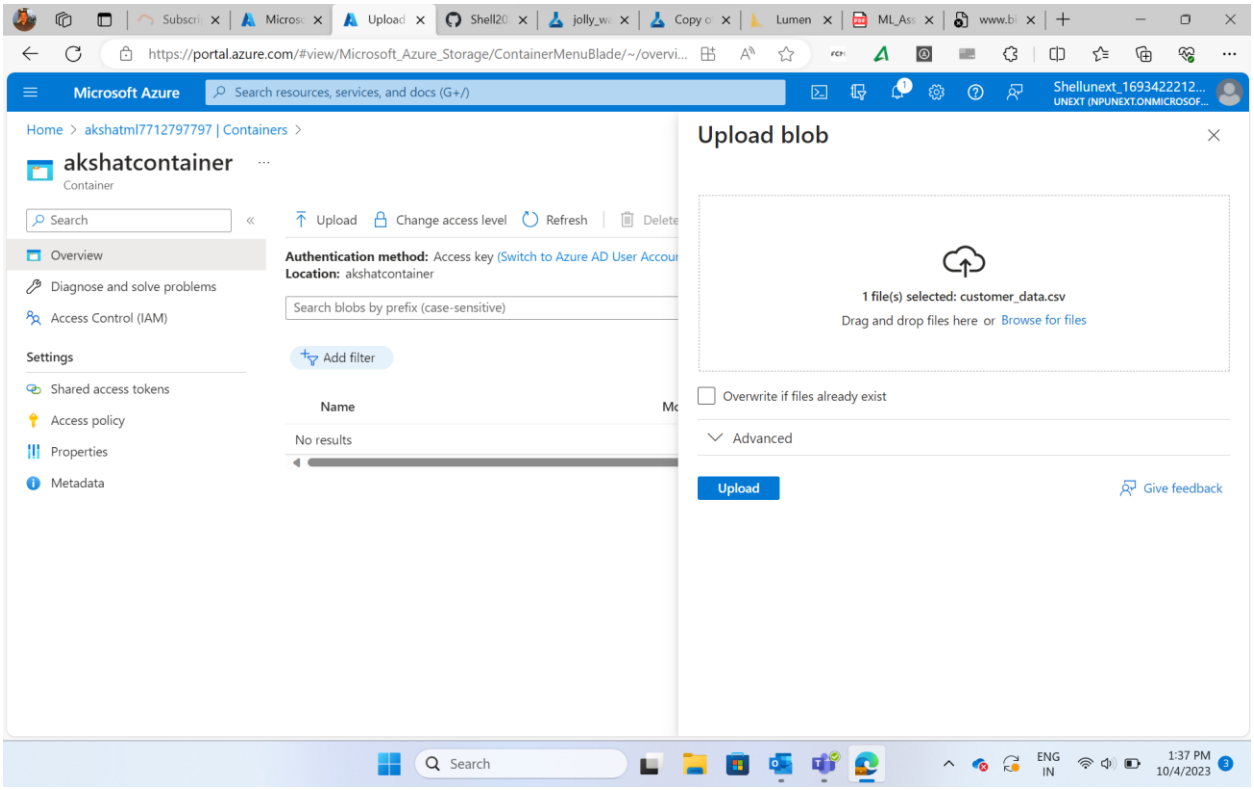
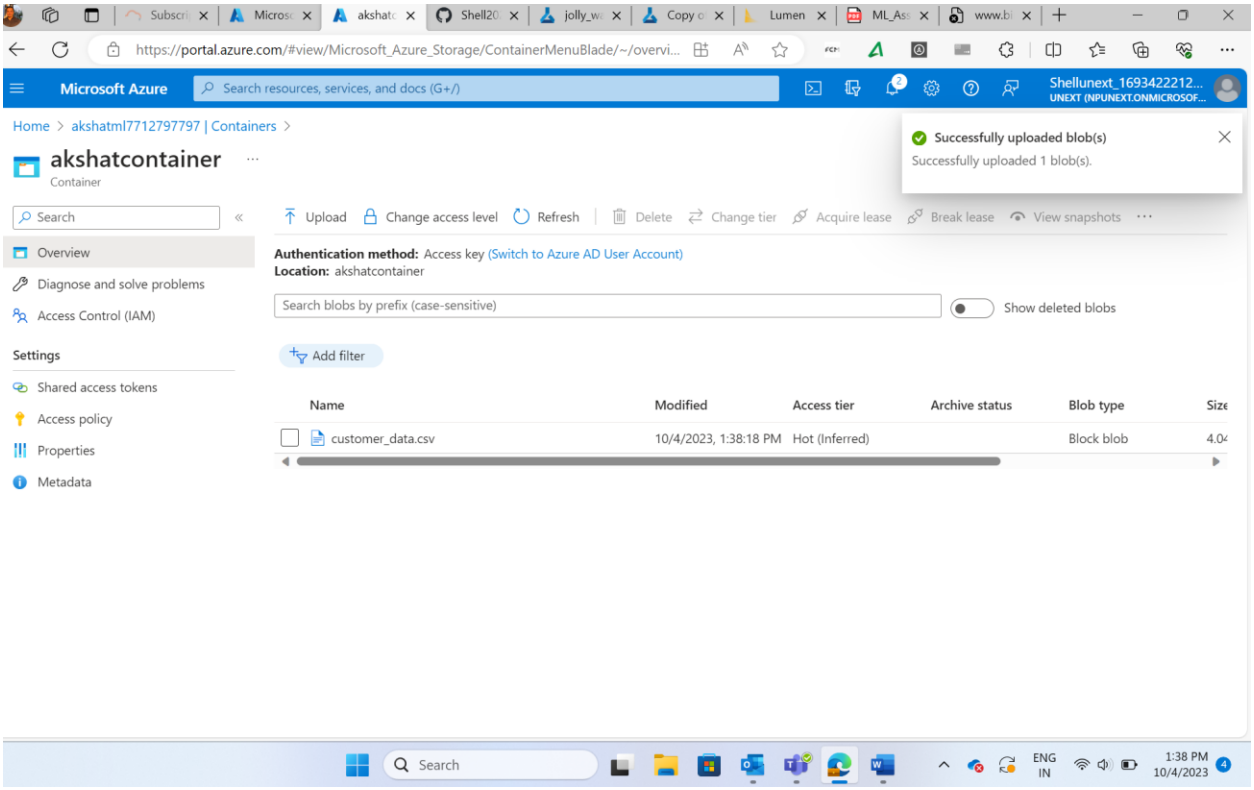
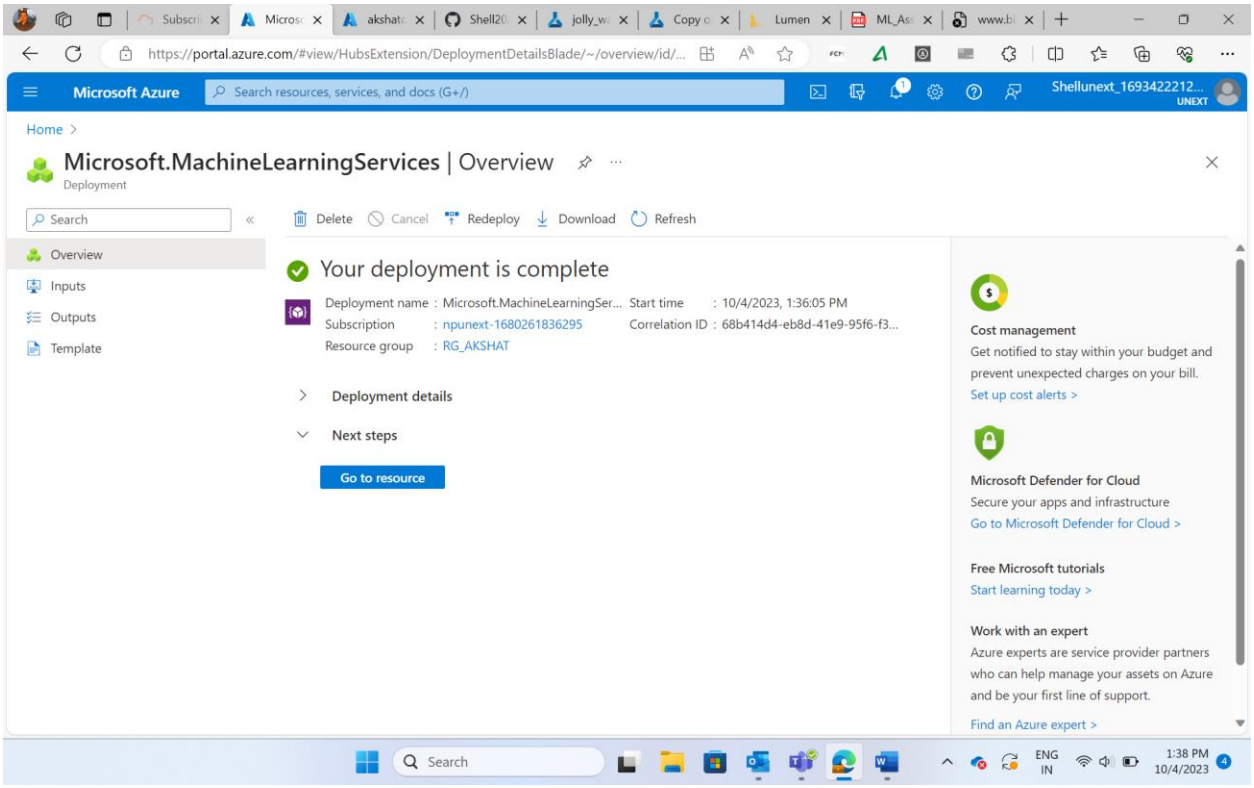
Uploading file to the blob storage account

