

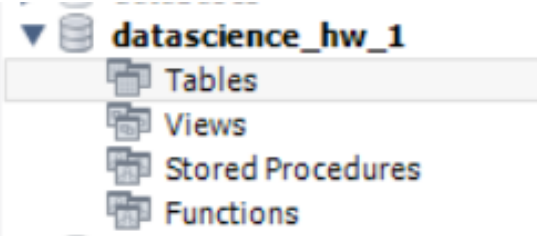
데이터사이언스개론

Python 실습 REPORT

컴퓨터공학과 201811259 배수빈

◆ HW_1 :

엑셀파일 db_score.xlsx 를입력파일로받아서, dataframe 으로 읽어
들이고, 이 데이터들을 MySQL 에 db_score 테이블을 생성하고 모두
투플로 추가하는 python 프로그램을 작성할 것.

개발 설명	<p>Pandas 라이브러리와 create_engine 라이브러리를 import 받아 활용한다.</p> <p>pandas 라이브러리를 활용해서 엑셀파일을 dataframe 으로 저장하고, 해당 dataframe 을 DB 에 저장하기 위해서 create_engine 라이브러리를 활용한다.</p> <p>tosql() 함수를 통해 'db_score' 이름의 테이블을 생성해서 해당 테이블에 데이터 프레임의 내용들을 저장한다.</p>
소스 코드	<pre>import pandas as pd from sqlalchemy import create_engine conn = pymysql.connect(host='localhost', user='root', password='mypassword', db='datascience_hw_1') xl_file = 'db_score.xlsx' df = pd.read_excel(xl_file) engine = create_engine("mysql+pymysql://root:mypassword@ localhost/datascience_hw_1" .format(user="root", pw="your_password", db="pandas")) df.to_sql(con=engine, name='db_score', if_exists='replace')</pre>
실행 결과	<p>1. 해당 소스코드 실행 전 datascience_hw_1 에 테이블이 존재하지 않는 모습</p> 

2. 소스코드 실행 후 datascience_hw_1 에 테이블과 튜플이 생성된 모습

Query 1: `SELECT * from db_score`

	index	sno	attendance	homework	discussion	midterm	final	score	grade
0	1	7.73	19.37	0	35	30.1	92.2	A	
1	2	7.73	20	0	34.12	30.1	91.95	A	
2	3	7.73	19.37	0	34.3	29.75	91.15	A	
3	4	8	18.12	2	31.5	29.05	88.67	A	
4	5	8	20	2	33.25	21.35	84.6	A	
5	6	8	19.37	0	24.5	32.2	84.07	A	
6	7	8	19.37	2	23.8	28.7	81.87	A	
7	8	8	16.87	0	28.87	25.9	79.64	A	
8	9	8	19.37	0	29.75	21.7	78.82	A	
9	10	7.73	20	0	31.5	19.25	78.48	A	
10	11	8	19.37	0	29.57	21.35	78.29	A	
11	12	6.66	19.37	0	31.5	20.65	78.18	A	
12	13	8	19.37	0	29.75	21	78.12	A	
13	14	8	20	0	31.5	18.55	78.05	A	
14	15	7.73	18.12	0	28	23.1	76.95	A	
15	16	8	19.37	0	24.5	24.85	76.72	A	
16	17	7.46	20	0	30.62	18.55	76.63	A	
17	18	8	19.37	0	30.62	18.55	76.54	A	
18	19	8	19.37	0	24.5	24.5	76.37	A	
19	20	8	19.37	0	28	20.65	76.02	A	
20	21	8	19.37	0	35	12.95	75.32	B	
21	22	8	16.87	0	35	15.05	74.07	B	

◆ HW_2 :

db_score 테이블로부터 midterm(중간고사)와 final(기말고사)가 모두 20 점 이상인 학생의 학번, 중간고사, 기말고사를 학번순으로 출력하는 python 프로그램을 작성할 것.

개발 설명

* pymysql 라이브러리를 활용한다.

* 데이터베이스에 연결한 뒤,

```
"select * from db_score where midterm >= 20 and final >= 20"
```

쿼리문을 통해 중간점수와 기말점수가 모두 20 점 이상인 것들만 select 한 데이터들을 커서에 저장한다.

* 해당 데이터들의 sno, midterm, final 의 정보만 하나의 튜플로 생성하고 그 튜플을 리스트에 추가한다.

	<p>* 리스트를 학번에 대해서 오름차순 정렬을 해야 하기 때문에, 리스트에 저장된 각 튜플에 대해 학번을 저장하는 0 번째 인덱스에 대해서 sort 한다.</p> <p>* 학번과 중간점수, 기말점수가 저장되어 있는 각 튜플의 0 번째, 1 번째, 2 번째 인덱스의 정보를 출력한다.</p>
소스 코드	<pre>import pymysql conn = pymysql.connect(host='localhost', user='root', password=' mypassword ', db='datascience_hw_1') curs = conn.cursor(pymysql.cursors.DictCursor) sql = "select * from db_score where midterm >= 20 and final >= 20" curs.execute(sql) row = curs.fetchone() list = [] while row: list.append((row["sno"], row["midterm"], row["final"])) row = curs.fetchone() curs.close() conn.close() list.sort(key = lambda object : object[0]) for object in list: print(f"학번: {object[0]} / 중간: {object[1]} / 기말: {object[2]}")</pre>
실행 결과	<pre>학번: 1 / 중간: 35.0 / 기말: 30.1 학번: 2 / 중간: 34.12 / 기말: 30.1 학번: 3 / 중간: 34.3 / 기말: 29.75 학번: 4 / 중간: 31.5 / 기말: 29.05 학번: 5 / 중간: 33.25 / 기말: 21.35 학번: 6 / 중간: 24.5 / 기말: 32.2 학번: 7 / 중간: 23.8 / 기말: 28.7 학번: 8 / 중간: 28.87 / 기말: 25.9 학번: 9 / 중간: 29.75 / 기말: 21.7 학번: 11 / 중간: 29.57 / 기말: 21.35 학번: 12 / 중간: 31.5 / 기말: 20.65 학번: 13 / 중간: 29.75 / 기말: 21.0 학번: 15 / 중간: 28.0 / 기말: 23.1 학번: 16 / 중간: 24.5 / 기말: 24.85 학번: 19 / 중간: 24.5 / 기말: 24.5 학번: 20 / 중간: 28.0 / 기말: 20.65 학번: 27 / 중간: 22.75 / 기말: 22.05 학번: 29 / 중간: 26.25 / 기말: 22.4 학번: 37 / 중간: 24.5 / 기말: 20.3</pre>