

Tema 5: Excel para Big Data

1. Introducción a Excel para Big Data

Excel permite trabajar con grandes cantidades de datos gracias a sus numerosas funciones, así como a herramientas que permiten analizar dichos datos o manipularlos. El modelado de datos y las tablas dinámicas también son herramientas muy útiles que facilitan la comprensión y visualización de grandes cantidades de datos de manera sencilla.

2. Manejo de grandes conjuntos de datos

Consejos para trabajar con grandes volúmenes de datos en Excel:

- Utilizar tablas dinámicas que permitan filtrar y modificar la organización de los datos.
- Hacer uso de fórmulas estadísticas y matemáticas para llegar a conclusiones útiles y predecir los siguientes resultados.
- Hacer uso de las herramientas de autocompletado para agilizar el trabajo.
- Mostrar la información de manera visual siempre que se pueda, utilizando gráficos.

3. Herramientas de análisis para Big Data

Dentro del paquete de Microsoft, contamos con un par de herramientas que facilitan el trabajo de Excel, ya que están pensadas para trabajar con todas ellas a la vez.

Power Query es una herramienta de preparación de datos que permite importar datos de diversas fuentes directamente en Excel. De esta manera, puedes transformar los datos directamente en Power Query antes de pasarlos a Excel.

Power Pivot es una herramienta de modelado de datos que permite crear modelos de datos avanzados directamente dentro de Excel. Así, puedes combinar varias tablas y crear relaciones entre ellas, permitiendo hacer análisis multidimensionales y crear cálculos complejos.

Por otra parte, contamos con otras herramientas que pueden integrarse con Excel para realizar el trabajo de forma más eficiente.

Microsoft Power BI permite conectar, modelar y visualizar datos de diversas fuentes. Esto permite que, una vez modelados los datos, puedas crear informes interactivos y paneles de control dinámicos, que pueden integrarse con Excel para analizarlo más profundamente.

SQL Server Integration Services (SSIS) es una plataforma de integración de datos de Microsoft que permite crear y administrar procesos de extracción, transformación y carga de datos. Permite importar grandes cantidades de datos desde otras fuentes para transformarlas y realizar análisis complejos. Una vez listos los datos, pueden exportarse a Excel.

4. Limitaciones y consideraciones

Debemos tener en cuenta que, aunque Excel es una buena herramienta para analizar grandes cantidades de datos, también tiene sus limitaciones, razón por la cual en ocasiones es más adecuado hacer uso de otras herramientas más adecuadas para ello.

Al trabajar con Excel nos vamos a encontrar con limitaciones de memoria, dado que es un programa de escritorio y va a funcionar mejor o peor en función del equipo en el que se esté realizando el trabajo.

También presenta un tamaño de archivo limitado pues, si bien es cierto que parece tener una extensión infinita, dicha extensión está limitada a un millón de filas y 16 mil columnas. Esto es suficiente para muchos trabajos, pero no para grandes empresas.

Por otra parte, una clara limitación de Excel es la obligación de trabajar con datos estructurados, pues no ofrece soluciones para procesar datos no estructurados o semiestructurados. En caso de necesitar trabajar con estos datos, no queda más remedio que acudir a otras herramientas más potentes.

5. Recursos adicionales y conclusiones

Son muchos los cursos, videos o tutoriales que se pueden encontrar en la red para aprender a utilizar Excel dentro del análisis de datos. Aquí están algunos de los que yo he utilizado hasta la fecha:

- Lista de reproducción del tutorial de Excel para principiantes de Alex The Analyst:

<https://www.youtube.com/playlist?list=PLUaB-1hjhk8Hyd5NiPQ9CND82vNodlFF5>

- Curso gratuito de freecodecamp sobre cómo analizar y visualizar grandes bases de datos con Microsoft Excel usando tablas dinámicas y diagramas (How to Analyze and Visualize Large Datasets with Microsoft Excel Using Pivot Tables and Charts)

<https://www.freecodecamp.org/news/analyze-and-visualize-large-datasets-with-excel-pivot-tables-charts/>

- Curso de LinkedIn “Excel: Análisis, gestión y validación de datos (Office 365/Microsoft 365)”
<https://www.linkedin.com/learning/excel-analisis-gestion-y-validacion-de-datos-office-365-microsoft-365>

Conclusiones

Este módulo hemos comenzado analizando todo lo que Excel puede ofrecer a simple vista desde su interfaz de escritorio, así como una comparación con Excel online, llegando a la conclusión de que el Excel de escritorio es mucho más potente y aporta más herramientas para trabajar que el online, aunque para cosas pequeñas podemos acudir a este último dado que es gratuito.

Por otra parte, también hemos aprendido las numerosas formulas y funciones que ofrece Excel, de las cuales conocía muy pocas, y todo lo que podemos hacer con ellas para analizar los datos.

Una vez vistas estas, avanzamos a trabajar con tablas dinámicas y gráficos, viendo cómo se pueden ver los mismos datos de maneras diferentes y agruparlos en función de lo que queremos observar.

Por último, hemos aprendido a importar datos desde otro tipo de documentos, como CSV, Json, etc., utilizando la herramienta Power Query, y hemos visualizado brevemente cómo se utiliza Power Pivot. Estas herramientas pueden ser muy útiles en el trabajo de Excel, pues facilitan las transformaciones y el modelado de datos antes de analizarlo en Excel.

En resumen, ha sido un módulo muy completo en el que tener una primera visión de cómo utilizar Excel para analizar datos y aprender todo lo que puede ofrecer. No obstante, nos queda por delante mucho trabajo para ser expertas en su funcionamiento y utilizarlo de la forma más óptima posible.