

05. - Data Collection_CU_18_06_adminshp_v_01

June 13, 2023

#

CU18_Infraestructuras_eventos

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 de localización de empresas y administraciones

- En esta tarea se descargan los datos actualizados de establecimientos comerciales, educativos y de administraciones públicas (no sanitarias)
- En el piloto, se han descargado los datos en formato shp siguientes de la url <https://gestiona.comunidad.madrid/nomecalles/DescargaBDTCorte.icm#delimitacionesterritoriales>:
- Administración pública_ Agencia Tributaria.zip
- Administración pública_ Ayuntamientos, Consejerías, Ministerios, etc..zip
- Administración pública_ Embajadas y consulados.zip
- Administración pública_ Oficinas de empleo.zip
- Administración pública_ Seguridad Social.zip
- Administración Pública. Administración de Justicia.zip
- Colectivo empresarial por tamaño y actividad.zip
- Comercio_ Centros comerciales.zip
- Comercio_ Galerías de alimentación.zip
- Comercio_ Grandes superficies especializadas.zip
- Comercio_ Hipermercados.zip
- Comercio_ Mercadillos.zip
- Comercio_ Mercados de abastos.zip
- Comercio_ Otros servicios. Bancos.zip
- Educación_ Campus universitarios.zip
- Educación_ Centros educativos no universitarios. Centros privados.zip

- Educación_ Centros educativos no universitarios. Centros públicos.zip
- Educación_ Centros educativos no universitarios. Servicios educativos.zip
- Educación_ Centros educativos universitarios.zip
- Educación_ Colegios mayores.zip

Establecimientos de administración, colectivo empresarial, comercio y educación. Datos disponibles en formato shp

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 Packages to use

No aplica

```
[ ]: # library(www)
```

0.1.2 Paths

```
[ ]: # iPath <- "Data/Input/"
      # oPath <- "Data/Output/"
```

0.2 Data Load

No aplica

0.3 ETL Processes

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

No aplica

0.4 Synthetic Data Generation

No aplica

0.5 Fake Data Generation

No aplica

0.6 Open Data

Los archivos shp se descargaron de la fuente abierta como ficheros comprimidos .zip que se descomprimieron para tener la estructura de carpetas necesaria

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

- Los datos están en formato shp con las coordenadas y alguna información adicional, casi todo referido a la dirección
- Para cada tipo de infraestructura pueden venir unos campos u otros

0.8.1 Consideraciones para despliegue en piloto

- Los datos de origen fueron obtenidos de <https://gestiona.comunidad.madrid/nomecalles/DescargaBDTCorte>. en fichero zip con formato .shp y descomprimido en la carpeta Data/Input
- Si la información geográfica cambia, se debería actualizar el fichero y volver a ejecutar todos los procesos.

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 han descargado y descomprimido los archivos

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

- Se debe consolidar la información en un fichero CSV

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
```