Datathon metadata documentation

Where to find codes and populations for provinces?

Spain autonomous regions and provinces codes ISO 3166

https://en.wikipedia.org/wiki/ISO_3166-2:ES

List of subdivisions of Spain by population

https://en.wikipedia.org/wiki/List_of_subdivisions_of_Spain_by_population

Files metadata

File casos_hosp_uci_def_sexo_edad_provres.csv

Number of hospitalizations, number of admissions to ICU and number of deaths by sex, age and province of residence.

Variable	description
provincia_iso	ISO code of the province of residence. NC (not stated)
fecha	Cases: In cases prior to May 11, the date of diagnosis is used, in its absence the date of declaration to the community and, in its absence, the key date (date used for statistics by the ACs). In cases after May 10, in the absence of a diagnosis date, the key date is used. Hospitalizations, ICU admissions, deaths: hospitalized cases are represented by hospitalization date (date of hospitalization). represented by date of hospitalization (failing this, the date of diagnosis, and in the absence of the ICU cases by date of admission to ICU (failing this, the date of diagnosis, and failing this, the key date), ICU cases by date of admission to ICU (failing this, the date of diagnosis, and failing this, the key date). date of diagnosis, and failing this, the key date) and deaths by date of death (failing this, the key date) and ICU cases by date of admission to the ICU (failing this, the key date). (failing this, the date of diagnosis, and failing this, the key date).
sex	Sex of cases: H (male), M (female), NC (not

	stated).
grupo_edad	Age group to which the case belongs: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, ≥80 years. NC: not reported.
num_casos	Number of reported cases confirmed as having a positive diagnostic test for active infection (PDIA) as (PDIA) as set out in the Early Detection Strategy, surveillance and control of COVID-19 and in addition to cases reported before May 11 that required hospitalization, ICU admission or hospitalization, ICU admission or death with a clinical diagnosis of COVID-19, according to the definitions of COVID-19. 19, according to the case definitions in force at the time.
num_hosp	Number of hospital admissions
num_uci	Number of ICU admissions
num_def	Number of deaths

Here you can find the official data or at the folder I have shared with you.

File google_trends_symptoms_behaviour_policy_spain_level.csv

Relative number of searches to each term

Variable	description
Semana	Week of the year
Search term	Relative volumen search value from 0 to 100

The sample extracted for you contains multiple terms at the same time.

week of the year - Semana,
cough - tos: (España),
fever - fiebre: (España),
travel - viajar: (España),
lockdown - confinamiento: (España)

Variable groups
symptoms - fever, cough
behaviour - travel
policy - lockdown

You could drill down your search to autonomous regions changing the code region.

Spain.

https://trends.google.com/trends/explore?date=2020-01-01%202021-06-18&geo=ES&q=tos,fiebre,viajar,dolor%20de%20cabeza

Castilla-La Mancha autonomous region. ISO code region highlighted in yellow.

https://trends.google.com/trends/explore?date=2020-01-01%202021-06-18&geo=ES-CM&q=tos,fiebre,viajar,dolor%20de%20cabeza

Try to look for new search terms such as curfew, "toque de queda" in spanish, or new symptoms.

You can use packages in R and python to download data.

Mobility data - Institute of statistics - Until June 2020

INE_daily_mobility_percentage_population.csv

Data at the daily level of provinces,no shapefile is needed.

Variables	Description
Zonas de movilidad	Name of province
Periodo	Date
Total	Percentage of the population that has moved

Official access to data in the folder shared with you data is there

INE_daily_mobility_between_pobA_pobB_202002_202006.zip

Data at the daily level of districts, shapefile shapefiles_celdas_marzo2020.zip is needed to join cells with municipalities and provinces.

Variables	Description
CELDA_ORIGEN	Identifier of the area of residence from which the flows are coming from
NOMBRE_CELDA_ORIGEN	Name of the area of residence from which the flows come from
CELDA_DESTINO	Identifier of the destination area to which the flows arrive
NOMBRE_CELDA_DESTINO	Name of the destination area to which the flows reach
FLUJO	The value of flows, i.e. number of people leaving an area of residence and arriving in a destination area.

Official access to data in the folder shared with you data is there.

Mobility data - Ministry of transportation - Until May 2021

Official access to data, data needs to be downloaded, it isn't at the folder shared with you.

maestra1-mitma-municipios

The master matrix 1 or trip matrix contains the number of trips and traveler-km for each day and each combination of origin, destination, origin_activity, destination_activity, residence, time period and distance (by ranges).

They are compressed text files, with fields separated by '|' (vertical bar) and numeric values have '. numeric values have '. (period) as decimal separator. The "from" and "to" data refer to the district code or a grouping of these in the case of sparsely populated areas (zoned areas). In the case of sparsely populated areas (zoning used and can be downloaded as a Shapefile file "zonificacion-distritos.zip").

Folder route for daily files

maestra1-mitma-municipios -> fichero_diarios

Example name file for daily data.

20200221_maestra_1_mitma_municipio.txt.gz

Variable	description
fecha	date
origen	from of trip statistics
destino	to of trip statistics
periodo	hour of the day
distancia	Group of trips by segments 2-5 km, 5-10km, and so on
viajes	number of trips
viajes_km	total number of kilometers

relacion_municipios_mitmal.csv

File containing the relationship between the districts and the zoning used. In case you want to use this mobility data you need to use the shapefile to decode zones to municipalities.

maestra2-mitma-municipios

This folder contains the matrices of trips per person per day and per full month, updated daily. updated daily. Master matrix 2 or person trip matrix contains the number of persons who have made trips on each date and combination of census tract and range of trips. who have made trips on each date and combination of municipalities and trip range (0, 1, 2 or more than 2 trips).

Folder route for daily files

maestra2-mitma-municipios -> fichero_diarios

20200221_maestra_2_mitma_municipio.txt.gz

Variable	description
fecha	date
distrito	Municipality
numero_viajes	trip range (0, 1, 2 or more than 2 trips).
personas	Number of people moving

zonificacion_municipios.zip

It contains a shapefile with the zoning used corresponding to municipalities or to a grouping of these in the case of sparsely populated areas.