assn2

May 22, 2023

[1]:
```
#Aishwarya kelgandre Roll no.73 batch T3
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
s1 =pd.Series(range(1,10,1))
s1
import pandas as pd
import numpy as np
student = pd.read_csv("E:\\TRINITY ACADEMY OF ENGINEERING PUNE\\TE␣
 ↪2022-23\\assignment\\dsbda\\csv\\StudentsPerformance.csv")
student.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   gender                       1000 non-null   object
 1   race/ethnicity               1000 non-null   object
 2   parental level of education  1000 non-null   object
 3   lunch                        1000 non-null   object
 4   test_preparation_course      1000 non-null   object
 5   math_score                   1000 non-null   int64 
 6   reading_score                1000 non-null   int64 
 7   writing_score                1000 non-null   int64 
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

[2]:
```
student.isnull().sum()
```

[2]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test_preparation_course        0
math_score                     0
reading_score                  0
writing_score                  0
```

```
dtype: int64
```

[3]:
```
student['math_score'].fillna(int(student['math_score'].mean()), inplace=True)
student.isnull().sum()
```

[3]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test_preparation_course        0
math_score                     0
reading_score                  0
writing_score                  0
dtype: int64
```

[43]:
```
student['reading_score'].fillna(method ='pad',inplace=True)
student.isnull().sum()
```

[43]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test_preparation_course        0
math_score                     0
reading_score                  0
writing_score                  0
dtype: int64
```

[12]:
```
student['writing_score'].fillna(int(student['writing_score'].median()),␣
 ↪inplace=True)

student.isnull().sum()
```

[12]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        0
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

[13]:
```
from numpy.random import seed
from numpy.random import randn
from numpy import mean
from numpy import std
```

```
seed(1)

data=5*randn(10000)+50
print('mean=%.3f stdv=%.3f' %(mean(data), std(data)))
```

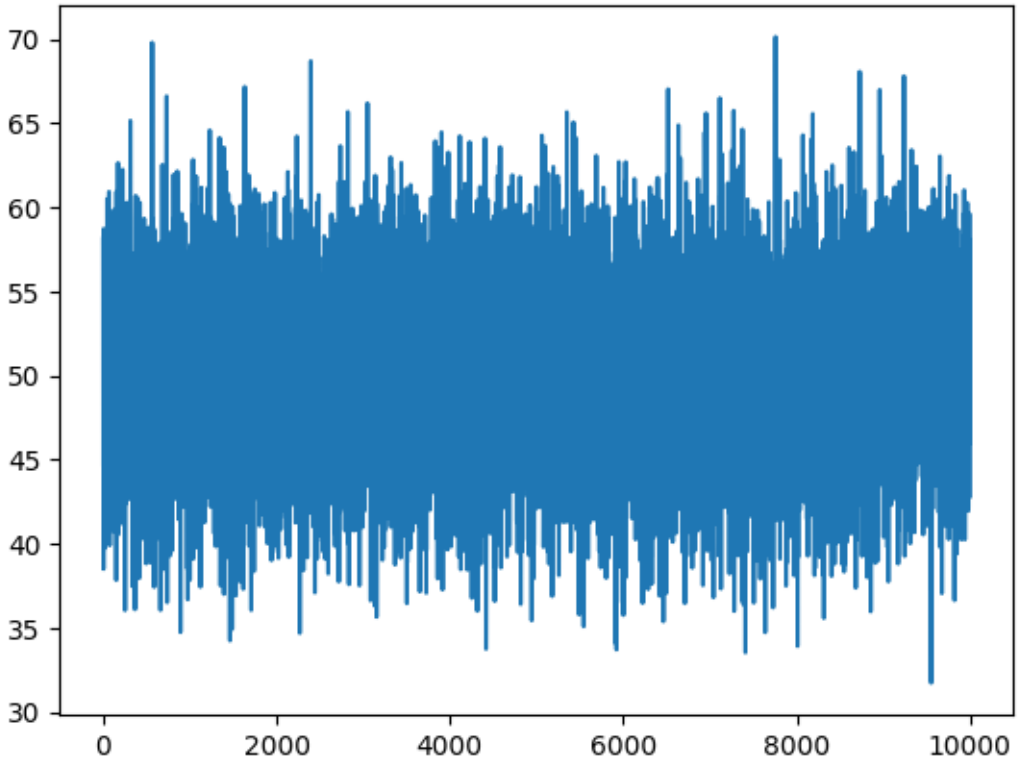
```
mean=50.049 stdv=4.994
```

[15]:
```
data_mean = mean(data)
data_std = std(data)
cut_off = data_std * 3
lower = data_mean - cut_off
upper = data_mean + cut_off



outliers=[x for x in data if x<lower or x > upper]
outliers
```

[15]:
```
[65.15428556186015,
 69.79301352018982,
 66.60539378085183,
 34.73117809786848,
 34.23321274904475,
 34.91984007395351,
 67.1633171589778,
 34.679293219474495,
 68.70124451852294,
 65.67523670043954,
 66.19171598376188,
 33.73482882511691,
 65.66014864070253,
 65.06377284118616,
 34.0469182658796,
 33.6969245211173,
 67.02151137874486,
 65.59239795391275,
 66.49270261640393,
 65.74492012609815,
 33.525707966507426,
 34.72183379792847,
 70.1342452227369,
 33.90433947188079,
 65.55945915508362,
 68.06638503541573,
 66.99057828251213,
 67.80436660352774,
 31.717799503726024]
```

[16]:
```
import matplotlib.pyplot as plt
plt.plot(data)
```

[16]:
```
[<matplotlib.lines.Line2D at 0x1a9e4efec90>]
```



[17]:
```
from numpy.lib.function_base import percentile
q25=percentile(data,25)
q75=percentile(data,75)
IQR=q75-q25
cut_off_IQR= IQR * 2
lower=q25-cut_off_IQR
upper= q75 +cut_off_IQR
```

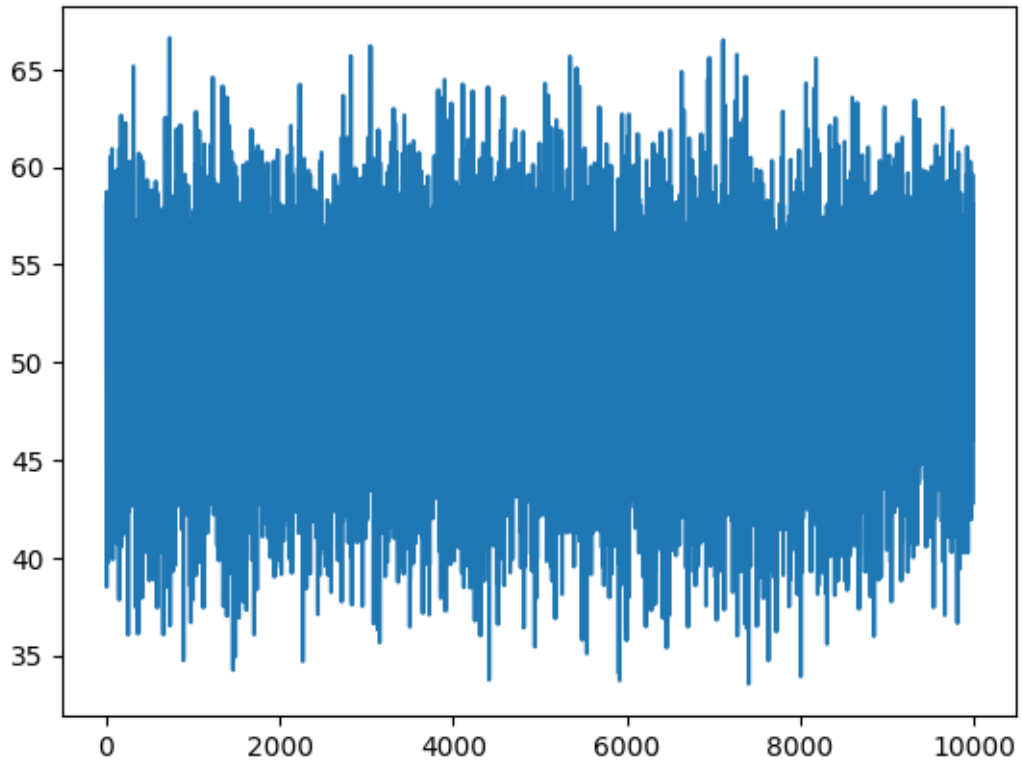
[18]:
```
outliers_IQR = [x for x in data if x < lower or x > upper]
outliers_IQR
```

[18]:
```
[69.79301352018982,
 67.1633171589778,
 68.70124451852294,
 67.02151137874486,
 70.1342452227369,
 68.06638503541573,
```

```
66.99057828251213,
67.80436660352774,
31.717799503726024]
```

[19]:
```
outliers_removed=[x for x in data if x>=lower and x<=upper]
plt.plot(outliers_removed)
```

[19]:
```
[<matplotlib.lines.Line2D at 0x1a9e713ec90>]
```



[ ]: