

Trabajo Práctico Especial

Calidad del Agua del Río de la Plata



Integrantes:

brizzali@alumnos.exa.unicen.edu.ar

milopez@alumnos.exa.unicen.edu.ar

marriaga@alumnos.exa.unicen.edu.ar

Introducción

En el presente informe se desarrollará un análisis sobre la calidad del agua del Río de la Plata a partir del dataset proporcionado por la cátedra de Fundamentos de la Ciencia de Datos. En el mismo se volcaron los resultados de las muestras tomadas a lo largo del año 2023 en distintas zonas del río.

El objetivo de este informe es, a través del análisis de las muestras, determinar qué factores influyen en la contaminación.

Materiales

Dentro del [dataset](#), se encuentra el conjunto de datos que incluye 169 mediciones en diferentes sitios del Río de La Plata y en diferentes épocas del año.

Descripción de las variables

1. **sitios:** Localización específica donde se realizó el muestreo del agua.
2. **código:** Identificador único para cada muestra o estación de muestreo.
3. **fecha:** Fecha en la que se tomó la muestra de agua.
4. **año:** Año en que se realizó el muestreo.
5. **campaña:** Nombre o número de la campaña de monitoreo en la que se realizó el muestreo.
6. **tem agua:** Temperatura del agua en grados Celsius.
7. **tem aire:** Temperatura del aire en grados Celsius.
8. **od:** Oxígeno disuelto, medido en miligramos por litro (mg/L), esencial para la vida acuática.
9. **ph:** Medida de la acidez o alcalinidad del agua, en una escala de 0 a 14.
10. **olores:** Presencia de olores en el agua, que puede indicar contaminación.
11. **color:** Color del agua, que puede ser un indicador de la calidad del agua.
12. **espumas:** Presencia de espumas en la superficie del agua, que puede ser un signo de contaminación.
13. **mat susp:** Materia suspendida, que se refiere a partículas sólidas que flotan en el agua.
14. **colif fecales ufc 100 ml:** Unidades formadoras de colonias de coliformes fecales en 100 ml de agua, un indicador de contaminación fecal.
15. **escher coli ufc 100 ml:** Unidades formadoras de colonias de Escherichia coli en 100 ml de agua, otro indicador de contaminación fecal.

16. **enteroc ufc 100 ml:** Unidades formadoras de colonias de enterococos en 100 ml de agua, que también indican contaminación fecal.
17. **nitrate mg l:** Concentración de nitratos en miligramos por litro (mg/L), que puede indicar contaminación por fertilizantes.
18. **nh4 mg l:** Concentración de amonio en miligramos por litro (mg/L), que puede ser un indicador de contaminación orgánica.
19. **p total l mg l:** Fósforo total en miligramos por litro (mg/L), que incluye todas las formas de fósforo en el agua.
20. **fosf ortofos mg l:** Concentración de ortofosfatos en miligramos por litro (mg/L), que es un nutriente importante.
21. **dbo mg l:** Demanda biológica de oxígeno en miligramos por litro (mg/L), que mide la cantidad de oxígeno requerido por microorganismos para descomponer materia orgánica.
22. **dco mg l:** Demanda química de oxígeno en miligramos por litro (mg/L), que mide la cantidad total de oxígeno requerido para oxidar materia orgánica e inorgánica.
23. **turbiedad ntu:** Turbidez del agua medida en unidades NTU (Nephelometric Turbidity Units), que indica la claridad del agua.
24. **hidr deriv petr ug l:** Hidrocarburos derivados del petróleo en microgramos por litro (µg/L), que indican contaminación por productos petroleros.
25. **cr total mg l:** Concentración total de cromo en miligramos por litro (mg/L), un metal pesado que puede ser tóxico.
26. **cd total mg l:** Concentración total de cadmio en miligramos por litro (mg/L), otro metal pesado que es tóxico en altas concentraciones.
27. **clorofila a ug l:** Concentración de clorofila a en microgramos por litro (µg/L), que indica la cantidad de fitoplancton en el agua.
28. **microcistina ug l:** Concentración de microcistinas en microgramos por litro (µg/L), que son toxinas producidas por ciertas algas.
29. **ica:** Índice de calidad del agua, que puede ser un valor calculado para evaluar la calidad general del agua.
30. **calidad de agua:** Clasificación general de la calidad del agua basada en los parámetros medidos

Métodos y resultados

Análisis exploratorio de los datos

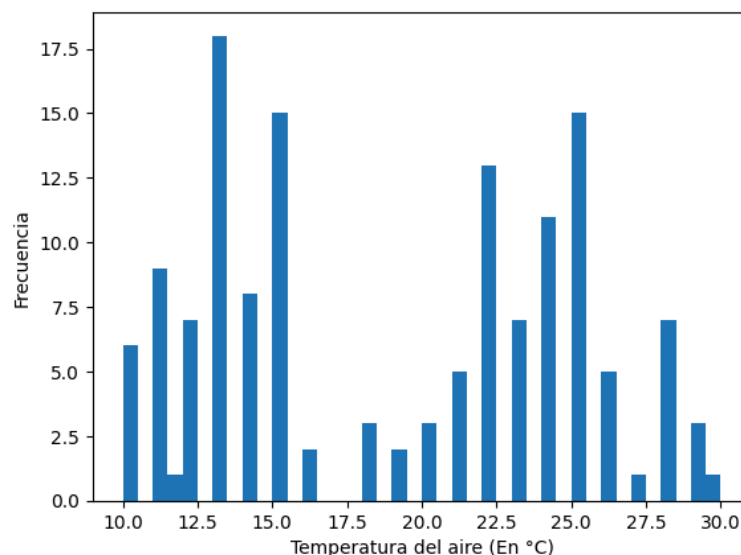
Realizamos un análisis univariado para reconocer el dominio del dataset

Sitios: Esta variable es cualitativa nominal y no tiene ningún valor nulo.

Campaña: Es una variable cualitativa nominal, podemos observar que las muestras fueron tomadas de manera equitativa, es decir, que hay igual cantidad de muestras en cada estación del año.

Temperatura del agua: Esta es una variable cuantitativa continua con una distribución normal. Como la media (19.7°C) y la mediana (19°C) son semejantes podemos decir que la temperatura más usual del Río de La Plata es de 19° . Para ser más específicos, luego del análisis resultó que en verano la temperatura media de las muestras tomadas fue de 25.15°C , en invierno de 14.13°C , en otoño aproximadamente 17.89°C y en primavera 21.57°C

Temperatura del aire: Esta variable es cuantitativa continua con una distribución bimodal, es decir que se pueden observar dos grupos claramente definidos. Uno de los grupos varía entre 10°C y 15°C y el otro entre 20°C y 26°C .



Oxígeno disuelto: Es una variable cuantitativa continua que tiene una distribución normal por lo tanto su media y mediana son iguales ($5,9 \text{ mg/L}$). El 21,3% de las muestras tomadas son nulas.

Acidez: Es una variable cuantitativa continua cuya media es de 7,38 y está levemente sesgada a la izquierda, su media y mediana no varían tanto y el 20% de los registros tomados son nulos.

Olores: Esta es una variable binaria cualitativa nominal. En el 82,2% de los casos no hay olor y el 10,7% de los registros son nulos.

Color: Es una variable binaria cualitativa nominal. En el 84% de las muestras carece de color y el 10,7% son registros nulos.

Espuma: Es una variable binaria cualitativa nominal. En el 86,4% no hay presencia de espuma y el 10,7% de las muestras no se registró esta variable.

Materia suspendida: Es una variable cualitativa nominal. En el 61.5% no hay presencia de presencia de partículas sólidas y el 10.7% de las muestras no se registró esta variable.

Colonias de coliformes fecales: Es una variable cuantitativa continua. Está sesgada a la derecha. Su media es de 7380 unidades/100 ml y su mediana es de 1250 unidades/100 ml. El 14,8% de los registros tomados son nulos. Además tiene un outlier de 150000 unidades/100 ml el cual fue tomado en el sitio "Perù Puente".

Escherichia coli: Es una variable cuantitativa continua. Está sesgada a la derecha, por lo tanto su media y mediana varían mucho ($\bar{X} = 5170,7$ y $mediana = 845$) y tiene presencia de un outlier cuyo valor es de 170000 unidades/100 ml y fue tomado en calle 14 y costa, donde hay una salida cloacal. Además tiene 14,8% de nulos.

Colonias de enterococos: Es una variable cuantitativa continua. En el histograma se ve que está sesgada a la derecha, es por ello que su media y mediana varían mucho ($\bar{X} = 1499,8$ y $mediana = 350$) y la kurtosis es de 14,3. Tiene un 14,8% de nulos.

Nitrato: Es una variable cuantitativa continua. Está sesgada a la derecha, por eso su media y mediana varían ($\bar{X} = 7,16$ y $mediana = 5,5$). Tiene un 15,4% de valores nulos y un outlier de 39,4 que fue tomado en "Playa Espigón de Pacheco".

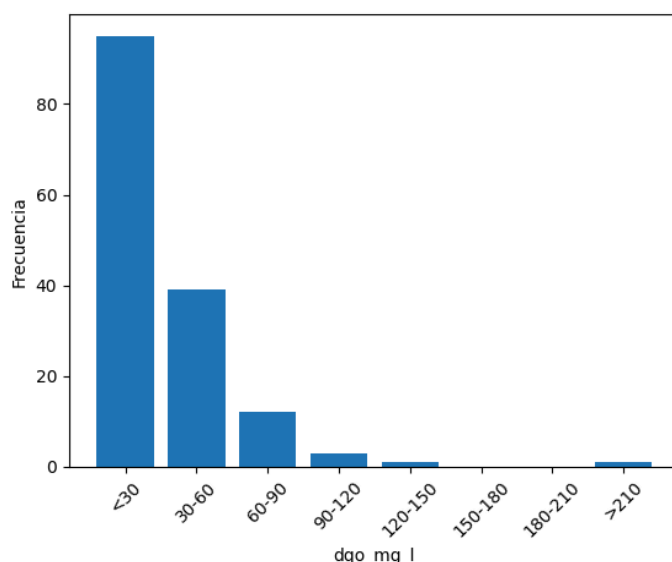
Amonio: Es una variable cuantitativa continua. Está muy sesgada a la derecha con una kurtosis de 8,6, la media vale 1,94 y la mediana 0,77 . Tiene 10,7% de valores nulos.

Fósforo: Es una variable cuantitativa continua. Está sesgada a la derecha, su media es de 0,5 y la mediana 0,35. Tiene el 11,2% de los valores nulos.

Ortofosfatos: Es una variable cuantitativa continua. Está sesgada a derecha, la kurtosis es de 12,5, su media vale 0,26 y la mediana 0,2 a simple vista pareciera que no varían tanto pero el rango que toma la variable va de 0,05 a 1,6. Tiene 13% de valores nulos.

Oxígeno requerido por microorganismos: Es una variable cuantitativa continua. Está sesgada a derecha con una kurtosis de 11. Tiene 11,2% de valores nulos.

Oxígeno requerido para oxidar materia: Es una variable cuantitativa continua que la transformamos en una discreta porque las muestras que se tomaron tenían como valor “<50” y “<30” y otros valores sueltos. Tiene 10,7% de valores nulos.



En el gráfico podemos observar que está sesgada a derecha y que la mayoría de las muestras demandan poco oxígeno.

Turbiedad: Es una variable cuantitativa continua. En el histograma se ve sesgada a la derecha, por eso su media varía de la mediana ($\bar{X} = 35,7$ y $mediana = 25$). Además tiene un outlier que es de 432 NTU en el sitio “Toma de agua Club de Pesca”. Tiene 10,7% de valores nulos.

Hidrocarburos derivados del petróleo: Es una variable cuantitativa continua cuyos valores son todos “<0,10”, estos registros están muy por debajo de la cantidad de hidrocarburos en el agua que se consideran perjudiciales en Argentina (50mg/L). Tiene 11,2% de valores nulos.

Cromo: Es una variable cuantitativa continua. En el histograma se ve que está sesgada a derecha con una kurtosis de 74,78, su media es de 0,089 y su media de 0,0025. Esto se da porque tiene dos outliers (que valen 6 y 7). Tiene 10,7% de valores nulos.

Cadmio: Es una variable cuantitativa continua. Tiene 10,7% de valores nulos, el 65,1% de los valores es 0,001 y el otro 24,3% es 1.

Clorofila: Es una variable cuantitativa continua. Está sesgada a derecha con una kurtosis de 90,89 donde su media vale 22,6 y la mediana es de 5, esto se debe a que tiene un outlier que vale 740,93. Tiene 12,4% de valores nulos.

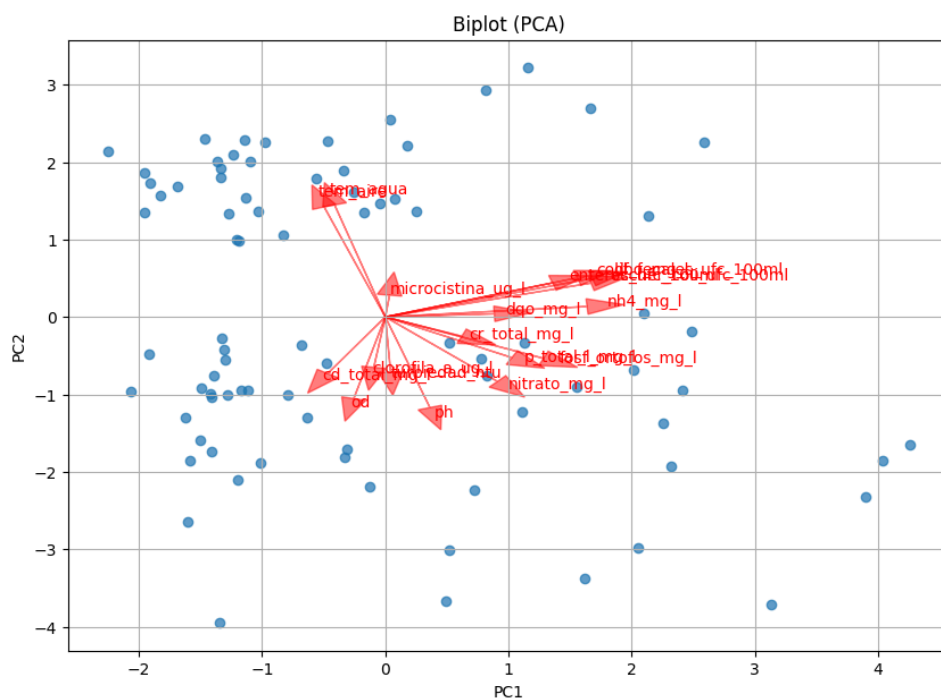
Microcistina: Es una variable cuantitativa continua. Está sesgada a derecha, su media vale 0,64 y la mediana 0,075. Tiene 12,4% de valores nulos.

Índice de calidad del agua: Es una variable cuantitativa discreta con distribución normal, donde su media y mediana no varían tanto ($\bar{X} = 37,2$ y $mediana = 37$), y con una desviación de 5,19 es decir que, en promedio, los valores se desvían 5,19 de la media. Tiene 18,9% de valores nulos.

Calidad del agua: Es una variable cualitativa ordinal que define que tan deteriorada se encuentra el agua. Tiene 18,9% de valores nulos.

Hipótesis

Como punto de partida, realizamos una reducción de dimensionalidad de las variables cuantitativas del dataset, tomando como variable objetivo *ica*. A partir de esto elaboramos un biplot, en el cual se puede observar cuales son las variables que más afectan al índice de calidad de agua. Decidimos que estas serán las variables que estudiaremos a lo largo del informe.

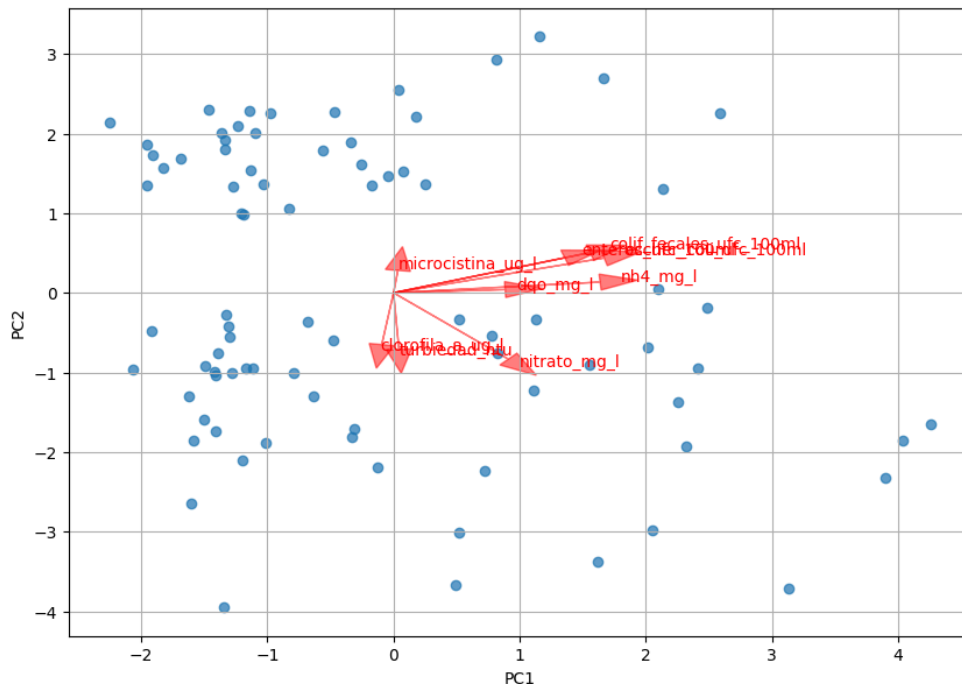


Componentes principales y variables originales representadas en biplot

Para analizar el gráfico debemos prestar atención a las variables que se encuentran más paralelas a los ejes. Vemos que *nh4_mg_l*, *colif_fecales_ufc_100ml*, *enteroc_ufc_100ml*, *escher_coli_ufc_100ml* se encuentran casi paralelas a la primera componente principal, al igual que *dco_mg_l*, aunque esta, al tener menor largo, explica menos varianza. Mirando la segunda componente principal, las variables más paralelas a este eje son *turbiedad_ntu*, *clorofila_a_ug_l*, *microcistina_ug_l*. Podemos observar que las

últimas dos mencionadas forman un ángulo de 180° lo cual puede indicar una correlación inversa. Aunque estas muestran poca longitud, es decir, explican poca varianza, las utilizaremos para el análisis.

Mostramos sólo las variables que más información aportan a las componentes principales, y las que utilizaremos para el análisis para poder centrarnos en ellas:



Componentes principales y variables originales de interés representadas en biplot

Decidimos incluir la variable *nitrato_mg_l*, ya que, aunque no aporta tanta información a las componentes principales, nos pareció interesante para indagar en ella a lo largo del informe.

Por otro lado, viendo que la *turbiedad_ntu* tiene un alto peso sobre la segunda componente, analizamos su correlación con las demás variables haciendo un mapa de calor, a través del cual concluimos que deberíamos analizar la relación entre las variables *clorofila_a_mg_l* y *dco_mg_l* con *turbiedad_ntu*.

Hipótesis 1: Los sitios con mayor concentración de nitrato tienen zonas de cultivos e industrias cercanas.

El nitrato se encuentra en los fertilizantes que utilizan los campos en sus cultivos y en los residuos de las industrias como mataderos y textiles. En Argentina, una concentración mayor a 10 mg por litro de nitrato es considerada perjudicial para la salud del ser humano. Decidimos analizar las muestras con valores más altos de nitrato y ver si fueron tomadas cerca de industrias o cultivos.

Dado que hay treinta y tres muestras con valores de nitrato mayores a 10 mg por litro, nos concentramos en analizar las muestras con concentraciones más significativas.

- Playa Espigon de Pacheco / Muelle Martinez:
Esta muestra con una concentración de **39,4 mg** de nitrato se encuentra a 2,5 km de la Fábrica De Alambre y Tejido y a 3,6 km de la Fabrica de Articulos Domisanitarios - SUR-MER.
- Río Tigre (100 metros antes del Río Luján):
Esta muestra con una concentración de **23,1 mg** de nitrato se encuentra a 1,2 km de la Fábrica Argentina de Hélices S. A. y a 1,7 km de la fábrica Wonderland.
- Playa La Bagliardi:
Esta muestra con una concentración de **21,5 mg** de nitrato tiene una gran zona verde alrededor la cual probablemente contenga cultivos. Además se encuentra a 870 metros de Cloaca General.
- Arroyo El Gato:
Esta muestra con una concentración de **20,7 mg** de nitrato se encuentra a 1,2 km de la fábrica de maquinaria Carpiargento SA, a 1,6 km de la fábrica textil Unicosarg y a 1,7 km de la fábrica de indumentaria RANDOM.INDUMENTARIA.
- Río Luján y Canal San Fernando:
Esta muestra con una concentración de **20,2 mg** de nitrato se encuentra a orillas de la fábrica de Sidra Del Valle.

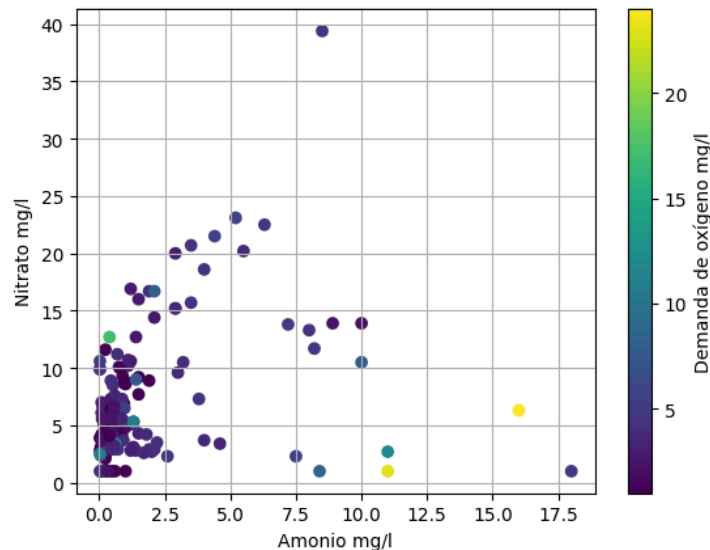
Podemos comparar estas muestras con una realizada en Costanera Hudson Calle 63 la cual tiene una concentración mínima de nitrato (1 mg por litro) y podemos ver que no se encuentra cerca de ninguna fábrica o cultivo. Esta información nos muestra que la contaminación por nitrato parece ser producida por las fábricas y cultivos que se encuentran cerca de las muestras.

Para confirmar esto, se debería hacer un análisis más detallado con un test de hipótesis. Categorizando las variables según su nivel de nitrato y si poseen fábricas cerca, se podría hacer un test de Chi-cuadrado que verifique si existe una dependencia entre las variables 'sitio' y 'nitrato'.

Hipótesis 2: Si el valor de amonio es más alto que el de nitrato, la contaminación orgánica ha sido reciente.

Sería interesante analizar esta hipótesis ya que, investigando el dominio, encontramos que la sustancia química nh_4 (Amonio) se transforma en nitrito y luego en nitrato a través de la oxidación. El amonio puede ser un indicador de contaminación orgánica, es por eso que, midiendo los valores de amonio y nitrato, se puede calcular aproximadamente que tan reciente fue esta contaminación. Es decir que si existen mayores niveles de amonio que de nitrato significa que este aún no se ha oxidado y, por lo tanto la contaminación orgánica ha sido reciente.

