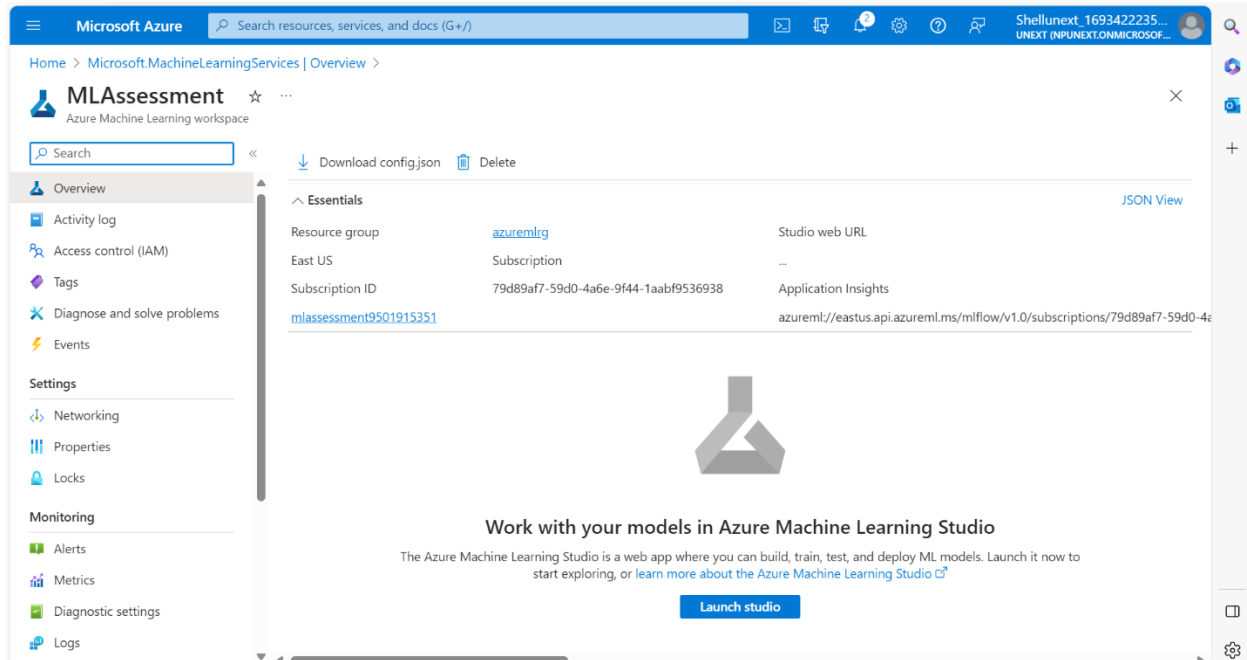


# Azure ML Hands on Assessment



ML Studio Workspace created

Microsoft Azure Search resources, services, and docs (G+/)

Home > Storage accounts > mlassessment9501915351 | Containers >

**azureml-blobstore-a9e48ff9-f55b-40fc-9f53-44cd684fdbce** Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots

**Overview**

Diagnose and solve problems

Access Control (IAM)

**Settings**

Shared access tokens

Access policy

Properties

Metadata

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

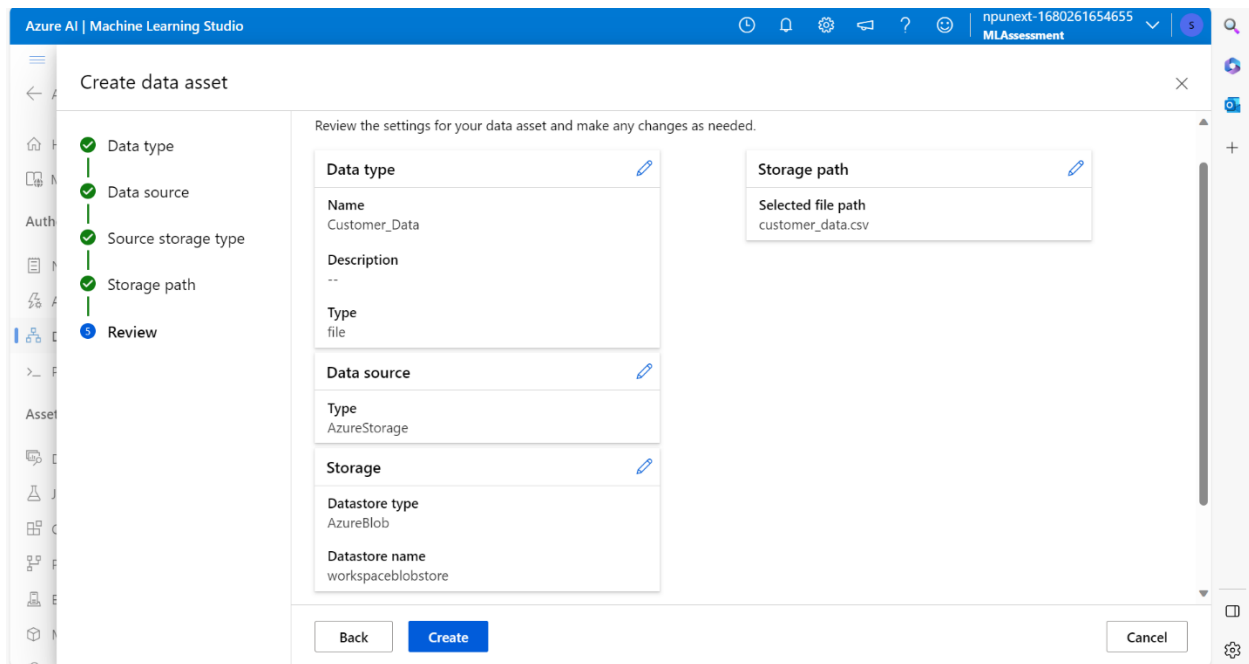
**Location:** azureml-blobstore-a9e48ff9-f55b-40fc-9f53-44cd684fdbce

Search blobs by prefix (case-sensitive) Show deleted blobs

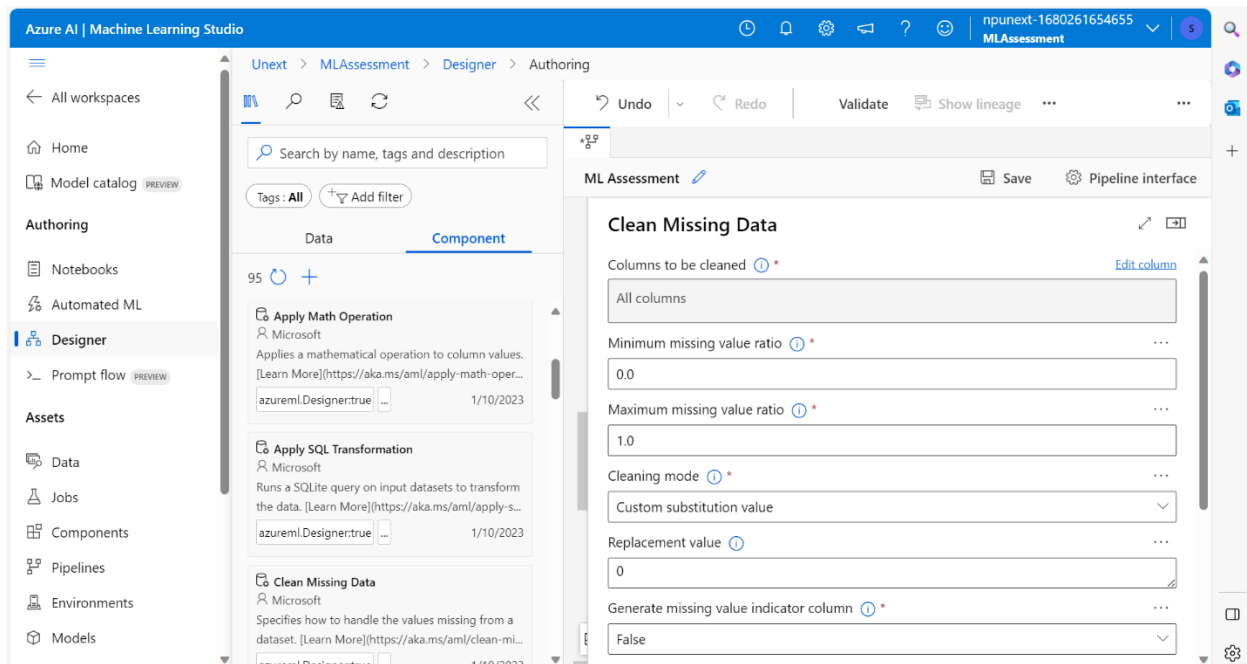
Add filter

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> customer_data.csv	10/4/2023, 1:42:52 PM	Hot (Inferred)		Block blob

