**Assignment #8**

**Objective**

Learn principles of data cleaning and wrangling with Python.

**Due**

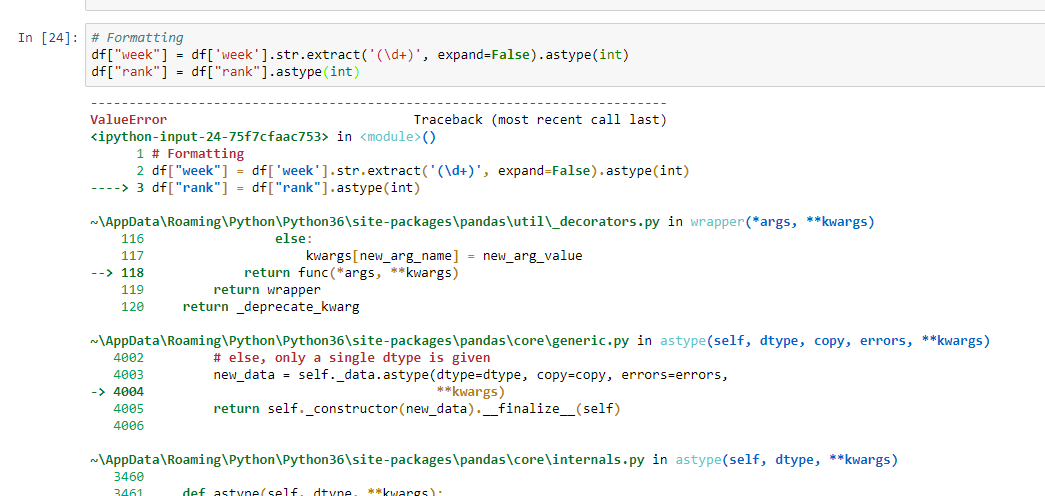
Before at 9am on Feb 26th. Commit your completed Jupyter notebooks to your GitHub repository. And when completed let Dr. Ficklin know via Slack that it is available. Please indicate the amount of time the assignment took to complete.

**Background**

This assignment is meant to introduce you to principles for organizing data (data cleaning) and wrangling to make it easier to work with. A lot of this builds on Pandas.

**Tasks**

* Read the paper titled “Tidy Data” by Hadley Wickham. It has been posted on Slack. When the article provides code examples they are typically in R, rather than Python. You can ignore the code if it does not make sense to you and focus on the principles.
* Complete the online tutorial for Tidy: <http://www.jeannicholashould.com/tidy-data-in-python.html>
  + Use a new Jupyter notebook to enter the code examples provided in the tutorial.
  + The data has been provided to you in a ZIP file on Slack.
  + Take some time to try to understand exactly what each line of code is doing. Unlike previous tutorials there is not a lot of explanations for Python code. It is assumed you know most of the syntax.
  + If you find you are unfamiliar with some of the functions used, take some time to look them up in the online Pandas documentation (you can often just do a Google search on the function name to get there).
* Complete the online tutorial for Python data wrangling: <https://www.springboard.com/blog/data-wrangling/>
  + Begin a new Jupyter notebook for this tutorial.
  + Notes:
    - The author tells you to look at Julia Evan’s Pandas Cookbook. You only need to do that if you want another reference for Panda.
    - The author tells you to use Anaconda. In this case, use your Jupyter notebook instead.
  + The input data file is available for you on Slack. No need to go hunting for it. You can find more information about this data set from the Kaggle website: <https://www.kaggle.com/c/titanic>
* Once you have completed both tutorials, post your two completed notebooks to your GitHub repository in an Assignment08 folder.
* Second section of Tidy tutorial does not work for me:



Third portion, I get the error “DataFrame has no attribute “sort”

Why is it that it sometimes uses “ “ and other times uses ‘’, ie

df = df.drop(['year',"month","day", "day\_raw"], axis=1)

df = df.dropna()

Second tutorial- cannot export to excel