**Lab Overview**

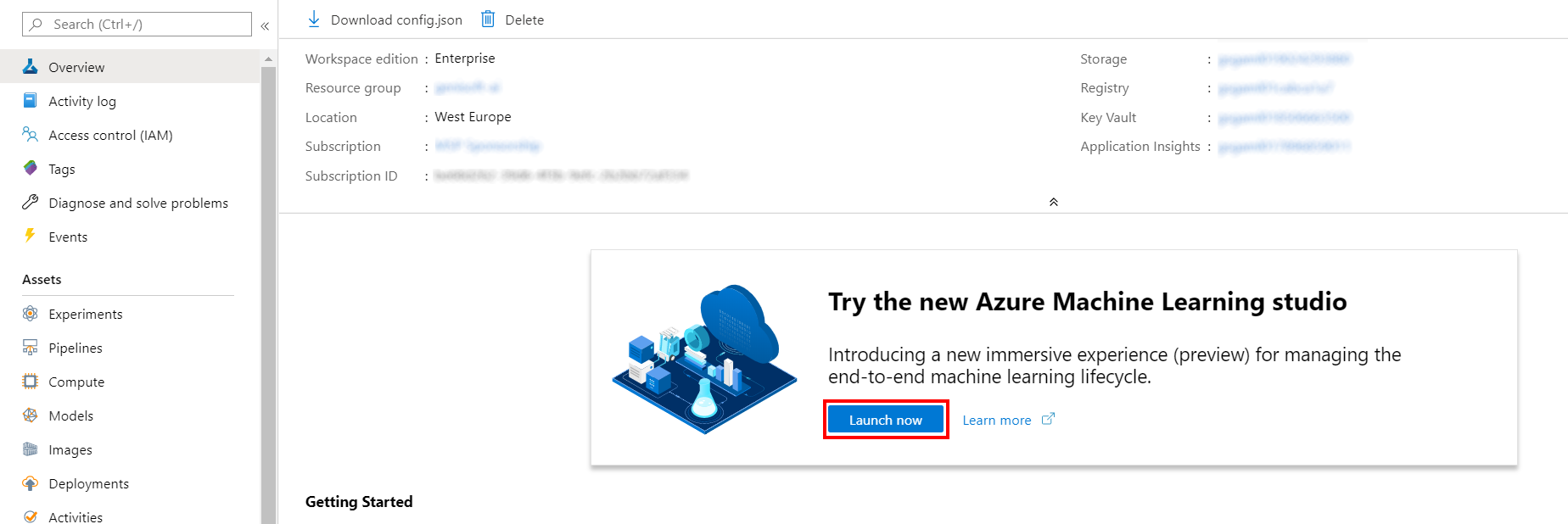
[Azure Machine Learning designer](https://docs.microsoft.com/en-us/azure/machine-learning/service/concept-designer) (preview) gives you a cloud-based interactive, visual workspace that you can use to easily and quickly prep data, train and deploy machine learning models. It supports Azure Machine Learning compute, GPU or CPU. Machine Learning designer also supports publishing models as web services on Azure Kubernetes Service that can easily be consumed by other applications.

In this lab, we will be using the Weather Dataset that has weather data for 66 different airports in the USA from April to October 2013. We will cluster the dataset into 5 distinct clusters based on key weather metrics, such as visibility, temperature, dew point, wind speed etc. The goal is to group airports with similar weather conditions. We will do all of this from the Azure Machine Learning designer without writing a single line of code.

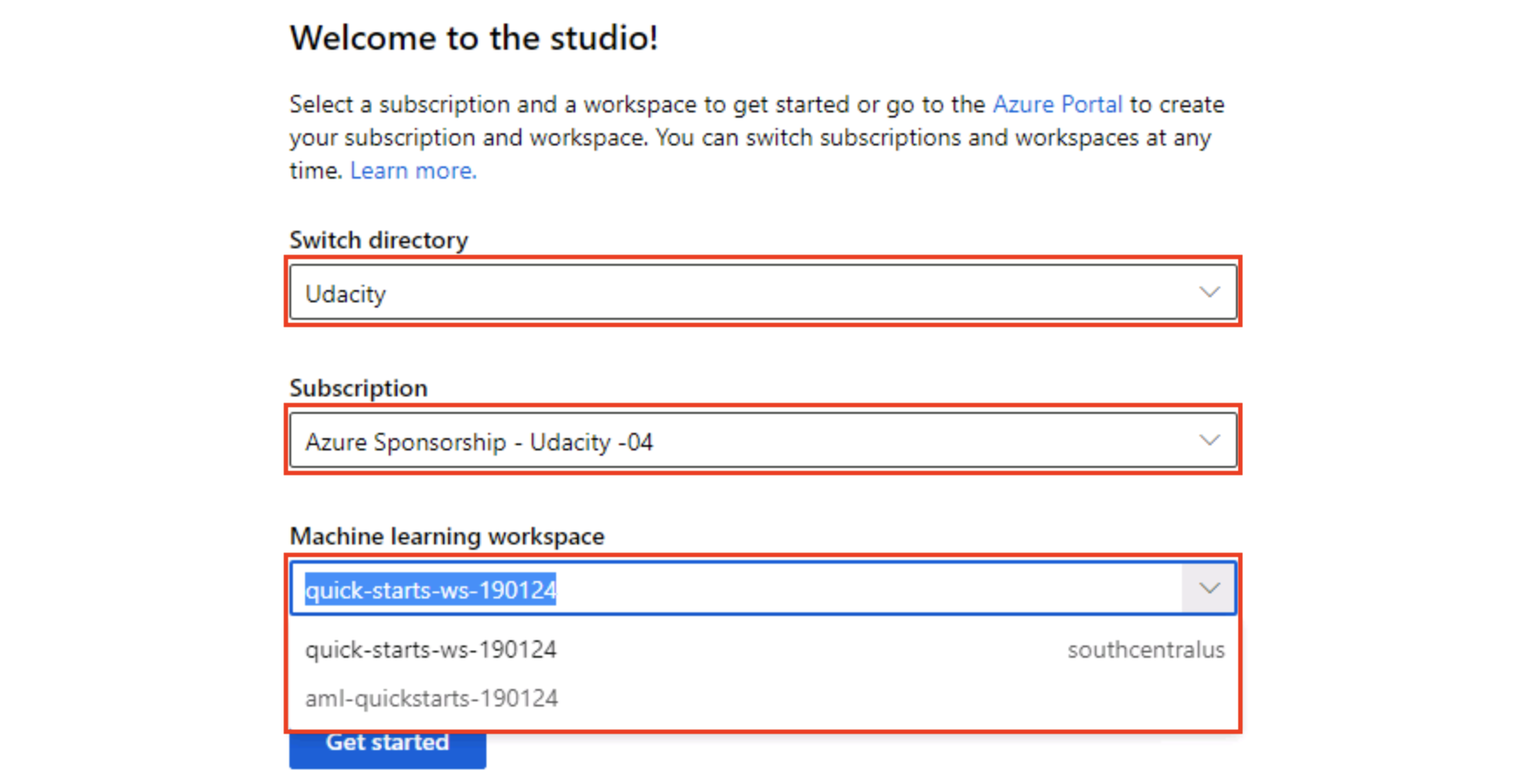
**Exercise 1: Create New Training Pipeline**

Task 1: Open Pipeline Authoring Editor

1. In [Azure portal](https://portal.azure.com/), open the available machine learning workspace.
2. Select **Launch now** under the **Try the new Azure Machine Learning studio** message.

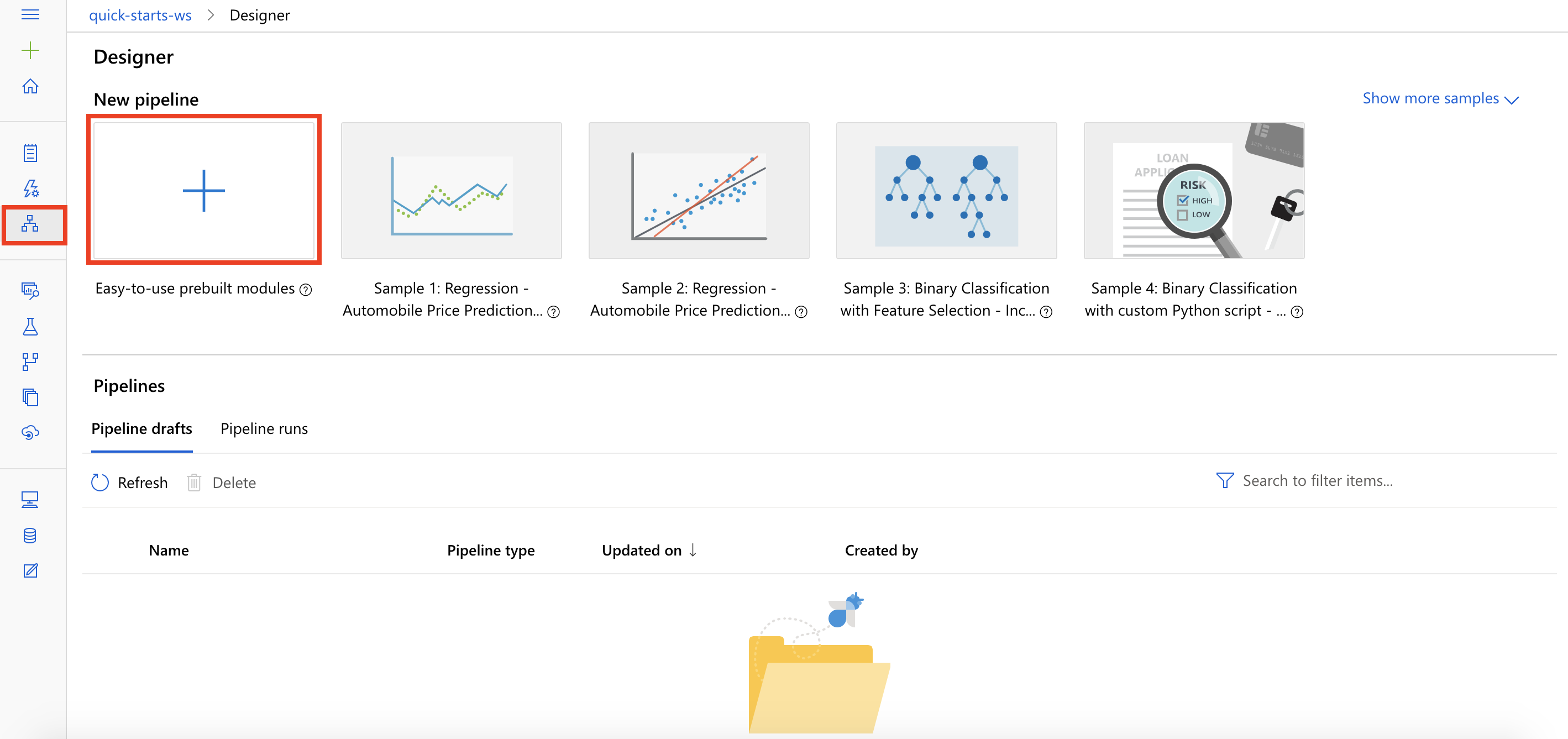


1. When you first launch the studio, you may need to set the directory and subscription. If so, you will see this screen:



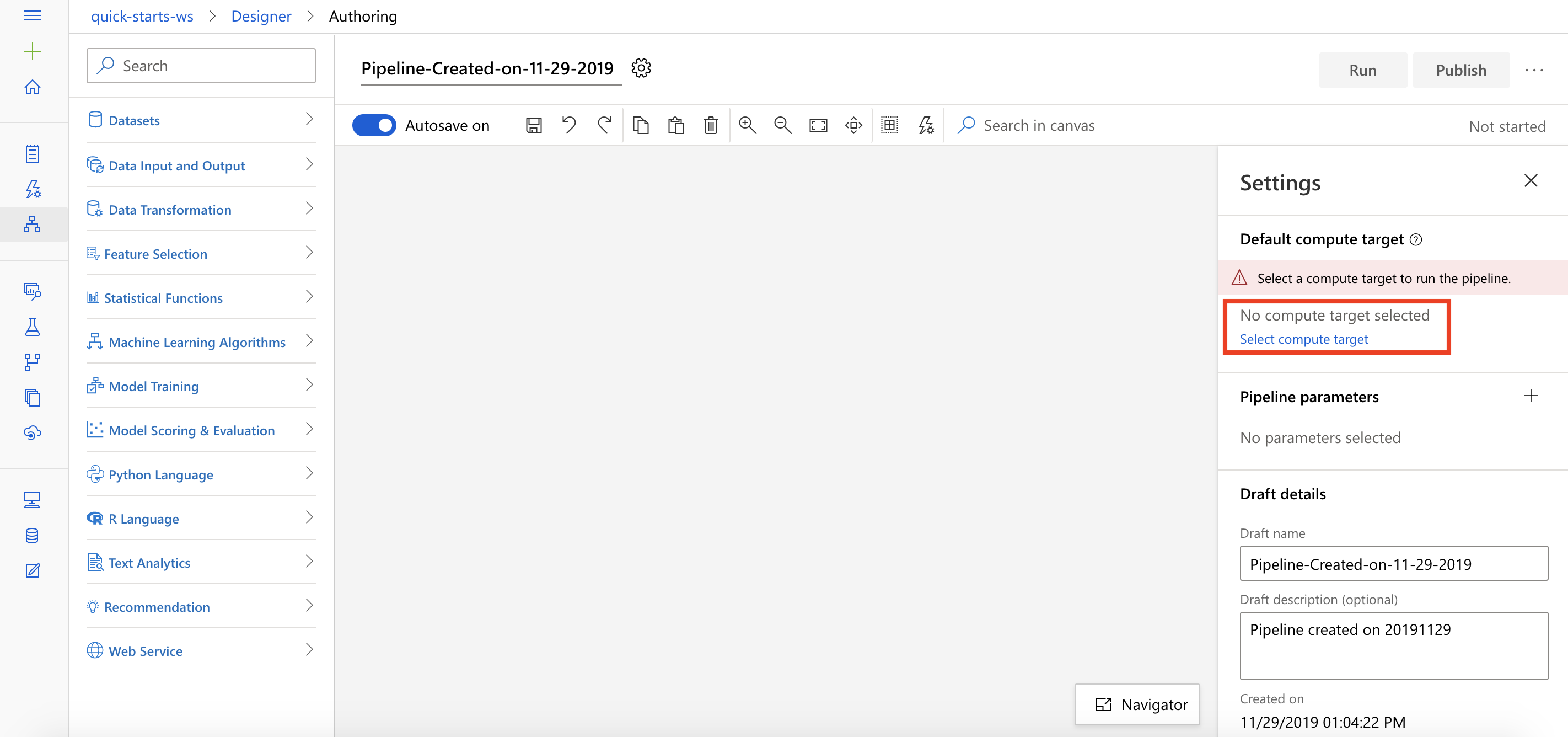
For the directory, select **Udacity** and for the subscription, select **Azure Sponsorship**. For the machine learning workspace, you may see multiple options listed. **Select any of these** (it doesn’t matter which) and then click **Get started**.

