

Exercise 3 Predict 420 Summer 2017 Section 56

Zeeshan Latifi

1. sqlite3 command-window shows only one table before running the Ipython Notebook script

```
Zeeshan — sqlite3 ~/Desktop/PREDICT 420/Week 3/Sync Session 3-5/Exercis...
Last login: Sun Jul  9 12:34:41 on ttys000
[Zeeshans-MacBook-Pro:~ Zeeshan$ sqlite3 '/Users/Zeeshan/Desktop/PREDICT 420/Week
3/Sync Session 3-5/ExercisePractice3/xyz.db'
SQLite version 3.16.0 2016-11-04 19:09:39
Enter ".help" for usage hints.
[sqlite> .tables
xyztrans
```

2. First Run for the Ipython Notebook script with no error

How many unique customer records do you now have? By the way, note that you could have limited your examination to just one or more columns, for example just ACCTNO, customer account number, by providing ACCTNO as an argument or by using it to define a Series:

```
In [27]: xyzcust10rev1.duplicated('ACCTNO').sum()
Out[27]: 292
```

```
In [28]: xyzcust10rev1.ACCTNO.duplicated().sum()
Out[28]: 292
```

When there are duplicates of a record, which of them do you think `.drop_duplicates()` retains? Now that we've checked for, and have removed, duplicate customer records, from the customer records, let's write them into a new table in xyztrans.db.

```
In [29]: xyzcustUnDup.to_sql('xyzcust', engine)
```

Did it create the table in xyz.db? Check:

```
In [30]: pd.read_sql_table('xyzcust', engine).columns
Out[30]: Index([u'index', u'ACCTNO', u'ZIP', u'ZIP4', u'LTD_SALES', u'LTD_TRANSACTIONS',
              u'YTD_SALES_2009', u'YTD_TRANSACTIONS_2009', u'CHANNEL_ACQUISITION',
              u'BUYER_STATUS', u'ZIP9_SUPERCODE'],
              dtype='object')
```

should produce the columns of the DataFrame you wrote to the db. Remember that "engine" refers to the SQLite3 DB by way of defining the connection using SQLAlchemy's `create_engine` method. How many tables are there now in xyz.db? And, what are their names? Do

```
In [31]: xyzMetaData.tables.keys()
Out[31]: [u'xyztrans']
```

Another way to look at the metadata of an RDB using SQLAlchemy is by using the "inspect" method:

```
In [32]: xyzMetaData
Out[32]: MetaData(bind=Engine(sqlite:///xyz.db))

In [33]: from sqlalchemy import inspect
In [34]: insp=inspect(engine)
In [35]: insp.get_table_names()
Out[35]: [u'xyzcust', u'xyztrans']
```

3. sqlite3 command-windows shows two tables after you run the Ipython Notebook script

```
Zeeshan — sqlite3 ~/Desktop/PREDICT 420/Week 3/Sync Session 3-5/Exercis...
Last login: Sun Jul  9 12:34:41 on ttys000
[Zeeshans-MacBook-Pro:~ Zeeshan$ sqlite3 '/Users/Zeehan/Desktop/PREDICT 420/Week
3/Sync Session 3-5/ExercisePractice3/xyz.db'
SQLite version 3.16.0 2016-11-04 19:09:39
Enter ".help" for usage hints.
[sqlite> .tables
xyztrans
[sqlite> .tables
xyzcust  xyztrans
sqlite> ]
```

4. Second Run for the Ipython Notebook script that shows error

```
In [69]: xyzcust10rev1.ACCTNO.duplicated().sum()
```

```
Out[69]: 292
```

When there are duplicates of a record, which of them do you think `.drop_duplicates()` retains? Now that we've checked for, and have removed, duplicate customer records, from the customer records, let's write them into a new table in `xyztrans.db`.

```
In [70]: xyzcustUnDup.to_sql('xyzcust', engine)
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-70-fec6c0e0f199> in <module>()
----> 1 xyzcustUnDup.to_sql('xyzcust', engine)

/Users/Zeehan/Library/Enthought/Canopy_64bit/User/lib/python2.7/site-packages/pandas/core/generic.py in to_sql
1199     l(self, name, con, flavor, schema, if_exists, index, index_label, chunksize, dtype)
1200         sql.to_sql(self, name, con, flavor=flavor, schema=schema,
-> 1201             if_exists=if_exists, index=index, index_label=index_label,
1202             chunksize=chunksize, dtype=dtype)
1203     def to_pickle(self, path):

/Users/Zeehan/Library/Enthought/Canopy_64bit/User/lib/python2.7/site-packages/pandas/io/sql.py in to_sql(frame
e, name, con, flavor, schema, if_exists, index, index_label, chunksize, dtype)
468     pandas_sql.to_sql(frame, name, if_exists=if_exists, index=index,
469                       index_label=index_label, schema=schema,
-> 470                       chunksize=chunksize, dtype=dtype)
471
472

/Users/Zeehan/Library/Enthought/Canopy_64bit/User/lib/python2.7/site-packages/pandas/io/sql.py in to_sql(self
, frame, name, if_exists, index, index_label, schema, chunksize, dtype)
1145         if_exists=if_exists, index_label=index_label,
1146         schema=schema, dtype=dtype)
-> 1147     table.create()
1148     table.insert(chunksize)
1149     if (not name.isdigit() and not name.islower()):

/Users/Zeehan/Library/Enthought/Canopy_64bit/User/lib/python2.7/site-packages/pandas/io/sql.py in create(self
)
584         if self.exists():
585             if self.if_exists == 'fail':
-> 586                 raise ValueError("Table '%s' already exists." % self.name)
587             elif self.if_exists == 'replace':
588                 self.pd_sql.drop_table(self.name, self.schema)

ValueError: Table 'xyzcust' already exists.
```

Did it create the table in `xyz.db`? Check:

```
In [71]: pd.read_sql_table('xyzcust', engine).columns
```

```
Out[71]: Index([u'index', u'ACCTNO', u'ZIP', u'ZIP4', u'LTD_SALES', u'LTD_TRANSACTIONS',
u'YTD_SALES_2009', u'YTD_TRANSACTIONS_2009', u'CHANNEL_ACQUISITION',
u'BUYER_STATUS', u'ZIP9_SUPERCODE'],
dtype='object')
```

5. sqlite3 command-window shows drop table for xyzcust

```
Zeeshan — sqlite3 ~/Desktop/PREDICT 420/Week 3/Sync Session 3-5/Exercis...
Last login: Sun Jul  9 12:34:41 on ttys000
[Zeeshans-MacBook-Pro:~ Zeeshan$ sqlite3 '/Users/Zeehan/Desktop/PREDICT 420/Week
3/Sync Session 3-5/ExercisePractice3/xyz.db'
SQLite version 3.16.0 2016-11-04 19:09:39
Enter ".help" for usage hints.
[sqlite> .tables
xyztrans
[sqlite> .tables
xyzcust  xyztrans
[sqlite> DROP TABLE xyzcust;
[sqlite> .tables
xyztrans
sqlite> 
```

6. Third Run for the Ipython Notebook script that shows NO error

```
In [108]: xyzcustUnDup=xyzcust10rev1.drop_duplicates()
xyzcustUnDup.duplicated().sum()
```

```
Out[108]: 0
```

How many unique customer records do you now have? By the way, note that you could have limited your examination to just one or more columns, for example just ACCTNO, customer account number, by providing ACCTNO as an argument or by using it to define a Series:

```
In [109]: xyzcust10rev1.duplicated('ACCTNO').sum()
```

```
Out[109]: 292
```

```
In [110]: xyzcust10rev1.ACCTNO.duplicated().sum()
```

```
Out[110]: 292
```

When there are duplicates of a record, which of them do you think `.drop_duplicates()` retains? Now that we've checked for, and have removed, duplicate customer records, from the customer records, let's write them into a new table in xyztrans.db.

```
In [111]: xyzcustUnDup.to_sql('xyzcust', engine)
```

Did it create the table in xyz.db? Check:

```
In [112]: pd.read_sql_table('xyzcust', engine).columns
```

```
Out[112]: Index([u'index', u'ACCTNO', u'ZIP', u'ZIP4', u'LTD_SALES', u'LTD_TRANSACTIONS',
u'YTD_SALES_2009', u'YTD_TRANSACTIONS_2009', u'CHANNEL_ACQUISITION',
u'BUYER_STATUS', u'ZIP9_SUPERCODE'],
dtype='object')
```

should produce the columns of the DataFrame you wrote to the db. Remember that "engine" refers to the SQLite3 DB by way of defining the connection using SQLAlchemy's `create_engine` method. How many tables are there now in xyz.db? And, what are their names? Do

```
In [113]: xyzMetaData.tables.keys()
```

```
Out[113]: [u'xyztrans']
```

Another way to look at the metadata of an RDB using SQLAlchemy is by using the "inspect" method:

```
In [114]: xyzMetaData
```

```
Out[114]: MetaData(bind=Engine(sqlite:///xyz.db))
```

```
In [115]: from sqlalchemy import inspect
```

```
In [116]: insp=inspect(engine)
```

```
In [117]: insp.get_table_names()
```

```
Out[117]: [u'xyzcust', u'xyztrans']
```