

# *Máster Data Science UOC*



---

## Práctica Web Scrapping

---

*Autor:*

Kyrylo MOROZOV

*Curso 2020/2021*

Fecha: 09/11/2020

**1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.**

Se trata de una red Social en continuo auge. El creciente uso de este tipo de webs, ha aumentado el tiempo en el que la sociedad hace uso de éstas. Esto conlleva que éstas dispongan de información útil que puede ser usada por empresas de marketing.

**2. Definir un título para el dataset. Elegir un título que sea descriptivo.**

Instagram Dominant Color posts Torrevieja

**3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).**

El Dataset extraído contiene información acerca de los posts de instagram en una región, en concreto Torrevieja. La finalidad de estos dataset es obtener los colores dominantes de las fotografías en esta zona.

**4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente**

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

- URL post
- Nick Usuario
- Nombre Completo
- Fecha publicación
- Likes
- Followers usuario
- Texto Post
- Hashtags
- Color Dominante 1
- Color Dominante 2

**6. Agradecimientos.** Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

No se ha hecho uso de ninguna fuente externa.

**7. Inspiración.** Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Como ya se ha mencionado al principio, las redes sociales pueden ser unas fuentes de información muy útiles pues éstas incluyen los gustos y las preferencias de la sociedad. Esto puede ser usado para un estudio de preferencias a la hora de vender productos y ofrecer servicios.

**8. Licencia.** Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

☐ Released Under CC0: Public Domain License

**9. Código.** Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

<https://github.com/kirilmn13/InstagramScraper/tree/main>

**10. Dataset.** Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

<https://zenodo.org/record/4263896#.X6kQxfNKiU1>

**Todo el Trabaja ha sido Realizado por Kyrylo Morozov**