```python
import pandas as pd
column_names = ["age", "workclass", "fnlwgt", "education", "education-
num", "marital-status","occupation", "relationship", "race", "sex",
"capital-gain","capital-loss","hours-per-week", "native-country",
"income"]
data = pd.read_csv('adult.csv', names = column_names)
data
```

	age	workclass	fnlwgt	education	education-num \
0	39	State-gov	77516	Bachelors	13
1	50	Self-emp-not-inc	83311	Bachelors	13
2	38	Private	215646	HS-grad	9
3	53	Private	234721	11th	7
4	28	Private	338409	Bachelors	13
...	...	...	...	...	...
32556	27	Private	257302	Assoc-acdm	12
32557	40	Private	154374	HS-grad	9
32558	58	Private	151910	HS-grad	9
32559	22	Private	201490	HS-grad	9
32560	52	Self-emp-inc	287927	HS-grad	9

	marital-status	occupation	relationship	race \
0	Never-married	Adm-clerical	Not-in-family	White
1	Married-civ-spouse	Exec-managerial	Husband	White
2	Divorced	Handlers-cleaners	Not-in-family	White
3	Married-civ-spouse	Handlers-cleaners	Husband	Black
4	Married-civ-spouse	Prof-specialty	Wife	Black
...	...	...	...	...
32556	Married-civ-spouse	Tech-support	Wife	White
32557	Married-civ-spouse	Machine-op-inspct	Husband	White
32558	Widowed	Adm-clerical	Unmarried	White
32559	Never-married	Adm-clerical	Own-child	White
32560	Married-civ-spouse	Exec-managerial	Wife	White

	sex	capital-gain	capital-loss	hours-per-week	native-country \
0	Male	2174	0	40	United-States
1	Male	0	0	13	United-

```
States
2          Male             0             0            40   United-
States
3          Male             0             0            40   United-
States
4        Female             0             0            40
Cuba
...         ...           ...           ...           ...
...
32556    Female             0             0            38   United-
States
32557      Male             0             0            40   United-
States
32558    Female             0             0            40   United-
States
32559      Male             0             0            20   United-
States
32560    Female         15024             0            40   United-
States

       income
0       <=50K
1       <=50K
2       <=50K
3       <=50K
4       <=50K
...       ...
32556   <=50K
32557    >50K
32558   <=50K
32559   <=50K
32560    >50K

[32561 rows x 15 columns]

data

        age            workclass  fnlwgt   education  education-num  \
0        39            State-gov   77516   Bachelors             13
1        50     Self-emp-not-inc   83311   Bachelors             13
2        38              Private  215646     HS-grad              9
3        53              Private  234721        11th              7
4        28              Private  338409   Bachelors             13
...     ...                  ...     ...         ...            ...
32556    27              Private  257302   Assoc-acdm            12
32557    40              Private  154374     HS-grad              9
32558    58              Private  151910     HS-grad              9
32559    22              Private  201490     HS-grad              9
32560    52         Self-emp-inc  287927     HS-grad              9
```

	marital-status	occupation	relationship	race
0	Never-married	Adm-clerical	Not-in-family	White
1	Married-civ-spouse	Exec-managerial	Husband	White
2	Divorced	Handlers-cleaners	Not-in-family	White
3	Married-civ-spouse	Handlers-cleaners	Husband	Black
4	Married-civ-spouse	Prof-specialty	Wife	Black
...	...	...	...	...
32556	Married-civ-spouse	Tech-support	Wife	White
32557	Married-civ-spouse	Machine-op-inspct	Husband	White
32558	Widowed	Adm-clerical	Unmarried	White
32559	Never-married	Adm-clerical	Own-child	White
32560	Married-civ-spouse	Exec-managerial	Wife	White

