In [24]:
```
#histogram with bank data
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [26]:
```
file_path=r"C:\Users\Mrityunjay\Desktop\Data science naresh it\Class notes by me\ba
bank_df=pd.read_csv(file_path,sep=";")
bank_df
```

Out[26]:

	age	job	marital	education	default	balance	housing	loan	contact	da
0	30	unemployed	married	primary	no	1787	no	no	cellular	
1	33	services	married	secondary	no	4789	yes	yes	cellular	
2	35	management	single	tertiary	no	1350	yes	no	cellular	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	
...	...	...	...	...	...	...	...	...	...	
4516	33	services	married	secondary	no	-333	yes	no	cellular	
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown	
4518	57	technician	married	secondary	no	295	no	no	cellular	
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular	
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular	

4521 rows × 17 columns

In [28]:
```
num_colm=[keys for keys,values in dict(bank_df.dtypes).items() if values!="object"]
num_colm
```

Out[28]:
```
['age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous']
```

In [30]:
```
len(bank_df["age"])
```

Out[30]:
```
4521
```

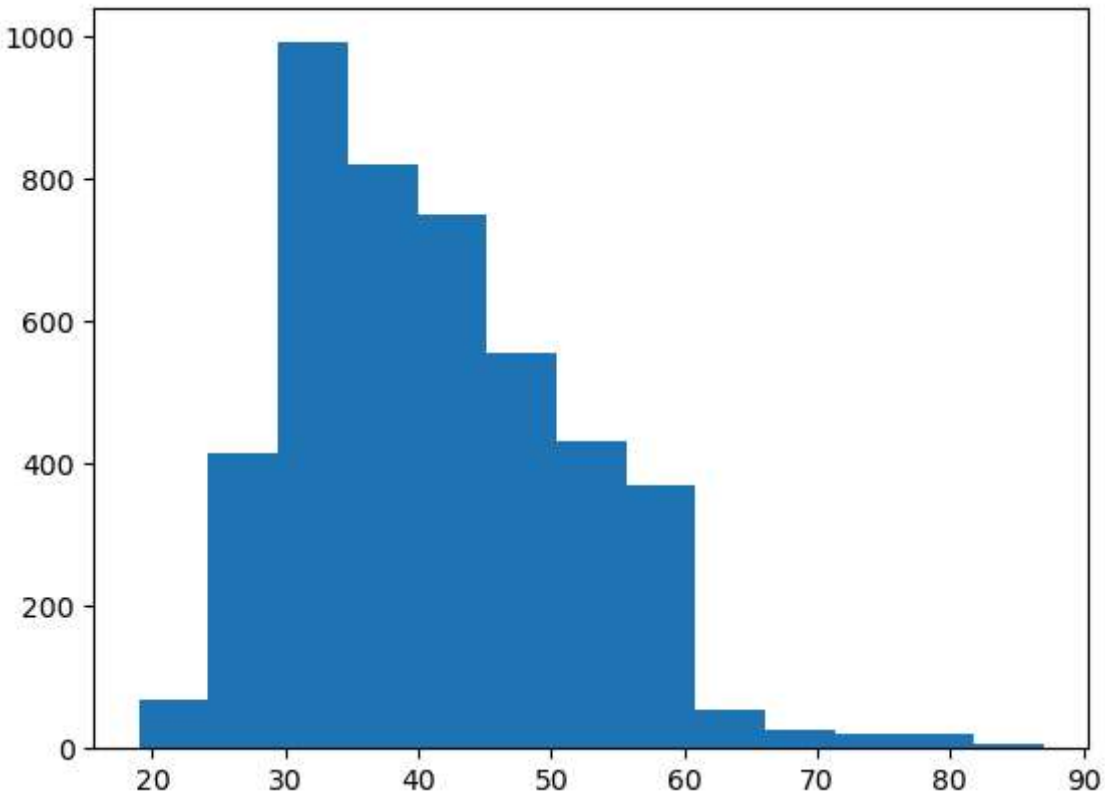
In [32]:
```
#1024
2**10
```

Out[32]:
```
1024
```

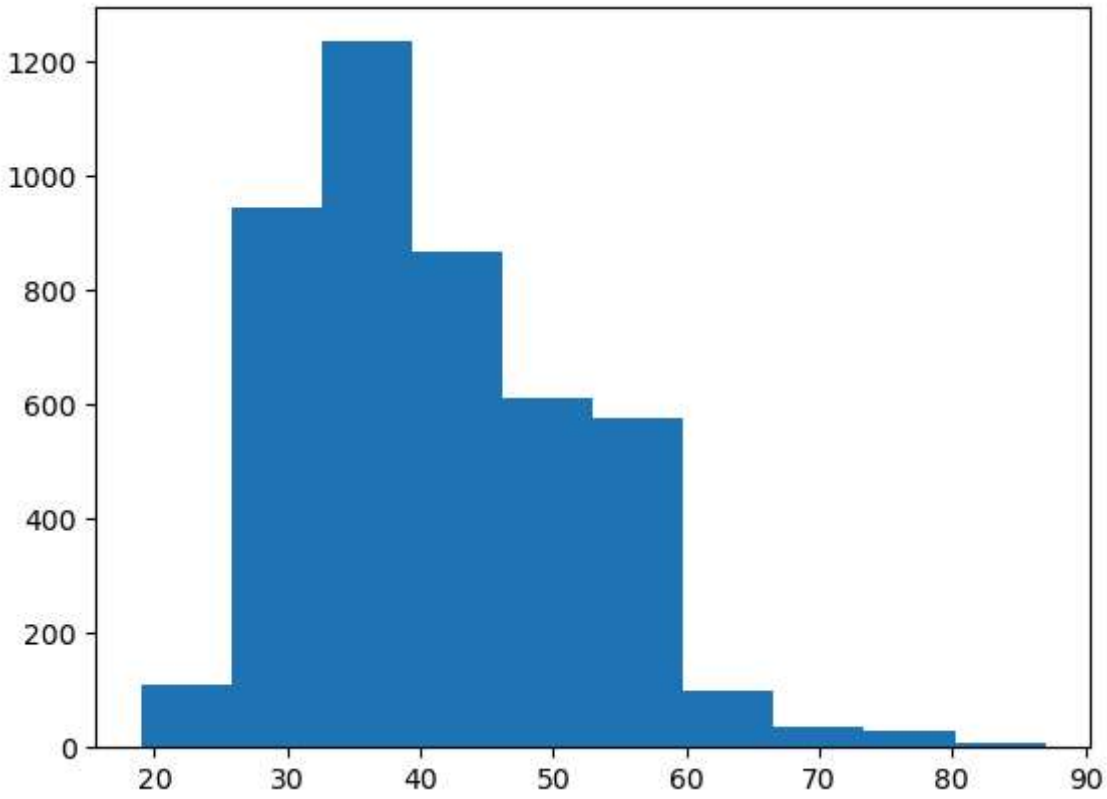
In [34]:
```
#how many bins 2^k
2**12,2**13
```

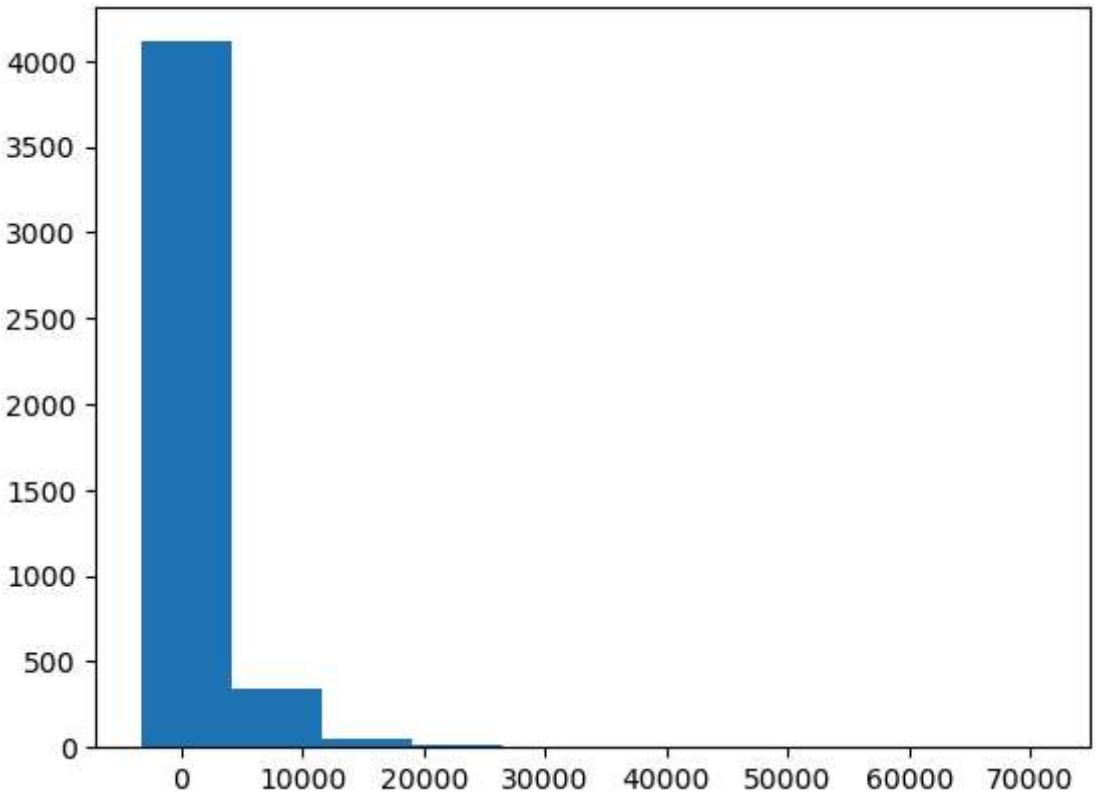
Out[34]: (4096, 8192)

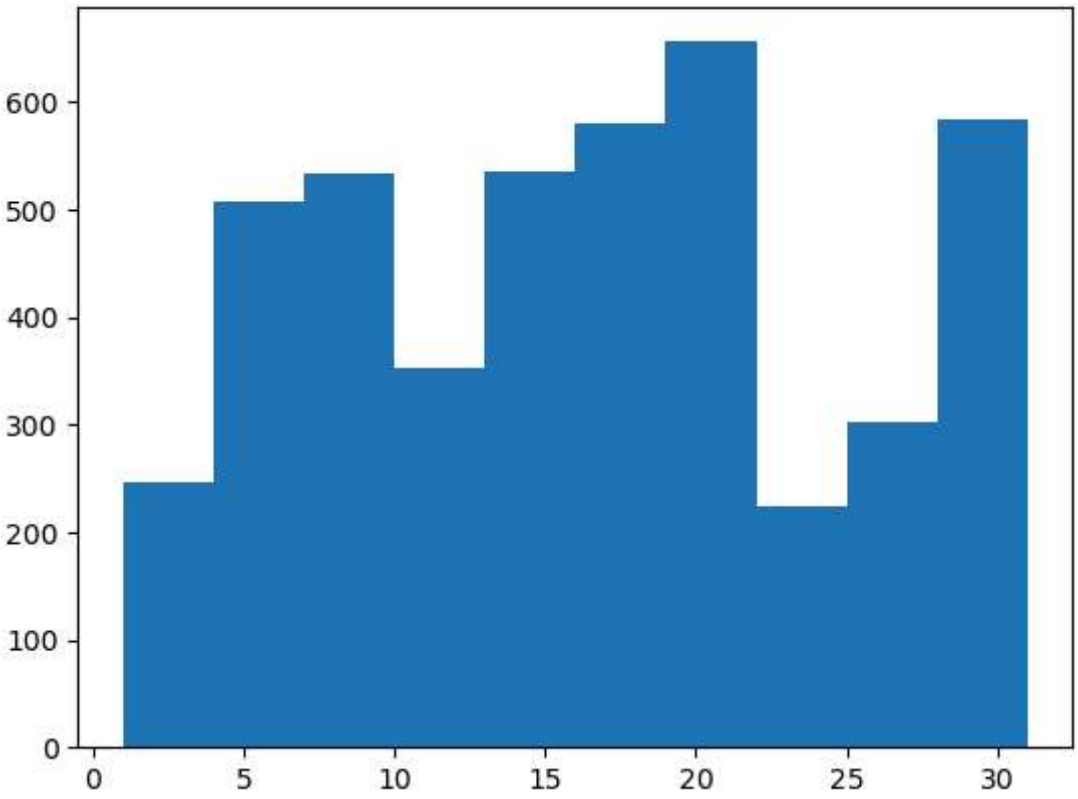
In [36]:
```
#histogram
age_data=bank_df["age"]
plt.hist(age_data,bins=13)
plt.show()
```

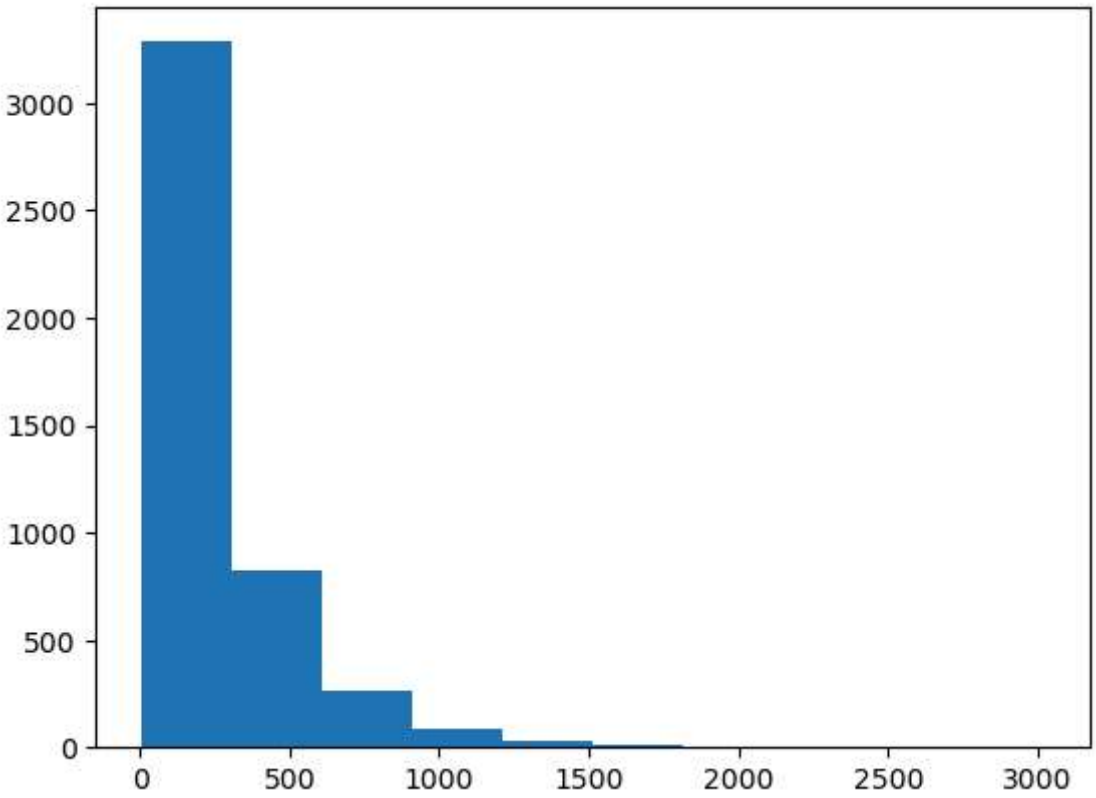


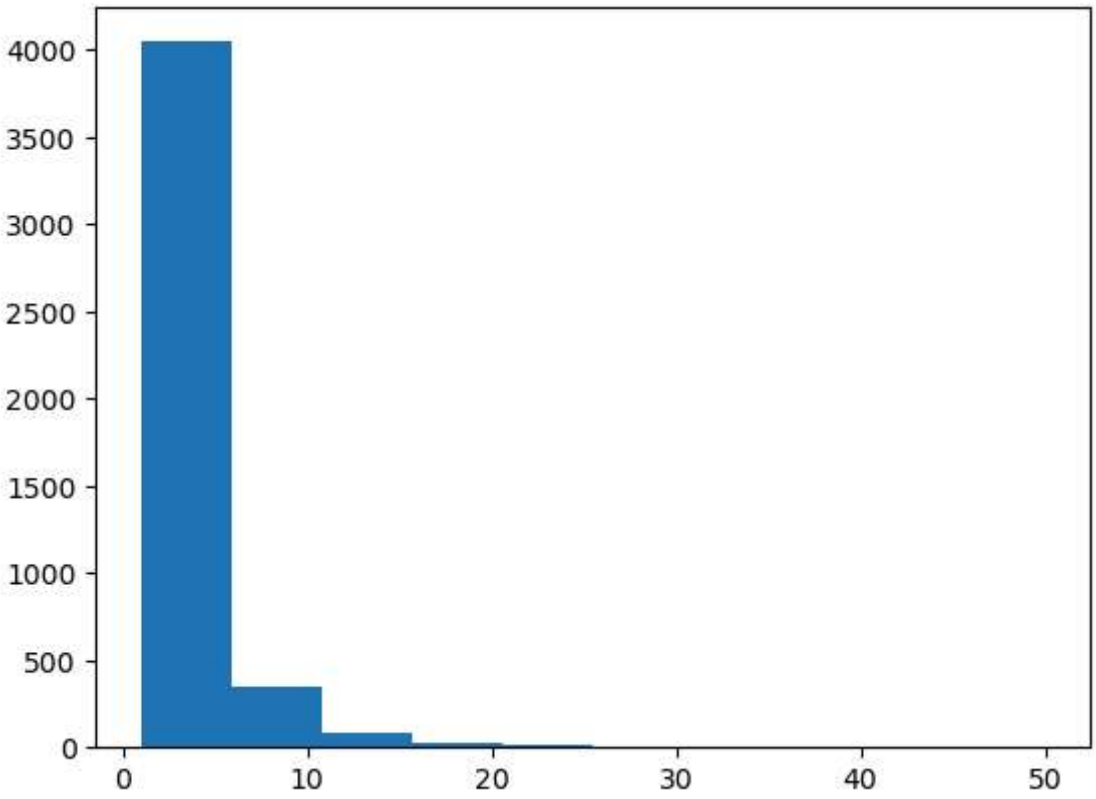
In [38]:
```
for i in num_colm:
    data=bank_df[i]
    plt.hist(data)
    plt.show()
```

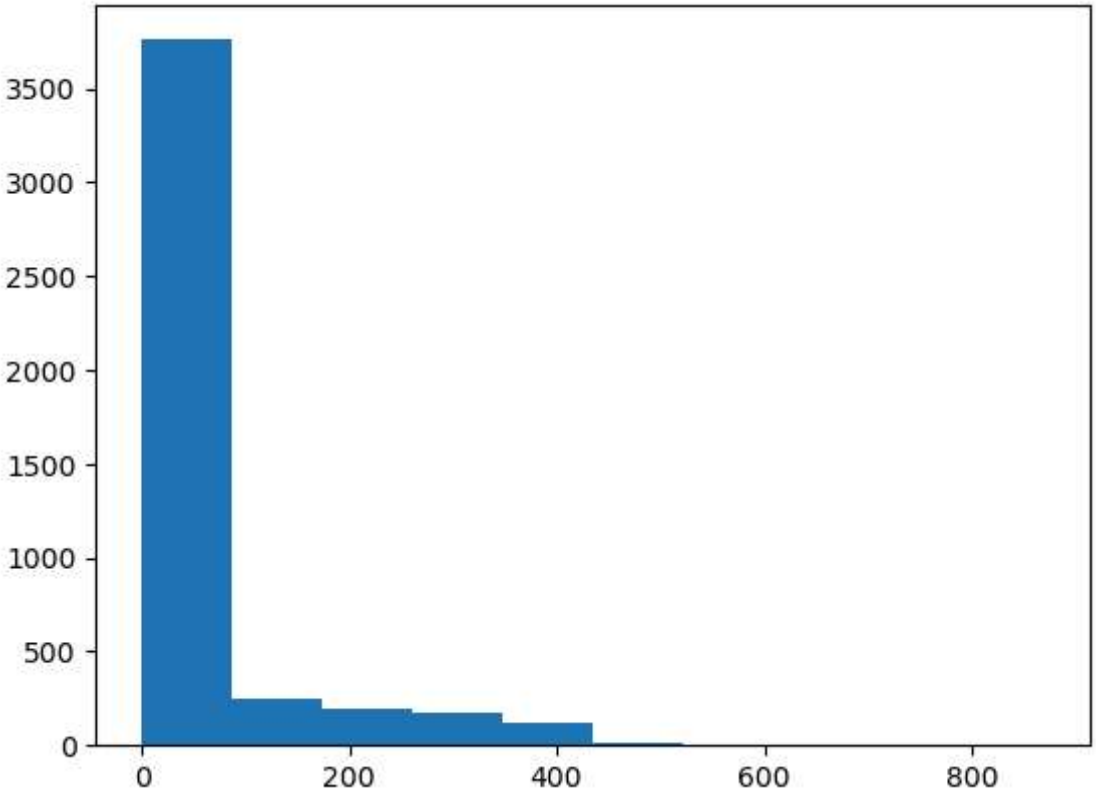

1200
1000
800
600
400
200
0
20
30
40
50
60
70
80
90

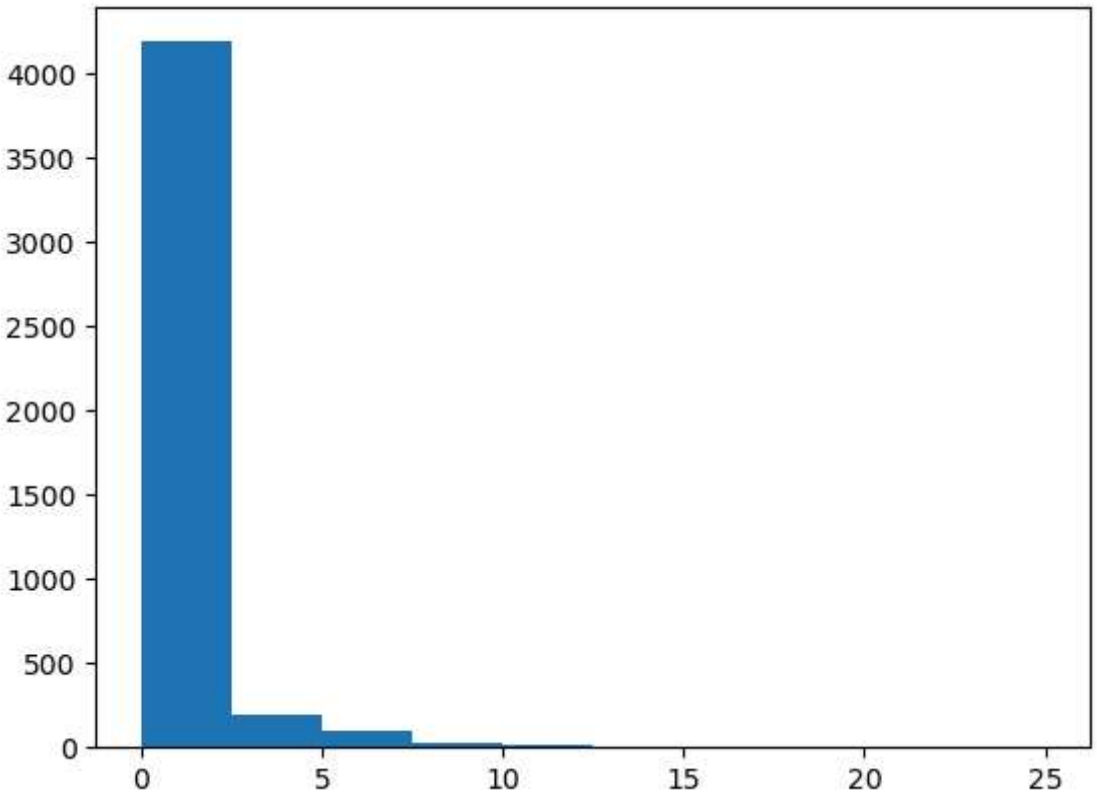

4000
3500
3000
2500
2000
1500
1000
500
0
0
10000
20000
30000
40000
50000
60000
70000


600
500
400
300
200
100
0
0
5
10
15
20
25
30


3000
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
3000


4000
3500
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50


3500
3000
2500
2000
1500
1000
500
0
0
200
400
600
800



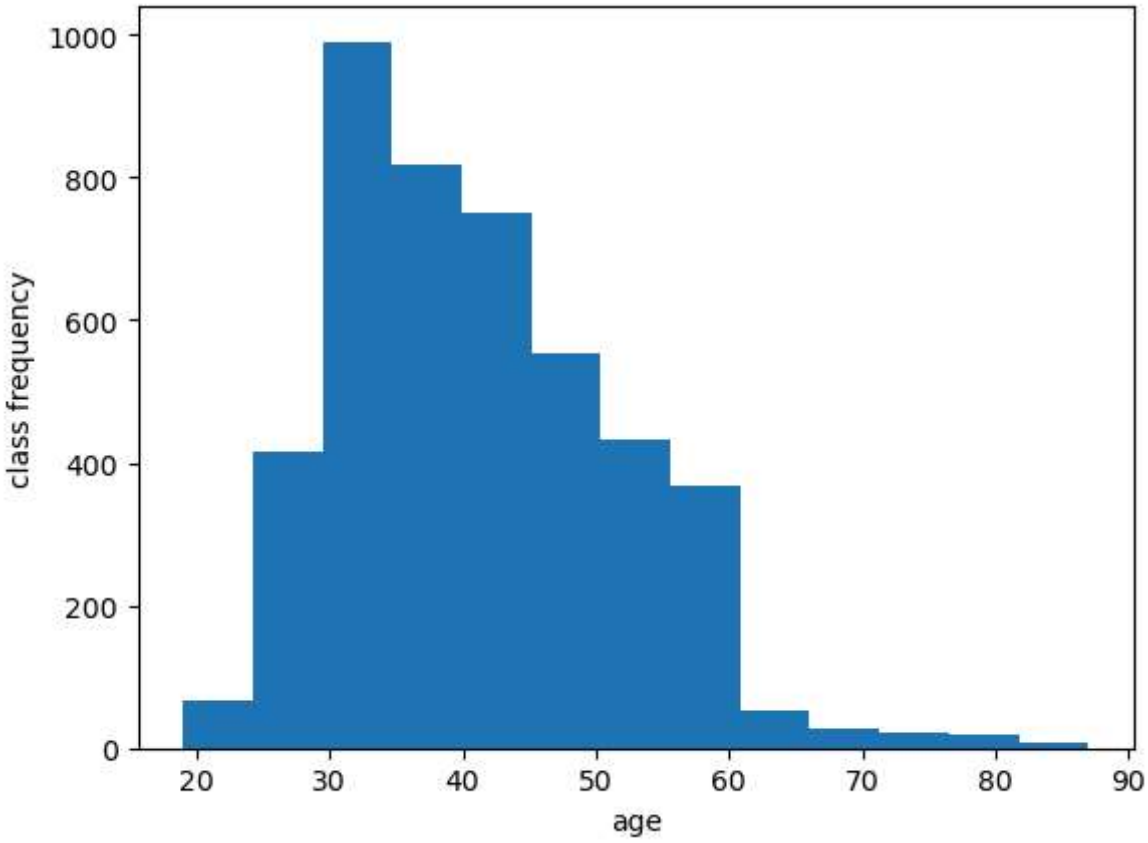
In [40]:
```
