

# Desafío Integrador - 1er Parte

## Introducción

Usted se ha incorporado recientemente al equipo liderado por el Ingeniero Ibañez, y se le ha asignado la tarea de analizar los datos de vacunación de COVID-19 que fueron provistos por el Ministerio de Salud de la Nación para su tratamiento.

Como parte de su labor, deberá familiarizarse con los datos disponibles, los recursos proporcionados y la información relacionada. Su objetivo es analizar los datos y generar los informes requeridos por el Ministerio de Salud de la Provincia del Chaco.

## Material provisto:

Se le proporcionará una parte del archivo base, el cual se dividió para facilitar su procesamiento (15 millones de registros aprox.). La fuente original es un archivo .csv. A continuación, se detalla la descripción de cada uno de los campos:

- **sexo**: Indica el género de la persona que recibió la vacuna. Los valores pueden ser 'M' (masculino) o 'F' (femenino).
- **grupo\_etario**: Define el rango de edad de la persona vacunada. Los valores suelen estar en intervalos como '18-29', '40-49', '50-59', etc., o categorías especiales como '<12' para menores de 12 años.
- **jurisdiccion\_residencia**: Indica la provincia o región donde reside la persona que recibió la vacuna, por ejemplo, 'Buenos Aires', 'Salta', 'La Pampa', etc.
- **jurisdiccion\_residencia\_id**: Es un código numérico que identifica de manera única la jurisdicción de residencia de la persona. Por ejemplo, '06' para Buenos Aires, '66' para Salta.
- **depto\_residencia**: Nombre del departamento o localidad dentro de la jurisdicción de residencia de la persona vacunada, por ejemplo, 'Capital', 'Baradero'.
- **depto\_residencia\_id**: Código numérico que identifica el departamento o localidad de residencia dentro de la jurisdicción. Este código varía dependiendo de la jurisdicción.
- **jurisdiccion\_aplicacion**: Provincia o región donde se aplicó la vacuna, que puede ser distinta de la jurisdicción de residencia. Ejemplos: 'Buenos Aires', 'Córdoba'.
- **jurisdiccion\_aplicacion\_id**: Código numérico que identifica de manera única la jurisdicción donde se aplicó la vacuna. Por ejemplo, '06' para Buenos Aires, '14' para Córdoba.
- **depto\_aplicacion**: Departamento o localidad dentro de la jurisdicción donde se aplicó la vacuna, por ejemplo, 'Capital', 'Baradero'.
- **depto\_aplicacion\_id**: Código numérico que identifica el departamento o localidad donde se aplicó la vacuna. Este código varía según la jurisdicción.
- **fecha\_aplicacion**: Fecha en la que se administró la vacuna. El formato es año-mes-día ('YYYY-MM-DD').

- **vacuna**: Nombre de la vacuna que fue administrada. Ejemplos: 'Sinopharm', 'Pfizer', 'Moderna ARNm 020 mg mL', etc.
- **cod\_dosis\_generica**: Código numérico que identifica la dosis genérica que fue aplicada (por ejemplo, '3' para segunda dosis, '14' para refuerzos).
- **nombre\_dosis\_generica**: Describe la dosis aplicada. Ejemplos: '2da', 'Refuerzo', '1ra'.
- **condicion\_aplicacion**: Indica la condición de la persona al recibir la vacuna, como '18 a 39 años SIN Factores de Riesgo', 'CON Factores de Riesgo', '60 o más años', entre otros.
- **orden\_dosis**: Especifica el número de dosis que ha recibido la persona en total, por ejemplo, '1' para la primera dosis, '2' para la segunda, '3' para un refuerzo.
- **lote\_vacuna**: Número de lote o identificación de la vacuna aplicada, que es utilizado para rastrear el origen y la producción de la dosis administrada.
- **id\_persona\_dw**: Es un identificador único para cada persona en la base de datos, probablemente generado automáticamente para preservar la privacidad. Tiene un formato de número flotante (con decimales) pero actúa como identificador único.

Archivos:

[Modelo Muestra](#)

[Conjunto\\_datos\\_1](#)

[Conjunto\\_datos\\_2](#)

## Solicitud

Dado que los datos fueron provistos por terceros, se ha informado que podrían existir inconsistencias. Por lo tanto, antes de realizar cualquier análisis, se deberá identificar posibles errores en los datos y tratarlos adecuadamente.

Además, se solicitará que en el informe final se incluya un archivo separado con los registros erróneos o inconsistentes. Este archivo deberá incluir una columna adicional llamada "OBSERVACIÓN" en la cual se explique el problema detectado. El archivo debe estar en formato .csv.

## Análisis Solicitado

El informe debe incluir lo siguiente:

### Estadísticas Descriptivas

1. **Distribución por Género:**
  - Calcular cuántas personas de cada género (**sexo**) han recibido la vacuna.
  - **Ejemplo de resultado:**
    - Masculino: 1,500,000
    - Femenino: 1,700,000
2. **Vacunas Aplicadas por Tipo de Vacuna:**

- Determinar cuántas personas han recibido cada tipo de vacuna (**vacuna**) y calcular la proporción respecto al total.
- **Ejemplo de resultado:**
  - Pfizer: 40%
  - Moderna: 35%
  - AstraZeneca: 25%

## Agrupación y Conteo

### 6. Dosis por Jurisdicción de Residencia:

- Agrupar los datos por **jurisdiccion\_residencia** y contar cuántas personas han sido vacunadas en cada provincia.
- **Ejemplo de resultado:**
  - Buenos Aires: 2,000,000 dosis
  - Córdoba: 500,000 dosis
  - Salta: 300,000 dosis

## Pedido Especial

El Ministro ha solicitado información específica sobre:

1. Cuántas personas han recibido la segunda dosis en cada jurisdicción.
2. Cuántas personas mayores de 60 años han recibido dosis de refuerzo.

**Tiempo Límite de entrega: Lunes 14 23:59. Se entrega en el campus del informatorio, subiendo un archivo .rar con su nro de dni de nombre**

## Contenido del .rar:

- Código fuente (documentado) **OBLIGATORIO**
- Documento explicando lo que realizaron. **DESEABLE**
- Archivo de los resultados solicitados. **OBLIGATORIO**
- Archivos de entrada. **NO (solo indicar que conjunto de datos utilizaron)**

## Se evaluará

- Resolución de lo solicitado.
- Reutilización del proyecto para análisis de otros conjuntos de datos con la misma estructura. (generalización).
- Incorporación de conceptos dados en clase.
- Presentación a término.