
Big Data or Big Analytics?

Allan Gray
Partner, End-to-End Analytics

*III Congreso Internacional
de Ingeniería Informática*

10 Agosto 2017



■ **Introducción**

- Acerca de End-to-End Analytics

■ **Big Data, Analytics, y Machine Learning**

- Qué son, realmente?
- Donde está el valor económica? Un punto de vista contraria

■ **Casos de Éxito en América Latina**

- Agroindustria: Subasta de Terrenos
- Retail: Detección de Ruptura
- Ropa Deportiva: Reabastecimiento de Tiendas

▪ **Introducción**

- Acerca de End-to-End Analytics

▪ **Big Data, Analytics, y Machine Learning**

- Qué son, realmente?
- Donde está el valor económica? Un punto de vista contraria

▪ **Casos de Éxito en América Latina**

- Agroindustria: Subasta de Terrenos
- Retail: Detección de Ruptura
- Ropa Deportiva: Reabastecimiento de Tiendas

Acerca de End-to-End Analytics



- Empresa consultora especializada en analítica de cadena de abastecimiento
- Clientes incluyen #1 o #2 en firmas globales en los rubros:
 - Automotora
 - Bebidas y Alimentos
 - Bienes de Consumo
 - Productos Médicos
 - Búsqueda internet
 - Semiconductores
 - Cosméticos
 - Tecnología
- Publicamos extensamente en:



Empresa

- Fundada in 2005
- Basada in Palo Alto, CA

Equipo

- 55 profesionales
- Antecedentes:
 - McKinsey & Co.
 - Procter & Gamble
 - Hewlett Packard
 - Nestlé
- etc.

Otras Credenciales

- Patentes múltiples
- Enseñamos en Stanford & UC Berkeley

▪ Introducción

- Acerca de End-to-End Analytics

▪ Big Data, Analytics, y Machine Learning

- Qué son, realmente?
- Donde está el valor económica? Un punto de vista contraria

▪ Casos de Éxito en América Latina

- Agroindustria: Subasta de Terrenos
- Retail: Detección de Ruptura
- Ropa Deportiva: Reabastecimiento de Tiendas

Qué Significa “Big Data”?

“Conjuntos de datos que son tan grandes o complejos que el software tradicional de procesamiento de datos es inadecuado para tratar con ellos”

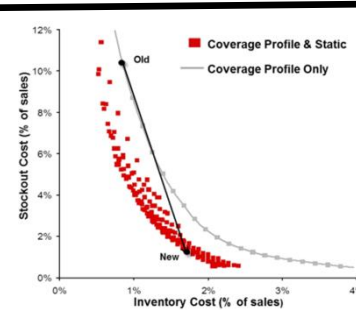


Volumen Velocidad Variedad

Qué Significa “Analytics”?

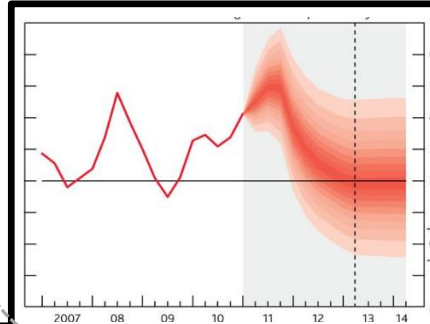
Nuestro enfoque

Prescriptivo



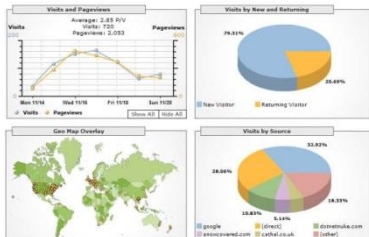
¿Qué deberíamos hacer?

Predictivo



¿Qué pasará?

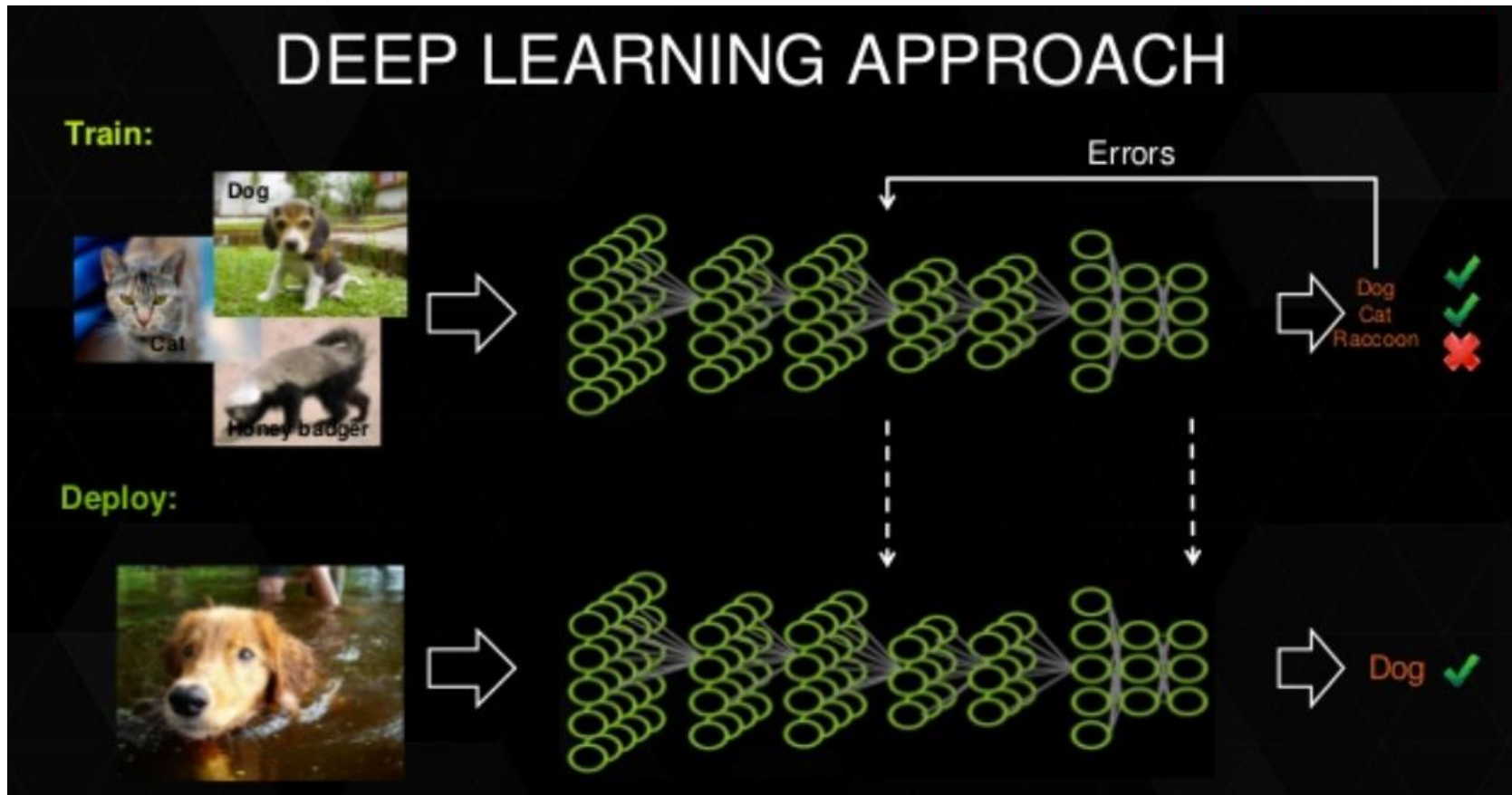
Descriptivo



¿Qué pasó?

