

## Informe de Análisis Multivariado del Régimen de Afiliación

### 1. Introducción

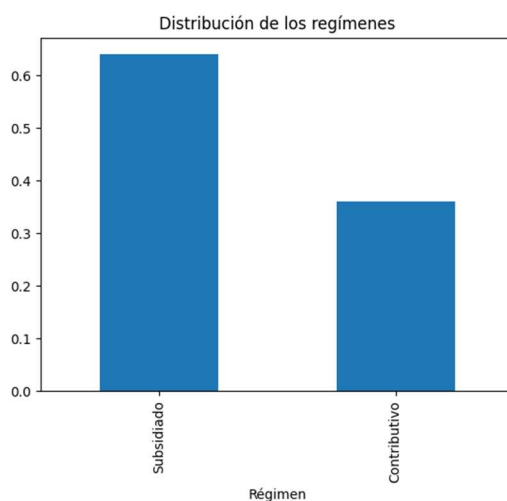
El presente análisis busca identificar los factores que explican la pertenencia de los individuos a los distintos regímenes del sistema de salud (Subsidiado y Contributivo), utilizando un conjunto de variables categóricas relacionadas con el tipo de afiliación, características socioeconómicas y entidad administradora.

Dado que todas las variables son categóricas, se aplicaron métodos estadísticos y de aprendizaje interpretables como:

- **Coefficiente de asociación (Cramer's V)** para determinar la fuerza de relación entre variables.
- **Análisis multivariado (SVD sobre codificación One-Hot)** para explorar patrones de agrupamiento.
- **Árbol de decisión** para identificar las reglas explicativas del régimen.
- **Análisis descriptivo** de la distribución general de los regímenes.

### 2. Distribución general de los regímenes

La base de datos presenta una mayor proporción de afiliados al régimen subsidiado ( $\approx 64\%$ ) frente al contributivo ( $\approx 36\%$ ).



- El régimen subsidiado tiene una participación predominante, lo cual refleja una mayor cobertura de la población en condiciones de vulnerabilidad o con bajos niveles del Sisbén.

- La menor proporción contributiva indica una base reducida de trabajadores formales o personas con capacidad de cotizar.
- Este patrón es coherente con la composición socioeconómica de los grupos y con los resultados del árbol de decisión, donde las variables económicas y de afiliación explican fuertemente el régimen.

### 3. Matriz de asociación (Cramer's V)

El coeficiente de asociación mostró las siguientes relaciones más fuertes con el régimen:

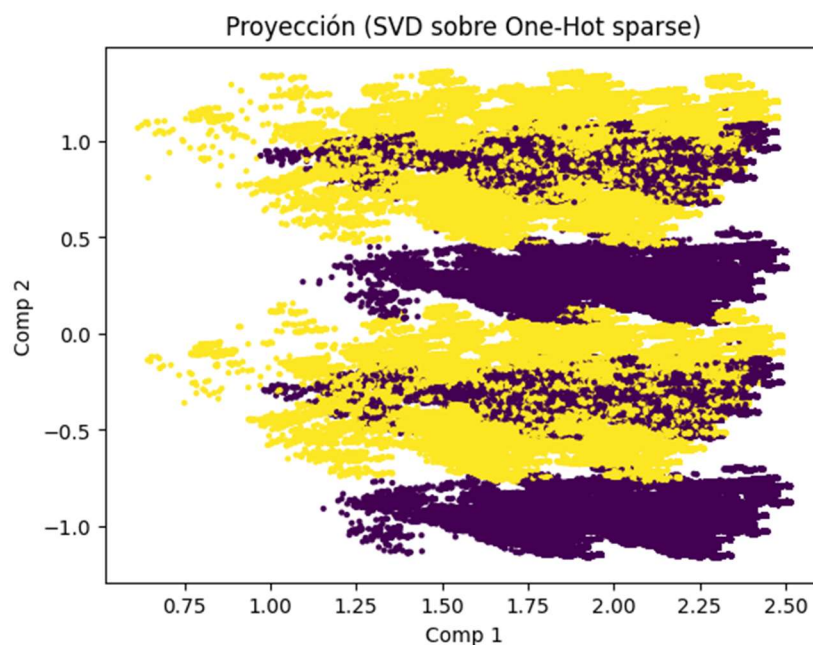
	0
Código de la entidad	0.999983
Tipo de afiliado	0.766364
Nombre de la entidad	0.673055
Nivel del Sisbén	0.578552
Estado del afiliado	0.217868
Municipio	0.134788
Zona de Afiliación	0.125388
Condición del beneficiario	0.121343
Grupo etario	0.117238
Departamento	0.092710
Género	0.008943

Variable	Cramer's	Interpretación
Código de la entidad	0.9999	Asociación casi perfecta: las EPS operan prácticamente bajo un único régimen.
Tipo de afiliado	0.766	Fuerte relación: los cotizantes y miembros familiares pertenecen al contributivo, mientras los beneficiarios y cabezas de familia al subsidiado.
Nombre de la entidad	0.673	Alta relación institucional: refuerza la segmentación por EPS.
Nivel del Sisbén	0.579	Fuerte relación socioeconómica: niveles bajos → subsidiado.
Estado del afiliado / Zona / Condición del beneficiario	0.12–0.21	Relaciones secundarias.
Género / Departamento / Grupo etario	<0.10	Sin influencia relevante.

El régimen está determinado principalmente por variables institucionales (entidad) y socioeconómicas (tipo de afiliado, nivel del Sisbén).

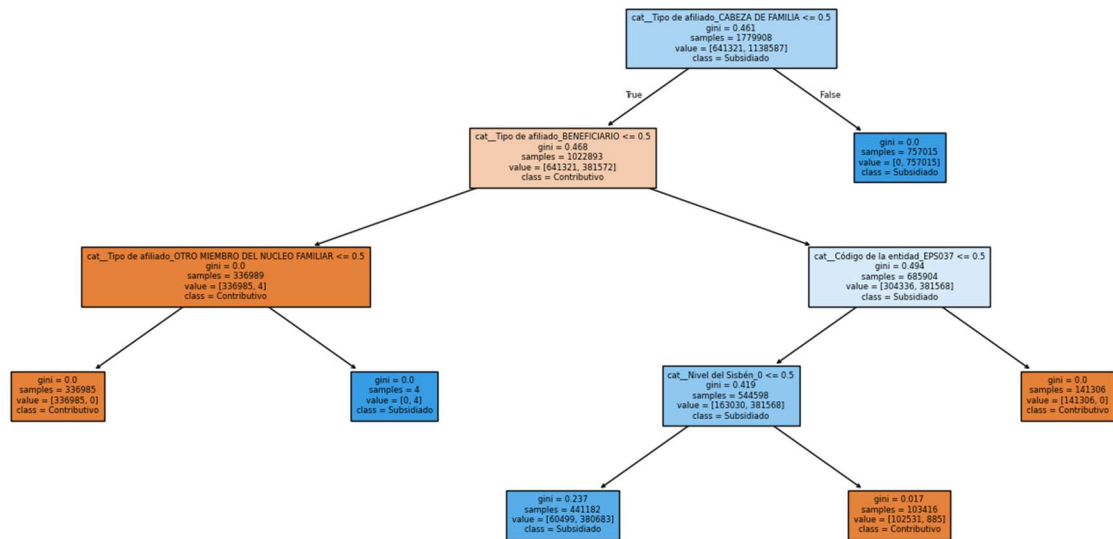
#### 4. Análisis multivariado (SVD sobre One-Hot sparse)

La proyección multivariada mostró dos grupos de puntos diferenciados por color (régimen). Las bandas horizontales en el gráfico indican la existencia de estratos o grupos categóricos bien definidos, donde cada régimen se agrupa según combinaciones de variables.



- Se observa una clara separación entre los afiliados de régimen contributivo y subsidiado, lo que confirma que las variables analizadas contienen suficiente información para distinguir ambos grupos.
- Las zonas de superposición entre colores reflejan perfiles mixtos, por ejemplo, afiliados en transición o entidades que gestionan ambos regímenes.
- En conjunto, el análisis evidencia que los regímenes se distribuyen de forma estructurada y no aleatoria.

## 5. Árbol de decisión explicativo



El modelo de árbol identificó las siguientes reglas de clasificación:

### 1. **Variable raíz:** Tipo de afiliado

Es el principal determinante del régimen.

- *Cabeza de familia* → Subsidiado
- *Cotizante / Miembro del núcleo familiar* → Contributivo

### 2. **Segunda variable clave:** Código de la entidad

Determina si la EPS pertenece al régimen subsidiado o contributivo.

- *EPS específicas (p. ej. EPS037)* → Contributivo
- *Otras EPS* → Subsidiado

### 3. **Tercer nivel:** Nivel del Sisbén

Reafirma la dimensión socioeconómica.

- *Niveles bajos (0, 1, 2)* → Subsidiado
- *Niveles altos o no aplicables* → Contributivo

El árbol logra una separación casi perfecta (gini cercano a 0 en la mayoría de las hojas), demostrando que con solo tres variables se puede explicar la mayor parte de la variabilidad del régimen.

## 6. Conclusiones generales

Aspecto	Resultado principal
Predominio del régimen	Subsidiado ( $\approx 64\%$ ), indicando mayor población en condiciones de vulnerabilidad.
Variables más influyentes	Tipo de afiliado, Código/Nombre de la entidad y Nivel del Sisbén.
Dimensión socioeconómica	Determina la pertenencia al régimen con alta precisión.
Dimensión institucional	Las EPS se asocian fuertemente con un régimen específico.
Dimensión demográfica	Género, edad o departamento no influyen significativamente.

## 7. Recomendaciones

- Utilizar las tres variables clave para construir modelos predictivos o dashboards de control de afiliación, optimizando la clasificación automática.
- Realizar un análisis geográfico complementario (por municipio o departamento) para identificar territorios con mayor proporción de subsidiados.
- Incorporar variables adicionales de ingreso, ocupación o formalidad laboral para mejorar la comprensión del fenómeno.