# Item-Based Scheme Recommendation

## Approach

This custom approach leverages item-based collaborative filtering using Jaccard similarity.  
  
Key Steps:  
- Load and preprocess stockist-product-scheme data.  
- Group products by scheme from training data.  
- Convert scheme-product mapping into a binary matrix using MultiLabelBinarizer.  
- Compute pairwise Jaccard similarity between schemes.  
- Recommend top-N similar schemes for each given test scheme based on product overlap.

### Key Findings and Business Insights

- Similar schemes tend to group together based on product overlaps.  
- This method helps in suggesting complementary or substitute schemes.  
- Aids the marketing team in identifying bundled offerings with overlapping product bases.  
- Product-based logic provides explainable recommendations that can be validated by domain teams.

### Assumptions, Limitations, and Edge Cases

- Assumes historical scheme-product associations are reliable indicators for future bundling.  
- Edge case: Schemes with only one or unique product will have fewer matches.  
- Cold-start limitation for new products or schemes not seen in training.  
- Only product co-occurrence is used — does not account for geography, pricing, or partner profile.

### Feature Descriptions

|  |  |
| --- | --- |
| Feature | Description |
| Partner\_id | Unique identifier for the channel partner or stockist. |
| Product\_id | Name or ID of the product involved in the scheme. |
| Geography | Region where the partner operates (e.g., East, West, North). |
| Stockist\_Type | Type of the partner (Retailer, Wholesaler, Distributor). |
| Scheme\_Type | The marketing scheme associated with the transaction (e.g., Cashback, Seasonal Offer). |
| Sales\_Value\_Last\_Period | Monetary value of sales made during the last period. |
| Sales\_Quantity\_Last\_Period | Number of units sold in the last period. |
| MRP | Maximum Retail Price of the product. |
| Growth\_Percentage | Growth in sales compared to previous periods (in %). |
| Discount\_Applied | Discount percentage applied to the product. |
| Bulk\_Purchase\_Tendency | Indicates if the stockist tends to purchase in bulk (Low, Medium, High). |
| New\_Stockist | Boolean flag to indicate whether the stockist is newly onboarded. |
| Feedback\_Score | Satisfaction score or feedback provided by the partner. |

### References

- Jaccard Similarity: https://en.wikipedia.org/wiki/Jaccard\_index  
- MultiLabelBinarizer: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MultiLabelBinarizer.html  
- Jaccard Score: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard\_score.html  
- pandas GroupBy: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html