

```
In [3]: import pandas as pd
import numpy as np

In [14]: import pandas as pd
numbers = [23.4,45.6,43.4,34.2,34.7,67.5,56.4]
pop = pd.Series(list(numbers))
pop

Out[14]:
0    34.2
1    67.5
2    34.7
3    23.4
4    56.4
5    43.4
6    45.6
dtype: float64

In [15]: pop.name=" population of seven nation"

In [16]: pop

Out[16]:
0    34.2
1    67.5
2    34.7
3    23.4
4    56.4
5    43.4
6    45.6
Name: population of seven nation, dtype: float64

In [17]: # selecting elements
pop[9]

Out[17]:
34.2

In [18]: pop[4]

Out[18]:
56.4

In [19]: pop.index=['kenya','uganda','zambia','rwanda','Tanzania','somalia','Djibouti']

In [20]: pop

Out[20]:
kenya    34.2
uganda   67.5
zambia   34.7
rwanda   23.4
Tanzania 56.4
somalia  43.4
Djibouti 45.6
Name: population of seven nation, dtype: float64

In [22]: #pandas indexing and conditional selection
pop['zambia']

Out[22]:
34.7

In [23]: pop['kenya']

Out[23]:
34.2

In [24]: pop.iloc[0]

Out[24]:
34.2

In [25]: pop.iloc[-1]

Out[25]:
45.6

In [30]: pop['kenya':'uganda']

Out[30]:
kenya    34.2
uganda   67.5
Name: population of seven nation, dtype: float64

In [31]: #conditional selection "boolean arrays"
pop>37

Out[31]:
kenya    False
uganda    True
zambia   False
rwanda   False
Tanzania  True
somalia   True
Djibouti  True
Name: population of seven nation, dtype: bool

In [32]: pop<70

Out[32]:
kenya    True
uganda    True
zambia    True
rwanda    True
Tanzania  True
somalia    True
Djibouti  True
Name: population of seven nation, dtype: bool

In [33]: pop[pop>37]

Out[33]:
uganda    67.5
Tanzania  56.4
somalia   43.4
Djibouti  45.6
Name: population of seven nation, dtype: float64

In [34]: pop.mean()

Out[34]:
43.600000000000001

In [35]: pop.std()

Out[35]:
14.802589863488977

In [38]: pop>pop.mean()-pop.std()/2

Out[38]:
kenya    False
uganda    True
zambia   False
rwanda   False
Tanzania  True
somalia    True
Djibouti  True
Name: population of seven nation, dtype: bool

In [39]: pop[pop>pop.mean()-pop.std()/2]

Out[39]:
uganda    67.5
Tanzania  56.4
somalia   43.4
Djibouti  45.6
Name: population of seven nation, dtype: float64

In [40]: #modifying series
pop['kenya']=67.9

In [41]: pop

Out[41]:
kenya    67.9
uganda   67.5
zambia   34.7
rwanda   23.4
Tanzania 56.4
somalia  43.4
Djibouti 45.6
Name: population of seven nation, dtype: float64

In [42]: pop[pop>37]=100

In [43]: pop

Out[43]:
kenya    100.0
uganda   100.0
zambia   34.7
rwanda   23.4
Tanzania 100.0
somalia  100.0
Djibouti 100.0
Name: population of seven nation, dtype: float64

In [40]: #Data Frames in pandas
import pandas as pd
df=pd.DataFrame(
{'employees':['mike','cynthia','sunshines','robin','kamau','ruger','otile','maina','Mejja','kiprotich'],
'salary':[75000,53000,34000,50000,34000,45000,45700,34000,23000,34600]},
'job description':['manager','assistant','clerk','secretary','gateman','driver','store keeper','intern','consultant','delivery'],
'rating':['good','good','fair','fair','fair','fair','good','excellent','good','good','fair']}
,columns=['employees','salary','job description','rating'])

In [41]: df

Out[41]:
   employees  salary job description  rating
0      mike  75000      manager      good
1  cynthia  53000    assistant      good
2  sunshines 34000      clerk      fair
3      robin 56000    secretary      fair
4      kamau 34000    gateman      fair
5      ruger 45000      driver      good
6      otile 45700  store keeper  excellent
7      maina 34000      intern      good
8      Mejja 23000    consultant      good
9  kiprotich 34600    delivery      fair

In [42]: df.iloc[2]

Out[42]:
employees      sunshines
salary         34000
job description  clerk
rating          fair
Name: 2, dtype: object

In [43]: df['employees']

Out[43]:
0      mike
1  cynthia
2  sunshines
3      robin
4      kamau
5      ruger
6      otile
7      maina
8      Mejja
9  kiprotich
Name: employees, dtype: object

In [44]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   employees             10 non-null    object
1   salary                10 non-null    int64
2   job description        10 non-null    object
3   rating                10 non-null    object
dtypes: int64(1), object(3)
memory usage: 448.0+ bytes

In [45]: df.shape

Out[45]:
(10, 4)

In [46]: df.describe()

Out[46]:
           salary
count  10.000000
mean   43430.000000
std    14966.632259
min    23000.000000
25%    34000.000000
50%    39800.000000
75%    51175.000000
max    75000.000000

In [47]: df['salary'].mean()

Out[47]:
43430.0

In [48]: #indexing ,selecting
df.iloc[0]

Out[48]:
employees      mike
salary         75000
job description  manager
rating          good
Name: 0, dtype: object

In [49]: df.iloc[-1]

Out[49]:
employees      kiprotich
salary         34600
job description  delivery
rating          fair
Name: 9, dtype: object

In [50]: df['salary']=100000

In [51]: df

Out[51]:
   employees  salary job description  rating
0      mike  100000      manager      good
1  cynthia  100000    assistant      good
2  sunshines 100000      clerk      fair
3      robin 100000    secretary      fair
4      kamau 100000    gateman      fair
5      ruger 100000      driver      good
6      otile 100000  store keeper  excellent
7      maina 100000      intern      good
8      Mejja 100000    consultant      good
9  kiprotich 100000    delivery      fair

In [52]: df['place']='Nakuru','Eldoret','Makueni','Nairobi','Nandi','Narok','Emali','Kitale','Marakwet','Naivasha'

In [53]: df

Out[53]:
   employees  salary job description  rating  place
0      mike  100000      manager      good  Nakuru
1  cynthia  100000    assistant      good  Eldoret
2  sunshines 100000      clerk      fair  Makueni
3      robin 100000    secretary      fair  Nairobi
4      kamau 100000    gateman      fair  Nandi
5      ruger 100000      driver      good  Narok
6      otile 100000  store keeper  excellent  Emali
7      maina 100000      intern      good  Kitale
8      Mejja 100000    consultant      good  Marakwet
9  kiprotich 100000    delivery      fair  Naivasha

In [56]: df['age']=20,23,30,32,34,35,19,45,37,31

In [57]: df

Out[57]:
   employees  salary job description  rating  place  age
0      mike  100000      manager      good  Nakuru   20
1  cynthia  100000    assistant      good  Eldoret   23
2  sunshines 100000      clerk      fair  Makueni   30
3      robin 100000    secretary      fair  Nairobi   32
4      kamau 100000    gateman      fair  Nandi     34
5      ruger 100000      driver      good  Narok     35
6      otile 100000  store keeper  excellent  Emali     19
7      maina 100000      intern      good  Kitale     45
8      Mejja 100000    consultant      good  Marakwet  37
9  kiprotich 100000    delivery      fair  Naivasha  31

In [58]: df.describe()

Out[58]:
           salary           age
count    10.0    10.000000
mean   100000.0    30.600000
std      10.0     8.071899
min    100000.0    19.000000
25%    100000.0    24.750000
50%    100000.0    31.500000
75%    100000.0    34.750000
max    100000.0    45.000000

In [59]: from matplotlib import pyplot as plt

In [60]: A=df['salary']

In [61]: B=df['age']

In [62]: A

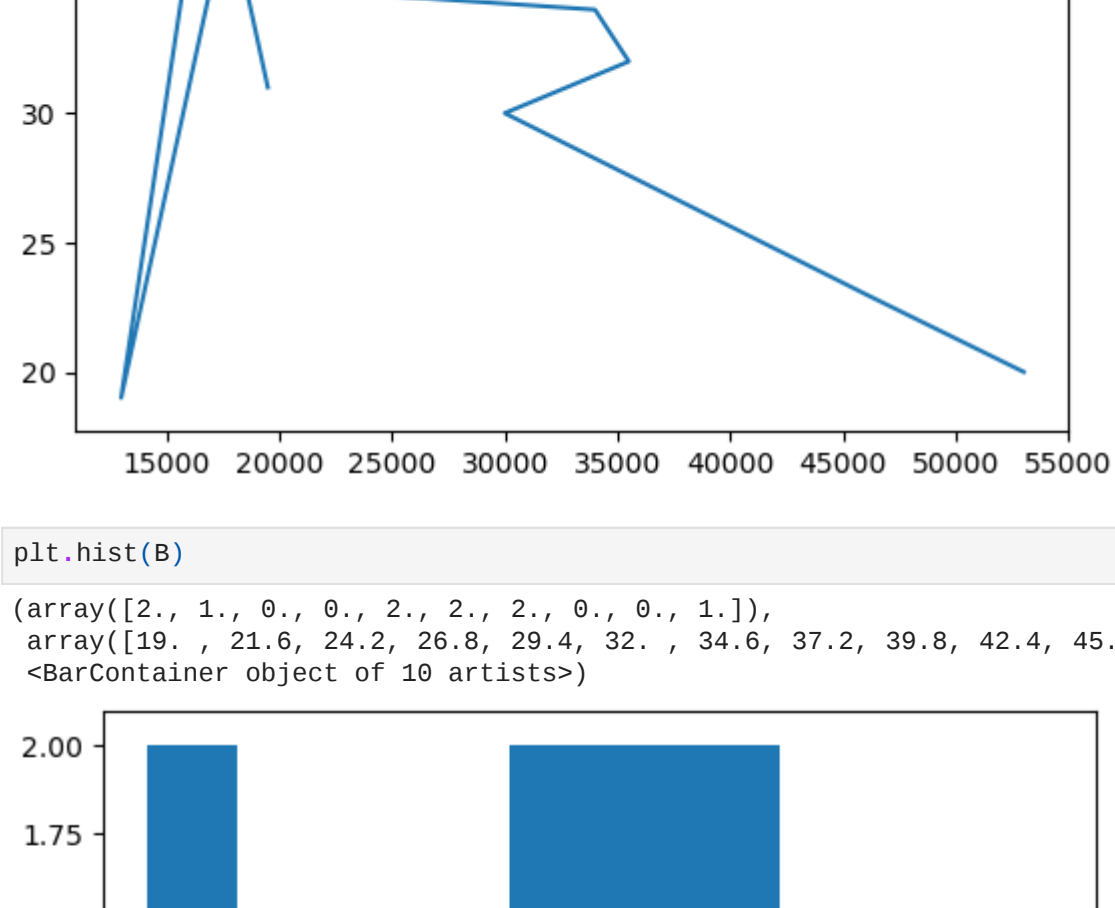
Out[62]:
0    100000
1    100000
2    100000
3    100000
4    100000
5    100000
6    100000
7    100000
8    100000
9    100000
Name: salary, dtype: int64

In [63]: df['salary']=53000,46000,30000,35500,34000,17000,13000,17500,18000,19500

In [64]: A=df['salary']

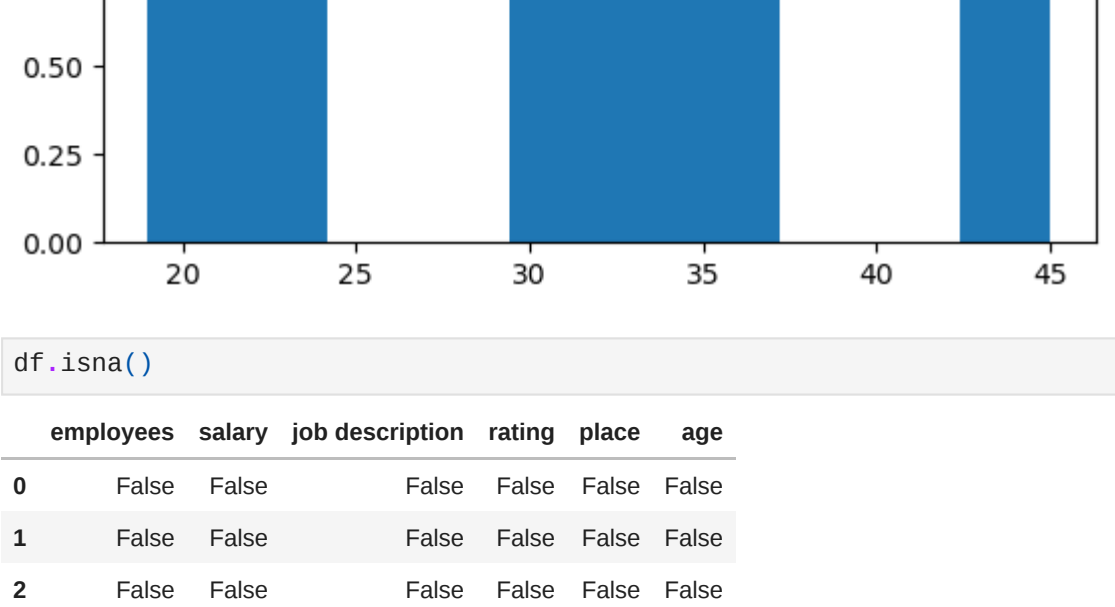
In [65]: plt.plot(A,B)

Out[65]:
<matplotlib.lines.Line2D at 0x297b9868f40>



In [66]: plt.hist(B)

Out[66]:
(array([2., 1., 0., 0., 2., 2., 2., 0., 0., 1.]),
array([19., 21.6, 24.2, 26.8, 29.4, 32., 34.6, 37.2, 39.8, 42.4, 45. ]),
<BarContainer object of 10 artists>)



In [38]: df.isna()

Out[38]:
   employees  salary job description  rating  place  age
0      False    False             False  False  False  False
1      False    False             False  False  False  False
2      False    False             False  False  False  False
3      False    False             False  False  False  False

In [ ]:
```