**Project Report: Penguin Species Clustering Using Azure ML**

**Name: Muhammad Munawar Khan  
Roll No: NUM-BSCS-2022-13  
Course: Coursera – Machine Learning with Microsoft Azure**

**Objective**

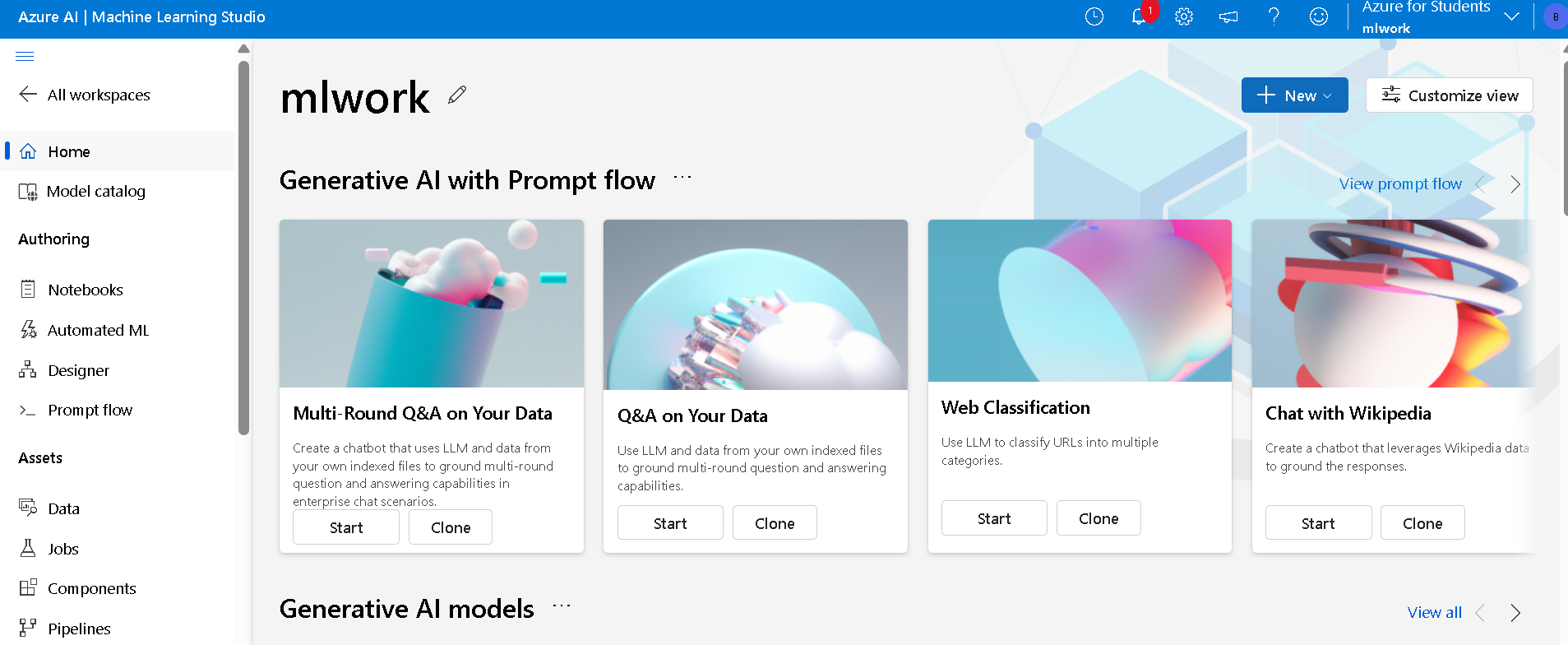
To build and deploy an unsupervised clustering machine learning model using Azure Machine Learning Designer that groups penguins into distinct species based on their physical measurements and characteristics.

**Part 1: Create an Azure Machine Learning Workspace**

Created a new Azure Machine Learning workspace in the Azure portal.

Selected a globally unique workspace name, region, and associated it with a resource group.

Accessed the workspace using Azure ML Studio at [ml.azure.com](https://ml.azure.com).  
 **Screenshot: Workspace dashboard in Azure ML Studio**



