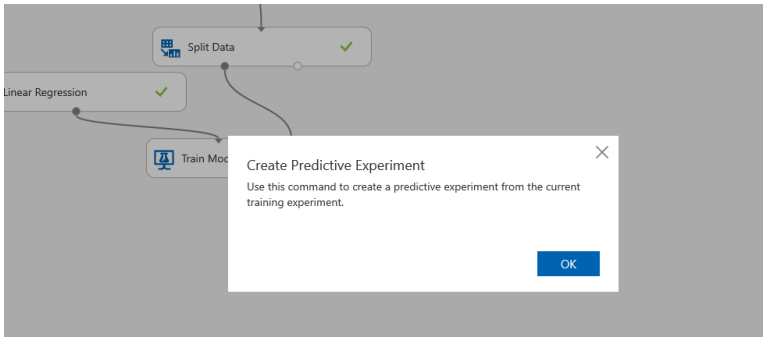
Train Predictive Model

10. available clusters and click **Run**. (*Note: that your instructor should indicate which cluster to be used*)



The experiment will now execute each module in order starting from the first module in the workflow. When the experiment is done executing, the words *Finished running* will display in the top right corner of the Canvas. Notice the Train Model module has a *green check mark* indicating it completed successfully.

12. When the experiment finishes running, **IF** a **CREATE SCORING EXPERIMENT COMMAND** box pops up, click the **X** at the top right corner of this box to close it.

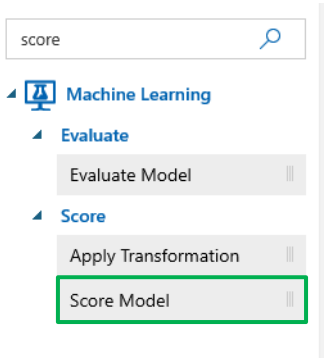


Test the Predictive Model

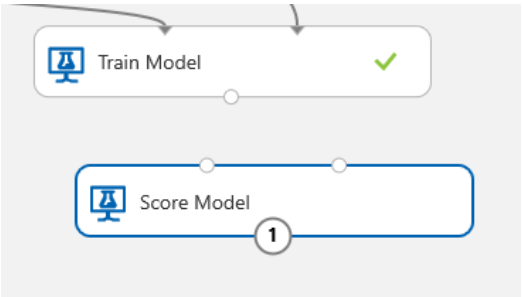
Next, we will use the test dataset we created to test our newly trained model. This will be done using our new model to predict the MPG for each row in the test dataset.

Test Predictive Model

1. In the search box at the top of the **Modules** pane, type the word **score**.

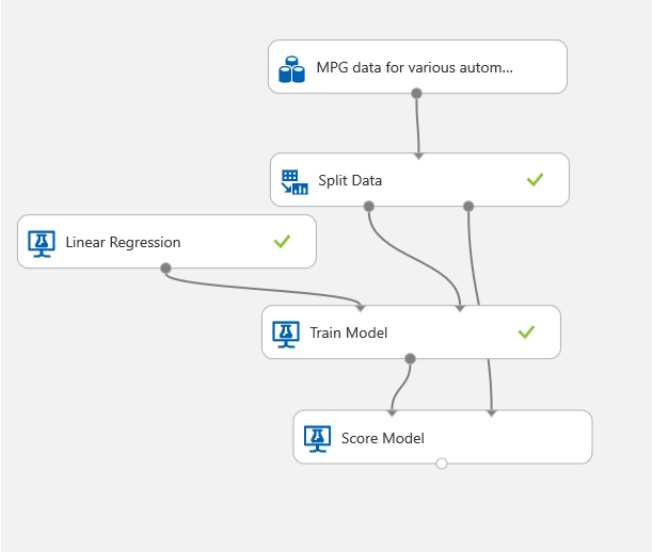


2. Find the **Score Model** module, and click and drag it onto the **Canvas** under the **Train Model** module.



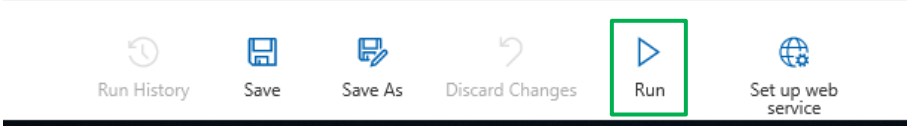
3. Connect the **output port** on the **Train Model** module to the **left input port** on the **Score Model** module.

