

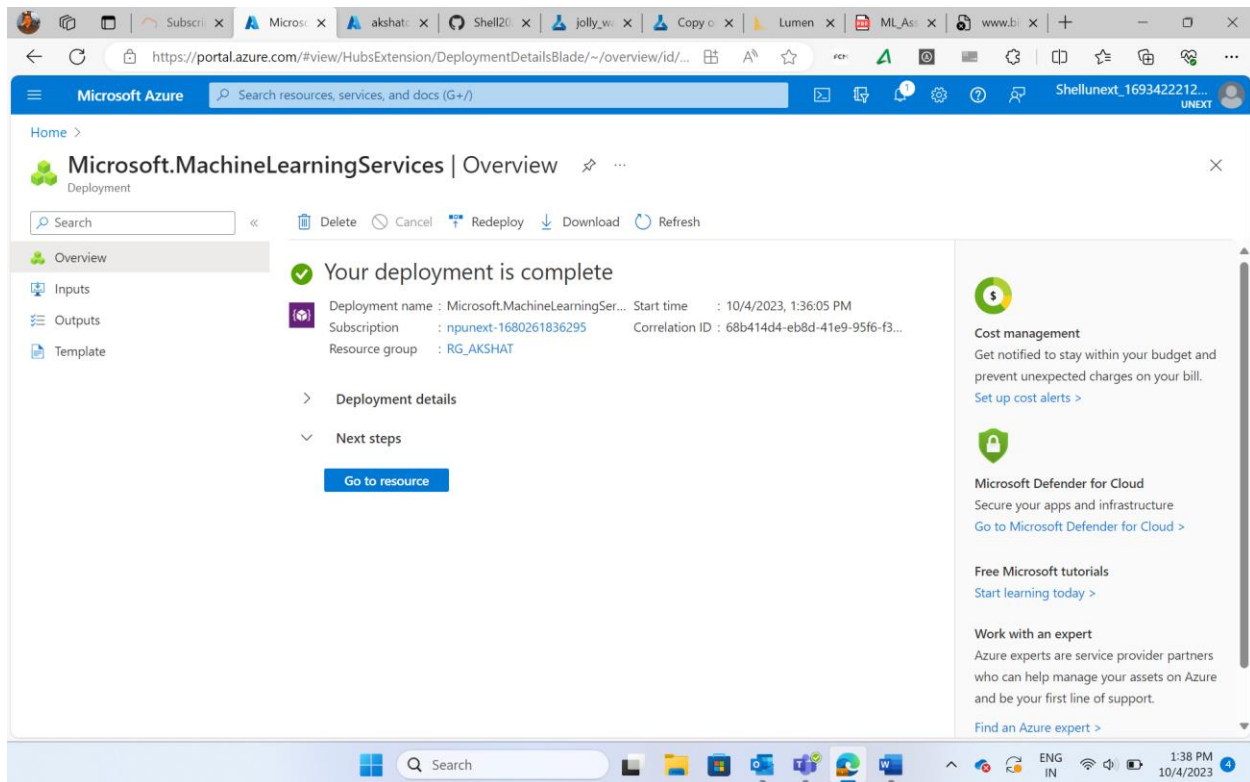
Uploading file to the blob storage account

The screenshot shows the Microsoft Azure portal interface. On the left, the navigation pane displays the 'akshatcontainer' container. The main area shows the 'Upload blob' dialog box. The dialog box has a title bar 'Upload blob' and a close button. Inside, it shows a cloud icon and the text '1 file(s) selected: customer_data.csv'. Below this, there is a checkbox for 'Overwrite if files already exist' and an 'Advanced' section. At the bottom, there is an 'Upload' button and a 'Give feedback' link. The background shows the 'akshatcontainer' container page with a search bar and various settings.

The screenshot shows the Microsoft Azure portal interface after the successful upload. A notification box at the top right states 'Successfully uploaded blob(s)' and 'Successfully uploaded 1 blob(s)'. The main area shows the 'akshatcontainer' container page. The 'Upload' button is now disabled. The 'Search blobs by prefix (case-sensitive)' field is empty. The 'Show deleted blobs' toggle is turned off. Below the search bar, there is a table with the following columns: Name, Modified, Access tier, Archive status, Blob type, and Size. The table contains one row: 'customer_data.csv', '10/4/2023, 1:38:18 PM', 'Hot (Inferred)', 'Block blob', and '4.0K'. The background shows the 'akshatcontainer' container page with a search bar and various settings.

Name	Modified	Access tier	Archive status	Blob type	Size
customer_data.csv	10/4/2023, 1:38:18 PM	Hot (Inferred)		Block blob	4.0K

Machine learning service is created



Microsoft Azure

Home > Microsoft.MachineLearningServices | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name : Microsoft.MachineLearningSer... Start time : 10/4/2023, 1:36:05 PM

Subscription : npunext-1680261836295 Correlation ID : 68b414d4-eb8d-41e9-95f6-f3...

Resource group : RG_AKSHAT

> Deployment details

< Next steps

Go to resource

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

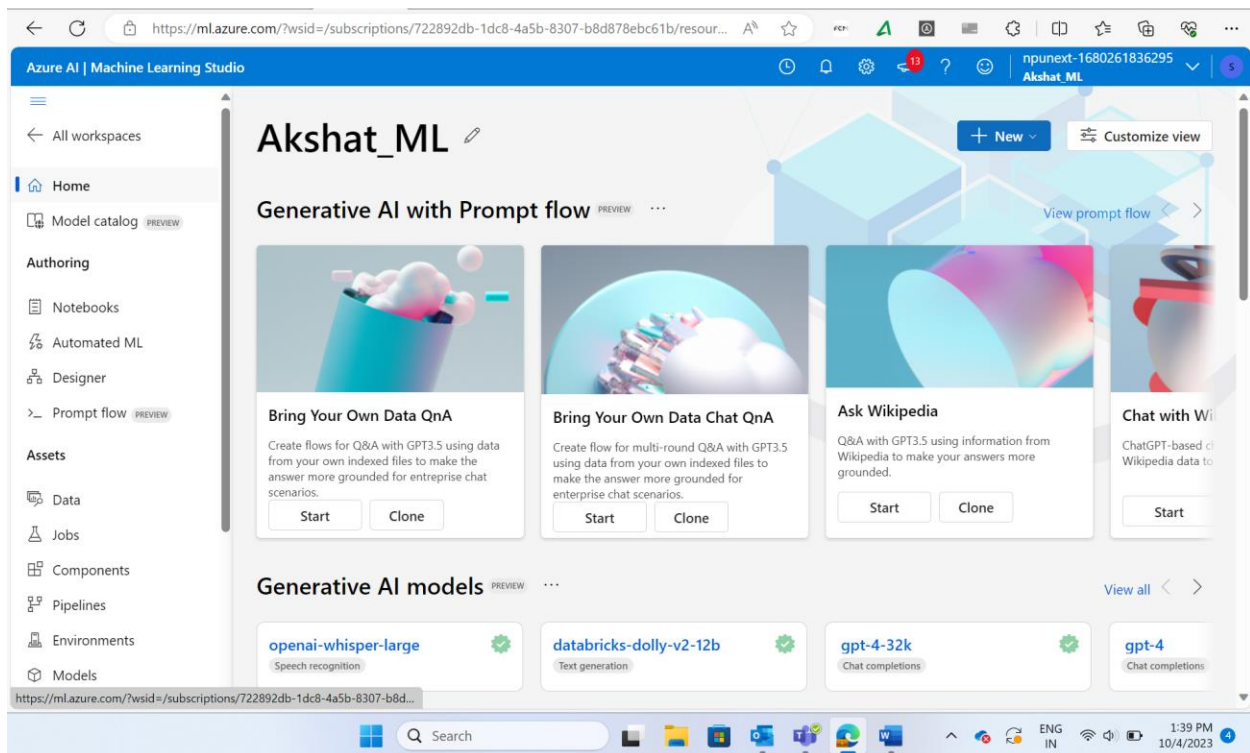
Start learning today >

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 >

ML Studio is launched



Azure AI | Machine Learning Studio

npunext-1680261836295 Akshat ML

All workspaces

Home

Model catalog

Authoring

Notebooks

Automated ML

Designer

Prompt flow

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Akshat_ML

+ New

Customize view

Generative AI with Prompt flow

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 enterprise 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 using Wikipedia data to answer questions.

Start

Generative AI models

View all

openai-whisper-large

Speech recognition

databricks-dolly-v2-12b

Text generation

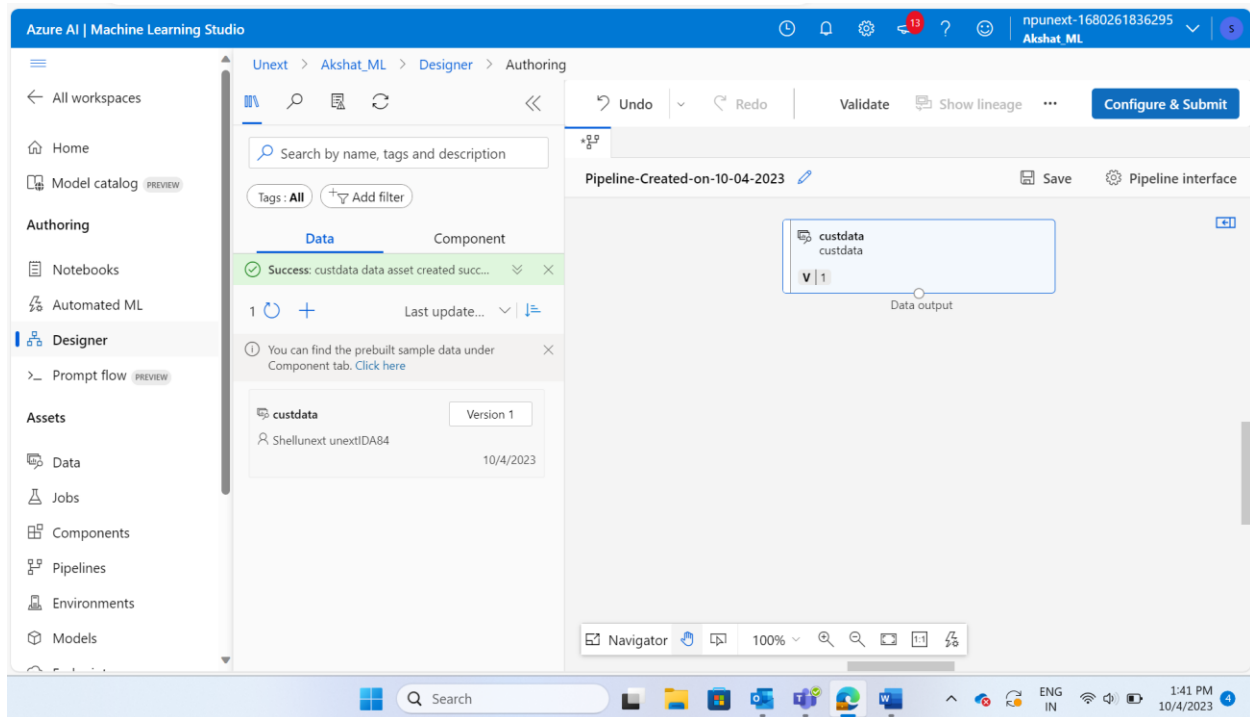
gpt-4-32k

Chat completions

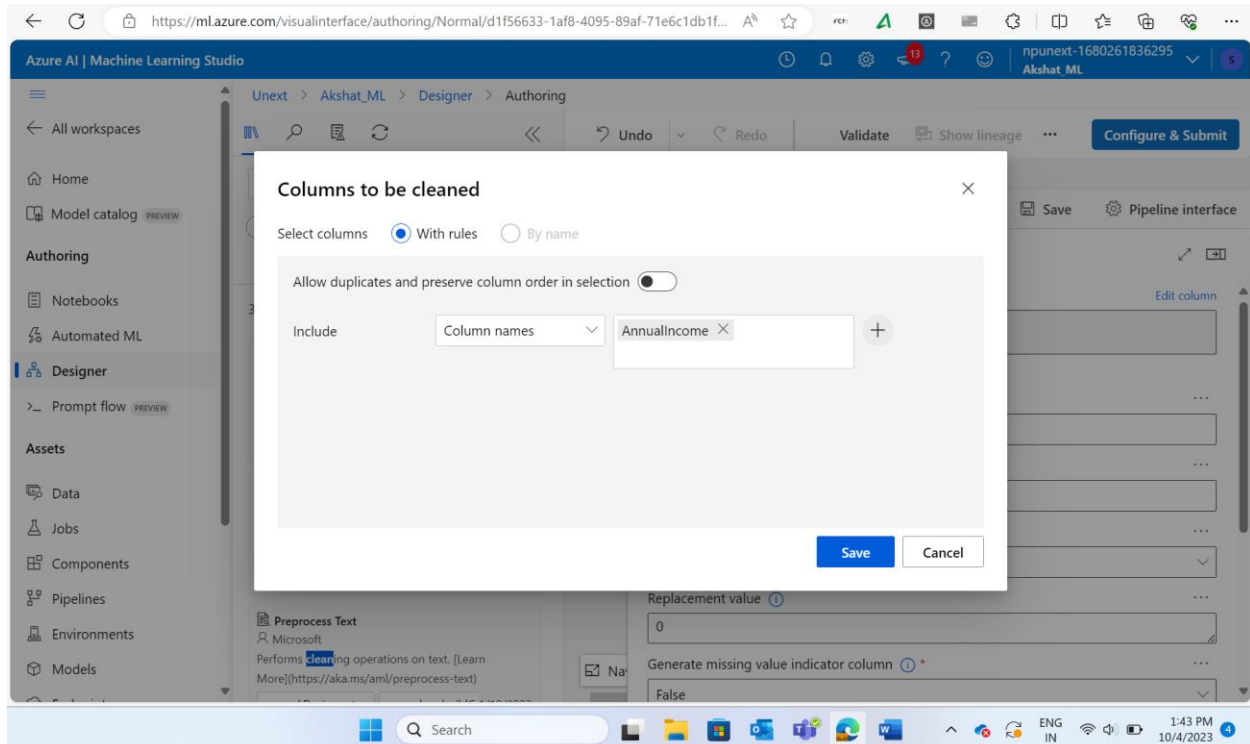
gpt-4

Chat completions

Dataset is loaded in ML Studio



Cleaning activity – cleaning the annual income and spending score column since these have some empty rows like missing values so replacing the missing values with 0



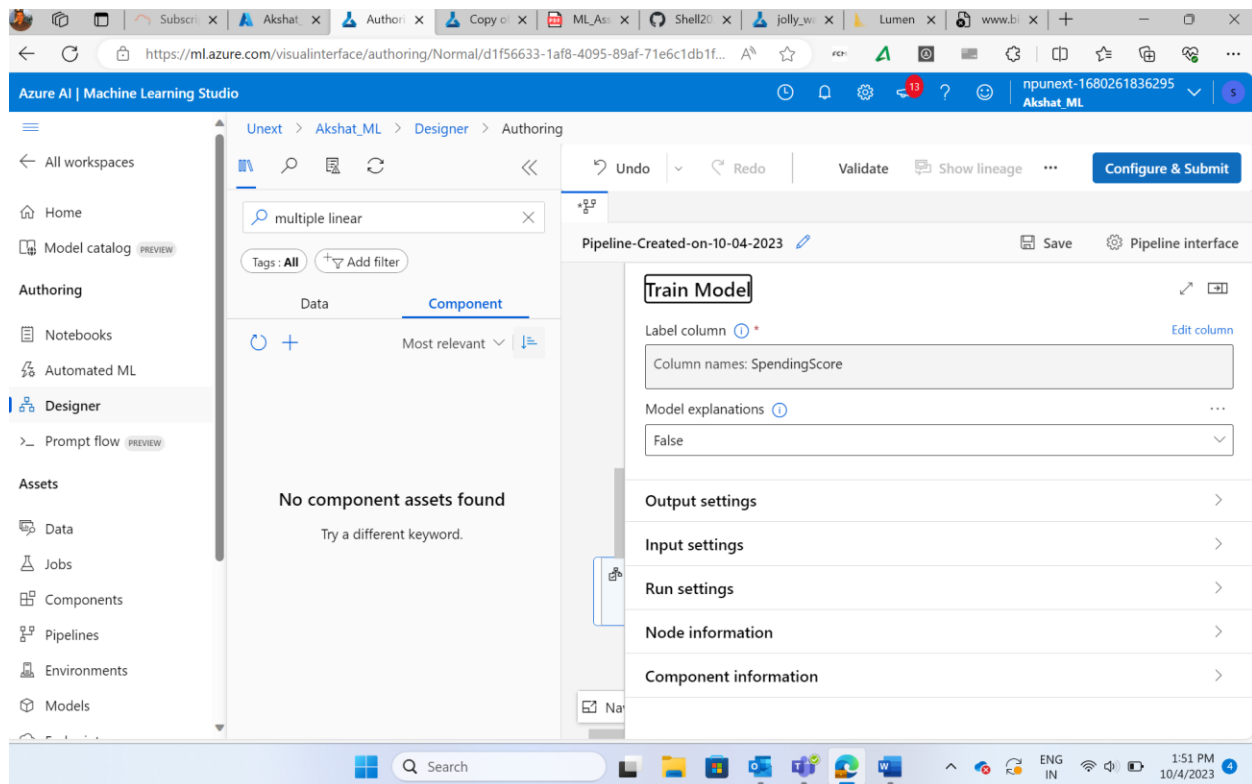
Splitting data into 70-30 ratio

The screenshot displays the Azure Machine Learning Studio interface. The left sidebar shows the navigation pane with 'Designer' selected. The main workspace is titled 'train model' and shows a search for 'train model' with results for 'Train Model', 'Train Clustering Model', and 'Train PyTorch Model'. The 'Split Data' component is selected, and its configuration panel is open on the right. The configuration panel shows the following settings:

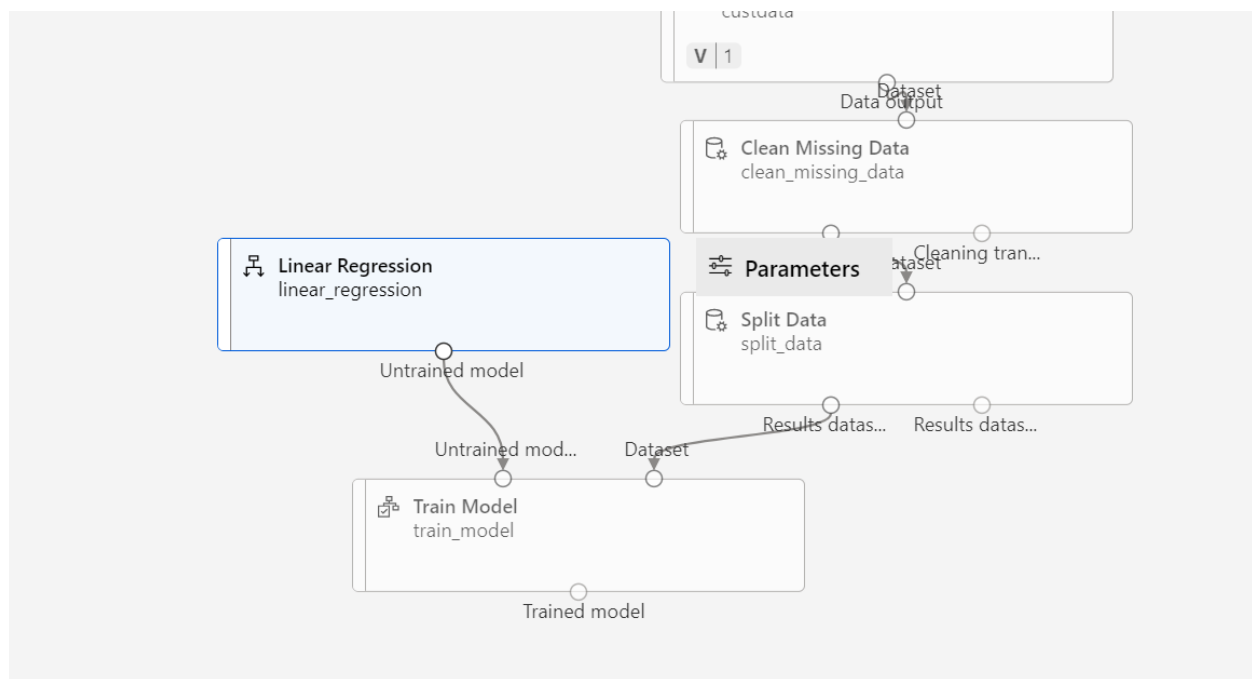
- Splitting mode: Split Rows
- Fraction of rows in the first output dataset: 0.7
- Randomized split: True
- Random seed: 0
- Stratified split: False

The 'Output settings' and 'Input settings' sections are also visible but not expanded.

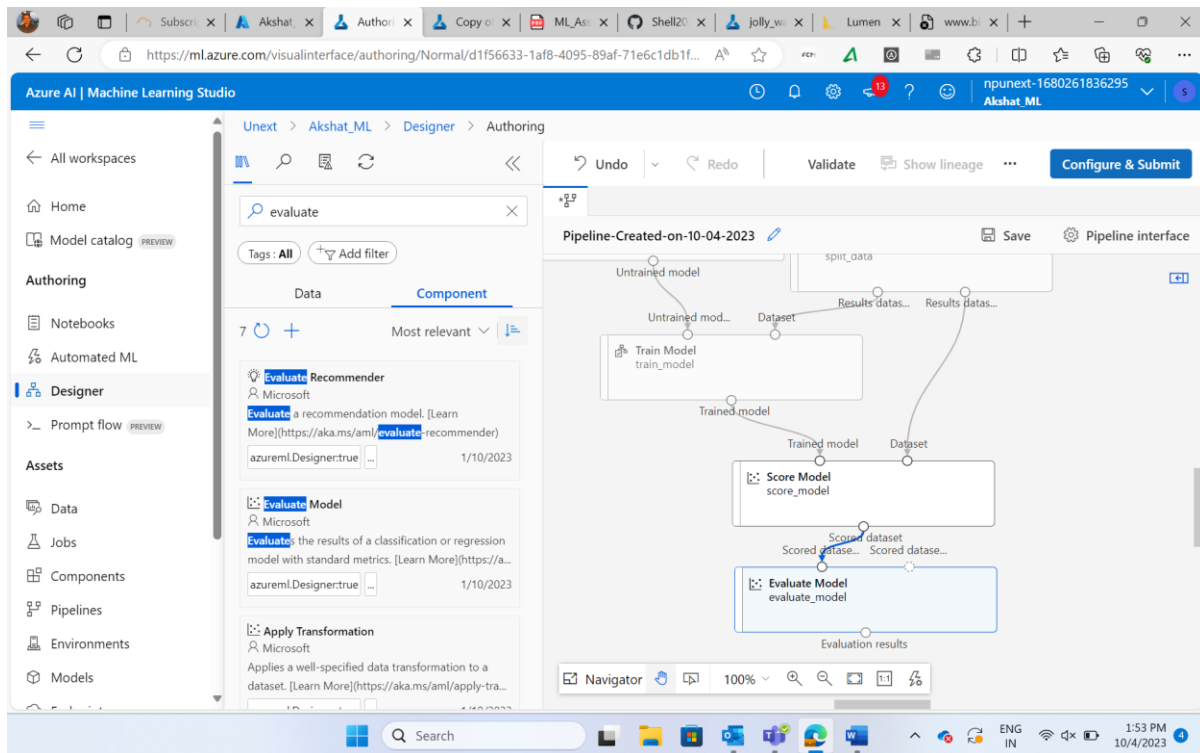
Setting column of train model as SpendingScore



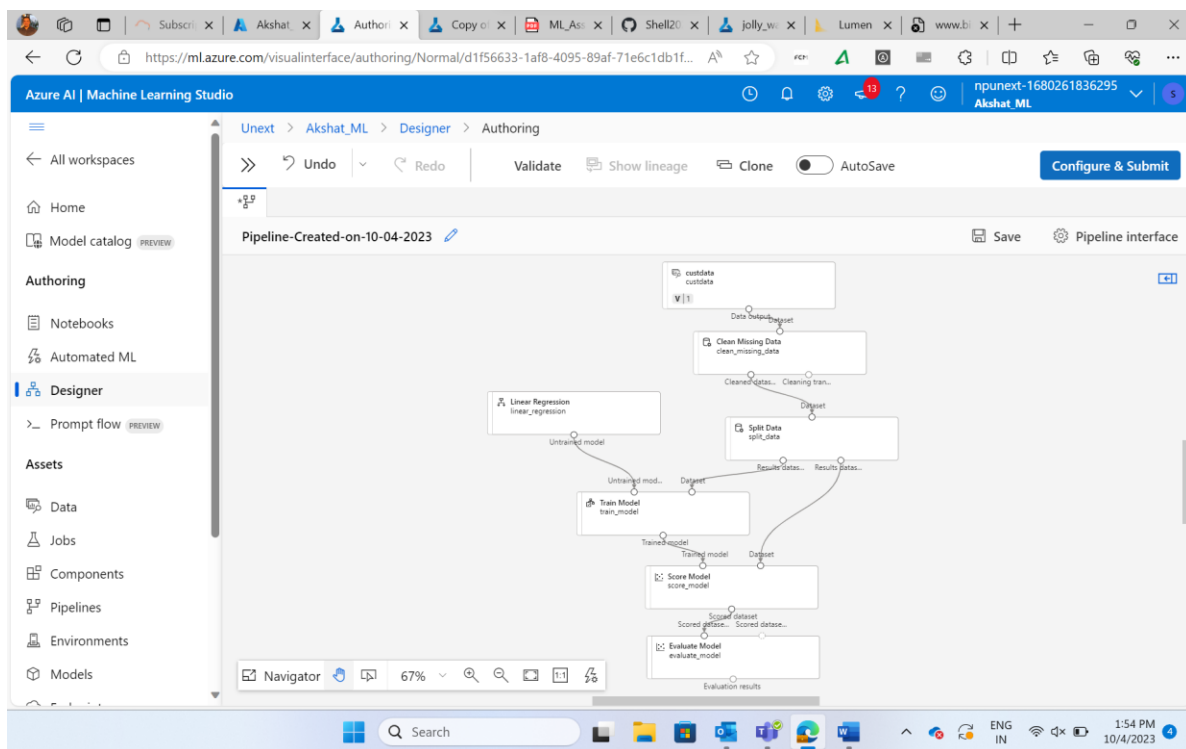
Using linear regression for the untrained model



