MACHINE LEARNING: ALGORITMOS NO SUPERVISADOS

Análisis de Canasta | Reglas de Asociación

Instructor: José Nelson Zepeda

San Salvador, Diciembre 2018

Análisis de Canasta

Concepto

Objetivo

Metodología

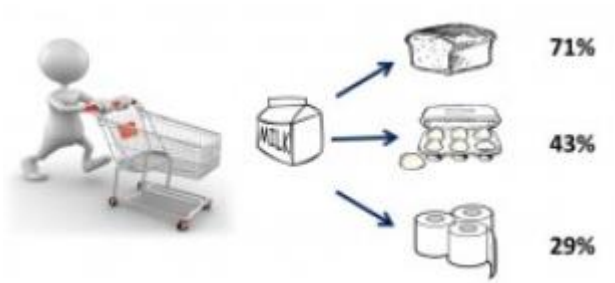
Definiciones

Ejemplos

Conceptos Básicos: Análisis de Canasta

¿Qué es el análisis de canasta?

El Análisis de canasta es un una metodología muy utilizada ya que permite describir asociaciones entre diferentes items. Este método permite que fácilmente que por ejemplo que identifiquemos las asociaciones propias en un lanzamiento de un nuevo producto, y conocer cual producto juega como rol apalancador y cual de soporte, de tal manera que permita de una mejor forma describir la causalidad entre productos a analizar.



CASO DE UN AUTOSERVICIO

- Familias con hijos pequeños (en cuyos tickets encontramos pañales, fórmula para bebés, toallas húmedas, alimento para bebés, etc.)
- Familias grandes (tickets con mayor cantidad de piezas por ítem en múltiplos del promedio de piezas)
- Clientes con inclinación a productos saludables (leche light, productos orgánicos, alimentos adicionados con antioxidantes, sal baja en sodio, etc.)
- Entre otros

CASO DE UNA CADENA DE FARMACIAS

- Clientes con padecimientos crónicos que requieren medicamentos de forma permanente (diabéticos, hipertensos, etc.)
- Clientes con hijos pequeños (Medicamentos pediátricos, fórmula, pañales, etc.)
- Clientes que son adultos mayores (complementos alimenticios, pañales para adultos, medicamentos geriátricos, etc.)
- Entre otros

¿Qué datos se pueden capturar de un ticket?

- Lugar de la transacción o punto de venta (POS)
- Fecha y hora de la transacción
- Descripción de cada artículo comprado incluyendo precio
- Cantidad comprada de cada articulo
- Valor total del ticket o factura
- Forma de pago

Si la descripción de cada artículo es completa, es decir incluye un ID único al cual se le pueda asociar una familia de productos, costo, SKU, etc., podremos empezar a buscar patrones en los datos y contestar de forma más acertada cuestionamientos asociados a los productos o bien al punto de venta.

¿Qué preguntas se pueden contestar?

Un análisis de canasta se enfoca en contestar preguntas relacionadas con los aspectos siguientes:

1. Efectividad de precios y promociones y sus medios de comunicación (folletos, cupones, etc.)

2. Portafolio y surtido

3. Acomodación del producto en el punto de venta.

¿Qué preguntas se pueden contestar?

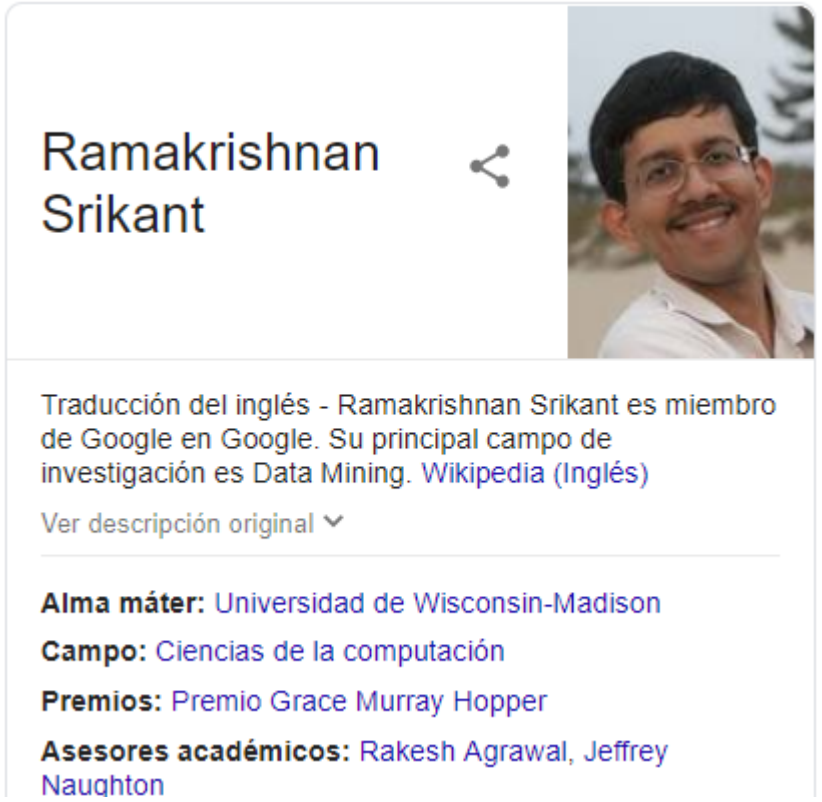
Un análisis de canasta se enfoca en contestar preguntas relacionadas con los aspectos siguientes:

1. Efectividad de precios y promociones y sus medios de comunicación (folletos, cupones, etc.)

2. Portafolio y surtido

3. Acomodación del producto en el punto de venta (Planograma)

Reglas de Asociación



En el año 1994, Sikrant y Agrawal, presentaron un algoritmo cuya función es identificar las asociaciones entre elementos. El algoritmo se vuelve vital cuando las posibles combinaciones entre elementos alcanzan una cantidad considerable y generar todas las reglas por medio de un trabajo manual seria extremadamente complejo

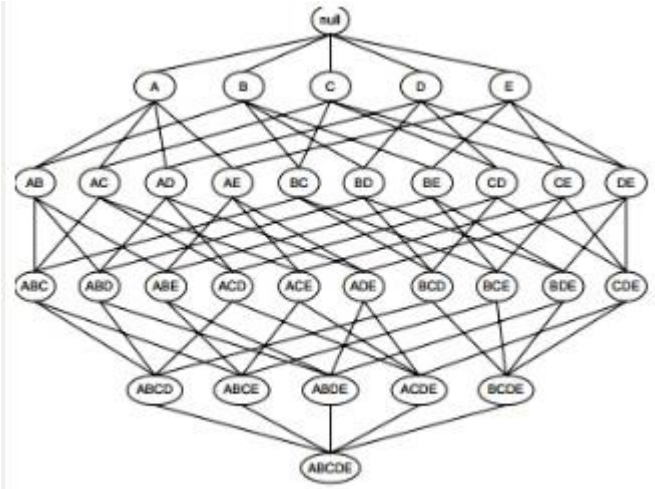
Reglas de Asociación

Las reglas de asociación son reglas que indican cierta relación entre sus conjuntos, sin que esto implique causalidad.

Generar estas reglas de asociación es un proceso muy costoso computacionalmente hablando ya que las reglas implican las combinaciones de productos.

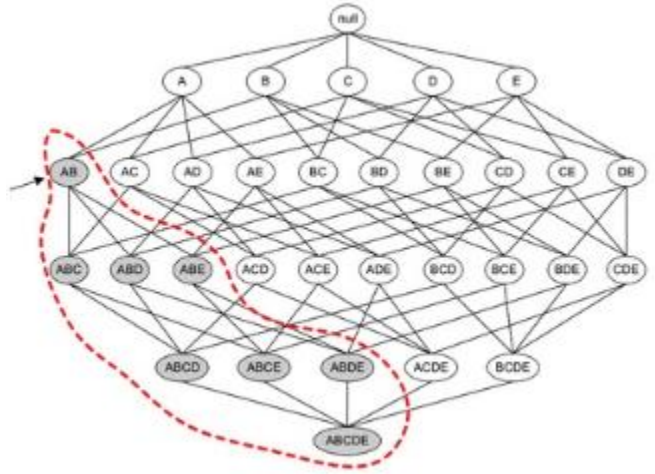
Gráficamente se puede observar la complejidad de los resultados, y entra en juego la necesidad de podar o "prunning" para eliminar todos aquellos conjuntos que no son frecuentes.

http://ferminpitol.blogspot.com/2014/05/reglas-de-asociacion-algoritmo-apriori.html



La forma de generar reglas de asociación consta de dos pasos:

1. Generación de combinaciones frecuentes: cuyo objetivo es encontrar aquellos conjuntos que sean frecuentes en la base de datos y a la vez considerando un umbral pre-establecido.
2. Generación de reglas: A partir de los conjuntos frecuentes se generan las reglas las cuales están basadas en el índice de confianza.



Reglas de Asociación: Conceptos

- **Elementos u Objetos (Items)**: dependiendo de la industria y el campo de aplicación, los elementos pueden ser pacientes, eventos, productos, clientes.
- **Transacción**: Es una operación identificada con un identificador único y que contiene como mínimo 1 elemento.
- **Conjunto de elementos (Itemset)**: Un grupo de elementos que se pueden encontrar en una o varias transacciones.
- **Soporte (Support)**: Probabilidad de encontrar un elemento o un conjunto de elementos en una transacción. Se estima por el número de veces que un elemento o conjunto de elementos se encuentra en todas las transacciones disponibles. Por ser una probabilidad este valor se encuentra entre 0 y 1.
- **Regla (Rule)**: Una regla define una relación entre dos conjuntos de elementos (Itemsets) X e Y que no tienen elementos en común. X -> Y significa que, si tenemos el elemento X en una transacción, entonces podemos tener Y en la misma transacción.

Reglas de Asociación: Conceptos

- **Soporte de una regla (Support of a Rule)**: Probabilidad de encontrar elementos o conjunto de elementos en una transacción. Se estima por el número de veces que ambos elementos o conjuntos de elementos se encuentran en todas las transacciones disponibles. Por ser una probabilidad este valor se encuentra entre 0 y 1.
- **Confianza de una regla (Confidence of a Rule)**: Probabilidad de encontrar un elemento o conjunto de elementos Y en una transacción, sabiendo que el elemento o conjunto de elementos X está en la transacción. Se estima por la frecuencia correspondiente observada (número de veces que X e Y se encuentran en todas las transacciones, dividido por el número que se encuentra X). Este valor se encuentra entre 0 y 1.
- **Importancia de una regla (lift of a rule)**: La importancia de una regla, que es simétrica (importancia(X->Y)=importancia(Y->X)), es el soporte del conjunto de elementos que agrupa X e Y, dividido por el soporte de X y el soporte de Y. Este valor puede ser cualquier número real positivo. Una lift mayor que 1 indica un efecto positivo de X en Y. un valor de 1 significa que no hay efecto, y es como si los elementos o conjuntos de elementos fueran independientes. Una lift menor que 1, significa que hay un efecto negativo de X en Y o viceversa, como si fueran excluyentes entre sí.

Reglas de Asociación: Conceptos

$$Rule: X \Rightarrow Y$$

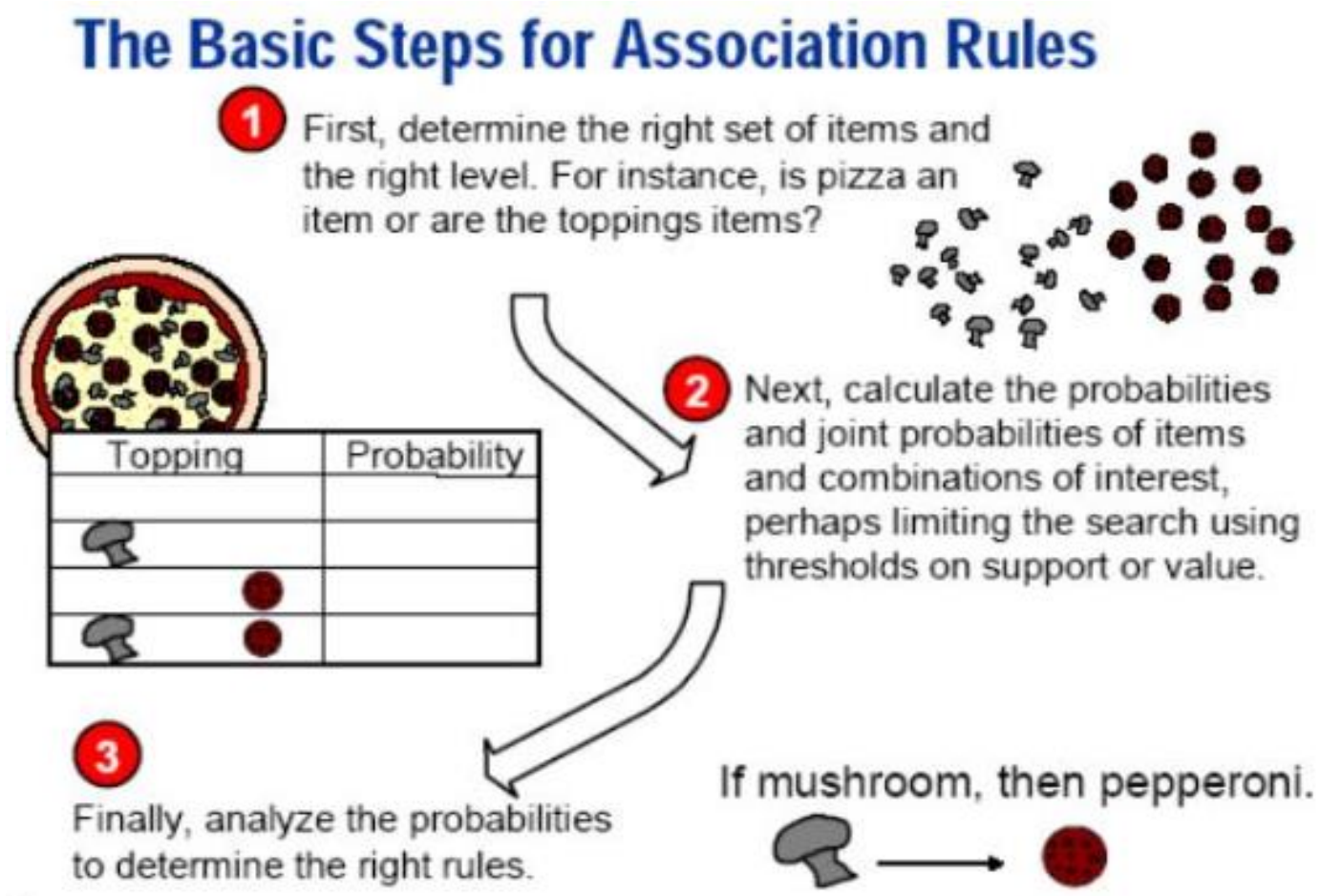
$$Supprt = \frac{Frequency(X,Y)}{N}$$

$$Confidence = \frac{Frequency(X,Y)}{Frequency(X)}$$

$$Lift = \frac{Support}{Support(X) \times Support(Y)}$$

Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B\&C \Rightarrow D$	1/5	1/3	5/9

