

# SQL을 이용한 데이터 분석 프로젝트

# 1. 코로나 데이터 분석하기

## (1) 데이터 수집하기

- OWID(Our World In Data)에서 제공하는 데이터 사용
- 2020/01/01~2021/02/28까지 전 세계 데이터 대상 분석
- 2개 테이블 생성
  - 국가 정보 테이블
  - 코로나 데이터 테이블

# 1. 코로나 데이터 분석하기

## (1) 데이터 수집하기

- covid19\_country(국가 정보) 테이블

칼럼명	데이터 타입	칼럼 설명
countrycode	VARCHAR(10)	국가 코드(기본 키)
countryname	VARCHAR(80)	국가명
continent	VARCHAR(50)	대륙명
population	DOUBLE	인구
population_density	DOUBLE	인구 밀도
median_age	DOUBLE	평균 연령
aged_65_older	DOUBLE	65세 이상 인구 비율
aged_70_older	DOUBLE	70세 이상 인구 비율
hospital_beds_per_thousand	INT	1000명당 병실 침대 수

# 1. 코로나 데이터 분석하기

## (1) 데이터 수집하기

- covid19\_data(코로나 데이터) 테이블

칼럼명	데이터 타입	칼럼 설명
countrycode	VARCHAR(10)	국가 코드(기본 키)
issue_date	DATE	발생일(기본 키)
cases	INT	확진자 수
new_cases_per_million	DOUBLE	100만 명당 확진자 수
deaths	INT	사망자 수
icu_patients	INT	중환자 수
hosp_patients	INT	병원 입원 환자 수
tests	INT	검사자 수
reproduction_rate	DOUBLE	감염재생산지수
new_vaccinations	INT	백신 접종자 수
stringency_index	DOUBLE	방역 지수로, 0에서 100까지의 값이 저장되어 있고 100이 가장 높음

# 1. 코로나 데이터 분석하기

## (2) 데이터 정제하기

- 데이터 정제 : 데이터 클렌징(cleansing), 데이터를 분석하기 쉽도록 데이터 수집과 입력 과정에서 발생한 오류로 생긴, 정확하지 않은 데이터를 걸러내는 작업
- 불필요한 데이터 삭제하기
- 숫자형 칼럼 값 NULL 처리하기

# 1. 코로나 데이터 분석하기

## (3) 데이터 분석하기

- 2020년 사망자 수 상위 10개국 조회하기
- 2020년 인구 대비 확진자 수와 사망자 수 비율 조회하기
- 우리나라의 월별 확진자 수와 사망자 수 조회하기
- 국가별, 월별 확진자 수와 사망자 수 조회하기
- 우리나라의 월별 누적 확진자 수와 사망자 수 조회하기
- 대륙별 사망자 수 상위 3개국 조회하기

## 2. 타이타닉 데이터 분석하기

### (1) 데이터 수집하기

- 타이타닉 사고 데이터를 대상으로 생존자, 사망자 등의 정보 분석(캐글 데이터)
- titanic\_data 테이블 생성

칼럼명	데이터 타입	칼럼 설명
passengerid	INT	승객 아이디(기본 키)
survived	INT	생존 여부(0: 사망, 1: 생존)
pclass	INT	객실 등급(1, 2, 3)
name	VARCHAR(100)	이름
gender	VARCHAR(50)	성별(male: 남성, female: 여성)
age	DOUBLE	나이
sibsp	INT	동반한 형제 및 배우자 수
parch	INT	동반한 부모 및 자녀 수
ticket	VARCHAR(80)	티켓 번호
fare	DOUBLE	요금
cabin	VARCHAR(50)	객실 번호
embarked	VARCHAR(20)	탑승 항구(C: 프랑스 세르부르, Q: 아일랜드 퀸즈타운, S: 영국 사우샘프턴)

## 2. 타이타닉 데이터 분석하기

### (2) 데이터 정제하기

- titanic\_data 테이블의 데이터 정제 내역

대상 칼럼	변경 값
survived	0 → 사망 1 → 생존
gender	Male → 남성 female → 여성
embarked	C → 프랑스 세르부르 Q → 아일랜드 퀸즈타운 S → 영국 사우샘프턴



## 2. 타이타닉 데이터 분석하기

### (3) 데이터 분석하기

- 성별 생존자 수와 사망자 수의 비율 조회하기
- 연령대별 생존자 수와 사망자 수의 비율 조회하기
- 연령대별, 객실 등급별 생존자 수와 사망자 수 조회하기
- 가족 동반과 미동반 시 생존자 수와 사망자 수의 비율 조회하기