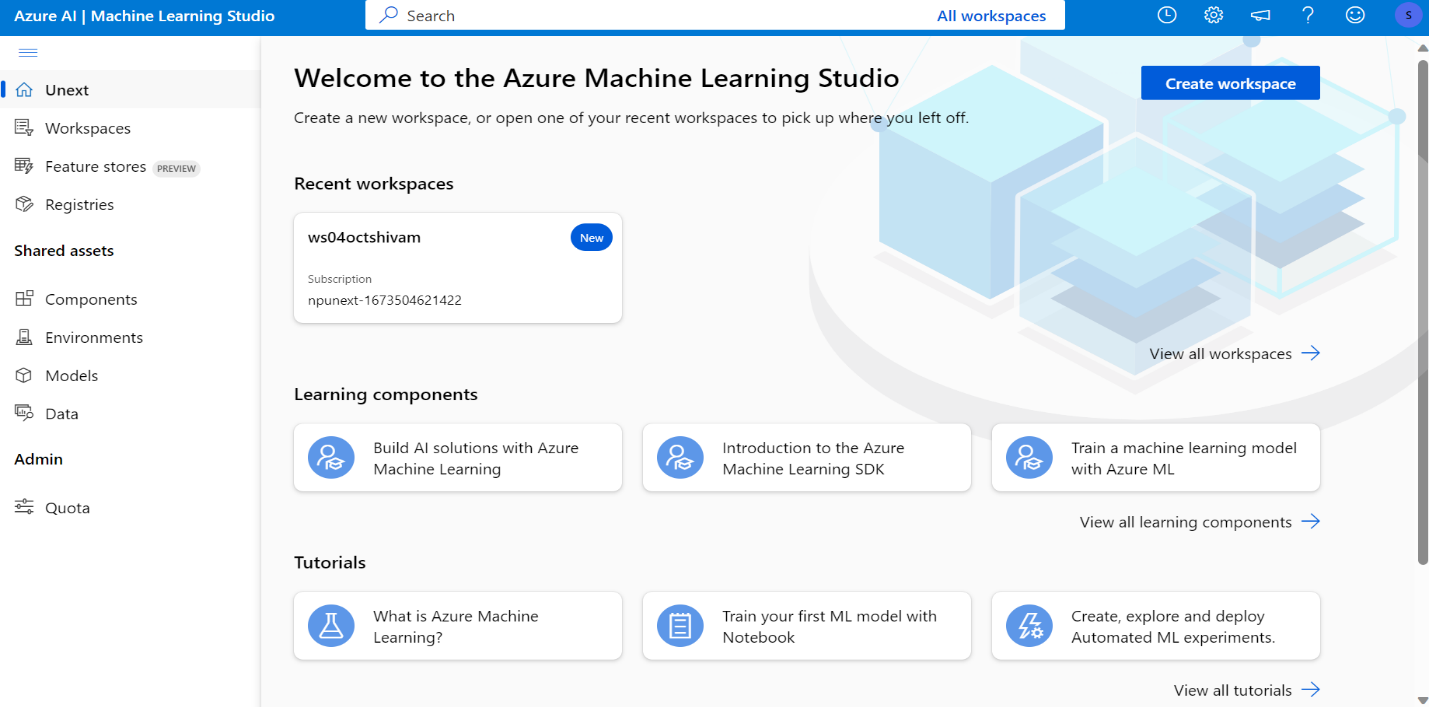
**Hands-On Assessment**

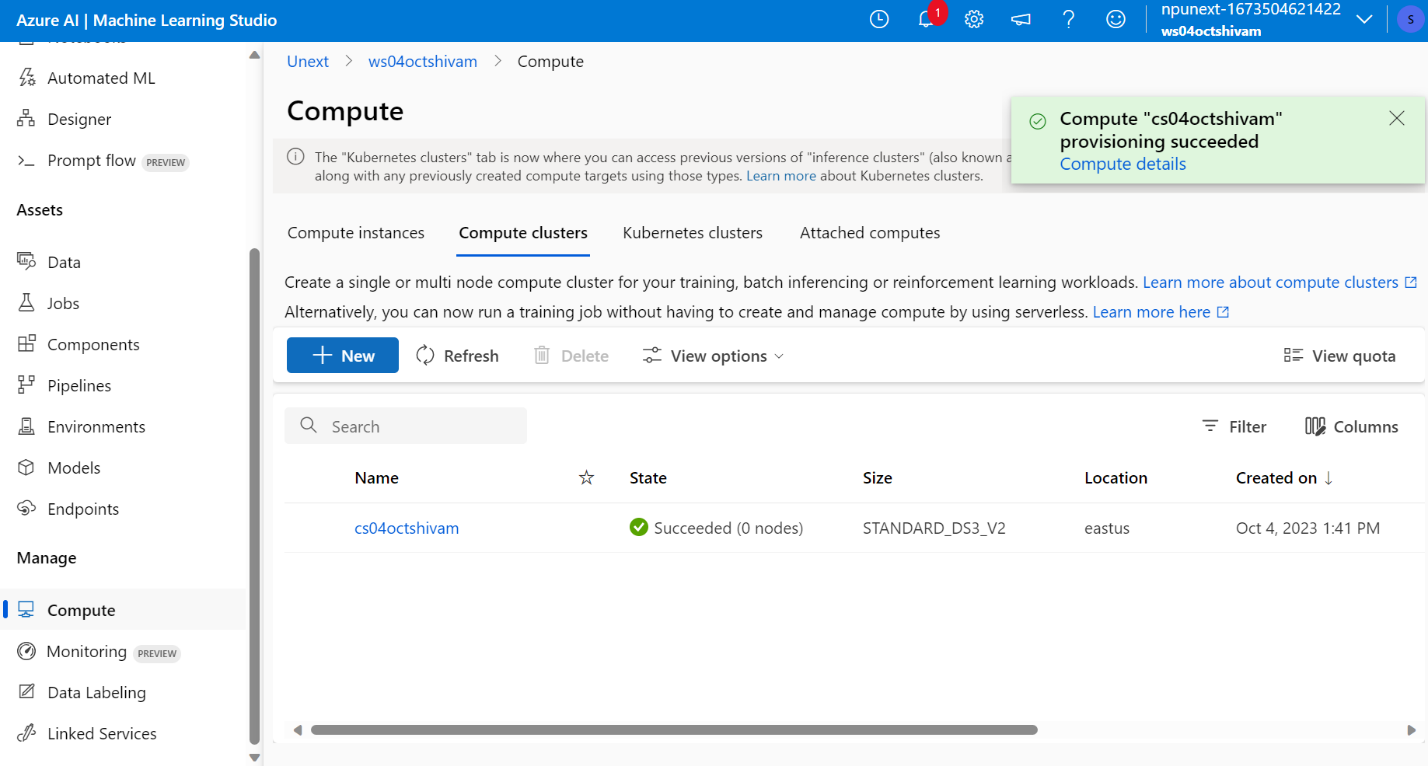
**ML**

**Solution:**

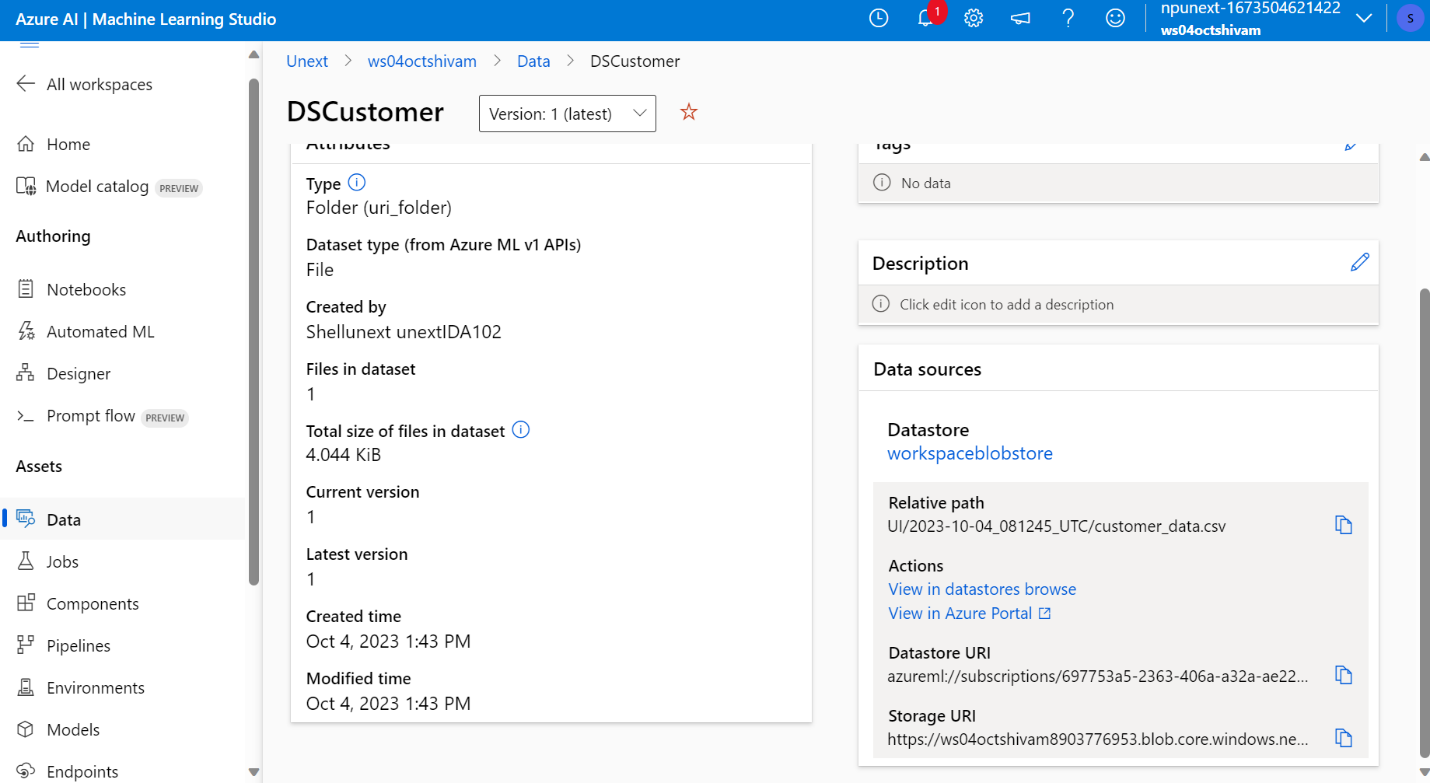
Workspace Created:



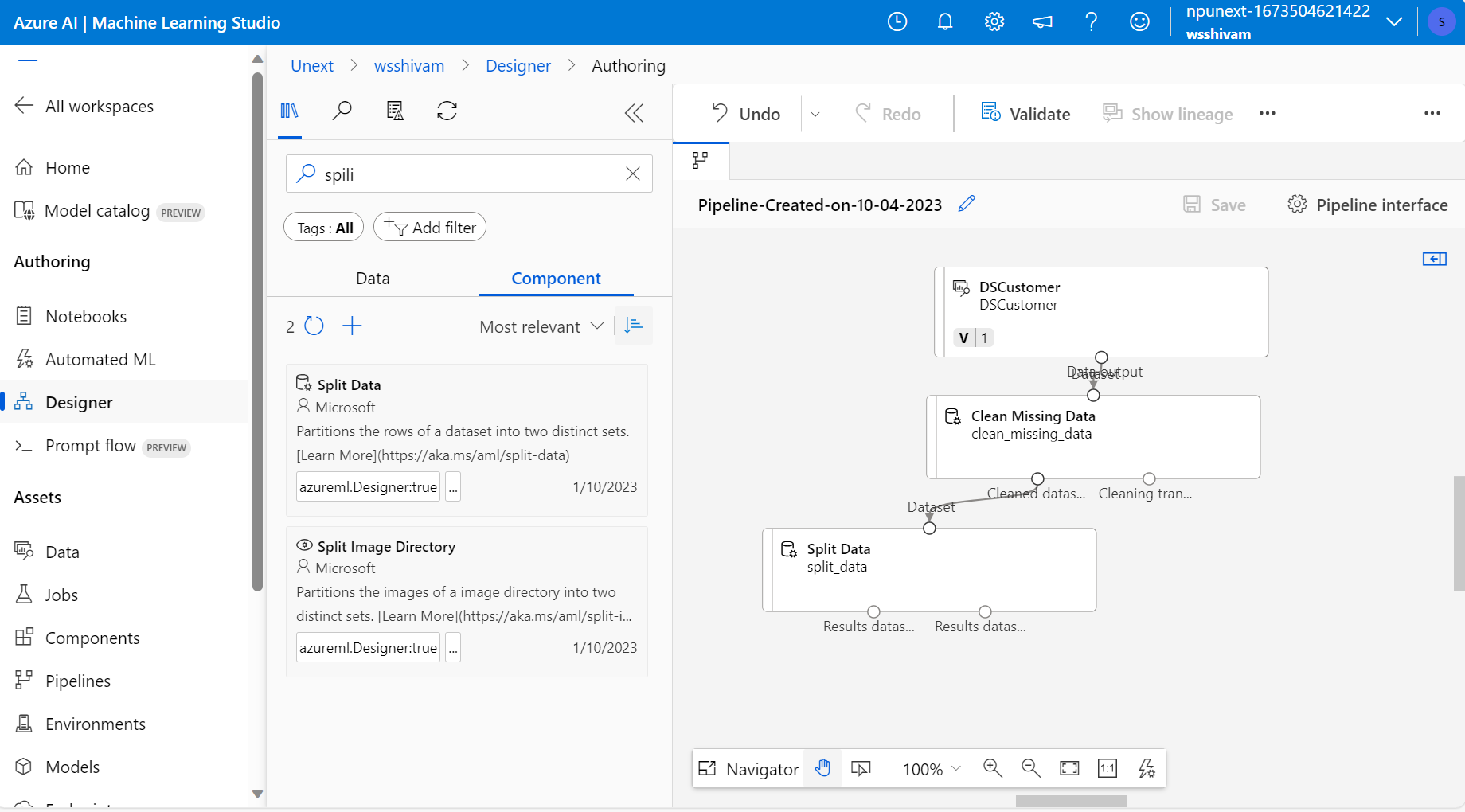
**Compute Cluster**



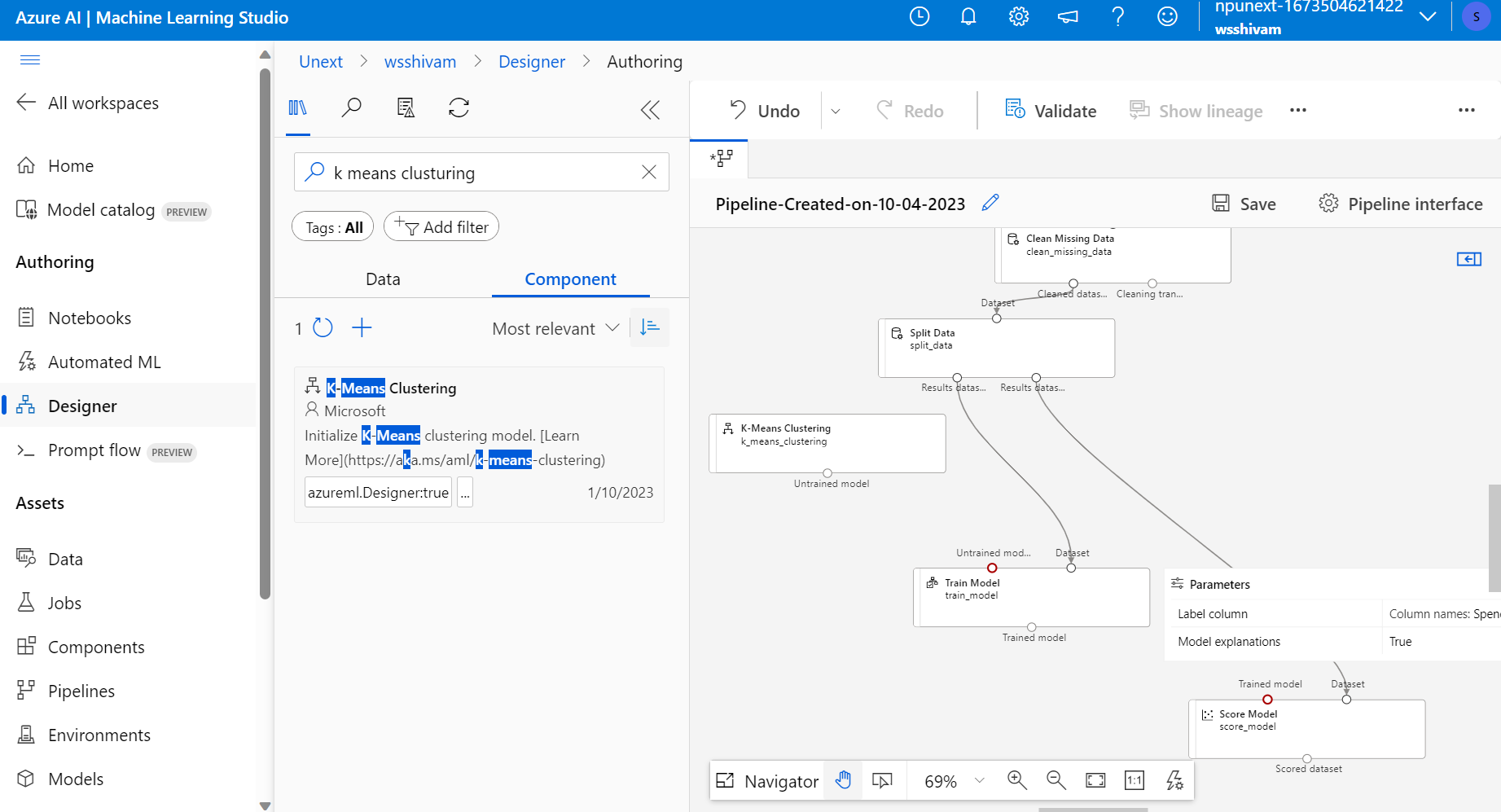
**DataSet Created and Loaded**

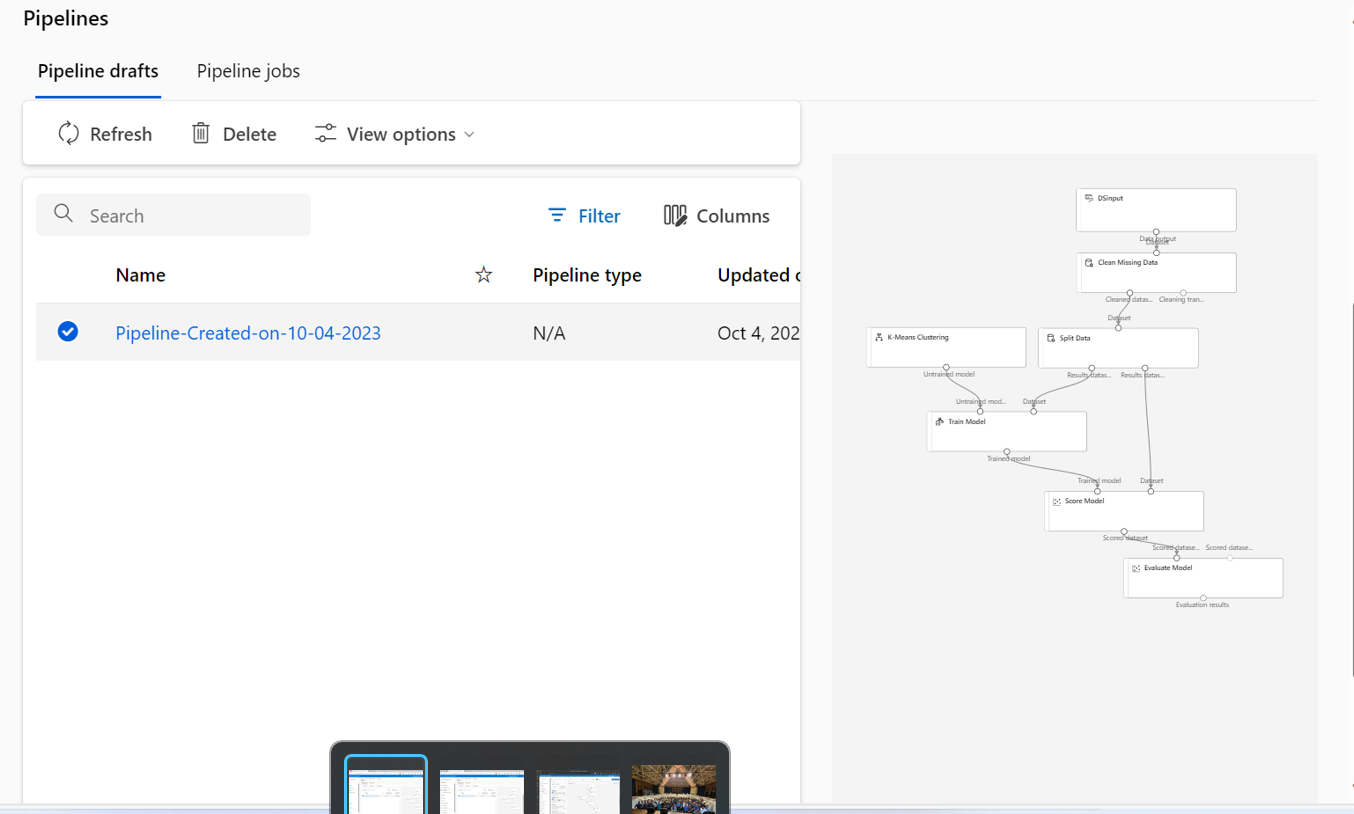


**Cleaning Missing Data and Spliting**

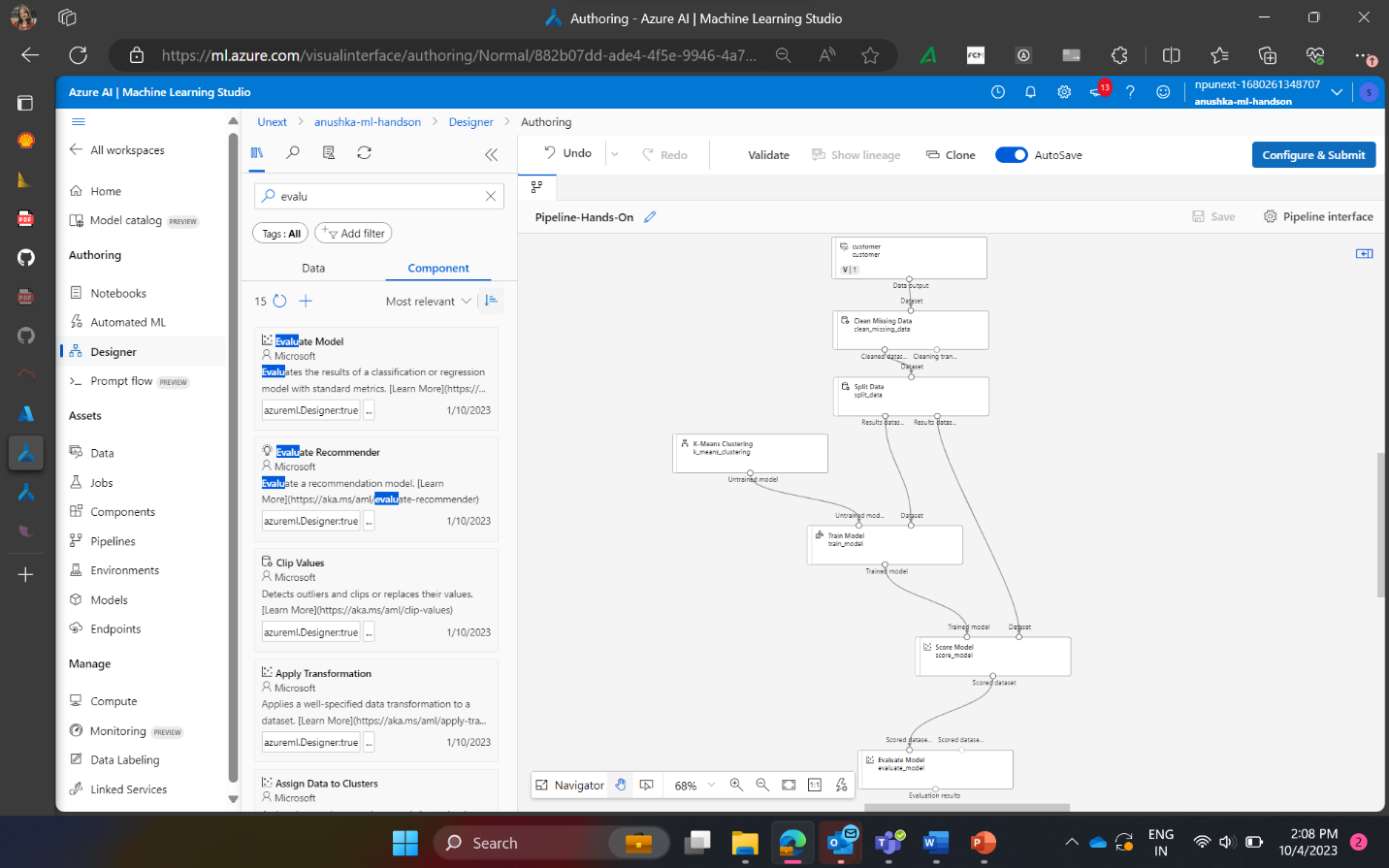


**Training the Dataset**



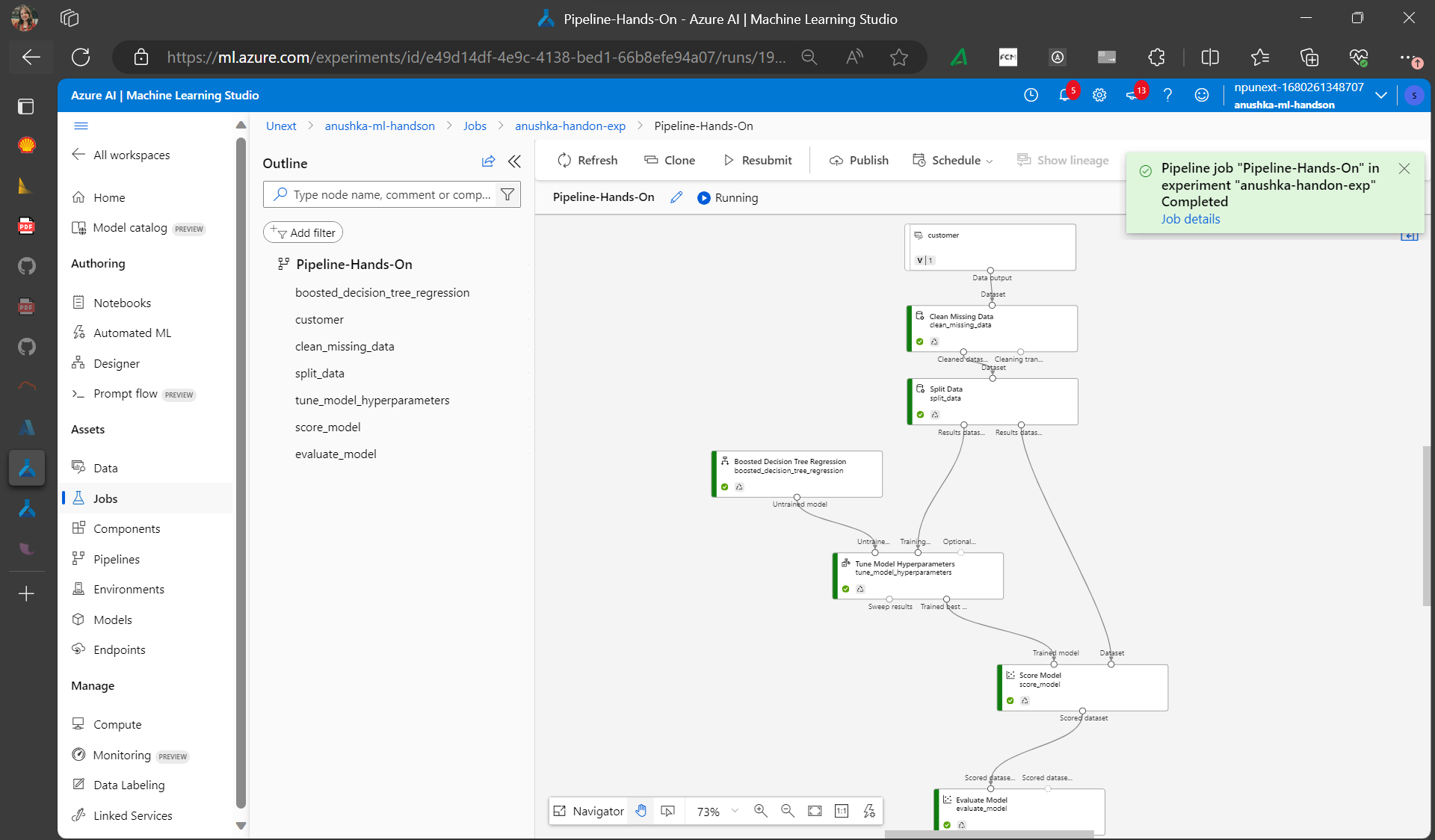


**Model Development Final Implementation**

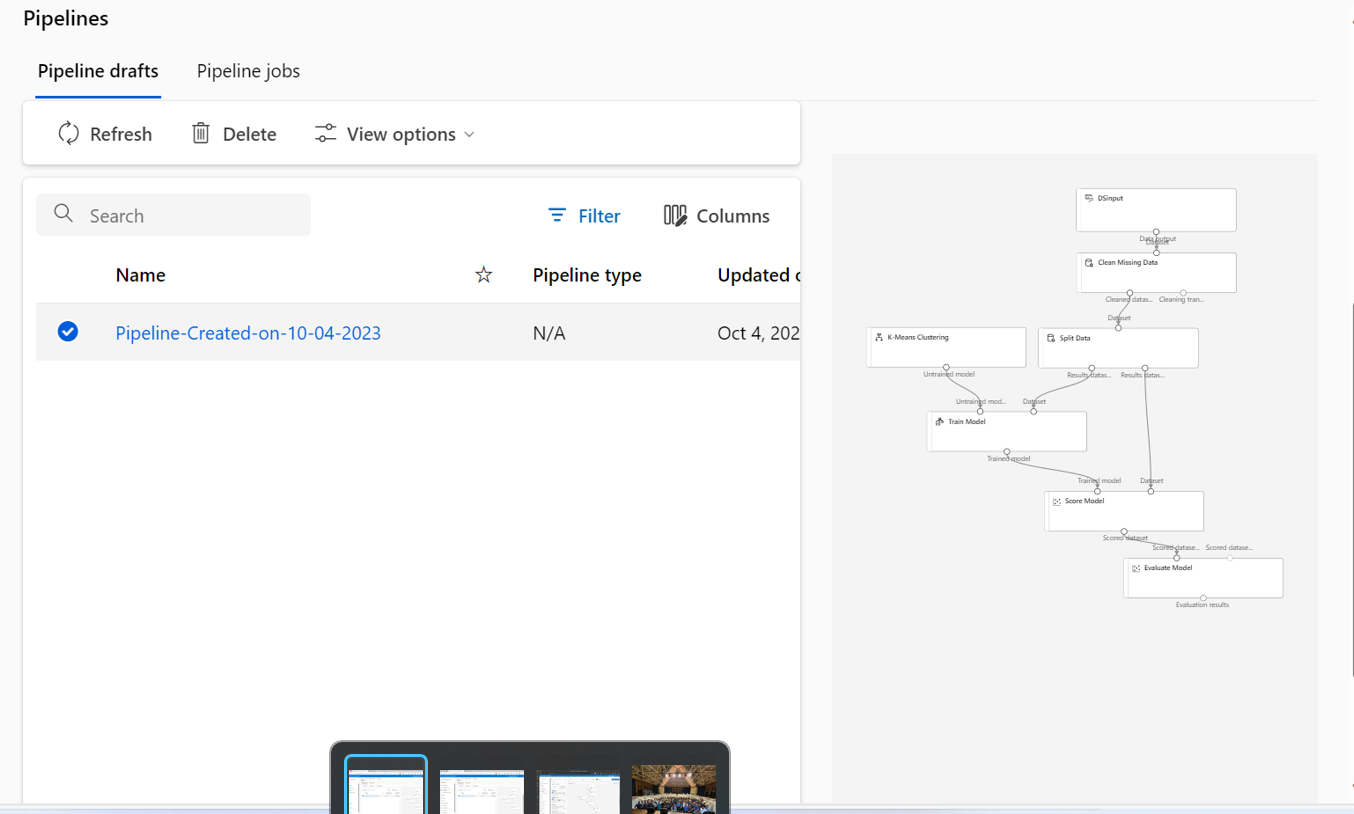


**A screenshot of a computer

Description automatically generated**

****

**HyperParameter Tuining**



**A screenshot of a computer

Description automatically generated**

Solutions:

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 :  Preparing a dataset for training a machine learning model using Azure Machine Learning involves several key steps. Below are the steps mentioned :

1. **Data Collection**: Gather the raw data from various sources, such as databases, files, APIs, or external datasets. Ensure that the data is comprehensive and representative of the problem you want to solve.
2. **Data Cleaning**: Clean the data to remove errors, inconsistencies, missing values, and outliers. This step is essential to ensure that the data is of high quality and suitable for training a machine learning model.
3. **Data Exploration**: Perform exploratory data analysis (EDA) to gain insights into the dataset's characteristics. Visualizations and statistical summaries can help you understand the data distribution, correlations, and patterns.
4. **Feature Engineering**: Create relevant features or transform existing ones to improve the model's performance. This step may involve encoding categorical variables, scaling numeric features, or generating new features from the existing ones.
5. **Data Splitting**: Divide the dataset into three subsets: training, validation, and testing. The training set is used to train the model, the validation set helps tune hyperparameters, and the testing set evaluates the model's generalization performance.
6. **Data Preprocessing**: Apply preprocessing techniques like normalization, standardization, or dimensionality reduction to make the data suitable for machine learning algorithms. This step ensures that all features have similar scales and that the data is in the right format.
7. **Data Transformation**: Depending on the specific ML algorithms used, you may need to transform the data into a specific format, such as one-hot encoding for categorical variables or sequence data for deep learning models.
8. **Data Pipeline**: Create a data pipeline that automates data preprocessing and transformation steps. Azure Machine Learning provides tools like Azure Machine Learning Dataflows to streamline this process.
9. **Data Versioning**: Track and manage different versions of your dataset to maintain reproducibility and traceability. Azure Machine Learning allows you to version datasets, making it easier to keep track of changes.
10. **Data Uploading**: Upload the preprocessed dataset to Azure Machine Learning's storage or workspace, making it accessible for model training and experimentation.

Ques: 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?

Solution:

Splitting the dataset into training and testing sets is a fundamental step in developing a machine learning model, and it serves several important purposes in model development and evaluation:

1. **Model Evaluation**: It allows you to assess how well your machine learning model generalizes to new, unseen data. The model's primary goal is to make accurate predictions on data it hasn't seen during training, and the testing set simulates this real-world scenario.
2. **Bias and Variance Evaluation**: By splitting the data, you can evaluate both bias and variance in your model. Bias refers to the error introduced by approximating a real-world problem, which may be complex, by a simplified model. Variance refers to the model's sensitivity to fluctuations in the training data.

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

Linear regression is a widely used machine learning and statistical technique for a variety of       reasons, and it's important to understand its strengths and common use cases:

1. **Simplicity**: Linear regression is one of the simplest and most interpretable machine learning algorithms. It's easy to understand and explain, making it accessible to both beginners and experts.
2. **Predicting Continuous Values**: Linear regression is primarily used for regression tasks, where the goal is to predict a continuous numerical value. For example, it can be used to predict stock prices, housing prices, or temperature.
3. **Linear Relationships**: When there is a linear relationship between the input features and the target variable, linear regression can work quite effectively. It's particularly useful when there is a roughly linear trend in the data.
4. **Interpretability**: Linear regression provides interpretable coefficients for each feature, which allows you to understand the impact and significance of each feature on the target variable.
5. **Efficiency**: Linear regression is computationally efficient and can handle large datasets and high-dimensional feature spaces with relative ease. It doesn't require extensive computational resources, making it suitable for various applications.
6. **Feature Importance**: It can help identify which features are most important for predicting the target variable, which is valuable for feature selection and engineering in more complex models.
7. **Regularization**: Regularized versions of linear regression, such as Ridge and Lasso regression, can handle multicollinearity (highly correlated features) and prevent overfitting.

Ques 4:

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

Solution:

1. **Hyperparameter Tuning**: When you split the data, you can also create a validation set, which is used for hyperparameter tuning. Hyperparameters are settings that control the model's learning process but are not learned from the data (unlike model parameters). You can adjust these hyperparameters based on the performance on the validation set to optimize the model's performance.Top of Form

**Grid Search**:

1. **Parameter Space Exploration**: Grid Search involves specifying a range of values or options for each hyperparameter that you want to tune. It creates a grid of all possible combinations of these hyperparameters.
2. **Model Training and Evaluation**: For each combination of hyperparameters in the grid, Grid Search trains a model using the training data and evaluates its performance using a validation dataset or cross-validation. It typically uses a performance metric (e.g., accuracy, F1-score, mean squared error) to assess model performance.

**Systematic and Exhaustive Search**: Grid Search is a systematic approach that explores the entire predefined hyperparameter space. It ensures that you consider a wide range of hyperparameter combinations, increasing the likelihood of finding the best configuration.

**Parameter Importance**: Grid Search provides insights into the importance of different hyperparameters. You can observe which hyperparameters have the most significant impact on model performance.

**Parallelization**: Grid Search can be parallelized, allowing you to explore multiple combinations simultaneously, which can significantly reduce the time required for hyperparameter tuning, especially on large compute clusters.Top of Form