Tutorial 5: Regression Analysis

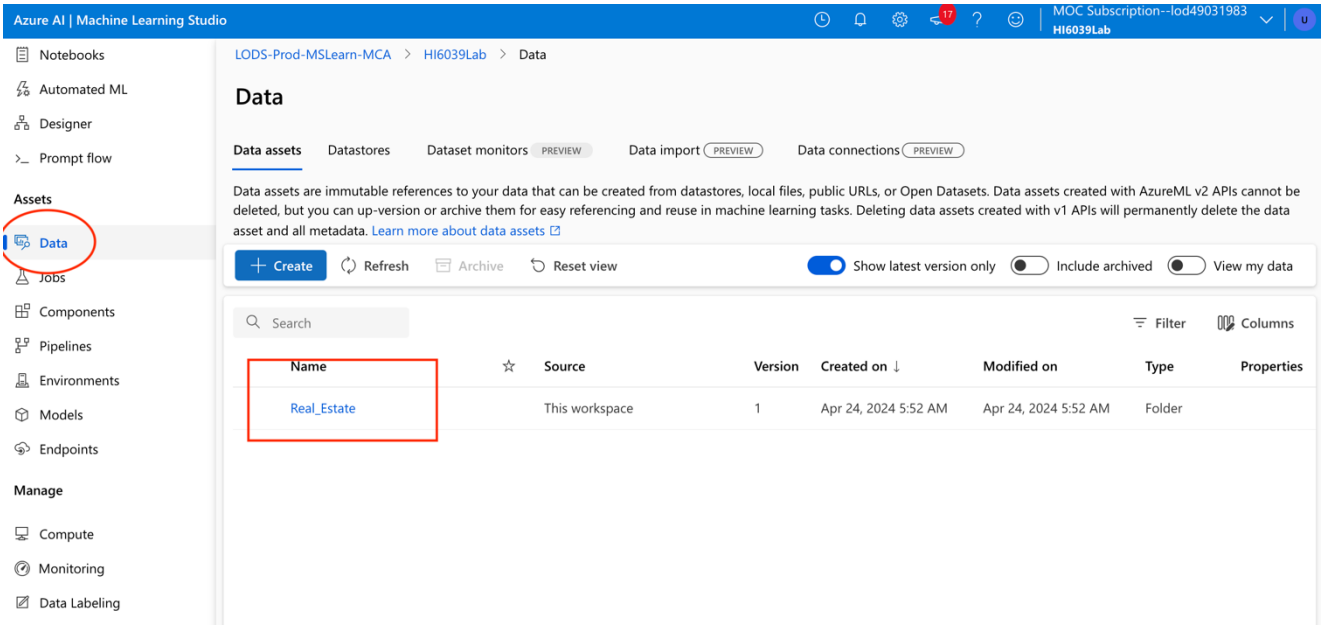
Before proceeding with the tasks ahead, make sure you have created ***resource***, ***workspace***, and ***compute*** in MS Azure Machine Learning. If you need guidance on this, please refer to Tutorial 4.

Create dataset

You need to download 'Real estate price prediction' dataset (Real estate.csv) from the following URL

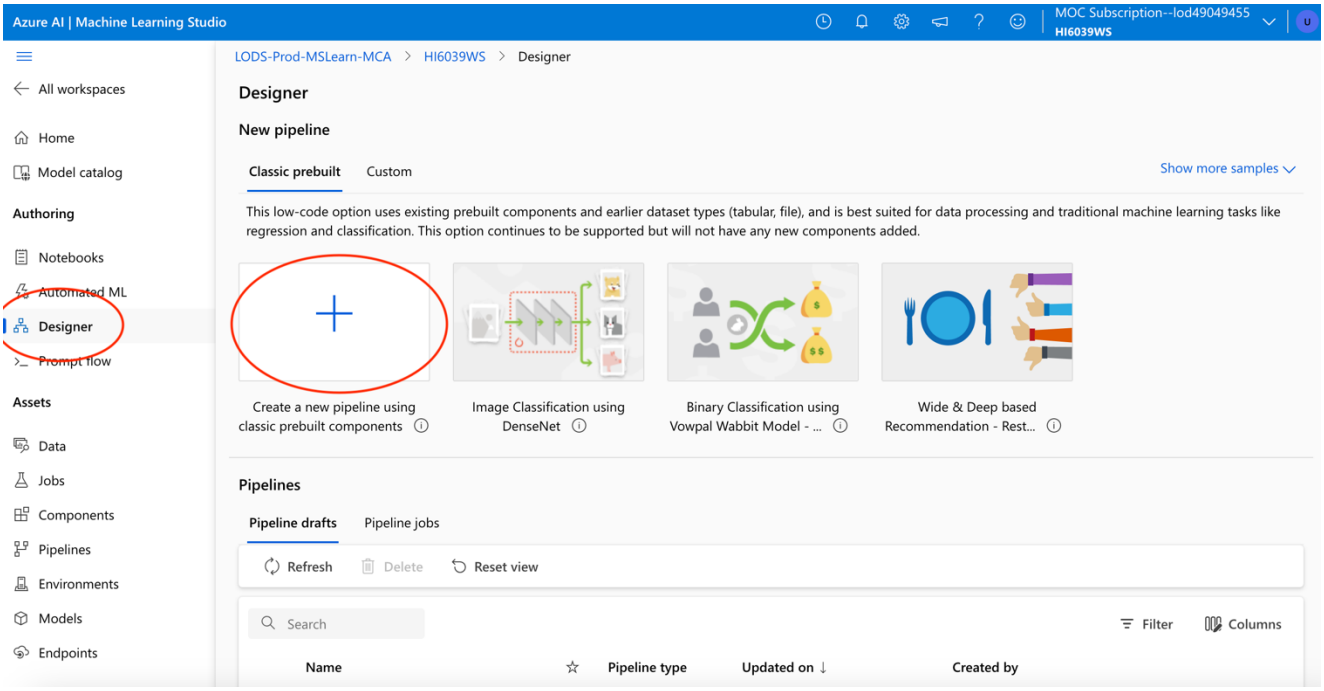
https://www.kaggle.com/datasets/quantbruce/real-estate-price-prediction

and create a dataset 'Real_Estate'

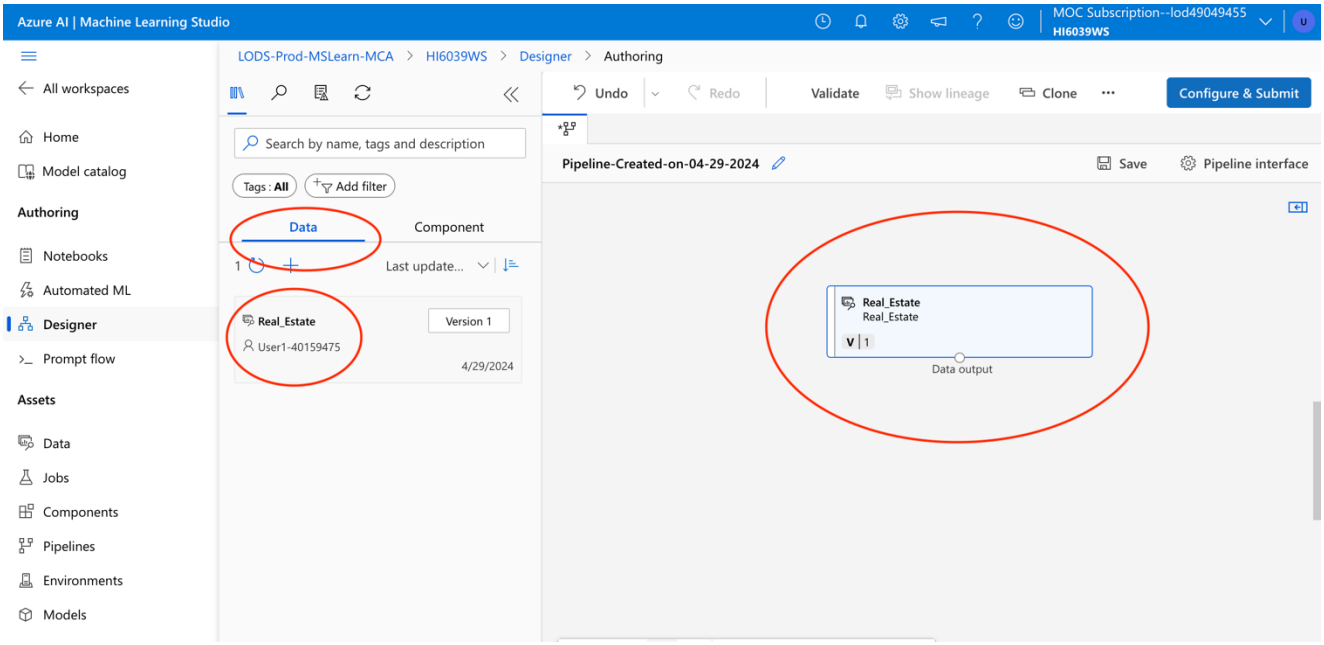


Regression Analysis

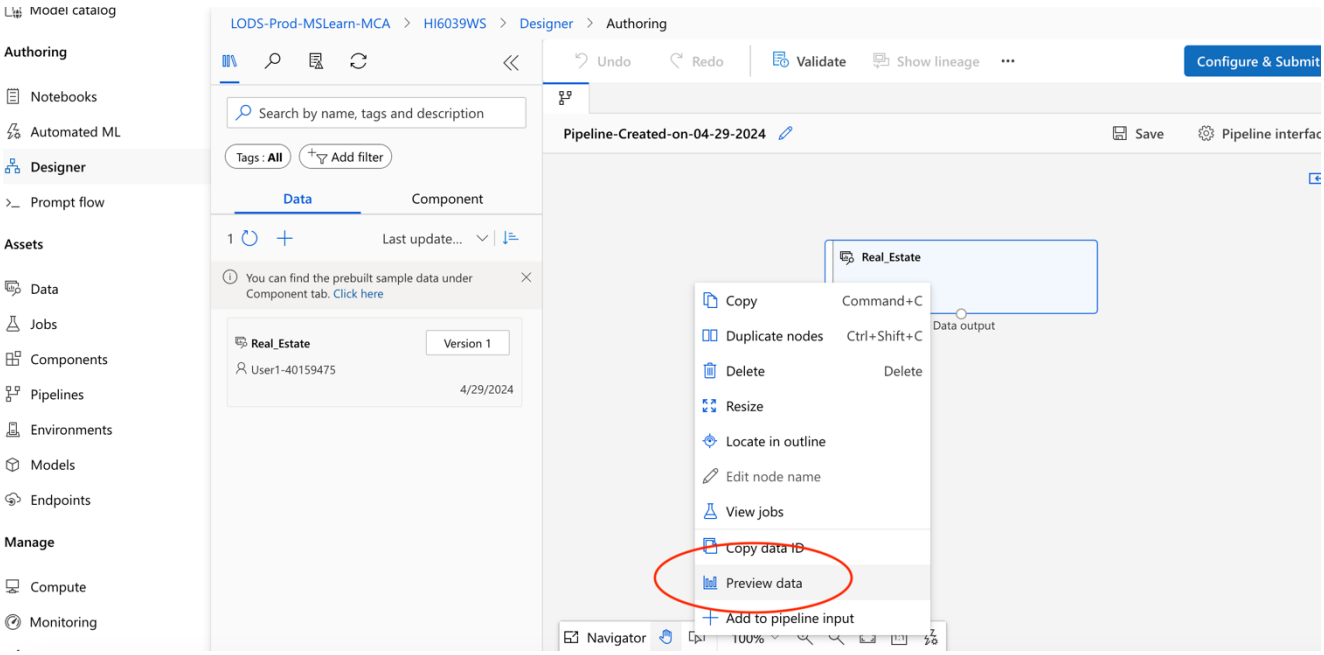
1. Select ‘Designer’ and click ‘Create a new pipeline using classic prebuilt components’.



