Assignment3

October 22, 2018

question1

In [19]:
```
import numpy as np
import pandas as pd
import math as ma
from datetime import datetime as dt
data= np.array([[1,2],[3,4],[4,5],[5,6]])
data= np.array([[1,2,3,4],[5,6,7,8]])
df= pd.DataFrame(data,columns=[['a','a','b','b'],[1,2,1,2]])
data= np.array([[np.datetime64('2017-06-05'),155,955,66,37.10,32.0,30.31],
      [np.datetime64('2017-06-06'),150,987,69,36.98,31.3,30.56],
      [np.datetime64('2017-06-07'),153,963,62,36.78,31.7,30.46],
      [np.datetime64('2017-06-08'),155,1000,61,36.11,31.2,30.11],
      [np.datetime64('2017-06-09'),156,1012,66,37.07,30.0,31.00]
               ])
df= pd.DataFrame(data,columns=[['','Price','Price','Price','Price to earnings ratio (P/
    ['date','Facebook','Google','Microsoft','Facebook','Google','Microsoft']])
df
```

Out[19]:

	Price				Price to earnings ratio (P/E)		\
	date	Facebook	Google	Microsoft	Facebook	Google	
0	2017-06-05	155	955	66	37.1	32	
1	2017-06-06	150	987	69	36.98	31.3	
2	2017-06-07	153	963	62	36.78	31.7	
3	2017-06-08	155	1000	61	36.11	31.2	
4	2017-06-09	156	1012	66	37.07	30	

	Microsoft
0	30.31
1	30.56
2	30.46
3	30.11
4	31

question 2

In [20]:
```
df=df.set_index([('','date')])
df.index.names=[('date','')]
```

```
        startDate=np.datetime64('2017-06-05')
        for i in range(5):
            print(df.loc[startDate+i])
```

```
Price                          Facebook       155
                               Google         955
                               Microsoft       66
Price to earnings ratio (P/E)  Facebook      37.1
                               Google          32
                               Microsoft    30.31
Name: 2017-06-05 00:00:00, dtype: object
Price                          Facebook       150
                               Google         987
                               Microsoft       69
Price to earnings ratio (P/E)  Facebook     36.98
                               Google        31.3
                               Microsoft    30.56
Name: 2017-06-06 00:00:00, dtype: object
Price                          Facebook       153
                               Google         963
                               Microsoft       62
Price to earnings ratio (P/E)  Facebook     36.78
                               Google        31.7
                               Microsoft    30.46
Name: 2017-06-07 00:00:00, dtype: object
Price                          Facebook       155
                               Google        1000
                               Microsoft       61
Price to earnings ratio (P/E)  Facebook     36.11
                               Google        31.2
                               Microsoft    30.11
Name: 2017-06-08 00:00:00, dtype: object
Price                          Facebook       156
                               Google        1012
                               Microsoft       66
Price to earnings ratio (P/E)  Facebook     37.07
                               Google          30
                               Microsoft       31
Name: 2017-06-09 00:00:00, dtype: object
```

question 3

```
In [21]: print('Mean prices:')
         print(df['Price'].mean())
         print('\nAvg price to earnings ratio (P/E): ')
         print(df['Price to earnings ratio (P/E)'].mean())
```

```
Mean prices:
Facebook     153.8
```

```
Google       983.4
Microsoft     64.8
dtype: float64

Avg price to earnings ratio (P/E):
Facebook     36.808
Google       31.240
Microsoft    30.488
dtype: float64
```

question 4

```
In [22]: studentsData= np.array([['John',20],
                                 ['Bob',30],
                                ['Suzan',22]
                                 ])
         students= pd.DataFrame(studentsData,columns=['Student Name','Age'])
         courseData= np.array([['CS',233,'Computer Science 233'],
                               ['CS',455,'Computer Science 455'],
                               ['ENGL',433,'English 433']
                              ])
         course= pd.DataFrame(courseData,columns=['dept','number','description'])
         takesData= np.array([['CS',233,'C'],
                              ['CS',455,'B'],
                              ['ENGL',433,'A']
                             ])
         takes= pd.DataFrame(takesData,columns=['dept','course num','grade'])
         print(students)
         print(course)
         print(takes)
```

```
  Student Name Age
0         John  20
1          Bob  30
2        Suzan  22
   dept number           description
0    CS    233  Computer Science 233
1    CS    455  Computer Science 455
2  ENGL    433           English 433
   dept course num grade
0    CS        233     C
1    CS        455     B
2  ENGL        433     A
```

question 5

```
In [23]: takes.index=['John','John','Suzan']
         takes['grade']=takes['grade'].transform(lambda x: abs(ord(x)-ord('A')-4))
```

```
takes= takes.reset_index()
takes= takes.rename(columns={'index':'name'})
totalPoints= takes.groupby('name')['grade'].sum()
totalClasses= takes.groupby('name')['name'].count()
gpa=totalPoints/totalClasses
gpa=gpa.reset_index()
gpa=gpa.rename(columns={0:'gpa'})
gpa= pd.merge(students,gpa,how='outer',left_on='Student Name',right_on='name')
gpa=gpa.fillna(0.0)
gpa[['Student Name','gpa']].set_index('Student Name')
```

Out[23]:

Student Name	gpa
John	2.5
Bob	0.0
Suzan	4.0

question 6

```
In [24]: totalClasses.index.name='Student Name'
         classes= totalClasses.reset_index()
         classes=classes.rename(columns={'name':'numClasses'})
         classes= pd.merge(classes,students, on='Student Name',how= 'outer')
         classes=classes.fillna(0)
         print(classes[classes['numClasses']==0]['Student Name'])
```

```
2    Bob
Name: Student Name, dtype: object
```

question 7

```
In [25]: dates= pd.PeriodIndex(start='2018-01-01',end='2018-12-31',freq="B")
         dates=pd.Series(dates)
         dates
```

Out[25]:

0	2018-01-01
1	2018-01-02
2	2018-01-03
3	2018-01-04
4	2018-01-05
5	2018-01-08
6	2018-01-09
7	2018-01-10
8	2018-01-11
9	2018-01-12
10	2018-01-15
11	2018-01-16
12	2018-01-17

```
13    2018-01-18
14    2018-01-19
15    2018-01-22
16    2018-01-23
17    2018-01-24
18    2018-01-25
19    2018-01-26
20    2018-01-29
21    2018-01-30
22    2018-01-31
23    2018-02-01
24    2018-02-02
25    2018-02-05
26    2018-02-06
27    2018-02-07
28    2018-02-08
29    2018-02-09
          ...
231   2018-11-20
232   2018-11-21
233   2018-11-22
234   2018-11-23
235   2018-11-26
236   2018-11-27
237   2018-11-28
238   2018-11-29
239   2018-11-30
240   2018-12-03
241   2018-12-04
242   2018-12-05
243   2018-12-06
244   2018-12-07
245   2018-12-10
246   2018-12-11
247   2018-12-12
248   2018-12-13
249   2018-12-14
250   2018-12-17
251   2018-12-18
252   2018-12-19
253   2018-12-20
254   2018-12-21
255   2018-12-24
256   2018-12-25
257   2018-12-26
258   2018-12-27
259   2018-12-28
260   2018-12-31
```

```
Length: 261, dtype: object
```

question 8

```
In [26]: days= pd.PeriodIndex(start='2018-01-01',end='2018-12-31',freq='D')
         dayNames=pd.DataFrame(days.strftime("%A"),columns=['DayName'])
         dayNames['Days']=days
         dayNames=dayNames.groupby('DayName').count()
         dayNames
```

```
Out[26]:            Days
         DayName
         Friday       52
         Monday       53
         Saturday     52
         Sunday       52
         Thursday     52
         Tuesday      52
         Wednesday    52
```

question 9

```
In [27]: from pandas_datareader import data
         googleProfit=data.DataReader('GOOG','yahoo',start='2017-01-01',end='2017-12-31')
         googleProfit= googleProfit['Close']-googleProfit['Open']
         googleProfit=googleProfit.reset_index()
         googleProfit=googleProfit.rename(columns={0:'difference'})
         googleProfit['Date']= googleProfit['Date'].transform(lambda x:x.strftime('%A'))
         googleProfit= googleProfit.groupby('Date')['difference'].sum().sort_values(ascending= F
         googleProfit
```

```
Out[27]: Date
         Wednesday    77.895081
         Tuesday      42.460205
         Monday       38.770020
         Thursday      5.105225
         Friday        1.340271
         Name: difference, dtype: float64
```