

Group Project Part 3 - ElasticSearch
of Systems and Methods for Big and Unstructured Data Course
(SMBUD)

held by
Brambilla Marco
Tocchetti Andrea

Group 14

Banfi Federico
10581441

Carotenuto Alessandro
10803080

Donati Riccardo
10669618

Mornatta Davide
10657647

Zancani Lea
10608972

Academic year 2021/2022



POLITECNICO
MILANO 1863

1 administration-vaccines-latest.csv

Field Name	Data Type	Description
area	String	Acronyms of the region of delivery
fornitore	String	Complete name of the supplier of the vaccine
data_somministrazione	Datetime	Administration date of the vaccines
fascia_anagrafica	String	Age group of the people administered with the vaccines
Sesso_maschile	Long	Number of vaccinations administered to males
Sesso_femminile	Long	Number of vaccinations administered to females
prima_dose	Long	Number of people administered with the first dose
seconda_dose	Long	Number of people administered with the second dose
pregressa_infezione	Long	Number of people administered with a dose after they have been infected
dose_addizionale_booster	Long	Number of people administered with an additional dose/recall
codice_NUTS1	String	Nomenclature of Territorial Units for Statistics code (Groups of regions)
codice_NUTS2	String	Nomenclature of Territorial Units for Statistics code (Regions)
codice_regione_ISTAT	Long	ISTAT code of a region
nome_area	String	Name of the region

2 dpc-covid19-ita-regions.csv

Field Name	Data Type	Description
data	Datetime	Date of observation
casi_da_screening	Long	Positive cases emerging from surveys and tests, planned at national or regional level
casi_da_sospetto_diagnostico	Long	Positive cases emerged from clinical activity
casi_testati	Long	Total number of people tested
codice_nuts_1	String	Nomenclature of Territorial Units for Statistics code (Groups of regions)
codice_nuts_2	String	Nomenclature of Territorial Units for Statistics code (Regions)
codice_regione	String	Code of the Region (ISTAT 2019)
deceduti	Long	Death
denominazione_regione	String	Name of the Region
dimessi_guariti	Long	Recovered
ingressi_terapia_intensiva	Long	Daily admissions to intensive care
isolamento_domiciliare	Long	Home confinement
lat	Double	Latitude
location	GeoPoint	Geographic Location
long	Double	Longitude
note	String	Notes
note_casi	String	Notes on the cases tested
note_test	String	Notes on the tests carried out
nuovi_positivi	Long	News amount of current positive cases
ricoverati_con_sintomi	Long	Hospitalised patients with symptoms
stato	String	Country of reference
tamponi	Long	Total amount of tests performed
tamponi_test_antigenico_rapido	Long	Antigenic swab tests performed
tamponi_test_molecolare	Long	Molecular swab tests performed
terapia_intensiva	Long	Intensive Care

totale_casi	Long	Total amount of positive cases
totale_ospedalizzati	Long	Total hospitalised patients
totale_positivi	Long	Total amount of current positive cases
totale_positivi _test_antigenico_rapido	Long	Current positive cases with an antigenic swab test
totale_positivi _test_molecolare	Long	Current positive cases with a molecular swab test
variazione_totale_positivi	Long	News amount of current positive cases

3 SARS-CoV-2-variants-data-in-EU-EEA.csv

Field Name	Data Type	Description
country_code	String	2-letter ISO country code
new_cases	Long	Weekly number of new confirmed cases
number_detections_variant	Long	Number of detections reported of the variant
number_sequenced	Long	Weekly number of sequences carried out
number_sequenced_known_variant	Long	Weekly number of known variant sequences carried out
percent_cases_sequenced	Double	100 x new_cases / number_sequenced
percent_variant	Double	100 x number_detections_variant / number_sequenced. No value given if valid_denominator == FALSE
source	String	Data source, either GISAID EpiCoV database or TESSy
valid_denominator	String	GISAID data: TRUE. TESSy data: FALSE if there are discrepancies in the data reported for a given week, such as where the sum of number_detections_variant across all variants exceeds number_sequenced (aggregate data), or where no sequences have been reported that are coded as 'wild type' (case-based data).
variant	String	Each VOC (" <i>variants of concern</i> ", i.e. those considered the most lethal), Other or UNK
year_week	String	Year and week of reference

4 owid-covid-data.csv

Field Name	Data Type	Description
total_cases	Long	Total confirmed cases of COVID-19
new_cases	Long	New confirmed cases of COVID-19
total_deaths	Long	Total deaths attributed to COVID-19
new_deaths	Long	New deaths attributed to COVID-19
location	String	Geographical location
date	Datetime	Date of observation
new_cases_per_million	Long	New confirmed cases of COVID-19 per 1,000,000 people
total_tests	Long	Total tests for COVID-19
new_tests	Long	New tests for COVID-19 (only calculated for consecutive days)
positive_rate	Long	The share of COVID-19 tests that are positive, given as a rolling 7-day average
total_vaccinations	Long	Total number of COVID-19 vaccination doses administered
people_vaccinated	Long	Total number of people who received at least one vaccine dose
people_fully_vaccinated	Long	Total number of people who received all doses prescribed by the vaccination protocol
total_boosters	Long	Total number of COVID-19 vaccination booster doses administered beyond the number prescribed by the vaccination protocol
new_vaccinations	Long	New COVID-19 vaccination doses administered (only calculated for consecutive days)