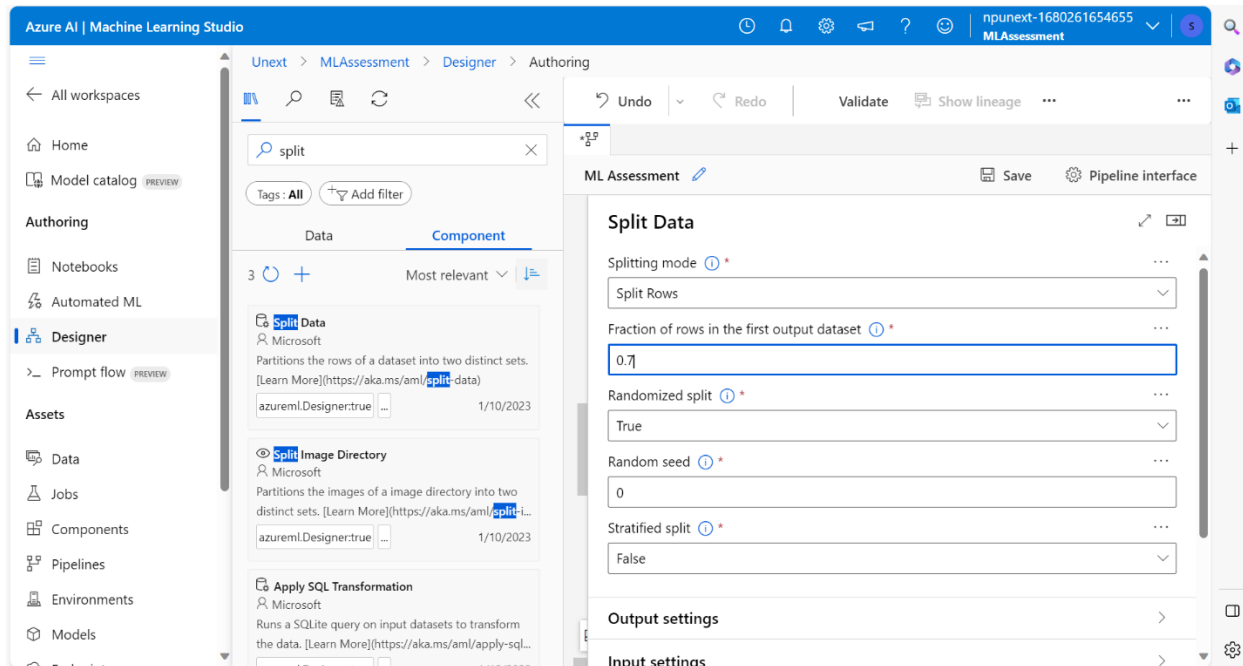
Required File uploaded to Blob



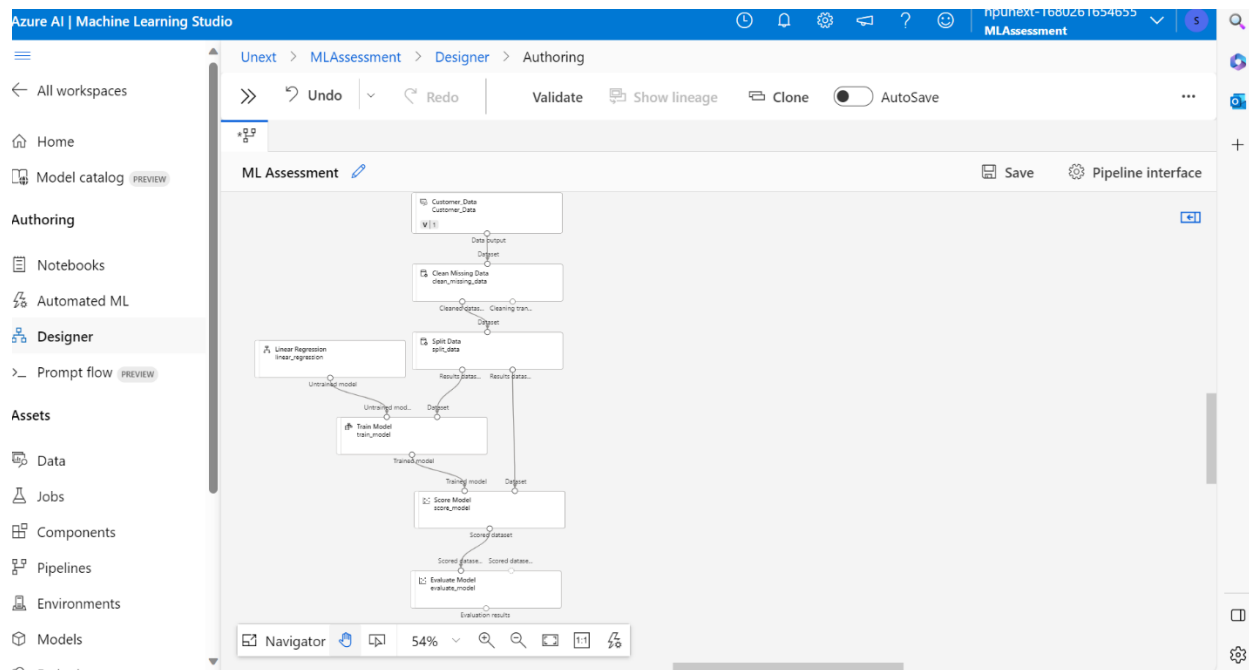
## Loading dataset in Azure ML



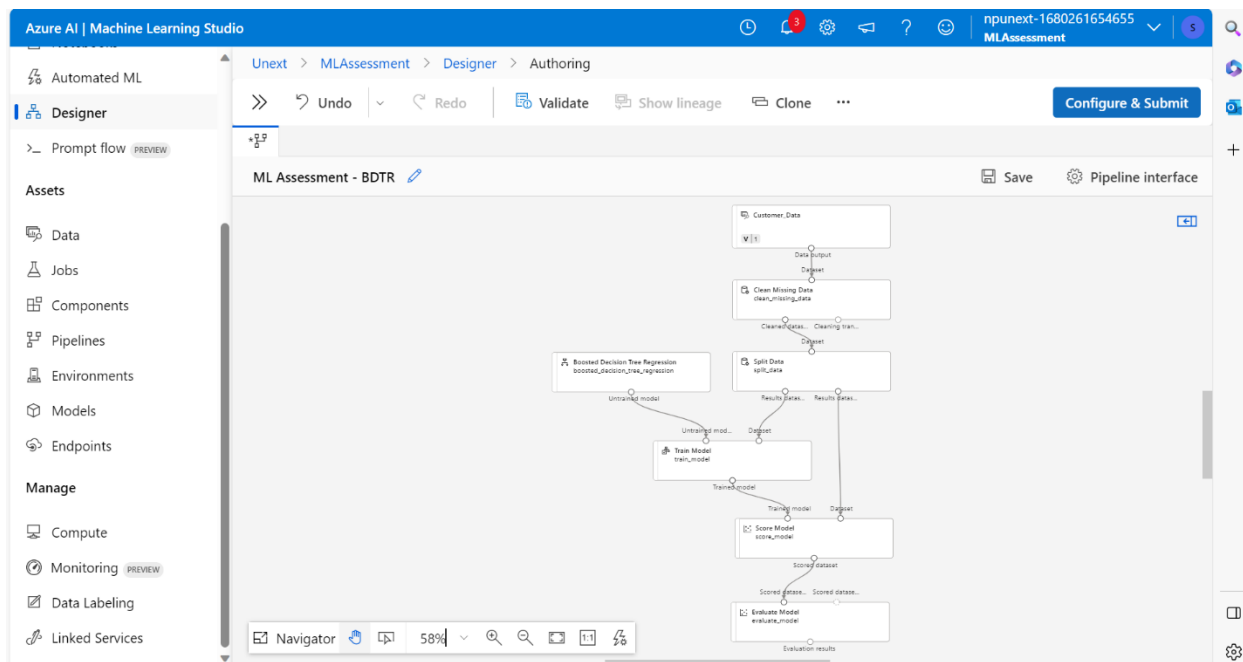
