

Market basket insights

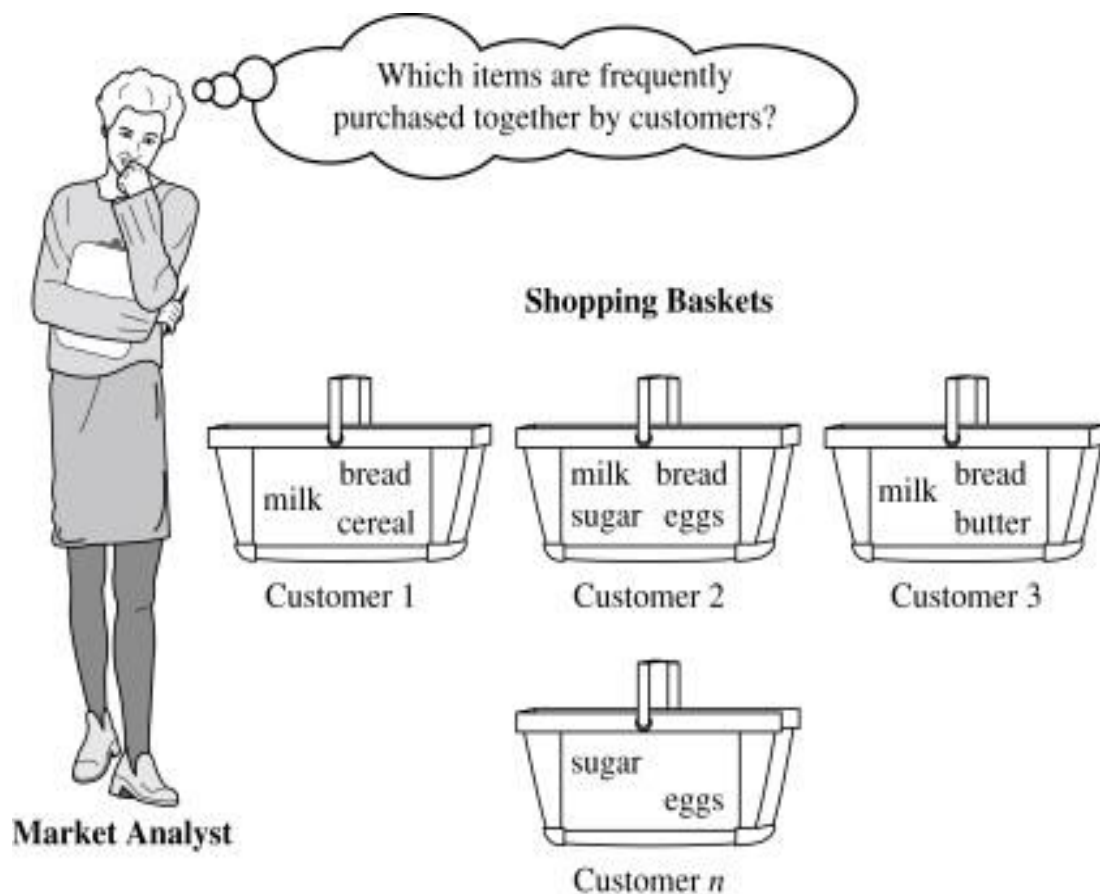
Phase 5: *submission document*

Project title : **market basket insights**

Phase 5: **project documentation and submission**

Topic

In this section you will document the complete project and prepare it for submission.



Market basket insights

Introduction:

Market Basket Analysis (MBA) is a technique used in data mining and business analytics. It explores the relationships between items that are frequently purchased together.

The analysis helps businesses understand customer behavior, improve sales strategies, and optimize product placement.

MBA relies on the concept of association rules, which are if-then statements that uncover patterns in transactional data. These rules are expressed in the form of “if item A is bought, then item B is likely to be bought as well.”

This information is valuable for businesses to make informed decisions about inventory management, product recommendations, and targeted marketing efforts.

Dataset

Link: <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

Here is a list of commonly used tools and software for conducting Market basket analysis:

Market basket analysis is a technique used in retail and e-commerce to understand which products are frequently purchased together. Here is a list of commonly used tools and software for conducting market basket analysis:

- ✓ **Microsoft Excel:** It can be used for basic market basket analysis, especially for smaller datasets.
- ✓ **R:** A programming language with various libraries (e.g., `arules`) that provides extensive support for market basket analysis.
- ✓ **Python:** Using libraries like `mlxtend`, `apyori`, or `Orange3` for implementing market basket analysis.
- ✓ **Tableau:** A data visualization tool that can be used for visualizing and exploring market basket analysis results.
- ✓ **IBM SPSS Modeler:** Offers comprehensive data mining and predictive analytics capabilities, including market basket analysis.
- ✓ **RapidMiner:** A data science platform that includes tools for various data mining tasks, including market basket analysis.

- ✓ **Weka:** A collection of machine learning algorithms and tools, including market basket analysis, with a graphical user interface.
- ✓ **KNIME:** An open-source data analytics platform that includes nodes for performing market basket analysis.
- ✓ **SAS Enterprise Miner:** Part of the SAS suite, it provides advanced analytics capabilities, including market basket analysis.
- ✓ **Orange:** An open-source data visualization and analysis tool, which includes a module for association rule mining (used in market basket analysis).
- ✓ **.Alteryx:** A data preparation and analytics platform that supports market basket analysis.
- ✓ **Google Analytics:** Can be used for basic market basket analysis for online retail platforms.
- ✓ **Apache Spark MLlib:** For large-scale market basket analysis on big data.
- ✓ **Hadoop:** When combined with tools like Apache Hive or Apache Pig, it can be used for market basket analysis on big data.

1.DESIGN THINKING AND PRESENT IN FORM

OF DOCUMENT

❖Introduction:

- Briefly introduce Design Thinking and its relevance in solving complex business problems.
- Explain the purpose of applying Design Thinking to Market Basket Analysis.

❖Empathize:

- Understand the perspectives of stakeholders, including customers, store managers, and data analysts.
- Conduct interviews, surveys, and observations to gather insights on shopping behaviors.

❖Define:

- Clearly define the problem or objective for the Market Basket Analysis.
- Specify the business goals, such as optimizing product placement, cross-selling, or improving customer experience.

❖Ideate:

- Brainstorm potential approaches and techniques for conducting effective Market Basket Analysis.

- Encourage creative thinking among team members to generate innovative solutions.

❖ **Prototype:**

- Develop a sample dataset and implement the chosen Market Basket Analysis algorithm (e.g., Apriori, FP-Growth).
- Visualize the results to demonstrate how associations between products are identified.

❖ **Test:**

- Evaluate the prototype results using historical transaction data.
- Analyze the associations and assess their relevance and potential business impact.

❖ **Implement:**

- Discuss how the insights from Market Basket Analysis can be applied in a real-world retail setting.
- Provide recommendations for adjusting product placement, promotions, or marketing strategies.

❖ **Reflection:**

- Reflect on the process, considering what worked well and areas for improvement.

- Address any challenges encountered during the application of Design Thinking in Market Basket Analysis.

❖ **Visuals:**

- Include charts, graphs, and visualizations to illustrate key findings and insights.
- Use diagrams to explain the Market Basket Analysis process and results.

❖ **Case Studies or Examples:**

- Include practical examples or case studies showcasing successful implementations of Market Basket Analysis using Design Thinking.

2.DESIGN INTO INNOVATION

Designing innovative solutions for market basket insights involves leveraging advanced techniques and technologies to extract deeper, more actionable information from customer transaction data. Here's a step-by-step approach:

1.Data Collection and Preprocessing:

- Gather transactional data. Each row should represent a unique transaction, and the columns should represent different items/products.

- Preprocess the data, ensuring it's in a format suitable for analysis (e.g., a list of items for each transaction).

2. Install Required Libraries:

You'll need to install and import libraries like `pandas` for data manipulation and `mlxtend` for the Market Basket Analysis.

```
```python
import pandas as pd

from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori,
association_rules
```
```

3. Data Integration:

- If you have data from multiple sources, you may need to integrate them into a single dataset using techniques like merging or joining.

4. Perform Market Basket Analysis:

- Transform the data into a suitable format for the analysis. This typically involves creating a one-hot

encoded matrix where each row represents a transaction and each column represents an item.

```
```python
Example of data transformation
te = TransactionEncoder()
te_ary = te.fit(transactions).transform(transactions)
df = pd.DataFrame(te_ary, columns=te.columns_)
```
```

- Apply the Apriori algorithm to find frequent itemsets.

```
```python
frequent_itemsets = apriori(df, min_support=0.01,
use_colnames=True)
```
```

- Generate association rules based on the frequent itemsets.

```
```python
rules = association_rules(frequent_itemsets,
metric="lift", min_threshold=1)
```
```

- Analyze the results and extract insights.

5. Interpret and Utilize Insights:

- The resulting rules will contain information about associations between items, along with metrics like support, confidence, and lift. These metrics can help you understand which items tend to be bought together and the strength of these associations.

