

## Ejercicio 2: Evaluación de la Función Pulmonar mediante Regresión Múltiple

### Descripción:

En este ejercicio, se trabajará con un conjunto de datos que simula la función pulmonar de pacientes y su relación con variables fisiológicas y ambientales. Se utilizará regresión múltiple para predecir la **Capacidad Vital Forzada (FVC)**, una medida clave en la evaluación de la salud pulmonar.

### Objetivos:

1. Comprender cómo factores como edad, altura, peso, género, tabaquismo y exposición a contaminantes afectan la función pulmonar.
  2. Aplicar técnicas de preprocesamiento de datos para manejar variables numéricas y categóricas.
  3. Construir y evaluar un modelo de regresión múltiple utilizando **scikit-learn**.
  4. Analizar la importancia de las variables en la predicción de FVC.
  5. Evaluar el desempeño del modelo con métricas adecuadas y análisis de residuos.
- 

### Instrucciones para el desarrollo:

#### 1. Carga y exploración de datos:

- Genera un conjunto de datos con aproximadamente 500 registros, siguiendo distribuciones fisiológicas realistas.
- Identifica las variables numéricas y categóricas.
- Realiza un análisis exploratorio con visualizaciones adecuadas.

#### 2. Preprocesamiento:

- Normaliza las variables numéricas.
- Codifica las variables categóricas utilizando **OneHotEncoder**.
- Separa la variable objetivo (FVC) de las características predictoras.
- Divide los datos en entrenamiento y prueba (por ejemplo, 75%-25%).

#### 3. Entrenamiento del modelo:

- Utiliza **regresión lineal múltiple** para predecir la FVC.

- Implementa un **Pipeline** en **scikit-learn** que incluya el preprocesamiento y el modelo.
4. **Evaluación del modelo:**
- Calcula métricas de desempeño: **MSE**, **RMSE**, **MAE** y **R<sup>2</sup>**.
  - Realiza validación cruzada para evaluar la estabilidad del modelo.
  - Analiza la distribución de residuos y su normalidad.
5. **Interpretación de coeficientes:**
- Extrae y analiza la importancia de las variables predictoras en la regresión.
  - Identifica qué factores tienen mayor impacto en la capacidad pulmonar.
6. **Visualización de resultados:**
- Representa gráficamente las relaciones entre FVC y las variables predictoras.
  - Muestra la comparación entre valores reales y predicciones.
7. **Prueba con nuevos datos:**
- Simula un nuevo paciente y utiliza el modelo para predecir su FVC.
  - Interpreta clínicamente el resultado.
- 

### Pistas y Recomendaciones:

- ✓ Utiliza **Seaborn** y **Matplotlib** para explorar las distribuciones y relaciones entre variables.
  - ✓ Recuerda que la **altura** y el **género** tienen un impacto significativo en la función pulmonar.
  - ✓ Evalúa si algunas variables categóricas tienen más influencia que otras en la predicción.
  - ✓ La normalidad de los residuos puede afectar la validez del modelo.
  - ✓ Usa **statsmodels** si quieres obtener valores p y análisis más detallado de los coeficientes.
-