Assignment2

October 15, 2018

problem 1

```
In [1]: import pandas as pd
        import numpy as np
        import math as ma
        #NOTE TO SELF:
        #not necessarily pertinent to this class, but this is more efficient than the way
        # you would normally do this b/c they have the same summation, but
        # the inner summation for the dp is known to be < but not == however
        # for the normal case the inner summation is both
        def isPrime(n,table):
            for i in table:
                if n % i==0:
                    return False
                if i > ma.sqrt(n):
                    break
            return True
        def getPrimes(n):
            table= np.array([2])
            for i in range(n-1):
                val=table[i]+1
                while not isPrime(val, table):
                    val+=1
                table= np.append(table,[val])
            return table
        output= getPrimes(10)
        series1= pd.Series(output)
        print(series1)
```

```
0     2
1     3
2     5
3     7
4    11
5    13
6    17
7    19
```

```
8    23
9    29
dtype: int64
```

problem 2

```
In [2]: for i in range(1,10,2):
            print(series1.iloc[i])
```

```
3
7
13
19
29
```

problem 3

```
In [3]: newIndices=[]
        for i in range(97,107):
            newIndices.append(chr(i))
        series1= pd.Series(output,newIndices)
        print(series1)
```

```
a     2
b     3
c     5
d     7
e    11
f    13
g    17
h    19
i    23
j    29
dtype: int64
```

problem 4

```
In [4]: for i in range(98,107,2):
            print(series1.loc[chr(i)])
```

```
3
7
13
19
29
```

problem 5

```
In [5]: df1Data= [[2,'Jason','Miller',42,4,25],
                  [5,'Jacob','Jacobson',52,24,94],
                  [10,'Tina','Ali',36,31,57],
                  [15,'Jake','Milner',24,2,62],
                  [20,'Amy','Cooze',73,3,70]]
        columns= ['id','first_name','last_name','age','preTestScore','postTestScore']
        df1= pd.DataFrame(df1Data,columns=columns)
        print(df1)
```

```
   id first_name last_name  age  preTestScore  postTestScore
0   2      Jason    Miller   42             4             25
1   5      Jacob  Jacobson   52            24             94
2  10       Tina       Ali   36            31             57
3  15       Jake    Milner   24             2             62
4  20        Amy     Cooze   73             3             70
```

problem 6

```
In [6]: df1= df1.set_index(['id'])
        print(df1)
```

```
   first_name last_name  age  preTestScore  postTestScore
id
2       Jason    Miller   42             4             25
5       Jacob  Jacobson   52            24             94
10       Tina       Ali   36            31             57
15       Jake    Milner   24             2             62
20        Amy     Cooze   73             3             70
```

problem 7

```
In [7]: print(df1['first_name'])
```

```
id
2     Jason
5     Jacob
10     Tina
15     Jake
20      Amy
Name: first_name, dtype: object
```

problem 8

```
In [8]: print(df1.loc[10,'age'])
```

```
36
```

problem 9

```
In [9]: print((df1['postTestScore']-df1['preTestScore']).mean())
```

```
48.8
```

problem 10

```
In [10]: df1.loc[10,'postTestScore']=np.nan
         df1.loc[15,'postTestScore']=np.nan
         print(df1)
```

```
   first_name last_name  age  preTestScore  postTestScore
id
2       Jason    Miller   42             4          25.0
5       Jacob  Jacobson   52            24          94.0
10       Tina       Ali   36            31           NaN
15       Jake    Milner   24             2           NaN
20        Amy     Cooze   73             3          70.0
```

problem 11

```
In [11]: print(df1.dropna())
```

```
   first_name last_name  age  preTestScore  postTestScore
id
2       Jason    Miller   42             4          25.0
5       Jacob  Jacobson   52            24          94.0
20        Amy     Cooze   73             3          70.0
```

problem 12

```
In [12]: df1=df1.reset_index()
         df1=df1.set_index(['first_name','last_name'])
         print(df1)
```

```
                      id  age  preTestScore  postTestScore
first_name last_name
Jason      Miller      2   42             4          25.0
Jacob      Jacobson    5   52            24          94.0
Tina       Ali        10   36            31           NaN
Jake       Milner     15   24             2           NaN
Amy        Cooze      20   73             3          70.0
```

problem 13

```
In [13]: print(int(df1.loc[('Tina','Ali'),'age']))
```

```
36
```