PythonPandasClass1-Completed

February 18, 2021

```
[ ]: # import necessary packages here
     import warnings
     warnings.filterwarnings('ignore')
```

1. Load the pandas library, and alias it so you can use `pd` instead of typing `pandas` each time you need to access the package.

2. Load the penguin and titanic datasets from GitHub. Store them in `penguins` and `titanic` respectively.

```
penguin_url = 'https://raw.githubusercontent.com/cmparlettpelleriti/CPSC392ParlettPelleriti/ma:
titanic_url = 'https://raw.githubusercontent.com/mwaskom/seaborn-data/master/titanic.csv'
print("done")
```

3. Print the head (first 5 or 6 rows) of each data set.

4. Store the sex column from the titanic dataset in the variable `sex`. Print it.

5. Store the 10th row from the penguin dataset in the variable `penguins_10`. Print it.

6. Grab the age of the 12th person in the titanic dataset and store it in the variable `passenger12`. Print it.

7. Store the mean age of titanic passengers in the variable `mean_age`, print it.

8. Create a new column in `penguins` called `bill_ratio` for the penguin dataset that contains the ratio of bill length to bill depth (in otherwords bill length divided by bill depth). Print out this column.

9. Create a new data frame called `lil_boys` that only contains penguins that have a body mass less than 4000 grams. Print the first 4 rows of this dataframe.

10. What is the mean bill length for each species of penguin? Store the means in the variables `bl_Adelie`, `bl_Chinstrap` and `bl_Gentoo` respectively. Print them out.

11. Which class from the titanic dataset had the highest proportion of surviors? Print out the proportion of people who survived for each class.

12. Is there a difference in the average fare paid between males and females in the titanic dataset? Use pandas to find your answer.

13. Create a dataframe called `small_penguin` that has all the rows from `penguins` but only the columns `bill_length_mm` ,`bill_depth_mm` , `flipper_length_mm` , and `body_mass_g`

```
[2]:  ### YOUR CODE HERE ###

      import pandas as pd
      import numpy as np

      penguin_url = 'https://raw.githubusercontent.com/cmparlettpelleriti/
       ↪CPSC392ParlettPelleriti/master/Data/penguins.csv'
      titanic_url = 'https://raw.githubusercontent.com/mwaskom/seaborn-data/master/
       ↪titanic.csv'

      titanic = pd.read_csv(titanic_url)
      penguins = pd.read_csv(penguin_url)
```

```
[4]:  # 3
      titanic.head()
```

```
[4]:     survived  pclass     sex   age  sibsp  parch      fare embarked  class  \
      0         0       3    male  22.0      1      0    7.2500        S  Third
      1         1       1  female  38.0      1      0   71.2833        C  First
      2         1       3  female  26.0      0      0    7.9250        S  Third
      3         1       1  female  35.0      1      0   53.1000        S  First
      4         0       3    male  35.0      0      0    8.0500        S  Third

           who  adult_male deck  embark_town alive  alone
      0    man        True  NaN  Southampton    no  False
      1  woman       False    C    Cherbourg   yes  False
      2  woman       False  NaN  Southampton   yes   True
      3  woman       False    C  Southampton   yes  False
      4    man        True  NaN  Southampton    no   True
```

```
[5]:  penguins.head()
```

```
[5]:     Unnamed: 0 species     island  bill_length_mm  bill_depth_mm  \
      0           0  Adelie  Torgersen            39.1           18.7
      1           1  Adelie  Torgersen            39.5           17.4
      2           2  Adelie  Torgersen            40.3           18.0
      3           3  Adelie  Torgersen             NaN            NaN
      4           4  Adelie  Torgersen            36.7           19.3

         flipper_length_mm  body_mass_g     sex  year
      0              181.0       3750.0    male  2007
      1              186.0       3800.0  female  2007
      2              195.0       3250.0  female  2007
      3                NaN          NaN     NaN  2007
      4              193.0       3450.0  female  2007
```

```
[6]:  #4

      sex = titanic["sex"]
      sex
```

```
[6]: 0        male
     1      female
     2      female
     3      female
     4        male
              …
     886      male
     887    female
     888    female
     889      male
     890      male
     Name: sex, Length: 891, dtype: object
```

```
[8]:  #5
      penguins_10 = penguins.iloc[9,]
      penguins_10
```

```
[8]: Unnamed: 0                  9
     species                Adelie
     island              Torgersen
     bill_length_mm             42
     bill_depth_mm            20.2
     flipper_length_mm         190
     body_mass_g              4250
     sex                       NaN
     year                     2007
     Name: 9, dtype: object
```

```
[10]:  #6

       passenger12 = titanic.iloc[11,]
       passenger12
```

```
[10]: survived             1
      pclass               1
      sex             female
      age                 58
      sibsp                0
      parch                0
      fare             26.55
      embarked             S
      class            First
```

```
who                      woman
adult_male               False
deck                         C
embark_town       Southampton
alive                      yes
alone                     True
Name: 11, dtype: object
```

[12]:
```python
#7

mean_age = titanic["age"].mean()
mean_age
```

[12]: 29.69911764705882

[13]:
```python
#8

penguins["bill_ratio"] = penguins["bill_length_mm"]/penguins["bill_depth_mm"]
penguins["bill_ratio"]
```

[13]:
```
0        2.090909
1        2.270115
2        2.238889
3             NaN
4        1.901554
           …
339      2.818182
340      2.403315
341      2.725275
342      2.673684
343      2.684492
Name: bill_ratio, Length: 344, dtype: float64
```

[14]:
```python
#9

lil_boys = penguins.loc[penguins.body_mass_g < 4000,]
lil_boys.head(4)
```

[14]:
```
   Unnamed: 0 species      island  bill_length_mm  bill_depth_mm  \
0           0  Adelie  Torgersen            39.1           18.7
1           1  Adelie  Torgersen            39.5           17.4
2           2  Adelie  Torgersen            40.3           18.0
4           4  Adelie  Torgersen            36.7           19.3

   flipper_length_mm  body_mass_g     sex  year  bill_ratio
0              181.0       3750.0    male  2007    2.090909
1              186.0       3800.0  female  2007    2.270115
```

```
2              195.0       3250.0  female  2007    2.238889
4              193.0       3450.0  female  2007    1.901554
```

[15]:
```python
#10

bl_Adelie = penguins.loc[penguins.species == "Adelie", "bill_length_mm"].mean()
bl_Chinstrap = penguins.loc[penguins.species == "Chinstrap", "bill_length_mm"].
 ↪mean()
bl_Gentoo = penguins.loc[penguins.species == "Gentoo", "bill_length_mm"].mean()

print(bl_Adelie, bl_Chinstrap, bl_Gentoo)
```

```
38.79139072847682 48.83382352941177 47.50487804878048
```

[19]:
```python
#11

classes = pd.unique(titanic.pclass)

for c in classes:
    survive_prop = titanic.loc[titanic.pclass == c, "survived"].mean()
    print("Class " + str(c) + " had " + str(survive_prop) + " survivors.")
```

```
Class 3 had 0.24236252545824846 survivors.
Class 1 had 0.6296296296296297 survivors.
Class 2 had 0.47282608695652173 survivors.
```

[21]:
```python
#12

titanic_males = titanic.loc[titanic.sex == "male","fare"].mean()

titanic_females = titanic.loc[titanic.sex == "female","fare"].mean()

print(titanic_males, titanic_females)
```

```
25.523893414211443 44.47981783439491
```

[24]:
```python
#13
small_penguin = penguins[["bill_length_mm" ,"bill_depth_mm" ,␣
 ↪"flipper_length_mm" , "body_mass_g"]]
small_penguin
```

[24]:
	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g
0	39.1	18.7	181.0	3750.0
1	39.5	17.4	186.0	3800.0
2	40.3	18.0	195.0	3250.0
3	NaN	NaN	NaN	NaN

4	36.7	19.3	193.0	3450.0
..	…	…	…	…
339	55.8	19.8	207.0	4000.0
340	43.5	18.1	202.0	3400.0
341	49.6	18.2	193.0	3775.0
342	50.8	19.0	210.0	4100.0
343	50.2	18.7	198.0	3775.0

[344 rows x 4 columns]