	sex	capital-gain	capital-loss	hours-per-week	native-country
0	Male	2174	0	40	United-States
1	Male	0	0	13	United-States
2	Male	0	0	40	United-States
3	Male	0	0	40	United-States
4	Female	0	0	40	Cuba
...	...	...	...	...	...
32556	Female	0	0	38	United-States
32557	Male	0	0	40	United-States
32558	Female	0	0	40	United-States
32559	Male	0	0	20	United-States
32560	Female	15024	0	40	United-States

```
       income
0        <=50K
1        <=50K
2        <=50K
3        <=50K
4        <=50K
...        ...
32556    <=50K
32557     >50K
32558    <=50K
32559    <=50K
32560     >50K

[32561 rows x 15 columns]
```

```python
#What is the central tendency of the age distribution?

data.describe()
```

```
                age        fnlwgt  education-num  capital-gain
capital-loss  \
count  32561.000000  3.256100e+04   32561.000000  32561.000000
32561.000000
mean      38.581647  1.897784e+05      10.080679   1077.648844
87.303830
std       13.640433  1.055500e+05       2.572720   7385.292085
402.960219
min       17.000000  1.228500e+04       1.000000      0.000000
0.000000
25%       28.000000  1.178270e+05       9.000000      0.000000
0.000000
50%       37.000000  1.783560e+05      10.000000      0.000000
0.000000
75%       48.000000  2.370510e+05      12.000000      0.000000
0.000000
max       90.000000  1.484705e+06      16.000000  99999.000000
4356.000000

       hours-per-week
count    32561.000000
mean        40.437456
std         12.347429
min          1.000000
25%         40.000000
50%         40.000000
75%         45.000000
max         99.000000
```
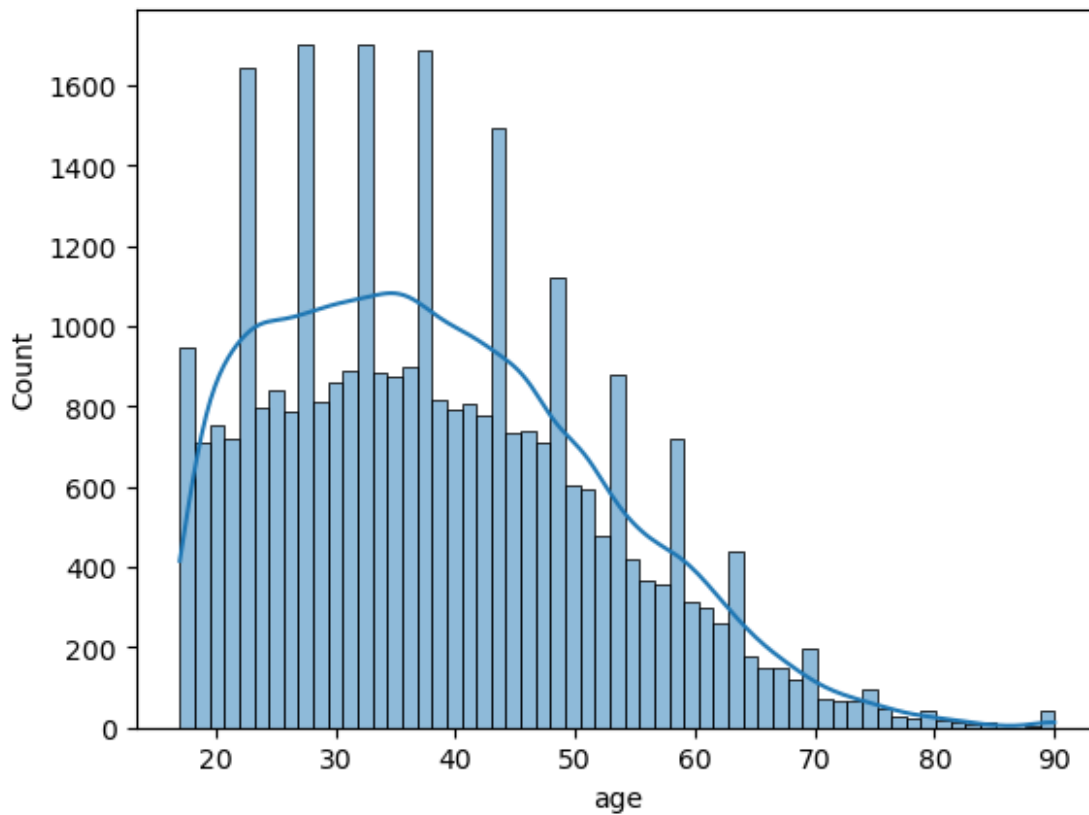
```python
import seaborn as sns
sns.histplot(data,x='age',kde=True)
```

```
<Axes: xlabel='age', ylabel='Count'>
```



```
#Is the age distribution skewed or symmetric?
#Ans: Thus Age Distribution is Right Skewed
#Above

age_mean = data['age'].mean()
age_mean

38.58164675532078

data.dtypes

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
income              object
dtype: object
```

```python
#Outliers
q1 = data['age'].quantile(0.25)
q3 = data['age'].quantile(0.75)
IQR = q3 - q1
upper_limit = q3 + 1.5*IQR
lower_limit = q1 - 1.5*IQR
out_data = data.loc[(data['age'] > lower_limit) & (data['age'] <
upper_limit)]
out_data
```

	age	workclass	fnlwgt	education	education-num \
0	39	State-gov	77516	Bachelors	13
1	50	Self-emp-not-inc	83311	Bachelors	13
2	38	Private	215646	HS-grad	9
3	53	Private	234721	11th	7
4	28	Private	338409	Bachelors	13
...	...	...	...	...	...
32556	27	Private	257302	Assoc-acdm	12
32557	40	Private	154374	HS-grad	9
32558	58	Private	151910	HS-grad	9
32559	22	Private	201490	HS-grad	9
32560	52	Self-emp-inc	287927	HS-grad	9

	marital-status	occupation	relationship	race \
0	Never-married	Adm-clerical	Not-in-family	White
1	Married-civ-spouse	Exec-managerial	Husband	White
2	Divorced	Handlers-cleaners	Not-in-family	White
3	Married-civ-spouse	Handlers-cleaners	Husband	Black
4	Married-civ-spouse	Prof-specialty	Wife	Black
...	...	...	...	...
32556	Married-civ-spouse	Tech-support	Wife	White
32557	Married-civ-spouse	Machine-op-inspct	Husband	White
32558	Widowed	Adm-clerical	Unmarried	White
32559	Never-married	Adm-clerical	Own-child	White

```
32560    Married-civ-spouse    Exec-managerial         Wife    White

          sex   capital-gain   capital-loss   hours-per-week   native-
country   \
0        Male          2174              0               40    United-
States
1        Male             0              0               13    United-
