

# GROUP E PROJECT REPORT

TITLE: MARKET BASKET ANALYSIS USING ASSOCIATION RULE MINING AND APRIORI THEOREM.

This project aims to analyze the concept of market basket analysis and its features such as association rules mining and apriori theorem and to perform market basket analysis on a sample data set. Market basket analysis is a crucial part of modern-day sales. It is an imperative tool that provides a clear edge to any retailer who possesses it in his armory. Market basket analysis is a real-life problem that be learnt to be solved with the help of this project. While, the project does not go in-depth into the concepts, it gives a clear understanding of the basics and sufficient knowledge for rudimentary real-life applications.

**Market Basket Analysis** determines what items are frequently bought together or placed in the shopping cart together by the customers. It establishes any possible relationship between different products by looking for combinations of products that frequently occur together. Association rules and apriori theorem are some of the tools that help facilitate in the process of obtaining results. **Association rules** defines specifics terms for each pair of products that help classify their relationship in a well-defined manner. In our project, we have discussed about **Antecedent** (This is an item/group of items that are typically found/occur in the datasets.), **Consequent** (items likely to occur if the antecedents occur.), **Support** (it gives the percentage of the data which contains both items A and B.), **Confidence** (It tells us the probability of B to occur provided that A has already occurred.) **Lift** (It indicates the strength of an association rule over the random occurrence of A and B.). **Apriori Theorem** is the most famous algorithm used to perform MBA. It helps in the formation of association values discussed above using 2 steps: **Join Step**: This step is used to generate  $(K+1)$  item set from a K-item set by joining each item with itself and **Prune Step**: This step is used to scan the total count of each of the items in the database/item set. If the item under evaluation does not meet the minimum support threshold required, then it is regarded as infrequent and thus is removed. These were the basic terms that were analyzed during the course of the project and that are mandatory prerequisites for our analysis of the sample data set.

The sample set is our hands is a famous Kaggle data set with 158 unique items comprising of 9835 records were possible associations could be made. Defining of support and confidence are mandatory in any real-life problems which are often very large in size, in order to reduce time and money spent. A **support of 3%** was decided to neglect items that occurred sparsely from our analysis. A **confidence of 20%** was chosen, such that only items with higher probability of simultaneous occurrence are considered.

With the help of python programming language, the dataset was analyzed and an output table was formatted containing the various association term values for 2 different products provided the support and confidence criterion is followed, using the inbuilt apriori tools.

Some results inferred from our analysis: There was a **total of 25 associations** that we were able to deduce with reasonable confidence. Whole milk, other vegetables, rolls/buns, root, vegetables, yogurt, soda were the consequent items which were obtained.

However, it has to be understood that MBA is a **cause and effect study** and not a correlation study, meaning that the results are not universal. It is an intuitive study that predicts exclusive to the **given conditions** only.

Market basket analysis is booming field with bright prospects, with the advent of large-scale machine learning and cheap large-scale data analysis. The retailer can use the inferences from MBA for **product placements, online recommendations, cross-selling, loss leader analysis, etc.** to further their business prospects, while also catering to the needs of the customer better. And this project hopes to help facilitate the same.