Qué Significa “Machine Learning”?

“Dar a las computadoras la capacidad de aprender sin ser programadas explícitamente”



Gartner Hype Cycle

Gartner Hype Cycle for Emerging Technologies, 2016



Source: Gartner
© 2016 Gartner, Inc. and/or its affiliates. All rights reserved.

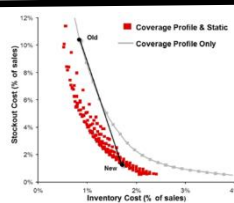
Gartner®

Donde está el Valor Económico para Empresas?

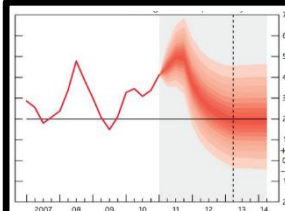
Tamaño de los Datos

	KB	MB	GB	TB
		Realidad		
				Percepción

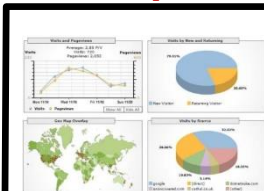
Prescriptivo



Predictivo



Descriptivo



■ Introducción

- Acerca de End-to-End Analytics

■ Big Data, Analytics, y Machine Learning

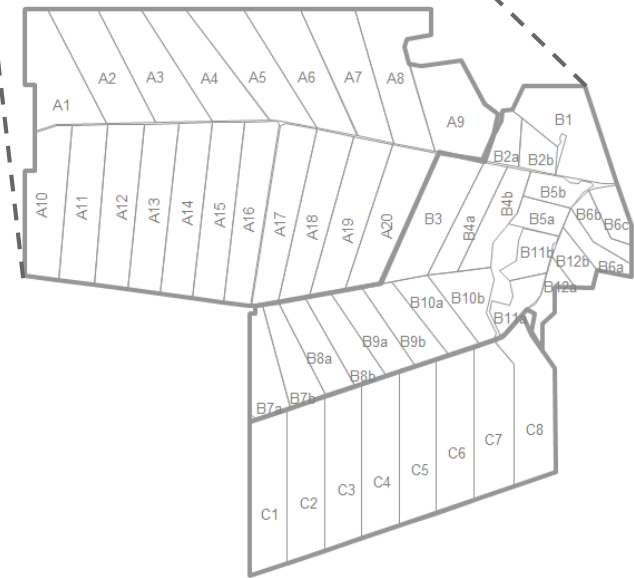
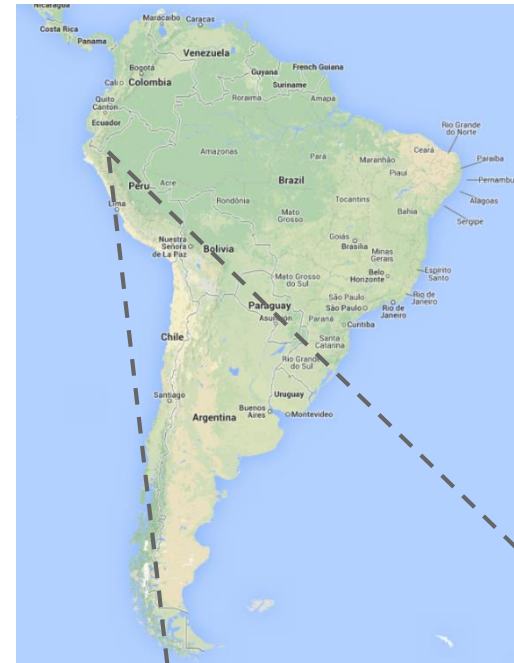
- Qué son, realmente?
- Donde está el valor económica? Un punto de vista contraria

■ Casos de Éxito en América Latina

- Agroindustria: Subasta de Terrenos
- Retail: Detección de Ruptura
- Ropa Deportiva: Reabastecimiento de Tiendas

Subasta de Terrenos: Situación

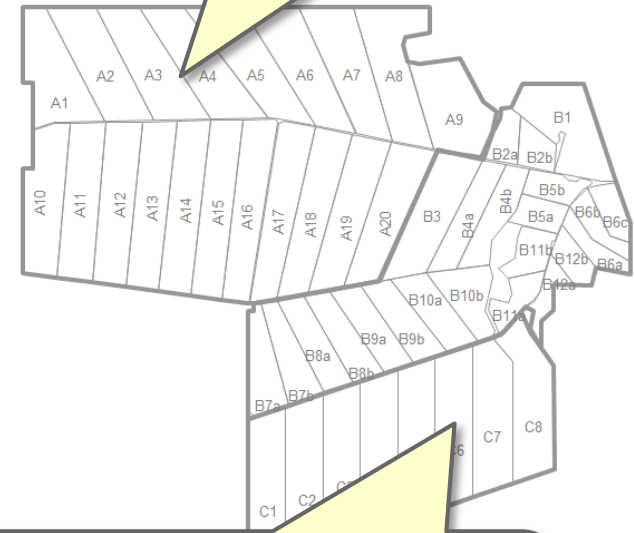
- Inversionista compró 38,000 hectáreas de tierra
- Irrigar & vender a agroindustria
- Opción simple:
 - Dividir en lotes
 - Subastar cada lote al postor que ofrece más
- Opción complicado:
“Subasta Combinatoria”...



Subasta de Terrenos : Subasta Combinatoria

- Subasta simple no maximizará los ingresos
- Porqué no?
 - Escala Eficiente Mínima
 - Adyacencia
- Solución: Subasta Combinatoria
 - Permitir ofertas para paquetes
 - Usado por FCC para vender espectro en EE.UU
- Pero... quien a ganado cada lote?
 - Encontrar la combinación de ofertas que maximice los ingresos

“Ofrecemos \$7,500 / há.
por 4 de los lotes A1 – A12”



“Ofrecemos \$8,200 / há.
por cualquier 3 lotes
adyacentes en el Sector C”

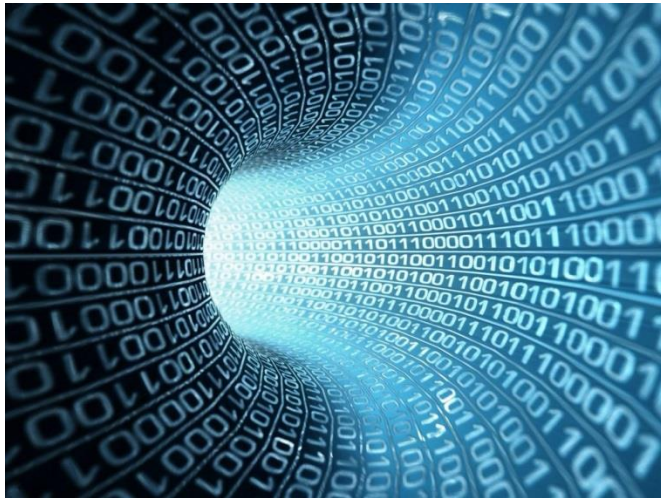
Subasta de Terrenos : “Too Big” Data

- Próximo problema: podría haber demasiado combinaciones
- Ejemplo:
 - 100 ofertas – cuántas combinaciones posibles?
 - Respuesta: 1.26×10^{30}
 - Imagínense una supercomputadora que puede evaluar 1,000,000,000,000 combinaciones por segundo
 - Imagínense que empezó cuando fue formada la Tierra, hace 4 bn. de años
 - Sería <10% terminado



Subasta de Terrenos : Intercambiando Big Data por Big Analytics

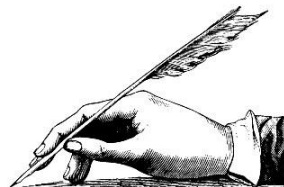
BIG data



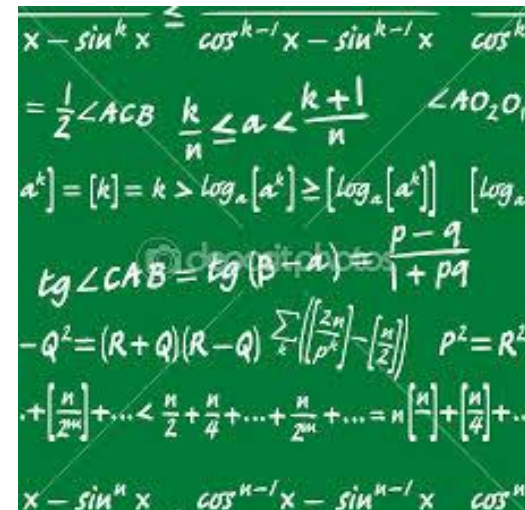
Normal Data

Ice Radar Location (UTM, Zone 6 - NAD 83, N50E23)						Ice Thickness Data						
Transmitter	Receiver		Mipoint		Position	Transmitter Receiver Separation	Ice Thickness					
Bearing	Nothing	Altitude	Bearing	Nothing	Altitude	Error	Ice Thickness					
(Degrees)	(Degrees)	(Meters)	(Degrees)	(Degrees)	(Meters)	(Meters)	(Meters)					
590.803	7.018.362	1.826	590.483	7.018.451	1.828	590.840	7.018.406	1.832	<1	86	1.80	180
590.805	7.018.213	1.880	590.875	7.018.270	1.888	590.880	7.018.242	1.883	<1	86	1.75	180
590.875	7.018.270	1.888	590.775	7.018.288	1.891	590.825	7.018.284	1.889	10	121	2.35	210
590.775	7.018.308	1.931	590.728	7.018.484	1.934	590.785	7.018.371	1.967	10	102	2.85	280
590.728	7.018.484	1.934	590.885	7.018.488	1.938	590.785	7.018.448	1.881	10	118	2.80	270
590.885	7.018.488	1.938	590.888	7.018.548	1.938	590.835	7.018.512	1.935	10	105	2.10	200
590.888	7.018.548	1.938	590.822	7.018.878	1.938	590.885	7.018.882	1.968	<1	87	1.35	130
590.822	7.018.878	1.938	590.488	7.018.954	1.938	590.685	7.018.954	1.958	<1	115	0.95	70
590.488	7.018.954	1.938	590.483	7.018.304	1.931	590.210	7.018.181	1.989	<1	89	1.25	120
590.483	7.018.304	1.931	590.887	7.018.288	1.932	590.140	7.018.230	1.911	<1	91	1.46	140
590.887	7.018.288	1.932	590.872	7.018.338	1.932	590.855	7.018.382	1.962	10	128	1.95	140
590.872	7.018.338	1.932	590.881	7.018.348	1.932	590.811	7.018.328	1.932	20	90	1.40	140
590.881	7.018.348	1.932	590.787	7.018.466	1.938	590.848	7.018.377	1.989	10	110	1.40	140
590.787	7.018.466	1.938	590.884	7.018.480	1.938	590.746	7.018.430	1.987	10	120	1.05	150
590.884	7.018.480	1.938	590.888	7.018.552	1.937	590.848	7.018.484	1.911	20	105	1.95	180
590.888	7.018.552	1.937	590.824	7.018.978	1.938	590.855	7.018.951	1.978	20	90	1.25	120
590.824	7.018.978	1.938	590.484	7.018.928	1.935	590.480	7.018.932	1.978	20	84	1.10	110
590.484	7.018.928	1.935	590.482	7.018.987	1.931	590.415	7.018.948	1.973	20	83	0.95	90
590.482	7.018.987	1.931	590.328	7.018.700	1.938	590.385	7.018.680	1.973	20	87	0.90	90

Small analytics

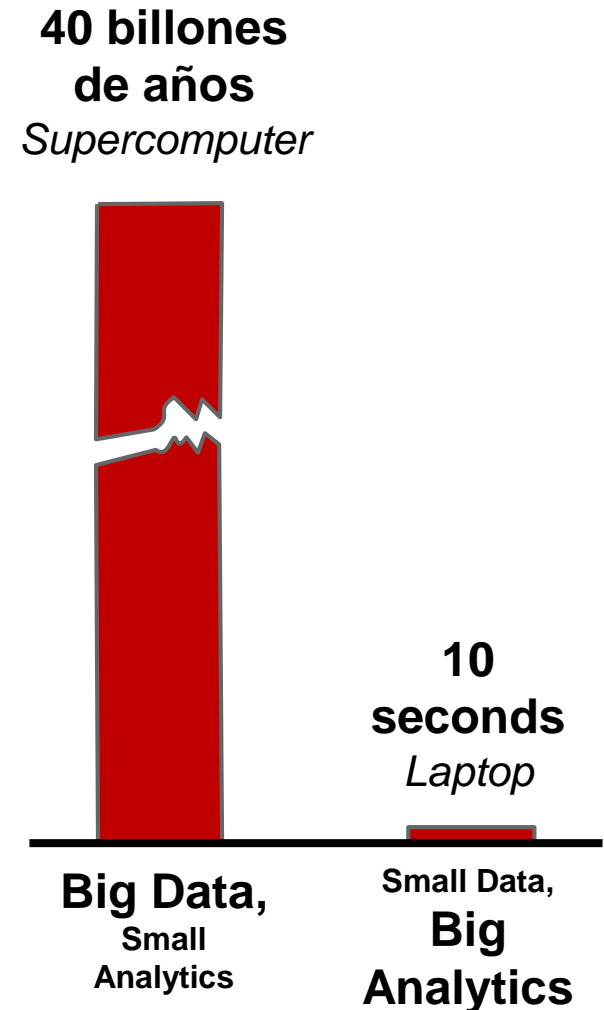


BIG analytics



Subasta de Terrenos: Results

- Qué significa Big Analytics aquí?
 - Formular como Mixed Integer Linear Program (MILP)
 - Código C# llamando Gurobi Optimizer (se podría usar CPLEX, etc.)
 - Cutting Planes, Implicit Enumeration & otros trucos
 - 10 segundos con un laptop Core i7...
 - ...vs. 40 billones de años con la supercomputadora



■ Introducción

- Acerca de End-to-End Analytics

■ Big Data, Analytics, y Machine Learning

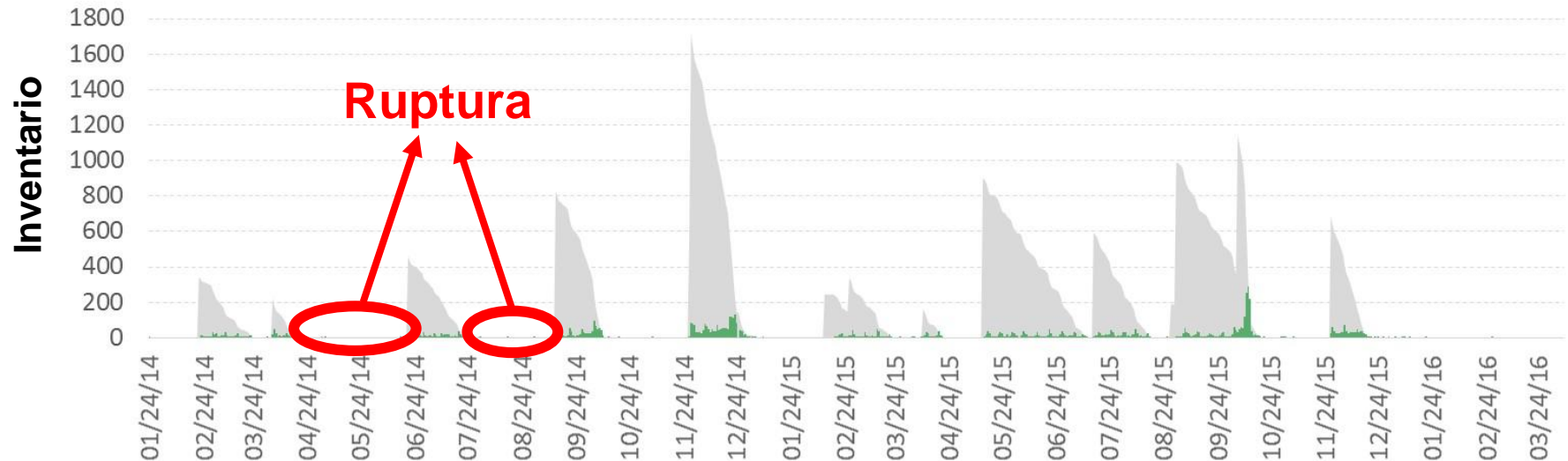
- Qué son, realmente?
- Donde está el valor económica? Un punto de vista contraria

■ Casos de Éxito en América Latina

- Agroindustria: Subasta de Terrenos
- Retail: Detección de Ruptura
- Ropa Deportiva: Reabastecimiento de Tiendas

Ruptura en Tiendas

- Queremos detectar ruptura en cualquier de 10,000 SKUs en cualquier de 2,000 tiendas...
- ... automáticamente sin tener que hacer un conteo manual
- Fácil, verdad? Busca $\text{Inventario}=0$ en el sistema...



Detectando Ruptura



Informe Priorizada

“Si solo tienes tiempo para revisar 10 ítems, mire las primeras 10 líneas de este informe”

Loja	Producto	Cor	Tamanho	Colecao	Lambda	Estoque Linx	Data Ultima Venda	Dias sem vend.	Perda Venda Estimada (Unit)	Perda Venda Estimada (R\$)
000068 - DIAMOND MALL	V15070194 - I	0215 - MESCLA	2 V15		0.76	1	9/3/2014	26	15 \$	3,253.19
000053 - MORUMBI SHOPPING	I14070112 - L	0003 - PRETO	3 PMT		0.86	1	9/8/2014	21	14 \$	2,454.72
000068 - DIAMOND MALL	V15170063 - I	0001 - BRANCO	3 PMT		0.42	6	8/15/2014	45	9 \$	2,350.23
000053 - MORUMBI SHOPPING	V15070215 - I	0004 - AZUL NOTURNO	2 V15		0.64	1	9/7/2014	22	11 \$	2,343.59
000053 - MORUMBI SHOPPING	P13070002 - I	0215 - MESCLA	3 PMT		0.50	1	8/16/2014	44	11 \$	2,151.00
000053 - MORUMBI SHOPPING	I14070112 - L	0003 - PRETO	2 PMT		0.45	1	8/16/2014	45	10 \$	1,813.50
000053 - MORUMBI SHOPPING	V15070194 - I	0215 - MESCLA	2 V15		0.41	1	8/27/2014	33	8 \$	1,796.14
000068 - DIAMOND MALL	P14070012 - I	0003 - PRETO	2 PMT		0.77	7	9/7/2014	22	10 \$	1,674.03
000053 - MORUMBI SHOPPING	A13210005 - I	0706 - OCEANO V14	1 PMT		0.59	4	8/11/2014	49	18 \$	1,586.07
000053 - MORUMBI SHOPPING	V15070215 - I	0004 - AZUL NOTURNO	2 V15		0.59	1	9/14/2014	15	7 \$	1,474.74
000053 - MORUMBI SHOPPING	P13070013 - I	0003 - PRETO	2 PMT		0.39	5	8/30/2014	30	9 \$	1,470.82
802402 - BT BELVEDERE	P14070134 - I	0003 - PRETO	2 PMT		0.36	3	9/5/2014	24	6 \$	1,356.47
802402 - BT BELVEDERE	P14070134 - I	0003 - PRETO	2 PMT		0.44	5	9/10/2014	19	6 \$	1,330.61
000068 - DIAMOND MALL	P13070013 - I	0003 - PRETO	2 PMT		0.36	6	9/12/2014	17	7 \$	1,319.57
000068 - DIAMOND MALL	A13210005 - I	0706 - OCEANO V14	1 PMT		0.59	1	8/27/2014	33	15 \$	1,303.99
000068 - DIAMOND MALL	A13210005 - I	0706 - OCEANO V14	1 PMT		0.59	1	8/22/2014	38	14 \$	1,238.40
000053 - MORUMBI SHOPPING	P13070013 - I	0003 - PRETO	2 PMT		0.39	4	8/30/2014	30	6 \$	1,222.71
000053 - MORUMBI SHOPPING	P13070013 - I	0003 - PRETO	2 PMT		0.39	5	8/22/2014	38	9 \$	1,216.35
000068 - DIAMOND MALL	V15170063 - I	0001 - BRANCO	3 PMT		0.42	3	8/18/2014	42	5 \$	1,202.33
000053 - MORUMBI SHOPPING	P14070013 - I	0003 - PRETO	2 PMT		0.39	2	9/9/2014	20	6 \$	1,185.14
000068 - DIAMOND MALL	P14580002 - I	0003 - PRETO	2 PMT		0.36	6	8/29/2014	31	7 \$	1,171.56
000068 - DIAMOND MALL	V14070012 - I	0003 - PRETO	2 PMT		0.36	1	8/9/2014	51	6 \$	1,141.52
000053 - MORUMBI SHOPPING	P13070013 - I	0003 - PRETO	2 PMT		0.39	1	8/16/2014	44	7 \$	1,129.04
000053 - MORUMBI SHOPPING	P13070013 - I	0003 - PRETO	2 PMT		0.39	2	8/9/2014	51	6 \$	1,077.75

“El sistema dice que tienes 6 unidades, pero han pasado 45 días desde la última venta, que tenía probabilidad de solo 0.2% si tuvieras inventario.

Estimamos venta perdida de 9 unidades y R\$2,350.”

▪ Introducción

- Acerca de End-to-End Analytics

▪ Big Data, Analytics, y Machine Learning

- Qué son, realmente?
- Donde está el valor económica? Un punto de vista contraria

▪ Casos de Éxito en América Latina

- Agroindustria: Subasta de Terrenos
- Retail: Detección de Ruptura

- Ropa Deportiva: Reabastecimiento de Tiendas

Caso de Éxito: Ropa Deportiva

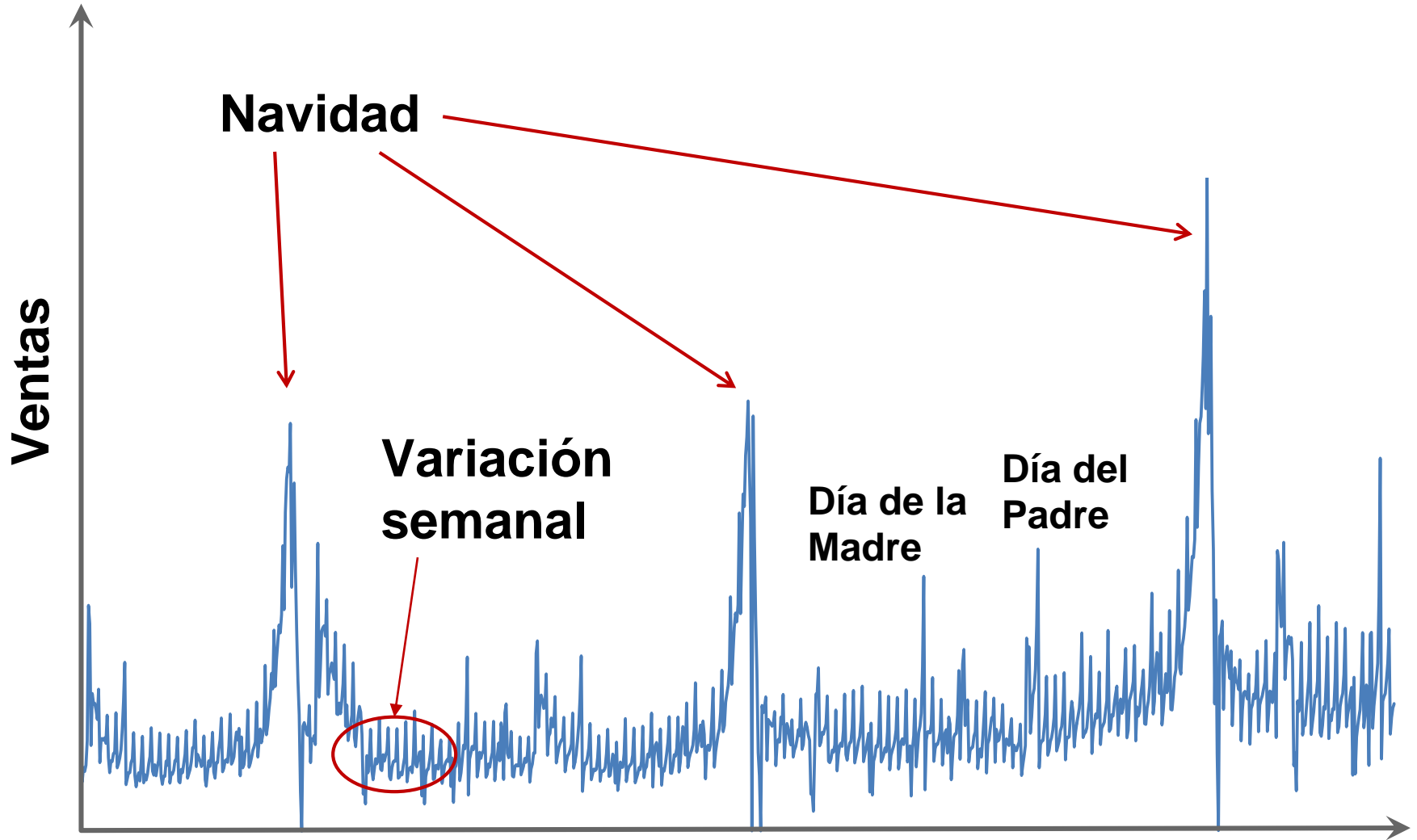
- Cadena latinoamericana
- ~200 tiendas
- ~5,000 SKUs
- Productos “fashion”,
 - Difíciles a pronosticar
 - Vida corta
- Altísimo niveles de faltantes (hasta 30%)
- Desafío: Mejorar pronósticos y reabastecimiento



Big Data? No Tanto

- **Historia de venta a nivel diario para 3 años, 200 tiendas, 5,000 SKUs**
 - Matriz con $365 \times 3 \times 5000 \times 200 = 1.1$ **billones** de elementos
 - p.e. en Excel (16 bytes por célula) sería **18GB**
 - Ni hablar de inventario, entradas, despachos, faltantes, etc.
- **Pero esa matriz es 99.4% sparse**
- **Y se puede guardar la venta de un día en 1 byte**
- **Resulta que representación sparse matriz en memoria usa solo 12MB en lugar de 18GB**

Historia de la Venta



Modelo Econométrico



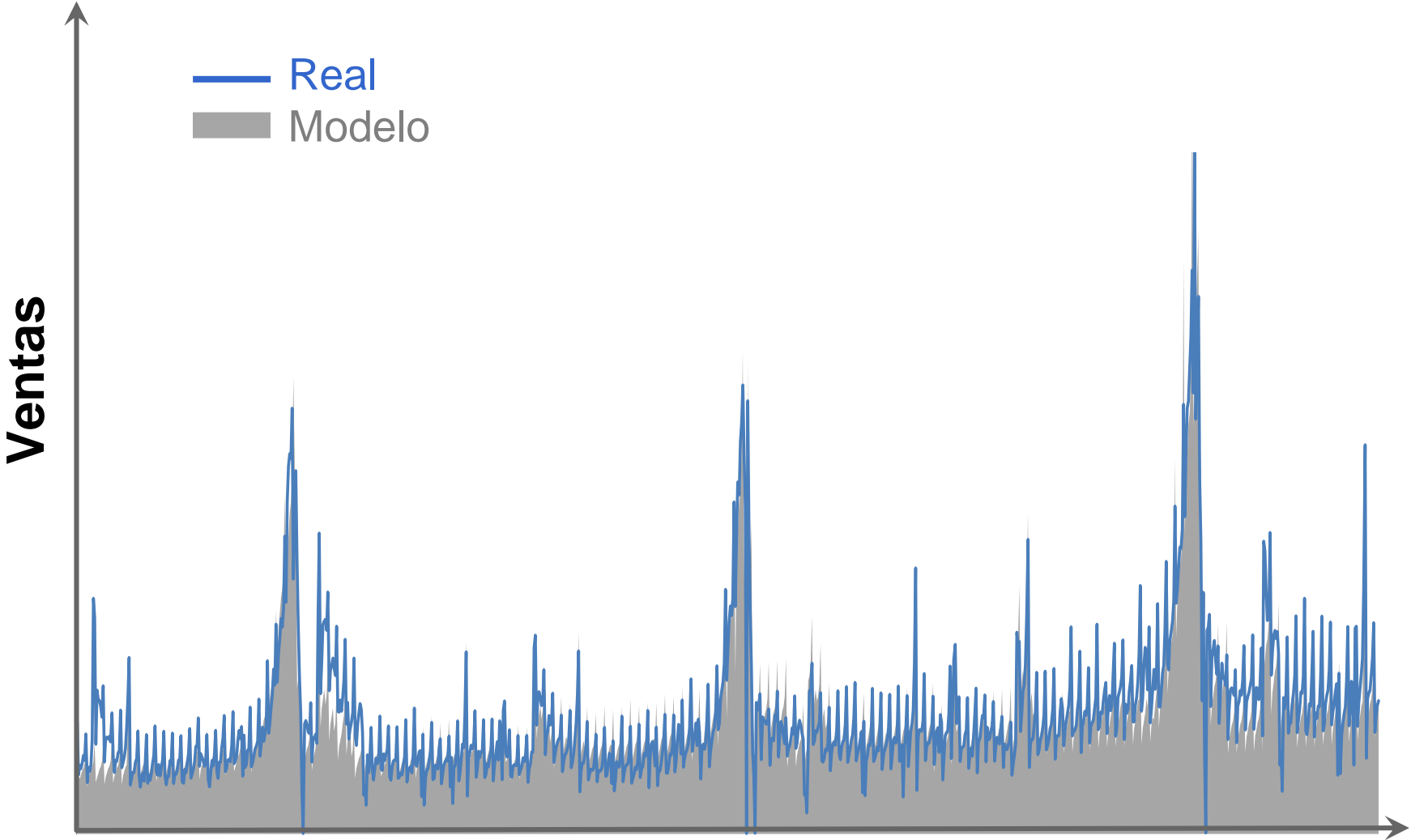
Log(Ventas) =	Variable	Coefficient	Effect	R-Sqd
	Constant	12.38		83.68%
	Day of Week = Sunday	-0.65	52%	
	Day of Week = Monday	-0.49	62%	
	Day of Week = Tuesday	-0.42	66%	
	Day of Week = Wednesday	-0.35	70%	
	Day of Week = Thursday	-0.32	73%	
	Day of Week = Friday	-0.24	79%	
	Christmas Day	-6.27	0%	
	New Year	-4.64	1%	
	Christmas Overhang	0.62	185%	
	After Christmas Exchanges	1.40	404%	
	Christmas wk	1.34	384%	
	Christmas wk-1	1.07	291%	
	Christmas wk-2	0.67	195%	
	Christmas wk-3	0.36	143%	
	Christmas wk-4	0.11	112%	
	Mother's Day	0.50	165%	
	Valentine's Day	0.69	199%	
	Father's Day	0.39	148%	
	Other Holidays	-0.09	92%	
	Days into most recent collection	0.00	100%	
	Num of Stores open	0.01	101%	
	Marathon % of chain \$ sales	2.32	1014%	
	Promotion Wk	0.51	166%	
	Promotion Wk+1	0.36	144%	
	Promotion Wk+2	0.12	113%	
	Promotion Wk+3	0.19	120%	
	Promotion Wk+4 (and on)	0.11	112%	

Modelo Econométrico

Log(Ventas) =

Variable	Coefficient	Effect	R-Sqd
Constant	12.38		83.68%
Day of Week = Sunday	-0.65	52%	
Day of Week = Monday	-0.49	62%	
Day of Week = Tuesday	-0.42	66%	
Day of Week = Wednesday	-0.35	70%	
Day of Week = Thursday	-0.32	73%	
Day of Week = Friday	-0.24	79%	
Christmas Day	-6.27	0%	
New Year	-4.64	1%	
Christmas Overhang	0.62	185%	
Christmas wk	1.34	384%	
Christmas wk	1.34	384%	
Christmas wk-1	1.07	291%	
Christmas wk-2	0.67	195%	
Christmas wk-3	0.36	143%	
Christmas wk-4	0.11	112%	
Father's Day	0.39	148%	
Valentine's Day	0.69	199%	
Father's Day	0.39	148%	
Other Holidays	-0.09	92%	
Days into most recent collection	0.00	100%	
Num of Stores open	0.01	101%	
Marathon % of chain \$ sales	2.32	1014%	
Promotion Wk	0.51	166%	
Promotion Wk+1	0.36	144%	
Promotion Wk+2	0.12	113%	
Promotion Wk+3	0.19	120%	
Promotion Wk+4 (and on)	0.11	112%	

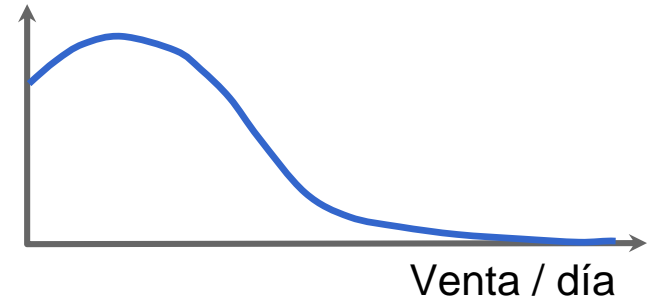
El Modelo Explica la Historia Muy Bien



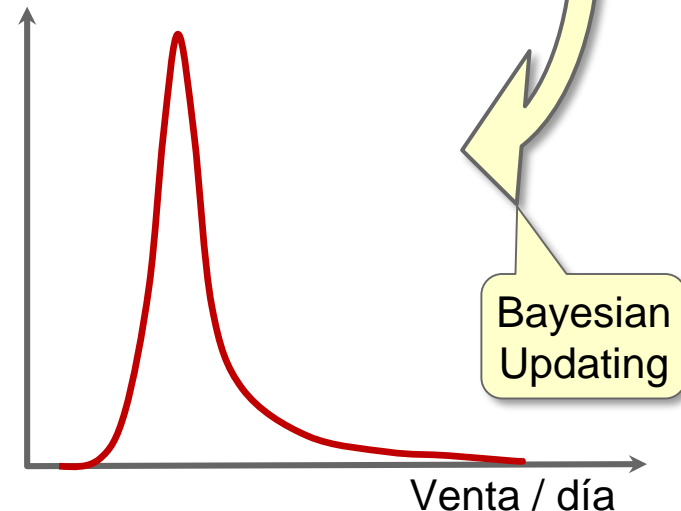
Acercamiento Pronósticos

- **Hacer segmentación estadística de los productos**
- **Crear rango pronostico inicial usando la historia de los productos del segmento**
- **Cada día:**
 - Normalizar la ultima observación usando los coeficientes del modelo econométrico
 - Actualizar el rango pronostico normalizado (“Bayesian updating”)
 - Reaplicar los coeficientes para reimponer estacionalidad, promociones, etc.

Rango inicial

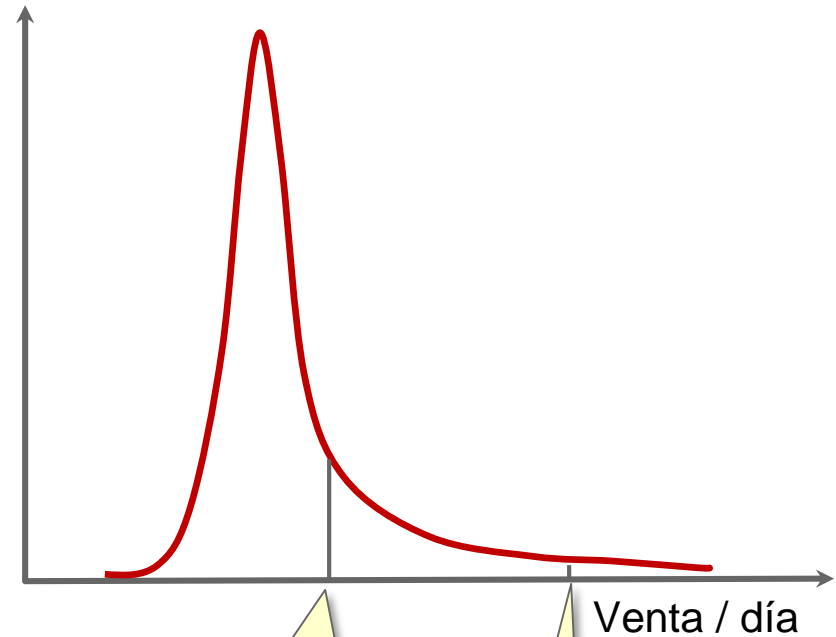


Rango después de 10 observaciones



Acercamiento Reabastecimiento

- Elegir nivel de servicio (¿qué percentil del rango de pronóstico a apuntar?)
 - Margen bruto...
 - ... vs. Costo inventario
- Cantidad a mandar = Pronóstico cumulativo durante lead time @ percentil deseado - Saldo - tránsitos
- Repetir para cada una de las 700,000 SKU-tiendas



Margen bajo,
Costo alto
=> 70% nivel
servicio

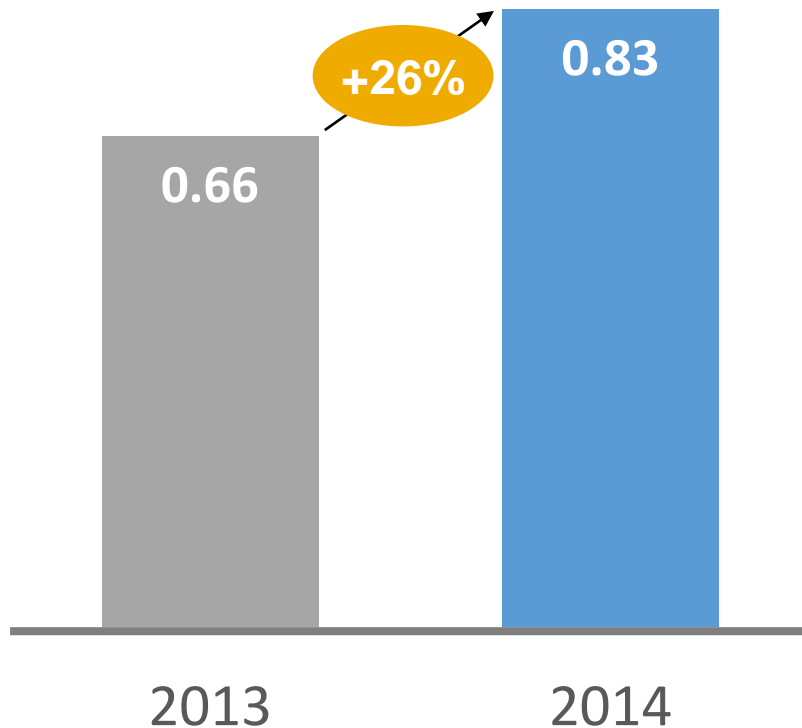
Margen alto,
Costo bajo
=> 98% nivel
servicio

