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

🔖 Bookmarks

🔖 Bookmark

Lab Assignment 2

This time, you're going to attempt to load your first csv dataset! Open up the starter code located in Module2/**assignment2.py**. Read through it and follow the directions to:

1. Load up Module2/Datasets/**tutorial.csv**

2. Print the entire dataframe, using `print df`

3. Use the `.describe()` method on the dataset

4. Slice the dataset using `[2:4, 'col3']`

Note: If Pandas complains about not being able to find the tutorial file, just use the full system path to it rather than a relative path. All paths mentioned in this course are in reference to the Module directory produced by unzipping the course material file.

Lab Questions

 (3/3 points)
Please enter a numeric value (e.g. 0, 1, 10.5, etc) which correctly answers the question(s) below:

3. Exploring Data

4. Transforming Data

5. Data Modeling

When you print the results of calling **.describe()** on your dataframe, what is the value displayed in the bottom right corner (col3 max)?

| 1.044722 | ✔ **Answer:** 1.044722 |

Which of the many indexing methods did you use to get [2:4,'col3'] working?

| df.loc ▾ | ✔ **Answer:** df.loc |

How many values are returned when you print the results of the [2:4,'col3'] indexing operation?

| 2 | ✔ **Answer:** 2 |

THE FIRST THING YOU SHOULD DO WHEN YOU ARE EXPERIMENTING WITH A NEW DATASET, IS YOU SHOULD LOAD IT UP. BOTH IN A TEXT EDITOR (SIZE PERMITTING) AS WELL AS BY PRINTING IT THROUGH PANDAS.

Use commands like **.describe()** to get an overall picture of your data. Its standard deviation, its min and max values, its quantiles, etc.

Until you get the hang of it, indexing is going to be pretty tough in Pandas. Luckily, if you're strong with NumPy, you can always just fall back onto NumPy style indexing. But spend time learning and playing around with indexing, so that later on when you're facing harder issues, you don't have indexing to worry about on top of that.

Lastly, for these problems, be sure you solve them *in order*, otherwise you may get incorrect results.

You have used 1 of 2 submissions

POWERED BY
OPENedX