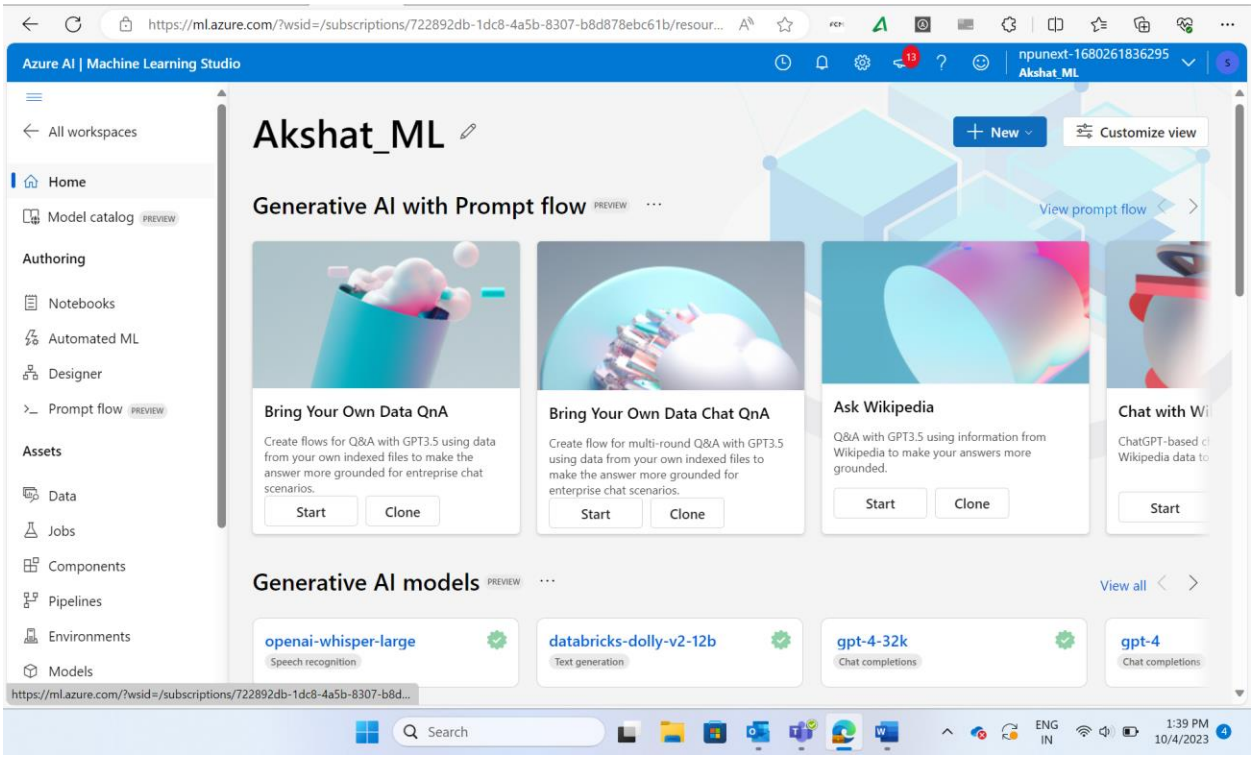




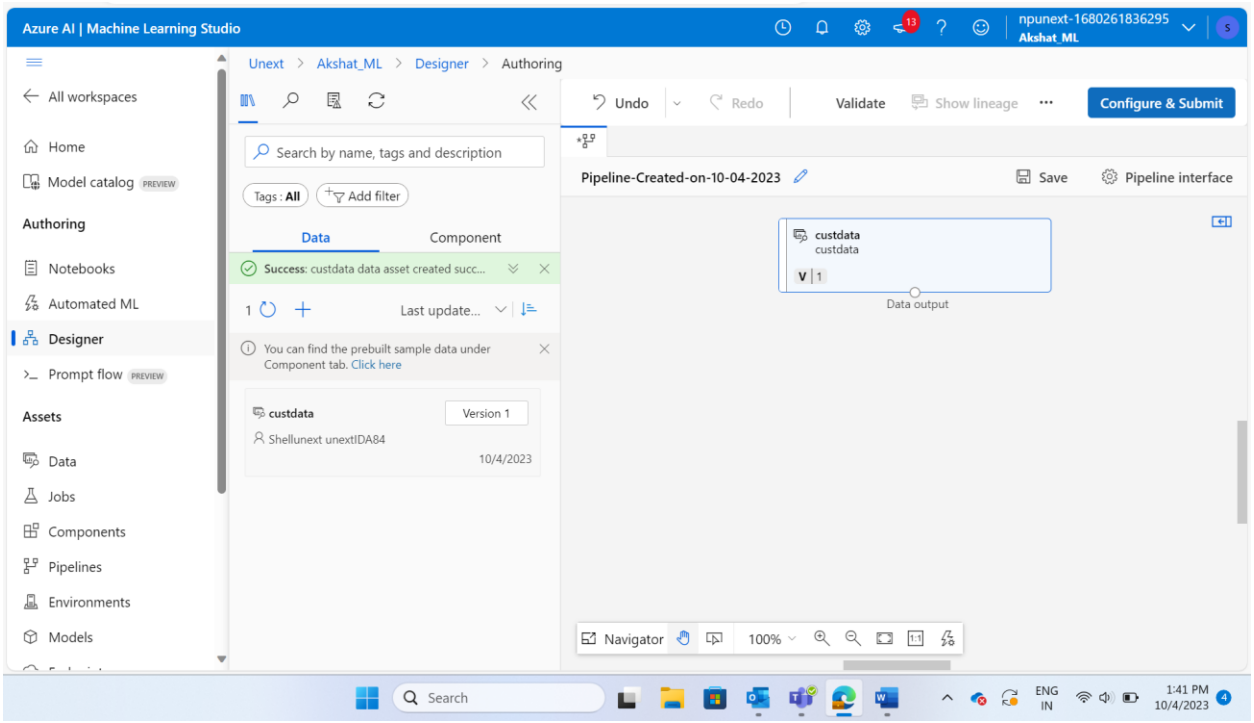
Machine learning service is created



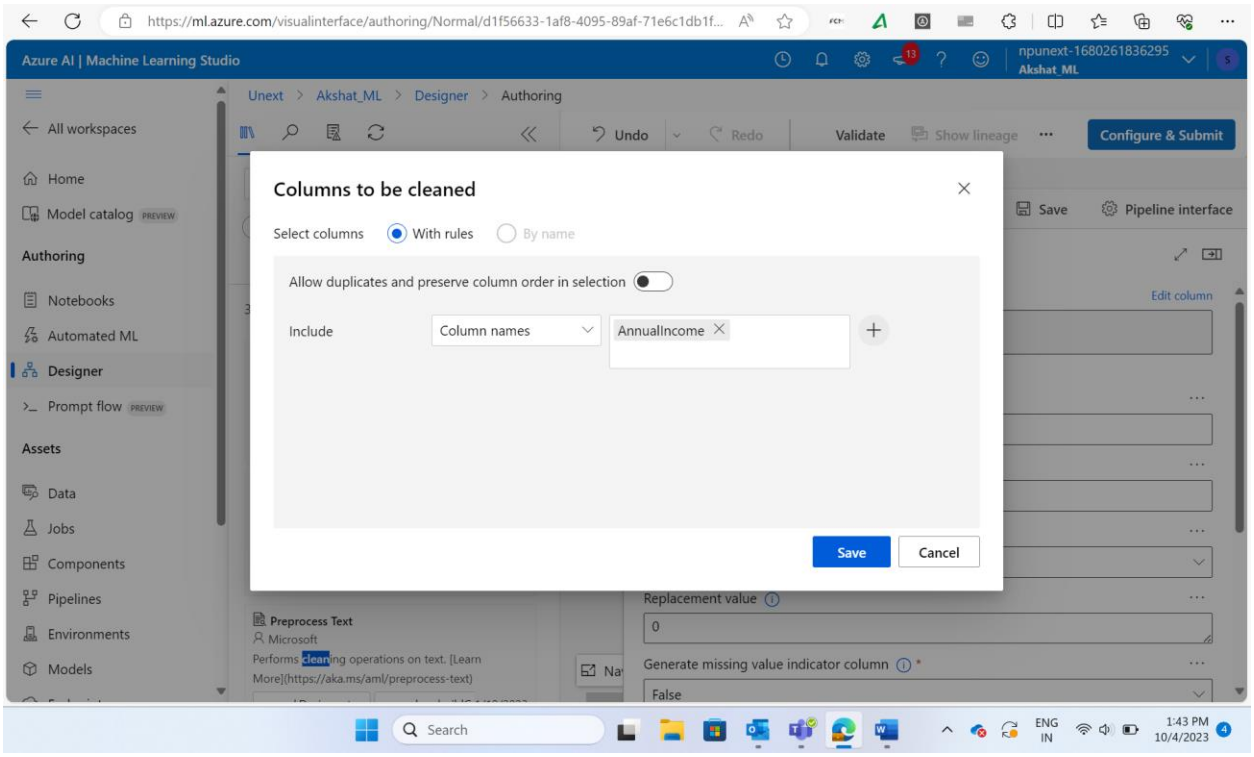
ML Studio is launched



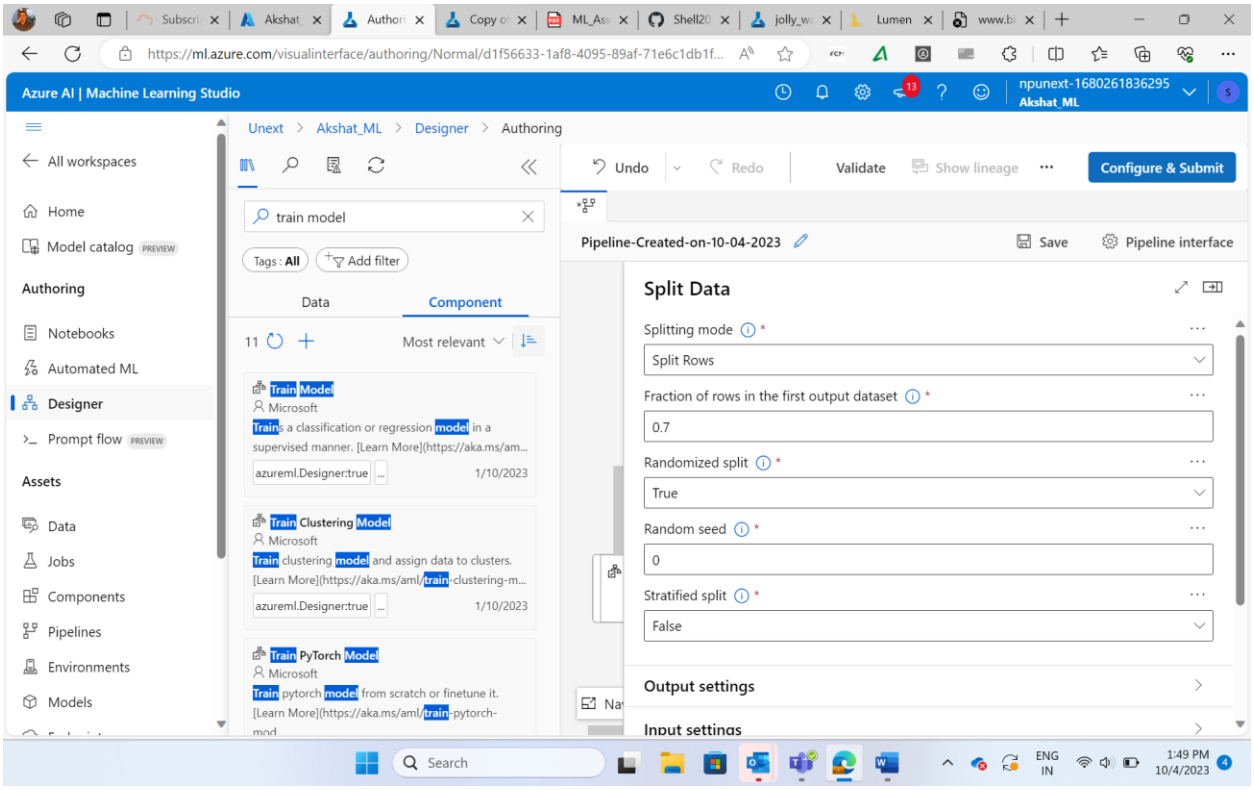
Dataset is loaded in ML Studio



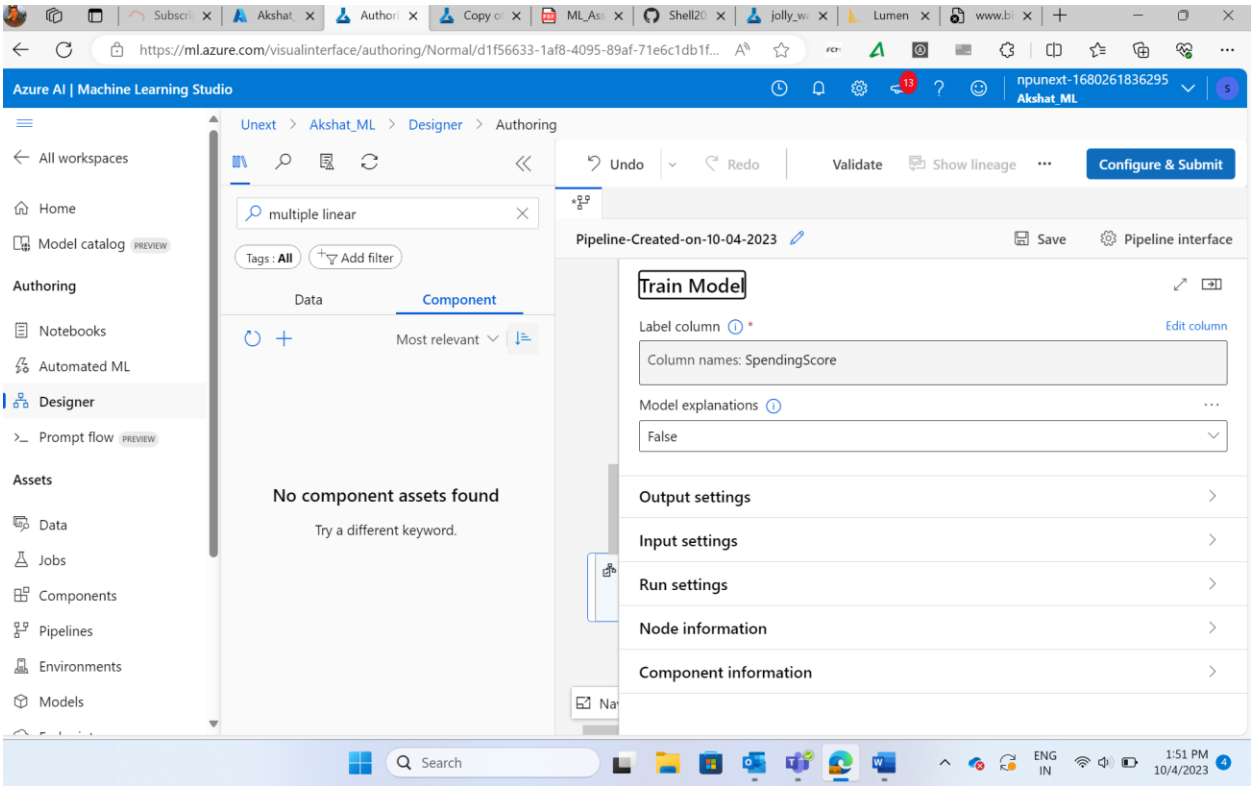
Cleaning activity – cleaning the annual income and spending score column since these have some empty rows like missing values so replacing the missing values with 0



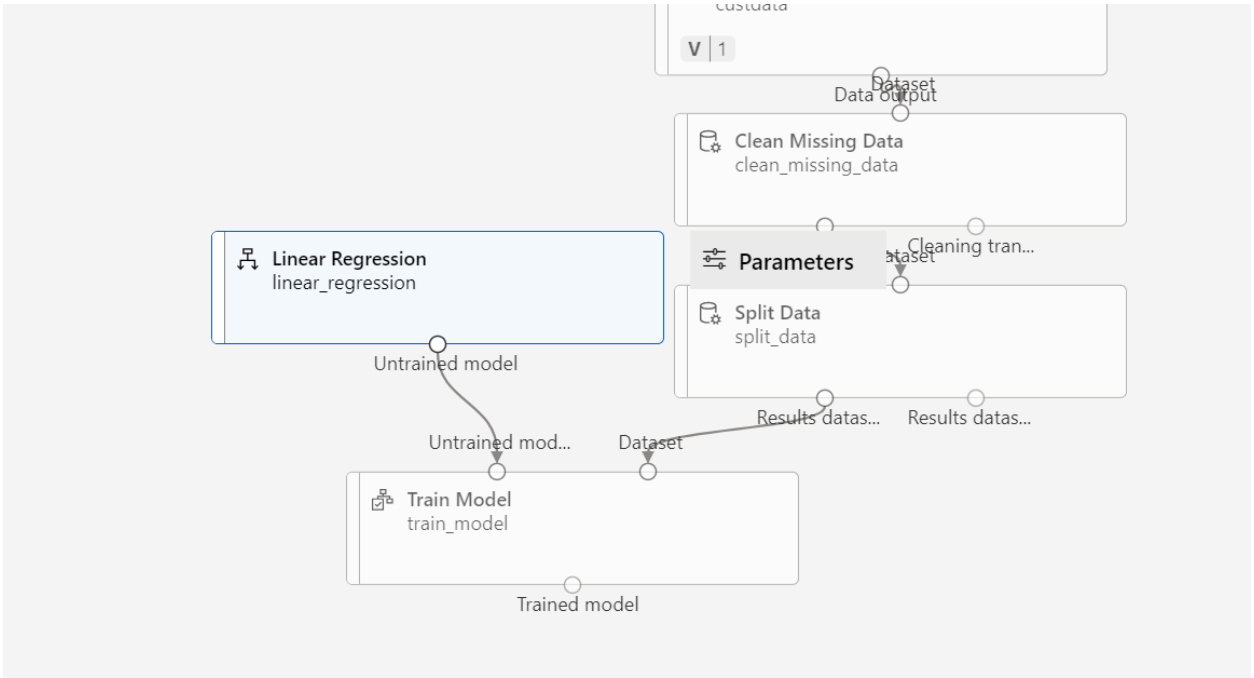
Splitting data into 70-30 ratio



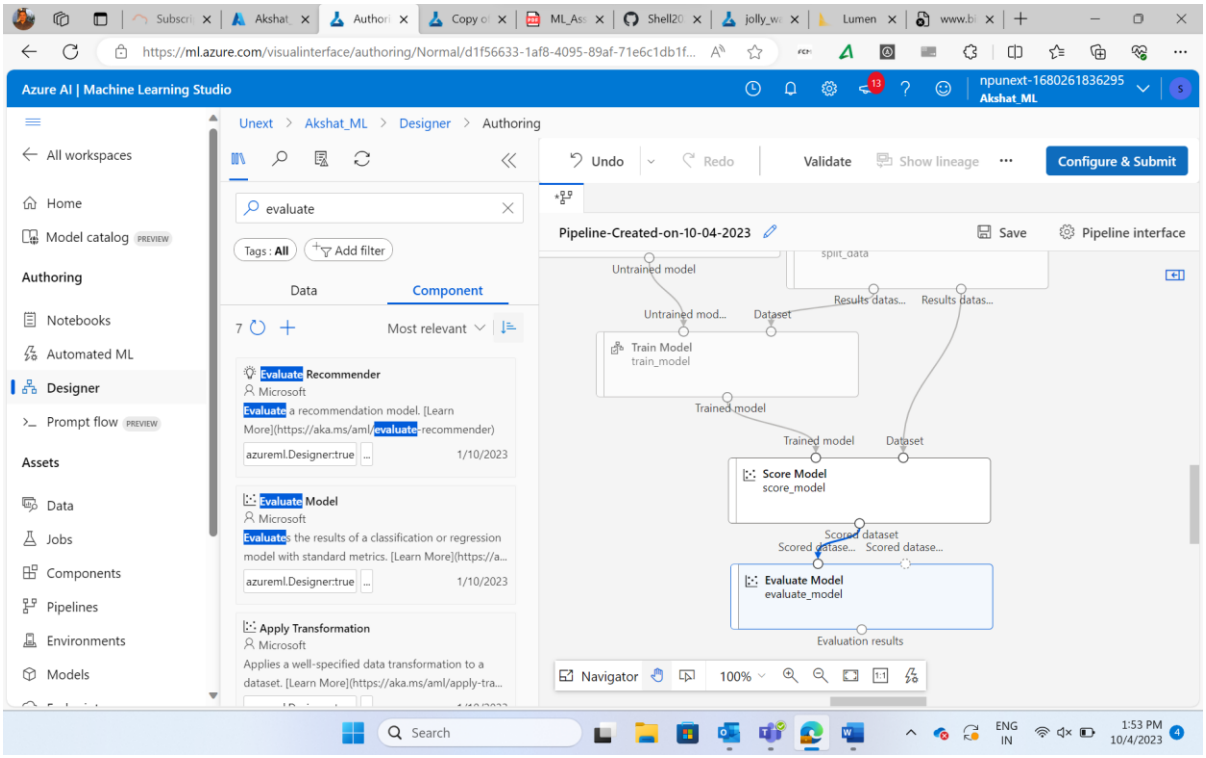
Setting column of train model as SpendingScore



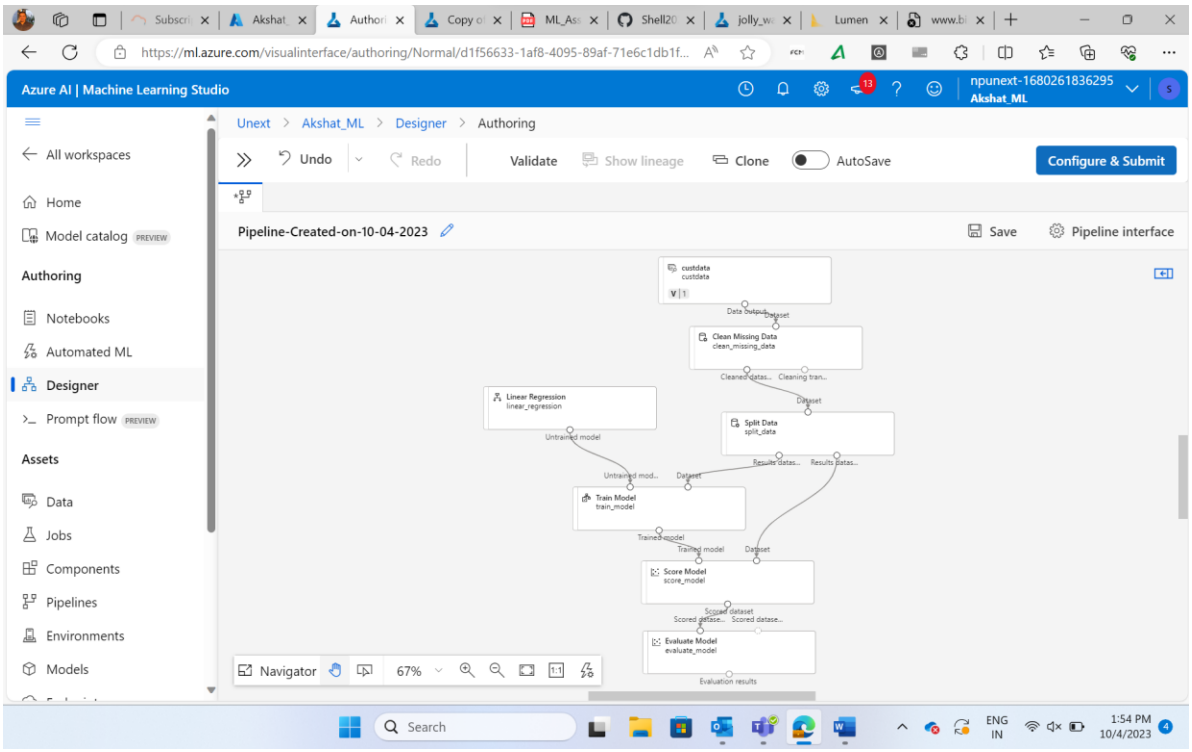
Using linear regression for the untrained model



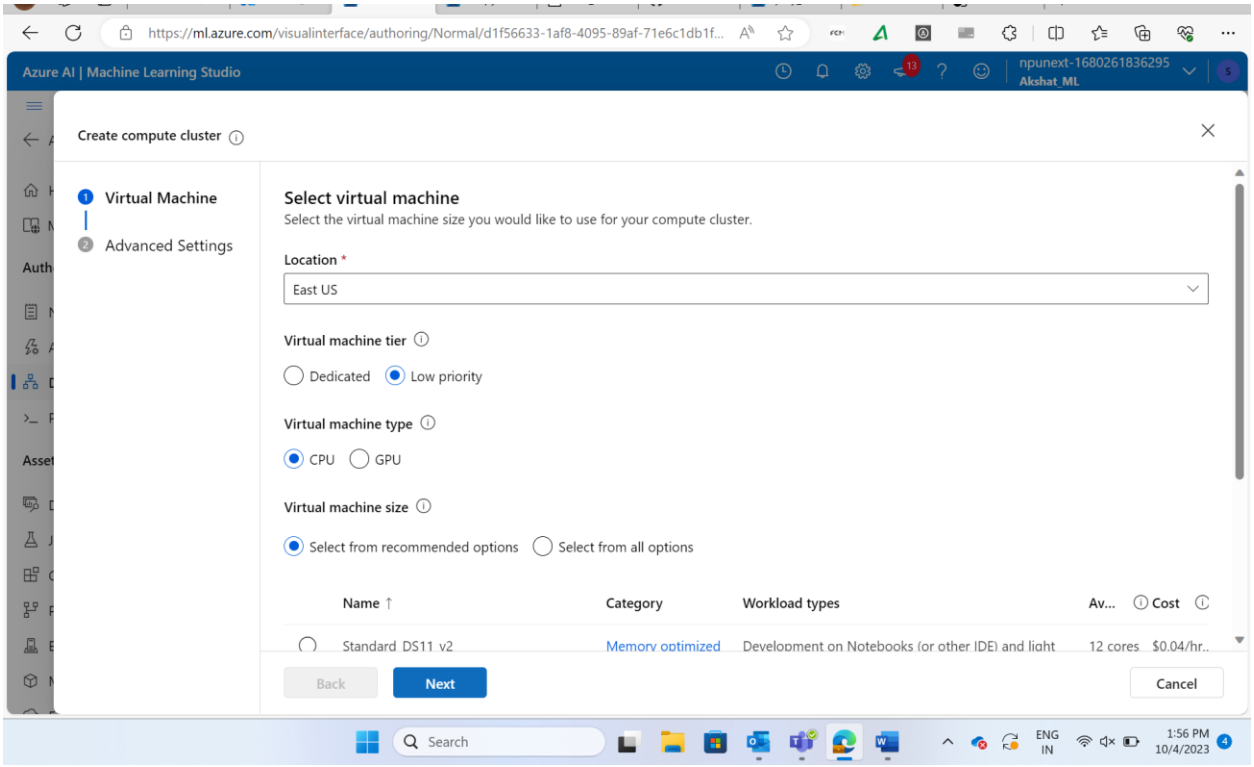
Adding ScoreModel and EvaluateModel components to tell us the results



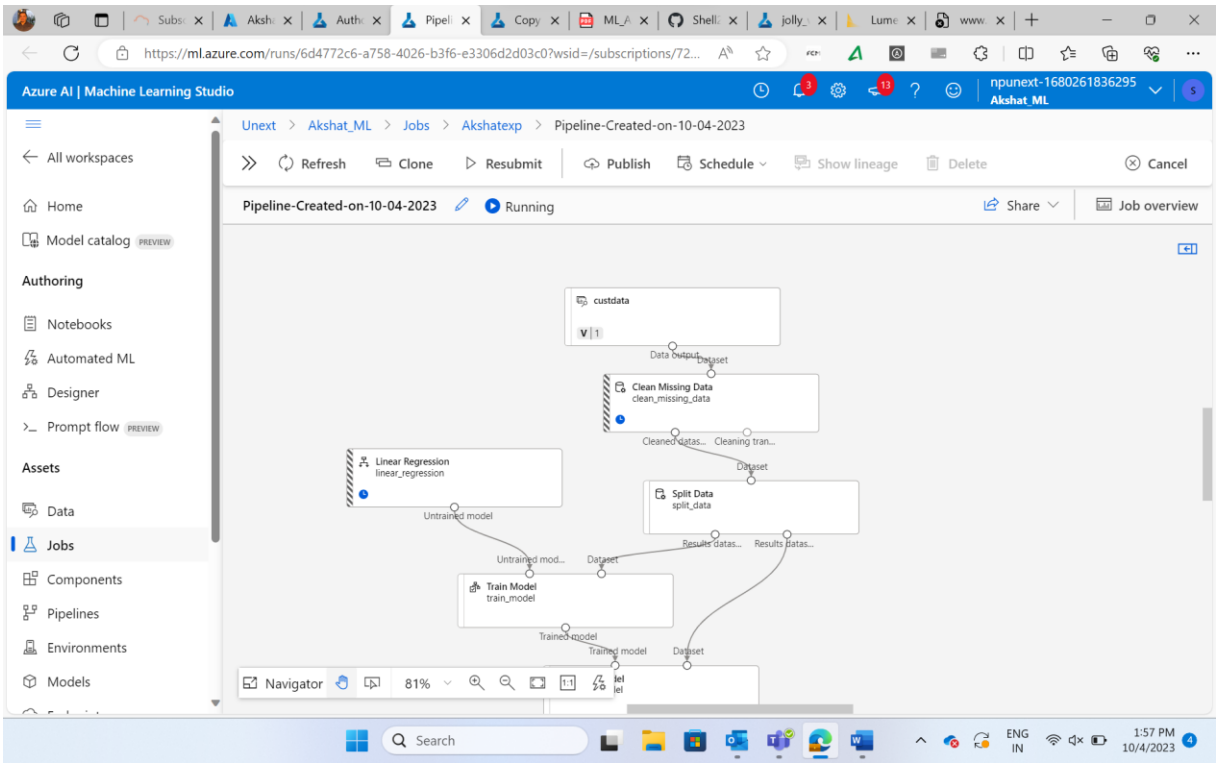
Final Pipeline looks like this



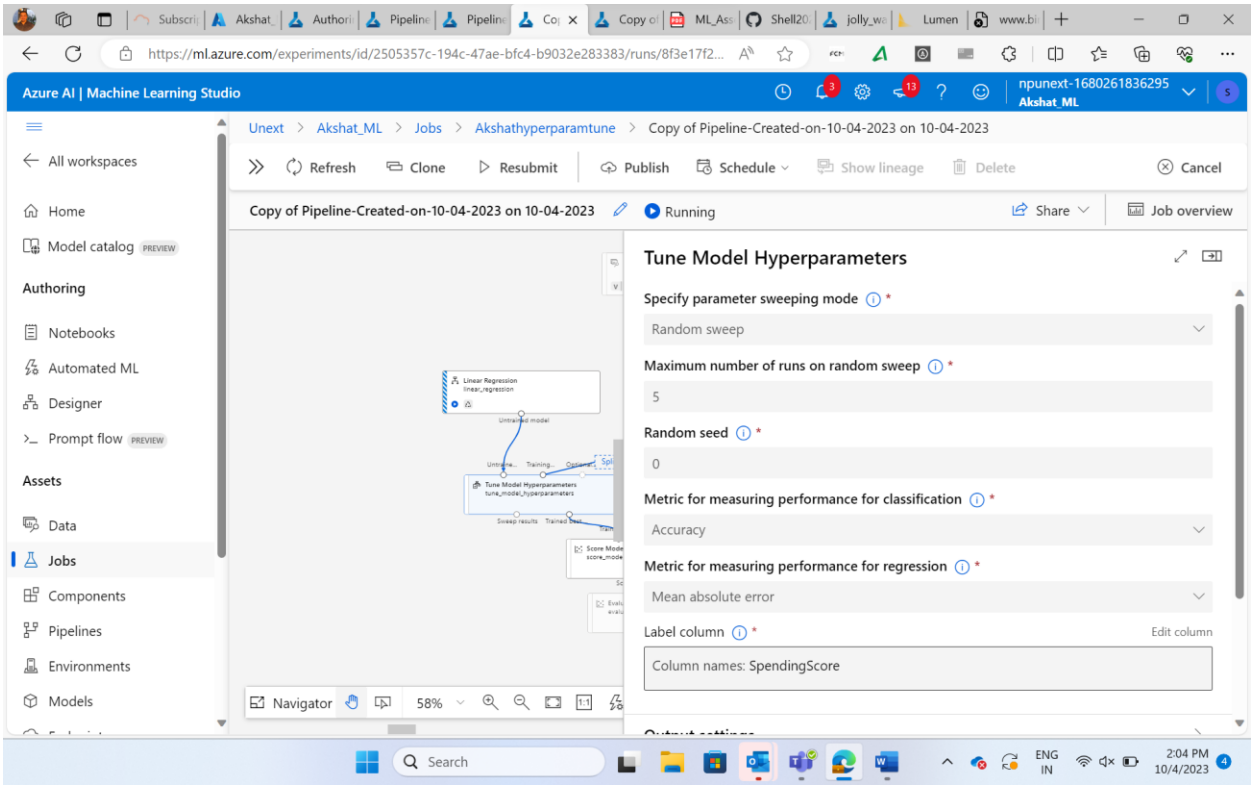
Creating compute cluster



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 datatset 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 patters 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.

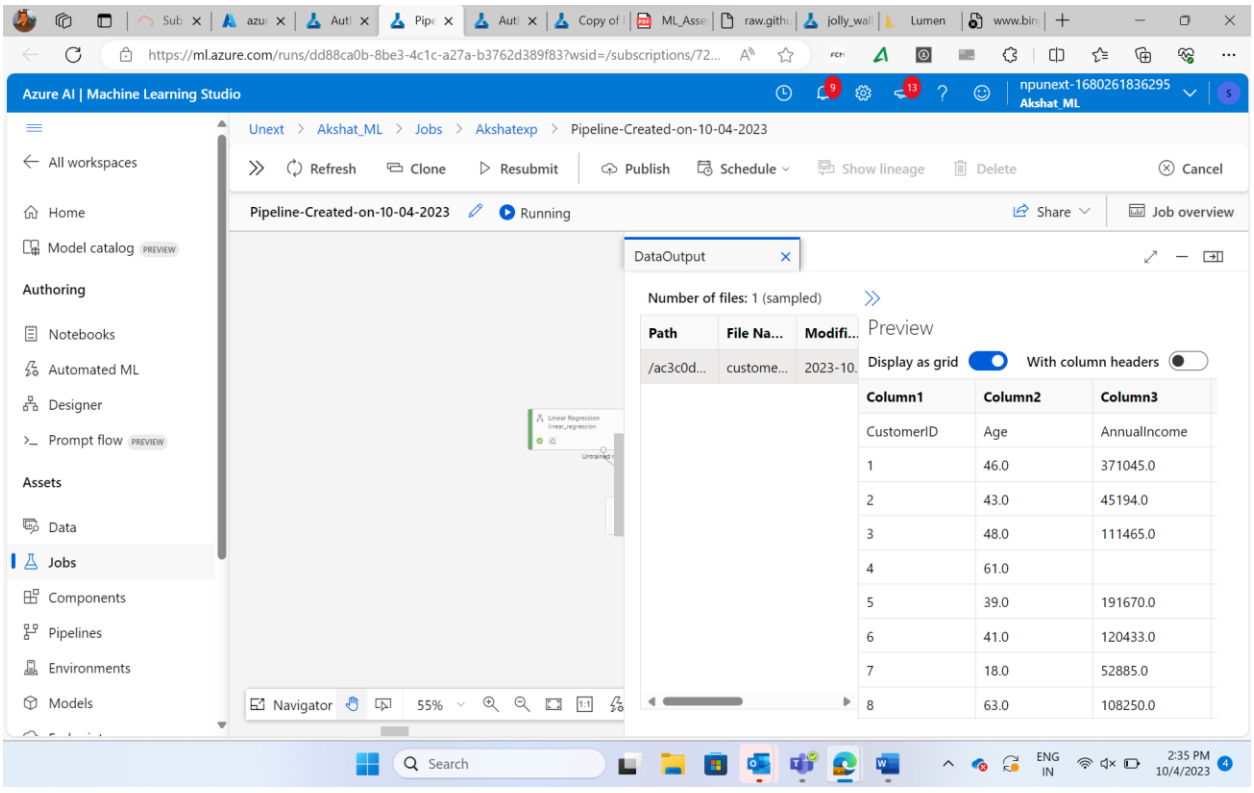
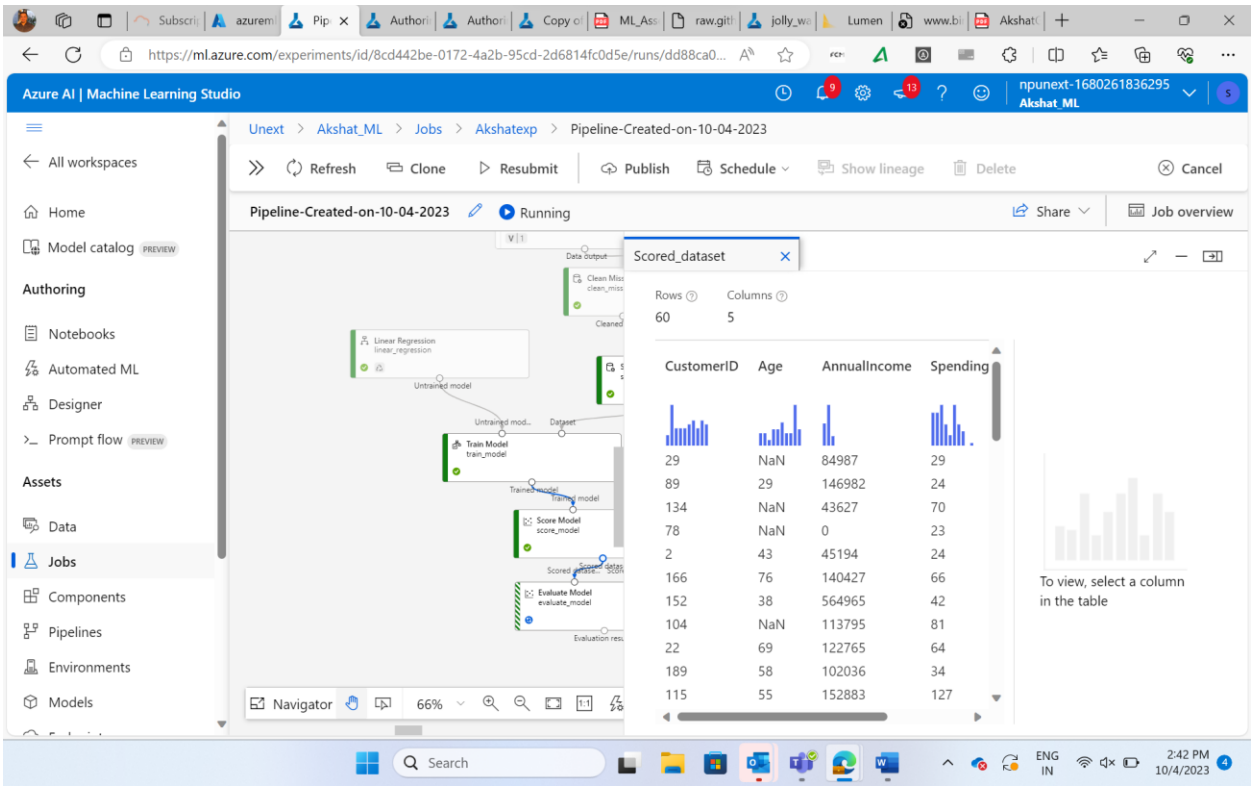
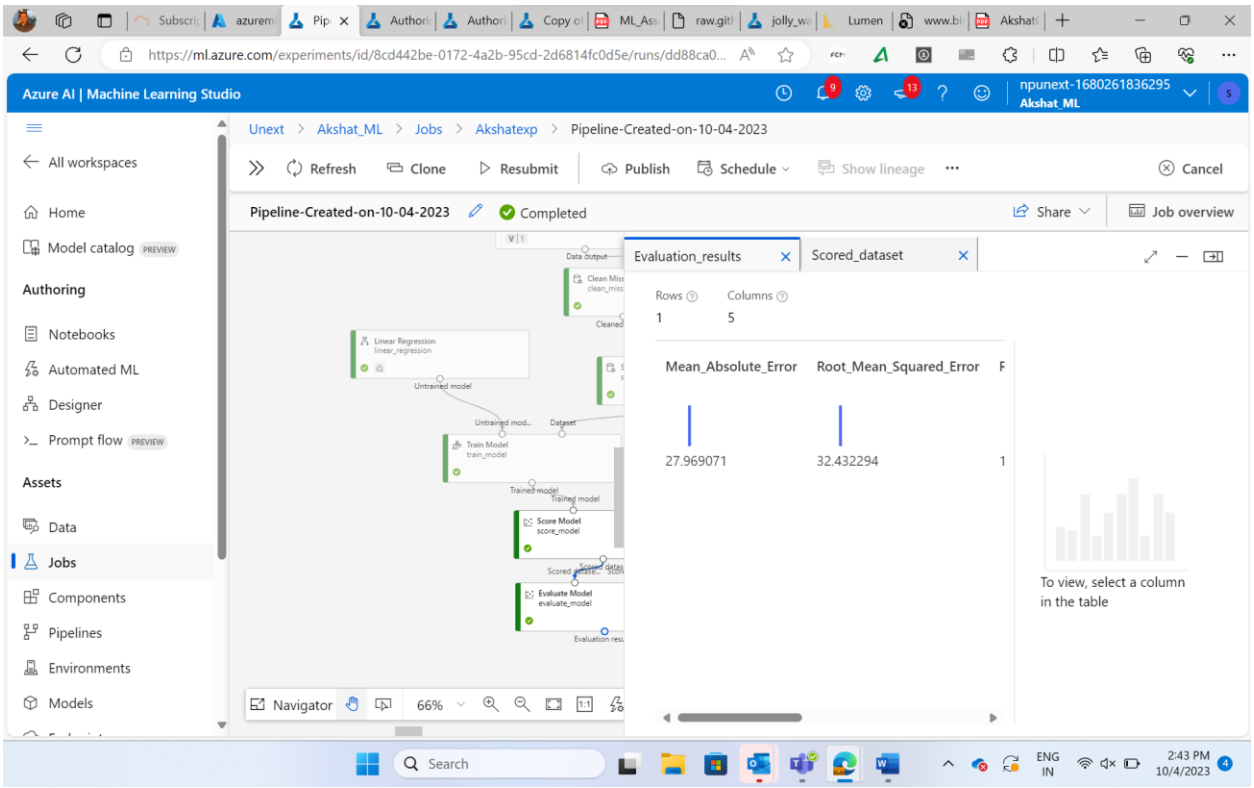
Linear regression is used as it is a supervised or a reinforcement learning which allows for linear regression to predict customer purchasing behavior.

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 -



Pipeline runs –