

Pautas para elegir un Dataset:

1. Cantidad de Columnas: **Más de 10 columnas.** Es recomendable elegir un dataset con un número considerable de columnas para obtener una variedad de información. Si el dataset tiene menos de 10 columnas, pero hay **varias solapas** en el archivo, también puede ser adecuado.
2. Cantidad de Filas/líneas: Aproximadamente 1000. Un dataset con **al menos 1000 líneas** o registros proporcionará suficiente información para realizar análisis significativos en Power BI.
3. Si el archivo de Excel contiene múltiples solapas, es importante analizar si estas solapas se pueden unir utilizando un **campo en común**.
4. **Campo Fecha:** Es recomendable que el dataset contenga al menos un campo de fecha. Esto permitirá realizar análisis temporales y generar visualizaciones basadas en el tiempo.
5. Evita elegir datasets en los que la mayoría de las columnas sean numéricas. Es importante tener información cualitativa que permita segmentar y analizar los datos de manera más completa.
6. Si se cuenta con información geográfica, asegúrense de que cuente con el campo Latitud y Longitud o en caso de tratarse de una ciudad, deben poner ciudad y país. Por ejemplo: "Córdoba, Argentina", "Texas, United State" o si quieren mostrar algo por barrio = "Almagro, Ciudad Autónoma de Buenos Aires, Argentina"

¿Qué es un modelo de datos? Explicado sencillamente, un modelo de datos se trata de varias tablas en donde cada una de ellas se une a otra tabla por medio de una relación, por medio de un campo en común.

Un ejemplo sencillo será el modelo de datos de un Local de ropa

1. Tabla Vendedor: Contiene información sobre los vendedores, como nombre, puesto de trabajo, años de antigüedad, etc. Tiene un campo único para cada fila, como el ID o **el legajo del empleado**.
2. Tabla Cliente: Contiene información sobre los clientes, como nombre, apellido, documento de identidad, fecha de nacimiento, país de residencia, ciudad de residencia, etc. También cuenta con un campo único para cada registro, como el número de **documento de identidad**.
3. Tabla Órdenes de compra: Es la tabla principal que muestra qué sucedió y quién lo realizó. Incluye información sobre la fecha de compra, el vendedor que la realizó, el cliente que la realizó, la fecha en que se efectuó, el monto pagado, el producto comprado y la sucursal en la que se realizó la compra. Tiene un campo único que distingue cada registro, como el **número de orden de compra**.
4. Tabla Sucursal: Contiene los datos de la sucursal, principalmente la ubicación geográfica en donde se encuentra. También cuenta con un campo único que hace único cada línea, **Número de Sucursal**

¿Qué debo hacer con esta tabla para poder crear un tablero? Unirlas entre sí.

- La tabla Vendedor se une a la tabla Órdenes de compra a través del campo del legajo del empleado.
- La tabla Cliente se une a la tabla Órdenes de compra a través del campo número de identidad del cliente.
- La tabla Sucursal se une a la tabla Órdenes de compra a través del campo número de sucursal.

Asegúrese que el dataset que elijan puedan hacer estas uniones. Posiblemente encuentren un Excel con una única pestaña/hoja/solapa, no te preocupes que más adelante en el curso vamos a ir viendo como podemos ir desarmando esa tabla única en varias más pequeñas y sencillas.

Página con Datas para utilizar:

<https://ourworldindata.org/>

<https://www.kaggle.com/datasets>

<https://data.oecd.org/>

<https://datos.bancomundial.org/>

<https://statistics.cepal.org/portal/cepalstat/index.html?lang=es>

<https://archive.ics.uci.edu/ml/index.php>

<https://herramientas.datos.gov.co/>

<https://oec.world/>

<https://atlas.cid.harvard.edu/>

<https://intelligence.weforum.org/>

<https://datos.gob.mx/>

<https://www.inegi.org.mx/siscon/>

<https://www.gov.co/>

<https://data.world/dcopendata>

<https://data.gov/>

<https://enigma.com/>