

Módulo 2: Entrevista con Itxaro Arteta - Animal Político

[00:00:12] **Sandra Crucianelli** En este curso masivo sobre periodismo de datos y periodismo de base de datos organizado por el Centro Knight, le damos la bienvenida a una entrevistada muy especial para nosotros. Se trata de Itxaro Arteta, del sitio Animal Político, y queremos entrevistarla porque su proyecto de periodismo de datos sobre mujeres desaparecidas en México en diciembre del año pasado ganó un concurso, un concurso muy importante organizado por Microsoft y el Centro Knight para capacitación y financiamiento para este importante proyecto. ¡Hola Itxaro, bienvenida!

[00:00:51] **Itxaro Arteta** Hola, muchas gracias por la invitación.

[00:00:54] **Sandra Crucianelli** Bueno, gracias a vos por participar. Contanos un poco de qué se trata este proyecto de periodismo de datos sobre mujeres desaparecidas.

[00:01:03] **Itxaro Arteta** Bueno, sí, para empezar, como quizás saben, en México, pues hay un problema fuerte de desaparición de personas actualmente y el Gobierno ha tratado de hacer distintas recopilaciones de la cantidad de personas que están desaparecidas, que se reportan en cada estado del país. Y entonces nosotros tuvimos, o bueno, se publicó hace un par de años una base de datos ya completa con todos los registros y con un montón de características incluso físicas que tienen que ver, por ejemplo con cosas como color de ojos, color de cabello, este tipo de cuestiones, el estado donde fue reportada desaparecida la persona, la edad que tenía la última vez que la vieron y esto. Entonces en realidad bueno, esa base de datos contiene más de 90 mil registros, son casi 100 mil, con como 70 columnas de características, algunas quizás no muy útiles, entonces la cuestión fue empezar a revisar eso. Y porque también las desapariciones actualmente en México se asocian bastante al problema de inseguridad, de narcotráfico que también hay. Pero al empezar a revisar un poco la base de datos, lo que yo me di cuenta al separar por hombres y mujeres es que los estados donde hay más desaparición de mujeres no son los que tienen un problema de crimen organizado, sino los que tienen problemas de violencia de género, de violencia intrafamiliar, de feminicidios, todo este tipo de cuestiones, ¿no? Entonces, pues ya también reportando un poco más, hablando con expertos, pues es lo que lo que también te dicen que quizá está más ligado a problemas de violencia contra las mujeres, y pues eso se refleja en los datos, ¿no? Y además también otro hallazgo muy, muy duro es que la mayoría de mujeres reportadas desaparecidas en México son menores de edad, menores de 18 años. Lo cual también habla justo de por ejemplo si tiene que ver con trata de personas que se las llevan, pues para utilizarlas y porque es más fácil convencerlas o engancharlas siendo menores de edad. Entonces pues ha sido explorar un poco los datos que nos están demostrando que sí es un problema diferente la desaparición de hombres que la desaparición de mujeres.

[00:03:21] **Sandra Crucianelli** Gracias por esta explicación que nos introduce en el tema. Ahora me gustaría hacerte una pregunta que tienen que ver con las características de esta base de datos. ¿Qué tipo de archivos nos encontramos en esta base de datos? ¿Son archivos que están en formatos abiertos o están en formatos cerrados?

[00:03:40] **Itxaro Arteta** Pues uno de los problemas, de los obstáculos que hemos tenido en la investigación es porque esta base que yo te comento se liberó en formato abierto hace dos años, en sí, digamos, es un Excel o un CSV. Pero después de eso se volvió a hacer..., la información que se ha seguido recopilando ya no está en datos abiertos. Entonces los datos actualizados están en una página de gobierno que uno puede consultar, pero ya no se puede descargar igual que esa que sí se liberó hace dos años

[00:04:18] **Sandra Crucianelli** Y están en una página web, pero no en un archivo como podría ser un PDF. ¿Y cómo hacen la extracción de los datos?

[00:04:24] **Itxaro Arteta** Efectivamente. Pues ahí hemos trabajado con la base pasada para la mayoría de cuestiones, por ejemplo de características. Y en cuestión de números, de cifras más actualizadas -- digo, lo que tiene que ver con el año 2019 - 2020 -- ha sido una extracción manual sólo de algunas cifras que se pueden porque de verdad de ir bajando como estado por estado, las cifras de entre hombres y mujeres, por edades, ¿no? Eso ha sido una parte de minería muy complicada.

[00:04:57] **Sandra Crucianelli** Muy importante. Y cuando hablás de extracción manual, ¿hablás de que hacen data entry manual? ¿Entrada de datos manual?

[00:05:06] **Itxaro Arteta** Sí, porque solamente puedes seleccionar algunas características. Por ejemplo, ver un periodo, un estado y que te arroje las desapariciones de hombres y mujeres por rangos de edad. Pero no puedes recolectar todos los demás datos como estos que te comentaba de las características físicas y demás. Entonces sólo lo hemos utilizado para cosas muy puntuales de tener la cifra total de desapariciones en los años más recientes.

[00:05:35] **Sandra Crucianelli** Y en cuanto a la información que pudieron descargar de la base de datos abiertos en formato CSV -- que son valores separados por comas -- o en formato de Excel, ahí ¿qué trabajo de minería de datos fue necesaria? Es decir, ¿ordenaron por variables? ¿Hicieron algún tipo operación matemática, algún tipo de cálculo? ¿Cómo manejaron la minería de dato de ese gran volumen de información?

[00:06:03] **Itxaro Arteta** Pues ha sido de todo un poco, desde cosas que parecen tontas casi, casi como el problema del manejo de las eñes y de los acentos que tenemos en español y que después cuando lo subes a un programa en inglés, quizá de repente se desconfigura. Entonces desde esta limpieza, por ejemplo. Para cruzar con otras bases de datos de otra información como las de delitos que hay en México, también hemos tenido que ver que estén iguales los nombres de los estados, por ejemplo. Y después pues ha sido mucho de explorar, explorar las posibilidades, explorar las opciones por ejemplo, para sacar totales, ¿no? Como que lo primero es, pues sí, los totales. Empezar a jugar con introducir la variable de edad, entonces eso. Al introducir la variable de edad te das cuenta de cómo el pico de desaparición de mujeres, no de hombres, está en los 15, 16, 17 años, ¿no? Después, por ejemplo, decir bueno, quiero saber cómo ha evolucionado en el tiempo, entonces como está separado por años también hemos jugado con la opción de cómo ha cambiado los lugares donde era más fuerte el problema a través de los últimos años.

[00:07:17] **Sandra Crucianelli** Eso me interesa mucho: encontrar patrones de comportamiento a través de una serie de datos que está contenida en una línea de tiempo. Esto, ¿ustedes lo han podido hacer?

[00:07:29] **Itxaro Arteta** Sí. Efectivamente es parte de lo que hemos hecho y que es interesante porque también se ve cómo el problema ha ido aumentando. Y es justo lo que te permiten los programas de visualización de datos y de manejo de grandes bases de datos como Power Bi, que es el que estamos usando, porque efectivamente tú ves una base de datos con más de 90 mil registros y pues tu mente solo ve números y más números, y es lo que puede ser complicado. Y en cambio a la hora de hacer esto un poco

automatizado, te das cuenta por ejemplo, hicimos una donde cruzábamos edades con años y entonces de repente ahí ves en una gráfica -- que se hace muy rápido -- puedes ver que a partir de 2014, por ejemplo, es cuando empezó a subir la desaparición de menores de edad. Digo en general ha sido una tendencia ascendente desde 2006, que es cuando empezó la guerra contra el narco, pero puedes ver cómo en 2014 es un año en el que hay ese cambio. Entonces es parte de lo que te permite estar explorando con estos programas los datos.

[00:08:32] **Sandra Crucianelli** Vos recién mencionaste una herramienta muy importante que es Power Bi, ¿es la primera vez que habías usado esta herramienta?

[00:08:40] **Itxaro Arteta** Sí, sí. La verdad es que yo había tomado un curso de visualización de datos, había usado otros, pero que son muy..., como que sólo tienes tus datos y los introduces y generas una gráfica.

[00:08:53] **Sandra Crucianelli** Y en este caso, ¿cómo podrías definirla para los periodistas que nunca han trabajado con ella? ¿Qué podrías decirles de tu experiencia trabajando con esta herramienta?

[00:09:06] **Itxaro Arteta** En principio es como un Excel superpoderoso, porque además las gráficas son muy útiles. Es mucho más amigable en ese sentido para hacerlo visual que Excel, ¿no? Yo había trabajado mucho con Excel, con filtros, con tablas dinámicas, un poco. Por ejemplo, las tablas dinámicas operan parecido, pero con esta gran posibilidad de hacer gráficas que manejen cantidades de información enormes y que además son muy fáciles de modificar en el momento, eso es lo que me gusta, que tú a lo mejor te imaginas una gráfica, pero cuando la haces, pues tampoco revela nada definitivo, ¿no? Y es muy fácil, como cambiarle una característica, por ejemplo, introducir la característica edad, entonces a lo mejor ahí es cuando se genera ese patrón que encuentras y que dices "ah, aquí hay un problema", que viendo miles de celdas no lo había visto.

[00:10:01] **Sandra Crucianelli** Itxaro, ¿y cuánto tiempo hace que estás trabajando con datos, con periodismo de datos?

[00:10:09] **Itxaro Arteta** Bueno, ¿con periodismo de datos en general o en este proyecto?

[00:10:16] **Sandra Crucianelli** En general.

[00:10:18] **Itxaro Arteta** Pues es que no sabría porque es algo que siempre me ha gustado mucho, yo les digo que soy muy, muy nerd porque porque sí me gusta, me gustan las bases de datos. Yo por ejemplo, hace pues ya casi diez años, trabajé en un lugar donde cubrí temas de pobreza. Entonces toda la información que se generaba de pobreza en México, que en realidad es bastante rica y bastante amplia, yo la trabajaba mucho. Ha sido por periodos porque después también he cubierto campañas electorales y eso no depende tanto.

[00:10:51] **Sandra Crucianelli** Bueno, pero hace casi una década que estás familiarizada con los datos.

[00:10:57] **Itxaro Arteta** Pues un poco sí. Y bueno, y aquí en donde trabajo actualmente se intenta mucho pues ir más allá de la nota del día. Justamente hacer un periodismo más

de investigación en los que constantemente recurrimos a buscar en bases de datos si hay más información.

[00:11:12] **Sandra Crucianelli** Yo te agradezco mucho esta entrevista, que nos hayas introducido a este trabajo que estás desarrollando en Animal Político que me parece súper interesante. Me quedan dos preguntas para hacerte. Una tiene que ver con tus herramientas digitales preferidas para aquella persona que está comenzando en periodismo de datos, está haciendo sus primeros pasos, ¿qué herramientas digitales o recursos pensás que son importantes para ejercer esta disciplina?

[00:11:42] **Itxaro Arteta** Pues suena muy básico, pero yo soy gran amiga de Excel. Creo que es muy útil justo perderle el miedo y experimentar y probar. Otra cosa, a mí me gustan mucho los mapas, me gusta mucho las visualizaciones en mapas, creo que son muy, muy interesantes. Entonces también hay muchos recursos, desde el Carto hasta, bueno, hasta empezar con un Google Maps donde simplemente con localizar algunos puntos te puedes dar cuenta de estos patrones, de que si algo pasa más en el norte o en el centro o este tipo de cuestiones. Umm no, no recuerdo ahorita cómo se llama, pero bueno, programas para generar gráficas animadas. Creo que también visualmente son muy llamativos. Creo que hay herramientas gratuitas que no, que no aprovechamos a veces y que pueden llamar mucho a nuestra audiencia. Incluso hacer un GIF en redes sociales que llama la atención, cuando tú ves como una rayita despega y te das cuenta de que ahí hay, ahí hay una nota.

[00:12:46] **Sandra Crucianelli** Es como que están vivas las visualizaciones, ¿no? Es la gran diferencia con la infografía, que si bien la infografía también puede ser interactiva en el caso de la visualización a partir de grandes volúmenes de datos con estas nuevas plataformas nos abren un panorama enorme a la hora de mostrar con gráficos los puntos centrales de nuestra historia. Finalmente, la última pregunta tiene que ver con, ¿qué consejo le podrías dar a un periodista que nunca ha trabajado bases de datos y quiere empezar a incursionar en esta temática?

[00:13:23] **Itxaro Arteta** Pues no tenerle miedo, pero sí saber que requieres tiempo, ¿no? Quizá la falta de práctica es lo único que a veces nos aleja de poder hacer este periodismo de datos. Cuando ya te familiarizas ya lo haces mucho más rápido. Pero primero, quizá como cuando aprendimos a usar una computadora y cuando aprendimos a usar Excel, es solo sentarte, estar probando, estar probando opciones, estar descubriendo también las posibilidades que tiene el programa que te ofrece y solamente dedicarle tiempo hasta que vas a encontrar una gran historia y un gran patrón en una base que puede parecer inmanejable al principio.

[00:14:03] **Sandra Crucianelli** Bueno, muchísimas gracias por estos consejos que sin duda son y serán de mucha utilidad para todos los participantes de este curso. Felicitaciones nuevamente por el proyecto y por todo el trabajo que estás haciendo y mis mejores deseos para que este proyecto escale mucho más todavía y puedan continuar poniendo luz allí donde hay sombra. ¡Muchísimas gracias!

[00:14:29] **Itxaro Arteta** Gracias a ti. Hasta luego.