



통계 기반 데이터 활용 개요

강의 내용 알아보기

강명호

데이터와 통계

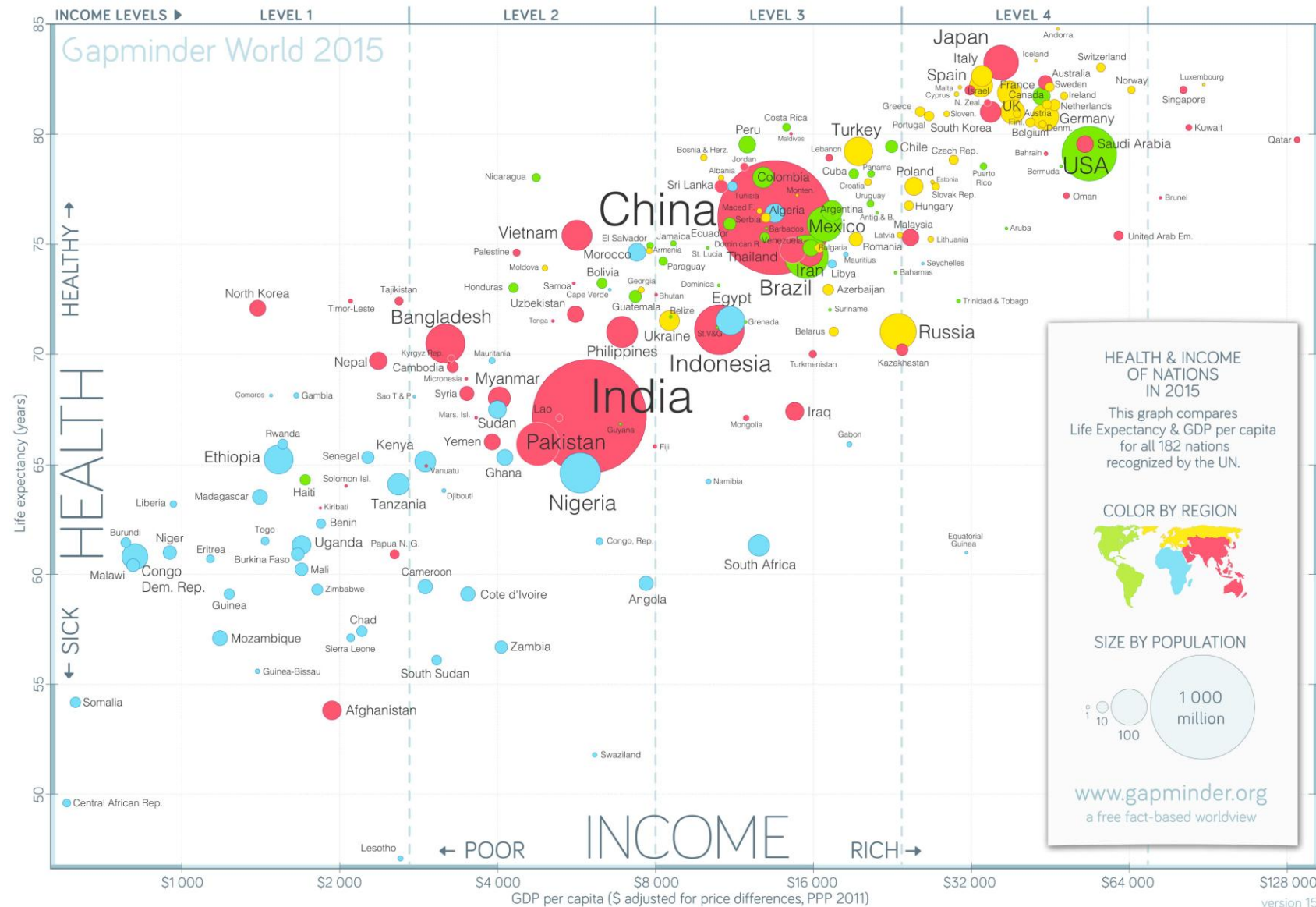
13개의 질문

1. 오늘날 세계 모든 저소득 국가에서 초등학교를 나온 여성은 얼마나 될까?
A: 20% B: 40% C: 60%
2. 세계 인구의 다수는 어디에 살까?
A: 저소득 국가 B: 중간 소득 국가 C: 고소득 국가
3. 지난 20년간 세계 인구에서 극빈층 비율은 어떻게 바뀌었을까?
A: 거의 2배로 늘었다. B: 거의 같다. C: 거의 절반으로 줄었다.
4. 오늘날 세계 기대 수명은 몇 세일까?
A: 50세 B: 60세 C: 70세
5. 오늘날 전 세계 1세 아동 중 어떤 질병이든 예방접종을 받은 비율은 몇 퍼센트일까?
A: 20% B: 50% C: 80%
6. 전 세계 30세 남성은 평균 10년간 학교를 다닌다. 같은 나이의 여성은 평균 몇 년간 학교를 다닐까?
A: 9년 B: 6년 C: 3년
7. 세계 인구 중 어떤 식으로든 전기를 공급받는 비율은 몇 퍼센트일까?
A: 20% B: 50% C: 80%



[Gapminder](https://www.gapminder.org/)

데이터와 통계

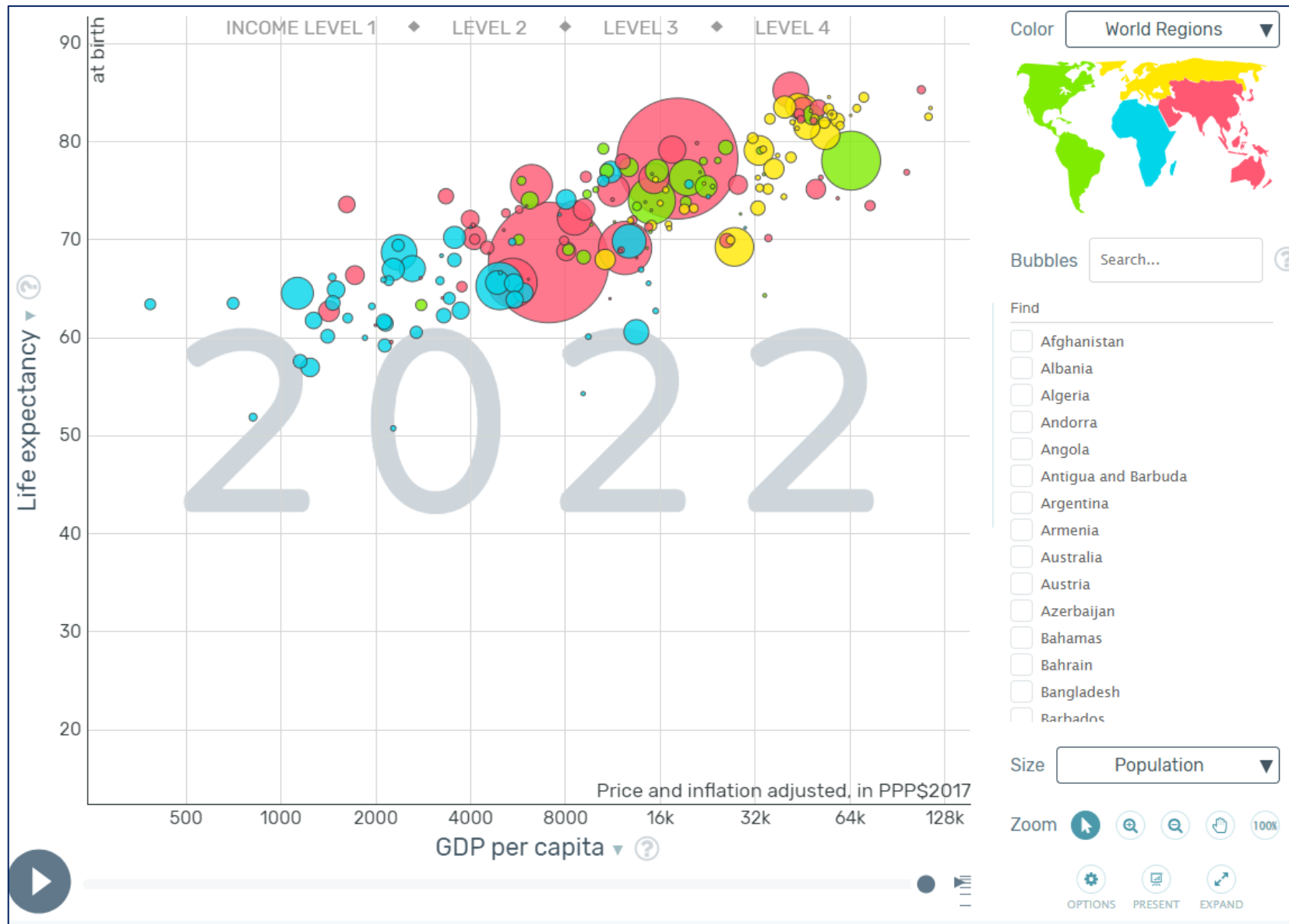


DATA SOURCES—INCOME: World Bank's GDP per capita, PPP (2011 international \$). Income of Syria & Cuba are Gapminder estimates. X-axis uses log-scale to make a doubling income show same distance on all levels. POPULATION: Data from UN Population Division. LIFE EXPECTANCY: IHME GBD-2015, as of Oct 2015. ANIMATING GRAPH: Go to www.gapminder.org/tools to see how this graph changed historically and compare 500 other indicators. LICENSE: Our charts are freely available under Creative Commons Attribution License. Please copy, share, modify, integrate and even sell them, as long as you mention: "Based on a free chart from www.gapminder.org".



Gapminder

데이터와 통계



Gapminder

통계 기본 개념

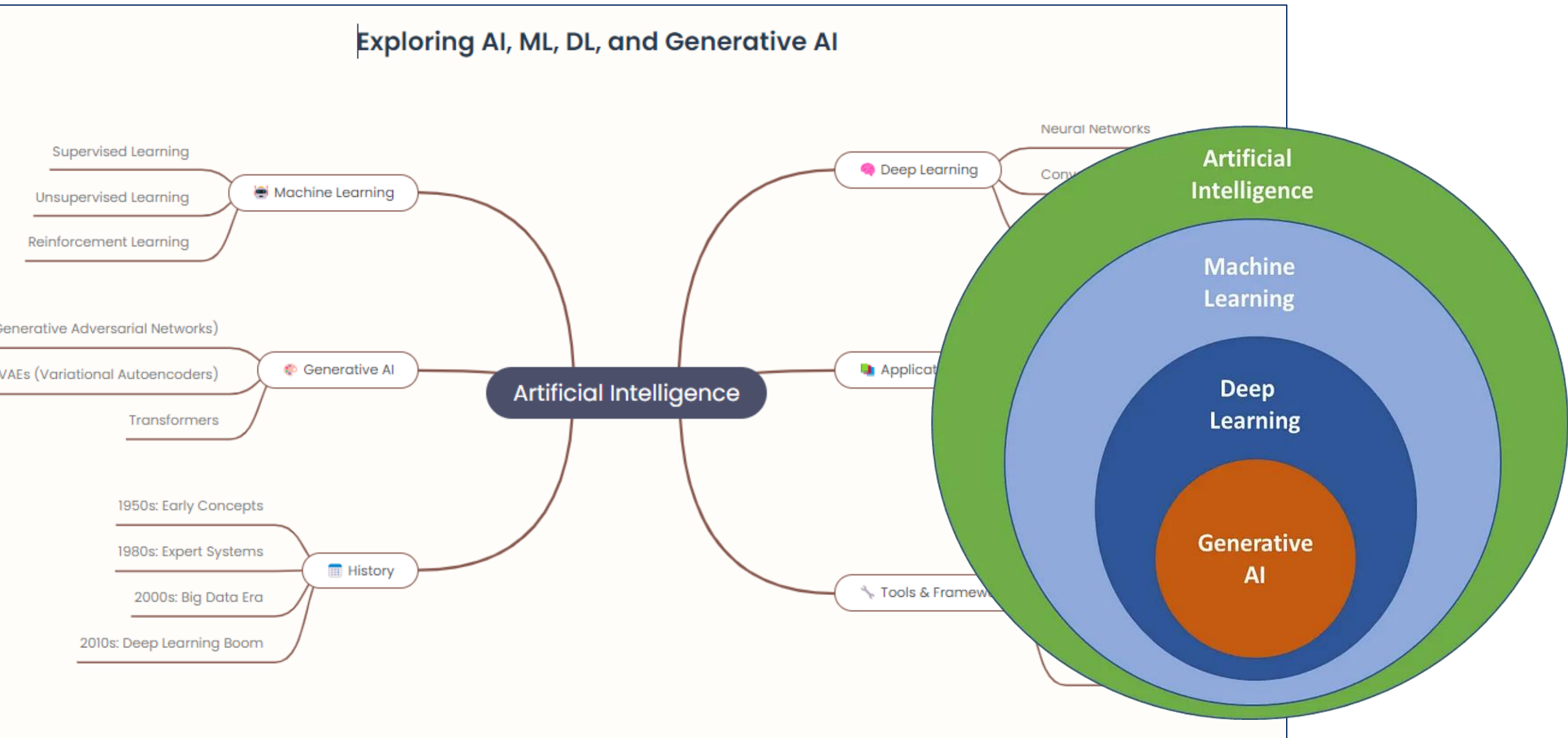
기술 통계

확률과 분포

추정과 가설 검정

상관분석

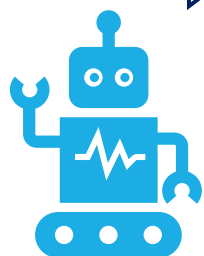
머신러닝 개요



머신러닝 개요

프로그램 기반 (기존 프로그래밍)

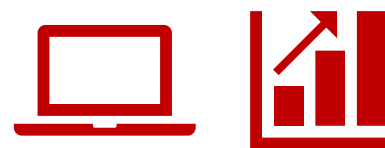
사람이 지시한
명령을 수행



- 사람이 **명시적인 프로그램을 지시**
- 기계는 프로그램에 기반하여 명령을 수행하고 결과 산출

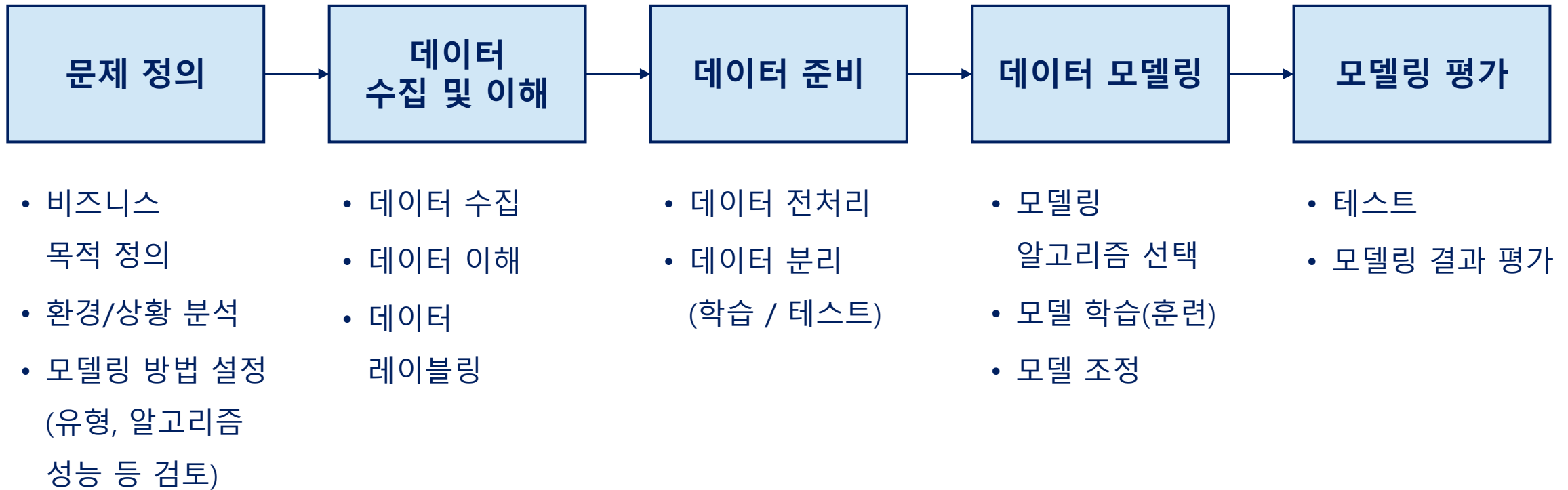
데이터 기반 (머신러닝)

데이터를
통해 학습

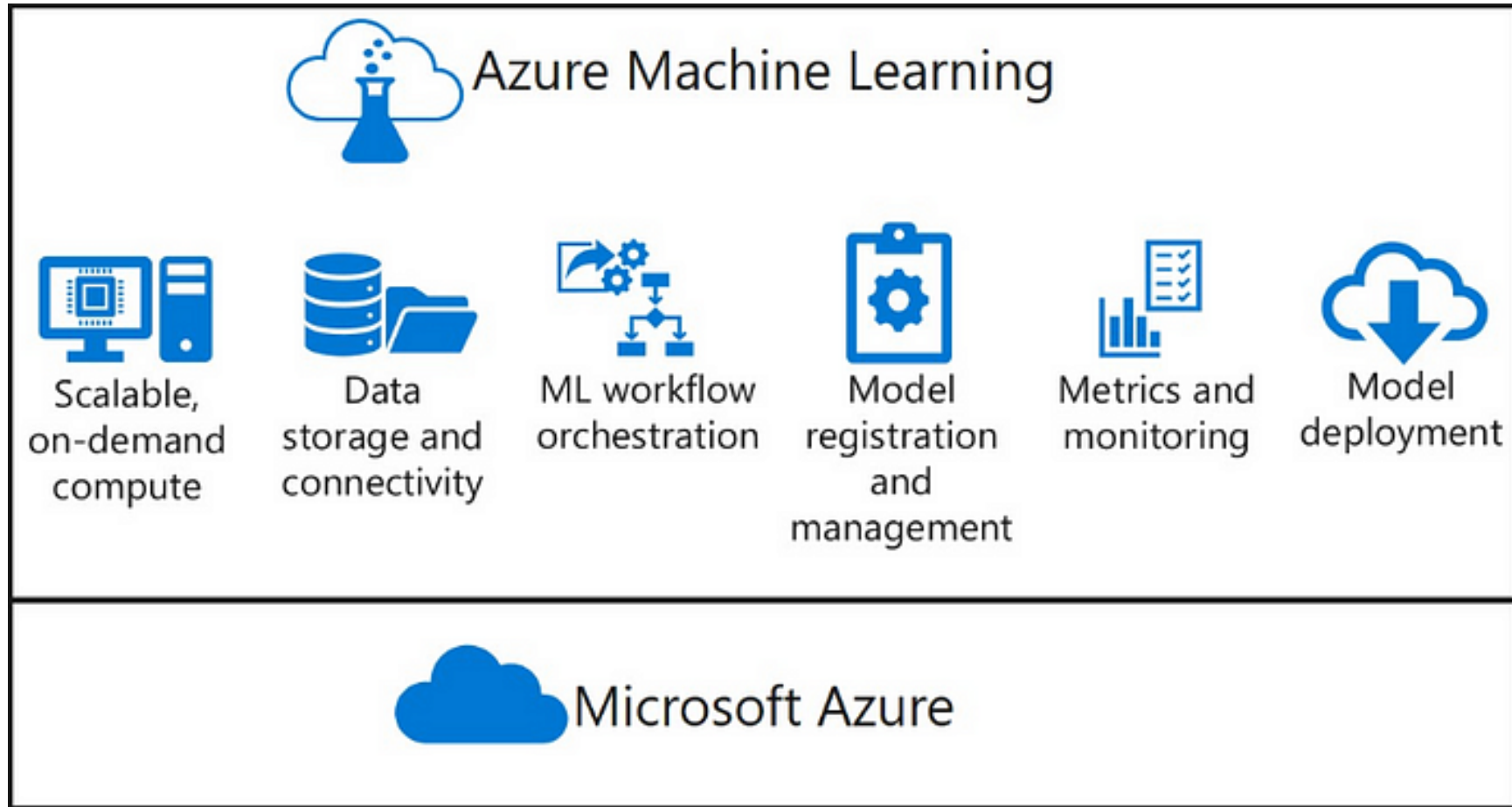


- 주어진 **데이터 속에서 규칙을 학습** (규칙적인 패턴을 발견)
- 학습한 규칙을 기반으로 **새로운 데이터에 대한 결과 예측**

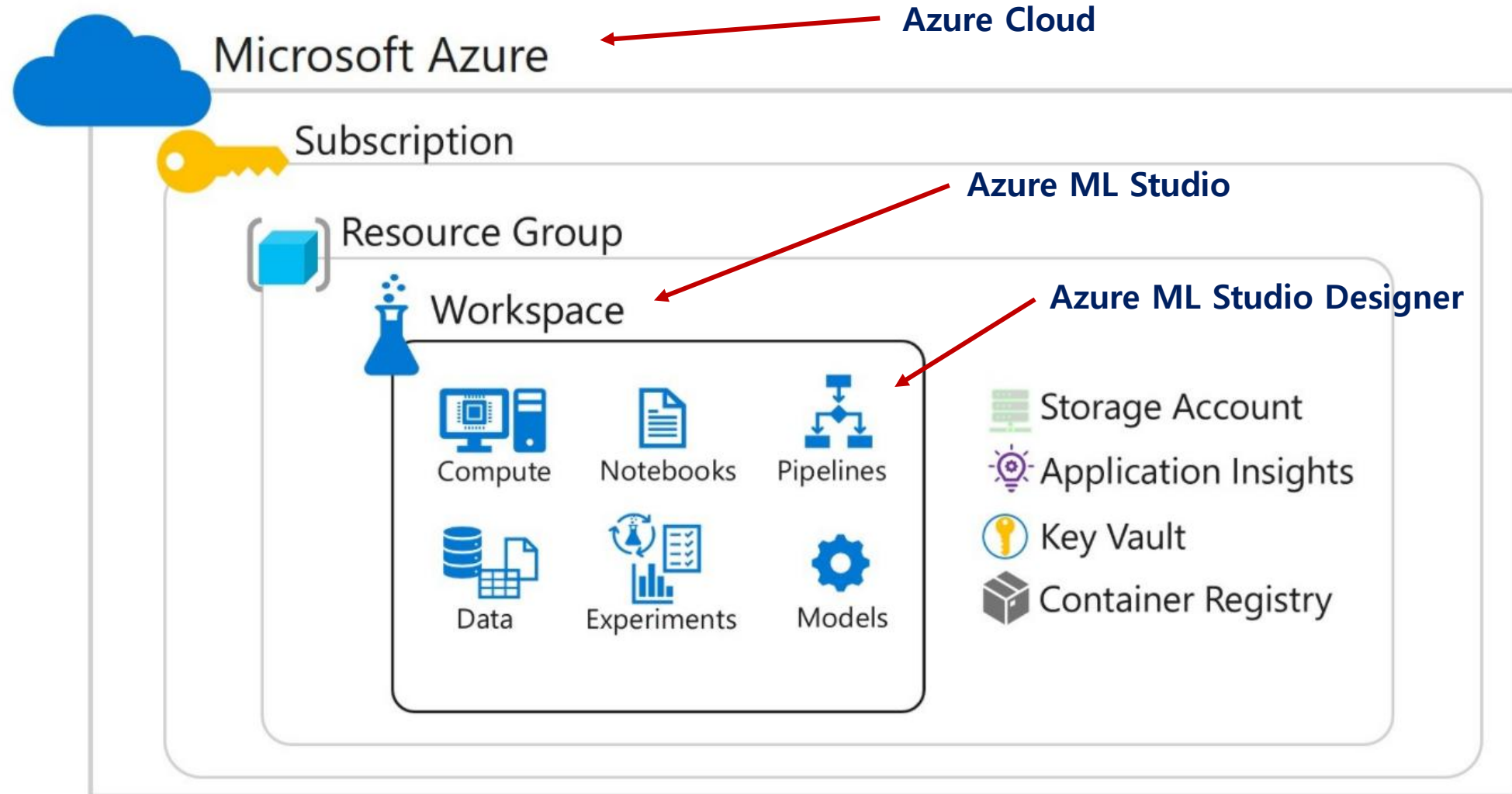
머신러닝 절차



Azure Machine Learning



Azure Machine Learning



Azure Machine Learning – 첫 화면

The screenshot displays the Azure AI Machine Learning Studio interface. The top navigation bar includes the title "Azure AI | Machine Learning Studio", a user profile for "고려대학교 W02-RocketLaunch", and several utility icons. The left sidebar contains a navigation menu with sections: "All workspaces", "Home", "Model catalog" (marked as PREVIEW), "Authoring" (containing Notebooks, Automated ML, and Designer), "Prompt flow" (marked as PREVIEW), "Assets" (containing Data, Jobs, Components, Pipelines, Environments, Models, and Endpoints), and "Manage".

The main content area is titled "W02-RocketLaunch" and features a "Generative AI with Prompt flow" section (marked as PREVIEW). Below this, there are three recommended workflows, each with a "Start" button:

- QnA with Your Own Data Using ...**: Q&A with GPT3.5 using domain knowledge from Faiss index to make the answer more grounded.
- Bring Your Own Data QnA**: Create flows for Q&A with GPT3.5 using data from your own indexed files to make the answer more grounded for enterprise chat scenarios.
- Ask Wikipedia**: Q&A with GPT3.5 using information from Wikipedia to make your answers more grounded.

