

## Caso Práctico

### **Problema 1**

En el archivo "**auto.csv**" se encuentran los siguientes datos de distintos automóviles:

- Cilindros.
- Cilindrada.
- Potencia.
- Peso.
- Aceleración.
- Año del coche.
- Origen.
- Consumo (mpg).

Es importante tener en cuenta que:

- El consumo está expresado en millas por galón (mpg); es decir, cuánta distancia recorre el automóvil con cada galón de combustible consumido, a diferencia de cómo se expresa en España —y en Europa, en general—, donde el consumo se mide en litros de combustible por cada 100 kilómetros recorridos (l/100 km).
- Las unidades de las características de los automóviles no se encuentran en el sistema internacional.

La variable *origen* es un código que identifica al país de origen.

### **Se pide**

- Crear un modelo para estimar el consumo de un vehículo a partir del resto de variables.

### **Problema 2**

En el archivo "**crime\_data.csv**" se muestra el número de crímenes por cada 100000 habitantes en cada uno de los estados de Estados Unidos, así como el porcentaje de la población urbana. Los crímenes se han agrupado en asalto, asesinato y violación

### **Se pide**

Segmentar el conjunto de datos utilizando *k-means* para obtener los centroides de cada clúster y el listado de los estados en cada uno de los clústeres. Para ello, se ha de encontrar el número óptimo de clúster en el que se divide el conjunto de datos.

**Observaciones:**

- Considerar la exploración de los datos al inicio de cada problema
- Ejecutar las tareas de transformación o procesamiento de las variables en caso de requerirlo
- Manejar conjuntos de datos para pruebas y entrenamiento si es necesario hacerlo
- Evaluar el desempeño del modelo de acuerdo a las métricas revisadas
- Optimizar o cambiar el modelo en caso de que sea necesario
- Documentar todo el proceso del Caso Práctico en un Notebook de Jupyter