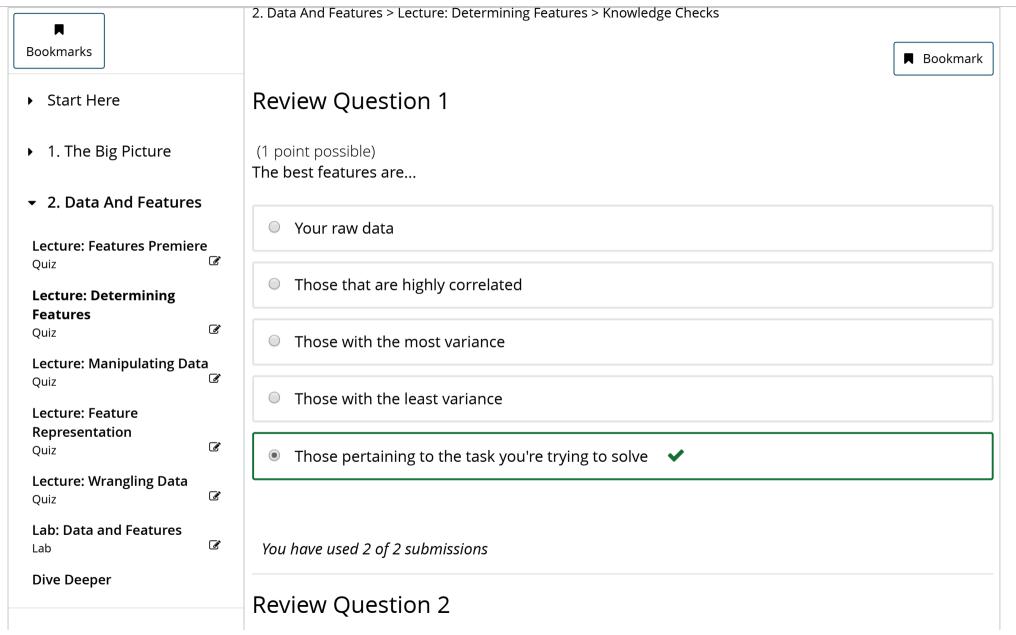


Microsoft: DAT210x Programming with Python for Data Science



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

(1 point possible)

Having been given a very wide (many features) dataset of patient medical records, let's say you want to trim it down to only those features that might have a correlation with a patient's allergies. Which **one** of the following features is **least likely** to be statistically relevant to a person's allergies?

- Patient Drug Prescriptions
- Patient Eye Color
- Patient DNA
- Patient Environment

You have used 1 of 2 submissions

Review Question 3

(1/1 point)

When building out your dataset initially, what are the three things should you focus on the most?

- Collecting more features than samples, so that each observation is documented in great detail.
- ✓ Collecting more samples than features so that the mathematics required for machine learning works out well.

- ✓ Collecting features, even if independently they don't do a great job at answering your dataset's question.
- Ensure you only collect features that directly answer your question without ambiguity.
- ✓ Let your intuition your question, and expertise in the domain of your issue, drive you to choosing appropriate features.



EXPLANATION

The two things that you should *not* be focusing on were both discussed in the video lecture, as well as in the reading materials.

You have used 1 of 2 submissions

© 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.















