# **Interview Questions:**

1. What is lift and why is it important in Association rules?
2. What is support and Confidence. How do you calculate them?
3. What are some limitations or challenges of Association rules mining?

1.

In association rule mining, lift is a metric that expresses how much higher the probability is to locate a rule's antecedent and consequent combined than if they were separate. It facilitates the discovery of significant correlations between dataset pieces. This is a thorough explanation:

Definition of Lift: Lift is defined as the ratio of the observed support of the antecedent and consequent appearing together to the support that would be expected if they were independent.

Mathematically, it is represented as:

Lift (𝐴 → 𝐵 ) = 𝑃 ( 𝐴 ∩ 𝐵 ) 𝑃 ( 𝐴 ) ⋅ 𝑃 ( 𝐵 )

Lift(A→B) = P(A)⋅P(B) P(A∩B)​

Where: 𝑃 (𝐴 ∩ 𝐵 ) P(A∩B) is the support of both A and B (i.e., how often A and B occur together).

P(A) and P(B) are the individual supports of A and B, respectively.

Lift's significance in association rules:

Lift is a measure of dependency that shows if a rule's antecedent and consequent occur together more often than would be predicted by chance. Lift = 1 denotes independence (lack of correlation). A positive association (things are more likely to be purchased together) is suggested if Lift > 1, while a negative association (items are less likely to be purchased together) is indicated if Lift < 1.  
  
Rules Filtering: Lift is used to eliminate unnecessary or erroneous rules. Since they show either no relationship or a very weak association between the objects, rules with low lift (near to 1) may not have much relevance.  
  
Strong correlations between goods are indicated by high lift values, which offer valuable insights into the behavior and preferences of customers. This supports strategic decision-making, such as product bundling, cross-selling, and targeted marketing campaigns.

2.ANSWER:

In association rule mining, support and confidence are essential measures that are used to measure the frequency and strength of links between objects in a dataset. Below is a description of each and a computation for it.

### Support:

* **Definition**: Support measures the frequency of occurrence of an itemset in the dataset. It indicates how popular an itemset is among all the transactions.
* **Calculation**:
  + A can be a single item or a set of items (itemset).
  + Support ranges from 0 to 1, where:
    - Support = 0 means the itemset never occurs.
    - Support = 1 means the itemset occurs in every transaction.
  + For a set of items {A, B}

Support(A→B)

=Transactions containing both A and B​/ Total number of transactions

### Confidence:

* **Definition**: Confidence measures the reliability of the rule. It indicates how often items in the consequent (B) appear in transactions that contain items in the antecedent (A).
* **Calculation**:

Confidence(A→B) =Support(A∩B)/Support(A)

* Confidence ranges from 0 to 1, where:
  + Confidence = 1 means that whenever itemset A occurs, itemset B also occurs.
  + Confidence = 0 means that itemset B never occurs when itemset A occurs.

3.

Although association rule mining is an effective method for identifying intriguing connections and patterns in huge datasets, there are a number of drawbacks and difficulties that scholars and practitioners should be aware of:  
  
High Computational Cost: When dealing with huge datasets including a significant number of unique items, the process of generating frequent item sets and association rules can be computationally costly. Algorithms with high memory and processing requirements include FP-Growth and Apriori.  
  
Memory Usage: When working with millions of transactions or high-dimensional data, it can be difficult to store and manipulate huge transaction datasets in memory, which can result in memory limitations and performance concerns.

Choosing the Best Parameters: The quality and applicability of rules that are found can be greatly impacted by the thresholds that are chosen for support, confidence, and other metrics (such lift). Iterative tuning and domain expertise are needed to find the ideal threshold balance between being too harsh and too lenient.  
  
Problems with Sparse Data: Many item sets in transactional data may have little support, which can result in sparse matrices or sparse data structures.

This sparsity can impact the quality of association rules and make it difficult to find significant patterns.  
  
Handling Big Itemsets:

The number of itemsets and association rules that can be created increases exponentially with the number of items in a dataset. This proliferation of rules may result in more computational demands and challenges with interpretation and analysis.

