

05. - Data Collection_CU_04_09_tuits_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 09. Datos de Twitter

- En este notebook se obtienen datos de Twitter
- Si se puede hacer código y la API de Twitter, bien
- Si no, se importa el archivo que se pueda descargar

Condiciones de los datos:

1. Una fila por semana y geografía que se pueda (municipio o comunidad autónoma)
2. De las semanas que se hayan simulado vacunas
3. Obtener tuits con la palabra “gripe”

La descarga se ha realizado manualmente siguiendo los siguientes pasos:

1. Crear una cuenta de Twitter Developer.
2. Solicitar la cuenta de Academic Level of the Twitter API.
3. Mediante el uso de Postman, se pueden hacer las consultas correspondientes (ver captura)

Ver notas de despliegue en producción

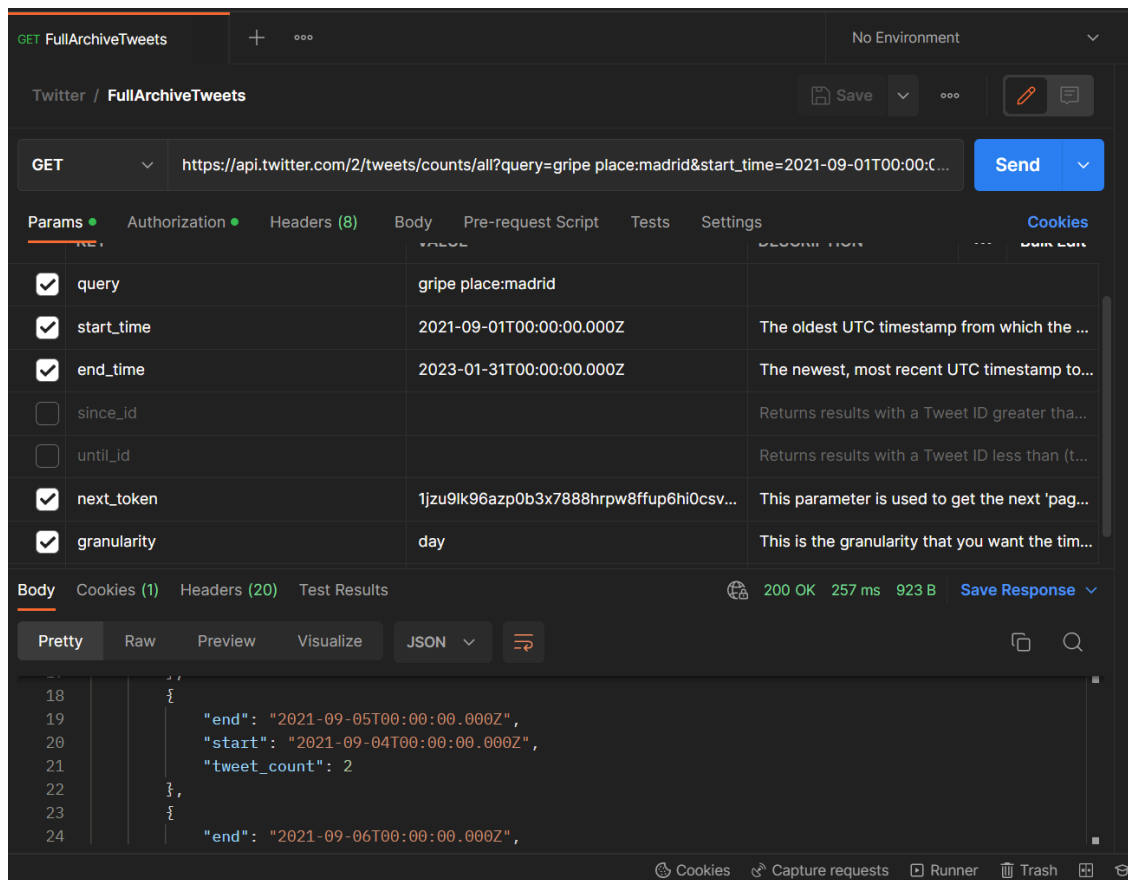


Table of Contents

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 Data Save

No aplica

0.5 Main Conclusions

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

Observaciones generales sobre los datos

- Se han obtenido con procedimientos no automatizados. Para la puesta en producción del caso se debe automatizar con algún lenguaje.

0.5.1 Consideraciones para despliegue en piloto

- Los datos a utilizar serán los resultantes del resto de tareas del proceso 05

0.5.2 Consideraciones para despliegue en producción

- Se debe crear el código para automatizar la actualización de los datos
- Se deben crear los procesos ETL en producción necesarios para que los datos de entrada estén actualizados

0.6 Main Actions

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

- Este notebook es informativo, no se ha realizado ninguna acción sobre los datos

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

- Se deben exportar los datos a formato csv

0.7 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
```