



BASKET ANALYSIS

DATA MINING

CONTENT

- 01** TEAM MEMBERS
- 02** PROJECT OVERVIEW
- 03** DATASET
- 04** DATA PREPROCESSING
- 05** ANALYSIS
- 06** EVALUATION
- 07** CONCLUSION

TEAM MEMBERS

Roll Number	Name
2017-miit-ece-035	Moe Myint Naing
2017-miit-ece-036	Naing Khant Win
2017-miit-ece-043	Pyae Phyo Tun
2017-miit-ece-44	Pyae Sone Khant Aung
2017-miit-ece-045	Sai Aung Soe Hein
2018-miit-cse-035	Oak Hla Gyi
2018-miit-ece-038	Myo Thinzar Kyaw

PROJECT OVERVIEW

Goal : To perform Basket Analysis using the Apriori algorithm to uncover patterns and relationships between items frequently purchased together.

By analyzing purchasing patterns, we can recommend products to customers based on the likelihood that they might be interested in buying them alongside other items.



USE CASES

Retail and E-Commerce: Product Bundling,Cross-Selling,Personalized Recommendations

Inventory Management : Stock Optimization,Demand Forecasting

Supermarkets and Grocery Chains : Product Placement Strategies,Promotional Campaigns



DATASET

Description



- The dataset titled "The Bread Basket" contains transactional data from a bakery located in the United Kingdom.
- It records all items sold at the bakery, providing a detailed view of customer purchases over time.
- This dataset is particularly useful for performing Market Basket Analysis to uncover patterns in customer buying behavior.

- Transactions: The dataset captures individual transactions, where each row represents an item purchased in a specific transaction.
- Time Period: The dataset covers sales data over a period from October 30, 2016, to April 9, 2017.

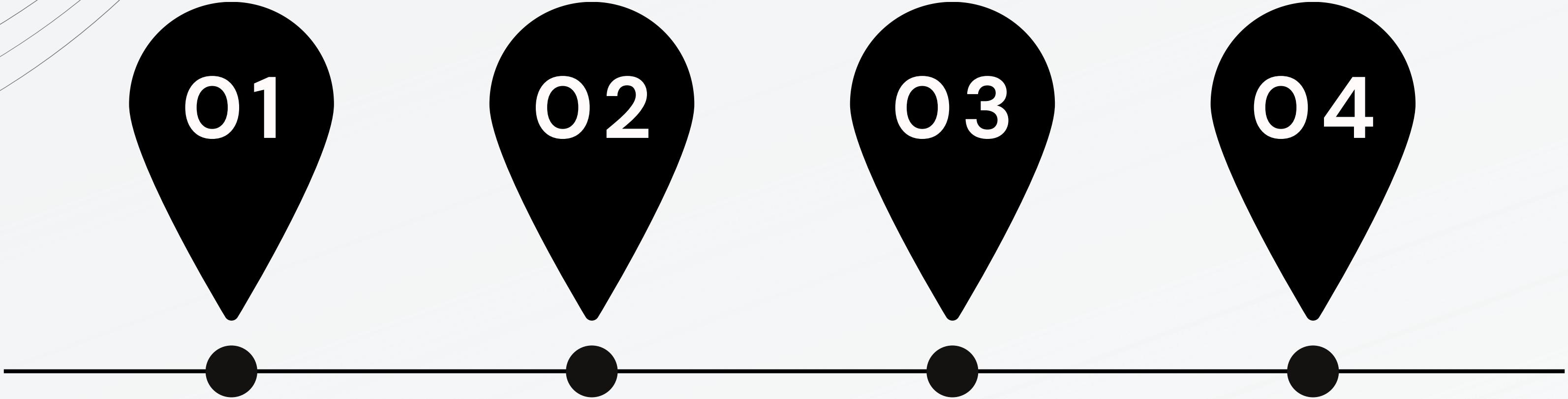
Vision



DATA PREPROCESSING

Preparation and transformation of raw data into a suitable format for modeling

Ensure that the data is clean, consistent, and ready for analysis



01

02

03

04

ADD COLUMNS

We use Pandas' method `to_datetime()`, and match the time formatting on what the `date_time` column is storing

