

본 자료와 관련 영상 콘텐츠는 저작권법 제25조 2항에 의해 보호를 받습니다.
본 콘텐츠 및 콘텐츠 일부 문구 등을 외부에 공개하거나, 요약해서 게시하지 말아주세요.
Copyright [잔재미코딩](#) Dave Lee

Sakila 데이터로 익히는 데이터 분석과 SQL

- 다운로드: <https://dev.mysql.com/doc/index-other.html> 에서 sakila database ZIP 파일로 다운받아서, 압축 해제

sakila database sql 파일 실행하기

- Mysql Workbench
 - File -> Open SQL Script -> sakila-schema.sql 선택 후 실행
 - File -> Open SQL Script -> sakila-data.sql 선택 후 실행

sakila 테이블 정보 참고

- <https://downloads.mysql.com/docs/sakila-en.pdf>

실습 환경 맞추기

- Mysql Workbench 에서 Schemes 탭 선택 후, refresh 버튼을 눌러 sakila 데이터베이스 있는지 확인해보기

실습

테이블 정보를 확인해보세요

```
SHOW TABLES
```

```
SELECT * FROM country
```

전체 sakila 데이터베이스 테이블 확인해보기

실제 현업에서 사용하는 데이터베이스에는 다양한 테이블이 있지만, 각 테이블에 대한 매우 친절하고, 상세한 설명이 없는 경우도 많습니다. 데이터를 보며 유추하기도 합니다.

- 전체 데이터에 대한 꼼꼼한 설명은 <https://downloads.mysql.com/docs/sakila-en.pdf> 에 있음

SQL SELECT 문법1 : LIMIT

- 결과중 일부만 데이터 가져오기
 - 예) 결과중 처음부터 10개만 가져오기
 - SELECT * FROM 테이블이름 LIMIT 10 (테이블 데이터 중 최상위에 있는 10개의 데이터만 가져오기)
 - SELECT * FROM 테이블이름 WHERE 조건문 LIMIT 1 (특정 조건에 맞는 데이터 중 최상위에 있는 1개의 데이터만 가져오기)

연습문제를 풀기 위해 이해가 필요한 정보

- film 테이블 : DVD 렌탈샵에서 렌탈할 수 있는 모든 영화 정보를 가지고 있는 테이블
 - 테이블의 각 컬럼은 실제 컬럼값과 함께 확인해야 이해가 쉬움
 - 보통 다양한 컬럼이 있으므로, 이 중에 필요한 컬럼만 이해하면 됨

특정 테이블의 컬럼과 컬럼값 확인을 위해 LIMIT 1 을 많이 사용함

```
SELECT * FROM film LIMIT 1
```

SQL SELECT 문법2: COUNT

- 결과 수 세기 (데이터 행의 수 세기)
 - 예) 결과 수 세기
 - SELECT COUNT(*) FROM 테이블이름 (테이블 전체 데이터 수 세기)
 - SELECT COUNT(*) FROM 테이블이름 WHERE 조건문 (특정 조건에 맞는 테이블 데이터 수 세기)

연습문제1 (직접 해보신 후에, 영상으로 확인하세요!)

영화 테이블(film) 에서 총 영화 수 알아내기

```
SELECT COUNT(*) FROM film
```

SQL SELECT 문법3: DISTINCT

- 특정 컬럼값 출력시 중복된 값을 출력하지 않음
 - 예) 유일한 컬럼값 확인하기
 - SELECT DISTINCT 컬럼명 FROM 테이블이름 (특정 컬럼에 들어가 있는 컬럼값 종류 확인하기)

- SELECT DISTINCT 컬럼명 FROM 테이블이름 WHERE 조건문 (특정 조건에 맞는, 특정 컬럼에 들어가 있는 컬럼값 종류 확인하기)

연습문제2 (직접 해보신 후에, 영상으로 확인하세요!)

영화 테이블(film) 에서 영화 등급 종류 알아내기, 각 영화의 영화 등급은 rating 컬럼에 들어 있음

```
SELECT DISTINCT rating FROM film
```

- 참고: 각 컬럼값 이해를 위한 배경 지식이 필요할 때가 있음 (이것을 도메인이라고 이야기함)
 - https://en.wikipedia.org/wiki/Motion_Picture_Association_film_rating_system

연습문제3 (직접 해보신 후에, 영상으로 확인하세요!)

영화 테이블(film) 에서 영화 release 연도 종류 알아내기, 각 영화의 release 연도는 release_year 컬럼에 있음

```
SELECT DISTINCT release_year FROM film
```

연습문제4 (직접 해보신 후에, 영상으로 확인하세요!)

영화 렌탈 테이블(rental) 에서 10개 데이터만 출력하기, rental 테이블은 DVD 를 언제, 누가 빌려갔고, 반환했는지에 대한 정보를 가지고 있음

```
SELECT * FROM rental LIMIT 10
```

연습문제5 (직접 해보신 후에, 영상으로 확인하세요!)

영화 렌탈 테이블(rental) 에서 inventory_id 가 367 인 로우(Row) 전체 출력하기, inventory_id 는 빌려간 DVD 의 ID 를 의미함

```
SELECT * FROM rental WHERE inventory_id = 367
```

테이블과 테이블 간의 관계

- inventory_id, customer_id 와 같이 다른 테이블에 상세 정보가 있고, 각 정보를 구별할 수 있는 ID 값으로 연결되어 있는 경우가 많음

하나의 테이블에 모든 정보를 담지 않습니다.

```
SELECT * FROM inventory LIMIT 1
SELECT * FROM customer LIMIT 1
SELECT * FROM address LIMIT 1
```

연습문제6 (직접 해보신 후에, 영상으로 확인하세요!)

customer 테이블에서 customer 수 알아내기

```
SELECT COUNT(*) FROM customer
```

SQL SELECT 문법4: SUM, AVG, MAX, MIN

- 특정 컬럼값을 기반으로
 - SUM(): 컬럼값의 합계
 - AVG(): 컬럼값의 평균
 - MAX(): 컬럼값의 최대값
 - MIN(): 컬럼값의 최소값
 - 예) 컬럼값 분석하기
 - SELECT SUM(컬럼명) FROM 테이블이름 (특정 컬럼에 들어가 있는 컬럼값의 합계 구하기)
 - SELECT AVG(컬럼명) FROM 테이블이름 WHERE 조건문 (특정 조건에 맞는, 특정 컬럼에 들어가 있는 컬럼값의 평균 구하기)

연습문제7 (직접 해보신 후에, 영상으로 확인하세요!)

payment 테이블에서 렌탈비용 합계, 평균, 최대값, 최소값 구하기, payment 테이블은 렌탈 비용을 포함한 정보를 담고 있는 테이블임

```
SELECT SUM(amount), AVG(amount), MAX(amount), MIN(amount) FROM payment
```

단계별로 진행하세요: 1단계 - payment 테이블 컬럼 확인하기

단계별로 진행하세요: 2단계 - payment 테이블 amount 컬럼에서 렌탈비용 합계, 평균, 최대값, 최소값 구하기

연습문제8 (직접 해보신 후에, 영상으로 확인하세요!)

영화 렌탈 테이블(rental) 에서 inventory_id 가 367 이고, staff_id가 1인 로우(Row) 전체 출력하기

```
SELECT * FROM rental WHERE inventory_id = 367 AND staff_id = 1
```

SQL SELECT 문법5: GROUP BY

- 특정 컬럼값을 기반으로 그룹핑하기
 - 예) SELECT rating FROM film GROUP BY rating (film 테이블의 rating 값을 그룹핑해라, 즉 rating 값별로 출력하므로, rating 값 종류를 확인할 수 있음)
 - 예) SELECT COUNT(*) FROM film GROUP BY rating (각 rating 값 종류별로, 몇 개의 데이터가 있는지를 확인)
 - 예) SELECT COUNT(*) FROM film WHERE 조건문 GROUP BY rating (특정 조건에 맞는 데이터 중 rating 값 종류별로, 몇 개의 데이터가 있는지를 확인)

연습문제9 (직접 해보신 후에, 영상으로 확인하세요!)

영화(film table)에 매겨진 등급(rating) 종류를 모두 출력하시요 (GROUP BY 를 사용하세요)

```
SELECT rating FROM film GROUP BY rating"
```

연습문제10 (직접 해보신 후에, 영상으로 확인하세요!)

영화(film table)에 매겨진 등급(rating) 종류에 따른 영화 갯수를 모두 출력하시요 (rating 값과 각 rating 값에 따른 영화 갯수를 출력하세요)

```
SELECT rating, COUNT(*) FROM film GROUP BY rating
```

연습문제11 (직접 해보신 후에, 영상으로 확인하세요!)

영화(film table)에서 영화가 PG 또는 G 등급의 영화 수를 각 등급별로 출력하세요 (rating 값과 각 rating 값에 따른 영화 갯수를 출력하세요)

```
SELECT rating, COUNT(*) FROM film WHERE rating = 'PG' OR rating = 'G'  
GROUP BY rating
```

연습문제12 (직접 해보신 후에, 영상으로 확인하세요!)

영화(film table)에서 영화가 G 등급인 영화 제목을 출력하세요

```
SELECT title FROM film WHERE rating = 'G'
```

연습문제13 (직접 해보신 후에, 영상으로 확인하세요!)

영화(film table)에서 영화가 PG 또는 G 등급인 영화 제목을 출력하세요

```
SELECT rating, title FROM film WHERE rating = 'PG' OR rating = 'G'
```

연습문제14 (직접 해보신 후에, 영상으로 확인하세요!)

영화(film table)에서 release 연도가 2006 또는 2007 연도이고, 영화가 PG 또는 G 등급인 영화 제목을 출력하세요

```
SELECT rating, release_year, title FROM film WHERE (rating = 'PG' OR
rating = 'G') AND (release_year = 2006 OR release_year = 2007)
```

처음부터 다음과 같은 데이터를 추출하고자 한다면, 단계를 나눠서 하나씩 작성해보세요

영화(film table)에서 release 연도가 2006 또는 2007 연도이고, 영화가 PG 또는 G 등급인 영화 제목을 출력하세요

SQL 과 데이터 분석

- 실제 요구사항은 여러 조건이 있을 수 있습니다.

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수와 rating (등급), 각 그룹별 평균 rental_rate (렌탈 비용) 출력하세요

접근 방법

- 단계별로 나눠서 SQL을 작성해보세요

연습문제15 (직접 해보신 후에, 영상으로 확인하세요!)

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수 출력하기 (각 등급별 갯수 출력하기)

```
SELECT COUNT(*) FROM film GROUP BY rating
```

연습문제16 (직접 해보신 후에, 영상으로 확인하세요!)

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수 출력하기 (등급과 각 등급별 갯수 출력하기)

```
SELECT rating, COUNT(*) FROM film GROUP BY rating
```

연습문제17 (직접 해보신 후에, 영상으로 확인하세요!)

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수와 각 등급별 평균 렌탈 비용 출력하기 (등급과 각 등급별 갯수, 각 등급별 평균 렌탈 비용 출력하기)

```
SELECT rating, COUNT(*), AVG(rental_rate) FROM film GROUP BY rating
```

SQL SELECT 문법6: ORDER BY

- 특정 컬럼값을 기준으로 데이터 정렬하기
 - ORDER BY 정렬할 기준 컬럼명 DESC|ASC
 - 예) SELECT * FROM film ORDER BY rating DESC (rating 값을 기준으로 내림차순으로 정렬해서 데이터를 가져와라)
 - 예) SELECT * FROM film ORDER BY rating ASC (rating 값을 기준으로 올림차순으로 정렬해서 데이터를 가져와라)
 - 예) SELECT * FROM film WHERE 조건문 ORDER BY rating ASC (특정 조건에 맞는 데이터를 rating 값을 기준으로 올림차순으로 정렬해서 데이터를 가져와라)
 - 예) SELECT COUNT(*) FROM film WHERE 조건문 GROUP BY 컬럼 ORDER BY COUNT(*) ASC (특정 조건에 맞는 데이터를 특정 컬럼값을 기준으로 그룹핑하되, rating 값을 기준으로 올림차순으로 정렬해서 데이터를 가져와라)
 - DESC 또는 ASC 를 안쓰면 ASC 로 정렬
- SQL 조건 순서

```
SELECT 컬럼  
FROM 테이블명  
WHERE 조건  
GROUP BY 컬럼  
ORDER BY 컬럼  
LIMIT
```

연습문제18 (직접 해보신 후에, 영상으로 확인하세요!)

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수와 각 등급별 평균 렌탈 비용 출력하기, 단 평균 렌탈비용이 높은 순으로 출력하기 (등급과 각 등급별 갯수, 각 등급별 평균 렌탈 비용 출력하기)


```
SELECT rating, COUNT(*), AVG(rental_rate) FROM film GROUP BY rating ORDER BY AVG(rental_rate) DESC
```

SQL SELECT 문법7: AS

- 표시할 컬럼명도 다르게 하기
 - 예) SELECT COUNT(*) AS total FROM film (film의 전체 데이터 갯수를 COUNT(*) 로 표시하지 말고, total 로 표시하기)

연습문제19 (직접 해보신 후에, 영상으로 확인하세요!)

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수와 rating (등급), 각 그룹별 평균 rental_rate (렌탈 비용) 출력하되, 영화 갯수와 평균 렌탈 비용은 각각 total_films, avg_rental_rate 으로 출력하고, avg_rental_rate 이 높은 순으로 출력하시오

```
SELECT rating, COUNT(*) AS total_films, AVG(rental_rate) AS avg_rental_rate FROM film GROUP BY rating ORDER BY AVG(rental_rate) DESC
```

연습문제20 (직접 해보신 후에, 영상으로 확인하세요!)

film테이블에서 rating (등급)으로 그룹을 묶어서, 각 등급별 영화 갯수와 rating (등급), 각 그룹별 평균 rental_rate (렌탈 비용) 출력하되, 영화 갯수와 평균 렌탈 비용은 각각 total_films, avg_rental_rate 으로 출력하고, avg_rental_rate 이 높은 순으로 출력하시오 (SQL 구문을 보기 좋게 여러 줄에 걸쳐서 써보기)

```
SELECT
    rating,
    COUNT(*) AS total_films,
    FLOOR(AVG(rental_rate)) AS avg_rental_rate
FROM film
GROUP BY rating
ORDER BY FLOOR(AVG(rental_rate)) DESC
```

실습

- 각 등급별 영화 길이가 130분 이상인 영화의 갯수와 등급을 출력해보세요

```
SELECT rating, COUNT(*) FROM film WHERE length >= 130 GROUP BY rating;
```

본 자료와 관련 영상 콘텐츠는 저작권법 제25조 2항에 의해 보호를 받습니다.

본 콘텐츠 및 콘텐츠 일부 문구 등을 외부에 공개하거나, 요약해서 게시하지 말아주세요.

Copyright [잔재미코딩](#) Dave Lee