

# Decisiones basadas en **datos**

---

Marcelo Ferreyra



# ¿Qué son los **Datos**?

- Los datos son una **colección de medidas** que tomamos de aquello que queremos representar
- No es lo mismo** analizar clientes de una empresa, pacientes de un hospital o accidentes de tránsito

# Datos de **clientes**

Cliente	Estado Civil	Edad	Última fecha de compra	Compra últimos 3 meses	Compra último año	Tendencia de compra	Medio de pago preferido
135	Casado	45	03-may	\$ 2.570	\$ 6.425	60%	Tarjeta Crédito
136	Casado	52	24-may	\$ 3.190	\$ 16.907	-25%	Tarjeta Débito
137	Casado	34	28-abr	\$ 1.850	\$ 6.290	18%	Efectivo
138	Soltero	25	17-may	\$ 987	\$ 2.961	33%	Tarjeta Crédito
139	Casado	29	12-may	\$ 1.256	\$ 7.285	-31%	Efectivo
140	Casado	23	26-abr	\$ 881	\$ 2.555	38%	Tarjeta Débito
141	Soltero	37	23-may	\$ 1.389	\$ 8.195	-32%	Tarjeta Débito
142	Casado	41	24-may	\$ 1.502	\$ 7.210	-17%	Tarjeta Crédito
143	Casado	33	29-abr	\$ 1.176	\$ 4.704	0%	Tarjeta Crédito
144	Soltero	20	22-may	\$ 542	\$ 2.764	-22%	Efectivo
145	Casado	39	20-may	\$ 4.103	\$ 20.515	-20%	Efectivo
146	Soltero	44	07-may	\$ 1.320	\$ 3.168	67%	Efectivo
147	Casado	31	16-may	\$ 992	\$ 1.984	100%	Tarjeta Débito
148	Casado	37	21-may	\$ 2.521	\$ 13.613	-26%	Tarjeta Crédito

# Datos de pacientes

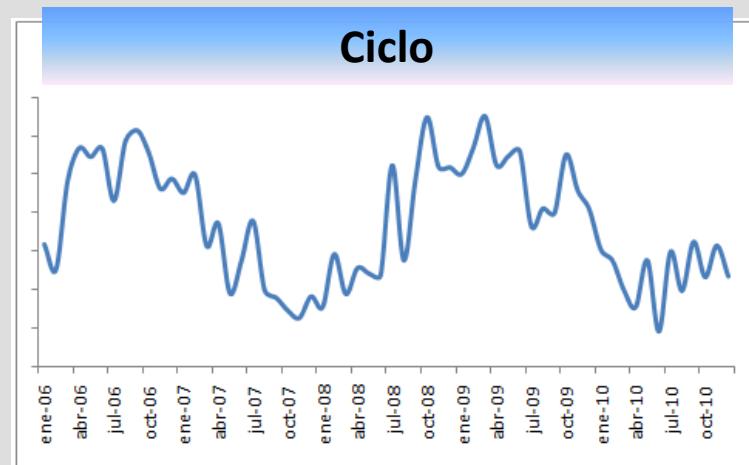
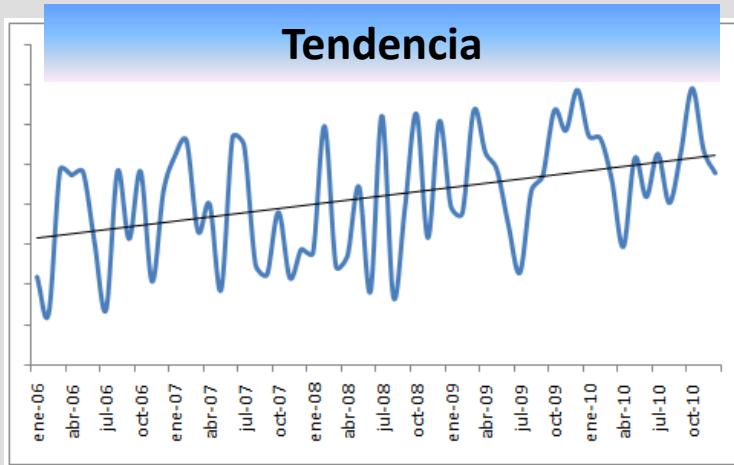
Historia Clínica	Edad	Altura	Peso	Sexo	Presión sanguínea en reposo	Colesterol mg/dl	Ritmo Cardiaco máx.
1001	70	1,65	78	M	130	322	109
1002	67	1,68	67	F	115	564	160
1003	57	1,75	103	M	124	261	141
1004	64	1,72	84	M	128	263	105
1005	74	1,67	65	F	120	269	121
1006	65	1,71	79	M	120	177	140
1007	56	1,81	90	M	130	256	142
1008	59	1,64	73	M	110	239	142
1009	60	1,73	75	M	140	293	170
1010	63	1,62	64	F	150	407	154



