Homework 7

For the following exercises, use the data frame created below.

```
import pandas as pd
df = pd.DataFrame({"var_1": [1, 2, 1, 1], "var_2": [1, 2, 3, 4], "var_3": ["a", "a", "b", "c"]})
```

1. Data frames and series have a method called `.isin()` that makes certain boolean masks easier to construct. Use `.isin()` to extract the rows of `df` where `var_3` is either `"a"` or `"b"`.

2. Does there appear to be any difference between `.mean()` and `df.agg(mean)`? Tinker with a few applications of each to `df` and read the help files to formulate your answer.

3. Write a new function called `cube_root` that takes a series and returns its cube root. Use that within `.agg()` to take the cube root of `var_2`.

4. A very common construction is to follow a `groupby` with an aggregation operation. Often it's just a single column of output that you're interested in, but you can get there by selecting the column at the beginning of the operation or at the end. Demonstrate both approaches using `df` and explain which you think is preferable and why.

5. When you studied numpy arrays, you learned that it makes a distinction between *viewing* a sub-array versus *copying* it (you used `.copy()`). Using `df`, determine if the same distinction applies with pandas data frames.

6. Consider the four different approaches to setting the value in the 1st row, 1st column to `2`. The final method is the preferred approach, but for the other three, see if you can figure out why you're getting those warning messages.

```
df[0:1]["var_1"] = 2
df["var_1"][0:1] = 2
df[["var_1"]][0:1] = 2
df.loc[0, "var_1"] = 2
```