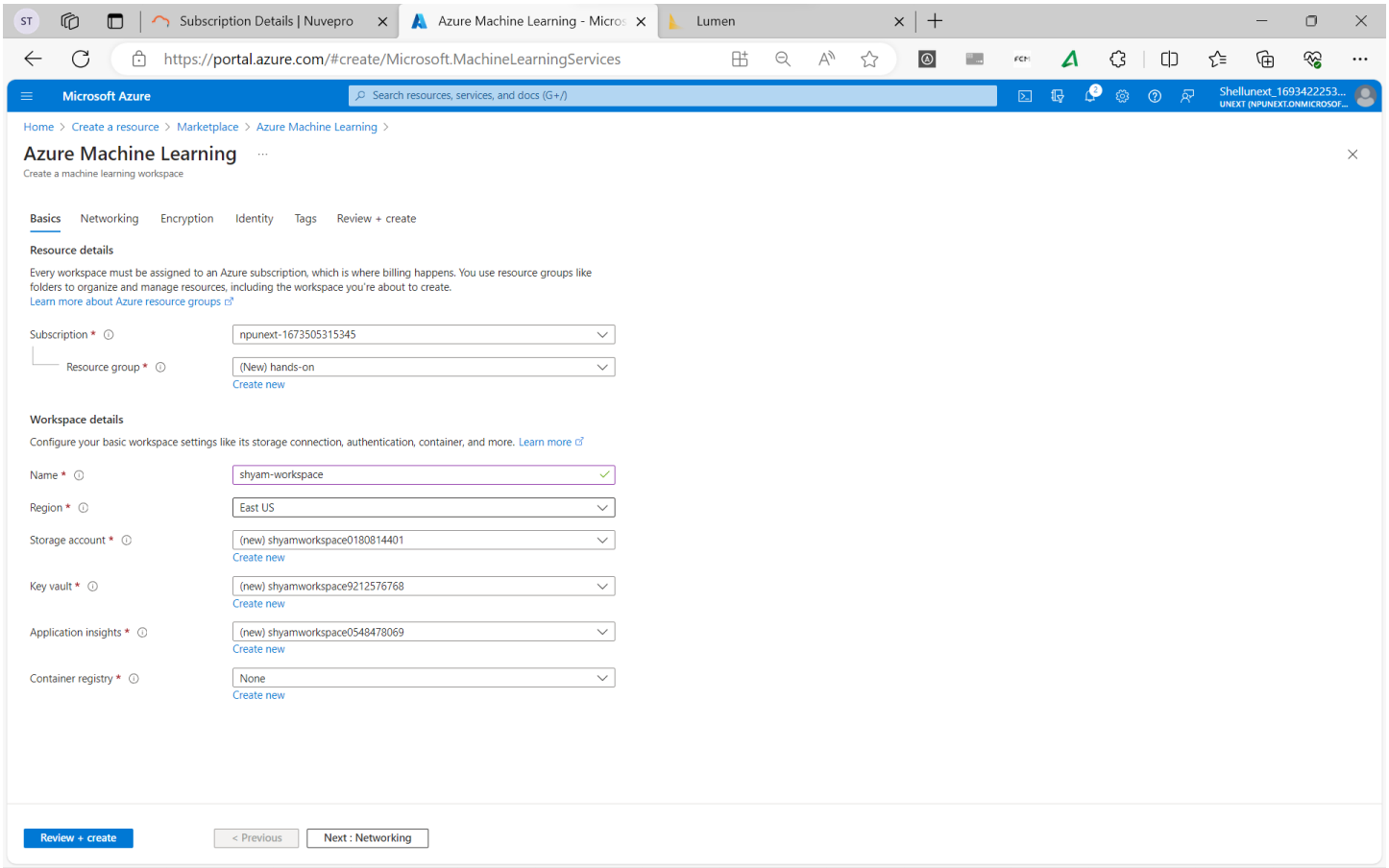
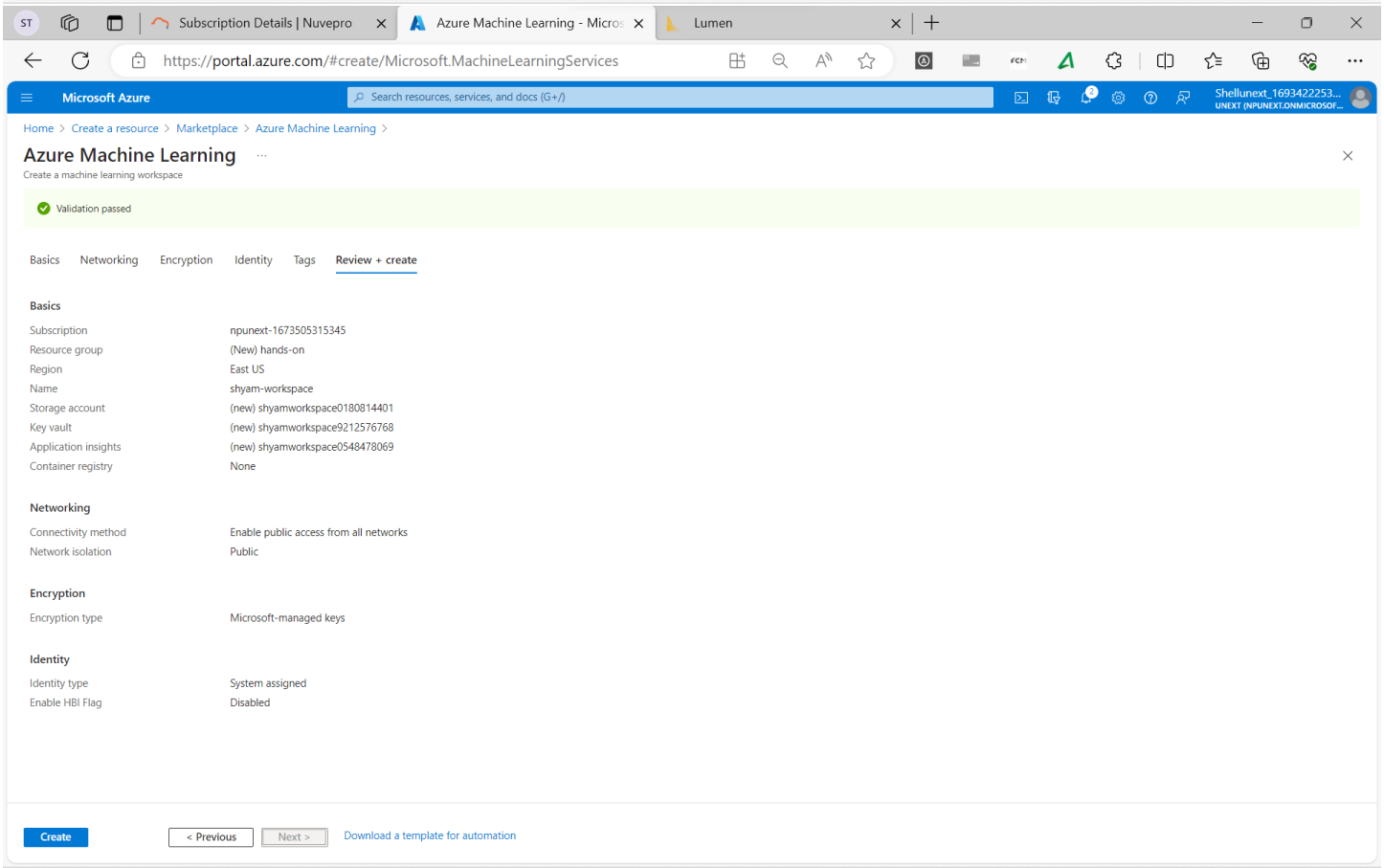
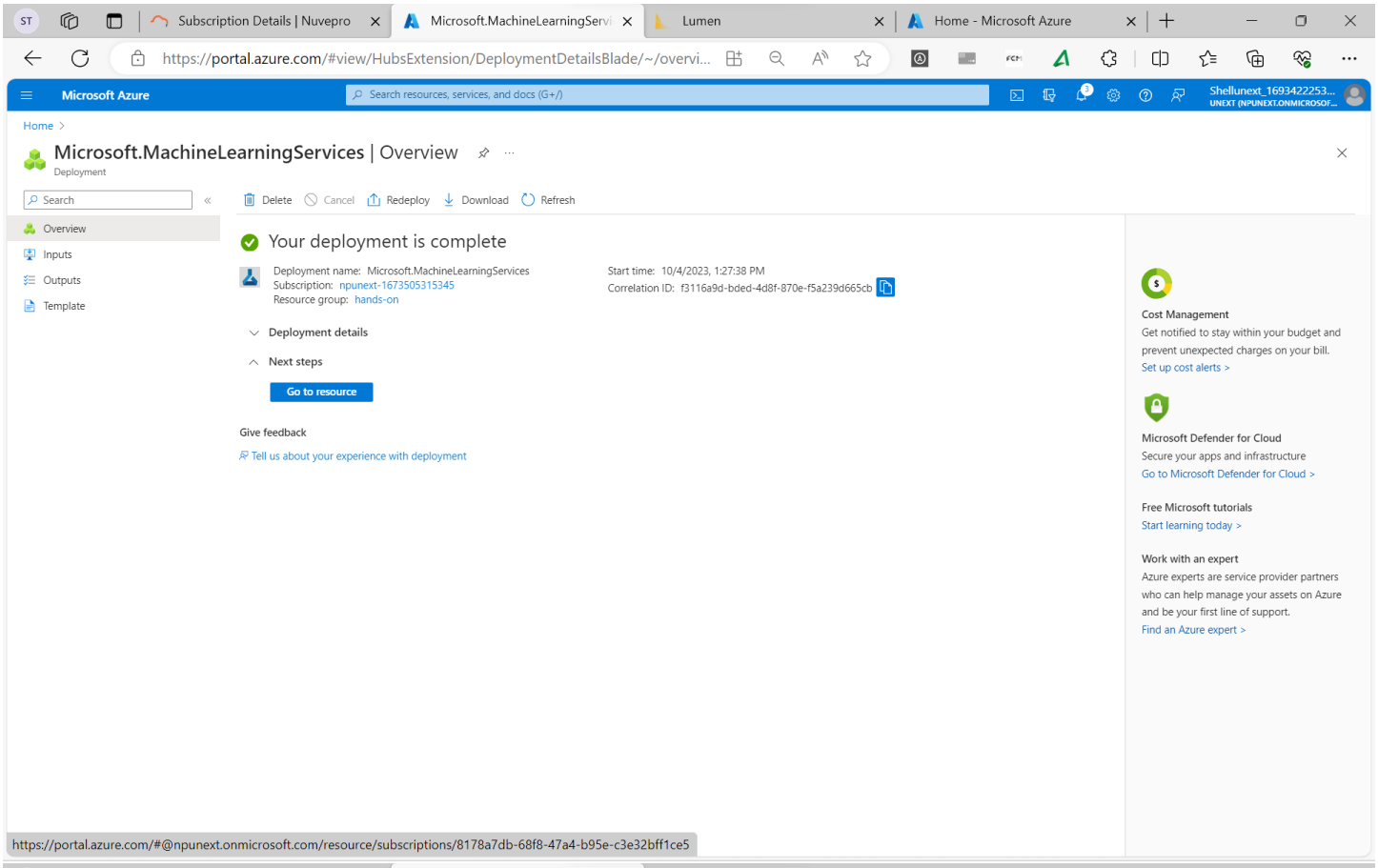
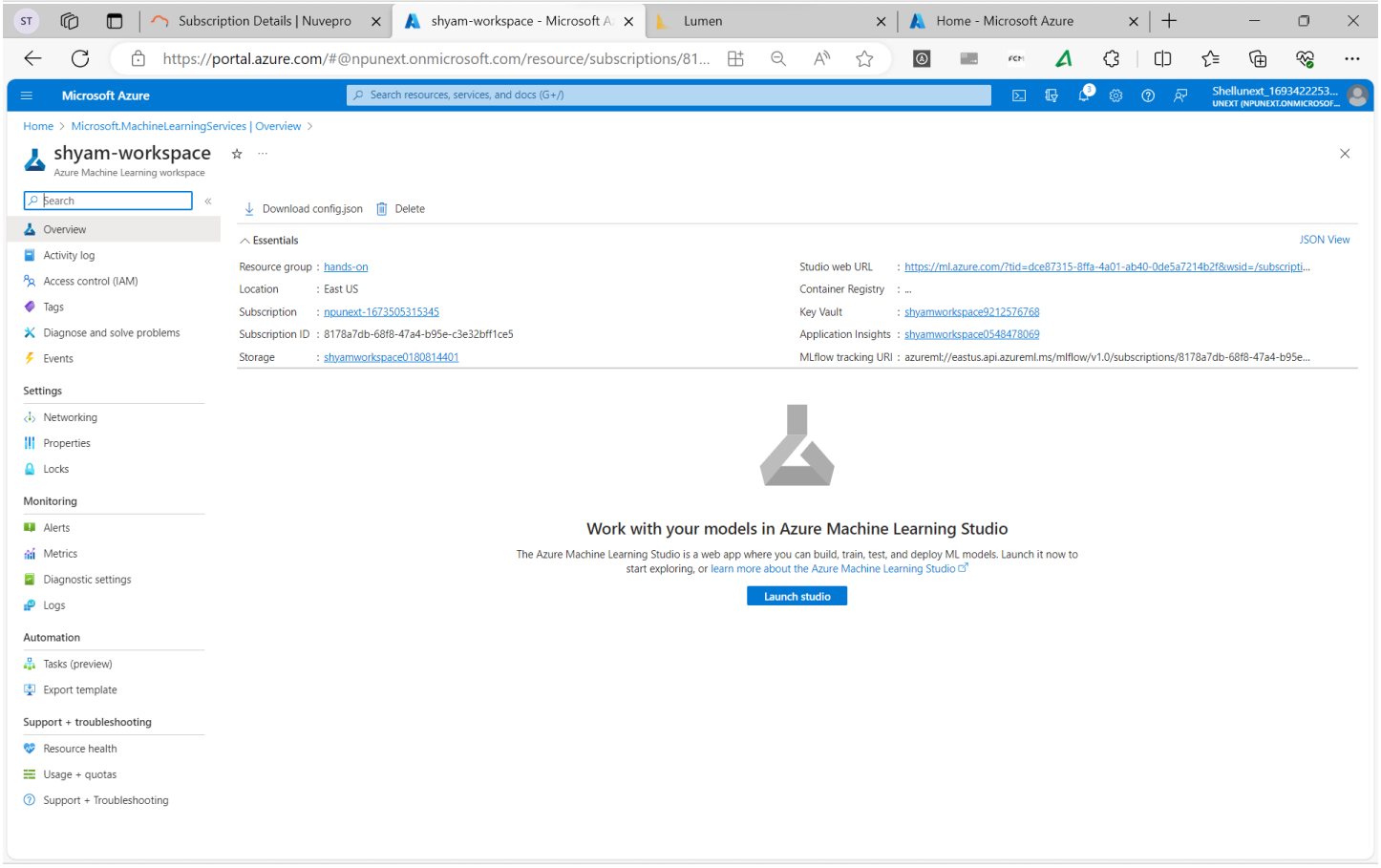
Name: Shyam Tibrewal

