The link to the dataset: [transaction\_data.csv - Google Drive](https://drive.google.com/file/d/1vE3xyM-hbZMB5RS8c3H13Ll1PnFXJ7Ui/view)   
  
We've all been purchasing more from online e-commerce sites recently. You've probably noticed an up-sell tool called ‘frequently bought together’ on most of these sites, such as Amazon, where it predicts what other goods will go well with the one you've just placed to your basket. Customers can add all of the things displayed under this feature to their basket or just the ones they need. Amazon accomplishes this using a process known as 'item-to-item collaborative filtering,' in which it runs suggestion algorithms based on the customer's item-search history to optimize the purchasing experience. It operates similarly in the case of offline stores.

Consider the simple example of bread and jam. If the shopkeeper notices an increase in bread sales, he can further upsell by lowering the price of jam, causing more consumers to purchase them together.

'Market Basket Analysis' refers to the full process of analysing client shopping behaviour. It is a strategy for analysing things based on the premise that if we buy one item, we are obligated to buy or not-buy a group (or single) items. For example, if a consumer purchases bread, the likelihood of him or her purchasing jam increases.

A -> B  
This equation is called the Association Mining Rule. This can be thought of as an IF-THEN relationship. If item A is bought by a customer then the chances of item B being bought by the same user in the same transaction is found out. Here A is called the Antecedent and B the consequent. Antecedents are primary item that are found in the basket and consequents are the items that are found with an antecedent/group of antecedents. The metrics for measuring association are:

**Support:** It tells us about the combination of items bought together frequently. It gives the part of transactions that contain both A and B. We can filter out the less frequently occurring items-sets using support.

**Confidence:**It tells us how frequently the items A and B are bought together, for the no. of times A is bought.

**Lift:** It indicates the strength of a rule over the randomness of A and B being bought together. It basically measures the strength of any association rule.

More the lift more is the strength of the rule. If the lift is 3 for A -> B then it means if we buy A, the chances of buying B is 3 times.

So the process is to make rules for each item-set to figure out the metrics of its association so as to decide whether to include it in the analysis or not. But consider a large data-set having millions of user and transactions, thereby producing a large amount of item-sets. So to make rules for all of these would be a humongous task. This is where Apriori algorithm enters the scene.

**Apriori algorithm**uses frequently bought item-sets to generate association rules. It is built on the idea that the subset of a frequently bought items-set is also a frequently bought item-set. Frequently bought item-sets are decided if their support value is above a minimum threshold support value.