Data Science Masters :Assignment 13

```
Read the following data set:
https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data
Task:
1. Create an sqlalchemy engine using a sample from the data set
2. Write two basic update queries
3. Write two delete queries
4. Write two filter queries
5. Write two function queries
```

In [73]:
```
import pandas as pd
from sqlalchemy import create_engine

df = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data",header=None, names = ['age'
df = df.drop(['prob'],axis=1)
df = df.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
```

In [83]:
```
# 1. Create an sqlalchemy engine using a sample from the data set
engine = create_engine('sqlite://', echo=False)
df.head(10).to_sql('adultdb', con=engine)
engine.execute("SELECT * FROM adultdb;").fetchall()
```

Out[83]:
```
[(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', 'Mal
e', 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, 4
0, 'Cuba'),
 (5, 37, 'Private', 284582, 'Masters', 14, 'Married-civ-spouse', 'Exec-managerial', 'Wife', 'White', 'Female', 0, 0, 4
0, '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')]
```

In [84]:
```
#2. Write two basic update queries
# Solution
# Update 1 query - Updating Male and Female to M and F resepectively...
engine.execute("""
UPDATE adultdb
SET sex = CASE
              WHEN sex='Male' THEN 'M'
              WHEN sex='Female' THEN 'F'
              ELSE sex
          END
""")
engine.execute("Select * from adultdb LIMIT 5;").fetchall()
```

Out[84]:
```
[(0, 39, 'State-gov', 77516, 'Bachelors', 13, 'Never-married', 'Adm-clerical', 'Not-in-family', 'White', 'M', 2174, 0,
40, 'United-States'),
 (1, 50, 'Self-emp-not-inc', 83311, 'Bachelors', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M',
0, 0, 13, 'United-States'),
 (2, 38, 'Private', 215646, 'HS-grad', 9, 'Divorced', 'Handlers-cleaners', 'Not-in-family', 'White', 'M', 0, 0, 40, 'Un
ited-States'),
 (3, 53, 'Private', 234721, '11th', 7, 'Married-civ-spouse', 'Handlers-cleaners', 'Husband', 'Black', 'M', 0, 0, 40, 'U
nited-States'),
 (4, 28, 'Private', 338409, 'Bachelors', 13, 'Married-civ-spouse', 'Prof-specialty', 'Wife', 'Black', 'F', 0, 0, 40, 'C
uba')]
```

In [85]:
```
# Update 2 Query -  Updating Bachelors as BE...
engine.execute("""
UPDATE adultdb
SET education = 'BE'
WHERE education = 'Bachelors';
""")
engine.execute("Select * FROM adultdb LIMIT 5;").fetchall()
```

Out[85]:
```
[(0, 39, 'State-gov', 77516, 'BE', 13, 'Never-married', 'Adm-clerical', 'Not-in-family', 'White', 'M', 2174, 0, 40, 'Un
ited-States'),
 (1, 50, 'Self-emp-not-inc', 83311, 'BE', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 0, 0, 1
3, 'United-States'),
 (2, 38, 'Private', 215646, 'HS-grad', 9, 'Divorced', 'Handlers-cleaners', 'Not-in-family', 'White', 'M', 0, 0, 40, 'Un
ited-States'),
 (3, 53, 'Private', 234721, '11th', 7, 'Married-civ-spouse', 'Handlers-cleaners', 'Husband', 'Black', 'M', 0, 0, 40, 'U
nited-States'),
 (4, 28, 'Private', 338409, 'BE', 13, 'Married-civ-spouse', 'Prof-specialty', 'Wife', 'Black', 'F', 0, 0, 40, 'Cuba')]
```

In [86]:
```
# 3. Write two delete queries
# Solution
# Delete 1 Query - Deleting records for country Cuba
engine.execute("""
DELETE FROM adultdb WHERE nativecountry = 'Cuba';
""")
engine.execute("Select * FROM adultdb;").fetchall()
```

