**Objetivo de Investigación:**

El objetivo de esta investigación es **desarrollar una regresión lineal multiple para predecir las calorías quemadas** durante las sesiones de ejercicio de los miembros de un gimnasio, utilizando el **Gym Members Exercise Dataset**. Este modelo predictivo buscará identificar patrones complejos y relaciones no lineales entre diversas características físicas, demográficas y de entrenamiento (como edad, peso, frecuencia cardíaca, tipo de ejercicio, duración de la sesión, entre otras) y el gasto calórico.

El objetivo específico es generar un modelo de regresión lineal que pueda predecir con alta precisión la cantidad de calorías quemadas en función de estas variables, permitiendo así la personalización de las rutinas de ejercicio y la mejora en la planificación de programas de entrenamiento, basados en un análisis más preciso del rendimiento físico de los miembros del gimnasio.

**Preprocesamiento de Datos**

En el código proporcionado, se llevaron a cabo diversas etapas de preprocesamiento para preparar los datos antes de entrenar el modelo de Machine Learning:

1. **Lectura de los datos**: El conjunto de datos se cargó desde un archivo CSV y se inspeccionaron las primeras filas, las estadísticas descriptivas y la información general sobre el dataset utilizando las funciones head(), info() y describe().
2. **Revisión de valores faltantes**: Se realizó un análisis de valores nulos con isnull().sum() para identificar si había columnas con valores faltantes que necesitaran ser eliminados o imputados.
3. **Transformación de variables categóricas**: Las columnas Gender y Workout\_Type se transformaron en variables binarias utilizando get\_dummies(). Esto es necesario para que los modelos de Machine Learning puedan trabajar con datos numéricos en lugar de categóricos.
4. **Escalado de características numéricas**: Se aplicó StandardScaler a las columnas numéricas para asegurar que todas las características estén en la misma escala (media 0 y desviación estándar 1), lo cual es importante para modelos que son sensibles a la escala, como los basados en distancias.
5. **Detección y manejo de valores atípicos (outliers)**: Se identificaron y eliminaron los valores atípicos en la columna BMI utilizando el rango intercuartílico (IQR). Esta técnica garantiza que los datos no contengan valores extremos que puedan afectar el rendimiento del modelo.
6. **División de datos en conjuntos de entrenamiento y prueba**: Los datos fueron divididos en un conjunto de características (X) y la variable objetivo (y), que es Calories\_Burned. A continuación, se dividió el dataset en entrenamiento y prueba usando train\_test\_split(), con un 80% de los datos para entrenamiento y un 20% para prueba, además también se dividió el dataset en entrenamiento y prueba usando 50% de los datos para entrenamiento y un 50% para prueba esto para un enfoque académico e investigación.
7. **Exportación de los conjuntos procesados**: Los conjuntos de entrenamiento y prueba se exportaron a archivos CSV para su posterior uso en la fase de modelado.

**Balanceo de** Datos

En este caso, **no se utilizó ninguna técnica de balanceo de datos** porque el objetivo es predecir una variable continua, Calories\_Burned, lo que implica que estamos trabajando con un problema de **regresión**. El balanceo de datos es comúnmente utilizado en problemas de **clasificación**, donde las clases pueden estar desbalanceadas (por ejemplo, en problemas de clasificación binaria o multiclase). Sin embargo, para un problema de regresión, donde la variable objetivo es continua, las técnicas de balanceo como **oversampling** o **undersampling** no son aplicables ni necesarias.

La variable Calories\_Burned no está dividida en clases discretas, sino que es un valor continuo que puede tomar una amplia gama de valores. Por lo tanto, las técnicas de balanceo de clases no son apropiadas en este caso, ya que no existe una categoría desbalanceada que corregir. En su lugar, el modelo de regresión se ajustará a los valores continuos de la variable objetivo sin necesidad de aplicar balanceo.