**PASOS CORRELACIÓN DE EEG**

**Preparación de datos:**

Lee todos los archivos parquet de la carpeta y combínalos en un solo DataFrame utilizando, por ejemplo, la biblioteca pandas de Python.

Asegúrate de que los datos estén limpios y consistentes. Esto incluye tratar los valores perdidos o atípicos, así como verificar que las columnas tengan el mismo nombre y tipo de datos en todos los archivos.

**Análisis exploratorio:**

Realiza un análisis exploratorio de los datos para tener una idea general de las características y las distribuciones de los datos. Esto puede incluir estadísticas descriptivas, gráficos de distribución, etc.

**Extracción de características:**

Para simplificar el análisis y mejorar la eficiencia computacional, puedes extraer características relevantes de los datos brutos del EEG, como la potencia espectral en diferentes bandas de frecuencia (delta, theta, alfa, beta y gamma), la coherencia entre pares de electrodos, etc.

**Reducción de dimensionalidad y visualización:**

Utiliza técnicas de reducción de dimensionalidad como PCA (Análisis de componentes principales) o t-SNE para visualizar los datos en 2D o 3D y observar si existen patrones o agrupaciones en los datos de pensamientos positivos.

**Clustering:**

Aplica algoritmos de clustering, como K-means, DBSCAN o HDBSCAN, para agrupar los datos en función de las similitudes en las características extraídas. Esto te permitirá identificar patrones comunes en los datos de pensamientos positivos.

**Correlaciones:**

Para examinar las correlaciones entre las características extraídas y entre los sujetos, puedes calcular la matriz de correlación utilizando la correlación de Pearson, Spearman o Kendall. Visualiza esta matriz utilizando un mapa de calor.

1. **Concatenación y correlación de Pearson:** Puedes concatenar los datos de todos los archivos manteniendo los canales en el mismo orden y luego calcular el coeficiente de correlación de Pearson entre los canales. Este coeficiente mide la relación lineal entre dos variables y se encuentra entre -1 y 1. Es uno de los métodos más utilizados para evaluar la correlación entre dos señales.
2. **Coherencia:** La coherencia es una medida de la correlación entre dos señales en el dominio de la frecuencia. Puedes calcular la coherencia para cada par de canales y luego promediarlas para obtener una medida global de la correlación entre los archivos.
3. **Sincronización de fase:** La sincronización de fase es una medida de la similitud entre las fases de las señales de EEG en diferentes canales. Puedes calcular la sincronización de fase para cada par de canales y luego promediarlas para obtener una medida global de la correlación.
4. **Análisis de componentes principales (PCA):** Realiza un PCA en los datos de EEG combinados de todos los archivos. Luego, calcula la correlación entre los componentes principales y los canales originales para evaluar la correlación entre los datos de EEG.
5. **Análisis de componentes independientes (ICA):** Similar al PCA, puedes realizar un ICA en los datos de EEG combinados de todos los archivos. Luego, calcula la correlación entre los componentes independientes y los canales originales para evaluar la correlación entre los datos de EEG.

**Guardar el patrón:**

Una vez que hayas identificado un patrón común en los datos, crea un nuevo DataFrame con los datos representativos del patrón de pensamientos positivos y guárdalo en un archivo parquet utilizando la función to\_parquet de pandas.