**Título del proyecto**

**Retiro del 10% de Fondo de Pensiones Chileno**”

**Objetivo del proyecto**

Presentar y desarrollar un flujo reproducible para un análisis exploratorio de datos (EDA) realizado en Python a partir de un conjunto de datos real, se recorren las etapas fundamentales del EDA: desde la preparación inicial de los datos (detección de valores faltantes, ajuste de tipos de variables y tratamiento de valores atípicos) hasta el análisis descriptivo y la exploración de relaciones entre variables.

La base para estudiar corresponde a una muestra del 10% de una de las bases que generan los resultados del estudio “Estudio sobre tasas de reemplazo en el sistema de pensiones chileno y sus proyecciones bajo distintos escenarios” publicada por la Superintendencia de Pensiones. La muestra corresponde a un 10% de afilados activos que, habiendo cumplido la edad legal de pensión entre junio de 2020 y diciembre de 2022, no se habían pensionado a diciembre de 2022 y realizaron todos o alguno de los “retiros del 10% de fondo de pensiones”. Para esta muestra se reporta el registro histórico previsional desde abril de 2007 en adelante.

**Estructura metodológica del proyecto**

1. Descripción de la base de datos:

-Información general sobre el dataset, columnas, tipos de datos y contexto del mismo.

2. Carga del dataset

- Importación desde fuente original manteniendo nombres de columnas y trazabilidad.

3. \*\*Lectura del Dataset y exploración inicial \*\*

- Revisión de estructura, tipos de datos y resumen estadístico,exploración inicial con

funciones básicas (.head(), .info(), .describe()), identificación de valores faltantes, ajuste

de tipos de variables y detección de valores atípicos.

4. Análisis Inicial de Datos (IDA)

4.1. Detección y tratamiento de valores faltantes o erróneos

4.2. Detección y tratamiento de valores atípicos (outliers)

- Variables Numéricas

- Variables Categóricas.

4.3 Reajuste de los tipos de variables

**5**. Análisis Exploratorio de Datos (EDA)

Visualización y análisis descriptivo de los datos, explorando relaciones entre variables y

patrones importantes.

5.1. Análisis Descriptivo

5.1.1. Univariado

-Variables numéricas

-Variables categóricas

5.1.2. Bivariado (Relacionar variables)

**Conclusión**

Este proyecto puede ayudar para enfocar mejoras en políticas públicas, hacer campañas de educación previsional mas efectivas y diseñar estrategias de segmentación apropiadas.

Lo que se encontró:

* Las personas entre 45 y 55 años son las que más retiran fondos, sobre todo en zonas urbanas.
* Las mujeres tienen una mayor proporción de retiro total (retira\_100) que los hombres, lo que podría reflejar diferencias en cómo ahorran o acceden a sus fondos.
* El saldo acumulado (saldo\_202212) varía bastante según la región, siendo más bajo en zonas extremas.
* La variable ind\_cumple permite identificar grupos con condiciones especiales, mostrando diferencias en el uso del dinero y el saldo disponible.