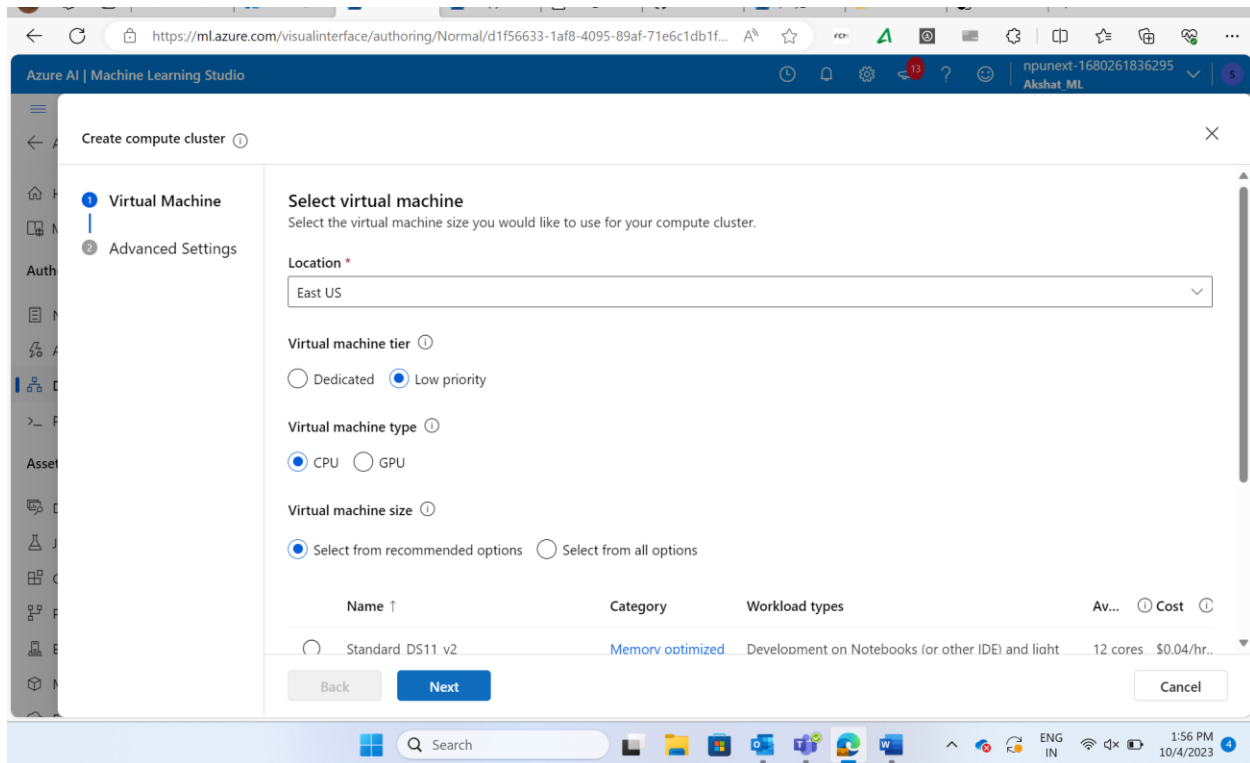
Adding ScoreModel and EvaluateModel components to tell us the results