States
2        Male             0              0               40    United-
States
3        Male             0              0               40    United-
States
4      Female             0              0               40
Cuba
...       ...           ...            ...              ...
...
32556  Female             0              0               38    United-
States
32557    Male             0              0               40    United-
States
32558  Female             0              0               40    United-
States
32559    Male             0              0               20    United-
States
32560  Female         15024              0               40    United-
States

       income
0       <=50K
1       <=50K
2       <=50K
3       <=50K
4       <=50K
...       ...
32556   <=50K
32557    >50K
32558   <=50K
32559   <=50K
32560    >50K

[32395 rows x 15 columns]

sns.boxplot(data=data,x='age')

<Axes: xlabel='age'>
```
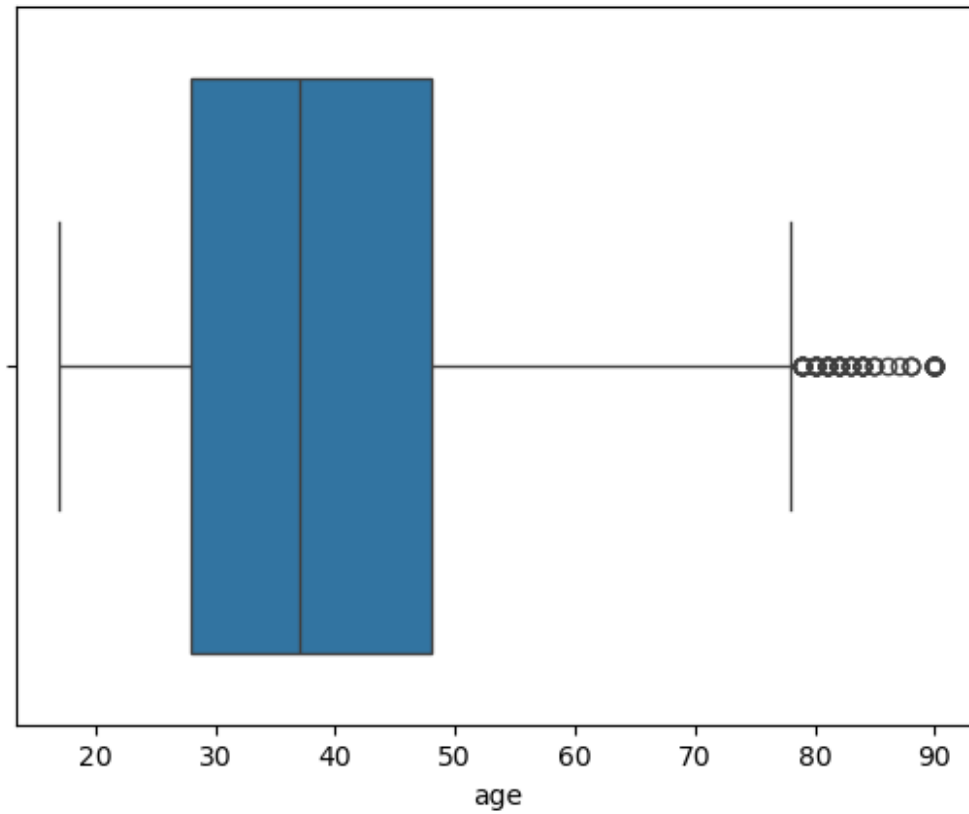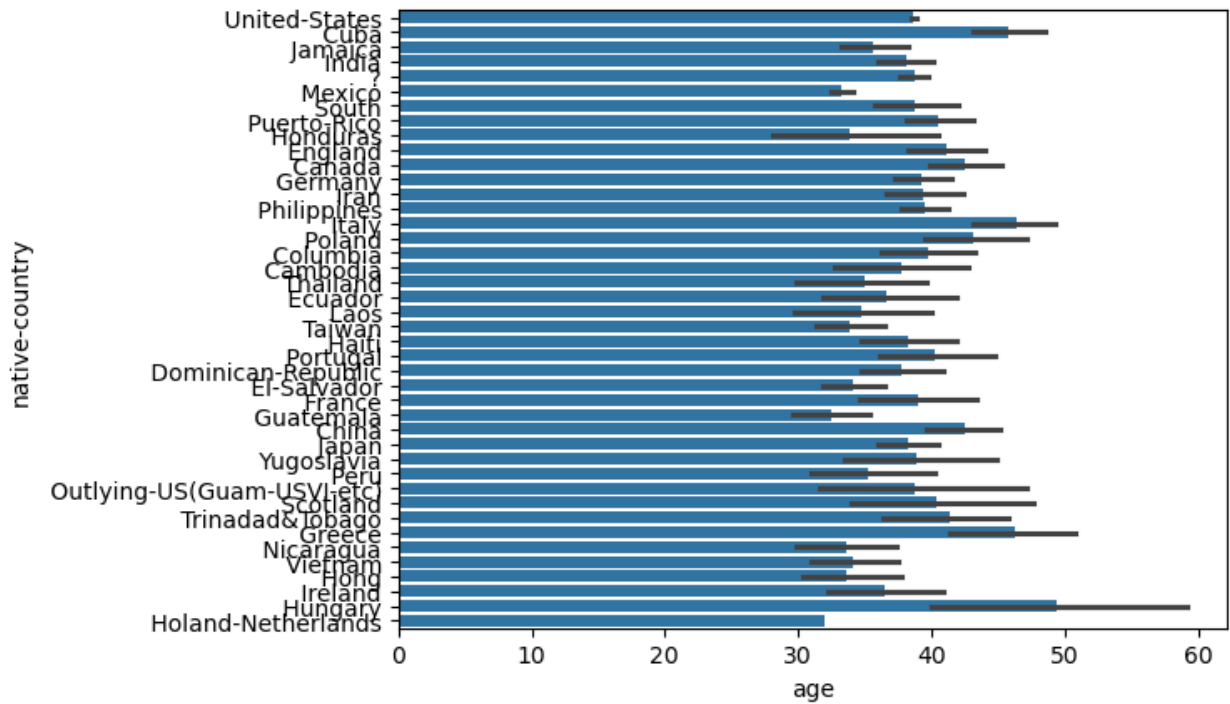
```
sns.barplot(data,x='age',y='native-country')
```
```
<Axes: xlabel='age', ylabel='native-country'>
```

```
sns.barplot(data,x='age',y='workclass')
<Axes: xlabel='age', ylabel='workclass'>
```