Los datos  
desordenados  
**no son** muy útiles

Pero si los ordenamos  
encontraremos **patrones**

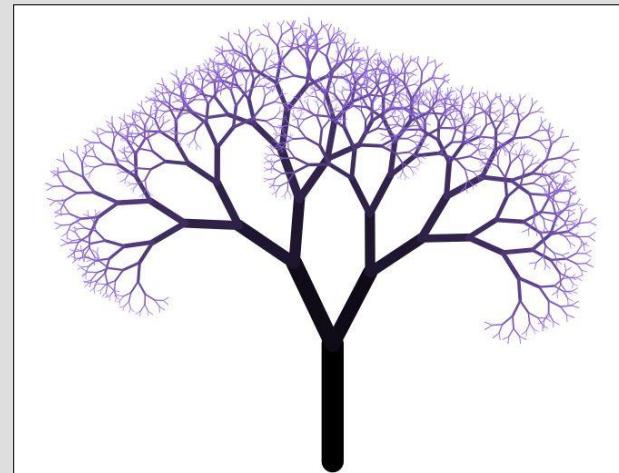
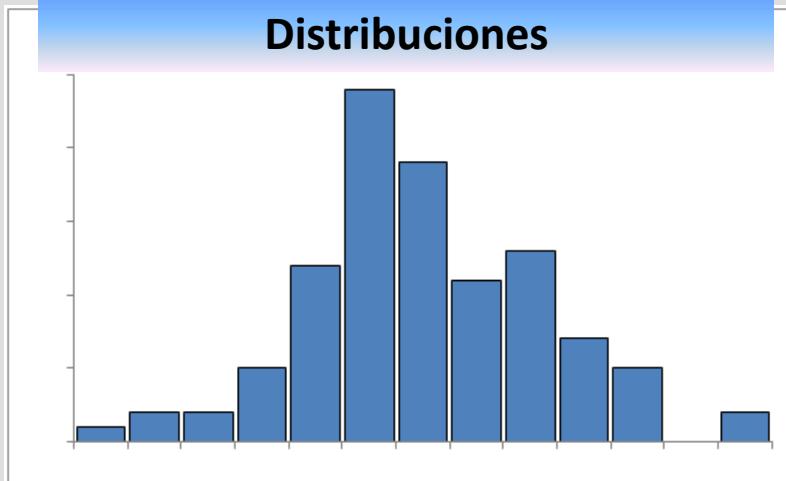
### Patrones temporales



Pero si los ordenamos  
encontraremos **patrones**

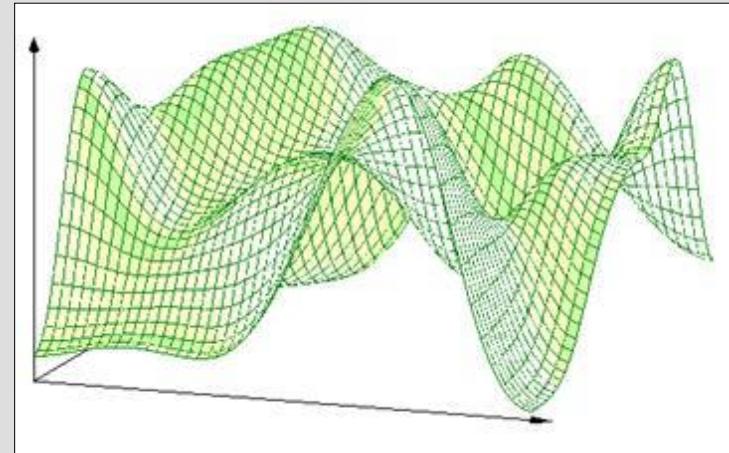
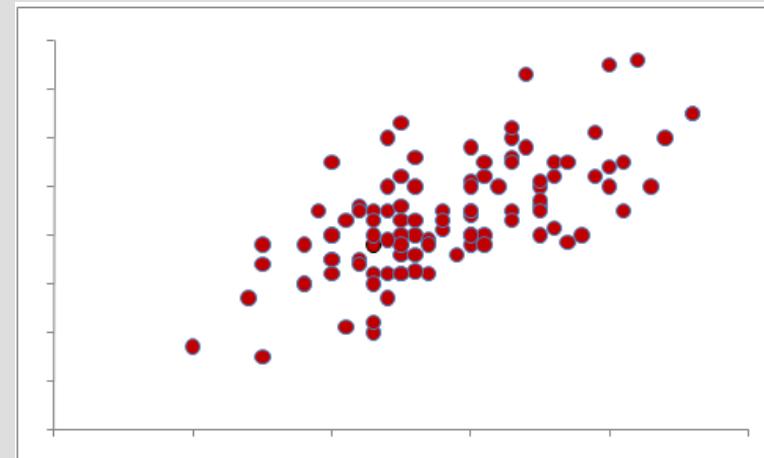
### Patrones estructurales

#### Distribuciones



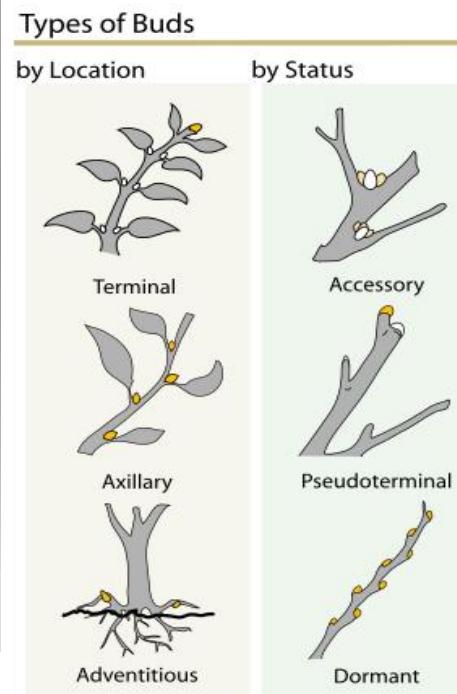
Pero si los ordenamos  
encontraremos **patrones**

### Patrones relacionales



# ¿Para qué nos sirven Los **patrones**?

## Clasificar



## Predecir



¿Qué tan **estables** son estos patrones?

*Depende*

**Del tipo de datos**



**Del horizonte de tiempo considerado**



# ¿Qué tan generales son esos **patrones**?

Es **imposible**  
predecir qué hará  
una persona

Pero **no es complicado**  
predecir qué hará un  
grupo de personas

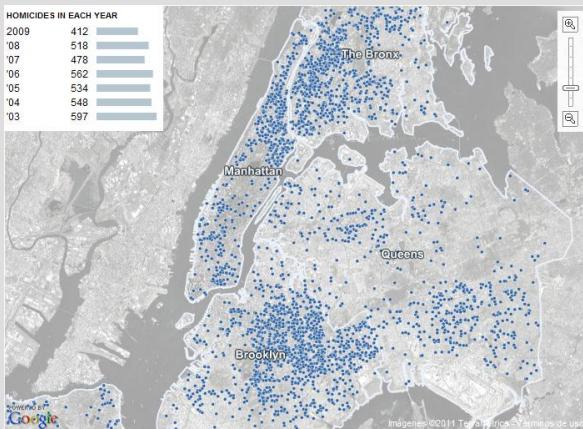


# Hasta ahora **vimos**

- Los datos contienen información en forma de **patrones**
- Usando patrones podemos **clasificar** y **predecir**
- Es **imposible** predecir qué hará **una persona**
- Pero es **possible** predecir, con cierto grado de certeza, qué hará **un grupo de personas**

# Caso 1: Jack Maple, NY, 90'

En el término de un año **redujo los robos** en subterráneos de **1.200** a **12**, utilizando mapas que llamó Tablas del Futuro



Creó el COMPSTAT, un **sistema de reducción del crimen** y mejora de la calidad de vida, basado en datos de crímenes e **información geográfica (GPS)**

## Caso 2: Hospital Italiano

En 2008, el Hospital Italiano llevó a cabo un proyecto para analizar sus datos en la búsqueda de patrones

---

### Objetivo:

Predecir qué pacientes podrían ver su estado de salud agravado durante el siguiente año.

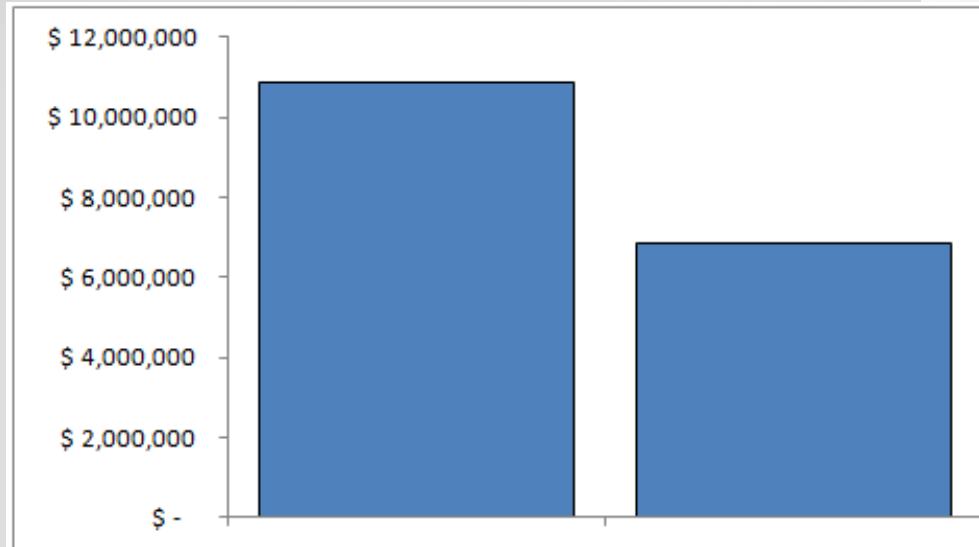
### Resultados:

Fue posible identificar pacientes de alto riesgo y explicar las razones: insuficiente prescripción o falta de adherencia al tratamiento.

## Caso 3: Riesgo Crediticio

**Uno de los objetivos de las empresas financieras es mantener al mínimo su exposición**

**Objetivo:** bajar la exposición identificando clientes riesgosos



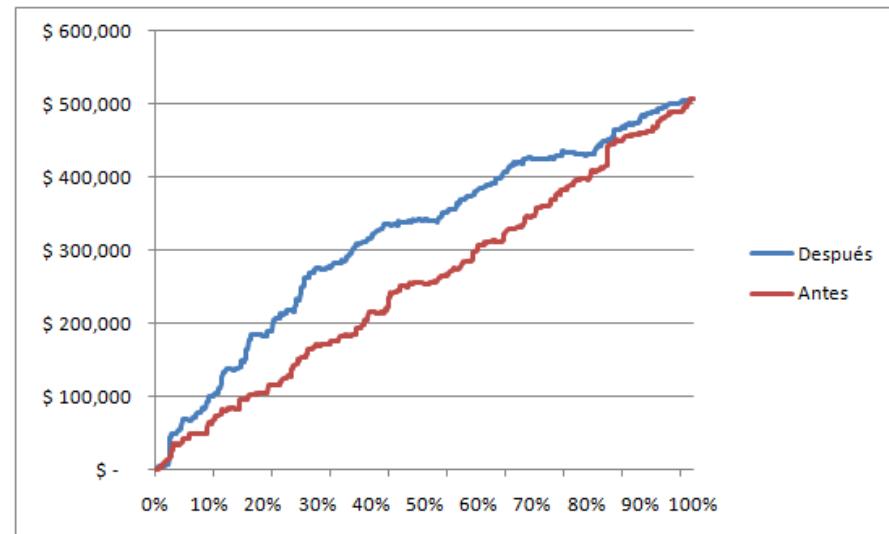
**Resultado:** bajó su exposición a **un 60%** e incentivó las compras de los clientes menos riesgosos mediante el incremento de sus límites de compra

## Caso 4: Optimización de flujo de caja

Una firma de abogados especializada en recuperar deudas, **mejoró su flujo de caja en un 48%** sobre el primer 20% de su cartera.

A su vez, contactó un **45% menos** de clientes para lograr recuperar los primeros 100.000\$, logrando bajar sus costos y optimizando el call center.

% Clientes	Antes	Después	Diferencia en \$	Diferencia en %
10%	\$ 61.678	\$ 100.603	\$ 38.925	63%
20%	\$ 123.843	\$ 183.548	\$ 59.705	48%
30%	\$ 186.007	\$ 275.218	\$ 89.211	48%
40%	\$ 248.171	\$ 315.775	\$ 67.604	27%
50%	\$ 310.335	\$ 342.405	\$ 32.070	10%



## Caso 5: Fuga de Clientes

**Un banco tenía una alta tasa de fuga de clientes, por lo que decidió averiguar a través de sus datos a qué se debía y cómo podía remediarlo**

**Objetivo:** analizar los datos en busca de patrones predictivos para contactar aquellos clientes con altas probabilidades de fuga

**Resultados:** el análisis permitió identificar dos factores muy relacionados con la fuga: la comisión de los servicios y la adhesión a los débitos automáticos.

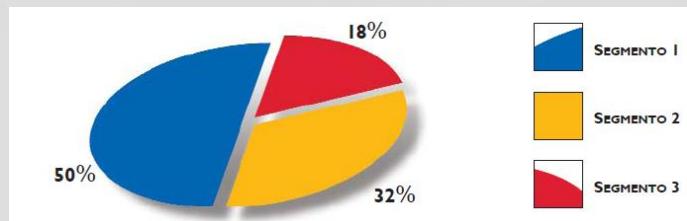
**Se logró un modelo capaz de identificar al 60% de los clientes riesgosos con 2 meses de anticipación**

# Caso 6: Segmentación de Clientes

Un supermercado deseaba segmentar sus clientes para encontrar hábitos de consumo.

