Day 03 - Working with `Pandas`

Sept. 15, 2020

CMSE

Administrative

- **Office Hours**: Posted on website and D2L
    - Zoom link: https://msu.zoom.us/j/97957751894 (https://msu.zoom.us/j/97957751894) password: help4me
- **Groups**: Reminder set your name to indicate your group number
    - We will stay out of your breakout rooms for about 10 minutes
    - This is to give you time to get started and read the assignment together
    - Request help if you need it during that time
- **Homework 1**: Now on EGR website and on D2L
    - Due by midnight Sept 25th; graded in about a week
- **Class Discussion**: We decided to change our approach to class discussion starting this week.
    - We will ask a random group to report on their work and results at the end of class.
    - We will tell you if your group is selected at the beginning of class.
    - We will close the breakout rooms at 11:55am for this and to answer additional questions.

Any questions?

From Pre-Class Assignment

Challenging bits

- I am confused on making the 3D array that represents color (Numpy arrays can be N-dimensional)
- I was struggling with the last question on how to print students who earn greater than 3.0 (slicing dataframes with logic)

You will get more practice with the second skill today

```python
import numpy as np

# 1D Array of length 10
print('\n1D Array of length 10')
a = np.zeros(10)
print(a,"\nSize: ", a.shape)

# 2D Array that is 3 by 3
print('\n2D Array that is 3 by 3')
b = np.zeros([3,3])
print(b,"\nSize: ", b.shape)
```

```
1D Array of length 10
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
Size:  (10,)

2D Array that is 3 by 3
[[0. 0. 0.]
 [0. 0. 0.]
 [0. 0. 0.]]
Size:  (3, 3)
```

```
In [2]:  # 3D Array that is 4 by 5 by 3
         print('\n3D Array that is 4 by 5 by 3')
         c = np.zeros([4,5,3])
         print(c,"\nSize: ", c.shape)
```

```
3D Array that is 4 by 5 by 3
[[[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]]
Size:  (4, 5, 3)
```

```
In [3]:  import pandas as pd

         # Create a dictionary of student records
         student_dict = {"name":["mac","dee","dennis","charlie","frank"],
                         "pid":[2081, 2082, 2083, 2084, 2085],
                         "total":[60,95,75,5,85],
                         "grade": [2,4,3,0,3.5]}

         # Create a dataframe from the dictionary
         student_df = pd.DataFrame(student_dict)

         student_df
```

Out[3]:

	name	pid	total	grade
0	mac	2081	60	2.0
1	dee	2082	95	4.0
2	dennis	2083	75	3.0
3	charlie	2084	5	0.0
4	frank	2085	85	3.5

```python
In [4]:   # Check the logic on grades
          student_df['grade'] >= 3.0

          # This generates binary values that can be used to slice dataframes
```

```
Out[4]:   0    False
          1     True
          2     True
          3    False
          4     True
          Name: grade, dtype: bool
```

```
In [5]:  # Slice the dataframe

         student_df[student_df['grade'] >= 3.0]
```

Out[5]:

	name	pid	total	grade
1	dee	2082	95	4.0
2	dennis	2083	75	3.0
4	frank	2085	85	3.5

Questions, Comments, Concerns?