3. Build loading and preprocessing the dataset for market basket insights

To build a dataset for market basket insights, you'll need to follow these steps:

➤ Data Collection:

- Obtain transactional data from a retail source.
This data should include information about

individual purchases, such as products bought and the transaction ID.

➤ **Data Cleaning:**

- Remove any duplicates or irrelevant information from the dataset.
- Handle missing values if any.

➤ **Data Exploration:**

- Explore the dataset to understand its structure and characteristics. This helps in making informed decisions during preprocessing.

➤ **Transaction Conversion:**

- Convert the data into a suitable format for market basket analysis. Usually, this involves creating a binary matrix where rows represent transactions and columns represent products. Each cell indicates whether a product was present in a given transaction.

➤ **Frequent Itemset Generation:**

- Use algorithms like Apriori or FP-Growth to identify frequent itemsets. These are combinations of products that occur frequently together.

➤ **Association Rule Mining:**

- Extract association rules from the frequent itemsets. These rules provide insights into which products are commonly bought together.
- **Filtering and Interpretation:**
 - Apply thresholds (like support, confidence, and lift) to filter out uninteresting or less relevant rules.
 - Interpret the remaining rules to gain actionable insights.
- **Data Visualization:**
 - Create visualizations (e.g., heatmaps, network diagrams) to present the findings in a clear and understandable manner.

Program:

In association rule mining for Market Basket Analysis (MBA), the input is typically a transaction dataset, where each row represents a transaction, and each column represents a product or item. The dataset should be structured in a way that allows the algorithm to identify which items were purchased together in each transaction.

Here's an example of a simple transaction dataset:

...

TransactionID Items

| | |
|---|---------------------------------------|
| 1 | Bread, Milk |
| 2 | 2 Bread, Diapers, Eggs |
| 3 | Milk, Diapers, Beer |
| 4 | 4 Bread, Milk, Diapers, Butter |
| 5 | Bread, Milk, Diapers |
| 6 | ... |

In this dataset, each row represents a transaction, and the 'Items' column lists the items bought in each transaction.

The output of association rule mining in MBA includes:

1. Frequent Itemsets:

These are sets of items that frequently occur together in transactions. For example, if the algorithm determines that "Bread" and "Milk" are frequently bought together, it forms a frequent itemset.

2. Association Rules:

These rules express relationships between different sets of items. They have the form $X \Rightarrow Y$, where X and Y are sets of items. The rule implies that if X is bought, then Y is also likely to be bought.

For example, if the algorithm discovers the rule `{Bread, Milk} => {Diapers}`, it means that customers who buy Bread and Milk are likely to also buy Diapers.

Each rule typically comes with metrics like support, confidence, and lift, which provide quantitative measures of the rule's significance and strength.

- **Support:** Measures how frequently the rule occurs in the dataset.
- **Confidence:** Indicates how often the rule has been found to be true.
- **Lift:** Shows how much more likely the items are bought together compared to being bought independently.

The rules are usually filtered based on these metrics to identify the most interesting and actionable ones.

So, after running association rule mining on the example dataset above, you might obtain output like:

...

Frequent Itemsets:

| | Support | itemsets |
|---|---------|------------|
| 0 | 0.8 | {Bread} |
| 1 | 1 | 0.6 {Milk} |

| | | |
|---|-----|----------------------------|
| 2 | 0.6 | {Diapers} |
| 3 | 3 | 0.4 {Eggs} |
| 4 | 0.2 | {Butter} |
| 5 | 5 | 0.4 {Beer} |
| 6 | 0.4 | {Bread, Diapers} |
| 7 | 7 | 0.4 {Bread, Milk} |
| 8 | 0.2 | {Milk, Diapers} |
| 9 | 9 | 0.2 {Bread, Milk, Diapers} |

Association Rules:

| | Antecedents | consequents | support | confidence | lift |
|-----|-------------|-------------|---------|------------|------|
| 0 | {Bread} | {Milk} | 0.4 | 0.5 | 1.25 |
| 1 | {Milk} | {Bread} | 0.4 | 0.67 | 1.25 |
| 2 | {Diapers} | {Bread} | 0.4 | 0.67 | 1.0 |
| ... | | | | | |
| ... | | | | | |

These are simplified examples, and in a real-world scenario, you would have a more extensive dataset with a larger number of items and transactions, resulting in more complex and potentially more valuable association rules.

Data visualisation :

Certainly! Below is an example Python program using Matplotlib to create a bar chart visualizing sales data for milk, butter, and bread.

```
```python
```

```
Install matplotlib using pip if you haven't already
```

```
!pip install matplotlib
```

```
import matplotlib.pyplot as plt
```

```
Sample data
```

```
products = ['Milk', 'Butter', 'Bread']
```

```
sales = [500, 300, 700]
```

```
Create a bar chart
```

```
plt.figure(figsize=(8, 6))
```

```
plt.bar(products, sales, color=['blue', 'green', 'red'])
```

```
plt.xlabel('Products')
```

```
plt.ylabel('Sales (in units)')
```



```
plt.title('Sales Data for Milk, Butter, and Bread')
plt.show()
...
```

In this example:

1. We import the ``matplotlib.pyplot`` module.
2. We define sample data: ``products`` contains the names of the products, and ``sales`` contains their corresponding sales figures.
3. We create a bar chart using ``plt.bar()`` with the products on the x-axis and sales on the y-axis. Each product is assigned a different color for better visualization.
4. We add labels to the x-axis (``xlabel``), y-axis (``ylabel``), and a title (``title``) to provide context to the chart.
5. Finally, we display the chart using ``plt.show()``.

When you run this program, it will generate a bar chart showing the sales data for milk, butter, and bread.

You can customize the colors, labels, and other aspects of the chart to suit your specific needs. Remember that this is a basic example, and Matplotlib provides a wide range

of options for creating more complex and detailed visualizations.

#### **4.Performing different activities like feature engineering,Model training, evaluation etc...**

Here's a step-by-step guide for performing activities like feature engineering, model training, and evaluation for market basket insights.

- **Feature Engineering:**
- **Transaction Level Features:**
  - Total number of items in a transaction.
  - Total cost of items in a transaction.
  - Time of day or day of the week of the transaction.
- **-Product Level Features:**
  - Frequency of each product being bought.
  - Average basket size when the product is bought.
  - Association strength of the product with other items (support, confidence, lift).
- **-Customer Level Features:**
  - Customer segmentation based on buying behavior.
  - Customer loyalty metrics.
- **Model Selection:**

- Common models for market basket analysis include Apriori, FP-Growth, and more advanced techniques like Neural Collaborative Filtering (for larger datasets).
- **Model Training:**
  - Train the chosen model on the preprocessed dataset.
  - Use a training-validation split to tune hyperparameters if applicable.
- **Evaluation:**
  - For market basket analysis, common metrics include:
    - **\*\*Support\*\***: The proportion of transactions that contain the itemset.
    - **\*\*Confidence\*\***: The probability of buying item B given that item A was bought.
    - **\*\*Lift\*\***: The ratio of the observed support to the expected support, given that the items are independent.
- **Model Interpretation:**
  - Understand the rules generated by the model.
  - Identify the most influential features and products.
- **Validation and Testing:**

- Use a holdout dataset or cross-validation for final evaluation.
- **Deployment:**
  - Implement the model in your production environment for real-time predictions.

## **Advantages of market basket insights**

Market basket insights offer several advantages for businesses in the retail and e-commerce sectors:

- ✓ **Cross-Selling Opportunities:**
  - Identify complementary products that are often bought together. This allows for strategic placement and promotion of these items.
- ✓ **Customer Segmentation:**
  - Understand different customer segments based on their purchasing behavior. This enables personalized marketing strategies.
- ✓ **Inventory Management:**
  - Optimize stock levels by knowing which products are frequently bought together. This reduces overstocking and stockouts.
- ✓ **Promotion Effectiveness:**

- Evaluate the impact of promotions or discounts on product sales. Understand which promotions drive higher basket sizes.

✓ **Optimized Store Layouts:**

- Arrange products in physical stores or online platforms to encourage cross-purchases and increase overall sales.

✓ **Price Optimization:**

- Set prices strategically for products that are commonly bought together to increase overall revenue.

✓ **Customer Experience Enhancement:**

- Offer product recommendations based on previous purchasing behavior, enhancing the customer's shopping experience.

✓ **Loss Prevention:**

- Detect unusual buying patterns or behaviors that might indicate fraudulent activity.

✓ **Marketing Campaign Efficiency:**

- Tailor marketing campaigns to specific customer segments to increase the likelihood of conversion.

✓ **Market Trend Analysis:**

- Identify emerging trends in customer preferences and adapt product offerings accordingly.
- ✓ **Cost Reduction:**
  - Reduce marketing expenses by targeting promotions more effectively based on insights gained from market basket analysis.
- ✓ **Competitive Advantage:**
  - Stay ahead of competitors by leveraging data-driven insights to enhance product offerings and customer experience.



- ✓ Overall, market basket insights provide valuable information that can lead to increased revenue, improved customer satisfaction, and more efficient business operations.

## **Disadvantages of market basket insights**

Certainly, here are some disadvantages of market basket insights:

### **❖ Limited Understanding of Customer Motivations:**

- Market basket analysis shows which items are frequently bought together, but it doesn't explain why these combinations occur. Understanding the underlying motivations or preferences of customers may require additional research.

### **❖ Lack of Context:**

- It focuses solely on transactional data, ignoring external factors such as promotions, seasonality, or specific events that may influence purchasing decisions.

### **❖ Overlooks Individual Preferences:**

- Market basket analysis treats all customers as a group, potentially missing out on the unique preferences and behaviors of individual customers.

### **❖ Static and Historical Perspective:**

- It provides insights based on past transactions and may not capture shifts or emerging trends in consumer behavior. It doesn't predict future behavior.

❖ **May Lead to Misinterpretation:**

- Misinterpreting the results of market basket analysis can lead to incorrect conclusions and potentially misguided business decisions.

❖ **Dependency on Data Quality:**

- Accurate and comprehensive transaction data is crucial for meaningful market basket insights. Inaccurate or incomplete data can lead to unreliable results.

❖ **Not Applicable to All Industries:**

- Market basket analysis is more relevant to industries where customers make multiple product choices in a single transaction, such as retail. It may not be as useful in industries with different purchasing dynamics.
- It often doesn't consider external factors like marketing efforts, economic conditions,



or competitor strategies, which can significantly impact buying behavior.

❖ **No Causality:**

- Market basket analysis identifies associations but doesn't establish causality. It doesn't explain why certain products are bought together.

❖ **May Result in Outdated Insights:**

- If not regularly updated with current data, market basket insights may become less relevant over time, especially in fast-changing markets.

**Application of market basket analysis**

Market basket analysis has a wide range of applications across various industries. Here are some common examples:

➤ **Retail and E-Commerce:**

- Recommender Systems: Online retailers use market basket analysis to suggest related or complementary products to customers.
- Inventory Management: Helps optimize stock levels and plan product assortments based on popular product combinations.

- Store Layout and Merchandising: Guides the placement of products in physical stores to encourage cross-category purchases.
- **Supermarkets and Grocery Stores:**
  - Promotions and Discounts: Helps design effective promotions and discounts on frequently purchased items or combinations to attract shoppers.
- **Online Services and Content Platforms:**
  - Content Recommendations: Streaming platforms use market basket analysis to suggest related movies, shows, or music based on user preferences.
- **Food and Beverage Industry:**
  - Menu Optimization: Restaurants and cafes can use market basket analysis to create optimized menu combinations and specials.
- **Telecommunications:**
  - Bundling Services: Helps determine which services (internet, TV, phone) are frequently subscribed to together, allowing for bundled offerings.
- **Healthcare:**

- Drug Prescriptions: Pharmacies can analyze prescriptions to identify commonly prescribed drugs and recommend related healthcare products.

➤ **Digital Marketing:**

- Cross-Channel Marketing: Helps identify which channels (email, social media, website) are frequently engaged with together, allowing for integrated marketing campaigns.

➤ **Financial Services:**

- Banking Products: Helps identify patterns in customers' use of banking services, allowing for targeted offers or bundling of financial products.

➤ **Travel and Hospitality:**

- Package Deals: Hotels and travel agencies can offer bundled packages based on popular combinations of accommodations, flights, and activities.

➤ **Gaming and Entertainment:**

- In-Game Purchases: Analyzes which virtual items or features are frequently bought

together within a game, informing pricing strategies and promotional offers.

➤ **Subscription Services:**

- Subscription Bundles: Analyzes subscriber behavior to create tailored bundles or offers that cater to specific customer segmentation .

**Conclusion :**

- In conclusion, market basket insights offer businesses a powerful lens through which to understand and optimize customer behavior. By delving into the patterns of products frequently purchased together, businesses can uncover a wealth of strategic opportunities.
- These insights pave the way for cross-selling, streamlined inventory management, and personalized marketing initiatives. They also guide businesses in creating a more intuitive and appealing shopping experience for customers.
- Furthermore, market basket insights equip businesses with the knowledge needed to make informed decisions. From product assortment to pricing strategies, these insights foster data-driven

approaches that lead to enhanced customer satisfaction and increased revenue.

- In essence, market basket insights empower businesses to align their offerings with customer preferences, ultimately fortifying their competitive edge in a dynamic market landscape. By placing the customer at the center of their strategies, businesses can forge deeper connections, drive growth, and flourish in an ever-evolving business environment.