4. Connect the **right output port** on the **Split** module to the **right input port** on the **Score Model** module.

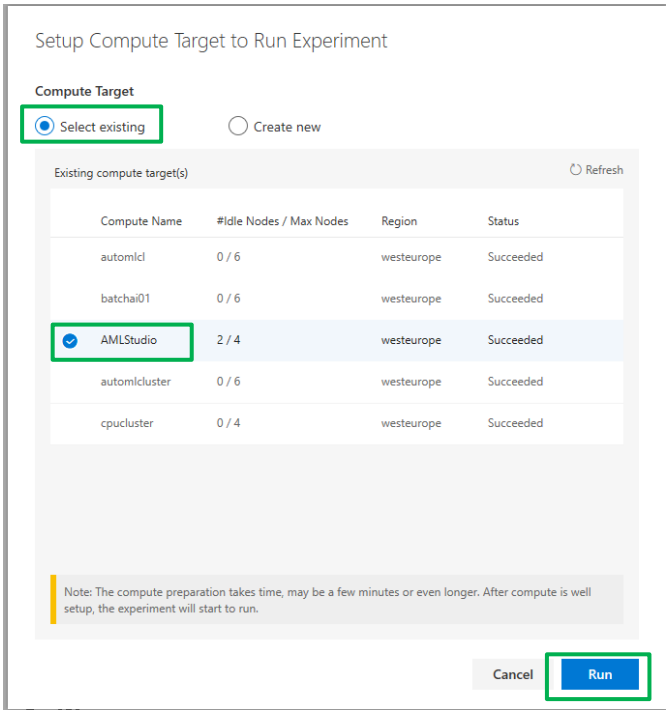


5. Click **RUN** at the bottom of the **Canvas**.

Test Predictive Model



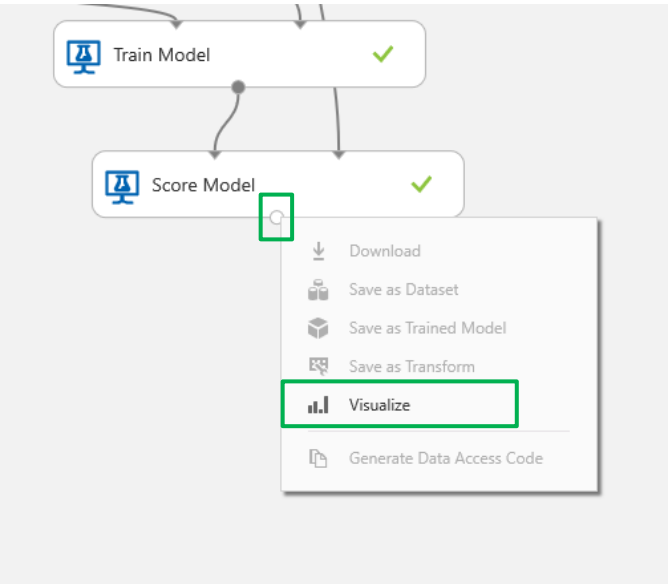
6. In the new window **Setup Compute Target to Run Experiment**, click **Select existing** and choose one of the available clusters and click **Run**. (*Note: that your instructor should indicate which cluster to be used*)



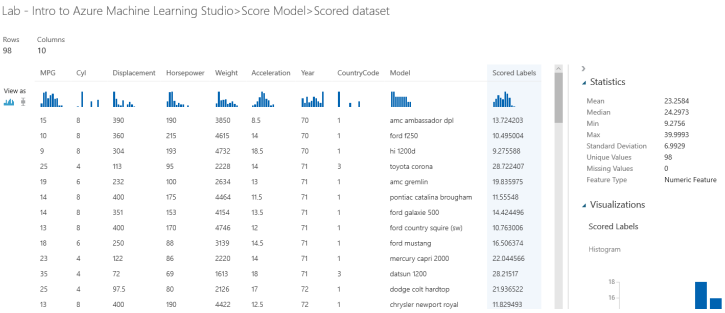
The experiment will now execute each module in order starting from the first module in the workflow. When the experiment is done executing, the words *Finished running* will display in the top right corner of the Canvas. Notice the Train Model module has a *green check mark* indicating it completed successfully.

7. After the experiment has finished running, click the **output port** on the **Score Model** module and select **Visualize** from the displayed menu (see illustration below).

Test Predictive Model



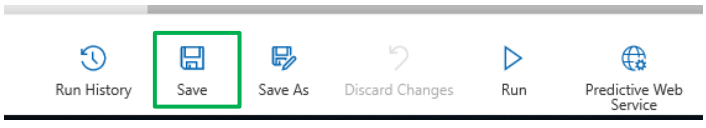
8. In the list of columns, scroll to the right until **Scored Labels** is visible, and click **Scored Labels** to select it.



The *Scored Labels* column represents the predicted MPG for each row in the test dataset. Notice the *Statistics* pane and histogram in the *Visualizations* pane on the right side of the *Visualize* dialogue box.

9. Click the **X** in the top right corner to close the **Visualize** dialogue box.

10. Click **SAVE** to save the experiment.



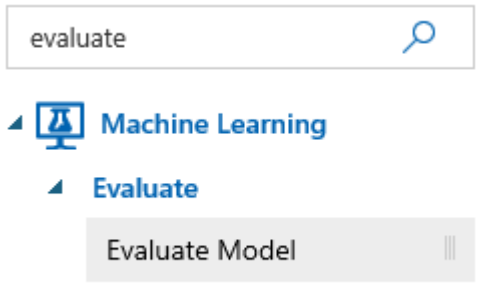
Evaluate Predictive Model

Evaluate the Test Results

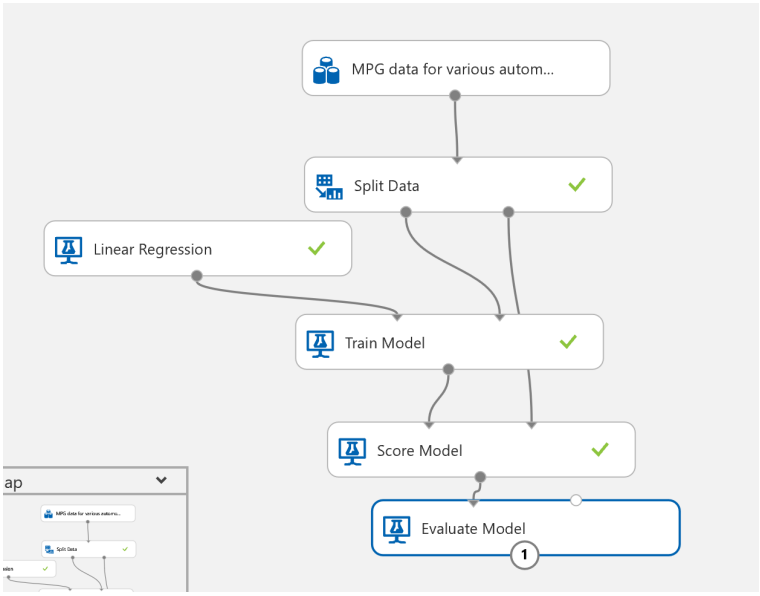
Finally, we will evaluate how well the model performed against the test dataset using a set of standard metrics for measuring regression model performance.

1. In the search box at the top of the **Modules** pane, type the word **evaluate**.

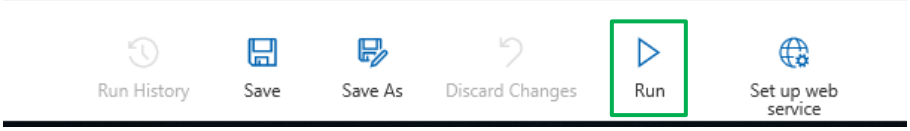
2. Find the **Evaluate Model** module, and click and drag it onto the **Canvas** below the **Score Model** module.



3. Connect the **output port** on the **Score Model** module to the **left input port** on the **Evaluate Model** module.



