

# Caso práctico | La ciencia de datos

## Introducción

El presente caso práctico está estructurado en dos partes o secciones, la primera dirigida a la temática de: Calidad y preparación de los datos, después la segunda se relaciona con: Herramientas de Data Viz y visualización. Ambas partes deben entregarse en su totalidad, cumpliendo con todos los parámetros requeridos en cada apartado para obtener y aprobar el módulo.

## Desarrollo

### Parte I: Calidad y preparación de los datos

1. **Ingestar los datos en Python**, el archivo: **car\_evaluation.xlsx**, que encontrarás adjunto en la sección de recursos del módulo, este contiene información sobre datos de vehículos.
2. **Realizar un análisis exploratorio de datos** del Dataset, especificando tus principales hallazgos. También deberás determinar si consideras necesario aplicar alguna transformación de datos complementaria.
3. **Renombrar las columnas del Dataset** a: **buying, maint, doors, persons, lug\_boot, safety y class**. Si ves necesario aplicar otra transformación de datos adicional, podrás hacerlo argumentando de tal modo tus motivos de elección.
4. **Mediante las bibliotecas de Matplotlib o Seaborn** de Python, confecciona al menos **3 visualizaciones de datos** con sus correspondientes interpretaciones.
5. Por último, generar una pequeña **conclusión** de la actividad efectuada junto con los temas vistos de **calidad y preparado de datos**, con el fin de resumir la primera parte de la práctica integradora del módulo.

### Extra

- Aplicar un Exploratory Data Analysis (EDA) utilizando la librería de Pandas Profiling.
- Enlace de interés sobre Car Evaluation Data Set:  
<https://www.kaggle.com/datasets/elikplim/car-evaluation-data-set>

## Parte II: Herramientas de Data Viz y visualización

1. **Importar a Power BI Desktop** la información del archivo: **Conjunto de Datos (Adquisición de vehículos).xlsx**, que encontrarás adjunto en la sección de recursos del módulo, este contiene la información sobre los datos de los vehículos preparados para tratar en esta herramienta.
2. **Configurar la primera fila** de la cabecera como “Encabezado”.
3. **Cambiar el tipo de datos** de las columnas de la siguiente forma:
  - a. Número entero: Identificador
  - b. Fecha: Fecha operación
  - c. Texto: Resto de las columnas
4. **Crear una columna condicional** en el objeto importado denominado “Rng\_Precios” con los siguientes apartados:
  - a. Cuando el valor es vhigh, el valor asignado debe ser “Mas de 50.000 €”.
  - b. Cuando el valor es high, el valor asignado debe ser “Entre 30.000 € y 49.999 €”.
  - c. Cuando el valor es med, el valor asignado debe ser “Entre 15.000 € y 29.999 €”.
  - d. De lo contrario “Menos de 14.999 €”.
5. La nueva columna creada “Rng\_Precios”, debe cambiarse a tipo de datos “Texto”
6. Sobre la columna “Tamaño maletero”, cambiaremos el contenido para que toda la información mostrada sea en mayúsculas.
7. En el campo “Nº de Puertas” tenemos los casos de 2, 3, 4 y 5more, vamos a excluir la parte de “More”, para ello extraeremos por la izquierda solamente el primer dígito de este campo.
8. Las siguientes **columnas no van a utilizarse en el informe**, para ello debemos ocultarlas de nuestro objeto importado en Power Query:
  - a. Identificador
  - b. Rango de precio
  - c. Nivel de Mantenimiento
  - d. N. Personas
  - e. Seguridad
  - f. Estado del Vehículo
9. En el campo **Tamaño maletero**, vamos a cambiar el texto de inglés a castellano, debemos sustituir los valores de la siguiente forma:
  - a. SMALL → PEQUEÑO
  - b. MED → MEDIANO
  - c. BIG → GRANDE
10. **Cambiar el nombre del objeto**, denominándolo “HEC\_Ventas\_Coches” y aplicar los cambios para empezar las visualizaciones
11. **Crear una nueva tabla** “TMedidas\_Ventas” dónde incluir de forma organizada las medidas “DAX”. En este objeto añadir una medida que permita contar el número de operaciones realizadas, dar formato de número a la misma en la parte de “Herramientas de medición”.
12. **Crear las siguientes visualizaciones** alrededor de esta medida de DAX, utilizar los tipos de visualizaciones más acordes a cada uno de los casos:
  - a. Información de ventas por número de puertas.

- b.** Información de ventas por tamaño del maletero.
  - c.** Información de ventas por rango de precios.
- 13. Cambiar la configuración de las visualizaciones** de tal manera que aprovechemos los diferentes parámetros de cada visualización “Título”, “Ejes”, “Leyendas”, “Valores”, “Colores”, etcétera.
- 14. Realizar pruebas cambiando las visualizaciones por otras** y sacar conclusiones respecto a dichas modificaciones.

## Requisitos de la entrega

Entregar un documento de Power BI (formato pbix) que el profesor pueda abrir a través de Power BI Desktop, este documento debe tener importados los datos y las sentencias DAX necesarias para el correcto funcionamiento de las visualizaciones incluidas.

**¡Importante!** El documento debe ejecutarse sin errores y debe cumplir con los criterios de visualización analizados en el curso.