**FUNDAMENTALS OF DATA SCIENCE**

**PROJECT PROPOSAL**

**Title: Market Basket Analysis**

Name: Dasaradh Sravanam

UUID: dsrvanam

U-number: U00830073

****

**Introduction:**

In a mundane way most of the customers tend to buy a selected set products together, the retailer in initial days identified the patterns and placed the product together, as a result the sales for products increased, proving their analysis of prediction.

This project is to perform a market analysis to understand a technique to know the associations between items at a market place. This is to seeks out the purchasing trends to gain useful customer information by determining which items are frequently bought together so that it helps to identify correlation and connection between items so that retailers can maximize their cross-selling.

**Vision:**

To understand the correlation between the items.

To predict the purchase behavior of the customer.

This helps in increase of business if the correlated items placed at same spot that makes customers to buy as if this item, then that item.

**Skillset:**

I have worked at TCS as data analyst and have experience in handling data. With this project I would like to extend my skills to Python and Data science related technologies.

**Rationales:**

I would like to attempt to do this project as it can be implemented real time, this project will help me in understanding Machine learning and data science concepts. I intend to use Apriori algorithm in this project. Apriori algorithm assumes that any subset of a frequent item set must be frequent.

Market basket analysis is a key technique used by large retailers like Amazon to analyze customer buying habits by finding associations between the different items that customers place in their shopping baskets. This project gives more insight for retailers such as Customer behavior analysis, Catalog design and Customized emails with add-on sales.

**Project Approach:**

The main objective of this project is to effectively predict the products that are likely to be brought together. Initially the dataset is loaded and then cleaned to eliminate null, duplicate (redundant). The data is further split into train and test data. The training data is data that will train the algorithms to create better model and will do performance measure and model tuning by applying this model to achieve good accuracy .