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) Task: Create a sql db from adult dataset and name it sqladb

1. Select 10 records from the adult sqladb
2. Show me the average hours per week of all men who are working in private sector
3. Show me the frequency table for education, occupation and relationship, separately
4. Are there any people who are married, working in private sector and having a masters degree
5. What is the average, minimum and maximum age group for people working in different sectors
6. Calculate age distribution by country
7. Compute a new column as 'Net-Capital-Gain' from the two columns 'capital-gain' and 'capital-loss'

In [2]:
```
# Import all the packages
import pandas as pd
import sqlite3 as sqllite
```

In [3]:
```
# Read the data from the given URL and see the top 5 records
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data')
df.head(5)
```

Out[3]:

	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K

In [4]:
```
# Rename the columns as per the description.
df.columns = ['age', 'workclass', 'fnlwgt', 'education', 'education-num', 'marital-status', 'occupation', 're
lationship' ,'race', 'sex' , 'capital-gain', 'capital-loss' , 'hours-per-week', 'native-country','class' ]

df = df.infer_objects()

#Strip the Object columns.
df_obj = df.select_dtypes(['object'])
df[df_obj.columns] = df_obj.apply(lambda x: x.str.strip())
```

In [5]:
```
# Print after rename.
df.head(5)
```

Out[5]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week
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

In [6]:
```
# Assign the Database and Table Names to local variables.
db_name = 'sqladb.db'
tbl_name = 'adult_names'
```

In [7]:
```
# Open Connection to SQLLIte and insert the data to SQL lite Table.
# Create a sql db from adult dataset and name it sqladb
con = sqllite.connect(db_name)
cur = con.cursor()

wildcards = ','.join(['?'] * len(df.columns))
data = [tuple(x) for x in df.values]

cur.execute("drop table if exists %s" % tbl_name)

col_str = '"' + '","'.join(df.columns) + '"'
cur.execute("create table %s (%s)" % (tbl_name, col_str))

cur.executemany("insert into %s values(%s)" % (tbl_name, wildcards), data)
con.commit()
```

In [8]:
```
# 1. Select 10 records from the adult sqladb
df = pd.read_sql_query("SELECT * FROM adult_names LIMIT 10", con)
df
```

Out[8]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week
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
5	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16
6	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45
7	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50
8	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40
9	37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0	80

In [9]:
```
# 2. Show me the average hours per week of all men who are working in private sector

average = pd.read_sql_query("SELECT CAST ([hours-per-week] as int) as Hours FROM adult_names where sex = 'Mal
e' and workclass ='Private'  ", con)
print(average.mean())
```

```
Hours    42.221226
dtype: float64
```

In [10]:
```
# 3. Show me the frequency table for education, occupation and relationship, separately.
education = pd.read_sql_query("SELECT education, count(education) as Frequency FROM adult_names GROUP BY educ
ation", con)
print(education)
```

```
       education  Frequency
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
```

In [11]:
```
occupation = pd.read_sql_query("SELECT occupation, COUNT(occupation) as Frequency FROM adult_names GROUP BY o
ccupation", con)
print(occupation)
```

```
           occupation  Frequency
0                   ?       1843
1        Adm-clerical       3769
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 [12]:
```
relationship = pd.read_sql_query("SELECT relationship, COUNT(relationship) as Frequency FROM adult_names GROU
P BY relationship", con)
print(relationship)
```

```
     relationship  Frequency
0         Husband      13193
1   Not-in-family       8304
2  Other-relative        981
3       Own-child       5068
4       Unmarried       3446
5            Wife       1568
```

In [13]:
```
#4. Are there any people who are married, working in private sector and having a masters degree
people = pd.read_sql_query("SELECT Count(*) as Count FROM adult_names where education = 'Masters' and workcla
ss ='Private' and [marital-status] like 'Married%'  ", con)
people
```

Out[13]:

	Count
0	540

In [14]:
```
#5. What is the average, minimum and maximum age group for people working in different sectors
agegroup = pd.read_sql_query("SELECT workclass, avg(age) as Average, min(age) as Min, max(age) as Max FROM ad
ult_names group by workclass ", con)
agegroup
```

Out[14]:

	workclass	Average	Min	Max
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

In [15]:
```
#6. Calculate age distribution by country
dfcountry = pd.read_sql_query("SELECT [native-country], age, count(*) as Distribution  FROM adult_names group
 by  [native-country], age ", con)
dfcountry
```

Out[15]:

	native-country	age	Distribution
0	?	17	2
1	?	18	8
2	?	19	5
3	?	20	10
4	?	21	11
5	?	22	12
6	?	23	6
7	?	24	14
8	?	25	11
9	?	26	18
10	?	27	15
11	?	28	19
12	?	29	12
13	?	30	19
14	?	31	18
15	?	32	17
16	?	33	13
17	?	34	24
18	?	35	18
19	?	36	23
20	?	37	22
21	?	38	20
22	?	39	19

	native-country	age	Distribution
23	?	40	12
24	?	41	22
25	?	42	24
26	?	43	14
27	?	44	10
28	?	45	17
29	?	46	15
...	...	...	...
1251	Vietnam	37	2
1252	Vietnam	38	1
1253	Vietnam	40	1
1254	Vietnam	41	1
1255	Vietnam	43	2
1256	Vietnam	44	3
1257	Vietnam	45	3
1258	Vietnam	46	1
1259	Vietnam	48	1
1260	Vietnam	50	1
1261	Vietnam	51	1
1262	Vietnam	52	1
1263	Vietnam	53	1
1264	Vietnam	54	1
1265	Vietnam	63	1
1266	Vietnam	70	1

	native-country	age	Distribution
1267	Vietnam	73	2
1268	Yugoslavia	20	1
1269	Yugoslavia	22	1
1270	Yugoslavia	25	1
1271	Yugoslavia	29	1
1272	Yugoslavia	31	1
1273	Yugoslavia	35	2
1274	Yugoslavia	36	1
1275	Yugoslavia	40	1
1276	Yugoslavia	41	2
1277	Yugoslavia	43	1
1278	Yugoslavia	45	1
1279	Yugoslavia	56	2
1280	Yugoslavia	66	1

1281 rows × 3 columns

In [16]:
```
#7 Compute a new column as 'Net-Capital-Gain' from the two columns 'capital-gain' and 'capital-loss'
df = pd.read_sql_query("SELECT * FROM adult_names", con)
df['Net-Capital-Gain'] = df['capital-gain' ] - df['capital-loss']
df
```

Out[16]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0
5	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0
6	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0
7	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0
8	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0
9	37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss
10	30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0
11	23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0
12	32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0	0
13	40	Private	121772	Assoc-voc	11	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0
14	34	Private	245487	7th-8th	4	Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0	0
15	25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0
16	32	Private	186824	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Male	0	0
17	38	Private	28887	11th	7	Married-civ-spouse	Sales	Husband	White	Male	0	0
18	43	Self-emp-not-inc	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0	0
19	40	Private	193524	Doctorate	16	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0
20	54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0	0

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss
21	35	Federal-gov	76845	9th	5	Married-civ-spouse	Farming-fishing	Husband	Black	Male	0	0
22	43	Private	117037	11th	7	Married-civ-spouse	Transport-moving	Husband	White	Male	0	2042
23	59	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0	0
24	56	Local-gov	216851	Bachelors	13	Married-civ-spouse	Tech-support	Husband	White	Male	0	0
25	19	Private	168294	HS-grad	9	Never-married	Craft-repair	Own-child	White	Male	0	0
26	54	?	180211	Some-college	10	Married-civ-spouse	?	Husband	Asian-Pac-Islander	Male	0	0
27	39	Private	367260	HS-grad	9	Divorced	Exec-managerial	Not-in-family	White	Male	0	0
28	49	Private	193366	HS-grad	9	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0
29	23	Local-gov	190709	Assoc-acdm	12	Never-married	Protective-serv	Not-in-family	White	Male	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
32530	30	?	33811	Bachelors	13	Never-married	?	Not-in-family	Asian-Pac-Islander	Female	0	0

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss
32531	34	Private	204461	Doctorate	16	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0
32532	54	Private	337992	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	Asian-Pac-Islander	Male	0	0
32533	37	Private	179137	Some-college	10	Divorced	Adm-clerical	Unmarried	White	Female	0	0
32534	22	Private	325033	12th	8	Never-married	Protective-serv	Own-child	Black	Male	0	0
32535	34	Private	160216	Bachelors	13	Never-married	Exec-managerial	Not-in-family	White	Female	0	0
32536	30	Private	345898	HS-grad	9	Never-married	Craft-repair	Not-in-family	Black	Male	0	0
32537	38	Private	139180	Bachelors	13	Divorced	Prof-specialty	Unmarried	Black	Female	15020	0
32538	71	?	287372	Doctorate	16	Married-civ-spouse	?	Husband	White	Male	0	0
32539	45	State-gov	252208	HS-grad	9	Separated	Adm-clerical	Own-child	White	Female	0	0
32540	41	?	202822	HS-grad	9	Separated	?	Not-in-family	Black	Female	0	0
32541	72	?	129912	HS-grad	9	Married-civ-spouse	?	Husband	White	Male	0	0
32542	45	Local-gov	119199	Assoc-acdm	12	Divorced	Prof-specialty	Unmarried	White	Female	0	0

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss
32543	31	Private	199655	Masters	14	Divorced	Other-service	Not-in-family	Other	Female	0	0
32544	39	Local-gov	111499	Assoc-acdm	12	Married-civ-spouse	Adm-clerical	Wife	White	Female	0	0
32545	37	Private	198216	Assoc-acdm	12	Divorced	Tech-support	Not-in-family	White	Female	0	0
32546	43	Private	260761	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0
32547	65	Self-emp-not-inc	99359	Prof-school	15	Never-married	Prof-specialty	Not-in-family	White	Male	1086	0
32548	43	State-gov	255835	Some-college	10	Divorced	Adm-clerical	Other-relative	White	Female	0	0
32549	43	Self-emp-not-inc	27242	Some-college	10	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0
32550	32	Private	34066	10th	6	Married-civ-spouse	Handlers-cleaners	Husband	Amer-Indian-Eskimo	Male	0	0
32551	43	Private	84661	Assoc-voc	11	Married-civ-spouse	Sales	Husband	White	Male	0	0
32552	32	Private	116138	Masters	14	Never-married	Tech-support	Not-in-family	Asian-Pac-Islander	Male	0	0

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss
32553	53	Private	321865	Masters	14	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0
32554	22	Private	310152	Some-college	10	Never-married	Protective-serv	Not-in-family	White	Male	0	0
32555	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0
32556	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0
32557	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0
32558	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0
32559	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0

32560 rows × 16 columns