

DESCRIPCIÓN DEL CONJUNTO DE DATOS A TRABAJAR EN EL PROYECTO DE TRATAMIENTO DE DATOS

MINERÍA DE DATOS

20 de abril de 2023

Grupo:	3CV15
--------	-------

Alumnos:

Proyecto No. 1. Tratamiento de datos

Generar un reporte en BIRT que considere los aspectos que se describen a continuación.

1) Descripción del conjunto de datos:

Progreso mundial de vacunación contra COVID-19

COVID-19 World Vaccination Progress

Vacunación diaria y total contra COVID-19 en el mundo de Our World in Data

Enlace de acceso:

<https://www.kaggle.com/gpreda/covid-world-vaccination-progress>

Los datos se recopilan diariamente del repositorio Our World in Data GitHub para covid-19, se fusionan y se cargan. Los datos de vacunación a nivel de país se recopilan y reúnen en un solo archivo. Luego, este archivo de datos se fusiona con el archivo de datos de ubicaciones para incluir información sobre las fuentes de vacunación. Se incluye un segundo archivo, con información de los fabricantes.

Créditos:

Gabriel Preda. Científico de datos

Primer archivo

Los datos (vacunaciones del país) contienen la siguiente información:

Atributo	Descripción	Tipo	Dominio
country	Este es el país para el que se proporciona la información de vacunación	Categorico	222 valores, nombres de países.
iso_code	Código ISO de cada país	Categorico	222 valores, un código por país
date	Fecha para la entrada de datos; para algunas de las fechas solo se tienen las vacunas diarias,	Numérico	01/12/2020 a 17/09/2021

	para otras, solo el total (acumulativo)		
total_vaccinations	Es el número absoluto de inmunizaciones totales en el país	Numérico	$(0, \infty)$
people_vaccinated	Una persona, según el esquema de inmunización, recibirá una o más (normalmente 2) vacunas; en un momento determinado, el número de vacunaciones puede ser mayor que el número de personas	Numérico	$(0, \infty)$
people_fully_vaccinated	Este es el número de personas que recibieron el conjunto completo de inmunización de acuerdo con el esquema de inmunización (típicamente 2), en un momento determinado, puede haber un cierto número de personas que recibieron una vacuna y otro número (menor) de personas que recibieron todas las vacunas del esquema	Numérico	$(0, \infty)$
daily_vaccinations_raw	Para una determinada entrada de datos, el número de vacunaciones para esa fecha/país	Numérico	$(0, \infty)$
daily_vaccinations	Para una determinada entrada de datos, el número de vacunaciones para esa fecha/país	Numérico	$(0, \infty)$
total_vaccinations_per_hundred	Relación (en porcentaje) entre el número de vacunaciones y la población total hasta la fecha en el país	Numérico	$(0, 100)$
people_vaccinated_per_hundred	Relación (en porcentaje) entre la población inmunizada y la población total hasta la fecha en el país	Numérico	$(0, 100)$
people_fully_vaccinated_per_hundred	Relación (en porcentaje) entre la población totalmente inmunizada y la población total hasta la fecha en el país	Numérico	$(0, 100)$
daily_vaccinations_per_million	Relación (en ppm) entre el número de vacunación y la población total para la fecha actual en el país	Numérico	$(0, \infty)$
vaccines	Número total de vacunas utilizadas en el país (actualizadas)	Categorico	CanSino, Covaxin, Moderna,

			Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V, Abdala, Soberana02, QazVac, Sinopharm/HayatVax, Johnson&Johnson.
source_name	Fuente de la información (autoridad nacional, organización internacional, organización local, etc.)	Nominal	Diferentes instituciones gubernamentales . (Ministry of Health, World Health Organization, COVID19 Vaccine information platform, etc.)
source_website	Sitio web de la fuente de información	Nominal	Cadena de texto, URL.

Segundo archivo

Hay un segundo archivo agregado recientemente (vacunas aplicadas de países por fabricante), con las siguientes columnas:

Atributo	Descripción	Tipo	Dominio
location	país	Categorico	222 valores, nombres de países.
date	fecha	Numérico	01/12/2020 a 17/09/2021
vaccine	Tipo de vacuna	Categorico	CanSino, Covaxin, Moderna, Oxford/AstraZeneca , Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V, Abdala, Soberana02,

			QazVac, Sinopharm/HayatV ax, Johnson&Johnson.
Total_vaccinations	Número total de vacunas	Numérico	número total de vacunas/a la fecha y tipo de vacuna

2) Responder las tareas propuestas por el autor:

Realice un seguimiento de la vacunación contra COVID-19 en el mundo, responda las siguientes preguntas:

- ¿Qué país está usando qué vacuna?
- ¿En qué país está más avanzado el programa de vacunación?
- ¿Dónde se vacunan más personas por día? ¿Pero en términos de porcentaje de toda la población?

3) Problemas iniciales:

- Un problema que presenta es que los nombres de las vacunas están en forma de lista. Hay que analizar la forma de identificar la cantidad de cada tipo de vacunas que se están aplicando en total en el conjunto de datos
- Algunos países no reportan cifras de vacunación todos los días, hay que ver cuáles son
- Hay que tener cuidado, ya que la cantidad de vacunas se van acumulando
- Se pueden eliminar los valores faltantes ya que algunos países no reportan cifras de vacunación todos los días, quizá filtrar a partir de un umbral los países de los cuales no se tengan suficientes datos si es que los hay. También sería recomendable un binning para pasar el valor de "vacunas diarias" de numérico a categórico.

4) Acciones a realizar:

- Asignar el continente al que pertenece cada país. Asignar el continente a todos los países. (<https://gist.github.com/kintero/7d1db891401f56256c79>)
- Reportar la cantidad total de vacunas que se aplicó en cada continente (Recategorizar)
- Elegir cinco países de cada continente. Elija a México y muestre los tipos y la cantidad de vacunas que se aplicaron en cada país (sugerencia: usar `country_vaccinations_by_manufacturer`)
- Reportar el porcentaje de la población que ha sido vacunada considerando la población del país.
- ¿Qué países son los que vacunaron más personas en cada continente.
- Crear un ciclo anidado en el cual reporte las vacunas por trimestre que se aplicaron en cada uno de los países (los cinco elegidos) de tres continentes
- Reportar qué lugar ocupa México en América y a nivel mundial en cantidad de vacunas aplicadas, considerando la cantidad de la población.

5) Elegir un país y realice lo siguiente:

- h) Elegir un país y ver el total de vacunas aplicadas por mes, para hacer esto se tiene que filtrar el país y transformar la columna "date" para que esté en meses en lugar de días, al hacer esto se tendrían que sumar las demás columnas de "vacunas diarias".
- i) Se podría tomar un intervalo de fecha, por ejemplo, del 01 de enero de 2021 al 01 de marzo de 2021 (a elegir), y dentro de ese intervalo ver que vacuna fue la más aplicada
- j) Crear binnings de las vacunas aplicadas por día.
- k) Analizar las variables que tiene el conjunto de datos y presentar aspectos descriptivos que es necesario resaltar. Use gráficas y tablas de frecuencias.

Revisar los siguientes países, al parecer son los que actualizan los datos más seguido: UK, USA, México, United Arab Emirates