**Project Title:**

Market-Basket Analysis

**Project Details:**

The task here is to find the association rules between the items using the Apriori Algorithm. Or in other words, you have to find out those items/itemset that customers bought together which helps the owner for store layout/marketing.  
And then show the relative Sales of the data in Tableau Dashboard.

**Tools Used :**

Google colab ,Iris dataset ,Machine Learning Algorithms.

**Algorithm used:**

* **MARKET Basket Analysis (MBA)** is an association analysis and is a popular data mining technique. It’s a kind of knowledge discovery in data (KDD) and this technique can be applied in various fields of work.
* In market basket analysis (also called **association analysis** or **frequent itemset mining**), you analyze purchases that commonly happen together. For example, people who buy bread and peanut butter also buy jelly. Or people who buy shampoo might also buy conditioner. What relationships there are between items is the target of the analysis. Knowing what your customers tend to buy together can help with marketing efforts and store/website layout.
* Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy.
* The discovery of these associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. The strategies may include:
  1. Changing the store layout according to trends
  2. Customer behavior analysis
  3. Catalog design
  4. Cross marketing on online stores
  5. What are the trending items customers buy
  6. Customized emails with add-on sales etc..
* **Apriori algorithm** is given by R. Agrawal and R. Srikant in 1994 for finding frequent itemsets in a dataset for boolean association rule. Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties.
* Apriori algorithm assumes that any subset of a frequent itemset must be frequent.The value of “frequent itemset” > than a threshold value(i.e. support). Its the algorithm behind Market Basket Analysis.
* Apriori assumes that: **All subsets of a frequent itemset must be frequent (Apriori propertry) .If an itemset is infrequent, all its supersets will be infrequent.** Say, a transaction containing {Grapes, Apple, Mango} also contains {Grapes, Mango}. So, according to the principle of Apriori, if {Grapes, Apple, Mango} is frequent, then {Grapes, Mango} must also be frequent.

The analysis of the association rules depend on five measures (but only first three are widely used)

1. Support: Support of the item x is nothing but the ratio of the number of transactions in which the item x appears to the total number of transactions.
   * **Support(Item A) = (Transactions containing Item A) / (Total transactions), range:[0,1]**
2. Confidence: Confidence refers to the likelihood that an item B is also bought if item A is bought. It can be calculated by finding the number of transactions where A and B are bought together, divided by total number of transactions where A is bought. A confidence of 0.5 would mean that in 50% of the cases where A were purchased, the purchase also included B. For product recommendation, a 50% confidence may be perfectly acceptable but in a medical situation, this level may not be high enough.
   * **Confidence(A → B) = (Transactions containing both (A and B)) / (Transactions containing A), range:[0,1]**
3. Lift: Lift(A → B) refers to the increase in the ratio of sale of B when A is sold. Lift (A → B) is nothing but the ‘interestingness’ or the likelihood of the item B being purchased when the item A is sold. A Lift of 1 means there is no association between products A and B. Lift of greater than 1 means products A and B are more likely to be bought together. Finally, Lift of less than 1 refers to the case where two products are unlikely to be bought together.
   * **Lift(A → B) = (Confidence (A → B)) / (Support (B)), range:[0,∞]**
4. Leverage: Leverage computes the difference between the observed frequency of A and C appearing together and the frequency that would be expected if A and C were independent. An leverage value of 0 indicates independence.
   * **levarage(A → C) = support(A → C) − support(A)×support(C), range:[−1,1]**
5. Conviction: A high conviction value means that the consequent is highly depending on the antecedent. For instance, in the case of a perfect confidence score, the denominator becomes 0 (due to 1 - 1) for which the conviction score is defined as 'inf'.Similar to lift, if items are independent, the conviction is 1.
   * **conviction(A → C) = (1−support(C)) / (1−confidence(A → C)), range:[0,∞]**