**Part 2: Create Compute Resources**

Compute Cluster

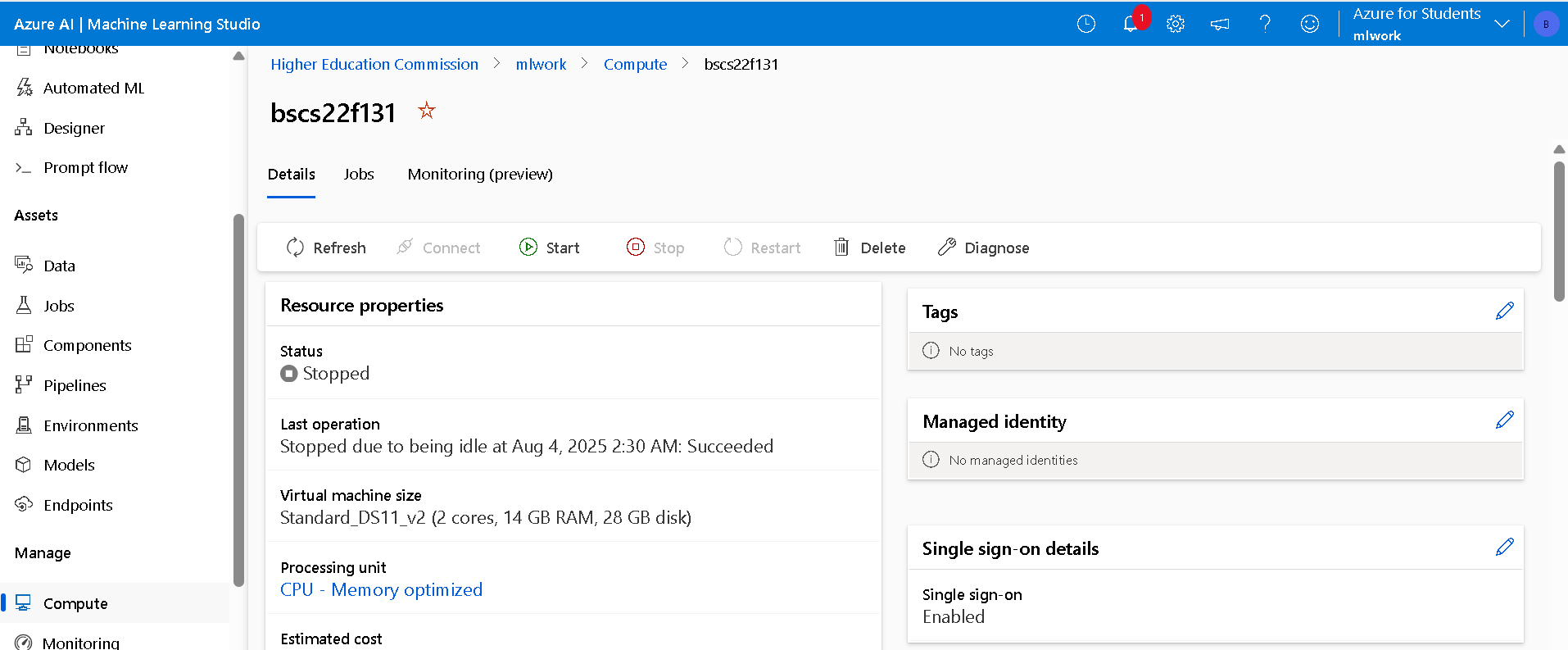
Name: penguin-cluster

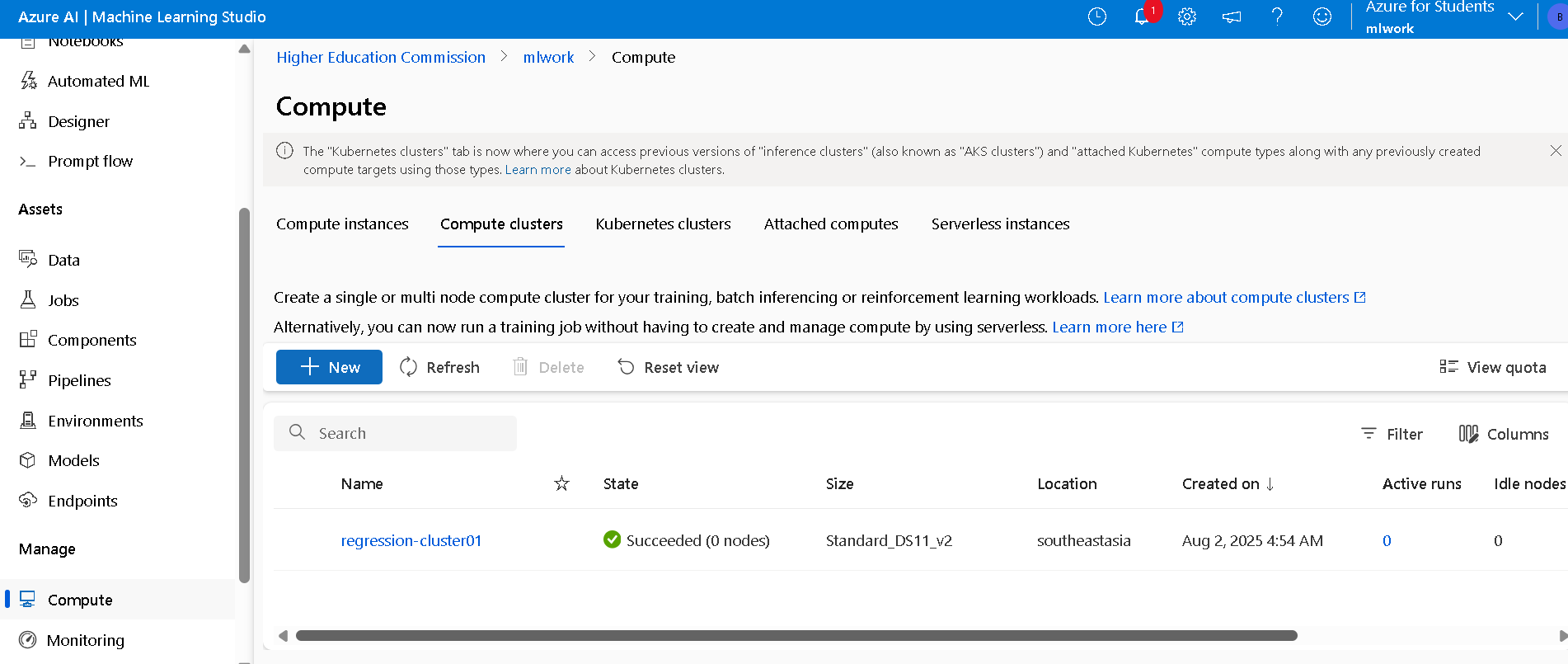
VM Type: Standard\_DS11\_v2

Minimum nodes: 0

Maximum nodes: 2

Purpose:Used for training, scoring, and deployment.  
**Screenshot: Created compute instance and compute cluster**





**Part 3: Explore and Prepare Data Using Azure ML Designer**

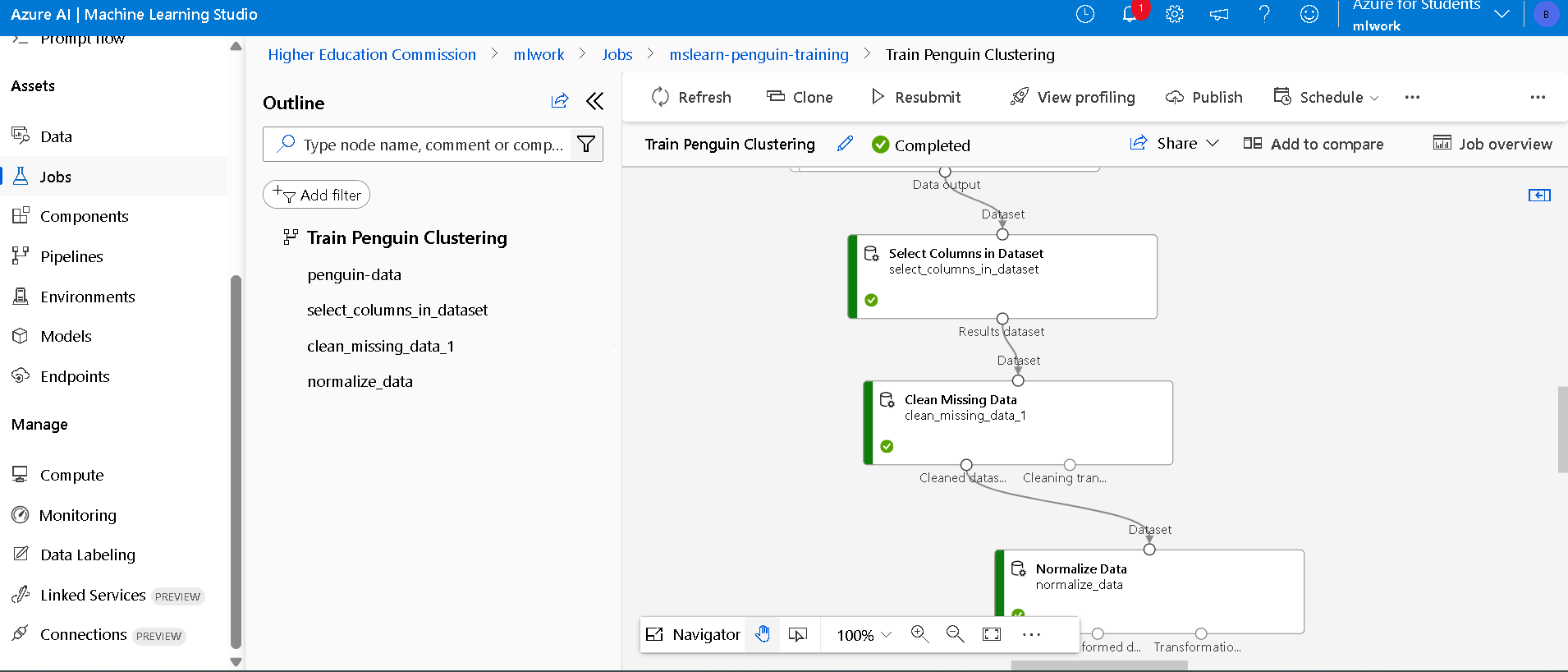
Imported the Palmer Penguins dataset using web access.

Applied preprocessing modules:

Select Columns in Dataset – Chose key numerical features such as Culmen Length, Culmen Depth, Flipper Length, Body Mass.

Clean Missing Data – Removed rows with null values.

Normalize Data – Applied MinMax normalization for consistency.  
**Screenshot: Data preprocessing pipeline**



**Part 4: Create and Run a Clustering Pipeline**

Designed the pipeline with the following modules:

Select Columns in Dataset

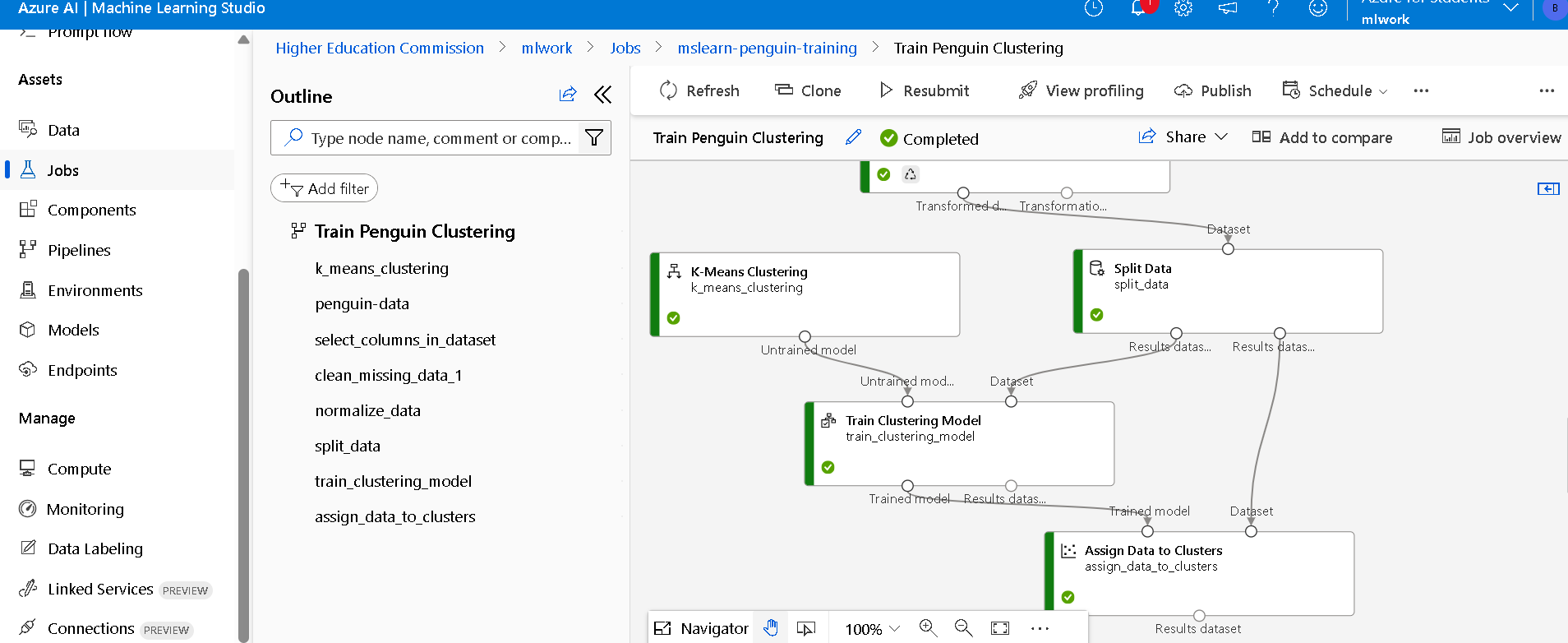
Clean Missing Data

Normalize Data

K-Means Clustering (Unsupervised) – Number of clusters set to 3.

Assign Data to Clusters – To assign each data point to a cluster.

Convert to CSV / Visualize Data – Optional for inspection.  
**Screenshot: Full clustering pipeline layout**

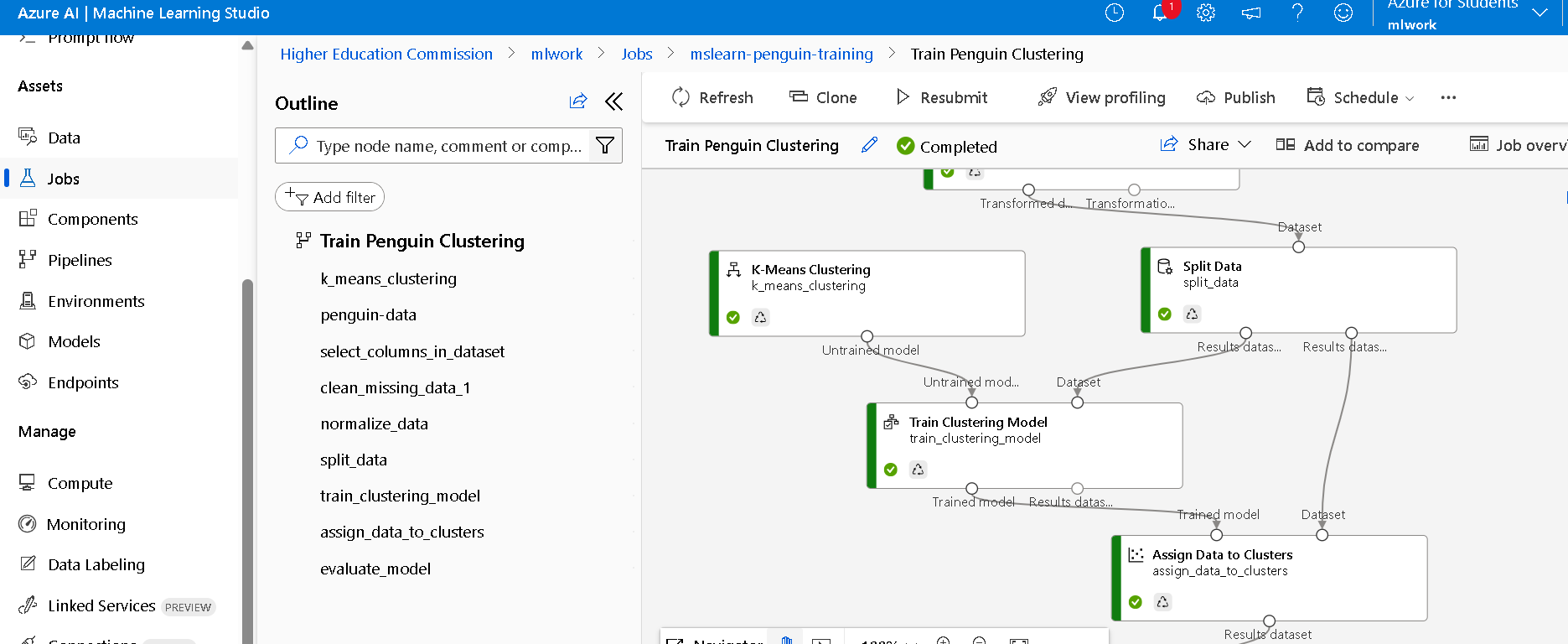


**Part 5: Evaluate Clustering Results**

Used Convert to Dataset + Visualize to explore the clustering assignments.

Verified cluster separation visually and checked feature distribution.

Observed that clusters closely align with actual penguin species groups.  
**Screenshot: Cluster assignment results with scatter plots**



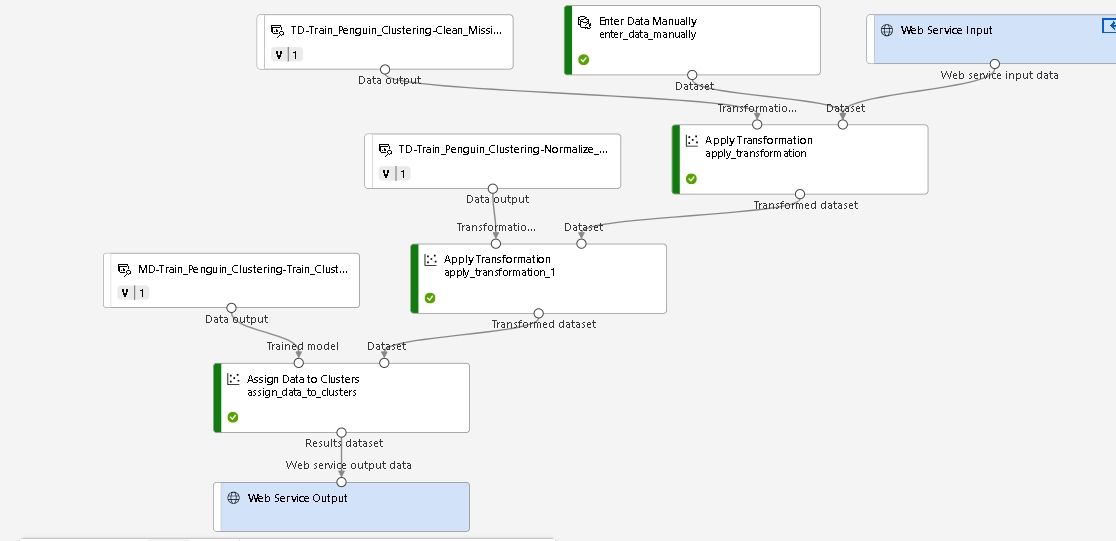
**Part 6: Create an Inference Pipeline**

Cloned the pipeline and made it suitable for real-time prediction:

Replaced data source with Enter Data Manually (excluding species column).

Removed unnecessary modules like evaluation visualizers.

**Screenshot: Inference pipeline**

****

**Part 7: Deploy a Predictive Service**

Deployed the inference pipeline as a real-time web service using Azure Container Instance (ACI).

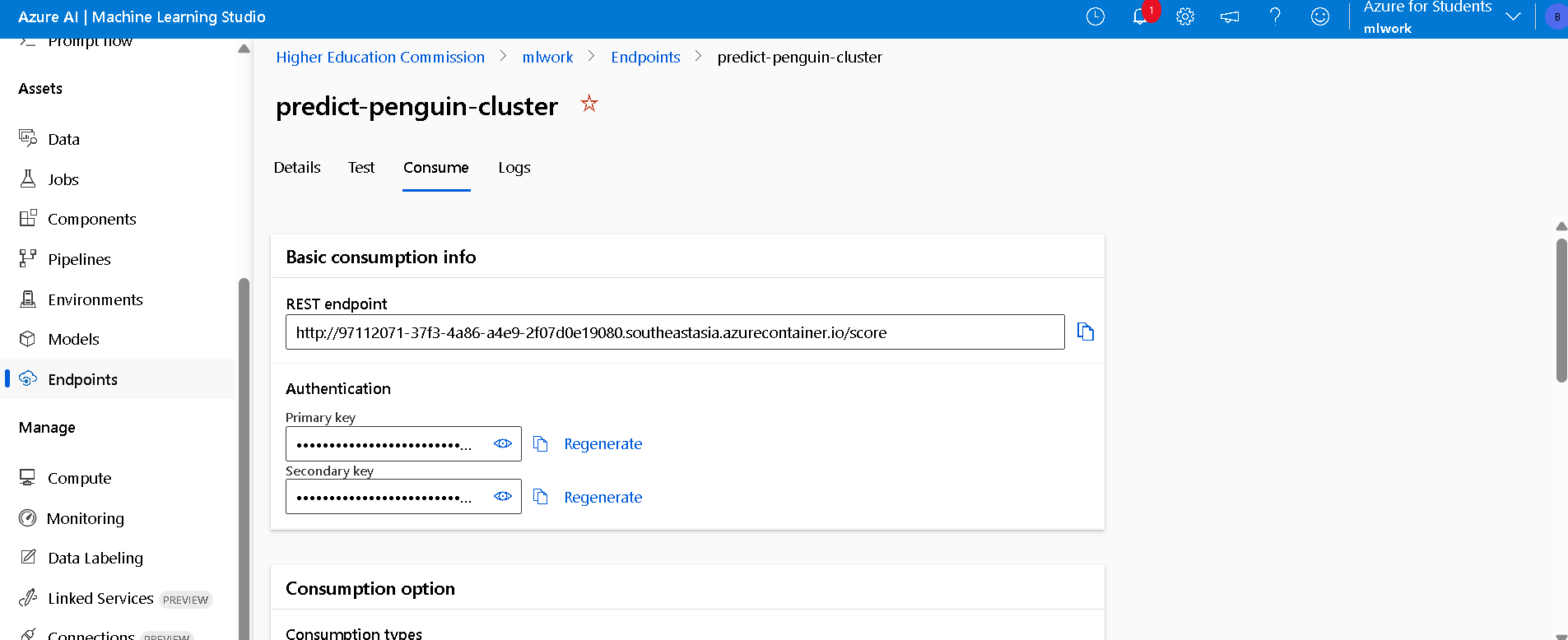
Faced access issue:  
"User does not have authorization to perform action 'Microsoft.ContainerInstance/containerGroups/write'"

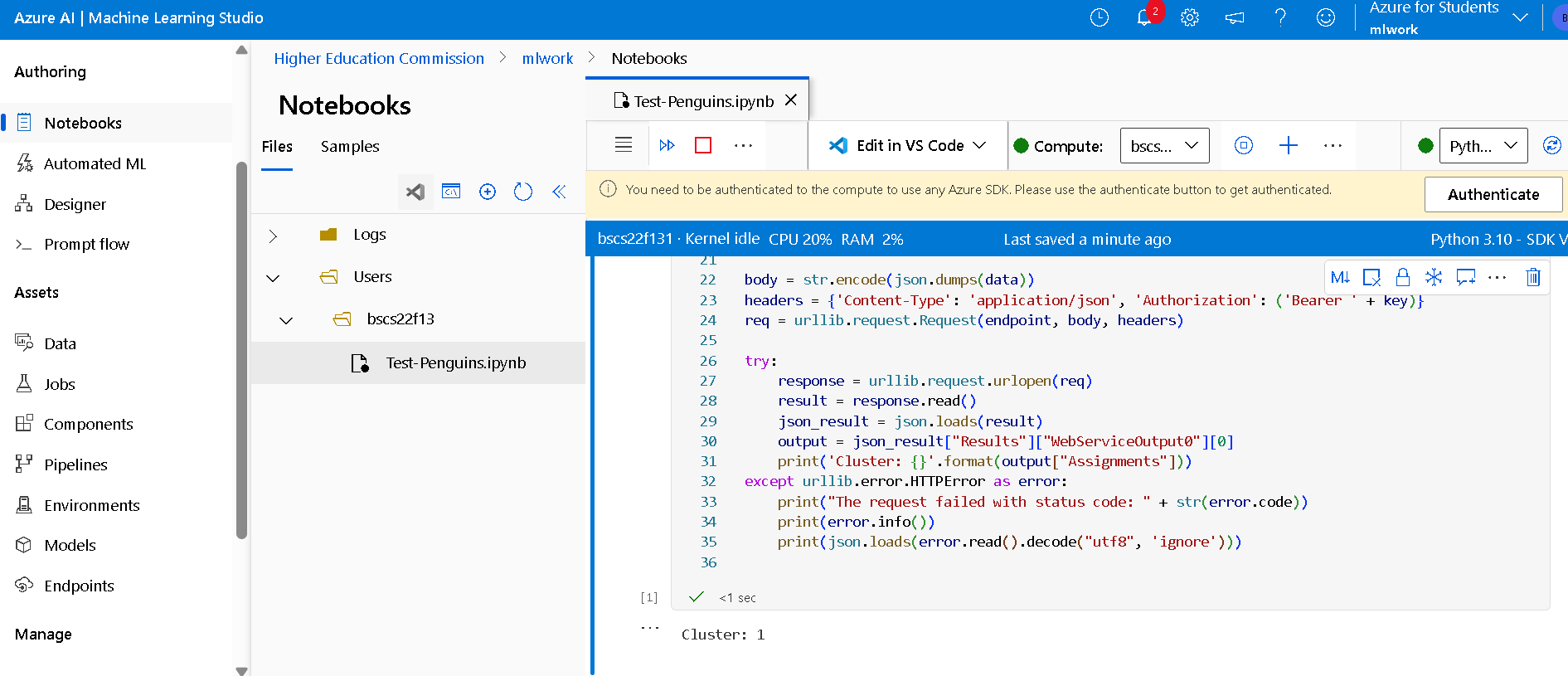
Resolution:

Requested Contributor access from Azure administrator

Alternatively deployed to penguin-cluster

Successfully tested deployment using REST endpoint.  
**Screenshot: Deployed web service endpoint**





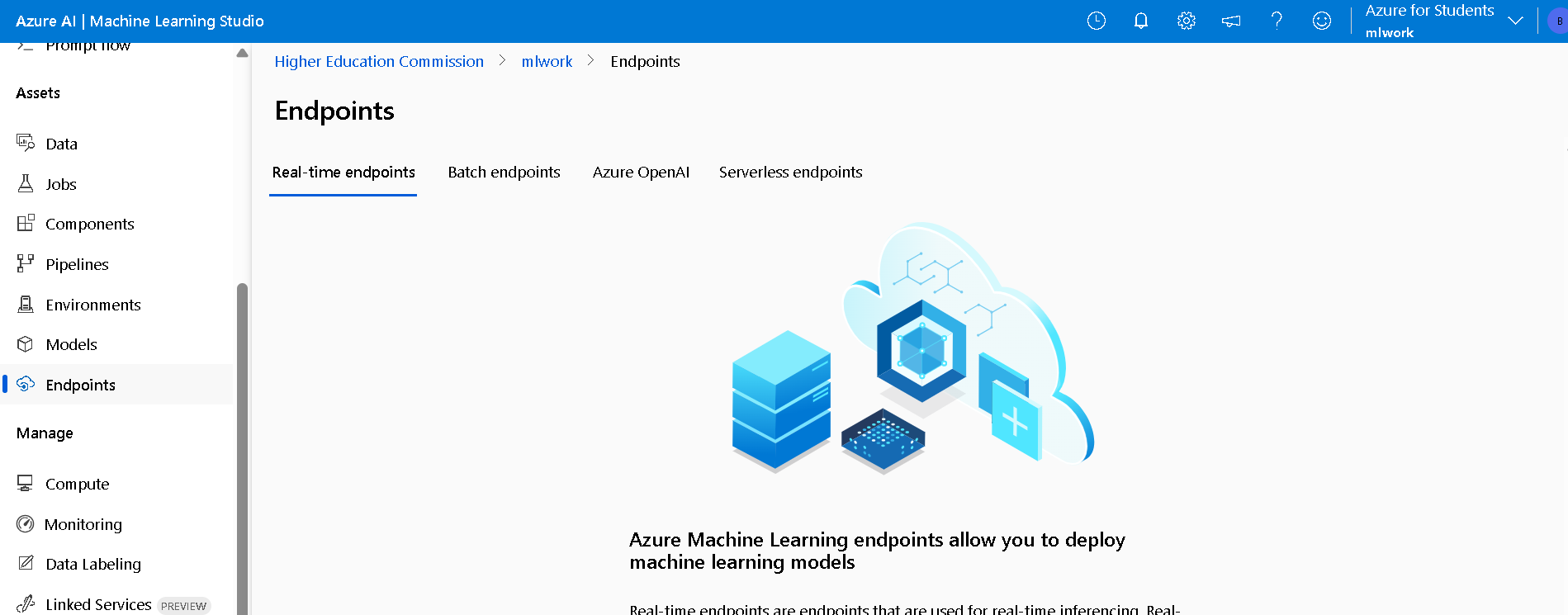
**Part 8: Clean-Up**

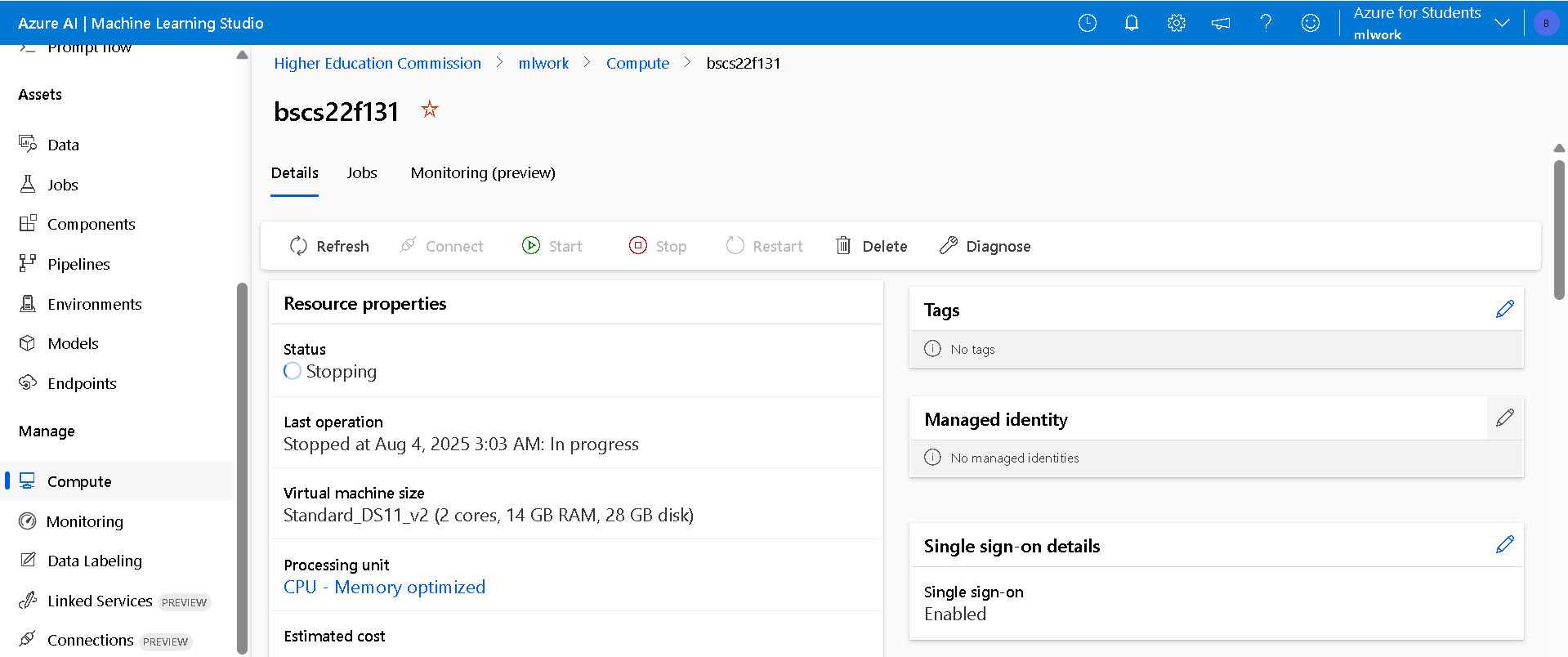
Deleted the following to avoid unnecessary charges:

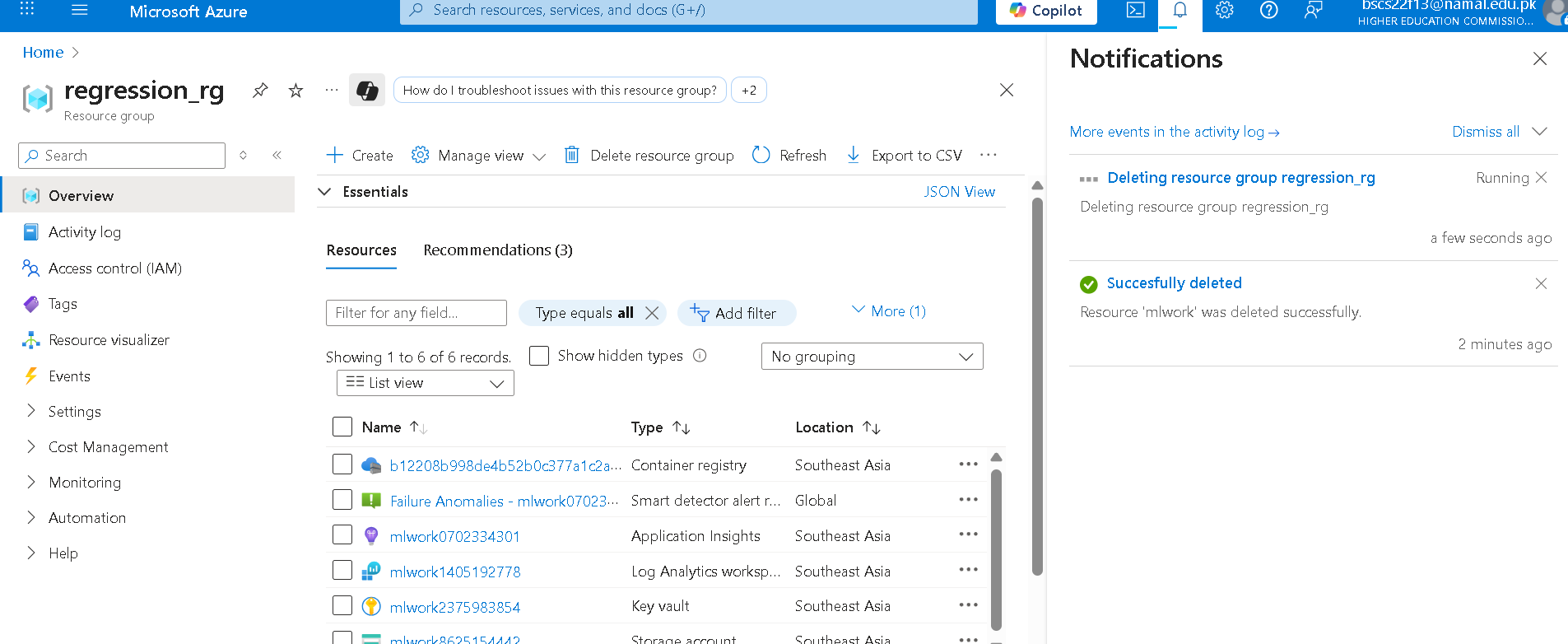
Real-time endpoint

Compute cluster

Azure ML Workspace (if not needed further)  
**Screenshot: Confirmation of deleted resources**







**Conclusion**

Successfully built, trained, and deployed a clustering model using the Palmer Penguins dataset in Azure ML.  
The project demonstrated how unsupervised learning can be used to group similar data points, and how the Azure ML Designer supports full model lifecycle development — from data preparation to real-time API deployment.

**Technologies Used**

Microsoft Azure ML Studio (Designer)

Azure Compute Clusters

Azure Container Instances (ACI)

Python (Execute Script Module)

K-Means Clustering (Azure ML built-in)