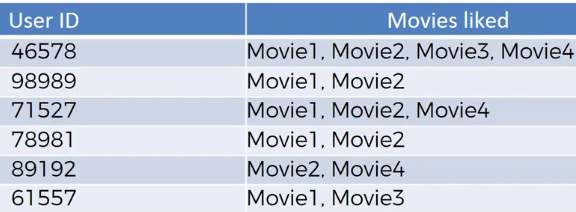
**Apriori**

* Ex: Fathers buying beer when sent out for diapers
* Ex: Supermarkets spreading out bread + milk b/c customers already buy them together so customers have to walk throughout entire store + have a higher chance of buying more products
* Apriori algorithm helps us get the info from data = about people who bought item 1 bought item 2, 3, 4… etc (**market basket analysis**) or who did 1 thing did thing 2
* This is association/rule learning of analyzing when things occurs in pairs/are combined together for some reason + discovering those rules

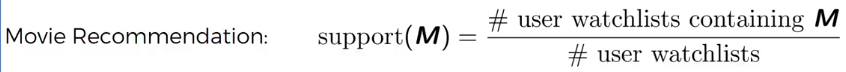


* Seems like those who like movie 1 will like movie 2, and those who like movie 2 will like movie 4
* Want to find strongest “rules” to build out business decisions based on rules found in the data
* **Market basket optimization**

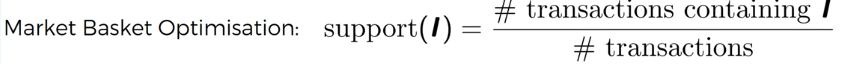


* Possible rules: burgers = FF’s, burgers = ketchup, burgers + FF’s = ketchup, fruits = veggies

* Apriori algorithm has 3 parts, all similar to Naïve Bayes classifier math:
* **Support**



* Support for movie M = # of users who watched movie M divided by total # of users



* Support for item I = # of users who bought item I divided by total # of users
* **Confidence**



* Confidence M1 is related to M2 = # of users who watched both / users who watched M1



* Confidence I1 is related to I2 = # of users who bought both / users who bought I1
* **Lift = confidence / support =** improvement in prediction





* Ex: Randomly suggest to a person to watch movie 2, what is the **likelihood** they’d like it?
* Say 10% like the movie, so likelihood = 10%
* Now, can we “prove” this result given some prior knowledge?
* Therefore, only suggest movie 2 to those who saw movie 1. Likelihood now could jump to 17.5%
* **Lift = 17.5/10 = 1.75**
* Performing the algorithm
* Quite slow b/c it goes through all combos of DP’s (not just pairs + triplets, but even more complex combos)
* Need to set up limitations
* 1) Set up minimum support + confidence 🡺 2) Take all subsets in transactions that have higher support than the minimum 🡺 3) Take all rules of these subsets that have higher confidence than minimum 🡺 4) sort rules by decreasing lift
* Those combinations with highest lift are those rules we base business decisions off of
* Recommender systems are good examples of where Apriori could work, but they use more and/or more complex algorithms + systems