Below these workflows is a "Generative AI models" section (marked as PREVIEW) with two model cards, each with a green checkmark icon:

- databricks-dolly-v2-12b**: Text generation
- openai-whisper-large**: Speech recognition

At the bottom, there is a "Notebook samples" section. The interface also includes buttons for "+ New" and "Customize view" in the top right, and "View prompt flow" and "View all" links.

Azure Machine Learning – 데이터 등록

Authoring

- 📄 Notebooks
- ⚙️ Automated ML
- 🧩 Designer
- >_ Prompt flow

Assets

📁 Data

- 🧪 Jobs
- 🧩 Components
- 🧩 Pipelines
- 🖨️ Environments
- 📦 Models
- 🔄 Endpoints

Manage

🖨️ Compute

Data

Data assets

Datastores

Dataset monitors PREVIEW

Data import PREVIEW

Data connections PREVIEW

Data assets are immutable references to your data that can be created from datastores, local files, public URLs, or Open Datasets. Data assets created with AzureML v2 cannot be deleted, but you can up-version or archive them for easy referencing and reuse in machine learning tasks. Deleting data assets created with v1 APIs will permanently delete the data asset and all metadata. [Learn more about data assets](#)

+ Create

🔄 Refresh

📁 Archive

↶ Reset view



Show latest version only



Include archived



View metadata

🔍 Search

≡ Filter

📄 Copy

Name



Source

Version

Created on ↓

Modified on

Type

adult_census_data

This workspace

1

Jun 19, 2024 8:41 PM

Jun 19, 2024 8:41 PM

Table

Azure Machine Learning – 컴퓨터

Designer

> Prompt flow

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Monitoring

Data Labeling

Compute

The "Kubernetes clusters" tab is now where you can access previous versions of "inference clusters" (also known as "AKS clusters") and "attached Kubernetes" compute types along with any previously created compute targets using those types. [Learn more](#) about Kubernetes clusters.

Compute instances

[Compute clusters](#)[Kubernetes clusters](#)[Attached computes](#)[Serverless instances](#)

Choose from a selection of CPU or GPU instances preconfigured with popular tools such as VS Code, JupyterLab, Jupyter, and RStudio, ML packages, deep learning frameworks, and GPU drivers. [Learn more about compute instances](#)

[+ New](#)[Refresh](#)[Start](#)[Stop](#)[Restart](#)[Schedule and idle shutdown](#)[...](#)[View quota](#)[Filter](#)[Columns](#)

Name



State

Idle shutdown ⓘ

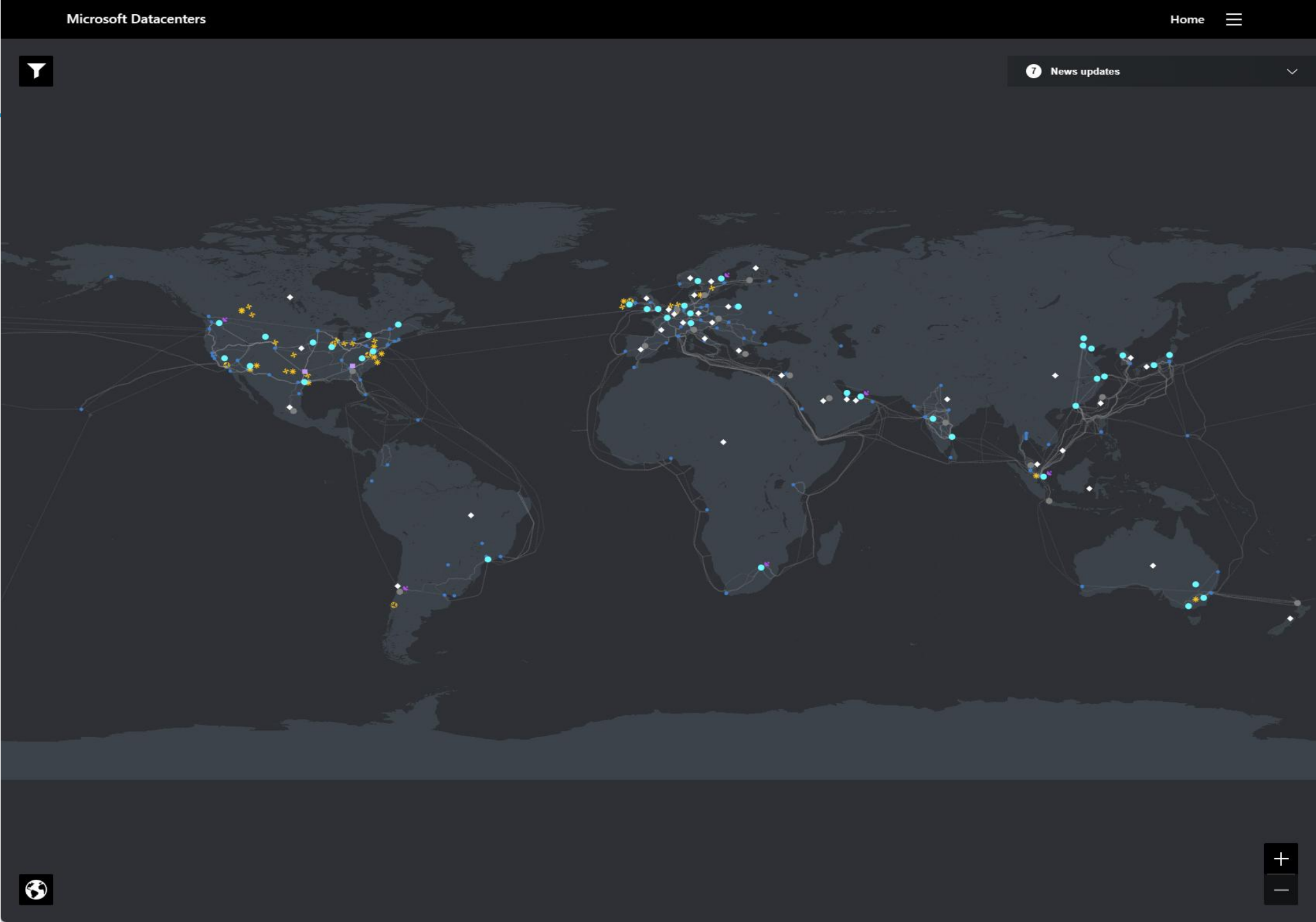
Applications ⓘ

A000Compute2

Stopped ⓘ

1 hour

JupyterLab Jupyter VS Code (Web) [PREVIEW](#)

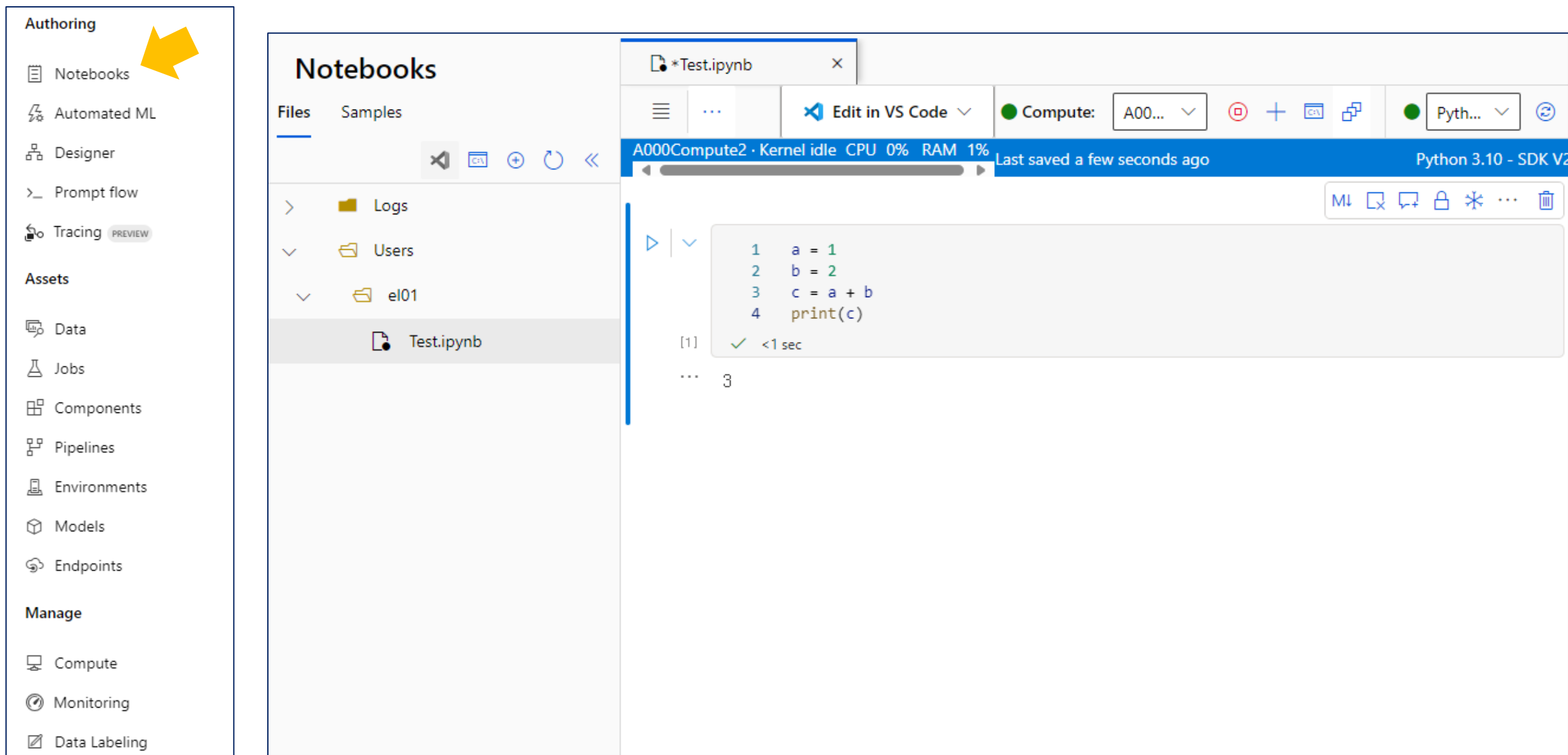


Azure Machine Learning – Designer

The screenshot displays the Azure Machine Learning Designer interface, which is used for building machine learning pipelines. The interface is divided into several sections:

- Authoring:** Contains links to Notebooks, Automated ML, Designer (highlighted with a yellow arrow), Prompt flow, and Tracing (PREVIEW).
- Assets:** Lists various assets including Data, Jobs, Components, Pipelines, Environments, Models, and Endpoints.
- Manage:** Includes links to Compute, Monitoring, and Data Labeling.
- Search and Filter:** A search bar at the top left allows searching by name, tags, and description. Below it, a filter section shows 'Tags: All' and an 'Add filter' button.
- Component Panel:** On the left, a list of components is shown under the 'Component' tab. The 'Data Transformation' section is expanded, showing components like 'Add Columns', 'Add Rows', and 'Apply Math Operation'. Each component has a description, a link to learn more, and a date.
- Pipeline Canvas:** The central area shows a pipeline titled 'Binary Classification with Feature Selection - Income Prediction'. The pipeline steps include:
 - filter_based_feature_selection:** Select 5 features by ChiSquared.
 - Two-Class Boosted Decision Tree:** A component labeled 'two_class_boosted_decision_tree'.
 - Train Model:** A component labeled 'train_model' that takes an 'Untrained model' and a 'Dataset' as input.
 - Score Model:** A component labeled 'score_model' that takes a 'Trained model' and a 'Dataset' as input.
 - Evaluate Model:** A component labeled 'evaluate_model' that takes a 'Scored dataset' and a 'Scored dataset' as input.
 - Apply Transform:** A component labeled 'apply_transformation' that takes a 'Trained model' and a 'Dataset' as input.
 - Select Columns Transform:** A component labeled 'select_columns_transform' that takes a 'Dataset with desired column...' as input.
- Configuration Panel:** On the right, a configuration panel for the 'Two-Class Boosted Deci...' model is shown. It includes settings for:
 - Create trainer mode:** SingleParameter.
 - Maximum number of leaves per tree:** 20.
 - Minimum number of samples per leaf no...:** 10.
 - Learning rate:** 0.2.
 - Number of trees constructed:** 100.
 - Random number seed:** (empty field).
 - Output settings:** (expandable).
 - Input settings:** (expandable).
 - Run settings:** (expandable).

Azure Machine Learning – Notebooks



The screenshot displays the Azure Machine Learning interface, specifically the Notebooks section. On the left sidebar, the 'Notebooks' option is highlighted with a yellow arrow. The main area shows a file explorer with folders 'Logs', 'Users', and 'el01', and a file 'Test.ipynb'. The 'Test.ipynb' file is open in a code editor, showing the following Python code:






```
1 a = 1
2 b = 2
3 c = a + b
4 print(c)
```

The output of the code execution is displayed below the code, showing a green checkmark and the text '[1] ✓ <1 sec'.








