

Conceptos básicos

- Una sola tabla de datos por archivo. **1**
- La primera fila puede contener exclusivamente el nombre de los campos (cabecera). Es opcional.
- Cada observación/registro es una línea. **2**
- Cada variable/campo es una columna.
- Todas las filas deben contener el mismo número de campos.
- Cada campo está separado del siguiente por un carácter singular: [“;”, [“;”, [“;”, [“;”] o [TAB]. Tener en cuenta el separador decimal para elegir el carácter separador.
- Codificación de caracteres UTF-8.

Diccionario de datos

- Cualquier información que no sea dato (esquemas, tablas, columnas, descripciones, comentarios o unidades de medida) debe incluirse en un diccionario de datos.
- Debe expresarse en un formato procesable, por ejemplo, JSON o JSON-LD, mediante un vocabulario estandarizado que permita definir cada una de las características de los elementos del archivo de datos. W3C recomienda un vocabulario para la descripción de datos tabulares.
- De las propiedades que existen para describir, tablas, filas, columnas o celdas, es recomendable usar, al menos:

Para las tablas { Título de tabla [“dc:title”]
Descripción [“dc:description”]
Publicador [“dc:creator”]
Ubicación del archivo que se describe [“url”]

Para las columnas { Nombre de columna[“name”]
Título de columna [“titles”]
Descripción[“dc:description”]
Tipos de datos: [“datatype”]

3 Los nombres de las columnas

- No deben repetirse
- Nombres cortos
- Evitar caracteres especiales

4 Estructura vertical

- Siempre que sea posible, el fichero debe actualizarse incorporando nuevas filas y manteniendo fijas las columnas.

5 Valores desconocidos

- A los valores “desconocidos” o “no obtenidos” se les asignar un código común. Por ejemplo: “NA”.
- El código que se utilice para indicar los valores desconocidos debe especificarse en el diccionario de datos.

6 No totales o subtotaes

- No deben incluirse columnas o filas que representen agrupamientos de otras.

7 Tipo de datos por columna

- Usar un tipo de datos por columna.
- Las unidades deben describirse en el diccionario de datos.

8 Valores estandarizados

- Usar vocabularios y normas estandarizadas.
- Utilizar la misma codificación y normalización para el mismo tipo de dato publicado en diferentes datasets.

9 Campos codificados

- Categorizar los datos para poder extraer patrones o aplicar filtros usando esquemas de conceptos y códigos con valores preestablecidos.

10 Campos de texto

- Valores que incluyen comas o saltos de línea deben ir entre comillas.
- El contenido entre comillas dentro de un valor debe ir doblemente entrecomillado.
- Los campos con valores numéricos que incluyen ceros a la izquierda significativos deben ser de tipo texto.

11 Campos numéricos

- No usar separadores de millares.
- El separador decimal depende de la configuración regional (En España usar “,”).
- Valores negativos irán precedidos del signo “-”. No usar paréntesis.
- El diccionario de datos debe expresar las unidades de medida o de moneda asociadas a valores numéricos.

12 Campos fecha

- Usar el formato ISO 8601: YYYY-MM-DD (forma abreviada) o YYYY-MM-DDTHH:MM:SS (forma extendida).
- Usar formato de 24 horas.
- Para indicar un mes es aconsejable ajustar al último día del mes (2019-09-30).

13 Campos teléfono

- Mantener siempre el mismo formato.
- Se recomienda incluir código de país.

14 Campos dirección postal

- Usar campos separados para codificar los elementos necesarios para localizar una dirección postal: tipo de vía; nombre de la vía; localidad; código postal.

15 Coordenadas geográficas

- Especificar latitud y longitud en columnas separadas y en grados decimales.

codigo_marca	modelo	año	consumo	aceleracion	cambio	descripcion	precio_venta	moneda	fecha_venta_1	promedio_ventas	codigo_vendedor	actividad_vendedor	tipo_via	nombre_via	localidad	c_postal	telefono	latitud	longitud
00034078236	Chevrolet Chevelle Malibu	1970	18	NA	M	Manual	23540,23	EUR	1970-01-31T15:10:00	2,51	45.11	Venta de automóviles y vehículos de motor ligeros	Calle	Automoción, 23, Bajo	Alicante	03559	+34-6760000	38,345171	-0,481492
000034078243	Buick Skylark 320	1970	15	NA	A	Automático	22189,00	EUR	1970-07-23T16:34:00	2,63	45.11	Venta de automóviles y vehículos de motor ligeros	Avenida	Industria, 33	Oviedo	33201	NA	43,345679	-5,844761
000034076256	Plymouth Satellite	1970	18	11	A	Automático	28362,65	USD	1973-05-01T11:13:34	3,42	45.19	Venta de otros vehículos de motor	Calle	Carrozas,2, Bajo	Madrid	28760	+34-6760003	40,416502	-3,702563
000034076258	Amc Rebel sst	1970	16	12	M	Manual	29200,00	USD	1972-11-12T10:00:01	NA	45.11	Venta de automóviles y vehículos de motor ligeros	Paseo	del Vagón,23	Huelva	11170	+34-6960004	43,345679	-6,940041
000034076321	Ford Torino	1970	17	10,5	M	Manual	31890,15	EUR	1969-03-12T12:43:36	2,54	45.19	Venta de otros vehículos de motor	Plaza	Berlina,1	Lugo	27294	+34-6760123	43,009923	-7,556020

codigo_marca;modelo;año;consumo;aceleracion;cambio;cambio_descripcion;precio_venta;moneda;fecha_venta_1;promedio_ventas;codigo_vendedor;actividad_vendedor;tipo_via;nombre_via;localidad;c_postal;telefono;latitud;longitud
 000034078236;Chevrolet Chevelle Malibu;1970;18;NA;M;Manual;23540,23;EUR;1970-01-31T15:10:00;2,51;45.11;Venta de automóviles y vehículos de motor ligeros;Calle;Automoción, 23, Bajo;Alicante;03559;+34-6760000;38,345171;-0,481492
 000034078243;Buick Skylark 320;1970;15;NA;A;Automático;22189,00;EUR;1970-07-23T16:34:00;2,63;45.11;Venta de automóviles y vehículos de motor ligeros;Avenida;Industria, 33;Oviedo;33201;NA;43,3456789;-5,8447609
 000034076256;Plymouth Satellite;1970;18;11;A;Automático;28362,65;USD;1973-05-01T11:13:34;3,42;45.19;Venta de otros vehículos de motor;Calle;Carrozas, 2, Bajo;Madrid;28760;+34-6760003;40,4165023;-3,7025631
 000034076258;Amc Rebel sst;1970;16;12;M;Manual;29200,00;USD;1972-11-12T10:00:01;NA;45.11;Venta de automóviles y vehículos de motor ligeros;Paseo;del Vagón,23;Huelva;11170; +34-6960004;43,3456789;-6,9400412
 000034076321;Ford Torino;1970;17;10,5;M;Manual;31890,15;EUR;1969-03-12T12:43:36;2,54;45.19;Venta de otros vehículos de motor;Plaza;Berlina,1;Lugo;27294;+34-6760123;43,0099231;-7,5560202