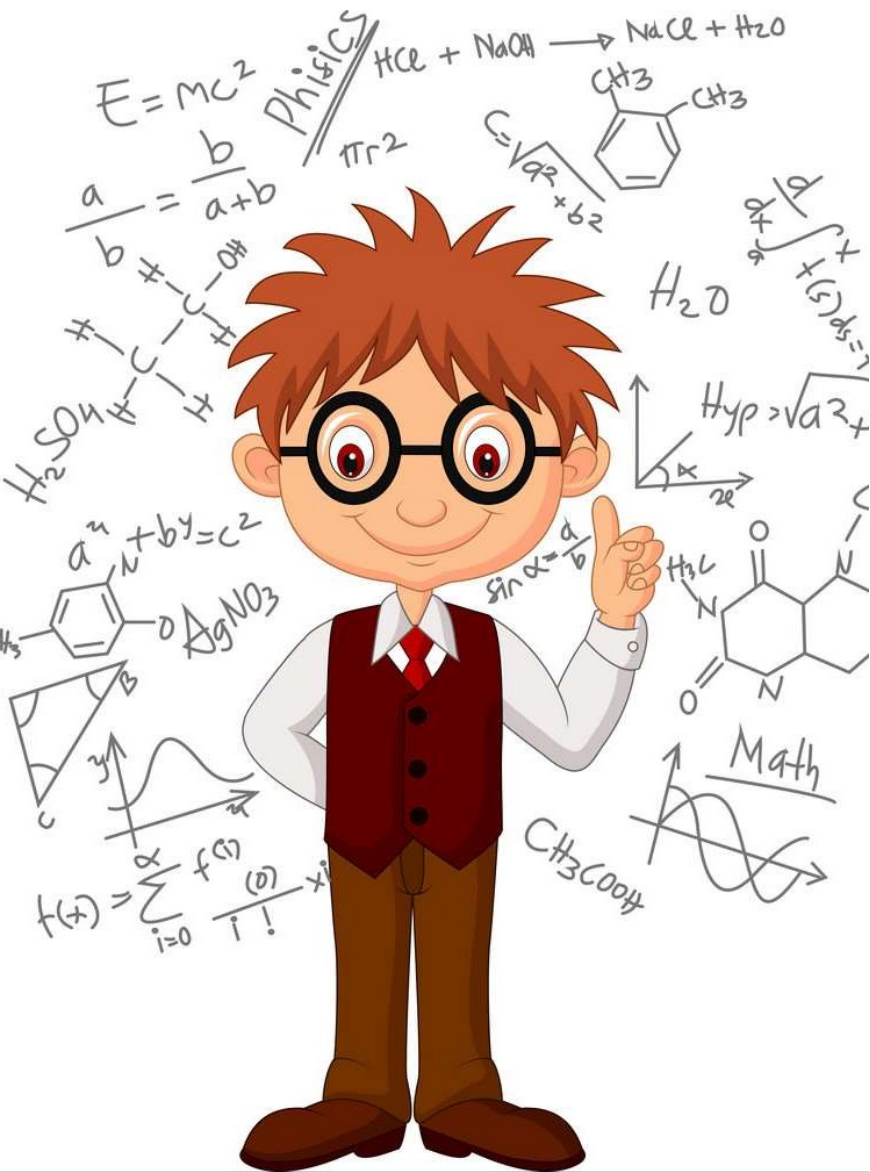


Predictive Analytics

Lecturer: Marina Iantorno

E-mail: miantorno@cct.ie





In today's class we will cover:

Association Rule:

☐ Apriori algorithm

Association Rule

There is a story in Machine Learning that connect two products of a grocery store/supermarket:



Do you think that could be any association between these two products?

Association Rule

A company started analysing the products they were selling, and performed some Machine Learning to check the products that people were purchasing, what were the communalities, the products that people may be interested in, and they had a sample of thousands of transaction, so, there were thousand of people's purchasing that were analysed in this study. They found out that in afternoon, between 6 and 9 pm, very often people who buy diapers also buy beer.

At the beginning it was so strange and the team tried to wonder why would that happen if those products are not connected, and even they hesitated whether this association was wrong while the maths were right.



Association Rule

But then, after they gave a thought about this, they realised that they had to put the eye on the person who was doing the transaction to get the answer, and that is how they associated the time with the father of the baby. For instance, it would be very likely that the mother is taking care of the baby, running out of pampers, and called the father to buy it when he is going back home. And after tired after work, stop at the supermarket, buys the diapers and also a pack of beers to enjoy in the evening, or to watch the game, because the matches usually happened in the evening as well.



