

### **Módulo 3: Entrevista con David Monroe - Microsoft**

[00:00:12] **Sandra Crucianelli** Bienvenidos a todos los participantes del nuevo curso masivo en línea sobre periodismo de datos del Knight Center en alianza con Microsoft. Le quiero dar la bienvenida a un invitado muy especial que tenemos, a David Monroe. ¡Hola, qué tal! ¿Cómo estás, David?

[00:00:28] **David Monroe** ¡Hola Sandra! Bien, ¿y tú?

[00:00:30] **Sandra Crucianelli** Gracias por aceptar esta entrevista. David, les cuento a todos los participantes del curso, es gerente de producto en Microsoft y nos va a hablar de una herramienta que es como la madre de todas las herramientas digitales que utilizamos en periodismo de datos, un periodismo de base de datos, y es Excel. David o David [en inglés], ¿qué nos podés decir de esta herramienta tan poderosa para quienes nunca han trabajado con ella?

[00:00:59] **David Monroe** Lo principal es que es muy, muy poderosa y que es muy fácil hacer análisis de datos, visualizar los datos, procesar los datos, de una forma que puede ser sencilla para análisis pequeños o donde estás trabajando con bases de datos de varios gigabytes de datos y donde estás haciendo resúmenes de millones de filas. Es, digamos, que se escala muy bien dependiendo de los escenarios que tengas y de cuántos datos necesites analizar y puedo platicar un poco más de cómo...

[00:01:37] **Sandra Crucianelli** Sí. Te cuento, en periodismo de datos trabajamos con volúmenes de datos grandes. Muchas veces hay que hacer data entry manual, es decir, carga manual de datos, porque los datos originales provienen de bases de datos o de documentos que están en formato cerrado, entonces no podemos importar los datos, y a veces esos documentos están en una fotografía, por ejemplo. Entonces tenemos que hacer carga de datos manuales. Y generalmente tanto cuando hacemos carga de datos manual como cuando hacemos técnicas de extracción de datos tenemos dos tipos de datos: los datos que son palabras y los datos que son números. Y generalmente analizamos variables que son cosas que cambian a lo largo del tiempo. ¿Cómo nos puede ayudar Excel a manejar esa información que generalmente es muy voluminosa?

[00:02:27] **David Monroe** Bueno, Excel tiene, y si quieres te puedo mostrar directamente en el producto.

[00:02:33] **Sandra Crucianelli** La famosa hoja de cálculo.

[00:02:36] **David Monroe** ¿Quieres que empecemos con escenarios más pequeños o vamos directamente a los más voluminosos?

[00:02:44] **Sandra Crucianelli** Yo creo que podemos trabajar con este ejemplo que nos están mostrando y contarle a la gente cómo se organiza una hoja de cálculo, qué son las columnas, qué son las filas y que nos expliques un poco la barra de herramientas, qué tipo de opciones nos da.

[00:02:59] **David Monroe** Claro. Entonces empecemos con la hoja en blanco. Algo que es genial de Excel es que toda la estructura de los datos, del archivo, es visible y está directamente en la unidad principal del programa. En inglés es "what you see is what you get", lo que ves aquí es lo que tenemos. Entonces los datos están organizados en formas de filas y columnas. Tenemos las filas, de uno hasta más o menos un millón de filas, y las

columnas. Entonces los datos los podemos traer directamente de un archivo de texto o de un archivo separado por comas. Y básicamente una vez tenemos los datos acá, sencillamente es como una tabla donde tenemos filas y columnas. Tiene más flexibilidad, pero también permite tener un poco más de estructura con esta función que es las tablas. Entonces, cuando digo acá insertar tabla en un rango normal, puedo convertirlo a una tabla para que tenga un poco más esa estructura de casi como una base de datos con las filas y columnas más estructurados. Entonces, una vez que tenemos los datos acá. Una de las funciones, de mis funciones favoritas es las tablas dinámicas que nos permiten hacer resúmenes y ver la información con diferentes cortes de la misma. Entonces en este momento tenemos ventas por categoría y producto para cada uno de los años, para tres años. Aquí puedo filtrar y ver solo algunos de ellos. Por ejemplo, si quiero ver 2016 para bicicletas de carretera, puedo ver ese dato específico.

[00:05:01] **Sandra Crucianelli** Esa sería la función filtro.

[00:05:03] **David Monroe** Esa es la función de filtro. Entonces es sencillamente para analizar los datos digamos de forma puntual, si quiero ver específicamente para estos datos. Entonces si acá le quito el filtro, puedo ver todo otra vez. Ves que aquí están azules porque está filtrado, quiere decir cuando los números están en azul. Ahora están otra vez en el color normal. Entonces si quiero ahora analizar el resumen como está mencionando las tablas dinámicas, puedo ir directamente y crear la tabla dinámica para estos datos. Entonces aquí ya me está proponiendo la fuente de datos, la tabla 1. Quiero crearlo en una hoja nueva de cálculo. Entonces aquí puedes ver que insertamos una nueva. Y ahora tengo los diferentes, las diferentes columnas de los datos en esta lista. Entonces puedo empezar a resumir los datos. Si quiero ver, por ejemplo, por año, simplemente lo traigo a filas. Ahora veo los diferentes años que tengo en los datos. Y si quiero ver, por ejemplo, el total de las ventas lo puedo traer a valores, y como este es un valor numérico, Excel asume que lo que queremos ver es las ventas, entonces trae el total de ventas.

[00:06:36] **Sandra Crucianelli** Claro, importantísimo es el manejo de las tablas dinámicas porque nos permiten encontrar patrones de comportamiento dentro de grandes volúmenes de datos. En este caso, como vos dijiste, trabajamos con una cantidad determinada de filas, pero podemos tener cientos de miles, hasta millón en algunos casos, hasta más de un millón de filas. Y las tablas dinámicas nos ayudan a ordenar que es lo que también estás explicando.

[00:07:04] **David Monroe** Y por ejemplo, si quiero ver los datos aquí divididos también por año y producto, puedo traer el producto a las columnas y básicamente desglosar la información de esta manera. Entonces es una función muy útil. Otra que..., bueno digamos que este es en cuanto a resumir y agregar los datos de diferentes formas. Otra que me encanta cuando estás empezando a hacer estos primeros análisis de la información son los mini gráficos.

[00:07:48] **Sandra Crucianelli** ¿Podemos ver, David, cómo se crea un gráfico?

[00:07:52] **David Monroe** Claro. Bueno, perdón, tal vez estoy saltando un poco entre los temas. Antes de que nos vayamos de las tablas dinámicas, algo que hemos agregado recientemente al producto y que es extremadamente útil es este análisis de datos. Entonces, cuando voy al inicio, me preguntabas también por la línea de comandos, aquí tengo las diferentes grupos. Entonces puedo..., este es el de inicio donde están las más usadas para insertar. Puedo insertar gráficos para dibujar directamente con una pluma en

la pantalla o pues si quiero también con el mouse. Formar la disposición de página, fórmulas, digamos que todo esto, y una muy útil es datos donde puedo ahorita podemos ir a ese detalle. Pero antes de movernos de las tablas dinámicas, esta es muy poderosa porque hago analizar datos y aquí te vamos a sugerir diferentes análisis de la información. Entonces esto son algoritmos que están procesándose en la nube y donde podemos analizar patrones, analizar outliers de datos que están, que sobresalen y ver resúmenes de esta manera. Entonces, por ejemplo, aquí vemos que categoría accesorios tiene un valor de ventas notablemente superior, puedo insertar este gráfico dinámico donde directamente me trae las ventas agregadas por categoría. Esta es una tabla dinámica directamente, pero también tiene un gráfico dinámico asociado a esa tabla. Entonces, en este caso podemos ver los accesorios comparados con el resto de categorías. Y si hago cambios en esta tabla dinámica, por ejemplo, quiero filtrarlo solo para mostrar accesorios. Ahora, esta gráfica dinámica, también actualiza el valor. Entonces esta es una forma muy sencilla con el análisis de datos de crear las tablas. Pero si quieres un poco más de flexibilidad o quieres crear gráficos directamente desde la tabla, eso sucede en la pestaña Insertar. Entonces, cuando escojo insertar, puedo ir a gráficos recomendados, donde también tenemos diferentes sugerencias para los análisis. O puedo escoger, por ejemplo, quiero ver mis ventas, año, categoría, una columna en 2D -- bueno, aquí los datos que escogí no tienen mucho sentido -- pero...

[00:10:49] **Sandra Crucianelli** Lo importante es que sepan que tienen que marcar las columnas, ¿cierto?, porque cuando utilizaste la función anterior, de analizar datos, ciertamente tenías marcadas, marcado todos los rangos de valores, ¿cierto?

[00:11:06] **David Monroe** Sí, sí. En el analizar datos...

[00:11:08] **Sandra Crucianelli** Lo mismo para insertar gráficos. Digamos hay que marcar, delinear el grupo de datos sobre el cual queremos o analizar datos o insertar un gráfico.

[00:11:19] **David Monroe** Exactamente.

[00:11:20] **Sandra Crucianelli** Bien. La función de analizar datos me parece genial. Me parece una herramienta poderosísima porque ya el programa interpreta lo que yo quiero ver. Muchas veces vemos tendencias, en periodismo analizamos patrones, básicamente cosas que suben, cosas que bajan, que cambian a lo largo del tiempo, y con estos gráficos -- que analizar datos nos tira -- nos ayuda a comprender mucho mejor dónde puede estar un parámetro de noticia. ¡Es fabuloso!

[00:11:54] **David Monroe** Exactamente. Y la idea es hacer esto más sencillo, de modo que aún si no eres un experto, puedas venir y empezar a analizar patrones de forma sencilla.

[00:12:07] **Sandra Crucianelli** Bien. ¿Qué otra función podrías considerar importante para que los periodistas de datos sepan? Por ejemplo, las funciones aritméticas, que vos sabés que los periodistas tienen bastante aversión por los números. Muchas veces los periodistas solemos decir que estudiamos periodismo porque no nos gustaban las matemáticas. Y sí, muchos dicen eso. Y sé que hay funciones que están automáticas, ¿no? Que no necesitamos insertar una fórmula, sino que podemos hacer algunos cálculos de manera automática. ¿Podrías mostrarnos un ejemplo?

[00:12:51] **David Monroe** Claro. Una de las opciones es cuando tienes una tabla puedes incluir directamente, por ejemplo, la fila de totales. Entonces, pues aquí está agregando. No, no es muy, no es un buen ejemplo con este...

[00:13:10] **Sandra Crucianelli** No, pero sí me lo da automático el total, ¿sí?

[00:13:13] **David Monroe** Sí, pero está, ay perdón, está simplemente agregando aquí, entonces, pues la suma del porcentaje, digamos que no tiene mucho sentido. Pero si quiero hacer cálculos, tengo por ejemplo opciones como autosuma. Entonces, y la verdad es que no uso mucho esta función seguido, así que no funcionó.

[00:13:49] **Sandra Crucianelli** Bueno, yo generalmente lo que hago es pinto todo y voy hasta abajo y me la da automática. Y tiene una función que es autocompletar también, que una vez que uno escribe una fórmula en una celda, Excel ya interpreta que necesitás lo mismo en el resto de las filas.

[00:14:10] **David Monroe** Sí, podemos ver ese, pero por ejemplo, el autosuma nos permite aquí tener el total. Puedo decirle que aquí sea la suma y que aquí sea el promedio.

[00:14:20] **Sandra Crucianelli** Sí.

[00:14:22] **David Monroe** Pero sí, lo que mencionas es muy útil. Por ejemplo, si quisiéramos extraer la primera palabra del producto, por ejemplo, podríamos hacer una fórmula o podemos también empezar a escribirlo cadenas, calcetines, luego mallas, luego pantalones. No, no pongo el ejemplo. Pero mencionabas también el autocompletar para...

[00:15:08] **Sandra Crucianelli** Calculo, por ejemplo, una diferencia, vamos a suponer que en la fila 3 en ventas tenés 3.700 y en la fila anterior, si quisieras calcular la diferencia entre las ventas de cadenas y las ventas de calcetines con colocar simplemente el signo igual, y completar una fórmula de resta de dos parámetros Excel automáticamente interpreta que querés hacer lo mismo en todas las filas. Entonces la función autocompletar es muy importante. Ahora una pregunta que tengo, vos hablaste recién de la capacidad de la hoja de cálculo que puede tener hasta un millón de filas. Nosotros en algunos proyectos periodísticos hemos trabajado con más millones de filas, ¿qué pasa cuando necesitamos una hoja de cálculo que tenga cinco o seis millones de filas ahí? ¿Qué ocurre en este caso?

[00:16:15] **David Monroe** En ese caso lo ideal es usar Power Query, el obtener y transformar datos, y el Power Pivot, que son dos de las herramientas de datos más poderosas en Excel. Entonces Power Query, que es obtener datos, nos permite conectarnos directamente a esos archivos que tienen varios millones de filas o directamente a las bases de datos. Puedo conectarme aquí a SQL, puedo conectarme a fuentes de Azure, puedo conectarme a Power Bi, y a diferentes servicios donde puedo tener esos varios millones de filas, y con Power Query fácilmente agregarlos a lo que necesite. Digamos que más allá de un millón no estamos hablando de datos que van a ser necesariamente tan útiles de ver cada una de las filas. Cuando estamos hablando de varios millones, probablemente estamos hablando de o un archivo de texto que viene de otra fuente o de una base de datos directamente. Entonces aquí lo más útil son esas agregaciones que podemos hacer en los datos.

[00:17:32] **Sandra Crucianelli** Lo más común que tenemos es en CSV, en formato de datos separados por comas. Por ejemplo, en pandemia ahora en los repositorios de datos abiertos nosotros manejamos uno para seguimiento de la pandemia, en el país donde yo vivo, y que al día de ayer tenía 10.5 millones de filas. Pero están todos CSV, a veces JSON, pero en general en CSV.

[00:18:03] **David Monroe** Pues no tengo acá un ejemplo con el archivo de texto directamente, pero cuando hago, cuando escojo CSV o archivo de texto puedo escoger directamente en Excel el archivo puedo tenerlo en mi computador y traer los datos directamente a Power Query y cargarlos a lo que llamamos el modelo de datos. Entonces esto es funcionalidad que ya es más cercana a una base de datos. No es simplemente la hoja de cálculo que tiene todo, sino es la hoja de datos y puedo, de hecho déjame mostrarte cómo funciona aquí, cargándolo de esa tabla que ya tengo en Excel. Entonces, en vez de ser esa tabla, puede ser conectando directamente con un archivo CSV. Y aquí lo que puedo hacer es procesar los datos de forma repetitiva. Entonces, digamos que ese archivo que tú estás usando para rastrear los casos de infección, lo actualizan todos los días.

[00:19:18] **Sandra Crucianelli** Claro.

[00:19:18] **David Monroe** Pero no necesitas los datos tal cual vienen sino necesitas, por ejemplo, limpiarlos y agregarlos. Entonces, digamos que no quiero tener la agregación de cada uno de los productos, sino sólo de las categorías. Aquí en Power Query, puedo hacer esa agregación para tener datos que sean mucho más agregados que cada línea detallada. Entonces aquí puedo escoger "agrupar por" y lo interesante es que este no es una sola acción que está sucediendo. No sé si puedes ver el Power Query o si todavía está mostrando sólo Excel. Entonces puedo decir quiero agregar...

[00:20:08] **Sandra Crucianelli** Nos está mostrando el Power Query.

[00:20:09] **David Monroe** Ah bueno, perfecto. Entonces quiero agregar por año. Este es puedes pensar casi como una tabla dinámica, pero donde estoy haciendo los pasos, de forma que puedo repetirlas más adelante. Entonces por año, por categoría y ¿qué quiero ver? Quiero ver la suma de las ventas, entonces total de ventas. Y quiero ver, por ejemplo, el promedio de clasificación. Le voy a decir promedio de clasificación. Entonces, cuando hago esto me da el agregado de esos datos inicialmente. Entonces, cuando digo listo, quiero cerrar y cargarlo. Quiero cerrar y cargarlo en este caso, también en mi modelo de datos y quiero pasarlo aquí como una tabla. Entonces esos cinco millones de datos que tenías de la fuente de datos oficiales, ahora los procesas. Y vienen a Excel como esa agregación que tengo acá.

[00:21:27] **Sandra Crucianelli** Es maravilloso.

[00:21:28] **David Monroe** Lo interesante es que si el archivo tiene el mismo nombre y simplemente lo reemplazas en tu disco duro, cuando yo le haga aquí actualizar, va a volver a aplicar todos esos pasos que yo definí con la nueva fuente de datos. Entonces, si mañana necesitas otra vez ver el resumen, le haces actualizar y te va a traer el nuevo total con los datos actualizados. Lo interesante es que esto lo puedo vincular, por ejemplo, a una fuente de datos en internet, en la nube y traer sin necesidad de descargar esta información.

[00:22:04] **Sandra Crucianelli** Ah, David, es una herramienta maravillosa. Yo te agradezco esta comunicación que hemos tenido, porque a todas las personas que nunca han usado Excel o que están haciendo sus primeros pasos por esta herramienta, me parece que es muy inspirador todo lo que vos contás y nos anima a querer probar y empezar a jugar también con los números y ver qué se puede hacer. Te digo la verdad, yo aprendí Excel sola, probando, abriendo, cargando números y haciendo clic en determinadas funciones que ni sabía que existían. Así que me parece que la curiosidad es algo muy bueno. ¿Qué consejo final podrías darle a todos aquellos periodistas, comunicadores, editores que bueno están haciendo sus primeros pasos con Excel o que nunca han trabajado con la herramienta?

[00:23:03] **David Monroe** Diría que lo que dijiste es muy cierto. Yo también aprendí por mi cuenta, explorando y lo bueno de Excel es que tiene una comunidad de usuarios muy grande. Excel es prácticamente un lenguaje de programación y es el más usado en el mundo. Entonces es simplemente cuestión de ir a internet, buscar si tienes una idea de un análisis que quieres hacer, es muy probable que alguien ya haya tenido que hacer ese análisis y haya documentado cómo lo hicieron. Entonces continuar ese trayecto de aprender y de tener ese conocimiento de la herramienta apoyándote en la comunidad. Tenemos en el Microsoft la comunidad Tech Community, pero también hay muchos otros sitios web donde hay recursos muy, muy útiles. Entonces es simplemente tener esa curiosidad y explorar, y saber que cuando tienes una idea de un análisis, buscar cómo alguien más lo resolvió ayuda mucho. Y hay una comunidad genial alrededor del producto.

[00:24:19] **Sandra Crucianelli** Sí, ciertamente hay una comunidad muy grande. También hay muchos tutoriales de ayuda dentro del mismo sitio de Microsoft. Hay tutoriales en... Yo recuerdo que una de las primeras cosas, hace muchos años, que quería aprender era cómo calcular una variación porcentual en Excel que no tiene una fórmula automática. Pero entonces encontré un montón de tutoriales en YouTube. Y en YouTube había muchas personas que habían colgado tutoriales "¿cómo transformar un archivo CSV en un Excel?", "¿cómo calcular una variación porcentual?". Y cuando yo me consumía todos esos tutoriales, porque muchas veces las cosas que uno no sabe obviamente otras personas ya pasaron por lo mismo y averiguaron la respuesta. Así que bueno David, un gusto hablar con vos. Muchas gracias por esta entrevista.

[00:25:11] **David Monroe** Gracias a ti.

[00:25:11] **Sandra Crucianelli** No, espero que sea inspiradora para los miles de participantes que están tomando este curso. Te agradezco muchísimo y, bueno, cuidense mucho y estén bien.

[00:25:24] **David Monroe** Genial. ¡Gracias!

[00:25:25] **Sandra Crucianelli** No, gracias a vos.