

**Microsoft: DAT210x Programming with Python for Data Science**

Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▼ **4. Transforming Data**

**Lecture: Transformations****Lecture: PCA**

Quiz

**Lab: PCA**

Lab

**Lecture: Isomap**

Quiz

**Lab: Isomap**

Lab

**Lecture: Data Cleansing**

Quiz



4. Transforming Data &gt; Lecture: Transformations &gt; Video



Bookmark

## Transformations

MOD17



## Dive Deeper

▶ 0:00 / 1:31

▶ 1.0x 🔊 🗒️ 📄 🗣️

## ▶ 5. Data Modeling

Download video

Download transcript

.srt

Your goal during the data gathering phase is to record as much working data about your observations as possible since you never know which feature is going to end up being the golden one that allows your machine learning algorithm to succeed. Due to this, there are usually a few redundant or even poor features in your dataset. Think back to those long word problems in grade school that were essentially a simple math question, but came filled with red herrings to throw you off; feeding an unfiltered soup of features to your machine learning algorithms is pretty similar to trying to get it to solve those word problems.

To be effective, many machine learning algorithms need the data passed to them be discerning, discriminating and independent. In this module, you're going to discover methods to get your data behaving like that using *transformers*. This will help improve your own knowledge of your data, as well as improve your machine learning algorithms' performance.

A transformer is any algorithm you apply to your dataset that changes either the feature count or feature values, but does not alter the number of observations. You can use transformers to mung your data as a pre-processing step to clean it up before it's fed to other algorithms. Another popular transformer use is that of dimensionality reduction, where the number of features in your dataset is intelligently reduced to a subset of the original.

Once you've used a few basic transformers, you will also learn about some data cleansing techniques that attempt to rectify problematic observations.



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