1. From the studio, select **Designer, +**. This will open a visual pipeline authoring editor.



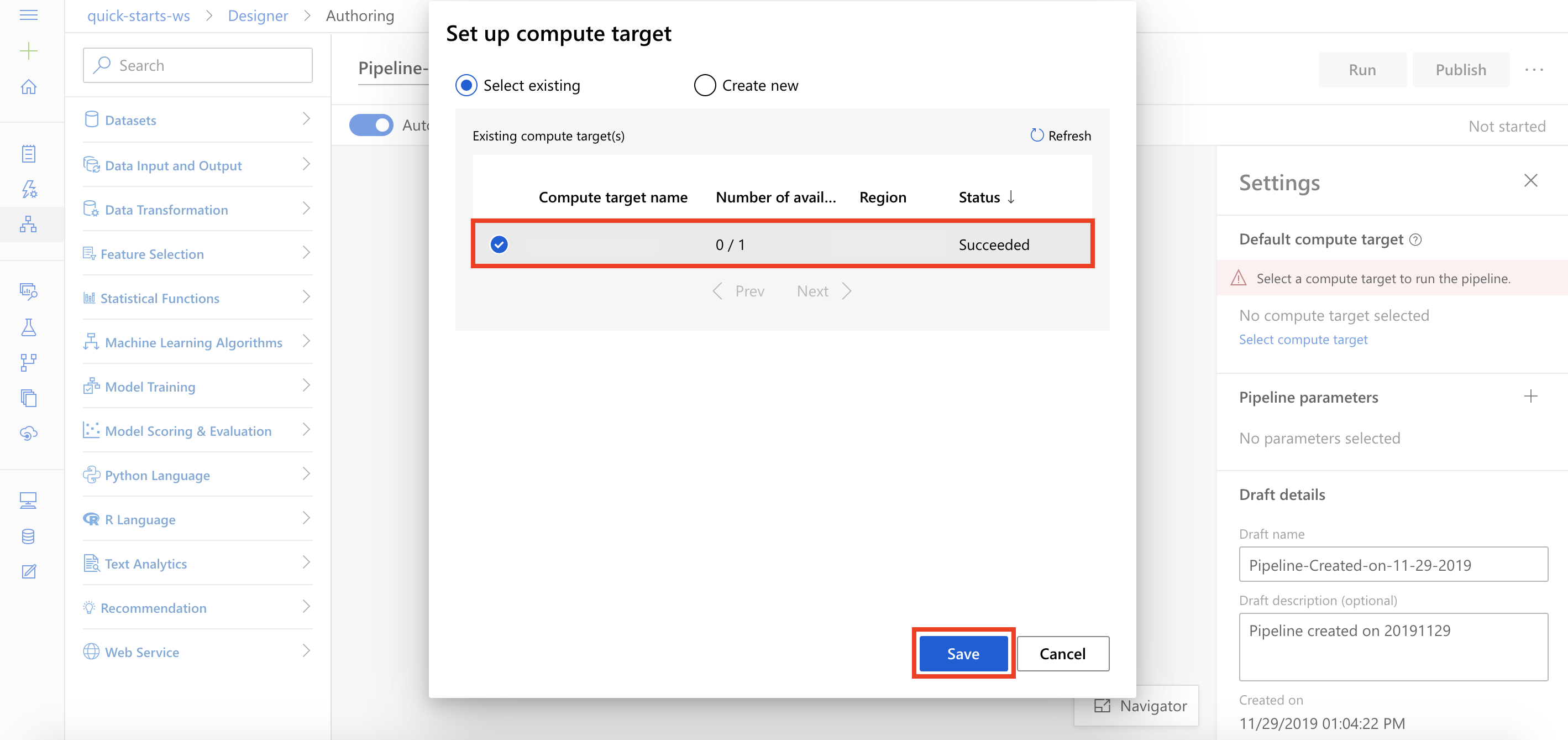
Task 2: Setup Compute Target

1. In the settings panel on the right, select **Select compute target**.



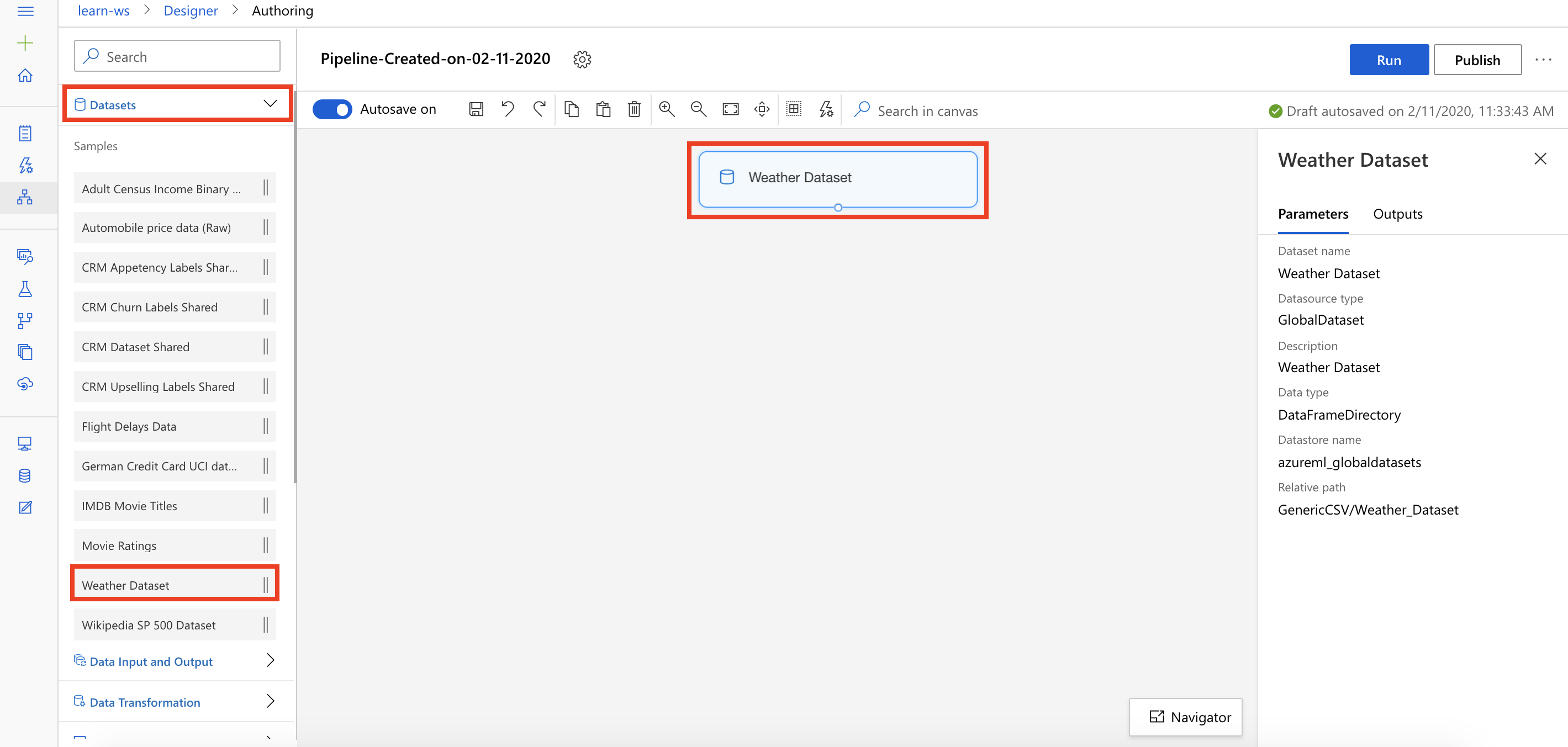
1. In the Set up compute target editor, select the available compute, and then select **Save**.

Note: If you are facing difficulties in accessing pop-up windows or buttons in the user interface, please refer to the Help section in the lab environment.



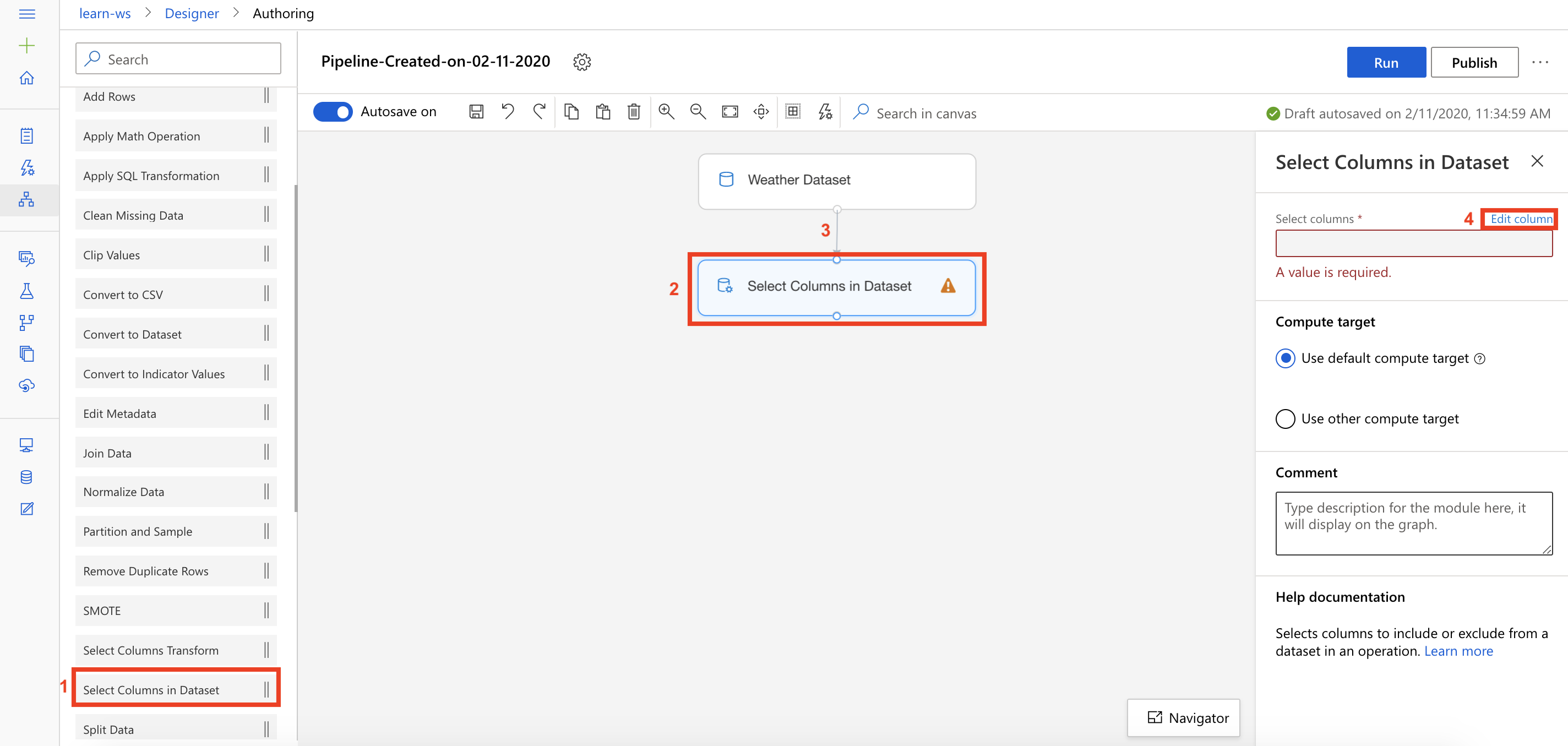
Task 3: Add Dataset

1. Select **Datasets** section in the left navigation. Next, select **Samples, Weather Dataset** and drag and drop the selected dataset on to the canvas.

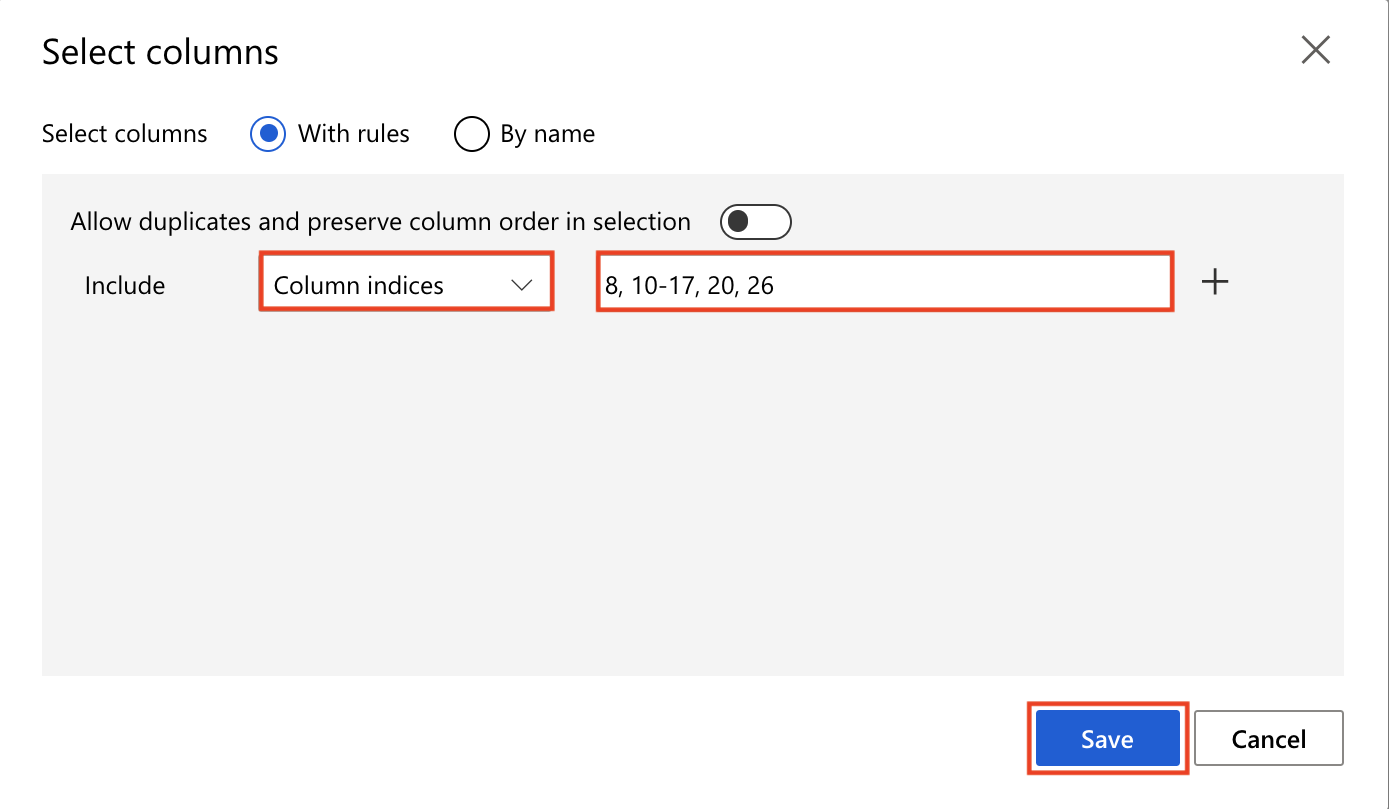


Task 4: Select Columns in Dataset

1. Select **Data Transformation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Select Columns in Dataset** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the Weather Dataset module to the Select Columns in Dataset module
   4. Select **Edit column** link to open the Select columns editor

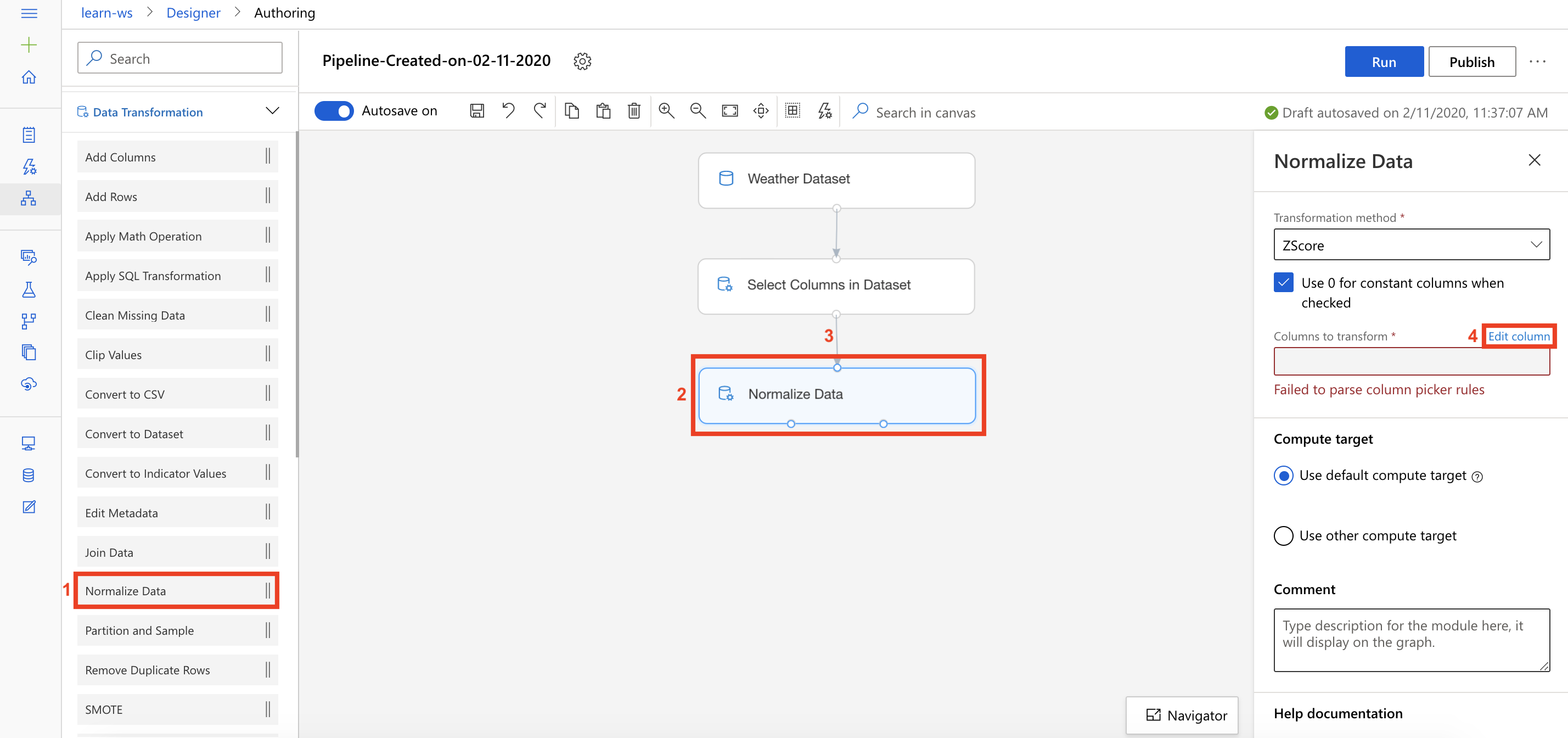


1. *Note that you can submit the pipeline at any point to peek at the outputs and activities. Running pipeline also generates metadata that is available for downstream activities such selecting column names from a list in selection dialogs.*
2. In the Select columns editor, follow the steps outlined below:
   1. Include: **Column indices**
   2. Provide column indices: **8, 10-17, 20, 26**
   3. Select **Save**

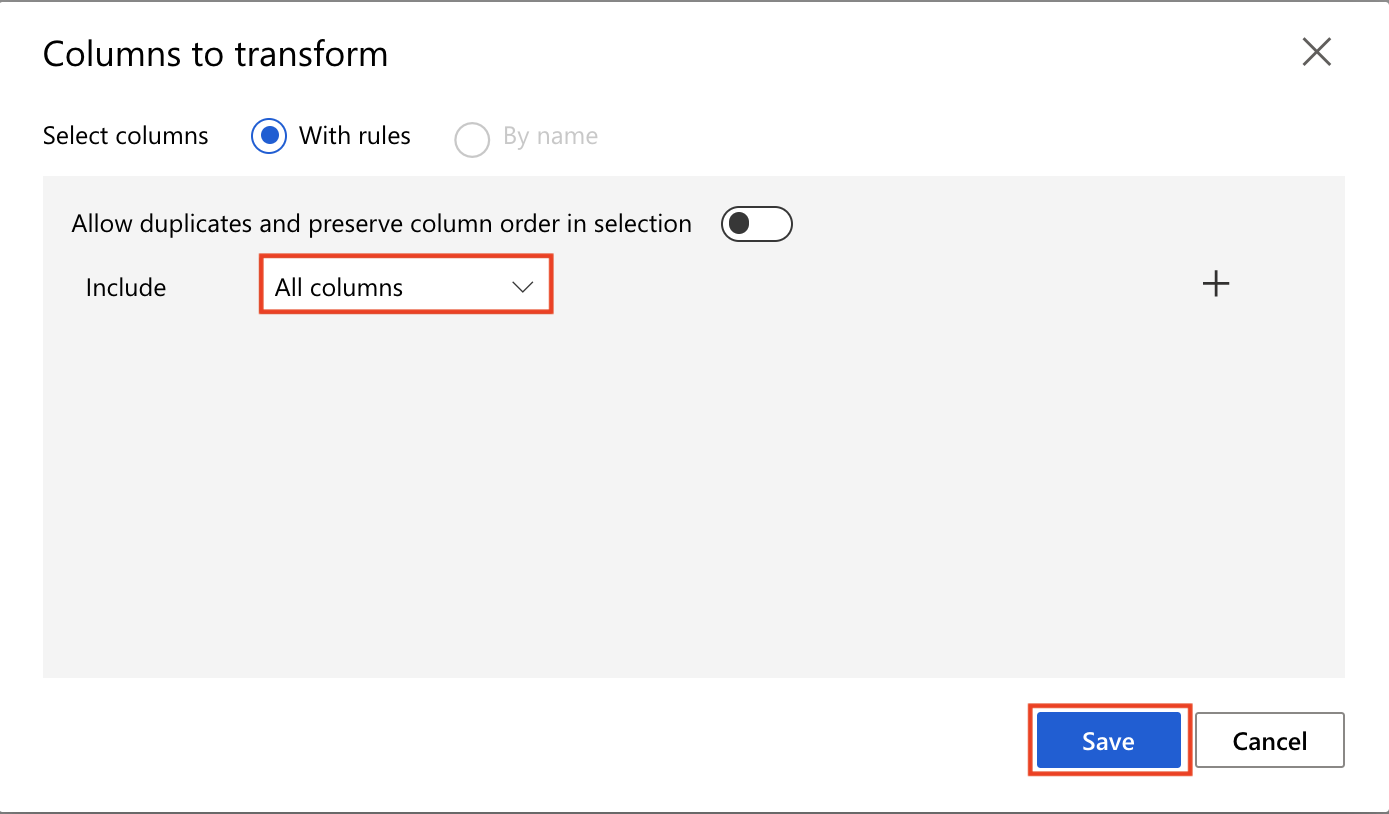


Task 5: Normalize Data

1. Select **Data Transformation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Normalize Data** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the Select Columns in Dataset module to the Normalize Data module
   4. Select **Edit column** link to open the Columns to transform editor

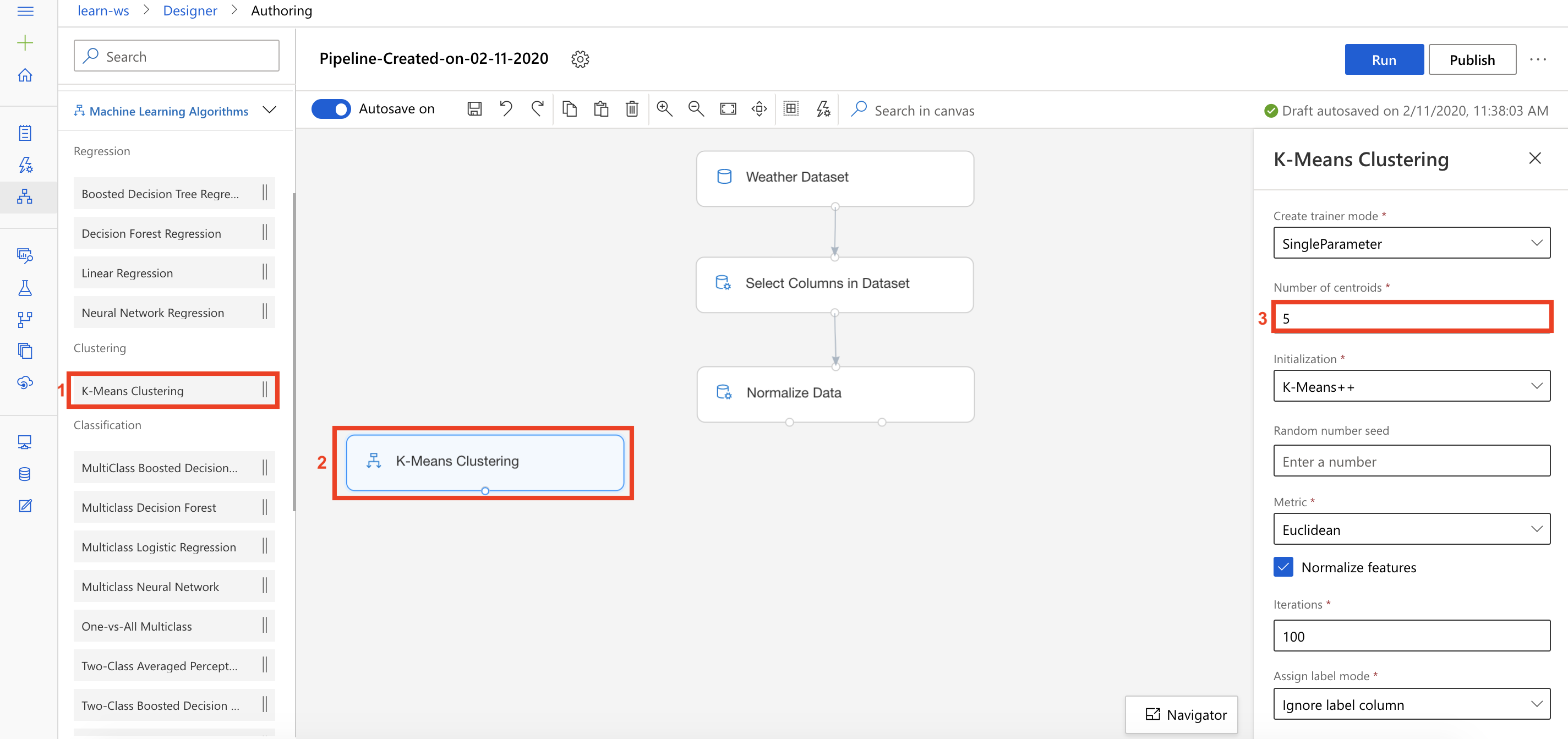


1. In the Columns to transform editor, follow the steps outlined below:
   1. Include: **All columns**
   2. Select **Save**



