



## Microsoft: DAT210x Programming with Python for Data Science



Bookmarks

► Start Here

► 1. The Big Picture

▼ 2. Data And Features

Lecture: Features Premiere

Quiz

Lecture: Determining  
Features

Quiz



Lecture: Manipulating Data

Quiz

Lecture: Feature  
Representation

Quiz



Lecture: Wrangling Data

Quiz



Lab: Data and Features

Lab



Dive Deeper

## 2. Data And Features &gt; Lecture: Determining Features &gt; Video



Bookmark

## Determining Features

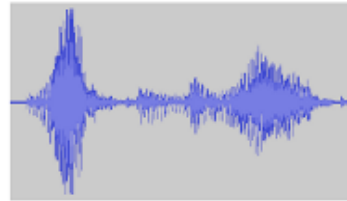
You've have a challenge in mind you want to solve, and you know that you need to collect a lot of samples along with features that describe them. You also know that these features can be continuous numeric values, or they can be categorically based. But which should you focus on? Features or samples? Also how exactly should you go about choosing features?

These are reasonable questions everyone has when they start amassing the data they need to solve an issue. The answer is, *it depends*. Just as in the example of Angie & Craig's lists mentioned in the Machine Learning section, your own intuition about the problem being tackled should really be the driving force behind what data you collect. The only unbreakable rule is that you need to ensure you collect as many features and samples as you possibly can.

If you ever become unsure which one you should focus on more, concentrate on collecting additional samples. A least during collection, try to make sure you have more samples than features because some machine learning algorithms won't function well if that isn't the case. This is also known as the curse of dimensionality. At its core, many algorithms are implemented as matrix operations, and without a  $\geq$  number of samples, a fully formed system of independent equations cannot be formed. You can always create *more* features based off of your existing features. Creating more *pseudo-samples* on the other hand isn't impossible, but might be a bit more difficult.

- ▶ 3. Exploring Data
- ▶ 4. Transforming Data
- ▶ 5. Data Modeling

## INPUT



"It is a curious thing, Harry, but perhaps those who are best suited to power are those who have never sought it. Those who, like you, have leadership thrust upon them, and take up the mantle because they must, and find to their own surprise that they wear it well."



## FEATURES


 $[x_0, x_1, x_2, \dots, x_i]$ 

 $[x_0, x_1, x_2, \dots, x_i]$ 

 $[x_0, x_1, x_2, \dots, x_i]$ 

## INTELLIGENCE



"What in the world..?"



—Albus Dumbledore



Cookie Monster

© All Rights Reserved



© 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.

POWERED BY  
OPENedX

