**DATA ASSIGNMENT 1**

1. What the data is.

The data is about the grocery shopping of my house.

1. What type of benefit you might hope to get from data mining.

From the data mining, I would at least know on which product am spending more money and which product is more necessary to buy for the house rather than spending in unnecessary things. I would also get to know how much am spending for one-time shopping and how can I reduce the money spend. I will also know when buying one item, which item is frequently brought together.

1. What type of data mining (classification, clustering, etc.) you think would be relevant.

Association rule mining is more relevant because Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets found in various kinds of databases such as relational databases, transactional databases, and other forms of repositories.

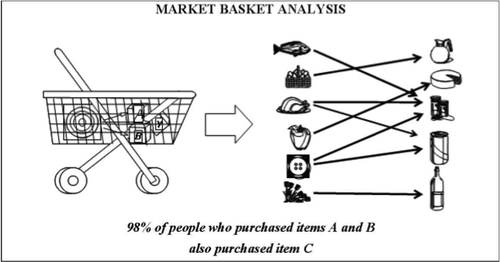
Association Rule Mining is sometimes referred to as “Market Basket Analysis”, as it was the first application area of association mining. The aim is to discover associations of items occurring together more often than you’d expect from randomly sampling all the possibilities.

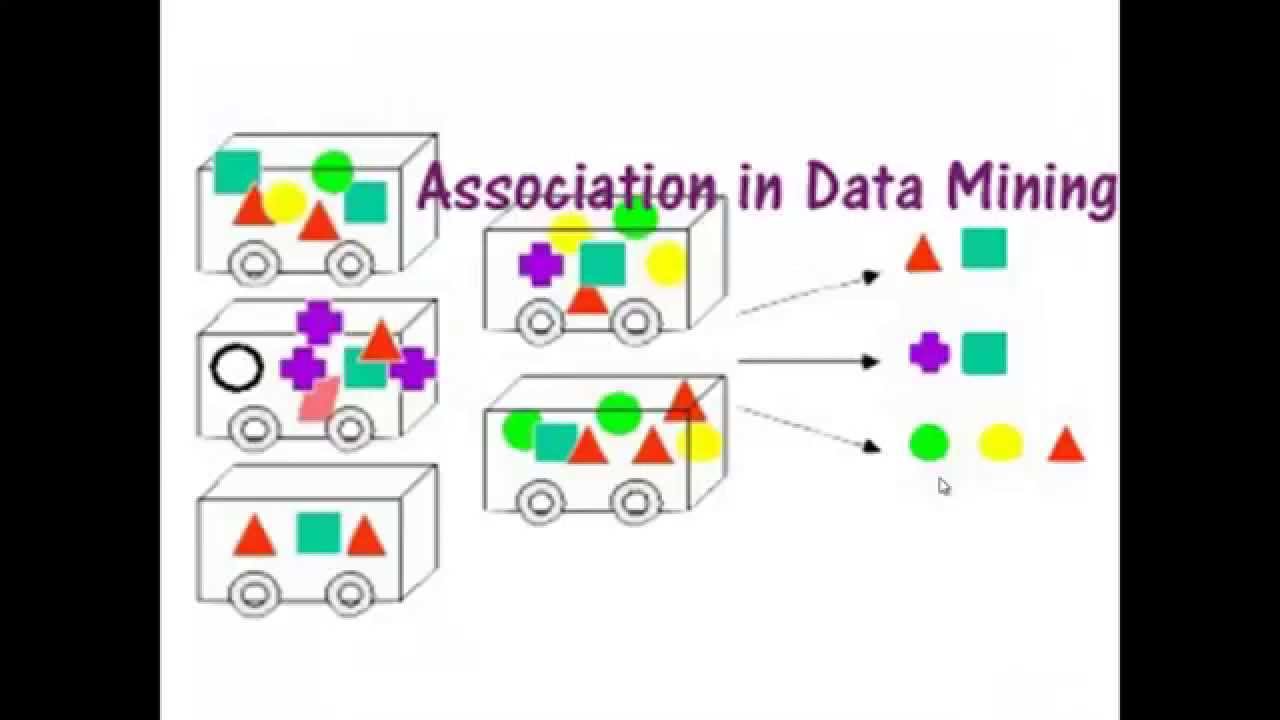
Suppose my grocery retail transactions database includes the following data:

* 1. Total number of transactions: 600,000
  2. Transactions containing milk: 7,500 (1.25 percent)
  3. Transactions containing sugar: 60,000 (10 percent)
  4. Transactions containing both milk and sugar: 6,000 (1.0 percent)

From the above figures, we can conclude that if there was no relation between milk and sugar (that is, they were statistically independent), then we would have got only 10% of milk purchasers to buy sugar too.

However, as surprising as it may seem, the figures tell us that 80% (=6000/7500) of the people who buy milk also buy sugar.  
This is a significant jump of 8 over what was the expected probability. This factor of increase is known as Lift – which is the ratio of the observed frequency of co-occurrence of our items and the expected frequency.





4. Name one type of data mining that you think would not be relevant, and describe briefly why not.

Regression wouldn’t be relevant because Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.

First, regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Second, in some situation’s regression analysis can be used to infer causal relationships between the independent and dependent variables.

We use [regression analysis](https://statisticsbyjim.com/glossary/regression-analysis/) to describe the relationships between a set of [independent variables](https://statisticsbyjim.com/glossary/predictor-variables/) and the [dependent variable](https://statisticsbyjim.com/glossary/response-variables/). [Regression analysis](https://statisticsbyjim.com/glossary/regression-analysis/) produces a [regression](https://statisticsbyjim.com/glossary/regression-analysis/) equation where the [coefficients](https://statisticsbyjim.com/glossary/regression-coefficient/) represent the relationship between each [independent variable](https://statisticsbyjim.com/glossary/predictor-variables/) and the dependent variable.