

Big data, big fails:

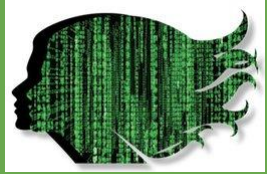
Desmitificando el Big Data en la industria

Jessica Barbosa

4 de marzo de 2019



Agenda



- Yo
- Contexto
- Algunos problemas
- Algunas recomendaciones

Yo



- OPI Analytics www.opianalytics.com
- Consultoría y producto
- ¡¡Estamos contratando!! :D

<https://opianalytics.com/cientifico-datos>

Contexto



- Empresa ya sabe que los datos dan valor
- Ya hay presupuesto asignado para hacer cosas con datos
- Suficientemente grande como para querer armar un área de analítica avanzada
- Quizá ya contrató alguna solución empresarial de *big data*
 - IBM, Oracle, SAS, etc.
- Suponemos situación de arriba hacia abajo
 - No necesito convencer a nadie de que los datos son útiles
 - No necesito conseguir presupuesto para contratar o capacitar
- Estamos suponiendo que ultrajefes ya compraron la idea y ya soltaron dinero para armar un área de analítica avanzada. ¿Qué se hace después?

Algunos problemas

Camino problemático

- Empresas deciden tener un área de analítica
- Contratan a mucha gente y le pagan mucho dinero
- Alguien elige los problemas que deben resolver
- Pasa el tiempo, no ven resultados
- Corren a todos / los presionan y se acaban yendo



Empresas deciden tener un área de analítica

- ¿Qué problemas son los importantes para el negocio?
 - No todo se soluciona analizando sentimientos en Twitter
- ¿Dónde voy a poner a esa área?
 - ¿Dentro de finanzas? ¿De marketing? ¿De TI?
- ¿Qué espero que resuelva? ¿Qué no puedo hacer hoy tal que necesito esa área?
 - Los silos de información no desaparecen mágicamente
 - Los datos no van a mejorar su calidad mágicamente
- Cambios sustanciales toman tiempo y muchas juntas



Contratan a mucha gente y le pagan mucho dinero

- Especialización
 - Modelado
 - Limpieza de datos
 - Desarrollo de software
 - Manejo de infraestructura
- Un posgrado no da superpoderes para manejar gente, planear proyectos, calcular presupuestos, diseñar estrategia....



Alguien elige los problemas que deben resolver

- Falta de datos
 - Todo el mundo cree que tiene datos limpios y de calidad
- Falta de infraestructura
 - Servidores locales: TI tiene tiempos de respuesta lentos y necesita todo bien definido
 - Nube: Calcular costos para presupuestar no es trivial; varía cada mes
- Área de analítica como de investigación: problemas sacados de artículos que están de moda
 - Chatbots
 - Análisis de redes sociales
- Moscas y bazookas
 - Problemas mal definidos y poco recurrentes (Excel)
 - Problemas bien definidos y recurrentes (Sistemas a cargo de TI)
 - ¡¡¡Y todo lo intermedio!!! (Analítica avanzada)



Pasa el tiempo y no ven resultados

- Choque cultural: intento y error
 - Falsas expectativas de que todo salga perfecto a la primera y rápido
- Problemas complicados que toman mucho tiempo
 - Alguien que haya visto los datos tiene que decir qué es factible producir
- Problemas irrelevantes para el negocio
 - Combinar cosas rápidas (ej. automatizar reportes mensuales) con cosas valiosas (tener datos integrados)
 - ¿Qué decisiones cruciales se toman que podrían tomarse usando datos?



Algunas recomendaciones

Recomendaciones: Crear un área de analítica

- Estrategia
 - Amarrar proyectos del área a cosas importantes para el negocio
 - MVPs y pruebas de concepto
 - Cambios sustanciales toman tiempo y muchas juntas pero son necesarios
- Choque cultural
 - Ayuda que otros jefes estén involucrados y convencidos
 - Desarrollo ágil, iteraciones, experimentos
- Contrapeso
 - Si el área responde a TI, ten a alguien de negocio a cargo del área de analítica
 - Si el área responde a Marketing, ten a alguien de TI a cargo
- Mantener viva el área
 - Expectativas del área
 - Comunicar resultados



Recomendaciones: Liderar un área de analítica

- Definición y selección de problemas
 - Mostrar resultados rápido en versiones
 - Cambios fuertes toman tiempo y juntas (ej. Integración de datos)
- Reclutamiento
 - Es caro y difícil contratar gente con experiencia
 - Juniors brillantes si y sólo si y hay un buen plan para que tengan formación completa
- Especialización de roles
 - Incluir gente que sepa administrar proyectos
- Procesos
 - Más gente no da más velocidad!!
 - Documentación, gobernanza, automatización
- Comunicación de resultados
 - No es magia! Expectativas realistas



Recomendaciones: Encontrar trabajo como data scientist

- ¿Qué se espera de mí?
 - El hecho de que yo tenga un posgrado no significa que entiendo / puedo / debo hacerlo TODO
 - ¿Cómo son mis entregables: PDFs, csv's, software productivo?
 - Mis jefes tienen que entender las diferencias entre hacer modelos, limpiar datos, generar sistemas productivos.
- ¿Quién hace administración de proyectos?
 - Si es mi primer trabajo, alguien que sepa más que yo me tiene que ayudar a organizarme, definir problemas, versionar resultados.
- ¿En qué tipo de proyectos trabaja el área?
 - Enfoque en el problema y no en que uses ciertos algoritmos o tecnologías
- ¿Qué mecanismos hay para que yo siga aprendiendo?
 - Todo se mueve rápido; debe haber planes para que el equipo se siga preparando



Recomendaciones: Ser data scientist en una empresa

- No es el fin del mundo si eres el único en el equipo
 - ¿Qué tanto depende la empresa de lo que produces?
- Un data scientist no puede ni debe hacer todo
 - ¿Cuáles son mis fortalezas? ¿Mis entregables se alinean con eso?
- Alguien tiene que definir el problema y ayudarte a ejecutar
 - Problemas factibles vs. imposibles
 - Proyectos largos deben tener entregables intermedios
- Comunicación de resultados
 - El lenguaje y las prioridades de la gente a la que le entregas resultados
 - No hacemos cosas sencillas; la gente tiene que entender las implicaciones de lo que entregamos



Referencias

Referencias

- **Data science as a team sport** - Angela Bassa en rstudio-conf-2019
 - <https://resources.rstudio.com/rstudio-conf-2019/data-science-as-a-team-sport>
- **Don't let Data Science become a scam** - Seth E Dobrin
 - <https://medium.com/ibm-analytics/dont-let-data-science-become-a-scam-d04840849249>
- **How to f**k up an Advanced Analytics team : A simple 20 step process** - Ritesh Mohan Srivastava
 - <https://www.linkedin.com/pulse/how-fk-up-advanced-analytics-team-simple-20-step-srivastava/>
- **How NOT to f**k an Advanced analytics team : A simple 10 step guide** - Ritesh Mohan Srivastava
 - <https://www.linkedin.com/pulse/how-fk-advanced-analytics-team-simple-10-step-guide-srivastava/>



En resumen...

- Tener un área de analítica no va a solucionar todos mis problemas
- No todo se soluciona a billetazos
- Para que rinda mi inversión necesito procesos y tiempo
- Ojo con los problemas / tecnologías de moda
- La gente que sabe de datos tendría que elegir y acotar los problemas
- Matices de la chamba de un data scientist
-