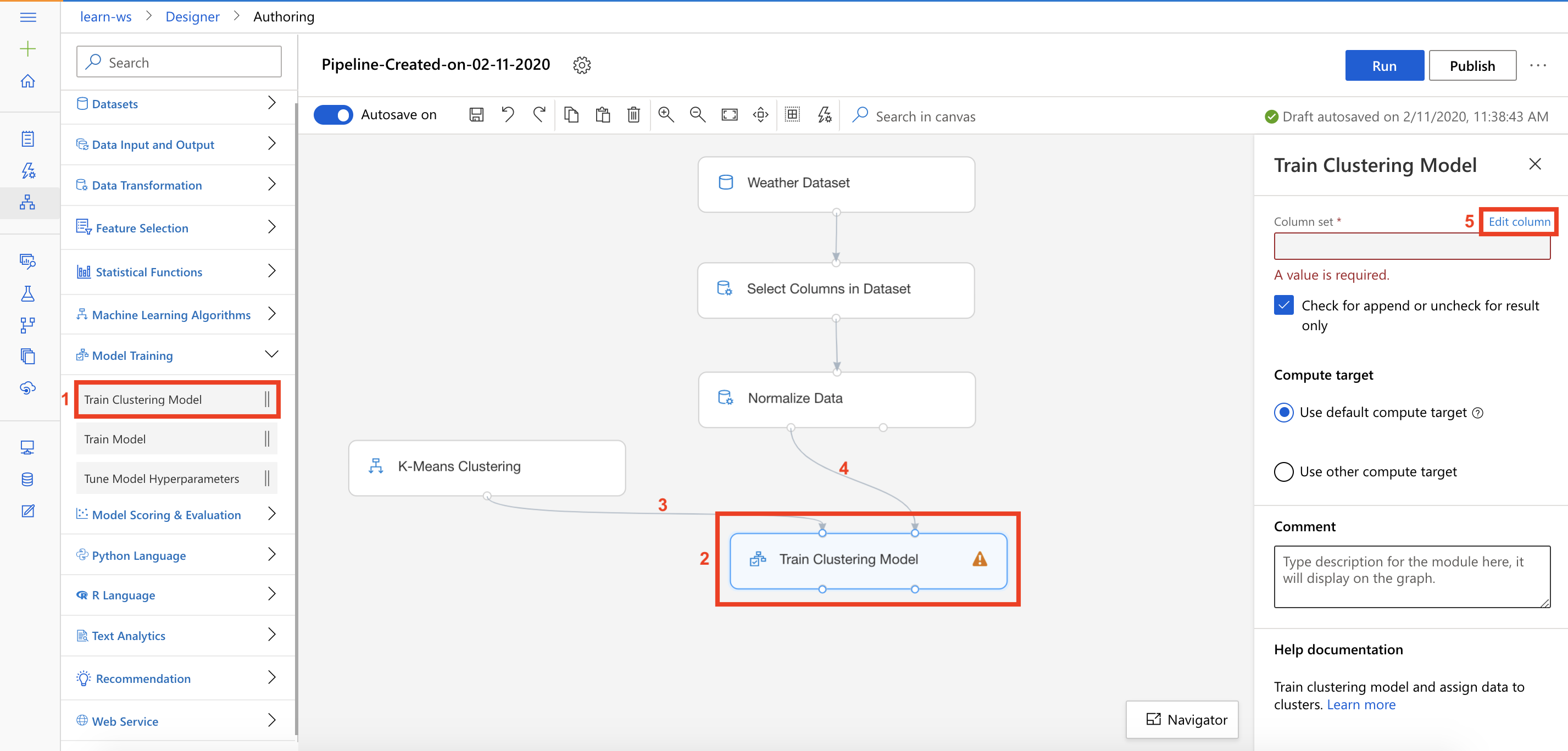
Task 6: Initialize K-Means Clustering Model

1. Select **Machine Learning Algorithms** section in the left navigation. Follow the steps outlined below:
   1. Select the **K-Means Clustering** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Number of centroids: **5**

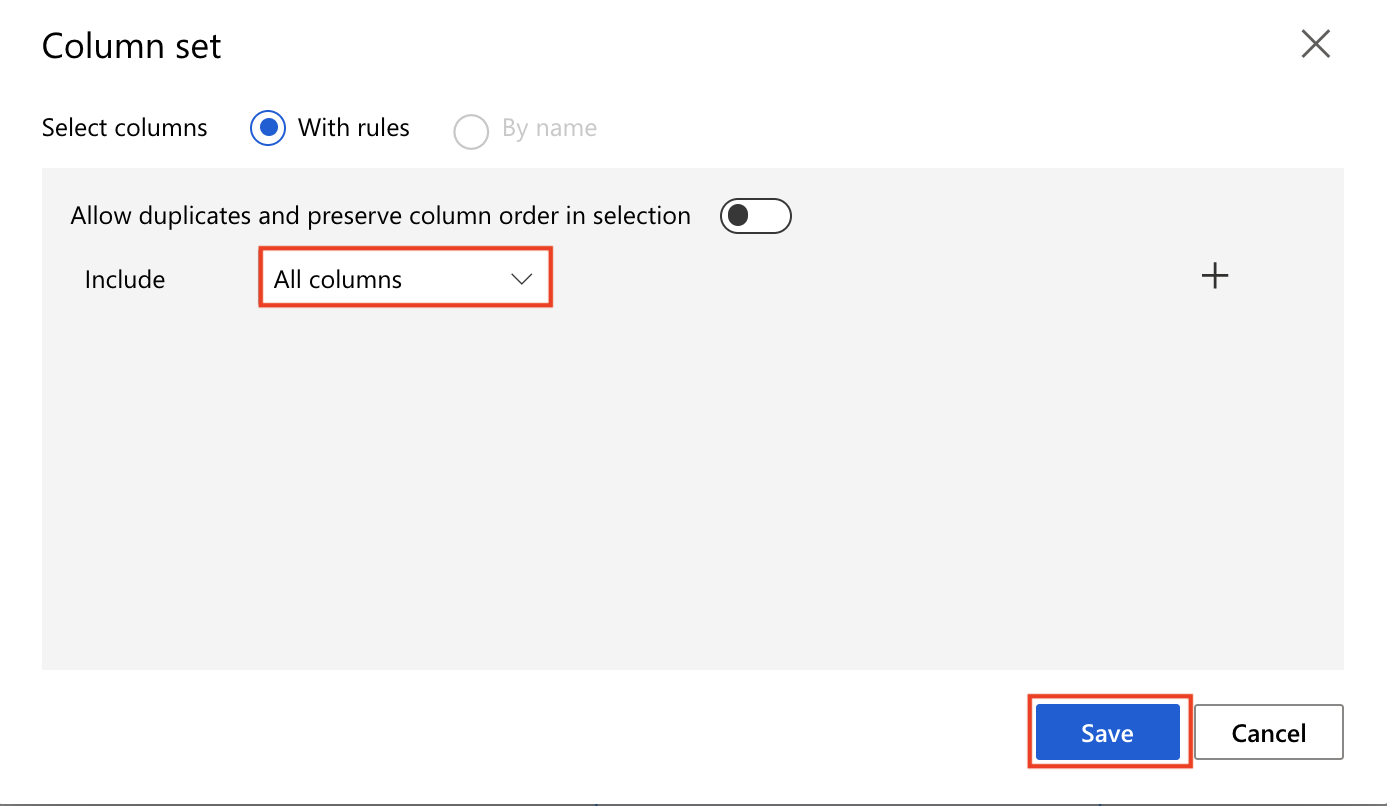


Task 7: Setup Train Clustering Model Module

1. Select **Model Training** section in the left navigation. Follow the steps outlined below:
   1. Select the **Train Clustering Model** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the K-Means Clustering module to the first input of the Train Clustering Model module
   4. Connect the first output of the Normalize module to the second input of the Train Clustering Model module
   5. Select the **Edit column** link to open the Column set editor

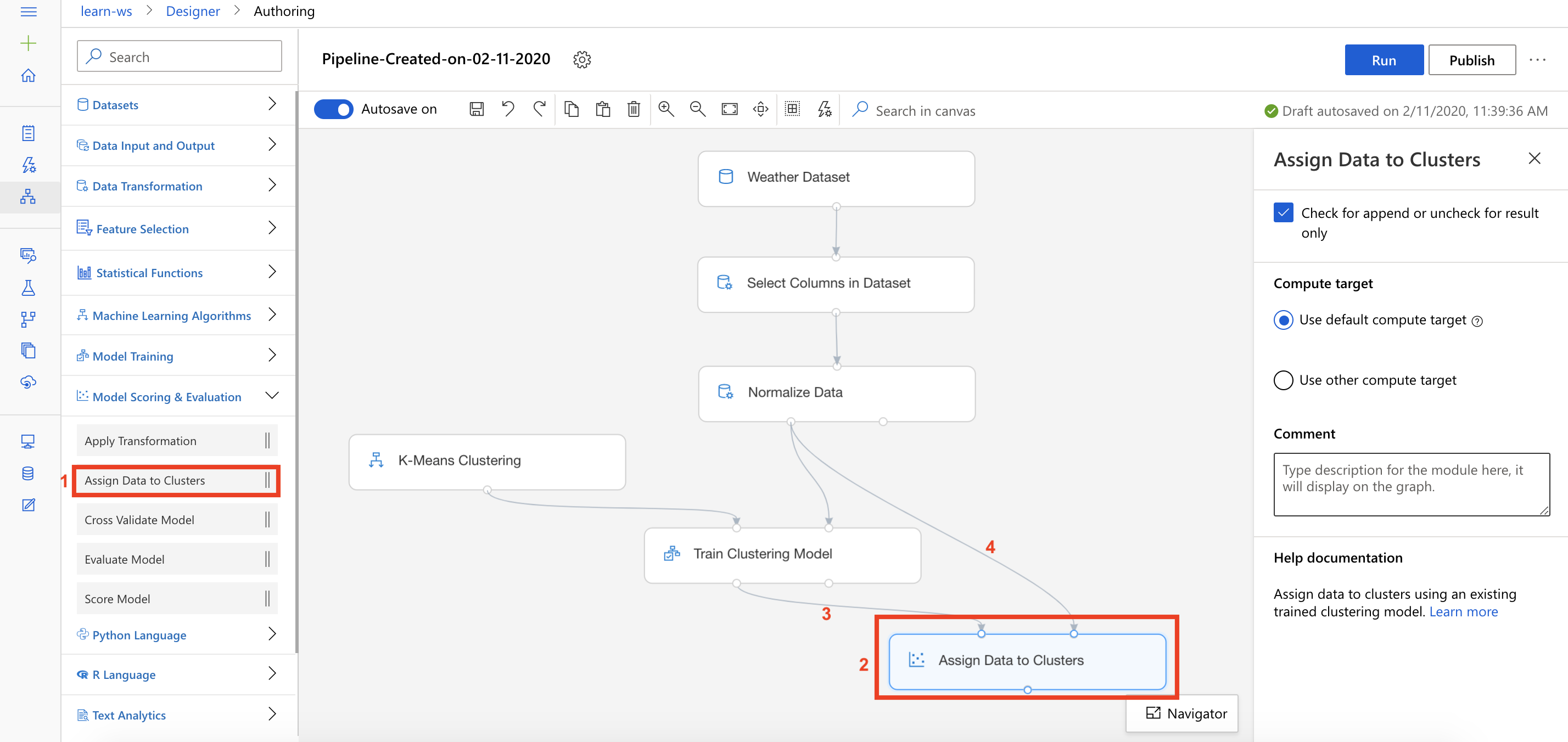


1. In the Columns set editor, follow the steps outlined below:
   1. Include: **All columns**
   2. Select **Save**



Task 8: Setup Assign Data to Clusters Module

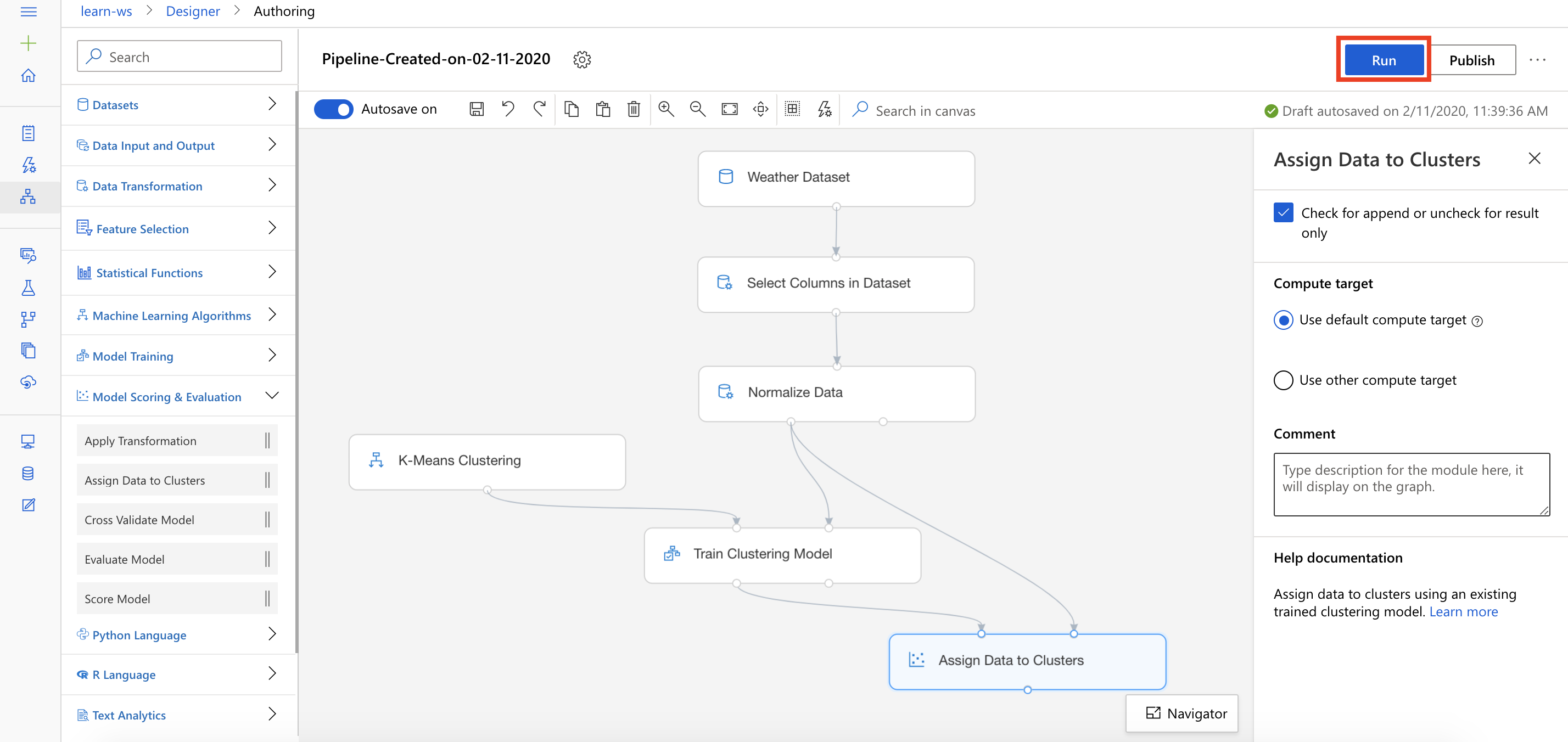
1. Select **Model Scoring & Evaluation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Assign Data to Clusters** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the first output of the Train Clustering Model module to the first input of the Assign Data to Clusters module
   4. Connect the first output of the Normalize Data module to the second input of the Assign Data to Clusters module



**Exercise 2: Submit Training Pipeline**

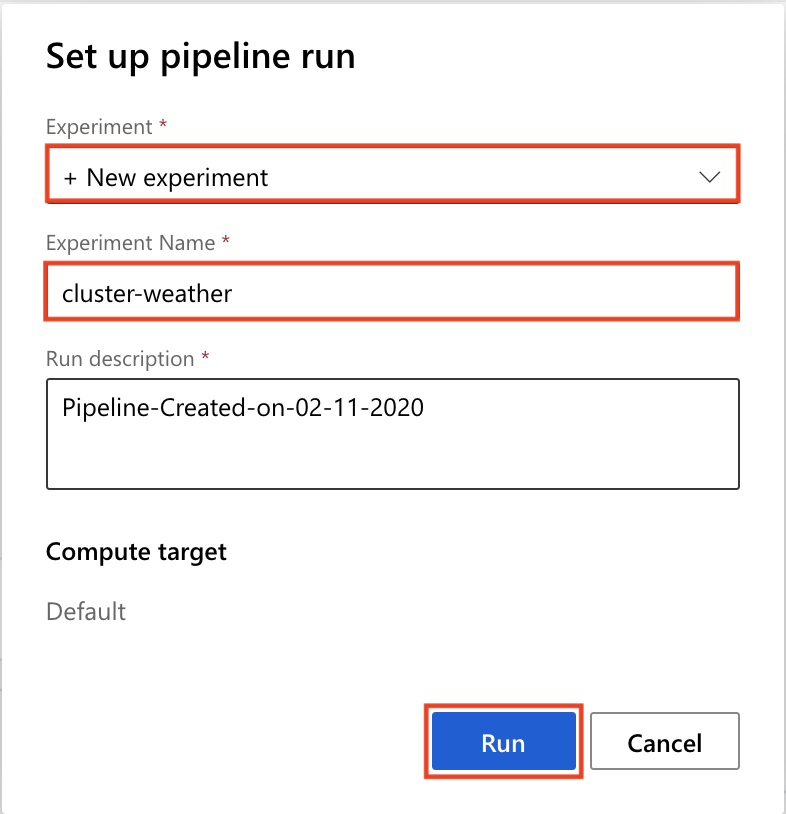
Task 1: Create Experiment and Submit Pipeline

1. Select **Submit** to open the Setup pipeline run editor.



Please note that the button name in the UI is changed from **Run** to **Submit**.

1. In the Setup pipeline run editor, select **Experiment, Create new** and provide New experiment name: **cluster-weather**, and then select **Submit**.

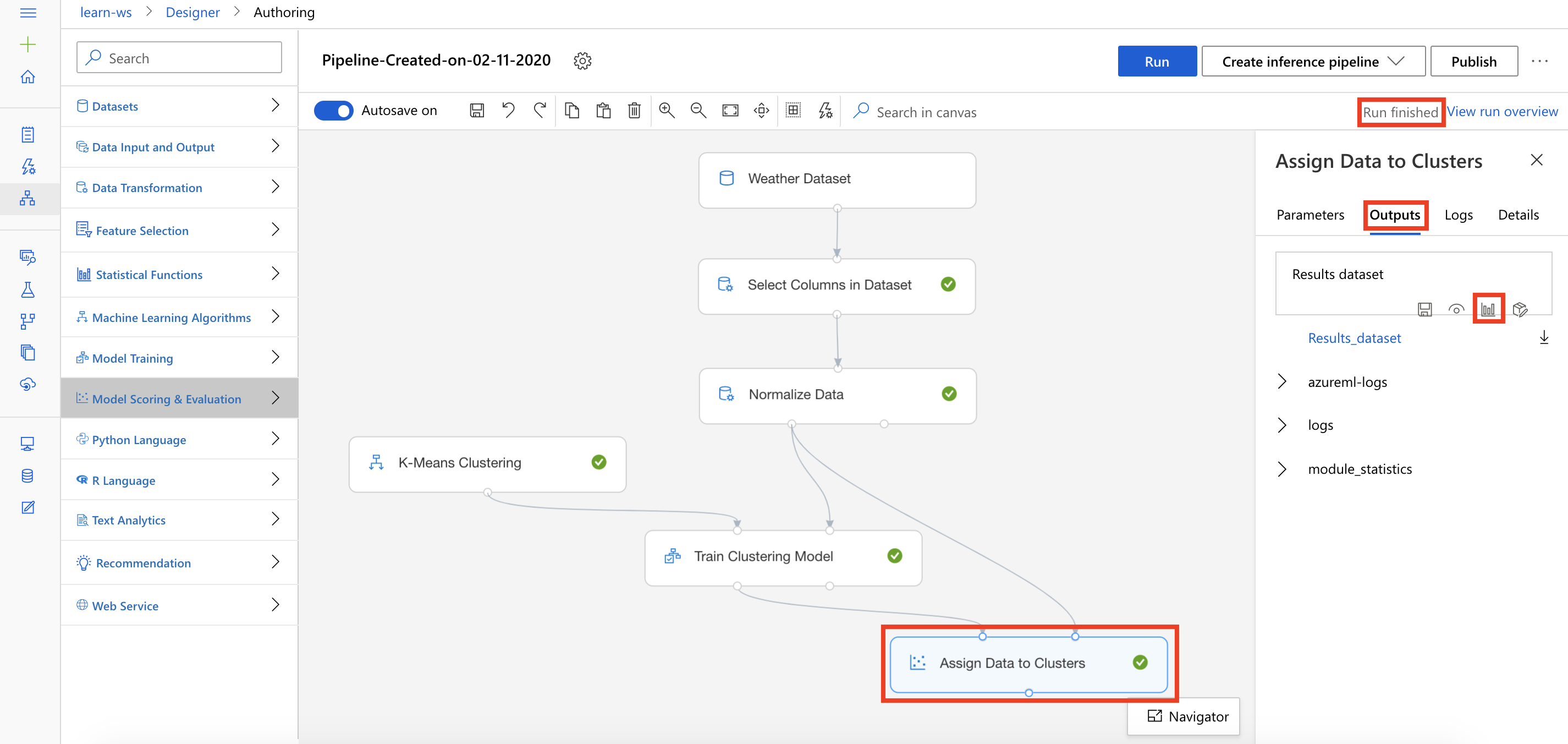


1. Wait for pipeline run to complete. It will take around **5 minutes** to complete the run.
2. While you wait for the model training to complete, you can learn more about the K-Means Clustering algorithm used in this lab by selecting [K-Means Clustering](https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/k-means-clustering#configure-the-k-means-clustering-module).

**Exercise 3: Visualize the Clustering Results**

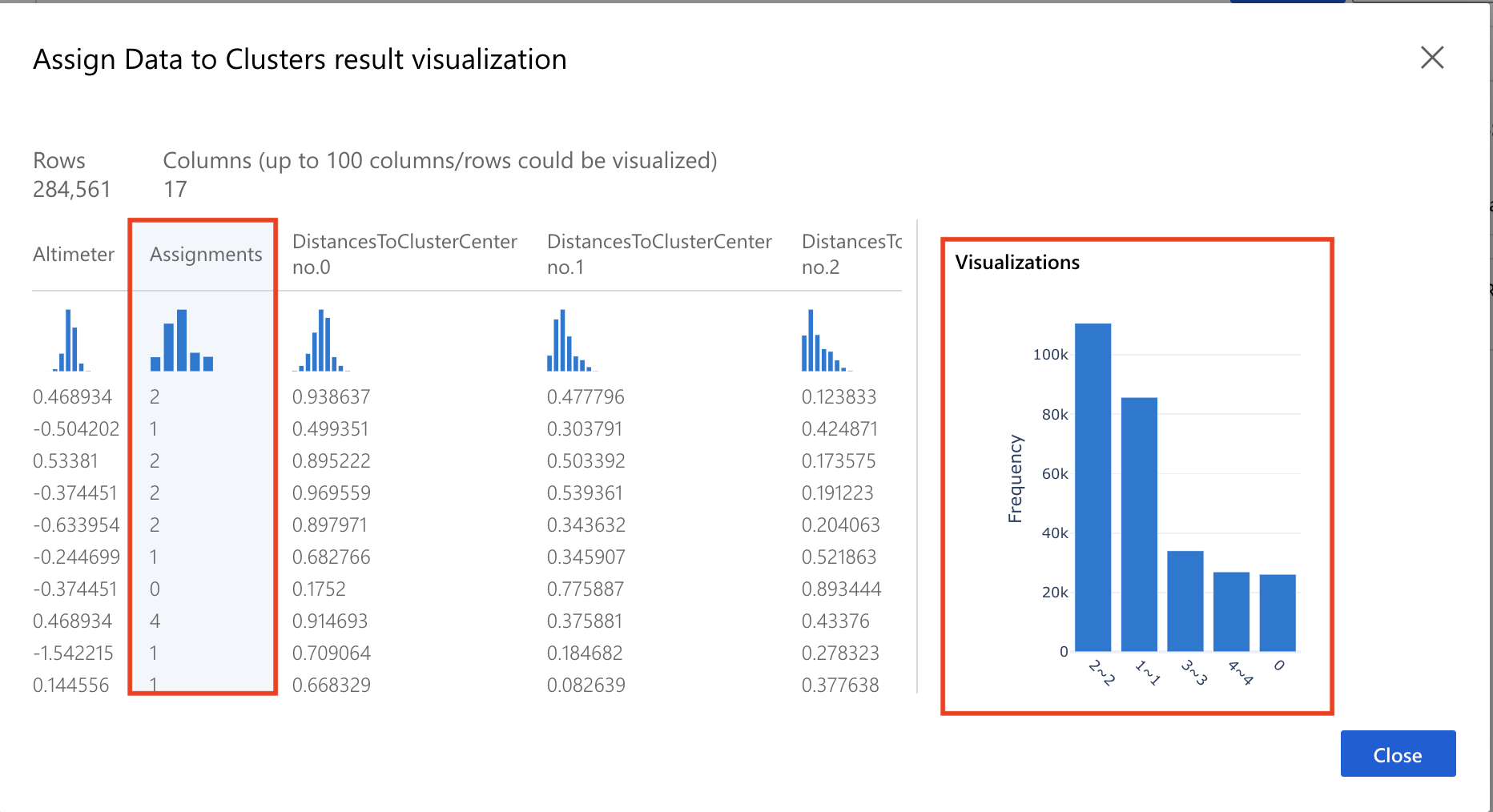
Task 1: Open the Visualization Dialog

1. Select **Assign Data to Clusters, Outputs, Visualize** to open the Assign Data to Clusters result visualization dialog.



Task 2: Evaluate Clustering Results

1. Scroll to the right and select **Assignments** column.
2. In the right-hand-side pane, scroll down to the **Visualizations** section.



1. From the results you can observe that each row (input) in the dataset is assigned to one of the 5 clusters: 0, 1, 2, 3, or 4. You can also see for each input, how far that input was from the various centroids. The cluster assignment is made based on the shortest distance between the input and cluster centroids. From the bar graph you can see the frequency distribution of all the inputs across the 5 clusters.

**Next Steps**

Congratulations! You have trained and evaluated your first clustering algorithm. You can continue to experiment in the environment but are free to close the lab environment tab and return to the Udacity portal to continue with the lesson.