

Descripción de la práctica

Esta práctica ha sido realizada individualmente por Javier Gómez de Diego.

- GitHub: <https://github.com/javi22gd/TyCVD>
- Video: <https://github.com/javi22gd/TyCVD/blob/master/Presentacion/Presentacion.zip>

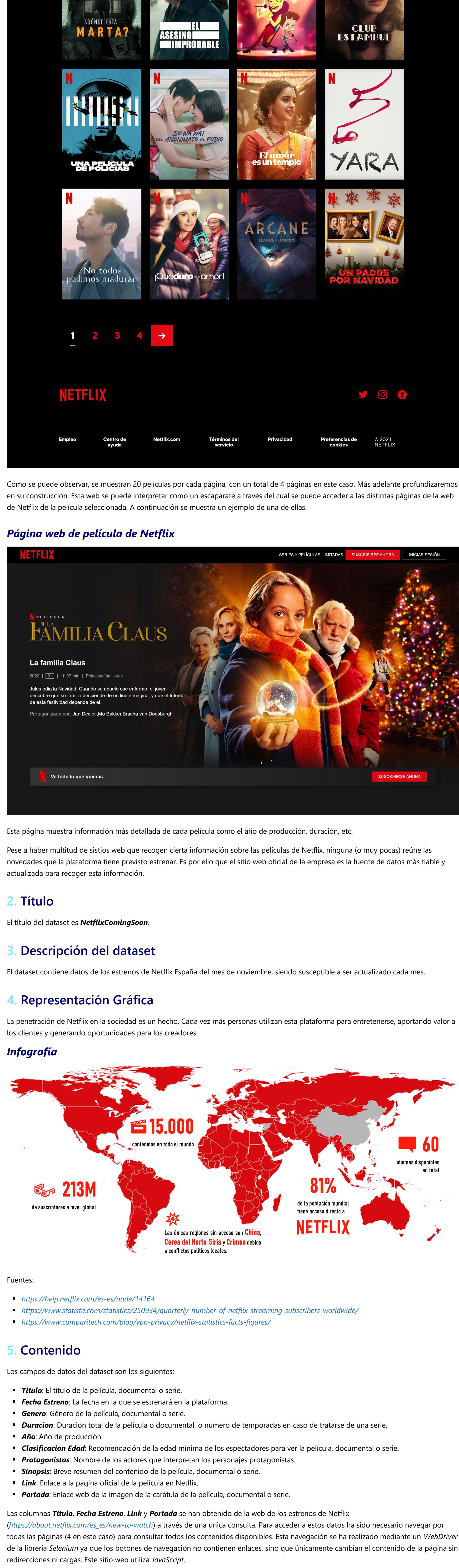
Índice

- 1. Contexto
- 2. Título
- 3. Descripción del dataset
- 4. Representación gráfica
- 5. Contenido
- 6. Agradecimientos
- 7. Inspiración
- 8. Licencia
- 9. Código
- 10. Dataset

1. Contexto

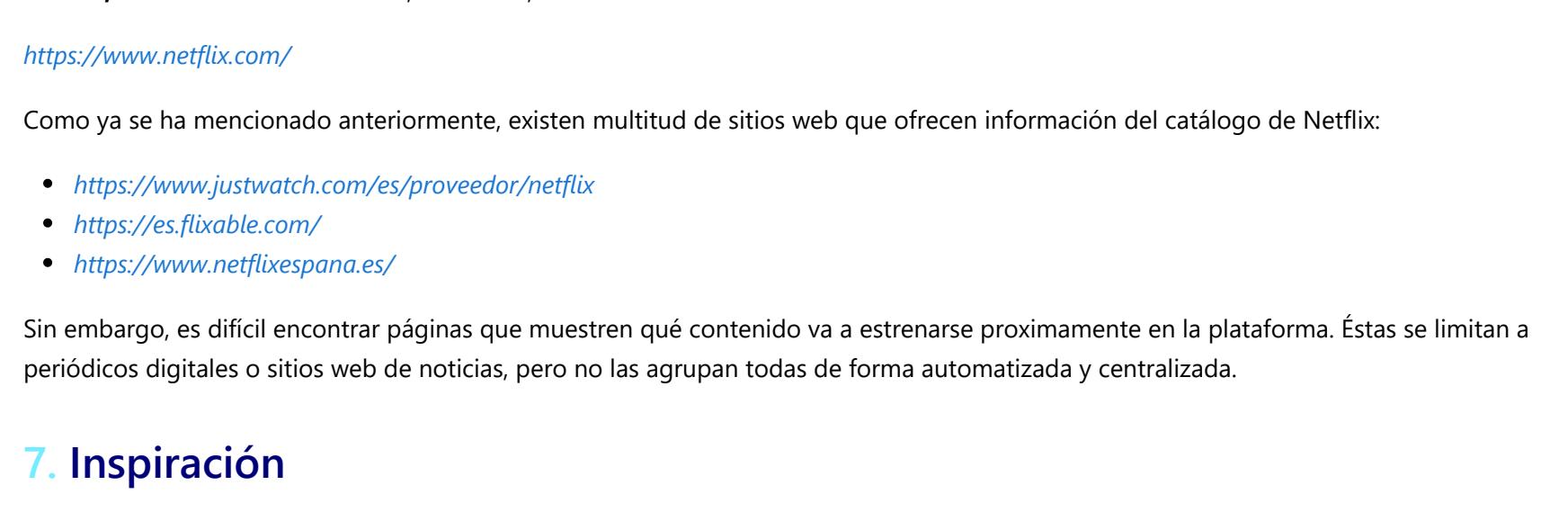
La plataforma de contenido audiovisual Netflix contiene un catálogo que va incorporando novedades constantemente. Este nuevo contenido es anunciado a través de su sitio web https://about.netflix.com/es_es/new-to-watch, en el que se incluyen los estrenos del mes. A continuación se puede ver una imagen del aspecto de esta página.

Web de novedades de Netflix



Como se puede observar, se muestran 20 películas por cada página, con un total de 4 páginas en este caso. Más adelante profundizaremos en su construcción. Esta web se puede interpretar como un escaparate a través del cual se puede acceder a las distintas páginas de la web de Netflix de la película seleccionada. A continuación se muestra un ejemplo de una de ellas.

Página web de película de Netflix



Esta página muestra información más detallada de cada película como el año de producción, duración, etc.

Pese a haber multitud de sitios web que recogen cierta información sobre las películas de Netflix, ninguna (o muy pocas) reúne las novedades que la plataforma tiene previsto estrenar. Es por ello que el sitio web oficial de la empresa es la fuente de datos más fiable y actualizada para recoger esta información.

2. Título

El título del dataset es *NetflixComingSoon*.

3. Descripción del dataset

El dataset contiene datos de los estrenos de Netflix España del mes de noviembre, siendo susceptible a ser actualizado cada mes.

4. Representación Gráfica

La penetración de Netflix en la sociedad es un hecho. Cada vez más personas utilizan esta plataforma para entretenerte, aportando valor a los clientes y generando oportunidades para los creadores.

Infografía

Fuentes:

- <https://help.netflix.com/es-es/node/14164>
- <https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/>
- <https://www.comparitech.com/blog/vpn-privacy/netflix-statistics-facts-figures/>

5. Contenido

Los campos de datos del dataset son los siguientes:

- **Título:** El título de la película, documental o serie.
- **Fecha Estreno:** La fecha en la que se estrenará en la plataforma.
- **Genero:** Género de la película, documental o serie.
- **Duración:** Duración total de la película o documental, o número de temporadas en caso de tratarse de una serie.
- **Año:** Año de producción.
- **Clasificación Edad:** Recomendación de la edad mínima de los espectadores para ver la película, documental o serie.
- **Protagonistas:** Nombre de los actores que interpretan los personajes protagonistas.
- **Sinopsis:** Breve resumen del contenido de la película, documental o serie.
- **Link:** Enlace a la página oficial de la película en Netflix, documental o serie.
- **Portada:** Enlace web de la imagen de la carátula de la película, documental o serie.

Las columnas **Título**, **Fecha Estreno**, **Link** y **Portada** se han obtenido de la web de los estrenos de Netflix (https://about.netflix.com/es_es/new-to-watch) a través de una única consulta. Para acceder a estos datos ha sido necesario navegar por todas las páginas (4 en este caso) para consultar todos los contenidos disponibles. Esta navegación se ha realizado mediante un WebDriver de la librería Selenium ya que los botones de navegación no contienen enlaces, sino que únicamente cambian el contenido de la página sin redirecciones ni cargas. Este sitio web utiliza JavaScript.

El resto de campos se han recogido accediendo directamente a las distintas páginas de cada película, documental o serie en Netflix, cuyo enlace se ha obtenido previamente. Para acceder a estas páginas ha sido necesario modificar la cabecera HTTPS de las consultas y espaciarlas temporalmente para evitar ser bloqueados por el servidor.

6. Agradecimientos

Todos los datos obtenidos pertenecen a:

Netflix, Inc. 100 Winchester Circle, Los Gatos, CA 95032 US

<https://www.netflix.com/>

Como ya se ha mencionado anteriormente, existen multitud de sitios web que ofrecen información del catálogo de Netflix:

- <https://www.justwatch.com/es/proveedor/netflix>
- <https://es.flixible.com/>
- <https://www.netflixespana.es/>

Sin embargo, es difícil encontrar páginas que muestren qué contenido va a estrenarse próximamente en la plataforma. Éstas se limitan a periódicos digitales o sitios web de noticias, pero no las agrupan todas de forma automatizada y centralizada.

7. Inspiración

El objetivo de este trabajo es facilitar la difusión de los próximos estrenos en Netflix, así como el conocimiento de los mismos a los espectadores. De esta forma, se puede obtener la información requerida en un solo documento en vez de tener que navegar a través de múltiples páginas en la web de Netflix, agilizando el proceso.

En comparación con las periódicos digitales que ofrecen redacciones publicitarias acerca de los próximos estrenos, a través de este conjunto de datos se puede acceder a la información de todo el contenido de forma más amplia, sencilla y eficiente.

8. Licencia

La licencia elegida es **CC BY-NC-SA 4.0** debido a las características del dataset:

- Se permite editar el dataset para ampliar y/o sustituir el contenido cuando se anuncien nuevos estrenos en la plataforma.
- No se puede dar un uso comercial al dataset ya que todos los datos pertenecen a Netflix, Inc., luego la decisión de uso comercial de los mismos le pertenece legítimamente a la compañía.
- Esta licencia se mantendrá en futuras ediciones de los datos.

9. Código

<https://github.com/javi22gd/TyCVD/blob/master/Code/Practica1.ipynb>

10. Dataset

DOI: [10.5281/zenodo.5652089](https://doi.org/10.5281/zenodo.5652089)

Contribuciones

- **Investigación previa:** J.G.D.
- **Redacción de las respuestas:** J.G.D.
- **Desarrollo del código:** J.G.D.

Los campos de datos del dataset son los siguientes:

- **Título:** El título de la película, documental o serie.
- **Fecha Estreno:** La fecha en la que se estrenará en la plataforma.
- **Genero:** Género de la película, documental o serie.
- **Duración:** Duración total de la película o documental, o número de temporadas en caso de tratarse de una serie.
- **Año:** Año de producción.
- **Clasificación Edad:** Recomendación de la edad mínima de los espectadores para ver la película, documental o serie.
- **Protagonistas:** Nombre de los actores que interpretan los personajes protagonistas.
- **Sinopsis:** Breve resumen del contenido de la película, documental o serie.
- **Link:** Enlace a la página oficial de la película en Netflix, documental o serie.
- **Portada:** Enlace web de la imagen de la carátula de la película, documental o serie.