

## Assignment Session 14

2. Problem Statement Read the following data set: <https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data> (<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 [1]: 1 import numpy as np
        2 import pandas as pd
        3 import sqlite3
        4
```

Reading dataset

```
In [2]: 1 adult_data_df= pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-d
```

```
In [3]: 1 adult_data_df.head()
```

Out[3]:

	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40
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

Insetring columns to make the dataset tidier

```
In [4]: 1 adult_data_df.columns =["age", "workclass", "fnlwgt", "education", "education_num
```

In [5]: 1 adult\_data\_df.head()

Out[5]:

	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	Bl
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Bl
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	Wh

Starting Engine Q1. Create an sqlalchemy engine using a sample from the data set

In [6]: 1 import sqlalchemy  
2 from sqlalchemy import create\_engine  
3 engine = create\_engine('sqlite:///sqladb', echo=False)

In [7]: 1 adult\_data\_df.to\_sql('sqladb', engine, if\_exists = 'replace')

In [8]: 1 connection = sqlite3.connect("sqladb")  
2 cursor = connection.cursor()

Q2. Write two basic update queries

In [9]: 1 cursor.execute('UPDATE sqladb SET fnlwgt = "1515" WHERE fnlwgt = " 83311"')  
2 print(pd.read\_sql\_query("SELECT \* FROM sqladb", connection).head(2))

	index	age	workclass	fnlwgt	education	education_num	\
0	0	50	Self-emp-not-inc	1515	Bachelors	13	
1	1	38	Private	215646	HS-grad	9	

  

	marital_status	occupation	relationship	race	sex	\
0	Married-civ-spouse	Exec-managerial	Husband	White	Male	
1	Divorced	Handlers-cleaners	Not-in-family	White	Male	

  

	capital_gain	capital_loss	hours_per_week	native_country	income
0	0	0	13	United-States	<=50K
1	0	0	40	United-States	<=50K

```
In [10]: 1 cursor.execute('UPDATE sqladb SET education = "B.A" WHERE education = " Bache
2 print(pd.read_sql_query("SELECT * FROM sqladb", connection).head())
```

	index	age	workclass	fnlwgt	education	education_num	\
0	0	50	Self-emp-not-inc	1515	B.A	13	
1	1	38	Private	215646	HS-grad	9	
2	2	53	Private	234721	11th	7	
3	3	28	Private	338409	B.A	13	
4	4	37	Private	284582	Masters	14	

  

		marital_status	occupation	relationship	race	sex	\
0		Married-civ-spouse	Exec-managerial	Husband	White	Male	
1		Divorced	Handlers-cleaners	Not-in-family	White	Male	
2		Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	
3		Married-civ-spouse	Prof-specialty	Wife	Black	Female	
4		Married-civ-spouse	Exec-managerial	Wife	White	Female	

  

		capital_gain	capital_loss	hours_per_week	native_country	income
0		0	0	13	United-States	<=50K
1		0	0	40	United-States	<=50K
2		0	0	40	United-States	<=50K
3		0	0	40	Cuba	<=50K
4		0	0	40	United-States	<=50K

Q3. Write two delete queries

```
In [11]: 1 cursor.execute('DELETE FROM sqladb WHERE fnlwgt = " 1515"')
2 print(pd.read_sql_query("SELECT * FROM sqladb", connection).head(2))
```

	index	age	workclass	fnlwgt	education	education_num	marital_status	\
0	1	38	Private	215646	HS-grad	9	Divorced	
1	2	53	Private	234721	11th	7	Married-civ-spouse	

  

		occupation	relationship	race	sex	capital_gain	\
0		Handlers-cleaners	Not-in-family	White	Male	0	
1		Handlers-cleaners	Husband	Black	Male	0	

  

		capital_loss	hours_per_week	native_country	income
0		0	40	United-States	<=50K
1		0	40	United-States	<=50K

```
In [12]: 1 cursor.execute('DELETE FROM sqladb WHERE age = " 53" AND fnlwgt = " 234721" a
2 print(pd.read_sql_query("SELECT * FROM sqladb", connection).head())
```

	index	age	workclass	fnlwgt	education	education_num	\
0	1	38	Private	215646	HS-grad	9	
1	3	28	Private	338409	B.A	13	
2	4	37	Private	284582	Masters	14	
3	5	49	Private	160187	9th	5	
4	6	52	Self-emp-not-inc	209642	HS-grad	9	

  

	marital_status	occupation	relationship	race	\
0	Divorced	Handlers-cleaners	Not-in-family	White	
1	Married-civ-spouse	Prof-specialty	Wife	Black	
2	Married-civ-spouse	Exec-managerial	Wife	White	
3	Married-spouse-absent	Other-service	Not-in-family	Black	
4	Married-civ-spouse	Exec-managerial	Husband	White	

  

	sex	capital_gain	capital_loss	hours_per_week	native_country	income
0	Male	0	0	40	United-States	<=50K
1	Female	0	0	40	Cuba	<=50K
2	Female	0	0	40	United-States	<=50K
3	Female	0	0	16	Jamaica	<=50K
4	Male	0	0	45	United-States	>50K

Q4 Write two filter queries

