MBA

La técnica de **Market Basket Analytics** (también conocida como **Market Basket Analysis**, o MBA) es un enfoque de análisis de datos utilizado para identificar **asociaciones o patrones de compra entre productos**. Su objetivo es responder a la pregunta: **“¿Qué productos se compran juntos frecuentemente?”**

**¿En qué consiste?**

Market Basket Analysis se basa en el análisis de **transacciones de compra** (como tickets de supermercados, ventas en línea, etc.), buscando combinaciones de ítems que ocurren juntas más a menudo de lo que se esperaría por azar.

Se apoya en una técnica llamada **“Reglas de asociación”**, siendo el algoritmo más común el **Apriori**.

## Ejemplo típico

Un supermercado analiza sus tickets de compra y descubre:

“El 65% de los clientes que compran pan y manteca también compran leche.”

Esto puede usarse para:

* Ubicar esos productos cerca unos de otros.
* Crear promociones del tipo “Combo pan + manteca + leche”.
* Sugerencias automatizadas en ecommerce.

## Métricas principales en MBA

1. **Soporte (Support)**
   * Frecuencia con que un conjunto de productos aparece en las transacciones.
   * Ejemplo: el 30% de los tickets incluyen “pan y leche”.
2. **Confianza (Confidence)**
   * Probabilidad de que se compre B dado que se compró A.
   * Ejemplo: el 70% de quienes compran pan también compran manteca.
3. **Lift**
   * Qué tan fuerte es la relación comparada con la independencia.
   * Un lift > 1 indica que los productos se compran juntos más de lo esperado por azar.

## Dataset simulado (5 tickets)

| **Ticket** | **Pan** | **Leche** | **Manteca** | **Cereal** | **Queso** |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 0 | 0 | 1 |
| 2 | 0 | 1 | 1 | 0 | 1 |
| 3 | 1 | 1 | 1 | 1 | 0 |
| 4 | 1 | 1 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 | 1 | 0 |

✅ Valores 1 = se compró el producto; 0 = no se compró

## 📊 Cálculo de métricas para una regla

### Ejemplo: **{Pan} ⇒ {Leche}**

#### 1. Soporte

Proporción de transacciones que contienen ambos: Pan y Leche.

* Aparece en ticket 1, 3 y 4 → 3 de 5  
  ➡️ **Soporte = 3 / 5 = 0.60**

#### 2. Confianza

Proporción de transacciones con Pan que también tienen Leche.

* Pan aparece en ticket 1, 3 y 4 → 3 veces
* En esas 3 veces, Leche también aparece → 3 veces  
  ➡️ **Confianza = 3 / 3 = 1.00**

#### 3. Lift

Comparación entre la confianza y la frecuencia de Leche.

* Leche aparece en ticket 1, 2, 3, 4 → 4 de 5 → **Soporte Leche = 0.80**  
  ➡️ **Lift = Confianza / Soporte(Leche) = 1.00 / 0.80 = 1.25**

## 📌 Interpretación

* **Soporte 0.60** → El 60% de los tickets contienen Pan y Leche.
* **Confianza 1.00** → Siempre que hay Pan, también hay Leche.
* **Lift 1.25** → La presencia de Pan incrementa la probabilidad de que se compre Leche un 25% más que si fueran independientes.

**Ejemplo matemático:**

## Supuestos

* Sea **A** un conjunto de productos (ej. {Pan})
* Sea **B** otro conjunto de productos (ej. {Leche})
* Sea **D** el conjunto de todas las transacciones (tickets)

Definimos:

* P(A): proporción de transacciones que contienen A
* P(B): proporción de transacciones que contienen B
* P(A∩B): proporción de transacciones que contienen **ambos**, A y B

## 📘 Fórmulas matemáticas

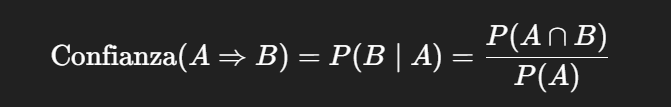
### 1. **Soporte** (Support)

Es la **frecuencia conjunta** con la que ocurren A y B en el total de transacciones.



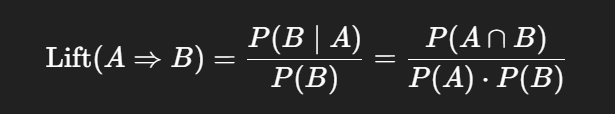
### 2. **Confianza** (Confidence)

Es la **probabilidad condicional** de que ocurra B dado que ocurrió A.



### 3. **Lift**

Mide si A y B ocurren juntos más seguido que si fueran eventos independientes.

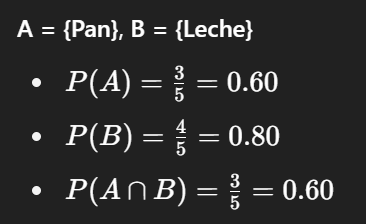


* **Lift > 1**: A y B se potencian.
* **Lift = 1**: A y B son independientes.
* **Lift < 1**: A y B se inhiben.

## ✅ Ejemplo con números

Usamos los valores anteriores para:

**A = {Pan}**, **B = {Leche}**



Ahora:

