STRINGFEST

ANALYTICS

Python + SQLAlchemy: Half-day workshop

Lesson 1: Combining the powers of SQL & Python

Lesson 2: Connecting to a database and returning results

Lesson 3: Sorting & filtering results

Lesson 4: Aggregating & calculating results

Lesson 5: Joining tables

Lesson 6: SQLAlchemy & pandas

Lesson 7: Capstone: creating a data pipeline

Recommended preparation: Introduction to Python, half-day workshop and Introduction to SQL, one-day workshop

Learning Objectives

- Student can compare and contrast the uses of Python and SQL for data analysis
- Student can connect to, retrieve from, and close a database
- Student can sort and filter query results
- Student can group by and create calculations from query results
- Student can join two or more tables
- Student can load query results into pandas DataFrames and analyze
- Student can create an end-to-end data pipeline using Python, SQLAlchemy and pandas

Lesson 1: Combining the powers of SQL & Python
Objective: Student can compare and contrast the uses of Python and SQL for data analysis
Description:
- Databases versus scripting languages
- Declarative versus procedural languages
- Object-oriented programming

Time: 20 minutes
Assets needed: none

Lesson 2: Connecting to a database and returning results
Objective: Student can connect to, retrieve from, and close a database
Description:
- Installing and loading the packages
- Connecting to a database
- Retrieving keys and items
- Retrieving records
- Closing the connection

Time: 45 minutes
Assets needed: Baseball database

Lesson 3: Sorting & filtering results
Objective: Student can sort and filter query results
Description:
- Conditional logic
- Filtering results
- Ordering results

Time: 35 minutes
Assets needed: Baseball database

Lesson 4: Student can group by and create calculations from query results
Objective: Student can create a data manipulation pipeline
- Grouping results
- Creating aggregated calculations
- Creating calculated fields

Time: 35 minutes
Assets needed: Baseball database

Lesson 5: Joining tables
Objective: Student can join two or more tables
- Inner joins
- Left outer joins

Time: 35 minutes
Assets needed: Baseball database

Lesson 6: SQLAlchemy & pandas
Objective: Student can load query results into pandas DataFrames and analyze
- Loading results to a DataFrame
- Summarizing and analyzing in pandas

Time: 30 minutes
Assets needed: Baseball database

Lesson 7: Capstone: creating a data pipeline
Objective: Student can create an end-to-end data pipeline using Python, SQLAlchemy and pandas
- Connect to, retrieve and close a database
- Load results to DataFrame
- Analyze, inspect and interpret results

Time: 35 minutes
Assets needed: Flights database

STRINGFEST
ANALYTICS

Lesson plan developed by George Mount. For more resources like this, visit stringfestanalytics.com