Cleaning missing values, by replacing them all with 0 in this case.



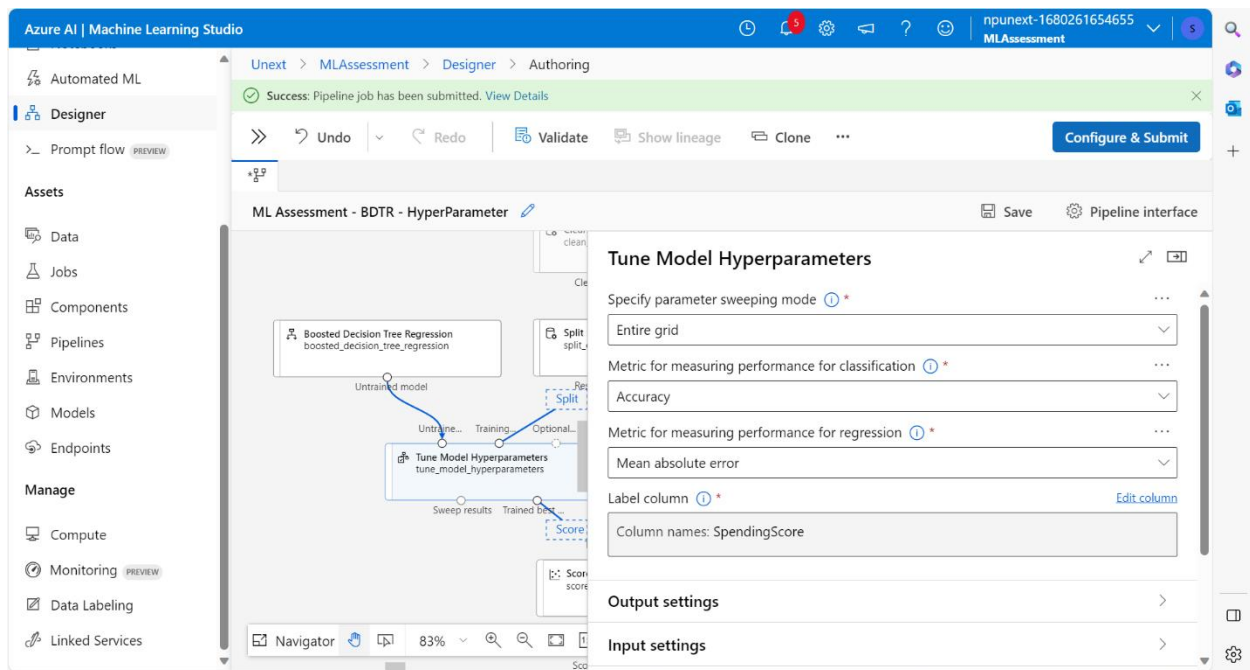
Splitting the data, 70% Train data and 30% Test data.