```
In [13]: 1 cursor.execute('SELECT DISTINCT * FROM sqladb WHERE workclass = " Private" AND
2 output = cursor.fetchall()
3 print(output)
```

```
[(141, 19, ' Private', 316868, ' Some-college', 10, ' Never-married', ' Other-s
ervice', ' Own-child', ' White', ' Male', 0, 0, 30, ' Mexico', ' <=50K'), (334,
17, ' Private', 270942, ' 5th-6th', 3, ' Never-married', ' Other-service', ' Ot
her-relative', ' White', ' Male', 0, 0, 48, ' Mexico', ' <=50K'), (2721, 19, '
Private', 275889, ' 11th', 7, ' Never-married', ' Handlers-cleaners', ' Own-chi
ld', ' White', ' Male', 0, 0, 40, ' Mexico', ' <=50K'), (3117, 19, ' Private',
130431, ' 5th-6th', 3, ' Never-married', ' Farming-fishing', ' Not-in-family',
' White', ' Male', 0, 0, 36, ' Mexico', ' <=50K'), (7753, 18, ' Private', 44482
2, ' 11th', 7, ' Never-married', ' Sales', ' Own-child', ' White', ' Female',
0, 0, 8, ' Mexico', ' <=50K'), (9959, 19, ' Private', 228238, ' HS-grad', 9, '
Never-married', ' Machine-op-inspct', ' Other-relative', ' White', ' Male', 0,
0, 40, ' Mexico', ' <=50K'), (13285, 18, ' Private', 333611, ' 5th-6th', 3, ' N
ever-married', ' Other-service', ' Other-relative', ' White', ' Male', 0, 0, 5
4, ' Mexico', ' <=50K'), (14156, 17, ' Private', 270942, ' HS-grad', 9, ' Never
-married', ' Other-service', ' Other-relative', ' White', ' Male', 0, 0, 35, '
Mexico', ' <=50K'), (14519, 18, ' Private', 79443, ' 9th', 5, ' Never-married',
' Machine-op-inspct', ' Own-child', ' White', ' Male', 0, 0, 40, ' Mexico', ' <
=50K'), (21785, 18, ' Private', 335065, ' 7th-8th', 4, ' Never-married', ' Sale
s', ' Own-child', ' White', ' Male', 0, 0, 30, ' Mexico', ' <=50K'), (22106, 1
7, ' Private', 438996, ' 10th', 6, ' Never-married', ' Other-service', ' Other-
relative', ' White', ' Male', 0, 0, 40, ' Mexico', ' <=50K'), (23874, 19, ' Pri
vate', 316797, ' 7th-8th', 4, ' Married-civ-spouse', ' Handlers-cleaners', ' Ow
n-child', ' White', ' Male', 0, 0, 45, ' Mexico', ' <=50K'), (24351, 18, ' Priv
ate', 184693, ' Some-college', 10, ' Never-married', ' Other-service', ' Own-ch
ild', ' White', ' Female', 0, 0, 20, ' Mexico', ' <=50K'), (24950, 19, ' Privat
e', 277695, ' 9th', 5, ' Never-married', ' Farming-fishing', ' Other-relative',
' White', ' Male', 0, 0, 16, ' Mexico', ' <=50K'), (27404, 18, ' Private', 4268
36, ' 5th-6th', 3, ' Never-married', ' Handlers-cleaners', ' Other-relative', '
White', ' Male', 0, 0, 40, ' Mexico', ' <=50K'), (27530, 19, ' Private', 47280
7, ' 1st-4th', 2, ' Never-married', ' Handlers-cleaners', ' Other-relative', '
White', ' Male', 0, 0, 52, ' Mexico', ' <=50K'), (27793, 19, ' Private', 22577
5, ' Some-college', 10, ' Never-married', ' Craft-repair', ' Not-in-family', '
White', ' Male', 0, 0, 45, ' Mexico', ' <=50K'), (28835, 19, ' Private', 31197
4, ' HS-grad', 9, ' Never-married', ' Craft-repair', ' Other-relative', ' Whit
e', ' Male', 0, 0, 25, ' Mexico', ' <=50K'), (29456, 19, ' Private', 340567, '
1st-4th', 2, ' Never-married', ' Handlers-cleaners', ' Not-in-family', ' Whit
e', ' Male', 0, 0, 55, ' Mexico', ' <=50K'), (29875, 19, ' Private', 311974, '
1st-4th', 2, ' Never-married', ' Machine-op-inspct', ' Other-relative', ' Whit
e', ' Male', 0, 0, 40, ' Mexico', ' <=50K')]
```

```
In [14]: 1 cursor.execute('SELECT DISTINCT * FROM sqladb WHERE education = " Masters" AN
2 output = cursor.fetchall()
3 print(output)
```

