

# Smart Market Insights Using Unsupervised Machine Learning

---

## Objective:

To explore customer purchasing patterns by performing unsupervised clustering and association rule mining on a retail product dataset.

---

## Dataset:

The dataset contains transaction-level data with the following columns:

- TransactionID: Unique transaction identifier
  - CustomerID: Unique customer identifier
  - Products: List of products purchased in a transaction
  - Timestamp: Date and time of transaction
- 

## Technologies Used:

- Python
  - Pandas, NumPy
  - Scikit-learn (TF-IDF, KMeans, silhouette score)
  - MLxtend (Apriori, association\_rules)
  - Matplotlib, Seaborn
- 

## Project Workflow:

### 1. Data Preprocessing:

- Removed null entries
- Cleaned product strings (lowercase, removed symbols)
- Converted product lists for transaction encoding

### 2. Market Basket Analysis:

- Applied **Apriori algorithm** with minimum support of 0.1
- Generated **association rules** using lift metric
- Identified frequently bought-together items

### 3. Clustering:

- Used **TF-IDF** to vectorize product combinations

- Applied **KMeans clustering** with cluster sizes ranging from 2 to 6
- Selected optimal number of clusters based on **silhouette score**

#### 4. Model Evaluation & Visualization:

- Plotted silhouette score for each cluster size
  - Visualized cluster distribution using bar charts
  - Displayed top 5 association rules sorted by lift
- 

#### Results:

- Best clustering result found with  $k = 5$  (5 = value with highest silhouette score)
  - Clear product groups formed based on similarity in purchases
  - Top association rules revealed strong item combinations (e.g., bread → butter)
- 

#### Conclusion:

This project demonstrates how machine learning can be used to extract valuable insights from retail transaction data. Clustering helps identify patterns in customer behavior, while association rules uncover product affinities. These findings can guide recommendation systems and sales strategies.

---

#### Future Scope:

- Customer segmentation
  - Product recommendation engine
  - Time-series purchase trend analysis
- 

**Author:** Priyanka P

**Tools:** Python, Jupyter Notebook, Scikit-learn, MLxtend, Pandas, Matplotlib