Machine Learning Use Cases

with Azure ML

Agenda

* Use Cases
* Azure ML Studio
* Create a Model
* Operationalize Model
* Getting Started

Why Azure ML ?

* Start in minutes
* No hardware or installation
* Only need a browser
* Use existing R or Python code
* One-click publishing
* Use what you need

Use Cases

* Data Cleaning and Preprocessing
* Build ML Models
* Operationalize your Model
* Technology Mashups
* Jupyter Notebooks

OOTB Algoritms

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

START

Finding unusual data points

Discovering structure

Three or more

Predicting categories

Predicting values

Two

ANOMALY DETECTION

- **One-class SVM** — >100 features, aggressive boundary
- **PCA-based anomaly detection** — Fast training

CLUSTERING

- **K-means**

MULTI-CLASS CLASSIFICATION

- Fast training, linear model — **Multiclass logistic regression**
- Accuracy, long training times — **Multiclass neural network**
- Accuracy, fast training — **Multiclass decision forest**
- Accuracy, small memory footprint — **Multiclass decision jungle**
- Depends on the two-class classifier, see notes below — **One-v-all multiclass**

REGRESSION

- **Ordinal regression** — Data in rank ordered categories
- **Poisson regression** — Predicting event counts
- **Fast forest quantile regression** — Predicting a distribution
- **Linear regression** — Fast training, linear model
- **Bayesian linear regression** — Linear model, small data sets
- **Neural network regression** — Accuracy, long training time
- **Decision forest regression** — Accuracy, fast training
- **Boosted decision tree regression** — Accuracy, fast training, large memory footprint

TWO-CLASS CLASSIFICATION

- **Two-class SVM** — >100 features, linear model
- **Two-class averaged perceptron** — Fast training, linear model
- **Two-class logistic regression** — Fast training, linear model
- **Two-class Bayes point machine** — Fast training, linear model

- Accuracy, fast training — **Two-class decision forest**
- Accuracy, fast training, large memory footprint — **Two-class boosted decision tree**
- Accuracy, small memory footprint — **Two-class decision jungle**
- >100 features — **Two-class locally deep SVM**
- Accuracy, long training times — **Two-class neural network**

© 2015 Microsoft Corporation. All rights reserved.

Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com

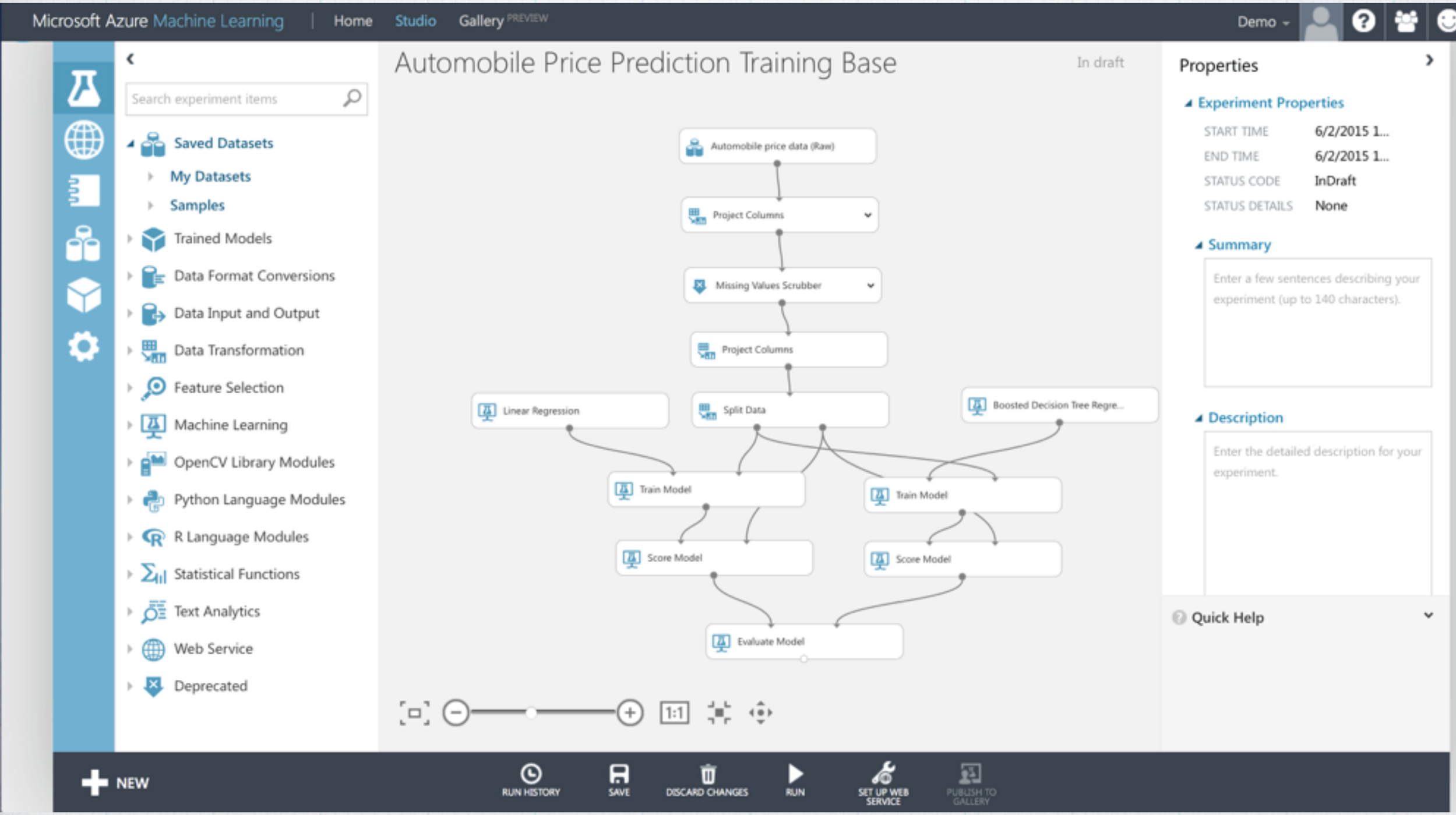
Download this poster: http://aka.ms/MLCheatSheet

Microsoft

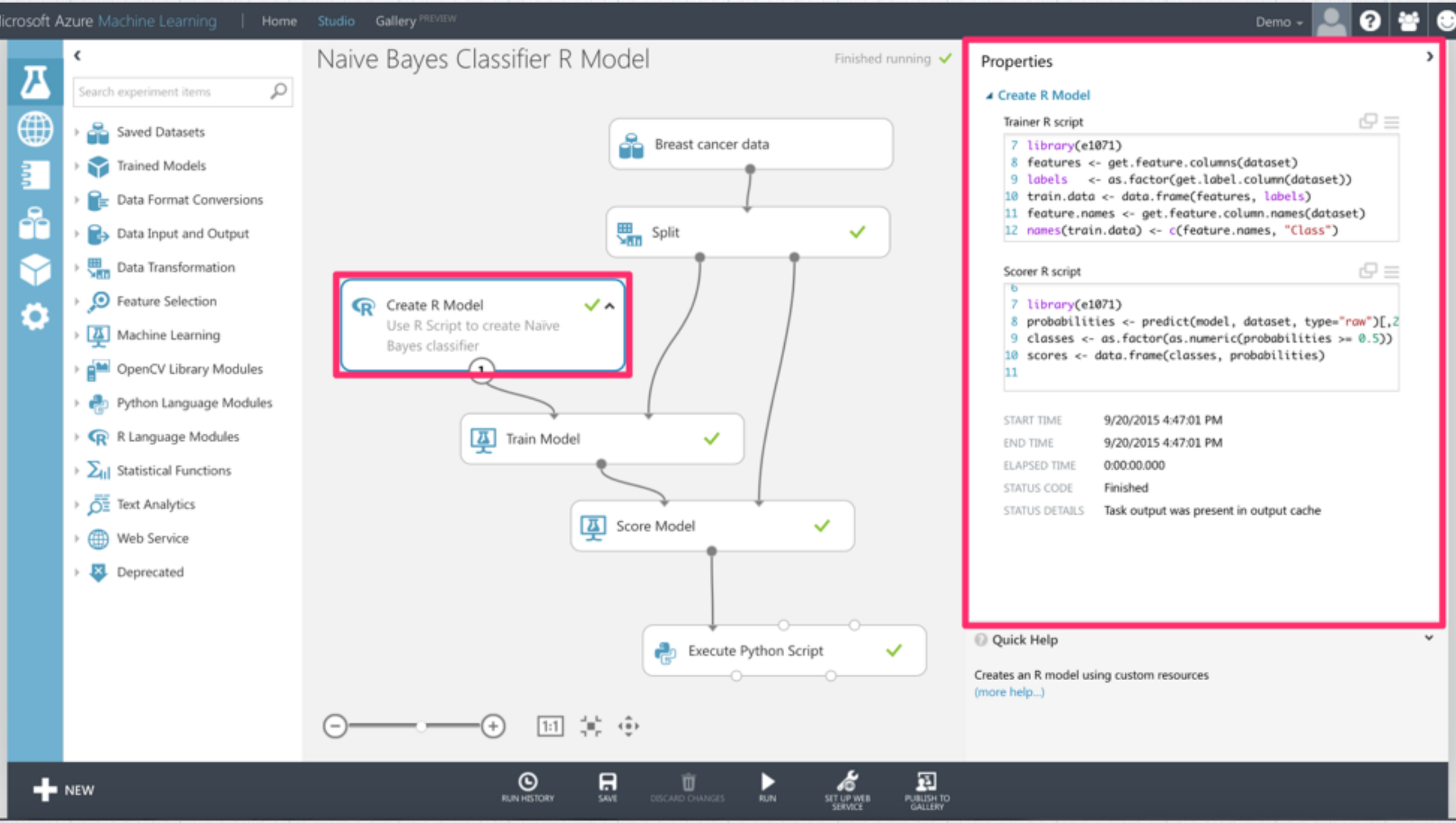
Custom Algorithms

* R
  * Execute R - Data Processing & Cleanup
  * Create R Model - Predictions
* Python
  * Execute Python - Data Processing & Cleanup

