# Estructura del Proyecto: Modelo de Agrupación del Desarrollo Socioeconómico de las Familias en Colombia y su Relación con la Conectividad a Internet, Basado en IA y Machine Learning

## 1. Comprensión del Negocio

Objetivo: Identificar patrones y agrupar a las familias en Colombia según su desarrollo socioeconómico y su nivel de acceso a internet, con el fin de descubrir insights que orienten políticas públicas, estrategias de inclusión digital y desarrollo social.

Objetivos Específicos:

* - Determinar la correlación entre la conectividad a internet y el desarrollo socioeconómico.
* - Identificar factores socioeconómicos clave que influyen en el acceso y uso de internet.
* - Crear un modelo de agrupación de familias para clasificar niveles de desarrollo y conectividad.

Stakeholders: Gobierno, ONGs, empresas de telecomunicaciones, e instituciones de desarrollo social.

## 2. Descripción de Variables

La base de datos se organiza en diferentes Dimensiones, cada una representando un aspecto clave del desarrollo socioeconómico y de la conectividad en las familias. Estas dimensiones incluyen:

1. Educación:

* - `P6160`: Indica si la persona sabe leer y escribir (Opciones: 1 = Sí, 2 = No).
* - `P8586`: Informa si la persona está estudiando actualmente (Opciones: 1 = Sí, 2 = No).
* - `P8587`: Nivel educativo más alto alcanzado (Opciones que van desde 1 = Ninguno hasta niveles superiores).

2. Conectividad a Internet:

* - Variables relacionadas con el tipo de conexión a Internet, velocidad de la conexión, y frecuencia de uso.
* - Opciones categorizadas que reflejan distintos niveles de acceso y calidad de la conectividad.

3. Situación Económica:

* - Incluye variables como los ingresos familiares, ocupación y otros indicadores financieros.
* - Proporciona datos cuantitativos y categóricos que permiten evaluar el bienestar de las familias.

4. Geografía:

* - Ubicación: Clasificación de zona (urbana o rural) y región, para entender la relación entre conectividad y acceso a servicios en distintas áreas.

## 3. Preparación de los Datos

Transformación: Procesar y estandarizar las variables clave para que sean compatibles con técnicas de agrupación, como la conversión de datos categóricos a numéricos y la normalización de valores.

Limpieza de Datos: Imputación de valores faltantes y tratamiento de datos anómalos.

Ingeniería de Características: Creación de nuevas variables basadas en combinaciones o transformaciones de las existentes.

División de Datos: Separación de datos en conjuntos de entrenamiento y prueba para validar la eficacia del modelo.

## 4. Modelado con KMeans

Algoritmo de Agrupación: El algoritmo KMeans se utilizará para agrupar a las familias en diferentes categorías de desarrollo socioeconómico y conectividad.

* - Número de Clústeres: Se probarán diferentes valores de K para encontrar la segmentación óptima, utilizando el método del 'Codo' (Elbow Method).
* - Distancia Euclidiana: KMeans utiliza la distancia euclidiana para definir la proximidad entre puntos y agrupar aquellos con características similares.
* - Optimización de Hiperparámetros: Ajustes y evaluaciones iterativas para maximizar la cohesión dentro de cada grupo.
* - Interpretación de Resultados: Análisis de las características de cada clúster para identificar patrones socioeconómicos y de conectividad.

## 5. Evaluación

Evaluación de Resultados: Medir la efectividad del modelo utilizando índices de cohesión y separación, como la silhouette score.

Validación con Stakeholders: Revisión de la segmentación para asegurar su utilidad práctica y alineación con objetivos del proyecto.

Iteración de Modelos: Ajustes en la fase de modelado en caso de que los resultados iniciales no cumplan con los estándares de calidad.

## 6. Despliegue

Entrega del Modelo: Desarrollo de una interfaz visual para visualizar los clústeres y analizar el desarrollo socioeconómico en función de la conectividad.

Implementación en Políticas Públicas: Generación de reportes y gráficos para la toma de decisiones.

Documentación y Capacitación: Documentación detallada del proyecto y capacitación para el equipo final sobre cómo interpretar los resultados.