In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import warnings
warnings.filterwarnings(action='ignore')
```

In [3]:
```
df=pd.read_csv(r"C:\Users\Shree\Downloads\data (1).csv")
df
```

Out[3]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relation
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husl
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husl
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husl
4	18	?	103497	Some-college	10	Never-married	?	Own-
...	...	...	...	...	...	...	...	
48837	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	
48838	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husl
48839	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unma
48840	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-
48841	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	

48842 rows × 15 columns

In [5]:
```
df.head()
```

Out[5]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband
4	18	?	103497	Some-college	10	Never-married	?	Own-child

In [7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48842 entries, 0 to 48841
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   age              48842 non-null  int64 
 1   workclass        48842 non-null  object
 2   fnlwgt           48842 non-null  int64 
 3   education        48842 non-null  object
 4   educational-num  48842 non-null  int64 
 5   marital-status   48842 non-null  object
 6   occupation       48842 non-null  object
 7   relationship     48842 non-null  object
 8   race             48842 non-null  object
 9   gender           48842 non-null  object
 10  capital-gain     48842 non-null  int64 
 11  capital-loss     48842 non-null  int64 
 12  hours-per-week   48842 non-null  int64 
 13  native-country   48842 non-null  object
 14  income           48842 non-null  object
dtypes: int64(6), object(9)
memory usage: 5.6+ MB
```

In [9]:
```
df.describe()
```

Out[9]:

	age	fnlwgt	educational-num	capital-gain	capital-loss	hours-w
count	48842.000000	4.884200e+04	48842.000000	48842.000000	48842.000000	48842.000
mean	38.643585	1.896641e+05	10.078089	1079.067626	87.502314	40.422
std	13.710510	1.056040e+05	2.570973	7452.019058	403.004552	12.391
min	17.000000	1.228500e+04	1.000000	0.000000	0.000000	1.000
25%	28.000000	1.175505e+05	9.000000	0.000000	0.000000	40.000
50%	37.000000	1.781445e+05	10.000000	0.000000	0.000000	40.000
75%	48.000000	2.376420e+05	12.000000	0.000000	0.000000	45.000
max	90.000000	1.490400e+06	16.000000	99999.000000	4356.000000	99.000

In [13]:
```
df.age.min()
```

Out[13]: 17

In [15]:
```
df.age.max()
```

Out[15]: 90

In [17]:
```
df.age.mean()
```

Out[17]: 38.64358543876172

In [19]:
```
df.age.median()
```

Out[19]: 37.0

In [21]:
```
df.age.count()
```

Out[21]: 48842

In [23]:
```
df.age.quantile(0.5)
```

Out[23]: 37.0

In [27]:
```
df.age.quantile(0.25)
```

Out[27]: 28.0

In [29]:
```
df.age.quantile(0.75)
```

Out[29]: 48.0

In [33]:
```
df.income.nunique()
```

Out[33]: 2

In [37]:
```
df.groupby(['income','age'])
```

Out[37]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000025B0C551100>

In [39]:
```
df.groupby(['age','income']).min()
```

Out[39]:

		workclass	fnlwgt	education	educational-num	marital-status	occupation	relatio
age	**income**							
17	**<=50K**	?	19752	10th	3	Married-civ-spouse	?	Hus
18	**<=50K**	?	20057	10th	3	Divorced	?	Hus
19	**<=50K**	?	20469	10th	1	Divorced	?	Hus
	>50K	?	142738	12th	4	Married-civ-spouse	?	N f
20	**<=50K**	?	19410	10th	1	Divorced	?	Hus
...	**...**	...	...	...	...	...	...	
88	**<=50K**	Private	30102	7th-8th	4	Divorced	Adm-clerical	Hus
	>50K	Self-emp-not-inc	263569	11th	7	Married-civ-spouse	Farming-fishing	Hus
89	**<=50K**	?	29106	10th	6	Married-civ-spouse	?	Hus
90	**<=50K**	?	39824	10th	2	Divorced	?	Hus
	>50K	?	46786	Assoc-acdm	9	Married-civ-spouse	?	Hus

142 rows × 13 columns

In [43]:
```
df.groupby(['age','income']).max()
```

Out[43]:

age	income	workclass	fnlwgt	education	educational-num	marital-status	occupation	relat
17	**<=50K**	State-gov	806316	Some-college	10	Widowed	Transport-moving	Un
18	**<=50K**	State-gov	761006	Some-college	14	Widowed	Transport-moving	
19	**<=50K**	Without-pay	1047822	Some-college	13	Separated	Transport-moving	
	>50K	Private	323605	Some-college	10	Never-married	Prof-specialty	Ov
20	**<=50K**	State-gov	745817	Some-college	14	Separated	Transport-moving	
...	**...**	...	...	...	...	...	...	
88	**<=50K**	Self-emp-not-inc	206291	Some-college	15	Widowed	Prof-specialty	Un
	>50K	Self-emp-not-inc	263569	11th	7	Married-civ-spouse	Farming-fishing	H
89	**<=50K**	Private	152839	Bachelors	13	Married-civ-spouse	Sales	H
90	**<=50K**	Self-emp-not-inc	347074	Some-college	15	Widowed	Transport-moving	Un
	>50K	Self-emp-not-inc	313986	Prof-school	15	Never-married	Sales	

142 rows × 13 columns

In [47]:
```
df.groupby(['age','income']).count()
```

Out[47]:

		workclass	fnlwgt	education	educational-num	marital-status	occupation	relation
age	**income**							
17	**<=50K**	595	595	595	595	595	595	
18	**<=50K**	862	862	862	862	862	862	
19	**<=50K**	1050	1050	1050	1050	1050	1050	
	>50K	3	3	3	3	3	3	
20	**<=50K**	1112	1112	1112	1112	1112	1112	
...	**...**	...	...	...	...	...	...	
88	**<=50K**	5	5	5	5	5	5	
	>50K	1	1	1	1	1	1	
89	**<=50K**	2	2	2	2	2	2	
90	**<=50K**	42	42	42	42	42	42	
	>50K	13	13	13	13	13	13	

142 rows × 13 columns

In [53]:
```
df.groupby('income')['age'].mean()
```

Out[53]:
```
income
<=50K    36.872184
>50K     44.275178
Name: age, dtype: float64
```

In [55]:
```
df.groupby('income')['age'].median()
```

Out[55]:
```
income
<=50K    34.0
>50K     43.0
Name: age, dtype: float64
```

In [57]:
```
df.groupby('income')['age'].quantile(0.5)
```

Out[57]:
```
income
<=50K    34.0
>50K     43.0
Name: age, dtype: float64
```

In [59]:
```
df.groupby('income')['age'].quantile(0.25)
```

Out[59]:
```
income
<=50K    25.0
>50K     36.0
Name: age, dtype: float64
```

In [63]:
```
df.groupby('income')['age'].quantile(0.75)
```

Out[63]:
```
income
<=50K    46.0
>50K     51.0
Name: age, dtype: float64
```

In [69]:
```
df.groupby('income')['age'].std()
```

Out[69]:
```
income
<=50K    14.104118
>50K     10.558983
Name: age, dtype: float64
```

In [71]:
```
df.groupby('income')['age'].count()
```

Out[71]:
```
income
<=50K    37155
>50K     11687
Name: age, dtype: int64
```

In [ ]: