Practice exercises week 11

Pandas:

1. File "StuInfo.txt" includes students information of Name, ID, Scores of 5 subjects and birthday, items in each line were splitted by tab mark, shown as follows (not include the fields name of the 1st row).

Name	ID	SE	Math	En	Fi	PE	Birthday
Vivian	210001	72	98	77	79	68	21/10/2002
Nancy	210002	83	61	74	83	89	7/8/2003
Wilson	210003	98	95	82	72	90	27/2/2003
James	210004	94	60	66	72	69	8/3/2004
Bruce	210005	83	62	86	72	74	28/10/2002
Taylor	210006	76	91	63	94	68	2/3/2002

(1). Turn it to pandas frame, shown as follows.

	Name	ID	SE	Math	En	Fi	PE	Birthday
0	Vivian	210001	72	98	77	79	68	21/10/2002
1	Nancy	210002	83	61	74	83	89	7/8/2003
2	Wilson	210003	98	95	82	72	90	27/2/2003
3	James	210004	94	60	66	72	69	8/3/2004
4	Bruce	210005	83	62	86	72	74	28/10/2002
5	Taylor	210006	76	91	63	94	68	2/3/2002
6	Kelly	210007	67	93	71	75	97	21/11/2002
7	Lane	210008	72	95	66	63	63	19/9/2002
8	Terry	210009	65	86	100	70	81	26/10/2002
9	Tina	210010	84	81	64	61	100	15/8/2002
10	Alexis	210011	63	77	76	60	96	20/9/2004
11	Benny	210012	78	84	96	72	99	24/4/2002
12	Kylie	210013	86	83	60	73	90	21/6/2004
13	Merry	210014	79	90	86	90	66	12/11/2003
14	Addison	210015	86	64	70	87	61	11/2/2003
15	Rachel	210016	87	88	97	100	75	5/3/2002
16	Yilia	210017	68	69	97	78	63	20/2/2002
17	Carl	210018	90	62	96	71	81	24/11/2004
18	Emma	210019	71	81	61	68	70	25/4/2002

(2). Calculate the average score and standard deviation of each subject (Use Pandas statistical functions).

```
*** average score and standard deviation ***
Average score of SE is: 78.75862068965517
Standard deviation of SE is: 11.201578882377612
Average score of Math is: 78.51724137931035
Standard deviation of Math is: 12.860850587442195
Average score of En is: 77.93103448275862
Standard deviation of En is: 12.983316312216004
Average score of Fi is: 76.48275862068965
Standard deviation of Fi is: 11.409960215453276
Average score of PE is: 79.55172413793103
Standard deviation of PE is: 12.391201391357589
```

Ref answer:

```
import pandas as pd
#(1)
ItemName=["Name","ID","SE", "Math","En","Fi","PE","Birthday"]
condic=dict()
dicval=[[],[],[],[],[],[],[],[]]
fstuinfo=open("StuInfo.txt","r")
fcontent=fstuinfo.readlines()
fstuinfo.close()
for sturec in fcontent:
    recitem=sturec.strip().split('\t')
    for i in range(0,len(ItemName)):
        condickey=ItemName[i]
        try:
            condicval=dicval[i].append(int(recitem[i]))
        except ValueError:
            condicval=dicval[i].append(recitem[i])
        condic[condickey]=dicval[i]
StuInfoFrame=pd.DataFrame(condic)
print(StuInfoFrame)
#(2)
print("*** average score and standard deviation ***")
for i in range(2,7):
    print("Average score of "+ItemName[i]+" is:",
          StuInfoFrame[ItemName[i]].mean())
    print("Standard deviation of "+ItemName[i]+" is:",
          StuInfoFrame[ItemName[i]].std())
```

2. File “VibrationData.csv” records the vibration acceleration signals in three directions(XYZ) at a certain position on the diesel engine, shown as follows:

	A	B	C	D	E	F
1	Time	DirectionX	Time	DirectionY	Time	DirectionZ
2	1.74E-05	-2.464599609	1.74E-05	6.405395508	1.74E-05	-2.416381836
3	5.64E-05	-2.174804688	5.64E-05	-13.42712402	5.64E-05	11.70776367
4	9.55E-05	-3.50402832	9.55E-05	-10.92236328	9.55E-05	-18.2668457
5	0.000134543	3.293945313	0.000134543	5.922241211	0.000134543	-7.789916992
6	0.000173606	12.06115723	0.000173606	-0.742797852	0.000173606	18.25866699
7	0.000212668	17.54418945	0.000212668	5.217407227	0.000212668	3.065185547
8	0.000251731	9.780883789	0.000251731	24.3605957	0.000251731	-15.39233398
9	0.000290793	-13.80981445	0.000290793	10.234375	0.000290793	22.29553223
10	0.000329856	-4.04296875	0.000329856	-21.82702637	0.000329856	-6.156005859
11	0.000368918	-16.89294434	0.000368918	-29.62768555	0.000368918	-18.5916748
12	0.000407981	-10.86877441	0.000407981	-12.01416016	0.000407981	8.157226563
13	0.000447043	-1.545410156	0.000447043	21.73400879	0.000447043	3.639160156
14	0.000486106	1.471435547	0.000486106	30.5291748	0.000486106	-11.37475586
15	0.000525168	1.972290039	0.000525168	-18.80310059	0.000525168	10.37463379

Calculate the mean of the absolute value and standard deviation of vibration

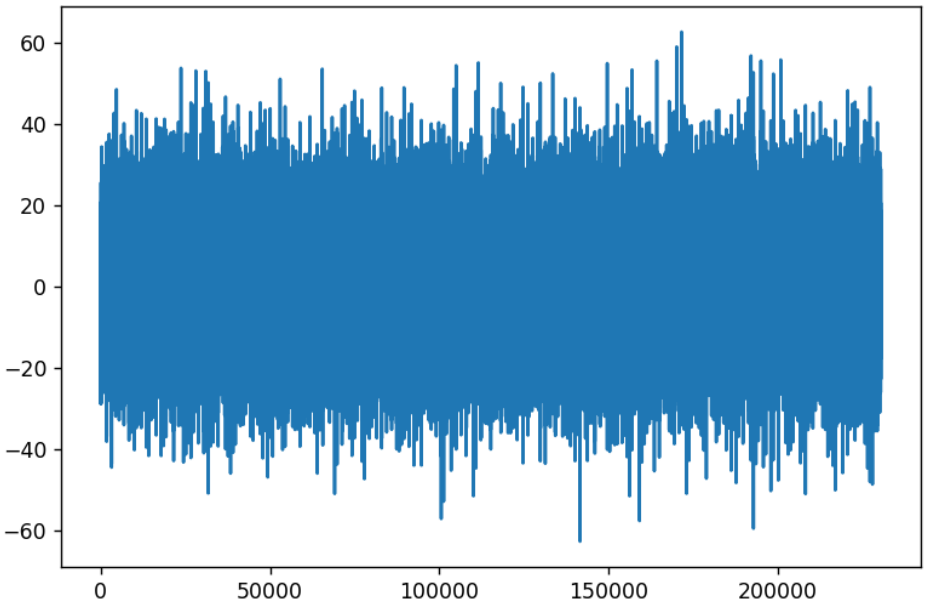
acceleration signal in each direction. If you need to calculate the mean of the absolute value, the following statements can be used as:

```
vibsig["DirectionX"].abs().mean()
```

Here is the running results.

```
DirectionX mean: 6.976225873160231
DirectionX std: 9.204280884045419
DirectionY mean: 13.737616280014041
DirectionY std: 18.852744449999935
DirectionZ mean: 9.899565050415664
DirectionZ std: 13.278071955905812
```

You could also plot the signal to view the vibration signal, shown as follow



Ref answer:

```
import pandas as pd
import matplotlib.pyplot as pt
vibsig=pd.read_csv("VibrationData.csv")
print("DirectionX mean:",vibsig["DirectionX"].abs().mean())
print("DirectionX std:",vibsig["DirectionX"].std())
print("DirectionY mean:",vibsig["DirectionY"].abs().mean())
print("DirectionY std:",vibsig["DirectionY"].std())
print("DirectionZ mean:",vibsig["DirectionZ"].abs().mean())
print("DirectionZ std:",vibsig["DirectionZ"].std())

pt.plot(vibsig["DirectionX"])
pt.show()
```

Algorithm (Search and Sort):

3. Search and sort

(1)

Define a function RandSort(), to generate 200 random different numbers between 1 and 1000, sort them by selection sort and return the sorted list in descending order. For example, if the function() be called and the running results shown as follows:

```
>>> print(RandSort())
 [995, 972, 970, 962, 950, 949, 940, 938, 927, 917, 913, 912,
 881, 880, 878, 877, 857, 842, 841, 832, 822, 808, 807, 795,
 789, 764, 722, 659, 643, 642, 633, 620, 619, 606, 591, 553,
 548, 546, 543, 536, 532, 525, 523, 501, 496, 478, 473, 463,
 461, 458, 444, 426, 415, 409, 406, 401, 383, 372, 371, 353,
 352, 321, 319, 308, 289, 276, 243, 242, 236, 232, 219, 202,
 194, 178, 169, 168, 160, 159, 142, 136, 133, 123, 114, 111,
 82, 63, 56, 54, 45, 41, 29, 21, 8]
```

(2)

Define two functions, LineSearch(dataList, dataToSearch) and BinarySearch(dataList, dataToSearch), Input a random number, to search it by linear/sequential search and binary/bisection search method in returned list by function RandSort() defined in question(1) and count the number of comparison times for get the result, then print search results and return the number of comparisons times. For example, if the function() be called and the running results shown as follows:

```
>>> List1=[10,9,8,7,6,5,4,3]
    print("Compare times of BinarySearch:",
          BinarySearch(List1,3))
    print("Compare times of LineSearch:",
          LineSearch(List1,3))
 Get it, in position 7
 Compare times of BinarySearch: 4
 Get it, in position 7
 Compare times of LineSearch: 8
>>> List1=[10,9,8,7,6,5,4,3]
    print("Compare times of BinarySearch:",
          BinarySearch(List1,668))
    print("Compare times of LineSearch:",
          LineSearch(List1,668))
 Can't find it.
 Compare times of BinarySearch: 4
 Can't find it.
 Compare times of LineSearch: 8
```

(3)
For the same data sequence, the two functions are called 100 times each, calculate the average comparison times of two different method. Here is one possible running result for reference.

```
Average compare times of linear search is: 93.14
Average compare times of binary search is: 7.56
```

Ref answer:

```
import random
#(1)
# define a subfunction for selection sort.
def SelectionSort(L1):
    listlen=len(L1)
    for i in range(0,listlen-1):
        maxi=i
        for j in range(maxi+1,listlen):
            if L1[j]>L1[maxi]:
                maxi=j
        swap=L1[i]
        L1[i]=L1[maxi]
        L1[maxi]=swap
    return L1

def RandSort():
    ls=[]
    i=0
    while i<=100:
        ls.append(random.randint(1,1000))
        s=set(ls)
        ls=list(s)
        if len(ls)<=100:
            i+=1
#     ls.sort(reverse=True)
    lsort=SelectionSort(ls)
    return(lsort)

print(RandSort())

#(2)
def LineSearch(dataList,dataToSearch):
    compareTimes=0
    for i in range(0,len(dataList)):
        compareTimes+=1
        if dataList[i]==dataToSearch:
```

```
                print("Get it, in position",i)
                return compareTimes
        else:
            print("Can't find it.")
            return compareTimes

def BinarySearch(dataList, dataToSearch):
        starPos=0
        endPos=len(dataList)-1
        midPos=int((starPos+endPos)/2)
        compareTimes=1
        while dataList[midPos]!=dataToSearch and starPos<=endPos:
            compareTimes+=1
            if dataToSearch>dataList[midPos]:
                endPos=midPos-1
            if dataToSearch<dataList[midPos]:
                starPos=midPos+1
            midPos=int((starPos+endPos)/2)
        if dataList[midPos]==dataToSearch:
            print("Get it, in position",midPos)
            return compareTimes
        else:
            print("Can't find it.")
            return compareTimes
List1=[10,9,8,7,6,5,4,3]
print("Compare times of BinarySearch:",
          BinarySearch(List1,3))
print("Compare times of LineSearch:",
          LineSearch(List1,3))
#(3)
comTimesOfLineSearch=0
comTimesOfBinSearch=0
for i in range(0,100):
        dataList=RandSort()
        findData=random.randint(1,1000)
        comTimesOfLineSearch+=LineSearch(dataList,findData)
        comTimesOfBinSearch+=BinarySearch(dataList,findData)
#print (i)
print("Average compare times of linear search is:",comTimesOfLineSearch/100)
print("Average compare times of binary search is:",comTimesOfBinSearch/100)
```