

05. - Data Collection_CU_04_06_sanitarias_v_01

June 8, 2023

#

CU04_Optimización de vacunas

Citizenlab Data Science Methodology > II - Data Processing Domain *** > # 05.- Data Collection

Data Collection is the process to obtain and generate (if required) necessary data to model the problem.

0.0.1 06. Datos sanitarios

- En este notebook se simulan los datos
- Cuando estén disponibles los datos reales, se debe implementar según corresponda

Condiciones de la simulación:

1. Simular datos para cada zona de salud
2. Las zonas son las 286 que están en el archivo cu_04_05_02_zonas.csv
3. Se simulan las siguientes variables para cada zona sanitaria:
 - **priesgo** - Proporción de población de riesgo. Sumar un 10% más algo de ruido a la proporción de mayores de 65 años (t4_3 en CU_04_05_05_02_indicadores_zonas.csv)
 - **triesgo** - Tasa de variación de la población de riesgo. Simular 1% más algo de ruido (desv. típica 0.005).
 - **tmayores** - Tasa de variación de la población mayor. Simular 2% más algo de ruido (desv. típica 0.01).
 - **capacidad** - Capacidad de la zona. Simular 25% de la población con algo de ruido (entero positivo)

Table of Contents

Settings

Data Load

ETL Processes

Import data from: CSV, Excel, Tab, JSON, SQL, and Parquet files

Synthetic Data Generation

Fake Data Generation

Open Data

Data Save

Main Conclusions

Main Actions

Acciones done

Acctions to perform

0.1 Settings

0.1.1 Encoding

Con la siguiente expresión se evitan problemas con el encoding al ejecutar el notebook. Es posible que deba ser eliminada o adaptada a la máquina en la que se ejecute el código.

```
[ ]: Sys.setlocale(category = "LC_ALL", locale = "es_ES.UTF-8")
```

```
'es_ES.UTF-8/es_ES.UTF-8/es_ES.UTF-8/C/es_ES.UTF-8/C'
```

No aplica

0.2 Data Load

No aplica

0.3 Synthetic Data Generation

- Se han generado los datos según la especificación utilizando herramientas de Microsoft (Power Query, Excel, Power BI)

0.4 Fake Data Generation

No aplica

0.5 Open Data

No aplica

0.6 Data Save

El archivo generado (output.xlsx) se ha copiado a la carpeta Data/Input, y será utilizado en otras tareas.

0.7 Main Conclusions

List and describe the general conclusions of the analysis carried out.

0.7.1 Prerequisites

No aplica

0.7.2 Configuration Management

No aplica

0.7.3 Data structures

Se mostrarán en las tareas de importación

Observaciones generales sobre los datos

- El archivo incluye los datos sanitarios y de vacunación
- En vez de tasas de variación de población riesgo y población mayor, se han generado proporciones
- La capacidad se ha generado por semanas, y en principio estaba pensada fija para todo el año

0.7.4 Consideraciones para despliegue en piloto

- Ninguna: se usarán los datos resultantes de la unión final

0.7.5 Consideraciones para despliegue en producción

- Se deben crear los procesos ETL en producción necesarios para que los datos de entrada estén actualizados
- Se deben conseguir los datos reales y sustituir los simulados

0.8 Main Actions

Acciones done Indicate the actions that have been carried out in this process

- Se ha guardado el archivo simulado

Acctions to perform Indicate the actions that must be carried out in subsequent processes

- Se deben calcular las tasas en base a las proporciones

0.9 CODE TO DEPLOY (PILOT)

A continuación se incluirá el código que deba ser llevado a despliegue para producción, dado que se entiende efectúa operaciones necesarias sobre los datos en la ejecución del prototipo

Description

- No hay nada que desplegar en el piloto, ya que estos datos son estáticos o en todo caso cambian con muy poca frecuencia, altamente improbable durante el proyecto.

CODE

```
[ ]: # incluir código
```