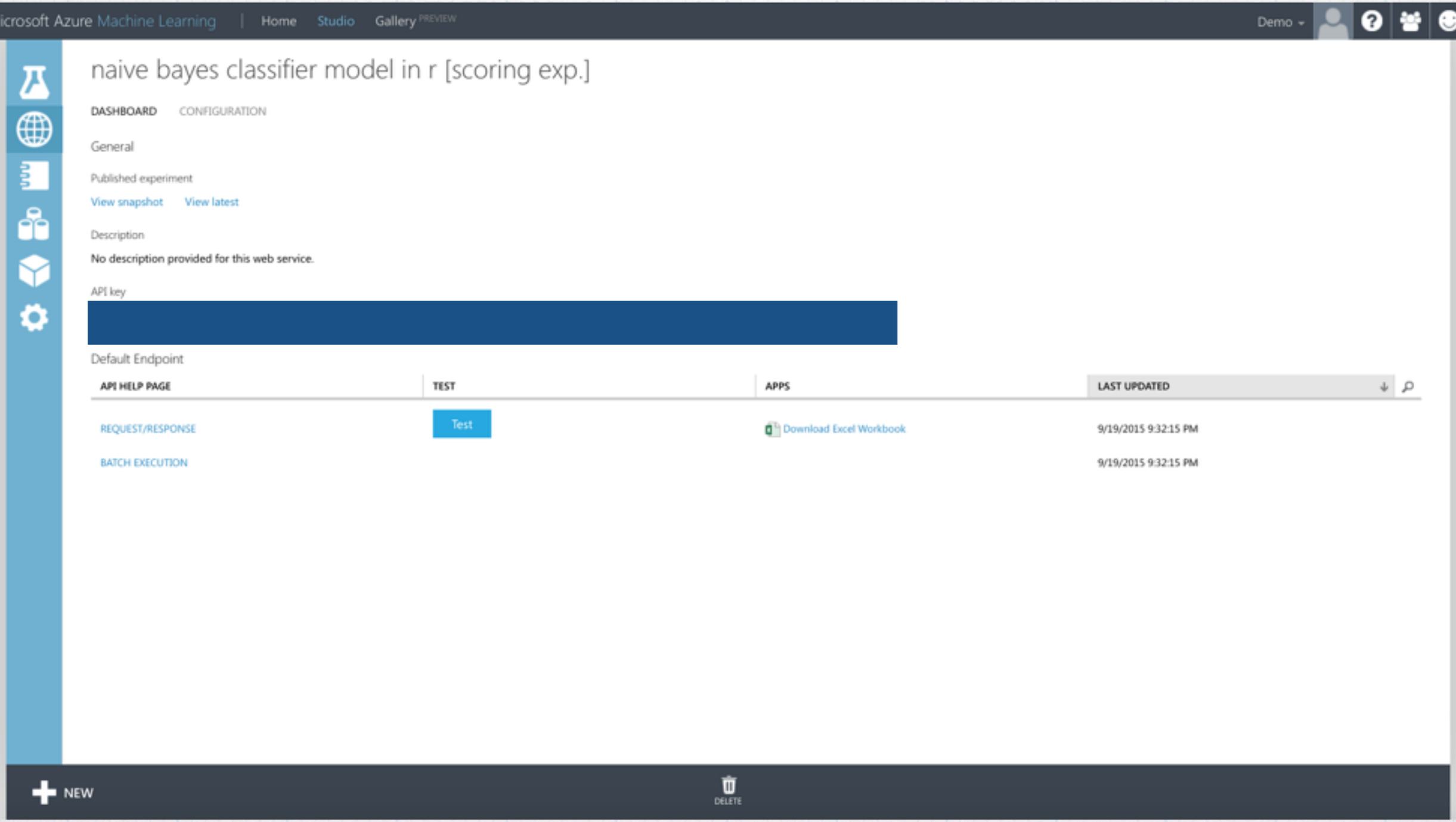
Demo: ML Studio



Demo: R Model

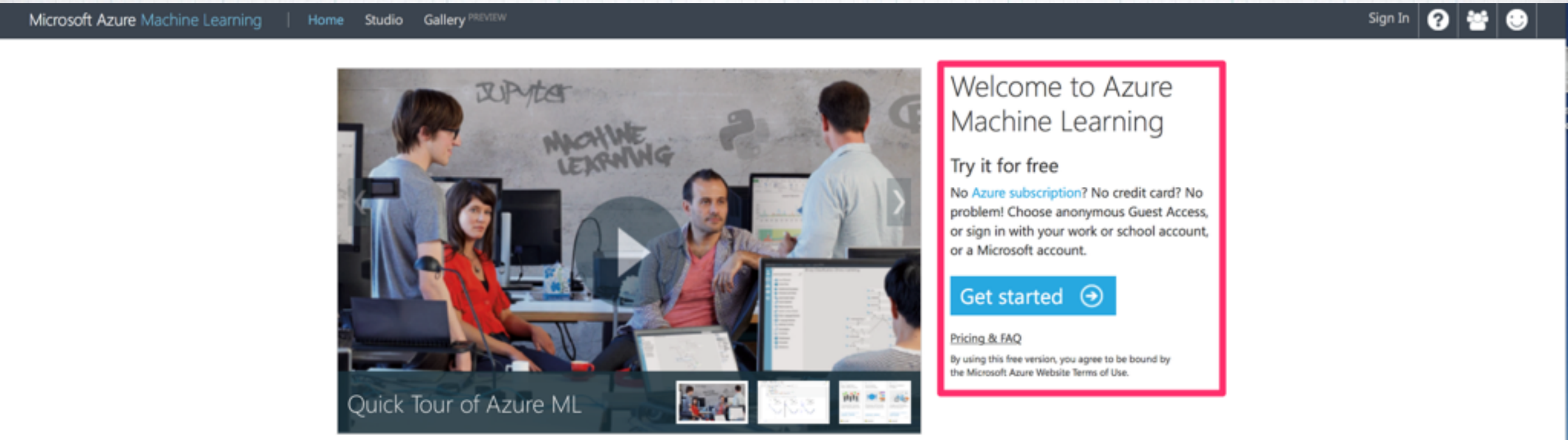

Microsoft Azure Machine Learning
Home
Studio
Gallery PREVIEW
Demo
Search experiment items
Saved Datasets
Trained Models
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Web Service
Deprecated
Naive Bayes Classifier R Model
Finished running
Breast cancer data
Split
Create R Model
Use R Script to create Naïve Bayes classifier
Train Model
Score Model
Execute Python Script
Properties
Create R Model
Trainer R script
7 library(e1071)
8 features <- get.feature.columns(dataset)
9 labels <- as.factor(get.label.column(dataset))
10 train.data <- data.frame(features, labels)
11 feature.names <- get.feature.column.names(dataset)
12 names(train.data) <- c(feature.names, "Class")
Scorer R script
7 library(e1071)
8 probabilities <- predict(model, dataset, type="raw")[,2
9 classes <- as.factor(as.numeric(probabilities >= 0.5))
10 scores <- data.frame(classes, probabilities)
START TIME 9/20/2015 4:47:01 PM
END TIME 9/20/2015 4:47:01 PM
ELAPSED TIME 0:00:00.000
STATUS CODE Finished
STATUS DETAILS Task output was present in output cache
Quick Help
Creates an R model using custom resources
(more help...)
NEW
RUN HISTORY
SAVE
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Demo: Publish a Service



Getting Started

https://studio.azureml.net



Webinars NEW!

Harness Predictive Customer Churn Models with Cortana Analytics Suite
Aired on August 18, 2015

Learn how to build a real-life churn model with Azure Machine Learning, make it enterprise-ready with Azure Data Factory, and deliver data insights with Power BI.

> Learn More

Introduction to Azure Data Factory
Aired on August 04, 2015

This session will help you jumpstart on understanding Data Factory capabilities, and the scenarios where Data Factory can be applied.

> Learn More

Energy Forecasting in Smart Grids using Azure Machine Learning
Aired on July 21, 2015

Learn about how we use Azure Stream Analytics to collect real time data; use Azure SQL to store data; use Azure Machine Learning to build a forecast model; use Azure Data Factory to automate the model and use PowerBI to visualize results on a dashboard.

> Learn More

Upcoming Webinars

Free vs. Standard

Pricing Details

Machine Learning is offered in two tiers: Free and Standard.

Features by tier are compared in the table below:

	FREE	STANDARD
Authentication	Microsoft account (does not require an Azure subscription or a credit card)	Requires Azure subscription
Max Number of Modules per Experiment	100	Unlimited
Max Experiment Duration	1 hour	None
Max Storage Space	10 GB	Unlimited - BYO
Execution / Performance	Single node	Multiple nodes
Staging Web API	Yes (Throttled performance)	Yes (Selectable performance)
Production Web API	No	Yes
SLA	No	Yes

Questions ?

Contact Info:

- mchenry19@gmail.com
- @CAMCHENRY
- http://cmchenry.com
- http://www.linkedin.com/in/cmchenry
- https://plus.google.com/+chrismchenry

Backups

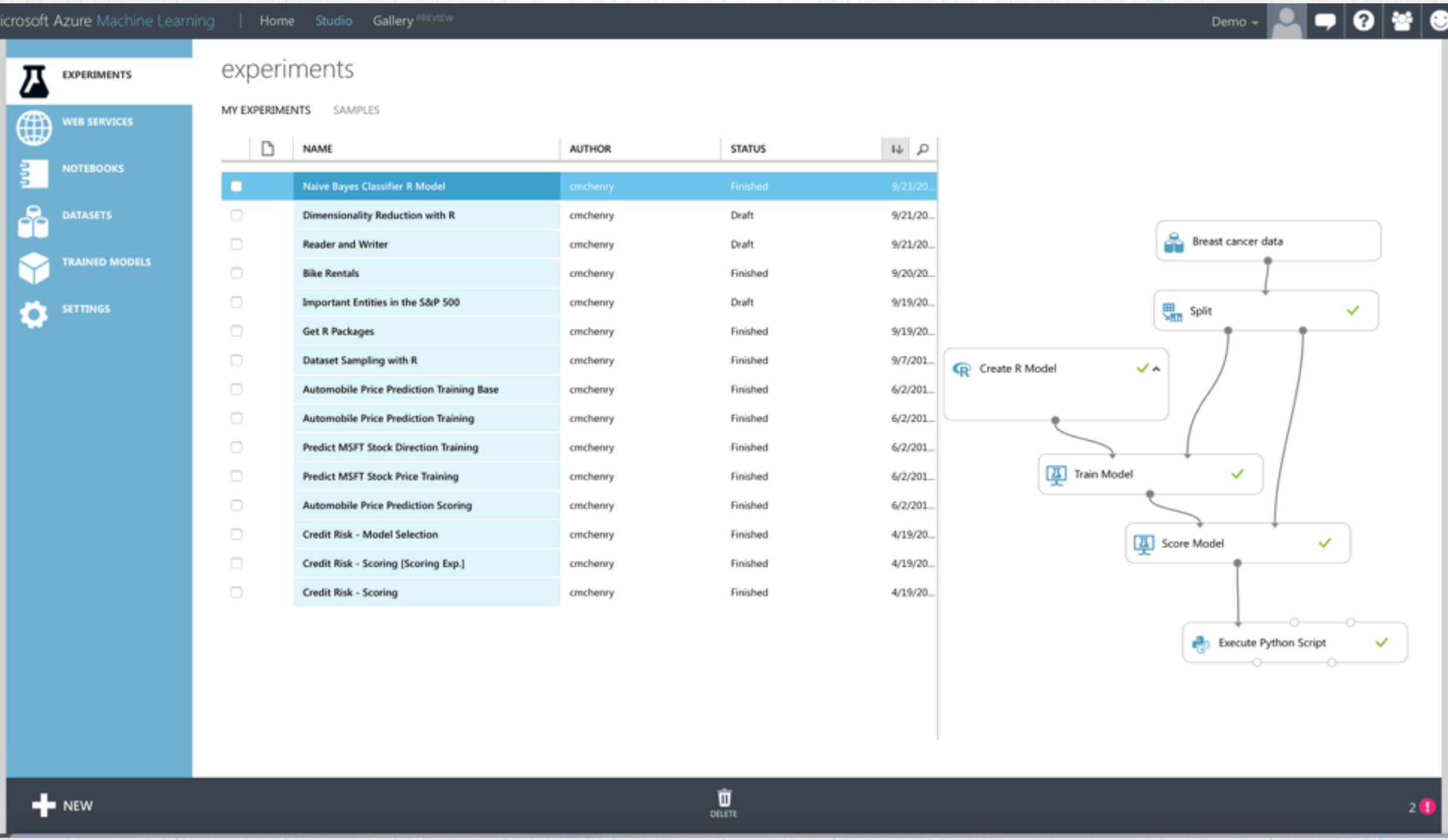
Pricing

Standard tier pricing

ML Seat Subscription		
	Monthly Fee	$9.99/ Seat/ Month
ML Studio Usage		
	Hourly	$1/Studio Experiment Hour
ML API Usage		
	Hourly	$2/Production API Compute Hour
	Transactions	$0.50/1,000 Production API Transactions

Hourly charges only apply to active use of the service. Where multiple meters are present they are applied concurrently.

Experiments



New Experiment

NEW

DATASET
MODULE
EXPERIMENT
NOTEBOOK PREVIEW

Search experiment templates

Microsoft Samples

VIEW MORE IN GALLERY

Blank Experiment

Experiment Tutorial

Sample 1: Download dataset from UCI: Adult 2 class dataset

Sample 2: Dataset Processing and Analysis: Auto Imports Regression

Sample 3: Cross Validation for Binary Classification: Adult

Two-Class Averaged Perceptron, Two-Class Logistic Regression,

Sample 4: Cross Validation for Regression: Auto Imports Dataset

Linear Regression, Boosted Decision Tree Regression, Poisson

Sample 5: Train, Test, Evaluate for Binary Classification: Adult

Two-Class Boosted Decision Tree

Sample 6: Train, Test, Evaluate for Regression: Auto Imports Dataset

Poisson Regression, Decision Forest Regression

Sample 7: Train, Test, Evaluate for Multiclass Classification: Letter

Multiclass Decision Jungle, One-vs-All Multiclass, Two-Class

Sample 8: Apply SQL transformation

Sample 9: Split, partition

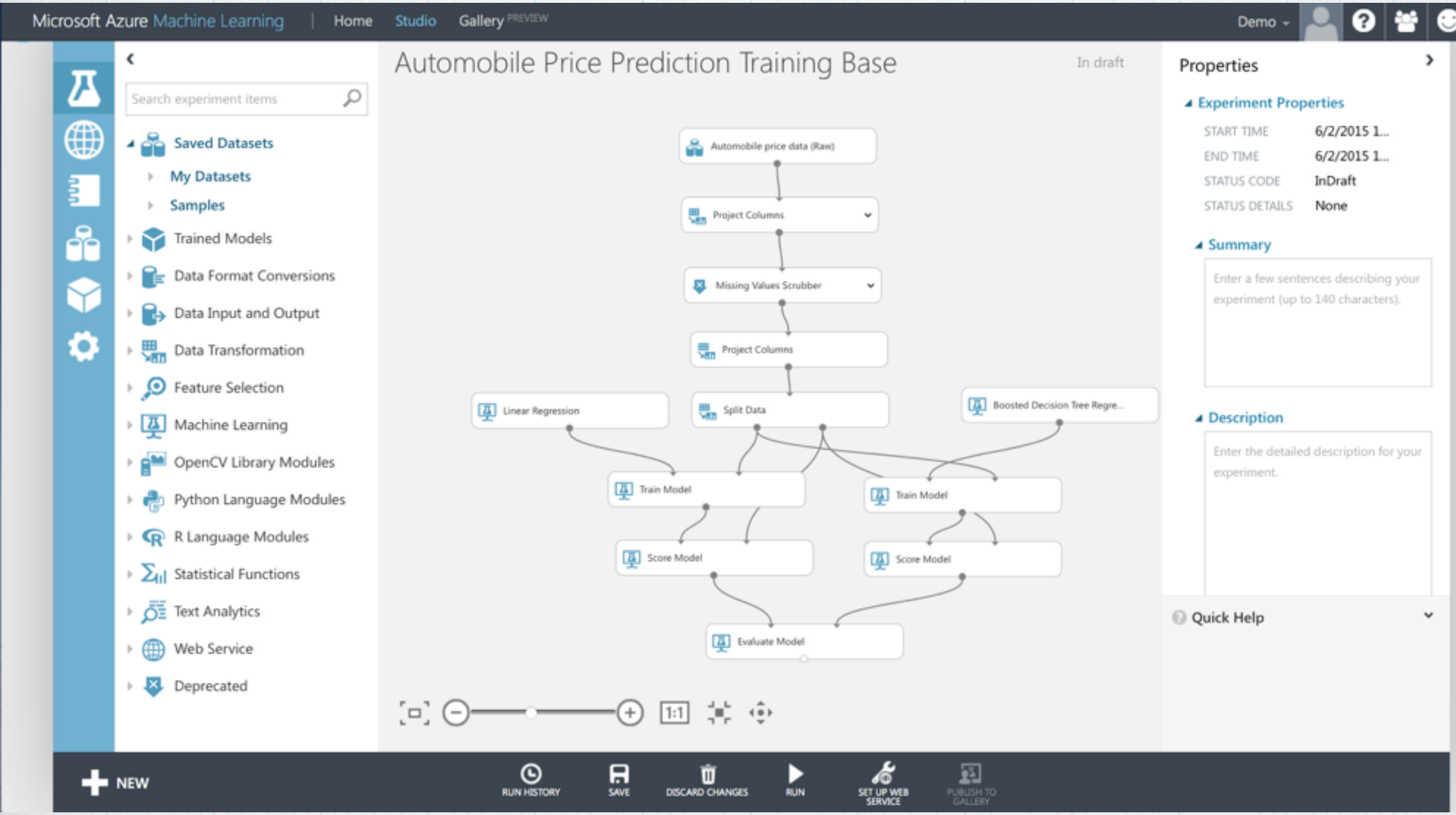
Anomaly Detection:

Binary Classfication:

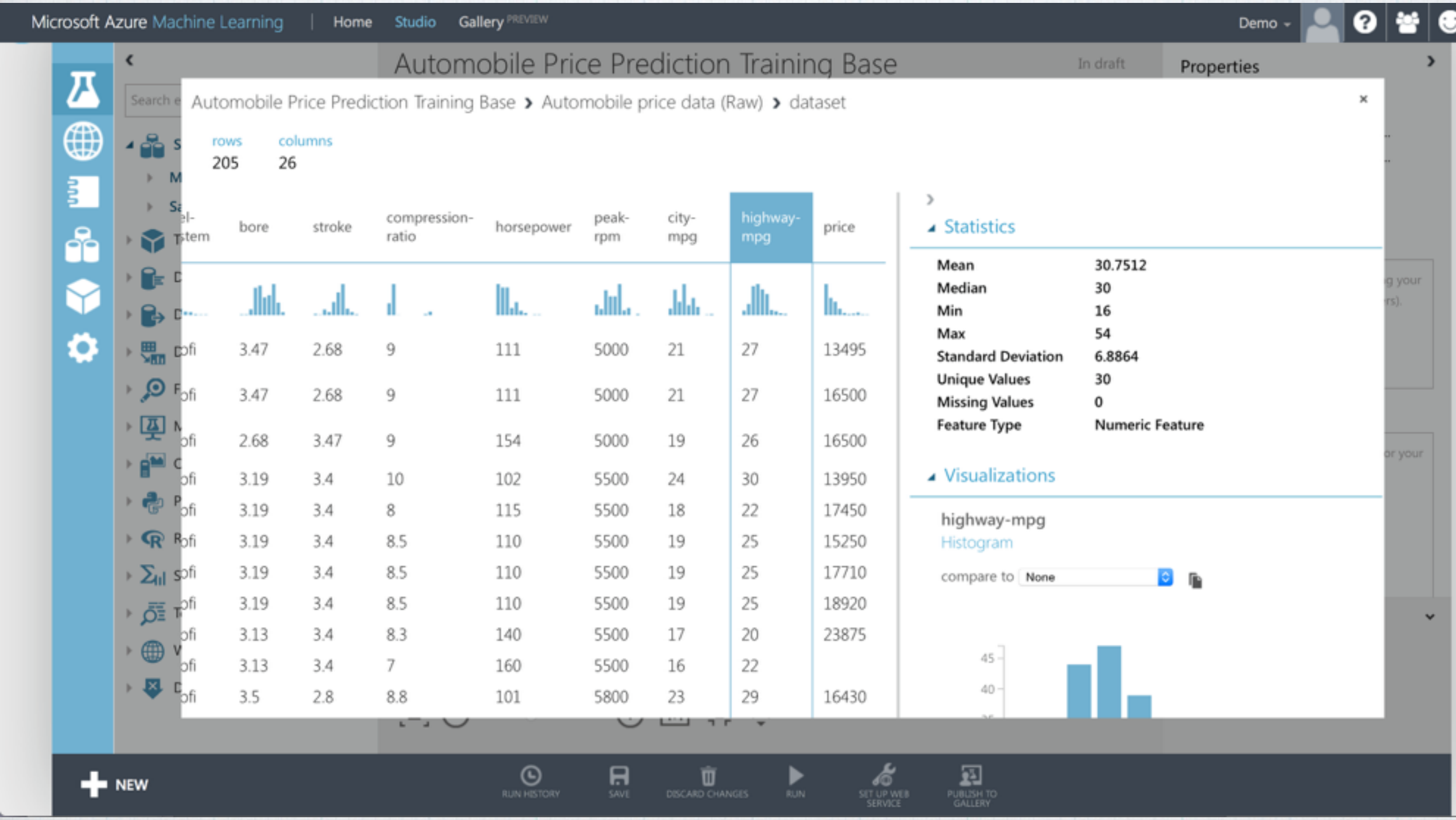
Binary Classification:

Binary Classification:

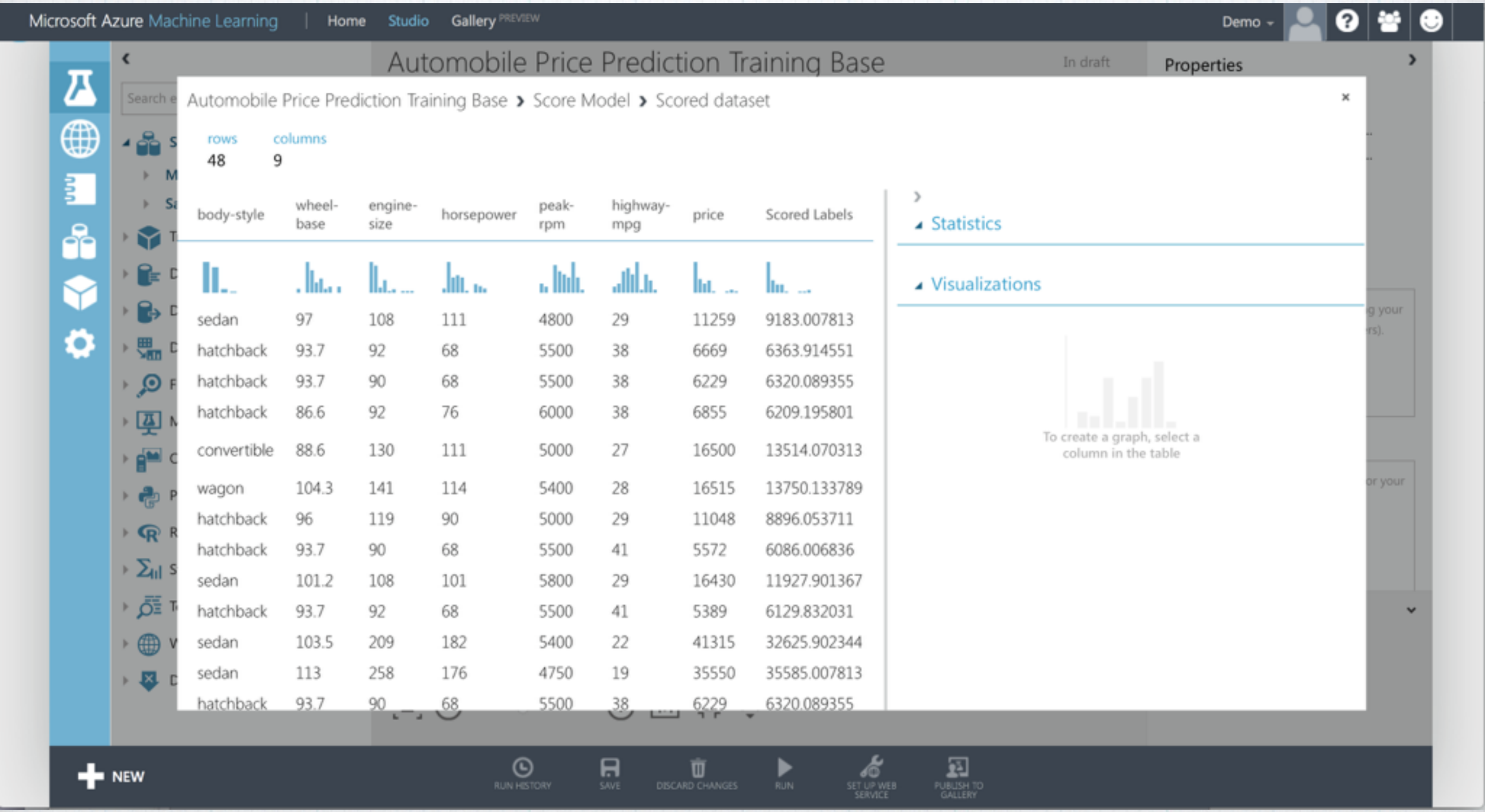
Regression Experiment



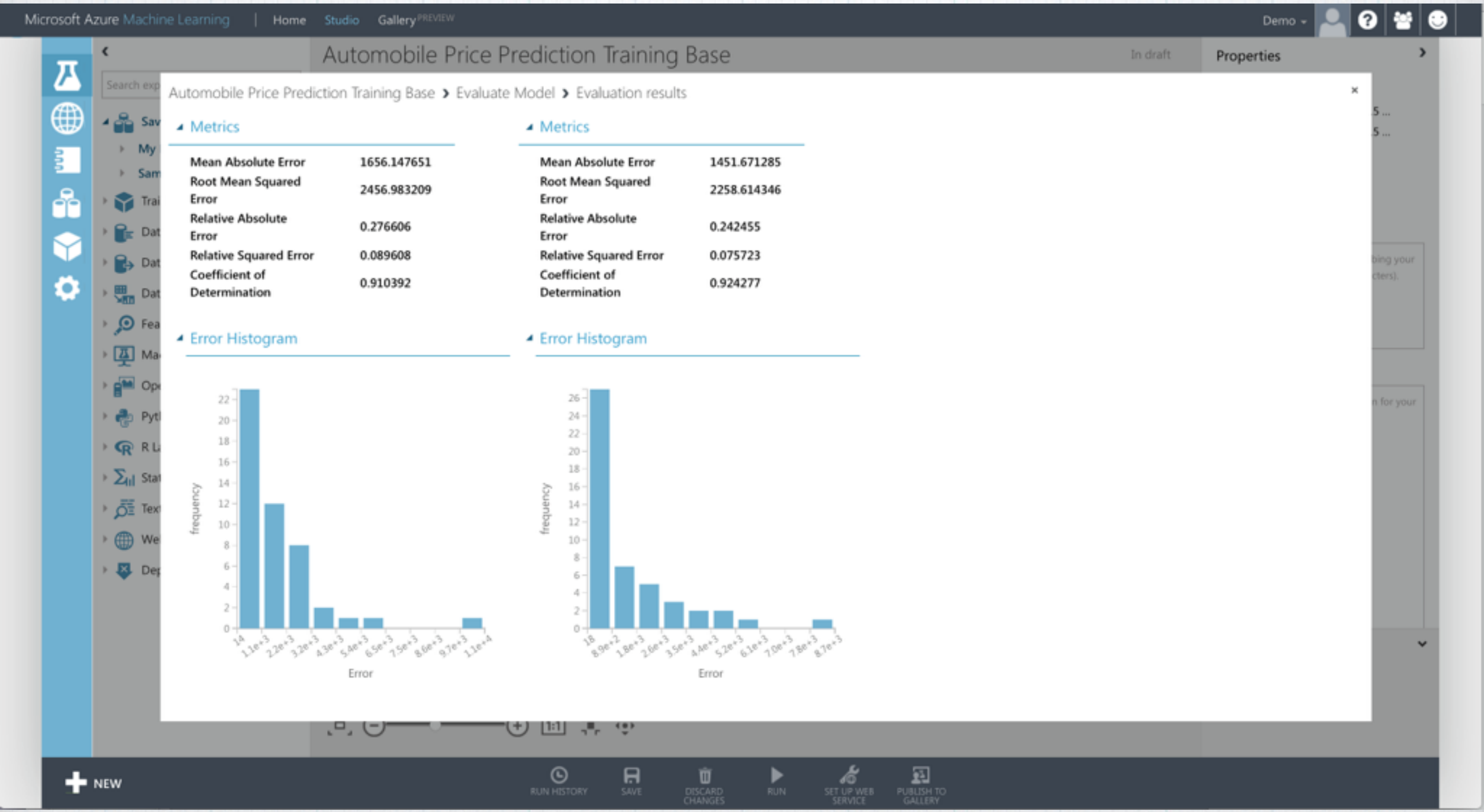
Visualize Data



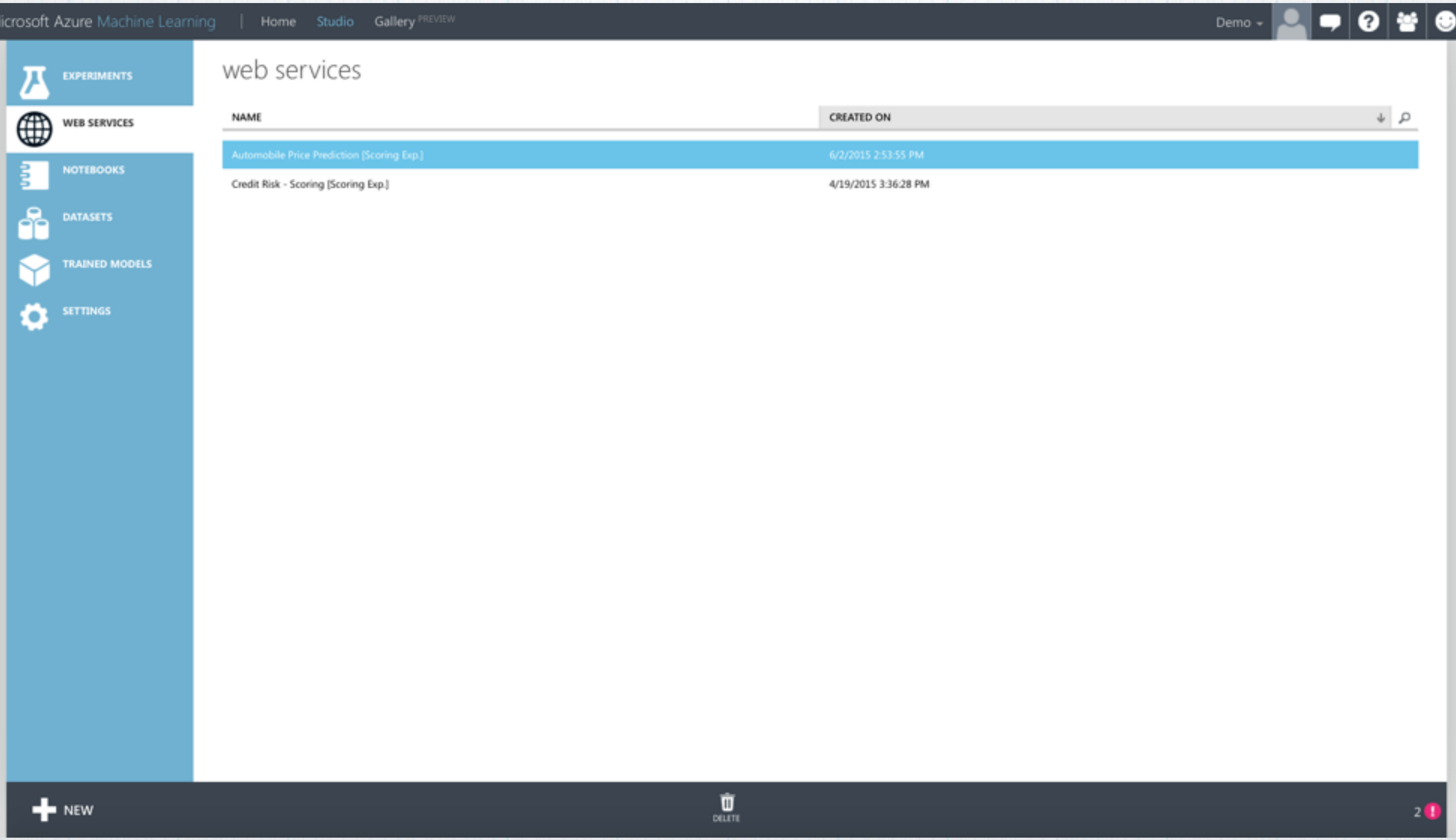
Score a Model



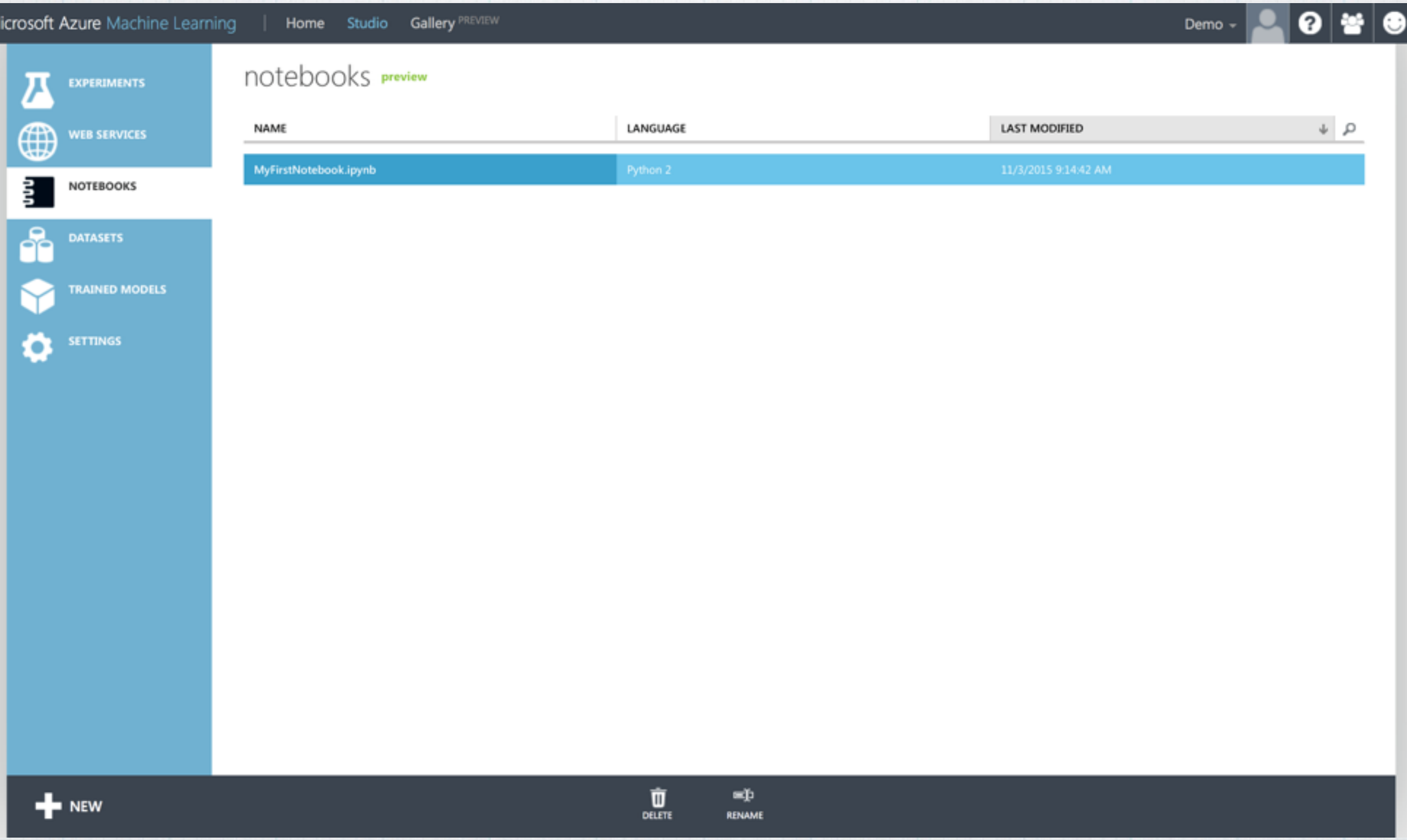
Evaluate Models



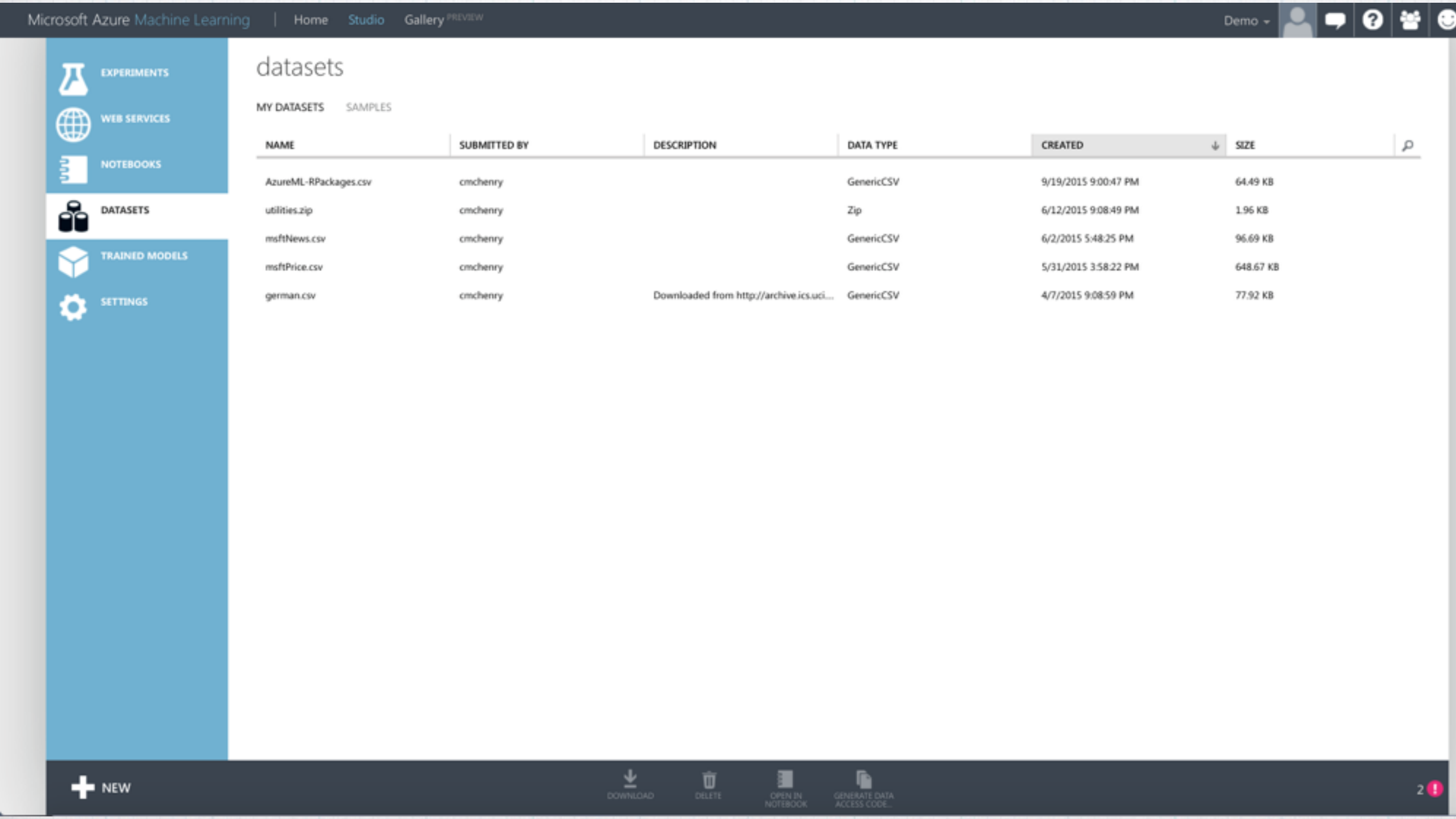
Web Services



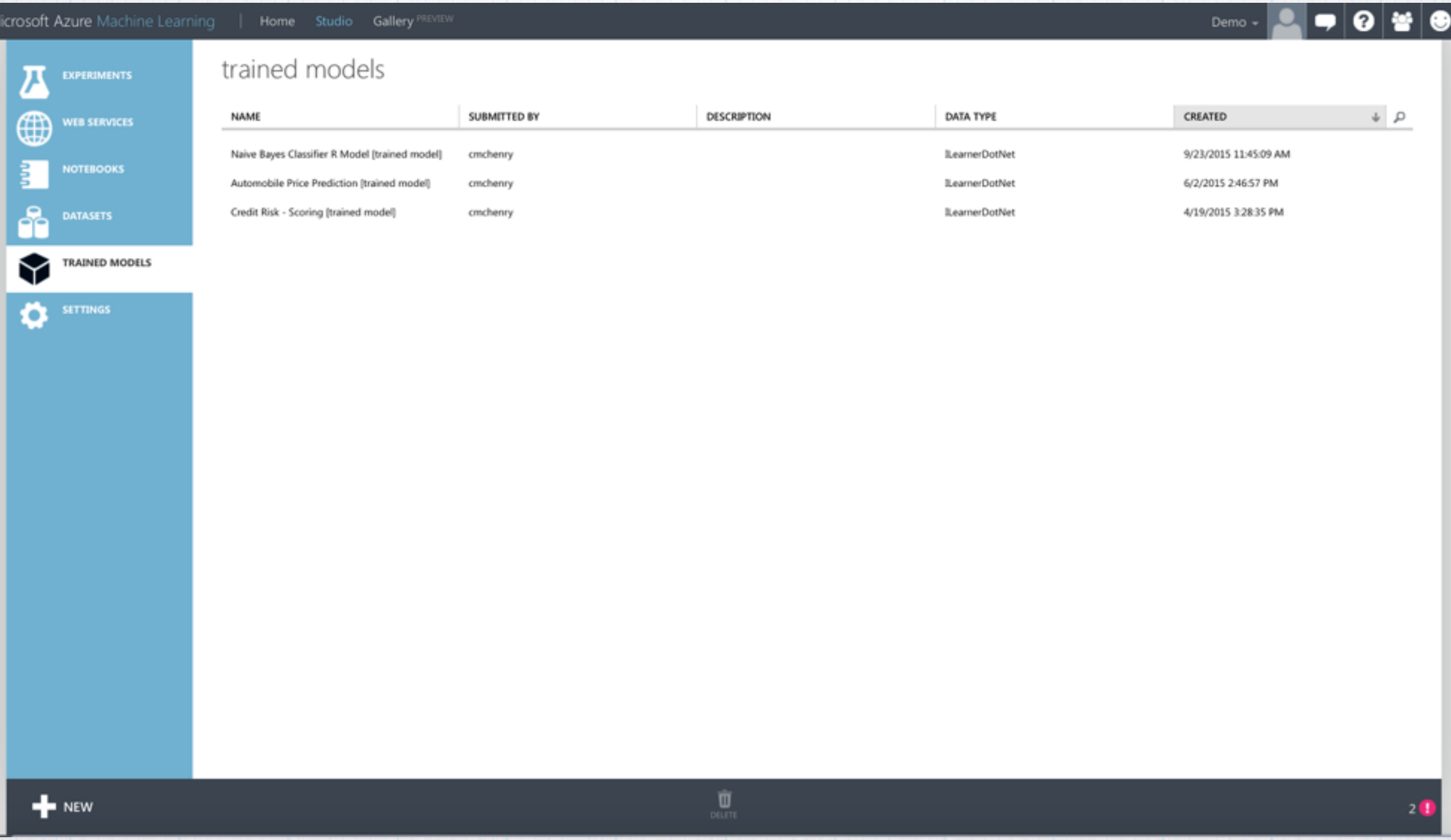
Notebooks


Microsoft Azure Machine Learning
Home
Studio
Gallery PREVIEW
Demo
EXPERIMENTS
WEB SERVICES
NOTEBOOKS
DATASETS
TRAINED MODELS
SETTINGS
notebooks preview
NAME
LANGUAGE
LAST MODIFIED
MyFirstNotebook.ipynb
Python 2
11/3/2015 9:14:42 AM
NEW
DELETE
RENAME

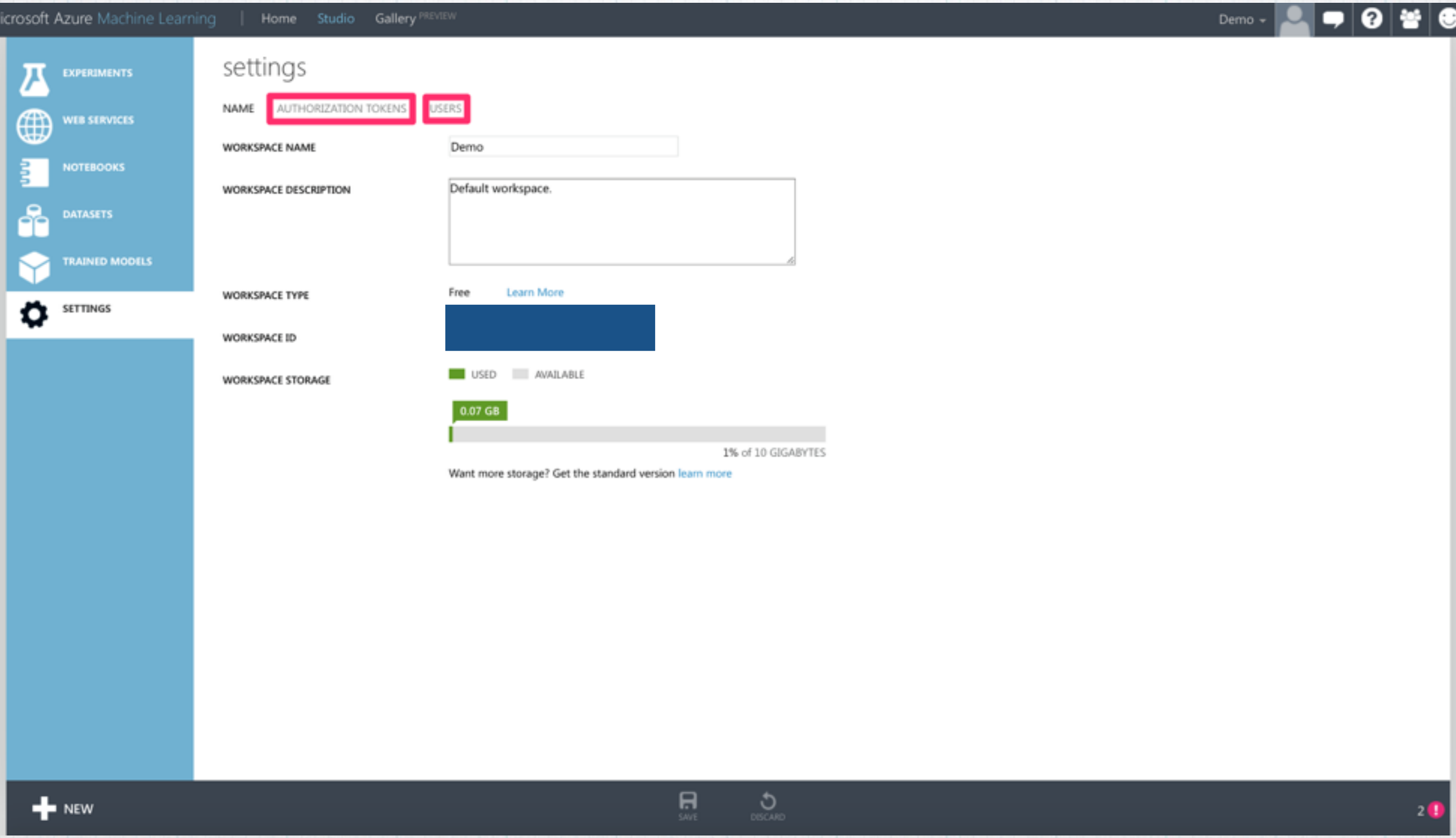
Datasets


Microsoft Azure Machine Learning
Home
Studio
Gallery PREVIEW
Demo
EXPERIMENTS
WEB SERVICES
NOTEBOOKS
DATASETS
TRAINED MODELS
SETTINGS
datasets
MY DATASETS
SAMPLES
NAME
SUBMITTED BY
DESCRIPTION
DATA TYPE
CREATED
SIZE
AzureML-RPackages.csv cmchenry GenericCSV 9/19/2015 9:00:47 PM 64.49 KB
utilities.zip cmchenry Zip 6/12/2015 9:08:49 PM 1.96 KB
msftNews.csv cmchenry GenericCSV 6/2/2015 5:48:25 PM 96.69 KB
msftPrice.csv cmchenry GenericCSV 5/31/2015 3:58:22 PM 648.67 KB
german.csv cmchenry Downloaded from http://archive.ics.uci.... GenericCSV 4/7/2015 9:08:59 PM 77.92 KB
NEW
DOWNLOAD
DELETE
OPEN IN NOTEBOOK
GENERATE DATA ACCESS CODE...
2

