

Tema 5

1. *Introducción a Excel para Big Data.*

Excel es una hoja de cálculo, no una herramienta diseñada para Big Data, por lo que es especialmente importante hacer una buena gestión de los datos cuando se usa para este fin. Es importante hacer un buen tratamiento previo de los datos, eliminando duplicidades y redundancias y simplificando los datos al máximo (sin perder información) antes de cargarlos en Excel.

También puede ser buena idea separar los datos que estén poco relacionados en varias tablas diferentes para trabajar siempre con la cantidad justa de datos y no realizar cálculos innecesarios, o incluso en distintas hojas de cálculo si el volumen de datos es muy grande. Luego, estas tablas pueden unirse si es necesario mediante consultas en Power Query, si están en el mismo archivo, o exportando los datos necesarios a otro documento.

En resumen, para aplicar Excel a conjuntos de datos más grandes y complejos hace falta una forma de trabajar más compartimentada que si se trabajara con una herramienta diseñada para el análisis de grandes cantidades de datos.

2. *Manejo de grandes conjuntos de datos.*

A la hora de trabajar con grandes volúmenes de datos, es conveniente tener en cuenta las siguientes prácticas:

- Fórmulas referenciadas. Usar referencias al aplicar fórmulas permite trabajar de una forma más general y ágil, pudiendo extender las operaciones al rango necesario con facilidad, independientemente del tamaño de este. Además, permite rastrear errores y actualizar más rápidamente la hoja de cálculo.
- Filtros y funciones adecuados. El uso de funciones que realizan operaciones matemático-estadísticas siguiendo un criterio condicional, como SUMAR.SI o CONTAR.SI.CONJUNTO, permiten realizar estas operaciones sin tener que copiar los valores deseados en una nueva columna, ayudando a evitar duplicidades. Por el contrario, añadir una columna con algún valor que sirva de filtro puede ayudar a seleccionar un conjunto de valores sin tener que

realizar un gran número de operaciones condicionales. Usar una u otra técnica dependerá de los datos concretos en cada caso.

- Propiedades de tabla. Convertir conjuntos de datos ordenados en tablas da la posibilidad de aplicar filtros para seleccionar los datos según diversos criterios sin tener que crear tablas nuevas o hacer operaciones adicionales.
- Macros. Las macros son muy útiles para automatizar procesos que se tienen que aplicar repetidamente a conjuntos de datos. Ahorran mucho tiempo si se trabaja habitualmente con el mismo tipo de formato y datos.
- Plantillas en M para el limpiado de datos. Cuando se trabaja frecuentemente con el mismo tipo y formato de datos, se pueden guardar los fragmentos de código en M que reflejan un conjunto de transformaciones sobre los datos importados en crudo. De esta manera, basta con copiar el código en el editor para realizar todas las transformaciones de una sola vez.

Estas prácticas facilitan el uso de Excel en todo su potencial y sacar provecho de los datos.

3. Herramientas de análisis para Big Data

Extraer con Excel grandes cantidades de datos procedentes de bases de datos es posible de forma eficiente gracias a la herramienta Power Query. Power Query es una herramienta de ETL que permite extraer datos de los principales tipos de bases de datos, transformarlos en un editor y cargarlos en una hoja de Excel para su análisis. Tiene la ventaja de estar integrada en Excel y permite realizar consultas y operaciones mediante una interfaz gráfica, sin necesidad de saber SQL.

Otra herramienta útil es Power Pivot, un complemento de Excel necesario para realizar análisis más profundo y modelado de datos. Permite optimizar tareas, crear modelos de datos más complejos que Excel, relacionar tablas y, en general, está pensado para cubrir las necesidades que surgen en el análisis de Big Data.

4. Limitaciones y consideraciones.

Excel cuenta con limitaciones que hacen que no sea la herramienta óptima para Big Data.

La primera y principal limitación de Excel es el número de datos que admite. El número de filas está limitado a algo más de un millón, lo cual es muy pequeño en el ámbito de Big Data.

Otra limitación importante es la velocidad de los cálculos. Excel no es especialmente rápido y el tiempo de ejecución de las operaciones aumenta al aumentar el número de datos hasta hacer el programa ineficiente y poco práctico.

En conclusión, Excel está preparado para un volumen bajo o medio de datos y un análisis genérico de los mismos. Para un análisis en profundidad, con cálculos y representaciones complejas de un volumen grande de datos, es más adecuada una alternativa como Power BI.

5. Recursos adicionales y conclusiones.

A lo largo del curso se ha visto cómo aprovechar las funcionalidades de Excel para crear tablas con las que manejar los datos de forma ágil y eficaz, cómo realizar con este programa el proceso de ETL. Además, se aprendió a hacer gráficos combinados y a usar formatos que faciliten la legibilidad de la información. También se han tratado las limitaciones de Excel y cómo superarlas.

Un recurso útil en este proceso ha sido la página web del propio programa:

<https://support.microsoft.com/en-us/excel>

Aquí es posible encontrar información acerca de todas las funciones disponibles en Excel, cómo realizar distintas tareas, cómo solucionar errores comunes y, en general, es un buen punto de partida para averiguar cómo hacer algo nuevo. Incluye información sobre Power Query y Power Pivot.

Para explicaciones más detalladas, se puede consultar:

<https://exceltotal.com/>

Contiene múltiples entradas sobre cómo realizar acciones concretas, redactadas, con todos los pasos y capturas de pantalla. Contiene también una lista de todas las funciones con sus nombres en español e inglés y los atajos de teclado.

Otra página web con tutoriales paso a paso es

<https://spreadsheeto.com/blog/>

Sus explicaciones son muy detalladas y didácticas, pero está en inglés, lo que hay que tener en cuenta porque el nombre de las funciones es

diferente. Tiene también cursos gratuitos de todos los niveles y de introducción a Power BI.

Además, por supuesto, existe una gran cantidad de video-tutoriales acerca de análisis de datos con Excel y resolviendo prácticamente cualquier duda que pueda surgir.