

KHULNA UNIVERSITY

190205 || 200240

Groceries Dataset Analysis Report

Submitted to:

Dr.Amit Kumar Mondal

Associate Professor

Computer Science And Engineering Discipline, Khulna University

Submitted By:

Sumaiah Binta Musa, istyaque Ahammed

Student ID:190205.200240

Computer Science And Engineering Discipline,Khulna University

# 1. Domain

The domain of this dataset is Retail and Grocery Transactions. It consists of transaction records from a grocery store where each row contains items purchased together by customers. This dataset can be used for market basket analysis, customer behaviour insights, and more.

# 2. Number of Samples

The dataset contains 317 samples. Each sample represents a unique transaction in which one or more items were purchased.

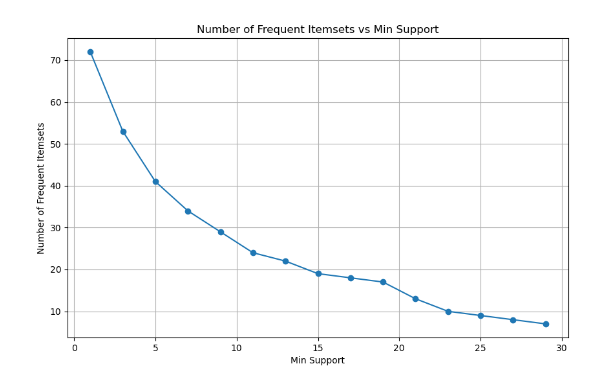
# 3. Number of Unique Categories (Items)

The dataset contains 7 unique grocery items. These items represent the different products available for purchase in the grocery store.

# 4. Dataset Size

The dataset is 6 KB in size, making it relatively small and easy to work with in terms of both data storage and processing requirements.

Graph:



Optimal min\_support:

The **optimal minimum support** is likely between **5 and 10**. This range balances capturing useful frequent itemsets without including too many rare or irrelevant ones. A **min support of 5 or 6** is a smart choice, providing a good number of patterns while avoiding overfitting.

The range **5-10** is optimal because:

In this range, you still get a good number of frequent itemsets (about 30-40), which is enough to discover meaningful patterns without overwhelming noise from rare combinations.

After this point (min support > 10), the number of frequent itemsets decreases slowly, meaning you start losing important patterns without gaining much in precision.

Too low a support (e.g., < 5) captures rare or irrelevant combinations, while too high a support misses out on useful but less frequent patterns.

Thus, **5-6** offers a good trade-off between finding valuable insights and keeping the model interpretable.