Reglas de Asociación: Pasos Básicos



Reglas de Asociación: Algoritmos

Algoritmo Apriori: Este algoritmo es el más conocido en el mundo de las reglas de asociación. Su estrategia se basa en los soportes de los diferentes conjuntos de elementos y luego por medio del uso de una función para generar candidatos realiza el cálculo del soporte.

Al igual que los algoritmos anteriores, el A priori recorre la base de datos en múltiples ocasiones, la primera iteración es importante por lo que se describe en el párrafo previo en donde calcula el soporte para cada conjunto o elemento identificando los elementos que tienen soporte mayor y menor, con esto se establece un valor de "soporte mínimo", este valor se compara con la siguiente iteración y de esa forma se van descartando elementos y se van convirtiendo los elementos frecuentes en reglas de asociación. El algoritmo se detiene cuando llega al conjunto vacío.

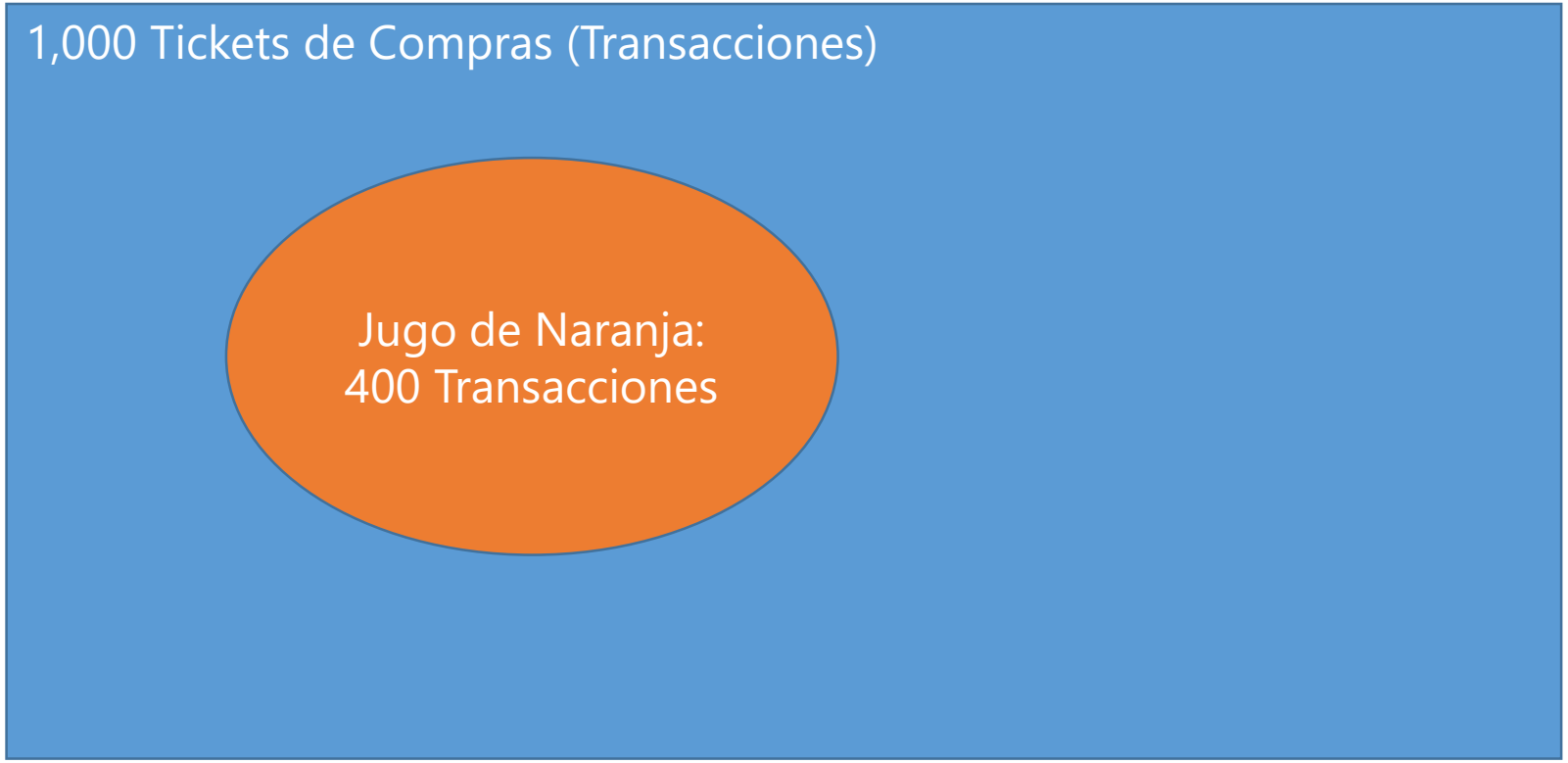
Algoritmo AIS

Algoritmo SETM (Set Oriented Mining

Algoritmo FP-Growth

Algoritmo Eclat

Reglas de Asociación: Ejemplo Jugo y Vodka

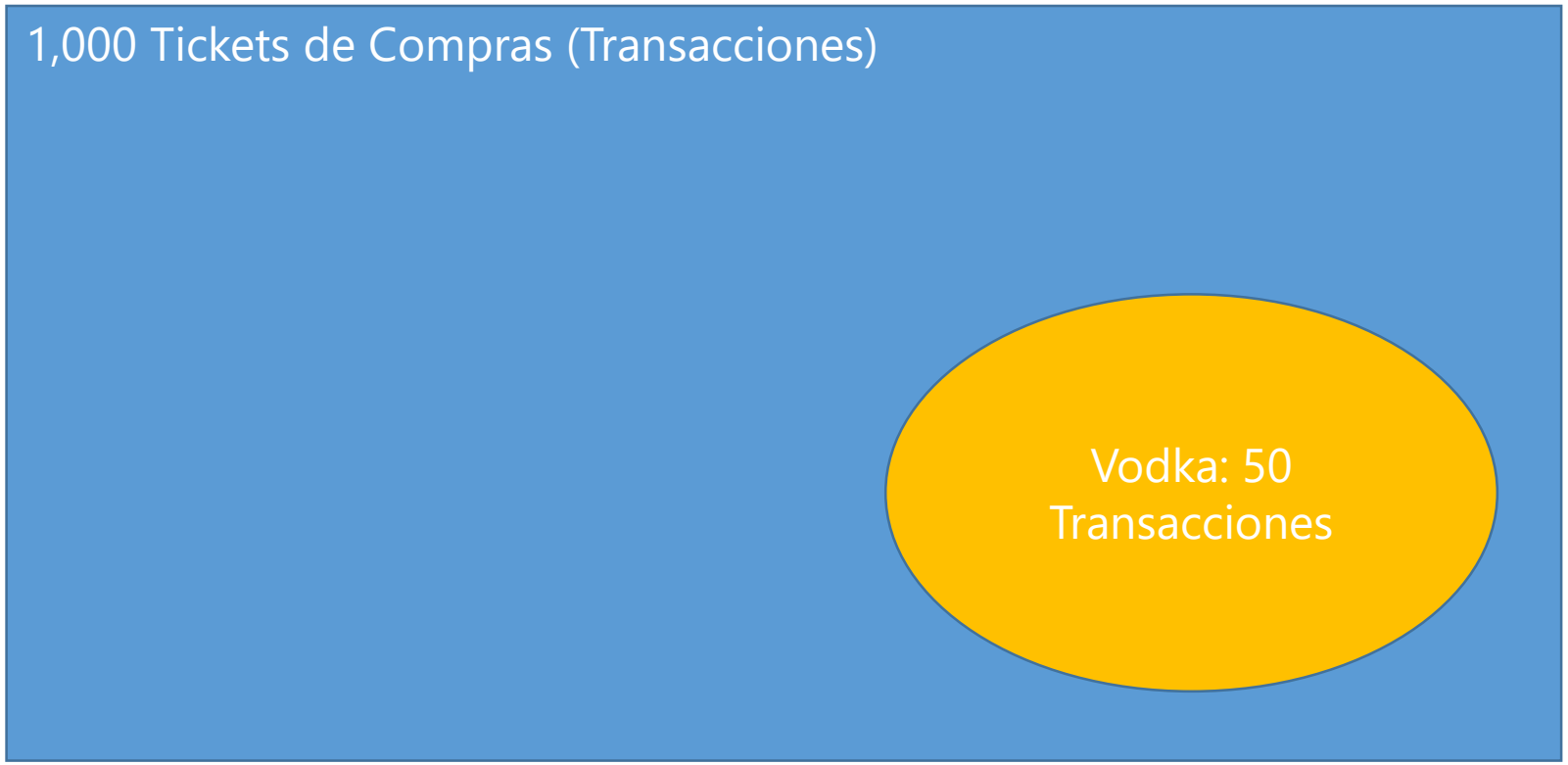


Soporte Jugo de Naranja=400/1000

Soporte Jugo de Naranja=0.4

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Vodka

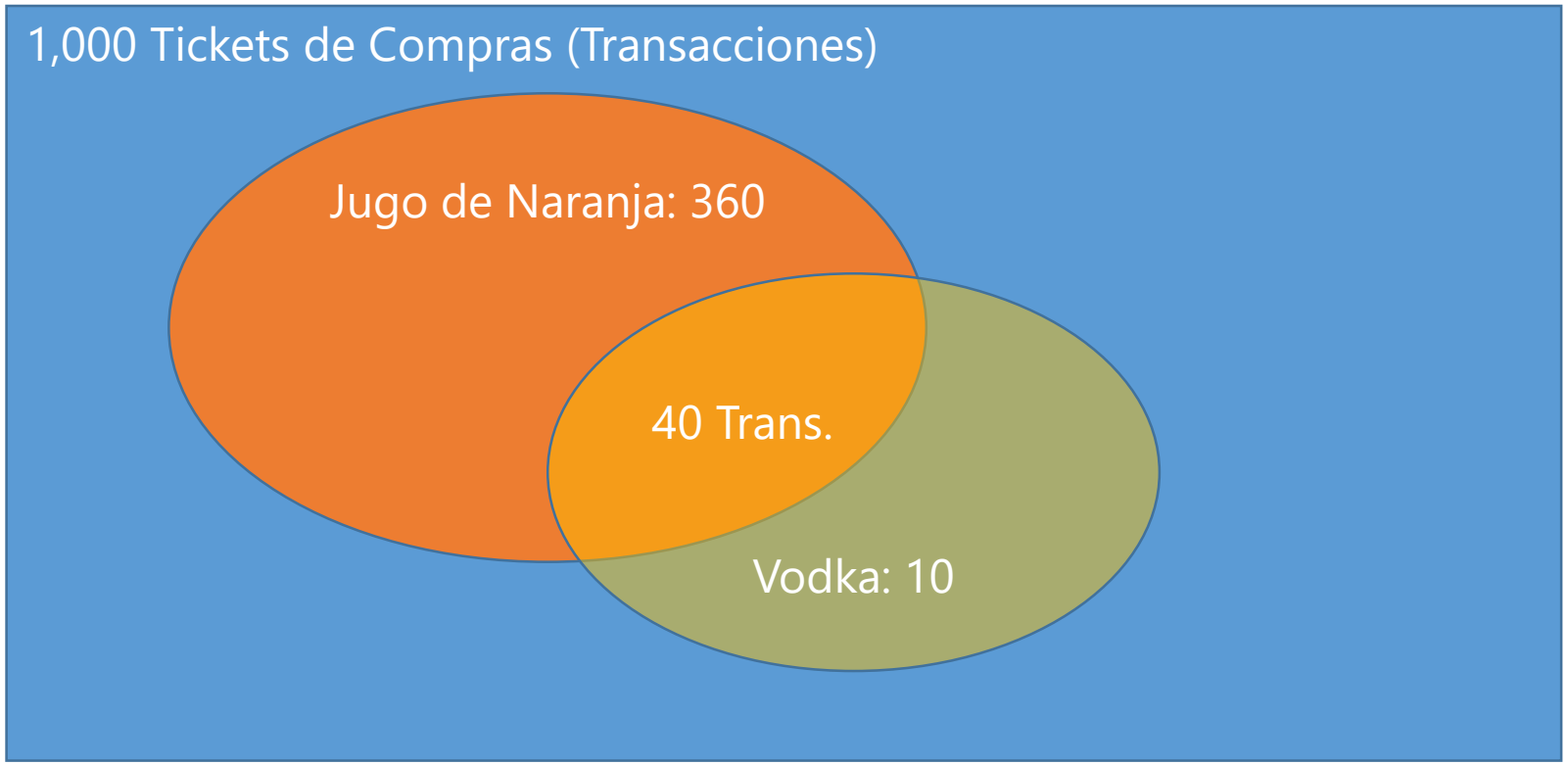


Soporte Vodka=50/1000

Soporte Vodka=0.05

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Vodka



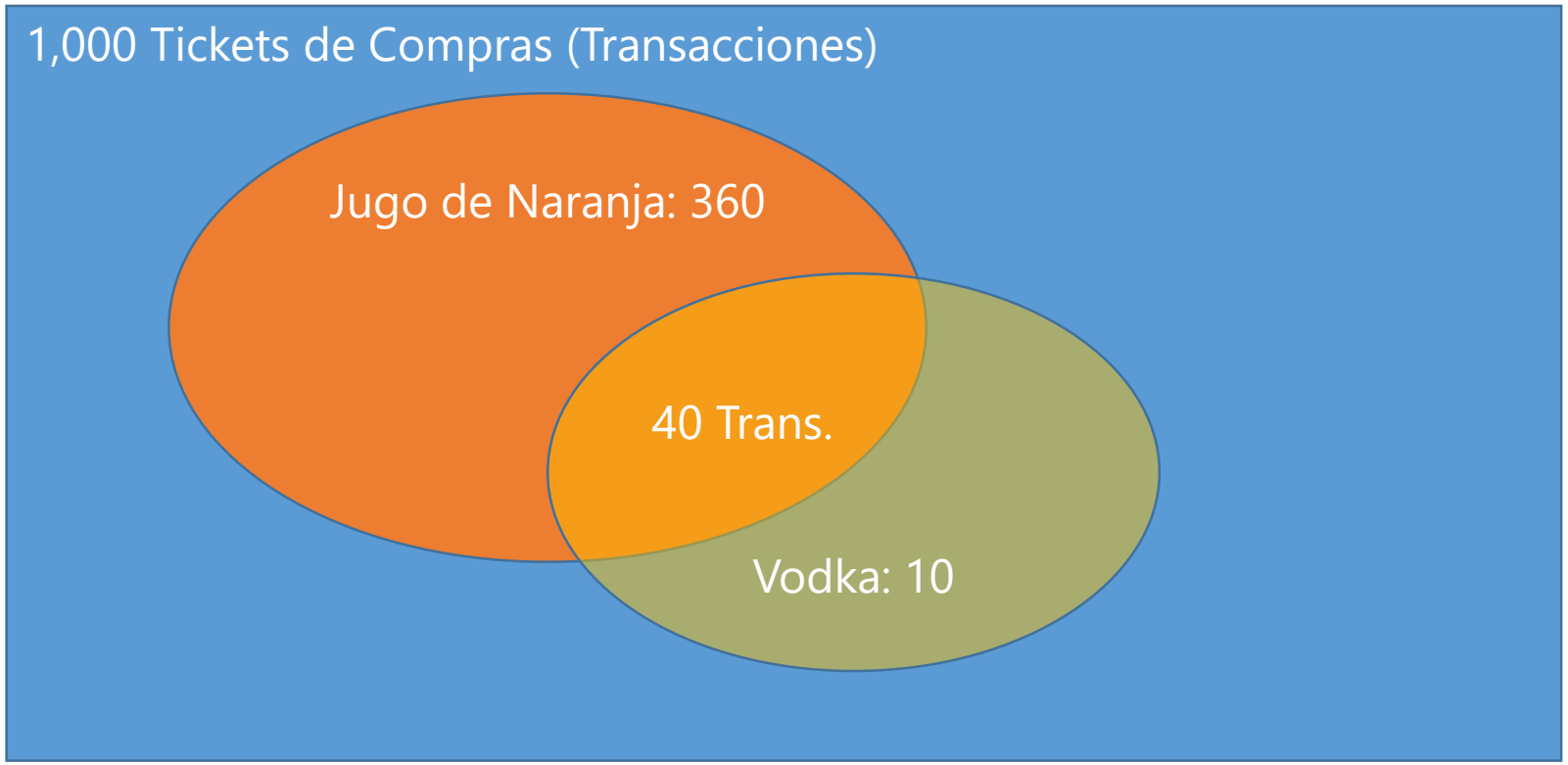
Soporte Jugo de Naranja **Y** Vodka:
40/1000 = 0.04

Confianza (A ->B) =
Soporte (A ->B) / Soporte (A)

Confianza (Vodka -> Jugo)=
Soporte (Vodka -> Jugo)/ Soporte (Vodka)
40/50=0.8

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Vodka



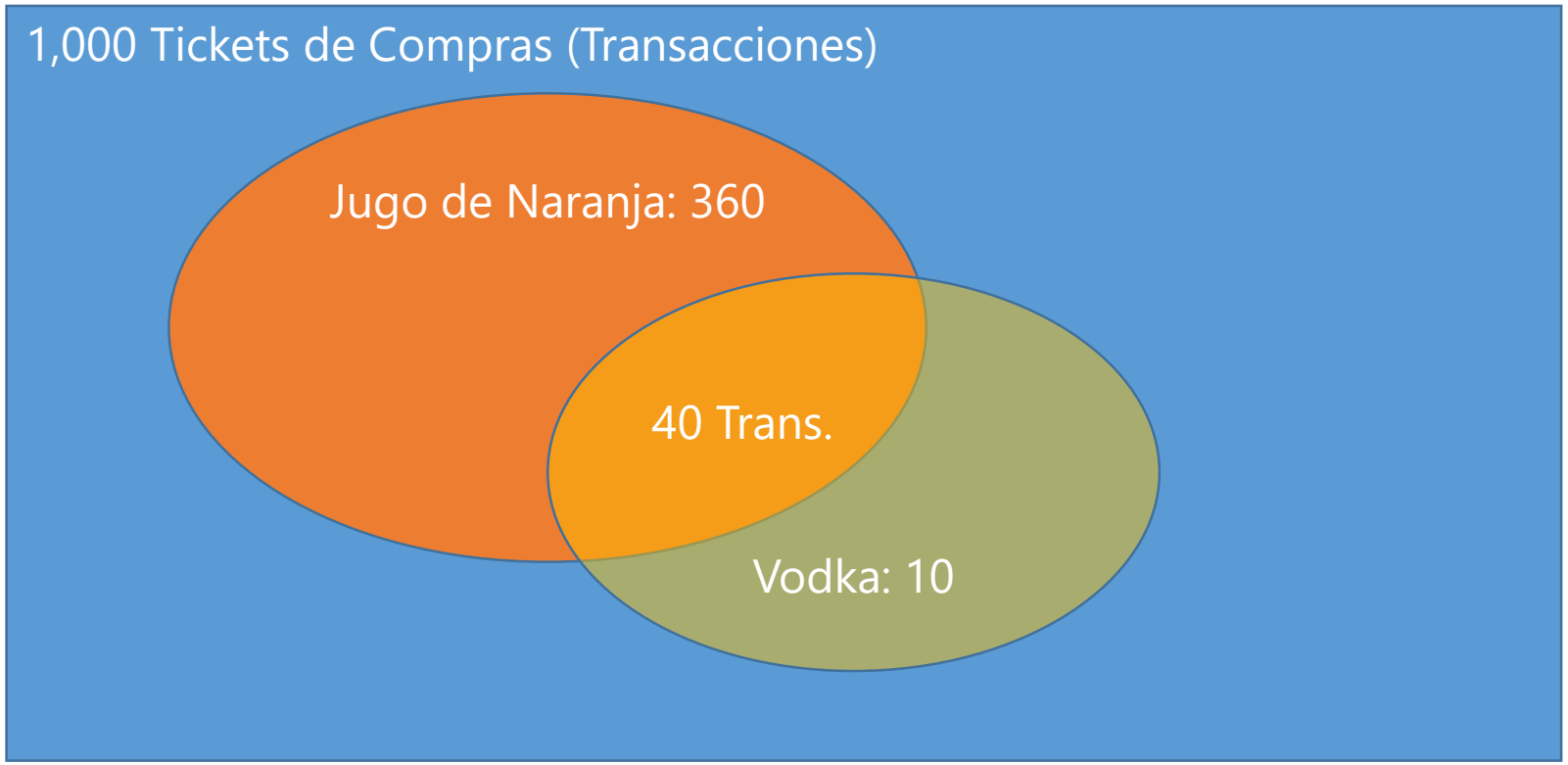
Soporte Jugo de Naranja **Y** Vodka:
40/1000 = 0.04

