

Práctica 1.

Carlos Lavado Mahia

Dionisio González Jiménez

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

En 2020, el virus SARS-COVID19 afectó a los ingresos de explotación de numerosos sectores, entre ellos el de la automoción; el cual tiene una importancia estratégica en España, en concreto podríamos decir que supone alrededor de un 11 % del PIB Nacional. Pero que con esta gran crisis general se ha reducido en un gran porcentaje.

Con el objetivo de estudiar el impacto de este suceso, hemos accedido a la web de la patronal de concesionarios FACONAUTO (faconauto.com). En ella, hemos comprobado, a través del archivo "robots.txt", que podemos hacer web crawling en la página con cualquier agente, siempre que se deje un delay de 60 segundos.

```
User-agent: *  
Crawl-delay: 60  
User-agent: MJ12bot  
Disallow:
```

A través de la sección de "Estadísticas", hemos podido observar los datos mensuales de matriculaciones de turismos, pudiendo hacerse una comparación entre los años 2019 y 2020 (también 2021, si consideramos los datos hasta febrero). Con la intención de poder generar un fichero para calcular las variaciones interanuales, así como la evolución mensual durante cada año, hemos investigado la forma de extraer los datos mediante la librería BeautifulSoup.

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

Comparativa de matriculaciones mensuales de turismos en España desde el 2019 hasta la actualidad.

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 describe el número de matriculaciones mensuales durante los últimos ejercicios (2019, 2020, 2021 hasta febrero). En él, se puede

observar que las matriculaciones, por lo general, son mucho menores en el 2020, año que coincide con la situación de pandemia generada por el virus SARS-COVID19. Sin embargo, hay algunos meses (como por ejemplo, julio o diciembre) para los cuales esta tendencia se revierte, lo que pone de manifiesto la recuperación del sector.

4. Representación gráfica. Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.

No sabemos exactamente a que se refiere y lo tenemos pendiente.
¿Debemos realizar un tipo de gráfico que muestre nuestro análisis?

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

- Los campos que incluye el dataset son: XXXX
- El periodo de tiempo de los datos es 2019, 2020, y enero y febrero de 2021.
- Los datos se han recogido a través de una araña creada con la librería BeautifulSoup, y después han sido almacenados a CSV con XXXX.

Nota: hemos indicado con xxx los campos ya que todavía no hemos generado el CSV y no sabemos que campos exactos contendrá dicho documento.

6. Agradecimientos. Presentar el propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis anteriores.

El propietario del conjunto de datos es la patronal de concesionarios. Este se encuentra, a su vez, en la web de la patronal, propiedad de la empresa 10DENCEHISPAHARD, S.L. En cuanto a análisis anteriores, cabe destacar que Faconauto ya recopila los datos registrados en formato PDF. Nuestra intención con este análisis, es el de poder disponer de los números en un formato que permita su comparativa y manipulación, como por ejemplo CSV.

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el ejercicio 6.

Como ya se ha comentado en el ejercicio anterior, este conjunto de datos puede resultar interesante para realizar cálculos con el número de matriculaciones mensuales en España. Si bien nuestra intención inicial es

la de hacer una comparativa intermensual e interanual, el dataset también puede ser utilizado, por ejemplo, para hacer predicciones.

Las preguntas principales que pretendemos responder son: ¿Ha afectado el COVID-19 a las matriculaciones mensuales de coches? ¿Existen variaciones interanuales significativas? Además, de otras preguntas que podremos hacer más específicas a partir del análisis realizado.

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

- ☐ Released Under CC0: Public Domain License
- ☐ Released Under CC BY-NC-SA 4.0 License
- ☐ Released Under CC BY-SA 4.0 License
- ☐ Database released under Open Database License, individual contents under Database Contents License
- ☐ Other (specified above) ☐ Unknown License

Seleccionamos la licencia CC0: Public Domain License, ya que no somos propietarios de los datos extraídos (estos corresponden a la patronal Faconauto), y no tenemos por tanto ningún interés en ser mencionados como tales. Tampoco podemos tener ninguna garantía de la veracidad de los datos, si bien estos han sido extraídos de una asociación oficial.

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

El código que por ahora hemos realizado basando en el análisis que queremos obtener está subido en el propio repositorio.

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

En proceso de elaboración del CSV.