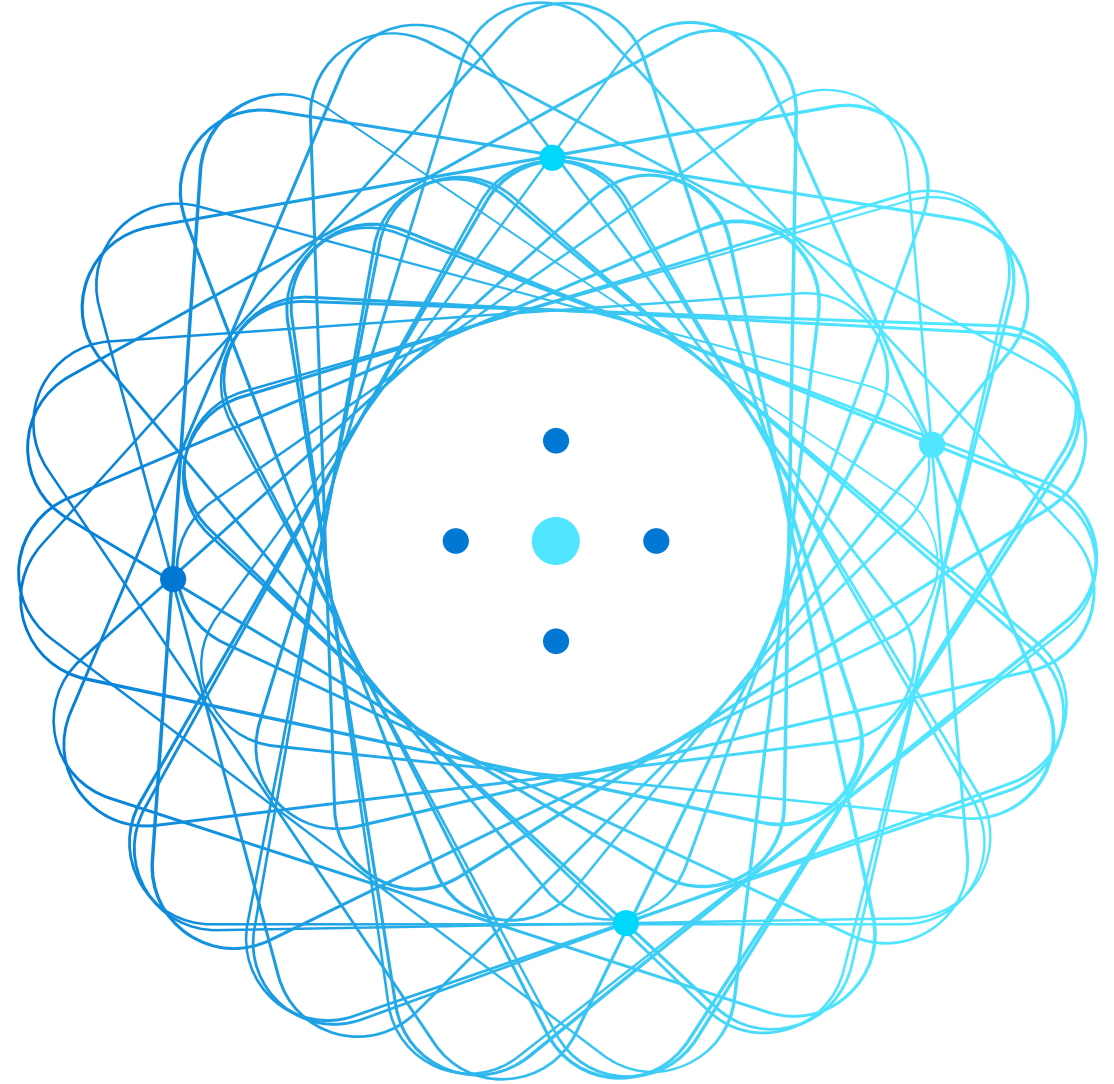


Use no-code machine learning with the Azure Machine Learning designer



Module Agenda

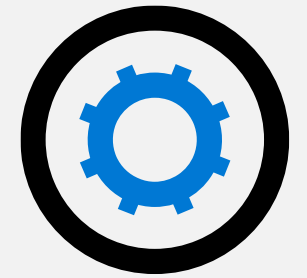


Explore data with the Azure Machine Learning designer



Train and compare models with the Azure Machine Learning designer

Explore data with the Azure Machine Learning designer



Explore data with NumPy and Pandas

NumPy is a Python library that offers mathematical functions to explore and analyze data.

Pandas is a Python library for data analysis and manipulation that mostly works with tables as *data frames*.

To explore data, you'll want to:

- **Clean data** to handle errors, missing values, and other issues.
- **Apply statistical techniques** to better understand the data.
- **Visualize** data to determine relationships between variables.

Clean data

Handle missing data:

- Delete data with missing rows
- Replace empty values with mean or median of the data
- Assign a new category to unknown categorical data

Python

```
# Delete missing rows
```

```
clean_dataset = dataset.dropna()
```

```
# Replace with the mean
```

```
mean_age = clean_dataset["Age"].mean()
```

```
clean_dataset["Age"].fillna(mean_age,  
                             inplace=True)
```

Visualize data

Continuous data are usually best viewed using either:

- An XY scatter plot, especially for relationships between two continuous features
- Histograms or box and whisker plots, to look at the spread of data

Categorical data are usually viewable in a similar way to continuous data, but with data viewed in groups. Categories appear as colors, or groups, in plots.

Matplotlib is a library commonly used to visualize data.

Train and compare models with the Azure Machine Learning designer



Understand the Azure Machine Learning designer

The **designer** is a drag-and-drop interface to build pipelines of components.

Built-in components include:

- Automated machine learning preprocessing capabilities.
- Responsible AI dashboard features.

Quickly build your own pipeline using *custom* components.

Understand a component

A **component** is a self-contained piece of code that does one step in a machine learning pipeline.

A component often has inputs and outputs.

Custom components can be registered to the workspace, available for anyone to use in their own pipelines.

Create a pipeline

A machine learning **pipeline** is the workflow for a full machine learning task. Components are the building blocks of a machine learning pipeline. When you're thinking of a component, it must be under the context of pipeline.

You can build a pipeline in the designer using the custom components registered in the workspace.

To run the pipeline, submit the pipeline as a job.

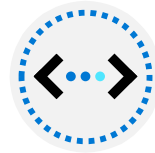
Exercise – Train a model with the designer

In this exercise, you will:

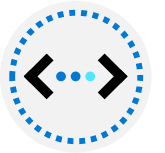
Task 1: Configure a new pipeline



Task 2: Create and run the pipeline



Task 3: Compare models in a pipeline



Instructions

Follow these instructions to complete the exercise:

1. View the exercise repo at <https://microsoftlearning.github.io/mslearn-azure-ml/>.
2. Complete the **Train a model with the Azure Machine Learning Designer** exercise.

