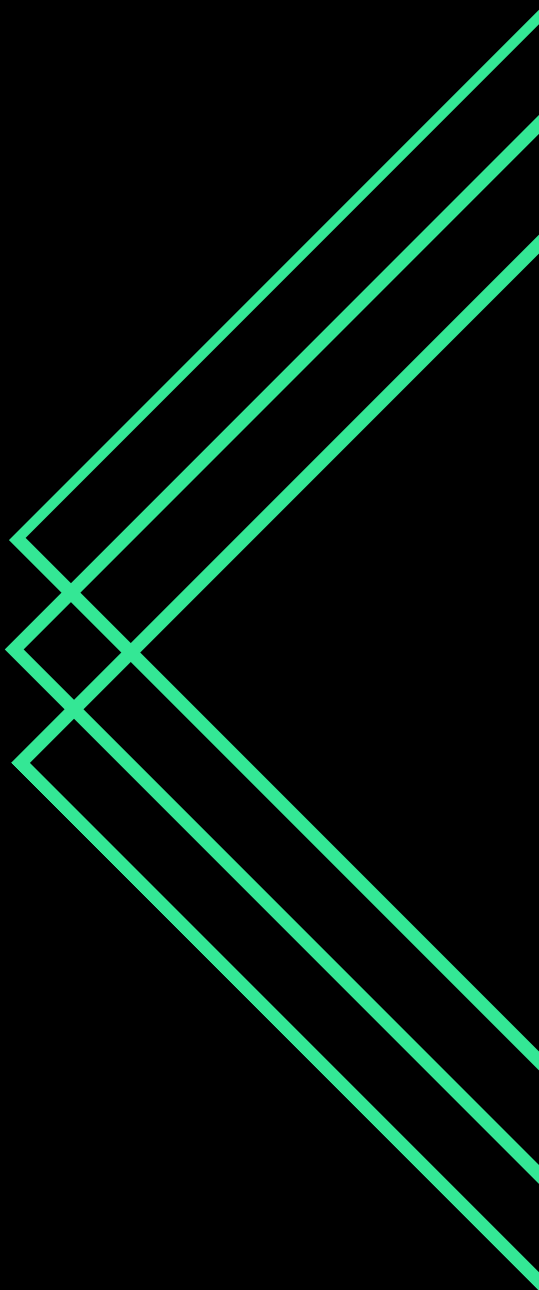
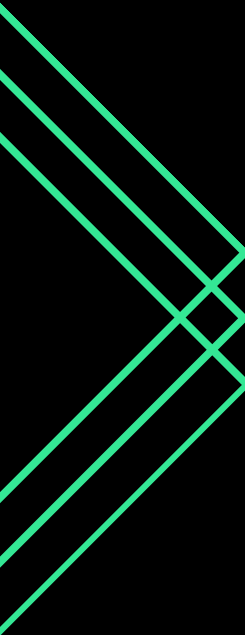


septiembre, 2023

# DESCRIPCIÓN DE DATOS

ESTIMACIÓN DE PENSIONES



Por: Víctor Sánchez

# Población

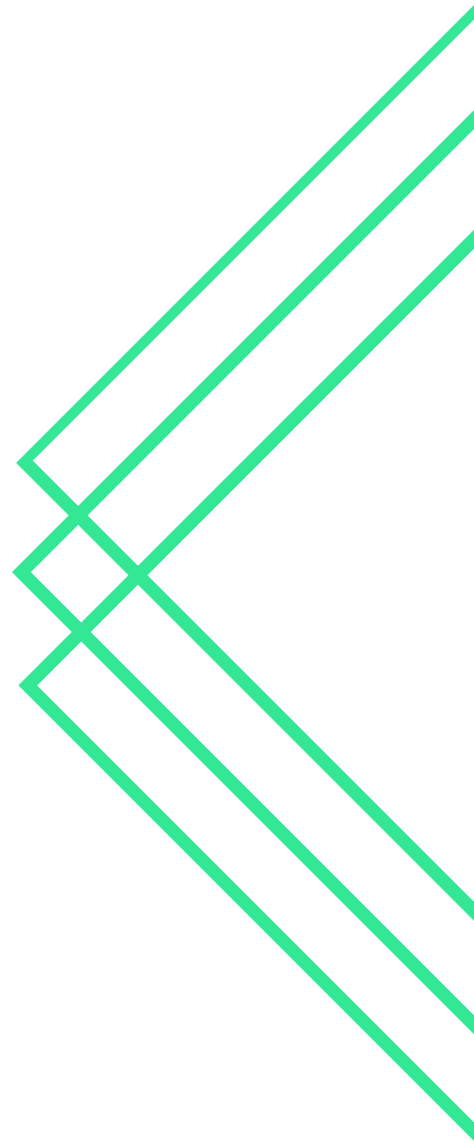
Trabajadores afiliados al Sistema de Ahorro para el Retiro (SAR) en México por debajo de la edad de jubilación, 65 años. De los cuales se dispone información necesaria para la estimación de su pensión.

## Variables y Características

En este apartado se describen cada uno de los datos del conjunto a estudiar:

- **Número de seguro social:** Clave alfanumérica única e intransferible asignada a trabajadores mexicanos asegurados con el fin de tener un registro de estos.
- **Género:** Variable binaria asociada a un atributo social característico e identificable de la persona de estudio.
- **Sexo:** Mismo sentido que la variable género pero numerica.
- **Fecha Nacimiento:** Dato en formato fecha YYYY-MM-DD
- **Salario:** Suma de dinero periódico que recibe un trabajador por el empleo de sus funciones y capacidades.
- **Saldo RCV:** Saldo de la subcuenta Retiro, Cesantía y Vejez del trabajador afiliado al Sistema de Ahorro para el Retiro.
- **Saldo VIV:** Saldo de la subcuenta Vivienda del trabajador afiliado al Sistema de Ahorro para el Retiro.
- **Saldo AV:** Saldo de aportaciones voluntarias del trabajador afiliado al Sistema de Ahorro para el Retiro.

Adicional a las características del conjunto descrito, se contemplarán otras variables univariadas tales como inflación, rendimiento y/o ajustes operativos.



# Hallazgos Iniciales

El conjunto no cuenta con todas las variables necesarias para la estimación de una pensión, sin embargo, mediante las mismas pueden ser definidas otras variables de interés, cómo:

1. Edad del trabajador, calculada a partir de la fecha de nacimiento (si existe).
2. Saldo total de la cuenta, que es la suma de las tres subcuentas del trabajador: RCV, AV y VIV.
3. Aportación voluntaria por mes\*

Adicional a lo anterior, existen datos faltantes en los campos:

1. Fecha de nacimiento
2. Salario

Para la primera variable se buscará definir una fecha a partir del NSS, mientras que para el salario se utilizarán métodos de imputación o algoritmos de agrupación, denominados clústers por su nombre en inglés.

Finalmente otras variables como el género cuenta con una clasificación de tres clases, por lo que se homologará con el contenido del campo sexo, y a partir de ello, se tomará una decisión.

\*Se definirá un criterio de monto de aportación voluntaria por mes en base al comportamiento o total de la cuenta de aportaciones y el tiempo activo del trabajador.