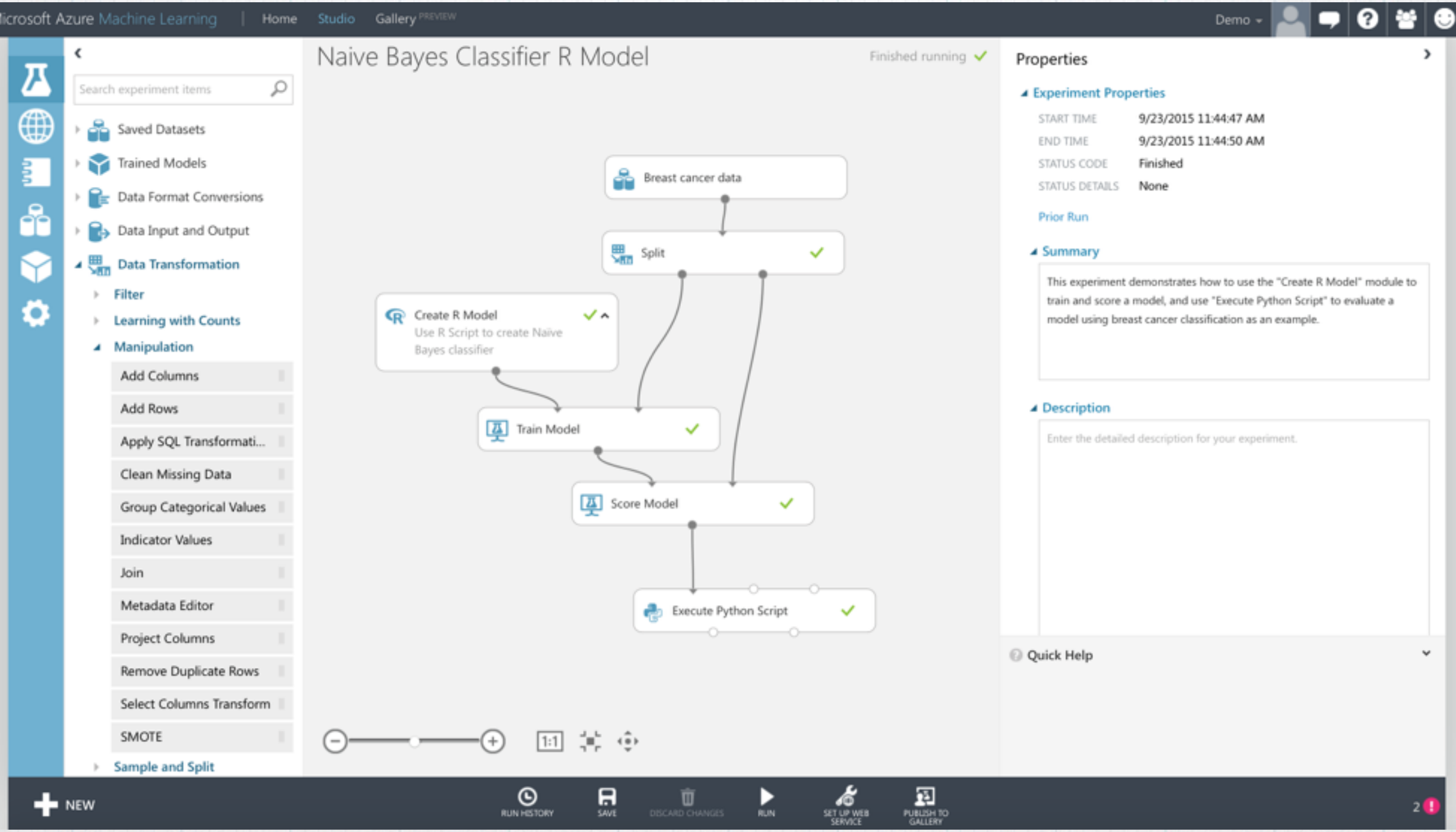
Trained Models



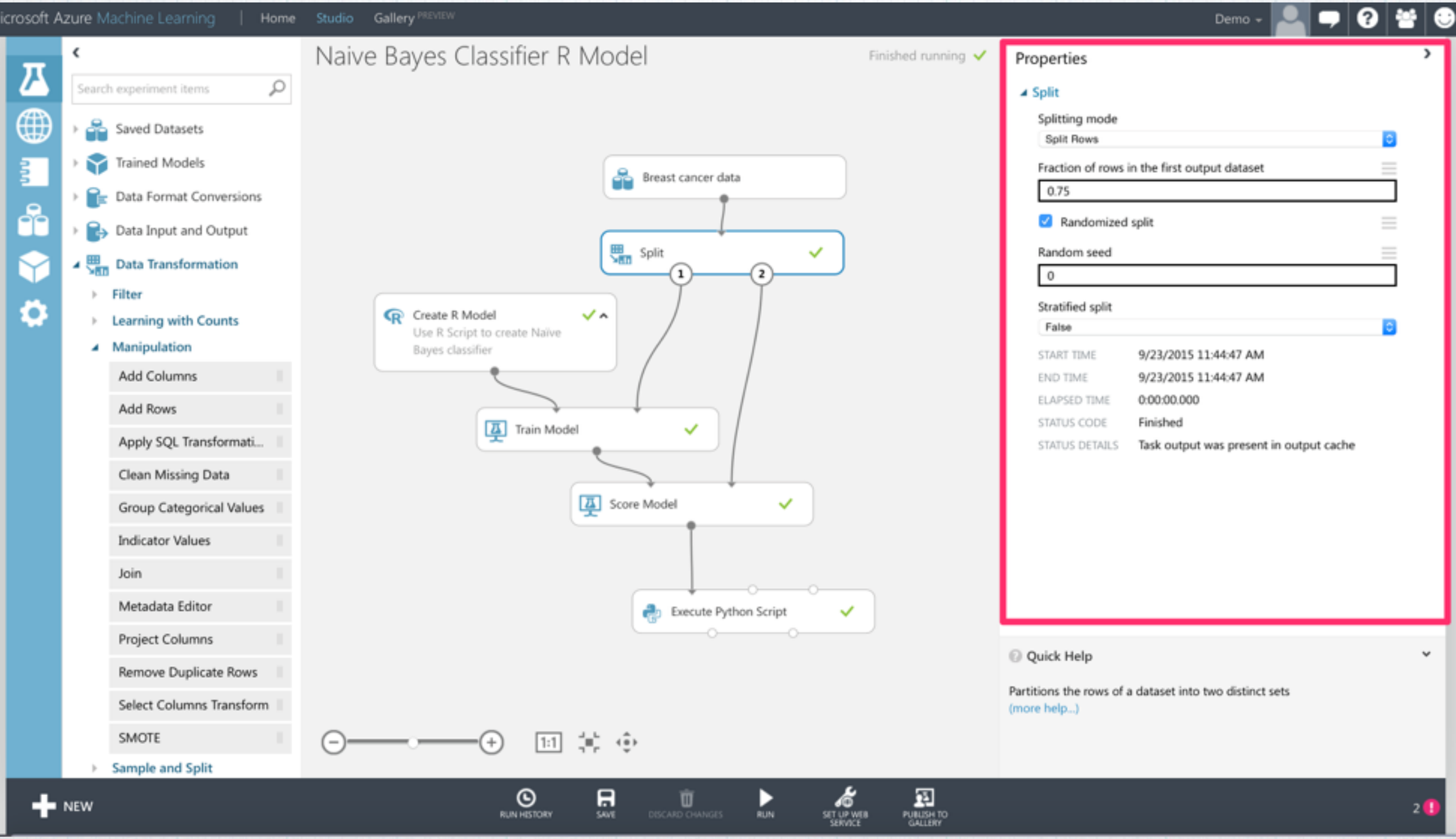
Settings


Microsoft Azure Machine Learning
Home
Studio
Gallery PREVIEW
Demo
EXPERIMENTS
WEB SERVICES
NOTEBOOKS
DATASETS
TRAINED MODELS
SETTINGS
settings
NAME
AUTHORIZATION TOKENS
USERS
WORKSPACE NAME
Demo
WORKSPACE DESCRIPTION
Default workspace.
WORKSPACE TYPE
Free
Learn More
WORKSPACE ID
WORKSPACE STORAGE
USED
AVAILABLE
0.07 GB
1% of 10 GIGABYTES
Want more storage? Get the standard version learn more
NEW
SAVE
DISCARD
2

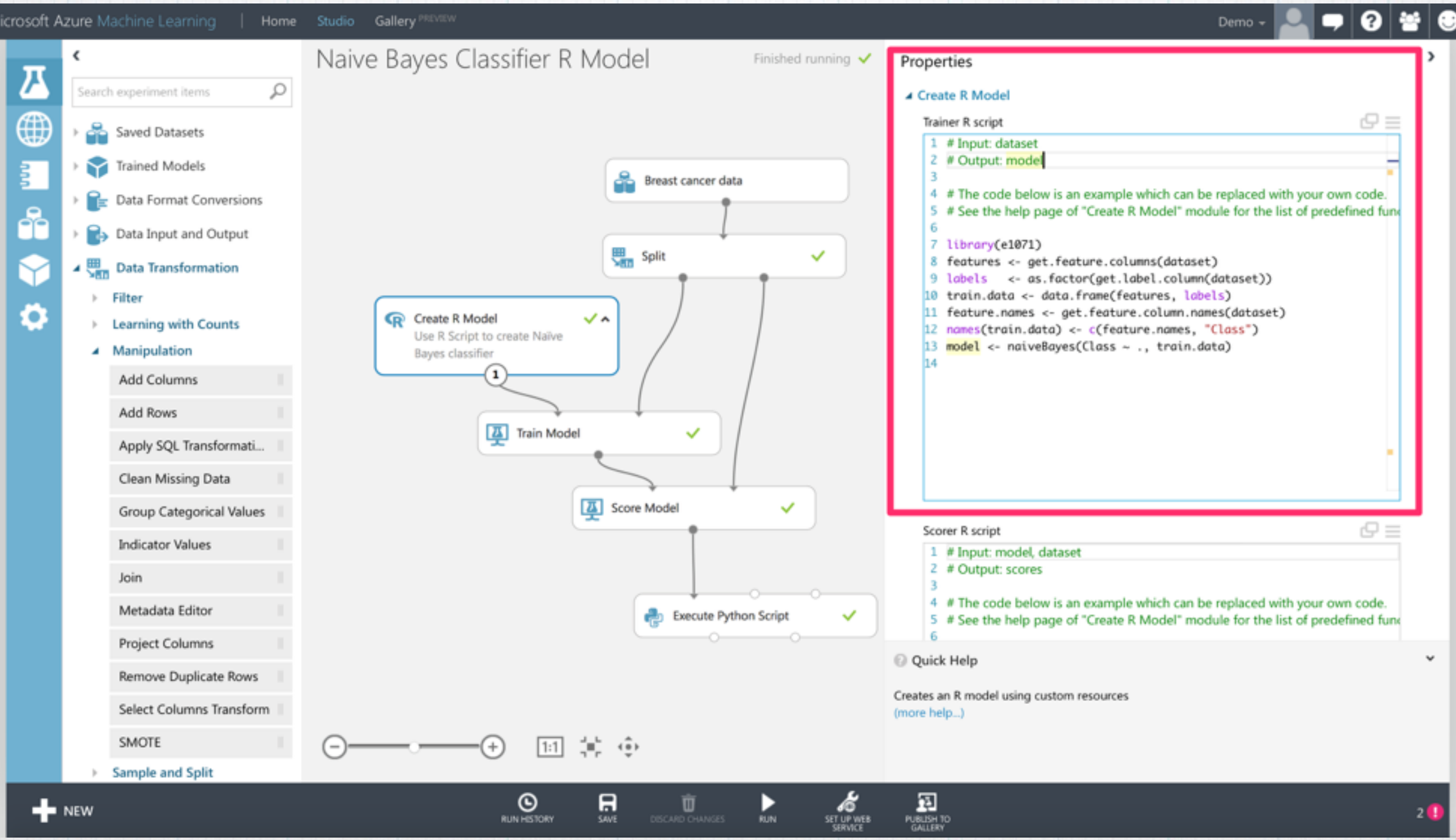
R Model


Microsoft Azure Machine Learning
Home
Studio
Gallery PREVIEW
Demo
Naive Bayes Classifier R Model
Finished running
Search experiment items
Saved Datasets
Trained Models
Data Format Conversions
Data Input and Output
Data Transformation
Filter
Learning with Counts
Manipulation
Add Columns
Add Rows
Apply SQL Transformati...
Clean Missing Data
Group Categorical Values
Indicator Values
Join
Metadata Editor
Project Columns
Remove Duplicate Rows
Select Columns Transform
SMOTE
Sample and Split
Breast cancer data
Split
Create R Model
Use R Script to create Naive Bayes classifier
Train Model
Score Model
Execute Python Script
1:1
Properties
Experiment Properties
START TIME 9/23/2015 11:44:47 AM
END TIME 9/23/2015 11:44:50 AM
STATUS CODE Finished
STATUS DETAILS None
Prior Run
Summary
This experiment demonstrates how to use the "Create R Model" module to train and score a model, and use "Execute Python Script" to evaluate a model using breast cancer classification as an example.
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY
2