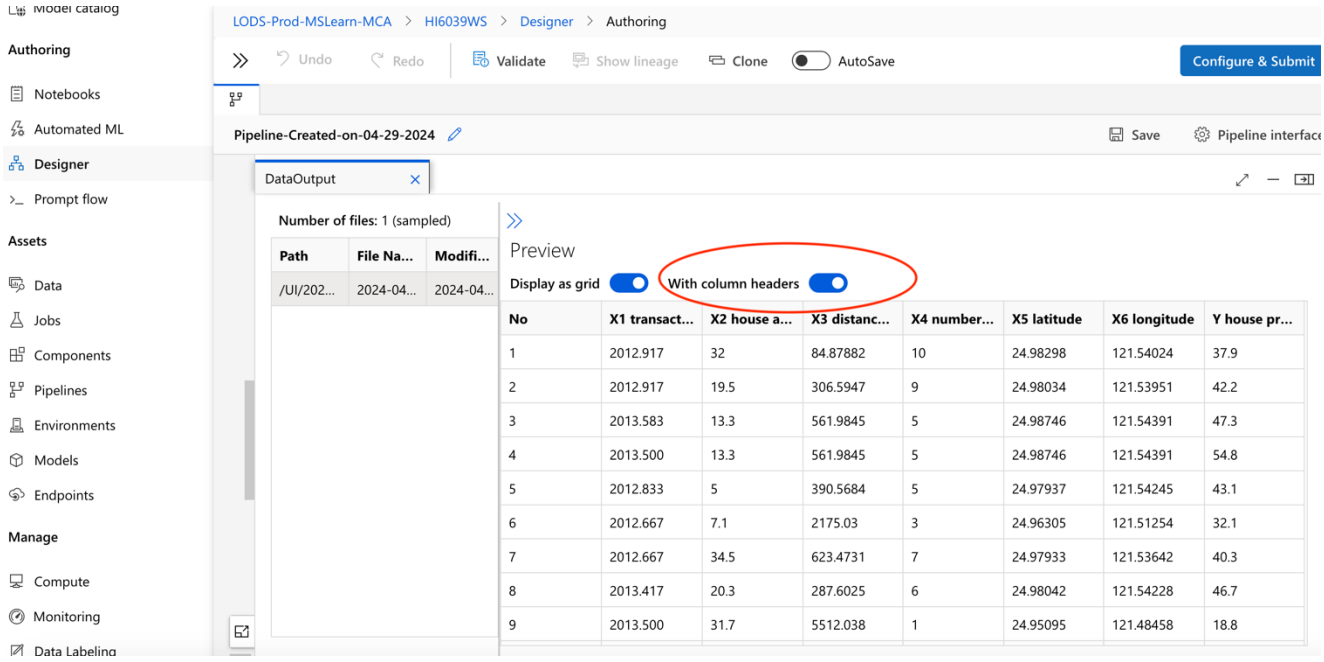
2. Drag the dataset ‘Real_Estate’ and drop it to the pipeline.



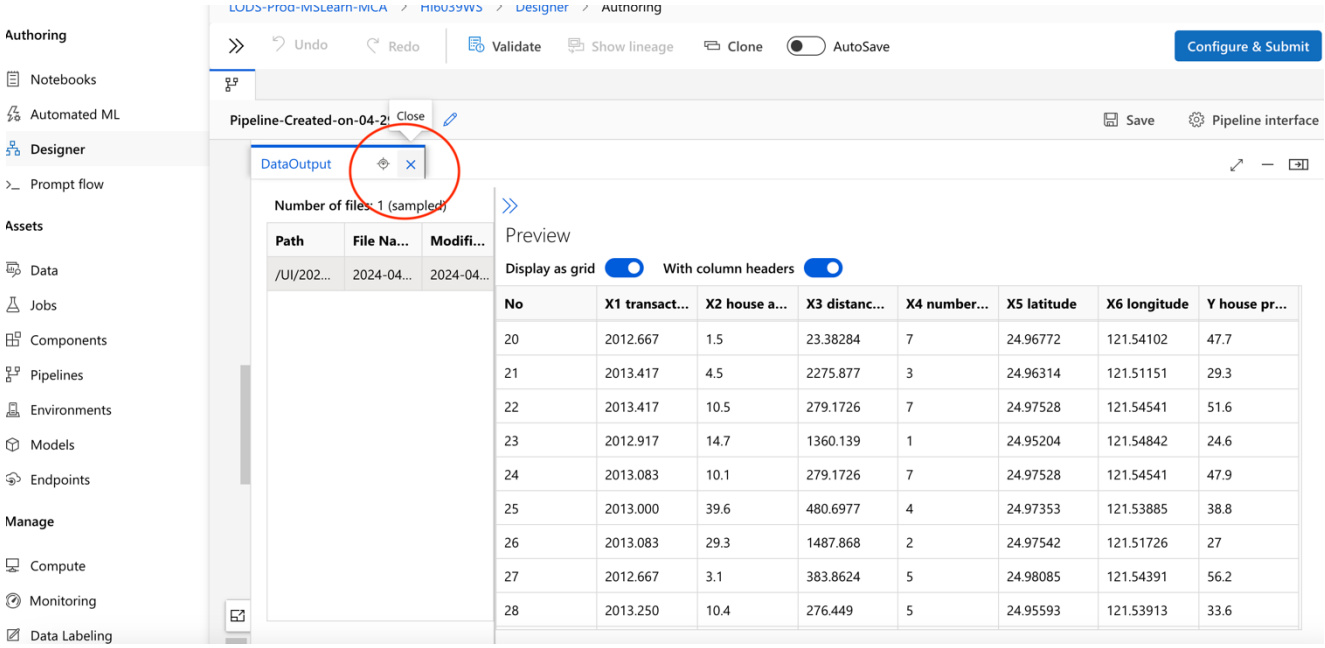
3. Right-click‘Real_Estate’ dataset and select ‘Preview data’.



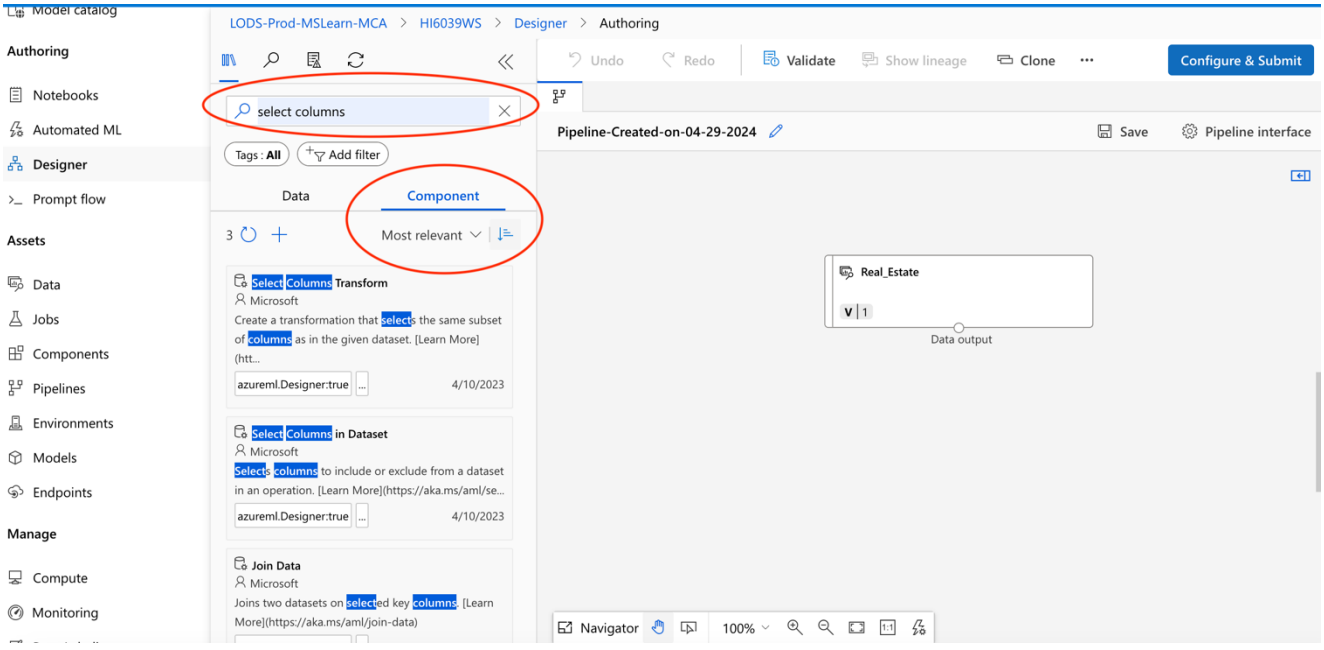
4. Turn on ‘With column headers’. Now, you can see the content of ‘Real_Estate’ dataset that contains 8 attributes.



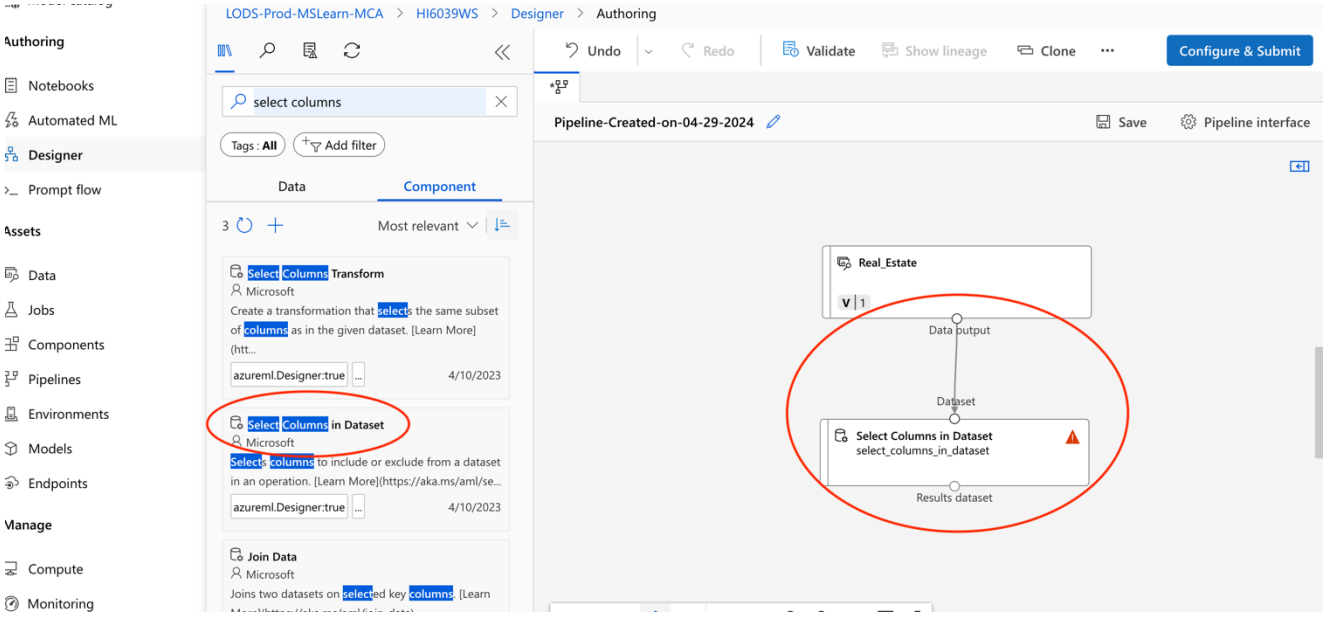
5. Close the ‘DataOutput’



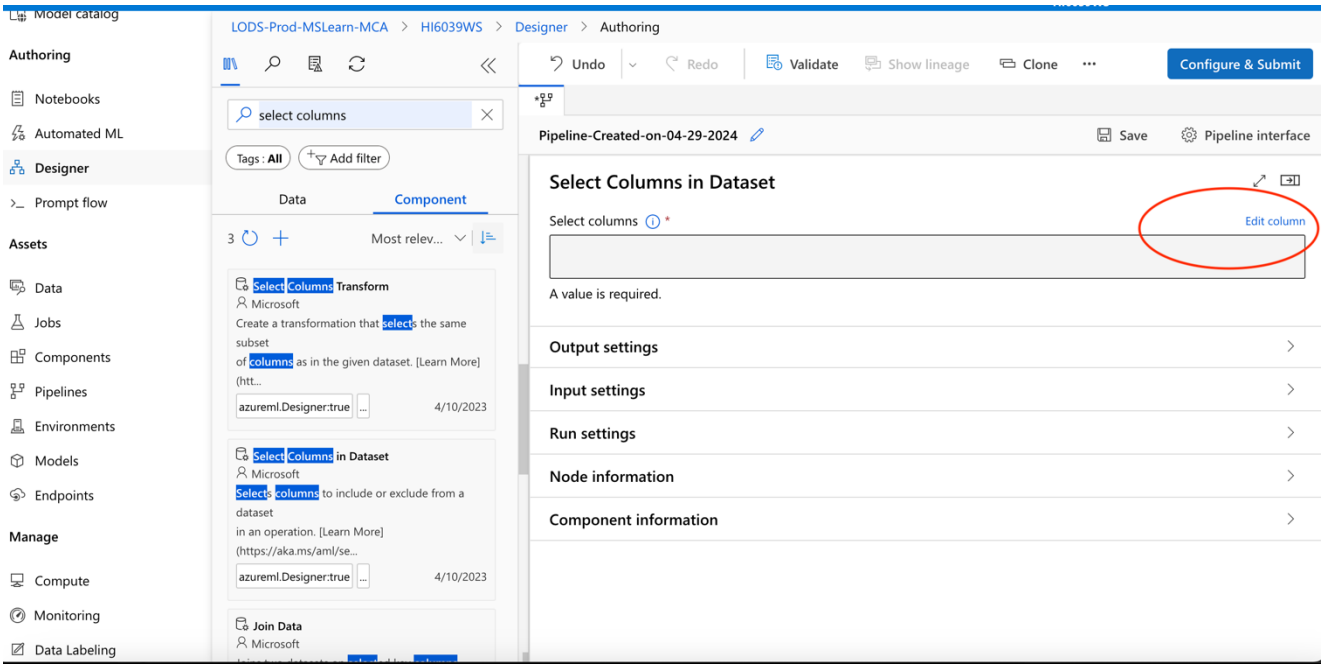
6. Select the 'Component' option and then type 'Select columns' into the search box.



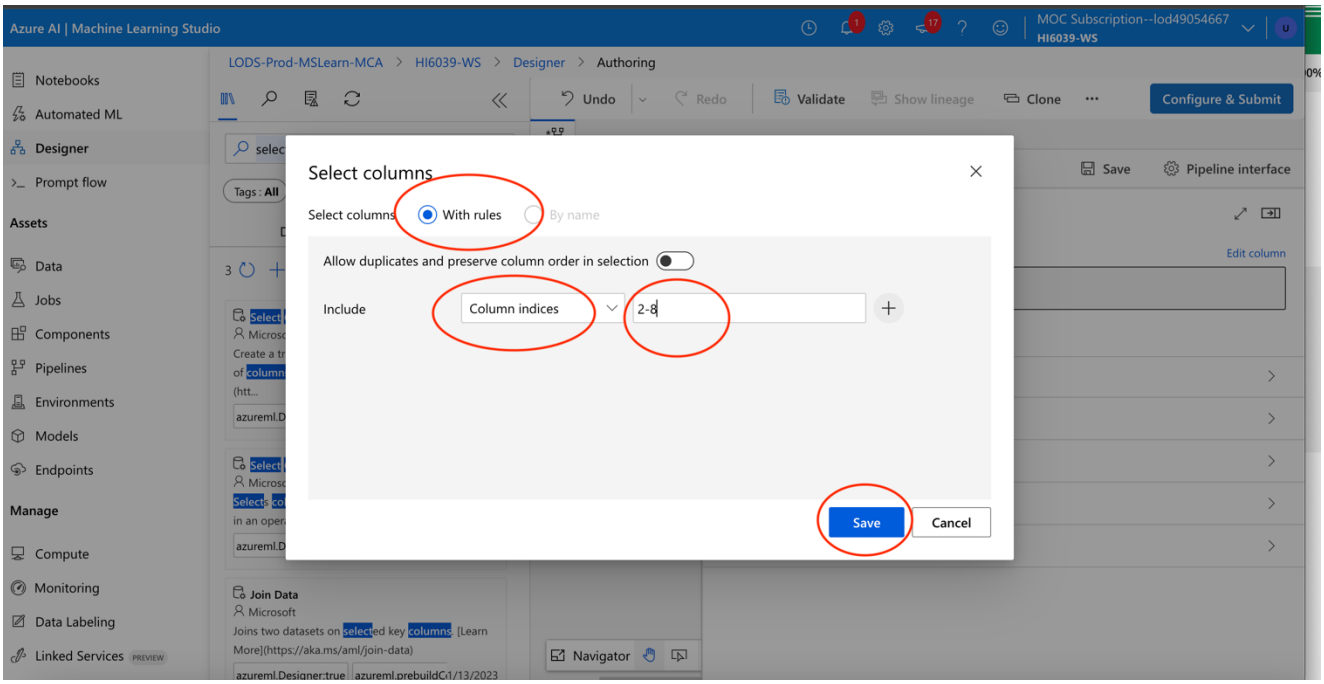
7. Drag ‘Select Columns in Dataset’ component and drop it to the pipeline. Then, connect ‘Real_Estate’ dataset to the ‘Selected Columns in Dataset’ component.



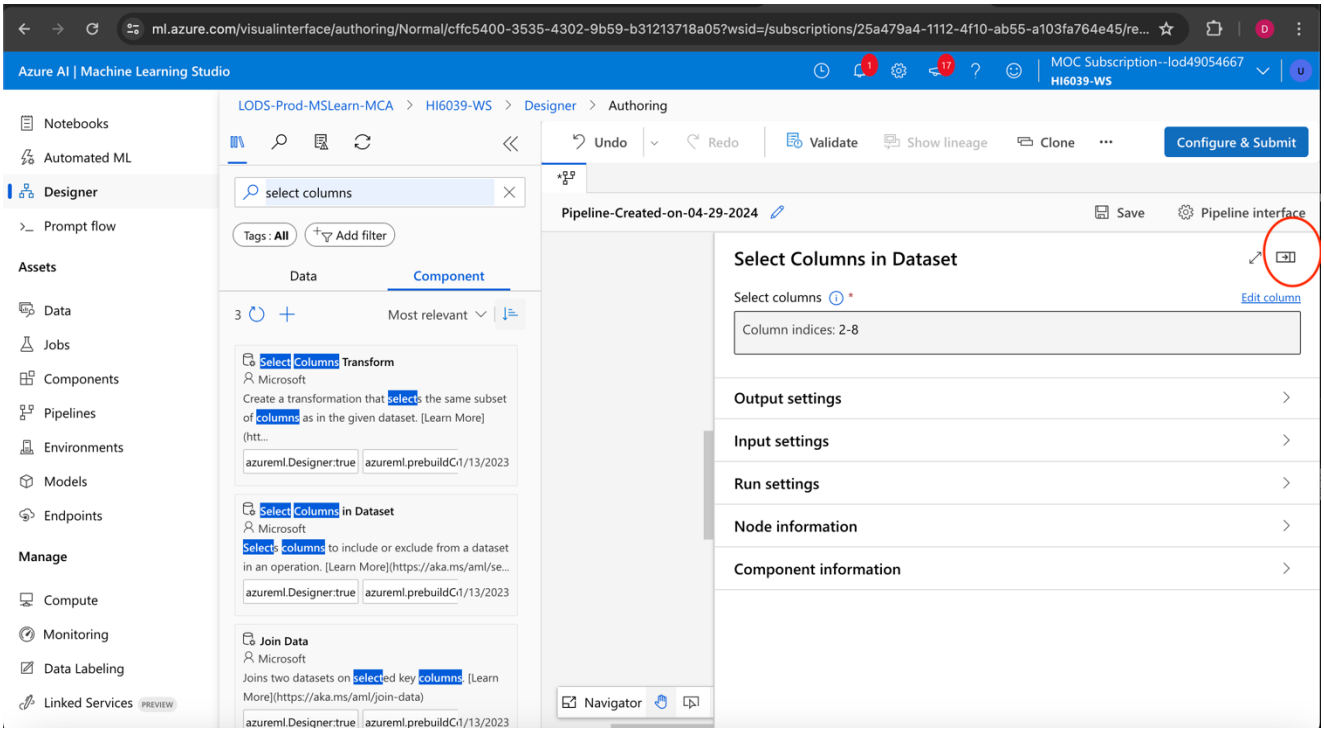
8. Double-click the ‘Selected Columns in Dataset’ component and select ‘Edit column’.



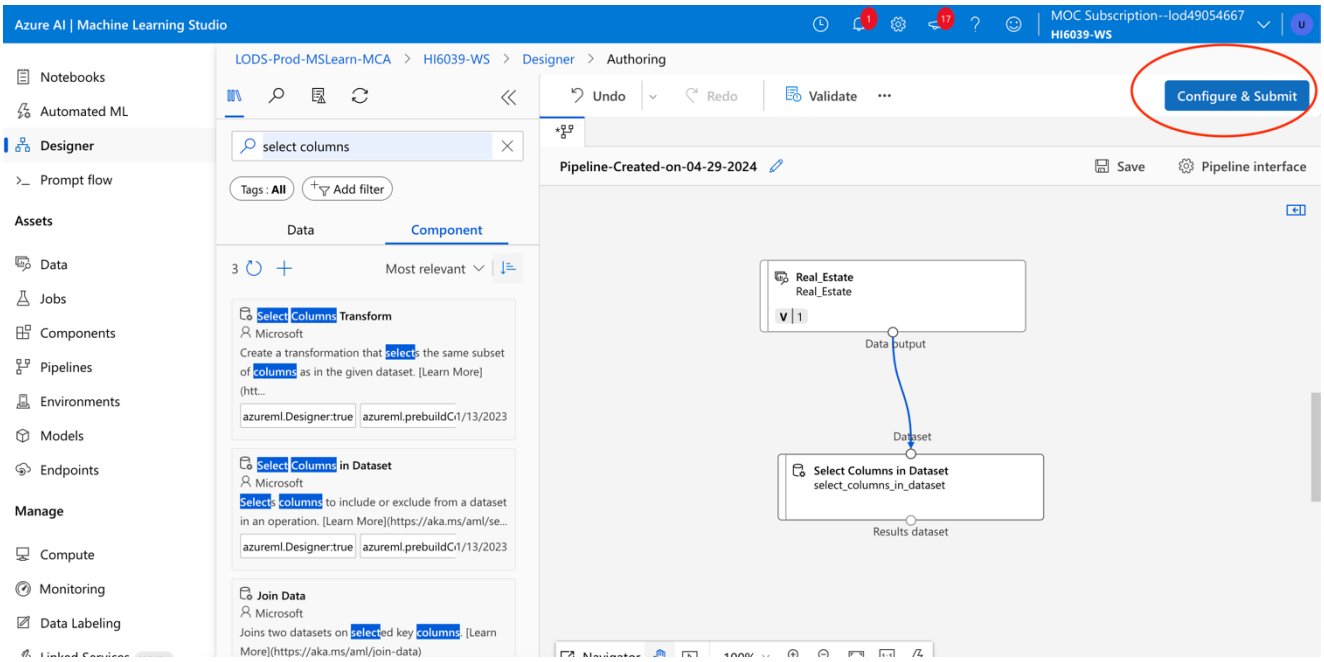
9. Select ‘With rules’, ‘Columns indices’, enter ‘2-8’, and click ‘Save’.



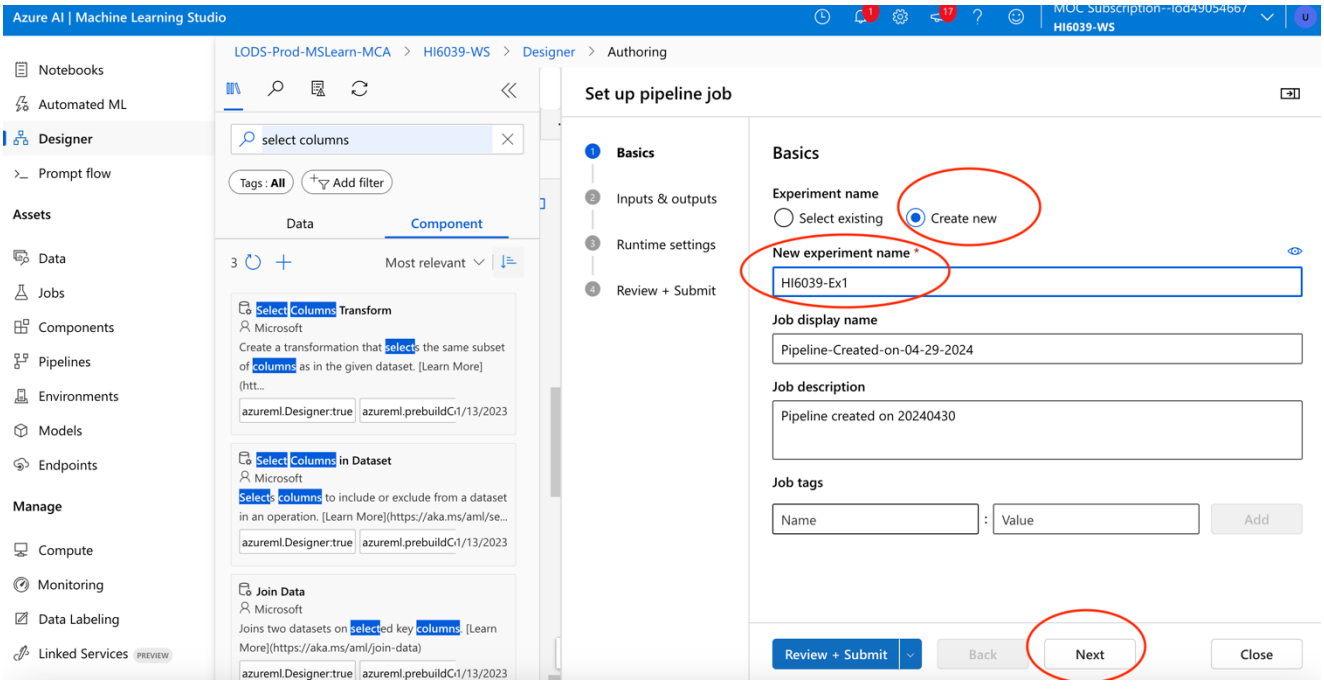
10. Click ‘Close’.



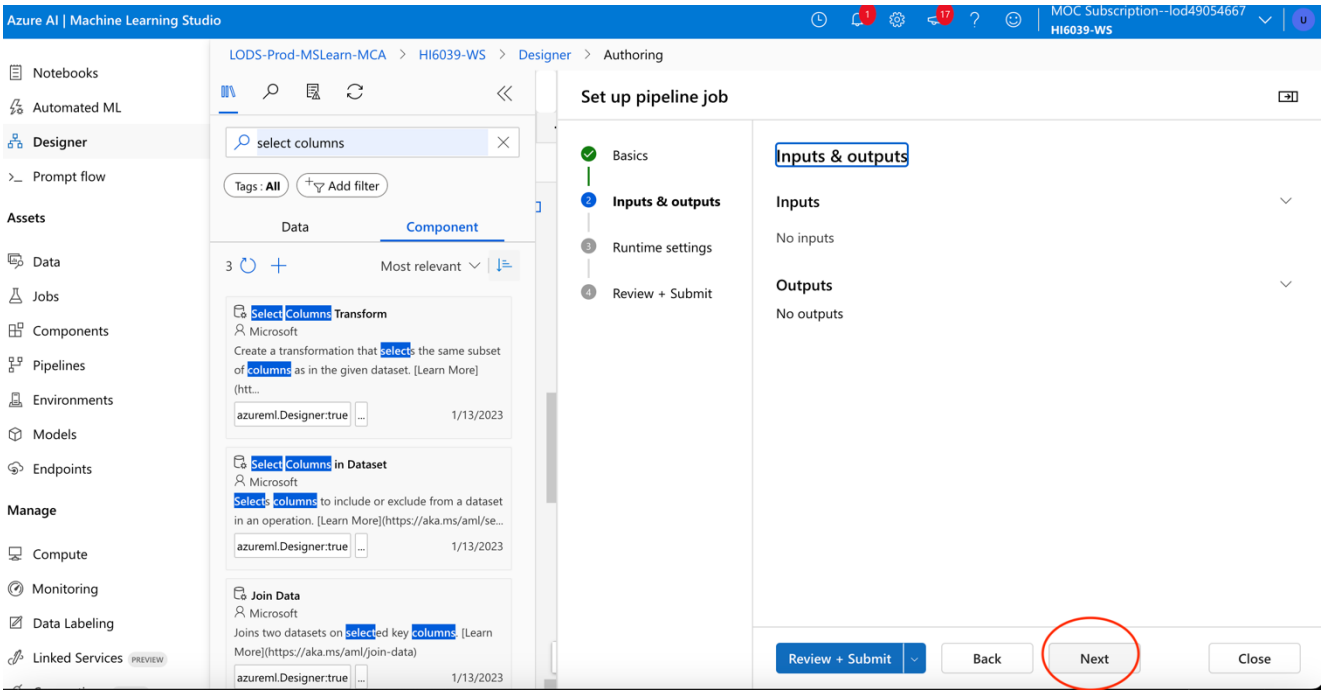
11. Click ‘Configure & Submit’.



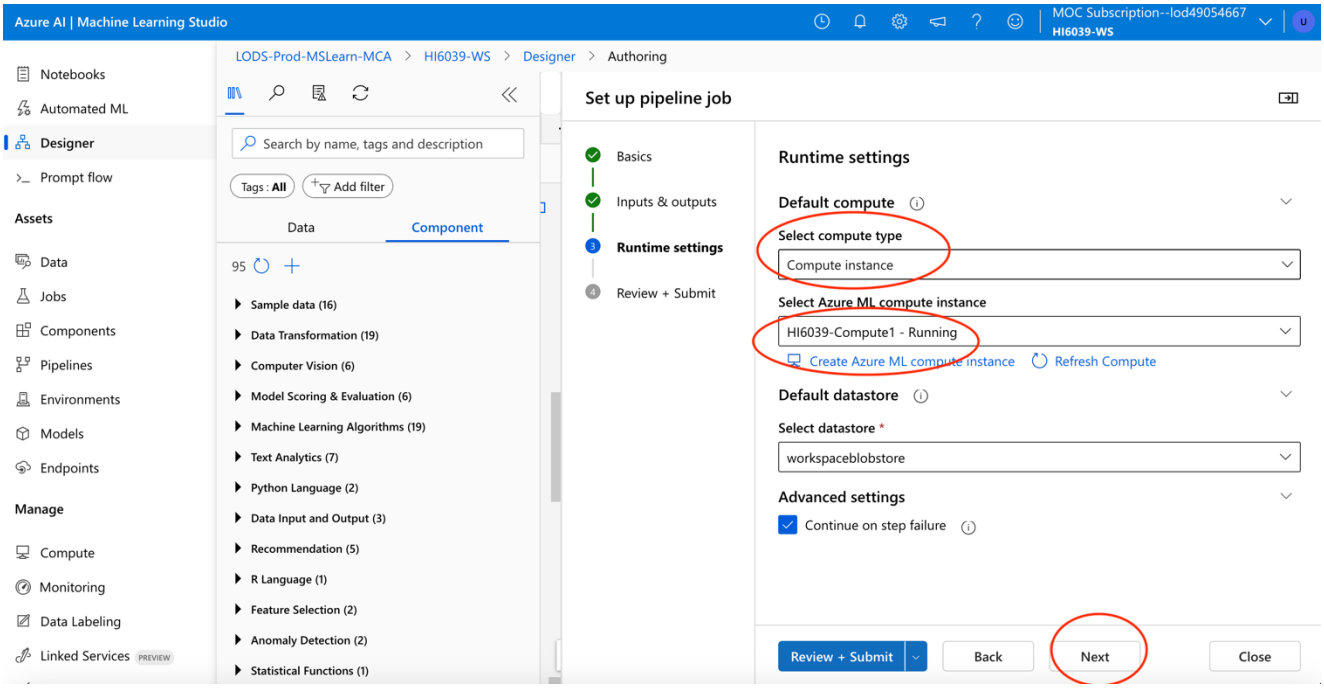
12. Select ‘Create new’, enter a new experiment name, and click ‘Next’



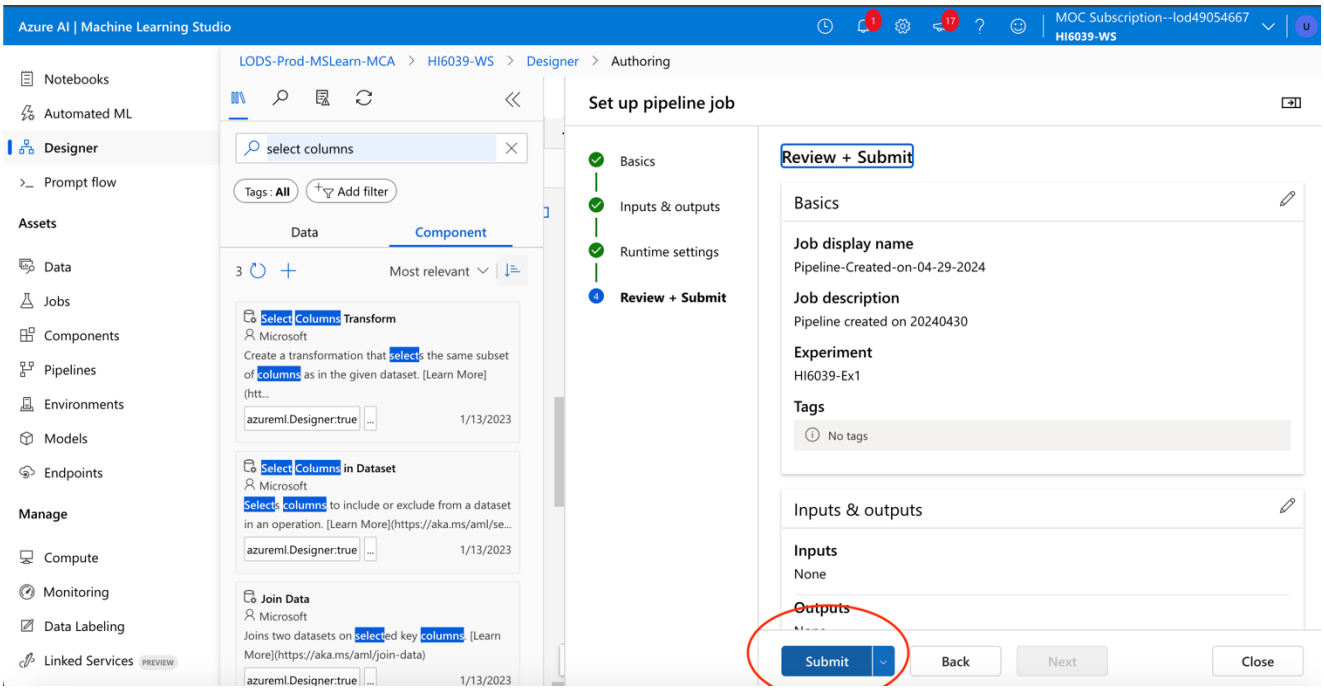
13. Click ‘Next’.



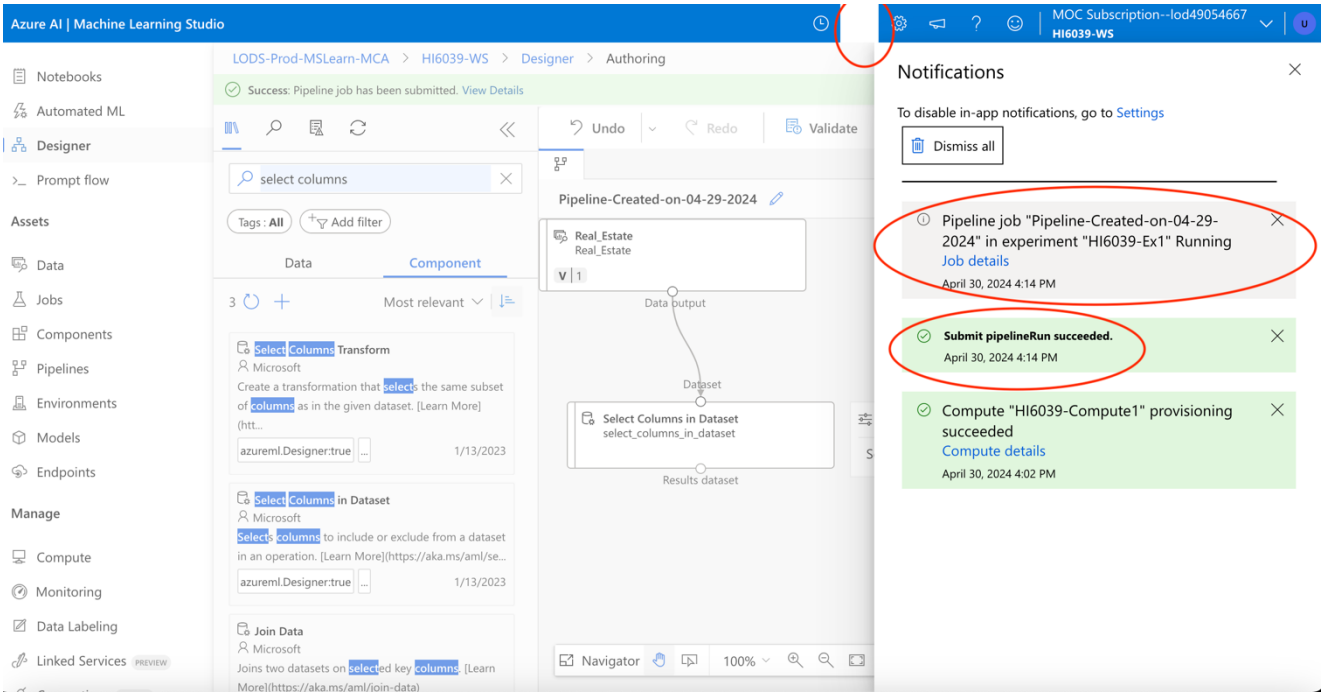
14. Select compute type ‘Compute instance’, select the compute you have already created, and select ‘Next’.



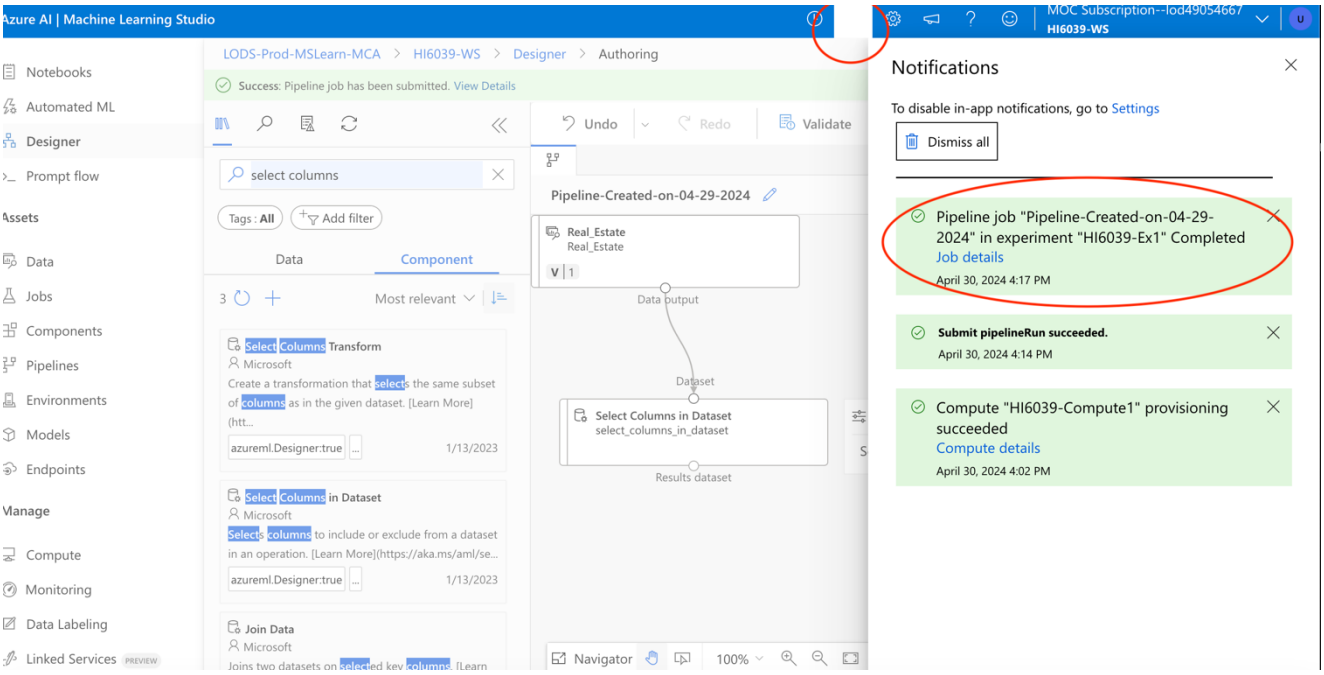
15. Click ‘Submit’.



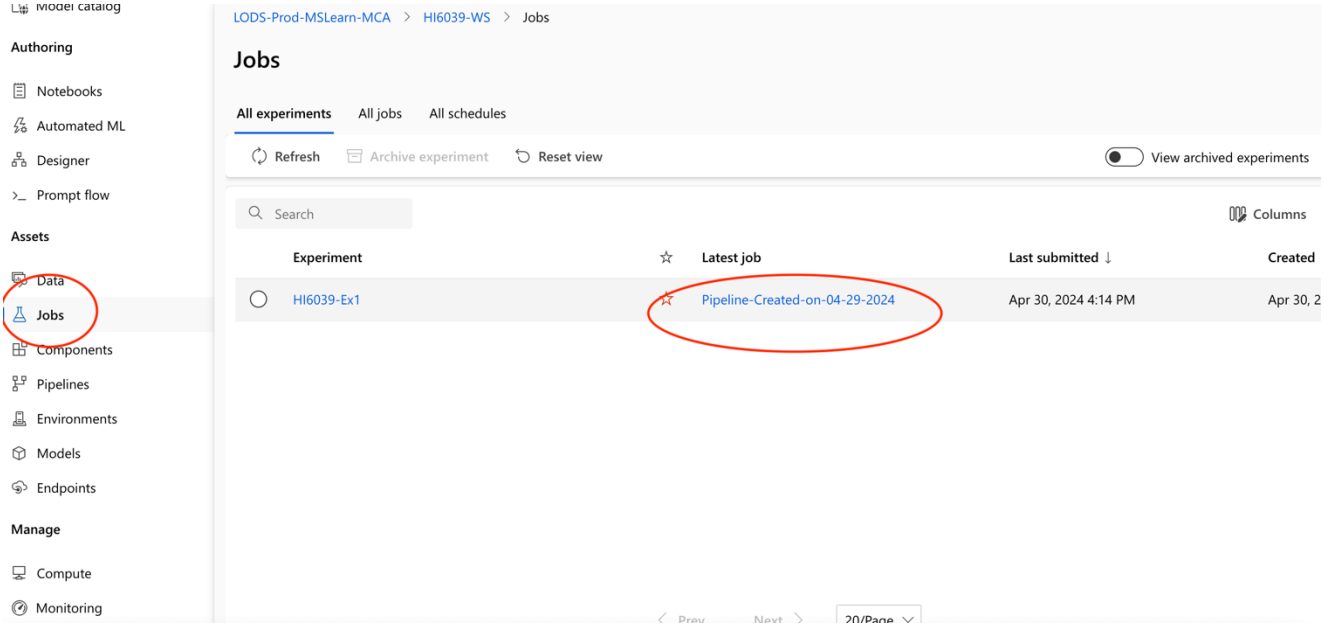
16. Now, checking ‘Notifications’, you can see the pipeline (a job) has been submitted and it is running.



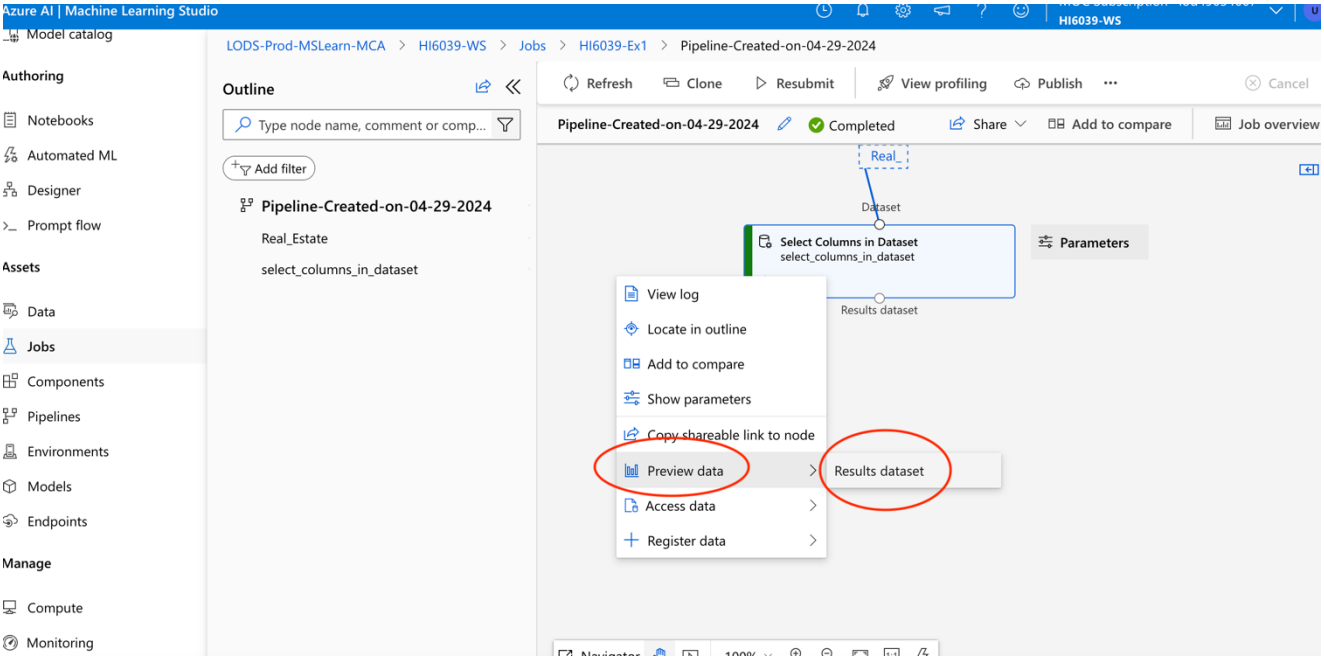
17. Waiting until the job has completed.



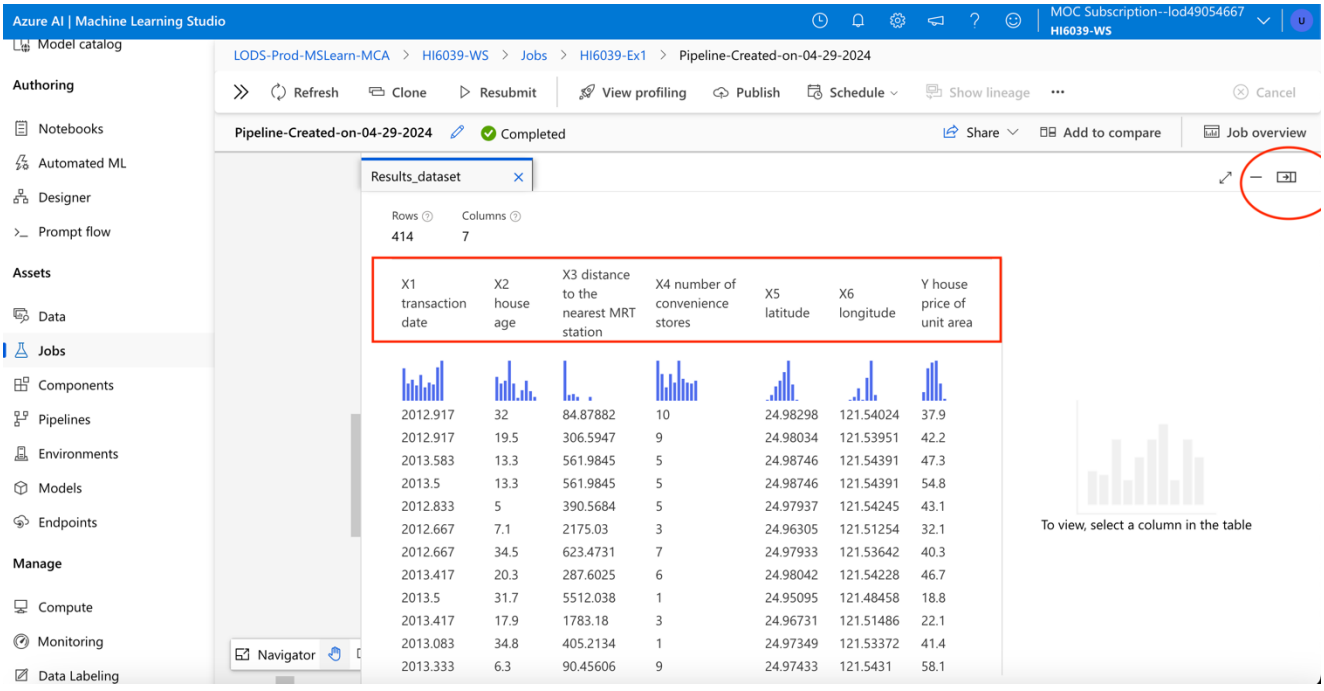
18. Select ‘Job’ and then select the pipeline (the latest job).



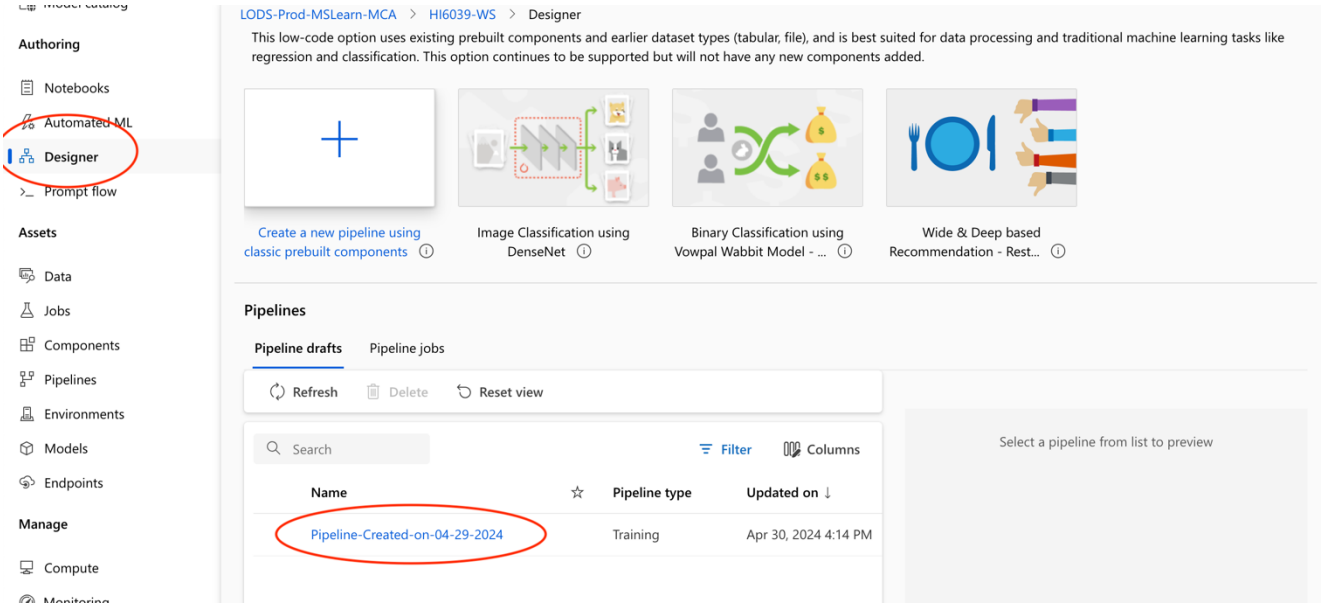
19. Right-click ‘Select Columns in Dataset’ component, select ‘Review data’, and select ‘Results dataset’



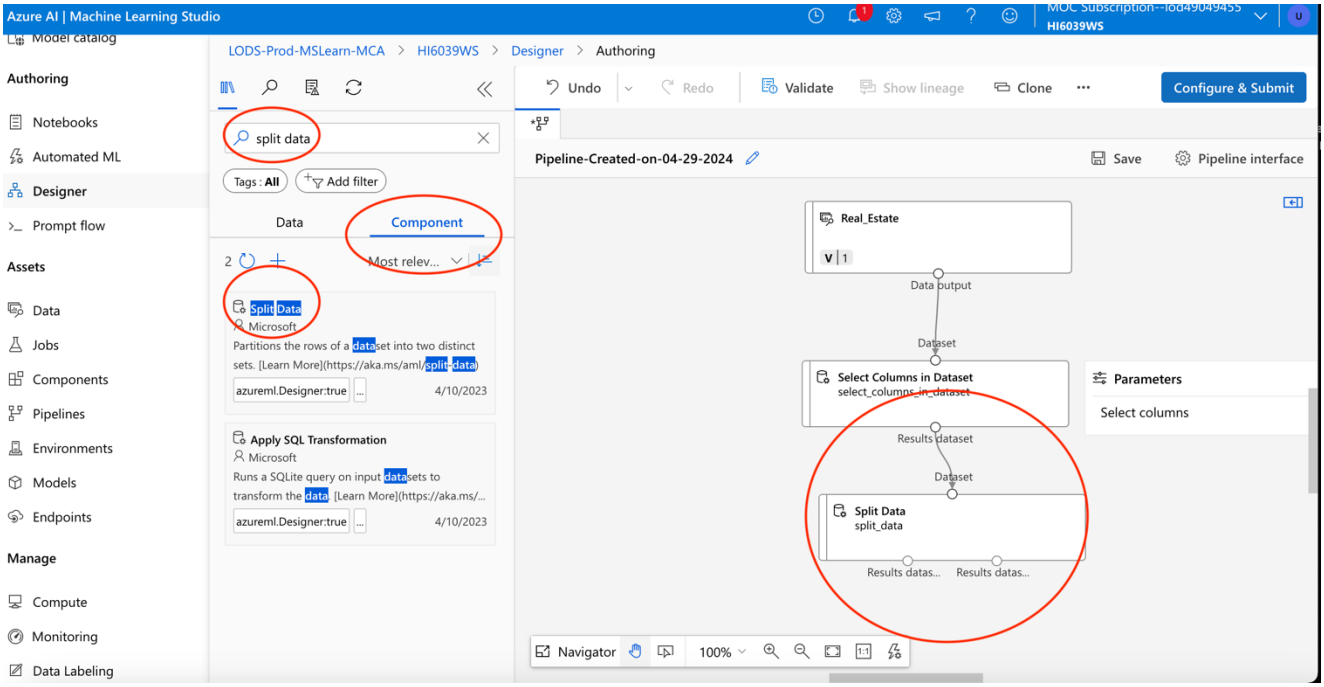
20. Now, you can see 7 selected attributes. Click ‘close’ to continue.



