

Ciencia de Datos

Conceptos Sobre Datos



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia



Without data you're just another person with an opinion.

—W. Edwards Deming, Data Scientist [1]



**“Yo tengo otros
datos”**



¿Cuánta información
se genera en 1 minuto?

2020 *This Is What Happens In An* Internet Minute



"Infographic: What Happens In An Internet Minute 2020", <https://www.allaccess.com/merge/archive/31294/infographic-what-happens-in-an-internet-minute>, marzo 2021



2021 *This Is What Happens In An Internet Minute*







Definición de Datos

Definición de Dato

- **"Dato" es un una colección de símbolos que representan un conjunto de medidas u observaciones acerca de un evento (Harry F. Foxwell)**



Características de los Datos

Características de un “buen” Dato

- **Exacto**
- **Relevante**
- **Representativo**
- **Bien definido**
- **Completo**
- **Granular**

Cauzas de un “mal” Dato

- **Dato mal definido (confuso)**
- **Tipo de dato no especificado**
- **Tipo de dato que no se ajusta a los estándares**
- **Mala recolección o grabación de datos**
- **Mala interpretación de los datos**
- **Error de copia ó transcripción de datos**
- **Datos fraudulentos**
- **Datos perdidos**
- **Datos fuera de rango**
- **Error de almacenamiento digital**
- **Agregar ceros a los datos (0014 en vez de 14)**
- **Colocar la letra "o" en vez del número "0"**
- **Transposición de caracteres (LA por AL)**
- **Inconsistencia de los datos (MEX ó MX)**

Así están las compañías...

https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards

Harvard Business Review

Data | **Only 3% of Companies' Data Meets Basic Quality Standards**

Data

Only 3% of Companies' Data Meets Basic Quality Standards

by Tadhg Nagle, Thomas C. Redman, and David Sammon

September 11, 2017

Conceptos sobre Datos

Datos

- **Tipos de Datos (según su estructura):**
 - **Estructurados**
 - **Semi-estructurados**
 - **No Estructurados**

Datos

- **Tipos de Datos (según su ubicación):**
 - **Local**
 - **HD, SSD, SD, etc.**
 - **En Red Local**
 - **Computadoras cercanas**
 - **Internet**
 - **Páginas estáticas**
 - **Páginas de servicios**
 - **Redes sociales**
 - **Video en streaming**
 - **Videoconferencias**
 - **Internet de las Cosas**



Datos

- **Tipos de Datos (según su formato):**
 - **txt**
 - **CSV**
 - **XLS**
 - **PDF**
 - **XML**
 - **JSON**
 - **HTML**

Datos

- **Tipos de Datos (según su contenido):**
 - **Nominales: Atributos cualitativos:**
 - **Ordinales: Atributos cuantitativos pero no comparables**
 - **Proporcional: Cantidades medibles**
 - **Intervalo: Cantidades medibles (sin haber un cero)**

Datos

- **Tipos de Datos (según su contenido):**

Analytical Data Type	Computational Data Type	Typical Statistics	Typical Visualizations
Nominal, categorical	Character, string	Counts, relative frequencies (%), modes	Bar chart, pie chart*
Ordinal	Character, string	Counts, relative frequencies (%), modes, percentiles, quartiles	Bar chart
Interval	Integer, floating point	Max, min, range, mean (arithmetic), median, mode, differences, std dev, addition, subtraction	Boxplot, histogram
Ratio	Integer, floating point	Max, min, range, median, mean (arithmetic and geometric), ratios, differences, std dev, addition, subtraction, multiplication, division	Boxplot, histogram

Datos

- **Almacenamiento de los Datos:**
 - **Archivos “txt”**
 - **Sistemas SQL**
 - **MySQL, PostgreSQL, SQL Server**
 - **Sistemas NoSQL**
 - **MongoDB**



Datasets

Datasets

- Un “Dataset” es un conjunto de datos que contiene los valores para cada una de las variables, como por ejemplo la altura y el peso de un objeto, que corresponden a cada miembro del conjunto de datos.
- Cada uno de estos valores se conoce con el nombre de dato. El conjunto de datos puede incluir datos para uno o más miembros en función de su número de filas.
- Conjuntos de datos tan grandes que aplicaciones tradicionales de procesamiento de datos no los pueden tratar se llaman big data.

Dataset

- **Características de un buen “Dataset”:**
 - **Localizable**
 - **Accesible**
 - **Interoperable**
 - **Reusable**

Dataset y Diccionarios/Metadatos

- Un “Dataset” requiere un “Diccionario/Metadatos” que describan el contenido del “Dataset.

Metadatos

- **Información a tomar en cuenta al construir Metadatos:**
 - **Nombres**
 - **Descripciones**
 - **Tipos de datos**
 - **Unidades de medida**
 - **Valores de rangos permitidos**
 - **Codificaciones especiales**



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia



rogelio.fe@morelia.tecnm.mx



rogeplus@gmail.com



xumarhu.net



[@rogeplus](https://twitter.com/rogeplus)



[https://www.youtube.com/
channel/UC0on88n3LwTKxJb8T09sGjg](https://www.youtube.com/channel/UC0on88n3LwTKxJb8T09sGjg)



[rogelioferreiraescutia](https://www.linkedin.com/in/rogelioferreiraescutia)

