**Market Basket Analysis**

**What is it?**

Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, if you are in an English pub and you buy a pint of beer and don't buy a bar meal, you are more likely to buy crisps (US. chips) at the same time than somebody who didn't buy beer.

The set of items a customer buys is referred to as an **itemset**, and market basket analysis seeks to find relationships between purchases.

Typically the relationship will be in the form of a rule:

IF {beer, no bar meal} THEN {crisps}.

The probability that a customer will buy beer without a bar meal (i.e. that the antecedent is true) is referred to as the **support** for the rule. The conditional probability that a customer will purchase crisps is referred to as the **confidence**.

The algorithms for performing market basket analysis are fairly straightforward ([Berry and Linhoff](https://www.albionresearch.com/ibook/0471179809) is a reasonable introductory resource for this). The complexities mainly arise in exploiting taxonomies, avoiding combinatorial explosions (a supermarket may stock 10,000 or more line items), and dealing with the large amounts of transaction data that may be available.

A major difficulty is that a large number of the rules found may be trivial for anyone familiar with the business. Although the volume of data has been reduced, we are still asking the user to find a needle in a haystack. Requiring rules to have a high minimum support level and a high confidence level risks missing any exploitable result we might have found.