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

🔖 Bookmarks

🔖 Bookmark

▸ 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

Review Question

(1/1 point)

A dataset on **oil wells** was made available to you in the form of an html table, which you've loaded into a Pandas dataset. While inspecting the dataset, you notice a *would-be* numeric features called 'depth' actually has a handful of entries that show up as '**?**' rather than as numbers.

What is the *best* course of action you should take to correct this?

○   No action is needed, as this was the desired result

○   Run a command to replace all the **?**'s with **0**'s, such as df.depth[df.depth=='?'] = 0

○   Update your .read_html() method so that **?** is included as a **na_value**, causing them to be converted to nans

◉   Run df.depth = pd.to_numeric(df.depth, errors='coerce') to convert the column to a numeric type and coerce non-numeric values to nans   ✔

3. Exploring Data

4. Transforming Data

5. Data Modeling

EXPLANATION

You certainly want to get rid of those pesky nans. And ideally, you'd do it as soon as you possible can. Of the methods above, which would be the first to execute?

You have used 1 of 3 submissions

POWERED BY
OPENedX