

Semana Tec 10025

Actividad 3 Mapas de calor y boxplots

12/01/2021



Herramientas computacionales: El arte de la analítica

Alison Magie Yáñez Dávila A01423011

Gilberto Huesca Juárez

Campus Cuernavaca

Instrucciones

En esta actividad, trabajarás con el conjunto de datos que seleccionaste. Tienes que replicar los pasos vistos durante la clase.

Realiza un código que haga lo siguiente:

1. Cargar los datos.
2. Selecciones dos columnas que al momento parezcan interesantes. Para éstas:
 1. Desplegar su histograma.
 2. Desplegar el diagrama de cajas y bigotes.
 3. Si hay outliers, quítalos y vuelve a desplegar el diagrama con los nuevos datos.
3. Despliega el mapa de calor de las correlaciones entre todas las variables numéricas. Escoge la visualización más adecuada y que aporte la mayor información posible para el análisis.

Realiza un pequeño reporte que muestre lo siguiente:

1. Describa el significado de las variables que seleccionaste.
2. Muestre su histograma y su diagrama de cajas y bigotes.
3. Muestre el mapa de calor de las correlaciones entre todas las variables numéricas.

Puedes agregar estos elementos al reporte anterior para ir mostrando el proceso completo.

Introducción

Los ríos son corrientes de agua que fluyen desde su nacimiento hacia otro río, lago o el mar. Los ríos son considerados una de las formaciones geográficas más importantes de la tierra, ya que son elementos vitales para todos los seres vivos no sólo por suministrar agua a ciudades o granjas, si no que también por que nos proporcionan alimentos e incluso nos permiten entretenernos y navegar, ya que se consideran poderosos agentes de transportación.

Sin embargo, hoy en día la contaminación de los ríos representa un gran problema a nivel mundial no sólo porque abre paso a enfermedades, sino que también porque la alteración de estos presenta un impacto en el mar, zonas costeras u otros ríos. Es por lo anterior que se han realizado estudios minuciosos sobre la calidad del agua.

A continuación, en el siguiente proyecto se mostrará un análisis estadístico sobre la calidad del agua en diferentes ríos de India, donde la investigación se basará en el promedio de los valores de calidad del agua obtenidos de algunos de los ríos de India.

Variables

- *Station code*

Variable que representa un código único para cada lugar.

El tipo de variable es Entero.

- *Locations*

Localizaciones de los ríos. Nombre de los ríos y dónde se encuentran ubicados.

El tipo de variable es String.

- *State*

Estado en donde corre el río. Los estados que se encuentran son Maharashtra, Bihar, Uttar Pradesh, Karnataka y Adhra pradesh.

El tipo de variable es String.

- *Temp*

Temperatura promedio del río. El valor mínimo registrado es 0 mientras que el máximo es de 33.8°.

El tipo de variable es Flotante.

- *DO*

Valor promedio del oxígeno disuelto. Esta variable nos permite dar una idea sobre la contaminación en el agua, debido a que un contenido en oxígeno significativamente menor no permite el desarrollo de especies acuáticas.

El valor mínimo registrado es 0.2mg/l mientras que el máximo es 16.3mg/l.

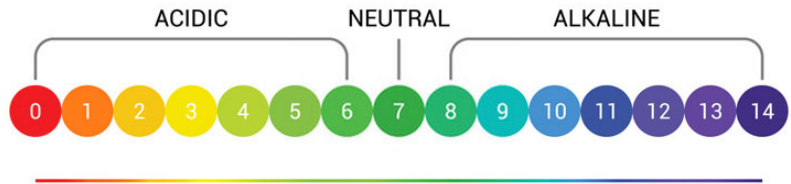
El tipo de variable es Flotante.

[OD] mg/L	Condición	Consecuencias
0	Anoxia	Muerte masiva de organismos aerobios
0-5	Hipoxia	Desaparición de organismos y especies sensibles
5-8	Aceptable	OD] adecuadas para la vida de la gran mayoría de especies de peces y otros organismos acuáticos.
8-12	Buena	
>12	Sobresaturada	Sistemas en plena producción fotosintética.

