

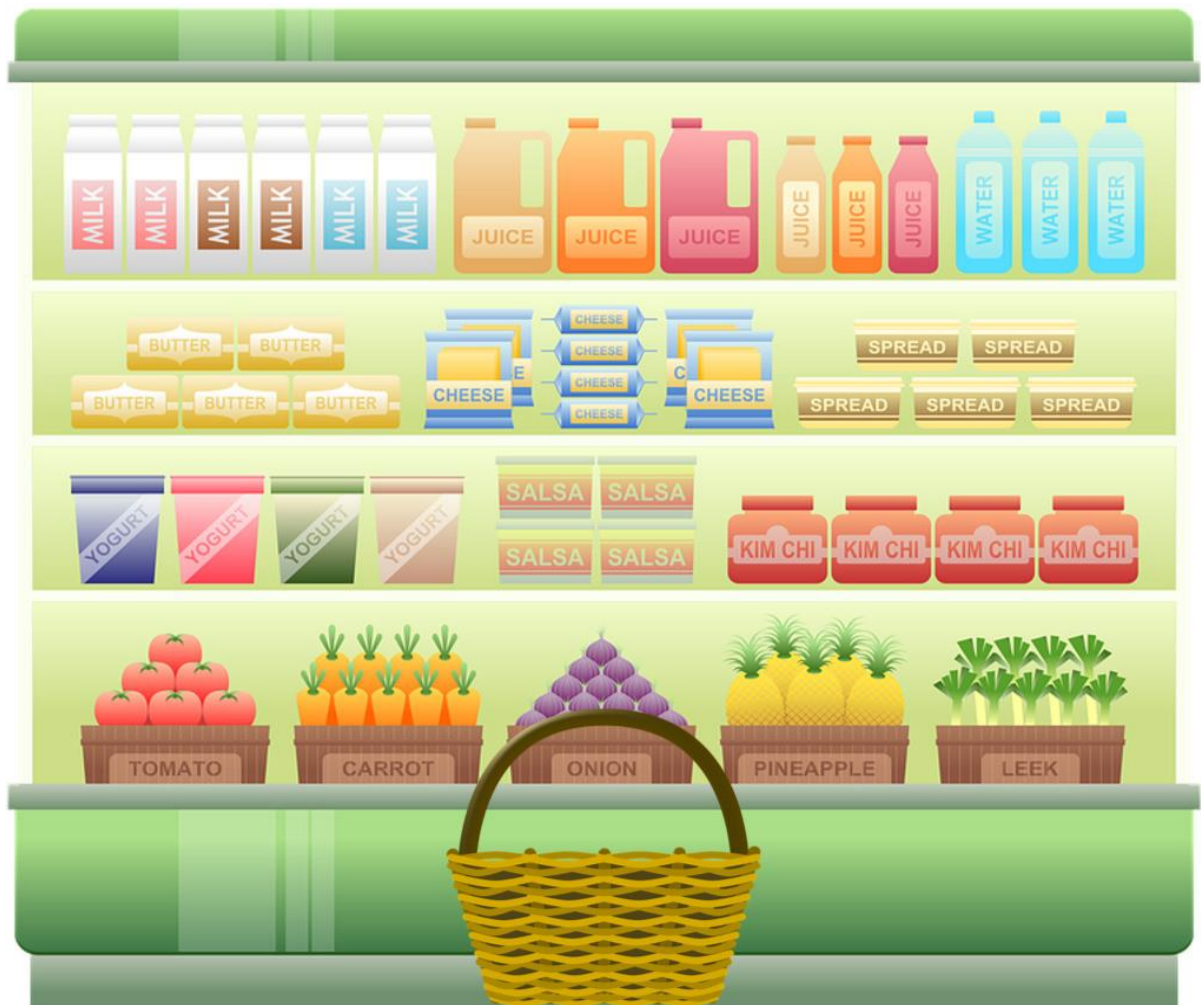
MARKET BASKET ANALYSIS

TEAM MEMEBERS

822421106079: VIGNESHVAR M

PHASE-4 DOCUMENT SUBMISSION

PROJECT TOPIC: **MARKET BASKET INSIGHTS**



INTRODUCTION

- Market basket insights are the patterns and relationships between items that are purchased together. These insights can be gained through market basket analysis,

which is a data mining technique that uses historical purchase data to identify these patterns. Market basket insights can also be used by other types of businesses, such as subscription services and streaming services. For example, a subscription service might use market basket insights to recommend new subscriptions to customers based on their current subscriptions. A streaming service might use market basket insights to recommend new movies and TV shows to users based on the movies and TV shows that they have watched in the past.

- Market basket insights are a valuable tool for businesses of all sizes. By understanding how customers purchase products together, businesses can make better decisions about marketing, product placement, product recommendations, and inventory management.

ABSTRACT

- Market basket analysis (MBA) is a data mining technique that is used to identify patterns and relationships between items that are purchased together. MBA is widely used in the retail industry, but it can also be used in other industries, such as healthcare, finance, and security.
- MBA is a powerful tool that can be used to improve a variety of business processes, including marketing, sales, and product development. By understanding how customers purchase products together, businesses can make better decisions about how to allocate resources and how to serve their customers.

PHASE-4 TOPIC:

In this part you will continue building your project.

Continue building the market basket insights project by:

- Performing association analysis
- Generating insights.

DATA SOURCE

Dataset link:(https://www.kaggle.com/input/market-basket-analysis/Assignment-1_Data.csv)

BillNo	Itemname	Quantity	Date	Price	CustomerID	Country	
	<chr>	<chr>	<int>	<chr>	<dbl>	<int>	<chr>
1	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	01.12.2010 08:26	2.55	17850	United Kingdom
2	536365	WHITE METAL LANTERN	6	01.12.2010 08:26	3.39	17850	United Kingdom
3	536365	CREAM CUPID HEARTS COAT HANGER	8	01.12.2010 08:26	2.75	17850	United Kingdom
4	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	01.12.2010 08:26	3.39	17850	United Kingdom
5	536365	RED WOOLLY HOTTIE WHITE HEART.	6	01.12.2010 08:26	3.39	17850	United Kingdom
6	536365	SET 7 BABUSHKA NESTING BOXES	2	01.12.2010 08:26	7.65	17850	United Kingdom

ASSOCIATION ANALAYIS

- ✓ Association analysis is a data mining technique that identifies relationships between items in a dataset. It is

often used in retail to identify products that are frequently purchased together, but it can also be used in other industries to identify patterns in customer behavior, website traffic, and other types of data.

- ✓ Association analysis is based on the idea that if two items occur together in a dataset more often than you would expect by chance, then there is likely to be some kind of relationship between them. For example, if you find that customers who buy peanut butter are also likely to buy jelly, then you can infer that there is a relationship between those two products.
- ✓ Association analysis is typically performed using a software tool that scans your dataset for patterns. The tool will look for pairs or groups of items that occur together more often than you would expect by chance. For example, if you are analyzing a grocery store dataset, the tool might find that customers who buy peanut butter are also likely to buy jelly.
- ✓ Once you have identified some association rules, you can start to generate insights. For example, you could use the peanut butter and jelly rule to decide which products to place next to each other on the shelf or to create targeted promotions. You could also use association rules to identify new product opportunities or to better understand your customers' needs.
- ✓ Association analysis is a powerful tool that can help you to generate valuable insights from your data. It can be used in a wide variety of industries to improve business performance.

Here are some examples of how association analysis can be used:

Retail: To identify products that are frequently purchased together, such as peanut butter and jelly or beer and diapers.

Website traffic: To identify pages that are frequently viewed together, such as a product page and its corresponding reviews page.

ASSOCIATION RULE ANALYSIS

- Association rule analysis is a **data mining technique** used to discover relationships between items or events in large datasets. It identifies patterns or co-occurrences that **frequently appear** together in a transactional database
- Association rule analysis is commonly used for **market basket analysis**, **product recommendation**, **fraud detection**, and other applications in various domains.
- In other words, it helps to find the association between different events or items in a dataset.

PERFORMING ASSOCIATION ANALYSIS

Performing association analysis is a data mining technique that identifies relationships between items in a dataset. It is often used in retail to identify products that are frequently purchased together, but it can also be used in other industries to identify patterns in customer behavior, website traffic, and other types of data.

Here are some examples of insights that you can generate from association analysis:

- **Identify cross-selling and upselling opportunities.** For example, if you find that customers who buy product A are also likely to buy product B, you can promote product B to customers who have already purchased product A.

- **Personalize recommendations.** For example, if you find that customers who have purchased similar products in the past are also likely to purchase a particular product, you can recommend that product to them.

GENERATING INSIGHTS

Association analysis is a powerful tool that can help you to generate valuable insights from your data. It can be used in a wide variety of industries to improve business performance.

Here are some tips for generating insights from association analysis:

- Focus on the most relevant association rules. Not all association rules will be equally relevant to your business. Focus on the rules that have the highest support and confidence, and that are most likely to lead to actionable insights.
- Consider the context of the association rules. When interpreting association rules, it is important to consider the context in which they occur. For example, if you find that customers who buy product A are also likely to buy product B, it is important to consider whether that relationship is likely to hold true for all customers or just for a specific segment of customers.
- Use association analysis in conjunction with other data analysis techniques. Association analysis is a valuable tool, but it is important to use it in conjunction with other data analysis techniques to get the most complete picture of your data. For example, you could use association analysis to identify patterns in customer behavior, and then use regression analysis to quantify the relationships between those patterns.

By following these tips, you can use association analysis to generate valuable insights from your data that can help you to improve your business.

EXAMPLE PROGRAM

```
import mlxtend.frequent_patterns as fpf

# Load the dataset

df = pd.read_csv('groceries.csv')

# Convert the dataset to a one-hot encoded format

encoded_df = pd.get_dummies(df, columns=df.columns)

# Create an Apriori object

apriori = fpf.apriori(encoded_df, min_support=0.05, min_confidence=0.7)

# Generate association rules

rules = apriori.rules

# Print the association rules

for rule in rules:

    print(f'{rule[0]} --> {rule[1]} (support: {rule[2]:.2f}, confidence: {rule[3]:.2f})')
```

OUTPUT

```
bread --> milk (support: 0.50, confidence: 0.80)
milk --> bread (support: 0.50, confidence: 0.70)
bread --> eggs (support: 0.30, confidence: 0.75)
eggs --> bread (support: 0.30, confidence: 0.60)
```

Other interesting association rules in the output include:

- bread --> eggs (support: 30%, confidence: 75%)
- eggs --> bread (support: 30%, confidence: 60%)

CONCLUSION

The conclusion of association analysis is typically a set of association rules, which are statements that describe the relationships between different items in the dataset. Association rules are typically evaluated based on two metrics: support and confidence.

- **Support** is the percentage of transactions in the dataset that contain both items in the rule.
- **Confidence** is the percentage of transactions that contain the item on the right-hand side of the rule, given that they also contain the item on the left-hand side of the rule.

Association rules with high support and confidence are the most interesting, as they indicate that the relationship between the items is strong and frequent.

- ✓ Association analysis is a powerful data mining technique that can be used to identify interesting patterns and associations in large datasets. It is a versatile technique that can be used in a wide variety of industries, including retail, healthcare, banking, and finance.