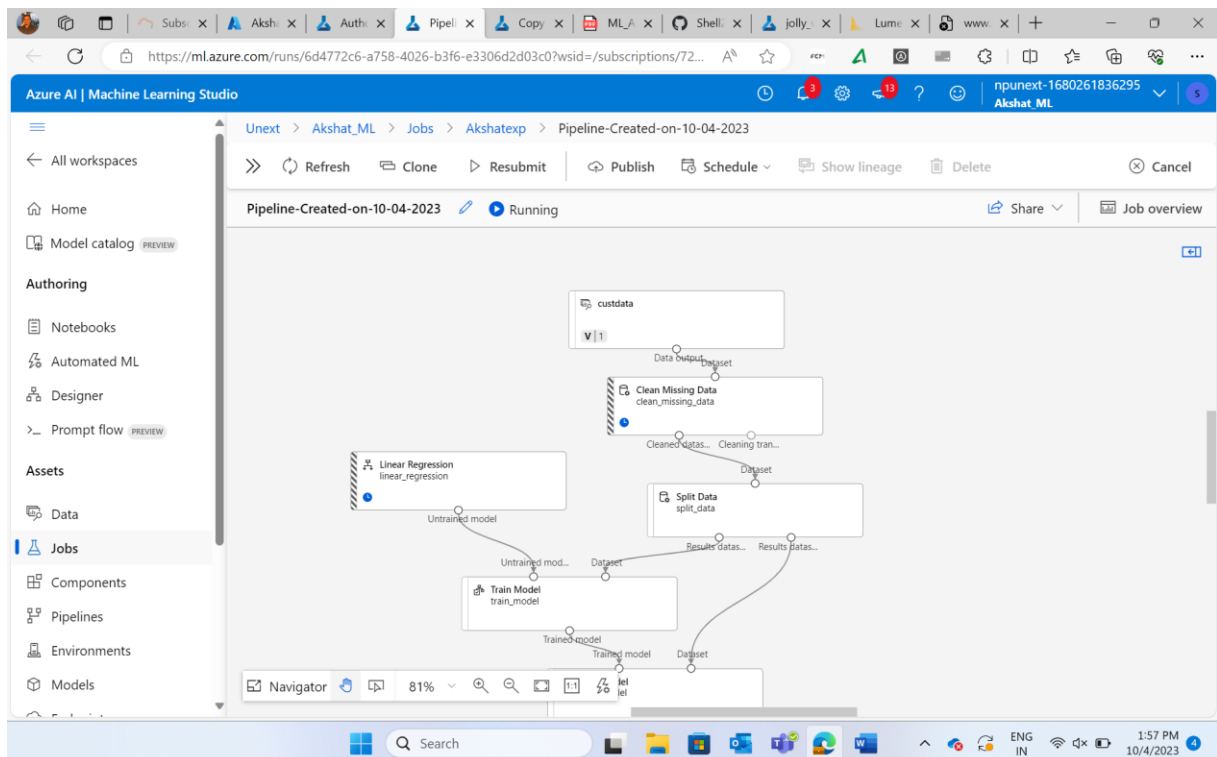
Final Pipeline looks like this



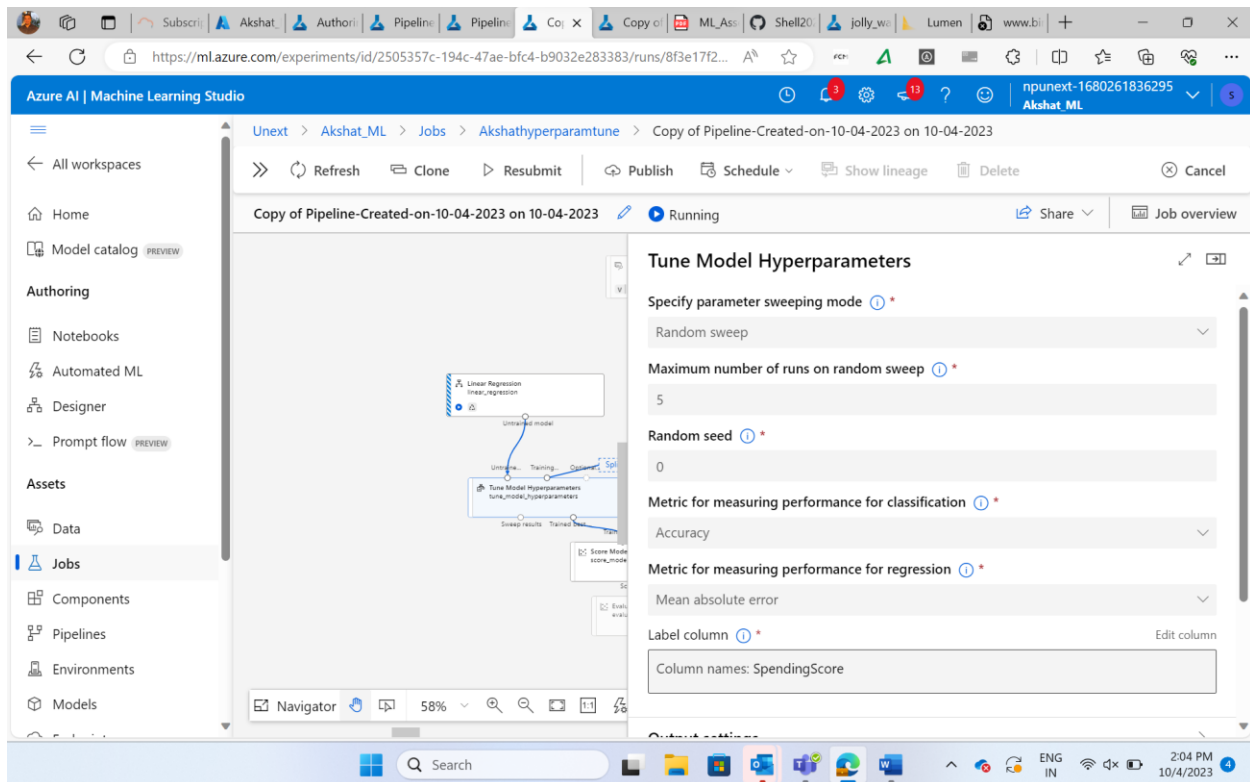
Creating compute cluster



Pipeline is running



Hyperparameter tuning –



Assessment Questions:

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.

Accessing the dataset using the blob storage account or web links

Importing the dataset into the pipeline

Removing the null and error values

Splitting the dataset into train and test data set with a specific ratio, usually 70-30.

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?

It is important to split the dataset into training and testing sets as a machine learning model is built after developing the model and training it on some original values so that it finds patterns in those values and help us in the required predictions. Hence the splitting of original dataset into train and test data set helps the train dataset to train the model and the test data set to check up with the original values of the test data.

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

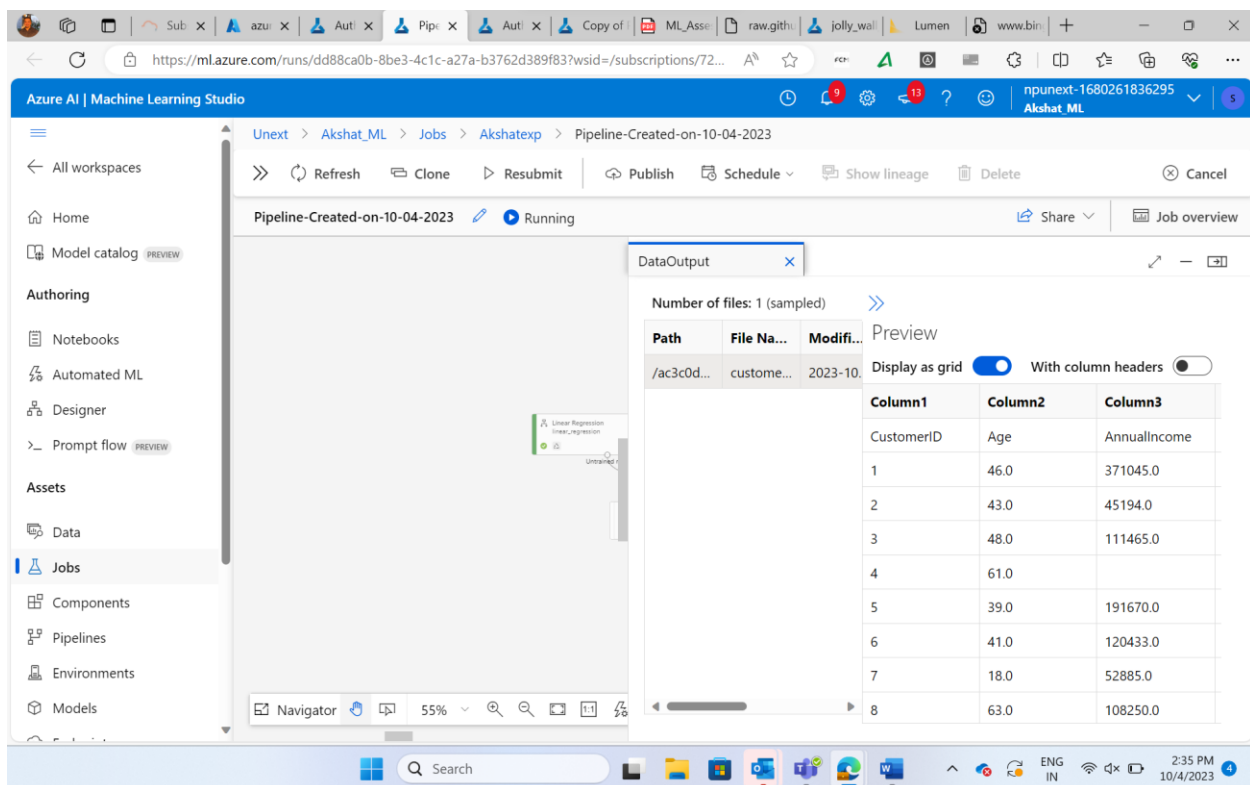
Linear regression is used as it is a supervised or a reinforcement learning which allows for linear regression to predict customer purchasing behavior.

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 uses hyperparameters that cannot be directly learned from the regular training. These are fixed before the actual training process begins. These express important properties of a model such as its complexity or how fast it should learn.

Grid Search is a technique in which model is evaluated for a range of hyperparameter values. It searches for the best set of hyperparameters from a grid of hyperparameter values which is its important benefit.

Data preview -



The screenshot displays the Azure AI Machine Learning Studio interface. The main workspace shows a pipeline named "Pipeline-Created-on-10-04-2023" in a "Running" state. A "DataOutput" window is open, displaying a preview of the data. The preview shows a table with 8 rows and 3 columns: CustomerID, Age, and AnnualIncome. The data is as follows:

CustomerID	Age	AnnualIncome
1	46.0	371045.0
2	43.0	45194.0
3	48.0	111465.0
4	61.0	
5	39.0	191670.0
6	41.0	120433.0
7	18.0	52885.0
8	63.0	108250.0

Pipeline runs –

Azure ML Studio interface showing a pipeline run in progress. The pipeline is named "Pipeline-Created-on-10-04-2023" and is currently "Running". The pipeline steps include "Linear Regression linear_regression", "Untrained model", "Train Model train_model", "Score Model score_model", and "Evaluate Model evaluate_model". The "Scored_dataset" output is displayed, showing a table with 60 rows and 5 columns: CustomerID, Age, AnnualIncome, Spending, and an unlabeled column. The table contains data for various customers, including their age, annual income, and spending. A histogram is also visible on the right side of the table.

CustomerID	Age	AnnualIncome	Spending	
29	NaN	84987	29	
89	29	146982	24	
134	NaN	43627	70	
78	NaN	0	23	
2	43	45194	24	
166	76	140427	66	
152	38	564965	42	
104	NaN	113795	81	
22	69	122765	64	
189	58	102036	34	
115	55	152883	127	

Azure ML Studio interface showing the same pipeline run completed. The pipeline is now "Completed". The pipeline steps are the same as in the previous screenshot. The "Evaluation_results" output is displayed, showing a table with 1 row and 5 columns: Mean_Absolute_Error, Root_Mean_Squared_Error, and three unlabeled columns. The table contains the evaluation metrics for the model. A histogram is also visible on the right side of the table.

Mean_Absolute_Error	Root_Mean_Squared_Error			
27.969071	32.432294			1