LOWER CASE

Apply `lower()` function with Pandas' `apply` method

WHITESPACES

Stripping out the whitespaces from "Item" column

PIVOT TABLE

data into a table where each row corresponds to a transaction, each column to an item, and the cells show the quantity of the item bought in that transaction

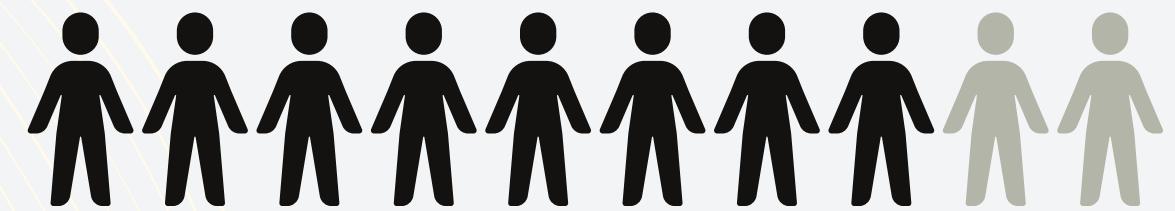
ANALYSIS

Apriori Algorithm
Association Rule

APRIORI

a popular method used in data mining for discovering frequent itemsets and generating association rules from a dataset based on two key concepts: frequent itemsets and association rules

80%



FREQUENT ITEMSETS

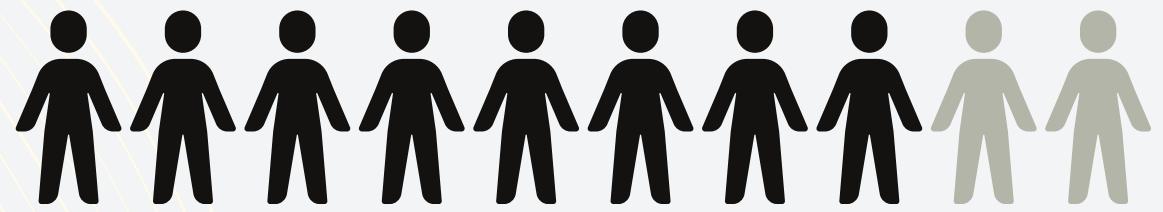
- An itemset is a collection of one or more items. An itemset is considered frequent if it appears in a dataset with a frequency greater than or equal to a user-defined minimum support threshold.
- The support of an itemset is the proportion of transactions in the dataset in which the itemset appears. For example, if "bread" appears in 20 out of 100 transactions, its support is 20%.



ASSOCIATION RULES

An association rule is an implication of the form $A \rightarrow B$, where A and B are itemsets, and A is called the antecedent while B is called the consequent. The rule suggests that when items in A are purchased, items in B are also likely to be purchased.

80%



METRICS FOR EVALUATION

Support

The support of a rule $A \rightarrow B$ is the proportion of transactions in the dataset that contain both A and B.

Confidence

This is the probability of seeing B in transactions under the condition that A is present. It is calculated as $\text{Support}(A \cup B) / \text{Support}(A)$.

Lift

This metric evaluates the strength of a rule over the random occurrence of B. It is calculated as $\text{Confidence}(A \rightarrow B) / \text{Support}(B)$. A lift greater than 1 indicates a positive correlation between A and B.



WORKING OF THE APRIORI ALGORITHM

Step 1: Frequent Itemset Generation

- Generate Candidate Itemsets: It starts by finding all itemsets that appear with the minimum support. First, it looks for individual items that meet the minimum support. Then, it generates larger itemsets by combining previously identified frequent itemsets.
- Prune Infrequent Itemsets: Any itemset that does not meet the minimum support is eliminated, reducing the number of candidate itemsets for the next iteration.



WORKING OF THE APRIORI ALGORITHM

Step 2: Association Rule Generation

- For each frequent itemset, the algorithm generates all possible association rules that have non-empty subsets.
- It calculates the confidence of each rule and prunes rules that do not meet the minimum confidence threshold.



EVALUATION

CONCLUSION

The goal is to discover patterns and associations between items that frequently occur together in transactions. The results of this analysis can be used to make recommendations or inform business strategies.

