wxprofessor / **acadgild.assignment.14.1**

Branch: master ▾ **acadgild.assignment.14.1** / **assignment.14.1.ipynb** Find file | Copy path


wxprofessor Add files via upload `34c25ba` on May 18

1 contributor

629 lines (628 sloc) 21.9 KB

Acadgild Assignment 14.1

Patrick L. Francis

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 [ ]:
```
# Import
import random
import numpy as np
import pandas as pd
import sqlite3
from Ipython.display import display
```

Read the Dataset and show the first few rows

In [71]:
```
adult_data_df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data')
display(adult_data_df.head(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

Insert some column names for cleanliness

In [72]:
```
adult_data_df.columns =["age","workclass","fnlwgt","education","education_num","marital_status","occupation","relationship","race","sex","capital_gain","capital_loss","hours_per_week","native_country","income"]
display(adult_data_df.head(3))
```

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0

Start Your Engines!

In [36]:
```
import sqlalchemy

from sqlalchemy import create_engine

engine = create_engine('sqlite:///sqladb', echo=False)
```

In [37]:
```
adult_data_df.to_sql('sqladb', engine, if_exists='replace')
```

In [38]:
```
connection = sqlite3.connect("sqladb")
cursor = connection.cursor()
```

Question 1: Write two basic update queries

In [45]:
```
cursor.execute('UPDATE sqladb SET fnlwgt = "7777" WHERE fnlwgt = " 83311"')
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    7777    Bachelors             13
1      1   38            Private  215646  High_School              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 [77]:
```
cursor.execute('UPDATE sqladb SET education = "High_School" WHERE education = " HS-grad"')
print(pd.read_sql_query("SELECT * FROM sqladb", connection).tail(7))
```

```
       index  age      workclass  fnlwgt      education  education_num  \
32550  32553   53        Private  321865        Masters             14
32551  32554   22        Private  310152   Some-college             10
32552  32555   27        Private  257302     Assoc-acdm             12
32553  32556   40        Private  154374    High_School              9
32554  32557   58        Private  151910    High_School              9
32555  32558   22        Private  201490    High_School              9
32556  32559   52   Self-emp-inc  287927    High_School              9

            marital_status          occupation    relationship    race  \
32550   Married-civ-spouse     Exec-managerial         Husband   White
32551        Never-married     Protective-serv   Not-in-family   White
32552   Married-civ-spouse        Tech-support            Wife   White
32553   Married-civ-spouse   Machine-op-inspct         Husband   White
32554              Widowed        Adm-clerical       Unmarried   White
32555        Never-married        Adm-clerical       Own-child   White
32556   Married-civ-spouse     Exec-managerial            Wife   White

           sex  capital_gain  capital_loss  hours_per_week  native_country  \
32550     Male             0             0              40   United-States
32551     Male             0             0              40   United-States
32552   Female             0             0              38   United-States
32553     Male             0             0              40   United-States
32554   Female             0             0              40   United-States
32555     Male             0             0              20   United-States
32556   Female         15024             0              40   United-States

       income
32550    >50K
32551   <=50K
32552   <=50K
32553    >50K
32554   <=50K
32555   <=50K
32556    >50K
```

Question 2: Write two delete queries

In [50]:
```
cursor.execute('DELETE FROM sqladb WHERE fnlwgt = " 7777"')
print(pd.read_sql_query("SELECT * FROM sqladb", connection).head(2))
```

```
   index  age workclass  fnlwgt    education  education_num  \
0      1   38   Private  215646  High_School              9
1      2   53   Private  234721         11th              7

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

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

In [52]:
```
cursor.execute('DELETE FROM sqladb WHERE age = " 38" AND fnlwgt = " 215646" and race = " White" and hours_
per_week > 39')
print(pd.read_sql_query("SELECT * FROM sqladb", connection).head(2))
```

```
   index  age workclass  fnlwgt   education  education_num  \
0      2   53   Private  234721        11th              7
1      3   28   Private  338409   Bachelors             13

        marital_status          occupation relationship    race      sex  \
0   Married-civ-spouse   Handlers-cleaners      Husband   Black     Male
1   Married-civ-spouse      Prof-specialty         Wife   Black   Female

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

Question 3: Write Two Filter Queries

In [58]:

```
cursor.execute('SELECT DISTINCT * FROM sqladb WHERE relationship = " Wife" AND age < 30 AND workclass = " Private" and native_country=" India"')
output = cursor.fetchall()
print(output)
```

```
[(891, 28, ' Private', 164170, ' Assoc-voc', 11, ' Married-civ-spouse', ' Adm-clerical', ' Wife', ' Asian-Pac-Islander', ' Female', 0, 0, 40, ' India', ' <=50K'), (30832, 25, ' Private', 110978, ' Assoc-acdm', 12, ' Married-civ-spouse', ' Adm-clerical', ' Wife', ' Asian-Pac-Islander', ' Female', 0, 0, 37, ' India', ' >50K')]
```

In [70]:

```
cursor.execute('SELECT * FROM sqladb WHERE relationship = " Unmarried" AND age > 21 AND age < 30 AND sex = " Female" AND native_country=" Mexico"')
output = cursor.fetchall()
print(output)
```

```
[(4561, 29, ' Private', 370494, 'High_School', 9, ' Never-married', ' Other-service', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (6520, 28, ' ?', 201844, 'High_School', 9, ' Separated', ' ?', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (8984, 28, ' ?', 196630, ' Assoc-voc', 11, ' Separated', ' ?', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (10606, 25, ' Private', 204219, 'High_School', 9, ' Never-married', ' Adm-clerical', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (12047, 22, ' Private', 353039, 'High_School', 9, ' Never-married', ' Craft-repair', ' Unmarried', ' White', ' Female', 0, 0, 36, ' Mexico', ' <=50K'), (13955, 23, ' Private', 218445, ' 5th-6th', 3, ' Never-married', ' Priv-house-serv', ' Unmarried', ' White', ' Female', 0, 0, 12, ' Mexico', ' <=50K'), (16345, 22, ' ?', 214238, ' 7th-8th', 4, ' Never-married', ' ?', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (24516, 22, ' Private', 213902, 'High_School', 9, ' Never-married', ' Adm-clerical', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (26675, 29, ' Private', 84366, ' 10th', 6, ' Married-spouse-absent', ' Adm-clerical', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (26717, 22, ' Private', 176321, ' 7th-8th', 4, ' Never-married', ' Other-service', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (27878, 28, ' Private', 261725, 'High_School', 9, ' Never-married', ' Other-service', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (29370, 24, ' Private', 86065, 'High_School', 9, ' Never-married', ' Transport-moving', ' Unmarried', ' White', ' Female', 0, 0, 40, ' Mexico', ' <=50K'), (31101, 27, ' Private', 363053, ' 9th', 5, ' Never-married', ' Priv-house-serv', ' Unmarried', ' White', ' Female', 0, 0, 24, ' Mexico', ' <=50K')]
```

Question 4: Write two function queries

In [79]:

```
def new_entry(db_file, new_data):

    query = "INSERT INTO sqladb VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?);"

    cursor.execute(query, list(new_data))
    cursor.close()
    connection.close()

new_entry('sqladb',('32557', '50','Private', '384675', 'HS-grad', '9', 'Divorced', 'Executive', 'Not-in-family', 'White', 'Male', '0', '0', '40', 'United-States', '>=50K'))
```

In [123]:

```
def age_check():
    connection = sqlite3.connect("sqladb")
    cursor = connection.cursor()
    cursor.execute('SELECT avg(age) FROM sqladb WHERE marital_status=" Never-married" AND sex=" Female"')
    output = cursor.fetchall()
    print(output)
    connection.close()

age_check()
```