

Datos FIC LL

Punto 1: confirmar que los análisis que estamos realizando tienen sentido y los criterios son correctos (si bien pueden estar incompletos).

Estamos realizando análisis de series de tiempo, por lo que nos interesa construir curvas de un cierto valor (resultado, eje y) por día (fecha, eje x), y por ahora estamos considerando una curva para cada análisis por área.

- Como valor consideramos el “Resultado” por “FechaExtraccion”
- Dado que estamos tratando de realizar análisis de indicadores (análisis) por área, agrupamos los datos válidos de un mismo día para un par (área, análisis) con algún criterio. Por ahora estamos considerando el valor Máximo por día para cada (área, análisis).
- Sólo estamos considerando las filas con estado “Informado”.
- Si una fila tiene columna signo “<”, lo consideramos 0. Si tiene “T” (Traces), lo corregimos como el LimiteAccion si es de tipo “DTX” o “AZA” o 0 en los demás. Ignoramos este campo si está vacío o es “>”.

Punto 2: Dudas acerca de datos o criterios adicionales.

- Suavizamiento de curvas: Promedio móvil semanal (o suavizado exponencial) parece razonable, pero hay casos con valores 0 intercalados diariamente entre datos mayores, quizás sea mejor omitir los 0 (o interpolarlos) entre datos mayores que cero y cercanos, antes de realizar suavizamientos.
- Respecto a la columna estado: ¿Qué significan exactamente las etiquetas INFORMADO/ANALIZADO/INICIADO? ¿Se podría extraer información relevante para algún análisis desde las otras?
- Respecto a la columna signo: ¿A qué corresponde?
- ¿La columna LOD es *Limit Of Detection*? ¿Qué significa exactamente? Hay filas con resultado (no cero) menor que LOD. Si el resultado es menor a LOD, el registro debería ser ignorado (criterio adicional de corrección a 0).
- ¿Qué análisis son relevantes para ellos? ¿Únicamente los que estamos considerando?
- Considerar también curvas para los otros indicadores incluyendo oxígeno, temperatura, etc. En esos casos, otras agrupaciones para diferentes curvas pueden ser válidas (max/min/mean).

Punto 3: Limpieza de datos.

Para efectos del análisis, estoy normalizando los textos pasando todo a mayúsculas y eliminando espacios y acentos. A continuación listo los valores observados con esas correcciones.

Estados observados:

-> "INFORMADO" (Siguiendo el diccionario, sólo considero filas con este estado. Esto es más para eliminar datos duplicados, pero podría cambiarse a un criterio para *priorizar* informado cuando haya varios, en lugar de simplemente omitir los demás, por si hubiera quedado mal reportado.)

-> "ANALIZADO"

-> "INICIADO"

Signos observados:

-> " "

-> ""

-> "T" (Si el registro pertenece al grupo "DTX" o "AZA", entonces corrijo el resultado al valor de la columna "LimiteAccion", de otro modo, lo corrijo a 0)

-> "<" (Siguiendo el diccionario, corrijo el resultado a 0)

-> "="

-> ">" (Notar que *hay* datos con este signo, aunque no debería)

Tipos de Análisis observados:

-> "QUIMICOS"

-> "FITOPLANCTON" (Podría ser relevante usar alguno de estos análisis, pero probablemente con datos adicionales para identificar un fitoplancton vs otro)

-> "TOXICOLOGICOSXLCMSMS" (Siguiendo el diccionario considero estas filas)

-> "MICROBIOLOGICOS" (Algunos de estos también podrían ser usados)

-> "TOXICOLOGICOS" (Siguiendo el diccionario considero estas filas)

Análisis observados:

-> "QUIMICOS"

-> "CD"

-> "ALDRIN"

-> "BHCALFA"

-> "BHCBETA"

-> "HEPTACHLOR"

-> "PESTICIDAS"

-> "HEPTACHLOREPOXIDO"

-> "PB"

-> "DDT"

-> "ENDRIN"

-> "LINDANO"

-> "DIELDRIN"

-> "HG"

-> "FITOPLANCTON"

-> "REDCUALITATIVA"

-> "INTEGRADACUANTITATIVA"

-> "SUPERFICIALCUANTITATIVA"

-> "TOXICOLOGICOSXLCMSMS"

-> "45OHYTX"

-> "AZA2"

-> "AZA3"

-> "HOMOYTX"

-> "AZA1"

-> "45OHHOMOYTX"

-> "PTX1"

-> "DTX1"

-> "DTX2"

-> "PTX2"

- > "YTX"
- > "AO"
- > "MICROBIOLOGICOS"
- > "ECOLIA"
- > "VIBRIOPARAHEMOLYTICUS"
- > "ECOLIB"
- > "SALMONELLA"
- > "TOXICOLOGICOS"
- > "VPM"
- > "VAM"
- > "VDMBE"

Grupos de análisis considerados:

- > YTX
- > 45OHHOMOYTX
- > 45OHYTX
- > HOMOYTX
- > YTX

- > PTX_AO
- > AO
- > PTX1
- > PTX2

- > AZA
- > AZA1
- > AZA2
- > AZA3

- > DTX
- > DTX1
- > DTX2

- > VAM
- > VAM

- > VPM
- > VPM