MARKET BASKET INSIGHTS: UNVEILING CUSTOMER BEHAVIOR THROUGH MARKET BASKET ANALYSIS

Phase 1: Problem Definition and Design Thinking Document

-Palli Muhammed Suhaib
B.E CSE 3rd yr. 5th SEM
311421104060
Meenakshi College of Engineering - 3114

PROBLEM DEFINITION:

- **Objective**: Conduct market basket analysis on a provided dataset.
- **Goal**: Uncover concealed patterns and relationships within product purchases.
- **Ultimate Aim**: Understand customer buying patterns and pinpoint cross-selling prospects for a real business.
- **Methodology**: Employ association techniques, notably the Apriori algorithm.
- **Outcome**: Identify frequently co-purchased items and formulate practical rules for the business's advantage.

DESIGN THINKING:

1. **Data Source Selection:**
    a. We will initiate the project by meticulously choosing a suitable dataset.
    b. The dataset should encompass transaction data, specifically lists of purchased products.
    c. Our criteria for selection will prioritize datasets that closely mirror the actual transactions of the business.
    d. This selection process aims to maintain the utmost relevance in our analysis.
2. **Data Preprocessing:** Data preparation is a pivotal phase in this analysis, involving the conversion of raw transaction data into an apt format for market basket analysis. This step encompasses several key tasks:
    a. Data Cleaning: Addressing inconsistencies and eliminating missing values within the dataset.
    b. Data Organization: Structuring the data into transaction lists or baskets, grouping items bought together in each transaction.

c. Data Encoding: Transforming the data into a binary matrix, with rows representing transactions and columns indicating products. The matrix utilizes binary values (1s and 0s) to signify whether a product was purchased or not.

3. **Association Analysis Using Apriori Algorithm:** We will utilize the Apriori algorithm on the preprocessed data to uncover frequent item sets and create association rules. The process involves the following steps:

    a. Minimum Support Threshold: Set a minimum support threshold to eliminate infrequent items from consideration.

    b. Frequent Item set Generation: Iteratively expand item set size to discover frequent combinations of items.

    c. Association Rule Derivation: Create association rules based on support and confidence metrics to identify relationships between items.

    d. Pruning of Irrelevant Rules: Remove irrelevant or redundant rules to ensure clarity and relevance in the generated insights

4. **Insights Generation:** After obtaining association rules, our focus shifts to extracting valuable insights regarding customer behavior and cross-selling potential. The insights generation phase involves a series of steps for a deeper understanding:

    a. Interpretation of Association Rules: Analyze the generated association rules to uncover meaningful patterns.

    b. Customer Behavior Understanding: Gain insights into how customers make purchase decisions and identify their preferences.

    c. Cross-Selling Opportunities: Discover opportunities to recommend complementary products based on purchase patterns.

    d. Data Visualization: Use visualizations to represent the discovered associations and insights effectively.

    e. Business Optimization: Leverage these insights to optimize business strategies and enhance customer experiences.