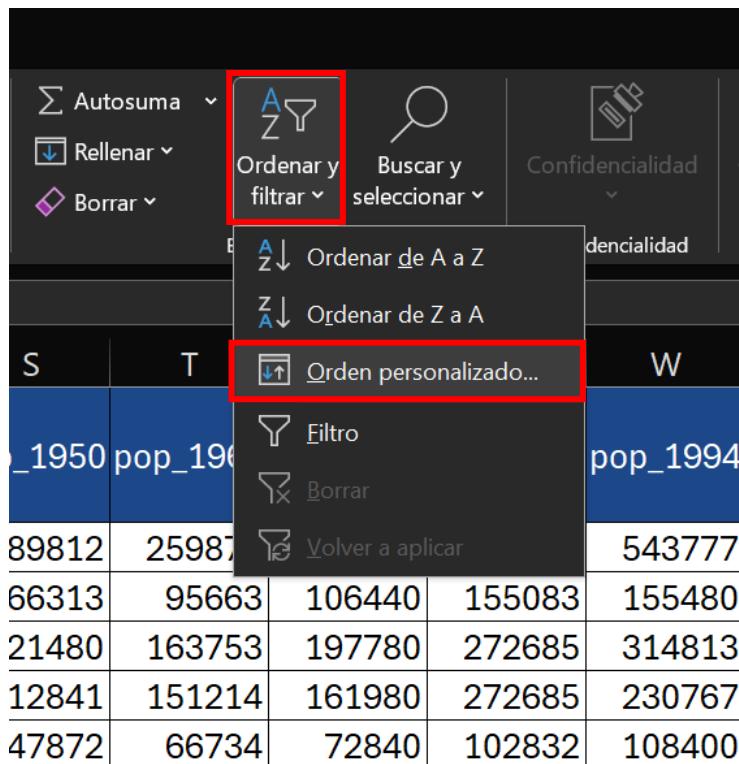


TUTORIAL

Tratamiento de tablas de datos

- Propósito de este tutorial:** Con este tutorial aprenderemos a manipular de forma correcta los datos contenidos en una tabla con información que nos servirán como insumos para realizar análisis espaciales que luego se representarán de forma visible en productos como (mapas, gráficos, tablas dinámicas, entre otros) para la toma de decisiones.
- Revisar los datos, previo a utilizarlos:** Es muy usual encontrarnos con bases de datos con muchos símbolos en lugar de algunas letras, información desordenada, diferentes fuentes, formatos, etc. y para poder trabajar con esas bases de datos es necesario hacerle ciertos tratamientos para tener datos de mayor calidad al momento de realizar análisis.
- Procedimientos para el tratamiento de tablas de datos:** Seguiremos los siguientes pasos para tener datos limpios y ordenados:
 - Ordenar los datos:** El primer paso es abrir el documento de Excel con la tabla de datos, una vez abierta seleccionamos en el menú inicio la opción de Ordenar y filtrar y luego en Orden personalizado.



S	T	W
1950	pop_196	pop_1994
89812	25987	543777
66313	95663	155480
21480	163753	314813
12841	151214	230767
47872	66734	108400

- Corregir errores de ortografía.** (En muchas ocasiones encontramos datos donde, por ejemplo, algunos signos de puntuación, letras como

la ñ son reemplazados por caracteres especiales (ñ®, ñí, etc.) Por tal razón se debe revisar la información y corregir estos errores).

GTM	Guatemala	11	Jutiapa	GT.JU	0		Departamento
GTM	Guatemala	12	Petén	GT.PE	0		Departamento
GTM	Guatemala	13	Quetzaltenango	GT.QZ	0		Departamento
GTM	Guatemala	14	Quiché	GT.QC	0		Departamento
GTM	Guatemala	15	Retalhuleu	GT.RE	0		Departamento
GTM	Guatemala	16	Sacatepéquez	GT.SA	0		Departamento
GTM	Guatemala	17	San Marcos	GT.SM	0		Departamento

c. **Coordenadas precisas:** Tener datos de coordenadas precisas de los lugares objeto de estudio es de mucha utilidad. Coordenadas precisas son esenciales para crear mapas precisos y actualizados, lo que es crucial para la planificación urbana, la gestión de recursos naturales y la respuesta a emergencias. Para verificar si las coordenadas de latitud y longitud son correctas podemos hacer uso de Google Earth Pro (copiamos y pegamos en la aplicación y verificamos si corresponden al lugar o ubicación con la que estamos trabajando).

Department	Latitud	Longitud	pop_1898
Alta Verapaz	15.63553	-90.13244	110936
Baja Verapaz	15.1012	-90.37108	50874
Chimaltenango	14.69391	-90.92038	61013
Chiquimula	14.69011	-89.41828	66823

d. **Estandarizar tipos de letra** (tamaño, mayúsculas, minúsculas) y formatos. Al momento de llevar la tabla a un programa SIG se recomienda quitar todos los formatos de la tabla.

- **Tabla con formato (ejemplo)**

OBJECTID	ID_0	ISO	NAME_0	ID_1	NAME_1	HASC_1	CCN_1	CCA_1	TYPE_1
1	94	GTM	Guatemala	1	Alta Verapaz	GT.AV	0		Departamento
2	94	GTM	Guatemala	2	Baja Verapaz	GT.BV	0		Departamento
3	94	GTM	Guatemala	3	Chimaltenango	GT.CM	0		Departamento
4	94	GTM	Guatemala	4	Chiquimula	GT.CQ	0		Departamento
5	94	GTM	Guatemala	5	El Progreso	GT.PR	0		Departamento

- **Tabla sin formato (ejemplo)**

OBJECTID	ID_0	ISO	NAME_0	ID_1	NAME_1	HASC_1	CCN_1	CCA_1	TYPE_1
1	94 GTM	Guatemala		1 Alta Verapaz	GT.AV	0			Departamento
2	94 GTM	Guatemala		2 Baja Verapaz	GT.BV	0			Departamento
3	94 GTM	Guatemala		3 Chimaltenango	GT.CM	0			Departamento
4	94 GTM	Guatemala		4 Chiquimula	GT.CQ	0			Departamento
5	94 GTM	Guatemala		5 El Progreso	GT.PR	0			Departamento

e. **Títulos:** Nombres de los títulos de las columnas que identifiquen con claridad los datos que contienen. En la siguiente table se muestran los datos de título que venían por default en la tabla de datos.

ID_1	NAME_1	HASC_1	CCN_1	CCA_1	TYPE_1	ENGTYPE_1	NL_NAME_1	VARNAME_1	Department	Latitud	Longitud
------	--------	--------	-------	-------	--------	-----------	-----------	-----------	------------	---------	----------

f. **Identificador:** El identificador (**ID**) es un elemento fundamental en la gestión de datos, ya que permite una identificación única y precisa de cada registro o elemento de datos.

i. **Ventajas de los identificadores:**

1. **Mejora la organización:** Los identificadores permiten ordenar y organizar los datos de manera eficiente.
2. **Facilita la búsqueda:** Los identificadores únicos permiten buscar y localizar rápidamente los datos necesarios.
3. **Evita duplicados:** Los identificadores únicos evitan la creación de registros duplicados.
4. **Mejora la integridad:** Los identificadores ayudan a mantener la integridad de los datos, ya que permiten identificar y corregir errores.

ID_1	NAME_1	HASC_1	CCN_1
1	Alta Verapaz	GT.AV	0
2	Baja Verapaz	GT.BV	0
3	Chimaltenango	GT.CM	0
4	Chiquimula	GT.CQ	0

g. **Depurar:** También es importante depurar la información que no sea relevante y que pueda causar ruido o confusión a la hora de realizar análisis en los sistemas de información geográfica. Las columnas que no tienen datos se pueden eliminar, así como columnas con información irrelevante.

ID_1	NAME_1	HASC_1	CCN_1	CCA_1	TYPE_1	ENGTYPE_1	NL_NAME_1	VARNAME_1	Department	Latitud	Longitud	pop_1898	pop_1940	pop_1950	pop_1964
1	Alta Verapaz	GT.AV	0		Departamento	Department			Alta Verapaz	15.63553	-90.13244	110936	282562	189812	259873
2	Baja Verapaz	GT.BV	0		Departamento	Department			Baja Verapaz	15.1012	-90.37108	50874	96182	66313	95663
3	Chimaltenango	GT.CM	0		Departamento	Department			Chimaltenango	14.69391	-90.92038	61013	177123	121480	163753
4	Chiquimula	GT.CQ	0		Departamento	Department			Chiquimula	14.69011	-89.41828	66823	144011	112841	151214

El ejemplo anterior es una forma de tratamiento de datos utilizando hojas de cálculo de Excel, existen diversas herramientas para hacer este tipo de ajustes, consideramos utilizar esta, por el acceso que todos tenemos a las herramientas de Office 365.