

# Relación entre variables meteorológicas e infectados por COVID

*Luis Ramón Mendoza Escareño*

LICENCIATURA EN FÍSICA | PROYECTO METEOROLOGÍA/CIENCIA DE DATOS

8 de Junio del 2023

---

## Abstract

En este proyecto, se buscó establecer correlaciones entre variables meteorológicas, como la velocidad del viento, temperatura y precipitación, y los casos de COVID-19. Dado que ya existen investigaciones al respecto, se intentó replicar los estudios para la Zona Metropolitana de Guadalajara. Se utilizaron bases de datos de un hospital donde se trataban casos de COVID-19, así como datos proporcionados por varias estaciones meteorológicas en la zona.

La parte técnica del proyecto incluyó la recopilación, transformación, estandarización y unificación de diversas bases de datos, tanto meteorológicas como de casos de COVID-19, en una única fuente de información. A partir de esta base de datos consolidada, se llevaron a cabo análisis estadísticos que incluyeron la generación de gráficas y la obtención de estadísticas.

Es importante destacar que este documento no pretende tener un valor científico, sino más bien ser una recopilación del trabajo realizado en este proyecto, los desafíos enfrentados y cómo se resolvieron.

---

# 1 Intento de mapeo de los casos de covid

El primer código generado se basó exclusivamente en la base de datos de COVID-19. En este código, se intentó determinar la ubicación utilizando el módulo de geopandas llamado "geolocator". Utilizando este módulo y los nombres de las localidades únicas, se generaron valores de latitud y longitud para cada una de ellas.

Una vez generados estos valores, se filtró la base de datos para incluir únicamente los casos positivos de COVID-19 y se representaron en un mapa utilizando los correspondientes "círculos de contagio" para cada estación.

Sin embargo, se encontró que el módulo de geolocator no proporcionó correctamente las ubicaciones. Esto se atribuye a que los nombres de las variables de "localización" son muy comunes tanto en México como en España, lo que condujo a la generación de coordenadas incorrectas.

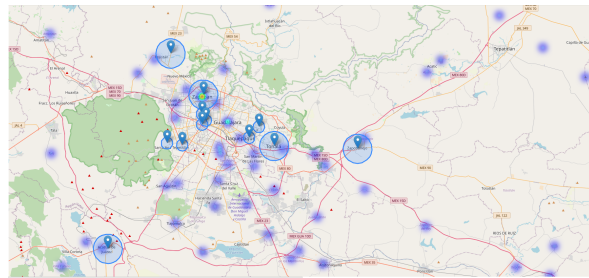


Figure 1: Mapa generado con la metodología antes mencionada, se observa que la mayoría de casos no se encuentran en la Zona deseada. En el mapa además se pueden observar los radios de contagio de 5km en las estaciones de interés.

El código comentado se encuentra en el siguiente link de: [google colab](#).

## 2 Ajuste de la base de datos COVID

En esta etapa, se utilizó nuevamente la base de datos de infectados por COVID-19, pero con una diferencia: el enfoque en el manejo de datos se centró en generar una base de datos que contenga la máxima cantidad de códigos postales. El objetivo es que a través de estos códigos postales, se encontrara manualmente la pertenencia de cada uno a las estaciones más cercanas, así permitiendo categorizar cada contagio en la estación correspondiente.

Para realizar esta asignación, se tuvo en cuenta un radio de 5 km alrededor de cada estación (como se muestra en el mapa mencionado anteriormente). Si un caso de infectado se encontraba dentro de este radio, se consideraba como un caso dentro del círculo de contagio. Los casos que no se encontraban dentro de estos radios se descartaron.

Además de la asignación de cada caso a una estación meteorológica, se generó un código para poder visualizar los casos generales de COVID-19, tanto en general como para cada estación, abarcando todo el período de contagio.

### 2.1 Ajuste de la base de datos de las estaciones meteorológicas

Para el próximo bloque se tomaron en cuenta información de estaciones meteorológicas vecinas al IAM, además de bases de datos del CONAGUA de diferentes zonas, debido a esto los métodos utilizados en estas dos difieren ligeramente.

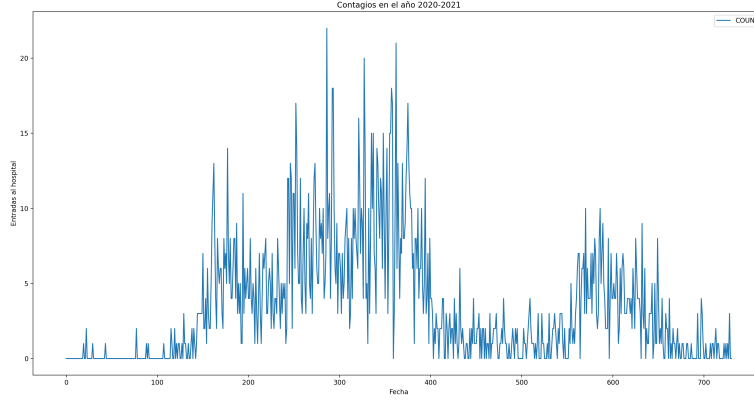


Figure 2: Entradas al hospital a lo largo del tiempo, en el código se puede restringir para observarse el comportamiento para una estación en específico.

## 2.2 Estaciones vecinas al IAM

Las bases de datos utilizadas contienen información de las siguientes estaciones: Universidad Autónoma de GDL (UAG), Tetlán, Preparatoria 20, Molinos Zapopan e IAM.

Para cada estación, se tienen registro de los valores de diversas variables en intervalos de 10 minutos. Con el objetivo de analizar las variables de interés, fue necesario obtener los valores máximos y mínimos por día para cada una de ellas.

Dada la falta de datos en varios días, se implementó un código que asigna el valor 0 a todas las variables en esos días sin información. Además de esto, se abordaron otros problemas de compatibilidad entre las bases de datos para asegurar que los datos pudieran ser representados gráficamente por día, facilitando así la identificación de posibles patrones.

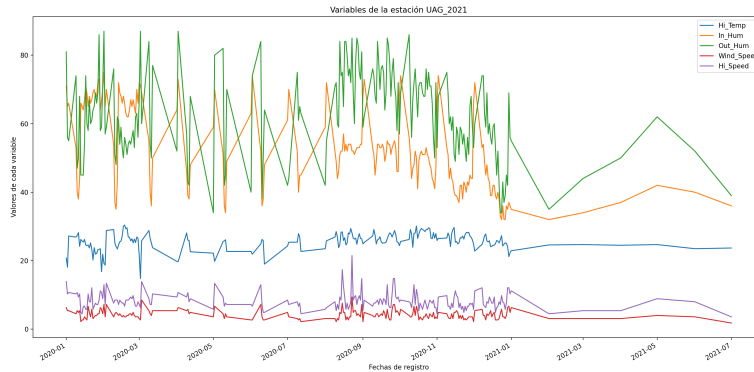


Figure 3: Estación de ejemplo UAG, en esta se observan las variables Temperatura, Velocidad del viento, Humedad, etc, a lo largo del periodo de observación, se observa que después de enero del 2021, existe una carencia de datos.

## 2.3 Estaciones de la CONAGUA

En esta única base de datos se cuenta con información de las estaciones: Acatlán de Juárez, Guadalajara, Tlaquepaque, Zapopan y Tonalá.

En este caso, se trabajó para asegurar que la estructura de la base de datos fuera similar a la anterior (donde se tenían datos desde 1970, con fechas en columnas y variables repetidas en filas). Se seleccionaron las variables meteorológicas relevantes y se resolvieron ciertos problemas adicionales para ajustar la base de datos.

El objetivo fue estandarizar la estructura y el contenido de la base de datos, de modo que fuera consistente con las anteriores y se pudiera realizar un análisis conjunto de los datos meteorológicos de todas las estaciones.

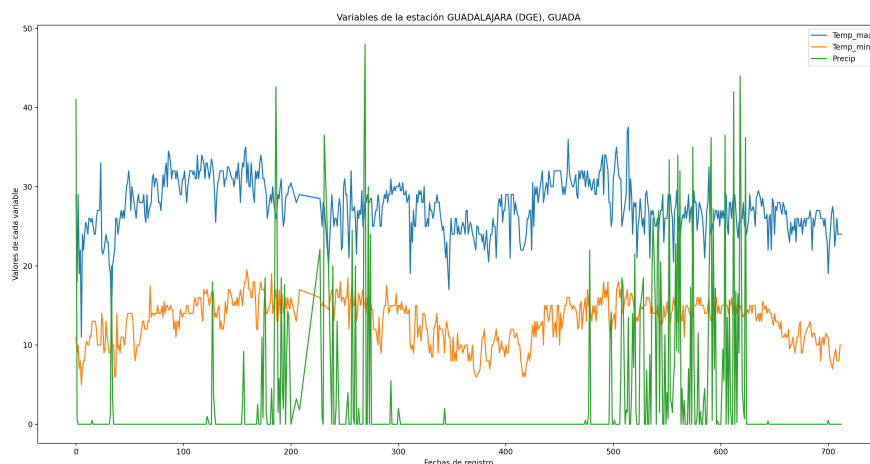


Figure 4: Estación de ejemplo Guadalajara, se observa como fueron variando a lo largo del tiempo, en este caso las fechas están datadas por días a partir del inicio del 2020 al final del 2021.

Para ambos se generó una celda que permite introducir la estación y las variables para observar su cambio a lo largo del tiempo.

Para ambos casos, el código que se generó está en el siguiente link de [google colab](#).

## 2.4 Unión de bases de datos, gráficos y estadísticos

Gracias a la integración de otros programas en Python, pudimos recuperar las bases de datos más relevantes de los códigos anteriores. El objetivo era unificarlas y observar posibles relaciones entre las variables.

Se realizaron filtros para seleccionar los valores correspondientes al periodo de estudio, enfrentándonos a desafíos como datos incompletos o formatos inadecuados en varias estaciones. Una vez que todas las bases de datos estuvieron en un formato compartido, se unieron utilizando variables, fechas y casos de COVID-19.

Además, se escalaron nuevamente los datos de contagios y las variables para visualizar su comportamiento de manera gráfica y comparar similitudes.

Finalmente, se desarrolló un código interactivo que permite observar las relaciones entre variables y contagios. Se generó un ciclo para crear todas las gráficas y se guardaron como imágenes.

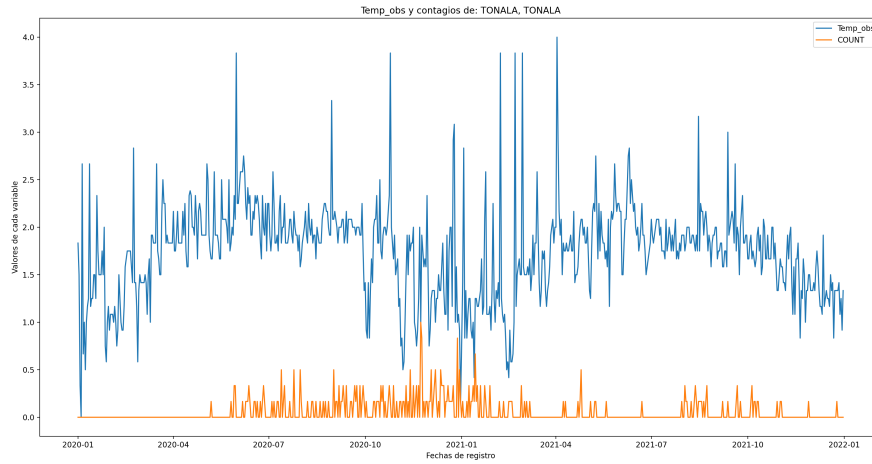


Figure 5: Estación de ejemplo Tonalá, Aquí se observa de mejor manera el comportamiento de los contagios en esta estación y la variable de la velocidad del viento promedio para cada día, de esta manera se tiene para todas las estaciones y todas las variables.

## 2.5 Estadísticos

Finalmente, se creó una matriz con estadísticas que relacionan las variables y los casos de COVID. Se generó un modelo de prueba utilizando un enfoque lineal entre las variables y los casos de COVID. A partir de este modelo, se obtuvieron importantes estadísticos como la Varianza, R, R cuadrado, MSE, entre otros, así como los índices de correlación que son independientes del modelo lineal generado.

La matriz resultante es la siguiente:

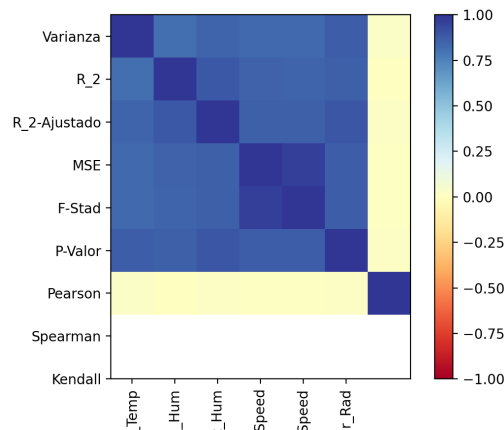


Figure 6: Matriz con estadísticos y correlaciones. Se observa que los índices de correlaciones prácticamente son nulos, de igual manera en el código se pueden observar los datos numéricos

De igual manera, se tiene la combinación de los contagios las variables, además de los estadísticos en el siguiente link de [google colab](#).

### 3 Aplicación a la estación IAM

Debido a los resultados insatisfactorios obtenidos para las estaciones en general, y considerando la gran cantidad de datos faltantes en algunas estaciones, se tomó la decisión de realizar un estudio enfocado exclusivamente en la estación "IAM". El objetivo de este estudio fue analizar el comportamiento de las variables y los contagios a lo largo de cada mes.

En este estudio, se calcularon la varianza y las correlaciones Kendall, Spearman y Pearson para cada mes. Además, se generaron gráficas con un formato más específico, brindando una visualización más clara de los resultados.

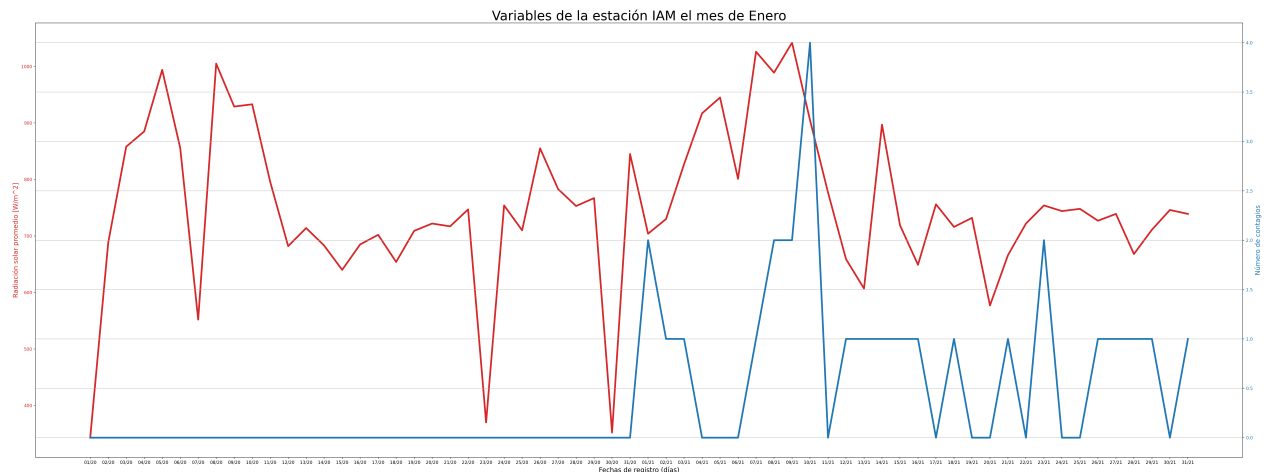


Figure 7: En la gráfica del ejemplo se observa el mes de enero la variable de la radiación IAM y el número de contagios, cada una está escalada en la gráfica para que sea bien observada, además de que está separada en el eje x por cada día. Este mismo formato se genera para cada variable para todos los meses del año 2020.

El código comentado de la estación sola de IAM, así como la generación de todos los datos, en el siguiente link de [google colab](#). Además, el valor de las correlaciones para cada mes se encuentra en el siguiente [link](#).

### 4 Resultados y conclusiones

En ambos casos, ya sea el de todas las estaciones y el del caso particular del IAM se observó una correlación prácticamente nula con todas las variables, cosa que generó un desinterés en el problema y dió por cerrada esta investigación.