- *pH*

Valor promedio del ph. El valor mínimo registrado es 6.3 mientras que el máximo es 14.7.

El tipo de variable es Flotante.



- Conductivity

Valor promedio de la conductividad eléctrica. Esta variable nos permite saber si las aguas superficiales son potables. El valor mínimo registrado es 39 μ S/cm, mientras que el máximo es 24062 μ S/cm.

El tipo de variable es Entero.

- *BoD*

Valor promedio de la demanda bioquímica de oxígeno. Este parámetro nos permite determinar el estado/ calidad del agua de los ríos.

El valor mínimo es 0.2 mg/l mientras que el máximo es 75.6 mg/l.

El tipo de variable es Flotante.

Demanda bioquímica de oxígeno (mg/L)

- Excelente ≤ 3 mg/L
- Buena calidad >3 y ≤ 6 mg/L
- Aceptable >6 y ≤ 30 mg/L
- Contaminada >30 y ≤ 120 mg/L
- Fuertemente contaminada >120 mg/L

- *Nitrate_N_Nitrite_N*

Valor promedio de nitrato-n y nitrito-n.

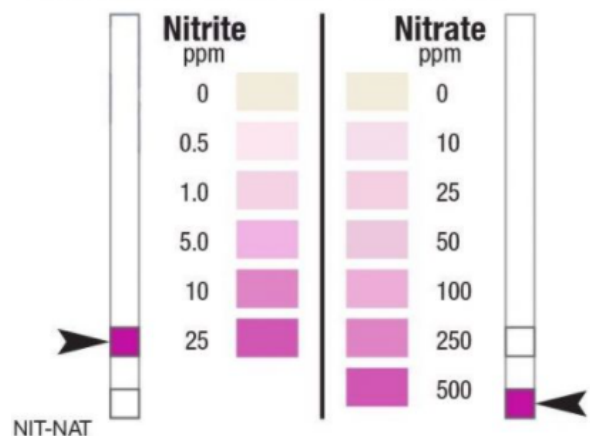
Los nitratos y nitritos en aguas naturales también nos ayudan a determinar el estado/calidad del agua.

Los niveles de nitrito superiores a 0,75 ppm pueden provocar estrés en peces, mayores a 5 ppm se considera tóxico para los seres vivos.

Por otra parte los niveles de nitrato entre 0 y 40 ppm son seguros, mientras que superior a 80 se considera tóxico.

Los rangos registrados en la dataset varían entre 0 ppm y 45.5ppm.

El tipo de variable es Flotante.



- *Fecal_Coliform*

Valor promedio de coliformes fecales, es decir, de un subgrupo de bacterias de coliformes. Su presencia indica que el agua está contaminada por excrementos o desechos de alcantarillas.

Los rangos registrados en la dataset varían entre 0 y 310417.

El tipo de variable es Flotante.

- *Total_Coliform*

Valor medio de coliformes. Estas son bacterias que usualmente son encontradas en el medio ambiente.

Los rangos registrados en la dataset varían entre 1 y 23816667.

El tipo de variable es Flotante.

Valoración	Agua	
	Coliformes fecales (NMP/100ml)	Coliformes totales (NMP/100ml)
No contaminado	0%-20% > 200	0%-20% > 1000
Contaminación Media	41% - 60 % > 200	41% - 60 % > 1000
Contaminación Alta	61% - 100 % > 200	61% - 100 % > 1000

Tabla 1. Valoración conceptual indicativa del grado de contaminación para los contaminantes microbiológicos. (Tomado y modificado de Marín, Garay *et al.*, 2001).

Análisis

Para el desarrollo del análisis decidí hacer uso de las variables BOD y Nitrate_N_Nitrite_N debido a que se encuentran directamente relacionadas con la calidad del agua, ya que la presencia o la falta de presencia de estos elementos en los ríos representan el grado de contaminación en el agua. Además de que los valores obtenidos de las variables deberían encontrarse relacionadas.

➤ *Demanda Bioquímica de Oxígeno (BOD)*

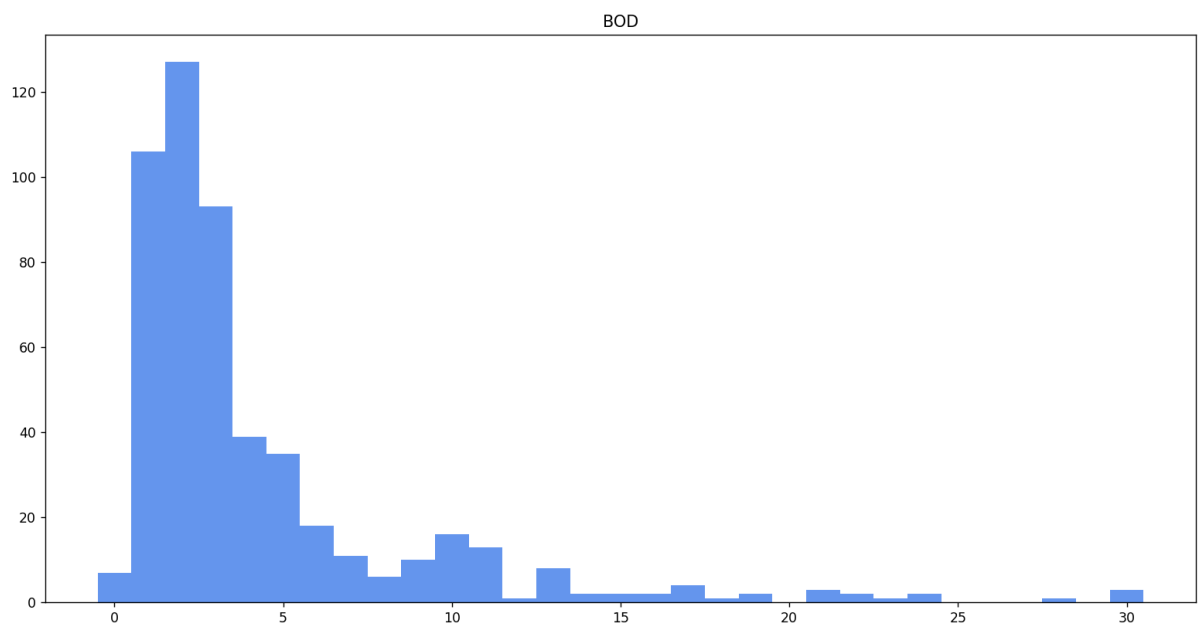
El BOD es la cantidad de oxígeno que los microorganismos, hongos y plancton consumen durante la degradación de sustancias orgánicas.

Mientras más altos se encuentren los valores, más contaminado se encontrará el río.

- **Histograma**

En el siguiente histograma se puede observar que la frecuencia de los datos recayó principalmente en un rango de 0 mg/L a 5 mg/L. Lo que quiere decir que el promedio de los datos obtenidos de los ríos de India indica que estos se encuentran en buen estado.

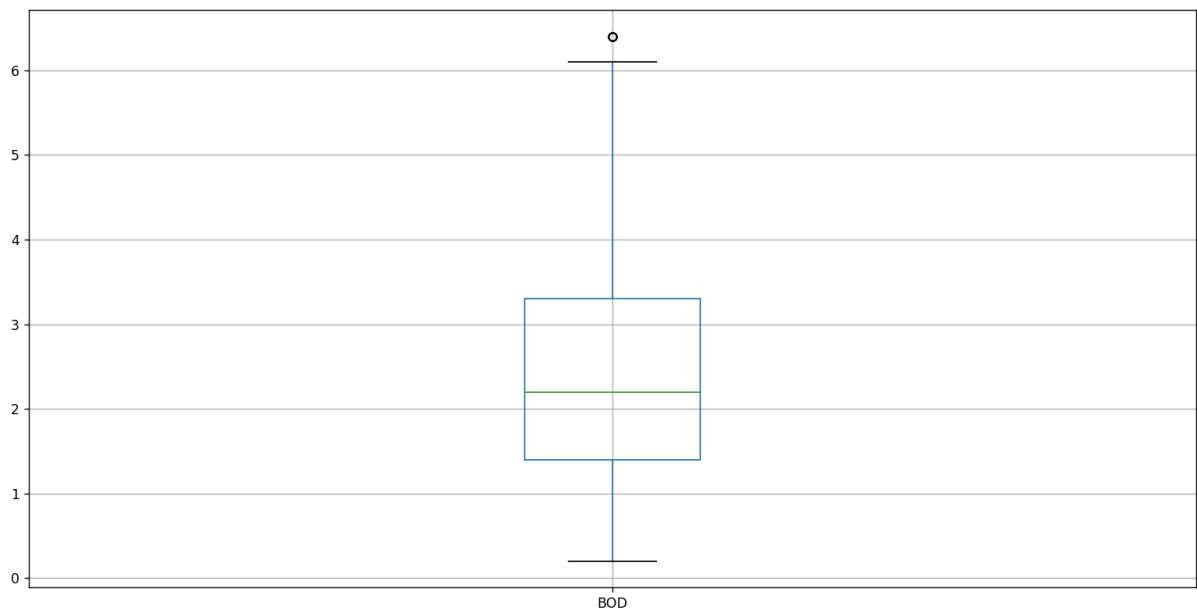
Por otra parte, existe una pequeña pequeña cantidad de BOD que se encuentra cercano a 30 mg/L, lo que quiere decir que hay una mínima cantidad de contaminación.