Association Rule

After that analysis, supermarkets adopted two different strategies to sell more:

- 1) Placing diapers and beer together, so it would be a temptation for the buyer and it would increase the sales (impulsive purchase).
- 2) Placing diapers and beer far from each other, so on this way the customer should walk across the whole store and could buy more products on that way from the diapers to the beer.

There many interesting Marketing strategies around this analysis, but here our question is, how to reach to this result?



Association Rule

Apriori algorithm

The apriori algorithm means something like: “People who bought this, also bought...”

Although this is not only an algorithm to use in stores, it is well known as “Market Basket Analysis”.

We can analyse one transaction and think that the person who bought “X” product may buy “Y” product on the same basket or on the same cart. For example, let’s say that when a customer buys tea or coffee, it is likely that also buys sugar or milk with it. This is an association.

If a person orders burger at a restaurant, probably also orders fries, and if the person orders both, probably also orders sauces like ketchup.



Association Rule

Apriori algorithm

There are several examples in the industry that use this algorithm and you probably use them periodically.

Best Value

Buy [Philips Norelco BG2020 Men's Bodygroom](#) and get [MANGROOMER Do-It-Yourself Electric Back Hair Shaver](#) at an **additional \$5.00 off** Amazon.com's everyday low price.



Total List Price: \$79.98
Buy Together Today: \$69.94
You Save: \$10.04

Buy both and save

Customers who bought this item also bought



[MANGROOMER Do-It-Yourself Electric Back Hair Shaver](#)
★★★★★
\$39.95



[Panasonic ER421KC Nose and Ear Hair Trimmer, Wet/Dry, Lighted](#)
★★★★★
\$14.99



[Tweezerman Deluxe Men's Grooming Kit](#)
\$19.99

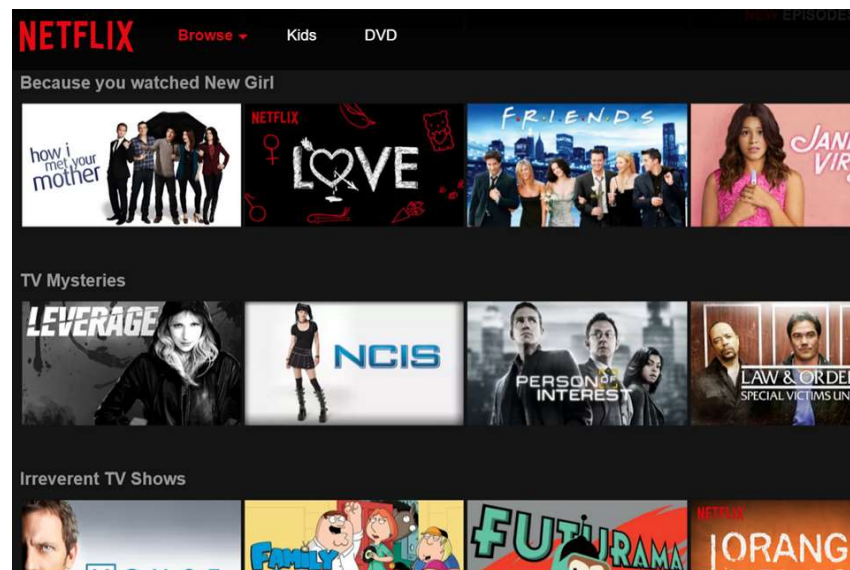


[Tweezerman Stainless Steel Ingrown Hair Splintertweezer](#)
★★★★★
\$8.97



[Brokeback Mountain \(Widescreen Edition\)](#)
★★★★★
\$14.87

► [Explore similar items](#) : [Health & Personal Care](#) (21) [DVD](#) (18) [Music](#) (3) [Books](#) (2)



Association Rule

Apriori algorithm

This algorithm has three parts: Support, Confidence and Lift.

The diagram illustrates the three metrics used to evaluate an association rule $X \Rightarrow Y$. Three blue arrows originate from the rule and point to the respective formulas:

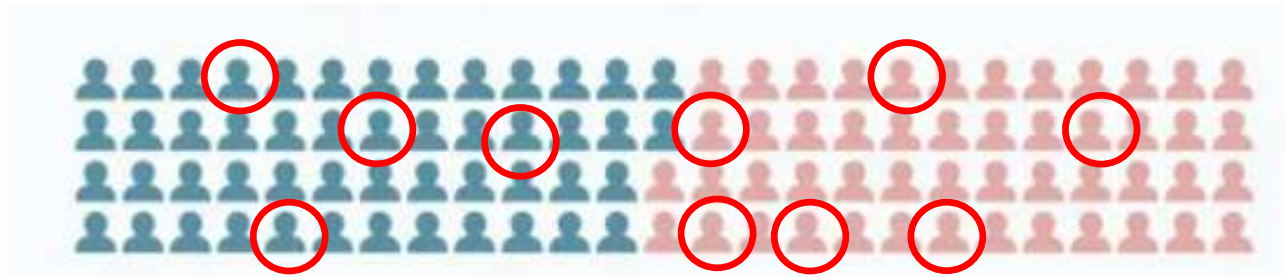
$$\begin{aligned} \text{Support} &= \frac{\text{Frequency}(X,Y)}{N} \\ \text{Confidence} &= \frac{\text{Frequency}(X,Y)}{\text{Frequency}(X)} \\ \text{Lift} &= \frac{\text{Support}}{\text{Support}(X) * \text{Support}(Y)} \end{aligned}$$

Let's explore them in details

Association Rule

Apriori algorithm

Support: Imagine that there are 100 people in a sample, and within those there are 10 who watched “Narcos”.

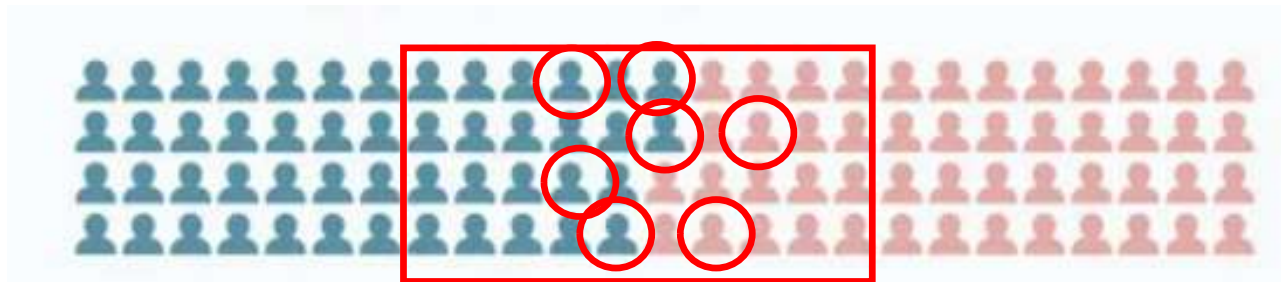


$$\text{Support} = \frac{\text{Frequency}(X,Y)}{N} \longrightarrow \text{Support} = 10/100 = 10\%$$

Association Rule

Apriori algorithm

- Confidence: Confidence is the number of people who have seen series 1 and series 2, divided those who watched 1. In other words, let's say that there are 40 people who have watched "El Chapo" and some of them are likely to watch "Narcos". Series 1 is "El Chapo", Series 2 is "Narcos".

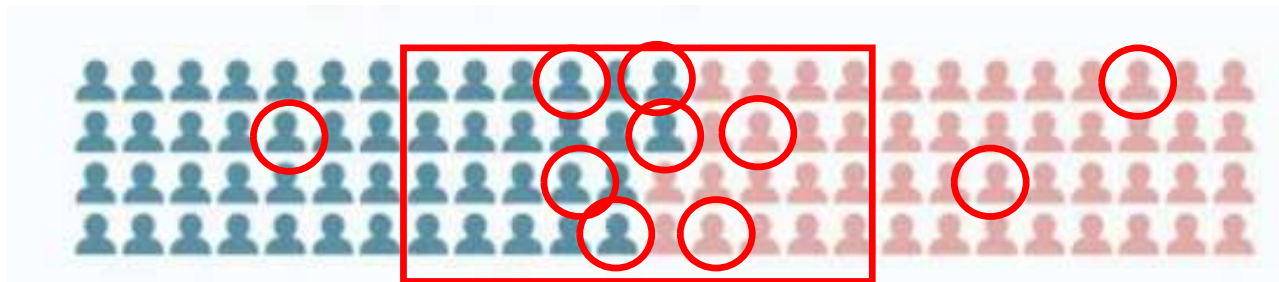


$$\text{Confidence} = \frac{\text{Frequency}(X,Y)}{\text{Frequency}(X)} \longrightarrow \text{Confidence} = 7/40 = 17.5\%$$

Association Rule

Apriori algorithm

- Lift: To calculate the lift we need to use the confidence and the support. What we calculate here is, if we randomly suggest a person to watch “Narcos”, what is the likelihood that they will like it? Can we improve this result with something that we already calculated?



$$Lift = \frac{Support}{Support(X) * Support(Y)} \longrightarrow Lift = 17.5\% / 10\% = 1.75$$

Note: The Lift is not the number of people who will like it, it will be analysed as a ranking against other lifts.



Practice
Time!

Practice Time

You will find on Moodle a file called “groseries.csv”, and it belongs to a store that records the products that are sold every week.

our mission here is to return the best association rules to offer the best deals to the clients.

Let's try this together on Python!

