



1000 XP ▶

Introduction to data for machine learning

45 min • Module • 9 Units

★★★★★ 4.7 (60)

[Rate it](#)

Beginner AI Engineer Data Scientist Student Azure

The power of machine learning models comes from the data that is used to train them. Through content and exercises, we explore how to understand your data, how to encode it so that the computer can interpret it properly, how to clean it of errors, and tips that will help you create models that perform well.

Learning objectives

In this module, you will:

- Visualize large datasets with Exploratory Data Analysis (EDA)
- Clean a dataset of errors
- Predict unknown values using numeric and categorical data

[Start >](#)[⊕ Save](#)

Prerequisites

Some familiarity with machine learning concepts, such as models and cost, is helpful, but not essential

This module is part of these learning paths

[Understand data science for machine learning](#)[Foundations of data science for machine learning](#)

Introduction

2 min

Good, bad, and missing data

3 min

Exercise - Visualize missing data

8 min

Examine different types of data

4 min

Exercise - Work with data to predict missing values

8 min

One-hot vectors

5 min

Exercise - Predict unknown values using one-hot vectors

10 min

Knowledge check

3 min

Summary

2 min