

Módulo 4: Video 2: Visualización de datos

[00:00:12] **Sandra Crucianelli** Hola a todos, espero que estén muy bien. Este es el último módulo de este curso de Introducción al periodismo de datos o periodismo de base de datos, como también se lo conoce, y vamos a continuar aprendiendo Excel. Es una de las herramientas preferidas que usamos para procesar grandes volúmenes de información, incluso para procesar volúmenes más pequeños. Siempre Excel nos da muy buenas respuestas para todos los interrogantes que tenemos cuando estamos trabajando con datos y muy especialmente cuando los datos son numéricos.

[00:00:52] **Sandra Crucianelli** Voy a compartir con ustedes una hoja de cálculo sobre la cual vamos a hacer algunos ejercicios. Por ejemplo, yo estoy trabajando sobre Excel Online aquí en este momento para que visualicen mejor los atributos de cada una de las operaciones matemáticas que vamos a aprender a hacer en esta hoja de cálculo. Aquí tengo una lista de países y aquí una lista de datos que corresponden a 2010 y datos que corresponden a 2020. Los datos podrían tener cualquier atribución, podrían ser personas de una determinada nacionalidad que ingresan a un país si estamos averiguando cómo es un proceso migratorio. Y en este caso yo coloqué países que son de América Latina con determinadas cantidades numéricas que se corresponden con dos años: el año 2010 y el año 2020. Entonces, primero vamos a aprender a cómo hacer sumas. Esto es muy sencillo, simplemente marcan la primera celda y la llevan hasta la última celda que tiene datos. Una vez que hicimos esto, vamos a ir acá a donde hay un símbolo que representa la suma. Es una letra griega, el es el símbolo de la sumatoria. Vamos a hacer clic y una vez que damos Enter ya tengo la suma perfectamente calculada. Lo mismo podría hacer con la columna que se corresponde con el 2020, ¿sí? Marco la suma y allí me queda el cálculo de manera perfecta. Yo puedo borrar este resultado y sabiendo que esta es una suma Excel interpreta que voy a querer hacer lo mismo en la columna al lado. Por lo tanto, si no quiero repetir el procedimiento, puedo marcar la columna donde tengo el resultado y simplemente estirar para tener la misma fórmula en la misma columna y así obtener la suma. Sería lo que está en verde en la columna B el resultado, la suma de todos los datos que aparecen en el año 2010. Y de la misma forma los datos del 2020.

[00:03:15] **Sandra Crucianelli** Ahora, supongamos que necesitamos hacer una resta, una diferencia neta entre dos valores. Yo aquí tengo en la columna C fila 2 el dato que se corresponde con México del 2020. Y aquí en la columna B fila 2, el dato que se corresponde con 2010. Supongamos que yo quiero restar un número y otro. Entonces, ¿qué voy a colocar? Siempre la fórmula comienza con un signo igual. Marco la fórmula de 2020 y le digo menos la fórmula del 2010 y simplemente con autocompletar yo completo las restas. Todas estas son restas que se pueden hacer simplemente con que ustedes hagan el cálculo en una sola celda, lo pueden estirar hacia abajo y van a obtener el mismo resultado. Ahora, yo voy a borrar los que aparecen de la celda 3 en adelante, marco la celda 2 y estiro la celda 1 de la columna B para obtener la diferencia neta. Y si ustedes observan aquí en este sector que es donde van colocadas las fórmulas, la fórmula es simplemente igual C2 menos B2 que sería el atributo de la columna C fila 2 menos el atributo de la columna B en su fila 2.

[00:04:49] **Sandra Crucianelli** Esto para calcular restas. Ahora, ¿cómo ordenamos de mayor a menor? En este caso vamos a suponer que queremos ver cuál es la mayor cantidad del 2010, a quién pertenece. Marcamos toda la tabla, vamos a la pestaña de datos, ordenar intervalo. Como mis filas tienen un dato encabezado, marco encabezado y le digo que me ordene sólo para 2010. Acá tengo dos posibilidades de la A a la Z, que sería del número más chiquito al número más grande, o de la Z a la A, que sería del

número más grande al más chico. Le pido ordenar y aquí, como ustedes pueden ver, tengo ordenado la fila de mayor a menor, la fila que se corresponde con el 2010 porque fue el encabezado que utilicé. Ahora también podría marcar las tres columnas y ordenar el 2020, ¿sí? Voy a la pestaña de datos, ordenar intervalos, marco el 2020 y le digo de mayor a menor, clic en ordenar y aquí voy a tener el año 2020 perfectamente ordenado de mayor a menor. Esto porque a veces necesitamos visualizar dónde se producen los extremos, los mayores números y los menores. Siempre es bueno tener una referencia de qué pasa, como siempre digo, en las puntas. Qué pasa en el extremo inferior y qué pasa en el extremo superior de una tabla. Pero también podría hacer otros cálculos, por ejemplo promedio para calcular promedio hay una fórmula automática. Yo marco todos los datos hasta la última que contiene datos. Voy acá al área donde tengo las fórmulas automáticas y marco promedio. Simplemente hago clic y tengo el promedio de 2010, aquí yo lo voy a marcar en roja para que lo vean. Y si quiero el promedio de 2020 podría haber hecho esto, marcar todos los atributos y la fórmula promedio. O también una vez que calculé el de la columna B, simplemente tomo la celda, la estiro y si la estiro, me calcula el promedio en la columna subsiguiente.

[00:07:14] **Sandra Crucianelli** Ahora veamos cómo se calcula una variación porcentual. Es decir, suponiendo que estos son cantidad de migrantes que van de un sitio a otro y que tienen una determinada nacionalidad, yo quiero qué variación porcentual hay entre el año 2010 y el año 2020. La fórmula, que la quienen aquí porque es muy importante que aprendan a calcular variaciones porcentuales, siempre comienza con el signo igual y la fórmula que deben aplicar para calcular variación porcentual es entre paréntesis el valor final o valor último, menos el valor inicial o el valor primario dividido el valor inicial. Y luego finalmente obtenemos el porcentaje haciendo clic en el ícono de porcentaje. Vamos a practicar. Yo voy a hacer aquí tengo marcada la celda de Honduras. Entonces voy a calcular la variación porcentual y digo: igual, abro paréntesis, valor final, marco el valor de 2020 sobre la celda, menos el valor inicial, marco el valor de 2010, cierro paréntesis, barra de división, el valor inicial y hago clic. Una vez que hago clic, autocompletar, se completa sola porque Excel interpreta que yo quiero hacer lo mismo en el resto de las celdas para ambas columnas. En este caso yo ya tenía formateada la celda para porcentaje y no tuve necesidad de aplicar el botón de porcentaje, pero si les apareciera como un cero coma, digamos como un número decimal, solo tienen que hacer clic aquí para formatear la celda como porcentaje. Esta es una operación matemática importantísima en periodismo de datos. El poder calcular las variaciones porcentuales porque podemos ver los cambios y a través de los cambios podemos encontrar noticias en esos cambios. Y muchas veces estas noticias son verdaderas primicias, cosas que alguien, nadie antes ha visto.

[00:09:29] **Sandra Crucianelli** Y la última fórmula que les quiero presentar es cómo calcular la tasa por cada 100 mil habitantes. Por ejemplo, aquí en este Excel en amarillo tengo los datos de cómo está evolucionando la pandemia de COVID. Las columnas en amarillo son los datos del scraping, que obtuvimos del scraping de la base de datos de la Universidad de Johns Hopkins. Y en las columnas, en celeste, están los cálculos que hemos hecho. Es decir, a partir de los datos de confirmados, fallecimientos y recuperados, podemos calcular la letalidad, el porcentaje de recuperación con relación a los confirmados, la incidencia -- que es una tasa, es decir, la cantidad de contagios por cada 100 mil habitantes --, la mortalidad es la cantidad de muertos por cada millón de habitantes. Y en este caso, en el caso de la tabla que nos ocupa y que estoy mostrándoles a modo de ejercicio, podemos calcular sobre la base de una..., si tenemos en la primera columna países, en la segunda columna tenemos las poblaciones que se corresponden con esos países y datos del 2020 supongamos, sigamos con el ejemplo de los migrantes, que estos son migrantes que han salido de sus países de origen y se han

dirigido a otros. Entonces aquí lo que vamos a hacer de la misma manera comenzamos con el signo menos. Como estamos trabajando con una tasa por cada 100 mil habitantes, voy a colocar primero el número 100 mil. Luego voy a multiplicar con el símbolo asterisco por la cantidad de residentes que se han ido, que es la celda C2, cierro y divido entre la población. Y así tengo lo que sería la tasa, la tasa de migrantes por cada 100 mil habitantes. Yo podría ordenarla ahora de mayor a menor para poder ver en dónde ha sido el mayor impacto. En este caso en Ecuador, vean que tiene la tasa más alta.

[00:11:41] **Sandra Crucianelli** Es decir, qué nos permite, voy a volver, aquí estamos nuevamente. ¿Qué nos permite una tasa por cada 100 mil habitantes? Nos permite comparar países, geografías que tienen distintos tamaños poblacionales y ponemos a todas estas geografías en una misma base poblacional. Podemos hacerlo por cada millón de habitantes, podemos hacerlo cada 100 mil. Esto depende de las cantidades poblacionales, del tamaño poblacional que tengan cada una de estas unidades geográficas. A veces estamos comparando ciudades chicas que tienen 15 mil habitantes o 20 mil y en ese caso deberíamos usar tasa cada 10.000 habitantes. O si son muy pequeñas, que tienen 6/7 mil habitantes, ese conjunto de ciudades podríamos usar una tasa cada mil habitantes. Lo importante es que cuando comparemos geografías, utilicemos la misma base poblacional para poder analizar un conjunto de ciudades o de países y poner a todos sobre la misma base poblacional.

[00:12:46] **Sandra Crucianelli** Espero que practiquen con estos ejercicios a sumar, a restar, a ordenar, a calcular promedios, a calcular variaciones porcentuales y tasas para una misma base poblacional. Estas son las operaciones matemáticas más básicas para comenzar en periodismo de datos.