In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
bank_df=pd.read_csv(r"F:\FSDS\Data Files\bank.csv",sep=';')
bank_df
```

Out[2]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
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
...	...	...	...	...	...	...	...	...	...	...	...
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

4521 rows × 17 columns

In [3]:

```
cat=bank_df.select_dtypes(include='object').columns
num=bank_df.select_dtypes(exclude='object').columns
```

In [4]:

```
cat
```

Out[4]:

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

In [5]:

```
num
```

Out[5]:

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

non outliers data

In [6]:

```
bal_data=bank_df['balance']
q1=round(np.quantile(bal_data,0.25),2)
q3=round(np.quantile(bal_data,0.75),2)
IQR=q3-q1
lb=q1-1.5*IQR
ub=q3+1.5*IQR
```

```
con1= bank_df['balance']>lb
con2=bank_df['balance']<ub
con3=con1&con2
count=len(bank_df[con3])
non_outliers_data=bank_df[con3]
non_outliers_data
```

Out[6]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	oct
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
5	35	management	single	tertiary	no	747	no	no	cellular	23	feb
...	...	...	...	...	...	...	...	...	...	...	...
4515	32	services	single	secondary	no	473	yes	no	cellular	7	jul
4516	33	services	married	secondary	no	-333	yes	no	cellular	30	jul
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

4015 rows × 17 columns

In [8]:

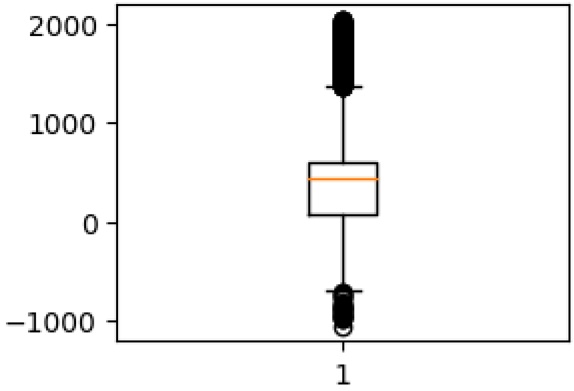
```
bal_data=bank_df['balance']
q1=round(np.quantile(bal_data,0.25),2)
q3=round(np.quantile(bal_data,0.75),2)
IQR=q3-q1
lb=q1-1.5*IQR
ub=q3+1.5*IQR
median=bal_data.median()
new_data=[]
for i in bal_data:
    if i<lb or i>ub:
        new_data.append(median)
    else:
        new_data.append(i)

bank_df['bal1']=new_data
```

In [12]:

```
import warnings
warnings.filterwarnings('ignore')

plt.subplot(2,1,1).hist(bank_df['bal1'])
plt.subplot(2,2,1).boxplot(bank_df['bal1'])
plt.show()
```



In [13]:

```
new_data=[]
for i in bal_data:
    if i<lb or i>ub:
        new_data.append(median)
    else:
        new_data.append(i)
bank_df['bal1']=new_data
```

In [14]:

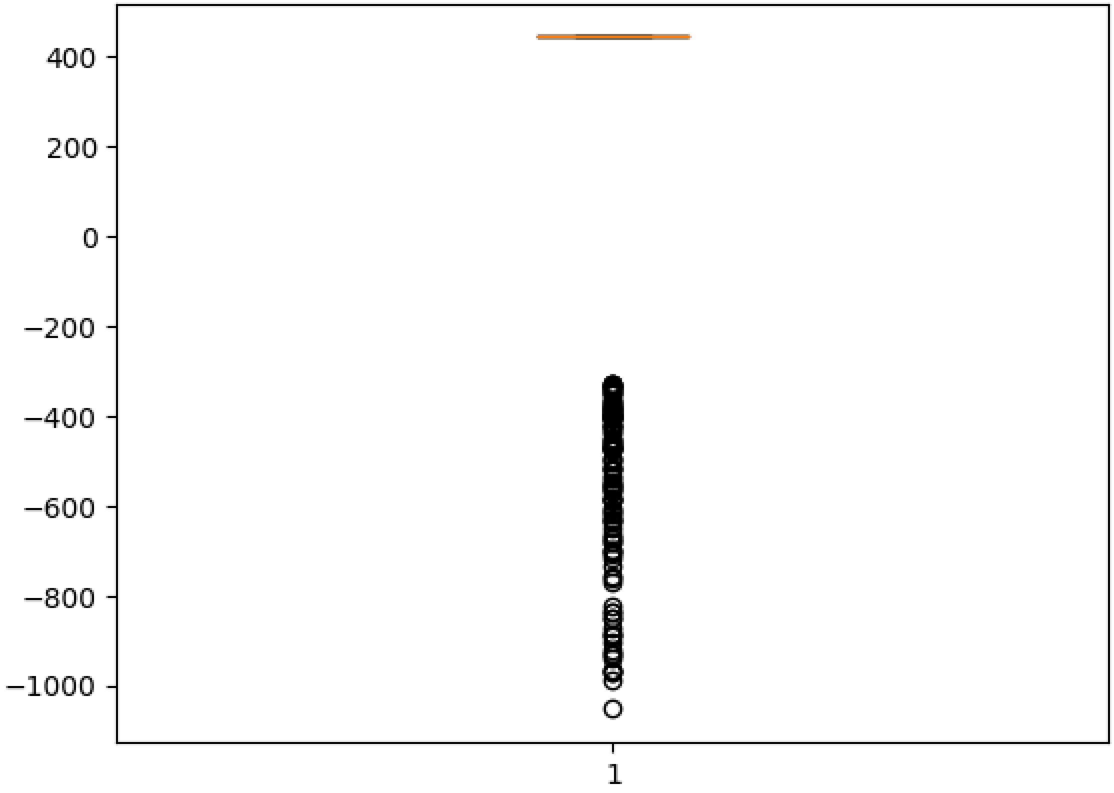
```
bal_data=bank_df['balance']
q1=round(np.quantile(bal_data,0.25),2)
q3=round(np.quantile(bal_data,0.75),2)
IQR=q3-q1
lb=q1-1.5*IQR
ub=q3-1.5*IQR
median=bal_data.median()
con=(bank_df['balance']<lb) |(bank_df['balance']>ub)
true=median
false=bank_df['balance']
bank_df['bal_2']=np.where(con, true, false)
```

In [15]:

```
plt.boxplot(bank_df['bal_2'])
```

Out[15]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1b21e2c0820>,
  <matplotlib.lines.Line2D at 0x1b21e2c3df0>],
 'caps': [<matplotlib.lines.Line2D at 0x1b21e2c1660>,
  <matplotlib.lines.Line2D at 0x1b21e2c18a0>],
 'boxes': [<matplotlib.lines.Line2D at 0x1b21e2c0a90>],
 'medians': [<matplotlib.lines.Line2D at 0x1b21e2c1b10>],
 'fliers': [<matplotlib.lines.Line2D at 0x1b21dd10e80>],
 'means': []}
```



In [ ]: