**Práctica 1, Web Scraping**

**Descripción**

El siguiente código en Python 3.6 tiene por objetivo aplicar las técnicas de Web Scraping para extraer datos de la web de la Organización Internacional de Fabricantes de Vehículos Motorizados OICA por sus siglas en inglés y generar como resultado un archivo de datos en formato CSV de las estadísticas de producción histórica de vehículos (desde el año 2000 hasta el 2021)

Esta práctica pertenece a la asignatura Tipología y ciclo de vida de los datos, correspondiente al Máster en Ciencia de Datos de la Universitat Oberta de Catalunya.

1. **Contexto**

Uno de los hechos más relevantes cuando se requiere analizar de datasets históricos es el proceso de descarga de cada uno de los años que se desean analizar, esta actividad puede tomar un lapso de tiempo considerable y más tratándose de varios periodos de tiempo, tradicionalmente es un proceso que consiste en visitar manualmente cada uno de los repositorios para recoger y descargar los datos que se necesiten. En este contexto donde los datos como la producción de vehículos de distintos países se conforman en datos importantes para analizar tendencias de crecimiento, o contrastes entre los distintos países, el uso de las herramientas que permitan acelerar la recopilación y posterior tratamiento de los mismos son indispensables.

La presente práctica tiene como objeto desarrollar un Script en el lenguaje de programación Python que permita la extracción anual de los datos correspondientes a la producción de vehículos, a partir de los datos publicados en la página web de la OICA <https://www.oica.net/production-statistics/>.

Se optado por el uso de lenguaje Python ya que tiene todas las herramientas para el manejo y procesamiento de datos, así como para llevar a cabo el proceso de escarbar dentro de páginas web, ya que cuenta con librerías y funciones que facilitan no solo este tipo de acciones si no también el posterior análisis de los datos obtenidos.

1. **Título**

“Estadísticas históricas de Producción de Vehículos en 41 países que reportan información en la OICA”

1. **Descripción de dataset**

El dataset se encuentra conformado por un conjunto de archivos que comprende un periodo de tiempo de 22 años y está constituido por las estadísticas relacionadas con, la producción de vehículos de un total de 41 países que han reportado estos datos a la Organización Internacional de Fabricantes de Vehículos Motorizados. Los datasets obtenidos son estáticos y anualizados, es decir que los valores presentados tienen un corte a final de cada año y no se los vuelve a actualizar hacia atrás del último periodo de tiempo; por lo que se tiene un conjunto de datos hasta el 2021, en vista de que el año 2022 aún se encuentra en curso.

1. **Representación gráfica**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | PRODUCCIÓN DE AUTOS |
|  |  |  |  |  |  |  |
| AÑO |  | ESTADISTICAS DE PRODUCCIÓN |  | PAIS |  | VEHÍCULOS COMERCIALIZADOS |
|  |  |  |  |
|  |  |  |  |  |  | TOTAL COMERCIALIZADO Y PRODUCIDO |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  | % DE VARIACIÓN |

1. **Contenido**

El dataset se encuentra conformado por un conjunto de archivos que comprende un periodo de tiempo de 22 años y está constituido por las estadísticas relacionadas con, la producción de vehículos cantidad de autos producidos, los vehículos comercializados, producción total y una variación porcentual de la producción, todo esto para alrededor de 40 países que reportan esta información, es importante mencionar que cada dataset se encuentra en diferentes menús, es decir que para cada año se tiene un dataset individualizado, adicionalmente es importante señalar que no se encuentra el mismo número de países en cada uno de los años seleccionados.

1. **Propietario**

El propietario del dataset es Organización Internacional de Fabricantes de Vehículos Motorizados, que es una organización fundada en 1919 y que tiene por misión defender los intereses de los fabricantes, ensambladores e importadores de vehículos agrupados en su federación. La organización lleva a cabo actividades en los campos de asuntos técnicos, comunicación y estadísticas de la industria automotriz.

Como se puede notar es una organización con un amplio periodo de vigencia y de desarrollo de sus actividades, cuyos datos permiten realizar el seguimiento y el análisis de una de las industrias que genera más fuentes de trabajo directos e indirectos y que dinamiza en gran medida la economía de un país, tal es el caso que esta industria aporta más de 430 000 millones de euros a los ingresos de los gobiernos de veintiséis países.[[1]](#footnote-1)

1. OICA [↑](#footnote-ref-1)