**Reto | Predicción de Contingencias Ambientales en el Área Metropolitana de la Ciudad de Monterrey**

**Indicaciones:**

* Para la entrega de esta actividad, subirás el enlace a tu libreta de Google Colab en la sección correspondiente en Canvas. Para obtenerlo, ubícate en la libreta desde tu *drive* y presiona *Get link* desde el menú contextual. No olvides darle permisos al archivo para poder evaluarlo (Opción: *Anyone with the link*).
* Sube también una versión en PDF de tu cuaderno al Canvas. Lo llamarás DS\_C7\_SC1\_NOMBRE (sin espacios).
* Tu reporte será evaluado con base al cumplimiento de los requerimientos y a su contenido, pero también por su presentación, por lo que **errores ortográficos o de redacción serán penalizados**. La fecha límite de envío del enlace de tu libreta estará establecida en la plataforma Canvas.
* El archivo de datos “Monterrey Pollution Data 2.csv” contiene las lecturas de una estación de monitoreo de calidad ambiental durante el año 2015 situada en el centro de la ciudad de Monterrey. Hay una lectura por hora. Cada renglón reporta las siguientes variables (Tabla 1):

|  |  |  |
| --- | --- | --- |
| **Abrev** | **Variable** | **Unidades** |
| Date | Fecha en que se tomó la lectura |  |
| Month | Mes del año |  |
| Day | Dia del mes |  |
| DayWeek | Día de la semana en texto |  |
| Weekday | Día de la semana (Domingo=1) |  |
| Hour | Hora del dia |  |
| CO | Monóxido de Carbono | Ppm |
| NO | Monóxido de Nitrógeno | Ppb |
| NO2 | Dióxido de Nitrógeno | Ppb |
| NOx | Suma de NO y NO2 | Ppb |
| O3 | Ozono | Ppb |
| PM10 | Particulas menores a 10 microns | g/m3 |
| PM2.5 | Particulas menores a 2.5 microns | g/ m3 |
| PRS | Presión | Mmhg |
| RAINF | Lluvia | Mm/hr |
| RH | Humedad Relativa | % |
| SR | Radiación Solar | KW/m2 |
| TOUT | Temperatura | DegC |
| WSR | Velocidad del Viento | Km/hr |
| WDV | Dirección del viento | Deg |

Tabla 1. Variables de contaminación, tiempo y clima reportados por la estación de monitoreo

El gobierno del estado de Nuevo León está preocupado particularmente por las variables O3 que no debe exceder 120 ppb y por partículas que diámetro menor a 2.5 micrones (PM2.5) que no debe exceder 40.5 g/m3. El resto de los contaminantes son considerados precursores.

1. **Para concretar este proyecto realiza los siguientes pasos:**
2. En algún entorno Spark, carga los datos y elimina las variables innecesarias.
3. Realiza un análisis de correlación y establece qué variables ambientales o de tiempo afectan la concentración de contaminantes O3 y PM2.5. Algunas correlaciones son negativas. Toma en cuenta que algunas correlaciones son positivas y algunas negativas indicando que los contaminantes incrementan y otras bajan.
4. Confirma estas relaciones usando gráficos de dispersión.
5. Crea modelos de regresión para predecir estos contaminantes con la menor cantidad de variables atributos posible. Puedes utilizar cualquier técnica: regresión lineal (o polinomial) multivariable, random forrests, gradient boost, o cualquier otra técnica que hayas investigado que funcione en Spark.
6. **Evalúa los modelos y escribe la reflexión de tu evaluación.**

|  |
| --- |
|  |

1. **Conclusiones, contesta lo siguiente y justifica tus respuesta:**
2. ¿Puedes decir que la contaminación por O3 o PM2.5 está ligada al tráfico vehicular?

|  |
| --- |
|  |

1. ¿Consideras que tendría efectos sobre la contaminación implantar un esquema de verificación vehicular?

|  |
| --- |
|  |

1. Con un reporte del pronóstico del clima dado en la mañana, ¿puedes predecir que habrá una contingencia ambiental debido a que los contaminantes en el aire rebasaron los límites permitidos por la norma? ¿Por qué razón (es)?

|  |
| --- |
|  |

1. En tus propias palabras, ¿cuáles consideras qué son las condiciones climáticas se deben cumplir para tener altos niveles de contaminación de O3? ¿Y PM2.5?

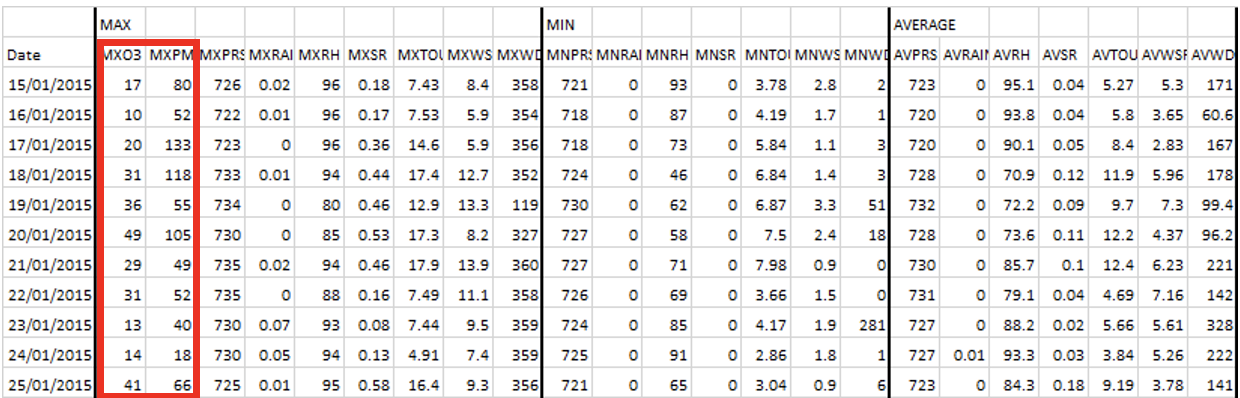
|  |
| --- |
|  |

**¡Has cumplido con el Reto!**

Si deseas reforzar tus conocimientos, puedes seguir el procedimiento siguiente:

Encontrarás que crear los modelos de predicción no es fácil. Para mejorar los niveles de predicción, toma en cuenta lo siguiente:

* Si los niveles de los contaminantes son bajos, no es relevante. Lo único que importa es cuando los niveles son altos.
* Modifica la tabla agrupando los datos por día y obteniendo el valor máximo para O3 y PM2.5 en cada día.
* Ahora obtén el mínimo, el promedio y el máximo de cada variable del clima, creando una tabla como la que sigue:



Ahora intenta crear nuevos modelos de regresión que, dadas los valores de variables del clima máximo, promedio y mínimas, predigan los valores máximos de los contaminantes y reflexiona:

¿Mejoraron los parámetros de desempeño de los modelos?

|  |
| --- |
|  |