

05. - Data Collection_CU_53_01_spi_raw_v_01

June 13, 2023

#

CU53_impacto de las políticas de inversión en sanidad, infraestructuras y promoción turística en el SPI

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 01. Descargar datos del SPI

- Datos por países de 60 indicadores y el índice que calculan. Se ha obtenido el siguiente fichero, que se copia a **Data/Input** para ser usado en el caso de uso:
 - 2022_social_progress_index_dataset-1663250262.xlsx

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.

```
[1]: 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'
```

0.1.2 Packages to use

No aplica

0.1.3 Paths

0.2 Data Load

No aplica

0.3 ETL Processes

No aplica

0.4 Synthetic Data Generation

No aplica

0.5 Fake Data Generation

No aplica

0.6 Open Data

No aplica

0.7 Data Save

No aplica

0.8 Main Conclusions

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

Observaciones generales sobre los datos

- El archivo obtenido es un libro de Excel .xlsx

0.8.1 Consideraciones para despliegue en piloto

- Tener en cuenta que el paso a producción los datos se deben comprar para uso comercial

0.8.2 Consideraciones para despliegue en producción

- Se deben crear los procesos ETL en producción necesarios para que los datos de entrada estén actualizados
- Datos obtenidos mediante donación para investigación. La empresa deberá comprarlos para usarlos en producción

0.9 Main Actions

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

- Se ha copiado manualmente el archivo 2022_social_progress_index_dataset-1663250262.xlsx a la carpeta Data/(Input)

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

- Se deben transformar los datos a csv con el formato estándar del proyecto

0.10 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
```