4521%2**13
```

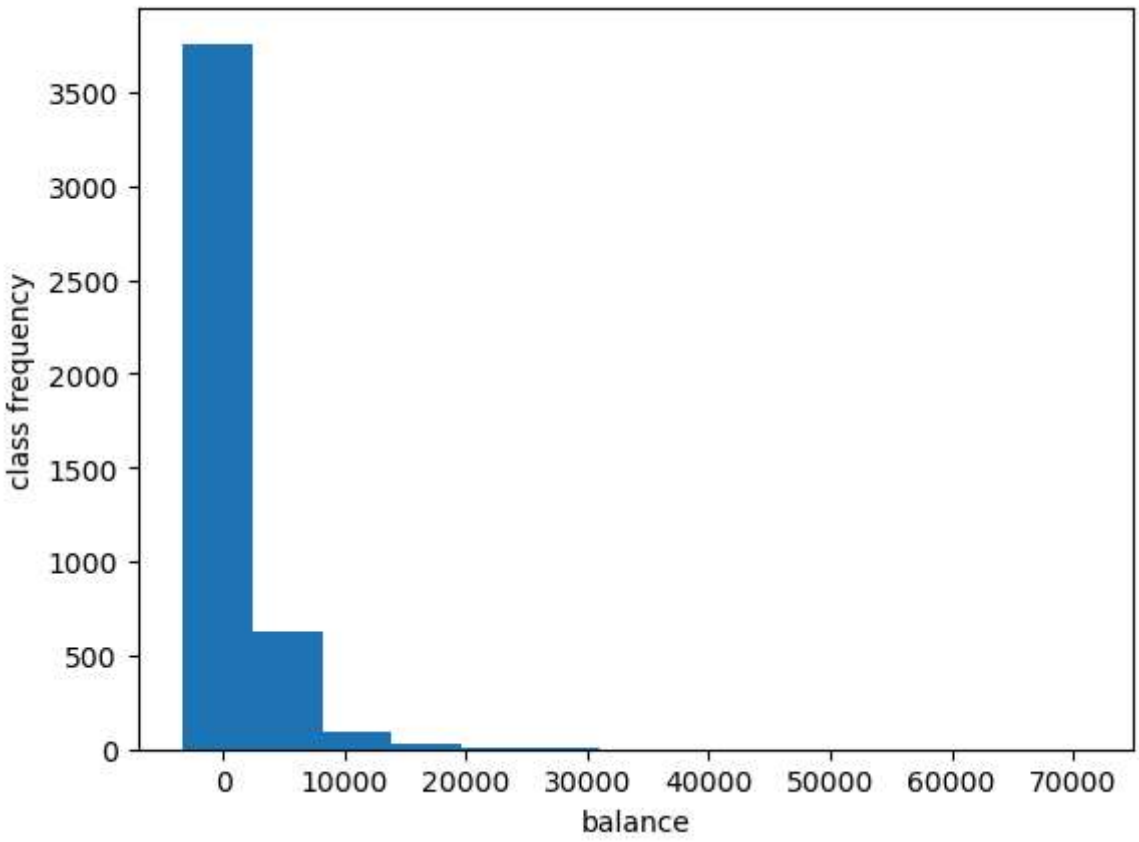
Out[40]: 4521

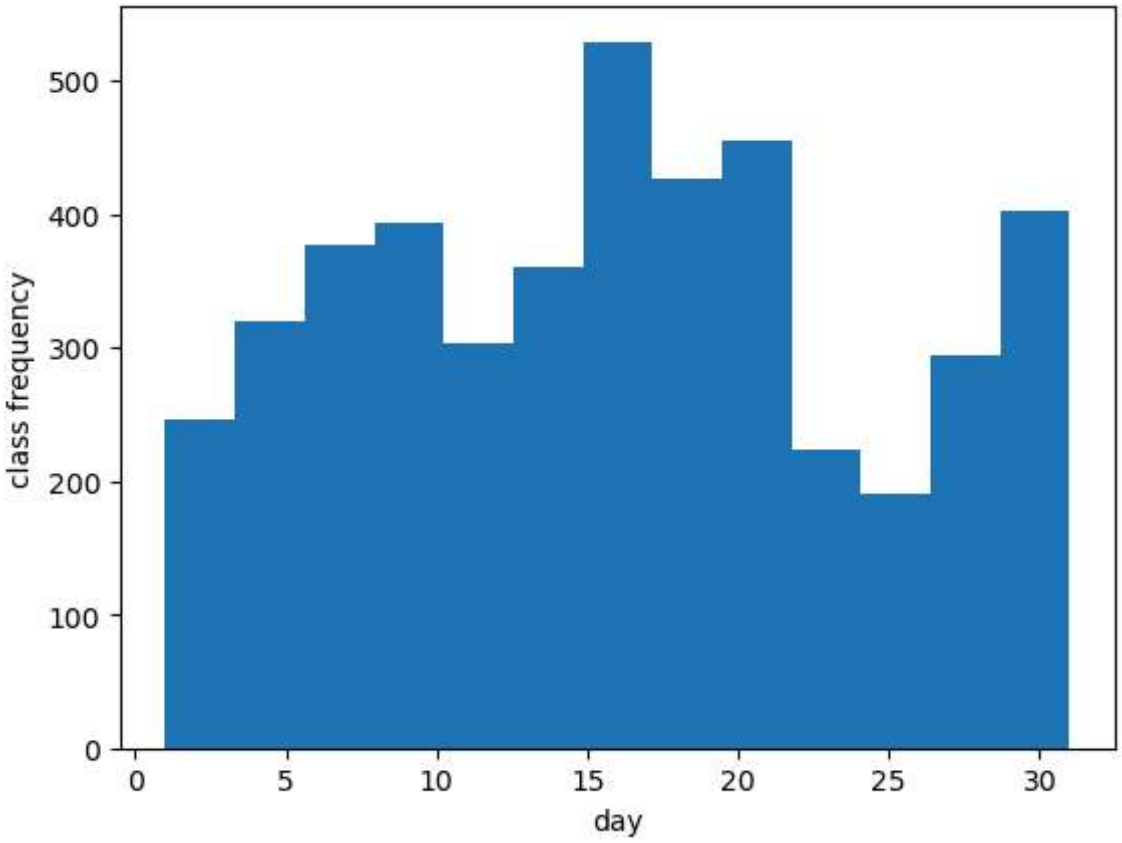
In [42]:
```
#how many bins 2^k
2**12,2**13
```

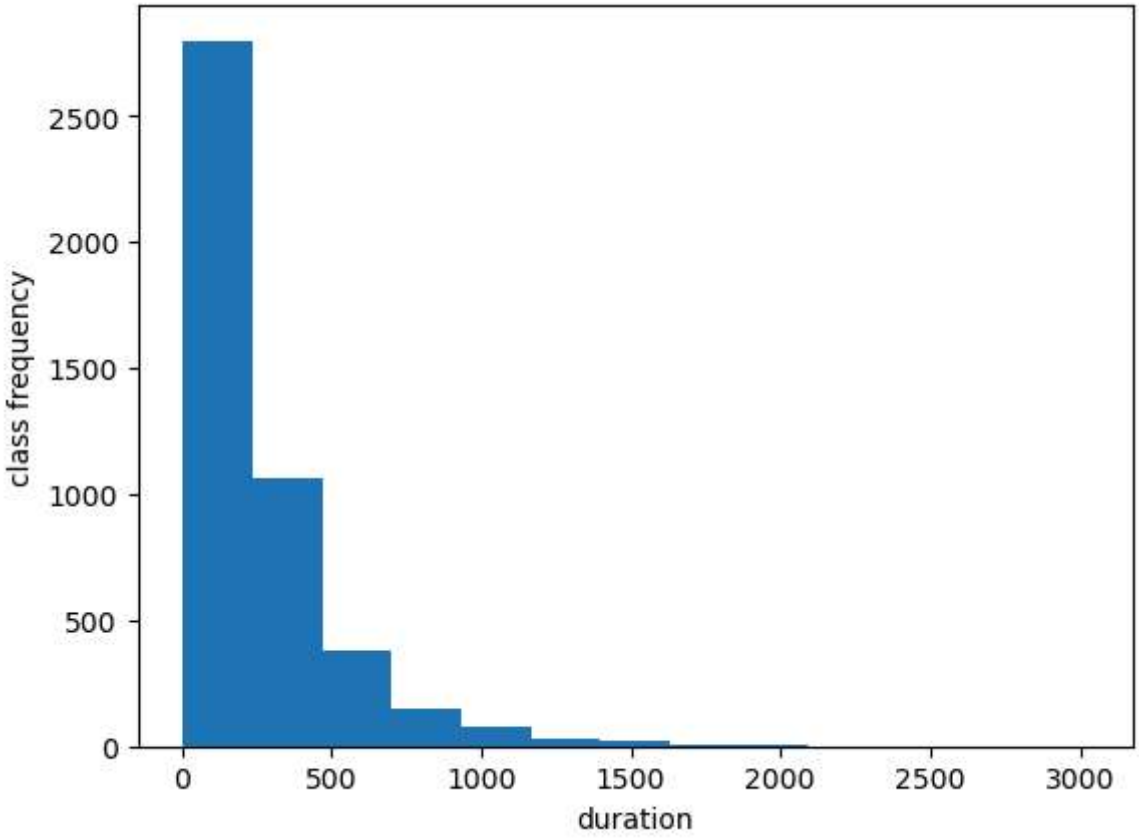
Out[42]: (4096, 8192)

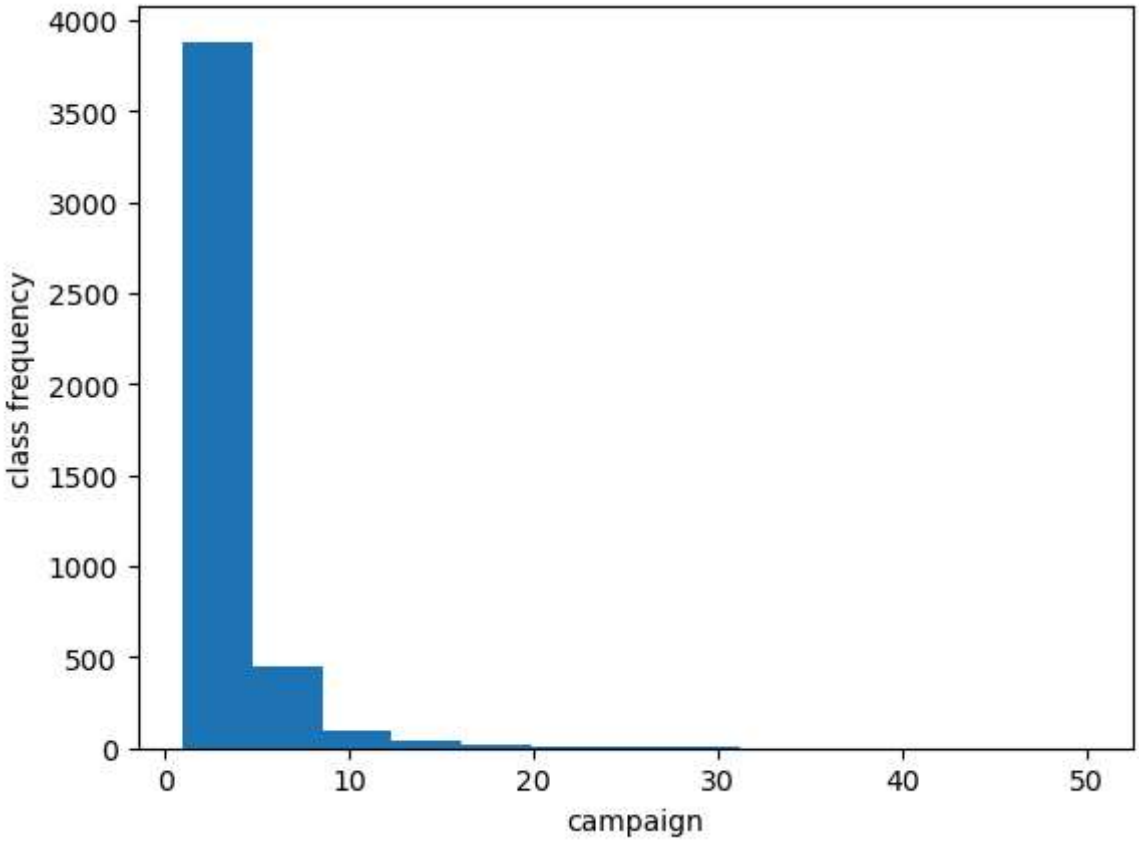
In [44]:
```
l=[1,2,3,4,5,13]

for i in num_colm:
    data=bank_df[i]
    n=len(data)


    for k in l:
        if n%2**k==n:
            plt.hist(data,bins=k)
            plt.xlabel(f"{i}")
            plt.ylabel("class frequency")
        plt.show()
```

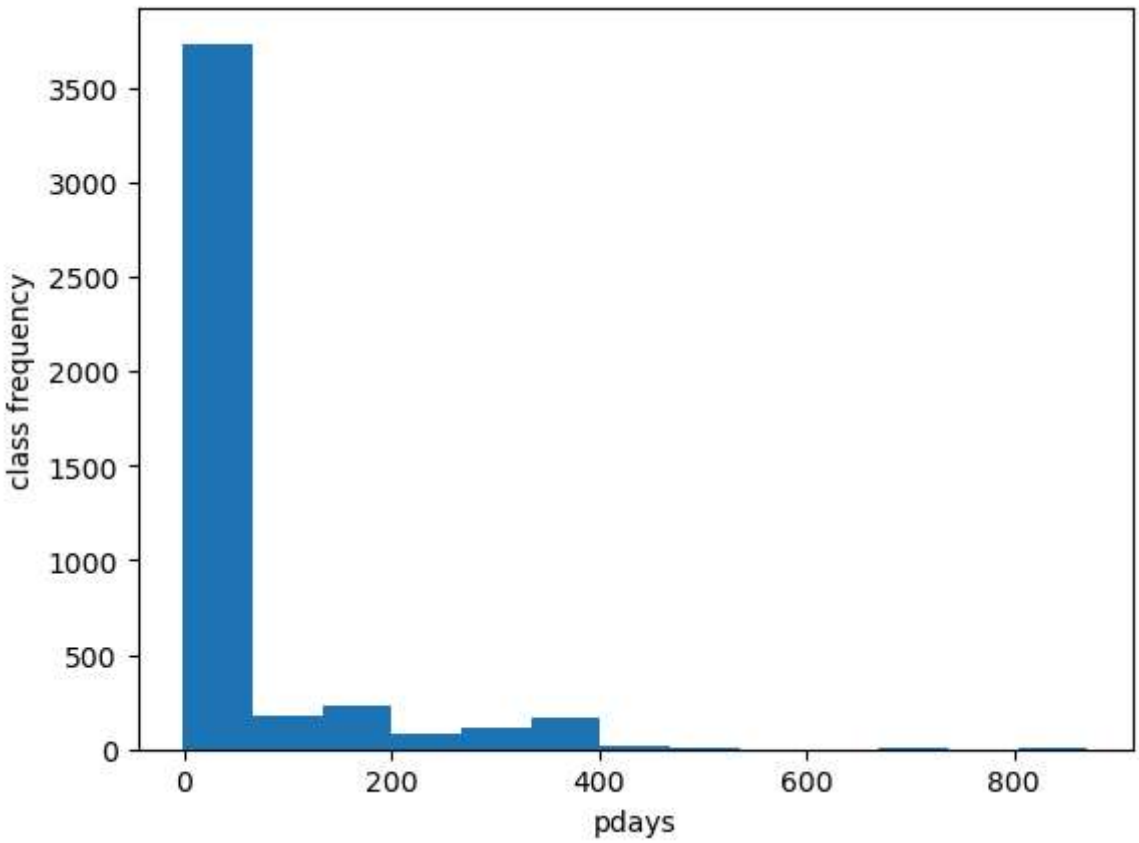

1000
800
600
400
200
0
class frequency
20
30
40
50
60
70
80
90
age

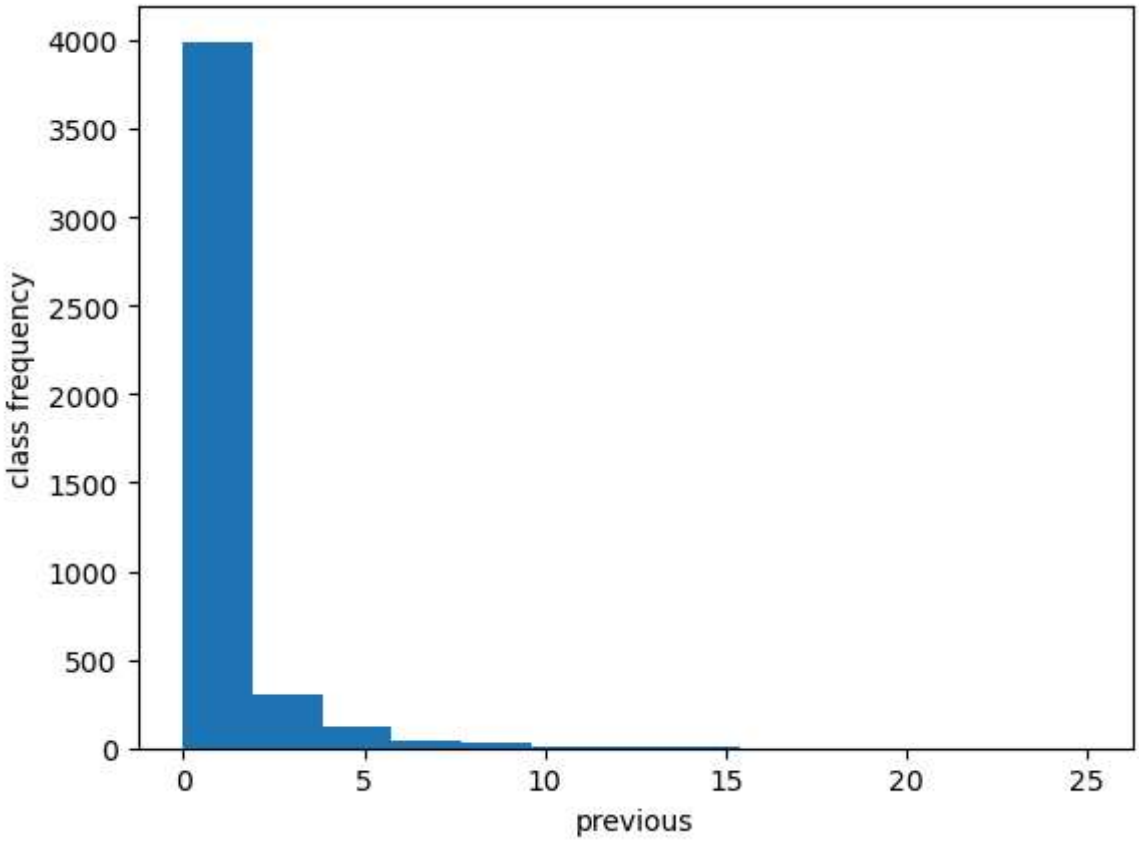

3500
3000
2500
2000
1500
1000
500
0
class frequency
0
10000
20000
30000
40000
50000
60000
70000
balance


class frequency
0
100
200
300
400
500
0
5
10
15
20
25
30
day

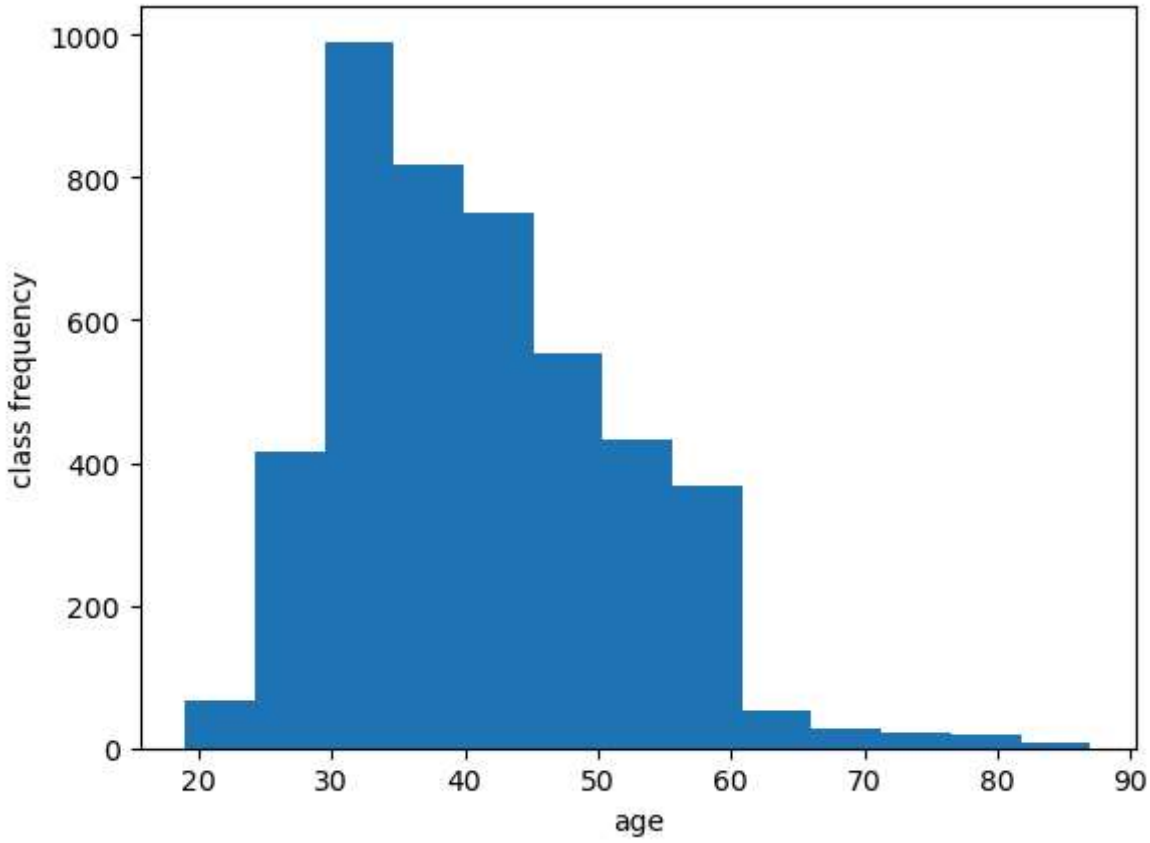

class frequency
0
500
1000
1500
2000
2500
0
500
1000
1500
2000
2500
3000
duration

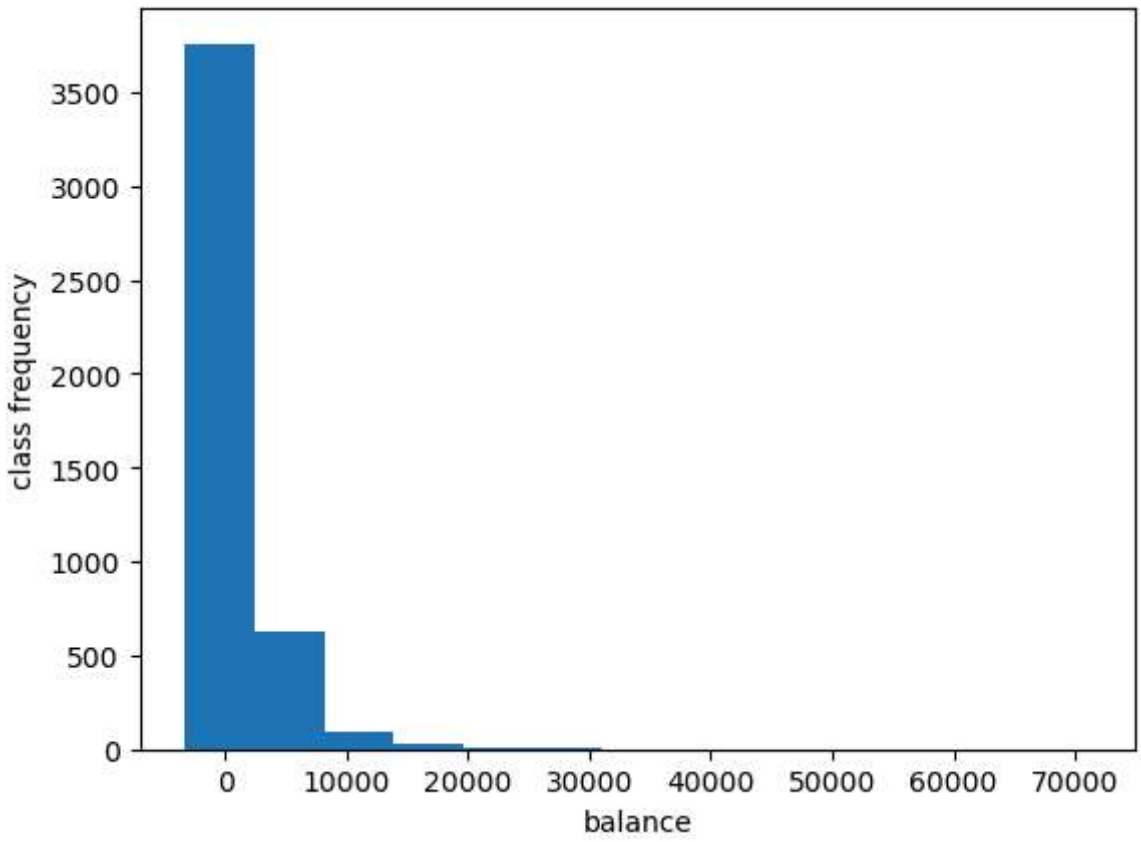

4000
3500
3000
2500
2000
1500
1000
500
0
class frequency
0
10
20
30
40
50
campaign

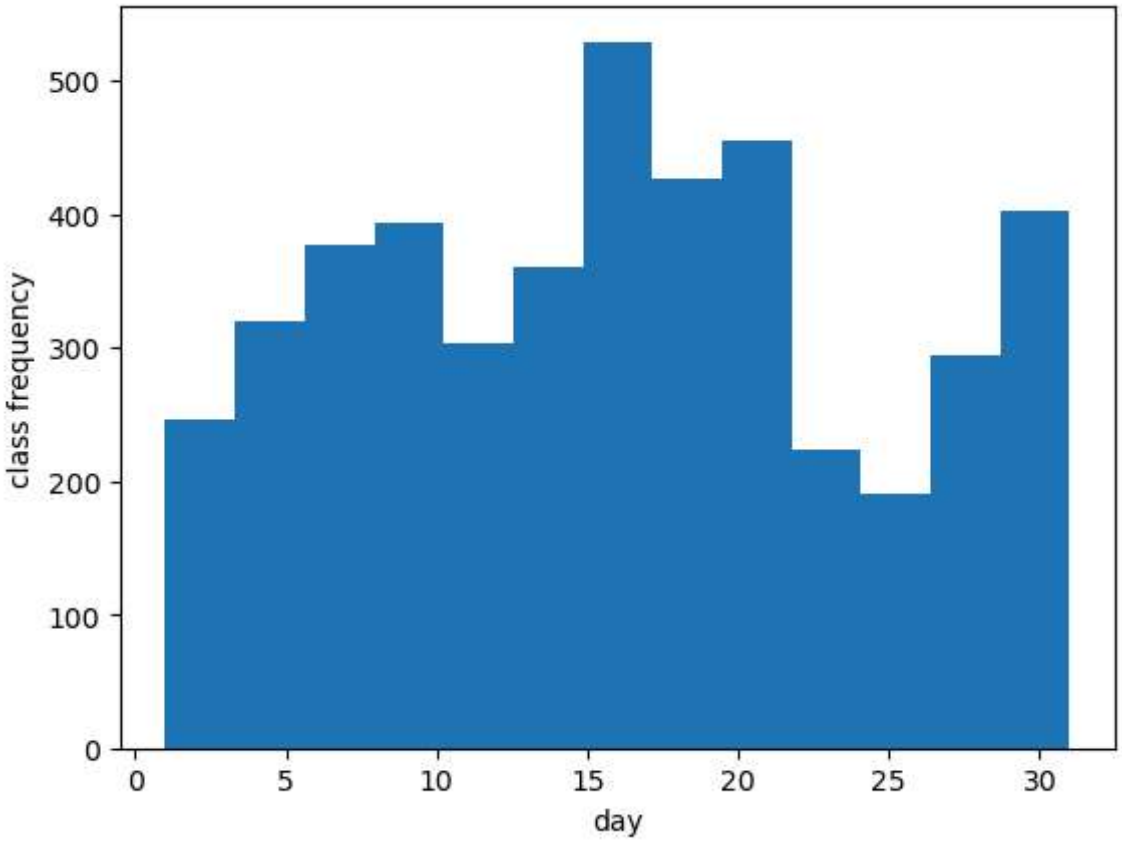

3500
3000
2500
2000
1500
1000
500
0
class frequency
0
200
400
600
800
pdays

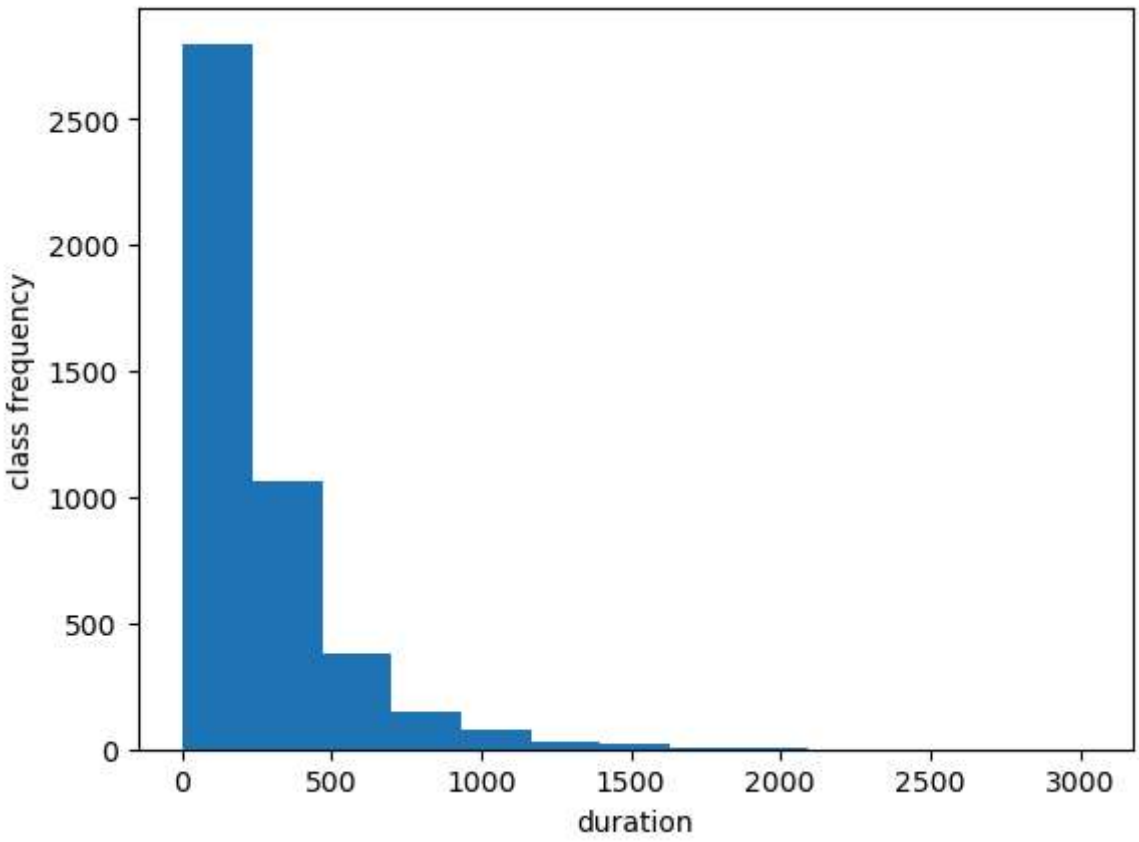


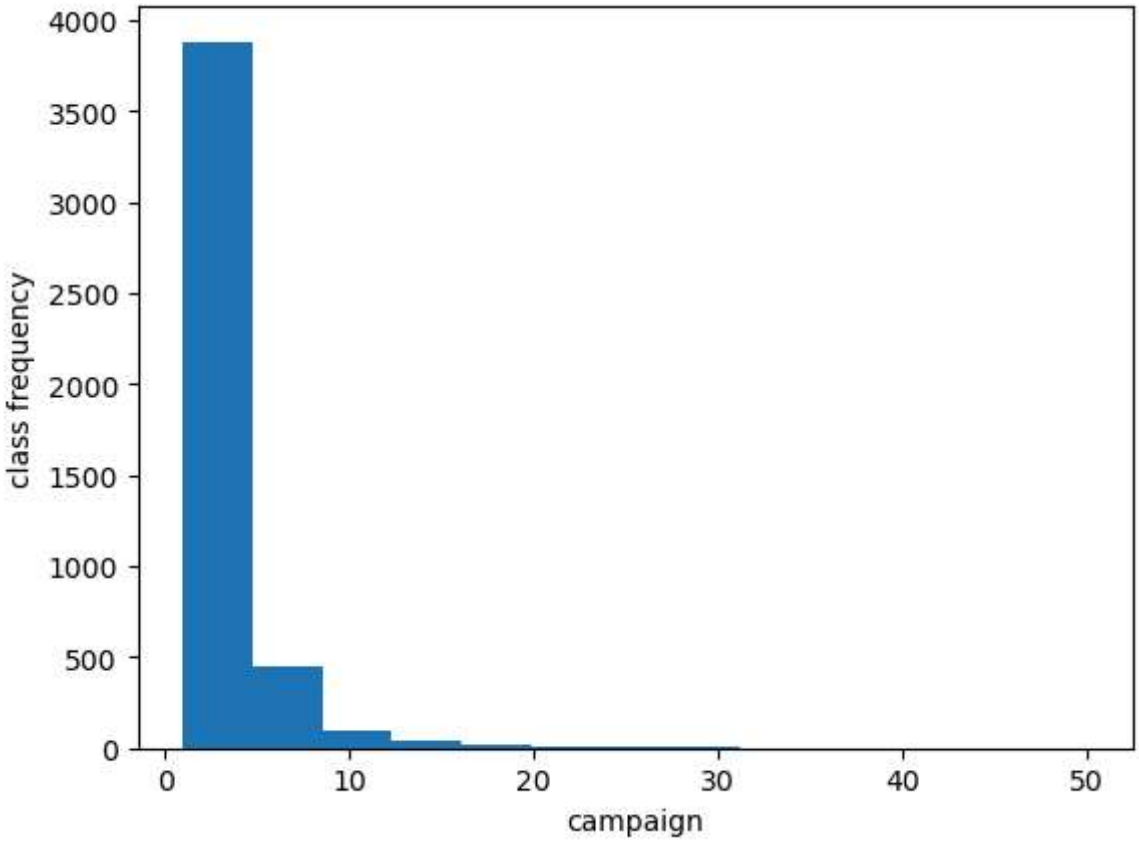
In [54]:
```
for i in num_colm:
    data=bank_df[i]
    n=len(data)


    k=0
    while True:
        k+=1
        if n%2**k==n:
            plt.hist(data,bins=k)
            plt.xlabel(f"{i}")
            plt.ylabel("class frequency")
            break
        plt.show()
```

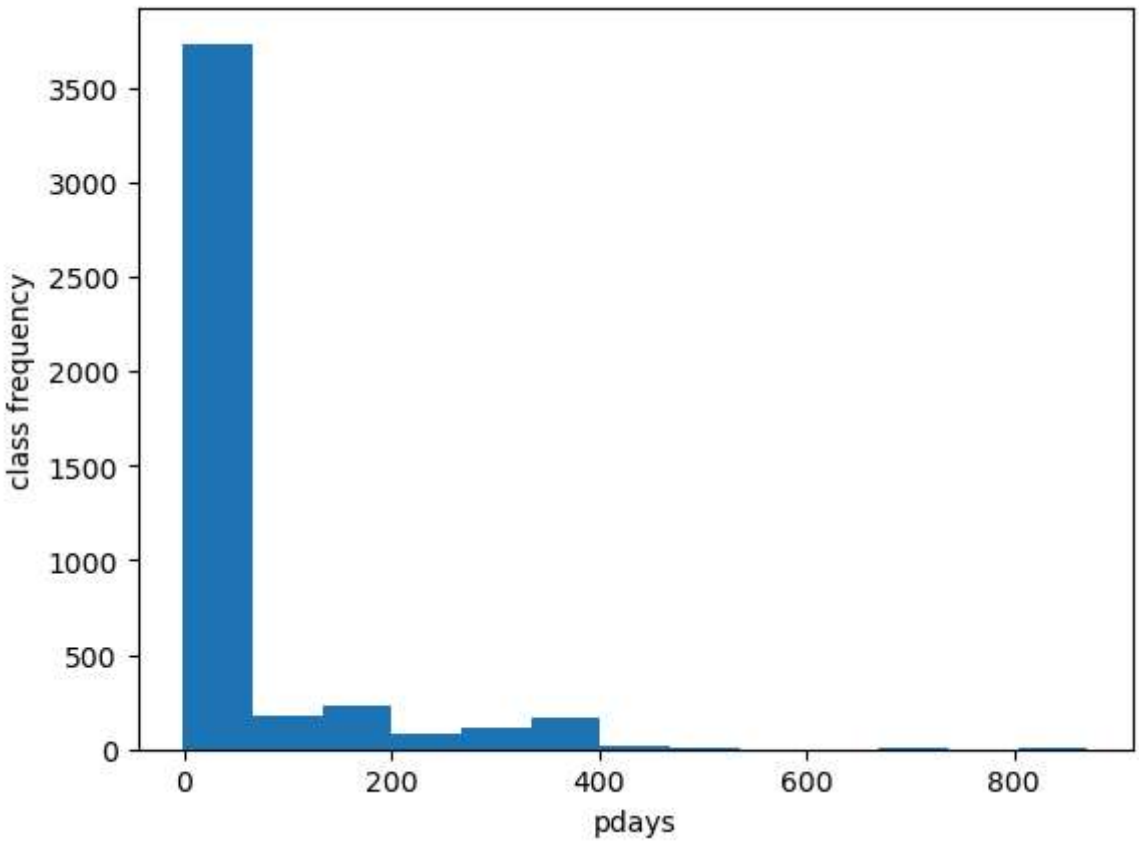

1000
800
600
400
200
0
class frequency
20
30
40
50
60
70
