



INTRODUCTION TO DATABASES IN PYTHON

Census Case Study



Census Case Study

- Preparing SQLAlchemy and the Database
- Loading Data into the Database
- Solving Data Science Problems with Queries



Part 1: Preparing SQLAlchemy and the Database

- Create an Engine and MetaData object

```
In [1]: from sqlalchemy import create_engine, MetaData
```

```
In [2]: engine = create_engine('sqlite:///census_nyc.sqlite')
```

```
In [3]: metadata = MetaData()
```



Part 1: Preparing SQLAlchemy and the Database

- Create and save the census table

```
In [4]: from sqlalchemy import (Table, Column, String,  
...:     Integer, Decimal, Boolean)
```

```
In [5]: employees = Table('employees', metadata,  
...:     Column('id', Integer()),  
...:     Column('name', String(255)),  
...:     Column('salary', Decimal()),  
...:     Column('active', Boolean()))
```

```
In [6]: metadata.create_all(engine)
```



INTRODUCTION TO DATABASES IN PYTHON

Let's practice!



INTRODUCTION TO DATABASES IN PYTHON

Populating the Database



Part 2: Populating the Database

- Load a CSV file into a values list

```
In [7]: values_list = []
```

```
In [8]: for row in csv_reader:
```

```
...:     data = {'state': row[0], 'sex': row[1],
...:              'age': row[2], 'pop2000': row[3],
...:              'pop2008': row[4]}
...:     values_list.append(data)
```



Part 2: Populating the Database

- Insert the values list into the census table

```
In [9]: from sqlalchemy import insert

In [10]: stmt = insert(employees)

In [11]: result_proxy = connection.execute(stmt,
...:      values_list)

In [12]: print(result_proxy.rowcount)
Out[12]: 2
```



INTRODUCTION TO DATABASES IN PYTHON

Let's practice!



Introduction to Databases in Python

Querying the Database



Part 3: Answering Data Science Questions with Queries

- Determine Average Age for Males and Females

```
In [13]: from sqlalchemy import select
```

```
In [14]: stmt = select([census.columns.sex,  
...:                  (func.sum(census.columns.pop2008 *  
...:                  census.columns.age) /  
...:                  func.sum(census.columns.pop2008)  
...:                  ).label('average_age')])
```

```
In [15]: stmt = stmt.group_by('census.columns.sex')
```

```
In [16]: results = connection.execute(stmt).fetchall()
```



Part 3: Answering Data Science Questions with Queries

- Determine the percentage of Females for each state

```
In [17]: from sqlalchemy import case, cast, Float

In [18]: stmt = select([
...:     (func.sum(
...:         case([
...:             (census.columns.state == 'New York',
...:             census.columns.pop2008)
...:         ], else_=0)) /
...:     cast(func.sum(census.columns.pop2008),
...:     Float) * 100).label('ny_percent')])
```



Part 3: Answering Data Science Questions with Queries

- Determine the top 5 states by population change from 2000 to 2008

```
In [19]: stmt = select([census.columns.age,  
...:                  (census.columns.pop2008-  
...:                  census.columns.pop2000).label('pop_change')  
...: ])
```

```
In [20]: stmt = stmt.order_by('pop_change')
```

```
In [21]: stmt = stmt.limit(5)
```



Introduction to Databases in Python

Let's practice!



INTRODUCTION TO DATABASES IN PYTHON

Congratulations!