



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 > Lecture: Determining Features > Feature Importance



Bookmark

Feature Importance

After coming up with a bunch of features to describe your data, you might be tempted to, very early on, investigate which of them deliver the most bang for their buck. Try not to fall into this trap by making too many assumptions about which features are truly relevant! There will be plenty of time and better tools for doing that later in the data analysis process. While out collecting, time spent deliberating whether or not you should move forward with a particular feature is time that's not adding more samples into your dataset. Moreover, if you were already aware of a single *golden feature* that completely solves your challenge and answers the question you had in mind, rather than approaching the problem through data analysis and machine learning, you could probably directly engineer a solution around that single attribute.

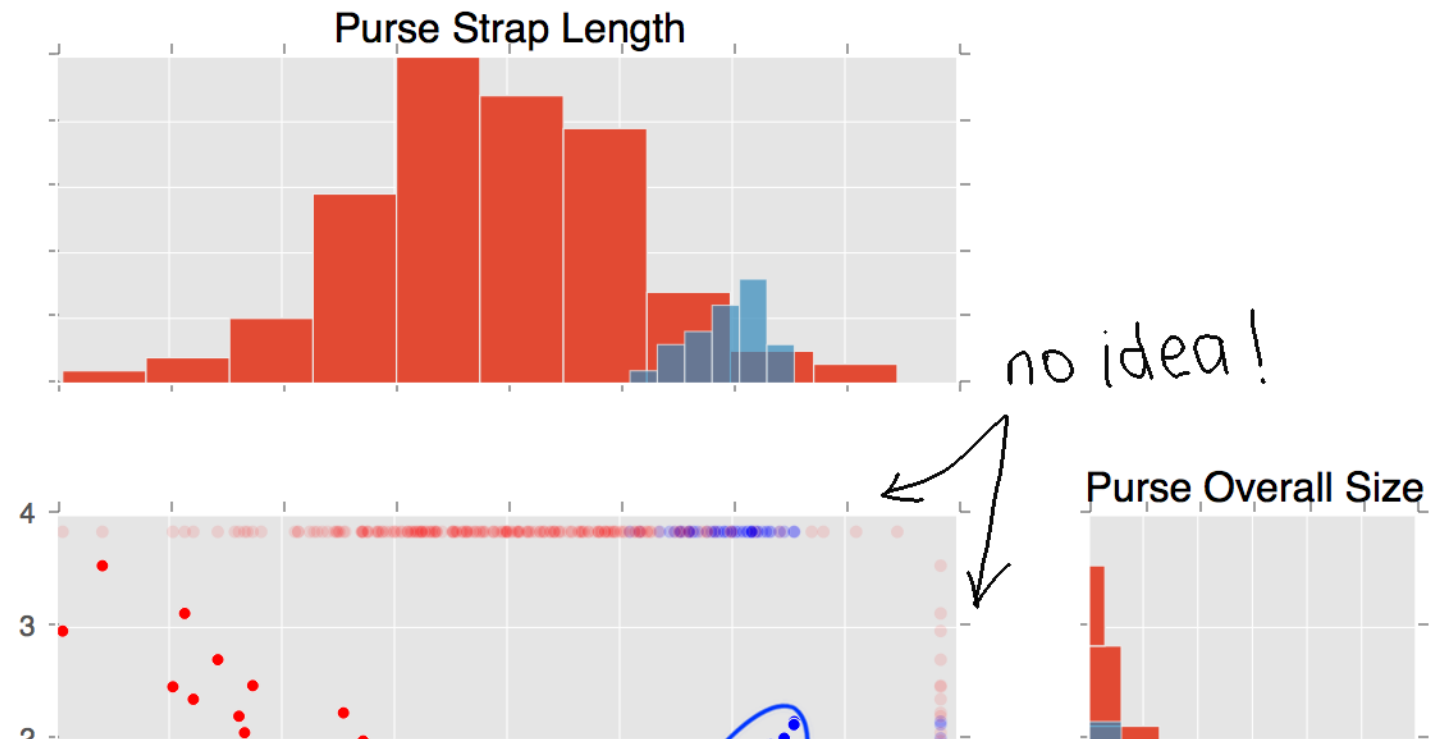
Data-driven problems can be so complex that even with your expertise in the problem's domain, you are still only vaguely aware of some noisy features that do not entirely capture the relationship you are trying to model. These subpar features might only partially help answer your question by providing marginal information. Make a list of all of them, and use them as your feature set.

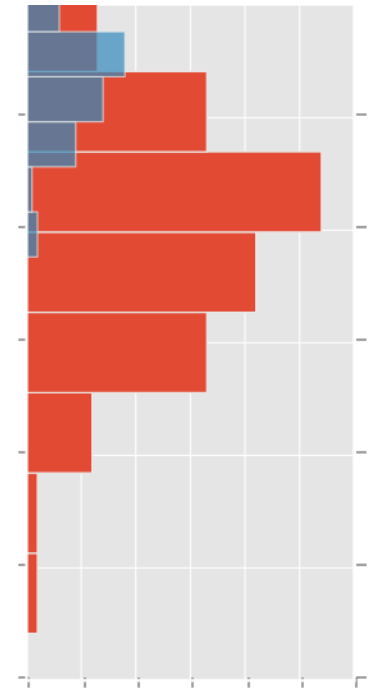
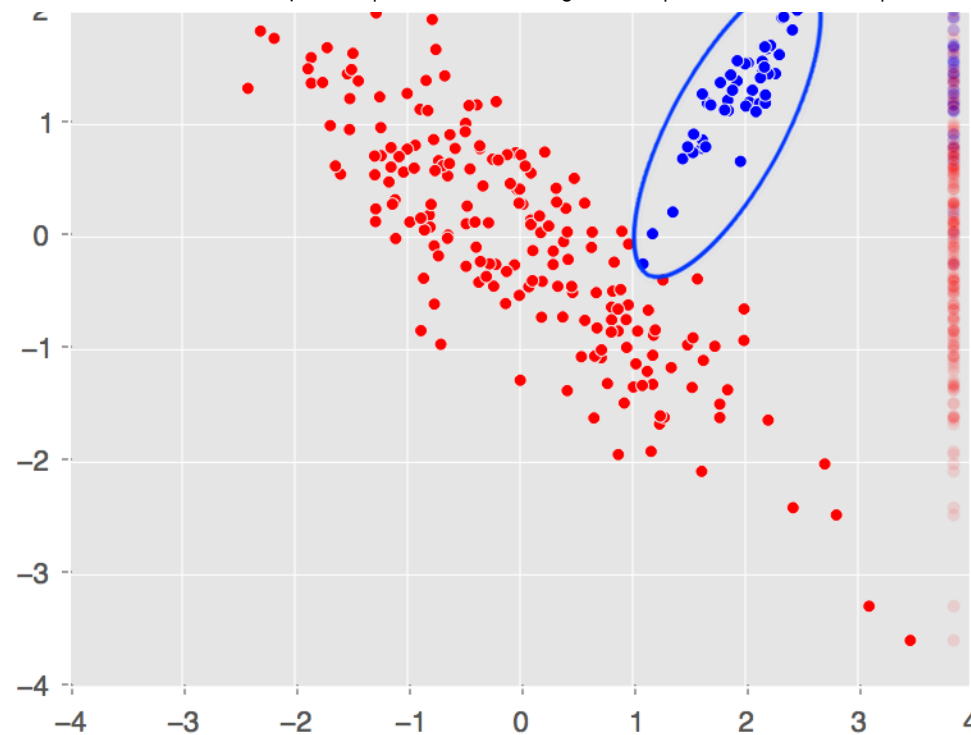
One of the beauties of machine learning is its ability to discover relationships within your data unbeknownst to you. Two or more seemingly weak features, when combined, might end up being exactly that golden feature you've been searching for. Unless you've collected as much data as possible before leaving the feature discovery part to machine learning, you will never have the opportunity to test for that.

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

Imagine you bought a purse for your wife. Curiously, she loves that *you* got it for her, but she *hates* the purse. What more, she hasn't ever been able to tell you exactly what makes her like a specific purse, but she sure knows it when she see's it. Being a data driven person, you've long since figured that there must be a method to the madness and you've actually been keeping some stats on all the purses you've ever bought for her throughout the course of your relationship. Two of the features you've been recording are the size of the purse and the length of the strap.

Individually, these features seem to be poor indicators of her preference in purses. She likes some small-sized purses and some large ones. She likes some purses with virtually no strap that sit snug under her armpit, yet she also likes others that hang below her hip. What gives? By combining these two features, you might realize that there truly is a succinct way of identifying purses your wife likes, each time every time. Just don't tell her the secret!





Yes, you bought her that many purses—whipped!

© 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.

