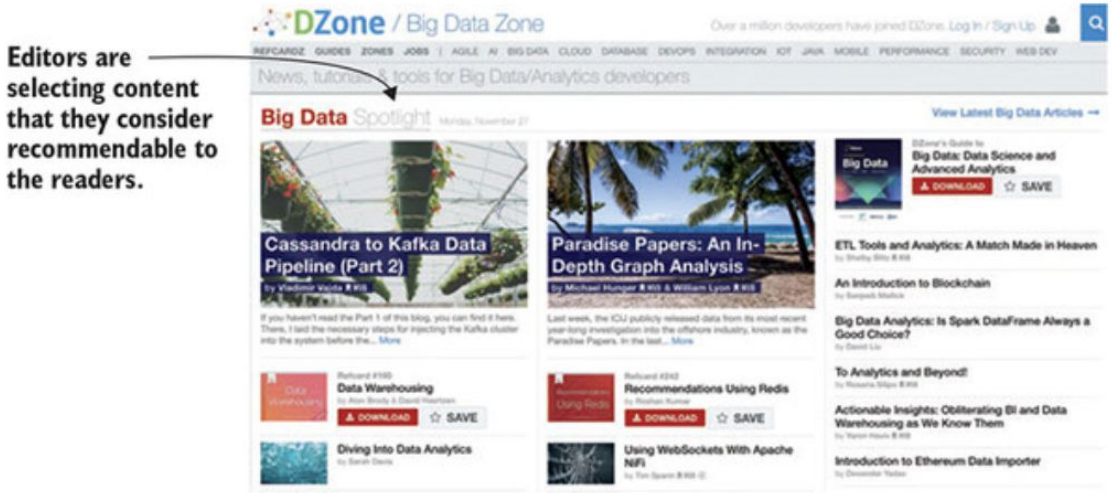
Non-Personalized Recommenders

Why non-personalized Recommendations?

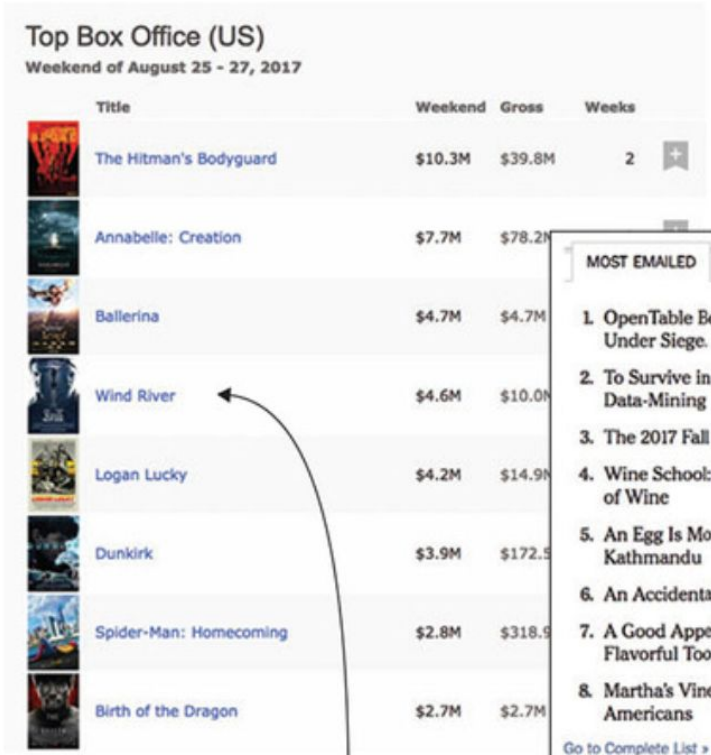
- Non-personalized recommendations can also show interesting content
- Provide recommendations when we have no data

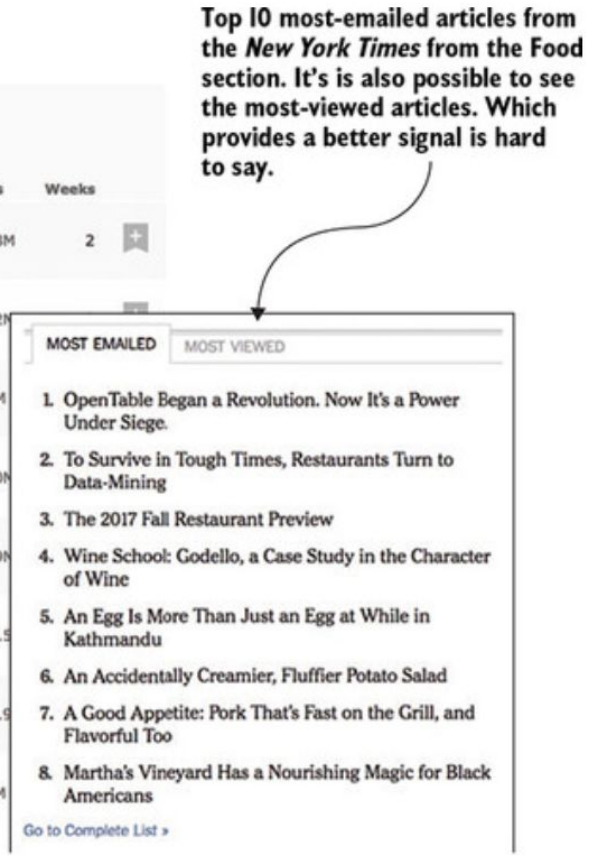


Consider ordering of Content

- Ordering by price is usually a bad idea
- Using recency keeps the website dynamic

Top 10 list


Top Box Office (US)
Weekend of August 25 - 27, 2017
Title
Weekend Gross
Weeks
The Hitman's Bodyguard
$10.3M
$39.8M
2
Annabelle: Creation
$7.7M
Ballerina
$4.7M
$4.7M
Wind River
$4.6M
Logan Lucky
$4.2M
Dunkirk
$3.9M
Spider-Man: Homecoming
$2.8M
Birth of the Dragon
$2.7M
$2.7M
Movies with the biggest box office take during a weekend in 2017 in the US. Don't know why earnings is a good measure of quality, but it often used.


Top 10 most-emailed articles from the New York Times from the Food section. It's is also possible to see the most-viewed articles. Which provides a better signal is hard to say.
MOST EMAILED
MOST VIEWED
1. OpenTable Began a Revolution. Now It's a Power Under Siege.
2. To Survive in Tough Times, Restaurants Turn to Data-Mining
3. The 2017 Fall Restaurant Preview
4. Wine School: Godello, a Case Study in the Character of Wine
5. An Egg Is More Than Just an Egg at While in Kathmandu
6. An Accidentally Creamier, Fluffier Potato Salad
7. A Good Appetite: Pork That's Fast on the Grill, and Flavorful Too
8. Martha's Vineyard Has a Nourishing Magic for Black Americans
Go to Complete List »

Frequently Bought Together

Total price: £43.49

Add all three to Basket

- **This item:** Frostfire Large 2 Person Instant Popup Tent £24.99
- New Set of 2 x 180cm Camping Yoga Roll Eva Foil Foam backed Sleeping Mat Mattress Tent Festival... £8.69
- Yellowstone Essential Mummy Sleeping Bag £9.81

Association Rules

Identify underlying relations between different items

Apriori algorithm is the most simple and straightforward

Support

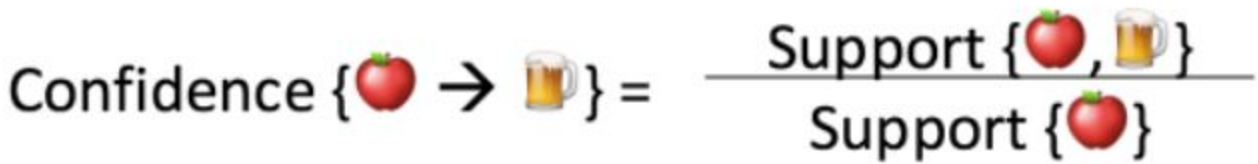
- Fraction of transactions that contain an itemset

Support {🍎} = $\frac{4}{8}$

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

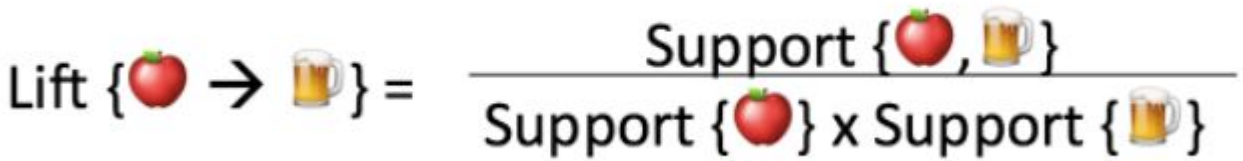
Confidence

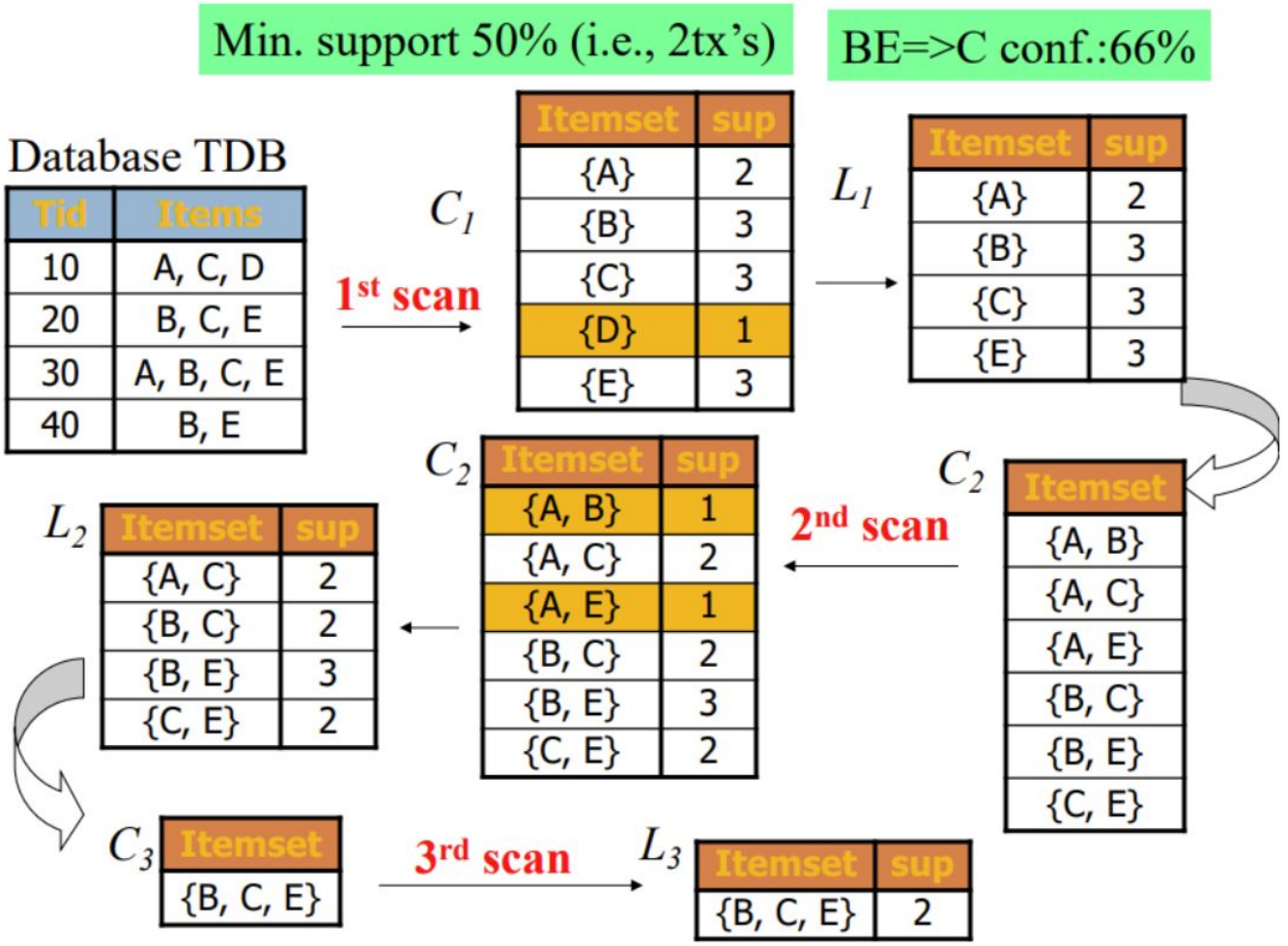
- Measures items in Y appear in transactions that contain X



Lift

- How likely item Y is purchased with item X, while controlling for the popularity of the items. Lift of above 1 is preferred




Min. support 50% (i.e., 2tx's)
BE=>C conf.:66%
Database TDB
Tid | Items
10 | A, C, D
20 | B, C, E
30 | A, B, C, E
40 | B, E
1st scan
C1
Itemset | sup
{A} | 2
{B} | 3
{C} | 3
{D} | 1
{E} | 3
L1
Itemset | sup
{A} | 2
{B} | 3
{C} | 3
{E} | 3
C2
Itemset
{A, B}
{A, C}
{A, E}
{B, C}
{B, E}
{C, E}
2nd scan
C2
Itemset | sup
{A, B} | 1
{A, C} | 2
{A, E} | 1
{B, C} | 2
{B, E} | 3
{C, E} | 2
L2
Itemset | sup
{A, C} | 2
{B, C} | 2
{B, E} | 3
{C, E} | 2
C3
Itemset
{B, C, E}
3rd scan
L3
Itemset | sup
{B, C, E} | 2

Steps for the Apriori Algorithm

- Computing the support for each individual item
- Deciding on the support threshold
- Selecting the frequent items
- Finding the support of the frequent itemsets
- Repeat for larger sets
- Generate Association rules
- Compute confidence and lift
- Store the results in a database

Shortcomings of Apriori

- The size of the itemset from candidate generation can be very large
- Lots of time wasted on counting the support since we have to scan the itemset database over and over again

Python Packages for Market Basket Analysis

- Apriori_python
- Efficient_Apriori
- Reference
  - https://towardsdatascience.com/apriori-association-rule-mining-explanation-and-python-implementation-290b42afdfc6

Summary

- Ordering the content
- Frequently Bought Together - Metrics suitable for your business
- Save the recommendations in a database
- Add versioning to the rules