

# Explore data

100 XP

3 minutes

Machine learning models must be trained with existing data. In this case, you'll use a dataset of historical bicycle rental details to train a model that predicts the number of bicycle rentals that should be expected on a given day, based on seasonal and meteorological features.

## Create a dataset

In Azure Machine Learning, data for model training and other operations is usually encapsulated in an object called a *dataset*.

1. View the comma-separated data at <https://aka.ms/bike-rentals> in your web browser.
2. In Azure Machine Learning studio, view the **Datasets** page. Datasets represent specific data files or tables that you plan to work with in Azure ML.
3. Create a new dataset **from web files**, using the following settings:
  - **Basic Info:**
    - **Web URL:** <https://aka.ms/bike-rentals>
    - **Name:** bike-rentals
    - **Dataset type:** Tabular
    - **Description:** Bicycle rental data
    - **Skip data validation:** *Do not select*
  - **Settings and preview:**
    - **File format:** Delimited
    - **Delimiter:** Comma
    - **Encoding:** UTF-8
    - **Column headers:** Only first file has headers
    - **Skip rows:** None
    - **Dataset contains multi-line data:** *Do not select*
  - **Schema:**
    - Include all columns other than **Path**
    - Review the automatically detected types
  - **Confirm details:**
    - Do not profile the dataset after creation
4. After the dataset has been created, open it and view the **Explore** page to see a sample of the data. This data contains historical features and labels for bike rentals.

**Citation:** *This data is derived from Capital Bikeshare and is used in accordance with the published data license agreement.*

---

## Next unit: Train a machine learning model