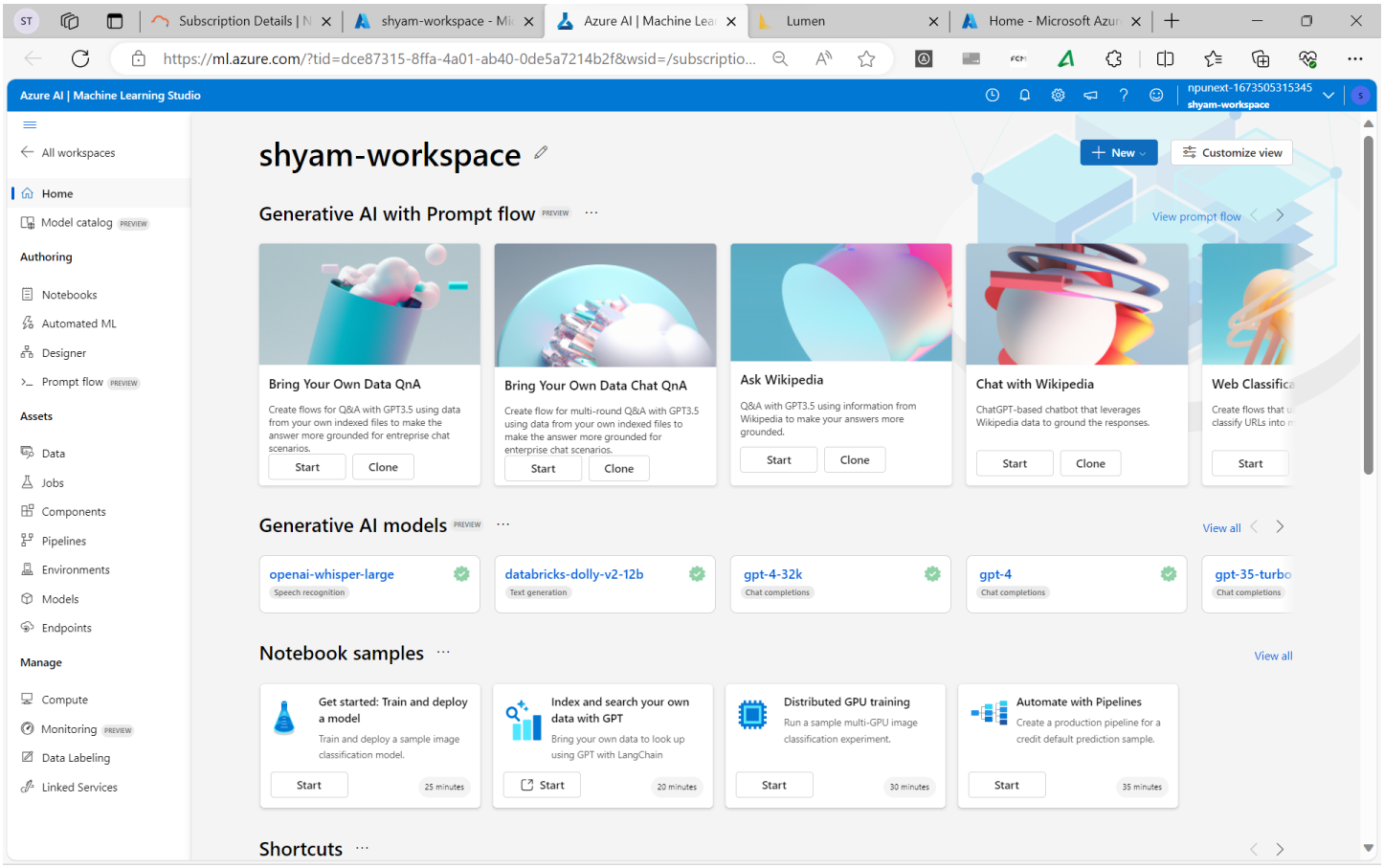
Creating Azure ML Workspace



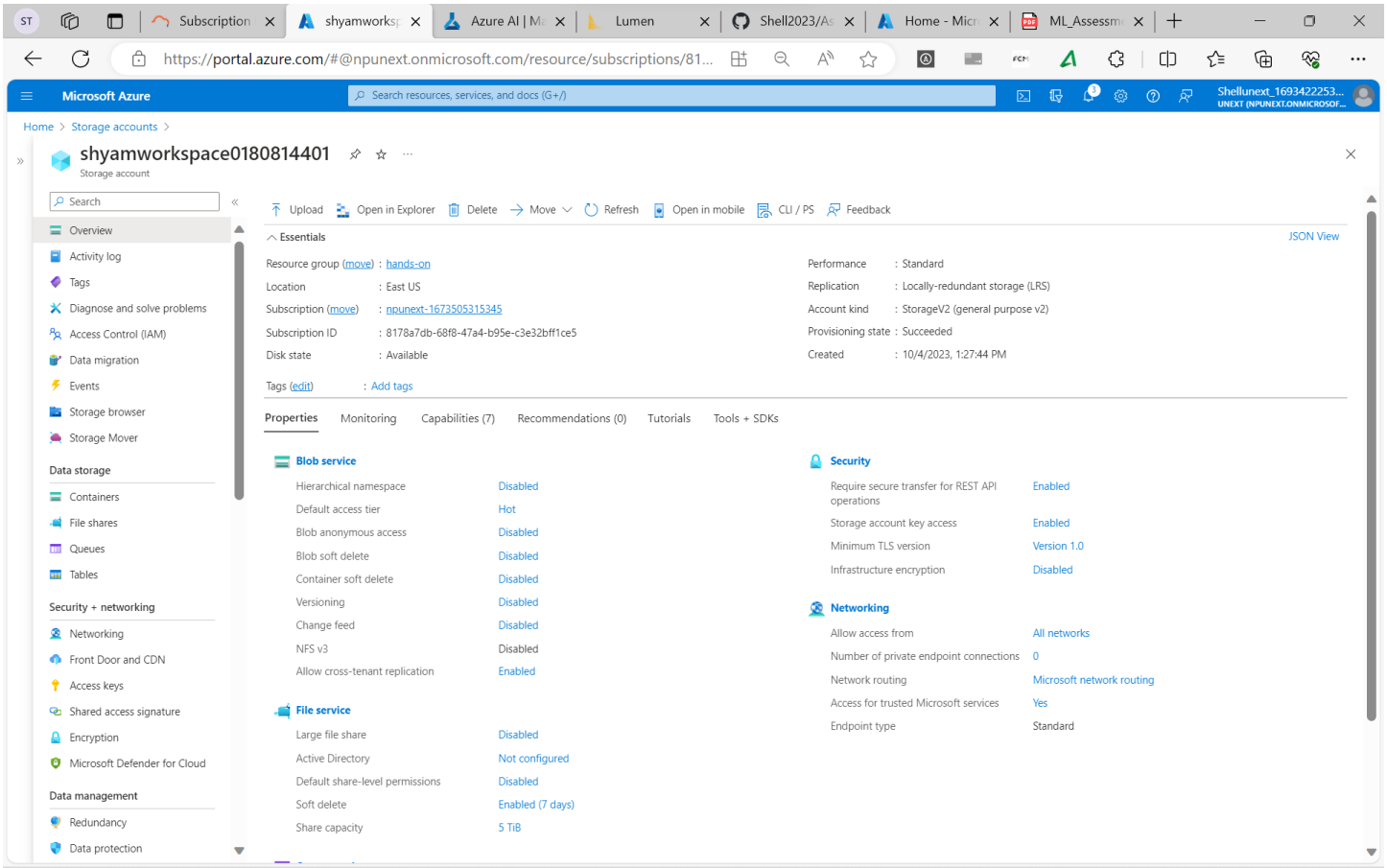


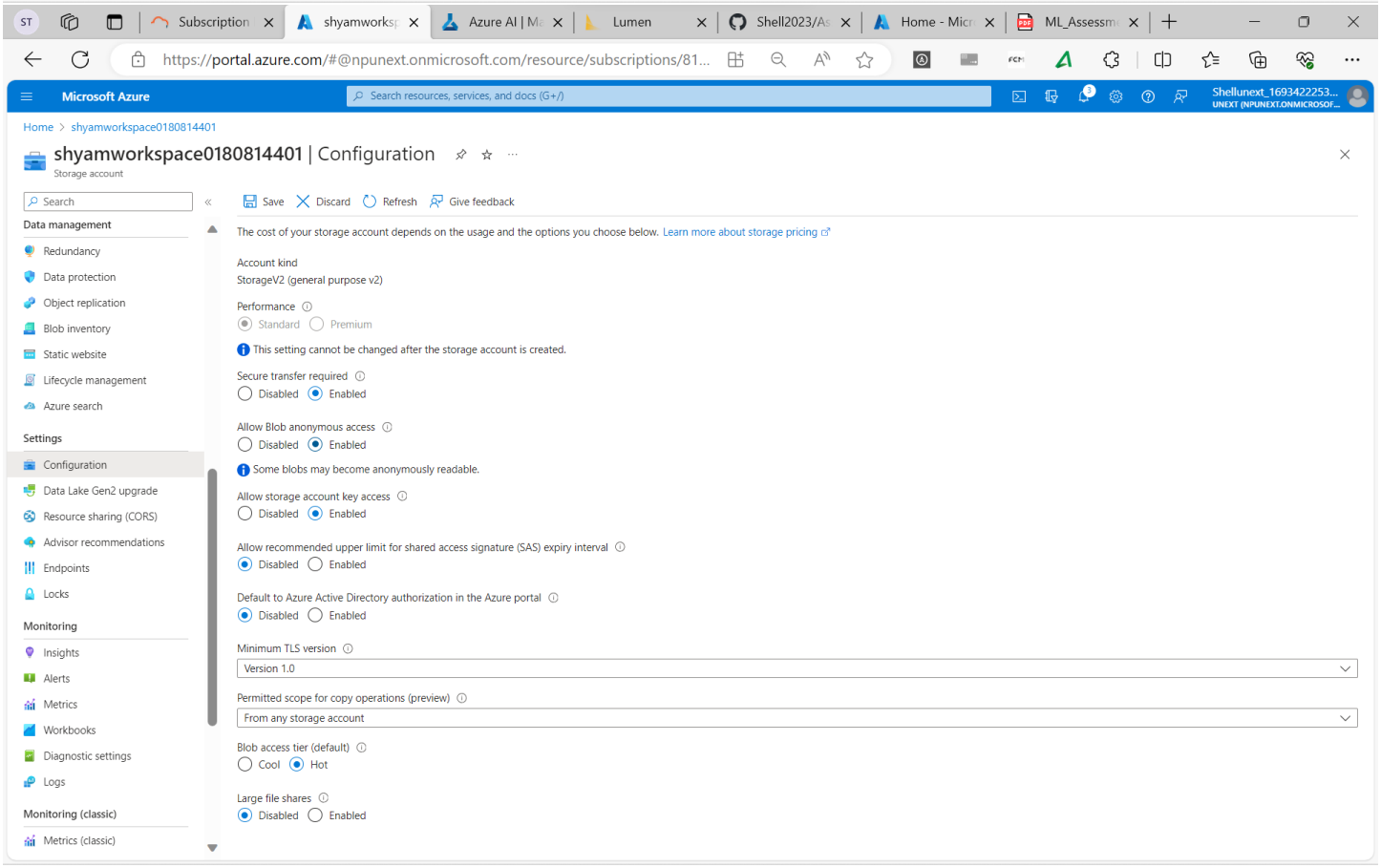

Subscription Details | Nuvepro
Microsoft.MachineLearningServi
Lumen
Home - Microsoft Azure
https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overvi...
Microsoft Azure
Search resources, services, and docs (G+/)
Shellunext_1693422253...
UNEXT (NPUNEXT.ONMICROSOF...
Home
Microsoft.MachineLearningServices | Overview
Deployment
Search
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Download
Refresh
Your deployment is complete
Deployment name: Microsoft.MachineLearningServices
Subscription: npunext-1673505315345
Resource group: hands-on
Start time: 10/4/2023, 1:27:38 PM
Correlation ID: f3116a9d-bded-4d8f-870e-f5a239d665cb
Deployment details
Next steps
Go to resource
Give feedback
Tell us about your experience with deployment
Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >
Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >
Free Microsoft tutorials
Start learning today >
Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >
https://portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/8178a7db-68f8-47a4-b95e-c3e32bff1ce5

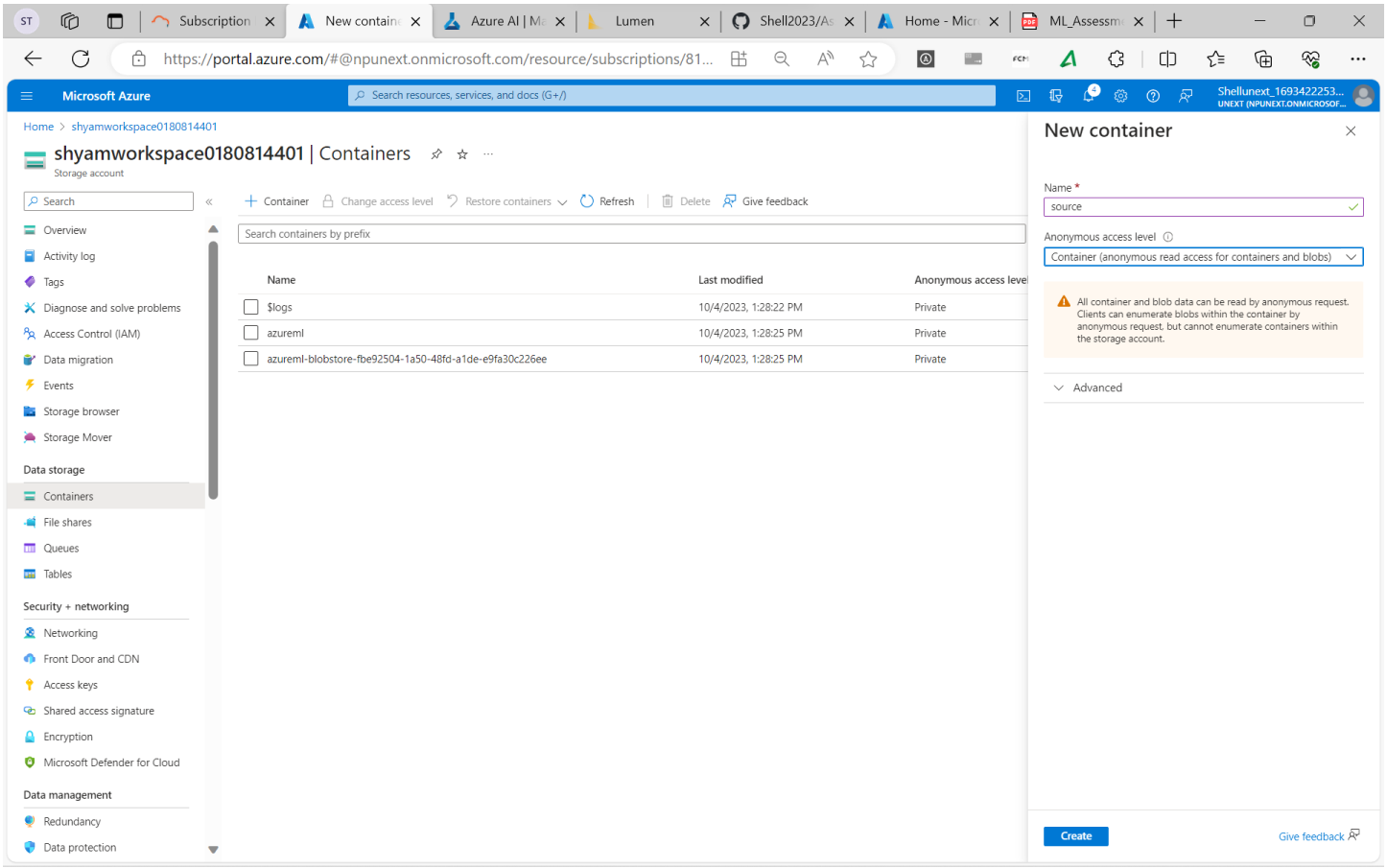

Subscription Details | Nuvepro
shyam-workspace - Microsoft A
Lumen
Home - Microsoft Azure
https://portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/81...
Microsoft Azure
Search resources, services, and docs (G+/)
Shellunext_1693422253...
UNEXT (NPUNEXT.ONMICROSOF...
Home
Microsoft.MachineLearningServices | Overview
shyam-workspace
Azure Machine Learning workspace
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Events
Settings
Networking
Properties
Locks
Monitoring
Alerts
Metrics
Diagnostic settings
Logs
Automation
Tasks (preview)
Export template
Support + troubleshooting
Resource health
Usage + quotas
Support + Troubleshooting
Download config.json
Delete
Essentials
JSON View
Resource group : hands-on
Location : East US
Subscription : npunext-1673505315345
Subscription ID : 8178a7db-68f8-47a4-b95e-c3e32bff1ce5
Storage : shyamworkspace0180814401
Studio web URL : https://ml.azure.com/?tid=dce87315-8ffa-4a01-ab40-0de5a7214b2f&wsid=/subscripti...
Container Registry : ...
Key Vault : shyamworkspace9212576768
Application Insights : shyamworkspace0548478069
MLflow tracking URI : azureml://eastus.api.azureml.ms/mlflow/v1.0/subscriptions/8178a7db-68f8-47a4-b95e...
Work with your models in Azure Machine Learning Studio
The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or learn more about the Azure Machine Learning Studio
Launch studio

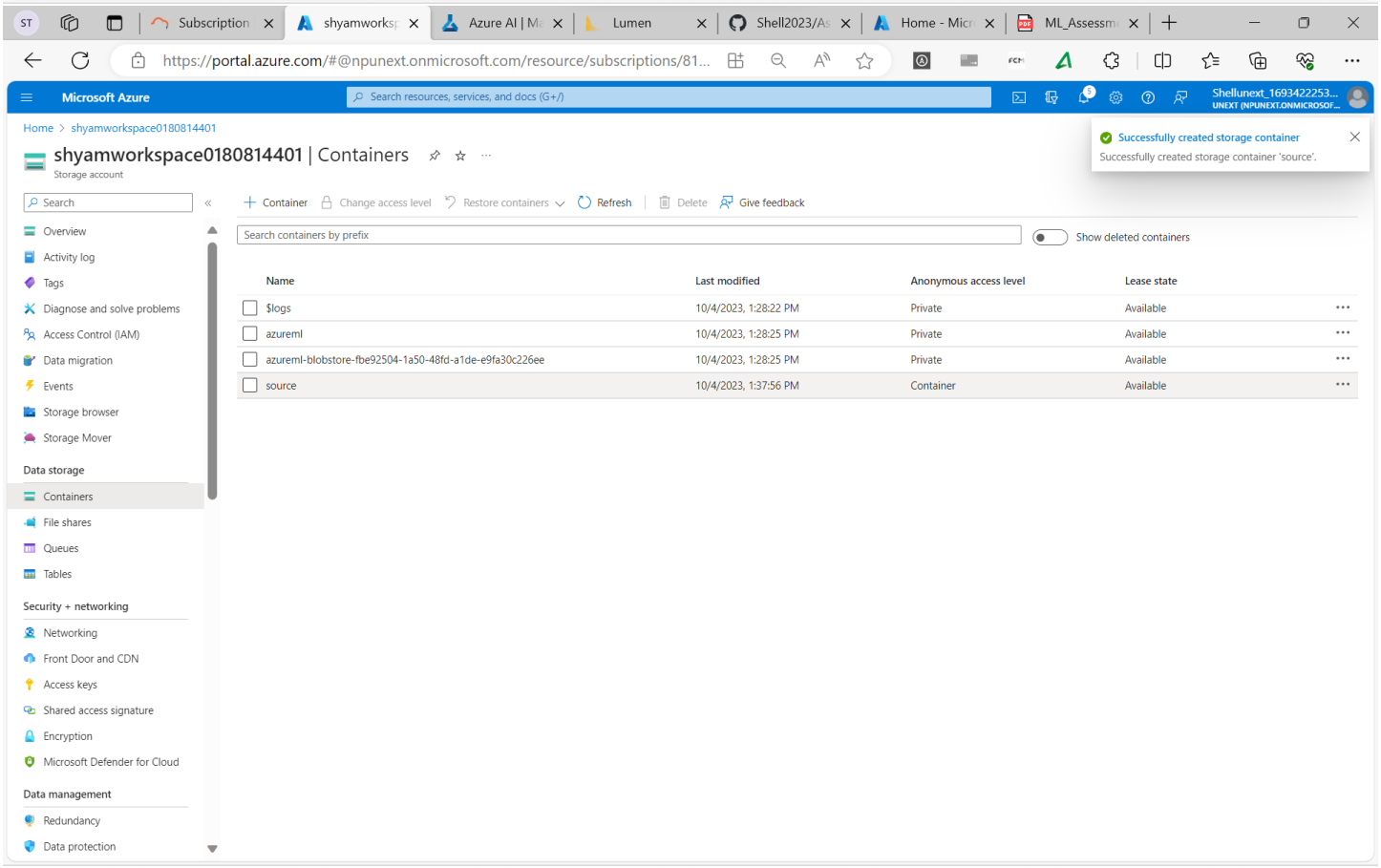

Azure AI | Machine Learning Studio
npunext-1673505315345
shyam-workspace
All workspaces
Home
Model catalog PREVIEW
Authoring
Notebooks
Automated ML
Designer
Prompt flow PREVIEW
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring PREVIEW
Data Labeling
Linked Services
shyam-workspace
New
Customize view
Generative AI with Prompt flow PREVIEW
View prompt flow
Bring Your Own Data QnA
Create flows for Q&A with GPT3.5 using data from your own indexed files to make the answer more grounded for entreprise chat scenarios.
Start
Clone
Bring Your Own Data Chat QnA
Create flow for multi-round Q&A with GPT3.5 using data from your own indexed files to make the answer more grounded for enterprise chat scenarios.
Start
Clone
Ask Wikipedia
Q&A with GPT3.5 using information from Wikipedia to make your answers more grounded.
Start
Clone
Chat with Wikipedia
ChatGPT-based chatbot that leverages Wikipedia data to ground the responses.
Start
Clone
Start
Generative AI models PREVIEW
View all
openai-whisper-large
Speech recognition
databricks-dolly-v2-12b
Text generation
gpt-4-32k
Chat completions
gpt-4
Chat completions
Notebook samples
View all
Get started: Train and deploy a model
Train and deploy a sample image classification model.
Start
25 minutes
Index and search your own data with GPT
Bring your own data to look up using GPT with LangChain
Start
20 minutes
Distributed GPU training
Run a sample multi-GPU image classification experiment.
Start
30 minutes
Automate with Pipelines
Create a production pipeline for a credit default prediction sample.
Start
35 minutes
Shortcuts

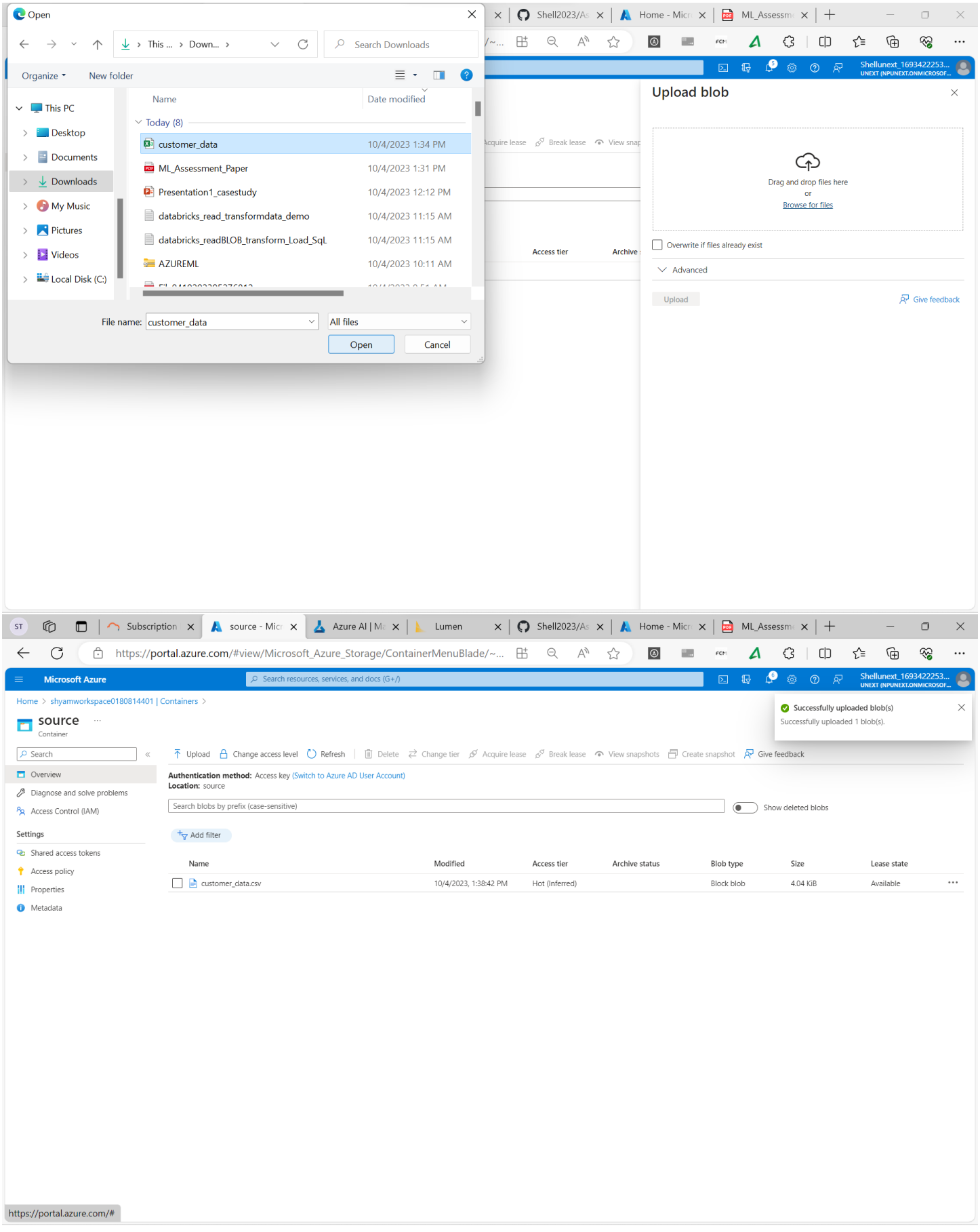
Accessing Data Storage Account Created By Azure ML and uploading the file.



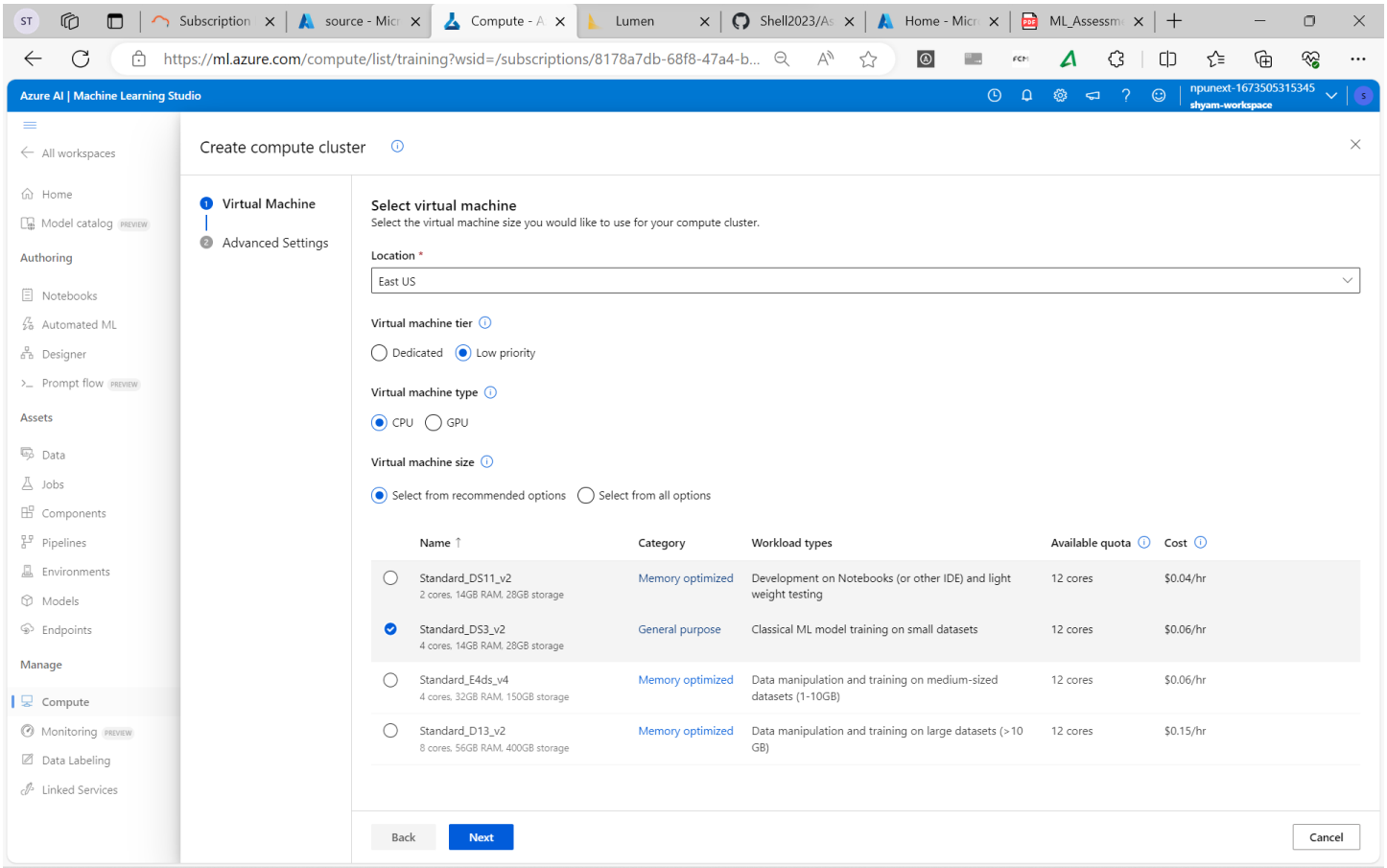


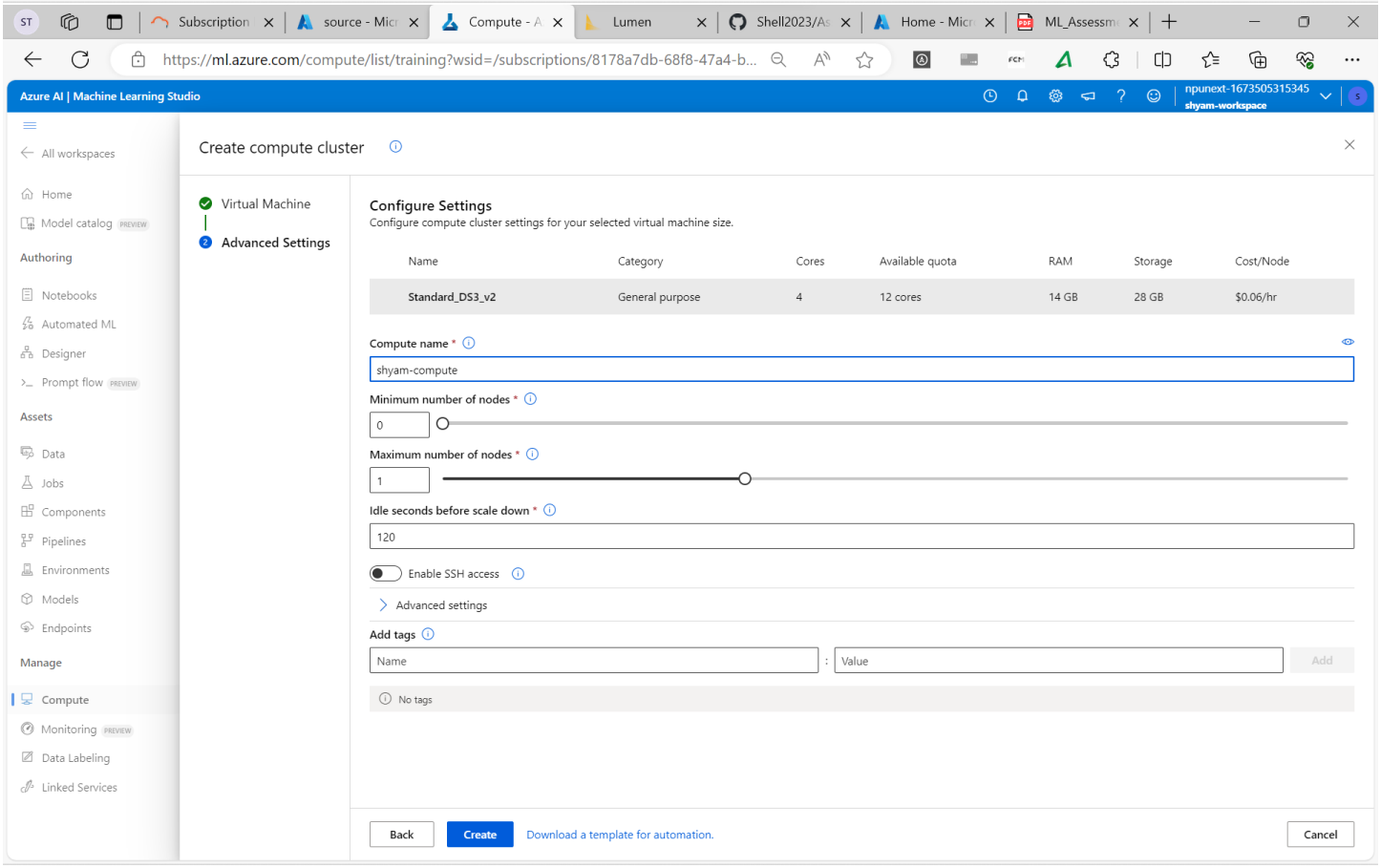

Microsoft Azure
Search resources, services, and docs (G+/)
Home > shyamworkspace0180814401
shyamworkspace0180814401 | Containers
Storage account
Search
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Storage Mover
Data storage
Containers
File shares
Queues
Tables
Security + networking
Networking
Front Door and CDN
Access keys
Shared access signature
Encryption
Microsoft Defender for Cloud
Data management
Redundancy
Data protection
+ Container
Change access level
Restore containers
Refresh
Delete
Give feedback
Search containers by prefix
Name
Last modified
Anonymous access level
$logs
10/4/2023, 1:28:22 PM
Private
azureml
10/4/2023, 1:28:25 PM
Private
azureml-blobstore-fbe92504-1a50-48fd-a1de-e9fa30c226ee
10/4/2023, 1:28:25 PM
Private
New container
Name *
source
Anonymous access level
Container (anonymous read access for containers and blobs)
All container and blob data can be read by anonymous request. Clients can enumerate blobs within the container by anonymous request, but cannot enumerate containers within the storage account.
Advanced
Create
Give feedback


Microsoft Azure
Search resources, services, and docs (G+/)
Home > shyamworkspace0180814401
shyamworkspace0180814401 | Containers
Storage account
Successfully created storage container
Successfully created storage container 'source'.
Search
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Storage Mover
Data storage
Containers
File shares
Queues
Tables
Security + networking
Networking
Front Door and CDN
Access keys
Shared access signature
Encryption
Microsoft Defender for Cloud
Data management
Redundancy
Data protection
+ Container
Change access level
Restore containers
Refresh
Delete
Give feedback
Search containers by prefix
Show deleted containers
Name
Last modified
Anonymous access level
Lease state
$logs
10/4/2023, 1:28:22 PM
Private
Available
azureml
10/4/2023, 1:28:25 PM
Private
Available
azureml-blobstore-fbe92504-1a50-48fd-a1de-e9fa30c226ee
10/4/2023, 1:28:25 PM
Private
Available
source
10/4/2023, 1:37:56 PM
Container
Available

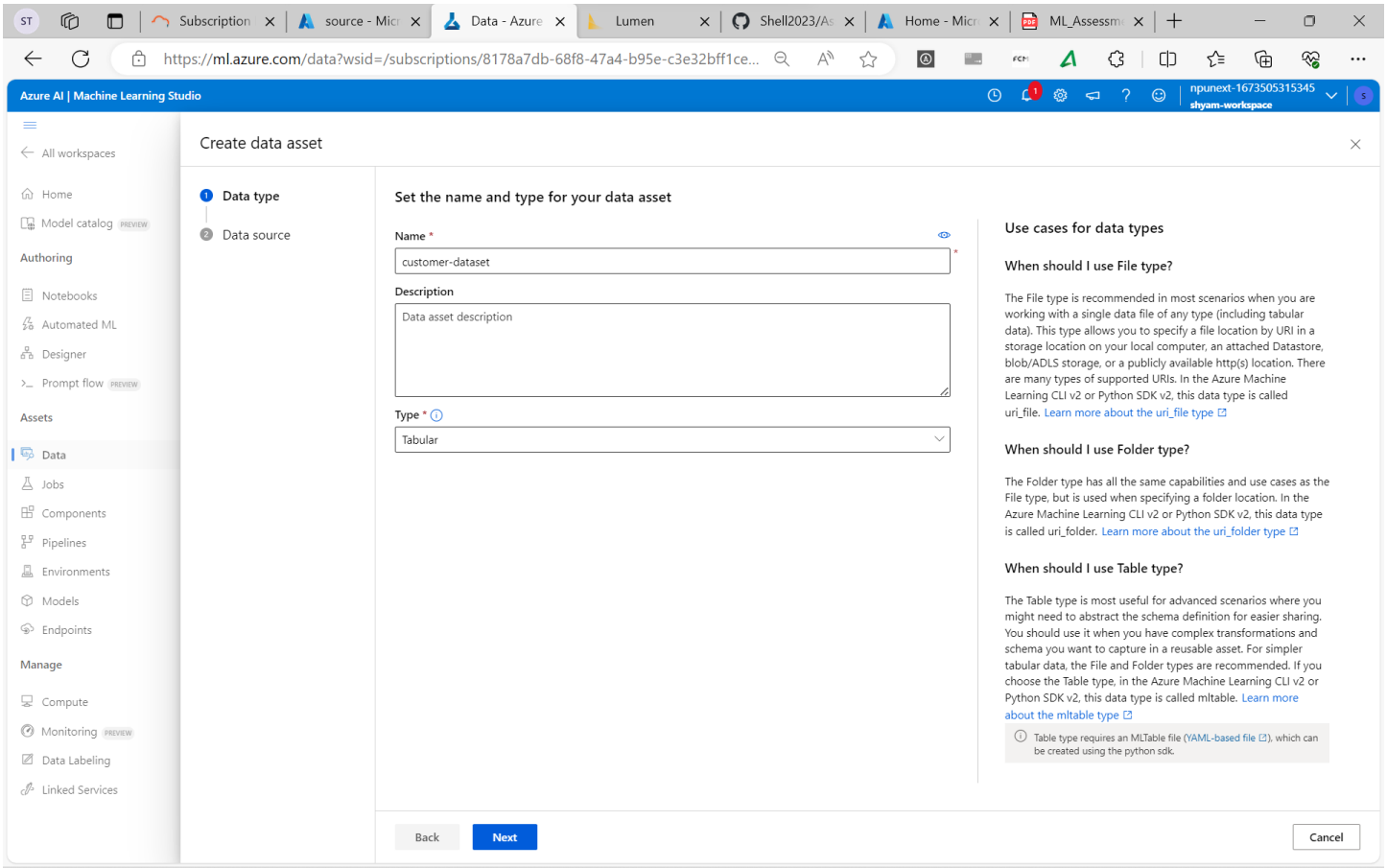

Open
This ... › Down... ›
Search Downloads
Organize
New folder
This PC
Desktop
Documents
Downloads
My Music
Pictures
Videos
Local Disk (C:)
Name
Date modified
Today (8)
customer_data 10/4/2023 1:34 PM
ML_Assessment_Paper 10/4/2023 1:31 PM
Presentation1_casestudy 10/4/2023 12:12 PM
databricks_read_transformdata_demo 10/4/2023 11:15 AM
databricks_readBLOB_transform_Load_SqL 10/4/2023 11:15 AM
AZUREML 10/4/2023 10:11 AM
File name: customer_data
All files
Open
Cancel
Upload blob
Drag and drop files here
or
Browse for files
Overwrite if files already exist
Advanced
Upload
Give feedback
Access tier
Archive
Microsoft Azure
Search resources, services, and docs (G+/)
Home > shyamworkspace0180814401 | Containers >
source
Container
Successfully uploaded blob(s)
Successfully uploaded 1 blob(s).
Search
Upload
Change access level
Refresh
Delete
Change tier
Acquire lease
Break lease
View snapshots
Create snapshot
Give feedback
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: source
Search blobs by prefix (case-sensitive)
Show deleted blobs
Add filter
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
customer_data.csv
10/4/2023, 1:38:42 PM
Hot (Inferred)
Block blob
4.04 KiB
Available
https://portal.azure.com/#

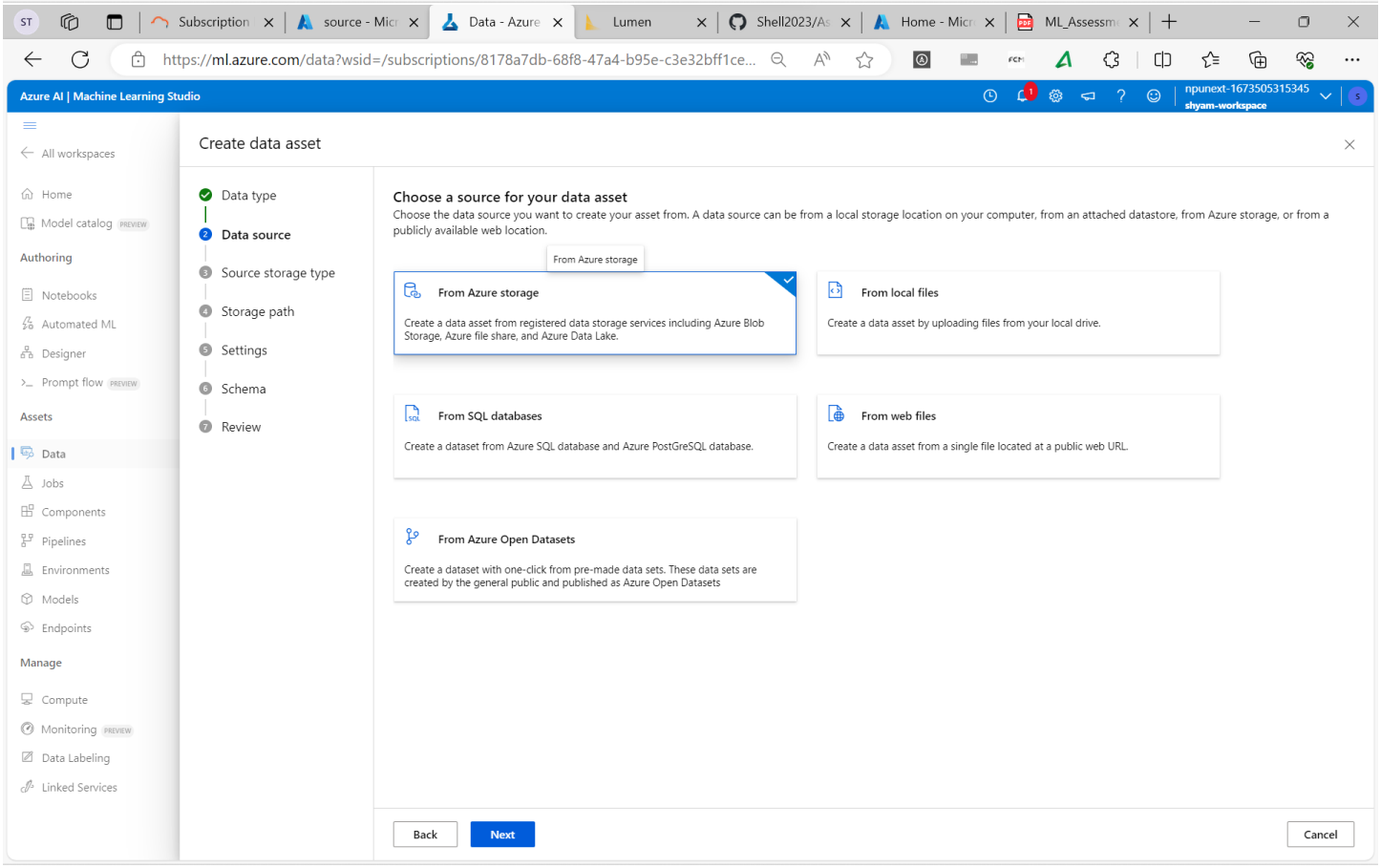
Create Compute Cluster

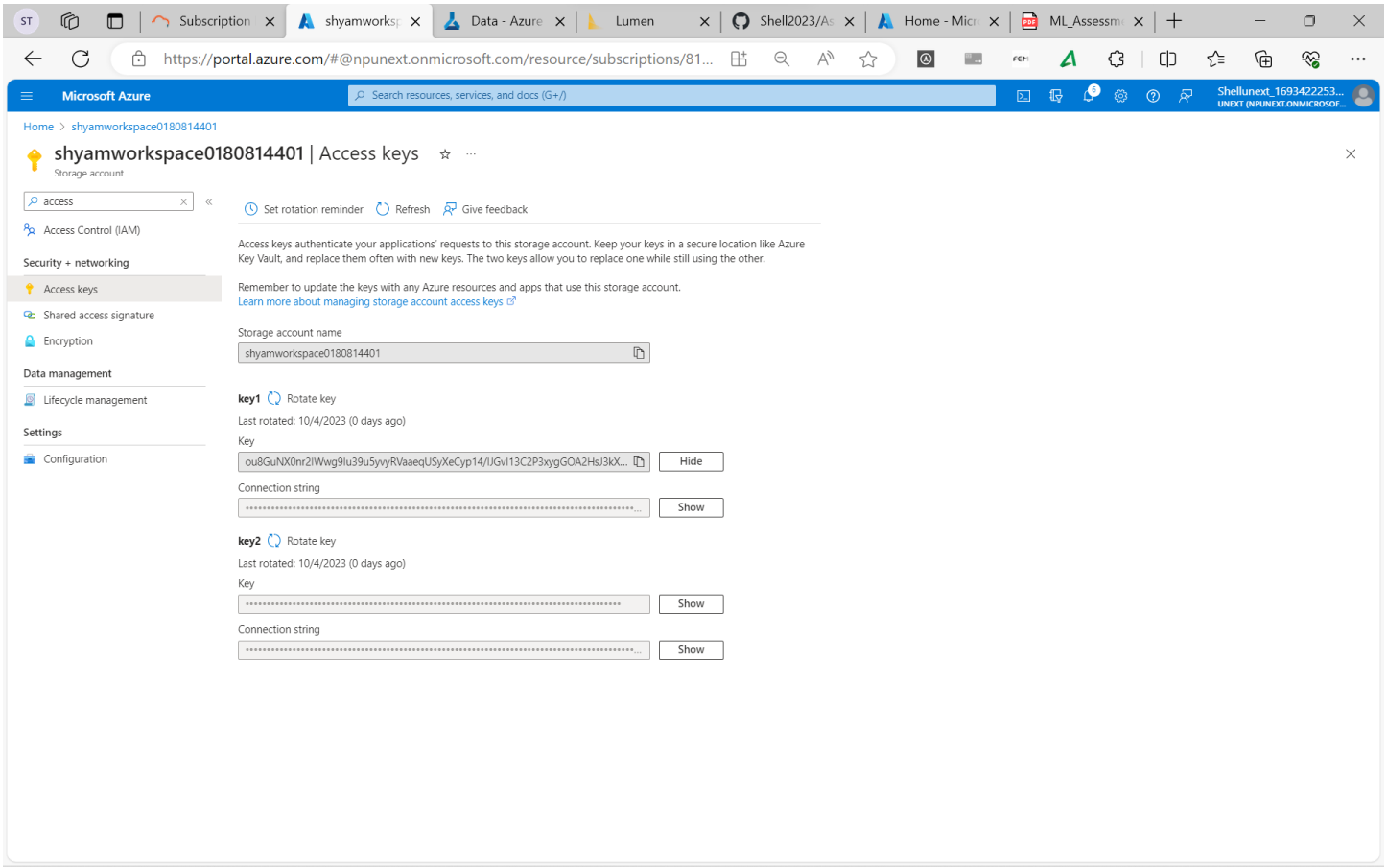


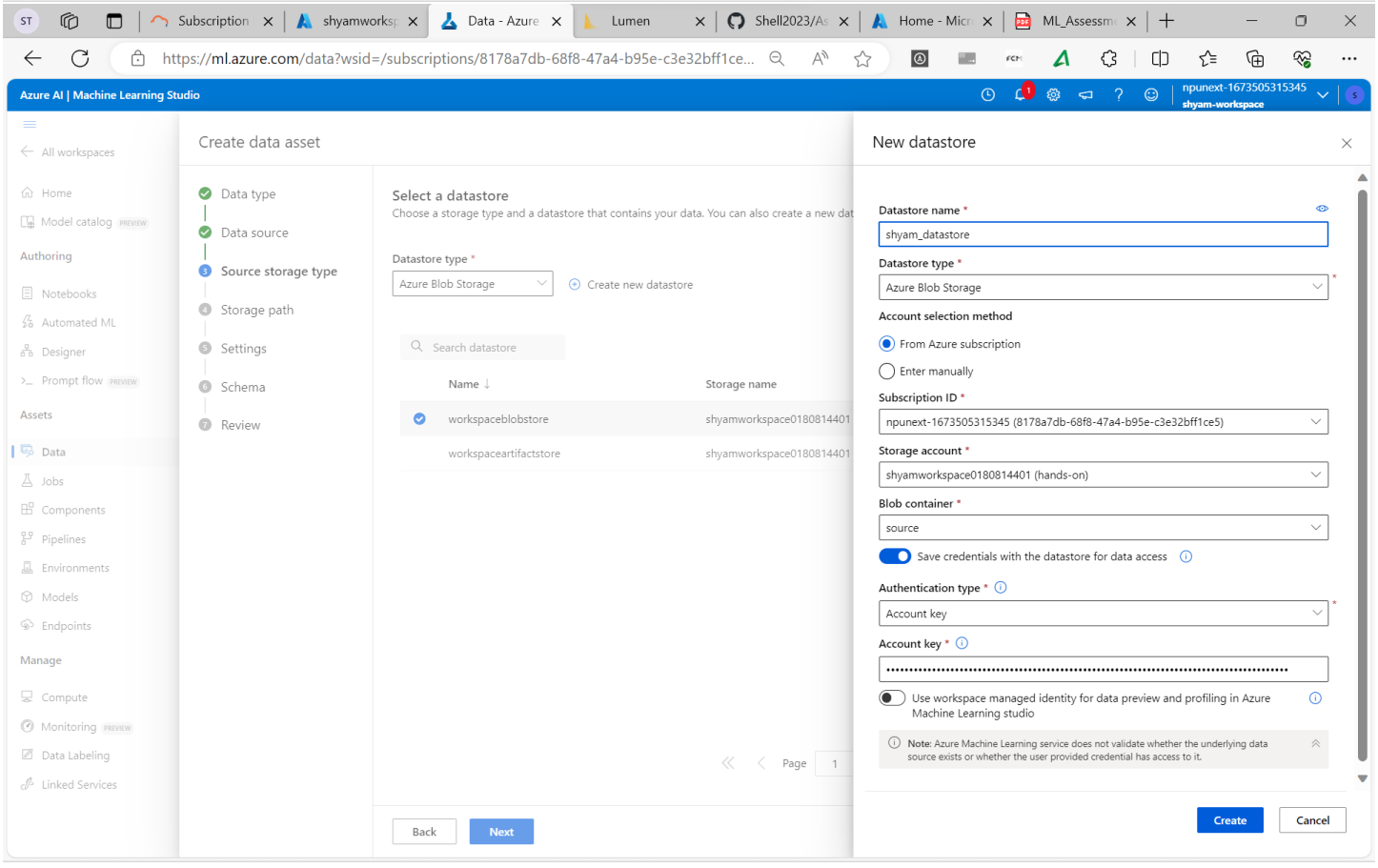


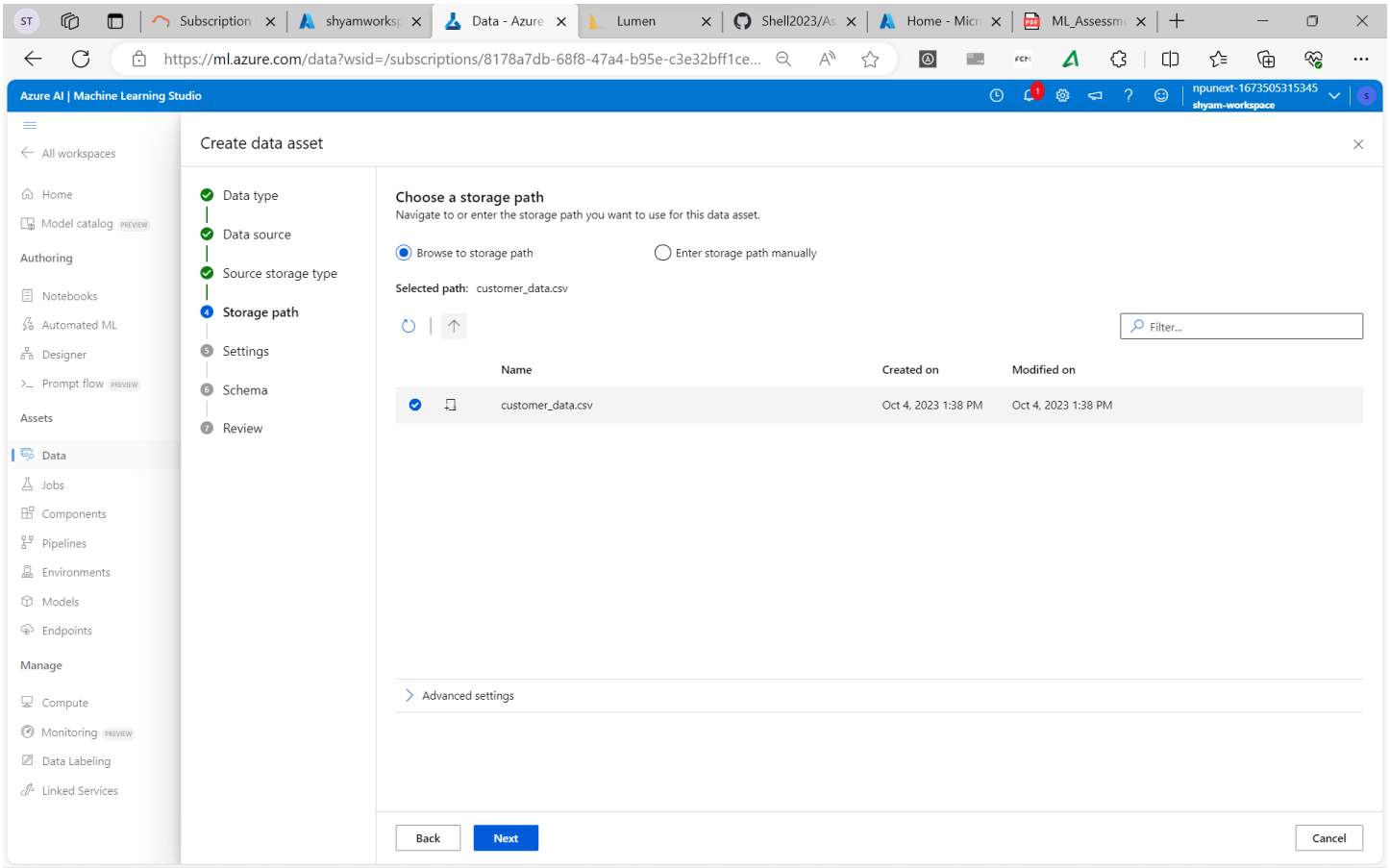
Creating Dataset

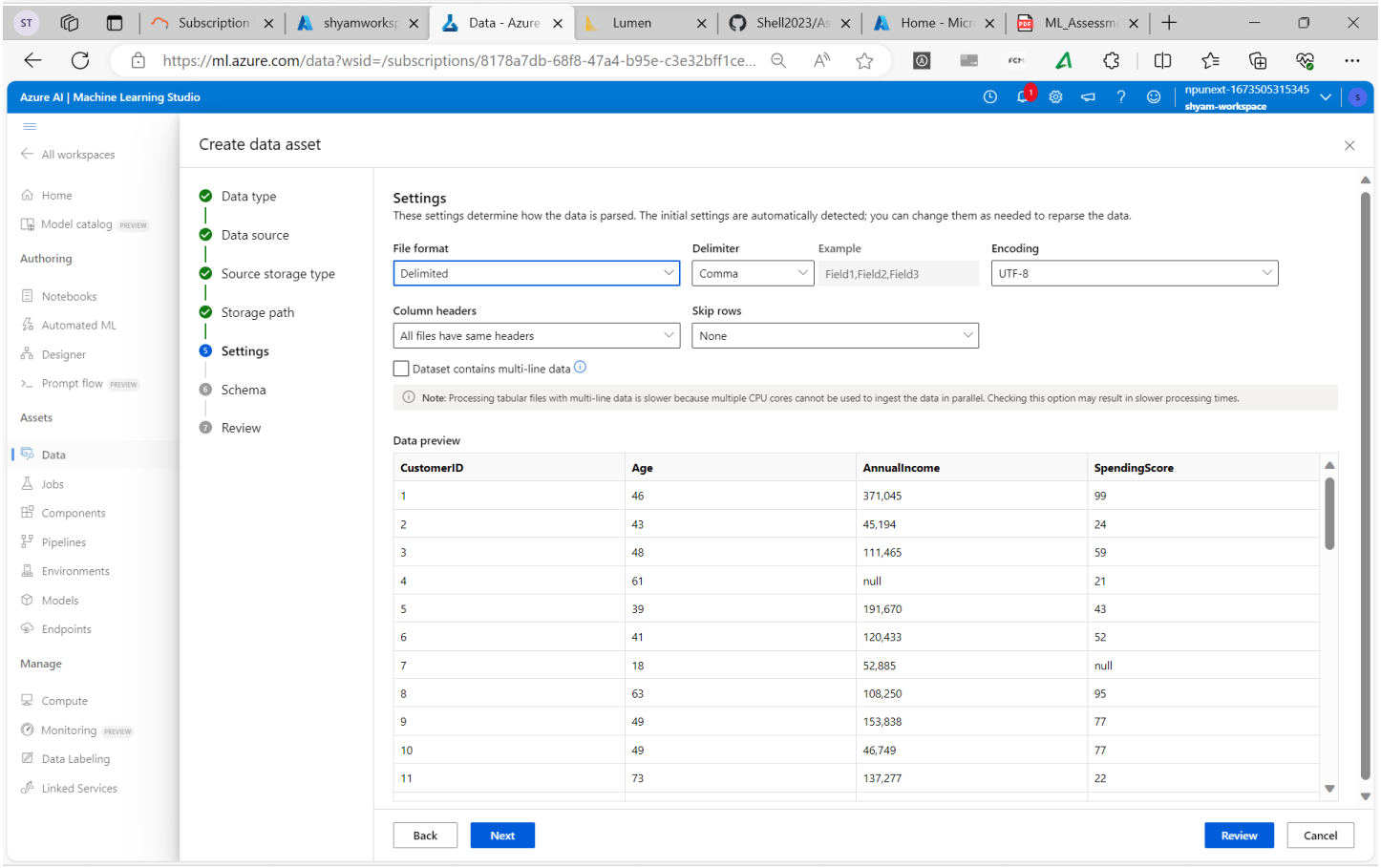


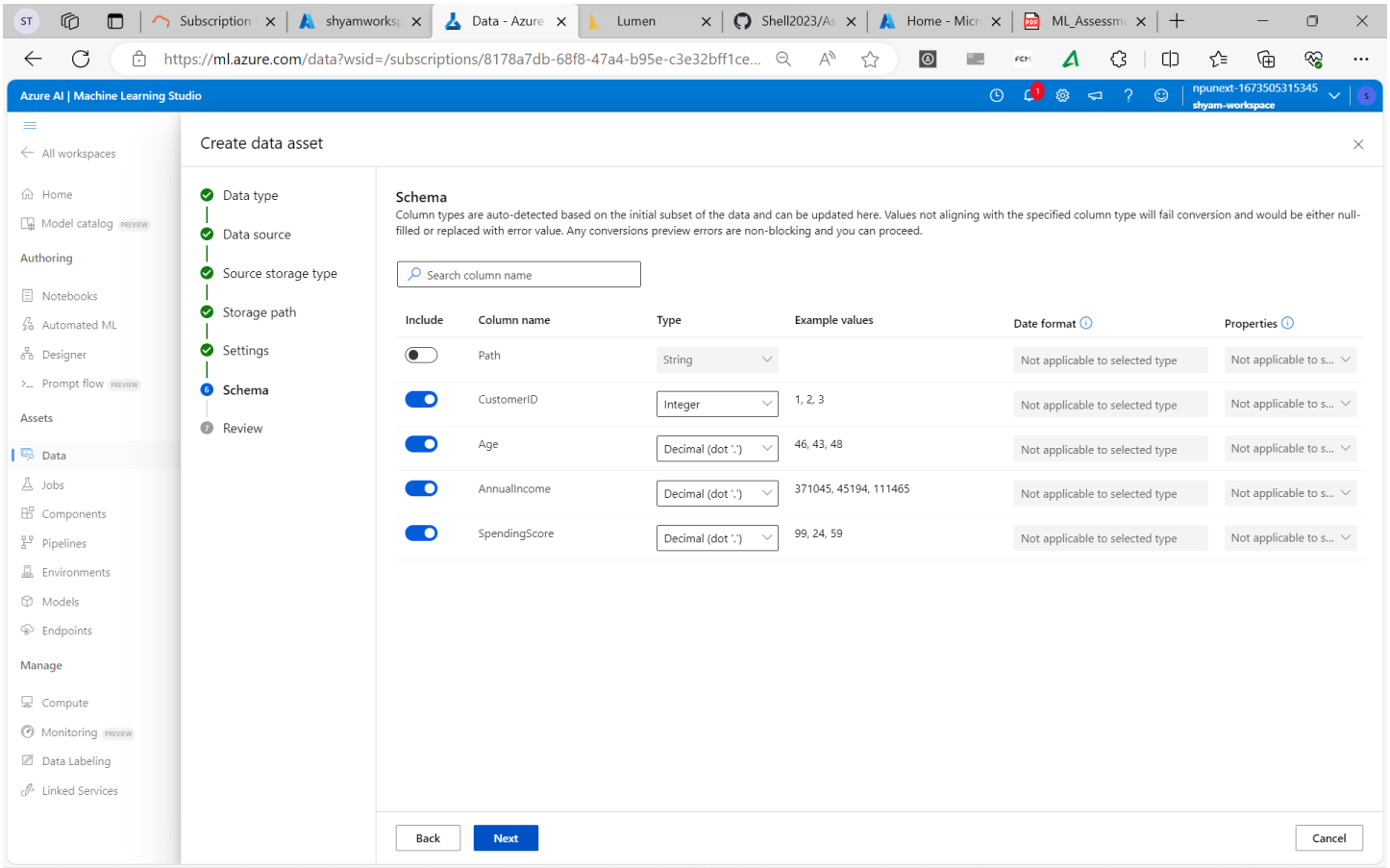


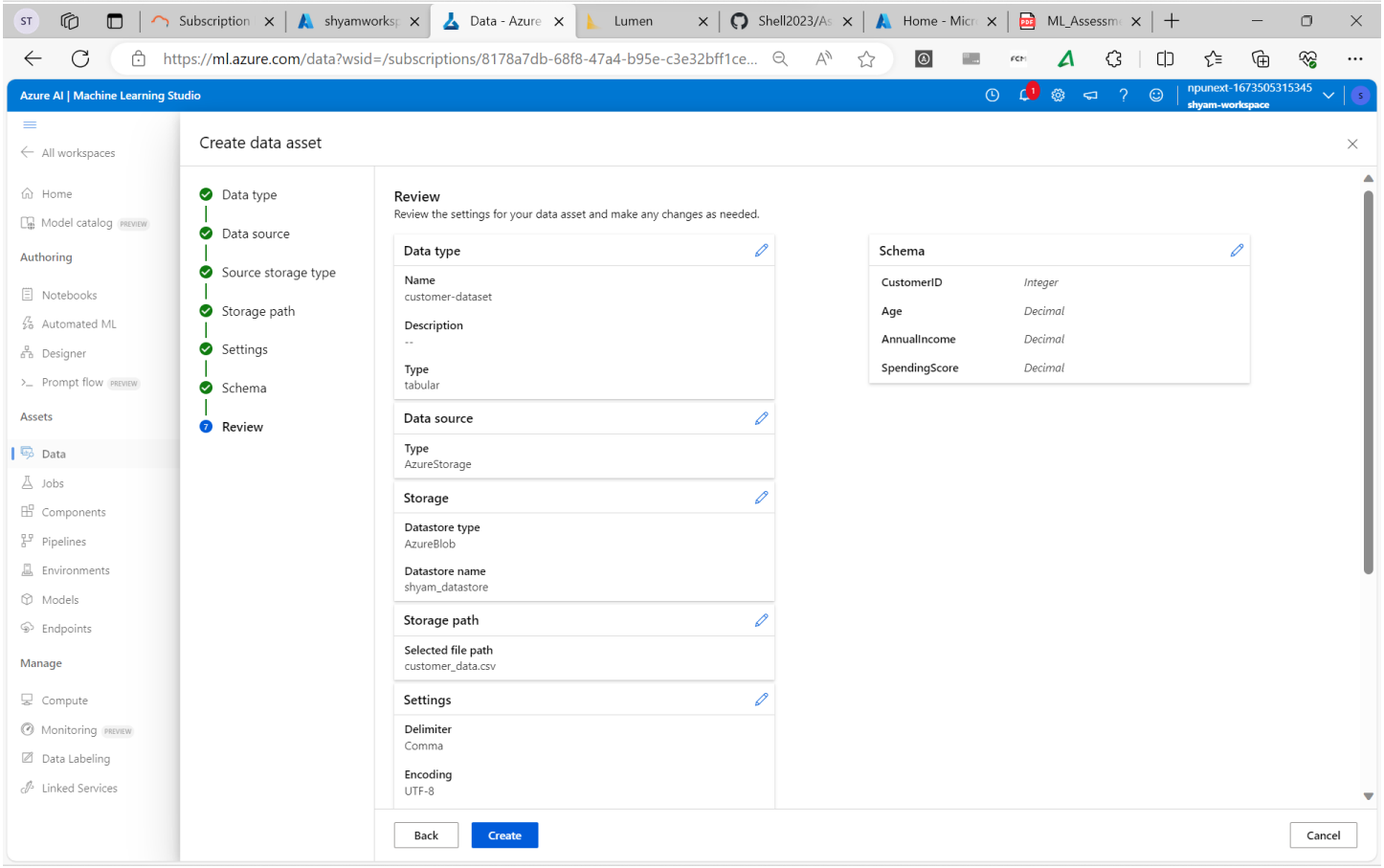

Microsoft Azure
Search resources, services, and docs (G+/)
Home > shyamworkspace0180814401
shyamworkspace0180814401 | Access keys
Storage account
access
Access Control (IAM)
Security + networking
Access keys
Shared access signature
Encryption
Data management
Lifecycle management
Settings
Configuration
Set rotation reminder
Refresh
Give feedback
Access keys authenticate your applications' requests to this storage account. Keep your keys in a secure location like Azure Key Vault, and replace them often with new keys. The two keys allow you to replace one while still using the other.
Remember to update the keys with any Azure resources and apps that use this storage account.
Learn more about managing storage account access keys
Storage account name
shyamworkspace0180814401
key1
Rotate key
Last rotated: 10/4/2023 (0 days ago)
Key
Hide
Connection string
Show
key2
Rotate key
Last rotated: 10/4/2023 (0 days ago)
Key
Show
Connection string
Show

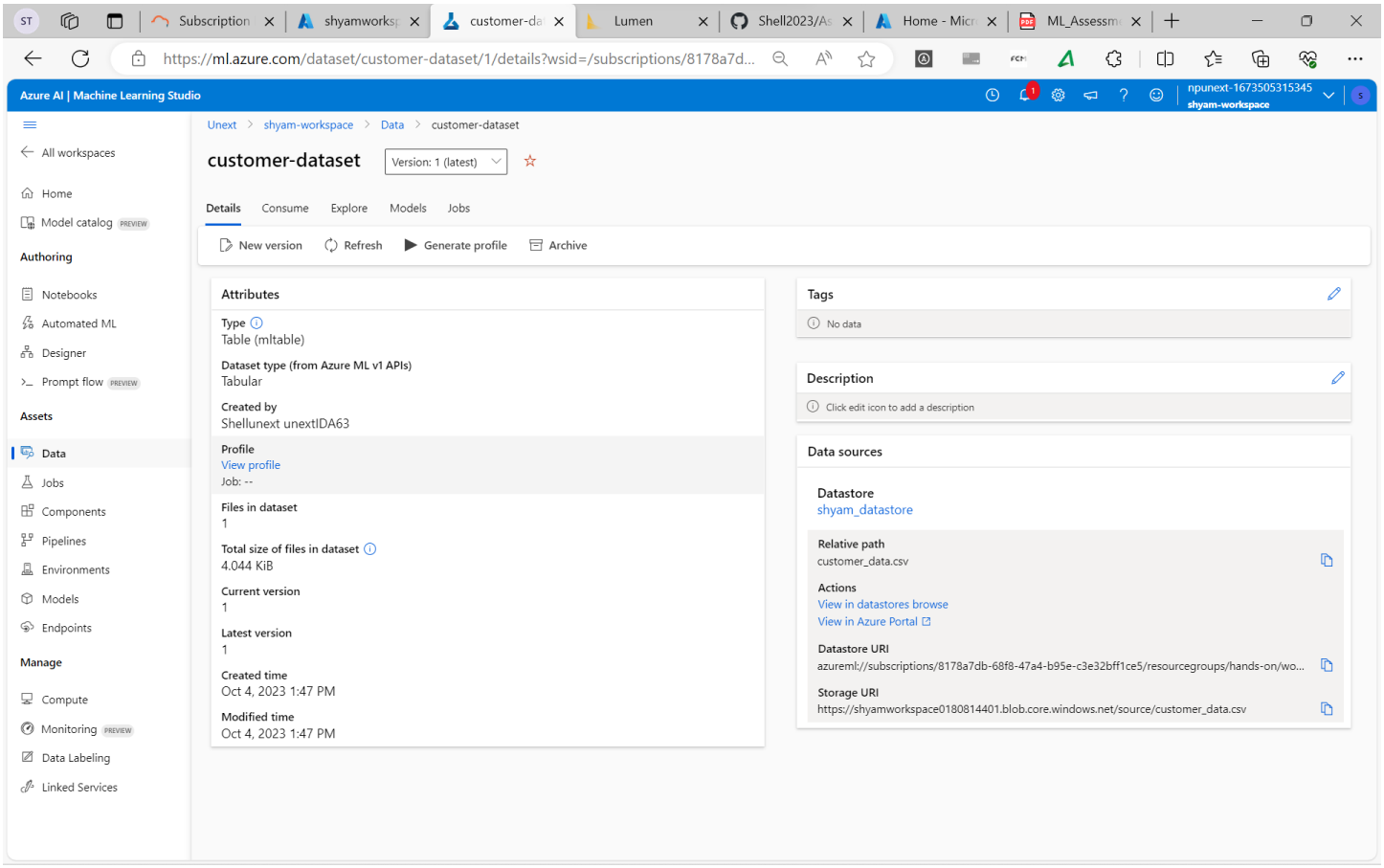

Azure AI | Machine Learning Studio
Create data asset
Data type
Data source
Source storage type
Storage path
Settings
Schema
Review
Select a datastore
Choose a storage type and a datastore that contains your data. You can also create a new dat
Datastore type *
Azure Blob Storage
Create new datastore
Search datastore
Name
Storage name
workspaceblobstore
shyamworkspace0180814401
workspaceartifactstore
shyamworkspace0180814401
Page 1
Back
Next
New datastore
Datastore name *
shyam_datastore
Datastore type *
Azure Blob Storage
Account selection method
From Azure subscription
Enter manually
Subscription ID *
npunext-1673505315345 (8178a7db-68f8-47a4-b95e-c3e32bff1ce5)
Storage account *
shyamworkspace0180814401 (hands-on)
Blob container *
source
Save credentials with the datastore for data access
Authentication type *
Account key
Account key *
Use workspace managed identity for data preview and profiling in Azure Machine Learning studio
Note: Azure Machine Learning service does not validate whether the underlying data source exists or whether the user provided credential has access to it.
Create
Cancel


Azure AI | Machine Learning Studio
Create data asset
Data type
Data source
Source storage type
Storage path
Settings
Schema
Review
Choose a storage path
Navigate to or enter the storage path you want to use for this data asset.
Browse to storage path
Enter storage path manually
Selected path: customer_data.csv
Filter...
Name
Created on
Modified on
customer_data.csv
Oct 4, 2023 1:38 PM
Oct 4, 2023 1:38 PM
Advanced settings
Back
Next
Cancel

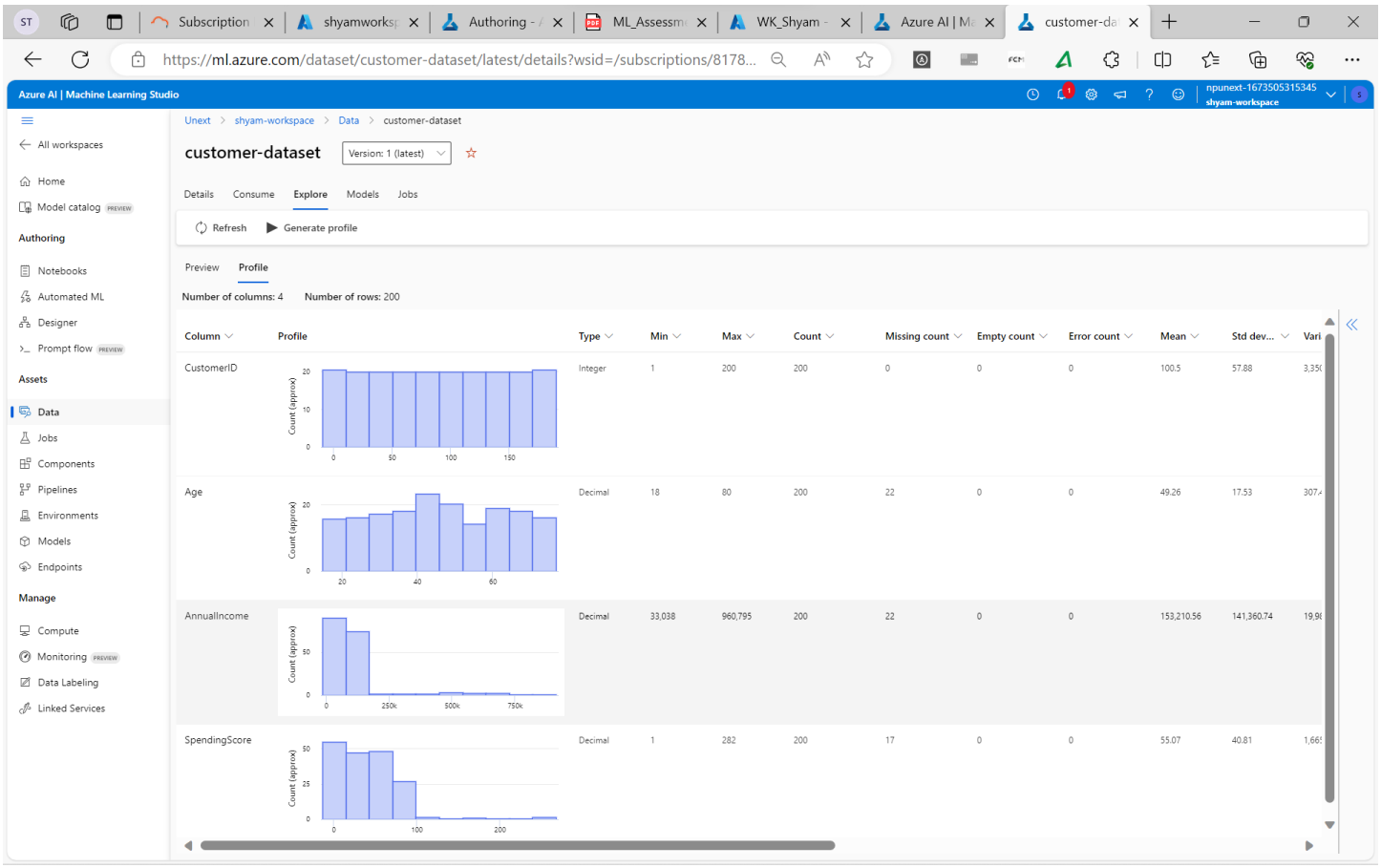

Azure AI | Machine Learning Studio
Create data asset
Settings
These settings determine how the data is parsed. The initial settings are automatically detected; you can change them as needed to reparse the data.
File format
Delimited
Delimiter
Comma
Example
Field1,Field2,Field3
Encoding
UTF-8
Column headers
All files have same headers
Skip rows
None
Dataset contains multi-line data
Note: Processing tabular files with multi-line data is slower because multiple CPU cores cannot be used to ingest the data in parallel. Checking this option may result in slower processing times.
Data preview
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
9 | 49 | 153,838 | 77
10 | 49 | 46,749 | 77
11 | 73 | 137,277 | 22
Back
Next
Review
Cancel

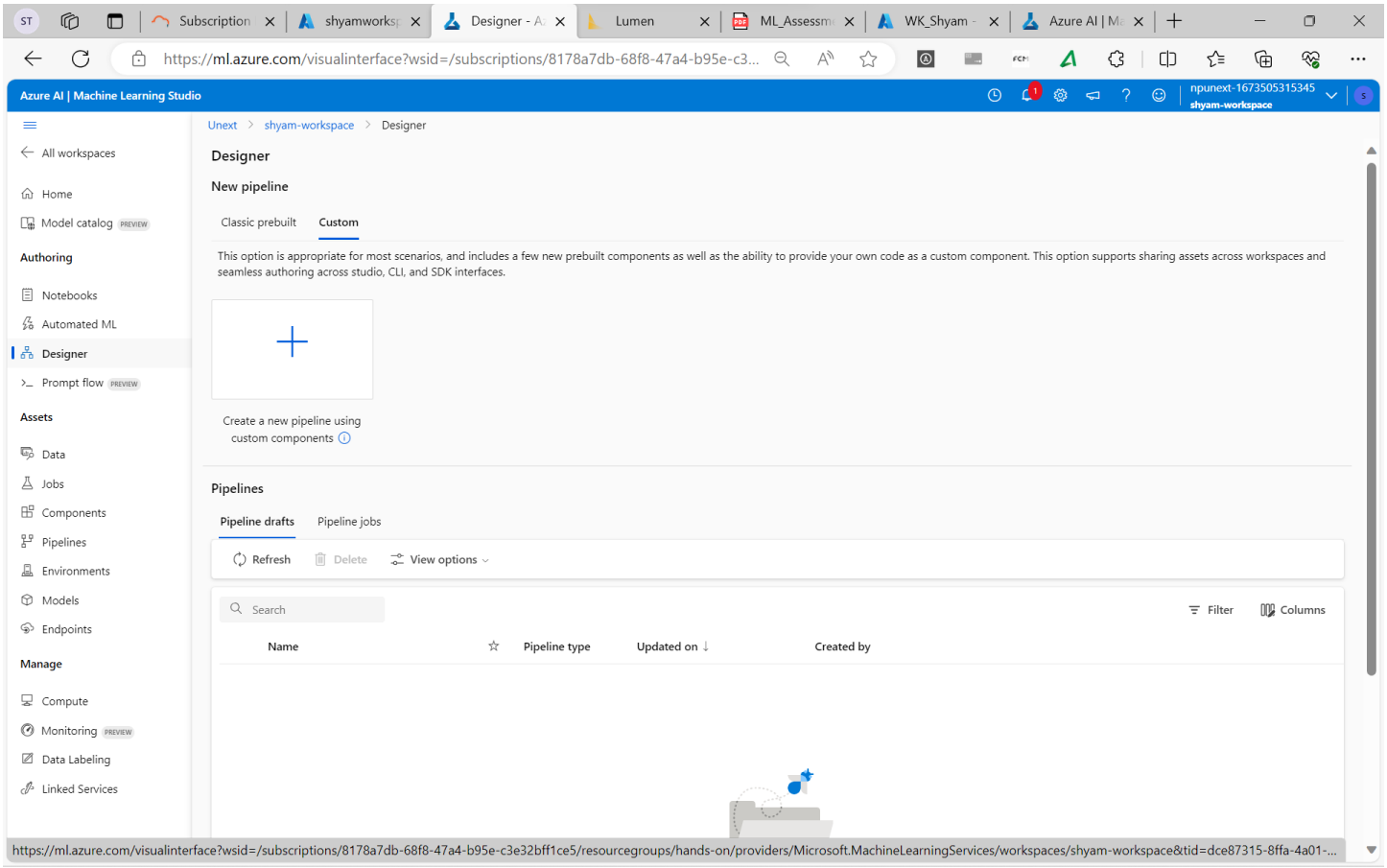

Azure AI | Machine Learning Studio
Create data asset
Data type
Data source
Source storage type
Storage path
Settings
Schema
Review
Schema
Column types are auto-detected based on the initial subset of the data and can be updated here. Values not aligning with the specified column type will fail conversion 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
Include
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 selected type
Back
Next
Cancel

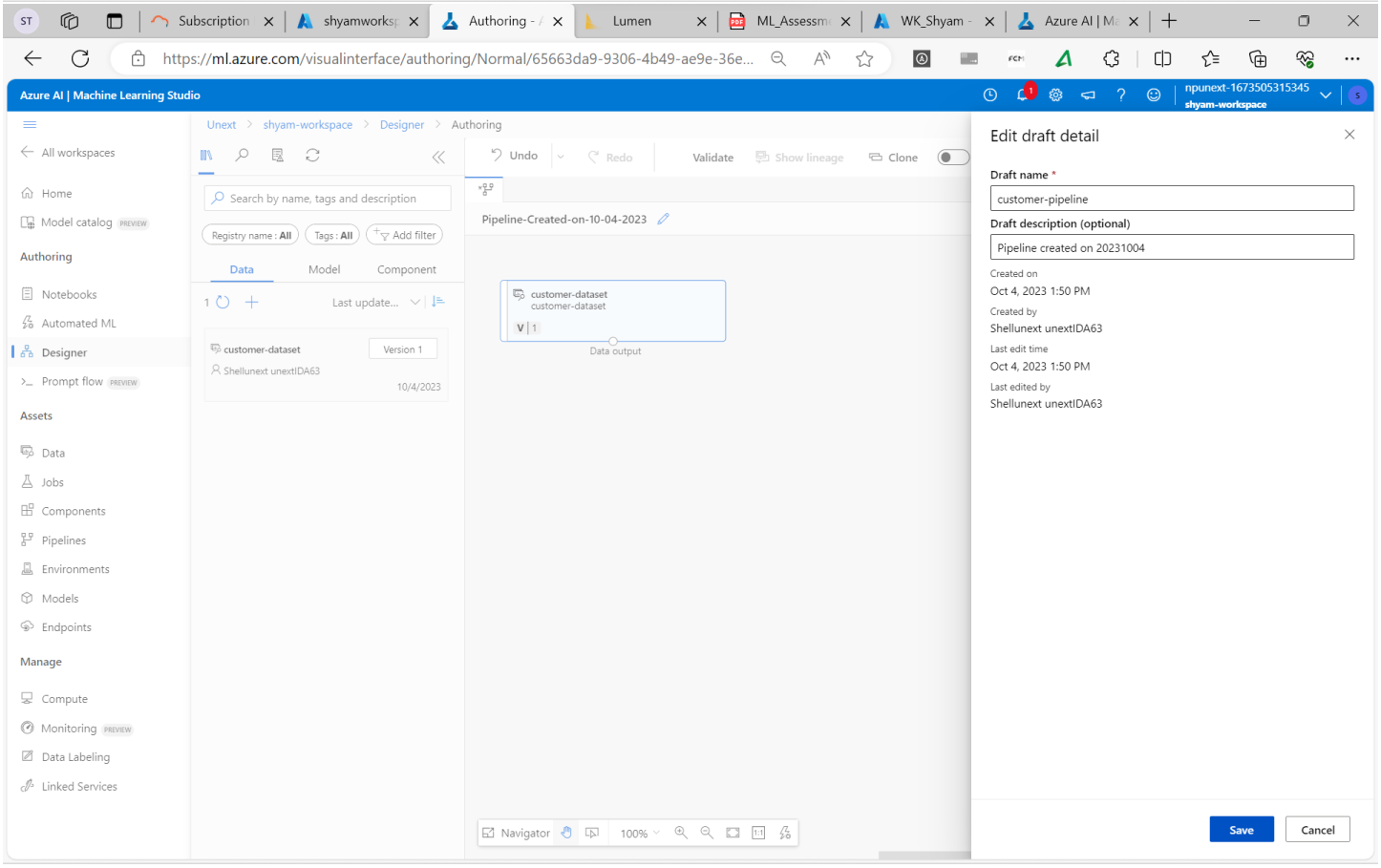

Azure AI | Machine Learning Studio
Create data asset
Review
Review the settings for your data asset and make any changes as needed.
Data type
Name
customer-dataset
Description
--
Type
tabular
Data source
Type
AzureStorage
Storage
Datastore type
AzureBlob
Datastore name
shyam_datastore
Storage path
Selected file path
customer_data.csv
Settings
Delimiter
Comma
Encoding
UTF-8
Schema
CustomerID Integer
Age Decimal
AnnualIncome Decimal
SpendingScore Decimal
Back
Create
Cancel

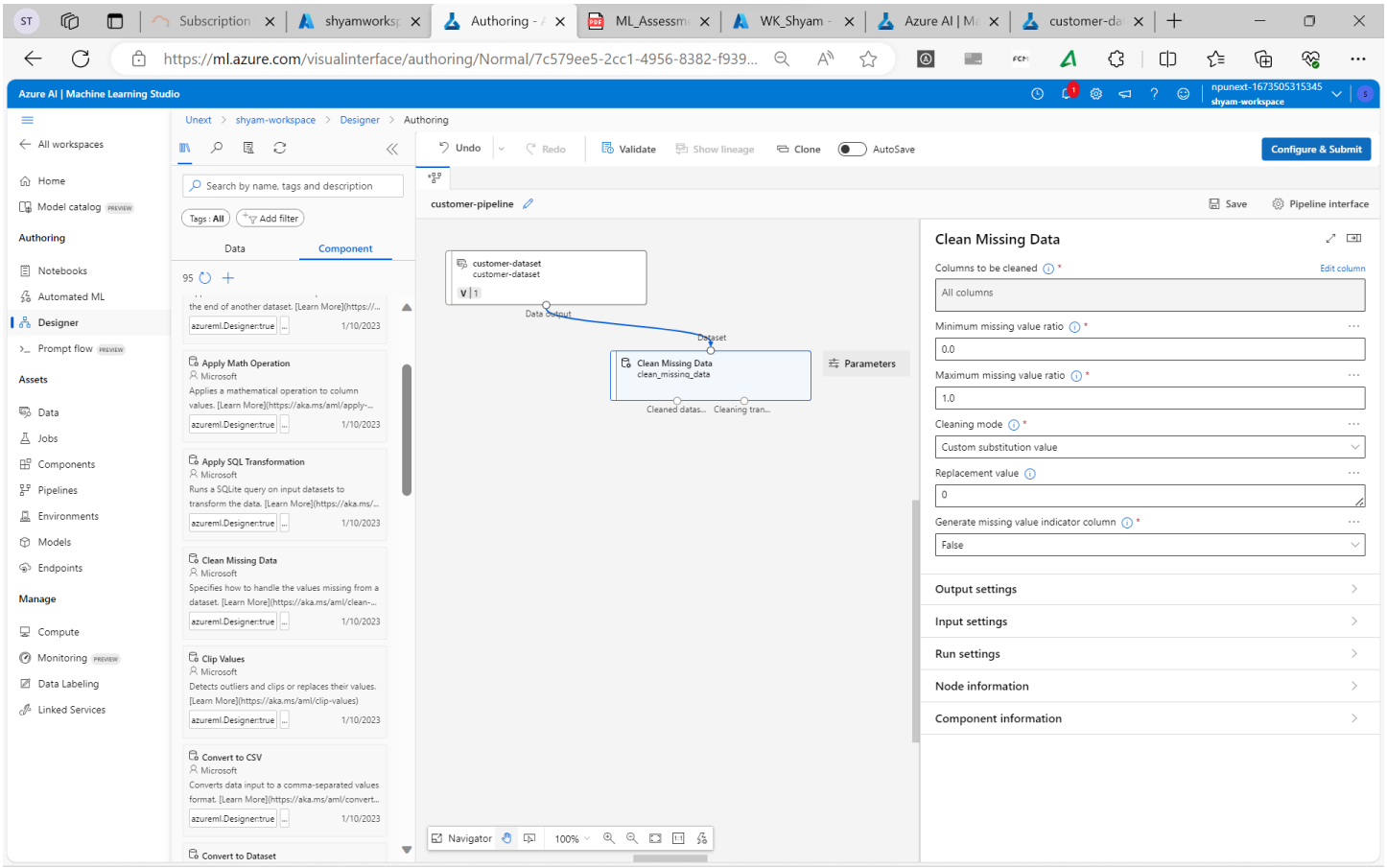


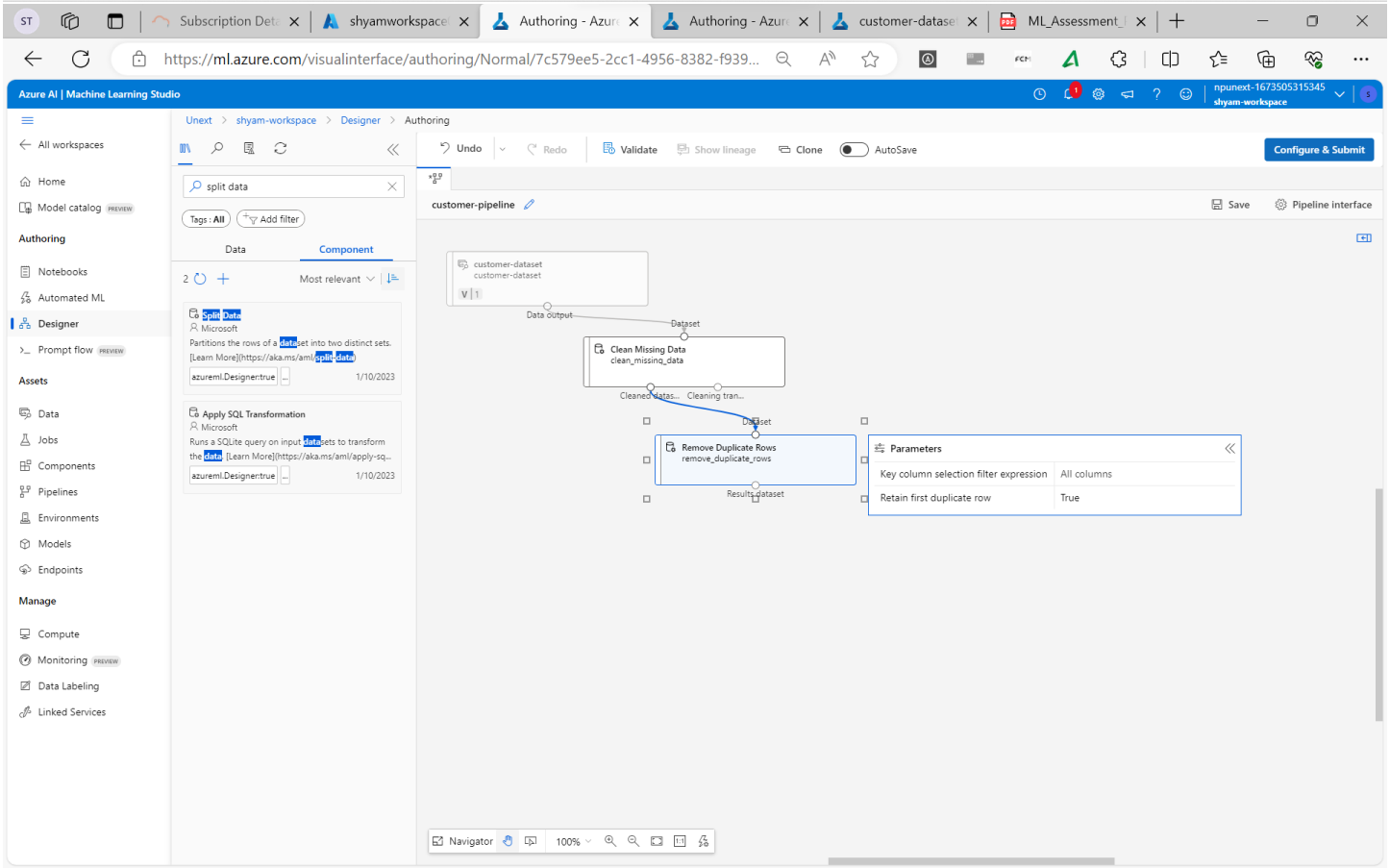
Data Transformation

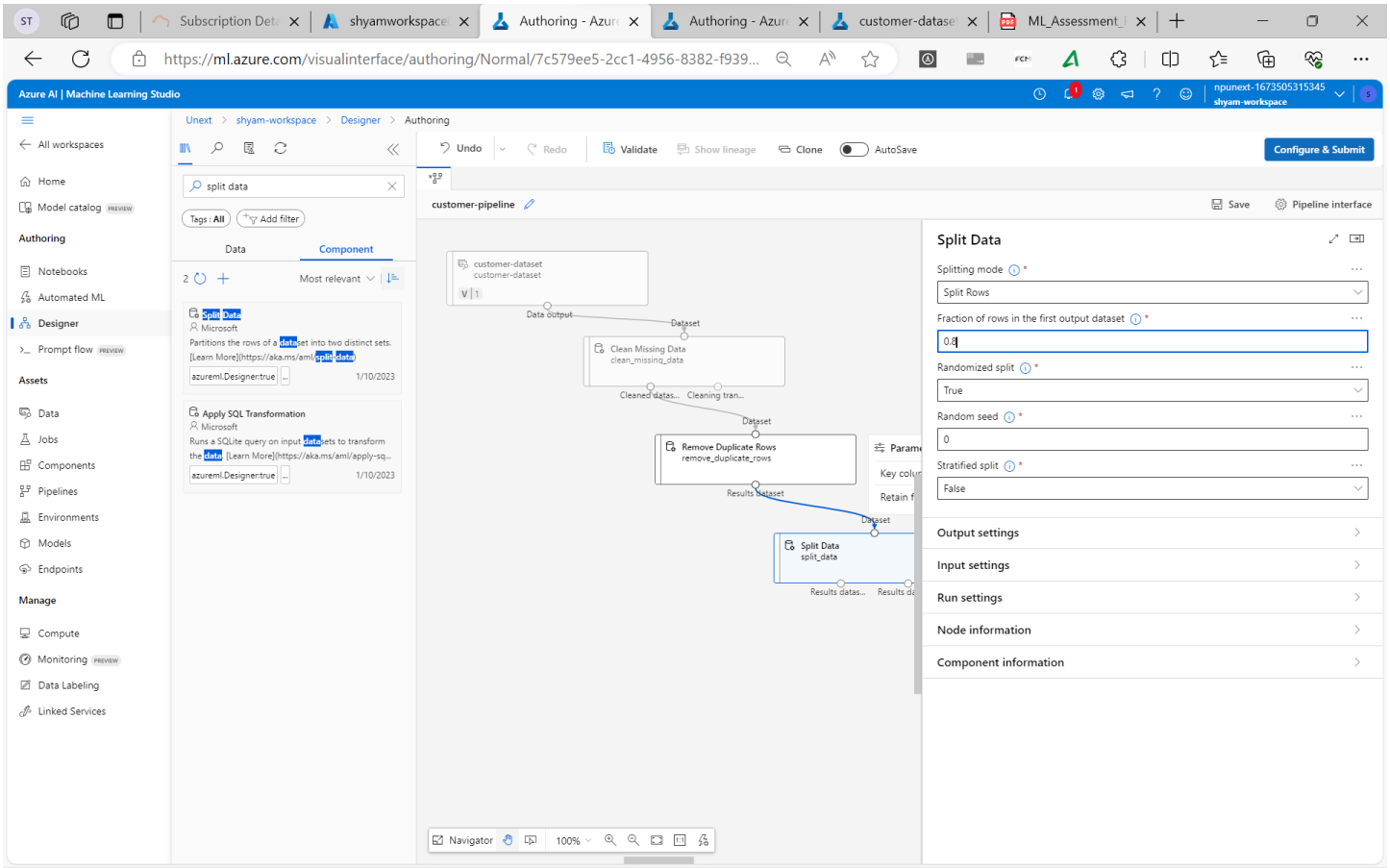


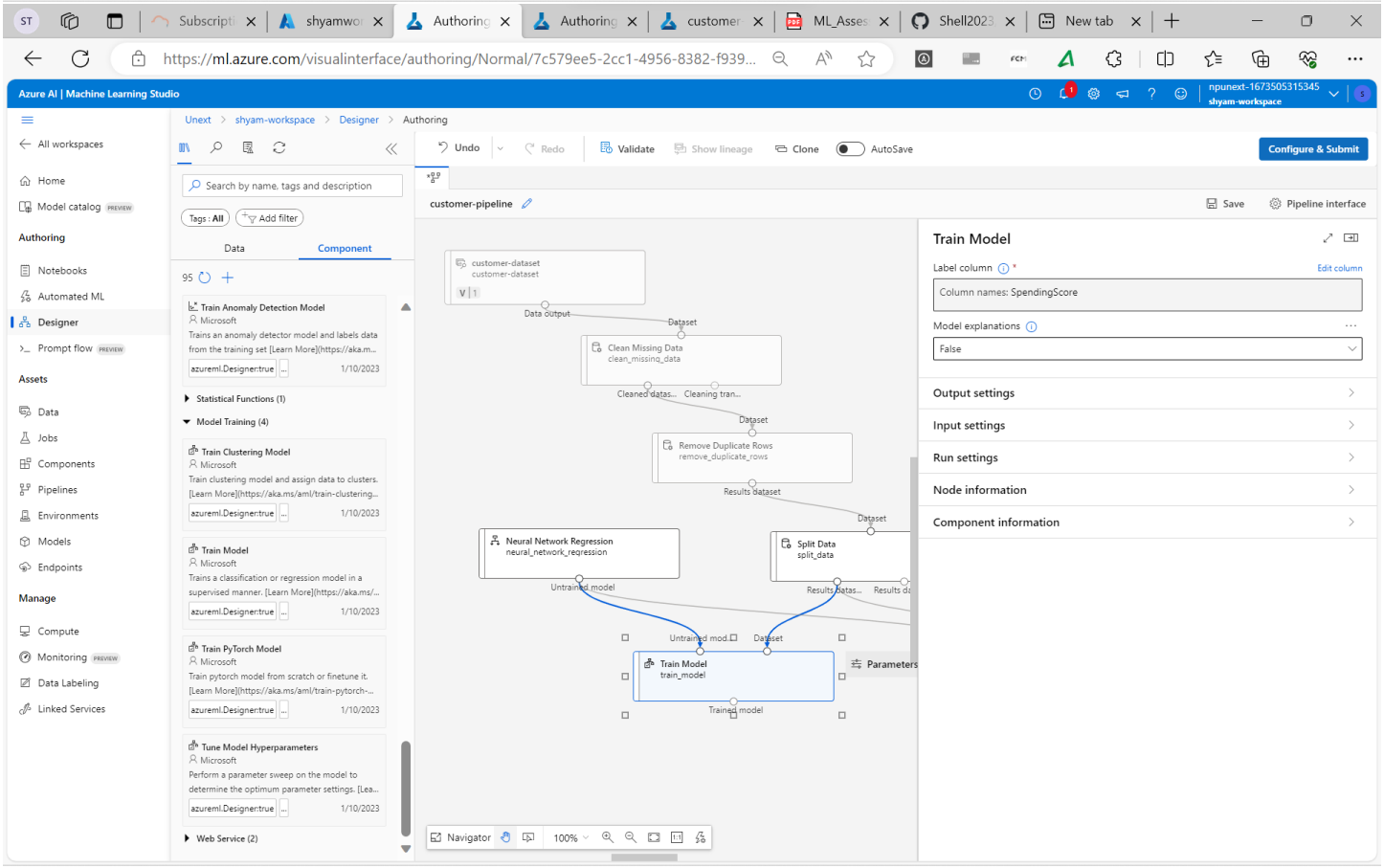

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer
Designer
New pipeline
Classic prebuilt
Custom
This option is appropriate for most scenarios, and includes a few new prebuilt components as well as the ability to provide your own code as a custom component. This option supports sharing assets across workspaces and seamless authoring across studio, CLI, and SDK interfaces.
Create a new pipeline using custom components
Pipelines
Pipeline drafts
Pipeline jobs
Refresh
Delete
View options
Search
Filter
Columns
Name
Pipeline type
Updated on
Created by
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services

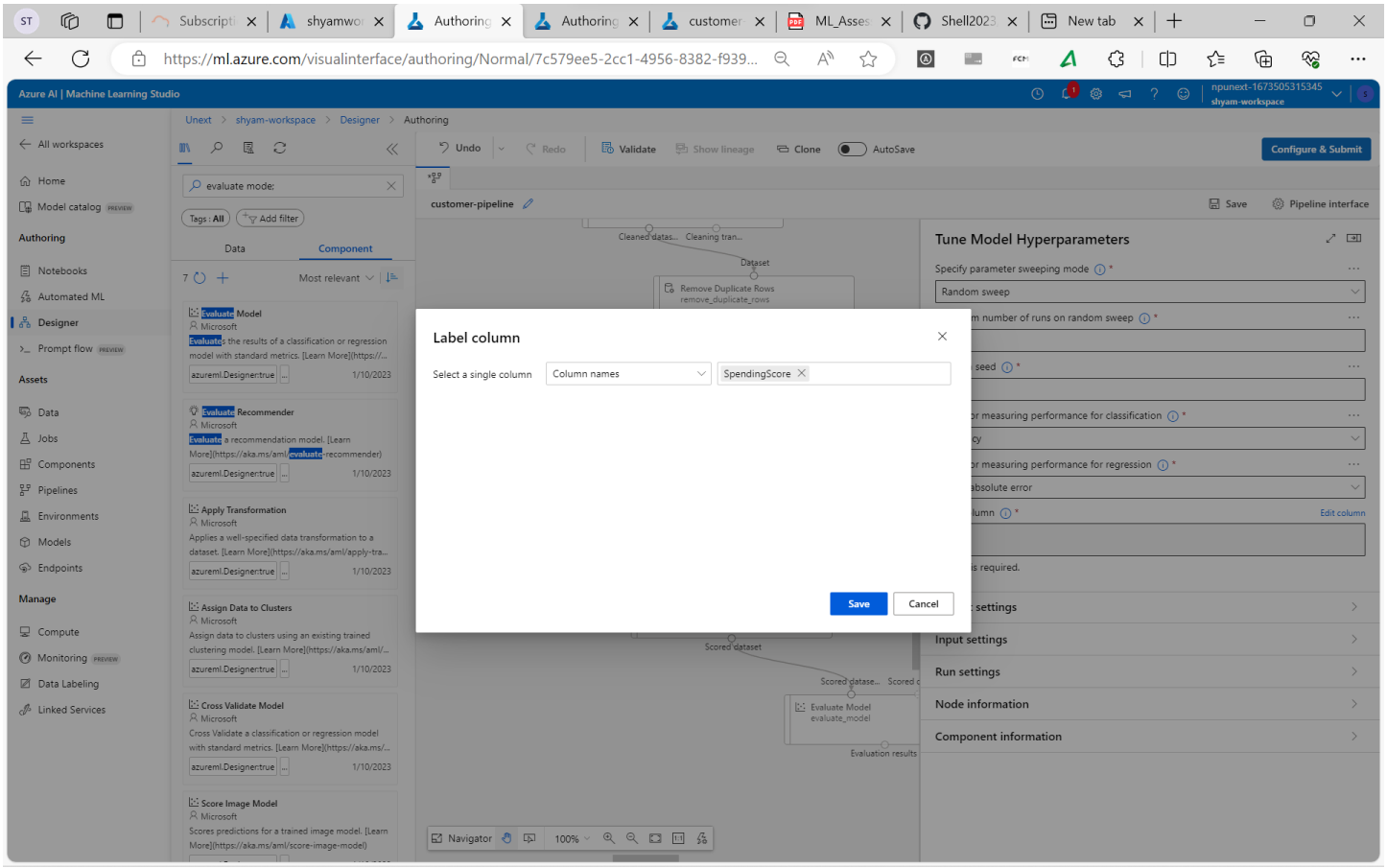

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
Search by name, tags and description
Registry name : All
Tags : All
Add filter
Data
Model
Component
customer-dataset
Version 1
10/4/2023
Pipeline-Created-on-10-04-2023
customer-dataset
Data output
Navigator
100%
Edit draft detail
Draft name *
customer-pipeline
Draft description (optional)
Pipeline created on 20231004
Created on
Oct 4, 2023 1:50 PM
Created by
Shellunext unextIDA63
Last edit time
Oct 4, 2023 1:50 PM
Last edited by
Shellunext unextIDA63
Save
Cancel

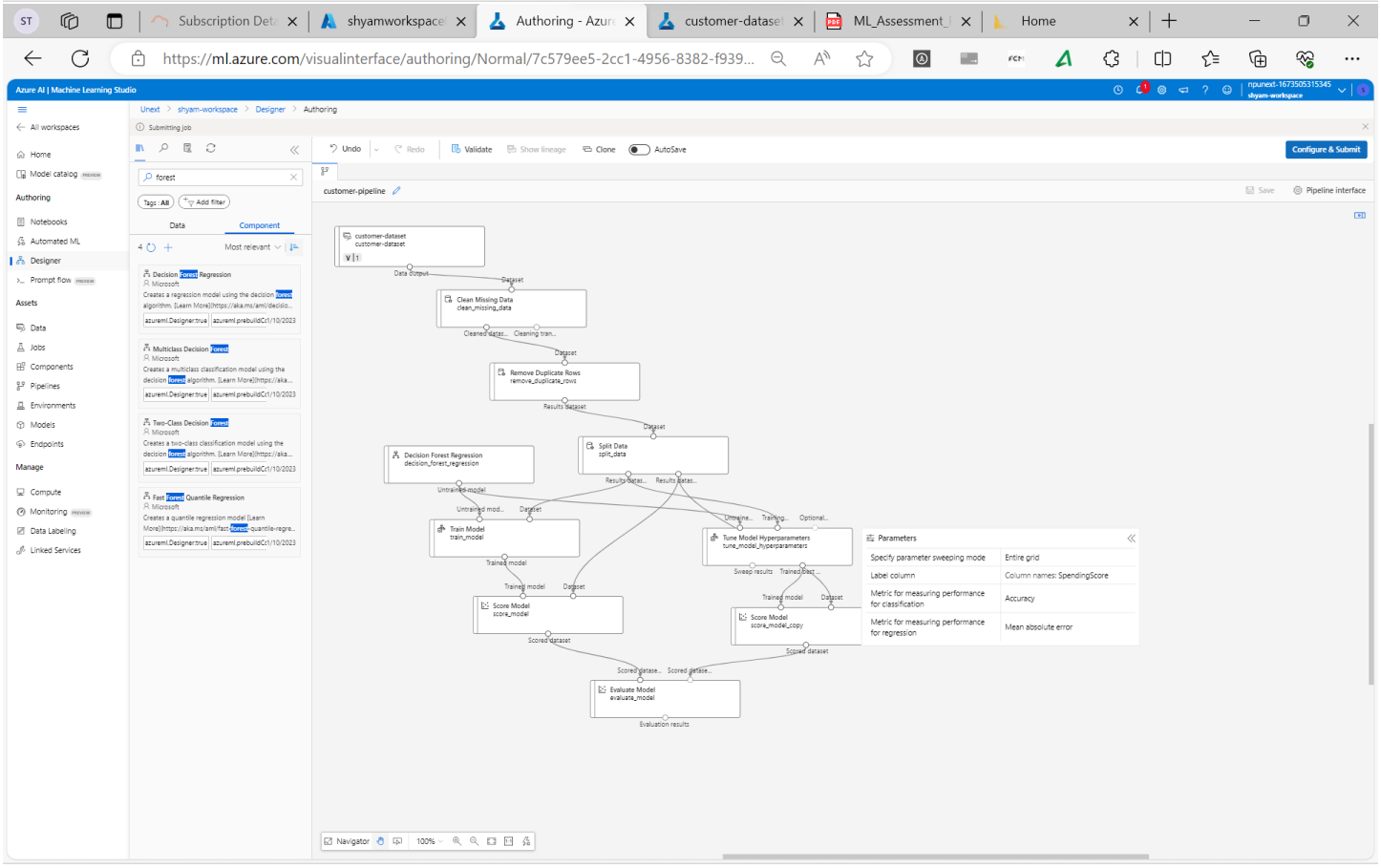

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
customer-pipeline
Save
Pipeline interface
customer-dataset
customer-dataset
Data output
Dataset
Clean Missing Data
clean_missing_data
Parameters
Cleaned datas...
Cleaning tran...
Clean Missing Data
Columns to be cleaned
Edit column
All columns
Minimum missing value ratio
0.0
Maximum missing value ratio
1.0
Cleaning mode
Custom substitution value
Replacement value
0
Generate missing value indicator column
False
Output settings
Input settings
Run settings
Node information
Component information
Navigator
100%

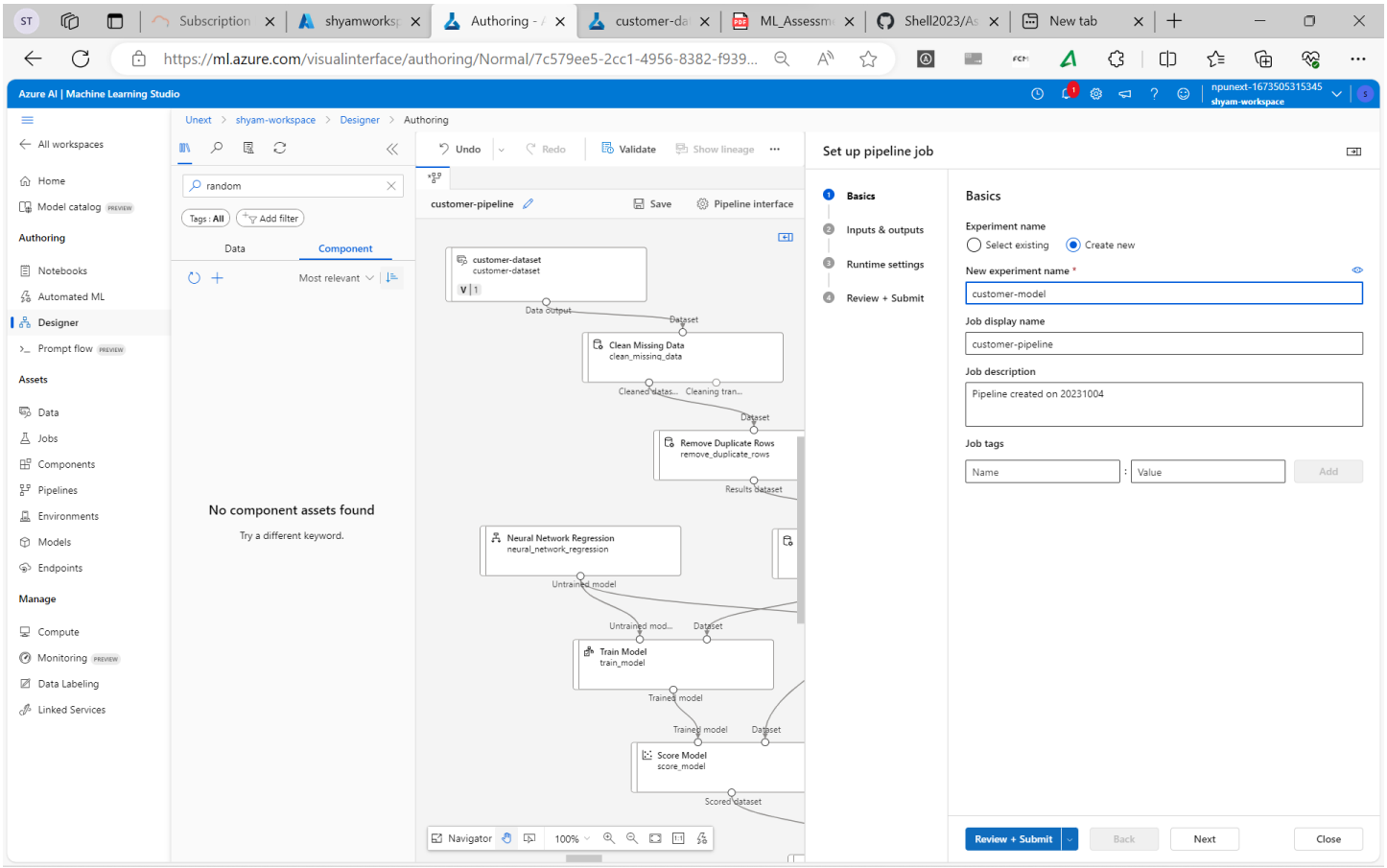

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
customer-pipeline
Save
Pipeline interface
customer-dataset
customer-dataset
Data output
Dataset
Clean Missing Data
clean_missing_data
Cleaned datas...
Cleaning tran...
Remove Duplicate Rows
remove_duplicate_rows
Results dataset
Parameters
Key column selection filter expression
All columns
Retain first duplicate row
True
Navigator
100%

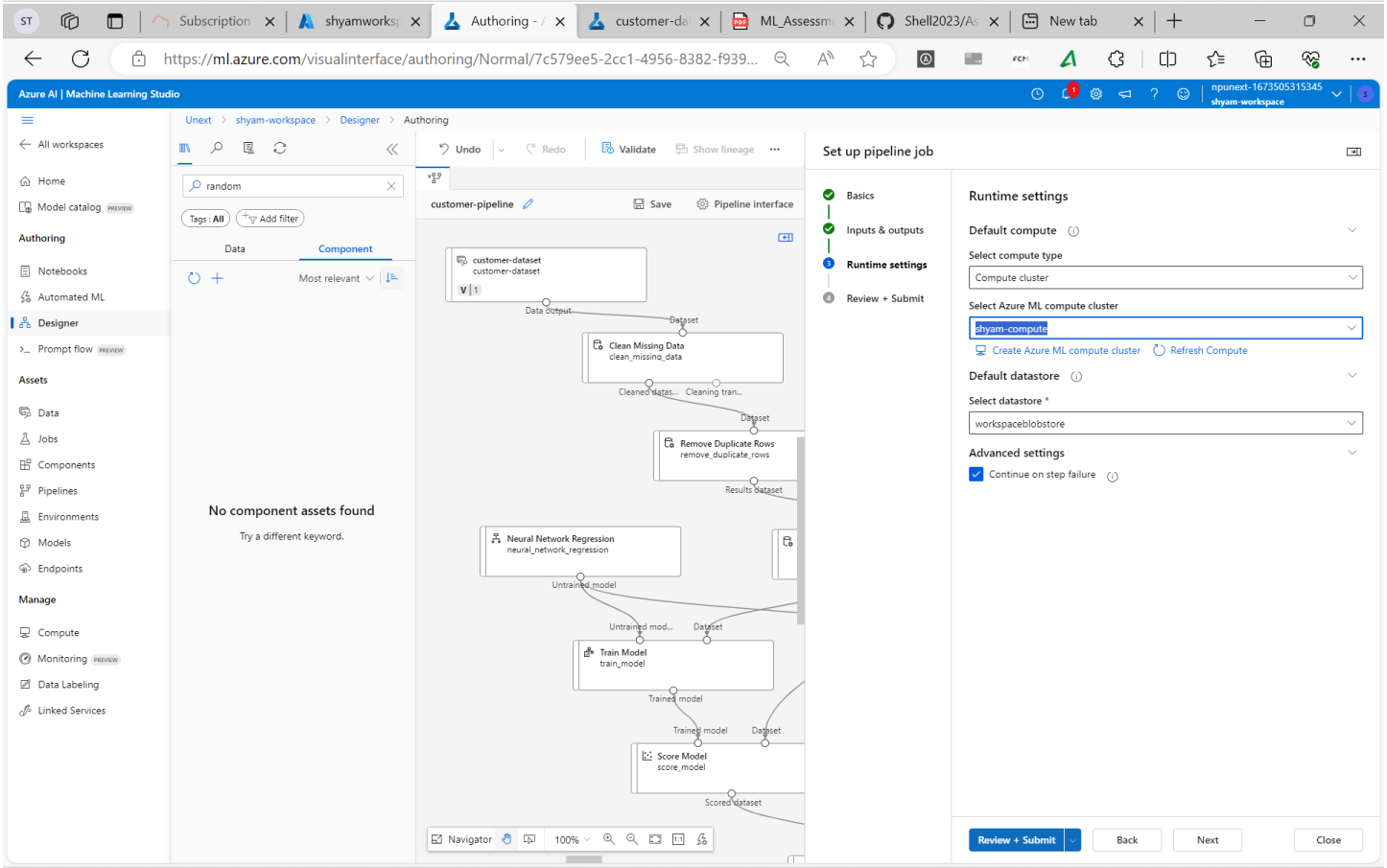

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
split data
Tags : All
Add filter
Data
Component
Most relevant
Split Data
Microsoft
Partitions the rows of a dataset into two distinct sets.
Apply SQL Transformation
Microsoft
Runs a SQLite query on input datasets to transform the data.
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
customer-pipeline
Save
Pipeline interface
customer-dataset
Data output
Clean Missing Data
clean_missing_data
Cleaned datas...
Cleaning tran...
Remove Duplicate Rows
remove_duplicate_rows
Results dataset
Split Data
split_data
Results datas...
Split Data
Splitting mode
Split Rows
Fraction of rows in the first output dataset
0.8
Randomized split
True
Random seed
0
Stratified split
False
Output settings
Input settings
Run settings
Node information
Component information
Navigator
100%

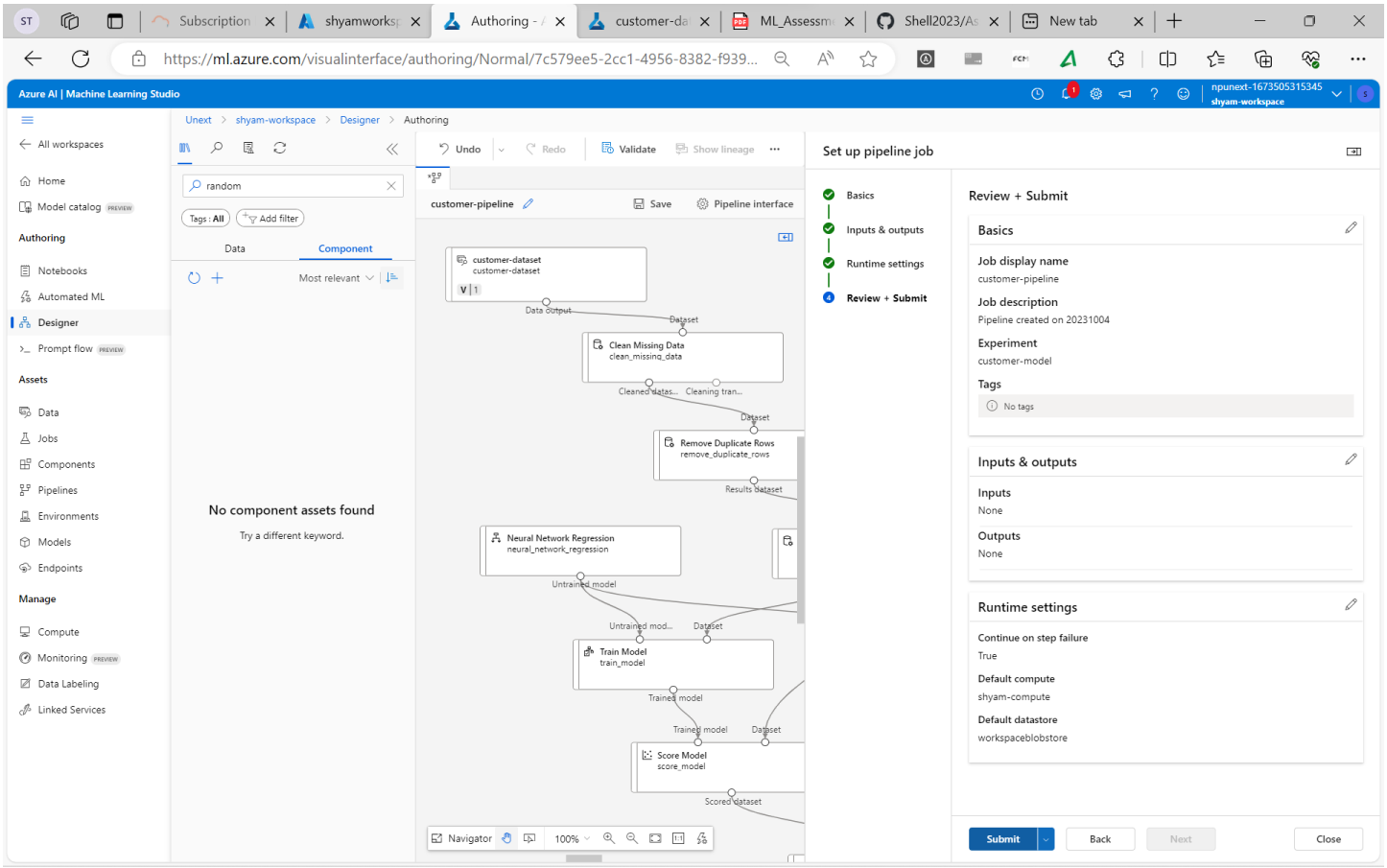

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
Search by name, tags and description
Tags : All
Add filter
Data
Component
Train Anomaly Detection Model
Microsoft
Trains an anomaly detector model and labels data from the training set
Statistical Functions (1)
Model Training (4)
Train Clustering Model
Microsoft
Train clustering model and assign data to clusters.
Train Model
Microsoft
Trains a classification or regression model in a supervised manner.
Train PyTorch Model
Microsoft
Train pytorch model from scratch or finetune it.
Tune Model Hyperparameters
Microsoft
Perform a parameter sweep on the model to determine the optimum parameter settings.
Web Service (2)
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
customer-pipeline
Save
Pipeline interface
customer-dataset
Data output
Clean Missing Data
clean_missing_data
Cleaned datas...
Cleaning tran...
Remove Duplicate Rows
remove_duplicate_rows
Results dataset
Neural Network Regression
neural_network_regression
Untrained model
Split Data
split_data
Results datas...
Train Model
train_model
Trained model
Parameters
Train Model
Label column
Edit column
Column names: SpendingScore
Model explanations
False
Output settings
Input settings
Run settings
Node information
Component information
Navigator
100%

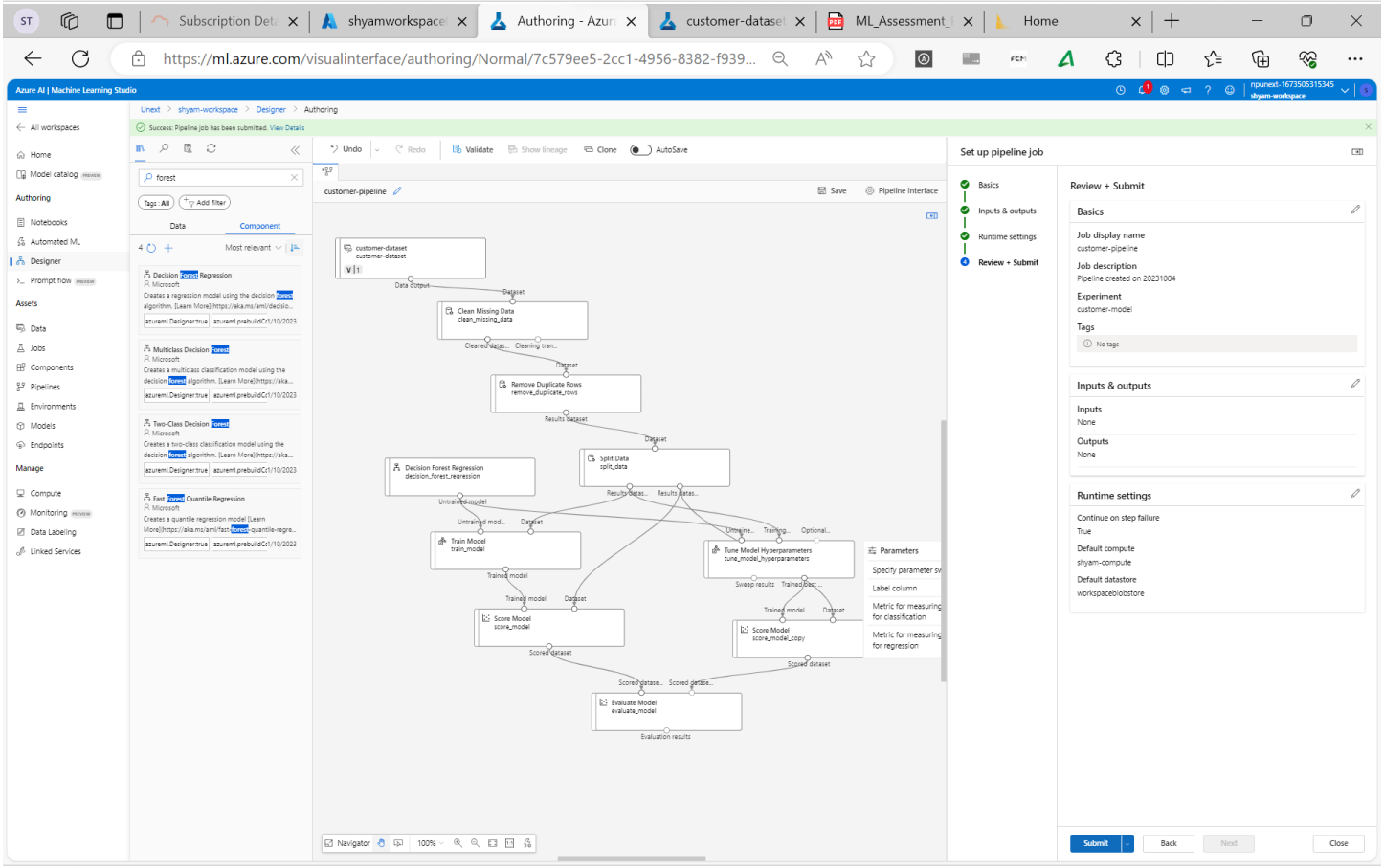

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
customer-pipeline
Save
Pipeline interface
Tune Model Hyperparameters
Specify parameter sweeping mode
Random sweep
Label column
Select a single column
Column names
SpendingScore
Save
Cancel
Input settings
Run settings
Node information
Component information
Evaluate Model
evaluate_model
Navigator
100%

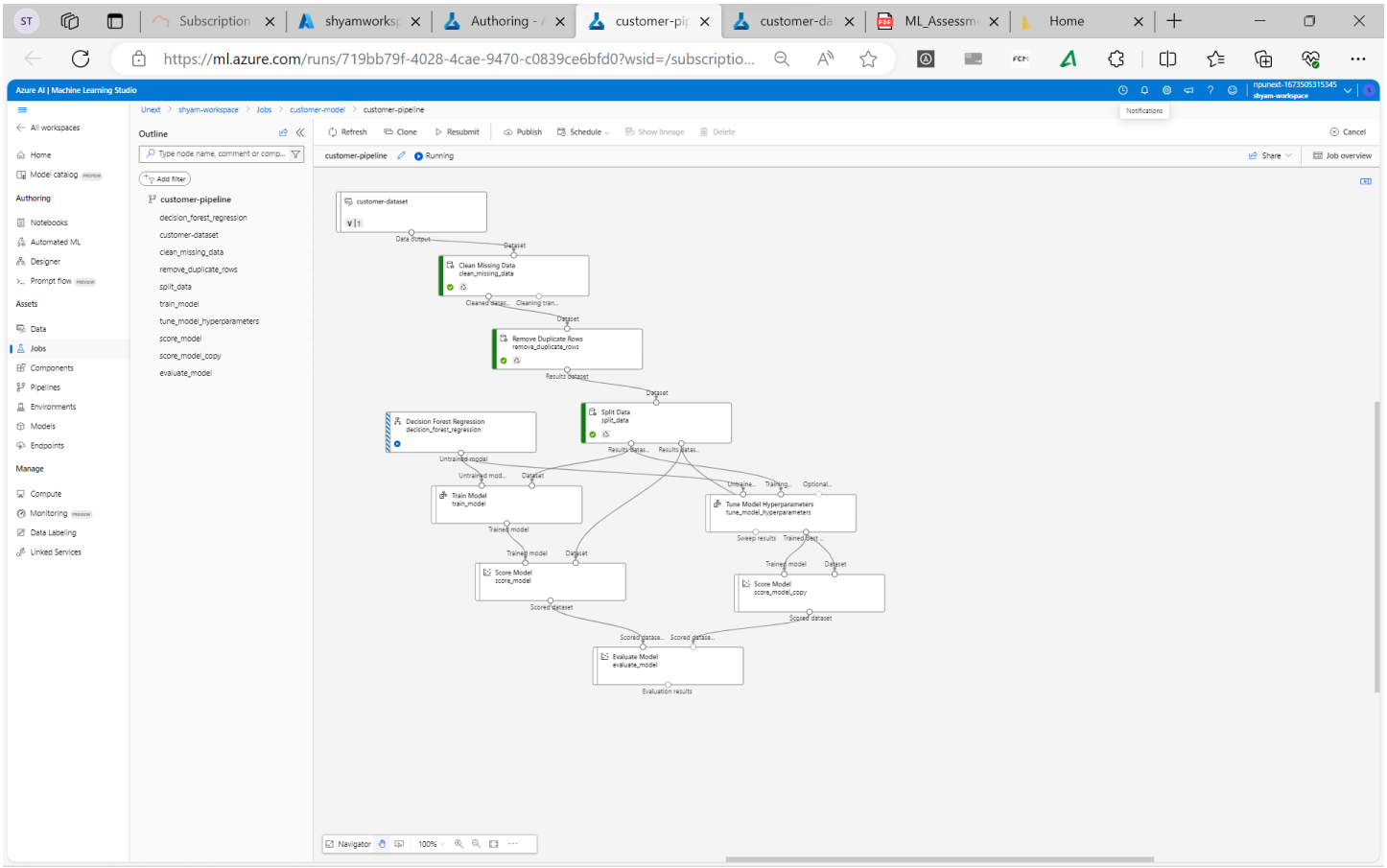

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
Undo
Redo
Validate
Show lineage
Clone
AutoSave
Configure & Submit
customer-pipeline
Save
Pipeline interface
customer-dataset
Clean Missing Data
clean_missing_data
Remove Duplicate Rows
remove_duplicate_rows
Split Data
split_data
Decision Forest Regression
decision_forest_regression
Train Model
train_model
Tune Model Hyperparameters
tune_model_hyperparameters
Score Model
score_model
Score Model
score_model_copy
Evaluate Model
evaluate_model
Parameters
Specify parameter sweeping mode
Entire grid
Label column
Column names: SpendingScore
Metric for measuring performance for classification
Accuracy
Metric for measuring performance for regression
Mean absolute error
Navigator
100%

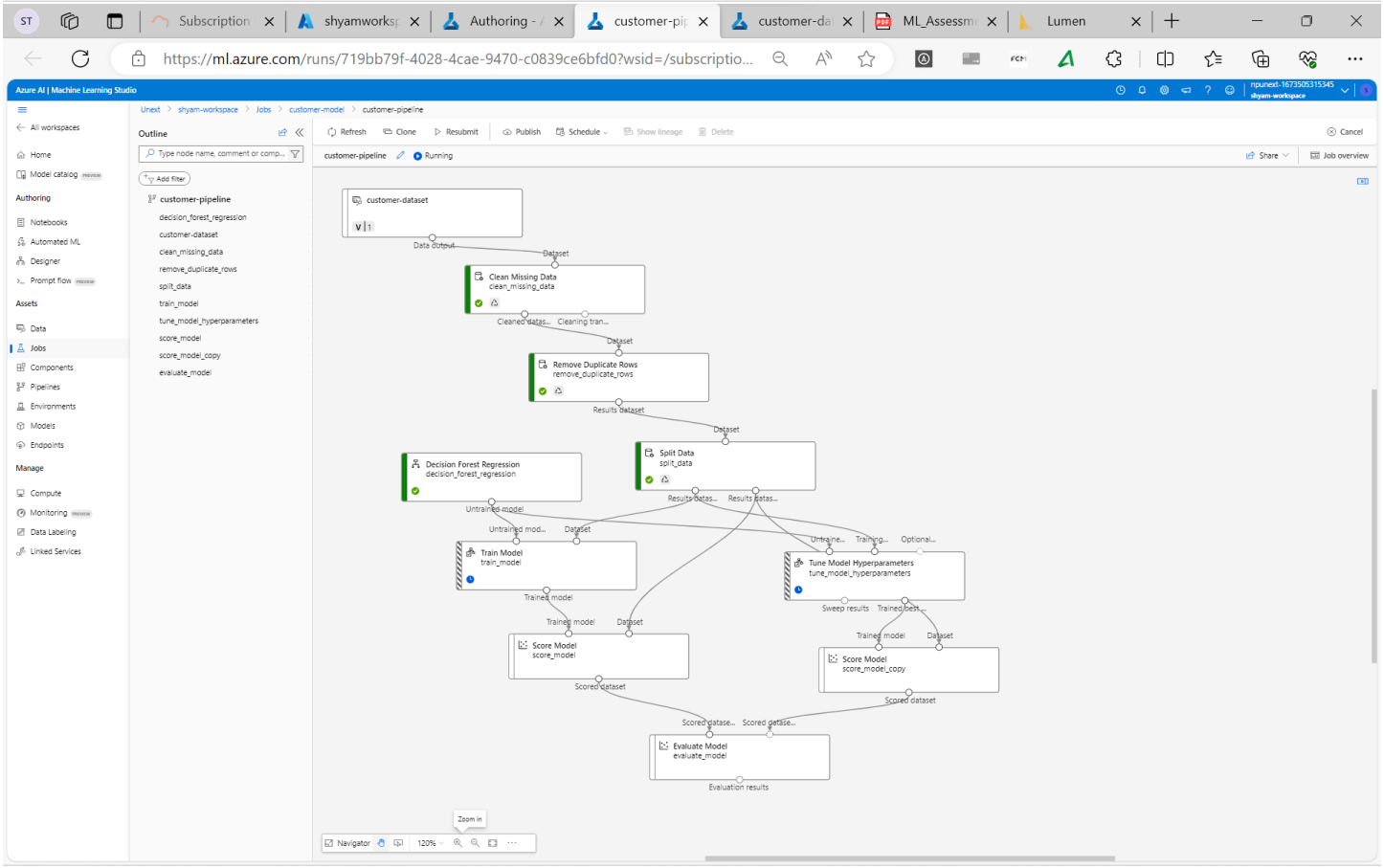

Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
random
Tags : All
Add filter
Data
Component
Most relevant
No component assets found
Try a different keyword.
Undo
Redo
Validate
Show lineage
customer-pipeline
Save
Pipeline interface
customer-dataset
Clean Missing Data
clean_missing_data
Remove Duplicate Rows
remove_duplicate_rows
Neural Network Regression
neural_network_regression
Train Model
train_model
Score Model
score_model
Navigator
100%
Set up pipeline job
Basics
Inputs & outputs
Runtime settings
Review + Submit
Experiment name
Select existing
Create new
New experiment name *
customer-model
Job display name
customer-pipeline
Job description
Pipeline created on 20231004
Job tags
Name
Value
Add
Review + Submit
Back
Next
Close


Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
No component assets found
Try a different keyword.
Set up pipeline job
Basics
Inputs & outputs
Runtime settings
Review + Submit
Runtime settings
Default compute
Select compute type
Compute cluster
Select Azure ML compute cluster
shyam-compute
Create Azure ML compute cluster
Refresh Compute
Default datastore
Select datastore *
workspaceblobstore
Advanced settings
Continue on step failure
Review + Submit
Back
Next
Close


Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Prompt flow
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring
Data Labeling
Linked Services
random
Tags : All
Add filter
Data
Component
Most relevant
No component assets found
Try a different keyword.
Undo
Redo
Validate
Show lineage
customer-pipeline
Save
Pipeline interface
customer-dataset
Clean Missing Data
clean_missing_data
Remove Duplicate Rows
remove_duplicate_rows
Neural Network Regression
neural_network_regression
Train Model
train_model
Score Model
score_model
Navigator
100%
Set up pipeline job
Basics
Inputs & outputs
Runtime settings
Review + Submit
Review + Submit
Basics
Job display name
customer-pipeline
Job description
Pipeline created on 20231004
Experiment
customer-model
Tags
No tags
Inputs & outputs
Inputs
None
Outputs
None
Runtime settings
Continue on step failure
True
Default compute
shyam-compute
Default datastore
workspaceblobstore
Submit
Back
Next
Close


Azure AI | Machine Learning Studio
Unext > shyam-workspace > Designer > Authoring
forest
Undo
Redo
Validate
Show lineage
Clone
AutoSave
customer-pipeline
Save
Pipeline interface
customer-dataset
Clean Missing Data
clean_missing_data
Remove Duplicate Rows
remove_duplicate_rows
Decision Forest Regression
decision_forest_regression
Split Data
split_data
Train Model
train_model
Tune Model Hyperparameters
tune_model_hyperparameters
Score Model
score_model
Evaluate Model
evaluate_model
Parameters
Label column
Set up pipeline job
Basics
Inputs & outputs
Runtime settings
Review + Submit
Review + Submit
Basics
Job display name
customer-pipeline
Job description
Pipeline created on 20231004
Experiment
customer-model
Tags
No tags
Inputs & outputs
Inputs
None
Outputs
None
Runtime settings
Continue on step failure
True
Default compute
shyam-compute
Default datastore
workspaceblobstore
Submit
Back
Next
Close


Azure AI | Machine Learning Studio
customer-pipeline
Running
customer-dataset
Clean Missing Data
clean_missing_data
Remove Duplicate Rows
remove_duplicate_rows
Split Data
split_data
Decision Forest Regression
decision_forest_regression
Train Model
train_model
Tune Model Hyperparameters
tune_model_hyperparameters
Score Model
score_model
Score Model
score_model_copy
Evaluate Model
evaluate_model


Azure AI | Machine Learning Studio
customer-pipeline
Running
customer-dataset
Clean Missing Data
clean_missing_data
Remove Duplicate Rows
remove_duplicate_rows
Split Data
split_data
Decision Forest Regression
decision_forest_regression
Train Model
train_model
Tune Model Hyperparameters
tune_model_hyperparameters
Score Model
score_model
Score Model
score_model_copy
Evaluate Model
evaluate_model

Assessment

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
   A. Removing rows which were empty as it would affect the model performance. Giving value 0 as age or mean would affect the performance.
   B. Removing duplicate rows

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?
   A. Assess model performance on unseen data.
   B. Avoids overfitting.
   C. Optimize model settings based on testing set results.
   D. Ensure model can handle real-world scenarios.

3. Describe a machine learning algorithm suitable for predicting customer purchasing behaviour in the given scenario. Explain why you chose this algorithm.
   1. Ensemble Learning: Random Forest combines multiple decision trees for robust predictions.
   2. High Accuracy: It offers reliable classification results, critical for predicting purchases accurately.
   3. Feature Importance: Provides insights into which customer attributes influence buying decisions.
   4. Versatile with Data: Handles both numerical and categorical customer data effectively.
   5. Reduced Overfitting: Aggregation of predictions from multiple trees reduces the risk of overfitting, ensuring generalization.

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

   Hyperparameter tuning involves optimizing the settings of a machine learning model's hyperparameters to improve its performance.
   Importance: Enhanced Model Performance: Hyperparameter tuning can significantly boost a model's accuracy and generalization by finding the best hyperparameter values.

   Technique - Grid Search:
   In grid search, you specify a range of hyperparameter values, and the algorithm systematically explores all possible combinations to find the best set.

   Benefits:
   - Exhaustively searches the hyperparameter space, ensuring no configuration is missed.
   - straightforward to implement
   - provides reproducible results, making it easy to document and share the optimal hyperparameters.