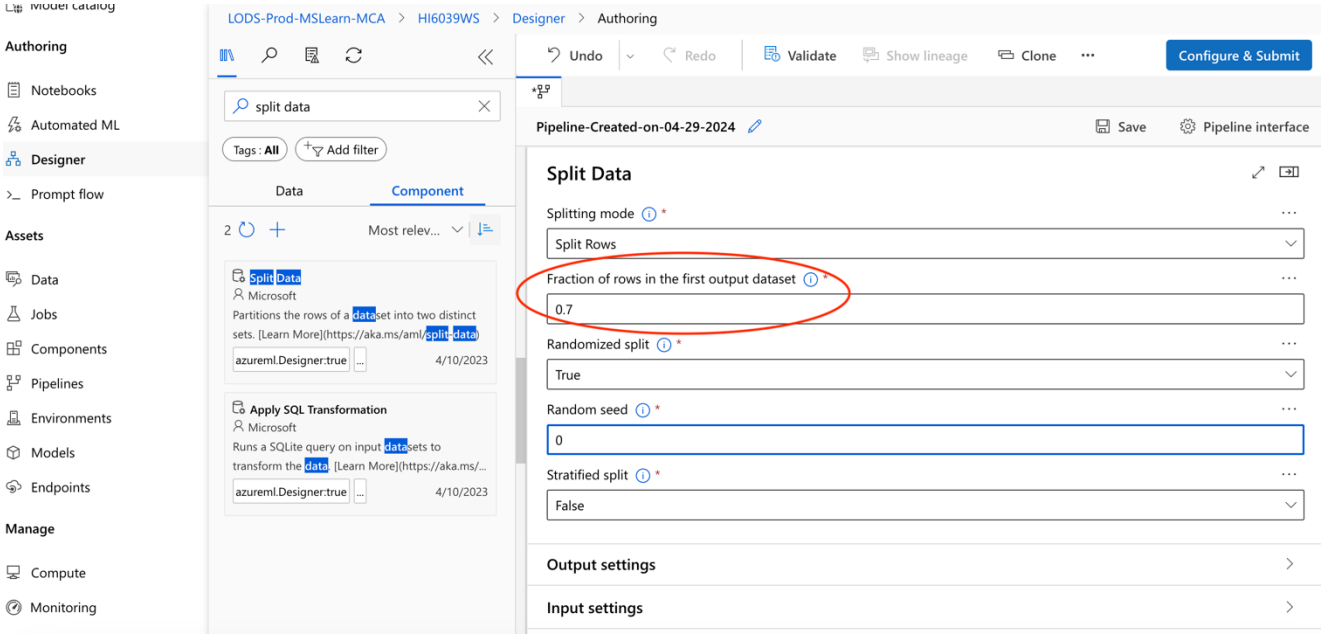
21. Click ‘Designer’ and click the pipeline.



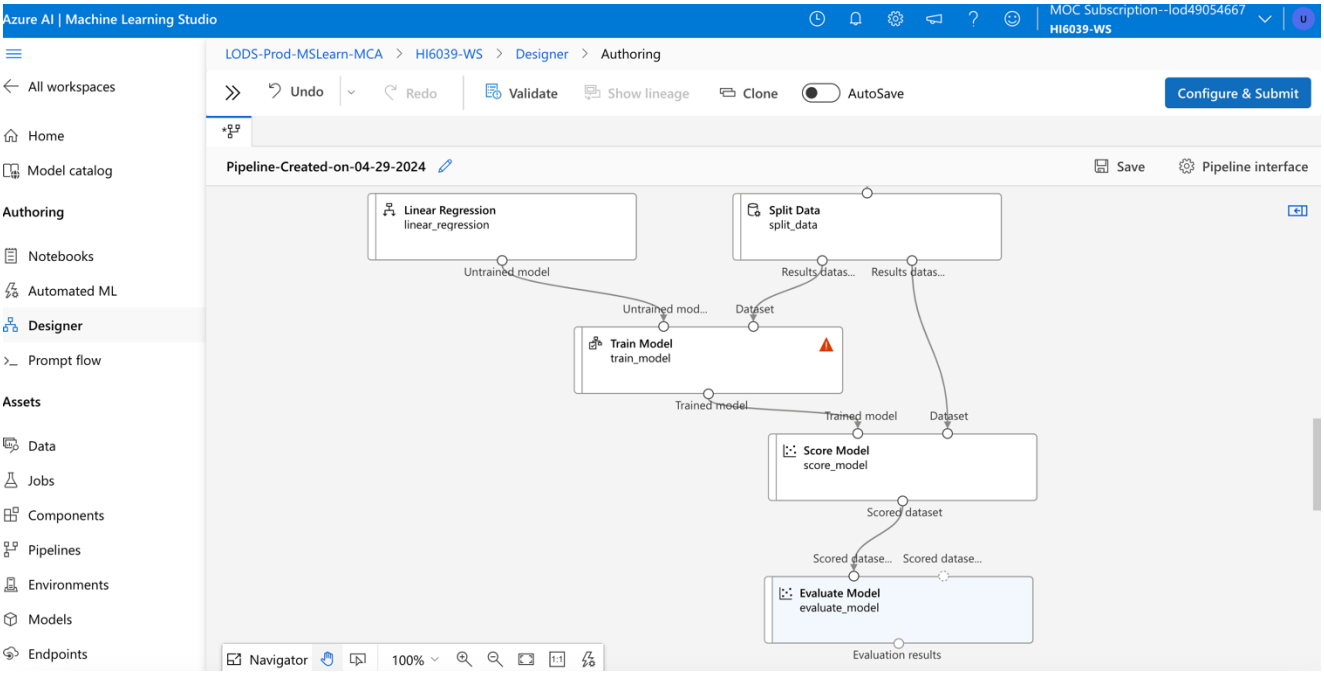
22. Search and add ‘Split Data’ component to the pipeline. Then, connect the ‘Selected Columns in Dataset’ to the ‘Split Data’ component.



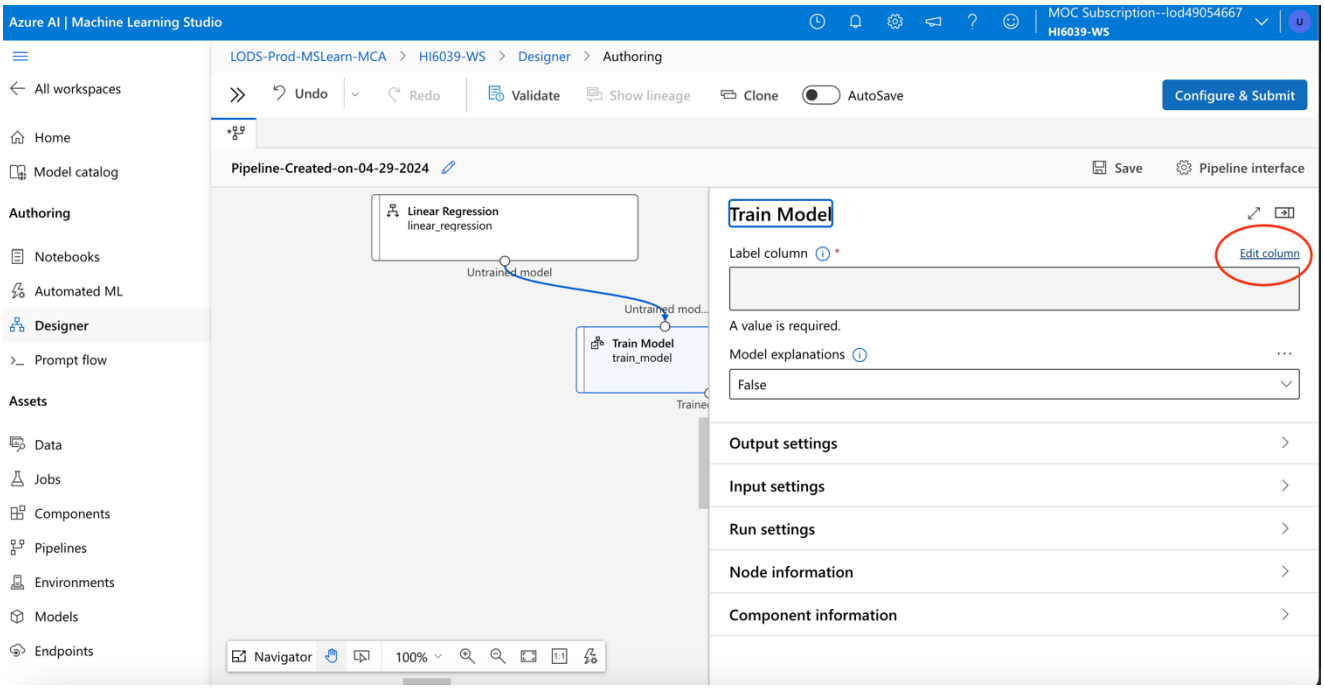
23. Double-click the ‘Split Data’ component and enter ‘0.7’ for the ‘Fraction of rows in the first output dataset’.



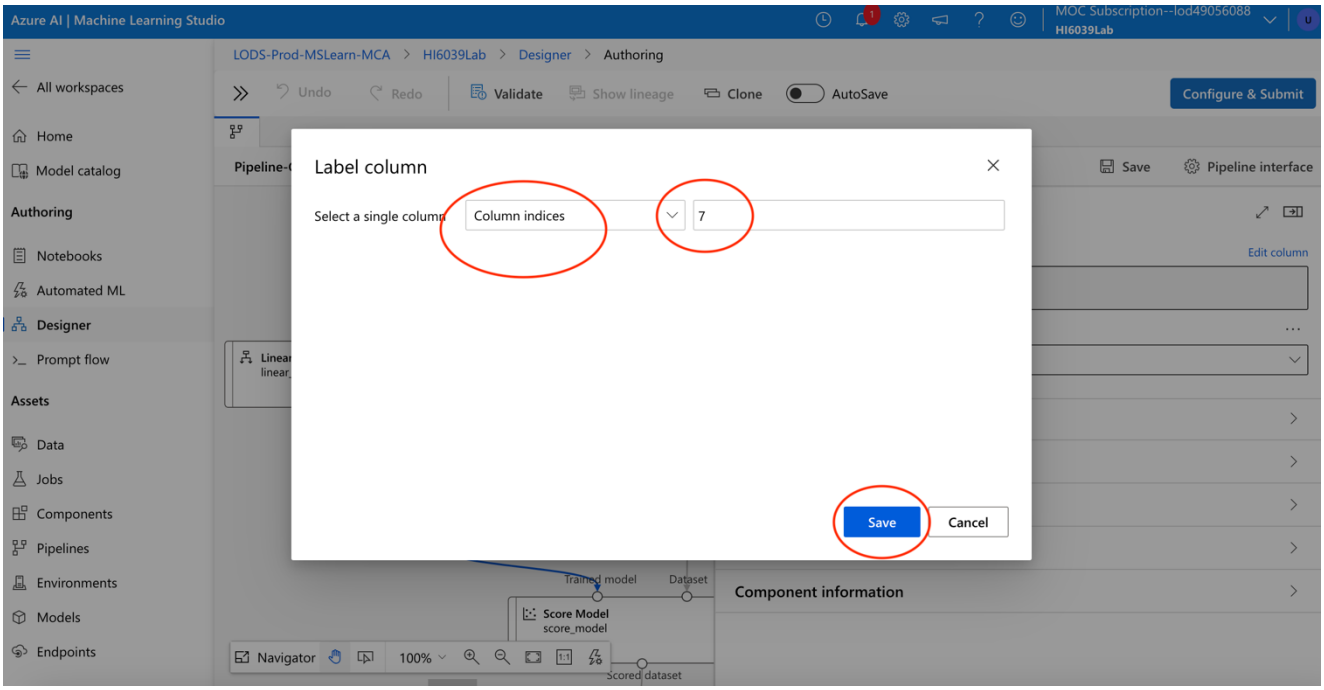
24. Search and add the following components to the pipeline: ‘Linear Regression’, ‘Train Model’, ‘Score Model’, and ‘Evaluate Model’. Then, connect them as shown in the screenshot below.



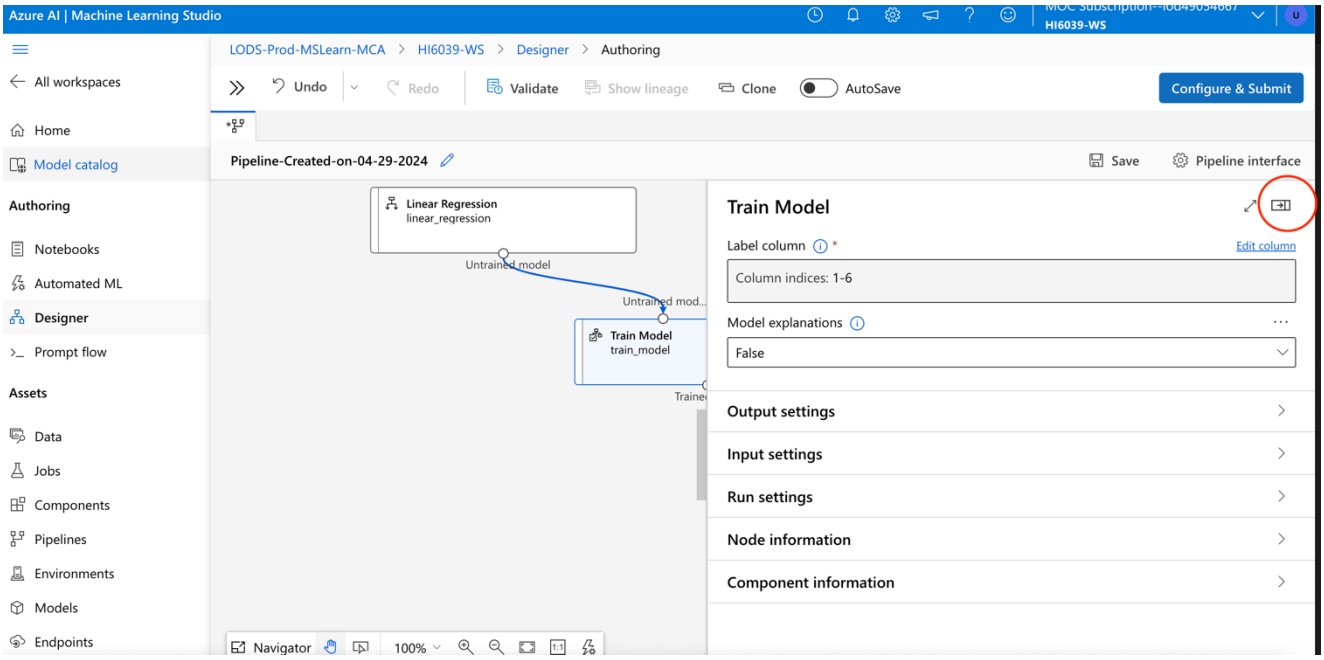
25. Double-click ‘Train Model’ component, and select ‘Edit column’.



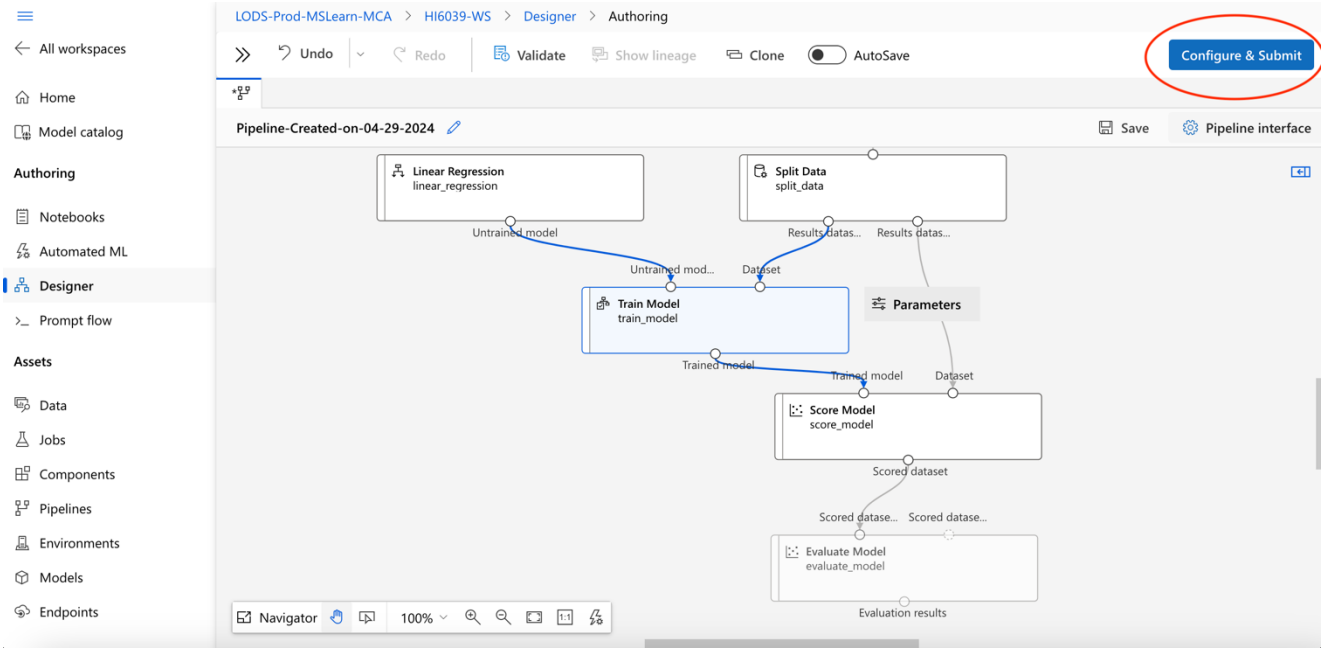
26. Select ‘Column indices’, enter ‘7’, and click ‘Save’.



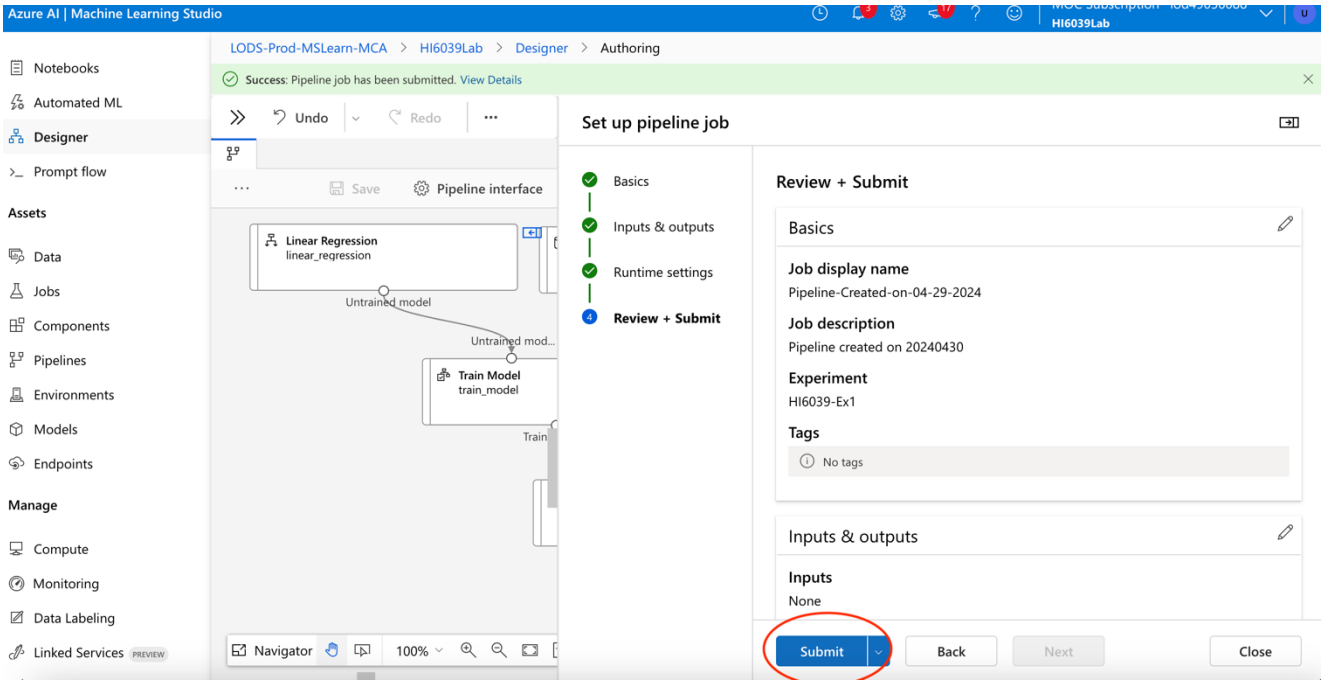
27. Click ‘Close’.



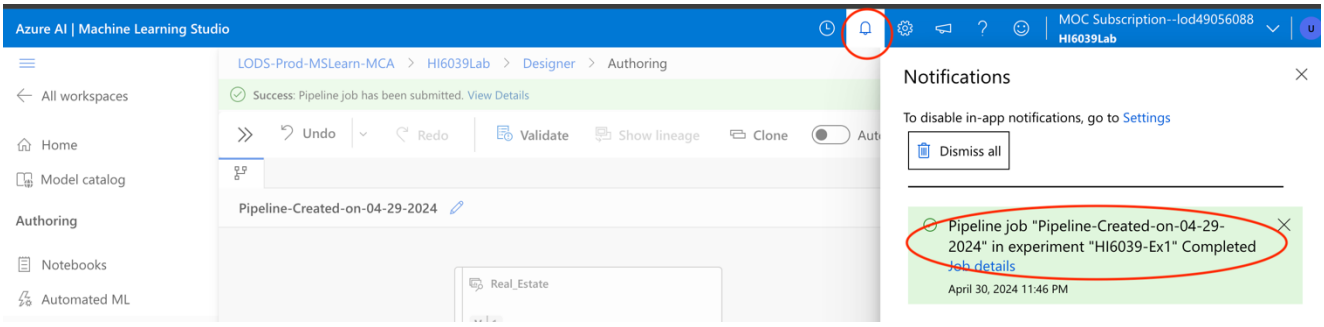
28. Click ‘Configure & Submit’.



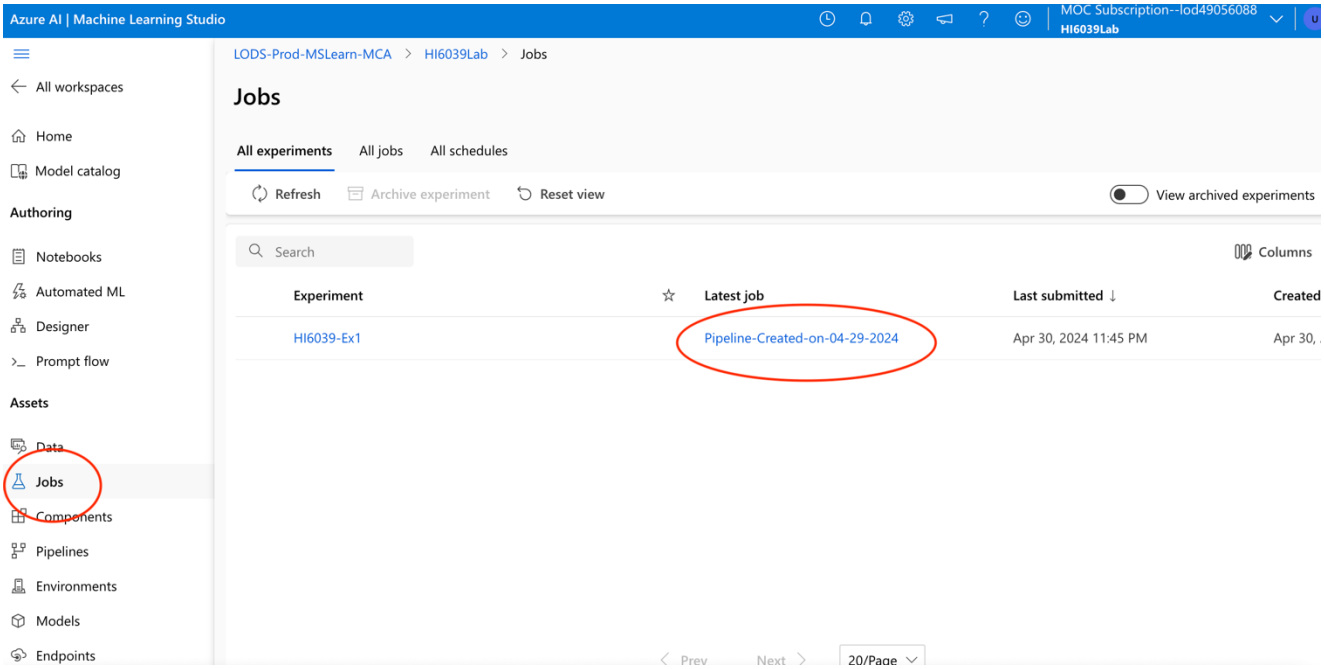
29. Click ‘Submit’.



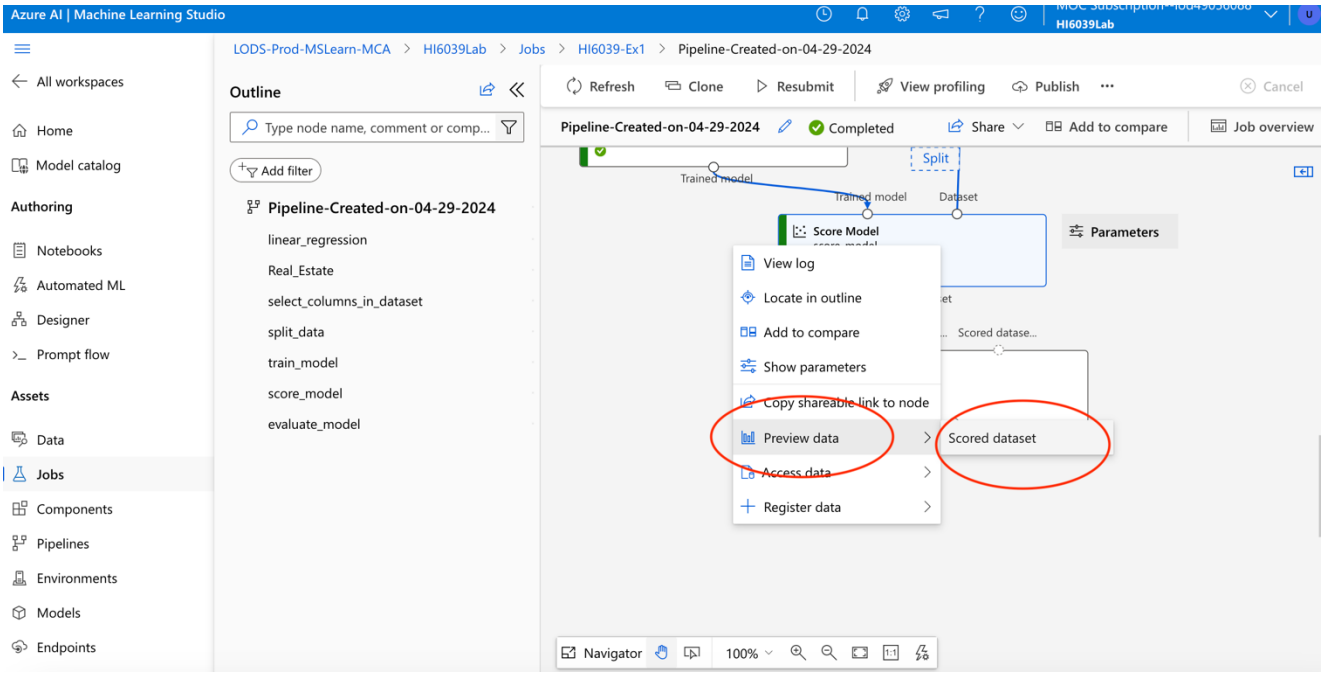
30. Waiting until the job has completed.



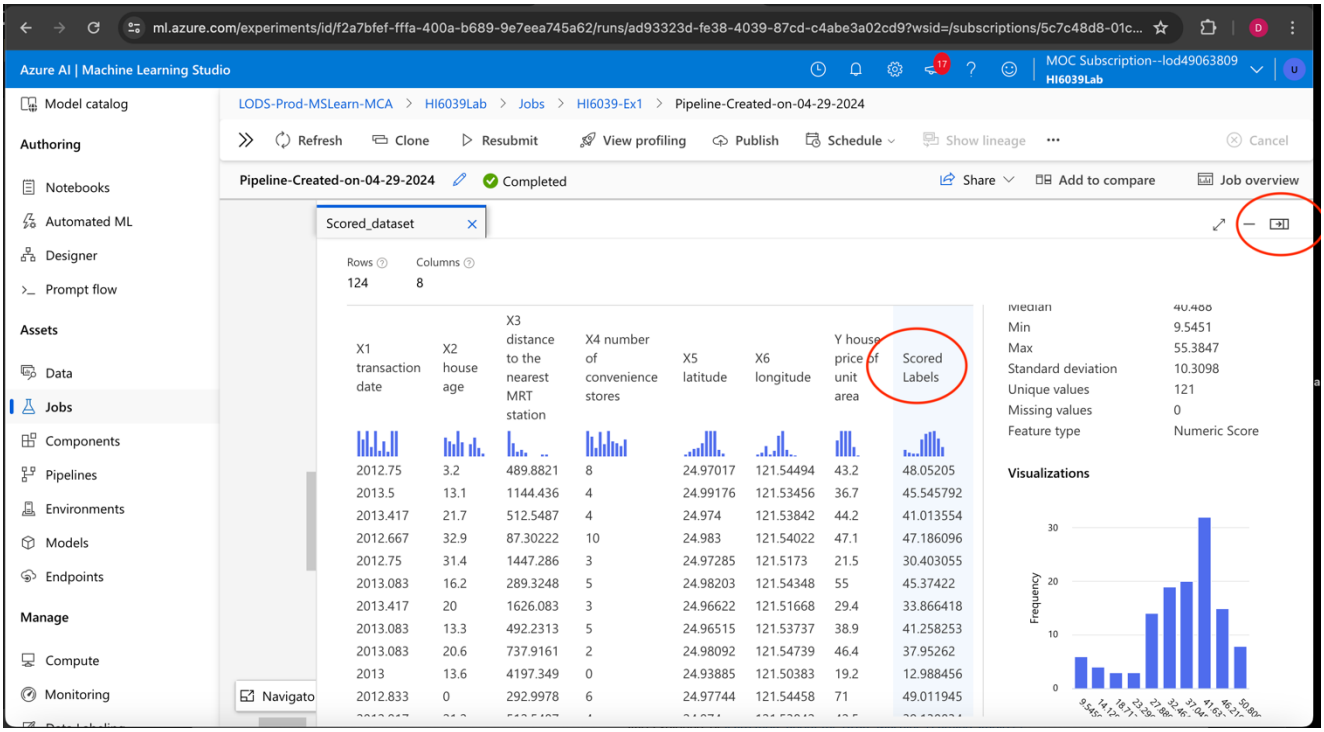
31. Click ‘Job’, then click the pipeline.



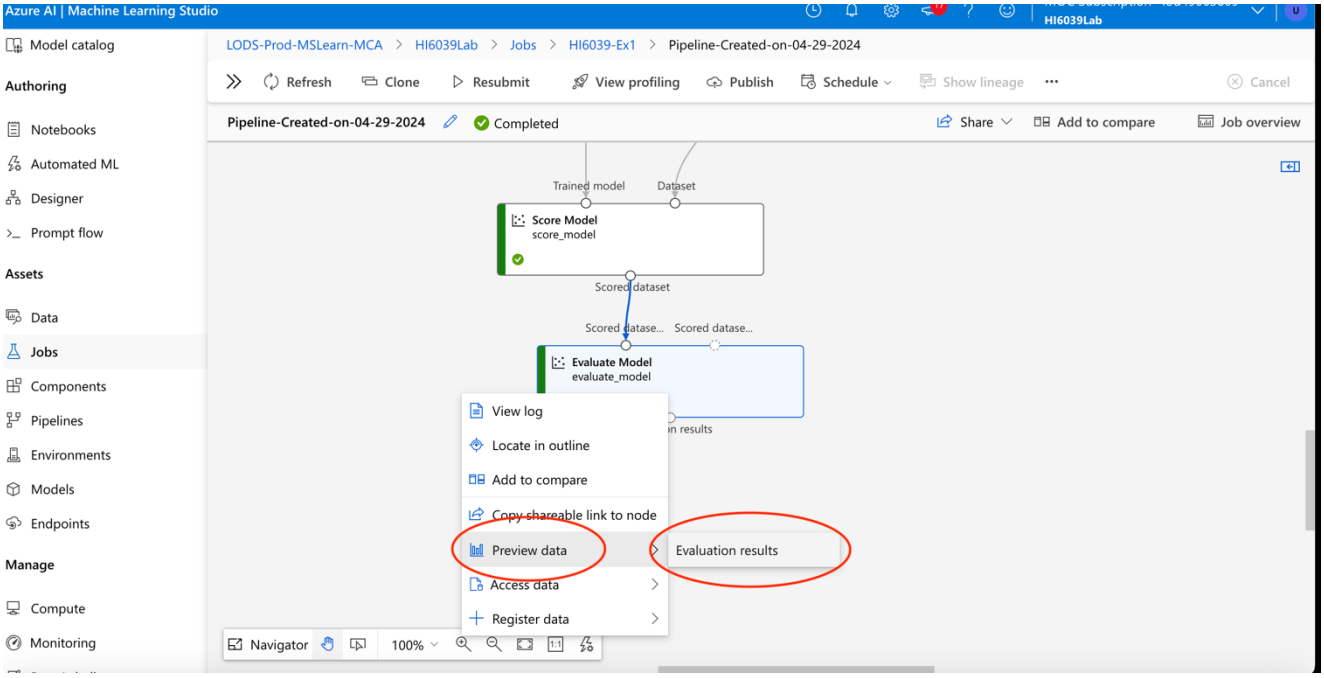
32. Right-click ‘Score model’ component, select ‘Preview data’, and select ‘Scored dataset’.



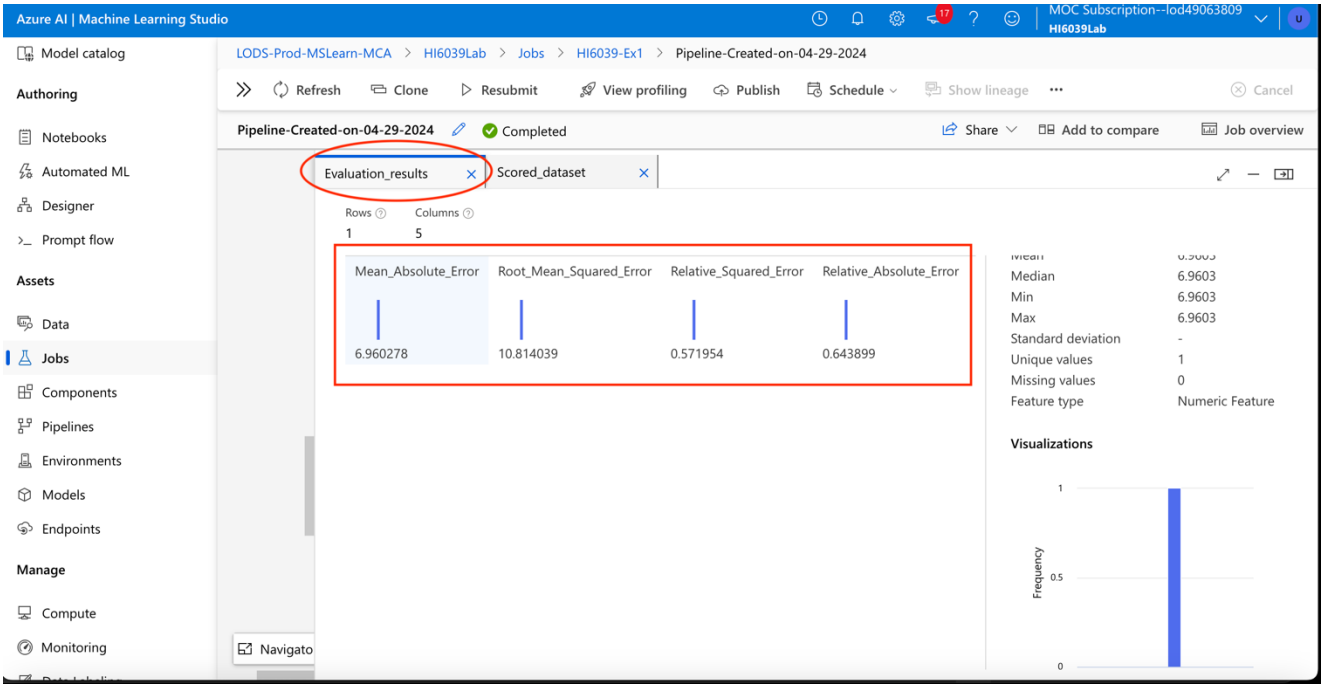
33. Now, you can see the predicted score for label/target attributes in the testing dataset. Click ‘Close’ to continue.



34. Right-click the ‘Evaluate Model’ component, select ‘Preview data’, and choose ‘Evaluation results’.



35. Now, you can see the performance of the predictive model.



---The end---