The interface also includes a top bar with the 'Notebooks' title, a 'Files' tab, and a 'Samples' tab. The 'Files' tab is active, showing the file explorer. The top bar also includes a 'Compute' section with a dropdown menu showing 'A00...' and a 'Python 3.10 - SDK V2' dropdown menu.

Azure Machine Learning – Job




Authoring

-  Notebooks
-  Automated ML
-  Designer
-  Prompt flow
-  Tracing PREVIEW

Assets

-  Data
-  Jobs
-  Components
-  Pipelines
-  Environments
-  Models
-  Endpoints

Manage

-  Compute
-  Monitoring
-  Data Labeling

Jobs


All experiments

All jobs

All schedules

 Refresh

 Archive experiment

 Reset view

☐ View archived experiments

 Search

 Columns

Experiment



Latest job

Last submitted ↓

A000-CrossValidation

cross_validation

Jun 20, 2024 8:38 PM

dataset_profile

good_bird_jcltys05

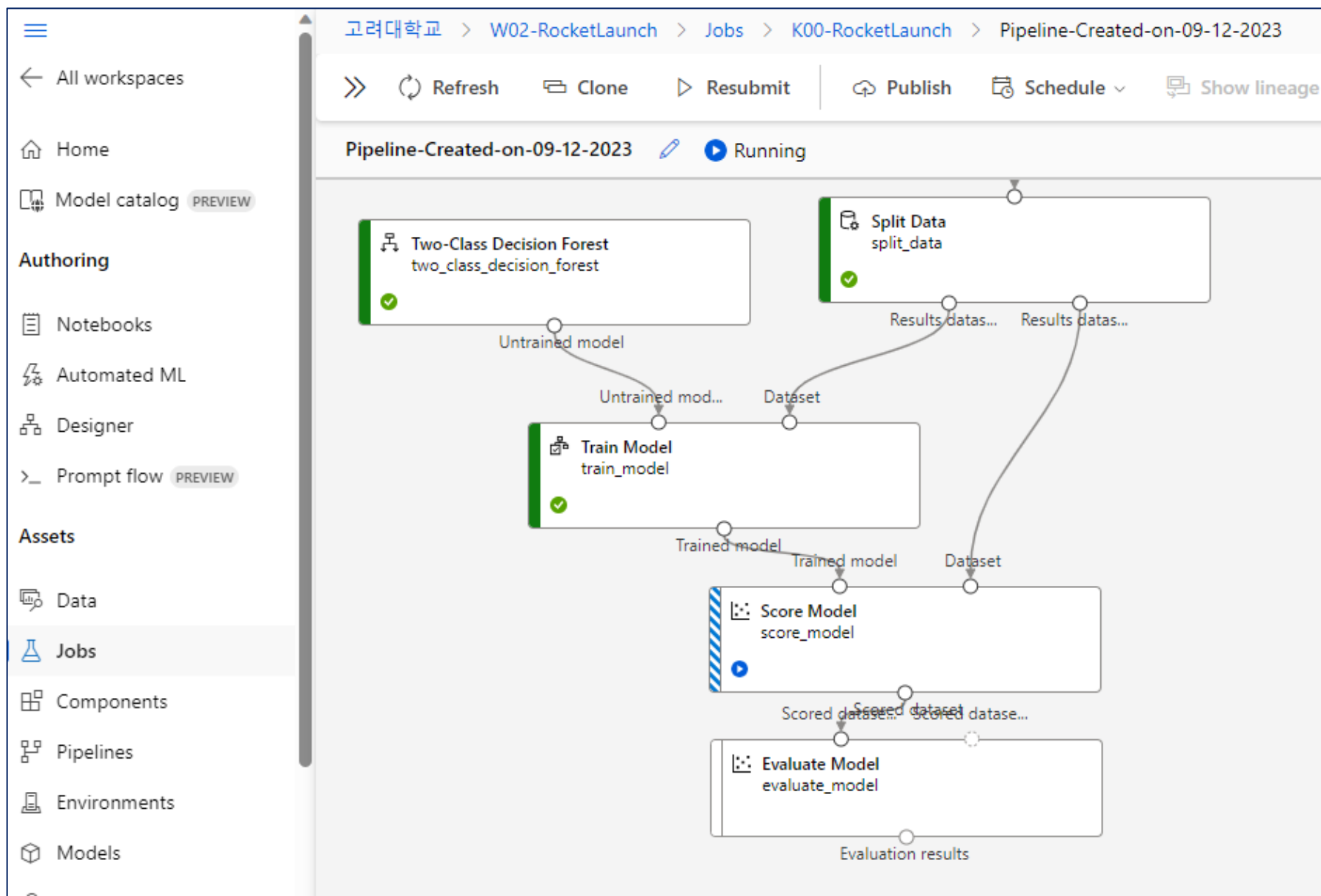
Jun 20, 2024 7:34 PM

A000-incomPrediction

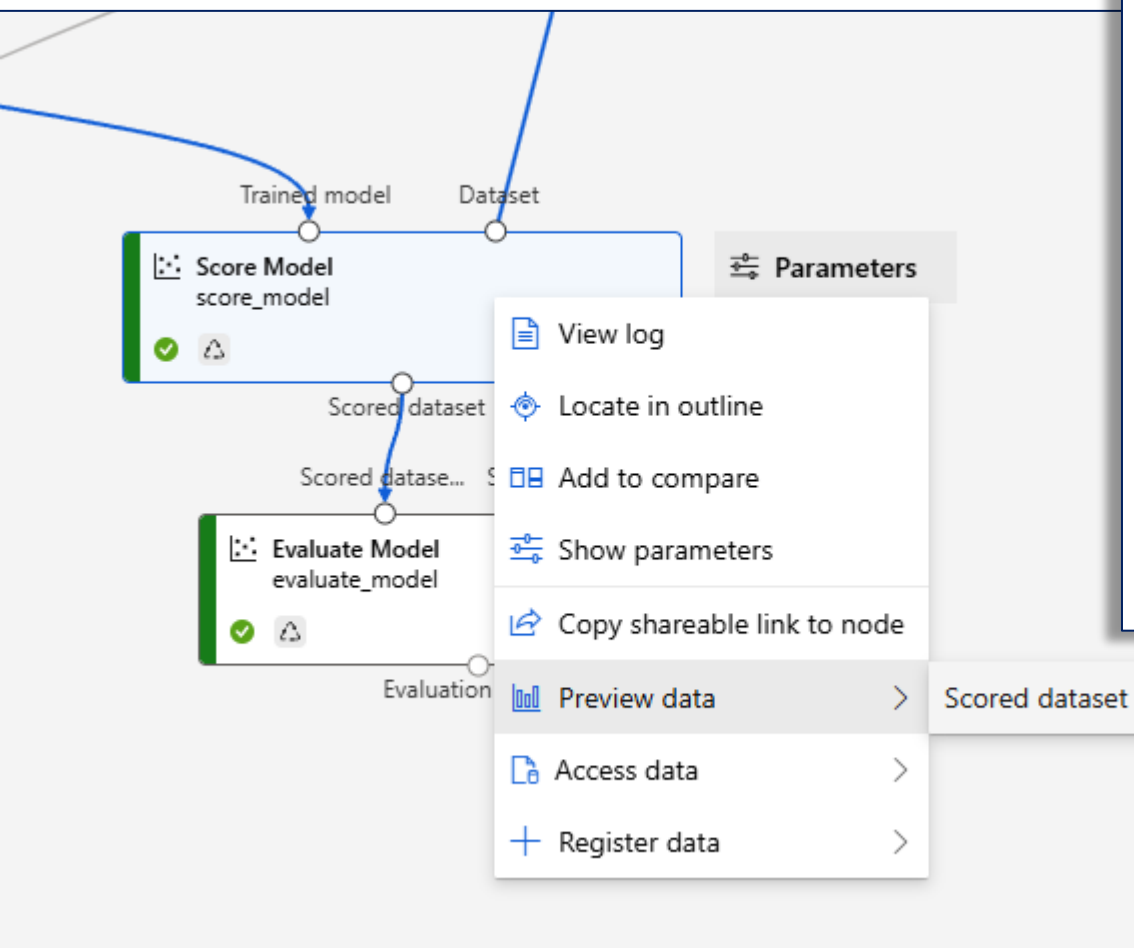
Pipeline-Created-on-06-19-2024

Jun 20, 2024 4:43 PM

Azure Machine Learning – Job



Azure Machine Learning – Job



Scored_dataset

Rows 219 Columns 36

workingday-1	workingday-1	weathersit-1	weathersit-2	weathersit-3	Scored Labels
1	1	1	0	0	208.95383
0	1	1	0	0	1760.468116
1	1	0	1	0	1052.466618
1	0	1	1	0	729.911063
1	0	1	0	0	-209.069685
1	1	0	0	0	962.069723
1	1	1	0	0	1105.020296
1	1	1	0	0	883.604318
0	1	1	0	0	1733.828867
0	0	0	1	0	1589.13771
1	1	1	0	0	1002.662357
0	0	0	1	0	1386.458338
1	0	1	1	0	-215.723783
0	1	1	0	0	471.531161
1	0	1	1	0	54.31919
0	1	1	0	0	1689.029751
1	0	1	1	0	503.282245

