In [9]:
```
import numpy as np
import pandas as  pd
```

In [10]:
```
adult=pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/adult
```

In [11]:
```
adult.columns=['age','workclass','fnlwgt','education','education_num','marital_sta
```

In [13]:
```
adult.head()
```

Out[13]:

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	ra
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	Wh
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	Wh
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Bla
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Bla
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	Wh

1. Select 10 records from the adult sqladb

In [4]:
```
from pandasql import sqldf
```

In [32]:
```
pysqldf = lambda q: sqldf(q, globals())
```

In [33]:
```
q='''SELECT * FROM adult LIMIT 10;'''
```

In [34]:
```
a = pysqldf(q)
print(a)
```

```
   age          workclass  fnlwgt      education  education_num  \
0   50   Self-emp-not-inc   83311      Bachelors             13
1   38            Private  215646        HS-grad              9
2   53            Private  234721           11th              7
3   28            Private  338409      Bachelors             13
4   37            Private  284582        Masters             14
5   49            Private  160187            9th              5
6   52   Self-emp-not-inc  209642        HS-grad              9
7   31            Private   45781        Masters             14
8   42            Private  159449      Bachelors             13
9   37            Private  280464   Some-college             10

           marital_status          occupation    relationship    race  \
0      Married-civ-spouse     Exec-managerial         Husband   White
1                Divorced   Handlers-cleaners   Not-in-family   White
2      Married-civ-spouse   Handlers-cleaners         Husband   Black
3      Married-civ-spouse      Prof-specialty            Wife   Black
4      Married-civ-spouse     Exec-managerial            Wife   White
5   Married-spouse-absent       Other-service   Not-in-family   Black
6      Married-civ-spouse     Exec-managerial         Husband   White
7           Never-married      Prof-specialty   Not-in-family   White
8      Married-civ-spouse     Exec-managerial         Husband   White
9      Married-civ-spouse     Exec-managerial         Husband   Black

        sex  capital_gain  capital_loss  hours_per_week  native_country   label
0      Male             0             0              13   United-States   <=50K
1      Male             0             0              40   United-States   <=50K
2      Male             0             0              40   United-States   <=50K
3    Female             0             0              40            Cuba   <=50K
4    Female             0             0              40   United-States   <=50K
5    Female             0             0              16         Jamaica   <=50K
6      Male             0             0              45   United-States    >50K
7    Female         14084             0              50   United-States    >50K
8      Male          5178             0              40   United-States    >50K
9      Male             0             0              80   United-States    >50K
```

2. Show me the average hours per week of all men who are working in private sector

In [35]:
```
pysqldf = lambda r: sqldf(r, globals())
r='''SELECT AVG(hours_per_week) FROM adult WHERE workclass =' Private' and sex=' M
a=pysqldf(r)
a
```

Out[35]:

	AVG(hours_per_week)
0	42.221226

3. Show me the frequency table for education, occupation and relationship, separately

In [42]:
```
pysqldf = lambda s: sqldf(s, globals())
s='''select distinct education,count(education) from adult group by education  ;''
a=pysqldf(s)
a
```

Out[42]:

	education	count(education)
0	10th	933
1	11th	1175
2	12th	433
3	1st-4th	168
4	5th-6th	333
5	7th-8th	646
6	9th	514
7	Assoc-acdm	1067
8	Assoc-voc	1382
9	Bachelors	5354
10	Doctorate	413
11	HS-grad	10501
12	Masters	1723
13	Preschool	51
14	Prof-school	576
15	Some-college	7291

same can be done for occupation and relationship separately

4. Are there any people who are married, working in private sector and having a masters degree

In [65]:
```
pysqldf = lambda w: sqldf(w, globals())
w='''SELECT count(*) FROM adult WHERE  marital_status=' Married-civ-spouse' and  w
a=pysqldf(w)
a
```

Out[65]:

	count(*)
0	531

5. What is the average, minimum and maximum age group for people working in different sectors

In [49]:
```
pysqldf = lambda t: sqldf(t, globals())
t='''SELECT workclass,AVG(age) Average ,MIN(age) Minimum ,MAX(age) Maximum

FROM adult GROUP BY workclass;'''
a=pysqldf(t)
a
```

Out[49]:

	workclass	Average	Minimum	Maximum
0	?	40.960240	17	90
1	Federal-gov	42.590625	17	90
2	Local-gov	41.751075	17	90
3	Never-worked	20.571429	17	30
4	Private	36.797585	17	90
5	Self-emp-inc	46.017025	17	84
6	Self-emp-not-inc	44.969697	17	90
7	State-gov	39.436392	17	81
8	Without-pay	47.785714	19	72

6. Calculate age distribution by country

In [51]:
```
pysqldf = lambda u: sqldf(u, globals())
u='''SELECT native_country,AVG(age) Average ,MIN(age) Minimum ,MAX(age) Maximum FR
a=pysqldf(u)
a
```

Out[51]:

	native_country	Average	Minimum	Maximum
0	?	38.725557	17	90
1	Cambodia	37.789474	18	65
2	Canada	42.545455	17	80
3	China	42.533333	22	75
4	Columbia	39.711864	18	75
5	Cuba	45.768421	21	82
6	Dominican-Republic	37.728571	18	78
7	Ecuador	36.642857	21	90
8	El-Salvador	34.132075	17	79
9	England	41.155556	17	90
10	France	38.965517	20	64
11	Germany	39.255474	18	74
12	Greece	46.206897	22	65
13	Guatemala	32.421875	19	66
14	Haiti	38.272727	17	63
15	Holand-Netherlands	32.000000	32	32
16	Honduras	33.846154	18	58
17	Hong	33.650000	19	60
18	Hungary	49.384615	24	81
19	India	38.090000	17	61
20	Iran	39.418605	22	63
21	Ireland	36.458333	23	68
22	Italy	46.424658	19	77
23	Jamaica	35.592593	18	66
24	Japan	38.241935	19	61
25	Laos	34.722222	19	56
26	Mexico	33.290824	17	81
27	Nicaragua	33.617647	19	67
28	Outlying-US(Guam-USVI-etc)	38.714286	21	63
29	Peru	35.258065	17	69
30	Philippines	39.444444	17	90

	native_country	Average	Minimum	Maximum
31	Poland	43.116667	17	85
32	Portugal	40.297297	19	78
33	Puerto-Rico	40.508772	17	90
34	Scotland	40.416667	18	62
35	South	38.750000	19	90
36	Taiwan	33.823529	20	61
37	Thailand	34.944444	19	55
38	Trinadad&Tobago	41.315789	17	61
39	United-States	38.655662	17	90
40	Vietnam	34.059701	19	73
41	Yugoslavia	38.812500	20	66

7. Compute a new column as 'Net-Capital-Gain' from the two columns 'capital-gain' and 'capital-loss'

In [56]:
```
pysqldf = lambda v: sqldf(v, globals())
v='''SELECT (capital_gain - capital_loss) 'Net-Capital-Gain',*  FROM adult order b
a=pysqldf(v)
a
```

Out[56]:

	Net-Capital-Gain	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relatior
0	99999	54	Self-emp-inc	166459	Prof-school	15	Married-civ-spouse	Prof-specialty	Hus
1	99999	52	Private	152234	HS-grad	9	Married-civ-spouse	Exec-managerial	Hus
2	99999	53	Self-emp-inc	263925	HS-grad	9	Married-civ-spouse	Sales	Hus
3	99999	52	Private	118025	Bachelors	13	Married-civ-spouse	Exec-managerial	Hus
4	99999	46	Private	370119	Prof-school	15	Married-civ-spouse	Prof-specialty	Hus
5	99999	43	Private	176270	Bachelors	13	Married-civ-spouse	Exec-managerial	Hus
6	99999	49	Private	159816	Bachelors	13	Married-civ-spouse	Prof-specialty	
7	99999	50	Private	171338	Some-college	10	Married-civ-spouse	Exec-managerial	Hus
8	99999	22	Self-emp-not-inc	202920	HS-grad	9	Never-married	Prof-specialty	Unma
9	99999	43	Self-emp-inc	172826	Some-college	10	Married-civ-spouse	Sales	Hus

In [ ]: