| Experiment No. 7 |
| --- |
| Implement frequent pattern mining algorithm(Apriori) |
| Date of Performance:22/09/2025 |
| Date of Submission: |

**Aim:** To implement Apriori algorithm

**Objective:** Develop a program to implement Apriori Algorithm on the given dataset

**Theory:**

Apriori is an algorithm for frequent item set mining and [association rule learning](https://en.wikipedia.org/wiki/Association_rule_learning) over transactional [databases](https://en.wikipedia.org/wiki/Databases). It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine [association rules](https://en.wikipedia.org/wiki/Association_rules) which highlight general trends in the [database](https://en.wikipedia.org/wiki/Database): this has applications in domains such as [market basket analysis](https://en.wikipedia.org/wiki/Market_basket_analysis).

**Procedure or algorithm description:**

Level-wise algorithm:

a. Let k = 1

b. Generate frequent itemsets of length 1

c. Repeat until no new frequent itemsets are identified

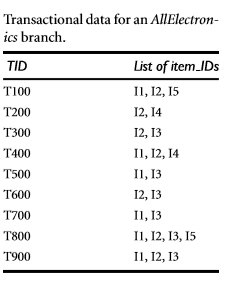
1. Generate length (k+1) candidate itemsets from length k frequent itemsets

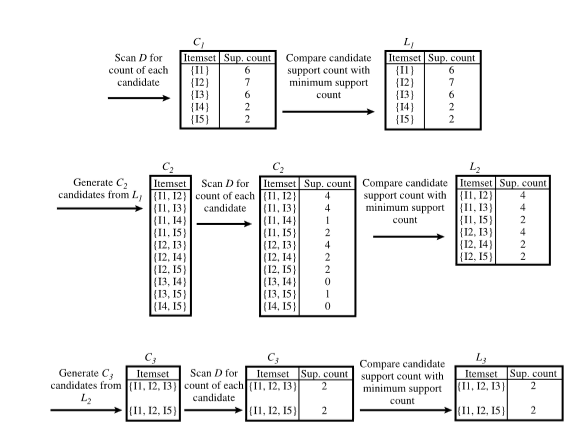
2. Prune candidate itemsets containing subsets of length k that are infrequent

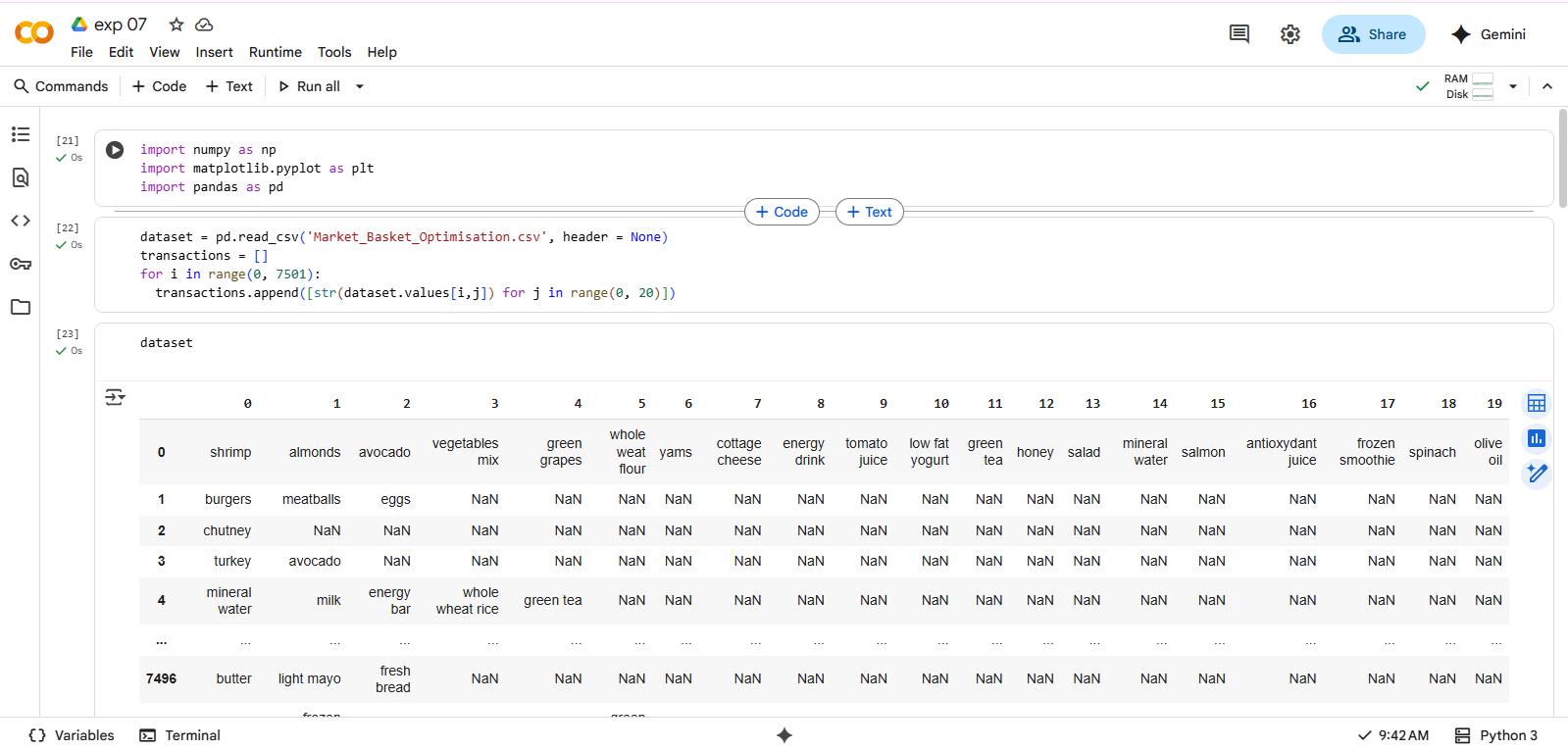
3. Count the support of each candidate by scanning the DB

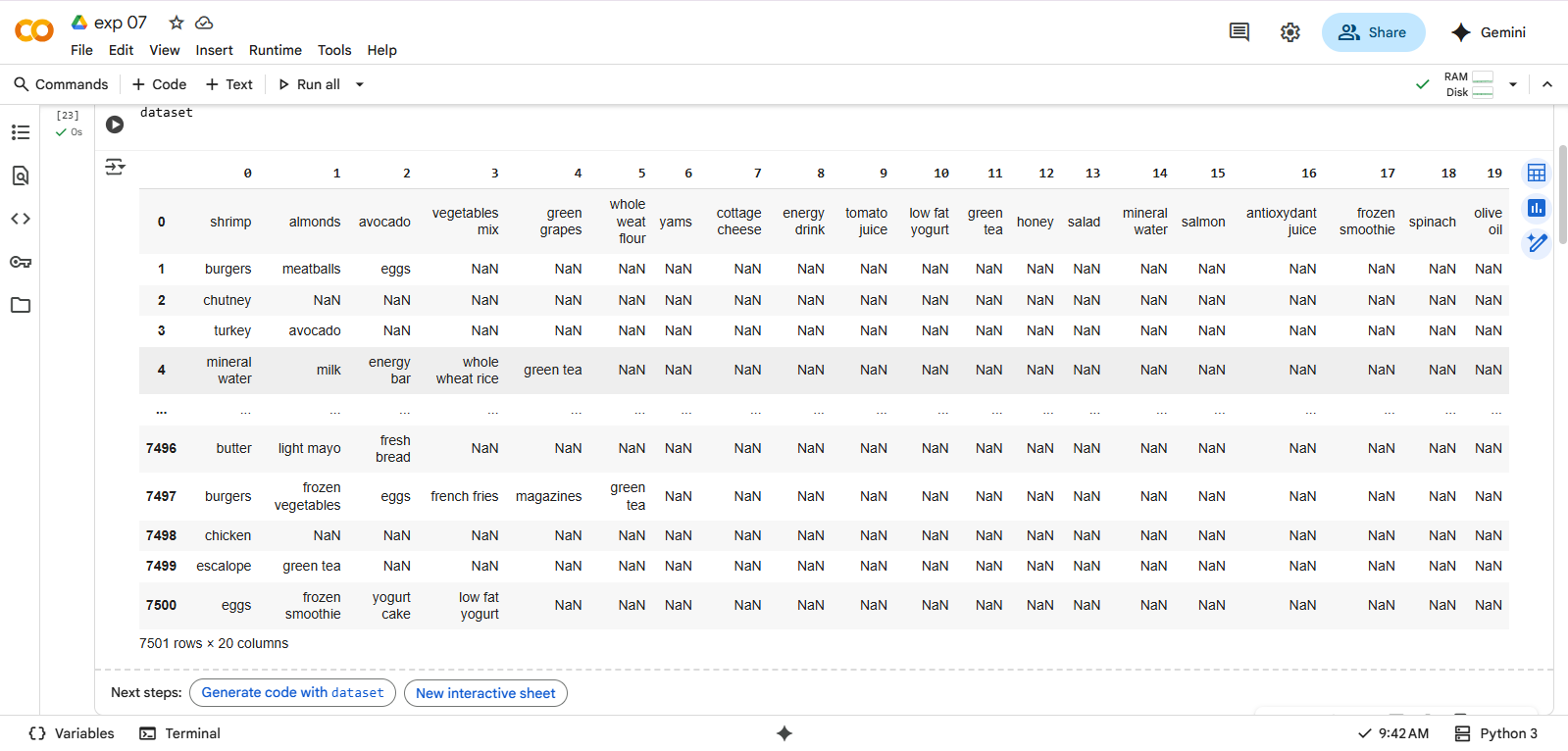
4. Eliminate candidates that are infrequent, leaving only those that are frequent

**Apriori Algorithm:**

****

****

**Code and output**:









**Conclusion**: Comment on the rules generated by the algorithm

The Apriori algorithm generates association rules by identifying frequent itemsets in a dataset and then deriving rules from these itemsets that meet predefined minimum support and confidence thresholds.

Characteristics of Rules Generated by Apriori:

* **Based on Frequent Itemsets:** Rules are derived only from itemsets that appear frequently in the dataset, as determined by the minimum support threshold. This ensures that the rules represent patterns that occur often enough to be considered significant.
* **Quantified by Support, Confidence, and Lift:** Each generated rule is associated with metrics:
  + **Support:** Indicates how frequently the entire itemset (antecedent and consequent) appears in the dataset.
  + **Confidence:** Measures the conditional probability of the consequent appearing given the presence of the antecedent.
  + **Lift:** Quantifies the strength of the association, indicating how much more likely the consequent is to occur when the antecedent is present, compared to its general occurrence. A lift value greater than 1 suggests a positive correlation.
* **Filtered by Thresholds:** Only rules exceeding the specified minimum support and confidence thresholds are considered "strong" and are outputted. This filtering helps to focus on the most relevant and reliable associations.
* **Actionable Insights:** The generated rules provide insights into relationships between items, which can be used for various applications such as market basket analysis, product recommendation systems, and understanding customer behavior. For example, a rule like "If a customer buys bread, they are also likely to buy milk" can inform store layout or promotional strategies.
* **Can Be Numerous:** In datasets with many items and transactions, the Apriori algorithm can potentially generate a large number of rules, which may require further analysis or visualization techniques to effectively interpret.
* **Subject to Threshold Impact:** The quality and quantity of generated rules are highly dependent on the chosen minimum support and confidence thresholds. Setting these too high may result in too few rules, while setting them too low can lead to an overwhelming number of less meaningful rules.