Data Source: https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset?phase=FinishSSORegistration&returnUrl=%2Fdatasets%2Fheeraldedhia%2Fgroceries-dataset%2Fversions%2F1%3Fresource%3Ddownload&SSORegistrationToken=CfDJ8OdV9jvuBrlOizyz5DdMpkz_QD2LunzIBNiJQVv-fd5W5s4slWe8P9l-uxgPL2KArC0rlIeSbKCx6K8nLZXdtVXu8xHKe8lEA7L3-cdJZaYT4hINFnWHJi6fJFq-A5oGcVcdSBSR52GZKGKPYg0I4ahnHkoVvChLHOZ05hCSFwqZ1Jsshk9cDZN51EvMhHDgT_B63PWUszPoQHlysmNNfteQiNUmKv8OH4Y3qOYAVlVtaF9QK5FRJRq1kymIq0sb4nCypcHeQwrDtulIHBgy4eE59sbh66m_x8C8cwkCUTsDYDYC4KU-6hWQNDI2xpV5P_TeH8DWHPCEa3yiWQM5WFF9dA&DisplayName=Lucas

The Apriori algorithm is a classic data mining algorithm used to identify frequent itemsets in a dataset and derive association rules. It's commonly applied in market basket analysis to discover which items are frequently purchased together.

Key Concepts:

1. **Itemset**: A collection of one or more items.
2. **Support**: The proportion of transactions that contain a particular itemset. This is calculated as:

   Support(X) = (Count of transactions containing X) / (Total number of transactions)
3. **Confidence**: The probability that item B is bought when item A is bought. This is calculated as:

   Confidence($A \rightarrow B$) = Support($A \cup B$) / Support(A)
4. **Lift**: The ratio of observed support to expected support if A and B were independent. This is calculated as:

   Lift($A \rightarrow B$) = Confidence($A \rightarrow B$) / Support(B)

Steps of the Apriori Algorithm:

1. **Generate Candidate Itemsets**:
   - Start with single items (1-itemsets).
   - For each itemset, calculate its support and discard itemsets that don't meet the minimum support threshold. These are considered **frequent itemsets**.
2. **Iterative Process**:
   - Combine the frequent itemsets from the previous step to form larger itemsets (e.g., 2-itemsets, 3-itemsets, etc.).
   - For each candidate itemset, calculate its support. If the support is above the threshold, it is added to the frequent itemsets. Otherwise, it is discarded.
3. **Generate Association Rules**:
   - Once frequent itemsets are identified, generate **association rules**. An association rule is of the form $A \rightarrow B$, meaning if A is bought, B is likely to be bought as well.

- The rule is considered valid if its confidence is greater than or equal to a specified threshold.