

한국의 COVID-19 데이터셋

<https://www.kaggle.com/kimjihoo/coronavirusdataset>

## 1) Case Data

Case: Data of COVID-19 infection cases in South Korea (자료 174개, 변수 8개)

변수명	변수 설명	변수 종류
case_id	the ID of the infection case	int
province	Special City / Metropolitan City / Province(-do) 특별시 / 광역시 / 도	string
city	City(-si) / Country (-gun) / District (-gu) 시 / 군 / 구	string
group	TRUE : group infection / FALSE : not group	bool
infection_case	the infection case (the name of group or other cases)	string
confirmed	the accumulated number of the confirmed 누적 확진 수	int
latitude	the latitude of the group (WGS84)	float
longitude	the longitude of the group (WGS84)	float

## 2) Patient Data

PatientInfo: Epidemiological data of COVID-19 patients in South Korea (자료 5165개, 변수 13개)

변수명	변수 설명	변수 종류
patient_id	the ID of the patient	int
sex	the sex of the patient	string
age	the age of the patient	string
country	the country of the patient	string
province	the province of the patient	string
city	the city of the patient	string
infection_case	the case of infection	string
infected_by	the ID of who infected the patient	int
contact_number	the number of contacts with people	int
symptom_onset_date	the date of symptom onset	date
confirmed_date	the date of being confirmed	date
released_date	the date of being released	date
deceased_date	the date of being deceased	date
state	isolated / released / deceased	string

### 3) Time Series Data

Time: Time series data of COVID-19 status in South Korea (자료 163개, 변수 7개)

변수명	변수 설명	변수 종류
date	YYYY-MM-DD (2020-01-20 ~ 2020-06-30)	date
time	Time (0 = AM 12:00 / 16 = PM 04:00)	int
test	the accumulated number of tests	int
negative	the accumulated number of negative results	int
confirmed	the accumulated number of positive results	int
released	the accumulated number of releases	int
deceased	the accumulated number of deceases	int

TimeAge: Time series data of COVID-19 status in terms of the age in South Korea (자료 1089개, 변수 5개)

변수명	변수 설명	변수 종류
date	YYYY-MM-DD (2020-03-02 ~ 2020-06-30)	date
time	Time (0 = AM 12:00)	int
age	the age of patients (0s, 10s, ... , 90s)	string
confirmed	the accumulated number of the confirmed	int
deceased	the accumulated number of deceases	int

TimeGender: Time series data of COVID-19 status in terms of gender in South Korea (자료 242개, 변수 5개)

변수명	변수 설명	변수 종류
date	YYYY-MM-DD (2020-03-02 ~ 2020-06-30)	date
time	Time (0 = AM 12:00)	int
sex	the gender of patients (male, female)	string
confirmed	the accumulated number of the confirmed	int
deceased	the accumulated number of deceases	int

TimeProvince: Time series data of COVID-19 status in terms of the Province in South Korea (자료 2771개, 변수 6개)

변수명	변수 설명	변수 종류
date	YYYY-MM-DD (2020-01-20 ~ 2020-06-30)	date
time	Time (0 = AM 12:00 / 16 = PM 04:00)	int
province	the province of South Korea (Seoul, Busan, ...)	string
confirmed	the accumulated number of the confirmed in the province	int
released	the accumulated number of the released in the province	int
deceased	the accumulated number of deceases in the province	int

#### 4) Additional Data

Region: Location and statistical data of the regions in South Korea (자료 244개, 변수 12개)

변수명	변수 설명	변수 종류
code	the code of the region	int
province	Special City / Metropolitan City / Province(-do)	string
city	City(-si) / Country (-gun) / District (-gu)	string
latitude	the latitude of the visit (WGS84)	float
longitude	the longitude of the visit (WGS84)	float
elementary_school_count	the number of elementary schools	int
kindergarten_count	the number of kindergartens	int
university_count	the number of universities	int
academy_ratio	the ratio of academies	float
elderly_population_ratio	the ratio of the elderly population	float
elderly_alone_ratio	the ratio of elderly households living alone	float
nursing_home_count	the number of nursing homes	int

Weather: Data of the weather in the regions of South Korea (자료 26271개, 변수 10개)

변수명	변수 설명	변수 종류
code	the code of the region	int
province	Special City / Metropolitan City / Province(-do)	string
date	YYYY-MM-DD (2016-01-01 ~ 2020-06-29)	date
avg_temp	the average temperature	float
min_temp	the lowest temperature	float
max_temp	the highest temperature	float
precipitation	the daily precipitation	float
max_wind_speed	the maximum wind speed	float
most_wind_direction	the most frequent wind direction (0~360)	int
avg_relative_humidity	the average relative humidity	float

SearchTrend: Trend data of the keywords searched in NAVER which is one of the largest portals in South Korea (자료 1642개, 변수 5개)

변수명	변수 설명	변수 종류
date	YYYY-MM-DD (2016-01-01부터 2020-06-29까지)	date
cold	the search volume of 'cold' in Korean language	float
flu	the search volume of 'flu' in Korean language	float
pneumonia	the search volume of 'pneumonia' in Korean language	float
coronavirus	the search volume of 'coronavirus' in Korean language	float

Policy: Data of the government policy for COVID-19 in South Korea (자료 61개, 변수 7개)

변수명	변수 설명	변수 종류
policy_id	the ID of the policy	date
country	the country that implemented the policy	float
type	the type of the policy	float
gov_policy	the policy of the government	float
detail	the detail of the policy	float
start_date	the start date of the policy	date
end_date	the end date of the policy	date