Out[86]:
```
[(0, 39, 'State-gov', 77516, 'BE', 13, 'Never-married', 'Adm-clerical', 'Not-in-family', 'White', 'M', 2174, 0, 40, 'United-States'),
 (1, 50, 'Self-emp-not-inc', 83311, 'BE', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 0, 0, 13, 'United-States'),
 (2, 38, 'Private', 215646, 'HS-grad', 9, 'Divorced', 'Handlers-cleaners', 'Not-in-family', 'White', 'M', 0, 0, 40, 'United-States'),
 (3, 53, 'Private', 234721, '11th', 7, 'Married-civ-spouse', 'Handlers-cleaners', 'Husband', 'Black', 'M', 0, 0, 40, 'United-States'),
 (5, 37, 'Private', 284582, 'Masters', 14, 'Married-civ-spouse', 'Exec-managerial', 'Wife', 'White', 'F', 0, 0, 40, 'United-States'),
 (6, 49, 'Private', 160187, '9th', 5, 'Married-spouse-absent', 'Other-service', 'Not-in-family', 'Black', 'F', 0, 0, 16, 'Jamaica'),
 (7, 52, 'Self-emp-not-inc', 209642, 'HS-grad', 9, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 0, 0, 45, 'United-States'),
 (8, 31, 'Private', 45781, 'Masters', 14, 'Never-married', 'Prof-specialty', 'Not-in-family', 'White', 'F', 14084, 0, 50, 'United-States'),
 (9, 42, 'Private', 159449, 'BE', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 5178, 0, 40, 'United-States')]
```

In [87]:
```
# Delete 2 Query -  Deleting a row where race is black and age above 50
engine.execute("""
DELETE FROM adultdb WHERE race = 'Black' AND age > 50;
""")
engine.execute("Select * FROM adultdb;").fetchall()
```

Out[87]:
```
[(0, 39, 'State-gov', 77516, 'BE', 13, 'Never-married', 'Adm-clerical', 'Not-in-family', 'White', 'M', 2174, 0, 40, 'Un
ited-States'),
 (1, 50, 'Self-emp-not-inc', 83311, 'BE', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 0, 0, 1
3, 'United-States'),
 (2, 38, 'Private', 215646, 'HS-grad', 9, 'Divorced', 'Handlers-cleaners', 'Not-in-family', 'White', 'M', 0, 0, 40, 'Un
ited-States'),
 (5, 37, 'Private', 284582, 'Masters', 14, 'Married-civ-spouse', 'Exec-managerial', 'Wife', 'White', 'F', 0, 0, 40, 'Un
ited-States'),
 (6, 49, 'Private', 160187, '9th', 5, 'Married-spouse-absent', 'Other-service', 'Not-in-family', 'Black', 'F', 0, 0, 1
6, 'Jamaica'),
 (7, 52, 'Self-emp-not-inc', 209642, 'HS-grad', 9, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 0,
0, 45, 'United-States'),
 (8, 31, 'Private', 45781, 'Masters', 14, 'Never-married', 'Prof-specialty', 'Not-in-family', 'White', 'F', 14084, 0, 5
0, 'United-States'),
 (9, 42, 'Private', 159449, 'BE', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 5178, 0, 40, 'U
nited-States')]
```

In [88]:
```
# 4. Write two filter queries
# Solution
# Filter 1 Query - Filtering private secotor employees
engine.execute("SELECT * FROM adultdb WHERE workclass = 'Private';").fetchall()
```

Out[88]:
```
[(2, 38, 'Private', 215646, 'HS-grad', 9, 'Divorced', 'Handlers-cleaners', 'Not-in-family', 'White', 'M', 0, 0, 40, 'Un
ited-States'),
 (5, 37, 'Private', 284582, 'Masters', 14, 'Married-civ-spouse', 'Exec-managerial', 'Wife', 'White', 'F', 0, 0, 40, 'Un
ited-States'),
 (6, 49, 'Private', 160187, '9th', 5, 'Married-spouse-absent', 'Other-service', 'Not-in-family', 'Black', 'F', 0, 0, 1
6, 'Jamaica'),
 (8, 31, 'Private', 45781, 'Masters', 14, 'Never-married', 'Prof-specialty', 'Not-in-family', 'White', 'F', 14084, 0, 5
0, 'United-States'),
 (9, 42, 'Private', 159449, 'BE', 13, 'Married-civ-spouse', 'Exec-managerial', 'Husband', 'White', 'M', 5178, 0, 40, 'U
nited-States')]
```

In [89]:
```
# Filter 2 Query - Filtering based on age <40
engine.execute("SELECT workclass,age,education FROM adultdb WHERE age < 40;").fetchall()
```

Out[89]:
```
[('State-gov', 39, 'BE'),
 ('Private', 38, 'HS-grad'),
 ('Private', 37, 'Masters'),
 ('Private', 31, 'Masters')]
```

In [90]:
```
# 5. Write two function queries
# Solution
# Function 1 Query - Calculating Mean value for age..
engine.execute("SELECT AVG(age) as Avg_ageValue FROM adultdb;").fetchone()
```

Out[90]:
```
(42.25,)
```

In [91]:
```
# Function 2 Query -  Finding edcuation distribution
engine.execute("""SELECT education, COUNT(education) FROM adultdb GROUP BY education;""").fetchall()
```

Out[91]:
```
[('9th', 1), ('BE', 3), ('HS-grad', 2), ('Masters', 2)]
```