Para ver gráficamente qué es lo que sucede con estas variables, plasmamos las muestras en el siguiente scatter plot y las coloreamos con la variable `dbo_mg_l`, que mide la cantidad de oxígeno requerido por microorganismos para descomponer materia orgánica, lo que significa que a mayor demanda de oxígeno, mayor contaminación de materia orgánica.



Relación entre nitrato y amonio con demanda biológica de oxígeno

Podemos ver en el gráfico que existen dos muestras con altos valores de Amonio y bajos de Nitrato, que contienen valores opuestos de demanda de oxígeno. Se podría analizar individualmente estas dos muestras para poder concluir que variables afectan la gran diferencia de demanda de oxígeno.

En general, no se puede ver una clara relación entre estas tres variables y para poder confirmar que la contaminación ha sido reciente se podría realizar algún test de hipótesis.

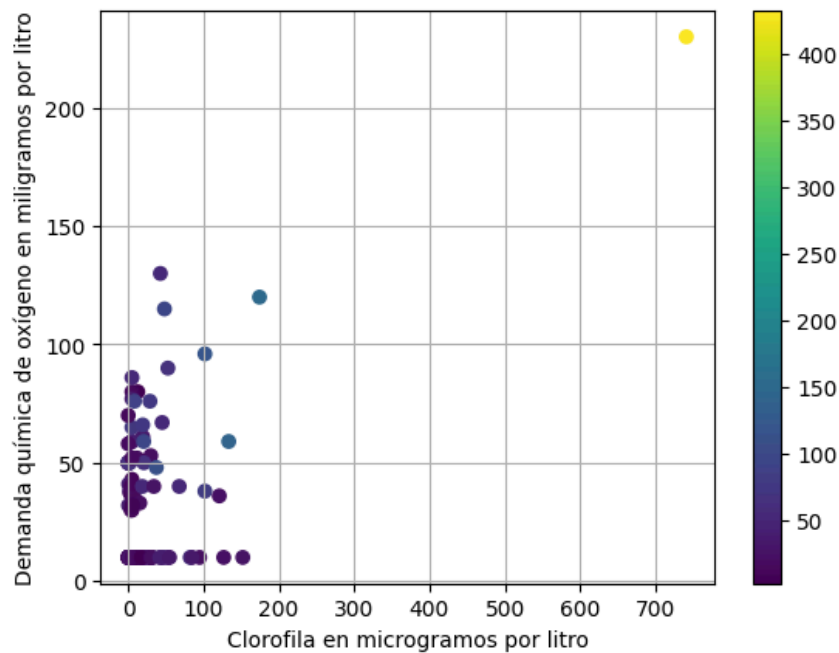
Hipótesis 3: A mayor clorofila y demanda química de oxígeno, más turbia está el agua

Estudiando la contaminación por materia orgánica, se nos ocurrió estudiar qué sucede con las variables que indican materia orgánica en el agua del Río de la Plata. La elección de estas variables se debe a que notamos una alta correlación de la clorofila y la demanda de oxígeno con turbiedad.

La clorofila es un conjunto de pigmentos que son los responsables de dar a las plantas su color verde y su función es absorber la energía lumínica. Un alto nivel de clorofila es un índice de fitoplancton en el agua.

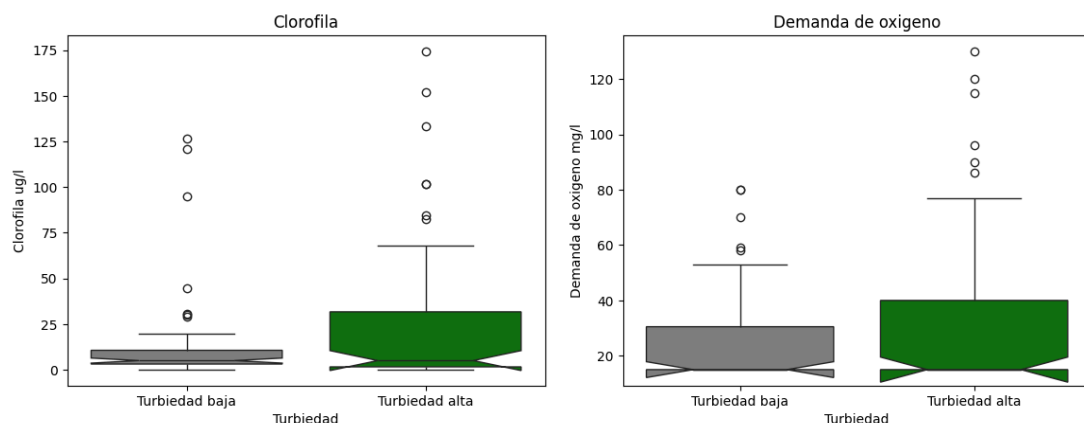
Por otro lado, una alta demanda química de oxígeno requerido para oxidar materia se presenta cuando existen altos valores de materia orgánica en el agua.

Realizamos un análisis para observar si altos índices de clorofila y de demanda de oxígeno requerida provocan turbiedad en el agua, ya que es posible que el agua más turbia contenga más algas, y por lo tanto más pigmento.



Relación entre clorofila y demanda química de oxígeno con turbiedad

Podemos observar que a mayor clorofila y mayor demanda de oxígeno, se producen mayores niveles de turbiedad en el agua. Vemos que existe un outlier que presenta los valores máximos de las tres variables. Esta muestra fue tomada en verano en el sitio “Toma de Agua Club de Pesca”. Además vemos que presenta un valor alto de microcistina (5 ug/l), esta toxina es producida por algunos tipos de algas, lo cual mantiene relación con los altos niveles de clorofila y demanda de oxígeno. Para el siguiente análisis decidimos eliminar el outlier para no alterar los resultados.



Los outliers que se ven en los gráficos fueron incluidos ya que son valores posibles de clorofila en el agua y creemos que son importantes para el análisis.

Realizamos un test de Kruskal-Wallis para cada variable cuantitativa por separado con un nivel de significancia de 0,05. La hipótesis nula es que no hay diferencia significativa entre los grupos, con niveles altos y niveles bajos de turbiedad.

Por otro lado, el p-valor para el test de demanda de oxígeno dio menor al nivel de significancia, por lo que se rechaza la hipótesis nula y afirmamos que los grupos son significativamente diferentes.

Hipótesis 4: La contaminación por materia fecal provoca olor.

Las colonias de coliformes fecales y escherichia coli son un indicador de contaminación por materia fecal. Esto puede producir olor en el ambiente.

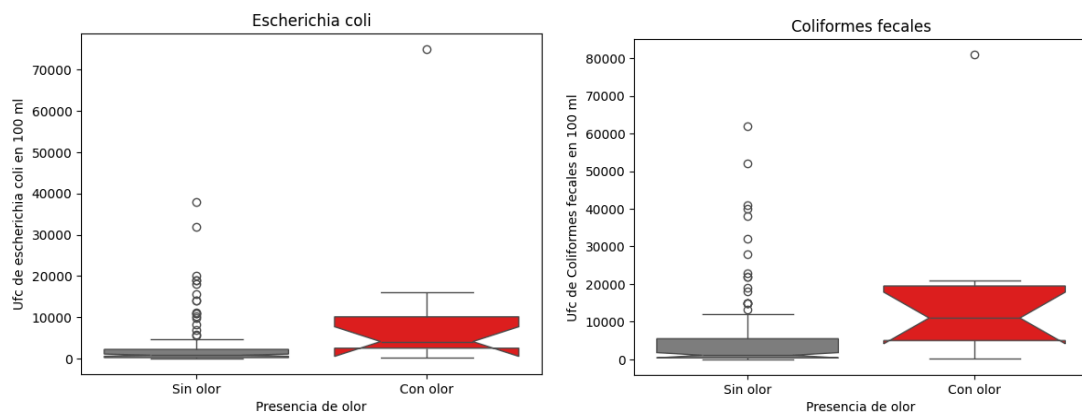
El gráfico nos muestra que existen registros con presencia de olor y baja contaminación fecal, y también, registros con alta contaminación fecal sin presencia de olor. Por lo tanto esto nos estaría indicando que no hay una clara relación entre estas variables. A continuación realizamos un test de hipótesis que determine si verdaderamente no hay relación, ya que los gráficos no son suficientes para refutar la hipótesis.

Decidimos eliminar los dos outliers que tienen altos valores de las variables estudiadas y que no tienen presencia de olor, para que no afecten significativamente el resultado de nuestro test.

Se realizó un test de Kruskal-Wallis para cada variable cuantitativa por separado con un nivel de significancia de 0,05. La hipótesis nula es que no hay diferencia significativa entre los grupos, con olor y sin olor.

El p-valor para ambos test dio por debajo del nivel de significancia lo cual nos lleva a rechazar la hipótesis nula y confirmar que existe diferencia significativa entre los grupos estudiados en esta hipótesis.

Para ver gráficamente esta diferencia entre ellos, realizamos un gráfico de box-plot por cada variable:

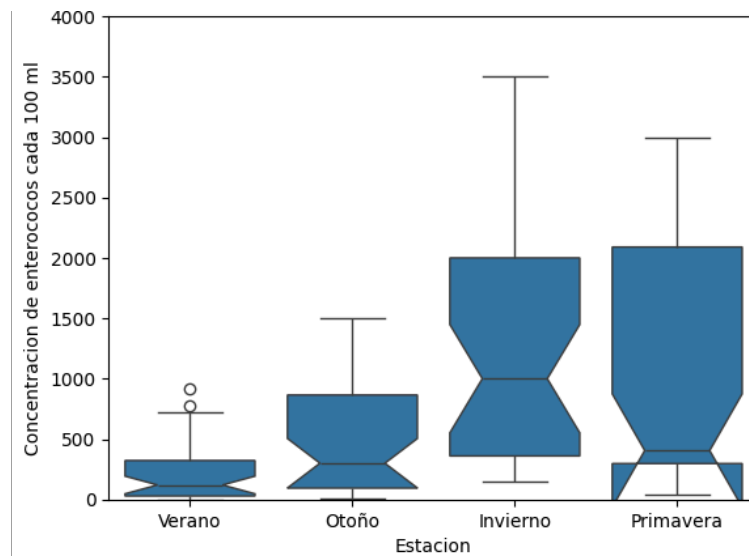


En conclusión esto refuta lo que analizamos en el gráfico de dispersión, comprobando así, que se puede aceptar la hipótesis de que la contaminación por materia fecal provocan la presencia de olor en el Río de la Plata.

Hipótesis 5: La estación del año afecta a la contaminación por enterococos.

Siguiendo la línea de análisis de la hipótesis anterior, pasamos a estudiar cómo influyen las estaciones del año a la contaminación fecal. Para esto usaremos la variable que registra concentración de enterococos, que es un indicador de presencia de materia fecal.

Se realizó el siguiente gráfico en el que se ve la diferente distribución de unidades colonias de enterococos en 100 ml de agua por estación del año. Se puede observar que las cantidades más altas de colonias se encuentran en invierno y primavera.

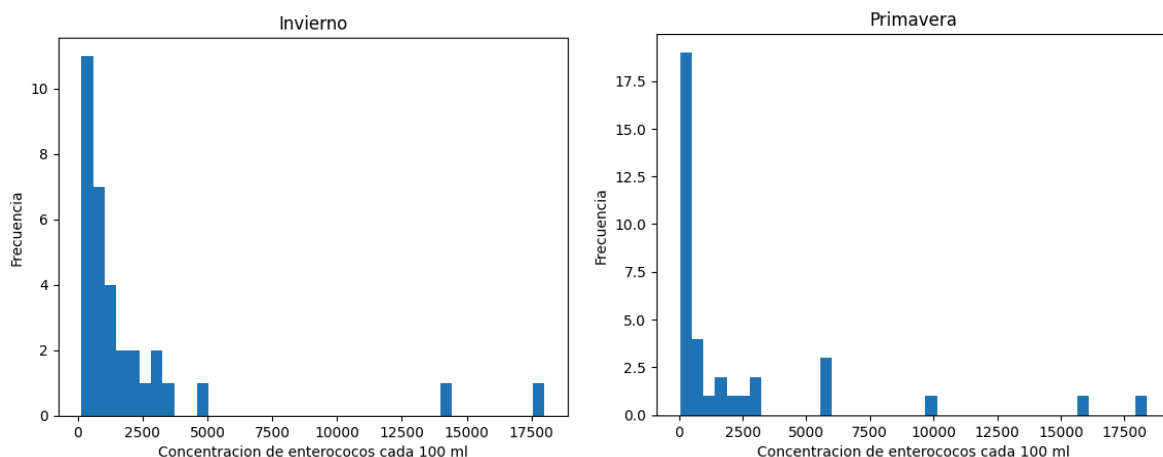


Distribución de enterococos en cada estación del año

Comprobamos si la diferencia entre estos grupos es significativa a partir del test de hipótesis Kruskal-Wallis, con un nivel de significancia de 0.05. Nuestra hipótesis nula fue que no hay diferencia significativa entre los grupos, la cual fue rechazada al obtener un P-valor menor que el nivel de significancia.

Con esto podemos confirmar nuestra hipótesis y afirmar que la estación del año, por ende la temperatura, afecta a la concentración de colonias de enterococos en el agua.

Decidimos ir más al detalle y analizar en qué estación se da la mayor concentración de esta bacteria. Viendo sus frecuencias podemos notar que en invierno hay más muestras con mayor concentración, pero la muestra con el valor máximo de enterococos fue tomada en primavera.



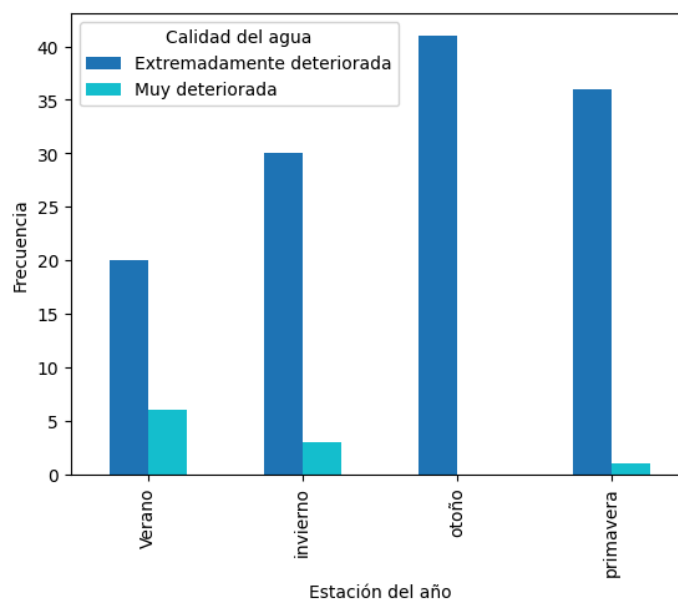
Un ejemplo de dos muestras tomadas en el mismo sitio pero en diferente estación es Perú Puente cuyo valor en invierno es de 18000 ufc y en primavera 10000 ufc.

Podemos concluir que, mayoritariamente, la estación en la que se encuentran concentraciones de colonias de enterococos más altas es invierno.

Hipótesis 6: La calidad del agua depende de la estación del año

Pensamos en esta hipótesis ya que opinamos que la época del año afecta a la contaminación del agua, dado que, por ejemplo, en primavera las algas crecen más y en verano, al hacer más calor, pueden aumentar los olores. Por otro lado, el volumen de producción de las industrias puede variar a lo largo del año al igual que la temporada de pesca. También es lógico pensar que en ciertas épocas del año la población utiliza en mayor medida el transporte náutico, es decir que hay mayor movimiento de barcos, por ende, mayor contaminación.

A continuación desarrollamos el análisis para estudiar esta hipótesis.

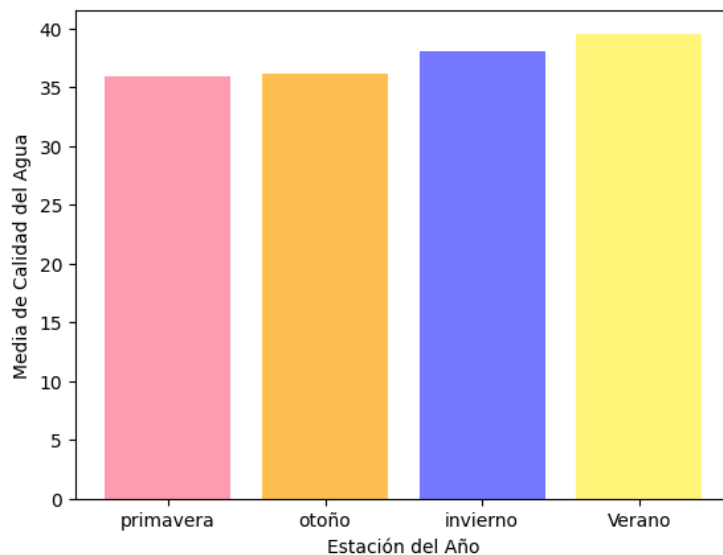


Frecuencia en las estaciones y la calidad del agua a través de gráfico de barra

En este gráfico de barras podemos observar cómo se distribuyen los datos a través de las estaciones del año y la calidad de agua indicada. Podemos ver que las barras con mayor frecuencia de datos corresponden a una calidad de agua extremadamente deteriorada. A través de esto deducimos que el agua del Río de la Plata se encuentra, en su mayor parte, en muy malas condiciones.

Para poder aceptar o rechazar nuestra hipótesis, realizamos un test de Chi-cuadrado con un nivel de significancia de 0.05, siendo la hipótesis nula que las dos variables categóricas son independientes. Dado que se obtuvo un p-valor menor a 0,05 se puede concluir que existe una relación significativa entre estas variables.

En conclusión, podemos afirmar que la calidad del agua del Río de la Plata depende de la estación del año y según los valores de *ica* tomados en cada estación, se ve que las más afectadas son otoño y primavera, parecido a la contaminación por enterococos. De todos modos se puede ver un comportamiento parecido durante todo el año y que el índice nunca supera el valor 50, lo que demuestra el mal estado del Río de la Plata.



Media de calidad del agua por estación del año

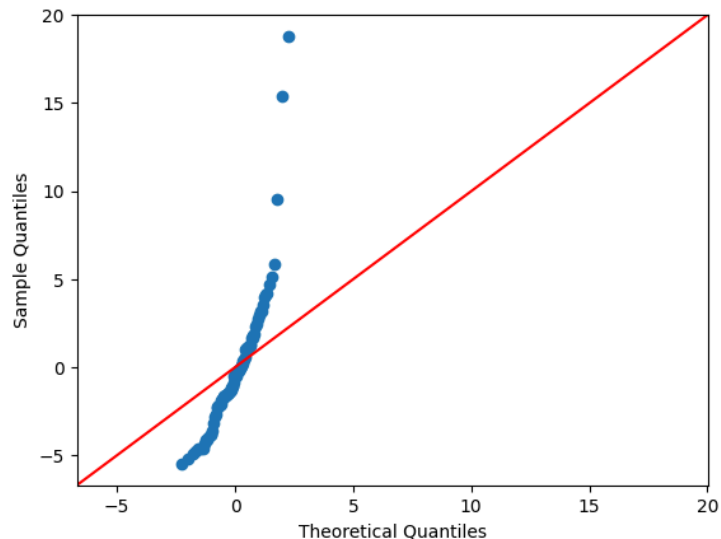
Hipótesis 7: Se puede predecir el índice de calidad del agua a partir de la contaminación por materia fecal, orgánica, y por la temperatura

Luego de todo el análisis realizado, creemos que se puede predecir el comportamiento del índice de calidad de agua a través de las variables estudiadas, es decir, de las que contribuyen a la contaminación por materia fecal, por materia orgánica y las que afectan a la temperatura en el Río de la Plata.

A continuación realizaremos un análisis adecuado, por regresión lineal, para determinar con exactitud si verdaderamente son estas las variables que más afectan a las condiciones del agua del Río de la Plata.

A través de la información que nos otorgó el modelo podemos decir que:

- $R^2 = 0.445$. Esto nos dice que el modelo nos explica un 44,5% de la varianza
- Prob (F-statistic) > 0.001. Esto nos muestra que el modelo no es significativo, es decir que estas variables no son muy útiles para predecir el *ica*.
- Coeficientes del modelo lineal: Vemos que las variables que mas influyen son: *od*, *enteroc_ufc_100ml*, *fosp_ortofos_mg_l*, *cd_total_mg_l* y *clorofila_a_ug_l*.
- P-valor de Omnibus y Jarque-Bera < 0.001. Esto nos indica que los residuos no son normales. Haremos un QQ-plot para analizar la normalidad a detalle:



Comportamiento de los residuos en gráfico QQ-Plot

Finalmente, concluimos que no se puede predecir el índice de calidad del agua con regresión lineal a través de las variables proporcionadas. Probablemente se podría estudiar esta variable con un análisis no lineal pero, en base a la técnica utilizada, rechazamos la hipótesis.

Conclusiones

A partir del dataset proporcionado por la cátedra y a lo largo de este informe, investigamos las posibles causas de contaminación del Río de la Plata, los agentes principales involucrados y cómo estos varían según la estación del año.

Creemos que un gran indicador de la calidad del agua es la turbiedad y el olor, que son influenciados por la contaminación por materia orgánica. A raíz de esto, estudiamos cómo se comporta esta materia orgánica y pudimos hallar que la turbidez del agua es directamente afectada por la demanda química de oxígeno, que es un indicador de presencia de materia orgánica. Por otro lado, también pudimos confirmar que el olor es consecuencia, en gran parte, de la contaminación por materia fecal, la cual se presenta en mayor medida en invierno y primavera, contrario a nuestras expectativas iniciales.

Al analizar otros indicadores del dominio, encontramos que el nitrato es un indicador de contaminación por industrias y cultivos, e indagando, localizamos varios sectores agrícolas e industriales cerca de las muestras con mayor cantidad de nitrato.

En cuanto a la calidad del agua, demostramos que es afectada por la época del año, siendo — la época con peor calidad de agua.

En un principio, creímos que la contaminación por materia orgánica y la temperatura eran los factores más influyentes en cuanto a la calidad del agua, pero no pudimos demostrarlo a través de los datos y las herramientas utilizadas.

Por último, aunque en general los datos no dieron los resultados esperados y no pudimos relacionarlos directamente con el índice de calidad a través de las herramientas a las que pudimos acceder, concluimos que las variables elegidas para estudiar el agua sí explican la contaminación en el Río de la Plata. Además, nos pareció interesante observar cómo se relacionan entre ellas y cómo una puede incidir en la aparición y/o cantidad de otra.

Referencias

Nitrato:

- [¿Qué lo produce?](#)
- [¿Qué cantidad es contraproducente?](#) - Apartado II.2.1

Hidrocarburos:

- [¿Qué cantidad es contraproducente?](#) - Artículo 32, anexo A

Olores:

- [¿Qué provoca olor en el Río de la Plata?](#)

Clorofila:

- [Características de la clorofila](#)
- [Valores de clorofila en agua](#)

Demanda Química de oxígeno:

- [Características de DQO](#)

Ortofosfatos:

- [¿Qué lo provoca?](#)

Amonio y nitrato:

- [¿Cómo se relacionan?](#)

Turbiedad:

- [¿Que valor se considera alto en agua de río?](#)

Índice de calidad de agua

- [¿Que valores indican una baja calidad?](#)