

05. - Data Collection_CU_04_08_google_trends_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 08. Datos de Google Trends

- En este notebook se obtienen datos de Google Trends
- Si se puede hacer código, bien
- Si no, se importa el archivo descargado directamente

Condiciones de los datos:

1. Una fila por semana y municipio (creo que está, si no de toda la comunidad)
2. De las semanas que se hayan simulado vacunas
3. Obtener tendencia de la palabra “gripe”

Se ha descargado el fichero con el siguiente código en python

```
from pytrends.request import TrendReq
import pandas as pd
```

```
# Iniciar sesión en Google Trends
pytrend = TrendReq()
```

```
# Especificar la palabra clave a buscar
keyword = 'gripe'
```

```
# Establecer la región de búsqueda
pytrend.build_payload(kw_list=[keyword], geo='ES-MD')
```

```
# Obtener los datos de tendencias de búsqueda por día
trend_daily = pytrend.get_historical_interest([keyword], year_start=2021, month_start=9, day_s
```

```
# Seleccionar solo los datos correspondientes al período septiembre 2021 - enero 2023
start_date = '2021-09-01'
end_date = '2023-01-31'
trend_daily = trend_daily.loc[start_date:end_date]
```

```
# Convertir los datos a un formato de marco de datos
trend_daily = pd.DataFrame(trend_daily.stack(), columns=['busquedas_gripe']).reset_index()

# Agrupar los datos por fecha y calcular el número de búsquedas por zona de la Comunidad de Ma
trend_daily = trend_daily.groupby(['date', 'geoName']).sum()

# Guardar los datos en un archivo CSV
trend_daily.to_csv('busquedas_diarias_gripe_zonas_CAM.csv')
```

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

No aplica

0.2 Data Load

If there are more than one input file, make as many sections as files to import.

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

- Se ha obtenido un archivo .xlsx que se transformará al estándar csv del proyecto en tareas posteriores

0.8.1 Consideraciones para despliegue en piloto

- Ninguna

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

0.9 Main Actions

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

- Se ha copiado el archivo .xlsx a la carpeta **Data/Input**

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

- Se deben guardar los dados en .csv
- Se deben unir los datos al resto de datos del caso de uso

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
```