



# Inteligencia de Negocios

Data WareHouse y Minería de Datos.

# Reglas de Asociación.

Minería de Datos

# Objetivo del método.

- ✓ “Determinar qué va con qué”.
- ✓ Encontrar combinaciones de variables cuya ocurrencia sucede a menudo: productos que se compran juntos, problemas médicos después de ciertos síntomas, etc.
- ✓ Medir la fuerza o importancia de estas combinaciones.
- ✓ Explotar el descubrimiento de estas reglas con el propósito de mejorar nuestro negocio.

# Objetivos del método

- ✓ Son un ejemplo de Data Mining No Supervisado, en el sentido que no estamos tratando de explicar ninguna variable dependiente.
- ✓ Estamos tratando de descubrir patrones de asociación entre las distintas variables (tienen que ser nominales)
- ✓ En general: SI condición ENTONCES resultado
- ✓ Requisito: Probabilidad Básica

# Preguntas típicas

- ✓ ¿Son los productos de limpieza de ventanas comprados cuando detergentes y jugo de naranja se compran juntos?
- ✓ ¿Se compra la soda típicamente cuando se compran bananos? ¿Importa la marca de la soda?
- ✓ Cómo afectan los factores demográficos del vecindario lo que los clientes compran? ¿Existen factores estacionales?

# Reglas de Asociación

Un ejemplo: Ventas de comestibles en un supermercado

La vista minable

Codigo	VinoA	VinoB	Gaseosa	Horchata	Bizcochos	Galletas	Chocolate
R1	true	false	true	false	false	true	false
R2	false	true	true	false	false	false	false
R3	false	false	false	true	true	true	false
R4	true	false	true	true	true	true	true
R5	false	false	false	false	false	true	false
R6	true	false	false	false	false	true	true
R7	false	true	true	true	true	false	false
R8	false	false	false	true	true	true	true
R9	true	false	true	false	true	false	true
R10	false	false	true	false	true	false	false

¿Cómo fue posiblemente la BD original?

# Reglas de Asociación

Un ejemplo: Ventas de comestibles en un supermercado

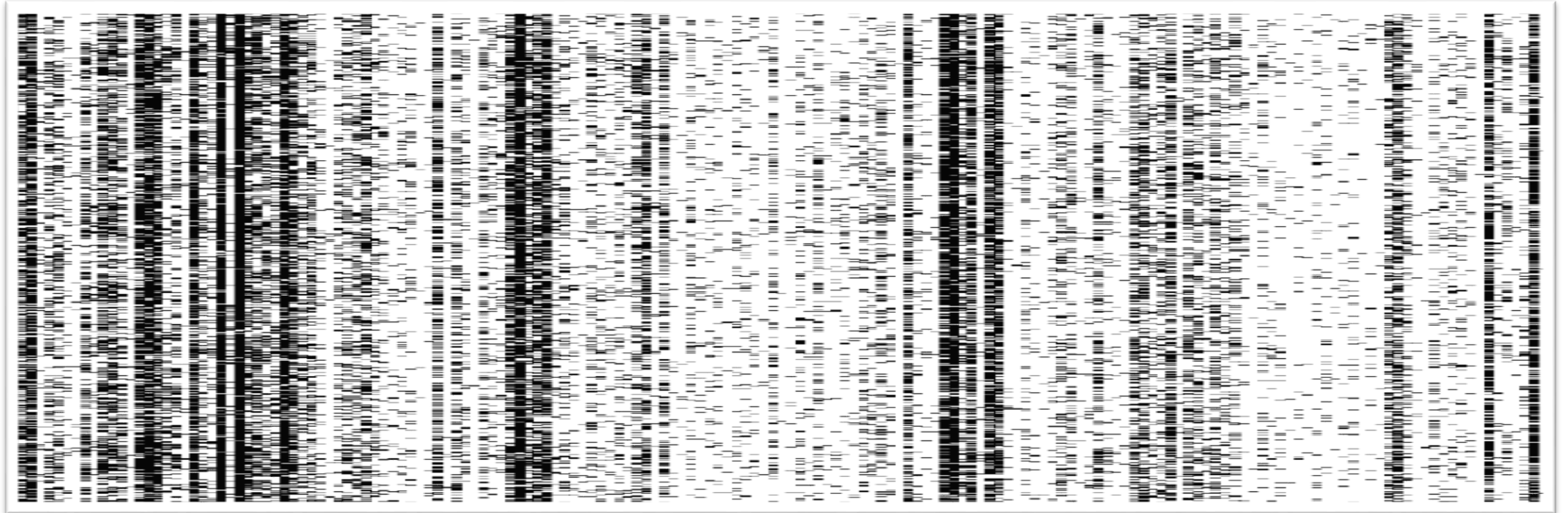
Ahora nuestro ejemplo en forma gráfica

Codigo	VinoA	VinoB	Gaseosa	Horchata	Bizcochos	Galletas	Chocolate
R1							
R2							
R3							
R4							
R5							
R6							
R7							
R8							
R9							
R10							

# Reglas de Asociación

Un ejemplo: Ventas de comestibles en un supermercado

Una muestra mayor





# Consideraciones – Reglas de Asociación

- ✓ El problema de este método visual es que, aún con un conjunto pequeño de ítems, el número de posibilidades es increíblemente alto.
- ✓ Necesitamos algún criterio para limitar el número de reglas, de tal forma que obtengamos las más valiosas.
- ✓ Para el punto anterior se utilizan métricas relacionadas a la calidad de las relaciones encontradas: **Soporte, Confianza y Lift.**

Soporte

# Soporte - Reglas de Asociación

- ✓ Se define como el número de veces o la frecuencia con que A y B aparecen juntos en transacciones en la base de datos, o cada uno por separado también.
- ✓ El primer requisito que podemos imponer para limitar el número de reglas es que tengan un soporte mínimo.
- ✓ Soporte mide la frecuencia con que ocurre una regla.

$$P(A \cap B)$$

# Reglas de Asociación

Volvamos al ejemplo de comestibles del supermercado, calculando algunos soportes tanto individuales como de grupos. Un método práctico es el algoritmo A Priori.

Codigo	VinoA	VinoB	Gaseosa	Horchata	Bizcochos	Galletas	Chocolate
R1							
R2							
R3							
R4							
R5							
R6							
R7							
R8							
R9							
R10							
	0.4	0.2	0.6	0.4	0.6	0.6	0.4

Confianza

# Confianza - Reglas de Asociación

- Es el cociente del soporte de la regla entre el soporte del antecedente.

$$\frac{P(A \cap B)}{P(A)}$$

- Es una probabilidad condicional de “B dado A”.
- Si SOPORTE mide la frecuencia, CONFIANZA mide la fortaleza de la relación.

# Reglas de Asociación

Por ejemplo:

“Chocolate -> Galletas” tiene una confianza de 0.75

Codigo	VinoA	VinoB	Gaseosa	Horchata	Bizcochos	Galletas	Chocolate
R1							
R2							
R3							
R4							
R5							
R6							
R7							
R8							
R9							
R10							

0.4

0.2

0.6

0.4

0.6

0.6

0.4

Lift



# Lift – Reglas de Asociación.

- ✓ En ocasiones, cuando hay un par de ítems que son relativamente frecuentes, el nivel de confianza será alto aún cuando no haya ninguna relación significativa entre ambas.
- ✓ Es una forma de identificar si las relaciones no han sido producto de la casualidad.
- ✓ Este análisis se corresponde con el Chi-Cuadrado de la estadística.

# Reglas de Asociación

- Su fórmula es:

$$\frac{P(A \cap B)}{P(A) * P(B)}$$

- Para establecer una relación realmente significativa debemos observar **Lift** por encima de 1.
- Valores superiores a 1 nos indican fortaleza de la relación (Ej: productos complementarios). Valores inferiores nos indican debilidad de relación (Ej: productos sustitutos).

# Ejercicio

# Enunciado

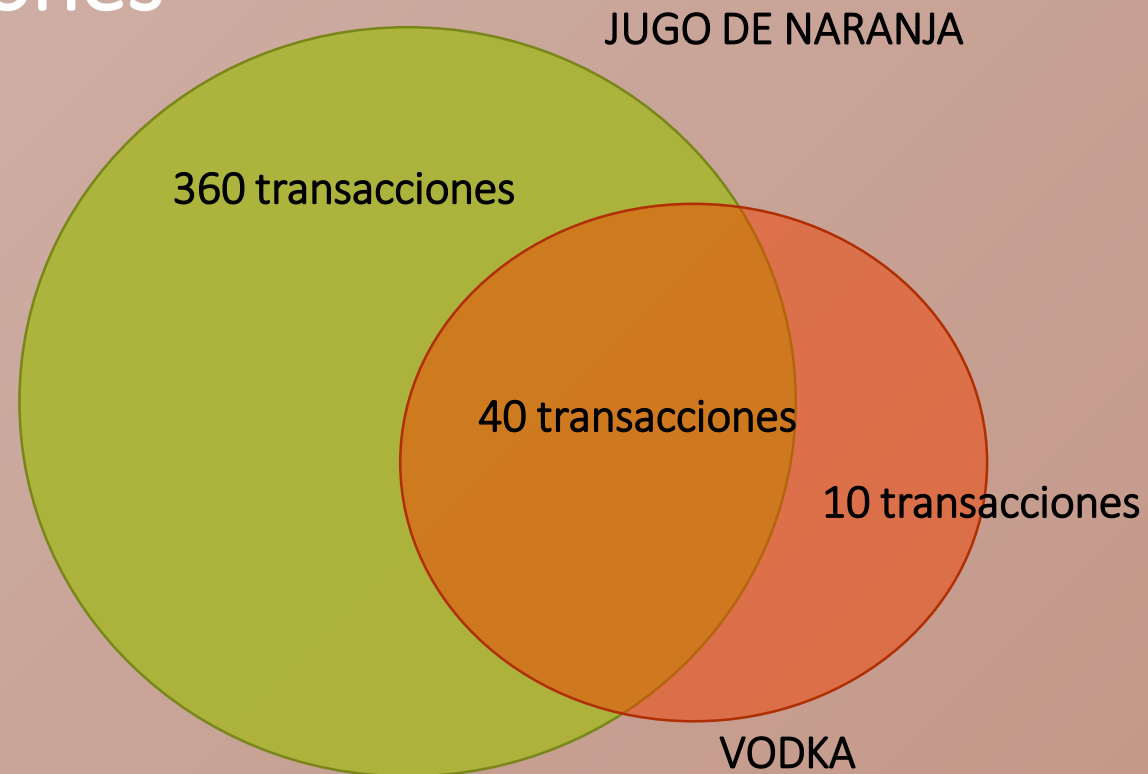
En un supermercado se hizo un análisis de compra de los clientes con 1000 transacciones. En esas 1000 transacciones se encontró que:

- ✓ En 400 transacciones la gente compró jugo de naranja
- ✓ En 50 transacciones la gente compró vodka
- ✓ Se encontró que en 40 transacciones se compró jugo de naranja y vodka

Encuentre el soporte de cada producto y la confianza de jugo => vodka y de vodka => jugo

# Representación gráfica

1000 transacciones



# LIFT

También dentro del mismo estudio se encontró comparó el jugo de naranja con el pan

	COMPRÓ PAN	NO COMPRÓ PAN	
COMPRÓ JUGO	280	120	400
NO COMPRÓ JUGO	420	180	600
	700	300	1000

# Resultados

Soporte (pan) = 0.7

Soporte (jugo) = 0.4

Soporte (pan => jugo) = soporte(jugo => pan) = 0.28

Confianza (pan => jugo) =  $0.28/0.7 = 0.4$

Confianza (jugo => pan) =  $0.28/0.4 = 0.7$

# Lift

$$\text{Lift (pan} \Rightarrow \text{jugo)} = \text{lift (jugo} \Rightarrow \text{pan)} = 0.28 / (0.7 * 0.4) = 1.00$$

Realicen el lift del vodka y el jugo

R//

$$\text{Lift (vodka} \Rightarrow \text{jugo)} = \text{lift (jugo} \Rightarrow \text{vodka)} = 0.04 / (0.4 * 0.05) = 2.00$$



# Consideraciones

# Reglas de Asociación

- ✓ En algunos casos podría ser posible identificar mejores reglas si se incluyen otras variables en el análisis. Si por ejemplo la compra de vino y de chocolates no ocurre simultáneamente ni es un descubrimiento explotable...
- ✓ ¿Podría llegar a serlo si consideramos esta relación para un segmento particular de clientes como los de altos ingresos? ¿O qué tal si incluimos el tiempo y se descubre que la regla tiene importancia para épocas de Navidad?

# Reglas de Asociación

- ✓ A veces puede ser útil modificar el nivel de abstracción para encontrar reglas más útiles: ¿Analizamos cada marca de vino con otros productos o lo hacemos simplemente como “vino”? ¿Nos quedamos con “chocolates” o analizamos cada caja-presentación como un producto diferente?
- ✓ ¿Qué otros factores externos a la empresa podrían considerarse en el análisis y que agreguen información útil? ¿el clima? ¿la economía? ¿acciones de la competencia?

# Aplicaciones de las Reglas de Asociación

Minería de Datos.

# Aplicaciones

- ✓ Las reglas de asociación son la técnica de Data Mining más sencilla conceptualmente.
- ✓ Las dificultades técnicas se refieren a su aplicación eficiente a grandes bases de datos. El descubrimiento de algoritmos que permitieron su aplicación a bases de datos grandes no son de interés para nosotros. Mas interesante para nosotros son la aplicación de las reglas en un contexto de negocios.

# Ventas Cruzadas

- ✓ Imaginen el caso de Amazon.com. Cuando usted pone el libro de «Introducción al Data Mining» en su canasta, Amazon sabe, por ejemplo, que 54% de los clientes que compraron ese libro compraron inmediata o posteriormente los libros X, Y, Z.
- ✓ Así que para tratar de adelantar o inducir nuevas compras, Amazon informará al cliente sobre este comportamiento.

# Up Selling

- ✓ Si sabemos que las personas que compran tablets de tipo A, típicamente compran fundas de tipo B, podríamos tratar de inducir a los clientes a comprar mejor fundas de tipo C que tienen un margen superior para nosotros.



# Promociones

- ✓ Si sabemos que el café y las rosquillas se compran juntas podemos inducir la compra de mas rosquillas ofreciendo descuentos al café. Otra manera de utilizar esta información sobre esta regla es que no queremos ofrecer descuentos simultáneos a café y rosquillas.
- ✓ Cuando los clientes compren mas café, respondiendo a nuestro descuento, esperamos que automáticamente compren mas rosquillas sin necesidad de ofrecerles un descuento.



# Otros

- ✓ Óptima colocación de los productos en los estantes de la tienda. Si yo sé que el producto A y el B tienden a comprarse juntos. Puedo ponerlos juntos para hacer mas fácil la compra... o puedo ponerlos lejos uno del otro para exponer a mi cliente a otros productos.
- ✓ Bundling: Reglas de asociación pueden ofrecer información valiosa en el diseño de paquetes de productos.

Gracias