Water Potability

El acceso al agua potable segura es fundamental para la salud, un derecho humano básico y un componente de políticas efectivas de protección de la salud.

*Descripción de las columnas del dataset*

1. **Valor de pH**:
   * El pH es un parámetro importante para evaluar el equilibrio ácido-base del agua. También es un indicador de la condición ácida o alcalina del agua. La Organización Mundial de la Salud (OMS) ha recomendado un límite máximo permisible de pH entre 6.5 y 8.5. Los valores de investigación actuales (6.52–6.83) están dentro del rango de las normas de la OMS.
2. **Dureza**:
   * La dureza se debe principalmente a sales de calcio y magnesio. Estas sales se disuelven de depósitos geológicos a través de los cuales el agua viaja. El tiempo que el agua está en contacto con material productor de dureza determina la cantidad de dureza presente en el agua cruda. Originalmente, la dureza se definió como la capacidad del agua para precipitar jabón debido al calcio y al magnesio.
3. **Sólidos disueltos totales (TDS)**:
   * El agua tiene la capacidad de disolver una amplia gama de minerales inorgánicos y algunos compuestos orgánicos, como potasio, calcio, sodio, bicarbonatos, cloruros, magnesio, sulfatos, etc. Estos minerales pueden afectar el sabor y la apariencia del agua. El valor deseable para TDS es de 500 mg/l, y el límite máximo es de 1000 mg/l para agua potable.
4. **Cloraminas**:
   * Las cloraminas son desinfectantes comunes utilizados en sistemas de agua pública. Se forman principalmente cuando se agrega amoníaco al cloro para tratar el agua potable. Los niveles de cloro de hasta 4 miligramos por litro (mg/L o 4 partes por millón) se consideran seguros en el agua potable.
5. **Sulfato**:
   * Los sulfatos son sustancias naturalmente presentes en minerales, suelos y rocas. Se encuentran en el aire ambiente, el agua subterránea, las plantas y los alimentos. La concentración de sulfato en el agua de mar es de aproximadamente 2700 mg/L. En la mayoría de las fuentes de agua dulce, los niveles de sulfato varían entre 3 y 30 mg/L, aunque en algunas ubicaciones geográficas se encuentran concentraciones mucho más altas (1000 mg/L).
6. **Conductividad**:
   * El agua pura no es un buen conductor de corriente eléctrica; más bien, es un buen aislante. El aumento en la concentración de iones mejora la conductividad eléctrica del agua. La conductividad eléctrica (EC) mide el proceso iónico de una solución que permite la transmisión de corriente. Según las normas de la OMS, el valor de EC no debe superar los 400 μS/cm.
7. **Carbono orgánico total (TOC)**:
   * El TOC en las aguas fuente proviene de la materia orgánica natural en descomposición (NOM) y de fuentes sintéticas. El TOC es una medida de la cantidad total de carbono en compuestos orgánicos en agua pura. Según la Agencia de Protección Ambiental de EE. UU., el valor de TOC en agua tratada o potable debe ser inferior a 2 mg/L, y en agua fuente (antes del tratamiento) debe ser inferior a 4 mg/L.
8. **Trihalometanos (THM)**:
   * Los THM son productos químicos que pueden encontrarse en el agua tratada con cloro. La concentración de THM en el agua potable varía según el nivel de material orgánico en el agua, la cantidad de cloro necesario para tratarla y la temperatura del agua tratada. Niveles de THM de hasta 80 ppm se consideran seguros en el agua potable.
9. **Turbidez**:
   * La turbidez del agua depende de la cantidad de materia sólida presente en estado suspendido. Es una medida de las propiedades de emisión de luz del agua y se utiliza para indicar la calidad de la descarga de residuos en relación con la materia coloidal. El valor promedio de turbidez obtenido para el Campus Wondo Genet (0.98 NTU) está por debajo del valor recomendado por la OMS de 5.00 NTU.
10. **Potabilidad**:
    * Indica si el agua es segura para el consumo humano, donde 1 significa “Potable” y 0 significa “No potable”.

Análisis del dataset (EDA)

En el dataset formado por las columnas descriptas anteriormente, buscaremos predecir si el agua es potable o no potable. Para ello seguiremos los siguientes pasos.

* ¿Como detectar y reemplazar valores nulos?
* ¿análisis de correlación de cada variable?
* ¿Del dataset, que cantidad de valores son potables y no potable? ¿En porcentajes?
* ¿Determinar los valores atípicos y el rango de confianza que tenemos en el dataset?
* ¿Comportamiento de cada variable para agua potable y no potable?