

Introduction: Data Transformations

Find out what you will learn in Data Transformations for Feature Analysis!

In this module you will learn how to prepare your data and features prior to running a machine learning model. This is a very essential step that, if skipped, can cause some grief down the line. It is usually performed right after or alongside Exploratory Data Analysis (EDA).

Why did we build this?

Let's take a moment to consider the consequence of not doing data transformations before a model implementation. An obvious one might be that the model code errors out. This is still a benign possibility as it tells you that *something* is wrong. The more insidious possibility is one where it doesn't cause a problem until much later. Multiple diagnoses of many poorly performing models later, it might become apparent that the data might've needed to be transformed in some way at the beginning! Data Transformations are thus an irreplaceable step for data scientist in their workflows.

What will you learn?

You're going to learn about:

- numerical data transformations such as standardizing, binning, scaling, etc.
- different ways of encoding categorical and non-numerical data.
- when, where and how to apply these methods when presented with a dataset

Who are my classmates?

Learning is social. Whatever you're working on, be sure to connect with the Codecademy community in the [forums](#). Remember to check in with the community regularly, including for things like asking for code reviews on your project work and providing code reviews to others in the [projects category](#), which can help to reinforce what you've learned.