

# Como ejecutar el script

---

Para construir la aplicación lo primero que debe hacer es crear un entorno virtual de python Para instalar virtualenv:

```
pip install virtualenv
```

Para generar un entorno virtual

```
virtualenv .venv
```

Una vez creado el entorno virtual debemos activarlo

Windows:

```
.\.venv\Scripts\activate.bat
```

Ubuntu:

```
source .venv/bin/activate
```

Una vez tenemos el entorno activado vamos a instalar en este todos los módulos necesarios

```
pip install -r requirements.txt
```

Ya tenemos configurado todo lo necesario para ejecutar el scrip, mediante el comando

```
python ./main.py
```

El script se encargará de refinar los datos en bruto, de todas las oleadas y redes sociales, de España e Italia, de forma que Gephi acepte dichos ficheros. Además, generará tres gráficas de frecuencia, una global, y otras dos, con los hashtags publicados en España e Italia.

## Configuración por defecto ejecutar el script

---

El código ha sido entregado con los datos en bruto, extraídos de las redes sociales. El script se encargará de refinarlos y pasar por todas las fases del procesado hasta generar los ficheros que soporta Gephi.

Si deseamos ejecutar también la extracción de datos, debemos:

Poner en el fichero credenciales.py, un usuario y contraseña válidos de Instagram, ya que son necesarios para que instaloader realice la extracción de datos. A continuación, debemos modificar el archivo main.py, y llamar a la función main, con el argumento scrap a 'True'.

## Código

---

El código consta de los siguientes ficheros: - **main.py**. Se encarga de llamar de forma ordenada a las distintas funciones que necesitamos para el refinamiento de los datos, así como carga en memoria, los hashtags relacionados con el covid, las distintas oleadas, y las compañías que vamos a estudiar.

- **get\_data\_instagram.py**. Hace uso del módulo Instaloader, para extraer los datos que necesitamos de cada compañía de Instagram.
- **get\_data\_twitter.py**. Hace uso del módulo Twint, para extraer los datos que necesitamos de cada compañía de Twitter.
- **functions.py**. Contiene distintas funciones que utilizamos, algunas de ellas son:
  - **split\_data\_in\_waves**. Divide las publicaciones en oleadas
  - **scarp**. Hace uso de get\_data\_instagram.py y get\_data\_twitter.py para extraer y agrupar la información correspondiente.
  - **convert\_json\_2\_graph**. Dado un vector de publicaciones, genera un grafo y lo almacena en un fichero
- **graph.py**. Implementa la clase Graph, la cual, aísla el tratamiento de los grafos, en su última etapa, y los almacena en ficheros.

- **hashtag.py**. Implementa la clase Hashtag, la cual nos ayuda a trabajar de forma eficiente con los grupos de hashtags.
- **graficos.py**. Se encarga de generar las gráficas de frecuencias.
- **constantes.py**. Agrupa algunas constantes globales
- **credenciales.py**. Contiene las credenciales de acceso a Instagram.
- **resources/companies.dat**. Agrupa todas las compañías que vamos a estudiar, y sus usuarios en Twitter e Instagram.
- **resources/covid\_hashtags.dat**. Agrupa los grupos de hashtags que vamos a estudiar.
- **resources/waves.dat**. Contiene las distintas oleadas con su fecha de inicio y fin, tanto en España como en Italia.