Scored Labels

Statistics

Mean 815.9717

Median 845.0384

Min -860.0815

Max 2099.8091

Standard deviation 583.376

Unique values 219

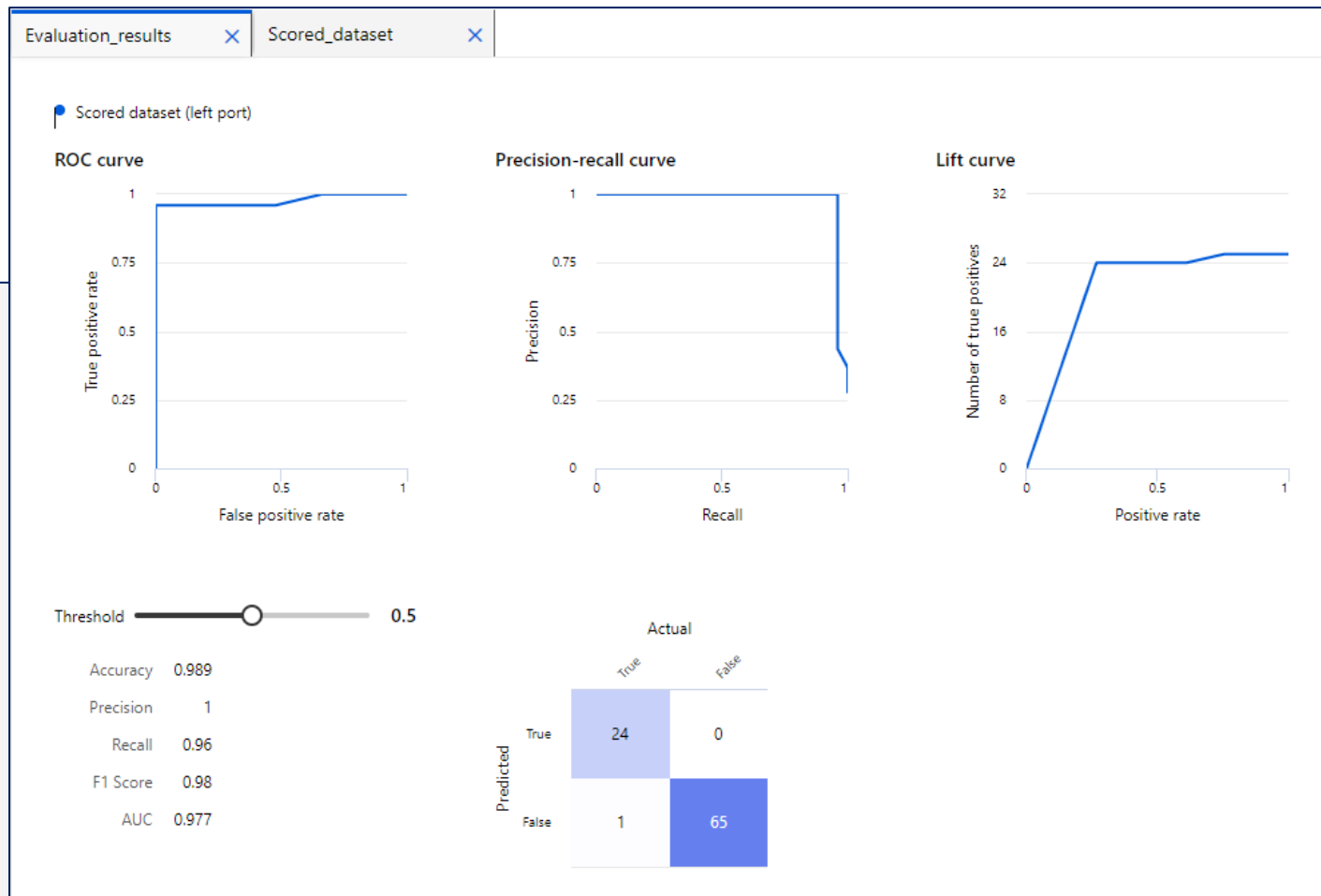
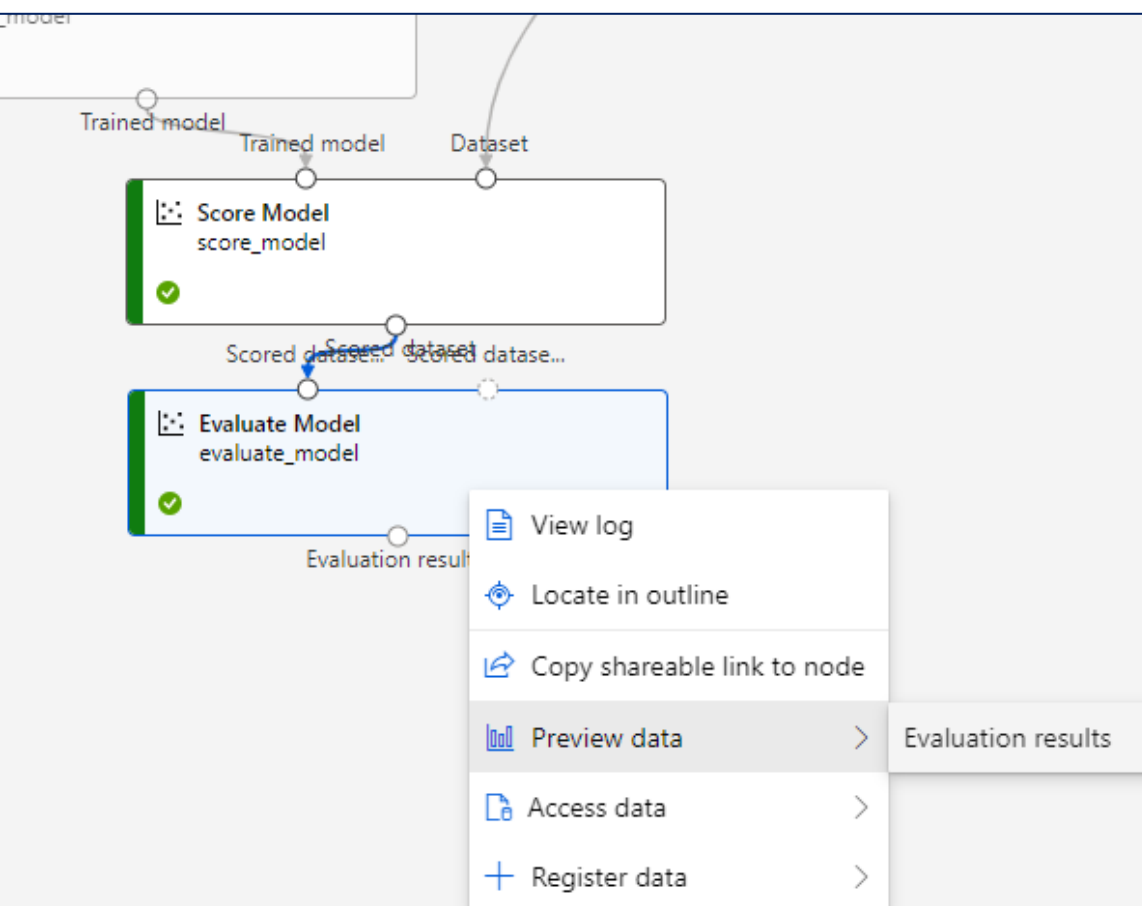
Missing values 0

Feature type Numeric S

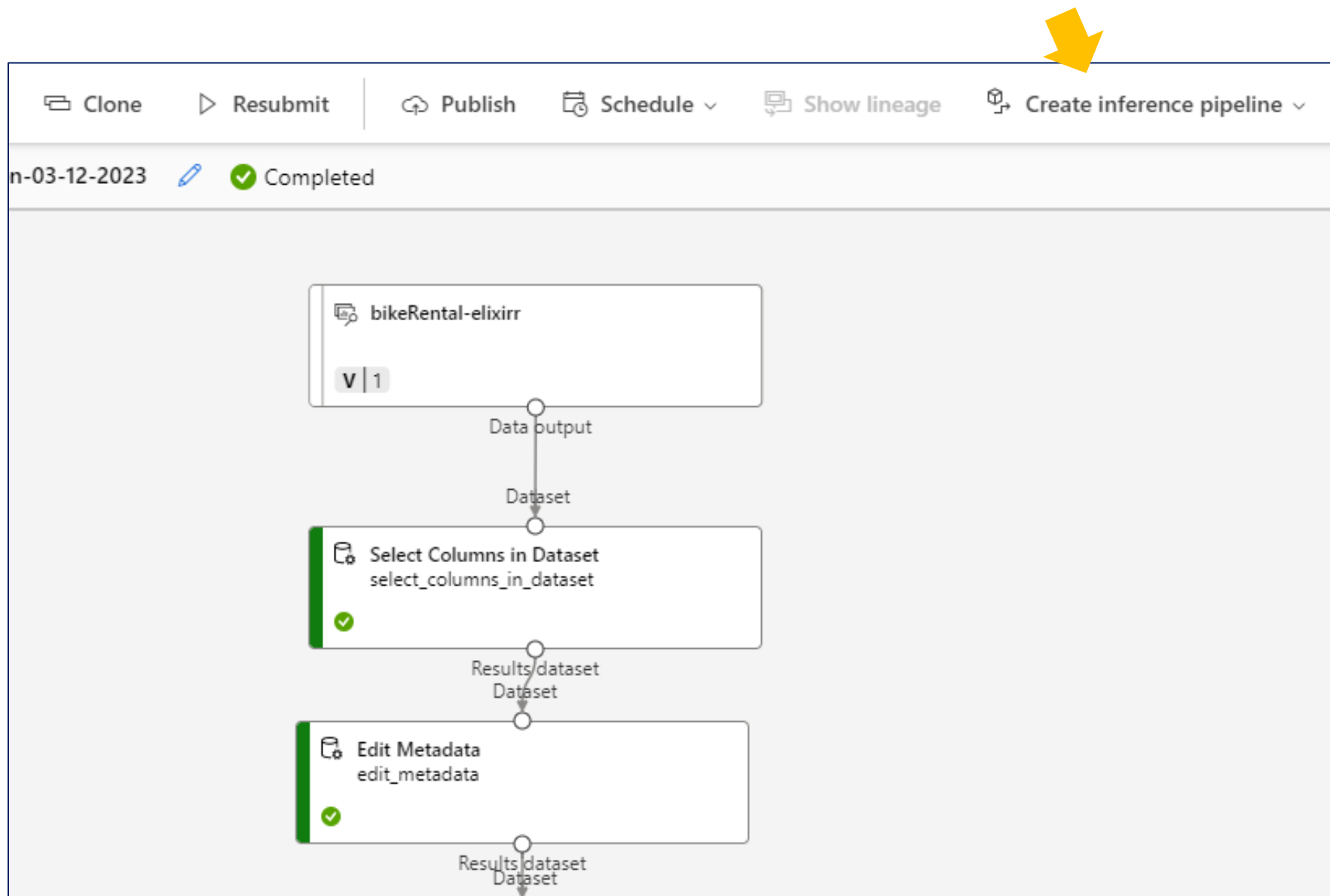
Visualizations

The histogram shows the frequency distribution of the 'Scored Labels'. The x-axis represents the 'Scored Labels' values, ranging from approximately -860 to 2100. The y-axis represents the 'Frequency', ranging from 0 to 50. The distribution is roughly bell-shaped, centered around 800-900, with a peak frequency of approximately 50.

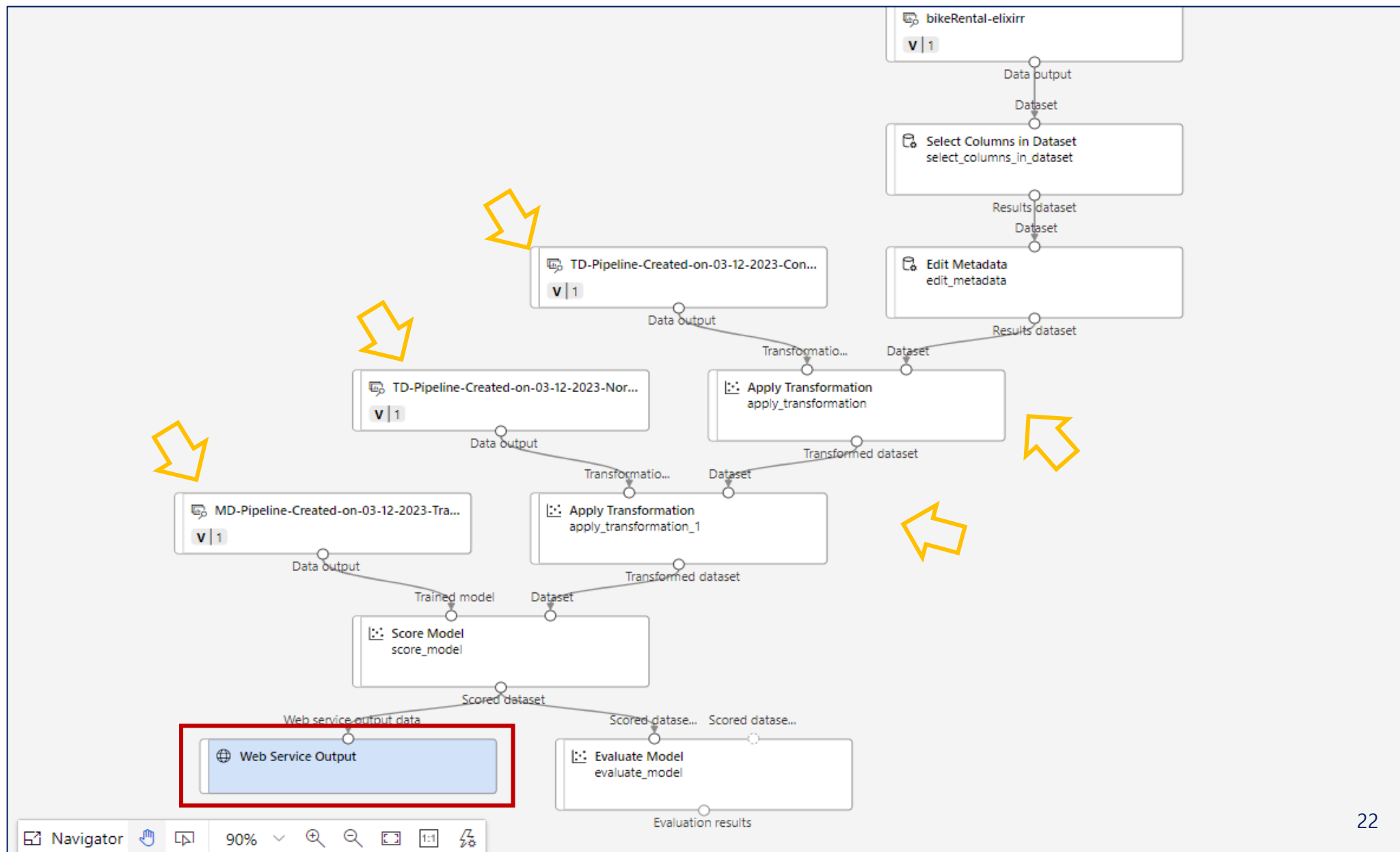
Azure Machine Learning – Job



Azure Machine Learning – 배포 및 유추



Azure Machine Learning – 배포 및 유추



Azure Machine Learning – 배포 및 유추

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The left sidebar contains a navigation menu with the following items: Microsoft, New, Home, Author, Notebooks, Automated ML, Designer, Assets, Data, Jobs, Components, Pipelines, Environments, Models, and Endpoints. The 'Endpoints' item is highlighted with a red box. The main content area shows the breadcrumb 'Microsoft > sdg-ws > Endpoints' and the title 'Endpoints'. Below the title, there are two tabs: 'Real-time endpoints' and 'Pipeline endpoints', with the latter being selected and highlighted with a red box. Below the tabs, there are controls for 'Refresh', 'Disable', 'Enable', and a toggle for 'View disabled'. A search icon and 'Sea' are also visible. A table lists the endpoints with columns for Name, Description, Modified on, and Modified by. The table contains one entry: 'My_New_Pipeline' with description 'My Published Pipeline D...' and modified on '11/11/2019, 3:41:03 PM'. Navigation links '< Prev' and 'Next >' are at the bottom of the table.


Microsoft Azure Machine Learning Studio

Microsoft > sdg-ws > Endpoints

Endpoints

Real-time endpoints **Pipeline endpoints**

Refresh Disable Enable | View disabled

Name ↓	Description	Modified on	Modified by
 My_New_Pipeline	My Published Pipeline D...	11/11/2019, 3:41:03 PM	

< Prev Next >

Azure Machine Learning – 배포 및 유추

Home

Model catalog PREVIEW

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

k00-bike-rental-inference ☆

Details Test **Consume** Deployment logs

Basic consumption info

REST endpoint

<http://2e3e01c8-d658-4a79-aa91-6d496a794899.koreacentral.azurecontainer.io/score>

Authentication

Primary key

.....



[Regenerate](#)

Secondary key

.....



[Regenerate](#)

Consumption option

Consumption types

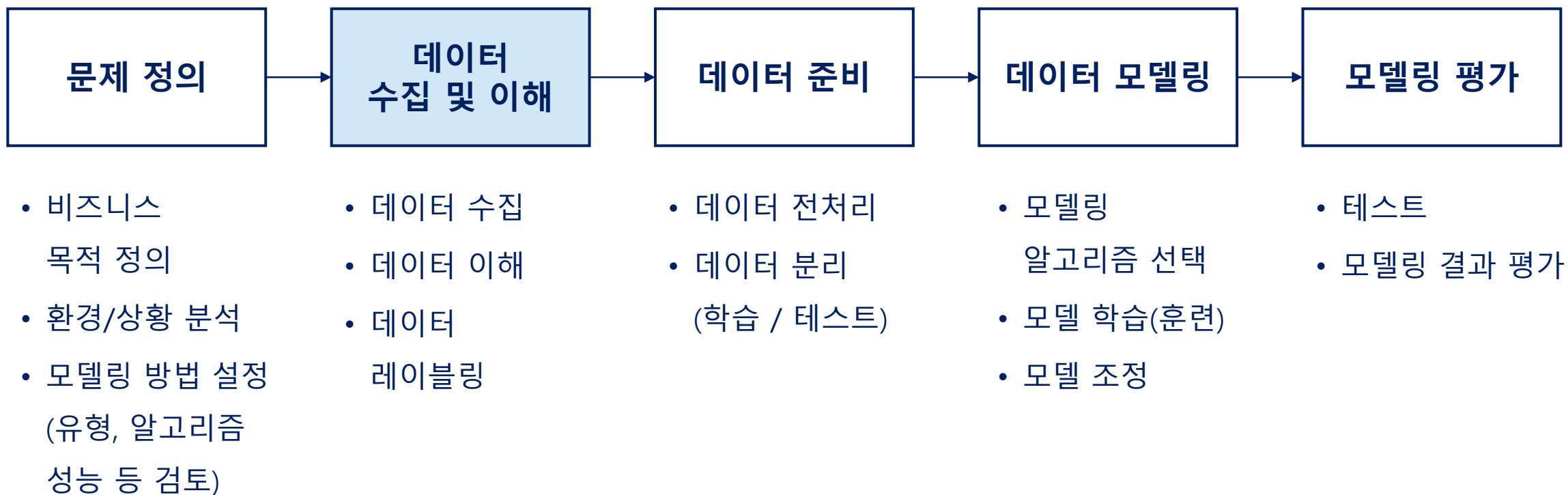
Python

C#

R

```
1 import urllib.request
2 import json
3 import os
4 import ssl
5
6 def allowSelfSignedHttps(allowed):
7     # bypass the server certificate verification on client side
8     if allowed and not os.environ.get('PYTHONHTTPSVERIFY', '') and getattr(ssl, '_create_unverified_context', None):
```


Azure Machine Learning – Designer : 머신러닝 절차



Azure Machine Learning – Designer : 데이터


Azure AI | Machine Learning Studio

Create data asset

1 Data type

2 Data source


Set the name and type for your data asset

Name * 

bike_rental_data *

Description

자전거 렌탈 데이터 for 실습

Type * 

Tabular

Back Next

Azure Machine Learning – Designer : 데이터

Azure AI | Machine Learning Studio

Create data asset

✓ Data type

✓ Data source

3 Destination storage type

4 File or folder selection

5 Settings

6 Schema

7 Review

Select a datastore

Choose a storage type and a datastore to upload your data to in the next step. You can also create a new datastore

Datastore type *

Azure Blob Storage

Create new datastore

Search datastore

Name ↓	Storage name
workspaceblobstore	w03bikerental4950615055
workspaceartifactstore	w03bikerental4950615055

<<

<

Page 1 of 1

>

>>

25/Page

Back

Next

27

Azure Machine Learning – Designer : 데 이 터

Create data asset

✓ Data type

✓ Data source

✓ Destination storage type

✓ File or folder selection

✓ Settings

6 Schema

7 Review

Schema
Column types are auto-detected based on the initial subset of the data and can be updated here. Values not aligning with the specified column type will fail conversion. Values would be either null-filled or replaced with error value. Any conversions preview errors are non-blocking and you can proceed.

Include	Column name	Type	Example values	Date format ⓘ	Properties ⓘ
<input type="checkbox"/>	Path	String ▾		Not applicable to s...	Not applicab
<input checked="" type="checkbox"/>	day	Integer ▾	1, 2, 3	Not applicable to s...	Not applicab
<input checked="" type="checkbox"/>	mnth	Integer ▾	1, 1, 1	Not applicable to s...	Not applicab
<input checked="" type="checkbox"/>	year	Integer ▾	2011, 2011, 2011	Not applicable to s...	Not applicab
<input checked="" type="checkbox"/>	season	Integer ▾	1, 1, 1	Not applicable to s...	Not applicab
<input checked="" type="checkbox"/>	holiday	Integer ▾	0, 0, 0	Not applicable to s...	Not applicab
<input checked="" type="checkbox"/>	weekday	Integer ▾	6, 0, 1	Not applicable to s...	Not applicab

Back

Next

Azure Machine Learning – Designer : 데이터

The screenshot displays the Azure Machine Learning Designer interface. The top toolbar includes icons for navigation and actions like Undo, Redo, Validate, Show lineage, Clone, and a blue 'Configure' button. Below the toolbar, a search bar and filter controls are visible. The main workspace shows a pipeline with a single component named 'bike_rental_data'. The component's output is labeled 'Data output'. On the right, a sidebar provides details for the selected component, including its parameters and outputs.

Search by name, tags and description

Tags: **All** [Add filter](#)

Data Component

1 [Refresh](#) [Add](#) Last update... [Dropdown](#) [Sort](#)

You can find the prebuilt sample data under Component tab. [Click here](#)

bike_rental_data Version 1

Myeongho Kang
고려대 자전거 렌탈 데이터
9/19/2023

Pipeline-Created-on-09-19-2023 [Edit](#)

bike_rental...

Parameters **Outputs**

Data name
[bike_rental_data](#)

ID
c881fb67-a759-49c2-83fb-f074aa467601

Data type
Tabular

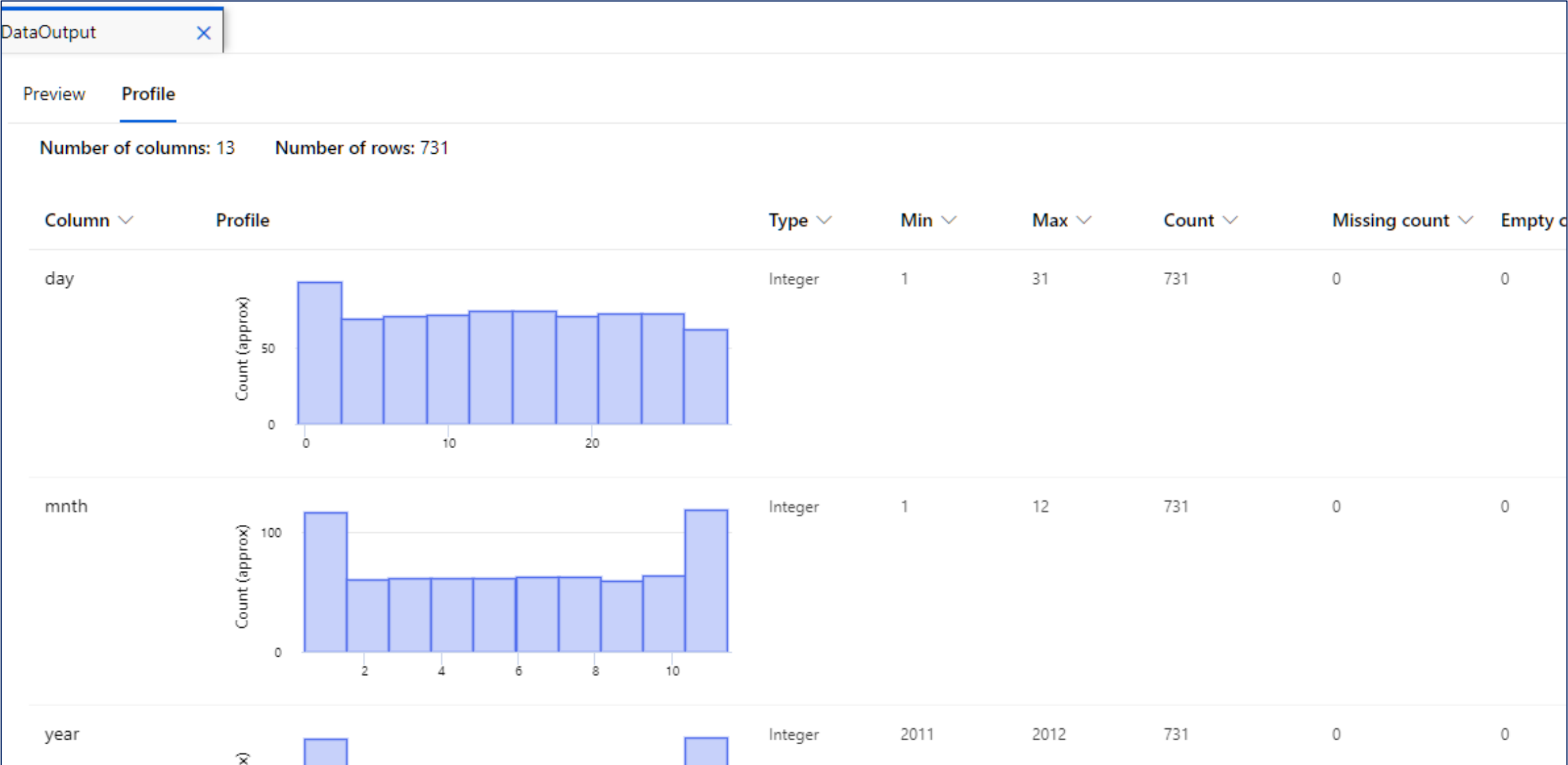
Description
고려대 자전거 렌탈 데이터

Datastore name
[workspaceblobstore](#)

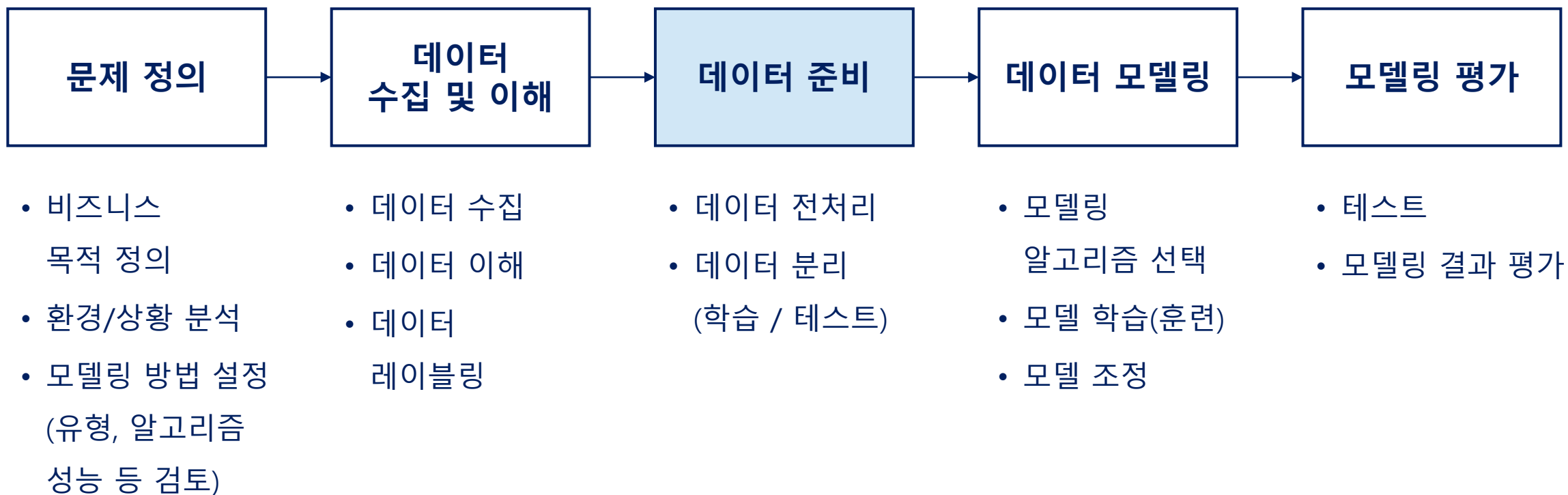
Relative path
UI/2023-09-19_053303_UTC/03 daily-bike-s

Created time
Sep 19, 2023 2:46 PM

Azure Machine Learning – Designer : 데이터



Azure Machine Learning – Designer : 머신러닝 절차



Azure Machine Learning – Designer : 데이터 준비

The screenshot displays the Azure Machine Learning Designer interface for a pipeline named "Pipeline-Created-on-09-19-2023". The interface is divided into several sections:

- Top Bar:** Includes navigation links (고려대학교 > W03-bikeRental > Designer > Authoring), action buttons (Undo, Redo, Validate, Show lineage, Clone), and a "Configure & Submit" button.
- Left Panel:** Contains a search bar with "select columns in", a "Tags: All" filter, and a "Component" tab. Under the "Component" tab, three components are listed: "Select Columns in Dataset" (Microsoft), "Join Data" (Microsoft), and "Select Columns Transform" (Microsoft). A yellow arrow points from the "Select Columns in Dataset" component in the list to its instance in the pipeline.
- Center Canvas:** Shows a pipeline diagram with two "Select Columns in Dataset" components. The first component takes "bike_rental_data" as input and outputs a "Data output". This output is connected to the second component, which also takes "bike_rental_data" as input and outputs a "Results dataset".
- Right Panel:** Displays the configuration for the selected "Select Columns in Dataset" component. It includes a "Select columns" field with an information icon and a red asterisk, indicating a required field. Below this field is a message: "A value is required." The panel also has tabs for "Parameters", "Output settings", "Input settings", "Run settings", "Node information", and "Component information".

Azure Machine Learning – Designer : 데이터 준비

The screenshot displays the Azure Machine Learning Designer interface for creating a data preparation pipeline. The interface is divided into three main sections: a left sidebar for component selection, a central canvas for the pipeline, and a right sidebar for component configuration.

Left Sidebar (Component Selection):

- Tags: All, Add filter
- Component tab is selected.
- Search bar: 1, +
- Sort: Most relevant
- Component: **Normalize Data** (Microsoft)
- Description: Rescales numeric data to constrain dataset values to a standard range. [Learn More](https://aka.ms/aml/...)
- Metadata: azureml.Designer:true, 1/13/2023

Central Canvas (Pipeline):

- Pipeline title: Pipeline-Created-on-09-19-2023
- Actions: Save, Pipeline interface
- Pipeline steps:
 - Dataset (Input)
 - Edit Metadata (edit_metadata)
 - Results dataset Dataset
 - Convert to Indicator Values (convert_to_indicator_values)
 - Results dataset Dataset, Indicator val...
 - Normalize Data (normalize_data)** (Highlighted with a yellow arrow)
 - Transformed d..., Transformatio...

Right Sidebar (Component Configuration):

- Component: **Normalize Data**
- Transformation method: ZScore
- Use 0 for constant columns when checked: True
- Columns to transform: Column names: temp,atemp,hum,windspeed
- Output settings: >

Azure Machine Learning – Designer : 데이터 준비

The screenshot displays the Azure Machine Learning Designer interface for a pipeline titled "Pipeline-Created-on-09-19-2023". The pipeline consists of three main components:

- bike_rental_data**: A dataset component that outputs a "Data output Dataset".
- Select Columns in Dataset**: A component that takes the "Data output Dataset" as input and outputs a "Results dataset Dataset".
- Edit Metadata**: A component that takes the "Results dataset Dataset" as input and outputs a "Results dataset". This component is highlighted in blue and has a yellow arrow pointing to it.

The left sidebar shows the search results for "edit metadata", listing the "Edit Metadata" component by Microsoft. The right sidebar shows the configuration for the "Edit Metadata" component, which is currently set to "Unchanged" for all fields.

Edit Metadata Configuration:

- Column: (Required)
- Data type: (Required)
- Categorical: (Required)
- Fields: (Required)
- New column names:

Azure Machine Learning – Designer : 데이터 준비

The screenshot displays the Azure Machine Learning Designer interface for creating a data preparation pipeline. The interface is divided into three main sections: a left sidebar for component selection, a central canvas for the pipeline, and a right sidebar for component configuration.

Left Sidebar (Component Selection):

- Search bar: `split data`
- Tags: `All`, `Add filter`
- Tab: `Component` (selected)
- Sort: `Most relevant`
- Components listed:
 - Split Data** (Microsoft): Partitions the rows of a dataset into two distinct sets. [Learn More](https://aka.ms/aml/split-data). `azureml.Designers: true`. 1/13/2023.
 - Apply SQL Transformation** (Microsoft): Runs a SQLite query on input datasets to transform the data. [Learn More](https://aka.ms/aml/apply-sq...). `azureml.Designers: true`. 1/13/2023.

Central Canvas (Pipeline):

The pipeline is titled `Pipeline-Created-on-09-19-2023`. It consists of the following steps:

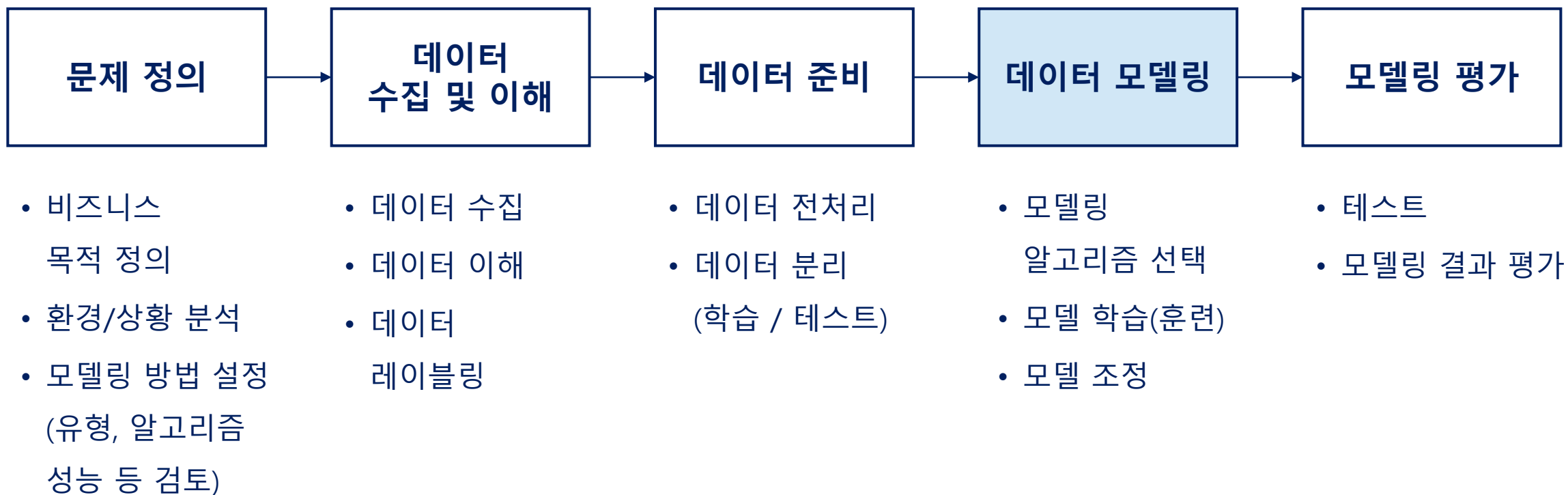
- Results dataset Dataset** (Input)
- Convert to Indicator Values** (`convert_to_indicator_values`)
- Results dataset Dataset** (Intermediate output)
- Indicator val... Dataset** (Intermediate output)
- Normalize Data** (`normalize_data`)
- Transformed d... Dataset** (Intermediate output)
- Split Data** (`split_data`) - This component is highlighted with a yellow arrow.
- Results datas...** (Final output)

Right Sidebar (Split Data Configuration):

Split Data

- Splitting mode: `Split Rows`
- Fraction of rows in the first output: `0.5`
- Randomized split: `True`
- Random seed: `0`
- Stratified split: `False`

Azure Machine Learning – Designer : 머신러닝 절차



Azure Machine Learning – Designer : 모델링

Virtual machine type ⓘ

☒ CPU ☐ GPU

Virtual machine size ⓘ

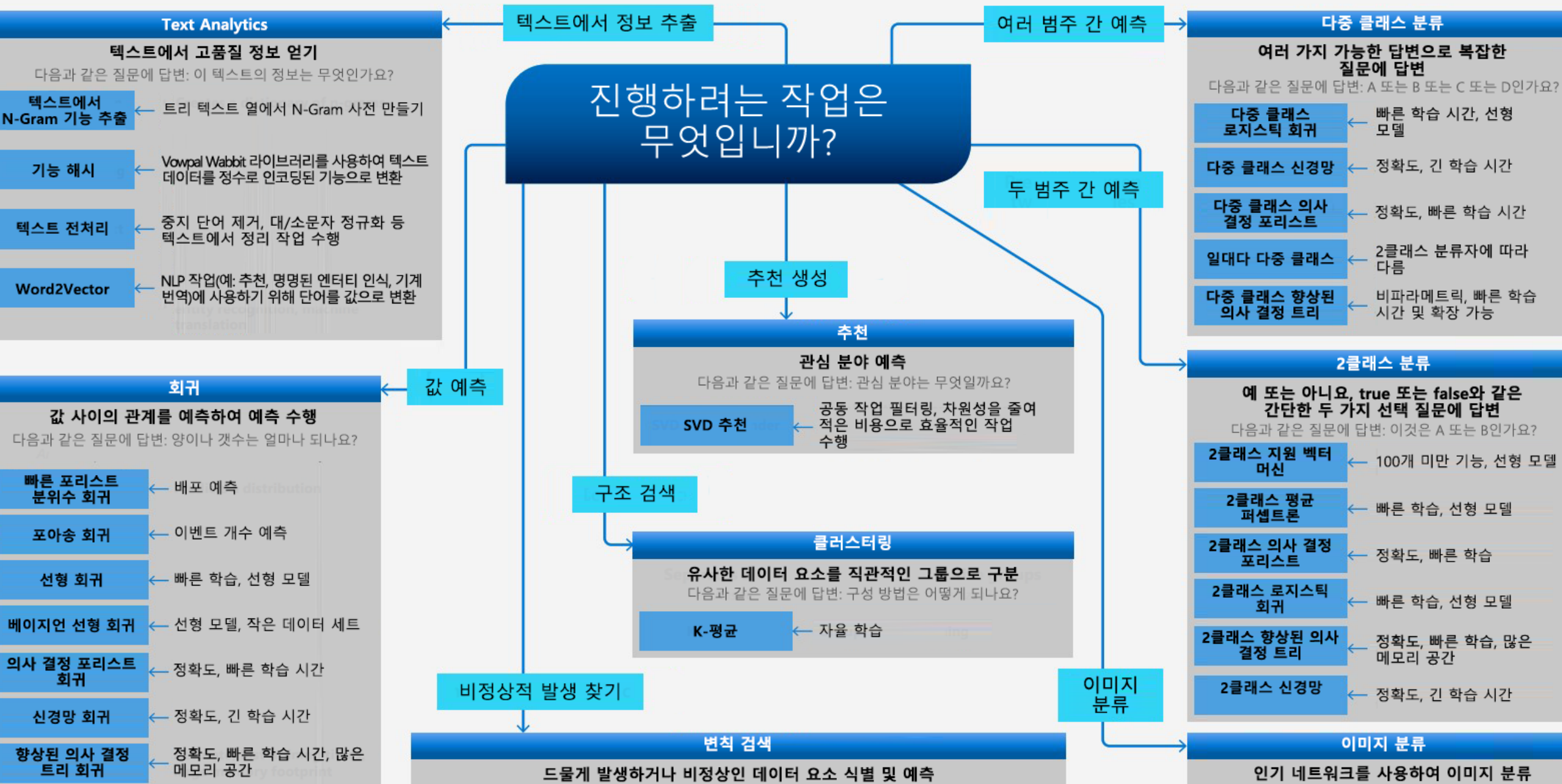
☐ Select from recommended options ☒ Select from all options

+ Add filter

Search by VM name...

Showing 200 VM sizes | Current selection: Standard_E4ds_v4

	Name ↑	Category	Available quota ⓘ	Cost ⓘ
<input type="radio"/>	Standard_E16ds_v4 16 cores, 128GB RAM, 600GB storage	Memory optimized	194 cores	\$1.38/hr
<input type="radio"/>	Standard_E20ds_v4 20 cores, 160GB RAM, 750GB storage	Memory optimized	194 cores	\$1.73/hr
<input type="radio"/>	Standard_E2ds_v4 2 cores, 16GB RAM, 75GB storage	Memory optimized	194 cores	\$0.17/hr
<input type="radio"/>	Standard_E32ds_v4 32 cores, 256GB RAM, 1200GB storage	Memory optimized	194 cores	\$2.77/hr



Azure Machine Learning – Designer : 모델링

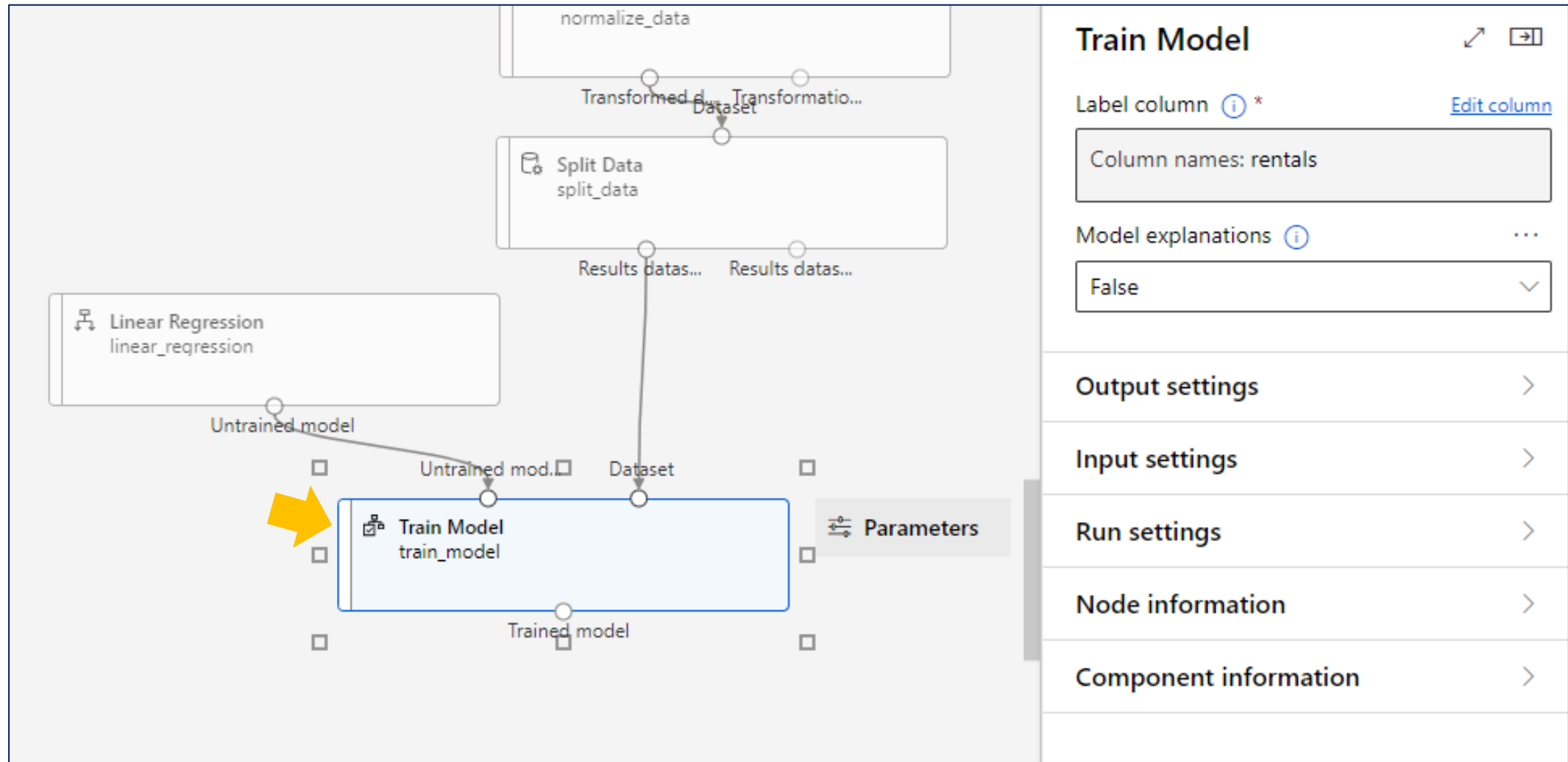
The screenshot displays the Azure Machine Learning Designer interface. On the left, a workflow is shown with three main steps: 'Convert to indicator values', 'Normalize Data', and 'Split Data'. The 'Convert to indicator values' step outputs 'Results datas...' and 'Indicator val... Dataset'. The 'Normalize Data' step outputs 'Transformed d...' and 'Transformatio... Dataset'. The 'Split Data' step outputs 'Results datas...' and 'Results datas...'. A yellow arrow points to the 'Linear Regression' model icon in the bottom left corner, which is labeled 'linear_regression' and 'Untrained model'.

On the right, the 'Linear Regression' settings panel is visible. It includes the following configuration options:

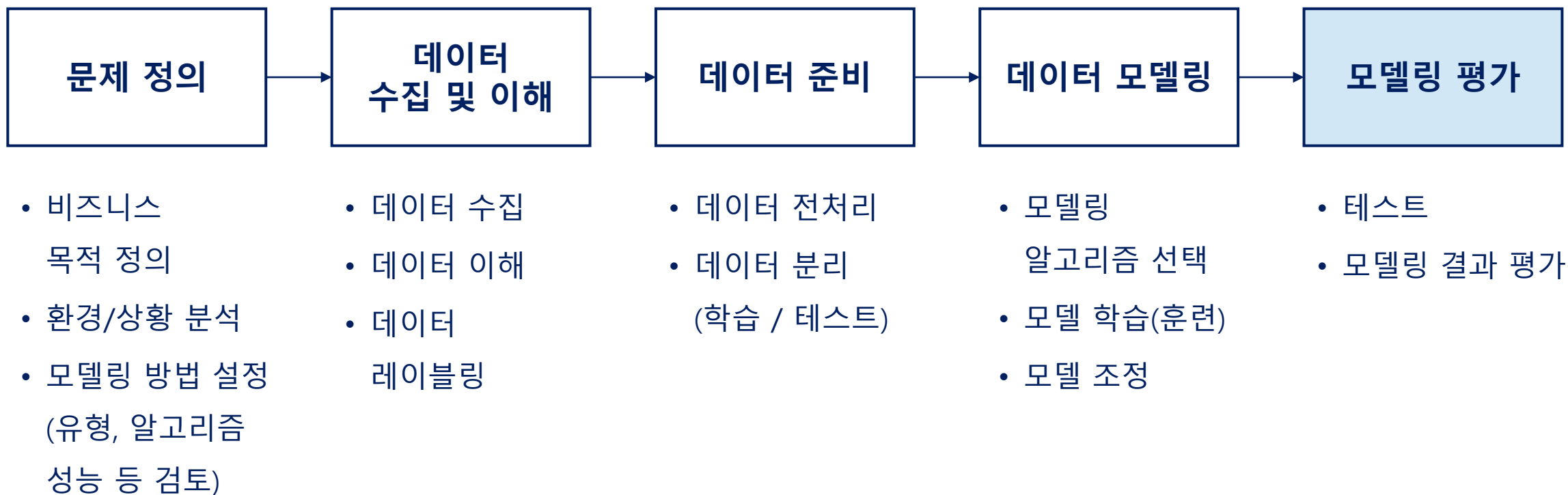
- Solution method**: Ordinary Least Squares
- L2 regularization weight**: 0.001
- Include intercept term**: True
- Random number seed**: (empty field)

Below these settings, there are three expandable sections: 'Output settings', 'Input settings', and 'Run settings'.

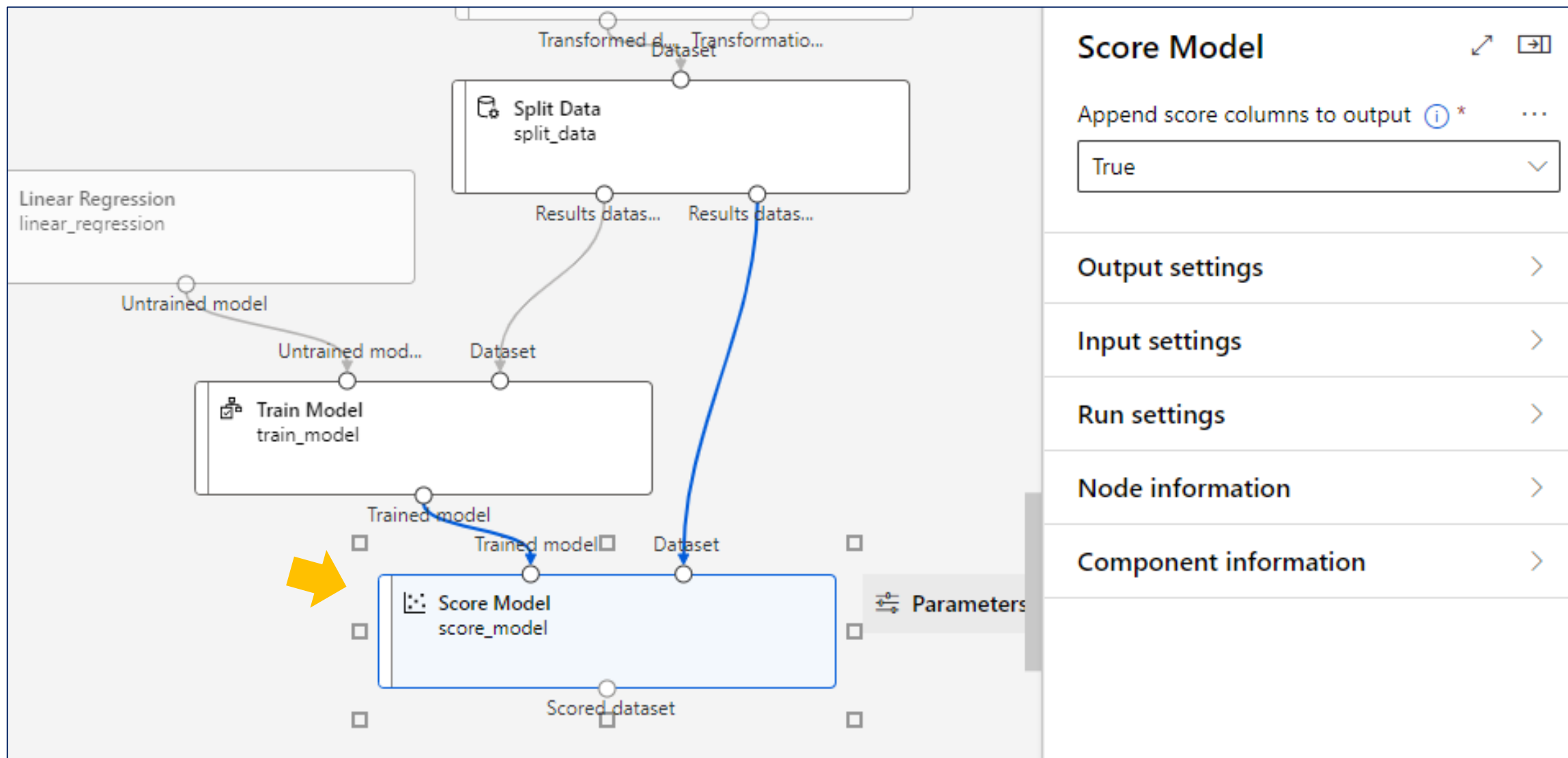
Azure Machine Learning – Designer : 모델링



Azure Machine Learning – Designer : 머신러닝 절차



Azure Machine Learning – Designer : 평가



Azure Machine Learning – Designer : 평가

The screenshot displays the Azure Machine Learning Designer interface. At the top, there are controls for 'Clone' and 'AutoSave' (a toggle switch), and a blue 'Configure & Submit' button. Below these are 'Save' and 'Pipeline interface' options. The main workspace shows a pipeline diagram. A 'train_model' node is at the top left. Its output, 'Trained model', connects to the 'Score Model' node (labeled 'score_model'). The 'Score Model' node also takes a 'Dataset' as input. Its output, 'Scored dataset', connects to the 'Evaluate Model' node (labeled 'evaluate_model'). A yellow arrow points to the 'Evaluate Model' node. The 'Evaluate Model' node's output is 'Evaluation results'. On the right side, there is a panel for the 'Evaluate Model' node, which currently shows 'No parameter'. Below this are expandable sections for 'Output settings', 'Input settings', 'Run settings', 'Node information', and 'Component information'.

Clone AutoSave **Configure & Submit**

Save Pipeline interface

train_model

Trained model

Trained model Dataset

Score Model
score_model

Scored dataset

Scored dataset... Scored dataset...

Evaluate Model
evaluate_model

Evaluation results

Evaluate Model

No parameter

Output settings >

Input settings >

Run settings >

Node information >

Component information >

Azure Machine Learning – Designer : 평가