Confianza (A ->B) =
Soporte (A ->B) / Soporte (A)

Confianza (Jugo->Vodka)=
Soporte (Jugo -> Vodka)/ Soporte (Jugo)
40/400=0.1

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Vodka

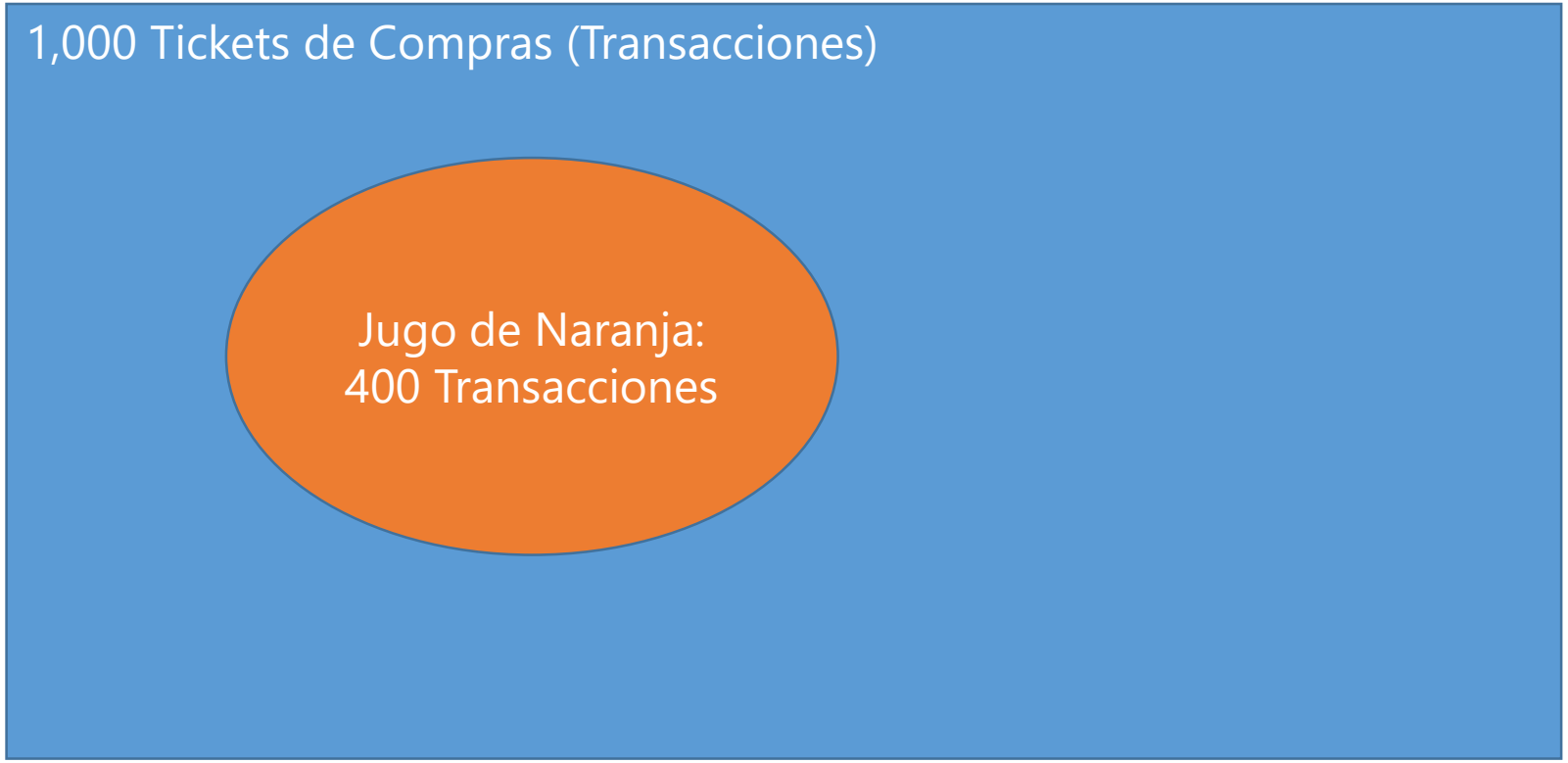


Lift (A ->B) =
Soporte (A ->B) / (Soporte (A) * Soporte (B))

Lift(Vodka -> Jugo)=
0.04/ (0.4*0.05)
0.04/0.02=2

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Pan

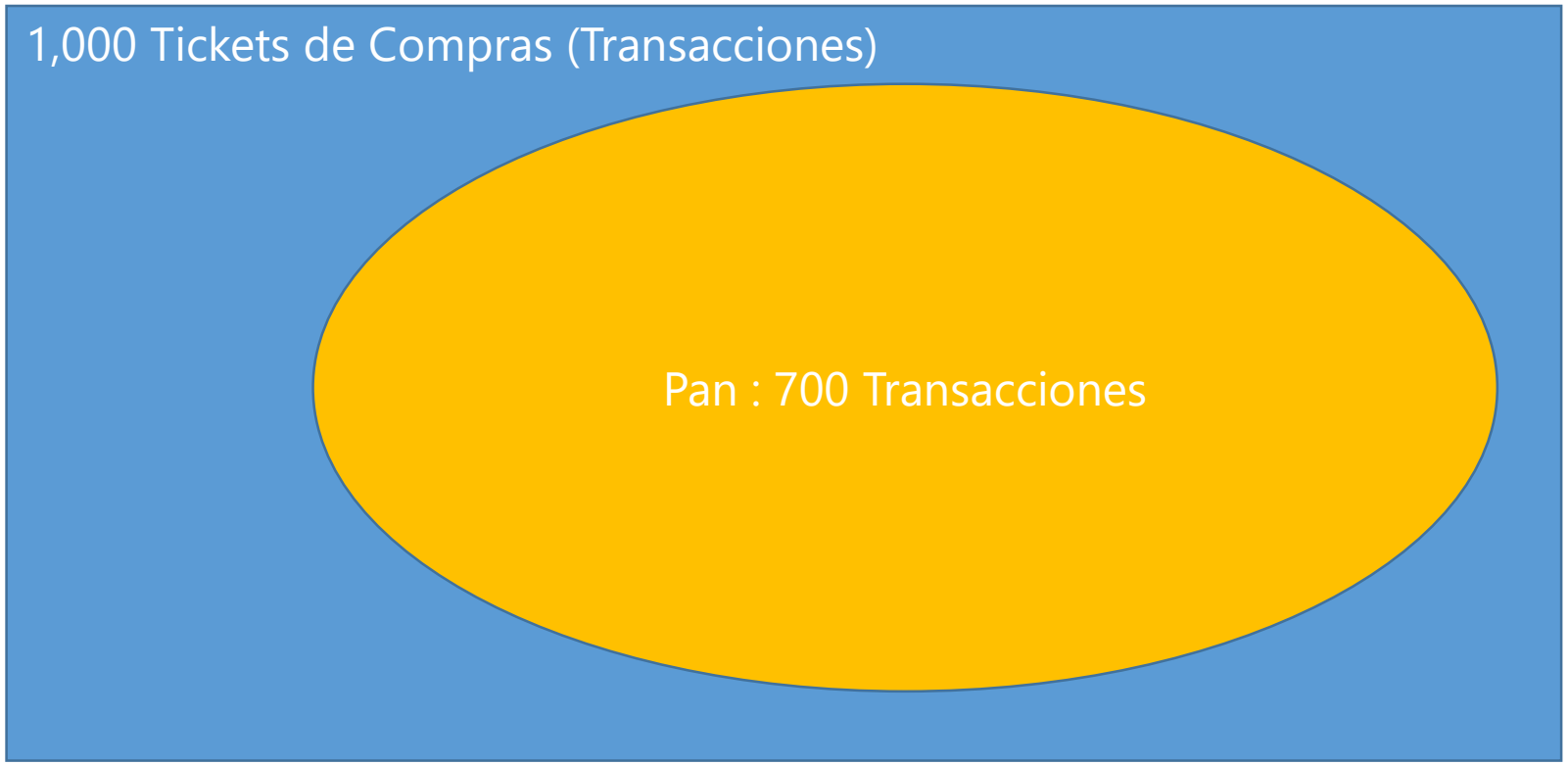


Soporte Jugo de Naranja=400/1000

Soporte Jugo de Naranja=0.4

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Pan

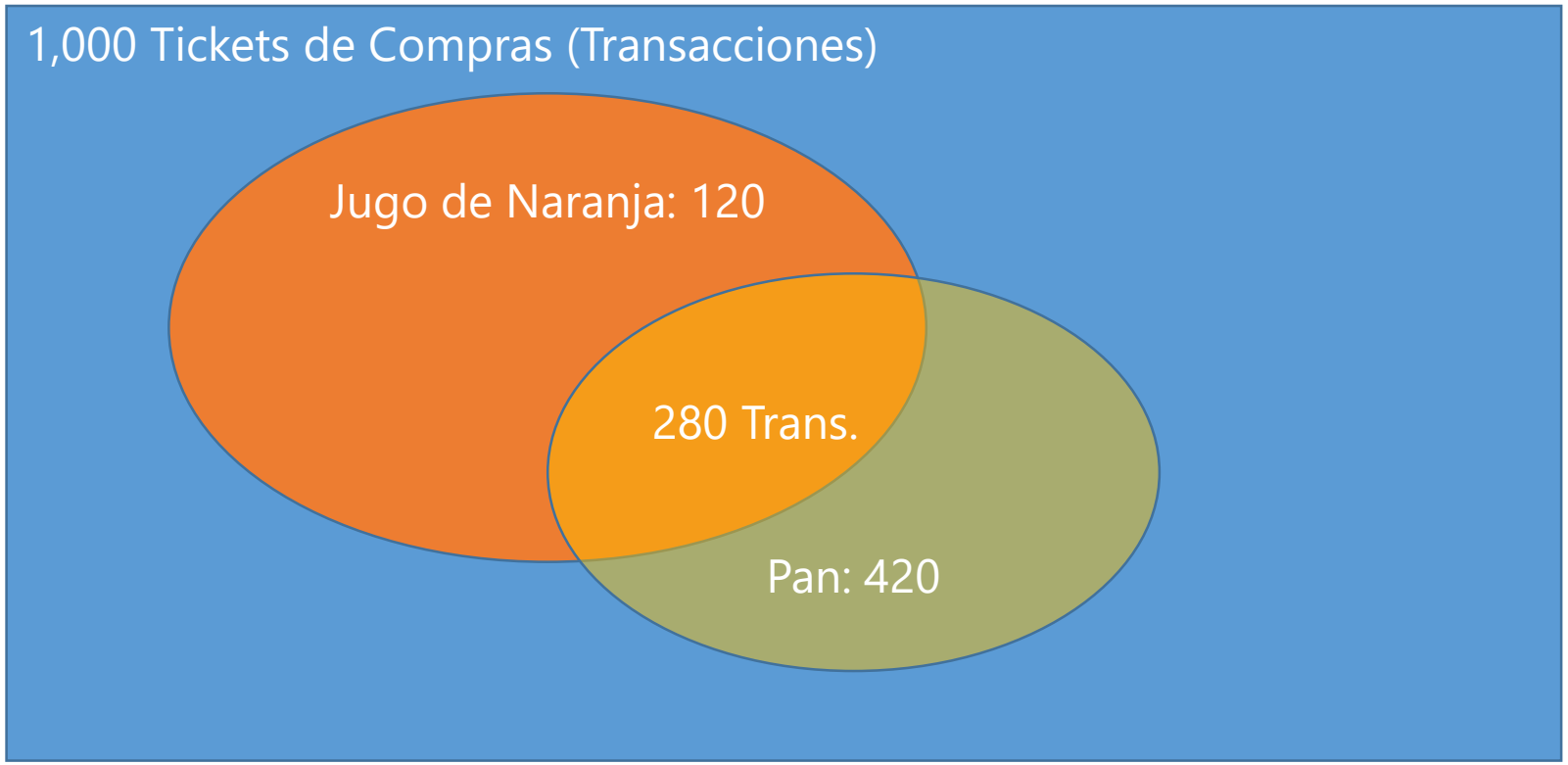


Soporte Pan=700/1000

Soporte Pan=0.7

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Pan



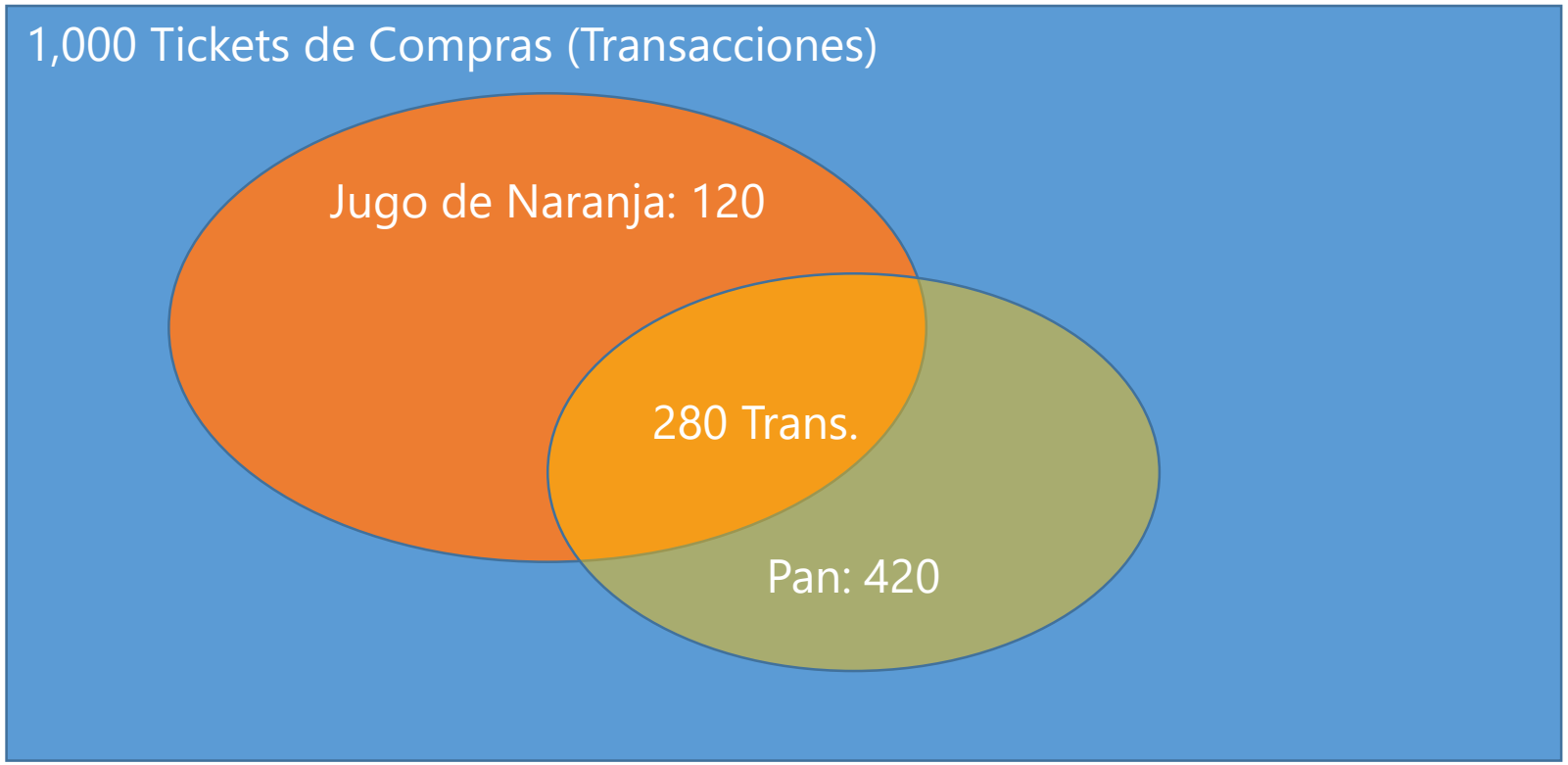
Soporte Jugo de Naranja **Y** Pan:
280/1000 = 0.28

Confianza (A ->B) =
Soporte (A ->B) / Soporte (A)

Confianza (Pan -> Jugo)=
Soporte (Pan -> Jugo)/ Soporte (Pan)
280/700=0.4

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Pan



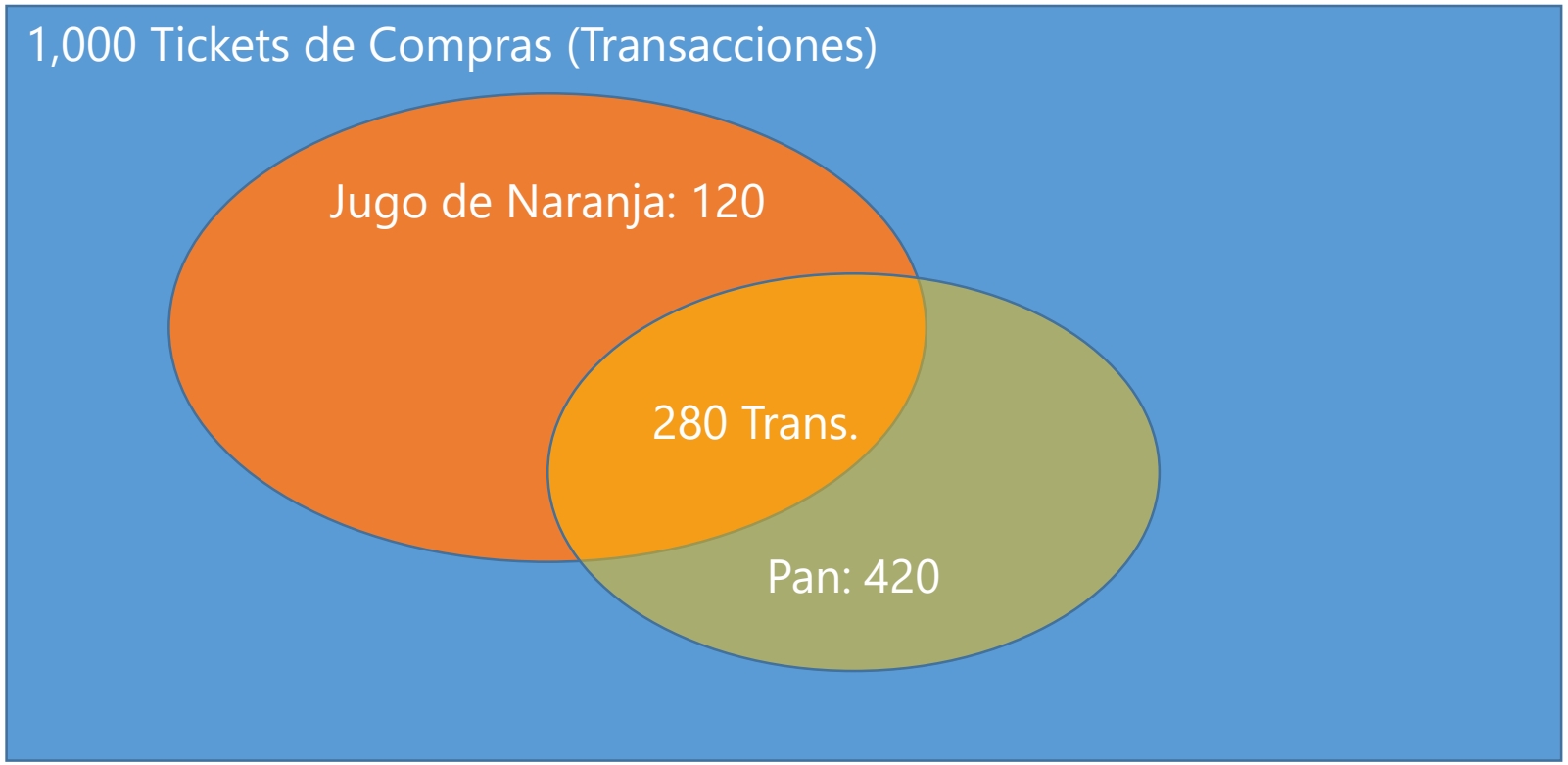
Soporte Jugo de Naranja **Y** Pan:
280/1000 = 0.28

Confianza (A ->B) =
Soporte (A ->B) / Soporte (A)

Confianza (Jugo -> Pan)=
Soporte (Jugo -> Pan)/ Soporte (Jugo)
280/400=0.7

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Reglas de Asociación: Ejemplo Jugo y Pan



Lift (A ->B) =
Soporte (A ->B) / (Soporte (A) * Soporte (B))

Lift(Pan -> Jug)=
0.28/ (0.7*0.4)
0.28/0.28=1

Ejemplo tomado del canal de INCAE en Youtube https://www.youtube.com/watch?v=i9-UfF2a38Q

Bibliografía

Reglas de asociación y algoritmo Apriori con R

By Joaquín Amat Rodrigo, 2018

Association Rule Mining Models and Algorithms

By Zhang, Chengqi, Zhang, Shichao, 2002

R Data Analysis Cookbook

by Kuntal Ganguly,2017