Association Rules Report

1. Objective

The purpose of this task is to apply **association rule mining** to synthetic market basket data. Using the **Apriori algorithm**, the goal is to identify frequently purchased item combinations and derive rules that can help guide business decisions such as product placement, cross-selling, and promotions.

2. Methodology

Step 1: Synthetic Data Generation

- A pool of 20 grocery-related items was defined (e.g., *milk*, *bread*, *butter*, *beer*, *diapers*, etc.).
- 30 random transactions (baskets) were generated, each containing between 3 and 8 randomly selected items.
- Business logic patterns were embedded to simulate realistic shopping behavior:
    - If **milk** is in a basket but **bread** is missing, add **bread**.
    - If **beer** is in a basket but **diapers** is missing, add **diapers**.
- This creates a more realistic dataset where certain items often appear together.

```
Sample Transactions:
['beer', 'milk', 'chicken', 'bananas', 'onions', 'butter', 'yogurt', 'bread', 'diapers']
['tomatoes', 'butter', 'coffee', 'bread', 'milk', 'onions', 'beer', 'fish', 'diapers']
['fish', 'milk', 'tomatoes', 'apples', 'coffee', 'beer', 'bananas', 'bread', 'diapers']
['chicken', 'milk', 'eggs', 'coffee', 'cheese', 'diapers', 'butter', 'bread']
['cheese', 'beer', 'butter', 'cereal', 'diapers']
```

Step 2: Transaction Encoding

- Data was transformed into a binary matrix format using TransactionEncoder:
    - Each row = a transaction.
    - Each column = an item.
    - Value = True if the item is present, False otherwise.
- This format is required by the Apriori algorithm.

Step 3: Frequent Itemset Mining (Apriori Algorithm)

- Applied Apriori with a minimum support threshold of 0.2 (items must appear in at least 20% of all transactions).
- The algorithm outputs frequent itemsets that meet the support threshold.

Step 4: Association Rule Generation

- Used association_rules() to generate rules from frequent itemsets:
  - Metric: Confidence (minimum 0.5).
  - Lift was calculated to measure the strength of the relationship beyond random chance.
- Rules were sorted by lift to identify the strongest relationships.

Step 5: Rule Analysis

- The top 5 rules by lift were saved as top5_rules.csv for quick review.
- Example Rule Analysis:
- If a customer buys {milk}, they are likely to also buy {bread}.
- Support: 0.50, Confidence: 0.85, Lift: 1.75.
- Business Implication: Place milk and bread together to encourage cross-sales.
- All rules were saved in association_rules.csv for full exploration.

3. Output Files

- association_rules.csv → All generated rules.

antecedents	consequents	antecedent support	consequent support	support	confidence
frozenset({'milk'})	frozenset({'diapers', 'bread'})	0.2	0.26666666666666666	0.2	1.0
frozenset({'diapers', 'bread'})	frozenset({'milk'})	0.26666666666666666	0.2	0.2	0.75
frozenset({'bread'})	frozenset({'diapers', 'milk'})	0.3333333333333333	0.2	0.2	0.6000000000000001
frozenset({'bread'})	frozenset({'milk'})	0.3333333333333333	0.2	0.2	0.6000000000000001
frozenset({'diapers', 'milk'})	frozenset({'bread'})	0.2	0.3333333333333333	0.2	1.0
frozenset({'milk'})	frozenset({'bread'})	0.2	0.3333333333333333	0.2	1.0

- top5_rules.csv → Top 5 rules ranked by lift.

1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage
2	frozenset({'milk'})	frozenset({'diapers', 'bread'})	0.2	0.26666666666666666	0.2	1.0	3.75	1.0	0.1466666666666
3	frozenset({'diapers', 'bread'})	frozenset({'milk'})	0.26666666666666666	0.2	0.2	0.75	3.75	1.0	0.1466666666666
4	frozenset({'bread'})	frozenset({'diapers', 'milk'})	0.3333333333333333	0.2	0.2	0.6000000000000001	3.0000000000000004	1.0	0.1333333333333
5	frozenset({'bread'})	frozenset({'milk'})	0.3333333333333333	0.2	0.2	0.6000000000000001	3.0000000000000004	1.0	0.1333333333333
6	frozenset({'diapers', 'milk'})	frozenset({'bread'})	0.2	0.3333333333333333	0.2	1.0	3.0	1.0	0.1333333333333

4. Business Insights

- Certain products (like *milk* & *bread*, *beer* & *diapers*) showed high lift values, meaning customers who buy one are much more likely than average to buy the other.
- These insights can be used to:
  - Create bundle discounts.
  - Optimize product placement in a store.
  - Improve recommendation systems in e-commerce.

5. Limitations

- Dataset is synthetic and limited to 30 transactions.
- Patterns may not fully represent real-world complexity.
- Increasing dataset size would produce more reliable results.

6. Conclusion

This analysis demonstrates the power of association rule mining in uncovering hidden patterns in transaction data. Even in a small, simulated dataset, strong product relationships emerged, highlighting potential strategies for boosting sales and customer satisfaction.