- **Diagrama de cajas y bigotes**

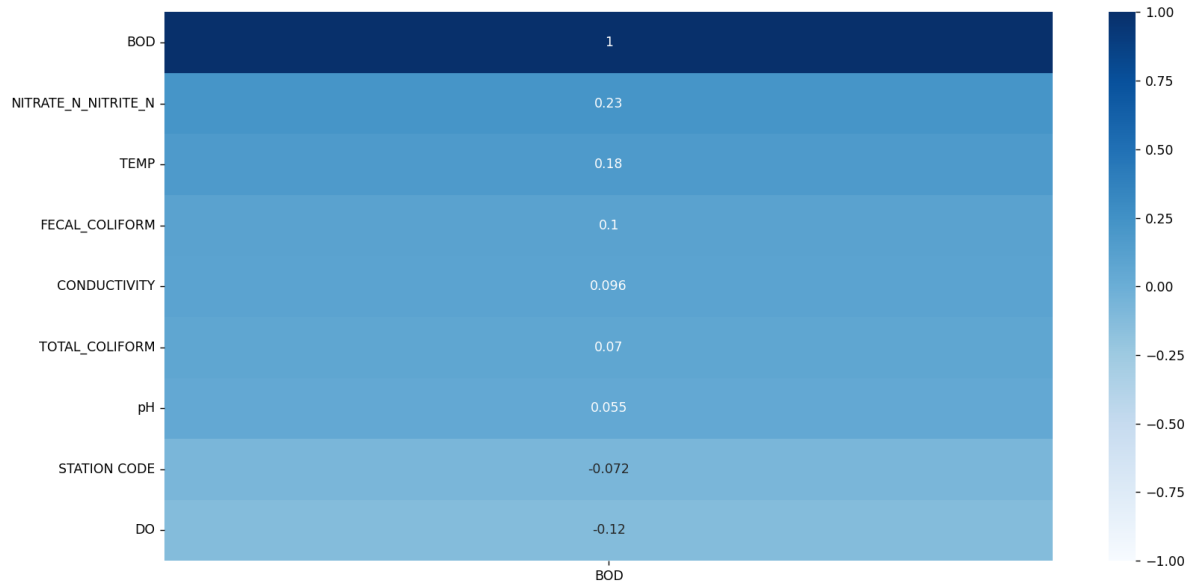
En el siguiente diagrama se puede observar los siguientes resultados:

- Valor mínimo: 0.2
- Valor Máximo: 6.09
- *Mediana*: Al tener un valor de 2.2, podemos decir que existe una distribución asimétrica positiva, esto debido a que se acerca al Q1. Quiere decir que el 50% central de la variable oscila entre los valores de 0.2 y 6.09.
- Rango : La separación de los valores corresponde a 5.89.
- Rango Intercuartil: La dispersión dentro de la caja corresponde a 2. Quiere decir que existe una variabilidad de los datos.
- Q1 (25%): 1.3
- Q2 (50%): 2.2
- Q3 (75%): 3.3



- **Mapa de calor**

En el siguiente mapa de calor podemos observar que el parámetro de BOD tiene una mayor relación con el parámetro de Nitrato y Nitrito. Por otra parte, el parámetro que se encuentra menos relacionado es DO.



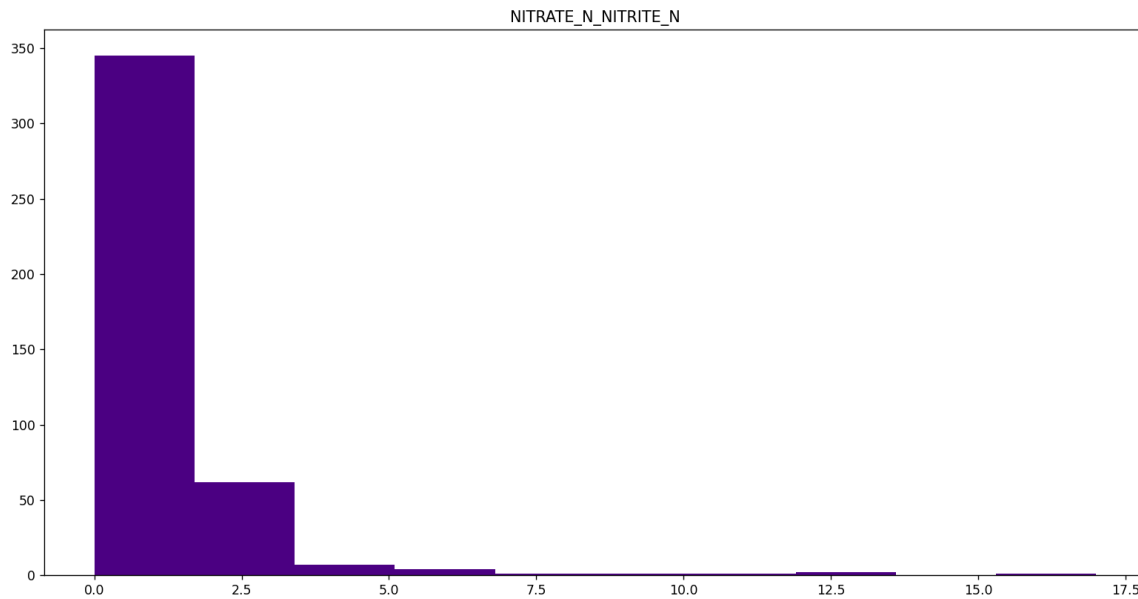
➤ Nitrato y Nitrito (Nitrate_N_Nitrite_N)

El Nitrato y nitrito son compuestos solubles que contienen nitrógeno y oxígeno. El estándar del nitrato es 10 ppm sin embargo hasta 40 ppm es considerado seguro, por otra parte el estándar del nitrito es 1 ppm.

Un alto contenido de nitrato y nitrito pueden causar problemas de salud crónicos como enfermedades del corazón o pulmones.

○ **Histograma**

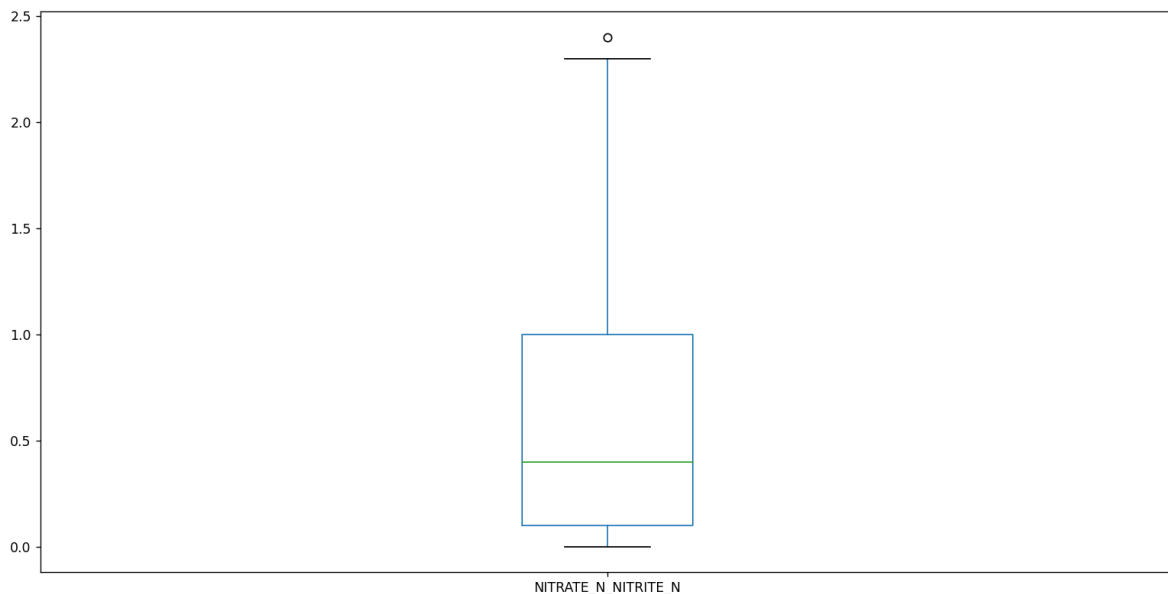
En el siguiente histograma se puede observar que la frecuencia de los datos recayó principalmente en un rango de 0 ppm a 2.5 ppm. Lo que quiere decir que el promedio de los datos obtenidos de los ríos de India indica que estos se encuentran en un estado aceptable.



- **Diagrama de cajas y bigotes**

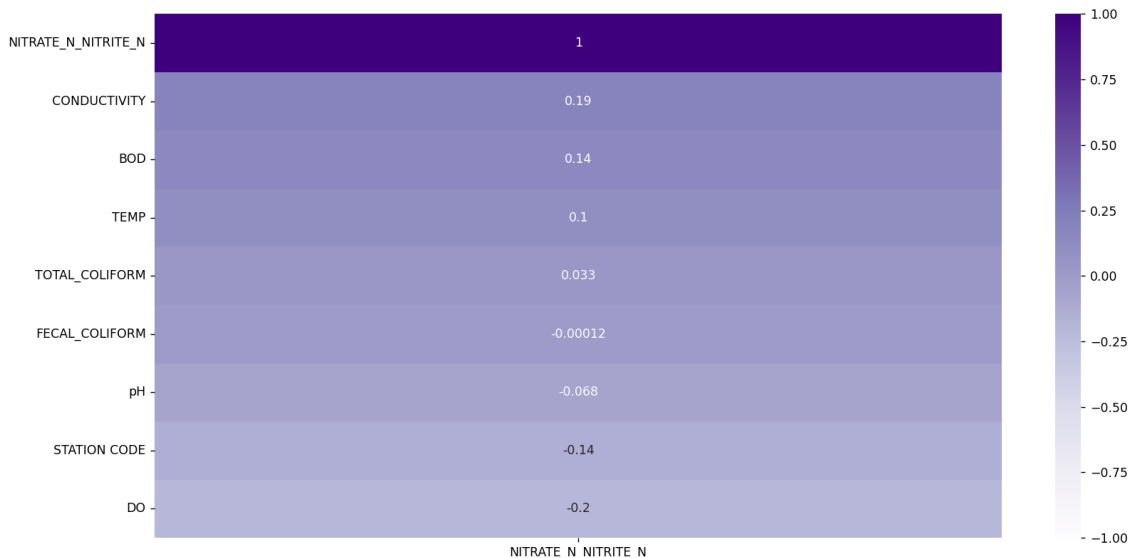
En el siguiente diagrama se puede observar los siguientes resultados:

- Valor mínimo: 0.023
- Valor Máximo: 2.203
- *Mediana*: Al tener un valor de 0.0406, podemos decir que existe una distribución asimétrica positiva, esto debido a que se acerca al Q1. Quiere decir que el 50% central de la variable oscila entre los valores de 0.0406 y 2.203.
- Rango : La separación de los valores corresponde a 2.18. Quiere decir que existe una variabilidad de los datos. Quiere decir que la variabilidad.
- Rango Intercuartil: La dispersión dentro de la caja corresponde a 0.914.
- Q1 (25%): 0.097
- Q2 (50%): 0.0406
- Q3 (75%): 1.011



- **Mapa de calor**

En el siguiente mapa de calor podemos observar que el parámetro de Nitrate_N_Nitrite_N tiene una mayor relación con el parámetro de conductividad. Por otra parte, el parámetro que se encuentra menos relacionado es DO.



Conclusiones

Tal como se observaron en los resultados anteriores las variables de BOD y Nitrate_N_Nitrite_N se encuentran relacionados debido a que se muestran valores que se encuentran de los rangos seguros, además de que las correlaciones que existen entre BOD y Nitrate_N_Nitrite_N en su gran mayoría concuerda con los parámetros más importantes para definir si un río se encuentra contaminado o no. Por ejemplo, en la variable BOD encontramos que una de sus correlaciones es Fecal_Coliform, lo cual tiene sentido porque existe una relación directamente exponencial entre estas debido a los compuestos químicos entre estos.

En cuanto al análisis del estado de los ríos, podemos concluir que se encuentran en estados medianamente aceptables, ya que existen valores en los límites o fuera de los rangos de seguridad en los datos.

Referencias

- Utcars, A. (2020, 29 mayo). *Water Quality Data*. Kaggle. Recuperado 12 de enero de 2022, de <https://www.kaggle.com/utcarshagrawal/water-quality-data/version/1>
- Río - NILSA. (s. f.). nilsa. Recuperado 12 de enero de 2022, de <https://www.nilsa.com/es/como-educamos/oferta-educativa/rio/>
- Valdivielso, A. (2020, 19 octubre). *¿Qué es un río?* iAgua. Recuperado 12 de enero de 2022, de <https://www.iagua.es/respuestas/que-es-rio>
- *análisis_aguas*. (s. f.). . Recuperado 12 de enero de 2022, de http://ficus.pntic.mec.es/ngom0007/analisis_aguas.html
- Andreo, M. (s. f.). *Demanda Biológica de Oxígeno (D.B.O.)*. . Recuperado 12 de enero de 2022, de [https://www.mendoza.conicet.gov.ar/porta/enciclopedia/terminos/DBO.htm#:~:text=Se%20expresa%20en%20mg%20%2F%20l,microorganismos%20para%20oxidarla%20\(degradarla\).](https://www.mendoza.conicet.gov.ar/porta/enciclopedia/terminos/DBO.htm#:~:text=Se%20expresa%20en%20mg%20%2F%20l,microorganismos%20para%20oxidarla%20(degradarla).)
- N. (s. f.). *DBO y DQO para caracterizar aguas residuales*. Nihon Kasetsu Europe | Monitoring & Water Clarification. Recuperado 12 de enero de 2022, de <https://nihonkasetsu.com/es/dbo-y-dqo-para-caracterizar-aguas-residuales/>
- *El medio ambiente en México*. (2014). semarnat.gob. Recuperado 12 de enero de 2022, de https://apps1.semarnat.gob.mx:8443/dgeia/informe_resumen14/06_agua/6_2_1.htm
- *nitratos y nitritos-lenntech*. (s. f.). Lenntech. Recuperado 12 de enero de 2022, de <https://www.lenntech.es/nitratos-y-nitritos.htm>
- Dirección de Salud Pública de Carolina del Norte. (s. f.). https://epi.dph.ncdhhs.gov/oe/docs/Las_Bacterias_Coliformes_WellWaterFactSt.pdf. Recuperado 12 de enero de 2022, de https://epi.dph.ncdhhs.gov/oe/docs/Las_Bacterias_Coliformes_WellWaterFactSt.pdf
- *Oxígeno Disuelto*. (s. f.). . Recuperado 12 de enero de 2022, de <https://www.ucm.es/data/cont/docs/952-2015-02-14-Oxigeno%20disuelto%20f.pdf>
- *Nitrato y Nitrito*. (s. f.). . Recuperado 12 de enero de 2022, de http://region8water.colostate.edu/PDFs/we_espanol/Nitrate%202012-11-15-SP.pdf