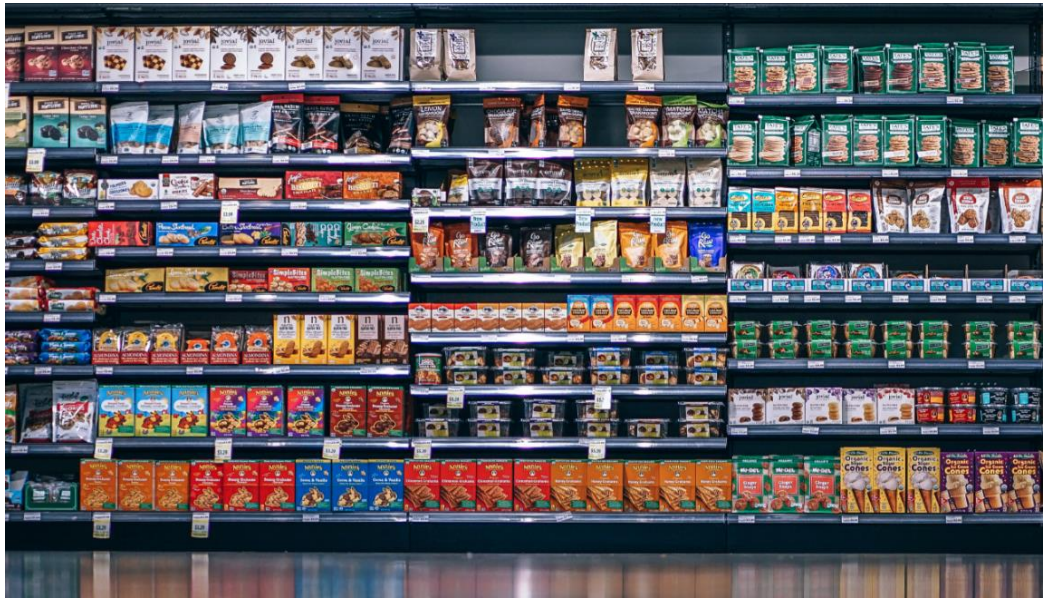
Association Rule Mining

DSA 6000: Data Science and Analytics, Fall 2019

Wayne State University

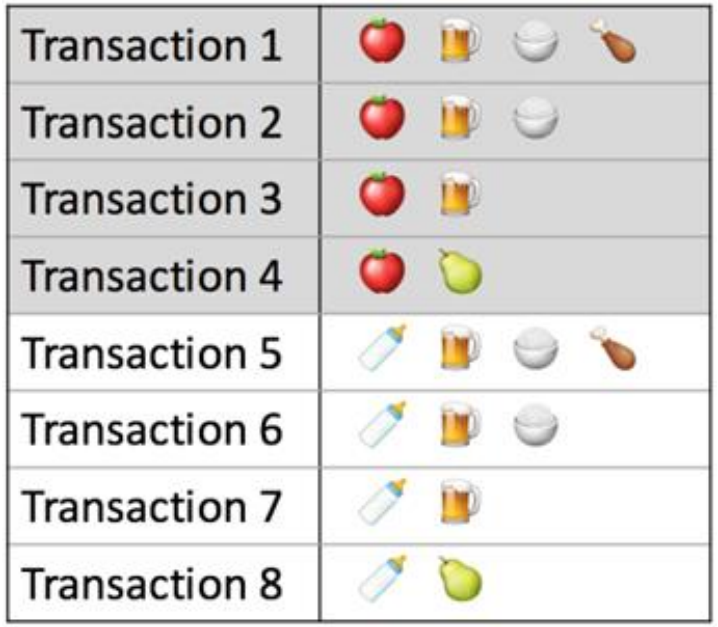
Association Rule Mining

- Discover interesting relations among variables in large datasets
- Typically applied to point-of-sale (POS) transaction data sets (usually huge). Also called Market Basket Analysis
- What items are frequently bought together?
- John watched movie X and Y, what other movie might John be interested in?



Association

- Measures co-occurrence, not causality
- Common metrics of association: Support, Confidence, Lift
- **Support (S)**: how popular an itemset S is, measured by the proportion of transactions in which an itemset appears.
  - The proportion of transactions in the dataset which contains the itemset S.



Support {🍎} = $\frac{4}{8}$

Support{Apple, Beer} = 3/8
Support{Chicken, Rice, Beer} = ?

The higher the support, the more frequently the itemset appears in customers' baskets

Confidence

- **Confidence (X->Y):** how likely item Y is purchased when item X is purchased, measured by the proportion of transactions with item X, in which item Y also appears.
- The proportion of the transactions that contains X which also contains Y.

Transaction 1	🍎 🍺 ⚪ 🍗
Transaction 2	🍎 🍺 ⚪
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	💉 🍺 ⚪ 🍗
Transaction 6	💉 🍺 ⚪
Transaction 7	💉 🍺
Transaction 8	💉 🍐

The higher the confidence, the stronger X is as an indicator of Y

$$\text{Confidence}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\}} = \frac{3/8}{4/8} = 0.75$$

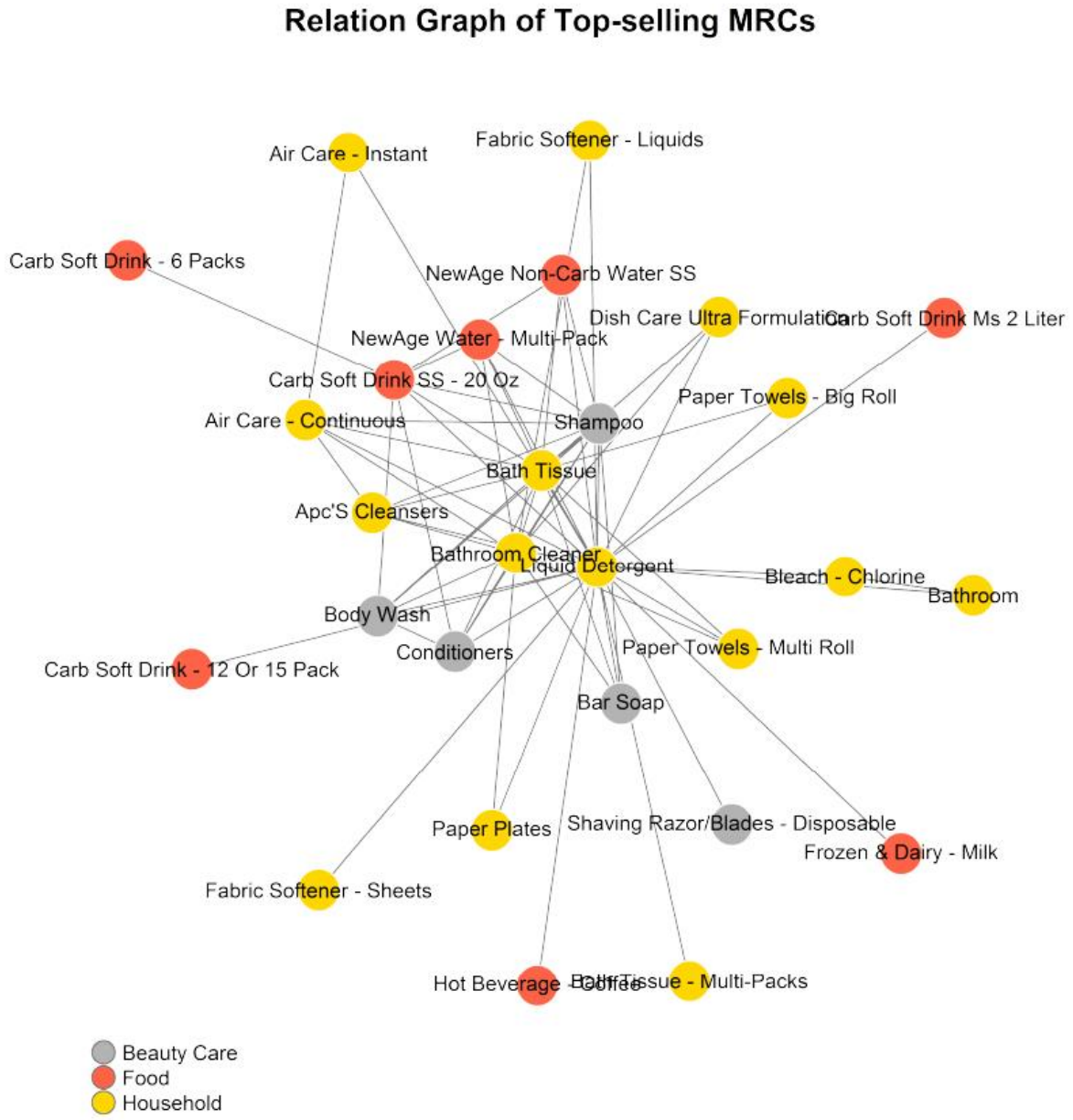
Lift

- Lift (X->Y): how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is.
- A lift value greater than 1 means that item Y is likely to be bought if item X is bought, while a value less than 1 means that item Y is unlikely to be bought if item X is bought.

Transaction 1	🍎 🍺 ⚪ 🍗
Transaction 2	🍎 🍺 ⚪
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	💉 🍺 ⚪ 🍗
Transaction 6	💉 🍺 ⚪
Transaction 7	💉 🍺
Transaction 8	💉 🍐

$$\text{Lift}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\} \times \text{Support}\{🍺\}} = 1$$

Relation graph of items



Algorithms

- Apriori and Eclat Algorithms
  - C program, very fast
  - http://www.borgelt.net/software.html
- "arules" package in R contains these algorithms
  - https://cran.r-project.org/web/packages/arules/index.html
- Examples
  - https://www.kaggle.com/msp48731/frequent-itemsets-and-association-rules