**Objetivo:** clasificar los clientes en base a sus comportamientos

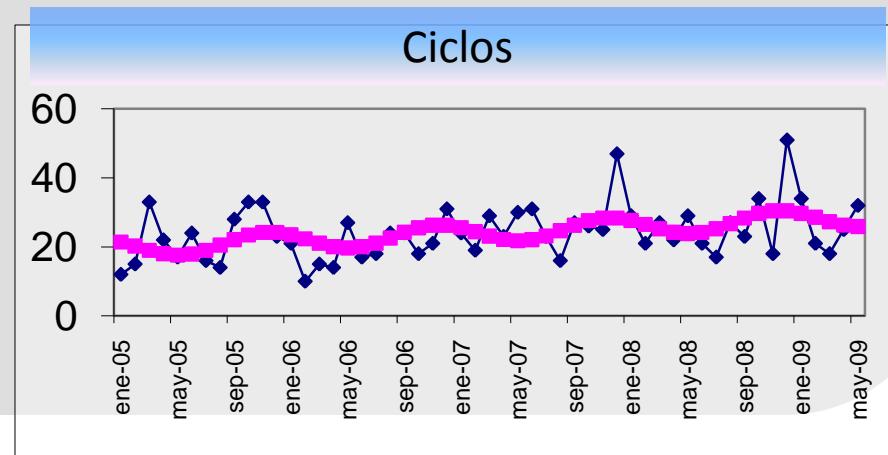
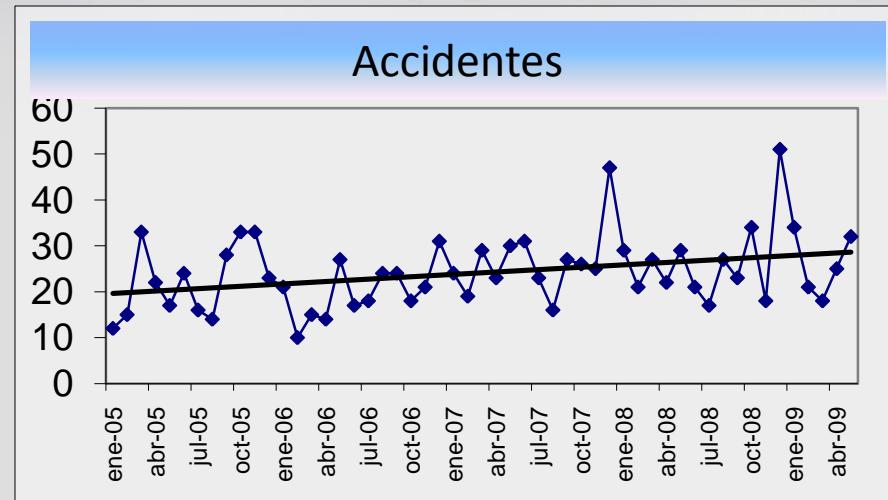
**Resultados:** encontró 3 segmentos y varias oportunidades de optimizar la oferta de sus productos



	Todos los clientes	Segmento 1	Segmento 2	Segmento 3
Valor promedio de una compra	114.81 \$	63.42 \$	180.55 \$	141.16 \$
Margen por compra	6.70 \$	4.24 \$	9.20 \$	9.18 \$
Margen en %	6.20%	6.60%	5.20%	6.80%
Visitas totales en el período	2.1	1.3	2.2	4.2
Productos por compra	29	13	48	39
Ofertas por compra	2.4	1.7	3.4	2.7

# Caso 7: Tránsito en Trenque Lauquen

El análisis de los datos de la guardia del hospital **sirvió para identificar una tendencia de colisiones** de tránsito en la ciudad, así como también un **ciclo anual** que muestra un incremento en la cantidad de colisiones durante el verano.



# Obteniendo ventajas de los datos

**Según Tomas Davenport, hay dos requisitos básicos para aprovechar los datos**

---

## Humano

**Se necesitan líderes que aprecien y entiendan lo que significa tomar decisiones en base a datos y que estén dispuestos a hacerlo**

## Técnico

**Se necesita contar con datos confiables, integrados y limpios**

# Obteniendo ventajas de los datos

## Los Datos y el Impacto en los Negocios

Reporte del MIT Center for Digital Business - Abril 2011

Andrew McAfee

Se analizó una muestra de 330 compañías de USA, de todas las industrias.

Las empresas con mayor uso de datos (+1  $\sigma$ ) en la toma de decisiones, son **4%** más productivas y **6%** más rentables que el resto.

Gracias