Read the following data set: https://archive.ics.uci.edu/ml/machine-learning-databases/adult/ (https://archive.ics.uci.edu/ml/machine-learning-databases/adult/)

Rename the columns as per the description from this file: https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.names (https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.names)

In [1]:
```
"""
Task:
Create a sql db from adult dataset and name it sqladb

"""
import numpy as np
import pandas as pd
import sqlite3 as db
from pandasql import sqldf
columns = ['age','workclass','fnlwgt','education','education-num','maritalStatus','occupation',
           'relationship','race','sex','capital-gain','capital-loss','hours-per-week','nativeCountry','Label']
sqladb = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data",names=columns)
pysqldb = lambda q: sqldf(q,globals())
```

1. Select 10 records from the adult sqladb

In [2]:
```
1 pysqldb("SELECT * FROM sqladb LIMIT 10")
```

Out[2]:

	age	workclass	fnlwgt	education	education-num	maritalStatus	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	nativeCountry	l
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<
5	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<
6	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	<
7	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	
8	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	
9	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40	United-States	

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

In [3]:
```
pysqldb(""" select sex,workclass,avg("hours-per-week") from sqladb where sex =' Male' and workclass =' Private' group by sex
```

Out[3]:

	sex	workclass	avg("hours-per-week")
0	Male	Private	42.221226

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

In [4]:
```
pysqldb("""SELECT education,count("education") FROM sqladb group by education""")
```

Out[4]:

	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	5355
10	Doctorate	413
11	HS-grad	10501
12	Masters	1723
13	Preschool	51
14	Prof-school	576
15	Some-college	7291

In [5]:
```
1 pysqldb("""SELECT occupation,count("occupation") FROM sqladb group by occupation""")
```

Out[5]:

	occupation	count("occupation")
0	?	1843
1	Adm-clerical	3770
2	Armed-Forces	9
3	Craft-repair	4099
4	Exec-managerial	4066
5	Farming-fishing	994
6	Handlers-cleaners	1370
7	Machine-op-inspct	2002
8	Other-service	3295
9	Priv-house-serv	149
10	Prof-specialty	4140
11	Protective-serv	649
12	Sales	3650
13	Tech-support	928
14	Transport-moving	1597

In [6]:
```
1 pysqldb("""SELECT relationship,count("relationship") FROM sqladb group by relationship""")
```

Out[6]:

	relationship	count("relationship")
0	Husband	13193
1	Not-in-family	8305
2	Other-relative	981
3	Own-child	5068
4	Unmarried	3446
5	Wife	1568

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

```
1
```

In [7]:
```
1 pysqldb("""SELECT count(*) FROM sqladb where ( maritalStatus =" Married-civ-spouse" or maritalStatus =" Married-spouse
```

Out[7]:

	count(*)
0	540

yes

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

In [8]:
```
1 pysqldb("""SELECT avg("age"),min("age"),max("age"),workclass FROM sqladb group by workclass""")
```

Out[8]:

	avg("age")	min("age")	max("age")	workclass
0	40.960240	17	90	?
1	42.590625	17	90	Federal-gov
2	41.751075	17	90	Local-gov
3	20.571429	17	30	Never-worked
4	36.797585	17	90	Private
5	46.017025	17	84	Self-emp-inc
6	44.969697	17	90	Self-emp-not-inc
7	39.436055	17	81	State-gov
8	47.785714	19	72	Without-pay

6.Calculate age distribution by country

In [9]:
```
1 pysqldb("""SELECT min("age"),max("age"),nativeCountry FROM sqladb group by nativeCountry """)
```

Out[9]:

	min("age")	max("age")	nativeCountry
0	17	90	?
1	18	65	Cambodia
2	17	80	Canada
3	22	75	China
4	18	75	Columbia
5	21	82	Cuba
6	18	78	Dominican-Republic
7	21	90	Ecuador
8	17	79	El-Salvador
9	17	90	England
10	20	64	France
11	18	74	Germany
12	22	65	Greece
13	19	66	Guatemala
14	17	63	Haiti
15	32	32	Holand-Netherlands
16	18	58	Honduras
17	19	60	Hong
18	24	81	Hungary
19	17	61	India
20	22	63	Iran
21	23	68	Ireland
22	19	77	Italy
23	18	66	Jamaica

	min("age")	max("age")	nativeCountry
24	19	61	Japan
25	19	56	Laos
26	17	81	Mexico
27	19	67	Nicaragua
28	21	63	Outlying-US(Guam-USVI-etc)
29	17	69	Peru
30	17	90	Philippines
31	17	85	Poland
32	19	78	Portugal
33	17	90	Puerto-Rico
34	18	62	Scotland
35	19	90	South
36	20	61	Taiwan
37	19	55	Thailand
38	17	61	Trinadad&Tobago
39	17	90	United-States
40	19	73	Vietnam
41	20	66	Yugoslavia

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

In [10]:
```
pysqldb(""" select "capital-gain","capital-loss",("capital-gain" - "capital-loss") as "Net-Capital-Gain" from sqladb"
```

Out[10]:

	capital-gain	capital-loss	Net-Capital-Gain
0	2174	0	2174
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	14084	0	14084
9	5178	0	5178
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	2042	-2042

	capital-gain	capital-loss	Net-Capital-Gain
24	0	0	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
...	...	...	...
32531	0	0	0
32532	0	0	0
32533	0	0	0
32534	0	0	0
32535	0	0	0
32536	0	0	0
32537	0	0	0
32538	15020	0	15020
32539	0	0	0
32540	0	0	0
32541	0	0	0
32542	0	0	0
32543	0	0	0
32544	0	0	0
32545	0	0	0
32546	0	0	0
32547	0	0	0
32548	1086	0	1086
32549	0	0	0

	capital-gain	capital-loss	Net-Capital-Gain
32550	0	0	0
32551	0	0	0
32552	0	0	0
32553	0	0	0
32554	0	0	0
32555	0	0	0
32556	0	0	0
32557	0	0	0
32558	0	0	0
32559	0	0	0
32560	15024	0	15024

32561 rows × 3 columns

In [ ]:
```
1
```