Data 624 Homework 10 Market Basket and Clusters

Enid Roman

2024-11-22

Imagine 10000 receipts sitting on your table. Each receipt represents a transaction with items that were purchased. The receipt is a representation of stuff that went into a customer's basket - and therefore 'Market Basket Analysis'.

That is exactly what the Groceries Data Set contains: a collection of receipts with each line representing 1 receipt and the items purchased. Each line is called a transaction and each column in a row represents an item. The data set is attached.

Your assignment is to use R to mine the data for association rules. You should report support, confidence and lift and your top 10 rules by lift.

Load the Data

Summary of Transactions

```
## transactions as itemMatrix in sparse format with
##  9835 rows (elements/itemsets/transactions) and
##  169 columns (items) and a density of 0.02609146
##
## most frequent items:
##       whole milk other vegetables      rolls/buns          soda
##            2513             1903            1809          1715
##          yogurt          (Other)
##            1372            34055
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
## 2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55   46
##   17   18   19   20   21   22   23   24   26   27   28   29   32
##   29   14   14    9   11    4    6    1    1    1    1    3    1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   4.409   6.000  32.000
##
## includes extended item information - examples:
##             labels
## 1 abrasive cleaner
## 2 artif. sweetener
## 3   baby cosmetics
```

The dataset consists of 9,835 transactions (rows) and 169 unique items (columns).

Density of the item matrix is 0.026, meaning that approximately 2.6% of all possible item-transaction pairs are populated (indicating sparsity).

The most frequently purchased items are:

Whole milk: Appears in 2,513 transactions. Other vegetables: Appears in 1,903 transactions. Rolls/buns: Appears in 1,809 transactions. Soda: Appears in 1,715 transactions. Yogurt: Appears in 1,372 transactions.

These items are key drivers of transaction frequency and may dominate associations and clusters.

Most transactions are small, containing 1–3 items:

2,159 transactions contain only 1 item. 1,643 transactions contain 2 items. 1,299 transactions contain 3 items.

The largest transaction contains 32 items, but such transactions are rare.

The median transaction length is 4 items, with a mean length of approximately 6 items.

The dataset includes specific labels for items such as:

Abrasive cleaner Artificial sweetener Baby cosmetics

These labels indicate the diversity of the product categories.

The dataset is sparse, with most transactions involving only a small subset of items.

Frequent items like "whole milk" and "other vegetables" are likely to appear in key clusters or association rules.

Most transactions are small, making it necessary to focus on highly frequent items or combinations for meaningful analysis.

Item Frequency Plot   The visualization below provides a clear visual representation of the most commonly purchased items in the dataset.

Top 20 Most Frequent Items
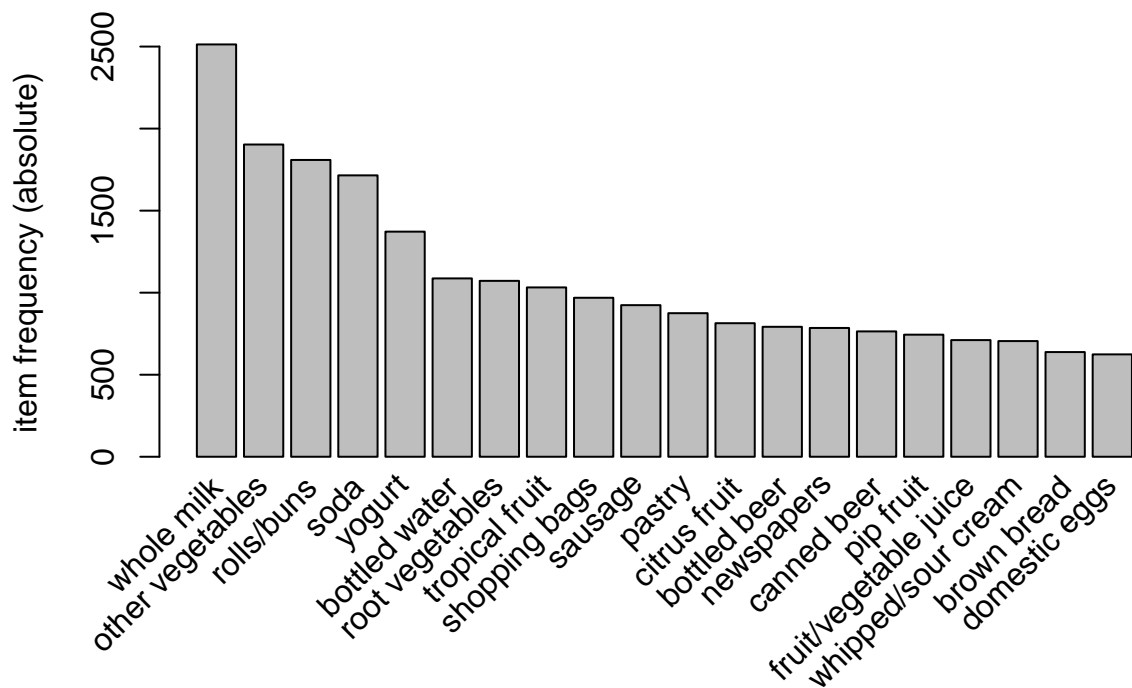
The top 5 most frequent items are:

Whole milk - Appears in the highest number of transactions (over 2,500). Other vegetables - Second most common, with around 1,900 occurrences. Rolls/buns - Approximately 1,800 occurrences. Soda - Around 1,700 occurrences. Yogurt - Over 1,300 occurrences.

Frequently purchased items also include:

Bottled water Root vegetables Tropical fruit Shopping bags Sausage

These items appear in 800–1,200 transactions, making them highly significant for market basket analysis.

Association Rules

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
```

```
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.01s].
## writing ... [410 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].


##      lhs                         rhs            support confidence coverage lift
## [1] {liquor, red/blush wine} => {bottled beer} 0.0019  0.90       0.0021   11.2
## [2] {cereals, curd}          => {whole milk}   0.0010  0.91       0.0011   3.6
## [3] {cereals, yogurt}        => {whole milk}   0.0017  0.81       0.0021   3.2
## [4] {butter, jam}            => {whole milk}   0.0010  0.83       0.0012   3.3
## [5] {bottled beer, soups}    => {whole milk}   0.0011  0.92       0.0012   3.6
##      count
## [1] 19
## [2] 10
## [3] 17
## [4] 10
## [5] 11
```

The Apriori algorithm was run with a minimum support threshold of 0.001 (absolute minimum support count = 9 transactions) and a minimum confidence threshold of 0.8. A total of 410 rules were generated.

Here are the top 5 rules generated (sorted by confidence):

{liquor, red/blush wine} => {bottled beer}

Support: 0.0019 Confidence: 0.90 Lift: 11.2 Insight: Transactions containing liquor and red/blush wine have a high probability (90%) of also including bottled beer, with strong association strength (lift > 11).

{cereals, curd} => {whole milk}

Support: 0.0010 Confidence: 0.91 Lift: 3.6 Insight: Customers buying cereals and curd often purchase whole milk together, making it a common pairing.

{cereals, yogurt} => {whole milk}

Support: 0.0017 Confidence: 0.81 Lift: 3.2 Insight: Whole milk is frequently purchased alongside cereals and yogurt, showing a strong breakfast-related pattern.

{butter, jam} => {whole milk}

Support: 0.0010 Confidence: 0.83 Lift: 3.3 Insight: Customers buying butter and jam often include whole milk in their basket, indicating another breakfast-related combination.

{bottled beer, soups} => {whole milk}

Support: 0.0011 Confidence: 0.92 Lift: 3.6 Insight: Even with unrelated items (bottled beer and soups), whole milk shows strong association, suggesting it's a frequent staple.

Support: Indicates the proportion of transactions containing the rule (ranges from 0.001 to 0.0019 for the top rules).

Confidence: High confidence values (0.81–0.92) suggest the rules are reliable predictors.

Lift: Ranges from 3.2 to 11.2, showing that the associations are stronger than random chance.

Whole Milk Dominance:

Whole milk appears in most of the rules, often associated with other staple or complementary items. Its strong lift values (>3.0) highlight its significant role in customer baskets.

Breakfast Patterns:

Items like yogurt, cereals, curd, butter, and jam commonly associate with each other and with whole milk.

Drinks and Pairings:

Liquor, wine, and bottled beer show strong associations, indicating complementary purchase patterns for beverages.

This analysis provides actionable insights for cross-selling strategies and better shelf arrangements to boost sales.

Sort rules by lift:

```
## set of 410 rules
```

The set of 410 rules generated by the Apriori algorithm is sorted by confidence in descending order. The rules are ranked based on the strength of the association between the antecedent and consequent items.

Setting the rules by confidence indicates the reliability of the rule when the antecedent occurs and helps identify rules that consistently predict the consequent.

Sorting helps prioritize the most actionable rules.

High-confidence rules are more reliable predictors.

Rules with high lift values indicate strong associations, useful for marketing strategies like cross-selling or promotions.

Analyze results: View top 10 rules by lift.

Table 1: Top 10 Association Rules Sorted by Confidence

	rules	support	confidence	coverage	lift	count
10	{rice,sugar} => {whole milk}	0	1	0	3.9	12
16	{canned fish,hygiene articles} => {whole milk}	0	1	0	3.9	11
33	{butter,rice,root vegetables} => {whole milk}	0	1	0	3.9	10
60	{flour,root vegetables,whipped/sour cream} => {whole milk}	0	1	0	3.9	17
62	{butter,domestic eggs,soft cheese} => {whole milk}	0	1	0	3.9	10
68	{citrus fruit,root vegetables,soft cheese} => {other vegetables}	0	1	0	5.2	10
126	{butter,hygiene articles,pip fruit} => {whole milk}	0	1	0	3.9	10
133	{hygiene articles,root vegetables,whipped/sour cream} => {whole milk}	0	1	0	3.9	10
135	{hygiene articles,pip fruit,root vegetables} => {whole milk}	0	1	0	3.9	10
143	{cream cheese,domestic eggs,sugar} => {whole milk}	0	1	0	3.9	11

The top 10 rules by lift provide valuable insights into customer purchase patterns and associations. Here are some key observations:

Whole Milk as a Key Item:

Whole milk frequently appears as the consequent (RHS), indicating it is often bought with other items.

Strong rules like {rice, sugar} => {whole milk} and {butter, root vegetables} => {whole milk} demonstrate its central role.

Breakfast and Essentials Patterns:

Associations like {butter, domestic eggs, soft cheese} and {cream cheese, sugar} suggest breakfast-related shopping patterns.

Diverse Item Pairings:

Items like hygiene articles, citrus fruits, and root vegetables show cross-category purchases with whole milk.

High Lift:

Rules with lift values >3.0 indicate strong associations compared to random co-occurrence. {citrus fruit, soft cheese} => {other vegetables} has the highest lift of 5.2.

Visualization of Top 10 Rules

Visualizing the top 10 rules by lift provides a clear overview of the association patterns between items. The graph highlights the relationships between antecedent and consequent items, showing the strength of the rules.

Top 10 Association Rules
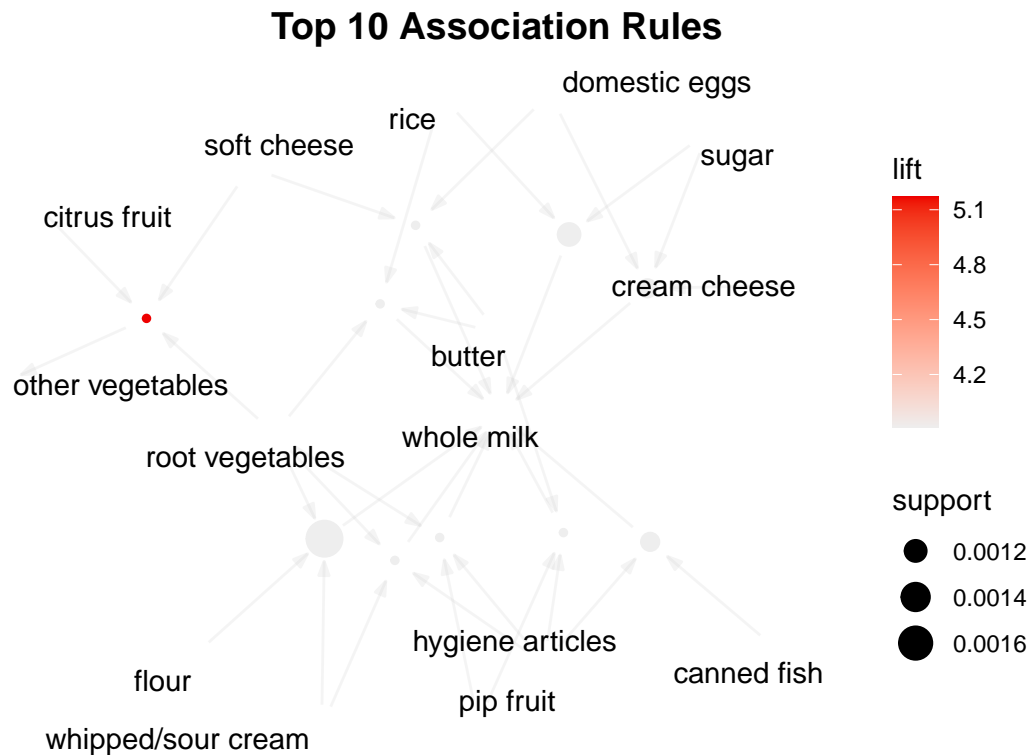
Nodes (Circles):

Each node represents an item in the dataset.

The size of the node corresponds to the support of the item (larger nodes have higher support).

Edges (Lines):

Edges represent association rules between items.

The thickness and color of the edges may indicate the strength of the association (e.g., lift).

Color Legend:

The lift of the rules is shown as a gradient (e.g., from lighter to darker red). Higher lift values represent stronger associations.

Support Legend:

The node size reflects the support of each item (how often it appears in the transactions).

Central Node: Whole Milk:

"Whole milk" appears to be a central item, frequently associated with multiple other items such as butter, root vegetables, and hygiene articles.

This indicates that "whole milk" is a key driver of transactions.

Strong Rules:

Edges connected to nodes like "canned fish," "cream cheese," and "hygiene articles" have high lift values (indicated by darker red), suggesting strong relationships with their consequents.

Clusters of Related Items:

Items such as "root vegetables," "whipped/sour cream," and "flour" cluster together, suggesting frequent co-occurrence.

Potential Uses of Visualization:

Marketing Strategy:

Use highly associated items for cross-promotions (e.g., discount "cream cheese" when "whole milk" is purchased).

Shelf Placement:

Place items with strong associations (e.g., "butter" and "whole milk") closer together.

Product Bundling:

Create bundles based on strong rules, such as pairing "canned fish" with "hygiene articles" and "whole milk."

Redundancies in Rules Remove redundant rules and sort by lift

Rules will be repeated sometimes. Redundancy means one item could be specified. You have the option of removing the item from the dataset. Otherwise, redundant rules generated can be removed. These repeated rules can be deleted.

Targeting Items We can target items that are most likely to be purchased together.

Lets limit the output. We could be interested in two types of targets, illustrated in an example of "whole milk".

If customers buy whole milk, what are they likely to buy?

Table 2: Top 5 Rules Targeting Whole Milk

	rules	support	confidence	coverage	lift	count
196	{rice,sugar} => {whole milk}	0	1	0	3.9	12
323	{canned fish,hygiene articles} => {whole milk}	0	1	0	3.9	11
1643	{butter,rice,root vegetables} => {whole milk}	0	1	0	3.9	10
1705	{flour,root vegetables,whipped/sour cream} => {whole milk}	0	1	0	3.9	17

	rules	support	confidence	coverage	lift	count
1716	{butter,domestic eggs,soft cheese} => {whole milk}	0	1	0	3.9	10

The rules were generated targeting transactions containing rhs = "whole milk":

Minimum Support: 0.001 Minimum Confidence: 0.08 (8%) Rules are sorted by confidence.

100% Confidence:

All top rules have a confidence value of 1.00, meaning that whenever the antecedent items are purchased, "whole milk" is always purchased.

This highlights the strong dependence of "whole milk" on these combinations.

Lift:

The lift value is 3.9 across all these rules, indicating that the probability of purchasing "whole milk" with these items is 3.9 times higher than random chance.

Combinations:

Some rules highlight staple or breakfast-related patterns: {butter, rice, root vegetables} {flour, root vegetables, whipped/sour cream} Others show cross-category combinations: {canned fish, hygiene articles}

Marketing Strategies:

Offer discounts or bundles on combinations like "butter + root vegetables" or "flour + whipped cream" with "whole milk."

Shelf Placement:

Place "whole milk" near frequently associated items to encourage complementary purchases.

Promotional Campaigns:

Highlight cross-category pairings like "canned fish and hygiene articles" to attract diverse customer needs.

Plot Visualization of Targeting Items Visualizing the top 5 rules targeting "whole milk" provides a clear overview of the association patterns between items. The graph highlights the relationships between antecedent items and "whole milk," showing the strength of the rules.

Top Rules Targeting to Whole Milk
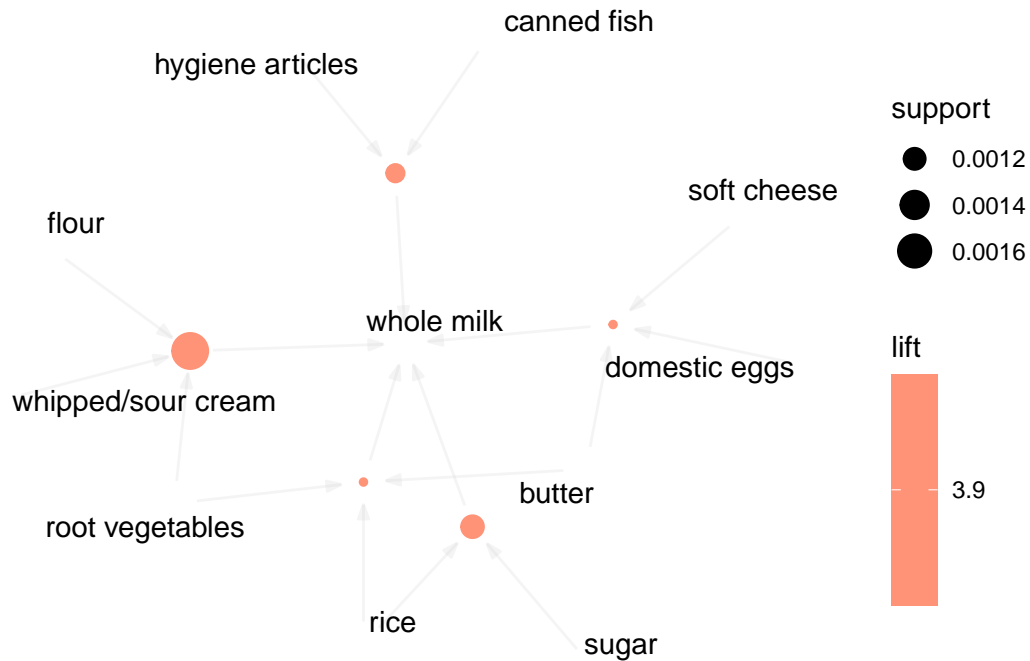
Nodes (Circles):

Each node represents an item (e.g., "whole milk," "butter," "sugar").

Size of Nodes

Larger nodes indicate items with higher support (i.e., items that appear more frequently in transactions).

Edges (Lines):

Edges represent association rules between items.

An arrow points from the antecedent (LHS) to the consequent (RHS) (in this case, "whole milk"). The thickness of edges and intensity of color represent the strength of the rule, often linked to the lift metric.

Legends:

Support: Node size correlates with how often the item appears in transactions.

For example, "whole milk" has the largest node because it is the most frequent item.

Lift: Edge color intensity represents the strength of the association.

Higher lift values mean a stronger association between items, indicating that the purchase of antecedents strongly predicts the consequent.

Whole Milk (Central Node):

As the consequent in all rules, "whole milk" is central in this graph.

Its strong connections with other items reflect its high frequency in transactions and its role as a common co-purchase.

Key Antecedent Items:

Root Vegetables and Whipped/Sour Cream:

Strong associations with "whole milk," often forming a shopping pattern for cooking or baking.

Sugar and Rice:

Frequently associated with "whole milk," likely as staples or complementary purchases.

Butter and Domestic Eggs:

Reflect breakfast-related or cooking-related patterns where these items co-occur with "whole milk."

Cross-Category Relationships:

Hygiene Articles and Canned Fish:

These are less intuitive but may reflect a pattern of "whole milk" being a consistent part of large or diverse shopping trips.

Central Role of Whole Milk:

"Whole milk" is a frequent purchase that complements many categories, making it a key product for cross-promotions.

Strong Pairings:

Items like "root vegetables," "butter," and "whipped cream" have strong links to "whole milk," showing patterns of meal preparation or staple shopping.

Lift Analysis:

Higher lift (e.g., 3.9) means that the likelihood of buying "whole milk" significantly increases when certain items are purchased (e.g., "sugar" and "rice").

This provides opportunities for marketing strategies like bundling these items with "whole milk."

Extra credit: do a simple cluster analysis on the data as well. Use whichever packages you like.

Cluster Analysis K-Means Clustering and Hierarchical Clustering

Converting the transactions to a binary matrix, dataframe, and scale the data.

K-Means Clustering

```
##    [1] 2 2 3 2 1 3 2 1 2 3 1 3 2 2 2 2 2 2 2 2 1 2 3 2 1 2 2 2 1 2 2 1 1 1 2 2 2
##   [38] 2 1 2 2 3 2 2 2 2 2 2 2 1 2 2 1 1 3 3 2 2 2 2 1 2 3 2 2 3 2 2 3 3 2 3 3 3
##   [75] 2 1 3 2 2 2 2 3 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 3 2 2 3 2 3 2 2 2 3 2 2 1
##  [112] 2 3 2 2 3 3 1 3 1 2 2 2 1 2 3 2 2 2 2 2 3 2 2 2 1 2 2 2 2 3 1 3 3 2 2 1 2
##  [149] 2 2 1 2 2 1 1 2 2 3 1 2 3 2 2 2 1 2 3 3 2 1 1 1 2 2 2 3 3 2 2 1 3 1 1 2 2
##  [186] 1 2 2 3 1 2 2 2 2 2 3 2 2 2 2 2 2 2 2 3 2 1 2 2 2 2 2 1 3 3 2 2 2 2 2 2 2
##  [223] 1 2 2 2 2 2 2 2 1 1 3 2 2 1 2 1 3 3 2 1 3 2 3 2 2 2 2 1 3 2 3 2 2 1 1 2 2 1
##  [260] 3 1 2 2 2 2 2 2 1 2 2 3 2 2 3 1 1 1 3 1 2 2 3 2 2 3 3 2 2 2 1 2 1 1 1 3 2
##  [297] 2 2 2 1 2 2 2 2 2 2 2 2 3 2 3 2 2 2 1 2 2 2 2 2 2 2 3 2 1 1 2 2 1 2 2 2 2
##  [334] 2 2 2 2 2 2 1 3 2 2 2 2 1 2 3 3 3 2 2 2 2 2 3 2 2 2 3 3 3 2 2 3 2 1 2 2 2
##  [371] 1 2 2 1 2 3 3 2 2 2 2 2 2 2 2 1 2 2 2 3 2 2 2 2 2 3 2 2 3 1 2 1 3 1 1 2 2
##  [408] 2 2 2 2 1 1 3 2 1 2 2 2 1 2 2 3 2 2 2 2 2 2 3 2 2 1 2 2 2 2 2 1 2 1 2 1 2
##  [445] 2 2 1 2 1 1 3 2 3 1 2 1 2 2 2 2 2 2 2 3 2 2 1 2 3 2 1 2 3 2 1 3 1 2 3 2 2
##  [482] 3 2 2 2 2 1 2 1 2 2 2 2 3 2 2 2 1 1 1 2 3 2 2 2 2 2 3 1 2 2 2 2 2 2 3 2 3
##  [519] 1 2 2 2 1 1 2 1 2 3 2 2 2 2 2 3 2 2 2 3 3 2 2 2 2 1 2 3 3 3 3 1 2 2 2 2 2
##  [556] 2 2 1 2 2 2 1 3 3 2 2 2 2 3 2 2 1 1 2 2 2 2 2 2 3 2 2 2 2 2 3 3 3 2 2 2 2
```

```
##  [593] 2 2 1 1 3 3 2 2 2 2 2 3 1 2 2 3 1 2 2 2 1 2 1 3 2 2 2 3 2 2 2 2 3 3 2 1 3
##  [630] 3 1 3 3 1 2 1 2 1 2 2 2 1 1 2 2 2 2 2 1 2 2 3 2 1 1 2 2 2 1 3 2 1 2 2 2 3
##  [667] 2 3 1 2 2 2 3 2 2 3 2 3 3 3 1 3 2 2 1 3 2 2 2 2 2 2 2 2 3 2 1 2 3 2 3 2 2
##  [704] 2 3 2 1 3 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 1 3 1 2 2 2 2 2 2 2 1
##  [741] 3 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1 3 1 2 1 2 2 3 3 2 1 2 2 1 3 1 1
##  [778] 1 2 1 2 3 3 2 3 2 2 2 2 1 2 3 3 2 2 3 1 3 2 2 3 1 2 3 2 2 2 2 3 3 3 3 2 2
##  [815] 2 2 3 1 2 2 1 3 3 1 2 3 2 1 2 3 3 1 3 2 2 3 2 2 2 3 3 2 3 1 2 3 1 1 1 1 3
##  [852] 2 1 2 1 3 2 2 2 1 1 1 2 2 3 2 2 2 3 2 2 3 2 3 3 2 2 2 2 2 2 2 2 2 2 3 2 2
##  [889] 2 1 2 2 1 3 1 3 3 2 2 3 2 1 3 2 1 2 2 3 2 3 2 2 2 2 2 2 2 2 2 2 3 3 3 1 2
##  [926] 2 2 3 2 2 3 3 2 2 2 2 3 2 2 2 2 2 3 2 2 1 3 2 1 1 1 2 3 2 2 2 2 3 3 2 2 1
##  [963] 1 2 2 2 3 2 1 2 2 2 2 2 1 2 2 2 1 1 1 2 2 2 3 3 2 2 1 3 1 2 2 2 2 1 1 2 2
## [1000] 1 2 2 2 1 2 2 2 2 3 3 2 1 3 1 2 2 1 2 3 3 2 2 3 2 2 2 1 2 2 2 3 2 2 2 2 3
## [1037] 2 2 2 3 2 1 1 2 1 1 1 2 2 3 2 3 3 2 2 2 3 2 3 2 2 2 3 1 2 2 1 2 2 2 2 2 2
## [1074] 1 1 2 1 2 2 2 2 3 2 1 3 2 2 2 2 1 1 1 3 2 2 2 2 2 2 2 2 1 1 2 3 1 2 2 1 2
## [1111] 2 2 2 2 1 2 2 2 2 3 2 1 1 1 2 2 1 2 2 2 2 2 2 1 1 2 2 1 2 1 2 2 2 2 1 2 2
## [1148] 1 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 3 1 1 1 1 2 2 2 1 2 2 2 2 2 2 2 1 2
## [1185] 1 2 2 3 3 3 2 2 2 2 2 3 1 3 2 1 2 2 2 1 3 3 2 3 1 3 1 3 2 2 3 1 1 1 1 2
## [1222] 1 2 2 2 2 1 1 1 3 2 2 2 2 1 3 3 2 2 2 2 2 1 1 1 2 2 3 2 2 2 2 3 2 3 1 2 2
## [1259] 2 2 3 2 1 2 2 2 1 3 2 3 2 1 2 2 1 2 2 2 3 2 2 1 2 1 1 2 2 2 2 3 3 2 2 1 2
## [1296] 2 2 2 2 2 1 1 3 2 1 2 3 1 1 2 1 2 1 2 1 2 2 2 1 2 2 2 1 1 3 2 2 2 2 2 3 1
## [1333] 3 2 3 2 2 3 1 1 2 2 2 3 2 1 3 3 2 2 2 1 2 1 2 2 2 1 3 2 1 2 2 3 2 1 2 2 2
## [1370] 2 2 1 2 3 1 2 2 2 2 2 2 3 2 2 3 1 2 1 2 2 2 3 1 2 2 2 2 2 2 3 2 2 2 2 2 1
## [1407] 3 2 2 3 3 2 2 1 2 2 2 1 2 3 2 2 2 2 2 3 3 2 2 2 3 3 2 3 2 3 2 2 2 2 2 2 3
## [1444] 3 1 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 1 2 2 3 2 1 2 2 3 2 1 2
## [1481] 3 2 3 2 2 2 1 2 1 1 2 2 3 2 3 2 1 2 2 2 3 2 2 2 2 2 1 1 3 2 2 1 2 1 3 2 1
## [1518] 2 3 2 2 2 3 2 2 1 2 2 2 2 2 2 2 1 3 3 2 3 1 2 1 1 1 1 2 2 2 2 3 2 1 3 2 2
## [1555] 1 2 2 1 1 2 2 3 2 2 2 3 2 2 2 2 2 2 2 2 2 1 2 3 2 2 2 2 2 2 2 2 2 2 2 1 2
## [1592] 1 2 1 2 2 2 2 2 1 2 2 1 2 2 2 1 3 2 2 2 2 3 2 2 2 3 2 2 2 1 1 1 2 2 2 2 2
## [1629] 3 2 2 2 2 2 2 2 2 2 1 2 3 2 2 2 2 2 2 2 2 3 1 3 2 1 2 2 2 1 3 2 1 1 3 3 3
## [1666] 1 3 2 2 2 3 3 2 2 2 1 2 1 2 1 2 1 2 2 2 2 3 3 2 2 2 2 3 2 1 1 2 1 2 2 2 2
## [1703] 2 1 3 1 1 2 2 3 1 2 3 2 2 2 1 2 1 2 1 2 3 1 2 2 3 3 3 2 2 2 2 2 2 2 1 2 2 3
## [1740] 3 2 2 1 2 2 2 2 2 2 2 3 2 2 1 2 1 2 1 2 2 2 2 3 2 3 3 1 2 2 1 2 2 3 1 1 2
## [1777] 2 2 2 1 3 2 2 2 2 3 3 2 2 3 3 2 2 3 3 2 3 3 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2
## [1814] 2 3 2 2 2 3 2 2 2 2 3 2 3 1 2 2 3 3 2 2 2 2 3 1 2 2 2 1 2 1 2 2 1 3 2 2 3
## [1851] 2 2 3 1 2 2 3 2 2 2 2 2 1 2 2 2 3 3 2 3 2 3 3 1 2 2 2 1 2 2 1 1 2 1 2 3
## [1888] 2 1 2 2 3 1 2 2 1 2 3 3 2 2 2 1 3 3 1 2 2 2 2 2 2 1 1 2 3 2 1 3 3 1 1 1 1
## [1925] 2 2 3 2 2 2 2 2 2 1 3 2 3 3 1 3 2 1 3 2 3 2 1 2 3 1 3 2 3 3 2 2 2 2 2 2 2
## [1962] 2 2 2 2 1 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 3 2 1 2 2 2 3 2 3 2
## [1999] 2 2 2 1 1 2 1 2 1 1 3 2 1 3 2 1 1 1 2 1 2 2 1 3 2 2 3 2 1 2 2 2 2 2 2 2 1
## [2036] 3 2 2 2 3 3 2 3 2 2 2 3 3 2 2 2 1 2 1 1 1 2 2 2 1 3 2 2 1 2 2 2 3 2 2 3 3
## [2073] 2 2 2 2 3 2 2 2 3 2 3 2 2 2 3 2 3 1 1 2 2 2 3 1 3 2 2 1 3 3 2 2 2 2 1 2 1 2
## [2110] 2 2 2 2 2 2 2 2 3 2 2 1 1 2 1 2 2 2 2 2 2 2 3 2 1 3 1 1 2 1 1 2 2 2 2 3 2 2
## [2147] 2 3 3 2 2 2 2 2 2 2 3 3 2 2 1 1 2 2 2 1 2 2 1 1 1 3 3 2 2 2 1 1 2 2 2 2 2
## [2184] 1 2 1 2 1 3 2 2 1 2 2 3 2 3 2 1 2 2 2 2 2 2 2 3 2 3 2 2 1 1 1 2 3 3 1 2 2
## [2221] 2 2 1 3 2 3 3 3 1 2 3 2 1 3 2 2 2 2 2 2 3 2 1 2 2 3 2 2 2 2 2 2 2 3 1 2 1 3
## [2258] 2 3 2 2 1 2 3 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 3 2 2 2 2 2 2 2 2 1 3 2 1 2 2
## [2295] 2 2 2 2 2 3 1 2 3 2 2 3 3 3 2 1 2 2 3 1 3 1 2 2 2 1 1 2 1 2 1 3 2 2 2 2 3 3 3
## [2332] 2 3 2 1 3 1 2 2 2 2 2 2 3 3 1 2 2 1 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 1 3 2 2
## [2369] 2 2 3 3 1 2 2 2 1 2 2 2 2 2 2 2 2 2 3 2 1 2 3 1 2 2 3 2 2 2 2 2 2 2 3 1 2 2
## [2406] 2 2 2 1 3 2 3 2 2 1 2 3 1 2 2 2 2 2 2 2 2 1 2 2 3 1 3 2 1 1 2 2 2 2 2 2
## [2443] 3 2 1 2 2 2 2 3 2 2 3 2 1 3 2 2 2 2 2 3 3 2 2 2 3 1 2 1 2 2 1 3 2 2 2 3 1
## [2480] 3 2 2 2 2 2 2 1 2 2 3 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 3 2 2 1 2 2 2 2 2 3 2 2
## [2517] 1 2 2 1 2 3 2 2 2 3 1 3 1 2 3 2 1 2 2 1 2 1 2 3 2 2 1 1 3 1 2 2 2 1 3 2 2
## [2554] 1 3 3 2 2 2 1 3 2 3 2 2 3 2 3 2 2 2 2 2 1 2 2 3 3 2 2 1 2 3 2 1 2 2 1 1 1
```

11

```
## [2591] 2 2 2 2 3 2 1 2 1 3 2 2 2 3 2 2 2 2 3 2 3 1 2 2 3 2 2 2 2 3 2 3 2 1 3 2 2
## [2628] 2 1 1 2 1 3 2 3 3 1 3 2 2 2 1 2 2 3 2 3 2 2 2 2 1 2 2 2 2 2 1 2 2 2 3 1 2
## [2665] 3 3 2 1 2 2 3 2 2 1 3 2 2 2 2 3 2 1 2 1 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 1 2
## [2702] 2 1 2 1 2 2 1 2 1 2 2 2 2 3 3 3 2 2 2 2 3 3 2 3 2 1 2 2 1 2 1 1 2 2 1 2
## [2739] 3 2 3 3 2 1 1 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 3 1 1 2 2 2 2 2 2 2 2 3 2 2
## [2776] 1 2 1 3 2 3 1 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
## [2813] 2 2 3 1 2 2 2 3 2 2 2 2 2 1 2 2 2 2 3 2 2 2 2 3 3 2 3 2 2 2 1 2 2 2 2 3 3
## [2850] 2 3 2 2 2 2 2 1 2 2 2 2 2 1 3 2 3 2 2 3 2 2 2 2 2 3 2 2 2 2 2 1 3 3 2 3 3
## [2887] 3 3 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 3 2 2 2 2 3 1 1 2 2 2 2 2 2 3 2 2 2 2 2
## [2924] 2 2 2 2 1 2 2 3 2 2 2 2 2 3 1 2 2 2 1 3 2 2 2 1 2 2 1 2 3 1 1 2 3 2 2 2
## [2961] 1 3 1 2 3 2 2 3 1 2 1 1 3 1 3 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 2 1 3
## [2998] 2 3 3 2 2 1 2 1 3 2 2 1 1 3 2 1 2 2 1 1 1 1 2 3 3 1 2 2 2 1 1 2 2 3 1 1 2
## [3035] 2 1 2 2 3 2 1 1 2 3 2 1 2 2 1 2 2 1 3 1 3 1 1 2 1 2 2 2 2 3 2 3 1 2 2 3 2
## [3072] 3 2 3 2 2 2 2 3 1 2 1 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2
## [3109] 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 1 3 1 2 2 2 2 2 2 2 2
## [3146] 3 1 3 3 1 1 2 1 1 2 2 1 2 3 2 2 1 2 2 2 2 2 1 3 1 2 2 1 3 2 3 2 1 3 2 2 2
## [3183] 3 2 2 3 2 2 2 1 2 2 1 2 1 3 1 3 1 2 2 2 2 1 1 2 1 2 1 1 2 2 2 1 1 1 3 1 2
## [3220] 1 1 1 3 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 3 1 3 1 2 3 2 2 1 2 2 2 2 2 2 3
## [3257] 3 2 2 1 1 2 1 2 2 1 2 2 1 2 3 2 2 2 1 2 2 2 2 2 2 3 3 2 1 1 2 2 3 2 2 1 2
## [3294] 2 3 1 2 1 2 3 3 2 3 2 2 1 2 3 2 2 2 2 2 2 2 3 2 1 1 2 2 2 2 2 2 2 2 2 2 1
## [3331] 1 3 2 2 1 1 2 1 1 3 2 2 3 2 1 2 2 2 2 1 2 2 2 2 3 2 2 1 2 2 2 2 2 3 2 2 2
## [3368] 2 2 2 2 3 2 2 3 2 2 2 2 2 1 3 1 1 2 2 2 2 2 1 3 3 2 2 2 1 2 2 2 2 2 2 2 3
## [3405] 3 3 2 2 2 3 3 2 2 2 3 2 2 1 2 1 2 3 2 1 2 2 2 1 3 2 2 2 2 2 3 2 3 2 2 2 3 2
## [3442] 2 3 2 2 2 2 2 3 2 1 2 2 3 2 2 2 2 2 1 2 2 2 2 2 1 2 3 3 3 2 2 2 2 1 3 2 2
## [3479] 2 2 2 2 2 2 2 2 3 2 2 3 3 2 1 2 3 3 2 2 2 2 1 2 1 2 2 3 2 3 3 2 2 2 2 2 1
## [3516] 2 2 2 3 2 2 2 2 2 2 2 1 2 2 2 2 3 2 2 2 2 2 1 2 2 3 2 2 3 2 3 3 3 3 2 2 1 2
## [3553] 2 2 2 1 3 3 2 1 2 1 2 2 2 3 2 3 2 2 2 2 2 2 2 3 1 2 2 2 2 2 2 2 2 3 2 2 2
## [3590] 3 2 2 3 1 2 3 2 2 2 2 3 2 2 2 2 2 2 1 1 2 3 2 2 3 3 2 2 2 2 2 2 2 1 2 2 1 2
## [3627] 2 1 2 1 1 1 2 3 2 2 2 2 2 3 3 3 2 3 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2
## [3664] 2 2 1 2 2 1 2 2 3 1 1 2 1 2 1 1 2 2 2 2 2 3 2 1 1 2 2 1 3 3 3 2 1 2 2 2 2
## [3701] 2 1 2 2 2 3 2 1 2 3 2 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 3 2 2 3 2 2 2 2 3 2 2
## [3738] 3 3 2 3 2 2 3 2 2 2 2 2 2 3 1 2 2 1 2 2 2 1 2 3 1 2 2 2 2 2 3 2 2 2 2 2 2
## [3775] 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 3 1 2 2 2 2 2 3 2 2 3 1 2 2 2 2 2 2 2 2 2 3
## [3812] 2 2 1 1 2 2 1 2 2 1 3 2 2 2 2 2 1 2 2 2 2 1 3 2 2 2 3 3 3 1 1 2 2 3 1 2 2
## [3849] 1 1 2 1 2 1 1 3 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 3 2 3 2 2 2 1 2 3 1
## [3886] 2 3 3 2 2 1 1 2 2 1 3 2 2 3 2 2 2 1 3 1 2 1 1 2 2 1 3 2 3 3 3 2 3 2 2 2 2
## [3923] 2 1 2 2 2 2 2 1 2 2 2 1 3 3 2 1 2 3 2 2 2 2 2 2 2 1 3 1 3 1 1 1 1 1 2 2 2
## [3960] 2 2 2 3 3 2 2 2 2 2 1 1 1 3 1 2 2 1 1 3 2 2 2 2 3 1 2 2 2 3 2 2 1 2 1 2 2
## [3997] 1 2 2 2 1 2 2 2 3 2 2 2 2 3 3 3 2 3 1 3 3 2 2 1 2 2 2 2 2 2 2 2 2 3 2 2 1 1
## [4034] 3 2 1 3 1 2 2 2 1 2 2 1 2 3 3 1 3 2 1 1 2 2 3 2 3 2 1 3 2 2 1 2 1 2 1 2 2
## [4071] 2 3 3 1 2 2 3 2 3 1 1 2 2 3 2 2 2 1 3 1 3 2 2 2 2 3 3 2 3 2 3 3 3 3 1 2 1
## [4108] 1 2 2 2 3 3 2 2 2 2 2 2 3 2 3 2 3 2 1 1 3 2 1 2 3 3 2 1 2 1 2 2 3 2 2 2 2
## [4145] 1 2 2 2 2 2 1 2 2 2 2 2 1 2 1 1 3 2 1 1 3 1 3 1 2 2 2 2 2 3 3 2 2 2 2 1 2
## [4182] 2 1 1 2 1 2 2 2 2 3 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 3 2 2 1 2 2 3 2
## [4219] 1 2 2 1 1 2 2 3 3 2 1 1 1 1 1 3 1 3 2 3 2 2 2 2 2 2 3 2 3 3 2 1 2 3 2 2 2
## [4256] 1 3 1 2 2 3 2 2 2 2 2 2 2 2 3 2 1 2 3 3 2 3 2 2 1 2 2 2 2 2 2 3 3 1 3 3 2
## [4293] 2 3 1 2 3 2 2 2 3 1 3 2 1 2 1 2 2 2 3 3 3 1 3 2 1 3 1 2 3 2 2 1 3 2 3 1 1
## [4330] 1 2 1 3 2 2 3 3 3 1 3 2 2 1 2 1 1 3 2 3 2 2 2 2 1 1 2 1 1 2 1 3 2 1 2 2 2
## [4367] 1 3 1 1 2 1 2 3 1 1 2 1 3 2 1 3 3 2 2 1 2 2 1 2 3 2 3 2 1 1 1 2 1 3 2 2 2
## [4404] 2 1 3 2 2 2 1 2 1 2 2 1 1 1 2 3 1 1 2 3 2 2 2 3 2 1 2 3 3 3 2 2 2 2 2 2 3
## [4441] 2 3 2 1 2 1 2 1 3 1 1 2 2 2 3 2 1 2 1 1 1 2 1 2 2 2 2 3 2 2 2 1 2 2 2 1 2
## [4478] 2 3 1 2 2 3 2 3 2 3 1 2 2 2 1 2 1 2 2 2 2 1 3 2 2 2 1 1 2 2 2 2 2 3 2 3 1
## [4515] 2 1 3 2 1 2 2 1 3 2 2 3 2 2 1 2 1 2 2 2 2 2 3 2 3 1 2 2 1 1 2 1 2 2 1 1 3
## [4552] 2 2 1 3 1 2 1 2 2 2 2 2 2 3 1 2 2 3 2 1 2 2 1 1 2 3 2 3 2 2 2 2 2 2 3 2 3
```

12

```
## [4589] 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 3 2 3 2 2 2 2 2 3 3 2 2
## [4626] 1 1 2 1 2 2 2 2 2 3 3 2 2 3 2 2 2 1 2 2 2 3 2 2 2 2 3 2 2 2 3 2 2 2 2 2 2 2
## [4663] 2 3 1 1 2 2 2 2 2 2 1 3 2 2 1 3 2 1 3 1 2 2 2 2 1 2 2 2 2 2 3 2 1 3 2 2 2
## [4700] 3 2 2 2 3 2 3 1 2 2 2 2 1 2 1 2 3 2 3 3 2 1 1 2 2 1 3 2 2 3 2 1 1 2 2 2 2
## [4737] 2 2 3 3 1 1 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 3 1 3 2 2 2 2 2 3 2 2 2 2 2 2 1
## [4774] 1 2 2 2 2 2 3 1 3 2 1 2 2 3 2 2 2 2 2 2 2 3 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2
## [4811] 2 2 2 2 2 3 2 2 2 2 1 2 2 2 2 2 2 3 3 2 2 2 2 3 2 2 3 2 1 2 1 1 3 2 2 2 1
## [4848] 1 3 2 2 2 2 2 3 3 2 2 2 2 2 2 3 2 2 1 2 1 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2
## [4885] 3 2 2 3 2 2 3 2 1 2 2 2 1 2 3 2 3 2 2 3 2 2 2 2 2 2 2 2 3 2 3 3 1 2 2 2 2
## [4922] 2 3 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 3 2 2 1 1 2 2 3 1 3 3 2 2 2 2 2 3 2
## [4959] 3 3 2 1 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 3 2 2 3 2 1 2 2 2 2 2 1 3 3 2 2 2
## [4996] 1 1 2 2 2 2 1 3 2 1 2 3 2 3 2 2 3 3 2 2 2 2 2 1 2 2 2 2 2 3 1 2 1 3 2 3 2
## [5033] 2 1 2 2 2 2 2 2 2 2 2 2 3 2 2 3 1 1 3 2 2 2 2 2 2 2 2 3 2 3 2 2 3 1 2 3
## [5070] 1 3 3 2 3 2 2 2 1 2 2 3 2 3 2 2 2 2 3 2 1 2 2 2 2 2 2 2 3 2 2 3 2 3 2 2 2
## [5107] 1 2 3 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 3 3 2 1 1 2 2 2 2 2 2 2 2
## [5144] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 3 1 1 1 2 1 2 1 2 2 1 1 2 1 2 2 3 2 2
## [5181] 2 2 2 2 2 2 2 2 1 2 2 3 2 3 1 1 2 2 3 1 2 2 2 2 1 1 2 2 3 2 3 2 2 2 3 1 2
## [5218] 2 2 1 2 2 2 2 2 3 2 2 1 2 2 2 2 2 2 2 1 2 1 2 3 1 2 1 1 1 2 2 2 2 2 2 2 1
## [5255] 2 2 3 3 2 2 2 3 2 1 2 2 2 2 3 1 3 2 2 3 2 2 2 2 2 2 3 2 2 3 1 2 1 2 2 2 2
## [5292] 3 2 3 1 1 3 1 2 3 2 2 2 2 2 1 2 2 3 2 3 3 1 2 2 1 2 2 2 2 2 3 3 2 3 2 1 2
## [5329] 2 3 3 2 2 2 2 1 2 1 3 3 2 2 3 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 3 1
## [5366] 3 2 2 2 2 2 2 2 1 3 2 2 2 3 2 2 2 2 1 3 3 2 2 2 2 2 1 2 2 2 2 1 3 2 2 2 3
## [5403] 2 2 2 1 2 1 2 2 2 2 2 1 2 2 2 1 3 2 2 2 1 1 1 2 1 2 2 3 2 2 2 3 3 3 2 2 2
## [5440] 1 3 2 2 3 3 2 3 2 3 2 3 3 2 2 2 1 2 2 2 3 2 1 2 1 2 2 3 3 2 1 2 3 2 1 2 2
## [5477] 1 3 1 2 2 1 1 1 2 1 2 2 3 2 1 3 2 1 2 2 2 2 3 3 1 2 2 2 1 2 2 2 3 3 3 3 3
## [5514] 2 1 2 1 1 2 3 2 2 2 2 1 2 1 1 1 1 2 2 1 2 1 2 3 1 2 3 2 1 2 2 2 2 2 1 1 1
## [5551] 1 1 1 2 3 2 2 3 1 2 2 2 2 2 2 2 2 3 3 2 3 3 2 2 2 1 2 3 2 2 2 1 2 1 2 2
## [5588] 2 2 2 2 3 2 3 1 1 2 3 2 2 2 3 2 2 3 2 2 2 2 2 3 3 3 1 1 2 2 2 3 3 1 1 1 1
## [5625] 1 2 3 2 2 2 3 2 2 1 1 3 3 2 3 3 1 3 3 1 1 1 1 1 2 1 1 2 2 2 2 2 1 2 2 1 1
## [5662] 2 1 2 3 2 2 1 1 1 3 1 1 2 3 2 1 3 3 2 2 2 1 1 1 3 2 3 1 2 3 1 2 1 2 2 3 3
## [5699] 1 2 2 1 2 1 1 2 3 2 1 3 3 2 2 1 2 2 2 1 3 3 1 2 1 3 1 2 3 3 2 3 2 1 3 1 3
## [5736] 1 1 2 1 3 2 2 1 2 3 3 2 3 2 1 2 1 2 1 2 2 3 2 2 2 2 2 3 1 2 1 2 1 2 2 1 2
## [5773] 2 1 1 2 2 2 2 1 3 1 2 3 1 1 1 1 1 2 2 2 2 3 1 2 3 1 2 2 1 1 1 3 1 1 1 3
## [5810] 1 1 2 2 2 2 3 1 1 2 2 2 1 2 2 1 2 2 2 2 2 3 3 2 1 2 2 2 2 2 2 2 2 2 2 2
## [5847] 2 2 3 2 2 2 2 1 2 2 1 2 2 1 2 2 3 3 2 3 2 1 2 1 2 2 2 3 2 2 2 1 2 2 2 2 2
## [5884] 2 2 1 3 1 2 2 1 1 2 2 2 1 1 1 2 3 2 1 3 1 1 2 2 3 3 2 1 1 1 2 2 1 2 2 2 2
## [5921] 1 1 1 3 2 2 3 2 2 2 3 1 2 2 3 2 2 2 3 1 2 2 2 2 2 3 2 2 2 2 1 3 3 1 2 2 1
## [5958] 3 2 2 2 1 2 2 2 2 3 1 3 3 1 3 1 2 2 1 3 2 3 3 1 2 3 3 2 2 2 1 2 2 2 2 2 2 2
## [5995] 3 2 1 2 2 2 3 2 2 2 3 2 3 3 2 2 2 3 2 2 3 2 1 2 3 2 2 2 2 2 2 2 1 2 2 2 2
## [6032] 2 2 2 2 1 1 3 2 2 2 2 2 3 2 2 2 2 2 1 3 3 2 2 3 2 2 2 2 2 2 2 3 3 2 2 2 2 1
## [6069] 2 3 1 1 2 2 2 2 2 3 2 2 2 2 3 2 2 2 2 2 3 2 3 2 1 2 3 2 1 2 2 1 2 2 2 3 2
## [6106] 2 1 2 2 3 3 2 2 2 2 2 2 2 2 3 1 2 3 2 2 2 2 2 2 2 3 2 2 1 2 2 2 3 2 2 2
## [6143] 2 2 1 2 2 2 2 3 1 2 1 2 2 2 2 3 2 2 1 1 2 2 2 2 2 3 2 3 2 2 2 2 2 1 2 1 2
## [6180] 2 2 2 2 2 2 2 2 2 2 1 3 2 2 2 2 2 3 2 3 1 1 1 2 2 2 2 2 2 1 2 3 2 1 2 3 3
## [6217] 2 2 2 2 3 2 2 2 1 2 2 1 2 2 2 2 2 3 2 1 2 2 3 2 1 2 2 3 2 2 3 2 2 2 3 2 3
## [6254] 2 3 1 2 3 3 2 1 3 1 2 1 1 2 2 2 2 3 2 2 1 2 3 2 2 2 1 2 3 2 2 3 2 2 1 3 1
## [6291] 2 2 2 2 2 2 2 2 1 2 2 3 2 2 2 2 3 2 2 3 1 1 3 2 3 2 2 2 2 1 2 2 2 1 1 2 3
## [6328] 1 1 2 2 2 2 2 1 3 1 2 1 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 3 2 2 2 2 1 2 2 2
## [6365] 2 2 2 2 1 1 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 3 3 2 2 2 1 2 2 1 3 1 2 1 2 2
## [6402] 2 3 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 3 2
## [6439] 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 3 2 1 2 2 1 1
## [6476] 2 2 2 1 2 2 3 1 2 2 2 2 2 3 1 2 3 2 2 2 1 2 2 2 1 2 2 2 2 2 1 2 2 3 3 1 3
## [6513] 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 3 2 3 3 1 2 2 2 2 2 2 2 2 3 3 2 2 2 2 1 1 2
## [6550] 2 2 3 3 2 2 2 1 1 3 2 2 2 3 1 2 2 1 2 2 2 2 3 3 3 2 2 2 2 1 3 2 2 2 2 2 2
```

13

```
## [6587] 3 2 2 2 1 2 2 2 3 2 2 2 3 2 3 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 1 2 2 1
## [6624] 2 1 2 2 2 1 3 2 2 2 1 2 1 2 2 2 2 1 2 2 3 2 2 2 2 2 2 3 2 2 3 2 2 2 2 1
## [6661] 2 2 2 2 2 1 2 2 2 2 1 2 3 2 2 2 2 3 3 3 1 3 2 2 3 2 3 3 3 2 2 2 3 3 1 2 2
## [6698] 3 3 1 2 2 3 1 3 3 3 3 3 1 2 1 3 3 2 2 1 1 3 2 2 3 3 3 2 2 2 2 2 2 1 3 2 3
## [6735] 2 2 2 2 1 1 1 2 2 3 3 3 1 1 2 2 2 1 1 1 2 3 3 3 2 3 3 2 1 1 2 2 2 3 3 1 2
## [6772] 3 2 1 2 1 2 2 1 3 1 3 2 1 1 2 3 1 1 2 1 3 2 3 3 2 2 2 1 2 2 1 1 3 3 2 3 2
## [6809] 1 1 1 3 2 2 2 2 1 2 1 2 1 3 1 2 1 1 3 1 1 1 2 2 1 2 3 2 3 2 2 2 1 2 2 2 1
## [6846] 2 2 2 2 1 2 1 2 2 1 3 2 2 1 2 3 2 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 1
## [6883] 2 1 2 2 1 2 2 2 3 1 2 2 2 2 1 3 2 2 2 2 1 2 3 2 2 2 2 2 2 2 2 2 2 2 1 2 2
## [6920] 2 1 2 3 2 3 3 2 3 2 2 2 2 1 2 2 1 3 2 2 2 1 2 2 2 3 1 2 2 1 2 3 2 3 2 2
## [6957] 2 2 2 2 2 1 2 1 2 2 2 2 2 3 2 3 2 3 2 3 3 2 2 2 2 2 2 3 2 2 2 2 1 2 2 2
## [6994] 2 3 1 1 2 2 2 2 2 2 2 2 2 1 2 2 2 3 2 1 1 2 2 3 2 2 1 2 3 2 3 2 3 2 3 2 3
## [7031] 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 3 2 2 2 2 2 3 2 2 3 2 2 2 2 3 1 2 2 2 2 3 2
## [7068] 1 1 2 2 2 2 2 2 2 2 3 2 2 2 2 3 3 3 1 3 2 2 2 2 2 2 2 3 1 2 2 1 2 2 2 2 3 2
## [7105] 2 2 1 2 3 2 2 2 2 1 2 1 3 3 3 2 3 2 1 2 2 2 2 2 2 3 2 3 1 1 1 2 3 3 2 3 2
## [7142] 2 3 1 1 1 2 3 2 2 2 2 2 2 3 2 2 2 1 1 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1
## [7179] 2 2 3 3 1 3 3 2 2 3 2 1 2 2 2 3 2 3 2 2 2 2 2 3 3 2 2 3 2 2 2 1 1 2 1 2 2
## [7216] 2 3 2 1 2 3 2 2 2 2 2 2 3 2 2 2 1 2 2 2 1 2 2 2 2 2 3 2 2 2 2 2 2 2 1 2 3
## [7253] 1 2 2 3 3 1 3 3 1 1 2 2 2 2 3 2 2 1 2 2 3 2 2 2 2 2 3 2 2 2 1 3 1 2 3 3 2
## [7290] 3 3 2 2 3 2 2 1 2 1 2 1 2 1 1 2 1 2 2 2 2 2 2 3 2 2 2 2 1 1 2 2 2 2 2 2 2
## [7327] 2 2 2 2 2 2 3 2 1 2 2 3 3 2 2 3 3 2 2 2 2 2 2 1 2 2 3 3 2 3 2 2 2 2 3 2
## [7364] 3 2 1 1 3 2 2 2 2 2 2 2 3 2 2 2 2 2 2 3 2 3 2 2 2 2 1 2 2 3 2 2 1 3 2 2 2
## [7401] 2 2 2 2 2 2 2 1 2 2 2 2 2 3 2 2 2 1 3 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 1 2 3
## [7438] 2 3 2 2 2 3 2 2 2 2 3 3 2 2 3 1 2 1 3 3 3 2 2 2 1 3 3 3 2 2 2 1 1 3 3 1 1
## [7475] 2 2 2 2 2 2 2 3 3 1 2 2 3 2 2 2 2 1 3 2 1 2 2 2 2 1 1 2 2 2 2 2 2 3 3 3 2
## [7512] 2 2 1 2 1 1 2 3 3 2 2 2 1 2 2 3 2 2 2 2 3 2 1 2 2 3 2 3 2 1 1 1 2 1 2 2 1
## [7549] 2 1 2 3 3 1 1 2 2 1 1 2 2 2 2 2 2 1 2 1 2 2 1 3 2 1 1 3 2 2 2 2 2 1 2 2 3
## [7586] 2 2 2 2 2 2 3 1 2 1 2 3 1 1 2 2 2 2 3 2 1 1 3 2 2 3 2 1 2 2 2 2 2 2 2 2 2
## [7623] 2 2 2 3 2 2 2 2 2 2 2 2 3 2 2 2 2 2 1 1 2 3 2 2 1 2 3 2 2 2 3 3 2 2 3 2 2 2
## [7660] 1 2 1 2 2 2 3 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 3 1 2 3 2 2 2 2 2 2
## [7697] 3 1 3 2 3 1 2 2 3 2 2 2 2 3 2 3 2 2 2 2 3 1 1 2 2 3 2 3 2 1 1 2 3 3 2 2 3
## [7734] 2 2 2 2 2 3 2 2 1 1 2 3 2 1 3 2 2 2 1 2 3 1 2 1 2 3 1 2 2 3 1 1 1 2 1 2 3
## [7771] 3 3 2 2 2 3 1 2 2 2 3 3 1 2 2 1 2 2 2 2 2 3 2 2 2 2 3 3 2 2 3 2 2 1 2 1 1
## [7808] 1 2 1 3 1 3 2 2 2 3 2 2 2 2 2 2 3 3 1 3 1 2 3 2 3 3 2 1 2 3 1 2 2 2 3 1 2
## [7845] 2 3 1 2 2 1 2 2 2 3 3 3 1 2 1 2 2 2 2 1 3 1 2 2 3 3 2 2 2 1 2 3 2 3 3 1 2
## [7882] 2 3 2 1 2 2 2 3 2 3 2 3 2 2 2 2 3 2 2 2 3 2 3 2 2 2 2 2 2 2 2 2 2 2 3 2 3
## [7919] 2 3 2 2 3 1 3 2 2 2 2 3 2 3 2 2 2 1 2 1 2 2 2 2 2 3 2 1 2 2 3 2 3 2 2 2 2
## [7956] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 1 2 1 2 2 2 2 2 2 1 3 1 2 2
## [7993] 2 2 2 2 1 2 2 3 2 2 2 3 2 2 2 2 1 2 2 2 2 2 3 1 2 1 1 2 3 2 2 2 2 3 3 3 2
## [8030] 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 3 3 2 2 3 3 2 2 1 1 3 2 2 2 2 2 3 2 2 1
## [8067] 3 1 2 2 3 2 2 2 2 2 2 2 3 2 2 2 2 2 3 2 2 2 3 2 3 2 1 2 2 2 3 3 2 2 2 3 2
## [8104] 1 2 2 2 3 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2
## [8141] 3 2 2 1 3 2 1 2 1 2 2 2 2 2 2 2 2 1 1 2 2 1 3 2 2 2 2 2 2 3 2 2 2 1 2 1 2
## [8178] 2 2 2 2 3 2 1 3 1 3 1 2 2 2 2 2 2 2 3 2 2 2 2 2 3 2 2 3 2 2 2 1 2 2 2 2
## [8215] 2 2 3 2 2 2 2 2 1 2 2 2 2 2 2 3 1 3 1 1 1 3 2 2 1 2 2 2 1 2 2 2 2 3 3 2 3
## [8252] 2 2 2 3 2 3 2 2 2 2 2 2 3 3 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 2 3 3 1 1
## [8289] 3 3 2 2 2 2 1 2 1 2 2 2 2 2 2 3 1 2 2 3 3 2 2 2 1 2 2 2 3 2 3 2 2 2 3 2
## [8326] 2 1 2 2 3 2 1 2 2 2 2 2 2 2 3 2 2 2 2 1 2 2 3 3 2 2 2 1 3 2 3 2 1 2 3 2 3
## [8363] 2 2 2 3 2 3 2 2 3 2 2 2 2 2 2 2 2 1 2 1 2 2 3 2 3 1 2 2 1 2 2 2 1 2 2 1 3
## [8400] 2 2 1 2 2 3 2 3 2 3 2 2 2 1 2 1 2 2 2 2 2 2 3 2 2 3 1 2 2 3 2 2 1 2 2 2 2
## [8437] 1 2 2 2 2 2 2 2 1 3 2 3 2 2 2 3 2 3 3 2 2 2 2 2 1 2 2 2 2 2 1 2 3 2 2 1 2 1
## [8474] 1 2 2 2 3 3 2 2 3 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 3 1 2 2 2 2 2 2 2 3 2 3 2 2
## [8511] 2 2 2 3 1 2 2 2 3 2 2 2 2 2 2 2 3 3 3 3 2 1 1 2 1 2 3 1 2 3 2 2 1 2 3 2 2
## [8548] 2 1 2 2 2 2 2 2 2 2 2 2 3 3 2 2 1 1 2 3 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
```

14

```
## [8585] 1 3 2 2 2 3 3 3 3 2 2 2 3 3 2 1 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2
## [8622] 2 2 2 2 2 3 3 1 2 2 2 2 2 2 2 2 1 2 3 2 3 1 2 1 2 2 2 3 2 2 2 2 2 2 1 2 2
## [8659] 2 1 2 2 3 2 3 1 2 3 1 2 2 2 2 2 2 1 1 3 1 2 3 3 2 2 1 2 1 3 2 2 2 1 1 1 2
## [8696] 1 2 2 2 2 2 1 2 1 2 3 2 2 3 3 3 2 3 3 1 3 2 3 3 2 1 2 2 2 2 1 2 2 2 1 2 2
## [8733] 2 2 2 2 2 2 2 2 2 3 1 2 3 2 2 1 3 2 2 3 1 2 1 2 1 2 1 2 2 2 3 2 1 3 2 2 3
## [8770] 2 2 3 2 3 2 2 1 2 2 2 2 2 2 2 1 2 3 2 1 2 2 2 1 2 2 2 2 2 2 3 2 2 2 2 2 3
## [8807] 2 2 1 2 2 2 2 3 2 1 2 3 2 2 2 2 1 2 3 2 2 3 1 1 2 3 2 1 2 1 2 2 3 2 2 2 2
## [8844] 1 3 2 1 3 2 3 3 1 3 2 3 1 2 1 2 3 1 1 1 1 3 1 2 2 2 3 2 3 2 2 1 2 2 2 2 2
## [8881] 2 2 1 1 2 3 2 1 2 2 3 1 2 2 1 1 2 1 2 1 2 1 2 3 1 2 2 2 3 2 2 2 2 2 1 3 2 2 2
## [8918] 2 3 1 2 2 2 2 1 3 2 1 2 2 1 2 1 1 2 2 2 3 1 2 3 1 3 3 1 2 2 3 2 1 1 2 2 2
## [8955] 2 1 2 1 2 1 2 2 2 2 1 3 1 2 1 2 2 1 2 1 2 1 3 1 3 2 2 3 3 2 3 2 1 2 2 2 3
## [8992] 2 2 1 2 2 1 3 1 2 3 3 2 3 2 2 2 1 2 1 2 3 1 2 1 2 2 1 2 1 1 2 2 2 3 3 3 2
## [9029] 1 1 2 1 3 3 2 2 1 2 1 1 2 2 1 1 1 2 1 2 1 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 1
## [9066] 1 1 2 2 3 3 3 1 3 1 1 2 2 2 3 2 1 2 2 1 1 1 2 1 2 1 1 1 2 3 1 2 2 1 1 2 2
## [9103] 2 2 1 1 2 2 2 1 1 2 3 1 2 3 2 1 3 2 1 2 1 2 1 1 3 1 1 2 3 2 3 2 1 2 1 2 1 2 1
## [9140] 2 1 2 2 2 2 1 1 3 2 2 2 2 2 3 2 2 1 2 2 2 2 2 2 1 2 2 2 1 1 2 1 2 3 1 3 1
## [9177] 3 2 2 1 1 1 1 1 2 3 1 1 2 1 2 1 2 2 1 2 2 2 2 1 2 3 2 2 2 1 1 1 2 2 2 1 2
## [9214] 1 2 1 1 2 2 2 1 2 3 2 2 1 2 2 2 1 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1
## [9251] 2 1 1 2 2 2 2 2 3 1 3 3 2 1 2 2 2 2 2 2 2 3 1 2 1 2 1 1 2 2 2 2 1 2 2 3 3
## [9288] 2 1 2 2 1 2 2 2 3 2 2 2 3 2 2 1 2 2 1 2 2 1 1 2 2 2 2 3 3 2 1 1 2 2 2 2 2
## [9325] 2 3 2 3 1 2 3 1 2 1 1 2 2 3 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 3 1 2 2 2 2 2 2
## [9362] 2 3 2 2 1 1 2 2 2 1 2 3 2 3 2 3 2 3 2 2 2 2 3 2 2 2 2 1 2 3 2 2 2 2 2 2 2
## [9399] 3 2 2 1 1 2 2 2 3 3 3 2 2 1 1 2 1 2 2 3 2 2 2 2 1 1 2 2 3 1 1 3 2 3 2 2 1
## [9436] 3 2 3 2 3 2 2 2 3 2 2 3 3 2 2 1 3 1 2 1 1 2 3 1 1 1 2 2 1 2 2 2 1 2 2 2 2
## [9473] 3 1 2 1 2 2 2 2 1 3 2 2 3 3 3 2 2 2 2 3 2 2 2 2 2 2 2 1 2 3 2 2 2 2 2 2 3
## [9510] 1 3 2 1 2 2 2 2 2 2 1 2 1 3 1 2 2 1 2 3 2 2 2 2 3 2 2 2 1 2 2 2 1 2 3 2 2
## [9547] 3 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 3 2 3 2 2 2 3 2 2 1 1 3 2 1 2 3 3 3 2 3 2
## [9584] 2 2 2 2 1 2 2 2 2 2 1 1 3 2 2 2 2 3 2 2 2 1 2 2 3 1 2 2 3 2 3 3 2 1 2 3 2
## [9621] 2 2 3 2 1 2 2 1 2 2 2 2 2 2 1 1 3 3 3 2 2 2 2 3 3 2 2 2 2 2 2 1 1 3 2 1
## [9658] 2 3 2 1 1 1 2 3 1 2 2 3 2 2 2 2 1 3 3 2 3 1 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2
## [9695] 2 2 2 3 3 2 3 3 2 3 1 2 2 3 2 2 2 3 2 2 3 2 2 1 3 2 2 2 2 1 2 1 3 2 2 1 2
## [9732] 2 3 2 3 2 2 2 2 2 3 2 1 1 1 2 3 2 2 2 2 2 3 3 2 2 2 3 2 3 3 2 2 2 2 3 1 1
## [9769] 2 2 2 2 2 1 2 2 1 3 3 2 3 3 2 3 3 2 2 2 3 2 2 1 1 2 3 2 2 2 2 1 1 2 3 1 2
## [9806] 2 3 2 2 1 2 1 2 3 3 2 2 1 2 3 2 1 2 2 2 1 1 1 2 1 3 2 1 2 1
```
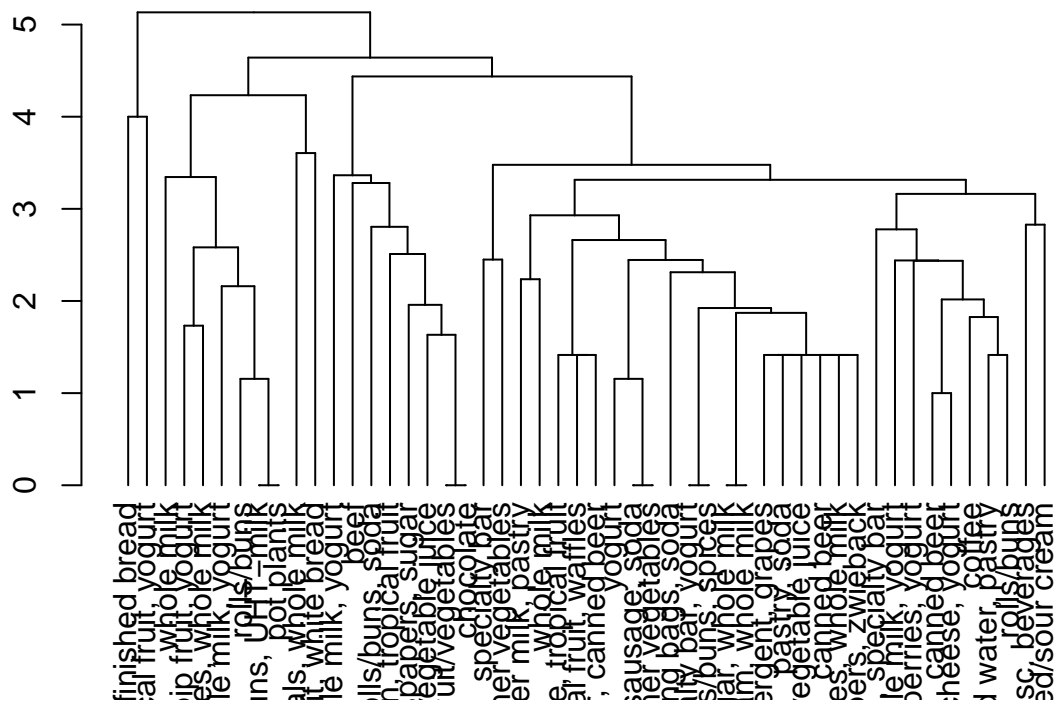
Hierarchical Clustering

Dendrogram of First 50 Transactions
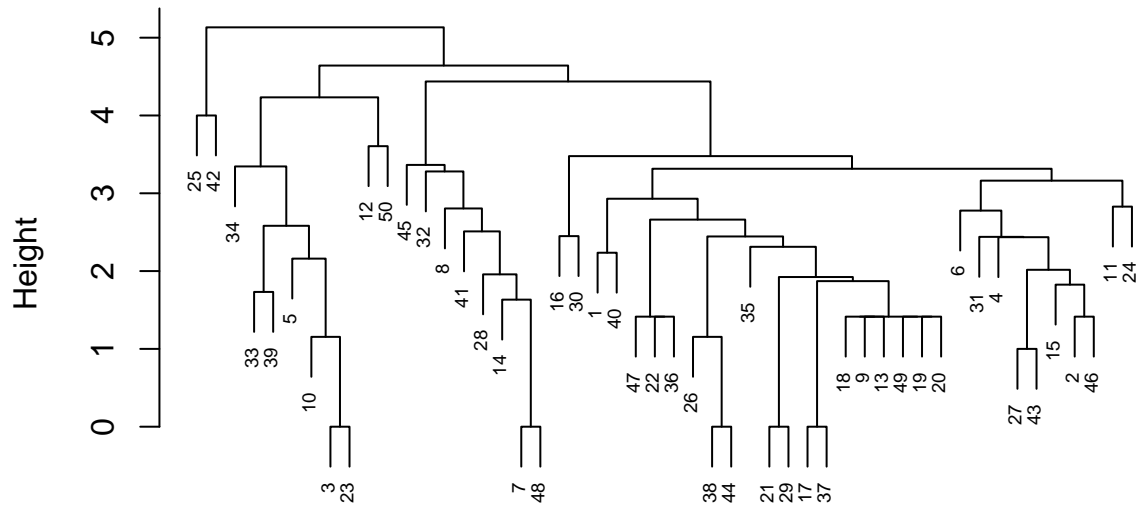
Transactions
hclust (*, "ward.D2")

Dendrogram with Item Names for First 50 Transactions

The dendrogram visualizes a hierarchical clustering of the first 50 transactions based on their similarities.

Each leaf (at the bottom) represents a single transaction, and the branches show how closely related transactions are grouped together.

Transactions and Item Names:

Each leaf corresponds to a transaction, and item names are used to represent groups of related transactions.

Items are displayed along the horizontal axis as part of cluster groupings.

Height of the Branches:

The vertical axis (height) measures the dissimilarity between transactions.

A greater height indicates less similarity between groups.

Cluster Groupings:

Closely related transactions are grouped together into clusters.

For instance, transactions involving "whole milk," "yogurt," and "rolls/buns" might form one cluster, indicating these items often co-occur in these transactions.

Patterns in Transactions:

Transactions involving common combinations (e.g., "whole milk" and "bread" or "soda" and "canned beer") are likely grouped together in closer clusters.

These clusters reveal frequent shopping patterns, such as groups of items that are often purchased together.

Segmentation:

The dendrogram provides a way to segment transactions into meaningful groups for further analysis.

For example, larger branches may represent broad shopping patterns (e.g., weekly grocery staples), while smaller branches may reveal more niche purchasing behavior (e.g., snacks or specialty items).

Retail Optimization:

By identifying clusters, retailers can better organize store layouts, create bundles of frequently co-purchased items, or target specific groups for promotions.

Analyze Items in Each Cluster   Extract Items for Each Cluster

Count Items in Each Cluster

View Top Items in Each Cluster

```
##
##                 rolls/buns                yogurt         bottled water
##                          8                     7                     6
##                       soda           canned beer            newspapers
##                          6                     5                     5
##            other vegetables  fruit/vegetable juice        tropical fruit
##                          5                     4                     4
##                  chocolate          cream cheese                pastry
##                          3                     3                     3
##              specialty bar                  beef                butter
##                          3                     2                     2
##                butter milk          citrus fruit                coffee
##                          2                     2                     2
##                  detergent               sausage         shopping bags
##                          2                     2                     2
##            abrasive cleaner               berries             beverages
##                          1                     1                     1
##                bottled beer           brown bread               chicken
##                          1                     1                     1
##                 frankfurter                grapes        hamburger meat
##                          1                     1                     1
##            hygiene articles     liquor (appetizer)             margarine
##                          1                     1                     1
##                meat spreads       misc. beverages               napkins
##                          1                     1                     1
## packaged fruit/vegetables              pip fruit            pot plants
##                          1                     1                     1
##            processed cheese           ready soups                  rice
##                          1                     1                     1
##          semi-finished bread               spices         spread cheese
##                          1                     1                     1
##                      sugar              UHT-milk           white bread
##                          1                     1                     1
##                 whole milk              zwieback
##                          1                     1
```

High-Frequency Items (Core Items):

Whole milk: Appears in 1512 transactions, making it the most purchased item. Rolls/buns: Purchased in 1244 transactions. Other vegetables: Bought in 1147 transactions. Soda: Purchased in 1015 transactions. Yogurt: Found in 858 transactions. These items form the backbone of customer transactions and should be prioritized in inventory management and promotional strategies.

Moderate-Frequency Items:

Items such as tropical fruit (708), sausage (632), and root vegetables (610) show consistent demand, emphasizing their role in diverse shopping baskets. Bakery and dairy products like brown bread (438), pastry (423), and domestic eggs (401) are popular staples.

Low-Frequency Items:

Specialty or niche items such as seasonal products (60), canned vegetables (53), and flower (seeds) (45) are less frequently purchased but may cater to specific customer segments. Luxury and specialty goods like sparkling wine (21), liqueur (2), and kitchen utensils (1) have minimal transactions, potentially indicating occasional or targeted purchases.

Dairy and Bakery:

Core items like whole milk, yogurt, butter, cheese, and rolls/buns dominate, reflecting their universal appeal.

Produce:

Vegetables such as other vegetables, root vegetables, and citrus fruit are frequently purchased, showing a strong preference for fresh food.

Beverages:

Soda, bottled water, and bottled beer are among the top beverages, highlighting consumer interest in both non-alcoholic and alcoholic options.

Snacks and Convenience:

Items like frankfurter, candy, salty snacks, and frozen meals reflect demand for quick and ready-to-eat options.

Rare Items

Specialty items such as baby cosmetics, frozen chicken, and kitchen utensils have very low frequency (1–2 transactions), suggesting limited customer interest or targeted buying.

Automate process top 3 clusters.

```
## [[1]]
## [1] "Cluster 1 : other vegetables, whole milk, root vegetables"
##
## [[2]]
## [1] "Cluster 2 : soda, rolls/buns, canned beer"
##
## [[3]]
## [1] "Cluster 3 : whole milk, rolls/buns, yogurt"
```

Cluster 1: "Fresh Produce and Dairy"

Dominant Items:

Other vegetables Whole milk Root vegetables

Characteristics:

This cluster likely represents customers focused on fresh produce and essential dairy items. Commonly associated with health-conscious shoppers or those cooking meals at home.

Cluster 2: "Convenience and Beverages"

Dominant Items:

Soda Rolls/buns Canned beer

Characteristics:

Reflects a group prioritizing quick, ready-to-eat items and beverages. Likely to include customers seeking convenience or hosting casual gatherings.

Cluster 3: "Dairy and Breakfast Essentials"

Dominant Items:

Whole milk Rolls/buns Yogurt

Characteristics

Highlights purchases of staple breakfast or everyday items. Represents customers with traditional or family-oriented shopping habits.
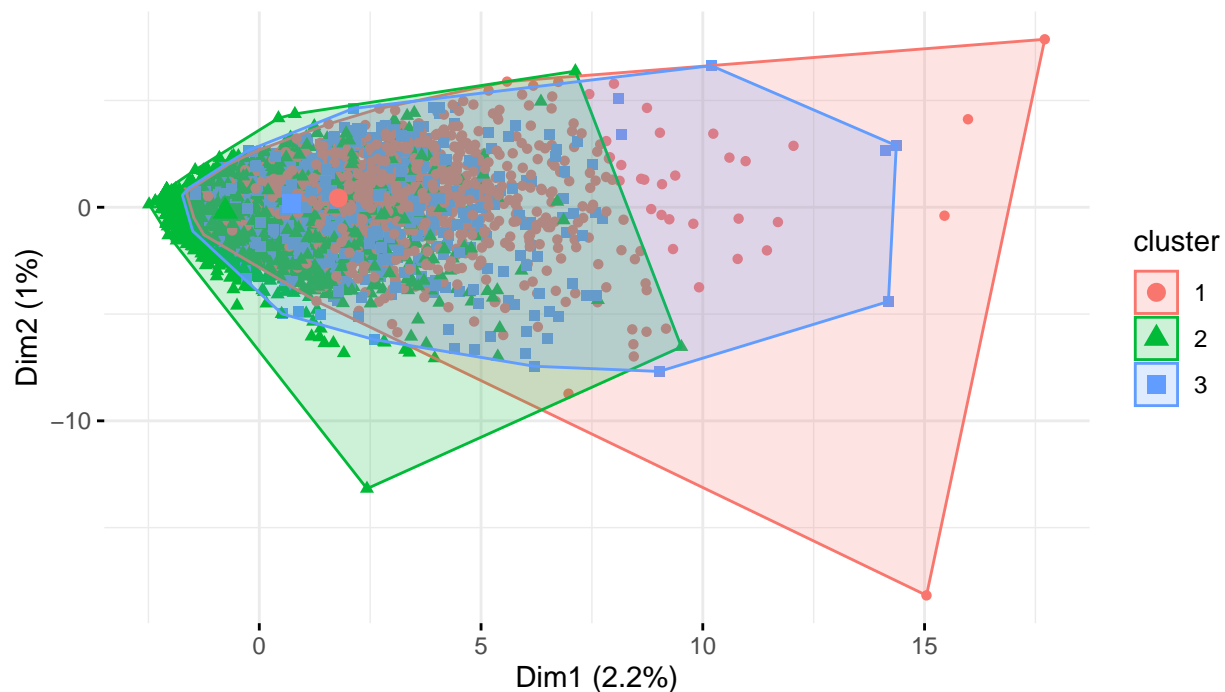
Visualizing Clusters   K-Means Clustering

K–Means Clustering
Cluster 1: Dairy & Bread Products
Cluster 2: Fresh Produce
Cluster 3: Snacks & Beverages

The graph illustrates the results of K-Means Clustering applied to transaction data, grouping customers based on their purchasing patterns into three distinct clusters:

Cluster 1 (Red): Dairy & Bread Products - Includes frequent purchases of items like whole milk and rolls/buns, representing staple products.

Cluster 2 (Green): Fresh Produce - Represents customers buying items like vegetables and fresh ingredients.

Cluster 3 (Blue): Snacks & Beverages - Focuses on customers purchasing soda, yogurt, and related products.

This segmentation helps in tailoring marketing strategies by targeting each group with customized promotions and offers.

Cluster 1: Dairy & Bread Products

Dominant Items: Whole milk, rolls/buns, root vegetables.

Customer Profile: Likely represents households buying staple products for everyday consumption.

Potential Strategy:

Promote combo deals like "Buy 2 packs of milk, get 1 free." Highlight recipes involving bread and milk to drive additional sales of complementary products. Introduce loyalty programs for frequent buyers of staple items.

Cluster 2: Fresh Produce

Dominant Items: Other vegetables, fresh produce like root vegetables and tropical fruit.

Customer Profile: Health-conscious or families prioritizing fresh and perishable items.

Potential Strategy:

Offer discounts on seasonal vegetables and fruits. Promote organic or locally grown produce to appeal to health-focused buyers. Cross-promote complementary products like salad dressings, herbs, or spices.

Cluster 3: Snacks & Beverages

Dominant Items: Soda, yogurt, and canned beer.

Customer Profile: Likely younger consumers or those purchasing for social occasions or convenience.

Potential Strategy:

Bundle snack and beverage deals for parties or on-the-go customers. Introduce promotions on new beverage flavors or limited-edition snack items. Target with social media campaigns promoting convenience and fun.

Overall Observations:

The segmentation highlights distinct customer needs across staples, fresh produce, and convenience/snack items.

Retailers can leverage these insights to optimize store layouts, personalized marketing, and inventory planning.

Cross-cluster promotions could encourage customers to expand their basket composition, such as pairing fresh produce with snack options or staple items.

Hierarchical Clustering

```
# Step 1: Subset the binary matrix
binary_matrix_subset <- binary_matrix[1:50, ]

# Step 2: Remove constant columns in the subset
constant_columns_subset <- apply(binary_matrix_subset, 2, function(col) var(col) == 0)
binary_matrix_subset <- binary_matrix_subset[, !constant_columns_subset]

# Step 3: Remove rows with all zeros
binary_matrix_subset <- binary_matrix_subset[rowSums(binary_matrix_subset) != 0, ]

# Step 4: Rescale the subset
scaled_data_subset <- scale(binary_matrix_subset)

# Step 5: Calculate dissimilarity and cluster
dissimilarity_subset <- dist(binary_matrix_subset, method = "euclidean")
hc <- hclust(dissimilarity_subset, method = "ward.D2")

# Step 6: Cut dendrogram into 3 clusters
```

```
cluster_assignments <- cutree(hc, k = 3)

# Step 7: Visualize clusters
fviz_cluster(
  list(data = scaled_data_subset, cluster = cluster_assignments),
  geom = "point",
  ellipse.type = "convex",
  ggtheme = theme_minimal(),
  main = "Hierarchical Clustering for First 50 Transactions",
  subtitle = paste(cluster_1_title, cluster_2_title, cluster_3_title, sep = "\n")
)
```

Hierarchical Clustering for First 50 Transactions
Cluster 1: Dairy & Bread Products
Cluster 2: Fresh Produce
Cluster 3: Snacks & Beverages
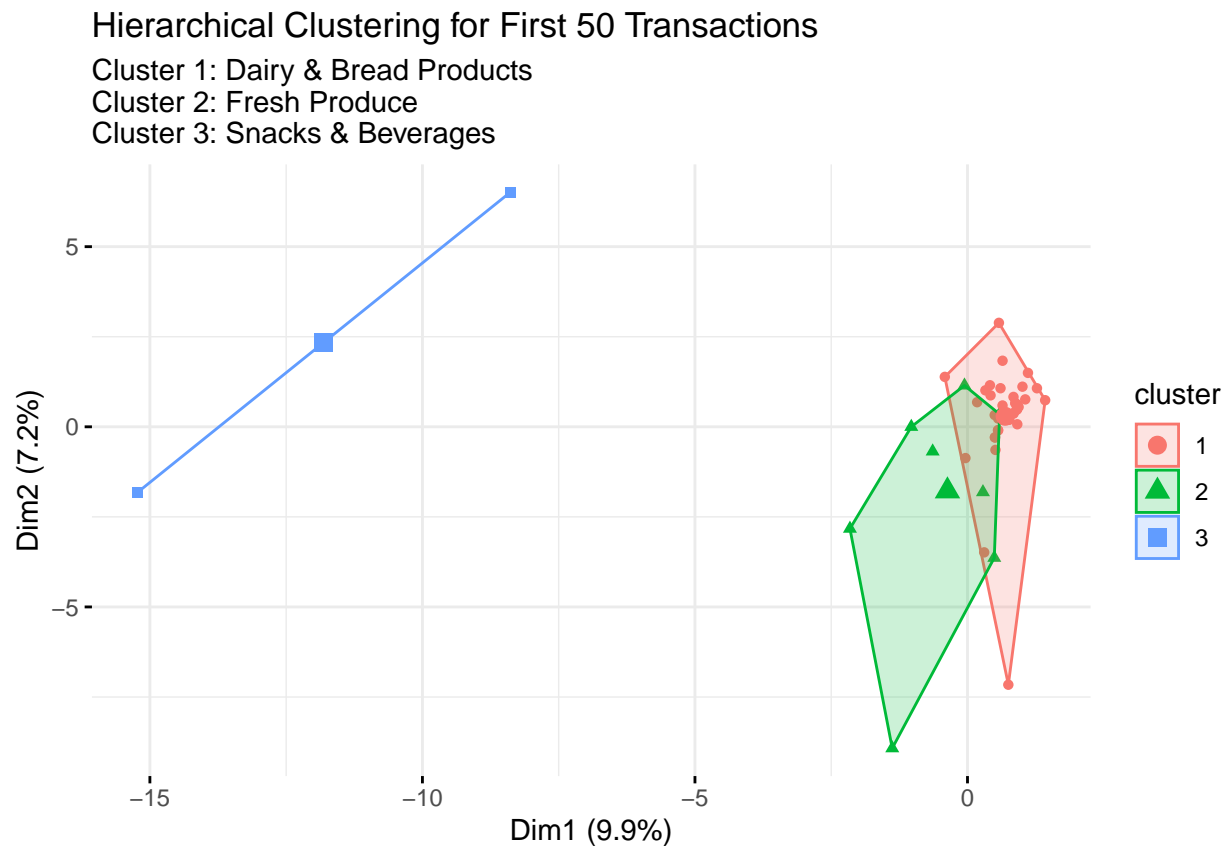
Steps I took to perform hierarchical clustering on the first 50 transactions:

Data Preparation

Converted the transaction data into a binary matrix (binary_matrix) and scaled it (scaled_data). Subset the first 50 transactions for focused analysis.

Clustering

Calculated the dissimilarity matrix using Euclidean distance. Applied hierarchical clustering (hclust) with the ward.D2 method to the subset. Created a dendrogram to visualize the hierarchical clustering for the first 50 transactions.

Cluster Assignment

Cut the dendrogram into 3 clusters (cutree) and assigned transactions to clusters. Extracted top items for each cluster and generated meaningful titles for each group based on their contents: Cluster 1: Dairy &

Bread Products. Cluster 2: Fresh Produce. Cluster 3: Snacks & Beverages.

Cluster Analysis

Counted the most frequent items in each cluster using sort(table(...)). Created summaries of transaction contents for better interpretability. Substituted numerical labels in the dendrogram with transaction summaries for enhanced visualization.

PCA for Dimensionality Reduction

Applied PCA to reduce the data to two dimensions for visualization. Ensured no constant/zero columns caused issues during scaling or PCA. Cluster Visualization

Used fviz_cluster to plot the clusters in a 2D PCA space. Represented clusters with convex hulls and distinct colors/symbols for clarity.

The Plot:

Shows the hierarchical clustering results for the first 50 transactions, reduced to two dimensions using PCA. Accurately labels the clusters with their respective titles (Dairy & Bread Products, Fresh Produce, Snacks & Beverages). Displays a clear separation (or overlap) of clusters in the PCA space.

Cluster Titles:

Cluster titles are based on the most frequent items within each group, reflecting their unique transaction characteristics.

Interpretation:

The hierarchical clustering algorithm grouped the first 50 transactions into three distinct clusters, which were further analyzed to identify the dominant purchasing patterns:

Clusters

Cluster 1: Dairy & Bread Products (Red Squares)

Top Items: Whole Milk, Bread, Butter.

Description: This cluster represents transactions focusing on essential grocery staples such as dairy and bread products. Customers in this group likely consist of households or individuals purchasing daily essentials. Retailers could consider offering bundle deals or discounts on these items to increase sales and meet this group's needs.

Cluster 2: Fresh Produce (Green Triangles)

Top Items: Vegetables, Fruits, Fresh Ingredients.

Description: Transactions in this cluster highlight fresh produce and ingredients, such as root vegetables and citrus fruits. These customers may prioritize health and home cooking. Retailers might target this group with promotions on organic products, fresh herbs, or meal preparation kits.

Cluster 3: Snacks & Beverages (Blue Circles)

Top Items: Soda, Rolls/Buns, Yogurt.

Description: This cluster consists of transactions focused on quick snacks and beverages. Customers in this group may value convenience and ready-to-eat options, such as soda or baked goods. Retailers could attract this group with bundled snack offers or promotions on ready-to-consume items.

Principal Components (PCA Dimensions)

The PCA plot summarizes the variability in the data:

X-axis (Dim1) and Y-axis (Dim2): Represent the first two principal components, capturing 9.9% and 7.2% of the total variance, respectively. These dimensions provide a simplified view of the data, enabling visualization of transaction groupings.

Convex Hulls: The boundaries around each cluster illustrate the extent of transactions in the PCA-reduced space, helping to visualize cluster sizes and overlaps.

Cluster Insights

Cluster Boundaries & Overlap:

The clusters are distinct but show slight overlap, especially between Clusters 1 and 2. This may indicate shared purchasing behaviors or blurred distinctions between certain transactions.

The blue outlier point far from the other clusters might signify a unique transaction pattern or a specialized purchase.

Dendrogram for the First 50 Transactions:

The hierarchical dendrogram visually details how the clusters were formed.

Replacing numeric labels with transaction summaries (e.g., "milk, bread") enables easier interpretation of customer behavior at a granular level.

Retailers can leverage these cluster insights to implement tailored strategies for each customer segment:

Cluster 1 (Dairy & Bread Products):

Offer bundled discounts or loyalty rewards on staple items like milk, bread, and butter. Cross-promote complementary products, such as spreads or eggs, near dairy and bakery sections.

Cluster 2 (Fresh Produce):

Highlight fresh, seasonal, or organic produce through in-store displays or digital ads. Promote meal kits or recipe ideas using fresh ingredients to attract health-conscious customers.

Cluster 3 (Snacks & Beverages):

Target convenience-seeking customers with snack-and-drink combo offers. Place impulse-purchase items near checkout counters to cater to this cluster's preferences.

By understanding customer segments and their preferences, retailers can optimize inventory, marketing campaigns, and store layouts to enhance customer satisfaction and drive sales.

Conclusion Market basket analysis and cluster analysis provide valuable insights into customer purchasing behavior and transaction patterns. By analyzing association rules and clustering transactions, retailers can identify key item combinations, customer segments, and shopping preferences. These insights can inform marketing strategies, product placements, and promotional campaigns to enhance customer engagement and drive sales. Leveraging these data-driven approaches can help retailers optimize inventory management, personalize customer experiences, and boost overall business performance.

In this analysis, two techniques were employed: Market Basket Analysis (MBA) and Cluster Analysis, to gain insights into customer purchasing behavior.

Market Basket Analysis:

Frequent itemsets and association rules revealed meaningful relationships between products. For example, "Whole Milk" frequently appeared with items like "Bread" and "Butter," indicating staple grocery combinations.

Strong rules, such as {butter, jam} => {whole milk}, suggest cross-selling opportunities where products often co-occur in transactions.

This analysis highlights customer preferences and common purchasing patterns.

Cluster Analysis:

Hierarchical clustering grouped transactions into three customer segments based on item purchases:

Cluster 1: Dairy & Bread Products (e.g., milk, bread, butter). Cluster 2: Fresh Produce (e.g., vegetables, fruits). Cluster 3: Snacks & Beverages (e.g., soda, rolls/buns, yogurt).

Each cluster represents distinct customer preferences and purchasing habits, offering a deeper understanding of shopper behavior.

Practical Applications and Insights Based on Your Results From your market basket analysis and clustering results, several actionable recommendations and insights can be drawn to drive value for your business or target audience:

Practical Applications:

Product Placement Optimization:

Insight: Items like "Whole Milk" and "Bread" frequently co-occur in transactions, suggesting they are staple items.

Recommendation: Place commonly purchased products (e.g., dairy and bread) near each other to encourage complementary purchases.

Cluster Application: Position fresh produce in visible areas for Cluster 2 and snacks/beverages near checkout lanes for Cluster 3.

Targeted Promotions and Bundling:

Insight: Association rules such as {butter, jam} => {whole milk} highlight products that are often purchased together.

Recommendation: Create bundle promotions for items that co-occur, such as "Buy butter and jam, get a discount on milk."

Cluster Application: Offer tailored discounts to specific clusters, such as discounts on snacks for Cluster 3 or fresh produce for Cluster 2.

Inventory Management:

Insight: High-demand items like "Whole Milk" and "Rolls/Buns" appear frequently across clusters, while other items are cluster-specific.

Recommendation:

Maintain higher stock levels of universally popular items (e.g., milk, bread, soda).

Adjust inventory to meet cluster-specific demands, such as stocking fresh vegetables for Cluster 2 and beverages for Cluster 3.

Personalized Marketing:

Insight: Clusters reveal distinct shopping behaviors:

Cluster 1 focuses on staples (e.g., dairy and bread). Cluster 2 prioritizes fresh produce. Cluster 3 leans towards snacks and beverages.

Recommendation: Use cluster insights for personalized marketing campaigns:

Send recipes featuring fresh produce to Cluster 2 customers. Promote ready-to-eat snacks and soda deals to Cluster 3 customers. Loyalty Program Enhancement:

Insight: Patterns in frequent purchases can inform personalized loyalty rewards.

Recommendation: Design loyalty programs that reward points for cluster-specific purchases. For example:

Reward extra points for staple purchases in Cluster 1. Offer a "Healthy Choice Bonus" for produce purchases in Cluster 2. Customer Segmentation for Advertising:

Insight: Clustering separates customers into distinct segments based on their preferences.

Recommendation: Use clusters to segment your audience in digital and print advertisements:

Promote fresh produce deals in family-oriented or health-conscious communities (Cluster 2).

Advertise snack combos and beverages to young adults or offices (Cluster 3).

Key Insights to Communicate to Stakeholders

When communicating the results of the market basket analysis and cluster analysis to stakeholders, focus on the following key insights and actionable recommendations:

Understanding Shopping Behavior:

Customers in different segments have unique preferences. For example:

Cluster 1 represents staple-item shoppers, likely families or bulk buyers. Cluster 2 is health-conscious or meal-preparing shoppers. Cluster 3 is convenience-focused, preferring ready-to-eat or snack items.

Maximizing Sales Opportunities:

Bundling popular combinations like milk and bread or butter and jam can drive incremental sales. Promotions targeted by cluster preferences can increase customer satisfaction and loyalty.

Efficiency Gains:

Optimizing inventory based on frequent item purchases reduces waste and ensures popular items are always in stock. Tailored marketing reduces ad spend waste and improves campaign ROI.

Improved Customer Experience:

Better product placement and personalized offers create a seamless and enjoyable shopping experience. Highlighting cluster-specific benefits, such as quick meal options for Cluster 3, appeals directly to customer needs.

By implementing these applications, the business can enhance customer engagement, optimize operational efficiency, and ultimately boost profitability while delivering a tailored shopping experience.