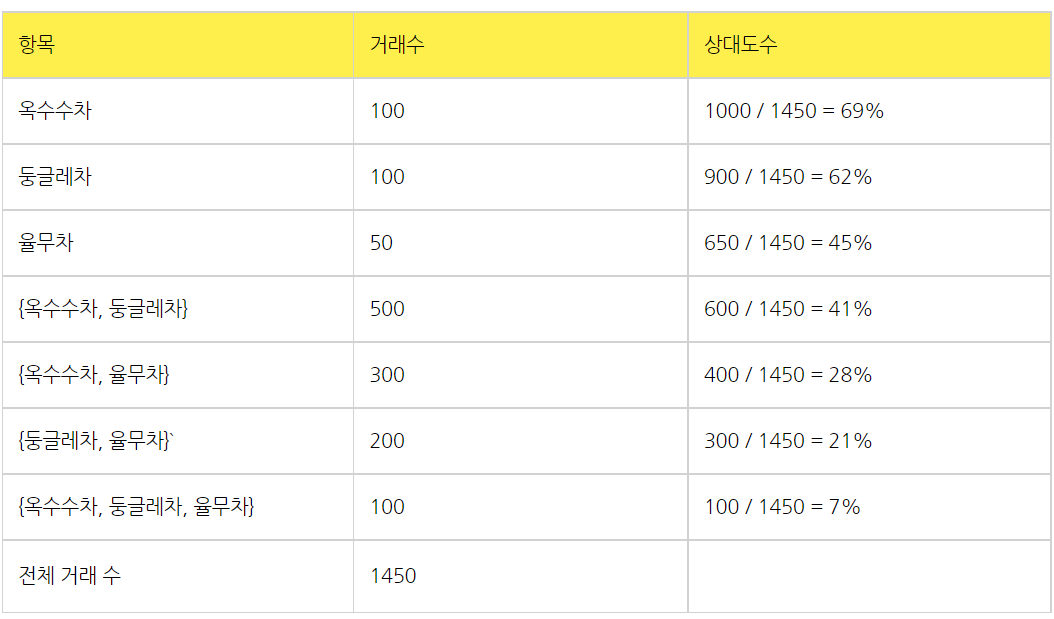
연관 규칙(Association rules) : 항목들이 얼마나 함께 발생하는지, 서로 얼마나 연관되어 있는지 분석하는 규칙 (Descriptive Analysis), 널리 알려진 예시로 마트에서 기저귀를 사가는 사람들은 맥주도 자주 사간다는 장바구니 분석 이라고도 불린다.

연관 규칙 분석을 위해 필요한 데이터 형태 :

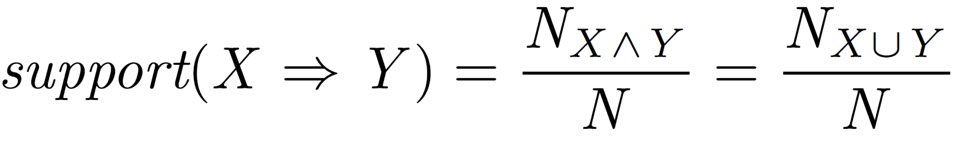


Or raw data 🡪 정리를 통해 위의 형태로 만들어야 함

{거래1 : 옥수수차} {거래2 : 옥수수차, 율무차}, {거래3 : 율무차}…{거래 1450 : 둥굴레차, 율무차}

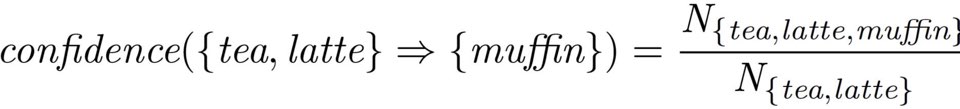
발생 빈도 측정 3가지 척도 : Support, Confidence, Lift

Support : 전체 경우의 수에서 두 아이템이 같이 나오는 비율. 분석 대상이 되는 항목들의 전체 발생수가 중요함 (이벤트 대상이 되는 전체 Mass)



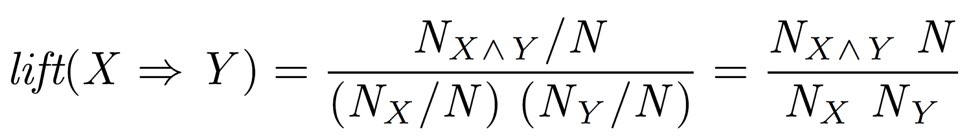


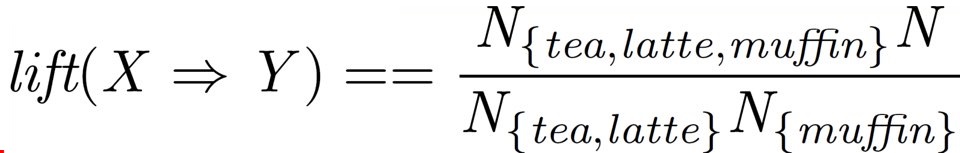
Confidence : X가 나온 경우중 X와 Y가 함께 나올 비율



차와 라떼를 주문한 고객이 머핀도 시킬 Confidence

Lift : X와 Y가 같이 나오는 비율을 X가 나올비율과 Y가 나올 비율의 곱으로 나눈 값





LIFT가 1보다 크면 X,Y는 양의 상관관계, 1이면 독립의 관계, 1 이하이면 음의 상관관계를 가진다고 본다. (아무것도 안했을 때 Y를 뽑을 확률보다 X를 뽑앗으면 그 다음에 Y뽑을 확률이 낮아진다)