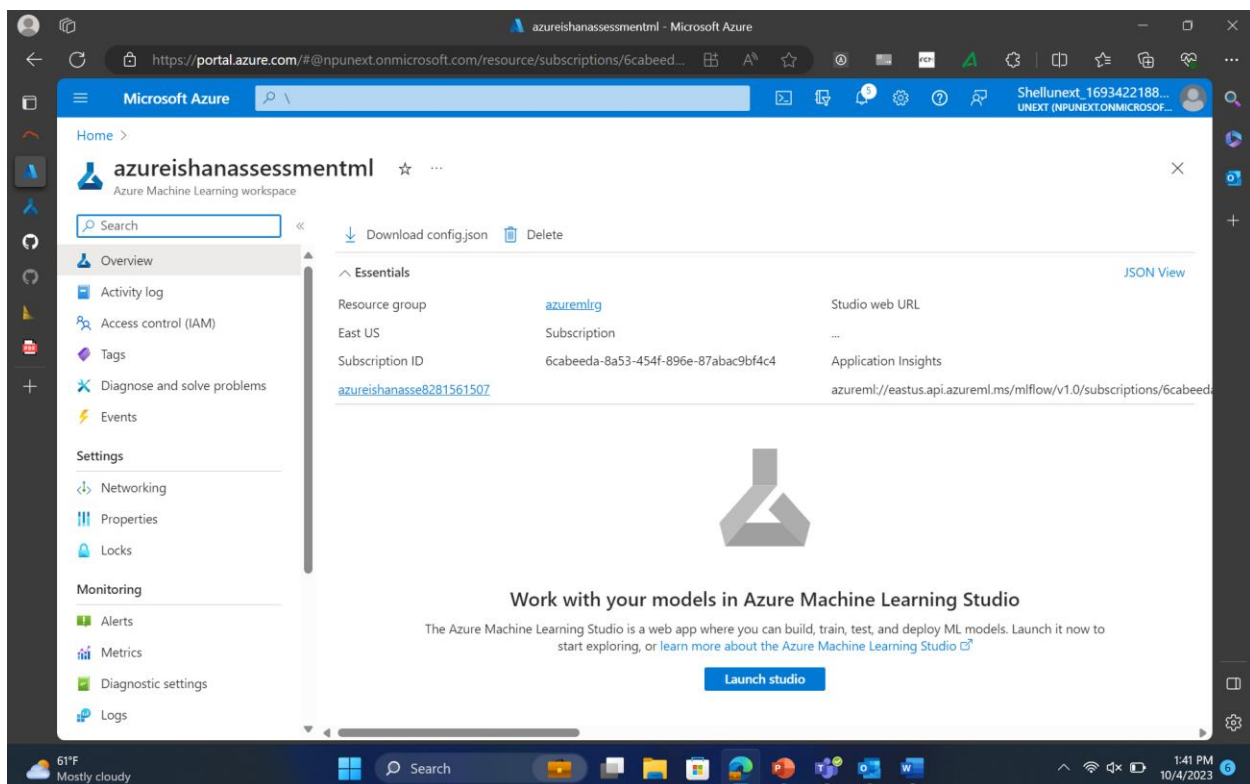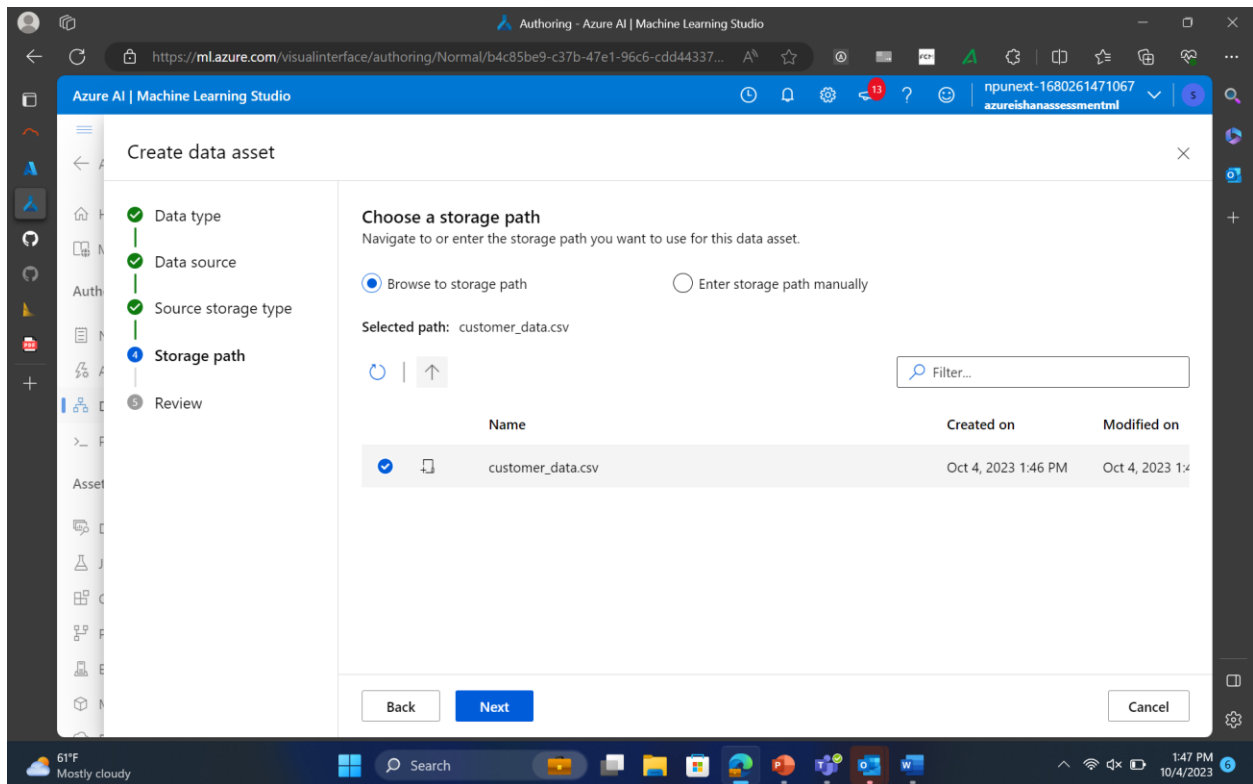