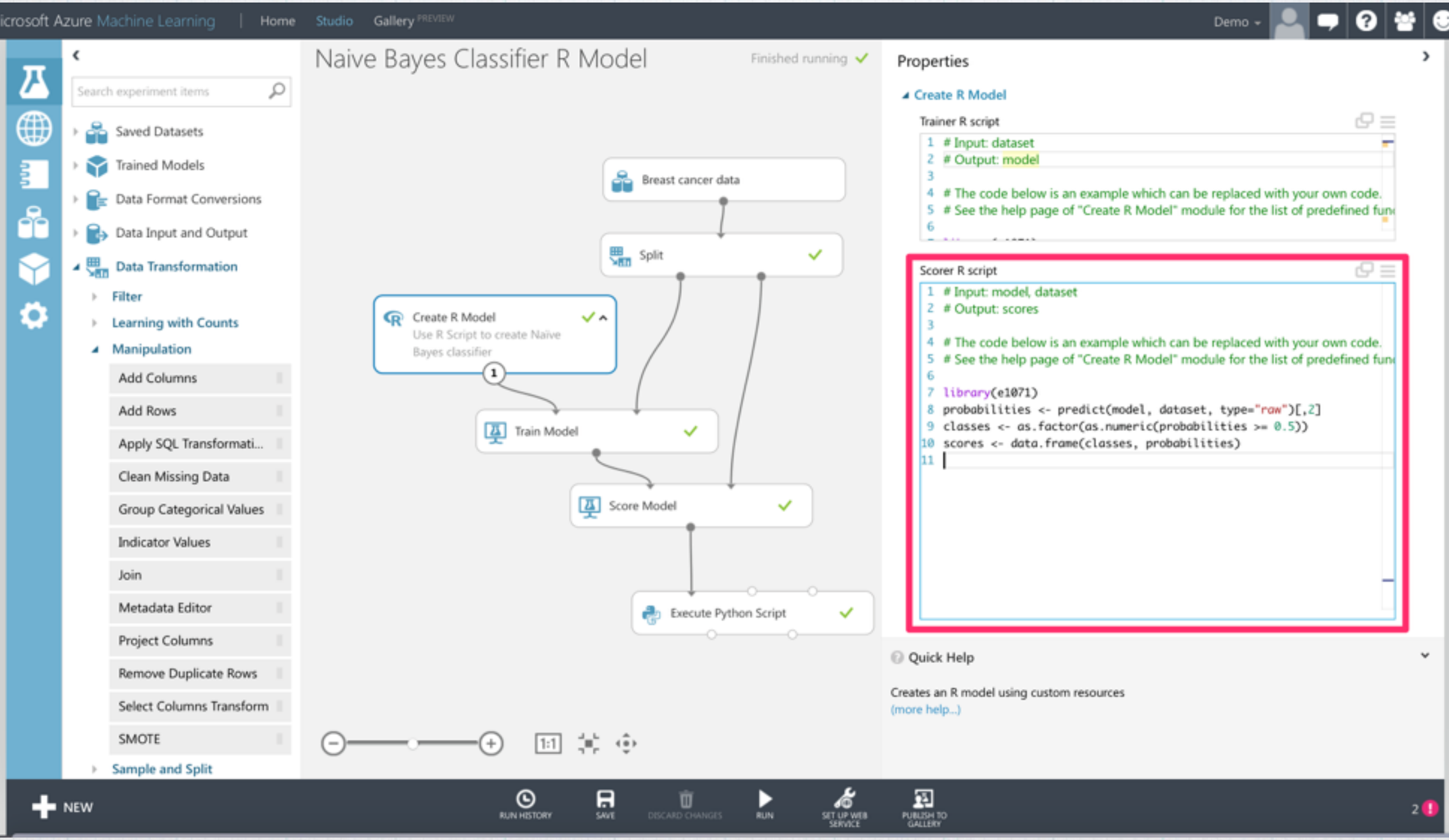
Split Module



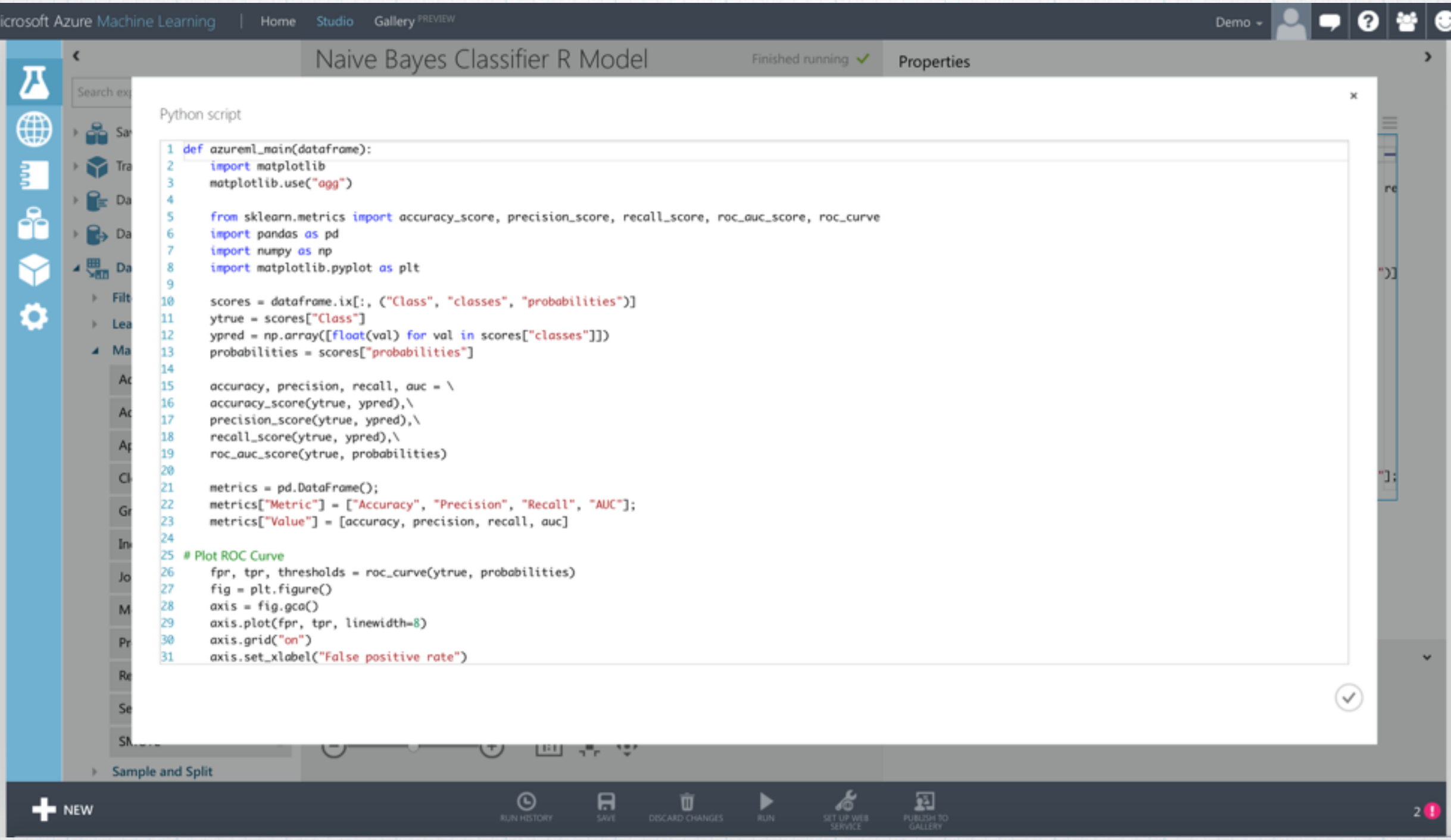
R Training Model



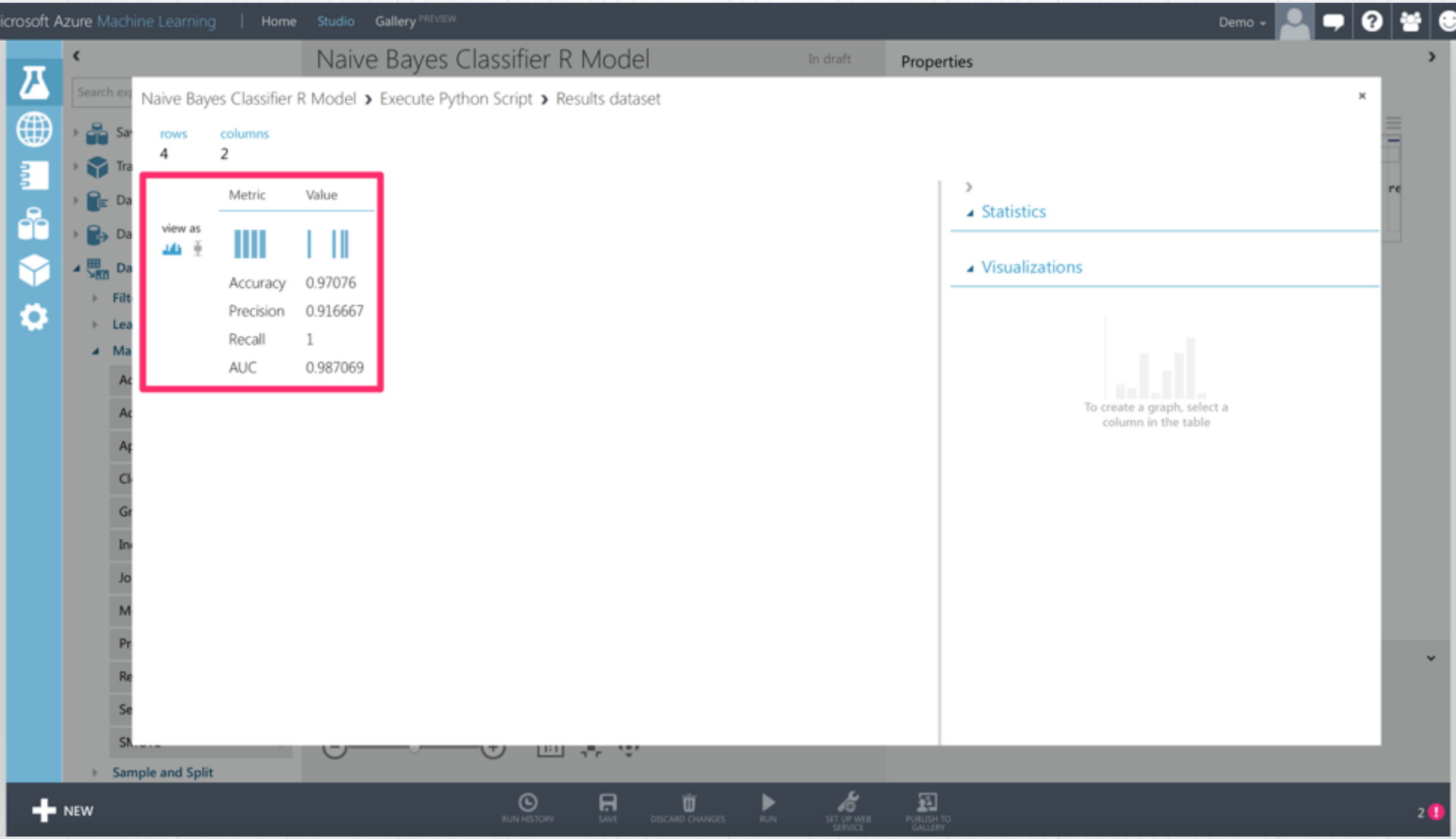
R Scoring Model



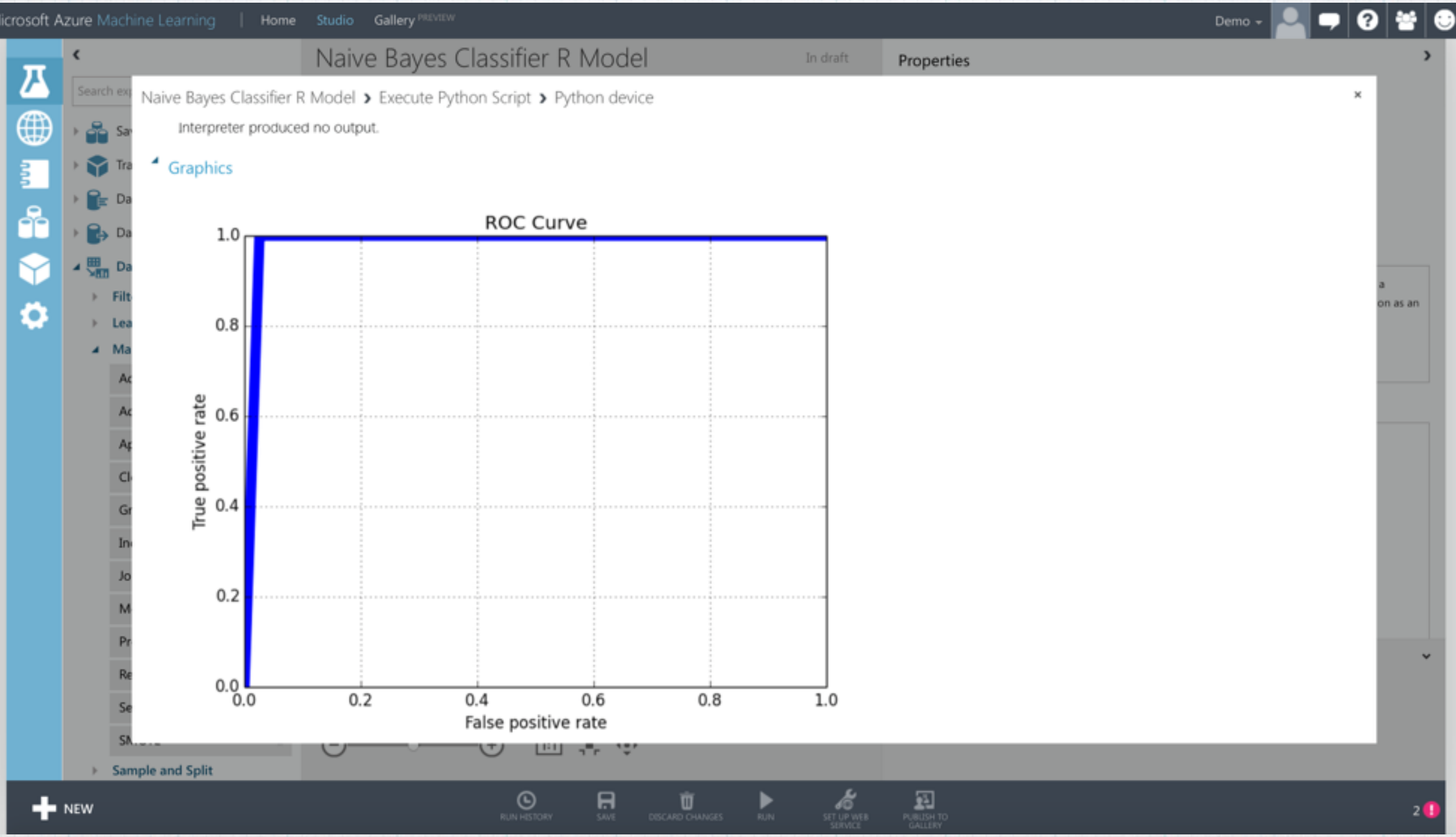
Python Eval of R Model



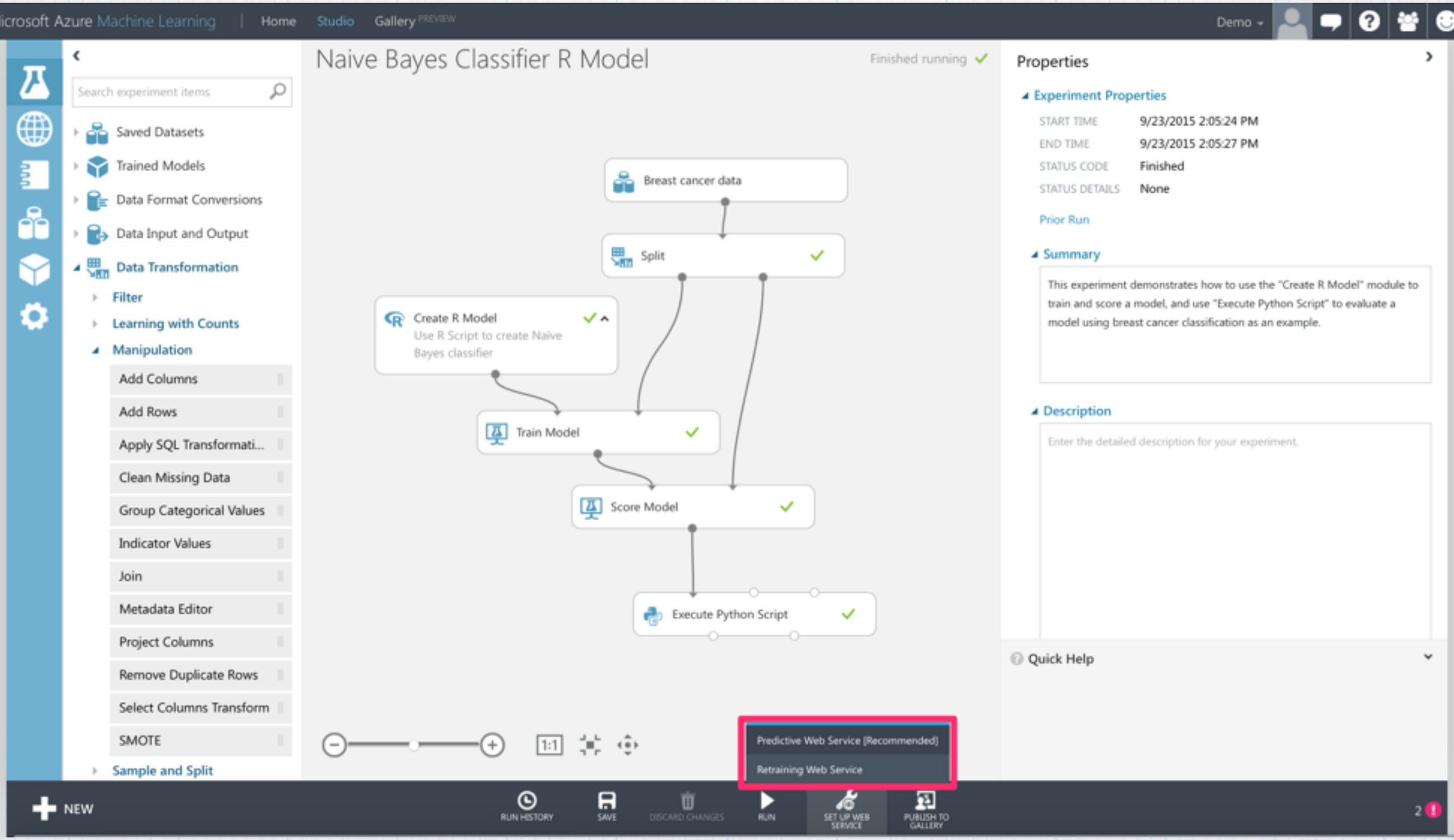
R Model Evaluate



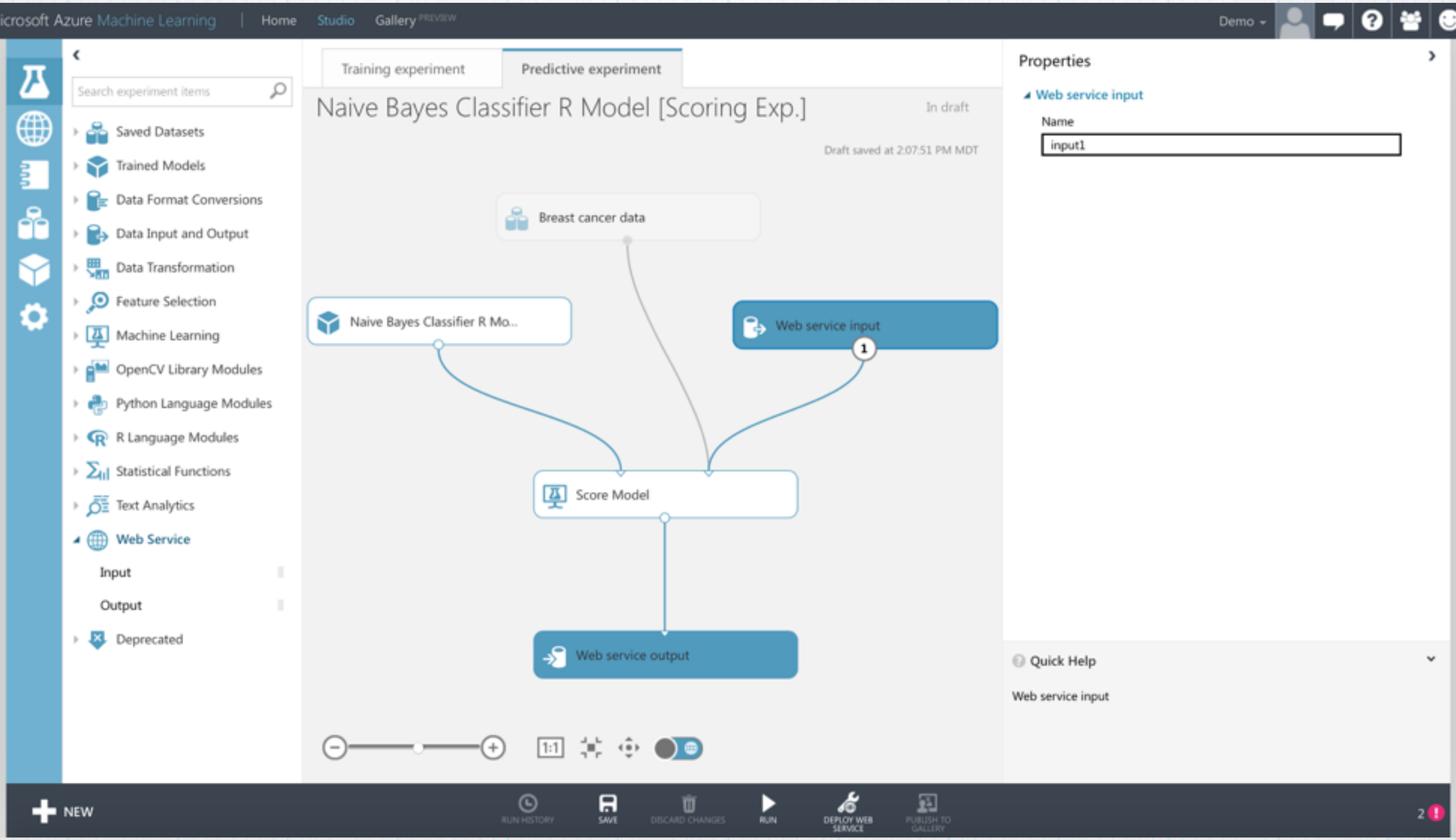
R Model ROC Curve



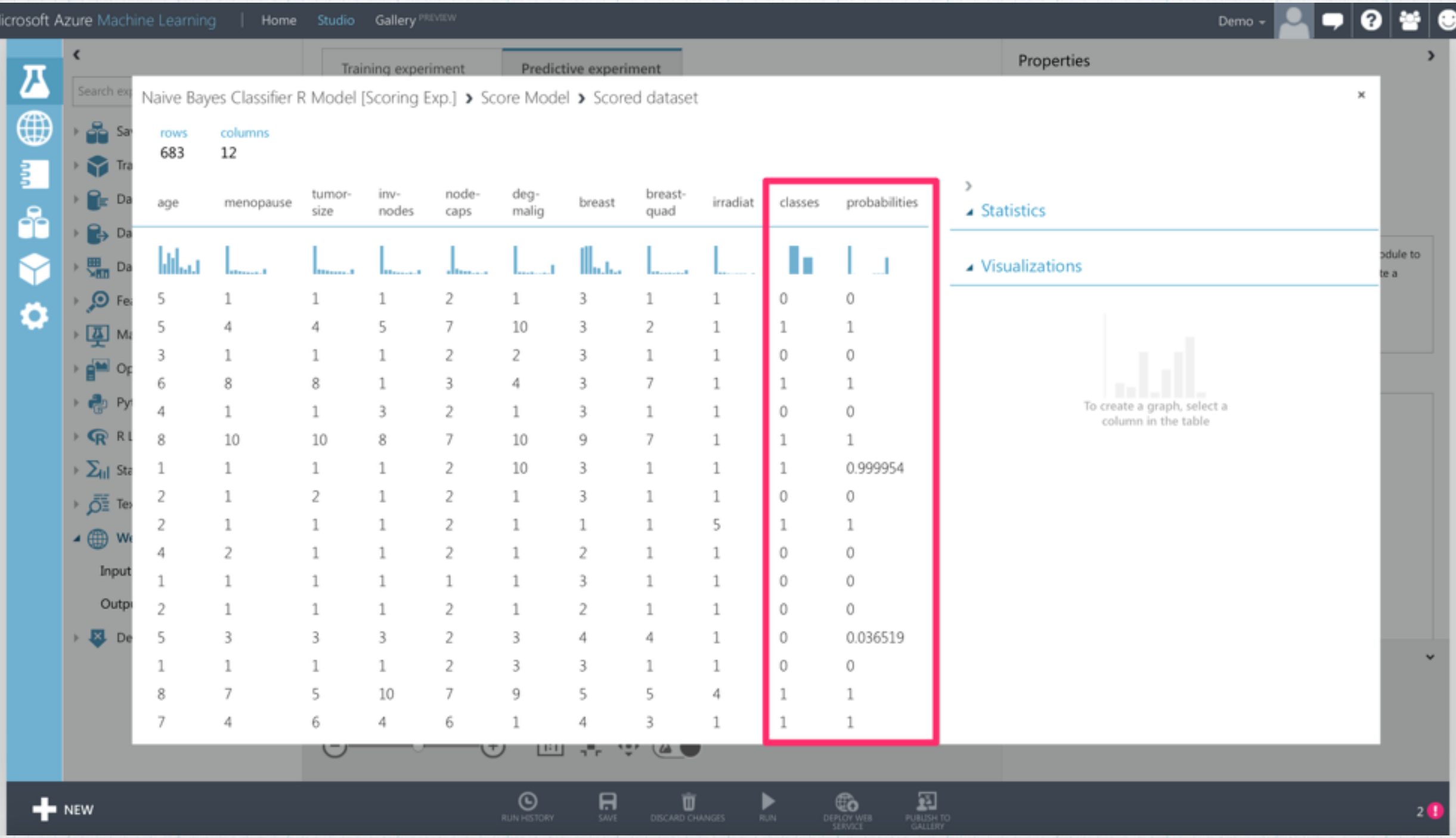
R Model Create Scoring



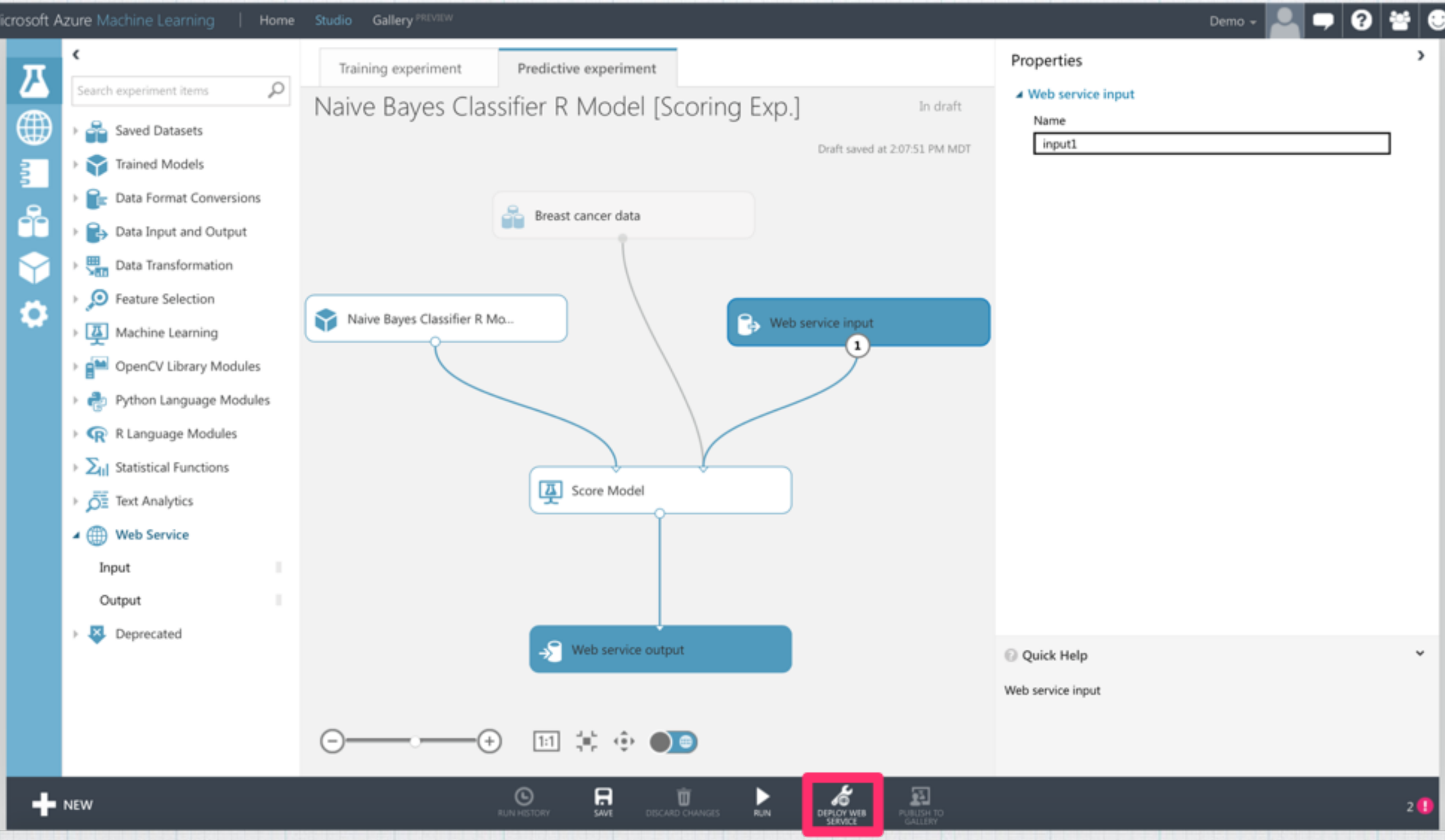
R Model Scoring Exp



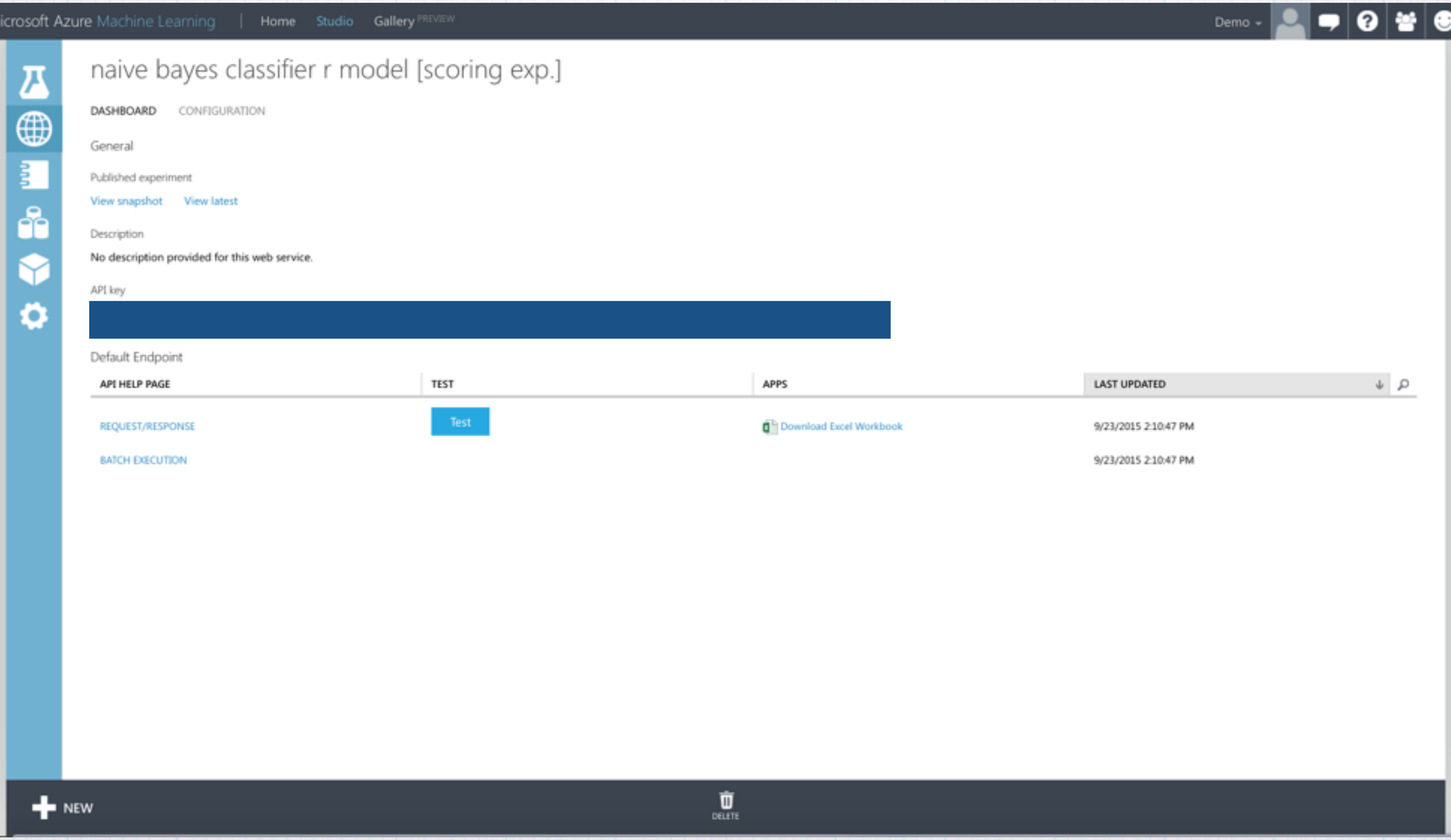
R Model Scoring Results



R Model Publish



R Model Service



R Model Req/Resp

Request Response API Documentation for Naive Bayes Classifier R Model [Scoring Exp.]

Updated: 09/23/2015 20:10

No description provided for this web service.