First pipeline created after following the above steps using Linear Regression Model and using Train Model.

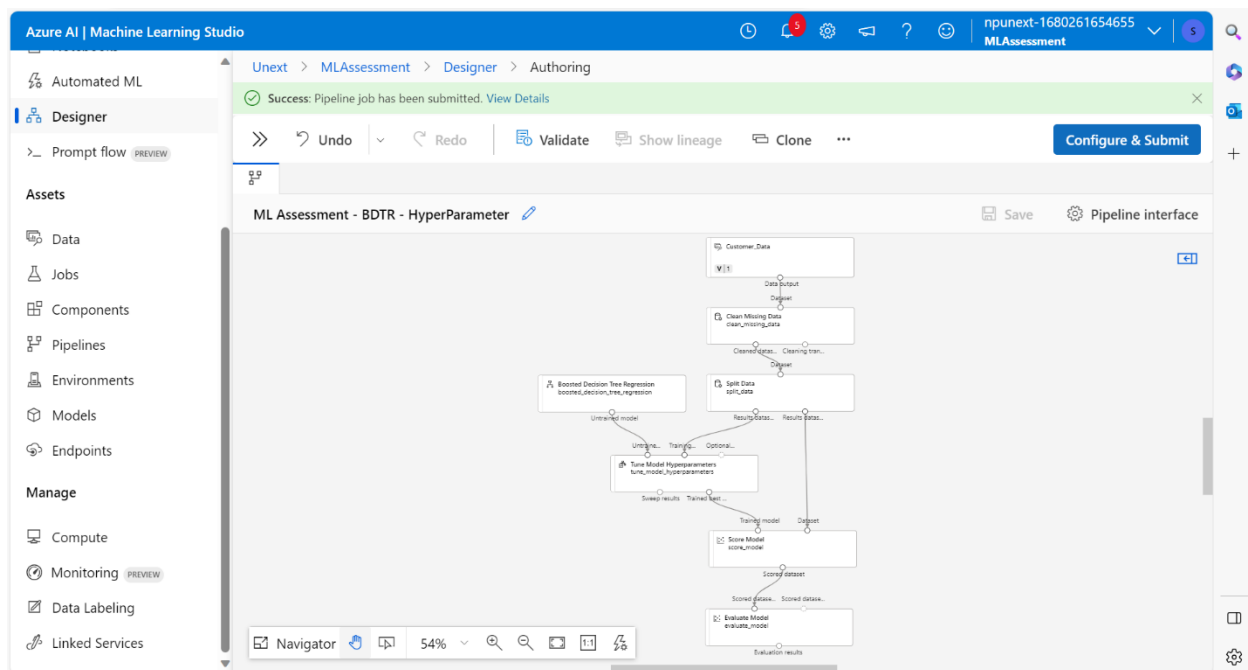


Second pipeline created using algorithm as Boosted Decision Tree Regressor and using Train Model.