Corre en
~2 minutos

Resultados

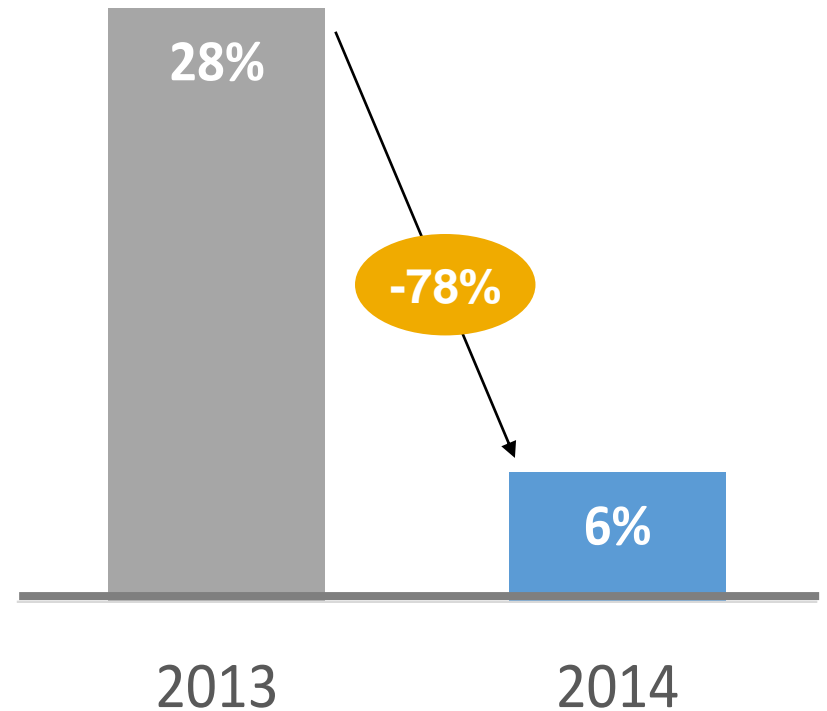
Eficiencia de Inventario

(Venta Mes \$ / Inventario \$)

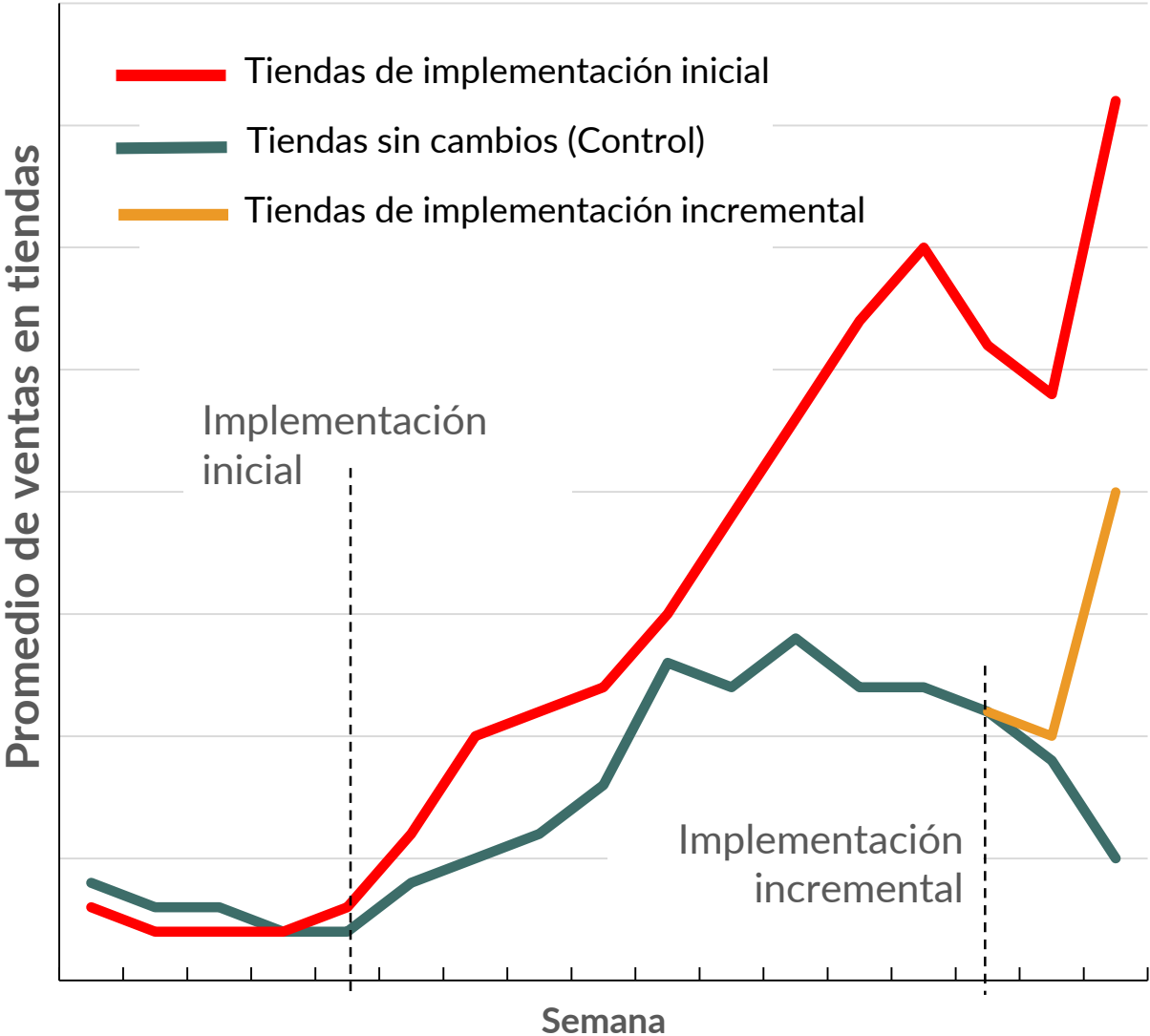


Faltantes

(Venta Perdida valorizada a margen bruto)



Más Resultados



Detalles de Contacto



Allan Gray (EE.UU.)

allan@e2eanalytics.com

+1-408-480-1143



Christian Ramirez (Perú)

christian@e2eanalytics.com

+51-956-103-131



Gustavo Froes (Brasil)

gustavo@e2eanalytics.com