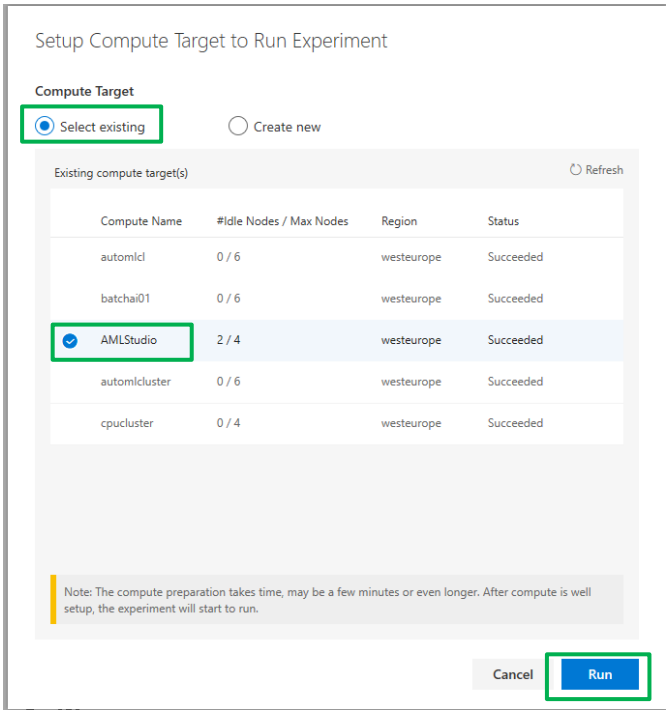
4. Click **RUN** at the bottom of the **Canvas**.



5. In the new window **Setup Compute Target to Run Experiment**, click **Select existing** and choose one of the

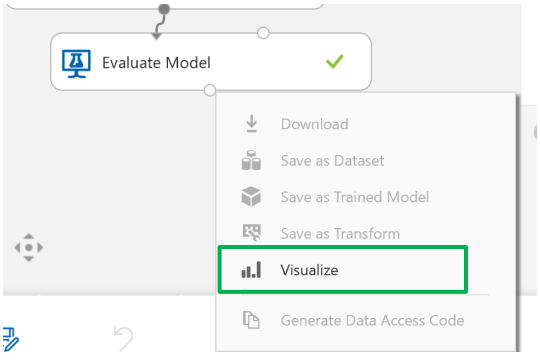
Evaluate Predictive Model

available clusters and click **Run**. (*Note: that your instructor should indicate which cluster to be used*)

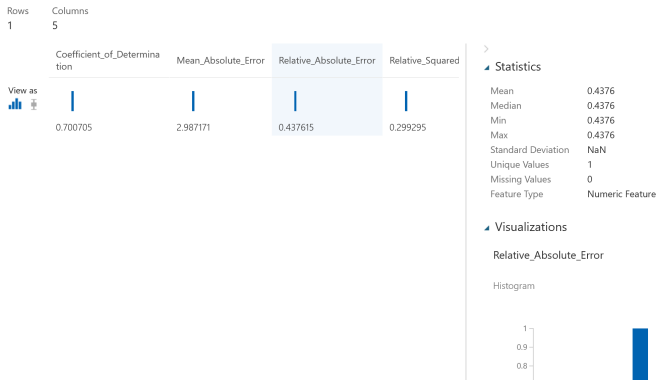


The experiment will now execute each module in order starting from the first module in the workflow. When the experiment is done executing, the words *Finished running* will display in the top right corner of the Canvas. Notice the Evaluate Model module has a *green check mark* indicating it completed successfully.

6. When the experiment has finished running, click the **output port** on the **Evaluate Model** module and select **Visualize** from the displayed menu.



Evaluate Predictive Model



The columns and values in the *Visualize* dialogue box represent common metrics for evaluating the performance of a Linear Regression model. The metrics are calculated using the results of the *Score Model* module. Many of the metrics are based on the Error, which is the difference between the Scored Labels (predicted value) and the actual values.

At this point, you can assess whether or not your model performs at a satisfactory level. If not, you could go back and tweak parameters, add new features, try a different machine learning algorithm, etc. to try and get the model performance to be more acceptable.

