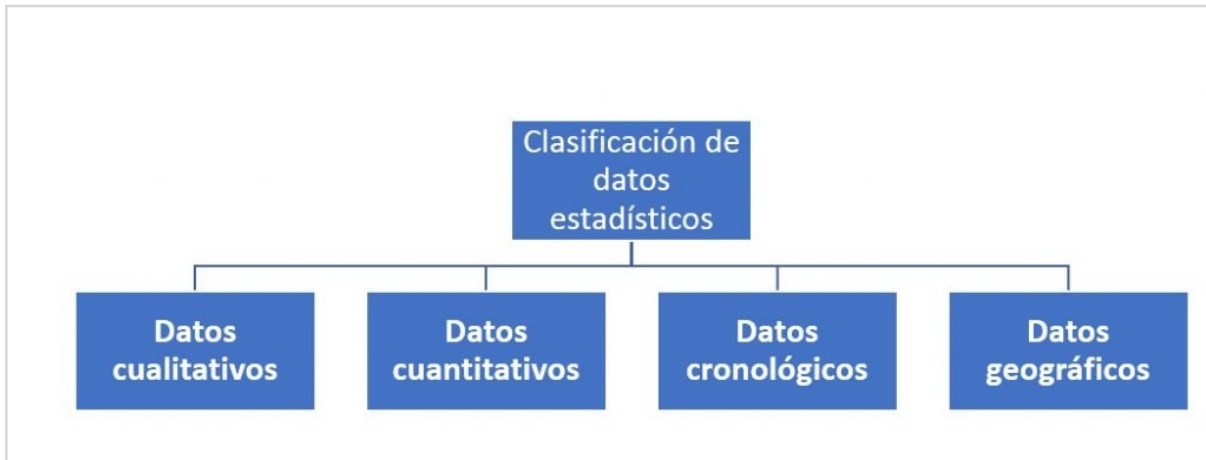


# Fundamentos: Entendiendo los datos

Tener una buena comprensión de los diferentes tipos de datos, sus características y fuentes de obtención, es crucial para realizar un análisis exploratorio de datos (EDA), ya que sólo se pueden usar ciertas mediciones estadísticas para tipos de datos específicos. También necesita saber con qué tipo de datos está tratando para elegir el método de visualización correcto.



## 1. ¿Qué son los datos?

Los datos son un conjunto de valores de variables cualitativas o cuantitativas sobre uno o más elementos que indica un valor que se le asigna a las cosas y se representan a través de una secuencia de símbolos, números o letras.

Aunque los términos "*datos*" e "*información*" a menudo se usan indistintamente, estos términos tienen significados distintos. Comúnmente los datos se transforman en información cuando se ven en contexto o en un análisis posterior.

## 2. Tipos de datos en estadística

Los tipos de datos son un concepto importante de la estadística, que debe entenderse, para aplicar correctamente las mediciones estadísticas a sus datos y, por lo tanto, para concluir correctamente ciertas suposiciones al respecto. Esta publicación de blog le presentará los diferentes tipos de datos que necesita conocer para realizar un análisis de datos exploratorio (EDA) adecuado, que es una de las partes más subestimadas de un proyecto de aprendizaje automático.

### 2.1 Datos cualitativos

Los datos cualitativos o categóricos representan características. Por lo tanto, puede

representar cosas como el género de una persona, el idioma, etc. Los datos categóricos también pueden tomar valores numéricos (Ejemplo: 1 para mujeres y 0 para hombres) pero estos no se pueden utilizar para realizar operaciones matemáticas. Existen dos tipos:

### Nominales

Los valores nominales representan unidades discretas y se utilizan para etiquetar variables que no tienen valor cuantitativo. Piensa en ellos como “*etiquetas*”. Ten en cuenta que los datos nominales no tienen orden. Por lo tanto, si cambiara el orden de sus valores, el significado no cambiaría.

Ejemplos:

- Sexo (masculino o femenino).
- Estado civil (soltero, casado, divorciado, viudo, etc.).
- Raza (asiático, afroamericano, etc.).

### Ordinales

Los valores ordinales representan unidades discretas y ordenadas. Por lo tanto, es casi lo mismo que los datos nominales, excepto que su orden es importante. Las escalas ordinales se utilizan generalmente para medir características no numéricas como la felicidad, la satisfacción del cliente, etc.

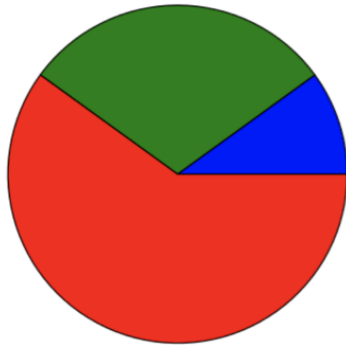
Un ejemplo son las escalas de Likert que miden la satisfacción:

<b>Ejemplo de la escala de Likert</b>				
Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo, ni en desacuerdo	De acuerdo	Totalmente de acuerdo
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

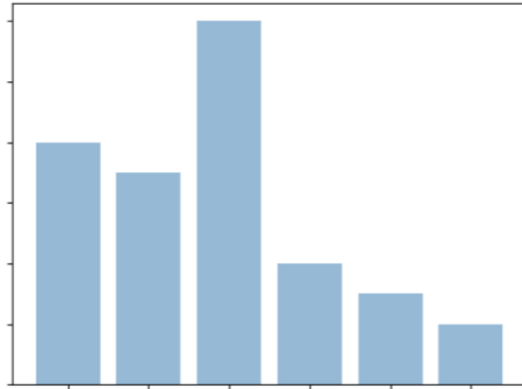
### Métodos de visualización

Para visualizar datos nominales, puede utilizar un gráfico circular o un gráfico de barras. Podemos resumir los datos con frecuencias, proporciones y porcentajes. Y puede visualizarlo con gráficos circulares y de barras.

## Pie Chart



## Bar Chart



### 2.2 Datos cuantitativos

Los datos cuantitativos o numéricos sirven para cuantificar elementos, comportamientos o cualquier variable que se haya definido para ser objeto de estudio.

#### Discretos

Hablamos de datos discretos si sus valores son distintos y separados. En otras palabras: hablamos de datos discretos si los datos solo pueden tomar ciertos valores. Este tipo de datos no se pueden medir, pero se pueden contar. Básicamente representa información que se puede categorizar en una clasificación.

Ejemplo:

- Número de hijos, adultos o mascotas en una familia.
- Número de habitantes.

#### Continuos

Los datos continuos representan mediciones y, por lo tanto, sus valores. no se pueden contar pero se pueden medir. Un ejemplo sería la altura de una persona, que puedes describir usando intervalos en la recta numérica real.

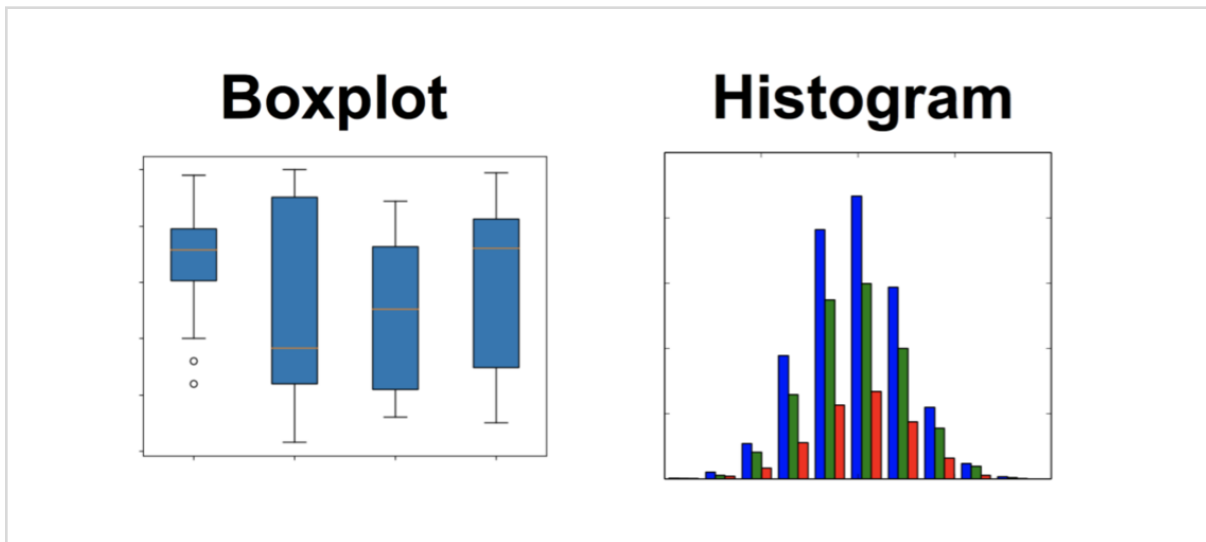
Los valores de intervalo también son datos continuos que representan unidades ordenadas que tienen la misma diferencia. Por tanto hablamos de datos de intervalo cuando tenemos una variable que contiene valores numéricos que están ordenados y donde conocemos las diferencias exactas entre los valores.

Ejemplo:

- Altura (Puedes mediar la altura en metros, centímetros. etc.).
- Edad (Puedes definir una edad en años, meses y hasta días).
- Temperatura.

### Métodos de visualización

Para visualizar datos continuos, se puede utilizar un histograma o un diagrama de caja. Con un histograma, se puede verificar la tendencia central, variabilidad y modalidad de una distribución.



### 2.3 Datos cronológicos

Los tipos de datos cronológicos registran tiempo y pueden servir para hacer análisis del histórico de un conjunto de datos así como aplicar filtros a hacer mediciones a través de un periodo determinado. Entre ellos tenemos:

- `date` - almacena una fecha del calendario (por ejemplo, "1993/10/22").
- `datetime` - registra un punto en el tiempo en cualquier grado de precisión, desde un año a una fracción de segundo. (Por ejemplo, "2018-01-10T08:14:00" ).
- `interval` almacena un lapso de tiempo; es decir, una duración (por ejemplo, "14:00-15:00").

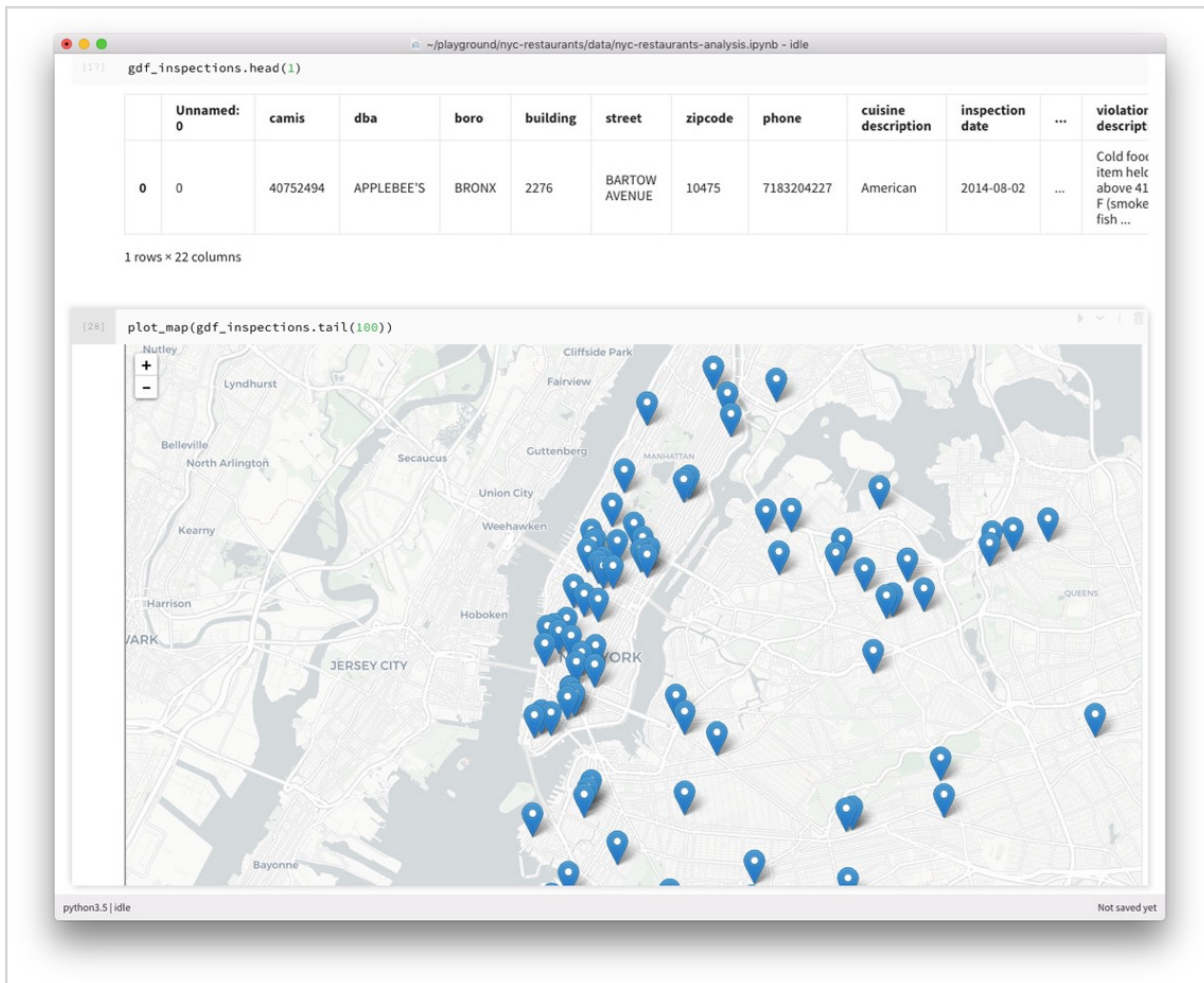
### 2.4 Datos geográficos

Los conjunto de datos geográficos están organizados de tal manera que permiten la realización de análisis y la gestión del territorio dentro de aplicaciones de Sistemas de Información Geográfica (SIG).

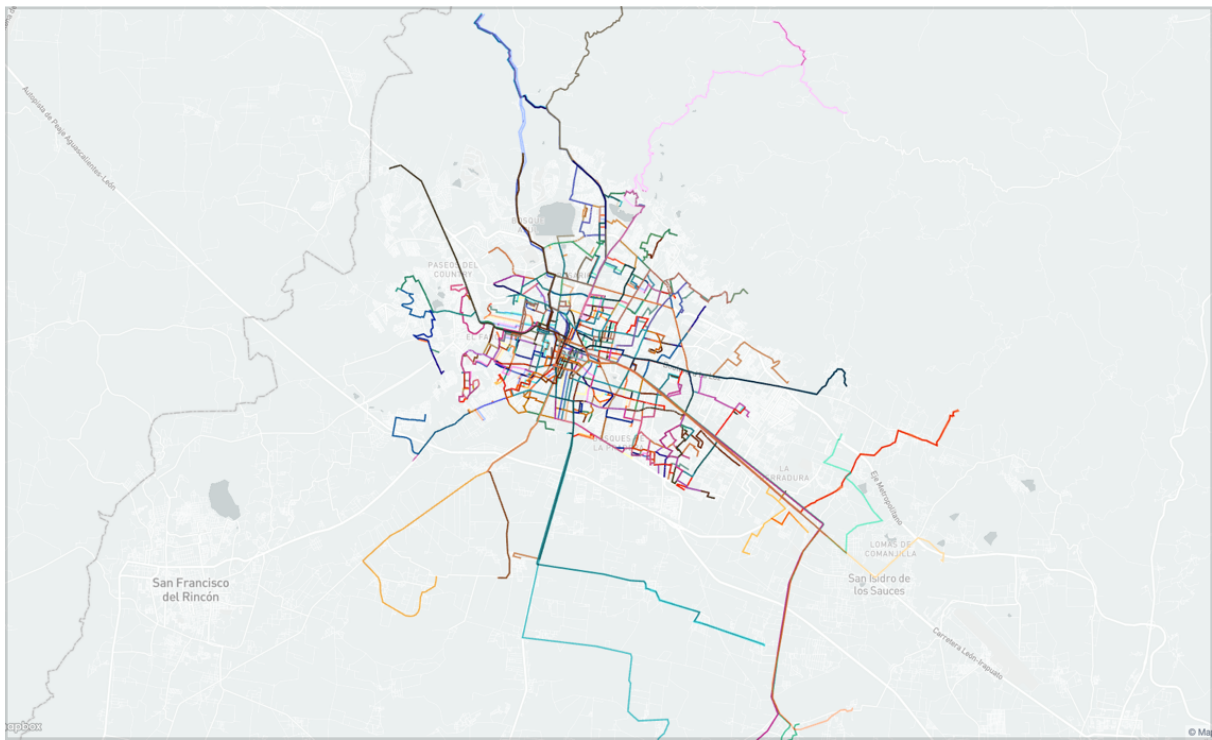
Los datos geográficos, presentan la información en representaciones subjetivas a través

de mapas y símbolos, que representan la geografía como formas geométricas, redes, superficies, ubicaciones e imágenes, a los cuales se les asignan sus respectivos atributos que los definen y describen. Un dato espacial es una variable asociada a una localización del espacio. Normalmente se utilizan datos vectoriales, los cuales pueden ser expresados mediante tres tipos de objetos espaciales:

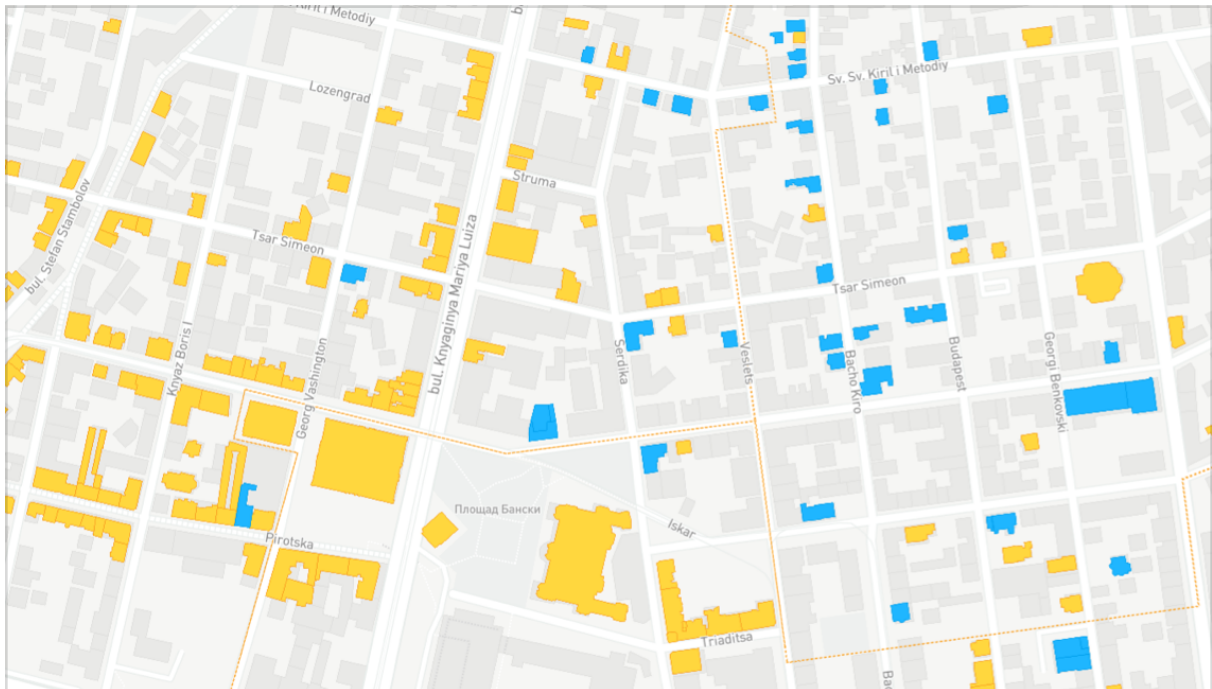
- **Puntos** - Se encuentran determinados por las coordenadas terrestres medidas por latitud y longitud (por ejemplo, ciudades).



- **Líneas** - Objetos abiertos que cubren una distancia dada y comunican varios puntos o nodos, también se le consideran como arcos (por ejemplo, carreteras y vías de trenes o rutas de transporte público).



- **Polígonos** - Figuras planas conectadas por distintas líneas u objetos cerrados que cubren un área determinada (por ejemplo países, regiones o lagos).



## 2.5 ¿Por qué los tipos de datos son importantes?

Los tipos de datos son un concepto importante porque los métodos estadísticos solo se pueden usar con ciertos tipos de datos. Debe analizar los datos continuos de manera diferente a los datos categóricos, de lo contrario, resultaría en un análisis incorrecto. Por lo tanto, conocer los tipos de datos con los que está tratando, le permite elegir el método de análisis correcto.





### 3. Agrupaciones de datos

Si bien ya vimos que los datos son observaciones o mediciones (sin procesar o procesadas) representadas como texto, números, etc. Existen diferentes formas en cómo se pueden agrupar estos datos y es importante saber la diferencia entre un conjunto de datos y una base de datos.

#### 3.1 Conjunto de datos

Un conjunto de datos o dataset, es una colección estructurada de datos que corresponde a los contenidos de una única tabla de datos o una única matriz de datos estadística, donde cada columna de la tabla representa una variable en particular, y cada fila representa a un registro determinado del conjunto de datos en cuestión.

Los formatos más comunes son:

- **Datos estructurados:** CSV, XLS, XML, JSON.
- **Datos geográficos:** KML, GeoJSON, SHP.

#### 3.2 Bases de datos

Las bases de datos, por ejemplo, agrupan los datos en estructuras lógicas y sistemáticas

para luego ser analizadas o procesadas según un propósito. Las bases de datos están formadas por una colección organizada de datos almacenados como múltiples conjuntos de datos, clasificados según un criterio y almacenados en un soporte (digital o no) con el fin de ser consultados y acceder a ellos de manera simple y rápida.

Las bases de datos más comunes son:

- Bases de datos relaciones (SQL).
- Bases de datos documentales (NoSQL).

## 4. Fuentes de datos

Los datos abiertos pueden provenir de cualquier fuente, ya sea gobierno, empresas privadas, sociedad civil, academia, etc.

### 4.1 Portales de datos abiertos

Son plataformas digitales que sirven para almacenar, compartir, conectar y visualizar bases de datos. Son el punto de acceso a la estrategia de una organización por abrir sus datos y un punto de encuentro entre la organización, las empresas, los ciudadanos, los desarrolladores informáticos y los periodistas. Usualmente los portales de datos abiertos contienen lo siguiente:

- **Catálogo de datos** - listado de todas las tablas publicadas.
- **Colecciones de datos** - conjuntos de tablas de datos (pueden tener un enfoque temático u orgánico).
- **Visualización de datos** - área donde se pueden publicar gráficos o vistas de datos.
- **Historias de datos** - narraciones o historias que presentan evidencias sobre el uso de los datos.
- **Sección para desarrolladores** - sección dedicada a los expertos que trabajan extrayendo información del portal.

Ejemplos: [Datos Abiertos de México - datos.gob.mx](https://datos.gob.mx), [Bienvenida - Datamx.io](https://datamx.io)

### 4.2 Metadatos

Los metadatos son datos que describen la estructura y las propiedades de otros datos. Los metadatos tienen la misma lógica que una ficha bibliográfica, en lugar de buscar un libro en todos los estantes de una biblioteca, una ficha te permite conocer el contenido de un catálogo de forma rápida.

Usualmente tiene información como una descripción del contenido del conjuntos de



datos, el nombre de la agencia responsable, la frecuencia de actualización, el número de veces que se ha visitado, la información técnica para conectarla con aplicaciones informáticas, etc.

## Listado de municipios con declaratoria de emergencia y/o desastre

### Datos y Recursos

**Listado de municipios con declaratoria de emergencia y/o desastre**

Archivo con los municipios declarados en estado de emergencia y en estado de desastre debido a los sismos del 07 de Septiembre y del 19 de Septiembre

[Descargar](#)[Más información](#)

### Metadatos

Campo	Valor
Email del publicador	escuadron@datos.mx
Frecuencia	irregular
GUID	municipios-con-declaratoria-de-emergencia-yo-desastre
Idioma	es
Nombre del publicador	Otras
Publicado	2018-04-03T23:37:51Z
Diccionario de Datos	<a href="#">Consultar</a>

### Etiquetas

[declaratoria](#)[emergencia](#)[municipios](#)[sismo](#)

### 4.3 Diccionarios o catálogo de datos

Los diccionarios nos ayudan a entender a qué se refieren los datos que pertenecen a cierto conjunto, en ellos se describe específicamente qué tipo de dato contiene cada columna y la escala en la que se está midiendo. Adicionalmente un catálogo nos puede ayudar a interpretar registro en donde sus valores estén codificados con claves, por ejemplo, claves para municipios, estatus, etc.

## Referencias

[Base de datos - Wikipedia, la enciclopedia libre](#)

[Datos abiertos - Wikipedia, la enciclopedia libre](#)

[¿Qué es un portal de datos abiertos y para qué sirve? - Abierto al Público](#)

[Estándar de Metadatos CDMX - Plaza Pública](#)

[Escala de Likert: Qué es y cómo utilizarla en tus encuestas.](#)

[Data Types in Statistics. Data Types are an important concept of... | by Niklas Donges |](#)

[Towards Data Science](#)

[IBM Knowledge Center](#)

[Datos Georreferenciados. Su importancia en censos, inventarios o muestreos](#)