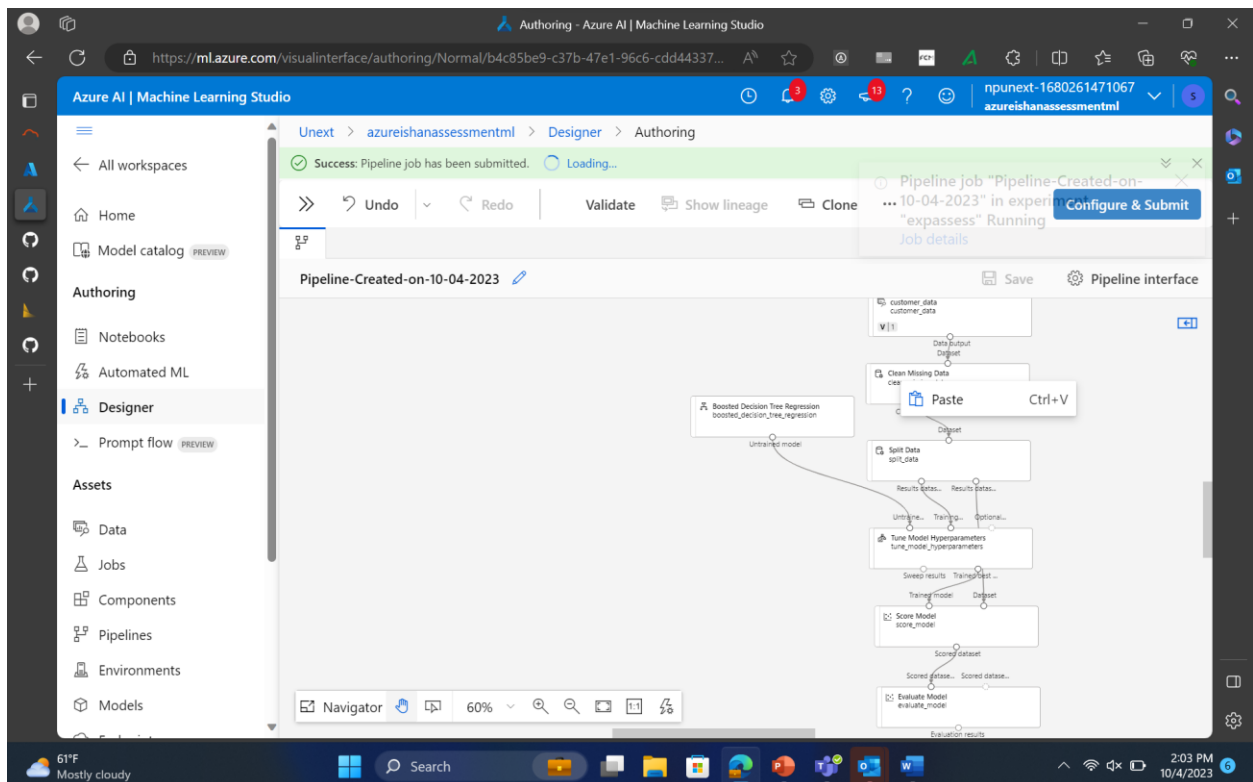1. Data uploaded in a blob storage

2. Azure ML workspace deployed
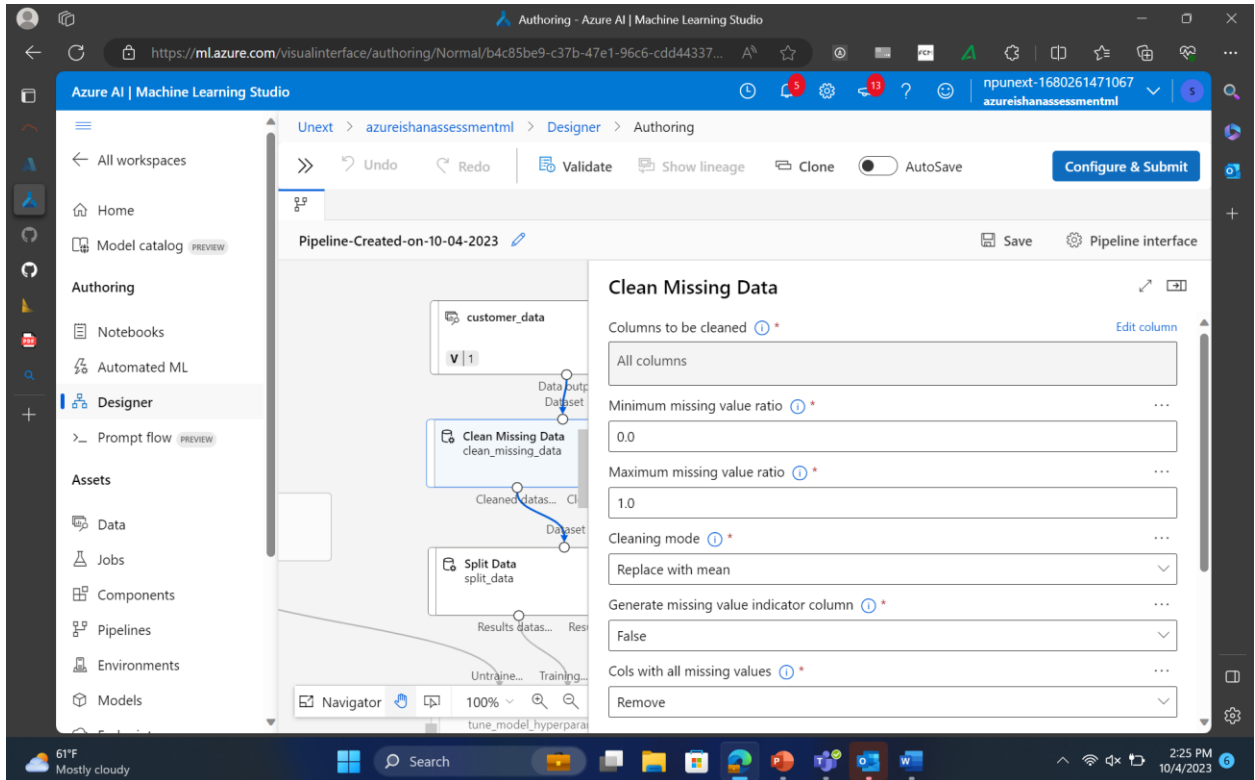

3. Data Asset uploaded
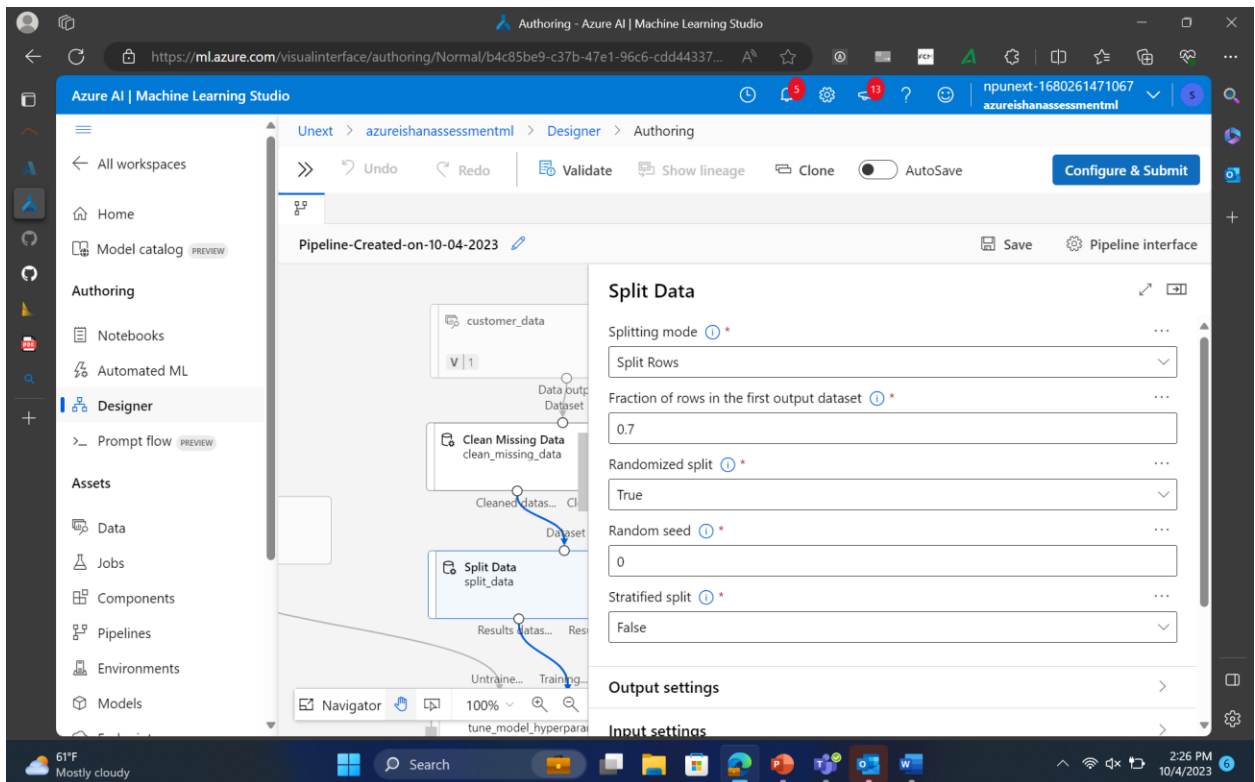
4. Applying the ML algorithm to predict the data.

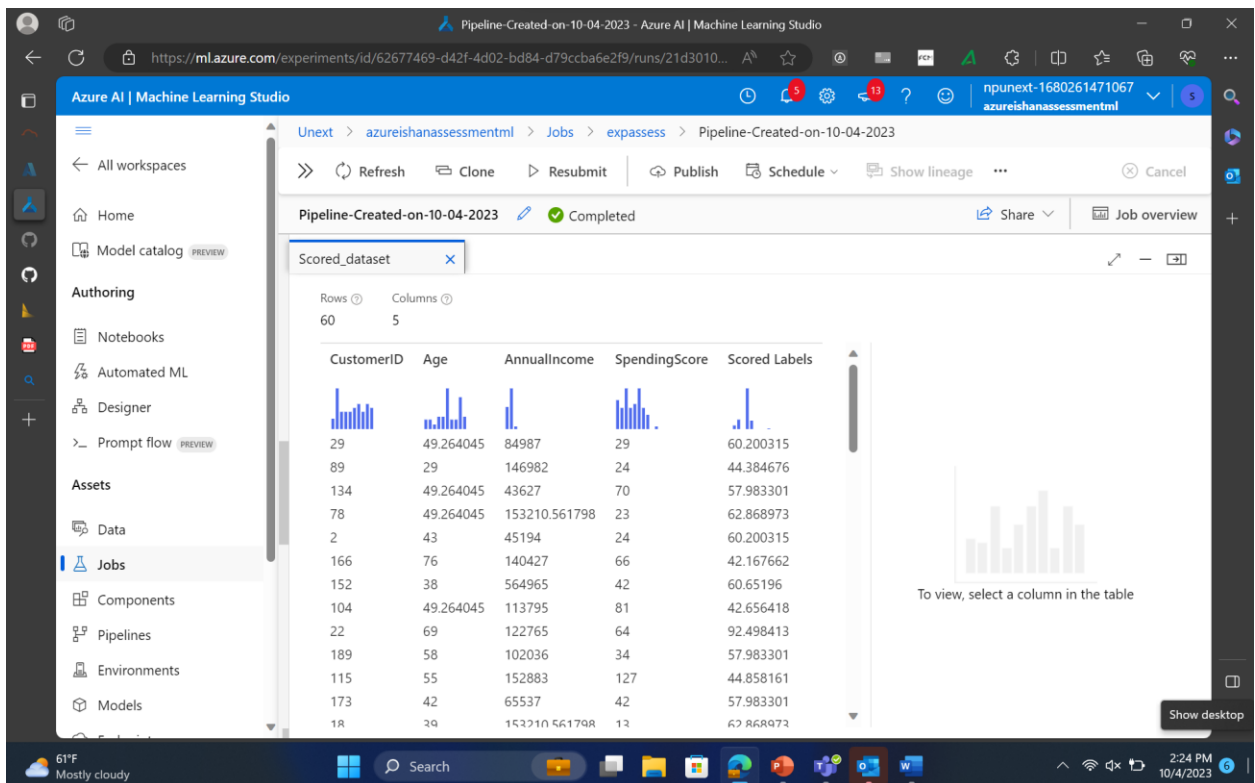THE STEPS ARE AS FOLLOWS:

1. Choosing the dataset

2. Cleaned the data by filling missing data with mean data.

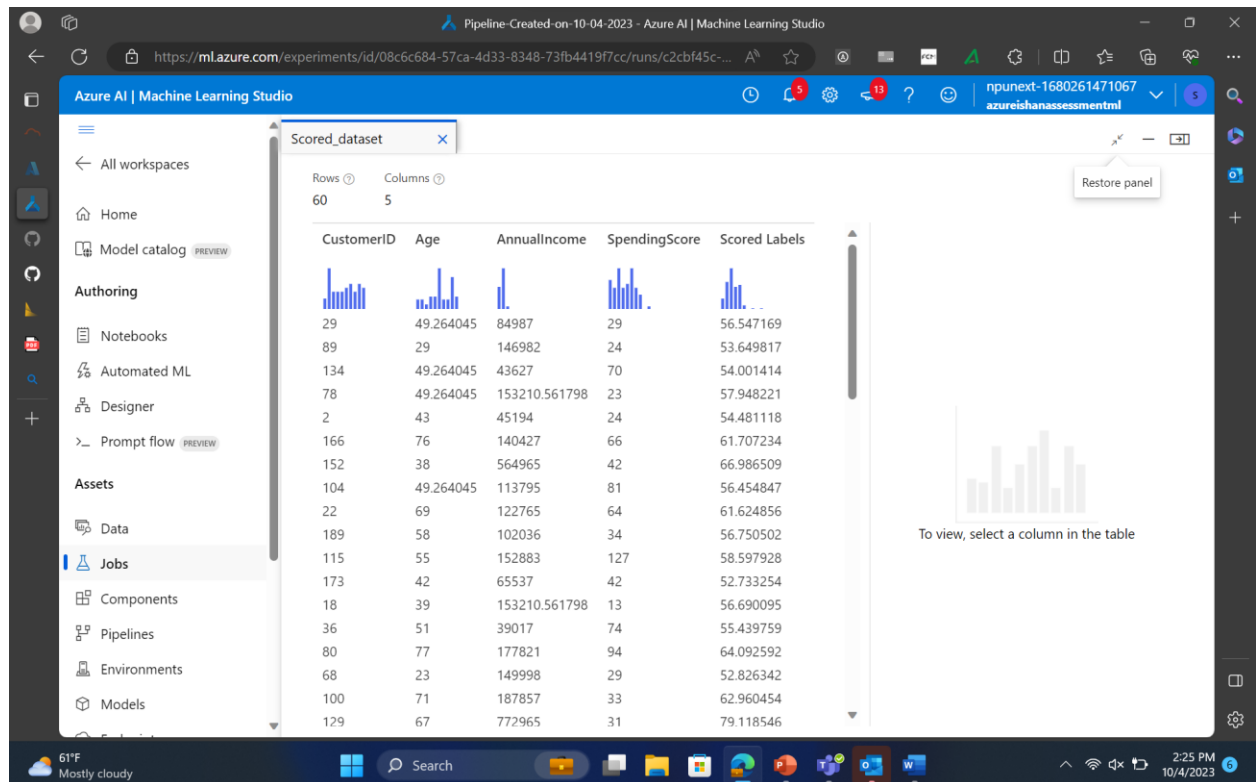3. Split dataset – 70% for training and the rest for testing.

4. Applying the algorithm – (Tried with both Decision tree Boosted and linear regression)

Using Boosted DT with hyperparameter tuning.

Rsquared error = 30

Using Linear Regression with hyperparameter tuning.

R squared error = 24

5. Scoring and evaluating the model.

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.

Ans. 1 The above detailed screenshots show all the steps.

Explanation:

1. Fetching the dataset
2. Cleaning the data : Imputing the missing data with custom/mean/median data or dropping thr data(unpreferred)
3. Splitting the dataset into train and test data. The train data helps the machine to learn for what kind of input, what output is expected and accordingly the model applied to the test data gets us the output using the ALGORITHM.
4. Further, certain tests such as r squared and least square approximation are used to test for error and optimize our model.

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?

Ans. 2 The train data helps the machine to learn for what kind of input, what output is expected and accordingly the model applied to the test data gets us the output using the ALGORITHM. It helps us to score and evaluate the model since the predicted value can be compared to the actual value and the accuracy can be found out.

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

Ans. 3 Since the column we are choosing for prediction is SpendingScore, which is a numerical value based on the Spent Amount, This can easily be solved by a regression algorithm such as linear regression, random forest or decision tree etc.

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

Ans 4. Sometimes, we are not aware of optimal values for hyperparameters which would generate the best model output. So, what we tell the model is to explore and select the optimal model architecture automatically. This selection procedure for hyperparameter is known as Hyperparameter Tuning.