**Qué es la Ciencia de Datos**

La ciencia de datos es un proceso, no un evento. Es el proceso de usar datos para entender diferentes cosas, para entender el mundo. Para mi es cuando tienes un modelo o hipótesis de un problema, y tratas de validar esa hipótesis o modelo con tus datos. La ciencia de datos es el arte de descubrir el conocimiento y las tendencias que se esconden tras los datos. Es cuando traduces los datos a una historia. Así que cuentas una historia para generar conocimiento. Y con este conocimiento, puedes hacer decisiones estratégicas para alguna compañía o institución. La ciencia de datos es un campo a procesos y sistemas para extraer datos de varios formas ya sea de manera estructurada o no. La ciencia de datos es el estudio de los datos. Así como las ciencia biológica es el estudio de la biología, la ciencia física el estudio de las reacciones físicas. Los datos son reales, tienen propiedades reales, y necesitamos estudiarlos si vamos a trabajar con ellos. La ciencia de datos involucra datos y algunos signos. La definición o el nombre se forjó en los 80s y 90s cuando algunos profesores estaban trabajando en los planes de estudio de estadística, y decidieron que seria mejor llamarlo ciencia de datos. ¿Pero que es la Ciencia de Datos? Yo vería la ciencia de datos como el intento de uno por trabajar con los datos, para encontrar respuestas a las preguntas que se están explorando. Más concretamente, es más sobre datos de lo que es sobre ciencia. Si tienes datos y tienes curiosidad, y trabajas con datos, los manipulas, los exploras, el ejercicio de realizar el análisis de datos, tratar de obtener algunas respuestas de ello, eso es ciencia de datos. La ciencia de datos es relevante hoy porque tenemos toneladas de datos disponibles. Solíamos preocuparnos sobre la falta de datos. y ahora tenemos un diluvio de datos. En el pasado, no teníamos algoritmos, ahora los tenemos. En el pasado el software era caro, ahora es de libre uso y gratis. En el pasado, no podíamos almacenar grandes cantidades de datos, ahora por una fracción del costo, podemos tener millones de conjuntos de datos a un costo muy bajo. Así que, las herramientas para trabajar con los datos, la variabilidad de los datos, la habilidad para almacenar y analizar datos, todo es barato, está disponible, es ubicuo y está aquí. Nunca ha habido una época mejor para ser un científico de datos.

**Los Muchos Caminos hacia la Ciencia de Datos**

La ciencia de los datos ni siquiera existía cuando yo estaba creciendo. No es algo con lo que haya despertado y dicho, quiero ser una científico de datos cuando sea grande. No, eso no existía. No sabía que estaría trabajando en la ciencia de los datos. Cuando crecí, no había ese campo llamado ciencia de datos. Y creo que realmente nuevo. La ciencia de datos no existió hasta el 2009, 2011. Alguien como DJ Patil o Andrew Gelman acuñaron el término. Antes de eso, eran solo estadísticas. Y yo no quería ser nada de eso. Yo quería estar en el negocio. Entonces encontré la ciencia de datos algo mucho mas interesante. Yo estudié estadística, así es como empecé.

Pasé por muchas diferentes etapas de mi vida en donde quería ser cantante y luego doctor. Y entonces me di cuenta que era buena en matemáticas. Así que escogí un área que se enfocaba en los análisis cuantitativos. Y desde entonces creo que quería trabajar con datos. No necesariamente ciencia de datos como la conocemos hoy. La primera vez que tuve contacto con la ciencia de datos, fue en mi primer año como ingeniero mecánico. Y las firmas de consultoría estratégica usan la ciencia de datos para tomar decisiones. Así que fue mi primer contacto con la ciencia de datos. Tenía un problema complicado que necesitaba resolver, y las técnicas usuales que teníamos no podían ayudarme a resolver el problema. Me gradué con un título en matemáticas en la peor época posible, justo después de la crisis económica, y tu realmente tienes que ser útil para obtener un trabajo. Así que fui a obtener un título en estadísticas. Y entonces trabajé los suficientes empleos que llamaban a científicos de datos que de pronto me convertí en uno. Mi licenciatura fue en negocios, y mi especialización en política, filosofía y economía. Y después hice una maestría en negocios analíticos en La Universidad de Nueva York en la escuela de negocios Stern. Cuando salí de mi licenciatura, la primera compañía a la que me uní, resultó que que estaban analizando datos de puntos de venta electrónicos para fabricantes minoristas. Y lo que estábamos haciendo era ciencia de datos. Pero realmente comenzamos a utilizar ese término mucho después. De hecho, diría que hace cuatro o cinco años es cuando empezamos a llamarlo analistas y la ciencia de datos. Tuve varias opciones para mi internado aquí en Canadá. Y una de las opciones era trabajar con la ciencia de datos. Yo solía trabajar en desarrollo de proyectos. Pero creo que fue una buena opción. Y entonces comencé mi internado como científico de datos. Soy ingeniero civil por capacitación, así que todos los ingenieros trabajan con datos. Diría que el uso convencional de ciencia de datos en mi vida empezó con la investigación de transporte. Empecé construyendo grandes modelos intentando pronosticar el tráfico en las calles, intentando determinar la congestión y emisión de gases de efecto invernadero o las emisiones de escape. Así que creo que así mi comienzo. Y comencé construyendo estos modelos cuando aun era un estudiante de la Universidad de Toronto.

Empecé trabajando con una gran cantidad de sets de datos, buscando muestras de hogares, digamos, 150,000 hogares de medio millón de viajes. Y so, también, estoy hablando de mitad de los 90s cuando esto era supuestamente una gran cantidad de set de datos, pero no en los términos de hoy en día. Pero así es como comencé. Y he continuado trabajando con esto. Y entonces me cambié a la Universidad McGill en donde era profesor de ingeniería de transporte. Y construí aun modelos de datos mas grandes que involucraban datos y análisis. Y por lo que diría, si, la investigación de transporte me trajo a la ciencia de datos. [MÚSICA] Soy todo oídos.

**Consejos Para los Científicos de Datos**

Mi consejo para un aspirante a científico de datos es que debe ser curioso, bueno para argumentar y crítico. La curiosidad es una cualidad obligatoria. Si no eres curioso, no sabrías qué hacer con los datos. El ser crítico es necesario porque si no tienes ideas preconcebidas sobre las cosas no sabrías por dónde empezar. Debes ser bueno argumentando pues si tienes esa habilidad, y debes abordar un tema, por lo menos puedes empezar en un punto y luego aprender de los datos, para luego modificar tus suposiciones e hipótesis y tus datos te ayudarán a aprender. Y podrías empezar en el punto equivocado. Podrías decir: "Yo pensé, yo creí ésto, pero ahora gracias a los datos yo sé ésto". Por lo tanto, esto permite un proceso de aprendizaje. Así, la curiosidad te ayuda a tomar una posición, una fuerte posición, para partir de ahí y seguir adelante. La otra cosa que el científico de datos necesita es algo de comodidad y flexibilidad con las plataformas de análisis: algunos programas, alguna plataforma de computación, pero eso es secundario. Lo más importante es la curiosidad y la capacidad de tomar posiciones. Una vez que has hecho esto, una vez que has analizado, entonces ya tienes algunas respuestas. Y esto es lo último que necesita un científico de datos: la capacidad para contar una historia. Que una vez que tengas tu análisis, una vez que tengas tus tablas, deberías ser capaz de contar una gran historia de ello. Porque si no logras contar una gran historia, tus hallazgos permanecerán ocultos, enterrados, y nadie los sabría. Tu ascenso a la trascendencia depende bastante de tu capacidad de contar grandes historias. Un punto de partida sería ver cuál es tu ventaja competitiva. ¿Quieres ser un científico de datos en cualquier campo o un campo específico? Porque, supongamos que quieres ser un científico de datos y trabajas para una empresa de IT o una empresa de Internet o basada en la web, entonces necesitas un conjunto diferente de habilidades. Y si quieres ser un científico de datos en la industria de la salud, entonces necesitas un conjunto diferente de habilidades. Así que primero averigua lo que te interesa, y cuál es tu ventaja competitiva. Tu ventaja competitiva no es necesariamente tu capacidad de análisis. Tu ventaja competitiva es tu comprensión de algún aspecto de la vida donde tus habilidades exceden las de los demás. Tal vez sean las películas, tal vez sea el comercio, tal vez es la salud, o quizás la computación. Una vez has averiguado donde se encuentra tu experiencia, entonces empiezas a adquirir habilidades analíticas. Qué plataformas aprender y esas plataformas, estas herramientas, deben ser específicas para el sector que te interesa. Y una vez que has adquirido algún dominio de las herramientas, lo siguiente debe ser aplicar tus habilidades a problemas reales, y mostrarle al resto del mundo lo que puedes hacer.

Ciencia de Datos: El Trabajo más Sexy del Siglo XXI

**Ciencia de Datos: El Trabajo más Sexy del Siglo XXI**

En el mundo basado en datos, los científicos de datos han surgido como un producto de moda. El objetivo es encontrar el mejor talento en ciencia de datos. Los expertos estiman que millones de empleos en ciencia de datos podrían quedar vacantes por falta de talento disponible. La búsqueda global de científicos expertos en datos no es simplemente una búsqueda de estadísticos o informáticos. De hecho, las empresas están buscando personas completas que posean la experiencia en el tema, cierta experiencia en programación y análisis de software y habilidades excepcionales de comunicación.

Nuestra huella digital se ha expandido rápidamente en los últimos 10 años. El tamaño del universo digital era de aproximadamente 130 billones de gigabytes en 1995. Para 2020, este número aumentará a 40 trilliones de gigabytes. Las empresas competirán por cientos de miles, sino millones, de nuevos trabajadores necesarios para navegar por el mundo digital. No es de extrañar que la prestigiosa *Harvard Business Review*califique a la ciencia de datos como "el trabajo más sexy del siglo XXI".

Un informe del Instituto Global McKinsey advierte sobre una gran escasez de talento para datos y análisis. "Para 2018, solo Estados Unidos podría enfrentar una escasez de 140,000 a 190,000 personas con profundas habilidades analíticas, así como 1.5 millones de gerentes y analistas con los conocimientos para usar el análisis de big data para tomar decisiones efectivas".

Debido a que la revolución digital ha tocado todos los aspectos de nuestras vidas, la oportunidad de beneficiarse al aprender sobre nuestros comportamientos es más que nunca. Con los datos correctos, los especialistas en marketing pueden echar un vistazo a nuestra formación de hábitos. La investigación en neurología y psicología revela cómo se forman los hábitos y las preferencias y los minoristas como Target se benefician de ello. Sin embargo, los minoristas solo pueden hacerlo si tienen científicos de datos trabajando para ellos. Por esta razón, es "como una carrera armamentista contratar estadísticos hoy en día", dijo Andreas Weigend, el ex jefe científico de Amazon.com.

Todavía existe la necesidad de convencer a los ejecutivos de C-suite de los beneficios de los datos y el análisis. Parece que la alta gerencia podría estar uno o dos pasos detrás de la gerencia media para estar informada sobre el potencial de la planificación basada en análisis. El profesor Peter Fader, que gestiona la *Customer Analytics Initiative*en Wharton, sabe que los ejecutivos llegan al C-suite sin tener que interactuar con los datos. Él cree que el cambio real sucederá cuando los ejecutivos estén bien versados en datos y análisis.

SAP, líder en datos y análisis, informó en una encuesta que el 92% de las empresas que respondieron en su muestra experimentaron un aumento significativo en sus tenencias de datos. Al mismo tiempo, tres cuartos identificaron la necesidad de nuevas habilidades de ciencia de datos en sus empresas. Accenture cree que la demanda de científicos de datos puede superar la oferta en 250,000 solo en 2015. Una encuesta similar de 150 ejecutivos realizada por KPMG en 2014 encontró que el 85% de los encuestados no sabían cómo analizar los datos. "La mayoría de las organizaciones no pueden conectar los puntos porque no entienden completamente cómo los datos y los análisis pueden transformar sus negocios", Alwin Magimay, jefe de digitales y analítica de KPMG UK, dijo en una entrevista en mayo de 2015.

Bernard Marr escribiendo para Forbes también plantea preocupaciones sobre el talento analítico insuficiente. "Simplemente no hay suficientes personas con las habilidades necesarias para analizar e interpretar esta información, transformándola de datos numéricos en bruto (u otros) en ideas procesables, el objetivo final de cualquier iniciativa impulsada por Big Data", escribió. Bernard cita una encuesta realizada por Gartner de líderes empresariales de los cuales más del 50% informó la falta de experiencia interna en ciencia de datos.

Bernard informó sobre Walmart, que recurrió al crowdsourcing para sus necesidades analíticas. Walmart se acercó a Kaggle para organizar una competencia para analizar sus datos propietarios. El minorista proporcionó datos de ventas de una lista restringida de tiendas y pidió a los competidores que desarrollaran mejores pronósticos de ventas basados en esquemas de promoción.

Dada la escasez de científicos de datos, los empleadores están dispuestos a pagar grandes cantidades por el talento. Michael Chui, director de McKinsey, lo sabe muy bien. La ciencia de datos "se ha vuelto relevante para todas las empresas... Hay una guerra por este tipo de talento", dijo en una entrevista. Tomemos a Paul Minton, por ejemplo. Estaba ganando $ 20,000 sirviendo mesas en un restaurante. Se había especializado en Matemáticas en la universidad. El Sr. Minton tomó un curso de programación de tres meses que lo cambió todo. Hizo más de $ 100,000 en 2014 como científico de datos para un startup web en San Francisco. "Seis cifras, desde el principio... Para mí, fue sorprendente", dijo el Sr. Minton.

¿Podría el Sr. Minton ser excepcionalmente afortunado, o son salarios tan altos la norma? La suerte tuvo poco que ver con eso; el New York Times reportó $ 100,000 como el salario base promedio de un ingeniero de software y $ 112,000 para científicos de datos.