

# Preparing Data for Machine Learning

Machine Learning is primarily about training models that you can use to provide predictive services to applications. In this exercise, you'll see how you can use Azure Databricks to prepare training data for machine learning purposes.

## Prerequisites

Before starting this lab, complete the **Getting Started with Azure Databricks** lab to set up your Azure Databricks environment and import the data and notebooks you require.

## Prepare Data for Machine Learning

In this exercise, you will learn how to load and manipulate data inside the Azure Databricks environment.

1. In a web browser, open your Azure Databricks workspace.
2. If your cluster is not running, on the **Compute** page, select your cluster and use the ► **Start** button to start it
3. In the Azure Databricks Workspace, using the command bar on the left, select **Workspace**. Then select **Users**, and ► **your\_user\_name**. Then in the folder named **02 - Training and Evaluating Machine Learning Models**, open the **1.0 Featurization** notebook.
4. Attach the notebook to your cluster. Then read the notes in the notebook, running each code cell in turn.

## Clean-up

If you're finished working with Azure Databricks for now, in Azure Databricks workspace, on the **Compute** page, select your cluster and select ■ **Terminate** to shut it down. Otherwise, leave it running for the next exercise.