7. Click the **X** in the top right corner to close the **Visualize** dialogue box.

8. Click **SAVE** to save your experiment

Exercises

Try the following exercises.

1. Add data cleansing modules to your experiment:

    a. Clean Missing Data (select an appropriate method for replacement).

    b. Remove Duplicate Rows

    In each case consult the help files in the module properties

2. Add in a feature selection method to your experiment. In Azure Machine Learning there are two types to explore:

    a. Filter based feature selection

    b. Permutation feature importance

    What are the differences between the two?

3. Your experiment has one model type – Linear regression. Add to your experiment another regression module (e.g. Boosted Decision Tree Regression) and compare the predictions on the out-of-sample test data with the linear regression model.

Conclusion

This concludes the *Intro to Azure Machine Learning* lab. To recap, you have successfully created and logged into an Azure Machine Learning Service, explored and visualized data in AML Visual Interface, and produced and saved a simple Linear Regression model that predicts automobile MPG.

You now have a model that can be used by an auto manufacturer to predict an acceptable MPG for any new automobile coming off the assembly line. As a next step, you could upload new data to be scored by your model, or you could even publish your model as a web service. A web service gives you the ability to pass individual rows in and get predicted values (MPG) out.

Terms of Use

© 2016 Microsoft Corporation. All rights reserved.
By using this Hands-on Lab, you agree to the following terms:

The technology/functionality described in this Hands-on Lab is provided by Microsoft Corporation in a "sandbox" testing environment for purposes of obtaining your feedback and to provide you with a learning experience. You may only use the Hands-on Lab to evaluate such technology features and functionality and provide feedback to Microsoft. You may not use it for any other purpose. You may not modify copy, distribute, transmit, display, perform, reproduce, publish, license, create derivative works from, transfer, or sell this Hands-on Lab or any portion thereof.

COPYING OR REPRODUCTION OF THE HANDS-ON LAB (OR ANY PORTION OF IT) TO ANY OTHER SERVER OR LOCATION FOR FURTHER REPRODUCTION OR REDISTRIBUTION IS EXPRESSLY PROHIBITED.

THIS HANDS-ON LAB PROVIDES CERTAIN SOFTWARE TECHNOLOGY/PRODUCT FEATURES AND FUNCTIONALITY, INCLUDING POTENTIAL NEW FEATURES AND CONCEPTS, IN A SIMULATED ENVIRONMENT WITHOUT COMPLEX SET-UP OR INSTALLATION FOR THE PURPOSE DESCRIBED ABOVE. THE TECHNOLOGY/CONCEPTS REPRESENTED IN THIS HANDS-ON LAB MAY NOT REPRESENT FULL FEATURE FUNCTIONALITY AND MAY NOT WORK THE WAY A FINAL VERSION MAY WORK. WE ALSO MAY NOT RELEASE A FINAL VERSION OF SUCH FEATURES OR CONCEPTS. YOUR EXPERIENCE WITH USING SUCH FEATURES AND FUNCITONALITY IN A PHYSICAL ENVIRONMENT MAY ALSO BE DIFFERENT.

FEEDBACK. If you give feedback about the technology features, functionality and/or concepts described in this Hands-on Lab to Microsoft, you give to Microsoft, without charge, the right to use, share and commercialize your feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies and services to use or interface with any specific parts of a Microsoft software or service that includes the feedback. You will not give feedback that is subject to a license that requires Microsoft to license its software or documentation to third parties because we include your feedback in them. These rights survive this agreement.

MICROSOFT CORPORATION HEREBY DISCLAIMS ALL WARRANTIES AND CONDITIONS WITH REGARD TO THE HANDS-ON LAB , INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. MICROSOFT

DOES NOT MAKE ANY ASSURANCES OR REPRESENTATIONS WITH REGARD TO THE ACCURACY OF THE RESULTS, OUTPUT THAT DERIVES FROM USE OF THE VIRTUAL LAB, OR SUITABILITY OF THE INFORMATION CONTAINED IN THE VIRTUAL LAB FOR ANY PURPOSE.

DISCLAIMER

This lab contains only a portion of the features and enhancements in Microsoft Azure Machine Learning. Some of the features might change in future releases of the product.