

```
In [88]: # read the data
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [89]: bank_data=pd.read_csv(r"F:\FSDS\Data Files\bank.csv", sep=';')
```

head

```
In [59]: bank_data.head()
```

```
Out[59]:
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	du
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	may	
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	

tail

```
In [60]: bank_data.tail()
```

```
Out[60]:
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	
4516	33	services	married	secondary	no	-333	yes	no	cellular	30	jul	
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown	9	may	
4518	57	technician	married	secondary	no	295	no	no	cellular	19	aug	
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular	6	feb	
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	3	apr	

```
In [61]: cat=bank_data.select_dtypes(include='object').columns
num=bank_data.select_dtypes(exclude='object').columns
```

```
In [62]: num
```

```
Out[62]: Index(['age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous'], dtype='object')
```

```
In [63]: cat
```

```
Out[63]: Index(['job', 'marital', 'education', 'default', 'housing', 'loan', 'contact',
               'month', 'poutcome', 'y'],
               dtype='object')
```

```
In [64]: bank_data.shape
```

```
Out[64]: (4521, 17)
```

- on numerical column we need to apply
- min
- max
- count
- len
- mean
- median
- 25%
- 50%
- 75%

```
In [65]: len(bank_data['balance'])
```

```
Out[65]: 4521
```

```
In [66]: bal_data=bank_data['balance']  
len(bal_data)
```

```
Out[66]: 4521
```

mean

```
In [67]: bank_data['balance'].mean()
```

```
Out[67]: 1422.6578190665782
```

```
In [68]: round(bank_data['balance'].mean(),2)
```

```
Out[68]: 1422.66
```

median

```
In [69]: bank_data['balance'].median()
```

```
Out[69]: 444.0
```

```
In [70]: round(bank_data['balance'].median(), 2)
```

```
Out[70]: 444.0
```

using numpy

```
In [71]: np.mean(bank_data['balance'])
```

```
Out[71]: 1422.6578190665782
```

```
In [72]: round(np.mean(bank_data['balance']))
```

```
Out[72]: 1423
```

```
In [73]: round(np.median(bank_data['balance']),2)
```

```
Out[73]: 444.0
```

```
In [74]: bank_data['balance'].mode()
```

```
Out[74]: 0    0
         Name: balance, dtype: int64
```

```
In [75]: import warnings
         warnings.filterwarnings('ignore')

         bal_data=bank_data['balance']
         bal_count=len(bal_data)
         bal_min=round(bal_data.min(),2)
         bal_max=round(bal_data.max(),2)
         bal_std=round(bal_data.std(),2)
         bal_mean=round(bal_data.mean(),2)
         bal_median=round(bal_data.median(),2)
```

```
In [76]: print(f'the count of bal is: {bal_count}')
         print(f'the min of bal is : {bal_min}')
         print(f'the max of bal is : {bal_max}')
         print(f'the std of bal is : {bal_std}')
         print(f'the mean of bal is: {bal_mean}')
         print(f'the median of bal is: {bal_median}')
```

```
the count of bal is: 4521
the min of bal is : -3313
the max of bal is : 71188
the std of bal is : 3009.64
the mean of bal is: 1422.66
the median of bal is: 444.0
```

```
In [77]: bal_data=bank_data['balance']
         bal_count=len(bal_data)
         bal_min=round(bal_data.min(),2)
         bal_max=round(bal_data.max(),2)
         bal_std=round(bal_data.std(),2)
         bal_mean=round(bal_data.mean(),2)
         bal_median=round(bal_data.median(),2)

         idx=['Count', 'Min', 'Max', 'Std', 'Mean', 'Median']
         data=[bal_count, bal_min, bal_max, bal_std, bal_mean, bal_median]
         pd.DataFrame(data, index=idx, columns=['balance'])
```

Out[77]:

balance	
Count	4521.00
Min	-3313.00
Max	71188.00
Std	3009.64
Mean	1422.66
Median	444.00

percentile

```
In [78]: bal_data=bank_data['balance']  
np.percentile(bal_data,25)
```

Out[78]: 69.0

```
In [79]: bal_data=bank_data['balance']  
np.quantile(bal_data,0.75)
```

Out[79]: 1480.0

```
In [80]: 25*4521/100, 50*4521/100, 75*4521/100
```

Out[80]: (1130.25, 2260.5, 3390.75)

```
In [81]: bal_data=bank_data['balance']  
bal_50p=np.percentile(bal_data,50)  
con=bal_data<bal_50p  
len(bal_data[con])
```

Out[81]: 2259

```
In [82]: bal_data=bank_data['balance']  
bal_75p=np.percentile(bal_data,75)  
con=bal_data<bal_75p  
len(bal_data[con])
```

Out[82]: 3390

```
In [83]: bal_data=bank_data['balance']  
bal_count=len(bal_data)  
bal_min=round(bal_data.min(),2)  
bal_max=round(bal_data.max(),2)  
bal_std=round(bal_data.std(),2)  
bal_mean=round(bal_data.mean(),2)  
bal_median=round(bal_data.median(),2)  
bal_25p=np.percentile(bal_data,25)  
bal_50p=np.percentile(bal_data,50)  
bal_75p=np.percentile(bal_data, 75)  
  
idx=['Count', 'Min', 'Max', 'Std', 'Mean', 'Median', '25%', '50%', '75%']
```

```
data=[bal_count, bal_min, bal_max, bal_std, bal_mean, bal_median, bal_25p, bal_50p, bal_75p]
pd.DataFrame(data, index=idx, columns=['balance'])
```

Out[83]:

balance	
Count	4521.00
Min	-3313.00
Max	71188.00
Std	3009.64
Mean	1422.66
Median	444.00
25%	69.00
50%	444.00
75%	1480.00

In [84]: num

Out[84]: Index(['age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous'], dtype='object')

In [85]:

```
for i in num:
    bal_data=bank_data[i]
    bal_data=bank_data['balance']
    bal_count=len(bal_data)
    bal_min=round(bal_data.min(),2)
    bal_max=round(bal_data.max(),2)
    bal_std=round(bal_data.std(),2)
    bal_mean=round(bal_data.mean(),2)
    bal_median=round(bal_data.median(),2)
    bal_25p=np.percentile(bal_data,25)
    bal_50p=np.percentile(bal_data,50)
    bal_75p=np.percentile(bal_data, 75)

    idx=['Count', 'Min', 'Max', 'Std', 'Mean', 'Median', '25%', '50%', '75%']
    data.append([bal_count, bal_min, bal_max, bal_std, bal_mean, bal_median, bal_25p, bal_50p, bal_75p])

print(data)
```

```
[4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0, [4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0], [4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0], [4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0], [4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0], [4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0], [4521, -3313, 71188, 3009.64, 1422.66, 444.0, 69.0, 444.0, 1480.0]]
```

In [92]:

```
for i in num:
    data=[]
    bal_data=bank_data[i]
    bal_data=bank_data['balance']
    bal_count=len(bal_data)
    bal_min=round(bal_data.min(),2)
```

```

bal_max=round(bal_data.max(),2)
bal_std=round(bal_data.std(),2)
bal_mean=round(bal_data.mean(),2)
bal_median=round(bal_data.median(),2)
bal_25p=np.percentile(bal_data,25)
bal_50p=np.percentile(bal_data,50)
bal_75p=np.percentile(bal_data, 75)

idx=['Count', 'Min', 'Max', 'Std', 'Mean', 'Median', '25%', '50%', '75%']
data.append([bal_count, bal_min, bal_max, bal_std, bal_mean, bal_median, bal_25p,
pd.DataFrame(data, index=num, columns=idx)

```

Out[92]:

	Count	Min	Max	Std	Mean	Median	25%	50%	75%
age	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0
balance	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0
day	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0
duration	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0
campaign	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0
pdays	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0
previous	4521	-3313	71188	3009.64	1422.66	444.0	69.0	444.0	1480.0

In [94]:

```

for i in num:
    data=[]
    bal_data=bank_data[i]
    bal_data=bank_data['balance']
    bal_count=len(bal_data)
    bal_min=round(bal_data.min(),2)
    bal_max=round(bal_data.max(),2)
    bal_std=round(bal_data.std(),2)
    bal_mean=round(bal_data.mean(),2)
    bal_median=round(bal_data.median(),2)
    bal_25p=np.percentile(bal_data,25)
    bal_50p=np.percentile(bal_data,50)
    bal_75p=np.percentile(bal_data, 75)

    idx=['Count', 'Min', 'Max', 'Std', 'Mean', 'Median', '25%', '50%', '75%']
    data.append([bal_count, bal_min, bal_max, bal_std, bal_mean, bal_median, bal_25p,
pd.DataFrame(data, index=num, columns=idx).T

```

Out[94]:

	age	balance	day	duration	campaign	pdays	previous
Count	4521.00	4521.00	4521.00	4521.00	4521.00	4521.00	4521.00
Min	-3313.00	-3313.00	-3313.00	-3313.00	-3313.00	-3313.00	-3313.00
Max	71188.00	71188.00	71188.00	71188.00	71188.00	71188.00	71188.00
Std	3009.64	3009.64	3009.64	3009.64	3009.64	3009.64	3009.64
Mean	1422.66	1422.66	1422.66	1422.66	1422.66	1422.66	1422.66
Median	444.00	444.00	444.00	444.00	444.00	444.00	444.00
25%	69.00	69.00	69.00	69.00	69.00	69.00	69.00
50%	444.00	444.00	444.00	444.00	444.00	444.00	444.00
75%	1480.00	1480.00	1480.00	1480.00	1480.00	1480.00	1480.00

In [93]: bank_data.describe()

Out[93]:

	age	balance	day	duration	campaign	pdays	previous
count	4521.000000	4521.000000	4521.000000	4521.000000	4521.000000	4521.000000	4521.000000
mean	41.170095	1422.657819	15.915284	263.961292	2.793630	39.766645	0.542579
std	10.576211	3009.638142	8.247667	259.856633	3.109807	100.121124	1.693562
min	19.000000	-3313.000000	1.000000	4.000000	1.000000	-1.000000	0.000000
25%	33.000000	69.000000	9.000000	104.000000	1.000000	-1.000000	0.000000
50%	39.000000	444.000000	16.000000	185.000000	2.000000	-1.000000	0.000000
75%	49.000000	1480.000000	21.000000	329.000000	3.000000	-1.000000	0.000000
max	87.000000	71188.000000	31.000000	3025.000000	50.000000	871.000000	25.000000



In []: