

Presentación Proyecto Gerardo Gil

Antes que nada, he de decir que mi idea original para este proyecto era comprobar correlaciones entre la actual vacunación y la incidencia de ciertas enfermedades y dolencias que se están asociando a las vacunas de la COVID-19 (equivocadamente o no), pero no he encontrado datos para corroborarlo en la realidad, así que me centraré exclusivamente en los fallecimientos, de los que sí he encontrado datos oficiales.

Después de hacer diversas pruebas, he decidido utilizar 4 datasets, porque con uno solo hubiera sido imposible llegar a alguna conclusión. Con cada uno de los datasets intentaré ahondar un poco más en el tema que estoy investigando, recogiendo las conclusiones extraídas del análisis del dataset anterior.

A continuación, paso a detallar las características de cada uno de los datasets.

1. VAERS

<https://vaers.hhs.gov/data/datasets.html>

Lo que hay aquí es en realidad un conjunto de datasets, que reflejan los eventos adversos reportados por todas las vacunas, desde 1990.

Lo que hago con ellos es una simple aproximación, sin intentar sacar ninguna conclusión (dado que cualquiera puede reportar efectos adversos, por lo que no podemos darle mucha credibilidad).

Se trata simplemente de comparar la cantidad de efectos adversos reportados en cada año, y por cada tipo de vacuna, solamente para tomar una imagen general de la situación, pero sin darle demasiada importancia.

Las columnas que voy a utilizar de estos datasets son las siguientes:

- **VAERS_ID** - Identificador único, que se usa para cruzar datos de las diferentes tablas.
 - **VAX_TYPE** - Tipo de vacuna (las vacunas similares son agrupadas).
 - **RECVDATE** - Fecha en la que el reporte fue recibido por VAERS.
 - **DIED** - Especifica si el receptor de la vacuna murió o no.
 - **DATEDIED** - Fecha de fallecimiento del receptor de la vacuna.
 - **VAX_DATE** - Fecha en la que el receptor fue vacunado.
-

2. MoMo

https://momo.isciii.es/public/momo/dashboard/momo_dashboard.html#datos

A estos datos ya podemos darle al menos cierta credibilidad, así que aquí haremos un análisis más serio.

Lo que intento hacer con ellos es buscar correlaciones entre los picos de fallecimientos y las campañas de vacunación, en toda la población en un principio, y por rangos de edad posteriormente.

Lo cierto es que los grupos de edad están poco logrados en este dataset, ya que incluyen a todos los menores de 65 años en el mismo grupo, pero peor es nada.

Aún así, he podido encontrar cosas interesantes aquí, aunque no definitivas.

Las columnas que voy a utilizar de este dataset son las siguientes:

- **ambito** - Especifica si el ámbito es nacional o comunitario.
 - **cod_gedad** - Código correspondiente al rango de edad (los valores posibles son: "all", "menos_65", "65_74" y "mas_74").
 - **fecha_defuncion** - Fecha de defunción.
 - **defunciones_observadas** - Número de defunciones observadas (incluye la corrección por retraso).
 - **defunciones_esperadas** - Número de defunciones esperadas (en base a un modelo predictivo).
-

3. CDC

<https://data.cdc.gov/NCHS/Weekly-Provisional-Counts-of-Deaths-by-State-and-S/muzy-jte6>

Con este dataset las cosas se ponen muy interesantes, porque discriminan por causa de muerte, aunque no por rango de edad.

Lo que intento hacer con este dataset es un poco lo mismo que con el anterior, pero poniendo el foco en cada una de las causas de fallecimiento. He de decir que aquí he encontrado cosas extremadamente interesantes, y preocupantes.

Las columnas que voy a utilizar de este dataset son las siguientes:

- **Jurisdiction** - Especifica si el ámbito es nacional o estatal.
 - **Year** - Año del fallecimiento.
 - **Week** - Semana del fallecimiento.
 - **All_Cause** - Número de defunciones por todas las causas.
 - **Natural_Cause** - Número de defunciones por causas naturales.
 - **Septicemia** - Número de defunciones por Septicemia.
 - **Malignant_Neoplasms** - Número de defunciones por Cancer.
 - **Diabetes_Mellitus** - Número de defunciones por Diabetes Mellitus.
 - **Alzheimer** - Número de defunciones por Alzheimer.
 - **Influenza_Pneumonia** - Número de defunciones por Gripe y Neumonía.
 - **Lower_Resp_Diseases** - Número de defunciones por Enfermedades Respiratorias crónicas.
 - **Other_Resp_Diseases** - Número de defunciones por otras Enfermedades Respiratorias.
 - **Nephritis_Nephrosis** - Número de defunciones por Nefritis, Síndrome Nefrótico y Nefrosis.
 - **Abnormal_Symptoms** - Número de defunciones por Síntomas, Signos y Hallazgos Clínicos Anormales, no clasificados.
 - **Heart_Diseases** - Número de defunciones por Ataques al Corazón.
 - **Cerebrovascular_Diseases** - Número de defunciones por Enfermedades Cerebrovasculares.
 - **COVID-19_Multiple** - Número de defunciones por COVID-19, como causa múltiple.
 - **COVID-19_Underlying** - Número de defunciones por COVID-19, como causa subyacente.
-

4. CDC

<https://data.cdc.gov/NCHS/AH-Monthly-Provisional-Counts-of-Deaths-for-Select/65mz-jvh5>

Este dataset es similar al anterior, pero además discrimina por grupos de edad (e incluye nada menos que 10 rangos). Esto me permite ahondar más en la investigación, comprobando la incidencia de cada una de las causas de fallecimiento en cada uno de esos rangos de edad.

Desgraciadamente, el último mes que incluye este dataset es setiembre de 2021, y no sé si va a volver a actualizarse, pero me ha resultado muy útil para confirmar o descartar las conclusiones obtenidas del dataset anterior. También es muy interesante lo que he encontrado aquí.

Las columnas que voy a utilizar de este dataset son las siguientes:

- **Year** - Año del fallecimiento.
- **Month** - Mes del fallecimiento.
- **AgeGroup** - Código correspondiente al rango de edad (los valores posibles son: "0-4 years", "5-14 years", "15-24 years", "25-34 years", "35-44 years", "45-54 years", "55-64 years", "65-74 years", "75-84 years" y "85 years and over").
- **All_Cause** - Número de defunciones por todas las causas.
- **Natural_Cause** - Número de defunciones por causas naturales.
- **Septicemia** - Número de defunciones por Septicemia.
- **Malignant_Neoplasms** - Número de defunciones por Cancer.
- **Diabetes_Mellitus** - Número de defunciones por Diabetes Mellitus.
- **Alzheimer** - Número de defunciones por Alzheimer.
- **Influenza_Pneumonia** - Número de defunciones por Gripe y Neumonía.
- **Lower_Resp_Diseases** - Número de defunciones por Enfermedades Respiratorias crónicas.
- **Other_Resp_Diseases** - Número de defunciones por otras Enfermedades Respiratorias.
- **Nephritis_Nephrosis** - Número de defunciones por Nefritis, Síndrome Nefrótico y Nefrosis.
- **Abnormal_Symptoms** - Número de defunciones por Síntomas, Signos y Hallazgos Clínicos Anormales, no clasificados.
- **Heart_Diseases** - Número de defunciones por Ataques al Corazón.
- **Cerebrovascular_Diseases** - Número de defunciones por Enfermedades Cerebrovasculares.
- **COVID-19_Multiple** - Número de defunciones por COVID-19, como causa múltiple.
- **COVID-19_Underlying** - Número de defunciones por COVID-19, como causa subyacente.

Para acabar la presentación, me gustaría aclarar que soy perfectamente consciente de lo espinoso del tema que he decidido tratar, pero creo que en momentos como éste es necesario enfangarse, cada uno dentro de sus posibilidades. Y sobretodo, pese a quien pese.

¿Predecir precios de viviendas o retrasos de vuelos? ¿Clasificar clientes? Sí, está muy bien todo eso, pero creo que ahora mismo hay temas mucho más importantes (e interesantes) en los que ahondar, y no seré yo quien me resista a hacerlo.

Saludos sinceros a Noel (y a todo el que lea esto), y hasta pronto!