### **Documentación del Proyecto: Análisis de Vacunación**

#### **1. Tratamiento inicial del archivo de datos**

El primer paso fue la lectura del archivo CSV con los registros de vacunación. Para esto, diseñamos una función que pudiera leer cada línea del archivo y dividirla en campos de acuerdo con las columnas correspondientes. Analizamos los encabezados para asegurarnos de que cada campo tuviera un significado claro y una estructura específica. Nos aseguramos de procesar estos datos de manera que coincidieran con los requisitos establecidos en las consignas.

#### **2. Identificación de datos necesarios para cada consigna**

Antes de implementar cualquier funcionalidad, revisamos las consignas y decidimos qué campos eran relevantes. Establecimos qué campos eran esenciales para responder a las preguntas que se nos plantearon, como la distribución de género, el tipo de vacuna, las dosis aplicadas por jurisdicción y las personas mayores de 60 años que recibieron refuerzos.

Esto nos permitió enfocarnos en la manipulación adecuada de los datos relevantes y evitar procesar información innecesaria. Definimos claramente qué tipo de validaciones y análisis debían realizarse sobre cada campo del archivo.

#### **3. Tratamiento de errores**

Para garantizar la calidad de los datos, implementamos una fase de validación de errores. Validamos campos críticos, como el género y los nombres de dosis, entre otros. Aquellos registros que no cumplieran con el formato esperado o que tuvieran valores inválidos fueron marcados como erróneos y almacenados en un archivo separado. Esta estrategia permitió asegurar que los datos utilizados en los análisis fueran consistentes y correctos, minimizando errores en los resultados.

#### **4. Desarrollo de funciones para cada consigna**

El siguiente paso fue construir funciones específicas para obtener la información solicitada en cada consigna. Por ejemplo:

* **Distribución por género**: Contamos las personas vacunadas por género utilizando un diccionario para sumar la cantidad de personas de cada categoría.
* **Vacunas aplicadas por tipo**: Usamos otro diccionario para contar el número de aplicaciones de cada tipo de vacuna.
* **Dosis por jurisdicción**: Implementamos un conjunto (set) para evitar duplicar personas y asegurar que cada persona fuera contada solo una vez por dosis.
* **Segundas dosis por jurisdicción**: Filtramos aquellos registros en los que el código de dosis era igual a '2' para identificar las segundas dosis.
* **Refuerzos para mayores de 60 años**: Realizamos una lógica adicional para detectar aquellos grupos etarios con edades mayores de 60 y sumar las aplicaciones de refuerzo.

Cada función fue diseñada para trabajar con los datos depurados, asegurando que las respuestas a las consignas fueran precisas y rápidas.

#### **5. Uso de herramientas vistas en clase**

Durante la implementación, utilizamos diversas estructuras de datos aprendidas en clase, como:

* **Diccionarios**: Fueron fundamentales para contar la cantidad de ocurrencias (vacunas por tipo, dosis por jurisdicción, etc.).
* **Conjuntos (sets)**: Los utilizamos para almacenar los identificadores únicos de personas, lo que nos permitió evitar contar varias veces a la misma persona.
* **Listas**: Para almacenar tanto los datos válidos como los registros erróneos.

Estas herramientas fueron claves para simplificar el proceso de manejo de grandes volúmenes de datos y realizar un análisis eficiente.