80
90
age

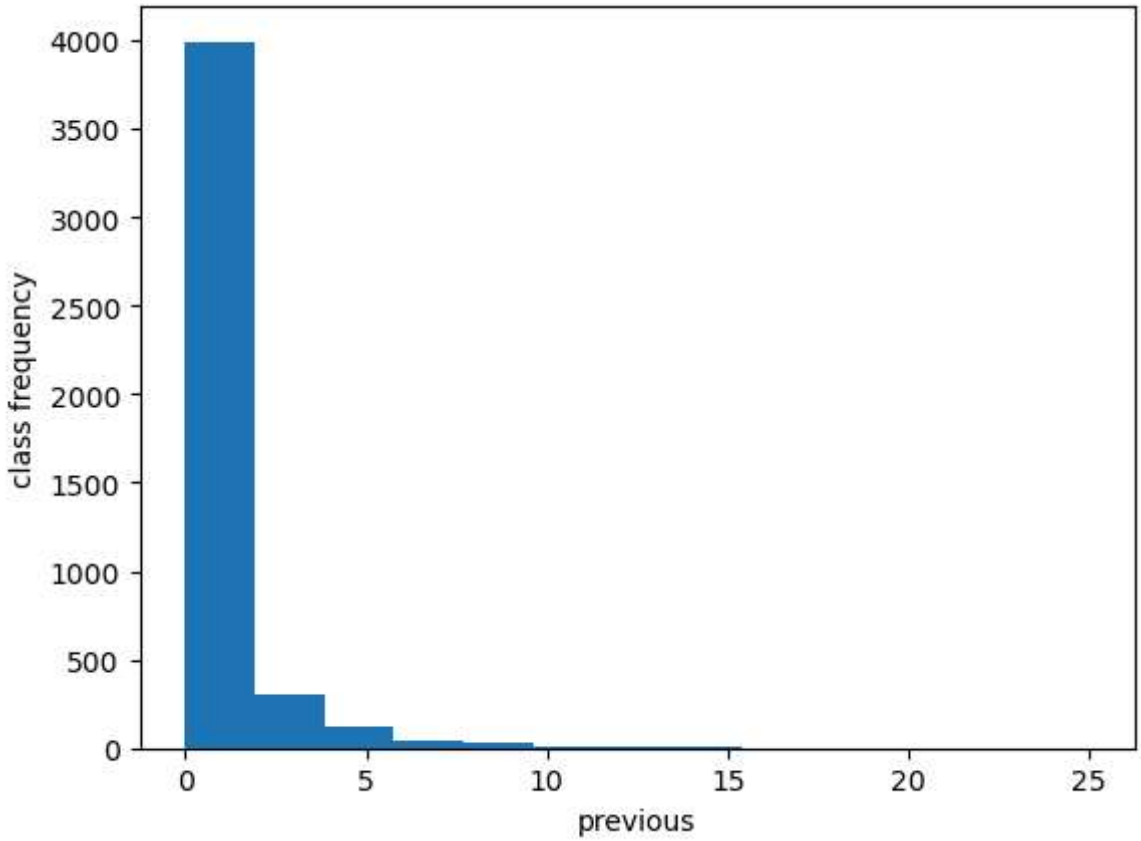

3500
3000
2500
2000
1500
1000
500
0
class frequency
0
10000
20000
30000
40000
50000
60000
70000
balance


500
400
300
200
100
0
class frequency
0
5
10
15
20
25
30
day


2500
2000
1500
1000
500
0
class frequency
0
500
1000
1500
2000
2500
3000
duration


4000
3500
3000
2500
2000
1500
1000
500
0
class frequency
0
10
20
30
40
50
campaign


3500
3000
2500
2000
1500
1000
500
0
class frequency
0
200
400
600
800
pdays


4000
3500
3000
2500
2000
1500
1000
500
0
class frequency
0
5
10
15
20
25
previous

In [ ]: