* **Objetivo del análisis:**

Este análisis tiene como objetivo demostrar las capacidades de PowerBI para explorar y visualizar grandes conjuntos de datos en el ámbito de la salud. Utilizando el Dataset de Framingham como caso de estudio, se realizará un análisis exploratorio con el fin de identificar los principales factores de riesgo asociados con las enfermedades cardiovasculares. Los resultados obtenidos permitirán seleccionar dicho Dataset como el óptimo para el posterior estudio riesgo individual de eventos cardiovasculares y, así, contribuir a la prevención y el tratamiento de estas enfermedades.

* **Conjunto de datos:**

Dado el carácter longitudinal y la riqueza de información del Dataset de Framingham, este estudio se propone analizar en profundidad las relaciones entre diversos factores de riesgo y la aparición de enfermedades cardiovasculares. Al seleccionar este conjunto de datos, se busca aprovechar la extensa investigación realizada a lo largo de décadas para obtener resultados robustos y generalizables.

Este Dataset está compuesto por:

1. **Variables Demográficas y Socioeconómicas**

* **age:** Edad del participante en años.
* **age\_groups:** Categorización de la edad en grupos (por ejemplo, jóvenes, adultos medios, ancianos).
* **gender:** Género del participante (masculino o femenino).
* **education:** Nivel educativo alcanzado por el participante.

1. **Factores de Riesgo Cardiovascular**

* **BMI:** Índice de Masa Corporal, una medida de la grasa corporal basada en la altura y el peso.
* **cigsPerDay:** Número promedio de cigarrillos fumados por día.
* **currentSmoker:** Indicador binario (sí/no) de si el participante es fumador activo.
* **diabetes:** Indicador binario (sí/no) de si el participante tiene diabetes.
* **diaBP:** Presión arterial diastólica (presión arterial cuando el corazón está relajado).
* **sysBP:** Presión arterial sistólica (presión arterial cuando el corazón se contrae).
* **totChol:** Nivel total de colesterol en sangre.
* **glucose:** Nivel de glucosa en sangre.
* **heartRate:** Frecuencia cardíaca en reposo.
* **heart\_rate\_groups:** Categorización de la frecuencia cardíaca en grupos (por ejemplo, bradicardia, taquicardia).
* **prevalentHyp:** Indicador binario (sí/no) de si el participante tiene hipertensión (presión arterial alta) al inicio del estudio.
* **prevalentStroke:** Indicador binario (sí/no) de si el participante ha sufrido un accidente cerebrovascular previo.

1. **Variable de Interés y Medicamentos**

* **TenYearCHD:** Variable objetivo binario (sí/no) que indica si el participante desarrolló enfermedad coronaria del corazón en los siguientes 10 años.
* **BPMeds:** Indicador binario (sí/no) de si el participante estaba tomando medicamentos para la presión arterial.
* **Limpieza de datos:**

En cuanto a la calidad del dato en nuestro Dataset, nos encontramos con un data bastante limpio, sólo con unos pocos valores faltantes en la columna 'BPMeds'.

La variable 'BPMeds' presentó un porcentaje reducido de valores faltantes (1,25%). Para garantizar la completitud del conjunto de datos y minimizar el impacto en los resultados del análisis, se optó por imputar los valores faltantes mediante la moda. Esta técnica, al reemplazar los valores faltantes por la categoría más común, permite mantener la distribución original de la variable y reduce el riesgo de introducir sesgos en el análisis.

También se han realizado varios cambios para facilitar el análisis en las columnas que indican valores numéricos para expresar distintos géneros, distintos grupos de edad o indicadores binarios (si/no) para expresar si las personas del estudio son o no fumadores, si presentan diabetes o hipertensión, entre otros.

* **Consideraciones finales:**

Dado el carácter longitudinal y la riqueza de información del dataset de Framingham, este estudio nos permitirá responder preguntas fundamentales para la investigación en enfermedades cardiovasculares. Entre ellas destacan:

* ¿Cuáles son los factores de riesgo más significativos asociados con el desarrollo de enfermedades cardiovasculares en la población estudiada?
* ¿Cuál es la combinación óptima de estos factores para predecir la probabilidad de eventos cardiovasculares a corto y largo plazo?
* ¿Podemos identificar subgrupos de población con mayor riesgo y, por lo tanto, priorizar las intervenciones preventivas?
* **Metodología:**

En este estudio, se utilizó Power BI debido a su interfaz intuitiva y a su amplia gama de visualizaciones, lo que facilitó la exploración y el análisis de los datos. Además, Power BI permitió crear paneles interactivos para comunicar de manera efectiva los resultados a un público no especializado, integrándose además en el estudio y ampliando los conocimientos expuestos en el estudio realizados durante el curso.

Durante el análisis exploratorio, se realizaron histogramas para evaluar la distribución de las variables numéricas, diagramas de caja para comparar grupos y gráficos de dispersión para explorar las relaciones entre las variables. Los resultados preliminares indicaron una asociación positiva entre la edad, el índice de masa corporal y la presión arterial sistólica, así como una mayor prevalencia de enfermedades cardiovasculares en los hombres fumadores.