Association rule mining may overlook implicit linkages or hidden associations between items that are not immediately noticed in the dataset, as it primarily concentrates on explicit co-occurrence patterns.  
  
Scalability: It's still difficult to scale association rule mining methods to function in huge data settings, distributed computing frameworks, or real-time data streams. For algorithms to effectively handle large-scale datasets, they must be modified or optimized.  
  
Interpreting the Results:

Although association rules shed light on item co-occurrences, it is frequently necessary to have domain expertise as well as a contextual awareness of the data and business context in order to evaluate and validate the practical meaning of identified rules.

**Mitigating Challenges:**

**Advanced Algorithms:**

To increase scalability and performance, research is being done on more effective algorithms (such as parallel algorithms and memory-efficient data structures).  
  
Domain Expertise:

Domain expertise aids in the interpretation of findings, the establishing of meaningful thresholds, and the removal of unnecessary regulations.  
  
Data Preprocessing: Noise and sparsity problems can be reduced with the use of efficient feature selection, data cleaning, and dimensionality reduction algorithms.  
  
Post-Processing: Methods like pattern visualization, rule aggregation, and pruning rules help refine and explain rules that have been found.

**INTERPRETATION AND ANALYSIS:**

SETTING THRESHOLD VALUES

#SUPPORT = 0.01

#CONFIDENCE=0.2

#LIFT=0.3

from apyori import apriori

rules = apriori(transactions = transactions2,

min\_support = 0.01,

min\_confidence = 0.2,

min\_lift = 3,

min\_length = 2, max\_length = 3)

results[0]

RelationRecord

(items=frozenset({'ground beef', 'herb & pepper'}),

support=0.015455950540958269,

ordered\_statistics=

[OrderedStatistic(items\_base=frozenset({'herb & pepper'),

items\_add=frozenset({'ground beef'}),

confidence=0.3292181069958848,

lift=3.179165898900559)])

results[1]

RelationRecord

(items=frozenset({'frozen vegetables', 'ground beef', 'spaghetti'}),

support=0.010046367851622875,

ordered\_statistics=

[OrderedStatistic

(items\_base=frozenset({'frozen vegetables', 'spaghetti'}),

items\_add=frozenset({'ground beef'}),

confidence=0.311377245508982,

lift=3.006881758870319)])

results[2]

RelationRecord

(items=frozenset ({'nan', 'ground beef', 'herb & pepper'}),

support=0.015455950540958269,

ordered\_statistics=

[OrderedStatistic (items\_base=frozenset({'herb & pepper'), items add=frozenset({'nan', 'ground beef'}),

confidence=0.3292181069958848,

lift=3.179165898900559),

OrderedStatistic

(items\_base=frozenset ({'nan', 'herb & pepper'}),

items add=frozenset({'ground beef'}),

confidence=0.3292181069958848,

lift=3.179165898900559)])

results [3]

RelationRecord

(items=frozenset ({'soup', 'mineral water', 'milk'}),

support=0.010046367851622875,

ordered\_statistics=

[OrderedStatistic

(items\_base=frozenset ({'soup', 'mineral water'}),

items add=frozenset({'milk'}),

confidence=0.4126984126984127, lift=3.1834977408747895)])

ANOTHER SET OF THRESHOLD VALUES:

from apyori import apriori

rules = apriori(transactions = transactions1,

min\_support = 0.003,

min\_confidence = 0.2,

min\_lift = 3,

min\_length = 2, max\_length = 2)

[RelationRecord(items=frozenset({'burgers', 'almonds'}), support=0.005795981452859351, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'almonds'}), items\_add=frozenset({'burgers'}), confidence=0.28571428571428575, lift=3.250235478806908)]),

RelationRecord(items=frozenset({'ground beef', 'blueberries'}), support=0.0030911901081916537, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'blueberries'}), items\_add=frozenset({'ground beef'}), confidence=0.3137254901960784, lift=3.029558091893474)]),

RelationRecord(items=frozenset({'chicken', 'light cream'}), support=0.004443585780525503, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'light cream'}), items\_add=frozenset({'chicken'}), confidence=0.2804878048780488, lift=4.508710801393729)]),

RelationRecord(items=frozenset({'cooking oil', 'light cream'}), support=0.0034775888717156105, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'light cream'}), items\_add=frozenset({'cooking oil'}), confidence=0.21951219512195122, lift=3.8127353085611393)]),

RelationRecord(items=frozenset({'escalope', 'mushroom cream sauce'}), support=0.0056027820710973725, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'mushroom cream sauce'}), items\_add=frozenset({'escalope'}), confidence=0.29591836734693877, lift=3.638179262203694)]),

RelationRecord(items=frozenset({'escalope', 'pasta'}), support=0.004829984544049459, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'pasta'}), items\_add=frozenset({'escalope'}), confidence=0.3164556962025316, lift=3.8906762079437143)]),

RelationRecord(items=frozenset({'olive oil', 'extra dark chocolate'}), support=0.0030911901081916537, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'extra dark chocolate'}), items\_add=frozenset({'olive oil'}), confidence=0.25396825396825395, lift=3.641384162159785)]),

RelationRecord(items=frozenset({'fromage blanc', 'honey'}), support=0.0032843894899536323, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'fromage blanc'}), items\_add=frozenset({'honey'}), confidence=0.2537313432835821, lift=5.031852233087436)]),

RelationRecord(items=frozenset({'parmesan cheese', 'frozen vegetables'}), support=0.005795981452859351, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'parmesan cheese'}), items\_add=frozenset({'frozen vegetables'}), confidence=0.3614457831325301, lift=3.7342183103672175)]),

RelationRecord(items=frozenset({'ground beef', 'herb & pepper'}), support=0.015455950540958269, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'herb & pepper'}), items\_add=frozenset({'ground beef'}), confidence=0.3292181069958848, lift=3.179165898900559)]),

RelationRecord(items=frozenset({'tomato sauce', 'ground beef'}), support=0.004829984544049459, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'tomato sauce'}), items\_add=frozenset({'ground beef'}), confidence=0.3246753246753247, lift=3.135297538282613)]),

RelationRecord(items=frozenset({'olive oil', 'light cream'}), support=0.0034775888717156105, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'light cream'}), items\_add=frozenset({'olive oil'}), confidence=0.21951219512195122, lift=3.1473549084521313)]),

RelationRecord(items=frozenset({'olive oil', 'whole wheat pasta'}), support=0.0071483771251932, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'whole wheat pasta'}), items\_add=frozenset({'olive oil'}), confidence=0.2781954887218045, lift=3.9887530460500273)]),

RelationRecord(items=frozenset({'shrimp', 'pasta'}), support=0.004829984544049459, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'pasta'}), items\_add=frozenset({'shrimp'}), confidence=0.3164556962025316, lift=4.3563156477242115)])]

The above are the obtained rules.

| **baseitem** | **additem** | **support** | **cofidence** | **lift** |
| --- | --- | --- | --- | --- |
| **0** | almonds | burgers | 0.005796 | 0.285714 | 3.250235 |
| **1** | blueberries | ground beef | 0.003091 | 0.313725 | 3.029558 |
| **2** | light cream | chicken | 0.004444 | 0.280488 | 4.508711 |
| **3** | light cream | cooking oil | 0.003478 | 0.219512 | 3.812735 |
| **4** | mushroom cream sauce | escalope | 0.005603 | 0.295918 | 3.638179 |
| **5** | pasta | escalope | 0.004830 | 0.316456 | 3.890676 |
| **6** | extra dark chocolate | olive oil | 0.003091 | 0.253968 | 3.641384 |
| **7** | fromage blanc | honey | 0.003284 | 0.253731 | 5.031852 |
| **8** | parmesan cheese | frozen vegetables | 0.005796 | 0.361446 | 3.734218 |
| **9** | herb & pepper | ground beef | 0.015456 | 0.329218 | 3.179166 |
| **10** | tomato sauce | ground beef | 0.004830 | 0.324675 | 3.135298 |
| **11** | light cream | olive oil | 0.003478 | 0.219512 | 3.147355 |
| **12** | whole wheat pasta | olive oil | 0.007148 | 0.278195 | 3.988753 |
| **13** | pasta | shrimp | 0.004830 | 0.316456 | 4.356316 |
| **14** | almonds | burgers | 0.005796 | 0.285714 | 3.250235 |
| **15** | blueberries | ground beef | 0.003091 | 0.313725 | 3.029558 |
| **16** | light cream | chicken | 0.004444 | 0.280488 | 4.508711 |
| **17** | light cream | cooking oil | 0.003478 | 0.219512 | 3.812735 |
| **18** | mushroom cream sauce | escalope | 0.005603 | 0.295918 | 3.638179 |
| **19** | pasta | escalope | 0.004830 | 0.316456 | 3.890676 |
| **20** | extra dark chocolate | olive oil | 0.003091 | 0.253968 | 3.641384 |
| **21** | fromage blanc | honey | 0.003284 | 0.253731 | 5.031852 |
| **22** | parmesan cheese | frozen vegetables | 0.005796 | 0.361446 | 3.734218 |
| **23** | herb & pepper | ground beef | 0.015456 | 0.329218 | 3.179166 |
| **24** | tomato sauce | ground beef | 0.004830 | 0.324675 | 3.135298 |
| **25** | light cream | olive oil | 0.003478 | 0.219512 | 3.147355 |
| **26** | whole wheat pasta | olive oil | 0.007148 | 0.278195 | 3.988753 |
| **27** | pasta | shrimp | 0.004830 | 0.316456 | 4.356316 |
| **28** | almonds | burgers | 0.005796 | 0.285714 | 3.250235 |
| **29** | blueberries | ground beef | 0.003091 | 0.313725 | 3.029558 |
| **30** | light cream | chicken | 0.004444 | 0.280488 | 4.508711 |
| **31** | light cream | cooking oil | 0.003478 | 0.219512 | 3.812735 |
| **32** | mushroom cream sauce | escalope | 0.005603 | 0.295918 | 3.638179 |
| **33** | pasta | escalope | 0.004830 | 0.316456 | 3.890676 |
| **34** | extra dark chocolate | olive oil | 0.003091 | 0.253968 | 3.641384 |
| **35** | fromage blanc | honey | 0.003284 | 0.253731 | 5.031852 |
| **36** | parmesan cheese | frozen vegetables | 0.005796 | 0.361446 | 3.734218 |
| **37** | herb & pepper | ground beef | 0.015456 | 0.329218 | 3.179166 |
| **38** | tomato sauce | ground beef | 0.004830 | 0.324675 | 3.135298 |
| **39** | light cream | olive oil | 0.003478 | 0.219512 | 3.147355 |
| **40** | whole wheat pasta | olive oil | 0.007148 | 0.278195 | 3.988753 |
| **41** | pasta | shrimp | 0.004830 | 0.316456 | 4.356316 |
| **42** | almonds | burgers | 0.005796 | 0.285714 | 3.250235 |
| **43** | blueberries | ground beef | 0.003091 | 0.313725 | 3.029558 |
| **44** | light cream | chicken | 0.004444 | 0.280488 | 4.508711 |
| **45** | light cream | cooking oil | 0.003478 | 0.219512 | 3.812735 |
| **46** | mushroom cream sauce | escalope | 0.005603 | 0.295918 | 3.638179 |
| **47** | pasta | escalope | 0.004830 | 0.316456 | 3.890676 |
| **48** | extra dark chocolate | olive oil | 0.003091 | 0.253968 | 3.641384 |
| **49** | fromage blanc | honey | 0.003284 | 0.253731 | 5.031852 |
| **50** | parmesan cheese | frozen vegetables | 0.005796 | 0.361446 | 3.734218 |
| **51** | herb & pepper | ground beef | 0.015456 | 0.329218 | 3.179166 |
| **52** | tomato sauce | ground beef | 0.004830 | 0.324675 | 3.135298 |
| **53** | light cream | olive oil | 0.003478 | 0.219512 | 3.147355 |
| **54** | whole wheat pasta | olive oil | 0.007148 | 0.278195 | 3.988753 |
| **55** | pasta | shrimp | 0.004830 | 0.316456 | 4.356316 |

#highest support value

{'almonds'}

In the same way we can identify the relationships with different products

(explained in the code file.)