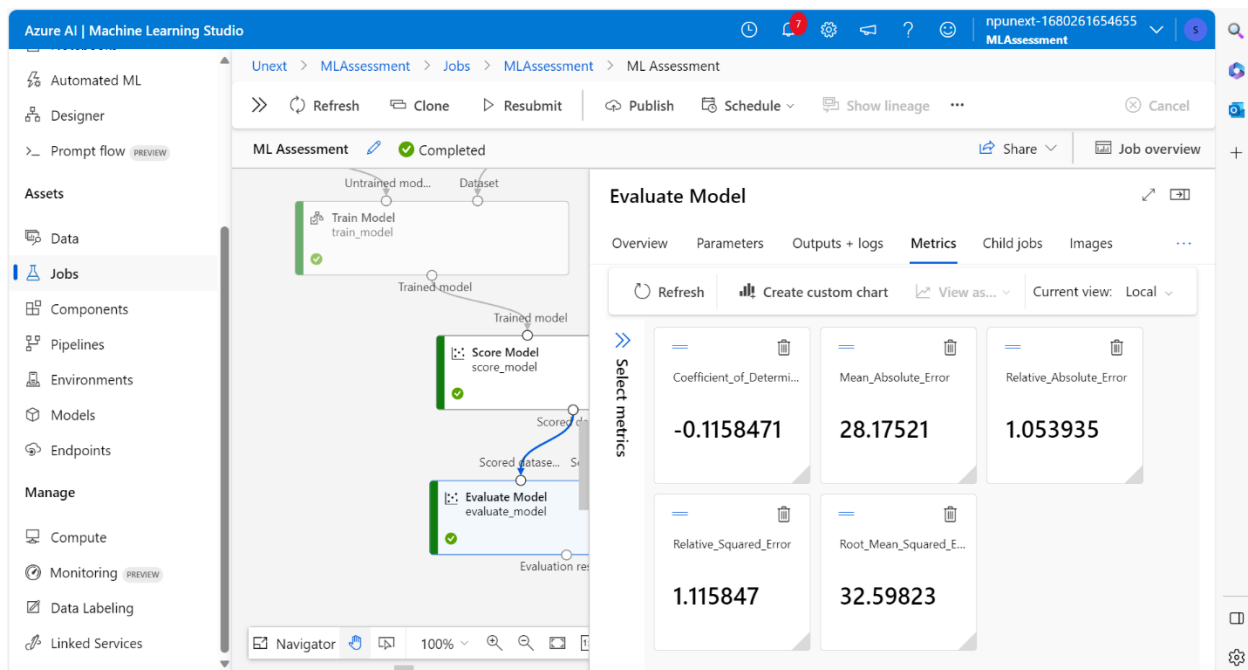


Creating 3<sup>rd</sup> pipeline with HyperParameter Tuning.  
Using Entire Grid in this case.

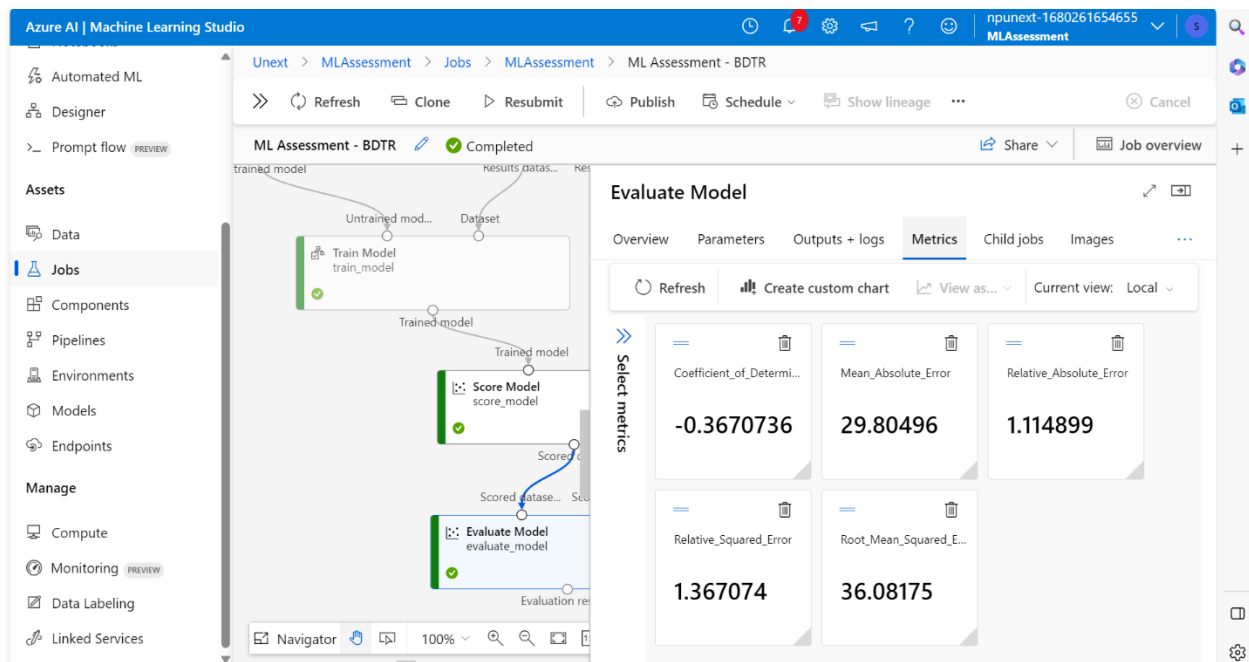




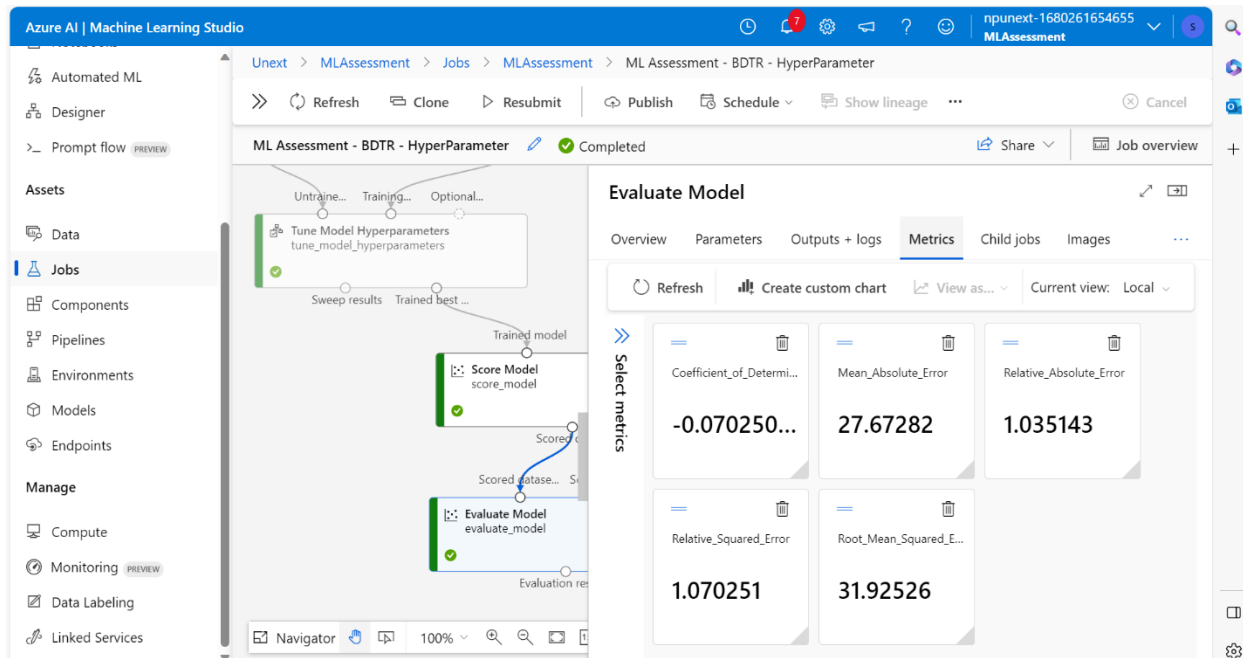
Third pipeline created with Boosted Decision Tree Regressor using HyperParameter Tuning.



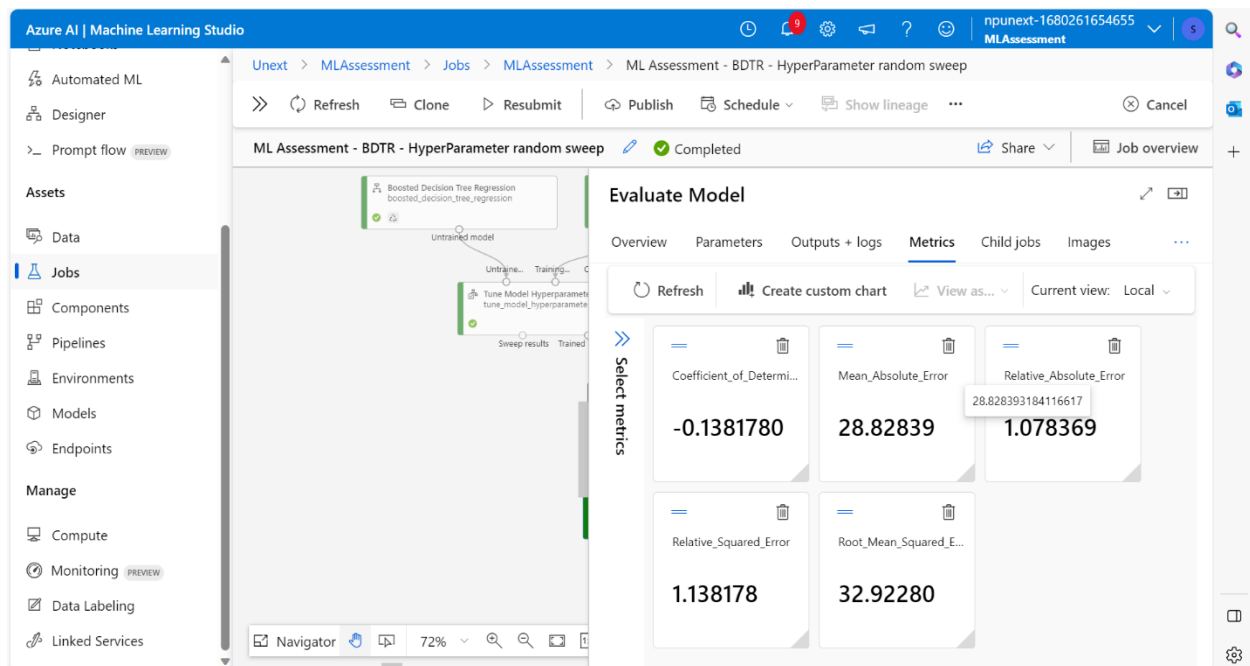
Scores using LR



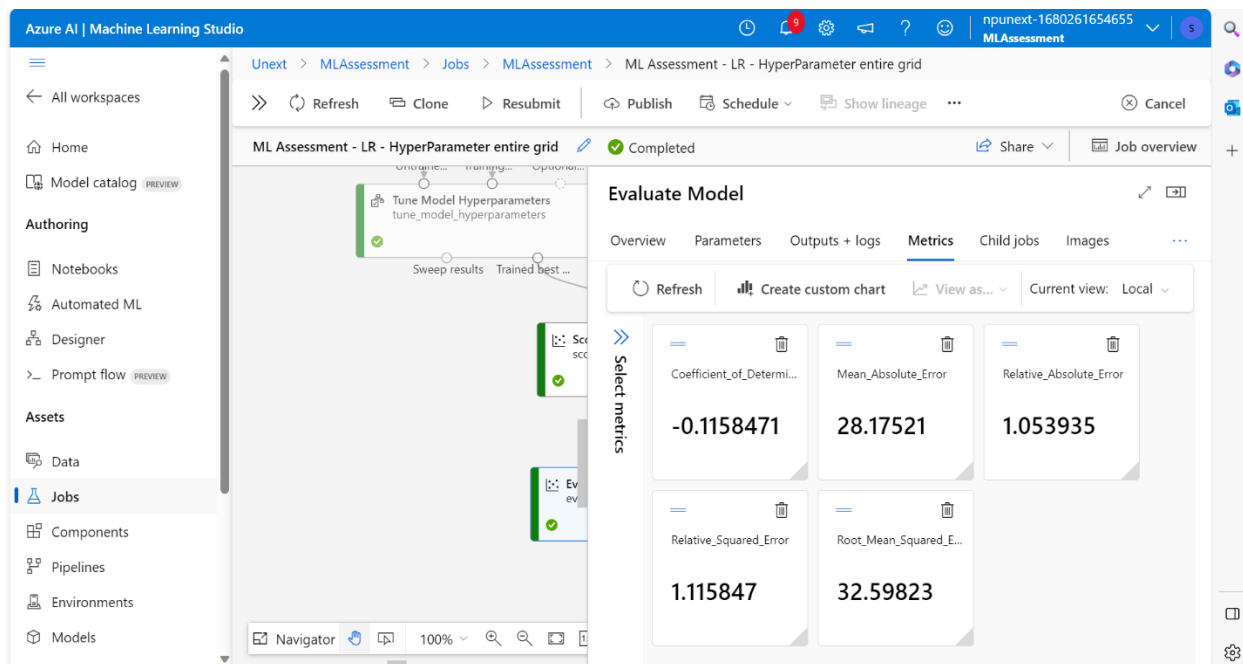
Scores using Boosted Decision Tree Regressor



Scores using Boosted Decision Tree regressor with  
HyperParameter Tuning (Entire Grid)



Score for Boosted Decision Tree Regressor using HyperParamater as Random Sweep (5)



Score with Linear Regression using HyperParameter as Entire Grid

**The best score was achieved using Boosted Decision Tree Regressor with HyperParameter Tuning as Entire Grid.**

## **Assessment Questions**

- 1) Preparing the dataset requires cleaning the Data and doing any necessary transformations before feeding it to the ML Model.

In our case, we have missing values in most of the columns, so I have filled the missing values with 0.

- 2) It is important to split the dataset into training and testing because we want the machine to be able to recognize that which data it should use to learn from and which data it should test itself on to see whether it learned well enough or not.

In more technical terms, Train dataset provides the algorithm with the data over which it iterates and learns the patterns. Whereas the Test data is what the model tests itself against to give us our evaluation metrics and accuracy score.

3) In this case, I have used Linear Regression and Boosted Decision Tree Regressor. I am getting a better metric score with Linear Regression.

Regression type algorithms are best when it comes to prediction problems.

4) HyperParameter Tuning is basically choosing what inputs to give to the train data so that it iterates over those chosen parameters accordingly.

It is helpful because we can control what the model is training on.

One example is grid sweep. In this we sweep over the entire dataset so that no data is missed.

If we want to save time, we can do random sweep and only iterate over some of the data.



