# Tutorial Paso a Paso: Gestión de Datos de Calidad del Agua en Power BI

**Ejercicio Integral - Embalse Gatún, Canal de Panamá**

## Preparación del Entorno de Trabajo

### Requisitos Previos

1. **Software necesario:**
   * Power BI Desktop (versión más reciente)
   * Microsoft Excel (para verificación de datos)
   * Acceso a los archivos de datos del ejercicio
2. **Archivos del ejercicio:**
   * datos\_bloque1\_problematicos.xlsx
   * datos\_bloque2\_terminologia.xlsx
   * datos\_bloque3\_validacion.xlsx
   * tablas\_referencia\_calidad\_agua.xlsx
   * Superficie.xlsx (datos originales para referencia)
3. **Configuración inicial:**
   * Crear carpeta de trabajo: C:\Ejercicio\_Calidad\_Agua\_Gatun
   * Copiar todos los archivos a esta carpeta
   * Verificar que Power BI Desktop esté actualizado

## BLOQUE 1: Importación y Estandarización de Datos (1.5 horas)

### Paso 1: Configuración del Proyecto Power BI

#### 1.1 Crear Nuevo Proyecto

1. **Abrir Power BI Desktop**
   * Hacer clic en el icono de Power BI Desktop
   * Seleccionar "Crear un informe en blanco"
2. **Configurar opciones del proyecto**

Archivo > Opciones y configuración > Opciones

**En "Carga de datos":**

* + ✅ Desmarcar "Detectar automáticamente relaciones entre tablas al cargar por primera vez"
  + ✅ Marcar "Actualizar o eliminar relaciones al actualizar datos"

**En "Power Query Editor":**

* + Tiempo de espera de consulta: 30 minutos
  + ✅ Marcar "Habilitar carga en segundo plano"

1. **Guardar proyecto**

Archivo > Guardar como

Nombre: "Ejercicio\_Calidad\_Agua\_Gatun\_v1.pbix"

Ubicación: C:\Ejercicio\_Calidad\_Agua\_Gatun\

#### 1.2 Configurar Opciones de Privacidad

Archivo > Opciones y configuración > Configuración de origen de datos

* Configurar nivel de privacidad como "Organizacional" para todos los archivos Excel

### Paso 2: Importación de Datos Problemáticos

#### 2.1 Importar Primer Archivo (Bloque 1)

1. **Iniciar importación**

Inicio > Obtener datos > Excel

1. **Seleccionar archivo**
   * Navegar a: C:\Ejercicio\_Calidad\_Agua\_Gatun\datos\_bloque1\_problematicos.xlsx
   * Hacer clic en "Abrir"
2. **Examinar estructura de datos**
   * Verificar que se detecte la hoja principal
   * Observar vista previa de datos
   * **IMPORTANTE:** NO hacer clic en "Cargar" todavía
3. **Abrir Editor de Power Query**
   * Hacer clic en "Transformar datos"
   * Esto abrirá el Editor de Power Query donde realizaremos las transformaciones

#### 2.2 Análisis Inicial de Problemas

**En el Editor de Power Query, examinar:**

1. **Problemas de nombres de columnas identificados:**

Problemas encontrados:

- 'ID ' (espacio al final)

- 'Alcalinidad Total' (nombre en español)

- 'Ca++' (formato químico)

- 'Clorofila-a ' (guión y espacio)

- 'Cloruros (mg/L)' (unidades en nombre)

- 'CONDUCTIVIDAD' (mayúsculas)

- 'coliformes\_totales' (minúsculas con guión bajo)

- 'Dureza del Agua' (espacios)

- 'E.coli' (sin espacio)

- Y muchos más...

1. **Problemas de datos identificados:**
   * Valores faltantes (celdas vacías)
   * Tipos de datos mixtos (texto y números mezclados)
   * Formatos inconsistentes (comas vs puntos decimales)

### Paso 3: Limpieza de Nombres de Columnas

#### 3.1 Corrección Sistemática de Nombres

**Método 1: Corrección Manual (Recomendado para aprendizaje)**

1. **Hacer doble clic en cada encabezado de columna y renombrar:**

Cambios a realizar:

'ID ' → 'Estacion\_ID'

'Alcalinidad Total' → 'Alcalinidad\_Total'

'Ca++' → 'Calcio\_Ca'

'Clorofila-a ' → 'Clorofila\_A'

'Cloruros (mg/L)' → 'Cloruros\_Cl'

'CONDUCTIVIDAD' → 'Conductividad'

'coliformes\_totales' → 'Coliformes\_Totales'

'Dureza del Agua' → 'Dureza\_Agua'

'E.coli' → 'E\_coli'

'K+' → 'Potasio\_K'

'Magnesio ' → 'Magnesio\_Mg'

'Nitratos-NO3' → 'Nitratos\_NO3'

'Sodio (Na)' → 'Sodio\_Na'

'O.D.' → 'Oxigeno\_Disuelto'

'Sat\_O2\_%' → 'Saturacion\_O2'

'Fosfatos PO4' → 'Fosfatos\_PO4'

'pH' → 'pH' (ya está correcto)

'Salinidad‰' → 'Salinidad'

'Sulfatos SO4--' → 'Sulfatos\_SO4'

'SDT' → 'Solidos\_Disueltos\_Totales'

'SST' → 'Solidos\_Suspendidos\_Totales'

'Temp°C' → 'Temperatura'

'Transparencia\_m' → 'Transparencia'

'NTU' → 'Turbidez'

1. **Verificar que todos los nombres estén estandarizados:**
   * Sin espacios al inicio o final
   * Sin caracteres especiales (excepto guión bajo)
   * Formato consistente: Primera\_Palabra\_Segunda\_Palabra
   * Sin unidades en los nombres

**Método 2: Corrección Automática (Avanzado)**

Seleccionar todas las columnas > Clic derecho > Quitar espacios > Recortar

#### 3.2 Documentar Cambios Realizados

**Crear tabla de log de cambios:**

1. En una hoja de Excel separada, documentar:

Nombre\_Original | Nombre\_Nuevo | Justificacion | Fecha\_Cambio

'ID ' | 'Estacion\_ID' | Eliminar espacio, clarificar propósito | [Fecha actual]

'Ca++' | 'Calcio\_Ca' | Estandarizar nomenclatura química | [Fecha actual]

### Paso 4: Corrección de Tipos de Datos

#### 4.1 Identificar Problemas de Tipos de Datos

1. **Examinar columnas con problemas:**
   * Columna pH: Contiene texto y números
   * Columna Temperatura: Contiene valores con "°C"
   * Otras columnas numéricas que aparecen como texto
2. **Verificar tipos de datos actuales:**
   * Observar iconos en encabezados de columnas
   * 📝 = Texto
   * 🔢 = Número entero
   * 🔢. = Número decimal
   * 📅 = Fecha

#### 4.2 Corregir Columna pH

1. **Seleccionar columna pH**
2. **Reemplazar comas por puntos:**

Inicio > Reemplazar valores

Valor que se va a buscar: ,

Reemplazar por: .

1. **Cambiar tipo de datos:**

Transformar > Tipo de datos > Número decimal

1. **Verificar que no hay errores:**
   * Si aparecen errores, revisar valores problemáticos
   * Usar "Quitar errores" si es necesario

#### 4.3 Corregir Columna Temperatura

1. **Seleccionar columna Temperatura**
2. **Reemplazar texto "°C":**

Inicio > Reemplazar valores

Valor que se va a buscar: °C

Reemplazar por: [dejar vacío]

1. **Cambiar tipo de datos a número decimal**

#### 4.4 Verificar Todas las Columnas Numéricas

**Columnas que deben ser numéricas:**

* Alcalinidad\_Total, Calcio\_Ca, Clorofila\_A, Cloruros\_Cl
* Conductividad, Coliformes\_Totales, Dureza\_Agua, E\_coli
* Potasio\_K, Magnesio\_Mg, Nitratos\_NO3, Sodio\_Na
* Oxigeno\_Disuelto, Saturacion\_O2, Fosfatos\_PO4, pH
* Salinidad, Sulfatos\_SO4, Solidos\_Disueltos\_Totales
* Solidos\_Suspendidos\_Totales, Temperatura, Transparencia, Turbidez

**Para cada columna numérica:**

1. Verificar tipo de datos
2. Corregir si es necesario
3. Verificar que no hay errores

### Paso 5: Manejo de Valores Faltantes

#### 5.1 Identificar Valores Faltantes

1. **Examinar cada columna para valores nulos**
2. **Documentar patrones de valores faltantes:**
   * ¿Están concentrados en ciertas estaciones?
   * ¿Afectan a parámetros específicos?
   * ¿Hay patrones sistemáticos?

#### 5.2 Estrategias para Valores Faltantes

**Opción 1: Mantener valores nulos (Recomendado)**

* Para análisis científicos, es mejor mantener la información sobre datos faltantes
* Power BI maneja automáticamente los valores nulos en cálculos

**Opción 2: Imputación simple (Solo si es justificable)**

Agregar columna > Columna personalizada

Nombre: Parametro\_Imputado

Fórmula: if [Parametro] = null then [Valor\_por\_defecto] else [Parametro]

**Opción 3: Marcar para revisión**

Agregar columna > Columna personalizada

Nombre: Requiere\_Revision

Fórmula: if [Parametro] = null then "Revisar" else "OK"

### Paso 6: Validaciones Básicas de Calidad

#### 6.1 Crear Validaciones de Rango

**Validación de pH:**

Agregar columna > Columna personalizada

Nombre: pH\_Valido

Fórmula: if [pH] >= 6 and [pH] <= 9 then "Válido" else "Revisar"

**Validación de Temperatura:**

Agregar columna > Columna personalizada

Nombre: Temperatura\_Valida

Fórmula: if [Temperatura] >= 20 and [Temperatura] <= 35 then "Válido" else "Revisar"

**Validación de Oxígeno Disuelto:**

Agregar columna > Columna personalizada

Nombre: OD\_Valido

Fórmula: if [Oxigeno\_Disuelto] >= 2 then "Válido" else "Crítico"

**Validación de Conductividad:**

Agregar columna > Columna personalizada

Nombre: Conductividad\_Valida

Fórmula: if [Conductividad] >= 10 and [Conductividad] <= 1500 then "Válido" else "Revisar"

#### 6.2 Crear Columnas de Metadatos

**Fecha de procesamiento:**

Agregar columna > Columna personalizada

Nombre: Fecha\_Procesamiento

Fórmula: DateTime.LocalNow()

**ID único de registro:**

Agregar columna > Columna de índice

Nombre: ID\_Registro

Desde: 1

Incremento: 1

**Fuente de datos:**

Agregar columna > Columna personalizada

Nombre: Fuente\_Datos

Fórmula: "Bloque1\_Problematicos"

### Paso 7: Aplicar y Cargar Datos

#### 7.1 Revisar Transformaciones

1. **Verificar panel "Pasos aplicados":**
   * Cada transformación debe estar listada
   * Verificar que no hay errores
   * Documentar pasos críticos
2. **Vista previa final:**
   * Verificar que todos los nombres de columnas están correctos
   * Confirmar que los tipos de datos son apropiados
   * Revisar que las validaciones funcionan correctamente

#### 7.2 Cargar Datos al Modelo

1. **Hacer clic en "Cerrar y aplicar"**
2. **Verificar carga exitosa:**
   * Los datos aparecen en el panel "Campos"
   * No hay errores de carga
   * El conteo de filas es correcto
3. **Renombrar tabla si es necesario:**

Clic derecho en tabla > Cambiar nombre > "Datos\_Bloque1\_Limpios"

## BLOQUE 2: Estandarización con Glosarios Ambientales (1 hora)

### Paso 8: Importar Datos con Problemas de Terminología

#### 8.1 Importar Segundo Archivo

1. **Obtener datos > Excel**
   * Seleccionar: datos\_bloque2\_terminologia.xlsx
   * Abrir en Editor de Power Query
2. **Examinar problemas de terminología:**
   * Columna ID: Variantes de nombres de estaciones
   * Columna Tipo\_Parametro: Inconsistencias en clasificación
   * Columna Calidad\_Agua: Términos no estandarizados

#### 8.2 Importar Tablas de Referencia

1. **Obtener datos > Excel**
   * Seleccionar: tablas\_referencia\_calidad\_agua.xlsx
   * Importar todas las hojas:
     + Estaciones\_Referencia
     + Estandares\_Parametros
     + Clasificacion\_Calidad

### Paso 9: Crear Glosarios Específicos

#### 9.1 Crear Tabla de Clasificación de Parámetros

**En Power BI (Vista de datos):**

Modelado > Nueva tabla

**Código DAX:**

Clasificacion\_Parametros =

DATATABLE(

"Parametro", STRING,

"Categoria", STRING,

"Subcategoria", STRING,

"Unidad\_Estandar", STRING,

"Codigo\_Parametro", STRING,

"Descripcion", STRING,

{

{"pH", "Fisicoquímico", "Acidez-Alcalinidad", "unidades pH", "PH", "Potencial de hidrógeno"},

{"Temperatura", "Fisicoquímico", "Térmica", "°C", "TEMP", "Temperatura del agua"},

{"Oxigeno\_Disuelto", "Fisicoquímico", "Oxigenación", "mg/L", "OD", "Concentración de oxígeno disuelto"},

{"Conductividad", "Fisicoquímico", "Mineralización", "μS/cm", "COND", "Conductividad eléctrica"},

{"Turbidez", "Físico", "Óptico", "NTU", "TURB", "Turbidez del agua"},

{"Clorofila\_A", "Biológico", "Productividad", "μg/L", "CHLA", "Concentración de clorofila a"},

{"E\_coli", "Bacteriológico", "Contaminación", "UFC/100mL", "ECOL", "Escherichia coli"},

{"Coliformes\_Totales", "Bacteriológico", "Contaminación", "UFC/100mL", "CTOT", "Coliformes totales"},

{"Nitratos\_NO3", "Químico", "Nutrientes", "mg/L", "NO3", "Nitrógeno como nitrato"},

{"Fosfatos\_PO4", "Químico", "Nutrientes", "mg/L", "PO4", "Fósforo como fosfato"},

{"Solidos\_Disueltos\_Totales", "Físico", "Sólidos", "mg/L", "TDS", "Sólidos disueltos totales"},

{"Solidos\_Suspendidos\_Totales", "Físico", "Sólidos", "mg/L", "TSS", "Sólidos suspendidos totales"},

{"Transparencia", "Físico", "Óptico", "m", "TRAN", "Transparencia del agua"},

{"Calcio\_Ca", "Químico", "Cationes", "mg/L", "CA", "Calcio"},

{"Magnesio\_Mg", "Químico", "Cationes", "mg/L", "MG", "Magnesio"},

{"Sodio\_Na", "Químico", "Cationes", "mg/L", "NA", "Sodio"},

{"Potasio\_K", "Químico", "Cationes", "mg/L", "K", "Potasio"},

{"Cloruros\_Cl", "Químico", "Aniones", "mg/L", "CL", "Cloruros"},

{"Sulfatos\_SO4", "Químico", "Aniones", "mg/L", "SO4", "Sulfatos"},

{"Alcalinidad\_Total", "Químico", "Alcalinidad", "mg/L", "ALK", "Alcalinidad total"},

{"Dureza\_Agua", "Químico", "Dureza", "mg/L", "HARD", "Dureza del agua"},

{"Salinidad", "Fisicoquímico", "Salinidad", "UPS", "SAL", "Salinidad"},

{"Saturacion\_O2", "Fisicoquímico", "Oxigenación", "%", "SATO2", "Saturación de oxígeno"}

}

)

#### 9.2 Crear Tabla de Rangos de Referencia

Rangos\_Referencia =

DATATABLE(

"Parametro", STRING,

"Valor\_Min\_Optimo", DOUBLE,

"Valor\_Max\_Optimo", DOUBLE,

"Valor\_Critico\_Bajo", DOUBLE,

"Valor\_Critico\_Alto", DOUBLE,

"Unidad", STRING,

"Referencia\_Estandar", STRING,

"Uso\_Recomendado", STRING,

{

{"pH", 6.5, 8.5, 6.0, 9.0, "unidades pH", "OMS", "Agua potable y ecosistemas"},

{"Temperatura", 20, 30, 15, 35, "°C", "EPA", "Ecosistemas tropicales"},

{"Oxigeno\_Disuelto", 5, 12, 2, 15, "mg/L", "EPA", "Vida acuática"},

{"Conductividad", 50, 500, 10, 1500, "μS/cm", "EPA", "Aguas naturales"},

{"Turbidez", 0, 4, 0, 50, "NTU", "OMS", "Agua potable"},

{"Clorofila\_A", 0, 10, 0, 50, "μg/L", "OECD", "Prevención eutrofización"},

{"E\_coli", 0, 100, 0, 1000, "UFC/100mL", "OMS", "Agua recreacional"},

{"Coliformes\_Totales", 0, 1000, 0, 10000, "UFC/100mL", "OMS", "Agua recreacional"},

{"Nitratos\_NO3", 0, 5, 0, 10, "mg/L", "EPA", "Prevención eutrofización"},

{"Fosfatos\_PO4", 0, 0.025, 0, 0.1, "mg/L", "EPA", "Prevención eutrofización"},

{"Solidos\_Disueltos\_Totales", 30, 300, 5, 1000, "mg/L", "EPA", "Aguas naturales"},

{"Transparencia", 1, 5, 0.3, 10, "m", "OECD", "Ecosistemas acuáticos"}

}

)

#### 9.3 Crear Glosario de Estaciones Expandido

Glosario\_Estaciones =

DATATABLE(

"Codigo\_Original", STRING,

"Codigo\_Estandar", STRING,

"Nombre\_Completo", STRING,

"Zona\_Embalse", STRING,

"Tipo\_Ambiente", STRING,

"Coordenada\_Lat", DOUBLE,

"Coordenada\_Lon", DOUBLE,

"Profundidad\_Promedio", DOUBLE,

"Caracteristicas\_Especiales", STRING,

{

{"ARN", "EST001", "Estación Arnulfo", "Central", "Zona de mezcla", 9.2129, -79.9077, 15, "Zona de confluencia de corrientes"},

{"Arnulfo", "EST001", "Estación Arnulfo", "Central", "Zona de mezcla", 9.2129, -79.9077, 15, "Zona de confluencia de corrientes"},

{"EST-ARN", "EST001", "Estación Arnulfo", "Central", "Zona de mezcla", 9.2129, -79.9077, 15, "Zona de confluencia de corrientes"},

{"Estación Arnulfo", "EST001", "Estación Arnulfo", "Central", "Zona de mezcla", 9.2129, -79.9077, 15, "Zona de confluencia de corrientes"},

{"BAT", "EST002", "Estación Batea", "Norte", "Influencia fluvial", 9.2500, -79.8750, 12, "Entrada de tributarios principales"},

{"Batea", "EST002", "Estación Batea", "Norte", "Influencia fluvial", 9.2500, -79.8750, 12, "Entrada de tributarios principales"},

{"EST-BAT", "EST002", "Estación Batea", "Norte", "Influencia fluvial", 9.2500, -79.8750, 12, "Entrada de tributarios principales"},

{"Estación Batea", "EST002", "Estación Batea", "Norte", "Influencia fluvial", 9.2500, -79.8750, 12, "Entrada de tributarios principales"},

{"BCI", "EST003", "Isla Barro Colorado", "Protegida", "Área de investigación", 9.1500, -79.8500, 20, "Estación de referencia científica"},

{"Barro Colorado", "EST003", "Isla Barro Colorado", "Protegida", "Área de investigación", 9.1500, -79.8500, 20, "Estación de referencia científica"},

{"EST-BCI", "EST003", "Isla Barro Colorado", "Protegida", "Área de investigación", 9.1500, -79.8500, 20, "Estación de referencia científica"},

{"Isla BCI", "EST003", "Isla Barro Colorado", "Protegida", "Área de investigación", 9.1500, -79.8500, 20, "Estación de referencia científica"}

}

)

### Paso 10: Aplicar Estandarizaciones

#### 10.1 Estandarizar Nombres de Estaciones

**En Editor de Power Query (tabla de datos Bloque 2):**

1. **Crear merge con glosario de estaciones:**

Inicio > Combinar consultas > Combinar como nueva

Tabla izquierda: datos\_bloque2\_terminologia

Columna izquierda: ID

Tabla derecha: Glosario\_Estaciones

Columna derecha: Codigo\_Original

Tipo de combinación: Combinación externa izquierda

1. **Expandir columnas del glosario:**
   * Seleccionar columnas: Codigo\_Estandar, Nombre\_Completo, Zona\_Embalse
   * Desmarcar "Usar nombre de columna original como prefijo"
2. **Crear columna de estación estandarizada:**

Agregar columna > Columna personalizada

Nombre: Estacion\_Estandarizada

Fórmula: if [Codigo\_Estandar] <> null then [Codigo\_Estandar] else [ID]

#### 10.2 Estandarizar Clasificaciones de Calidad

1. **Crear tabla de correspondencias de calidad:**

Agregar columna > Columna personalizada

Nombre: Calidad\_Estandarizada

Fórmula:

if Text.Upper([Calidad\_Agua]) = "EXCELENTE" or Text.Upper([Calidad\_Agua]) = "EXC" then "Excelente"

else if Text.Upper([Calidad\_Agua]) = "BUENA" or Text.Upper([Calidad\_Agua]) = "B" then "Buena"

else if Text.Upper([Calidad\_Agua]) = "REGULAR" or Text.Upper([Calidad\_Agua]) = "REG" then "Regular"

else if Text.Upper([Calidad\_Agua]) = "DEFICIENTE" or Text.Upper([Calidad\_Agua]) = "DEF" or Text.Upper([Calidad\_Agua]) = "MALA" then "Deficiente"

else "Sin clasificar"

1. **Crear códigos numéricos para calidad:**

Agregar columna > Columna personalizada

Nombre: Calidad\_Codigo

Fórmula:

if [Calidad\_Estandarizada] = "Excelente" then 4

else if [Calidad\_Estandarizada] = "Buena" then 3

else if [Calidad\_Estandarizada] = "Regular" then 2

else if [Calidad\_Estandarizada] = "Deficiente" then 1

else 0

#### 10.3 Estandarizar Tipos de Parámetros

Agregar columna > Columna personalizada

Nombre: Tipo\_Parametro\_Estandarizado

Fórmula:

if Text.Contains(Text.Upper([Tipo\_Parametro]), "FISICO") then "Fisicoquímico"

else if Text.Contains(Text.Upper([Tipo\_Parametro]), "BIO") then "Biológico"

else if Text.Contains(Text.Upper([Tipo\_Parametro]), "BACTER") then "Bacteriológico"

else "Sin clasificar"

### Paso 11: Implementar Sistema de Alertas

#### 11.1 Crear Medidas de Alerta por Parámetro

**En Vista de datos, crear medidas DAX:**

Alerta\_pH =

VAR pHPromedio = AVERAGE('datos\_bloque2\_terminologia'[pH])

RETURN

IF(

pHPromedio < 6.5 || pHPromedio > 8.5,

"⚠️ pH fuera de rango óptimo",

"✅ pH en rango normal"

)

Alerta\_Oxigeno =

VAR ODPromedio = AVERAGE('datos\_bloque2\_terminologia'[Oxigeno\_Disuelto])

RETURN

IF(

ODPromedio < 5,

"🚨 Oxígeno disuelto bajo - Riesgo para vida acuática",

IF(

ODPromedio < 7,

"⚠️ Oxígeno disuelto moderado",

"✅ Oxígeno disuelto adecuado"

)

)

Alerta\_Eutrofizacion =

VAR ClorofilaPromedio = AVERAGE('datos\_bloque2\_terminologia'[Clorofila\_A])

VAR FosforoPromedio = AVERAGE('datos\_bloque2\_terminologia'[Fosfatos\_PO4])

RETURN

IF(

ClorofilaPromedio > 20 || FosforoPromedio > 0.05,

"🚨 Riesgo alto de eutrofización",

IF(

ClorofilaPromedio > 10 || FosforoPromedio > 0.025,

"⚠️ Riesgo moderado de eutrofización",

"✅ Bajo riesgo de eutrofización"

)

)

#### 11.2 Crear Medida de Alerta General

Alertas\_Generales =

VAR AlertapH = [Alerta\_pH]

VAR AlertaOD = [Alerta\_Oxigeno]

VAR AlertaEutro = [Alerta\_Eutrofizacion]

RETURN

AlertapH & UNICHAR(10) & AlertaOD & UNICHAR(10) & AlertaEutro

### Paso 12: Crear Clasificación Automática

#### 12.1 Implementar Clasificación Trófica

Estado\_Trofico =

VAR ClorofilaPromedio = AVERAGE('datos\_bloque2\_terminologia'[Clorofila\_A])

VAR TransparenciaPromedio = AVERAGE('datos\_bloque2\_terminologia'[Transparencia])

VAR FosforoPromedio = AVERAGE('datos\_bloque2\_terminologia'[Fosfatos\_PO4])

VAR PuntuacionClorofila =

IF(ClorofilaPromedio <= 2.6, 1,

IF(ClorofilaPromedio <= 20, 2,

IF(ClorofilaPromedio <= 56, 3, 4)))

VAR PuntuacionTransparencia =

IF(TransparenciaPromedio > 4, 1,

IF(TransparenciaPromedio >= 2, 2,

IF(TransparenciaPromedio >= 0.5, 3, 4)))

VAR PuntuacionFosforo =

IF(FosforoPromedio <= 0.012, 1,

IF(FosforoPromedio <= 0.024, 2,

IF(FosforoPromedio <= 0.096, 3, 4)))

VAR PuntuacionPromedio = (PuntuacionClorofila + PuntuacionTransparencia + PuntuacionFosforo) / 3

RETURN

SWITCH(

TRUE(),

PuntuacionPromedio <= 1.5, "Oligotrófico",

PuntuacionPromedio <= 2.5, "Mesotrófico",

PuntuacionPromedio <= 3.5, "Eutrófico",

"Hipereutrófico"

)

#### 12.2 Crear Índice de Calidad del Agua

ICA\_Puntuacion =

VAR pHPuntos =

VAR pHValor = AVERAGE('datos\_bloque2\_terminologia'[pH])

RETURN

SWITCH(

TRUE(),

pHValor >= 6.5 && pHValor <= 8.5, 100,

pHValor >= 6.0 && pHValor < 6.5, 80,

pHValor > 8.5 && pHValor <= 9.0, 80,

pHValor >= 5.5 && pHValor < 6.0, 60,

pHValor > 9.0 && pHValor <= 9.5, 60,

40

)

VAR ODPuntos =

VAR ODValor = AVERAGE('datos\_bloque2\_terminologia'[Oxigeno\_Disuelto])

RETURN

SWITCH(

TRUE(),

ODValor >= 7, 100,

ODValor >= 5, 80,

ODValor >= 3, 60,

ODValor >= 2, 40,

20

)

VAR TurbPuntos =

VAR TurbValor = AVERAGE('datos\_bloque2\_terminologia'[Turbidez])

RETURN

SWITCH(

TRUE(),

TurbValor <= 1, 100,

TurbValor <= 4, 80,

TurbValor <= 10, 60,

TurbValor <= 25, 40,

20

)

VAR EcoliPuntos =

VAR EcoliValor = AVERAGE('datos\_bloque2\_terminologia'[E\_coli])

RETURN

SWITCH(

TRUE(),

EcoliValor <= 100, 100,

EcoliValor <= 200, 80,

EcoliValor <= 500, 60,

EcoliValor <= 1000, 40,

20

)

RETURN (pHPuntos \* 0.25 + ODPuntos \* 0.35 + TurbPuntos \* 0.25 + EcoliPuntos \* 0.15)

## BLOQUE 3: Validación y Análisis Avanzado (1 hora)

### Paso 13: Importar Datos para Validación

#### 13.1 Importar Tercer Archivo

1. **Obtener datos > Excel**
   * Seleccionar: datos\_bloque3\_validacion.xlsx
   * Abrir en Editor de Power Query
2. **Examinar problemas introducidos:**
   * Valores de pH fuera de rango
   * Temperaturas imposibles
   * Relaciones inconsistentes entre parámetros
   * Valores negativos donde no son posibles
   * Outliers extremos

### Paso 14: Implementar Validaciones Cruzadas

#### 14.1 Validación Conductividad-TDS

**En Editor de Power Query:**

Agregar columna > Columna personalizada

Nombre: TDS\_Estimado

Fórmula: [Conductividad] \* 0.64

Agregar columna > Columna personalizada

Nombre: Validacion\_Cond\_TDS

Fórmula:

let

Diferencia = Number.Abs([Solidos\_Disueltos\_Totales] - [TDS\_Estimado]),

PorcentajeError = Diferencia / [Solidos\_Disueltos\_Totales]

in

if PorcentajeError <= 0.2 then "Consistente"

else if PorcentajeError <= 0.5 then "Revisar"

else "Inconsistente"

#### 14.2 Validación Temperatura-Oxígeno

Agregar columna > Columna personalizada

Nombre: DO\_Saturacion\_Teorica

Fórmula:

let

T = [Temperatura]

in

14.652 - 0.41022 \* T + 0.007991 \* Number.Power(T, 2) - 0.000077774 \* Number.Power(T, 3)

Agregar columna > Columna personalizada

Nombre: Validacion\_Temp\_DO

Fórmula:

let

SaturacionReal = ([Oxigeno\_Disuelto] / [DO\_Saturacion\_Teorica]) \* 100

in

if SaturacionReal >= 80 and SaturacionReal <= 120 then "Normal"

else if SaturacionReal < 80 then "Subsaturado"

else "Sobresaturado"

#### 14.3 Validación Transparencia-Turbidez

Agregar columna > Columna personalizada

Nombre: Validacion\_Transp\_Turb

Fórmula:

let

TransparenciaAlta = [Transparencia] > 3,

TurbidezBaja = [Turbidez] < 2

in

if (TransparenciaAlta and TurbidezBaja) or (not TransparenciaAlta and not TurbidezBaja) then "Consistente"

else "Inconsistente"

### Paso 15: Detección de Outliers

#### 15.1 Calcular Z-Scores

**Para cada parámetro numérico clave:**

Agregar columna > Columna personalizada

Nombre: Z\_Score\_pH

Fórmula:

let

Media = List.Average(Table.Column(#"Paso anterior", "pH")),

DesviacionEstandar = List.StandardDeviation(Table.Column(#"Paso anterior", "pH"))

in

Number.Abs(([pH] - Media) / DesviacionEstandar)

**Repetir para otros parámetros:**

* Z\_Score\_Temperatura
* Z\_Score\_Oxigeno\_Disuelto
* Z\_Score\_Conductividad
* Z\_Score\_Clorofila\_A

#### 15.2 Identificar Outliers Multivariados

Agregar columna > Columna personalizada

Nombre: Outlier\_Multivariado

Fórmula:

if [Z\_Score\_pH] > 2.5 or [Z\_Score\_Temperatura] > 2.5 or [Z\_Score\_Oxigeno\_Disuelto] > 2.5 or [Z\_Score\_Conductividad] > 2.5 then "Outlier"

else if [Z\_Score\_pH] > 2 or [Z\_Score\_Temperatura] > 2 or [Z\_Score\_Oxigeno\_Disuelto] > 2 or [Z\_Score\_Conductividad] > 2 then "Sospechoso"

else "Normal"

### Paso 16: Análisis de Correlaciones

#### 16.1 Crear Medidas de Correlación

**En Vista de datos, crear medidas DAX:**

Correlacion\_Cond\_TDS =

VAR TablaCorr =

ADDCOLUMNS(

'datos\_bloque3\_validacion',

"X", [Conductividad],

"Y", [Solidos\_Disueltos\_Totales]

)

VAR SumX = SUMX(TablaCorr, [X])

VAR SumY = SUMX(TablaCorr, [Y])

VAR SumXY = SUMX(TablaCorr, [X] \* [Y])

VAR SumX2 = SUMX(TablaCorr, [X]^2)

VAR SumY2 = SUMX(TablaCorr, [Y]^2)

VAR N = COUNTROWS(TablaCorr)

VAR Numerador = N \* SumXY - SumX \* SumY

VAR DenominadorX = N \* SumX2 - SumX^2

VAR DenominadorY = N \* SumY2 - SumY^2

RETURN

IF(DenominadorX > 0 && DenominadorY > 0,

Numerador / SQRT(DenominadorX \* DenominadorY),

BLANK())

Correlacion\_Temp\_DO =

VAR TablaCorr =

ADDCOLUMNS(

'datos\_bloque3\_validacion',

"X", [Temperatura],

"Y", [Oxigeno\_Disuelto]

)

VAR SumX = SUMX(TablaCorr, [X])

VAR SumY = SUMX(TablaCorr, [Y])

VAR SumXY = SUMX(TablaCorr, [X] \* [Y])

VAR SumX2 = SUMX(TablaCorr, [X]^2)

VAR SumY2 = SUMX(TablaCorr, [Y]^2)

VAR N = COUNTROWS(TablaCorr)

VAR Numerador = N \* SumXY - SumX \* SumY

VAR DenominadorX = N \* SumX2 - SumX^2

VAR DenominadorY = N \* SumY2 - SumY^2

RETURN

IF(DenominadorX > 0 && DenominadorY > 0,

Numerador / SQRT(DenominadorX \* DenominadorY),

BLANK())

#### 16.2 Interpretar Correlaciones

Interpretacion\_Correlaciones =

VAR CorrCondTDS = [Correlacion\_Cond\_TDS]

VAR CorrTempDO = [Correlacion\_Temp\_DO]

RETURN

"Conductividad-TDS: " &

IF(CorrCondTDS > 0.8, "Fuerte positiva",

IF(CorrCondTDS > 0.5, "Moderada positiva", "Débil")) &

UNICHAR(10) &

"Temperatura-OD: " &

IF(CorrTempDO < -0.5, "Moderada negativa (esperada)",

IF(CorrTempDO < -0.3, "Débil negativa", "Anómala"))

### Paso 17: Sistema de Puntuación de Calidad de Datos

#### 17.1 Crear Medida de Calidad de Datos

Puntuacion\_Calidad\_Datos =

VAR CompletitudPuntos =

VAR CamposCompletos =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

NOT ISBLANK([pH]) &&

NOT ISBLANK([Temperatura]) &&

NOT ISBLANK([Oxigeno\_Disuelto]) &&

NOT ISBLANK([Conductividad])

)

)

VAR TotalRegistros = COUNTROWS('datos\_bloque3\_validacion')

RETURN (CamposCompletos / TotalRegistros) \* 100

VAR ConsistenciaPuntos =

VAR RegistrosConsistentes =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

[Validacion\_Cond\_TDS] = "Consistente" &&

[Validacion\_Temp\_DO] = "Normal" &&

[Validacion\_Transp\_Turb] = "Consistente"

)

)

VAR TotalRegistros = COUNTROWS('datos\_bloque3\_validacion')

RETURN (RegistrosConsistentes / TotalRegistros) \* 100

VAR RangoPuntos =

VAR RegistrosEnRango =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

[pH] >= 6 && [pH] <= 9 &&

[Temperatura] >= 20 && [Temperatura] <= 35 &&

[Oxigeno\_Disuelto] >= 0

)

)

VAR TotalRegistros = COUNTROWS('datos\_bloque3\_validacion')

RETURN (RegistrosEnRango / TotalRegistros) \* 100

VAR OutlierPuntos =

VAR RegistrosSinOutliers =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

[Outlier\_Multivariado] = "Normal"

)

)

VAR TotalRegistros = COUNTROWS('datos\_bloque3\_validacion')

RETURN (RegistrosSinOutliers / TotalRegistros) \* 100

RETURN (CompletitudPuntos \* 0.3 + ConsistenciaPuntos \* 0.3 + RangoPuntos \* 0.25 + OutlierPuntos \* 0.15)

#### 17.2 Crear Clasificación de Calidad

Clasificacion\_Calidad\_Datos =

VAR Puntuacion = [Puntuacion\_Calidad\_Datos]

RETURN

SWITCH(

TRUE(),

Puntuacion >= 90, "Excelente",

Puntuacion >= 80, "Buena",

Puntuacion >= 70, "Aceptable",

Puntuacion >= 60, "Regular",

"Deficiente"

)

### Paso 18: Crear Reportes de Validación

#### 18.1 Tabla de Registros Problemáticos

Registros\_Problematicos =

FILTER(

'datos\_bloque3\_validacion',

[Outlier\_Multivariado] = "Outlier" ||

[Validacion\_Cond\_TDS] = "Inconsistente" ||

[Validacion\_Temp\_DO] <> "Normal" ||

[pH] < 6 || [pH] > 9 ||

[Temperatura] < 20 || [Temperatura] > 35 ||

[Oxigeno\_Disuelto] < 0

)

#### 18.2 Resumen de Problemas por Tipo

Resumen\_Problemas =

VAR ProblemasRango =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

[pH] < 6 || [pH] > 9 ||

[Temperatura] < 20 || [Temperatura] > 35 ||

[Oxigeno\_Disuelto] < 0

)

)

VAR ProblemasConsistencia =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

[Validacion\_Cond\_TDS] = "Inconsistente" ||

[Validacion\_Temp\_DO] <> "Normal" ||

[Validacion\_Transp\_Turb] = "Inconsistente"

)

)

VAR ProblemasOutliers =

COUNTROWS(

FILTER(

'datos\_bloque3\_validacion',

[Outlier\_Multivariado] = "Outlier"

)

)

RETURN

"Problemas de rango: " & ProblemasRango & UNICHAR(10) &

"Problemas de consistencia: " & ProblemasConsistencia & UNICHAR(10) &

"Outliers detectados: " & ProblemasOutliers

## Paso 19: Crear Visualizaciones y Dashboard

### 19.1 Configurar Páginas del Reporte

**Crear las siguientes páginas:**

1. **Dashboard Principal** - Vista general de calidad
2. **Análisis por Estación** - Comparación entre estaciones
3. **Validación de Datos** - Problemas y outliers
4. **Correlaciones** - Relaciones entre parámetros
5. **Alertas y Monitoreo** - Sistema de alertas

### 19.2 Dashboard Principal

**Elementos a incluir:**

1. **Tarjetas de KPI:**
   * ICA Puntuación promedio
   * Estado trófico predominante
   * Número de estaciones monitoreadas
   * Puntuación de calidad de datos
2. **Gráfico de barras:**
   * ICA por estación
   * Filtros: Zona del embalse
3. **Mapa de calor:**
   * Parámetros vs Estaciones
   * Colores según cumplimiento de estándares
4. **Gráfico de líneas:**
   * Tendencias de parámetros clave (si hay datos temporales)

### 19.3 Página de Validación

**Elementos a incluir:**

1. **Tabla de registros problemáticos:**
   * Mostrar todos los registros con problemas
   * Columnas: Estación, Parámetro, Valor, Problema detectado
2. **Gráficos de dispersión:**
   * Conductividad vs TDS (con línea de regresión esperada)
   * Temperatura vs Oxígeno disuelto
3. **Histogramas:**
   * Distribución de Z-scores por parámetro
   * Identificación visual de outliers

### 19.4 Configurar Filtros y Segmentaciones

**Filtros globales:**

* Estación
* Zona del embalse
* Tipo de parámetro
* Clasificación de calidad

**Filtros específicos por página:**

* Página de validación: Tipo de problema
* Página de correlaciones: Parámetros a correlacionar

## Paso 20: Documentación y Entrega

### 20.1 Crear Documentación Integrada

**Página de documentación en Power BI:**

1. **Cuadro de texto con metodología:**

METODOLOGÍA DE PROCESAMIENTO

Bloque 1: Limpieza y estandarización

- Corrección de 24 nombres de columnas

- Validación de tipos de datos

- Implementación de controles básicos

Bloque 2: Estandarización terminológica

- Aplicación de glosarios específicos

- Clasificación automática de parámetros

- Sistema de alertas por umbrales

Bloque 3: Validación avanzada

- Validaciones cruzadas entre parámetros

- Detección de outliers multivariados

- Cálculo de índices de calidad compuestos

1. **Tabla de transformaciones aplicadas:**
   * Log de todos los cambios realizados
   * Justificación científica de cada decisión
2. **Referencias y estándares utilizados:**
   * OMS, EPA, OECD
   * Metodologías de análisis aplicadas

### 20.2 Verificación Final

**Checklist de verificación:**

* ✅ Todos los datos están cargados correctamente
* ✅ Las relaciones entre tablas están establecidas
* ✅ Las medidas DAX funcionan sin errores
* ✅ Las visualizaciones muestran datos coherentes
* ✅ Los filtros funcionan apropiadamente
* ✅ La documentación está completa
* ✅ El archivo se guarda correctamente

### 20.3 Exportar Resultados

**Opciones de exportación:**

1. **Archivo Power BI (.pbix):**
   * Guardar versión final del proyecto
2. **Reportes en PDF:**
   * Exportar cada página del dashboard
3. **Datos procesados:**
   * Exportar tablas limpias a Excel para uso futuro
4. **Documentación técnica:**
   * Crear documento Word con metodología completa

## Preguntas de Análisis Estadístico Adicional

### Análisis Descriptivo

#### Pregunta 1: Caracterización del Embalse

**¿Cuáles son las características fisicoquímicas predominantes del Embalse Gatún?**

**Análisis a realizar:**

Estadisticas\_Descriptivas =

SUMMARIZE(

'datos\_procesados',

"Parámetro", "Valor",

"Media", AVERAGE('datos\_procesados'[Valor]),

"Mediana", MEDIAN('datos\_procesados'[Valor]),

"Desviación\_Estándar", STDEV.P('datos\_procesados'[Valor]),

"Mínimo", MIN('datos\_procesados'[Valor]),

"Máximo", MAX('datos\_procesados'[Valor]),

"Coeficiente\_Variación", DIVIDE(STDEV.P('datos\_procesados'[Valor]), AVERAGE('datos\_procesados'[Valor])) \* 100

)

#### Pregunta 2: Variabilidad Espacial

**¿Existe variabilidad significativa entre las diferentes zonas del embalse?**

**Análisis ANOVA simplificado:**

F\_Estadistico\_pH =

VAR SSB = SUMX(

VALUES('datos\_procesados'[Zona\_Embalse]),

VAR ZonaActual = 'datos\_procesados'[Zona\_Embalse]

VAR MediaZona = CALCULATE(AVERAGE('datos\_procesados'[pH]), 'datos\_procesados'[Zona\_Embalse] = ZonaActual)

VAR MediaGeneral = AVERAGE(ALL('datos\_procesados'[pH]))

VAR N\_Zona = CALCULATE(COUNTROWS('datos\_procesados'), 'datos\_procesados'[Zona\_Embalse] = ZonaActual)

RETURN N\_Zona \* POWER(MediaZona - MediaGeneral, 2)

)

VAR SSW = SUMX(

'datos\_procesados',

VAR ZonaActual = 'datos\_procesados'[Zona\_Embalse]

VAR MediaZona = CALCULATE(AVERAGE('datos\_procesados'[pH]), 'datos\_procesados'[Zona\_Embalse] = ZonaActual)

RETURN POWER('datos\_procesados'[pH] - MediaZona, 2)

)

VAR df\_between = DISTINCTCOUNT('datos\_procesados'[Zona\_Embalse]) - 1

VAR df\_within = COUNTROWS('datos\_procesados') - DISTINCTCOUNT('datos\_procesados'[Zona\_Embalse])

VAR MSB = DIVIDE(SSB, df\_between)

VAR MSW = DIVIDE(SSW, df\_within)

RETURN DIVIDE(MSB, MSW)

### Análisis de Calidad Ambiental

#### Pregunta 3: Estado Trófico del Embalse

**¿Cuál es el estado trófico del embalse y qué factores lo determinan?**

**Índice de Estado Trófico de Carlson:**

IET\_Carlson =

VAR IET\_Clorofila =

VAR CHL = AVERAGE('datos\_procesados'[Clorofila\_A])

RETURN 10 \* (6 - ((0.92 - 0.34 \* LOG10(CHL)) / LOG10(2)))

VAR IET\_Transparencia =

VAR SD = AVERAGE('datos\_procesados'[Transparencia])

RETURN 10 \* (6 - ((1.77 - 0.253 \* LOG10(SD)) / LOG10(2)))

VAR IET\_Fosforo =

VAR TP = AVERAGE('datos\_procesados'[Fosfatos\_PO4]) \* 1000 // Convertir a μg/L

RETURN 10 \* (6 - ((0.42 - 0.218 \* LOG10(TP)) / LOG10(2)))

RETURN (IET\_Clorofila + IET\_Transparencia + IET\_Fosforo) / 3

#### Pregunta 4: Riesgo de Eutrofización

**¿Qué estaciones presentan mayor riesgo de eutrofización?**

**Índice de riesgo multifactorial:**

Indice\_Riesgo\_Eutrofizacion =

VAR RiesgoClorofila =

VAR CHL = AVERAGE('datos\_procesados'[Clorofila\_A])

RETURN

IF(CHL > 56, 4,

IF(CHL > 20, 3,

IF(CHL > 10, 2, 1)))

VAR RiesgoFosforo =

VAR P = AVERAGE('datos\_procesados'[Fosfatos\_PO4])

RETURN

IF(P > 0.096, 4,

IF(P > 0.024, 3,

IF(P > 0.012, 2, 1)))

VAR RiesgoTransparencia =

VAR SD = AVERAGE('datos\_procesados'[Transparencia])

RETURN

IF(SD < 0.5, 4,

IF(SD < 2, 3,

IF(SD < 4, 2, 1)))

VAR RiesgoNitrogeno =

VAR N = AVERAGE('datos\_procesados'[Nitratos\_NO3])

RETURN

IF(N > 10, 4,

IF(N > 5, 3,

IF(N > 2, 2, 1)))

VAR PuntuacionTotal = RiesgoClorofila + RiesgoFosforo + RiesgoTransparencia + RiesgoNitrogeno

RETURN

SWITCH(

TRUE(),

PuntuacionTotal >= 14, "Muy Alto",

PuntuacionTotal >= 11, "Alto",

PuntuacionTotal >= 8, "Moderado",

PuntuacionTotal >= 5, "Bajo",

"Muy Bajo"

)

### Análisis de Contaminación

#### Pregunta 5: Contaminación Bacteriológica

**¿Existe evidencia de contaminación antropogénica?**

**Análisis de indicadores bacteriológicos:**

Indice\_Contaminacion\_Fecal =

VAR Ecoli = AVERAGE('datos\_procesados'[E\_coli])

VAR ColiformesTotales = AVERAGE('datos\_procesados'[Coliformes\_Totales])

VAR Ratio\_Ecoli\_CT = DIVIDE(Ecoli, ColiformesTotales)

VAR ClasificacionEcoli =

IF(Ecoli <= 100, "Excelente",

IF(Ecoli <= 200, "Buena",

IF(Ecoli <= 500, "Regular",

IF(Ecoli <= 1000, "Deficiente", "Muy deficiente"))))

VAR ClasificacionRatio =

IF(Ratio\_Ecoli\_CT <= 0.1, "Contaminación antigua",

IF(Ratio\_Ecoli\_CT <= 0.3, "Contaminación mixta",

"Contaminación reciente"))

RETURN ClasificacionEcoli & " - " & ClasificacionRatio

### Análisis Multivariado

#### Pregunta 6: Factores Principales de Variabilidad

**¿Qué factores explican la mayor variabilidad en la calidad del agua?**

**Análisis de componentes principales simplificado:**

// Estandarización de variables

Variables\_Estandarizadas =

ADDCOLUMNS(

'datos\_procesados',

"pH\_std", DIVIDE([pH] - AVERAGE(ALL('datos\_procesados'[pH])), STDEV.P(ALL('datos\_procesados'[pH]))),

"DO\_std", DIVIDE([Oxigeno\_Disuelto] - AVERAGE(ALL('datos\_procesados'[Oxigeno\_Disuelto])), STDEV.P(ALL('datos\_procesados'[Oxigeno\_Disuelto]))),

"Cond\_std", DIVIDE([Conductividad] - AVERAGE(ALL('datos\_procesados'[Conductividad])), STDEV.P(ALL('datos\_procesados'[Conductividad]))),

"Temp\_std", DIVIDE([Temperatura] - AVERAGE(ALL('datos\_procesados'[Temperatura])), STDEV.P(ALL('datos\_procesados'[Temperatura]))),

"CHL\_std", DIVIDE([Clorofila\_A] - AVERAGE(ALL('datos\_procesados'[Clorofila\_A])), STDEV.P(ALL('datos\_procesados'[Clorofila\_A])))

)

// Componente principal 1 (pesos determinados por análisis previo)

PC1\_Score =

0.45 \* [pH\_std] +

0.40 \* [DO\_std] +

0.35 \* [Cond\_std] +

0.30 \* [Temp\_std] +

(-0.25) \* [CHL\_std] // Negativo porque alta clorofila indica menor calidad

#### Pregunta 7: Agrupación de Estaciones

**¿Es posible identificar grupos homogéneos de estaciones?**

**Análisis de conglomerados simplificado:**

Distancia\_Euclidiana\_Estaciones =

VAR Estacion1 = SELECTEDVALUE('datos\_procesados'[Estacion\_ID])

VAR Estacion2 = [Estacion\_Comparacion] // Parámetro de entrada

VAR Diff\_pH =

CALCULATE(AVERAGE('datos\_procesados'[pH]), 'datos\_procesados'[Estacion\_ID] = Estacion1) -

CALCULATE(AVERAGE('datos\_procesados'[pH]), 'datos\_procesados'[Estacion\_ID] = Estacion2)

VAR Diff\_DO =

CALCULATE(AVERAGE('datos\_procesados'[Oxigeno\_Disuelto]), 'datos\_procesados'[Estacion\_ID] = Estacion1) -

CALCULATE(AVERAGE('datos\_procesados'[Oxigeno\_Disuelto]), 'datos\_procesados'[Estacion\_ID] = Estacion2)

VAR Diff\_Cond =

CALCULATE(AVERAGE('datos\_procesados'[Conductividad]), 'datos\_procesados'[Estacion\_ID] = Estacion1) -

CALCULATE(AVERAGE('datos\_procesados'[Conductividad]), 'datos\_procesados'[Estacion\_ID] = Estacion2)

VAR Diff\_CHL =

CALCULATE(AVERAGE('datos\_procesados'[Clorofila\_A]), 'datos\_procesados'[Estacion\_ID] = Estacion1) -

CALCULATE(AVERAGE('datos\_procesados'[Clorofila\_A]), 'datos\_procesados'[Estacion\_ID] = Estacion2)

// Normalizar diferencias y calcular distancia euclidiana

RETURN SQRT(

POWER(Diff\_pH / 1, 2) + // pH normalizado por rango típico

POWER(Diff\_DO / 5, 2) + // DO normalizado

POWER(Diff\_Cond / 200, 2) + // Conductividad normalizada

POWER(Diff\_CHL / 10, 2) // Clorofila normalizada

)

### Análisis de Cumplimiento Regulatorio

#### Pregunta 8: Cumplimiento de Estándares

**¿Qué porcentaje de las muestras cumple con los estándares internacionales?**

**Análisis de cumplimiento multi-criterio:**

Porcentaje\_Cumplimiento\_General =

VAR RegistrosCumplen =

COUNTROWS(

FILTER(

'datos\_procesados',

[pH] >= 6.5 && [pH] <= 8.5 &&

[Oxigeno\_Disuelto] >= 5 &&

[Turbidez] <= 4 &&

[E\_coli] <= 100 &&

[Coliformes\_Totales] <= 1000 &&

[Nitratos\_NO3] <= 10 &&

[Fosfatos\_PO4] <= 0.05

)

)

VAR TotalRegistros = COUNTROWS('datos\_procesados')

RETURN DIVIDE(RegistrosCumplen, TotalRegistros) \* 100

**Cumplimiento por parámetro individual:**

Cumplimiento\_por\_Parametro =

UNION(

ROW("Parametro", "pH", "Cumplimiento\_%",

DIVIDE(COUNTROWS(FILTER('datos\_procesados', [pH] >= 6.5 && [pH] <= 8.5)), COUNTROWS('datos\_procesados')) \* 100),

ROW("Parametro", "Oxígeno Disuelto", "Cumplimiento\_%",

DIVIDE(COUNTROWS(FILTER('datos\_procesados', [Oxigeno\_Disuelto] >= 5)), COUNTROWS('datos\_procesados')) \* 100),

ROW("Parametro", "Turbidez", "Cumplimiento\_%",

DIVIDE(COUNTROWS(FILTER('datos\_procesados', [Turbidez] <= 4)), COUNTROWS('datos\_procesados')) \* 100),

ROW("Parametro", "E. coli", "Cumplimiento\_%",

DIVIDE(COUNTROWS(FILTER('datos\_procesados', [E\_coli] <= 100)), COUNTROWS('datos\_procesados')) \* 100)

)

## Conclusiones y Próximos Pasos

### Resultados Esperados del Ejercicio

Al completar este tutorial, los participantes habrán:

1. **Dominado técnicas de limpieza de datos** en Power BI
2. **Implementado sistemas de estandarización** basados en glosarios
3. **Desarrollado validaciones automáticas** para control de calidad
4. **Creado dashboards interactivos** para monitoreo ambiental
5. **Aplicado análisis estadísticos** a datos ambientales reales

### Competencias Desarrolladas

* **Técnicas:** Power Query, DAX, modelado de datos
* **Científicas:** Interpretación de parámetros de calidad del agua
* **Analíticas:** Detección de outliers, análisis de correlaciones
* **Comunicativas:** Visualización de datos ambientales

### Aplicaciones Futuras

Este enfoque puede ser aplicado a:

* Otros embalses y cuerpos de agua
* Diferentes tipos de monitoreo ambiental
* Sistemas de alerta temprana
* Reportes regulatorios automatizados

*Este tutorial proporciona una guía completa para la implementación práctica del ejercicio de gestión de datos de calidad del agua, combinando rigor científico con herramientas tecnológicas accesibles.*