In [227]:

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

from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import GaussianNB,MultinomialNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report
```

In [182]:

```python
sv('/home/hemang/TE/sem2/DSBDAL Practical/DSBDALExam DataSets/Adult/adult.csv')
```

Out[182]:

	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40
...	...	...	...	...	...	...	...	...	...	...	...	...	...
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38
32556	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40
32557	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20
32559	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40

32560 rows × 15 columns

In [183]:
```
df.columns = ['age','workclass','fnlwgt','education','education-num','marital-s
df
```

Out[183]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	c
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	
...	...	...	...	...	...	...	...	...	...	...	
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	
32556	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	
32557	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	
32559	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	

32560 rows × 15 columns

In [184]:
```
df.isnull().sum()
```

Out[184]:
```
age               0
workclass         0
fnlwgt            0
education         0
education-num     0
marital-status    0
occupation        0
relationship      0
race              0
sex               0
capital-gain      0
capital-loss      0
hours-per-week    0
native-country    0
salary            0
dtype: int64
```

In [185]:
```
df.shape
```
Out[185]:
```
(32560, 15)
```

In [186]:
```
df.dropna(inplace=True)

df.replace('?',pd.NA,inplace=True)
df
```
Out[186]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female
...	...	...	...	...	...	...	...	...	...	...
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female
32556	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male
32557	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male
32559	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female

32560 rows × 15 columns

In [187]:
```
df['age'].fillna(method='ffill',inplace=True)
df
```

Out[187]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	c
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	
...	...	...	...	...	...	...	...	...	...	...	
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	
32556	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	
32557	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	
32559	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	

32560 rows × 15 columns

In [188]:
```
df[df['age'] < 40]
```

Out[188]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female
7	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female
9	37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male
...	...	...	...	...	...	...	...	...	...	...
32550	32	Private	34066	10th	6	Married-civ-spouse	Handlers-cleaners	Husband	Amer-Indian-Eskimo	Male
32552	32	Private	116138	Masters	14	Never-married	Tech-support	Not-in-family	Asian-Pac-Islander	Male
32554	22	Private	310152	Some-college	10	Never-married	Protective-serv	Not-in-family	White	Male
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male

18323 rows × 15 columns

In [189]:
```
df['sex'].value_counts()
```

Out[189]:
```
 Male      21789
 Female    10771
Name: sex, dtype: int64
```

In [190]:
```
df.dtypes
```

Out[190]:
```
age                int64
workclass         object
fnlwgt             int64
education         object
education-num      int64
marital-status    object
occupation        object
relationship      object
race              object
sex               object
capital-gain       int64
capital-loss       int64
hours-per-week     int64
native-country    object
salary            object
dtype: object
```

In [191]:
```python
# from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['sex'] = le.fit_transform(df['sex'])
df['salary'] = le.fit_transform(df['salary'])
df
```

Out[191]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capi g
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	1	
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	1	
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	1	
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	0	
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	0	
...	...	...	...	...	...	...	...	...	...	...	
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	0	
32556	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	1	
32557	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	0	
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	1	
32559	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	0	15(

32560 rows × 15 columns

In [192]:
```python
df['salary'].value_counts()
```

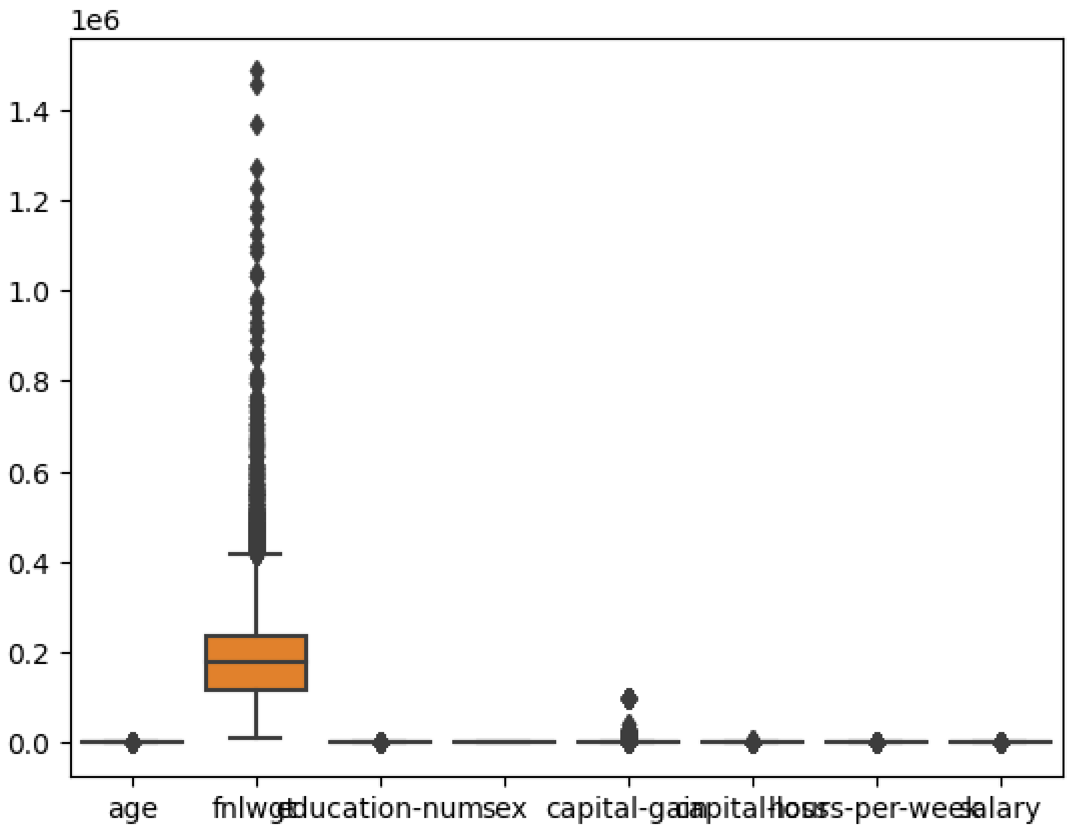
Out[192]:
```
0    24719
1     7841
Name: salary, dtype: int64
```

In [193]:
```
df.dtypes
```

Out[193]:
```
age                int64
workclass         object
fnlwgt             int64
education         object
education-num      int64
marital-status    object
occupation        object
relationship      object
race              object
sex                int64
capital-gain       int64
capital-loss       int64
hours-per-week     int64
native-country    object
salary             int64
dtype: object
```

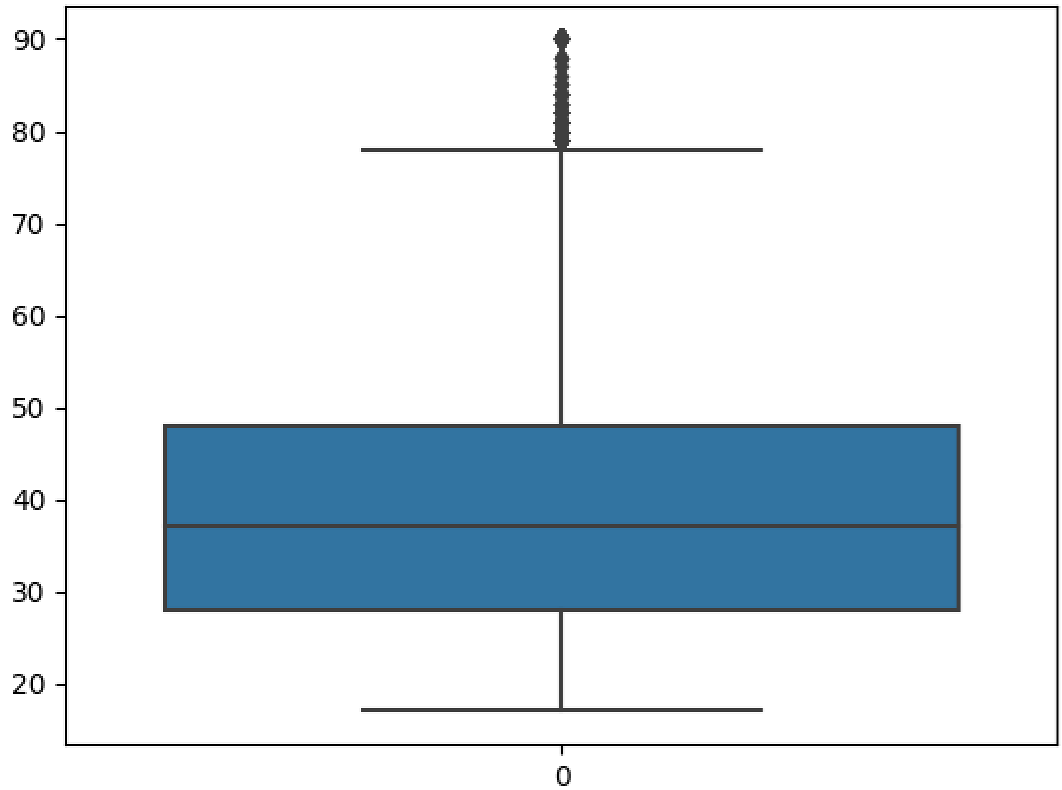
In [194]:
```
sns.boxplot(df)
```

Out[194]:
```
<AxesSubplot: >
```



In [195]:
```
sns.boxplot(df['age'])
```

Out[195]: <AxesSubplot: >

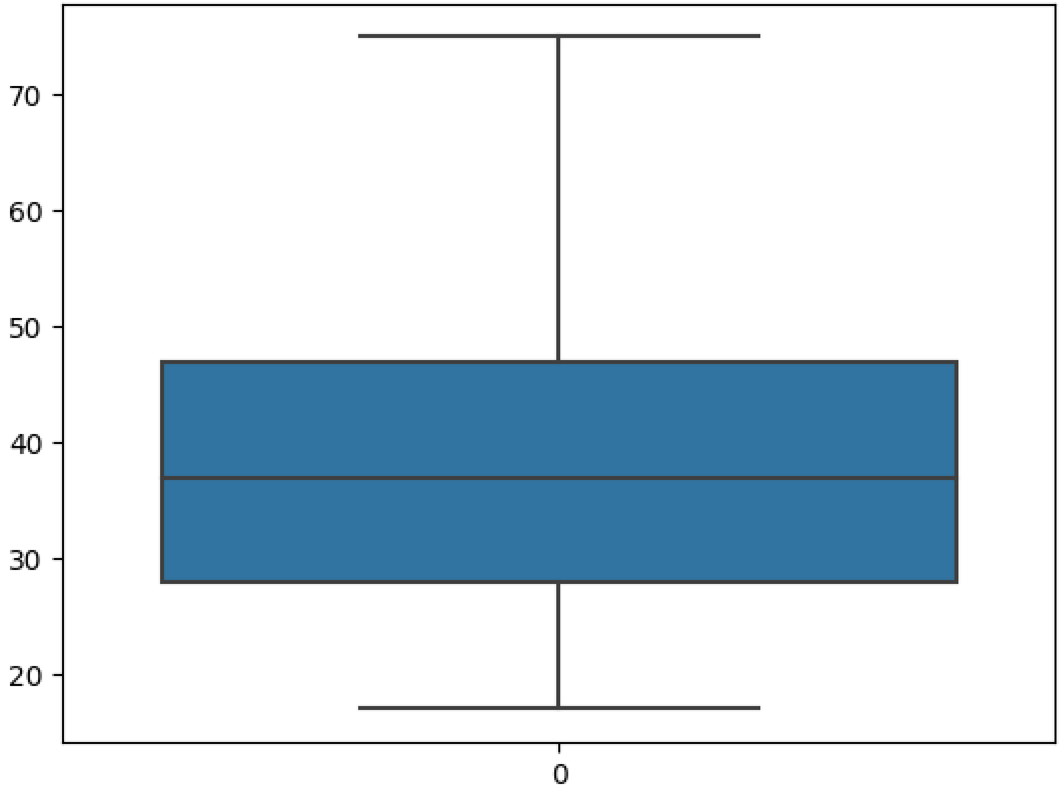


In [198]:
```
# outliers(df['age'])
Q3,Q1 = np.percentile(df['age'],[75,25])
IQR = Q3-Q1
lower = Q1-1.5*IQR
upper = Q3 + 1.5*IQR

def2 = df[(df['age']>=lower ) & (df['age'] <=upper)]
df = def2y
```

In [201]:
```
sns.boxplot(df['age'])
```

Out[201]: <AxesSubplot: >



In [205]:
```
df.reset_index()
df
```

Out[205]:

ıcation	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	salary
chelors	13	Married-civ-spouse	Exec-managerial	Husband	White	1	0	0	13	United-States	0
IS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	1	0	0	40	United-States	0
11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	1	0	0	40	United-States	0
chelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	0	0	0	40	Cuba	0
Aasters	14	Married-civ-spouse	Exec-managerial	Wife	White	0	0	0	40	United-States	0
...	...	...	...	...	...	...	...	...	...	...	...
Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	0	0	0	38	United-States	0
IS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	1	0	0	40	United-States	1
IS-grad	9	Widowed	Adm-clerical	Unmarried	White	0	0	0	40	United-States	0
IS-grad	9	Never-married	Adm-clerical	Own-child	White	1	0	0	20	United-States	0
IS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	0	15024	0	40	United-States	1

In [207]:
```
def2 = df[['age','education-num','capital-gain']]
def2
```

Out[207]:

	age	education-num	capital-gain
0	50	13	0
1	38	9	0
2	53	7	0
3	28	13	0
4	37	14	0
...	...	...	...
32555	27	12	0
32556	40	9	0
32557	58	9	0
32558	22	9	0
32559	52	9	15024

32319 rows × 3 columns

In [211]:
```
x = def2
y = df['salary']
```

In [209]:
```
xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.2)
```

In [221]:
```
reg = LogisticRegression()
reg.fit(x,y)
ypred = reg.predict(xtest)
```

In [223]:
```
mul = MultinomialNB()
mul.fit(x,y)
ypred2 = mul.predict(xtest)
```

In [225]:
```
knn = KNeighborsClassifier()
knn.fit(x,y)
ypred3 = knn.predict(xtest)
```

In [229]:
```
print(classification_report(ypred,ytest))
```

```
              precision    recall  f1-score   support

           0       0.96      0.82      0.88      5726
           1       0.34      0.71      0.46       738

    accuracy                           0.81      6464
   macro avg       0.65      0.77      0.67      6464
weighted avg       0.89      0.81      0.83      6464
```

In [230]:
```
print(classification_report(ypred2,ytest))
```

```
              precision    recall  f1-score   support

           0       0.96      0.79      0.87      5924
           1       0.21      0.62      0.32       540

    accuracy                           0.78      6464
   macro avg       0.59      0.71      0.59      6464
weighted avg       0.90      0.78      0.82      6464
```

In [231]:
```
print(classification_report(ypred3,ytest))
```

```
              precision    recall  f1-score   support

           0       0.91      0.85      0.88      5268
           1       0.48      0.63      0.55      1196

    accuracy                           0.81      6464
   macro avg       0.70      0.74      0.71      6464
weighted avg       0.83      0.81      0.82      6464
```

In [ ]: