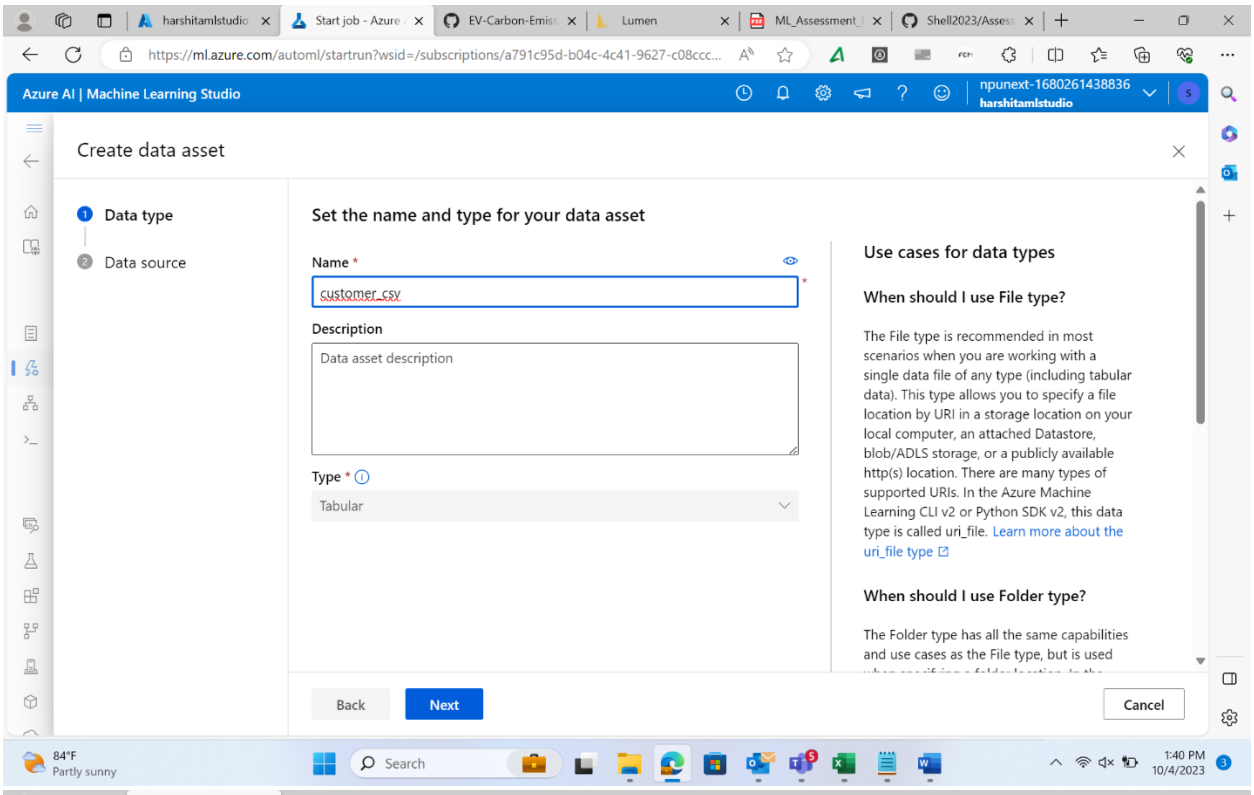
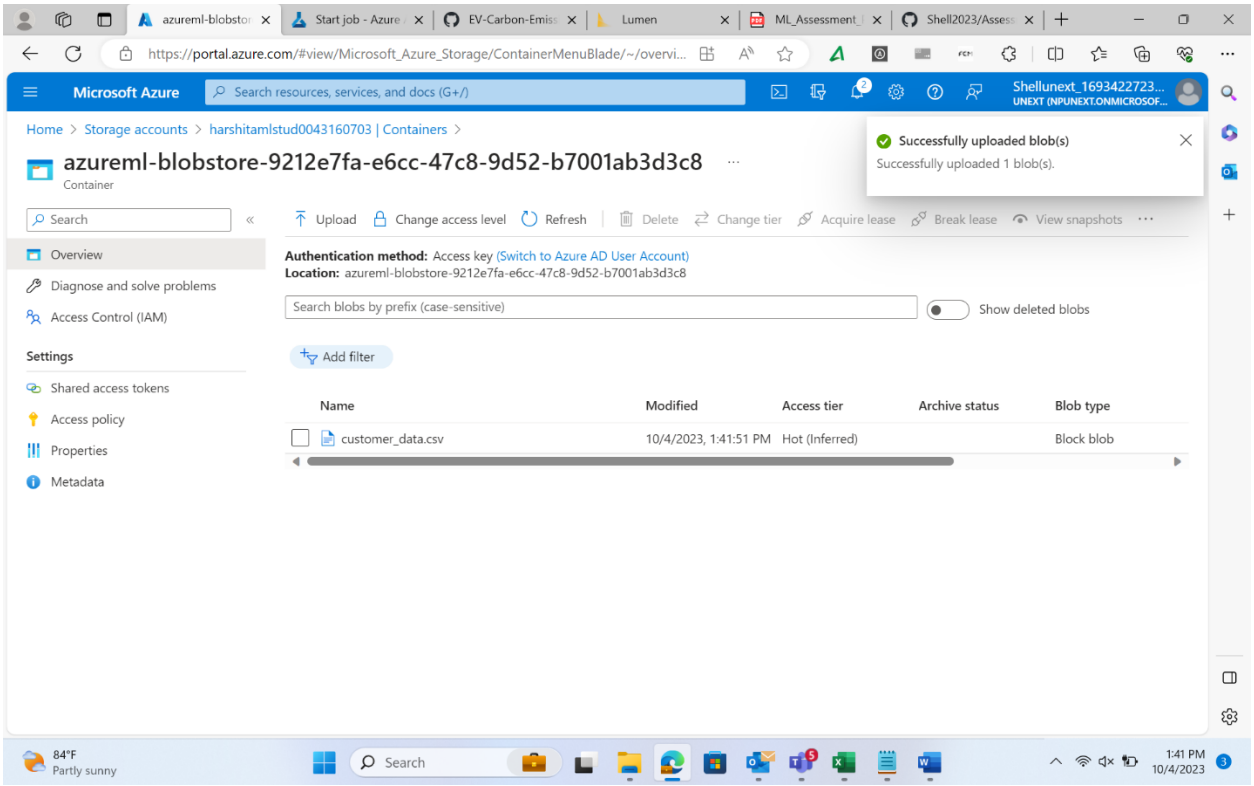
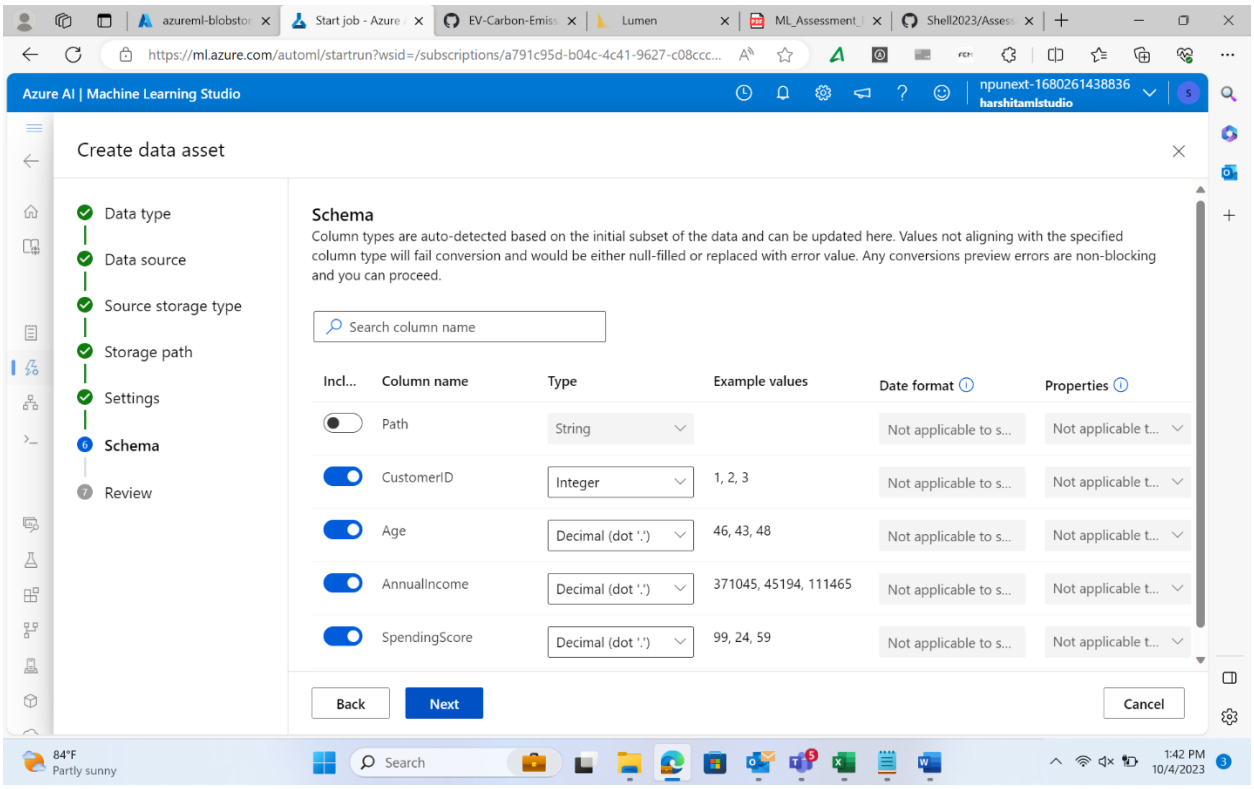
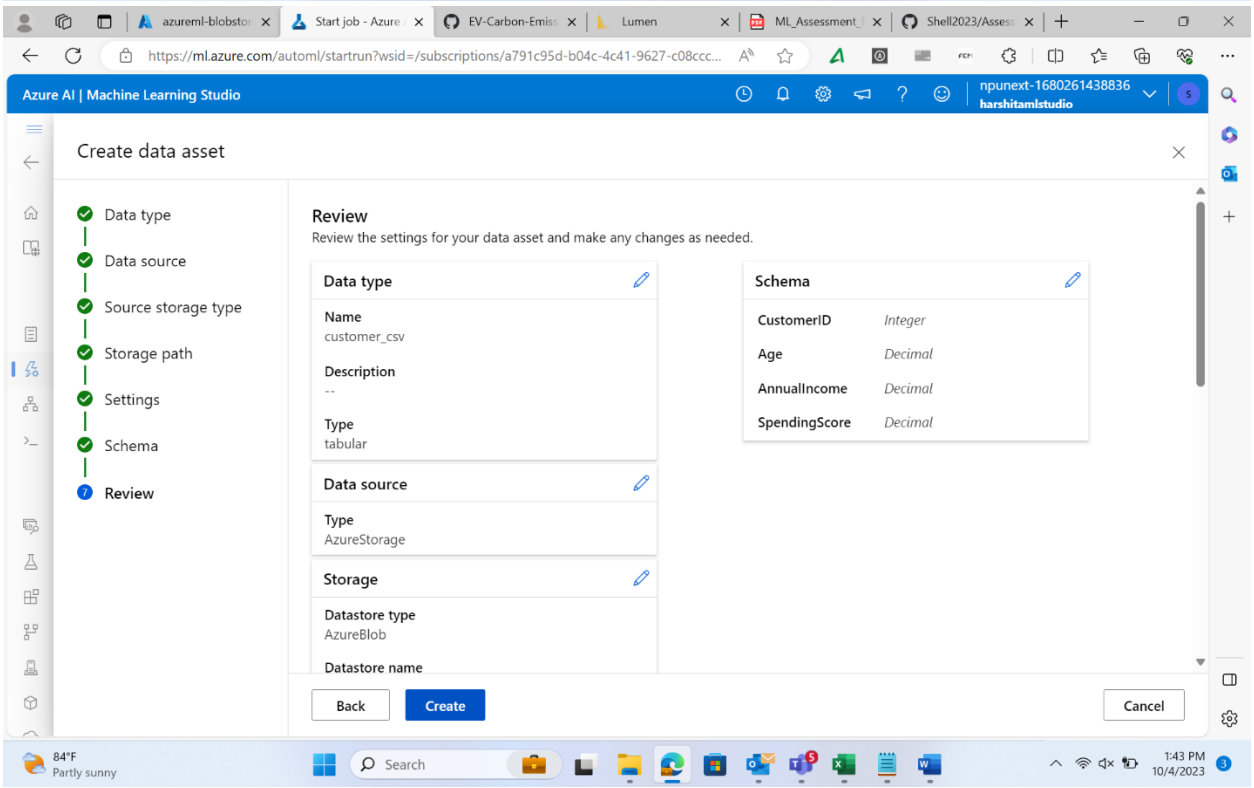
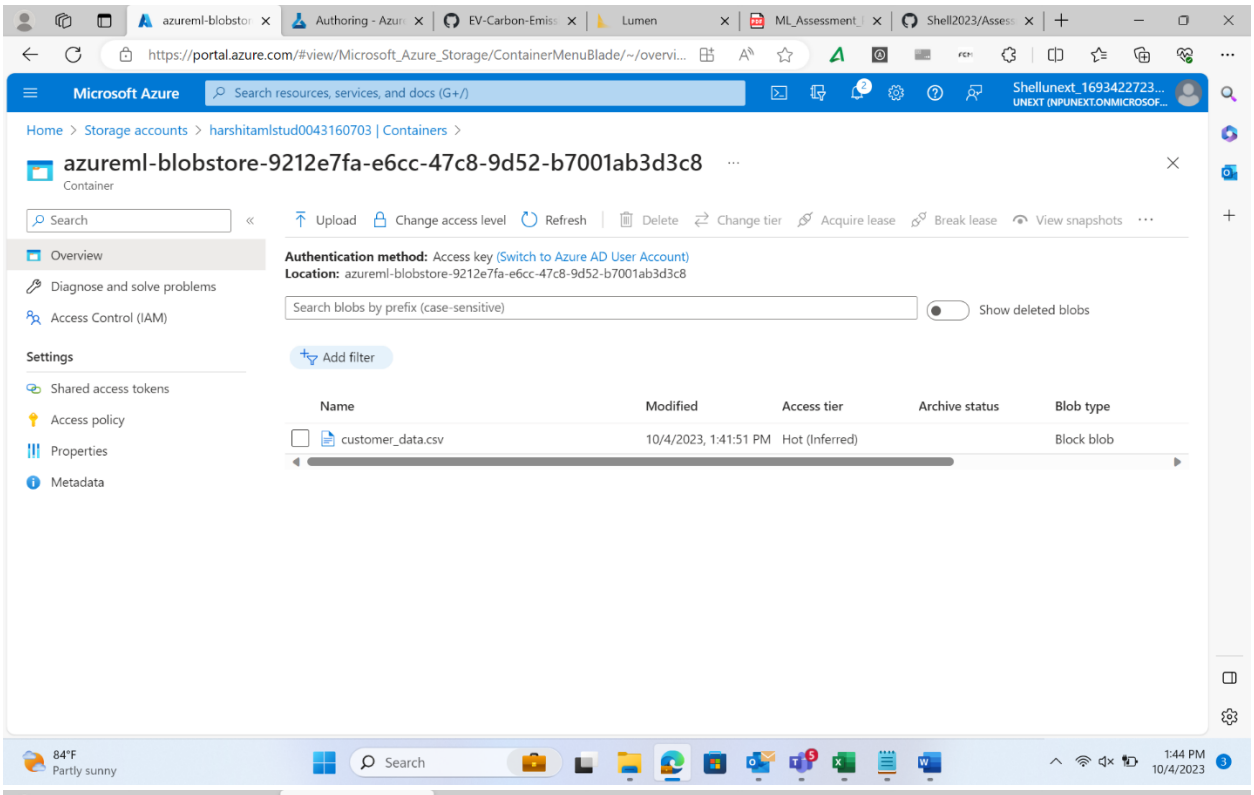
Hands on Assessment for Azure ML Studio

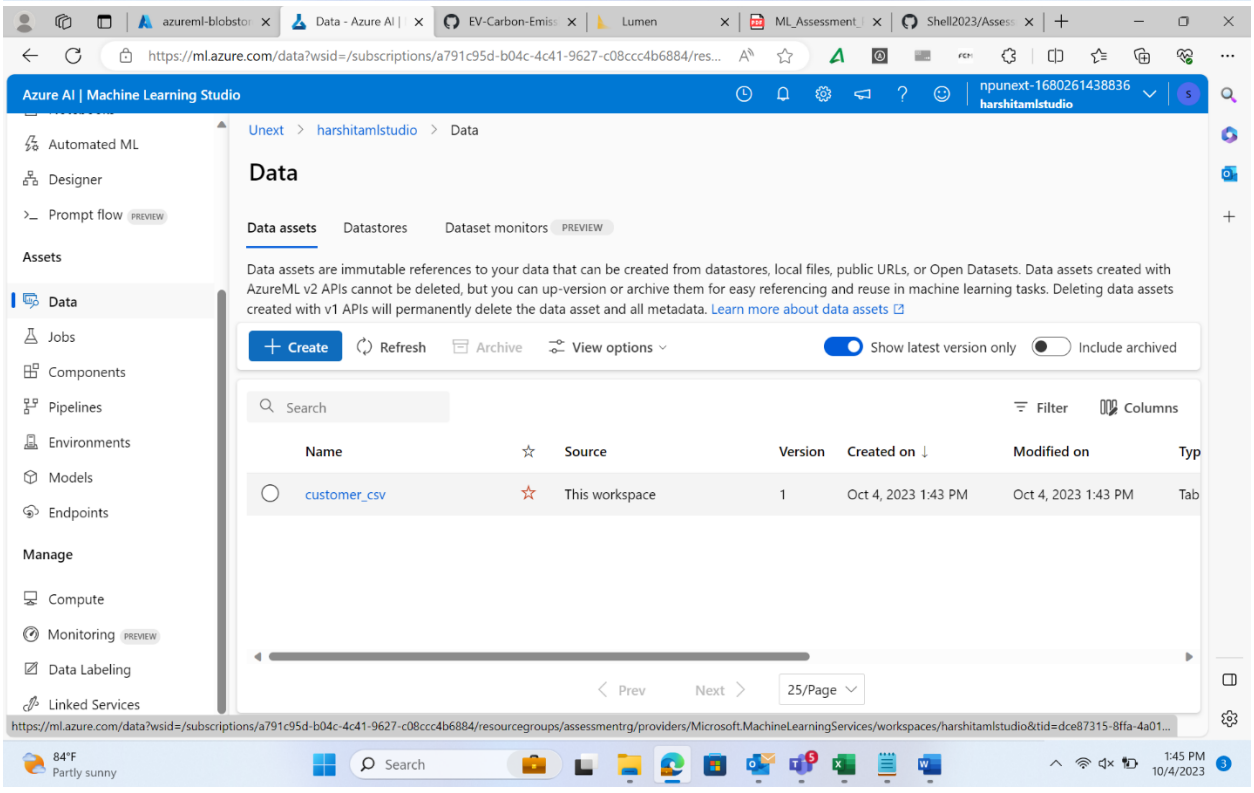


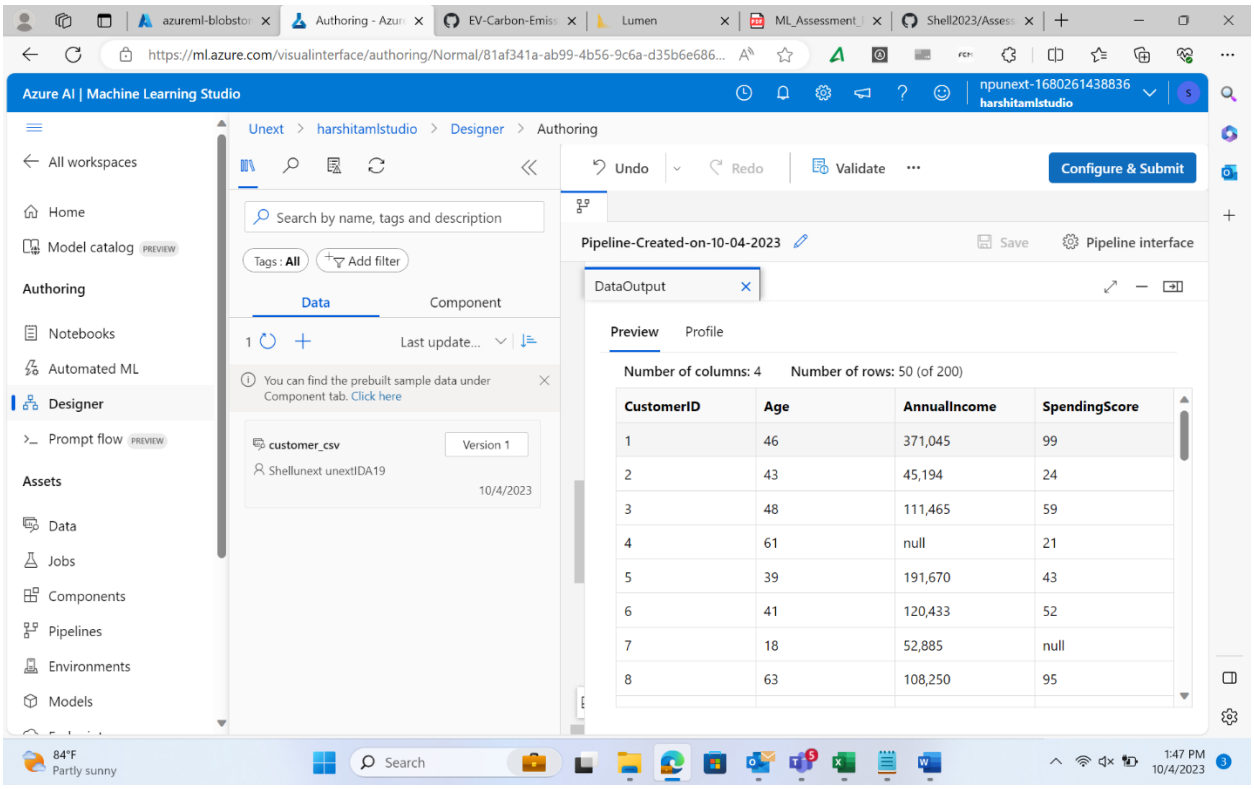


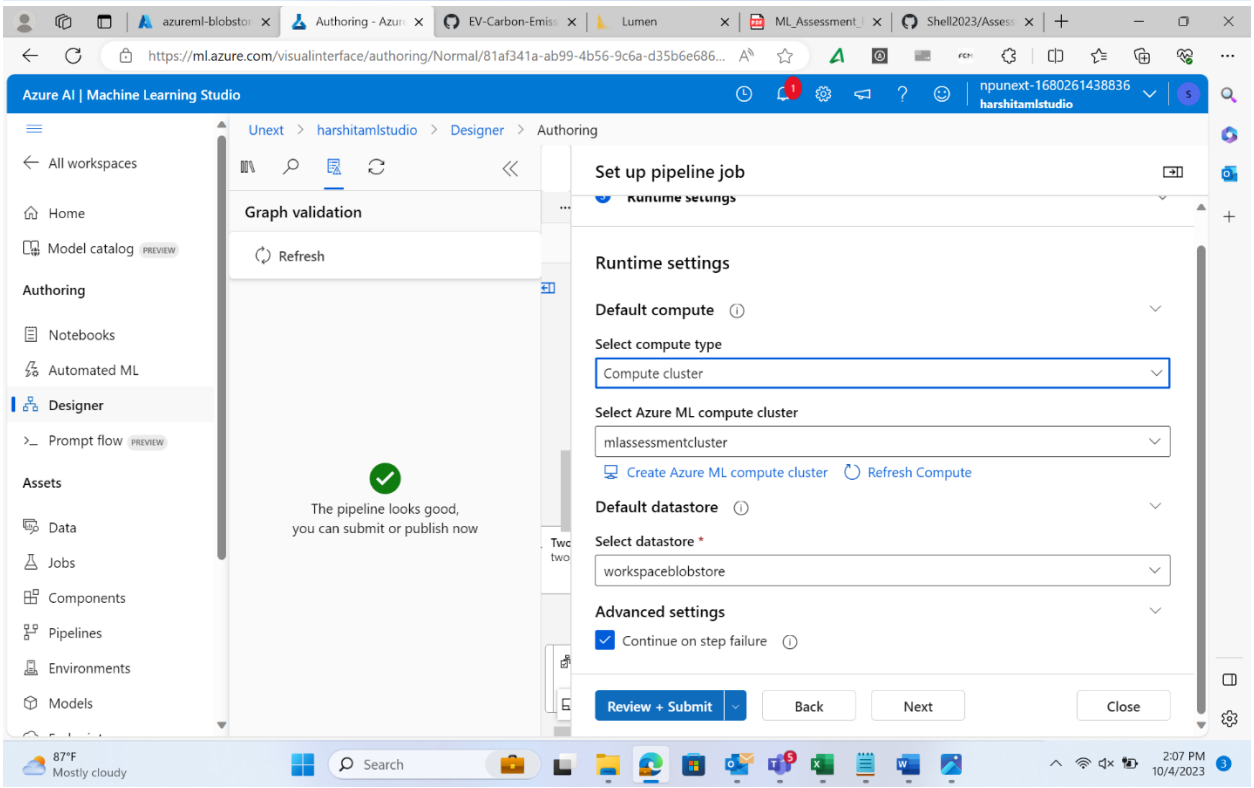

Azure AI | Machine Learning Studio
npunext-1680261438836
harshitamlstudio
Create data asset
Data type
Data source
Source storage type
Storage path
Settings
Schema
Review
Schema
Column types are auto-detected based on the initial subset of the data and can be updated here. Values not aligning with the specified column type will fail conversion and would be either null-filled or replaced with error value. Any conversions preview errors are non-blocking and you can proceed.
Search column name
Incl...
Column name
Type
Example values
Date format
Properties
Path
String
CustomerID
Integer
1, 2, 3
Age
Decimal (dot '.')
46, 43, 48
AnnualIncome
Decimal (dot '.')
371045, 45194, 111465
SpendingScore
Decimal (dot '.')
99, 24, 59
Not applicable to s...
Not applicable t...
Back
Next
Cancel

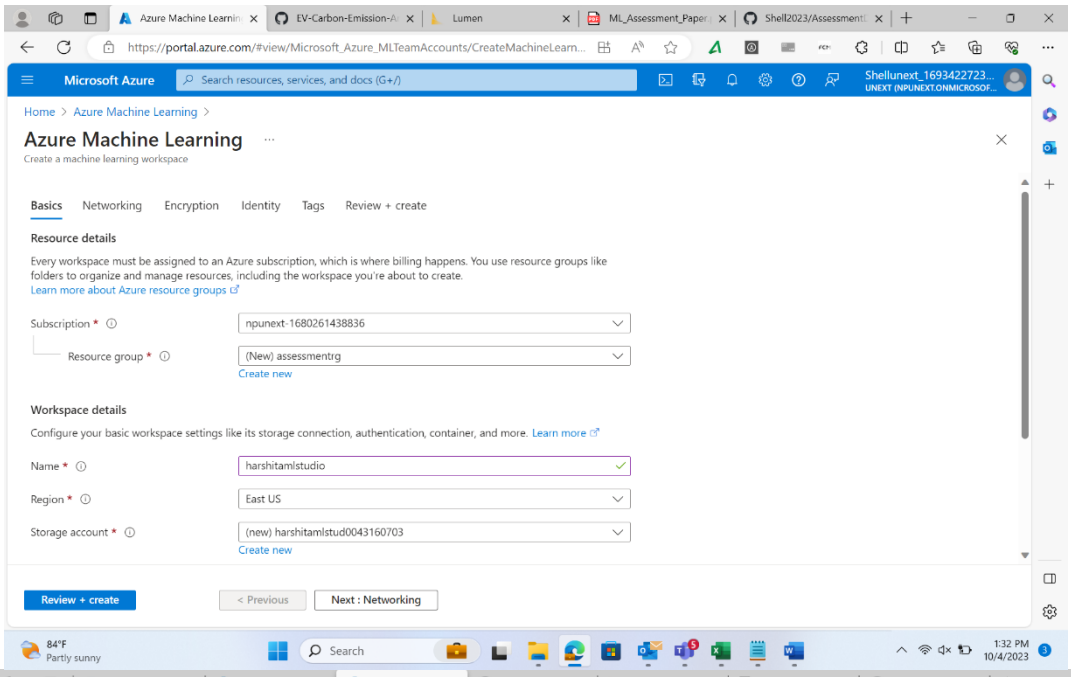

Azure AI | Machine Learning Studio
npunext-1680261438836
harshitamlstudio
Create data asset
Data type
Data source
Source storage type
Storage path
Settings
Schema
Review
Review
Review the settings for your data asset and make any changes as needed.
Data type
Name
customer_csv
Description
--
Type
tabular
Data source
Type
AzureStorage
Storage
Datastore type
AzureBlob
Datastore name
Schema
CustomerID Integer
Age Decimal
AnnualIncome Decimal
SpendingScore Decimal
Back
Create
Cancel

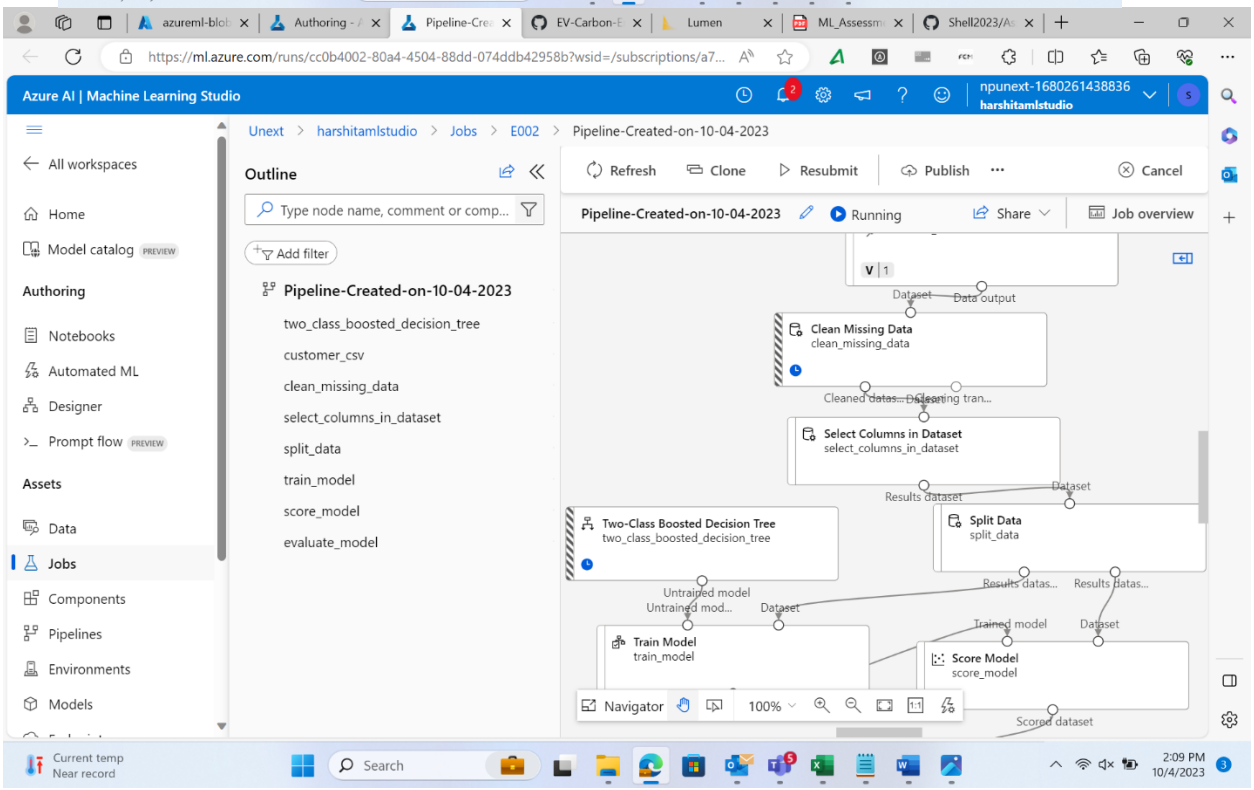

azureml-blobstor
Authoring - Azur
EV-Carbon-Emiss
Lumen
ML_Assessment_
Shell2023/Assess
https://portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overvi...
Microsoft Azure
Search resources, services, and docs (G+/)
Shellunext_1693422723...
UNEXT (NPUNEXT.ONMICROSOF...
Home > Storage accounts > harshitamlstud0043160703 | Containers >
azureml-blobstore-9212e7fa-e6cc-47c8-9d52-b7001ab3d3c8
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Access policy
Properties
Metadata
Upload
Change access level
Refresh
Delete
Change tier
Acquire lease
Break lease
View snapshots
Authentication method: Access key (Switch to Azure AD User Account)
Location: azureml-blobstore-9212e7fa-e6cc-47c8-9d52-b7001ab3d3c8
Search blobs by prefix (case-sensitive)
Show deleted blobs
Add filter
Name
Modified
Access tier
Archive status
Blob type
customer_data.csv
10/4/2023, 1:41:51 PM
Hot (Inferred)
Block blob
84°F
Partly sunny
Search
1:44 PM
10/4/2023

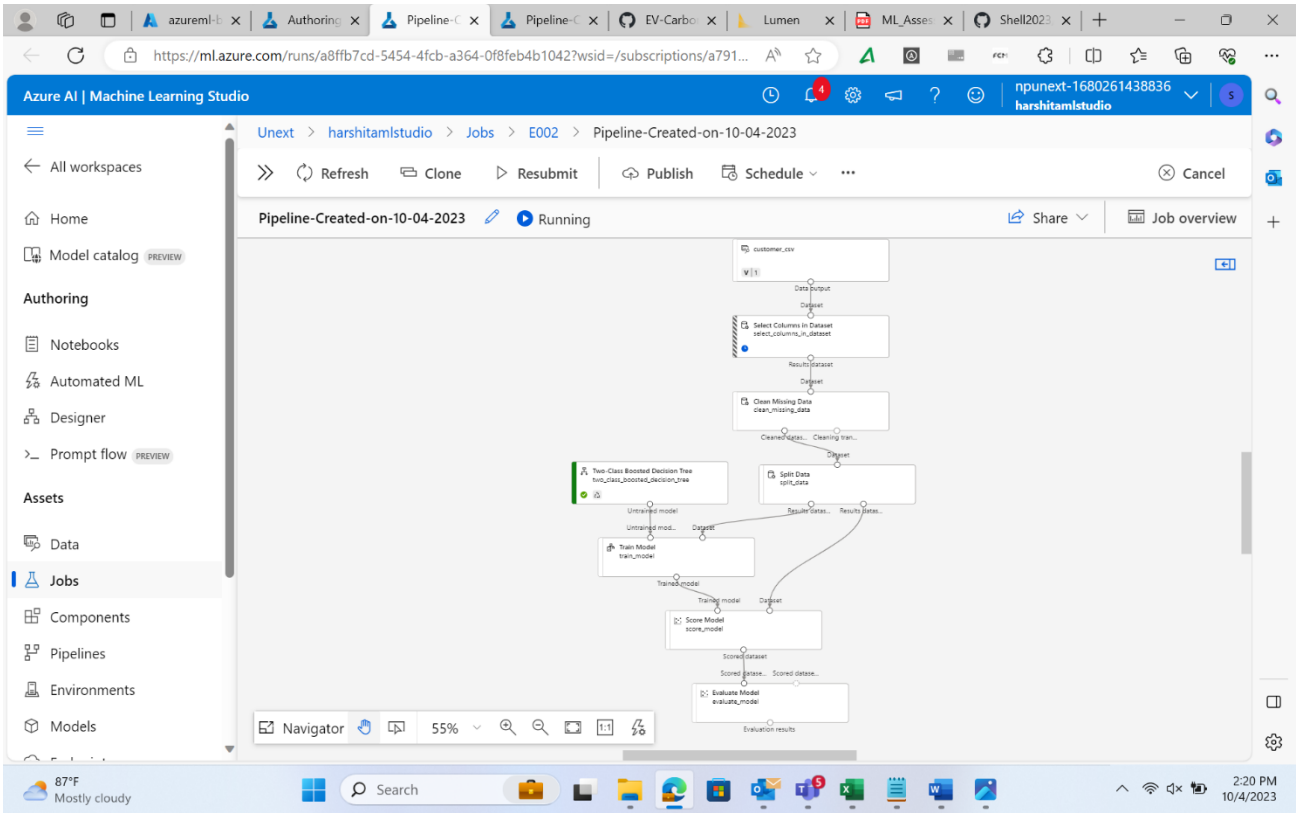

azureml-blobstor
Data - Azure AI |
EV-Carbon-Emiss
Lumen
ML_Assessment_
Shell2023/Assess
https://ml.azure.com/data?wsid=/subscriptions/a791c95d-b04c-4c41-9627-c08ccc4b6884/res...
Azure AI | Machine Learning Studio
npunext-1680261438836
harshitamlstudio
Automated ML
Designer
Prompt flow PREVIEW
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring PREVIEW
Data Labeling
Linked Services
Unext > harshitamlstudio > Data
Data
Data assets
Datastores
Dataset monitors PREVIEW
Data assets are immutable references to your data that can be created from datastores, local files, public URLs, or Open Datasets. Data assets created with AzureML v2 APIs cannot be deleted, but you can up-version or archive them for easy referencing and reuse in machine learning tasks. Deleting data assets created with v1 APIs will permanently delete the data asset and all metadata. Learn more about data assets
Create
Refresh
Archive
View options
Show latest version only
Include archived
Search
Filter
Columns
Name
Source
Version
Created on
Modified on
Typ
customer_csv
This workspace
1
Oct 4, 2023 1:43 PM
Oct 4, 2023 1:43 PM
Tab
Prev
Next
25/Page
https://ml.azure.com/data?wsid=/subscriptions/a791c95d-b04c-4c41-9627-c08ccc4b6884/resourcegroups/assessmentrg/providers/Microsoft.MachineLearningServices/workspaces/harshitamlstudio&tid=dce87315-8ffa-4a01...
84°F
Partly sunny
Search
1:45 PM
10/4/2023


Azure AI | Machine Learning Studio
Unext > harshitamlstudio > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Undo
Redo
Validate
Configure & Submit
Search by name, tags and description
Tags : All
Add filter
Data
Component
Last update...
You can find the prebuilt sample data under Component tab. Click here
customer_csv
Version 1
Shellunext unextIDA19
10/4/2023
Pipeline-Created-on-10-04-2023
Save
Pipeline interface
DataOutput
Preview
Profile
Number of columns: 4
Number of rows: 50 (of 200)
CustomerID
Age
AnnualIncome
SpendingScore
1 46 371,045 99
2 43 45,194 24
3 48 111,465 59
4 61 null 21
5 39 191,670 43
6 41 120,433 52
7 18 52,885 null
8 63 108,250 95


Azure AI | Machine Learning Studio
Unext > harshitamlstudio > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Graph validation
Refresh
The pipeline looks good, you can submit or publish now
Set up pipeline job
Runtime settings
Default compute
Select compute type
Compute cluster
Select Azure ML compute cluster
mlassessmentcluster
Create Azure ML compute cluster
Refresh Compute
Default datastore
Select datastore *
workspaceblobstore
Advanced settings
Continue on step failure
Review + Submit
Back
Next
Close


Microsoft Azure
Search resources, services, and docs (G+/)
Home > Azure Machine Learning >
Azure Machine Learning
Create a machine learning workspace
Basics
Networking
Encryption
Identity
Tags
Review + create
Resource details
Every workspace must be assigned to an Azure subscription, which is where billing happens. You use resource groups like folders to organize and manage resources, including the workspace you're about to create.
Learn more about Azure resource groups
Subscription *
npunext-1680261438836
Resource group *
(New) assessmentrg
Create new
Workspace details
Configure your basic workspace settings like its storage connection, authentication, container, and more. Learn more
Name *
harshitamlstudio
Region *
East US
Storage account *
(new) harshitamlstud0043160703
Create new
Review + create
< Previous
Next : Networking


Azure AI | Machine Learning Studio
npunext-1680261438836
harshitamlstudio
Unext > harshitamlstudio > Jobs > E002 > Pipeline-Created-on-10-04-2023
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Outline
Type node name, comment or comp...
Add filter
Pipeline-Created-on-10-04-2023
two_class_boosted_decision_tree
customer_csv
clean_missing_data
select_columns_in_dataset
split_data
train_model
score_model
evaluate_model
Refresh
Clone
Resubmit
Publish
Cancel
Pipeline-Created-on-10-04-2023
Running
Share
Job overview
Clean Missing Data
clean_missing_data
Select Columns in Dataset
select_columns_in_dataset
Two-Class Boosted Decision Tree
two_class_boosted_decision_tree
Split Data
split_data
Train Model
train_model
Score Model
score_model
Navigator
100%


Azure AI | Machine Learning Studio
npunext-1680261438836
harshitamlstudio
All workspaces
Home
Model catalog PREVIEW
Authoring
Notebooks
Automated ML
Designer
Prompt flow PREVIEW
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Unext > harshitamlstudio > Jobs > E002 > Pipeline-Created-on-10-04-2023
Refresh
Clone
Resubmit
Publish
Schedule
Cancel
Pipeline-Created-on-10-04-2023
Running
Share
Job overview
customer_csv
Select Columns in Dataset
Clean Missing Data
Two-Class Boosted Decision Tree
Split Data
Train Model
Score Model
Evaluate Model
Navigator
55%
87°F
Mostly cloudy
Search
2:20 PM
10/4/2023

1.What are the key steps involved in preparing the dataset for training a machine learning model using Azure Machine Learning? Briefly explain each step.

1.Importing the data either through blob storage or local files.

2.cleansing of data.(i.e replacing null values)

3.Transformation.

4.Splitting the data into train and test data.

2.Why is it important to split the dataset into training and testing sets when developing a machine learning model? How does this help in model evaluation?

Splitting the data into training and testing is all about training the model with the training data in the ratio of 70% i.e. 0.7 training data and 30% i.e. 0.3 testing data.The ratio may vary like 80% and 20%.More the training data ratio more is the accuracy.

3.Describe a machine learning algorithm suitable for predicting customer purchasing behaviour in the given scenario. Explain why you chose this algorithm

Two class boosted decision tree algorithm is being used.Decision tree because based on certain condition decision is being taken.

4.What is hyperparameter tuning, and why is it important in machine learning? Explain a technique used for hyperparameter tuning and its benefits

It is a method for optimizing the performance.It is important in order to get more accurate results.We will tune certain parameters.Tune model hyperparameters is the method we have in azure ML studio.