- Previous version of this API
- Submit a request
- Input Parameters
- Output Parameters
- ASP.Net App Template for RRS
- Sample Code

OData Endpoint Address

Request

Method	Request URI	HTTP Version

*Note: You may omit the **details** parameter from the query string. This would cause **ColumnTypes** to be omitted from the output*

Request Headers

Request Header	Description
Authorization:Bearer abc123	Required. Pass the API Key here. Obtain this key from the publisher of the API.
Content-Length	Required. The length of the content body.

R Model Sample Code

classes	Categorical	0, 1
probabilities	Numeric	

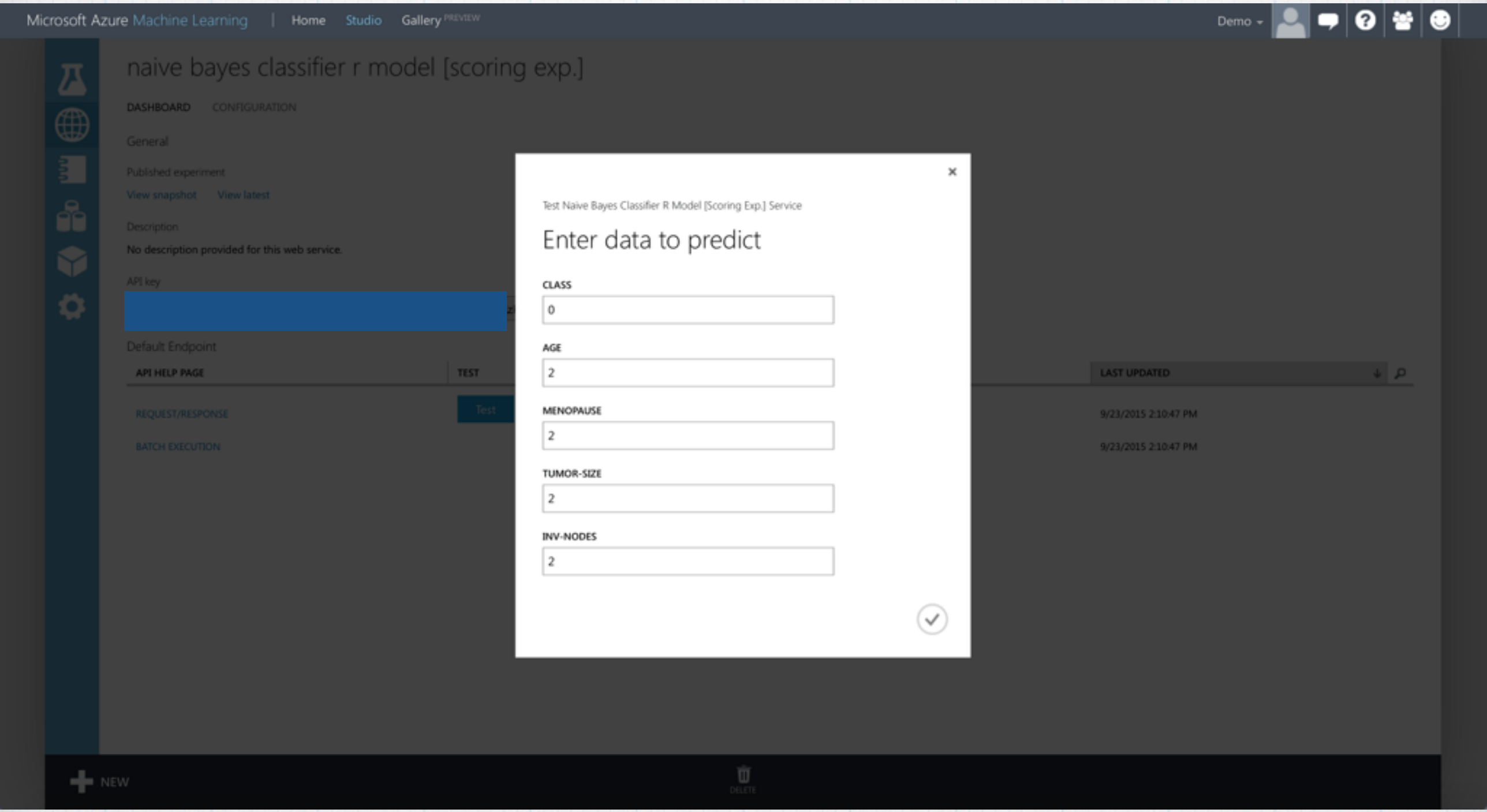
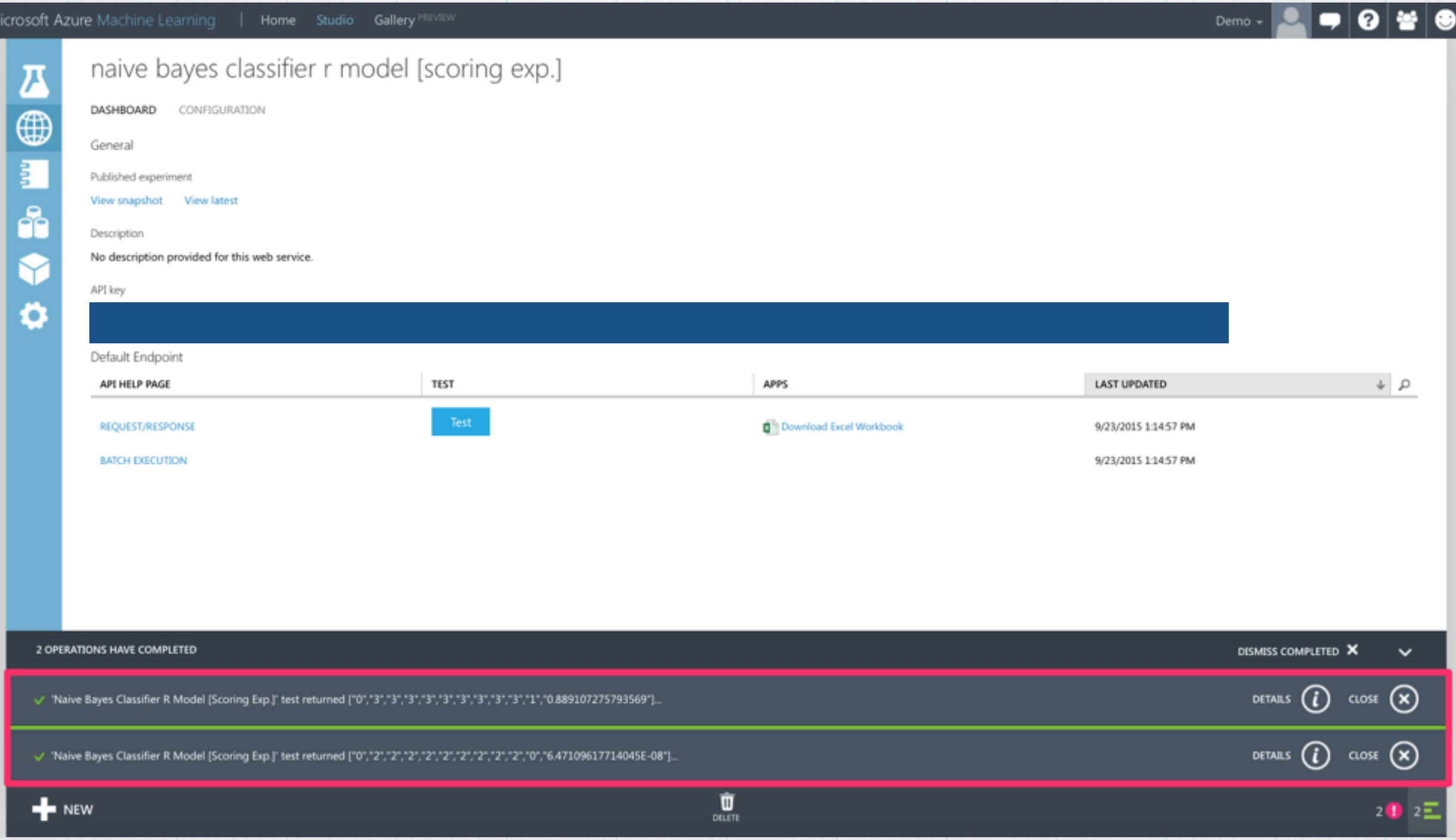
Sample Code

C# | Python | R

Select sample code

```
// This code requires the Nuget package Microsoft.AspNet.WebApi.Client to be installed.
// Instructions for doing this in Visual Studio:
// Tools -> Nuget Package Manager -> Package Manager Console
// Install-Package Microsoft.AspNet.WebApi.Client

using System;
using System.Collections.Generic;
using System.IO;
using System.Net.Http;
using System.Net.Http.Formatting;
using System.Net.Http.Headers;
using System.Text;
using System.Threading.Tasks;

namespace CallRequestResponseService
{

    public class StringTable
    {
        public string[] ColumnNames { get; set; }
        public string[,] Values { get; set; }
    }

    class Program
    {
        static void Main(string[] args)
        {
```

Invoking R Model



Testing a Web Service