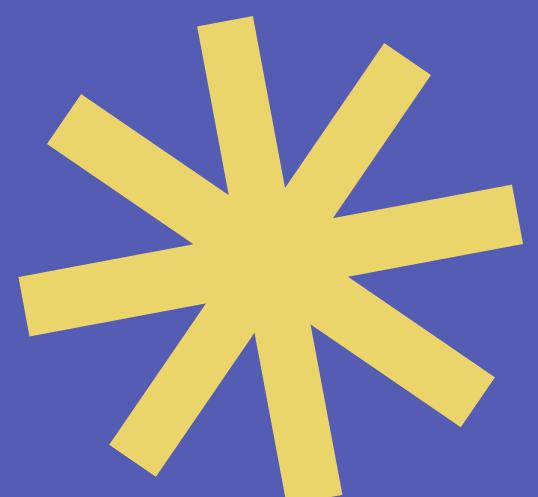


MARKET BASKET SIGHT



Data Mining: Market Basket Analysis with Apriori Algorithm

**Uncovering the secret behind why chocolates are always conveniently placed beside
in groceries**



“People that buy this will most likely want to buy that as well”

- The strategic placement of chocolates and other tempting treats near the checkout counters in grocery stores is not so much a secret as it is a well-established marketing and merchandising technique. It's a practice known as "impulse merchandising" or "impulse buying," and it's designed to capitalize on shoppers' impulsive tendencies. Here's why chocolates are conveniently placed beside the checkout counters in groceries

Data-driven strategies

- Huge retailers pivot on a detailed market basket analysis to uncover associations between items. Using this valuable information, they are able to carry out various strategies to improve their revenue:
- Associated products are placed close to each other, so that buyers of one item would be prompted to buy the other.
- Discounts can be applied to only one of the associated products

Core concepts illustration

- I will be illustrating three of the core concepts that are used in Association Rule Mining with some simple examples below. This will assist you in grasping the data mining process
- Let's say you have now opened up your own cafeteria. How will you utilize your data science skills to understand which of the items on your menu are associated?
- We can utilize three core measures that are used in Association Rule Learning, which are: Support, Confidence, and Lift.



1. Support

- Support is just the plain basic probability of an event to occur. It is measured by the proportion of transactions in which an item set appears. To put it in another way, $\text{Support}(A)$ is the number of transactions which includes A divided by the total number of transactions.

$$\text{Support}(\text{Cookie}) = \frac{3}{6}$$

2. Confidence

- The confidence of a consequent event given an antecedent event can be described by using conditional probability. Simply put, it is the probability of event A happening given that event B has already happened

3. Lift

- Lift is the observed to expected ratio (abbreviation o/e). Lift measures how likely an item is purchased when another item is purchased, while controlling for how popular both items are. It can be calculated by dividing the probability of both of the items occurring together by the product of the probabilities of the both individual items occurring as if there was no association between them.

The Apriori Algorithm Overview

- The Apriori Algorithm is one of the most popular algorithms used in association rule learning over relational databases. It identifies the items in a data set and further extends them to larger and larger item sets.
- However, the Apriori Algorithm only extends if the item sets are frequent, that is the probability of the itemset is beyond a certain predetermined threshold.



Apriori and Association Rules

- The Apriori Algorithm will be used to generate frequent item sets. We will be specifying the minimum support to be 6 out of total transactions. The association rules are generated and we filter for Lift value > 1.5.

Code:

```
frequent_itemsets = apriori(transactions, min_support= 6/len(basket), use_colnames=True,  
max_len = 2)rules = association_rules(frequent_itemsets, metric="lift", min_threshold = 1.5)  
display(rules.head())  
print("Rules identified: ", len(rules))
```

Output:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(UHT-milk)	(butter milk)	0.021386	0.017577	0.000601	0.028125	1.600131	0.000226	1.010854
1	(butter milk)	(UHT-milk)	0.017577	0.021386	0.000601	0.034221	1.600131	0.000226	1.013289
2	(cream cheese)	(UHT-milk)	0.023658	0.021386	0.000869	0.036723	1.717152	0.000363	1.015922
3	(UHT-milk)	(cream cheese)	0.021386	0.023658	0.000869	0.040625	1.717152	0.000363	1.017685
4	(soda)	(artif. sweetener)	0.097106	0.001938	0.000468	0.004818	2.485725	0.000280	1.002893

Visualizations

To visualize our association rules, we can plot them in a 3D scatter plot.
Rules that are closer to top right are the rules that can be the most meaningful to be further dived in

Code:

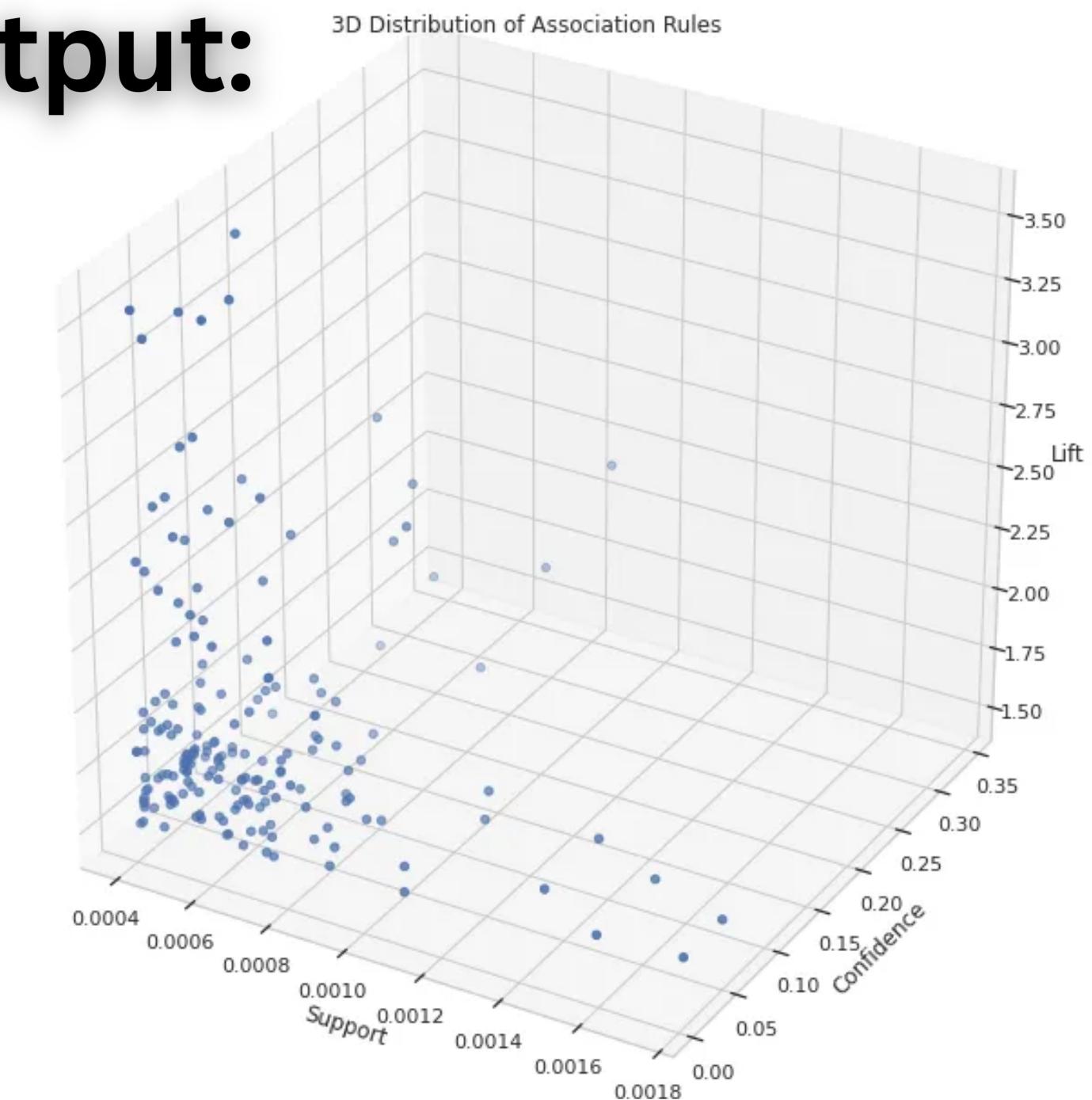
```
sns.set(style = "whitegrid")
fig = plt.figure(figsize=(12, 12))
ax = fig.add_subplot(projection = '3d')

x = rules['support']
y = rules['confidence']
z = rules['lift']

ax.set_xlabel("Support")
ax.set_ylabel("Confidence")
ax.set_zlabel("Lift")

ax.scatter(x, y, z)
ax.set_title("3D Distribution of Association Rules")
plt.show()
```

Output:



Business Application

- Let's say the grocery has bought up too much Whole Milk and is now worrying that the stocks will expire if they cannot be sold out in time. To make matters worse, the profit margin of Whole Milk is so low that they cannot afford to have a promotional discount without killing too much of their profits.



- One approach that can be proposed is to find out which products drive the sales of Whole Milk and offer discounts on those products instead.

Code:

```
milk_rules = rules[rules['consequents'].astype(str).str.contains('whole milk')]  
milk_rules = milk_rules.sort_values(by=['lift'], ascending = [False]).reset_index(drop = True)  
  
display(milk_rules.head())
```

Output:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(brandy)	(whole milk)	0.002540	0.157923	0.000869	0.342105	2.166281	0.000468	1.279957
1	(softener)	(whole milk)	0.002740	0.157923	0.000802	0.292683	1.853328	0.000369	1.190523
2	(canned fruit)	(whole milk)	0.001403	0.157923	0.000401	0.285714	1.809201	0.000179	1.178908
3	(syrup)	(whole milk)	0.001403	0.157923	0.000401	0.285714	1.809201	0.000179	1.178908
4	(artif. sweetener)	(whole milk)	0.001938	0.157923	0.000535	0.275862	1.746815	0.000229	1.162868

- For instance, we can apply a promotional discount on Brandy, Softener, Canned Fruit, Syrup and Artificial Sweetener. Some of the associations may seem counter-intuitive, but the rules state that these products do drive the sales of Whole Milk.

THANK YOU