**🧪 ¿Cómo harías la fusión de FTIR + Raman a nivel bajo?**

1. **Preprocesar** ambos espectros:
   * Corrección de línea base.
   * Normalización (muy importante para que los valores no estén en escalas muy distintas).
   * Suavizado si hace falta.
2. **Alinear los ejes X** si es necesario:
   * Raman Shift [cm⁻¹] y FTIR [cm⁻¹] deberían coincidir o interpolarse a un mismo rango común.
3. **Concatenar** los vectores de intensidades:
   * Para cada muestra: [FTIR | Raman] → un vector largo.
4. **Construir un DataFrame de espectros fusionados**.
5. **Aplicar análisis**:
   * PCA, HCA, KMeans, DBSCAN, Clasificación, etc.

| **Nivel** | **¿Qué implica?** | **Dificultad** | **Recomendado para vos** |
| --- | --- | --- | --- |
| **Nivel bajo** | Fusionar directamente las señales (concatenar los espectros). | Fácil | ✅ Ideal para empezar |
| **Nivel medio** | Fusionar después de extracción de características (por ejemplo, componentes PCA, picos, áreas). | Medio | Opcional más adelante |
| **Nivel alto** | Fusionar decisiones de modelos separados (por ejemplo, votación de clasificadores). | Avanzado | Más adelante |

**🧠 Tips muy importantes:**

* **Normalizá siempre** los datos antes de fusionar.
* **Interpolá** si los espectros no tienen la misma resolución de ejes.
* **Mantené el orden correcto de las muestras** entre FTIR y Raman.
* **Guardá siempre** los vectores fusionados aparte para reproducibilidad.
* **Documentá** qué preprocesamiento hiciste en cada técnica.

**🚀 ¿Cómo seguir después de la fusión?**

1. **Aplicar PCA** al DataFrame fusionado para ver si la separación mejora.
2. **Aplicar HCA o KMeans** sobre los vectores fusionados.
3. **Comparar** resultados contra usar solo FTIR o solo Raman.

| **Paso** | **Acción** |
| --- | --- |
| 1. | Leer FTIR y Raman. |
| 2. | Normalizar intensidades. |
| 3. | Alinear o interpolar los ejes si es necesario. |
| 4. | Concatenar espectros. |
| 5. | Aplicar PCA, HCA, KMeans, etc. |

|  |  |
| --- | --- |
| AnalGesicos.xlsx | Probablemente espectros **FTIR** de analgésicos |
| allspectra.xlsx | Probablemente espectros **Raman** (por nombre "allspectra") |

🔵 FTIR: mide absorción (cambios en dipolo).  
🟠 Raman: mide dispersión (cambios en polarizabilidad).

✅ Son **complementarios**, **ideales para Data Fusion**.

**Primer Paso: ¿Qué vamos a hacer?**

1. **Inspeccionar ambos archivos**:
   * ¿Qué columnas tienen?
   * ¿Qué son las filas?
   * ¿El eje X (wavenumber/cm⁻¹) está en la primera columna?
2. **Verificar compatibilidad**:
   * ¿Tienen mismo tipo de muestras? (nombres similares, o podemos mapear).
   * ¿Tienen el mismo rango de ejes? (o interpolamos).
3. **Preparar normalización**:
   * FTIR y Raman deben tener intensidades **normalizadas** para no darle peso injusto a uno.
4. **Fusionar**:
   * Para cada muestra, unir FTIR + Raman.

# Cosas importantes a analizar en tus datos:

| **Pregunta** | **¿Por qué importa?** |
| --- | --- |
| ¿Los espectros están como columnas o filas? | Hay que alinear |
| ¿Primer columna es Raman Shift o Wavenumber? | Necesitamos saber el eje |
| ¿Muestras coinciden en nombre? | Para fusionar correctamente |
| ¿Mismos puntos de eje? | Si no, interpolamos |

**¿Qué haremos después?**

✅ Si el **eje X coincide**: fusionamos directamente.  
✅ Si **no coincide**: hacemos una **interpolación** para hacerlos compatibles.

Luego:

1. **Normalizamos** espectros FTIR y Raman (por Min-Max o Z-Score).
2. **Fusionamos** concatenando vectores.
3. **Guardamos** un DataFrame final listo para PCA, HCA, KMeans.

# Resumen para tu flujo:

| **Etapa** | **Acción** |
| --- | --- |
| 1. Leer FTIR y Raman | pandas.read\_excel |
| 2. Inspeccionar estructura | .head() y .columns |
| 3. Ajustar formato | Transponer o interpolar si hace falta |
| 4. Normalizar intensidades | Escalar datos |
| 5. Fusionar | Concatenar espectros |
| 6. Analizar | PCA, HCA, Clustering |

el centímetro recíproco, cm−1, es una unidad de energía igual a la energía de un fotón con una longitud de onda de 1 cm. Esa energía asciende a aproximadamente 1.24×10−4 eV o 1.986×10−23 J.

acetylsalicylic-acid esta relacionado con la aspirina, es su nombre quimico

Pero si en tus espectros "zmol" aparece junto con "aspirina", "acetylsalicylic-acid" o compuestos relacionados, **puede estar representando una muestra basada en aspirina** o un derivado. Por ejemplo:

### Errores comunes en datafusion

### 🧨 ¿Por qué pasa?

La causa más común es que:

* Una o más columnas en df **tienen valores nulos (NaN)**, o
* El eje X fue convertido mal (ej. por tener strings vacíos o puntos no numéricos),
* O hubo un recorte/desfase accidental en la longitud.

Tipo de interpolacion que admite la librería que estoy usandio

|  | **Tipo de interpolación** |
| --- | --- |
| 'linear' | ✔️ actual, rectas entre puntos |
| 'nearest' | valor más cercano |
| 'cubic' | spline cúbico (suave pero más lento) |
| 'quadratic' | spline cuadrático |
| 'slinear' | spline lineal segmentado |

## 🧪 ¿Qué se recomienda en ciencia de datos espectrales?

| **Escenario** | **Recomendación** |
| --- | --- |
| Rango parcialmente común | Interpolar sobre el rango común |
| Rango completamente distinto | Concatenar sin interpolar ✅ |
| Visualización combinada | Mostrar cada espectro con su eje |
| Modelos de clasificación o clustering | Concatenar espectros interpolados individualmente |