```
[(377, 46, ' Self-emp-not-inc', 80914, ' Masters', 14, ' Divorced', ' Exec-manag
erial', ' Not-in-family', ' White', ' Male', 0, 0, 30, ' United-States', ' <=5
0K'), (568, 47, ' Private', 87490, ' Masters', 14, ' Divorced', ' Exec-managerial', ' Unmarried', ' White', ' Male', 0, 0, 42, ' United-States', ' <=50K'), (995, 45, ' Private', 89325, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 45, ' United-States', ' <=50K'), (1674, 43, ' Private', 37937, ' Masters', 14, ' Divorced', ' Exec-managerial', ' Unmarried', ' White', ' Male', 0, 0, 50, ' United-States', ' <=50K'), (2185, 47, ' Self-emp-not-inc', 77102, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (3159, 43, ' Local-gov', 247514, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Unmarried', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (3978, 48, ' Private', 96798, ' Masters', 14, ' Divorced', ' Sales', ' Not-in-family', ' White', ' Male', 0, 0, 35, ' United-States', ' <=50K'), (5282, 40, ' Private', 167265, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 43, ' United-States', ' <=50K'), (5304, 42, ' Federal-gov', 74680, ' Masters', 14, ' Divorced', ' Adm-clerical', ' Not-in-family', ' White', ' Male', 0, 2001, 60, ' United-States', ' <=50K'), (5647, 45, ' State-gov', 67544, ' Masters', 14, ' Divorced', ' Protective-serv', ' Not-in-family', ' White', ' Male', 0, 0, 50, ' United-States', ' <=50K'), (5925, 42, ' Private', 259643, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 4650, 0, 40, ' United-States', ' <=50K'), (8184, 37, ' Private', 312271, ' Masters', 14, ' Divorced', ' Exec-managerial', ' Not-in-family', ' White', ' Male', 0, 0, 65, ' United-States', ' <=50K'), (8389, 45, ' Local-gov', 195418, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' Black', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (8765, 42, ' Local-gov', 176716, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (11256, 37, ' Private', 372525, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Unmarried', ' White', ' Male', 0, 0, 48, ' United-States', ' <=50K'), (11817, 36, ' State-gov', 108320, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Unmarried', ' White', ' Male', 5455, 0, 30, ' United-States', ' <=50K'), (17227, 41, ' Local-gov', 33068, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 1974, 40, ' United-States', ' <=50K'), (22433, 45, ' Private', 168262, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (25378, 45, ' State-gov', 127089, ' Masters', 14, ' Divorced', ' Exec-managerial', ' Not-in-family', ' White', ' Male', 0, 0, 45, ' United-States', ' <=50K'), (25519, 48, ' State-gov', 49595, ' Masters', 14, ' Divorced', ' Protective-serv', ' Not-in-family', ' White', ' Male', 0, 0, 72, ' United-States', ' <=50K'), (25877, 45, ' Self-emp-not-inc', 246891, ' Masters', 14, ' Divorced', ' Craft-repair', ' Not-in-family', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (26218, 42, ' State-gov', 219682, ' Masters', 14, ' Divorced', ' Exec-managerial', ' Not-in-family', ' White', ' Male', 0, 0, 45, ' United-States', ' <=50K'), (29188, 40, ' Private', 56072, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 20, ' United-States', ' <=50K'), (29536, 46, ' Local-gov', 133969, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' American-Indian-Eskimo', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (29694, 47, ' Private', 606752, ' Masters', 14, ' Divorced', ' Tech-support', ' Not-in-family', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K'), (30270, 41, ' State-gov', 100800, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-family', ' White', ' Male', 0, 0, 35, ' United-States', ' <=50K'), (30373, 42, ' Sta
```

```
te-gov', 121265, ' Masters', 14, ' Divorced', ' Prof-specialty', ' Not-in-famil  
y', ' White', ' Male', 0, 0, 40, ' United-States', ' <=50K')]
```

Q5. Write two function queries

In [4]: 1 `import sqlite3`

In [5]: 1 `def age_check():`  
2  `connection = sqlite3.connect("sqladb")`  
3  `cursor = connection.cursor()`  
4  `cursor.execute('SELECT avg(age) FROM sqladb WHERE marital_status=" Divorced")`  
5  `output = cursor.fetchall()`  
6  `print(output)`  
7  `connection.close()`  
8  
9 `age_check()`  
10  
11

`[(44.306666666666665,)]`

In [6]: 1 `def weekly_hours():`  
2  `connection = sqlite3.connect('sqladb')`  
3  `cursor = connection.cursor()`  
4  `cursor.execute('SELECT avg(hours_per_week) FROM sqladb where occupation = "Teacher")`  
5  `output = cursor.fetchall()`  
6  `print(output)`  
7  `connection.close()`  
8 `weekly_hours()`  
9

`[(38.11081322609473,)]`