NAANMUTHALVAN

ARTIFICIAL INTELLIGENCE

PROJECT TITLE

MARKET BASKET INSIGHTS

PHASE 5

Problem Statement

The problem at hand is to perform market basket analysis to uncover hidden relationships and patterns in customer purchasing behavior. Specifically, the goal is to answer questions like:

- **Frequent Item Sets:** Identify what products are frequently purchased together, allowing businesses to optimize product placement and promotion strategies.

- **Association Rules:** Discover association rules, such as "if a customer buys product A, they are likely to buy product B," enabling cross-selling and targeted marketing efforts.

- **Recommendation Systems:** Develop recommendation systems that can suggest additional products to customers based on their current selections, increasing sales and customer satisfaction.

- **Inventory Management:** Optimize inventory and supply chain management by understanding which products tend to be bought together, ensuring efficient stocking and reducing out-of-stock situations.

- **Customer Segmentation:** Segment customers based on their purchasing patterns, enabling personalized marketing campaigns and customer-specific product recommendations.

- **Basket Size and Value Analysis:** Analyze the average size and value of customer baskets to gain insights into customer spending habits.

Data:

To tackle this problem, we need access to historical transaction data, which should include information such as:

- Date and time of each transaction
- Customer ID or some form of customer identifier
- List of products purchased in each transaction

Methods and Tools:

Several methods and tools can be employed to solve this problem, including:

- **Apriori Algorithm:** This is a classic algorithm for mining association rules and frequent item sets.
- **FP-growth Algorithm:** An alternative to Apriori for efficient frequent item set mining.
- **Machine Learning:** Advanced machine learning techniques can be used for recommendation systems and customer segmentation.
- **Data Visualization:** Tools like data visualization libraries and dashboards can be used to present the insights in an understandable manner.

Expected Outcome:

The expected outcome of this analysis is to provide actionable insights to the business, which can be used to enhance product offerings, marketing strategies, and overall operational efficiency. The insights should enable the business to increase revenue, improve customer satisfaction, and optimize their product inventory.

Success Metrics:

The success of the market basket analysis can be measured using various metrics, including:

- Increase in cross-sell and up-sell revenue
- Improved customer retention and satisfaction
- Reduction in out-of-stock instances
- Enhanced inventory turnover and management.

Design Thinking

Design thinking is a human-centered approach to problem-solving and innovation. It can be applied to the process of gaining market basket insights to ensure that the analysis is not only data-driven but also user-centric and actionable. Here's a design thinking process tailored to market basket insights:

1. Empathize: Understand Customer Needs

- Start by empathizing with the customers and stakeholders. Engage in conversations, surveys, and interviews to gain a deep understanding of their needs, pain points, and expectations when it comes to shopping and product recommendations.

2. Define: Clearly Articulate the Problem

- Synthesize the information gathered in the empathize phase to define the problem. Clearly articulate the goals and challenges related to market basket insights. For example, is it to increase sales, improve customer experience, or optimize inventory?

3. Ideate: Generate Insights and Solutions

- Brainstorm potential solutions for gaining market basket insights. This could include data analysis methods, algorithms, or recommendation systems. Encourage creative thinking and consider a variety of approaches to tackle the problem.

4. Prototype: Create Visual Representations

- Develop prototypes or visual representations of the proposed solutions. This might include mock-ups of user interfaces for presenting insights, sample data analysis results, or sketches of how recommendation systems might work.

5. Test: Gather Feedback and Validate

- Test the prototypes with real users or stakeholders. Collect feedback on the proposed solutions and iterate on them based on the input. Ensure that the insights generated are actionable and align with the defined problem.

6. Implement: Build the Solution

- Once a viable solution is identified and validated, proceed to implement it. This may involve developing data analysis models, recommendation systems, or visualization tools to extract and present market basket insights.

7. Learn: Continuously Improve

- Post-implementation, monitor the effectiveness of the solution in generating market basket insights. Gather user feedback and data on the impact of the insights on sales, customer satisfaction, and other relevant metrics. Use this information to make continuous improvements.

8. Iterate: Refine and Evolve

- The design thinking process is iterative. As you gather more insights and data, continue to refine and evolve your solutions. Be open to changes and improvements as the market and customer needs evolve.

9. Communicate: Share Insights and Actions

- It's essential to communicate the insights and actions derived from the market basket analysis with the relevant stakeholders, including marketing teams, inventory managers, and executives. Ensure that the insights are effectively communicated and acted upon.

Phases Of Development

The development of market basket insights involves several phases, from data collection and preprocessing to the implementation of analytical techniques and the presentation of actionable insights. Here are the typical phases in the development of market basket insights:

- **Data Collection:**
    - Gather historical transaction data from various sources, including point-of-sale systems, e-commerce platforms, or other relevant data repositories.
    - Collect data on customer transactions, including transaction timestamps, customer ID s, and lists of products purchased in each transaction.

- **Data Preprocessing:**
    - Clean and preprocess the data to handle missing values, duplicate entries, and outliers.
    - Transform the data into a suitable format for analysis, such as a transaction-item matrix.

- **Exploratory Data Analysis (EDA):**
    - Conduct EDA to gain an initial understanding of the dataset.
    - Explore summary statistics, visualize product frequencies, and identify patterns in the data.

- **Market Basket Analysis Techniques:**
    - Implement market basket analysis techniques, such as Apriori or FP-growth algorithms, to identify frequent item sets and association rules.
    - Determine key parameters, such as support and confidence levels, to define what is considered significant.

- **Association Rule Generation:**
    - Generate association rules that indicate the relationships between items/products in customer transactions.
    - Evaluate the strength and relevance of these rules based on support and confidence metrics.

- **Recommendation Systems:**
    - If building recommendation systems is part of the goal, develop collaborative filtering or content-based recommendation models.
    - Utilize the association rules and historical transaction data to make personalized product recommendations.

- **Customer Segmentation:**
    - Segment customers based on their purchasing behavior.
    - Cluster customers into groups with similar preferences and behaviors, allowing for targeted marketing efforts.

- **Visualization and Reporting:**
    - Create data visualizations, such as heat maps, network diagrams, or bar charts, to present market basket insights in an understandable and actionable format.
    - Generate reports or dashboards for stakeholders to access and interpret the results.

- **Actionable Insights:**
  - Extract actionable insights from the analysis, such as product bundling opportunities, cross-selling strategies, or inventory management recommendations.
  - Identify specific actions that the business can take to leverage the insights.

- **Evaluation and Validation:**
  - Evaluate the impact of implemented strategies based on the insights. Monitor changes in sales, customer satisfaction, and other relevant metrics.
  - Validate whether the actions taken have been effective in achieving the desired outcomes.

- **Continuous Improvement:**
  - Continuously refine and update the market basket analysis as new data becomes available.
  - Stay responsive to changes in customer preferences and market dynamics.

- **Knowledge Sharing:**
  - Share the insights and findings with relevant teams within the organization, including marketing, sales, and inventory management.
  - Ensure that actionable insights are communicated effectively and lead to informed decision-making.

- **Scaling and Integration:**
  - If successful, scale the use of market basket insights across the organization.
  - Integrate market basket insights into the business's decision-making processes and operations.

Data Preprocessing

The dataset used for market basket insights typically contains transaction data from a retail or e-commerce environment. This data is essential for analyzing customer purchasing behavior and extracting insights related to product associations.

1. Loading the Dataset

➢ The process of acquiring and bringing your transaction data into a suitable data structure for analysis.

➢ This data is typically a record of transactions made by customers in a store, restaurant, or e-commerce platform, where each transaction lists the items purchased by a customer.

➢ Obtaining the transaction data from the source, which can be in various formats, including CSV files, Excel spreadsheets, SQL databases, or other data storage systems.

SOURCE CODE:

```
#import the dataset
import pandas as pd
data=pd.read_csv("Market_basket_analysis_dataset.csv")
print(data)
```

OUTPUT:

	BILL NO	ITEM NAME	QUANTITY	DATE	PRICE	CUSTOMERID	COUNTRY
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
...	...	...	...	...	...	...	
10011	537237	ASSORTED COLOURS SILK FAN	2	12/6/2010 9:58	1.28	NaN	United Kingdom
10012	537237	EDWARDIAN PARASOL NATURAL	1	12/6/2010 9:58	12.72	NaN	United Kingdom
10013	537237	CARTOON PENCIL SHARPENERS	1	12/6/2010 9:58	2.13	NaN	United Kingdom
10014	537237	PARTY TIME PENCIL ERASERS	5	12/6/2010 9:58	0.43	NaN	United Kingdom
10015	537237	SWIRLY CIRCULAR RUBBERS IN BAG	3	12/6/2010 9:58	0.84	NaN	United Kingdom

2. Data Understanding and Exploration

Data understanding and exploration are critical steps in the process of gaining market basket insights. These steps involve getting to know the dataset, its characteristics, and the patterns within it. Here's how to approach data understanding and exploration for market basket insights:

➢ **Data Loading:**

Begin by loading the market basket dataset into your preferred data analysis environment, such as Python with libraries like Pandas or R.

SOURCE CODE:

```
#display the first few rows of the dataset to get an overview

print(data.head())
```

➢ **Basic Statistics:**

Compute basic statistics to understand the dataset's structure. Calculate summary statistics such as mean, median, and standard deviation for numerical variables (e.g., quantity, transaction amount).

SOURCE CODE:

```
#Get basic statistics about the dataset
summary=data.describe()
print(summary)
```

3. Data Cleaning

Data cleaning is a crucial step in preparing your dataset for market basket analysis. It involves handling missing data, removing duplicates, and ensuring the data is in the right format for analysis. Below are common data cleaning steps with example code using Python:

Handle Missing Values:

➢ Check and handle missing values in your dataset. In market basket analysis, the focus is on transaction data, so it's important to ensure that the data is complete.

SOURCE CODE:

```python
import pandas as pd

# Load your transaction data into a DataFrame

data = pd.read_csv('Market_basket_analysis_dataset.csv')

# Check for missing values

missing_values = data.isnull().sum()

print(missing_values)

# If there are missing values, you may choose to drop rows with missing data

data.dropna()
```

Remove Duplicates:

Check for and remove duplicate transactions or items within transactions, as duplicates can distort the analysis.

SOURCE CODE:

```
# Check for duplicates based on all columns

duplicates = data.duplicated()

print("Number of duplicate rows:", duplicates.sum())

# Remove duplicate rows

data = data.drop_duplicates()
```

Transaction Format:

Ensure that your data is in a suitable transaction format where each row represents a transaction and each column represents items purchased. If your data is not in this format, you may need to pivot or reshape it.

SOURCE CODE:

```
# Convert the data into a binary matrix

basket = data.groupby(['CustomerID',
'Itemname'])['Itemname'].count().unstack().fillna(0)

print(basket)
```

Association Analysis Techniques

➢ Association analysis techniques, such as Apriori and FP-growth, are commonly used in market basket analysis to discover patterns and relationships between items that are frequently purchased together. Below, I'll provide examples of how to perform association analysis using these techniques in Python.

➢ You can use the "**mlxtend**" library in Python, which provides easy-to-use implementations of these association analysis techniques.

Apriori Algorithm:

The Apriori algorithm finds frequent itemsets and generates association rules based on support, confidence, and lift.

SOURCE CODE:

```
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
# Create a binary matrix (0 or 1) of transactions
# Each row is a transaction, and each column is an item
basket = data.groupby(['CustomerID', 'Itemname'])['Itemname'].count().unstack().fillna(0)
# Find frequent itemsets with a minimum support
frequent_itemsets = apriori(basket, min_support=0.1, use_colnames=True)
# Generate association rules
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1.0)
# Display the association rules
print(rules)
```

FP-growth Algorithm:

The FP-growth algorithm is an efficient algorithm for frequent pattern mining.

SOURCE CODE:

```python
from mlxtend.frequent_patterns import fpgrowth
from mlxtend.frequent_patterns import association_rules

# Create a binary matrix (0 or 1) of transactions
# Each row is a transaction, and each column is an item
basket = data.groupby(['customerID',
'Itemname'])['Itemname'].count().unstack().fillna(0)

# Find frequent itemsets with a minimum support
frequent_itemsets = fpgrowth(basket, min_support=0.1, use_colnames=True)

# Generate association rules
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1.0)

# Display the association rules
print(rules)
```

Discovered association rules and their business implications

➢ Interpreting and understanding the discovered association rules in market basket analysis is crucial for deriving actionable insights for your business. Association rules typically consist of antecedents (the items in the left-hand side of the rule) and consequents (the items in the right-hand side of the rule), along with metrics like support, confidence, and lift. Here's how you can interpret and derive business implications from the discovered association rules with code:

➢ Let's assume you have already obtained the association rules from your data, and you have them stored in a Data Frame.

SOURCE CODE:

```
import pandas as pd
# Load the association rules into a DataFrame (you should replace this
with your actual rules)
association_rules_data =
{
    'Antecedents': [['item1'], ['item2'], ['item1', 'item2'], ...],
    'Consequents': [['item3'], ['item4'], ['item4'], ...],
    'Support': [0.1, 0.2, 0.05, ...],
    'Confidence': [0.8, 0.7, 0.9, ...],
    'Lift': [1.2, 1.4, 1.5, ...]
}
```

Now, let's interpret the association rules and discuss their business implications:

➢ **Support:** Support measures how frequently the rule occurs in the dataset. Higher support indicates that the rule is more common.

➢ **Business Implication:** Rules with high support suggest that the associated items are frequently bought together. These items could be placed close to each other in the store to facilitate easier shopping.

➢ **Confidence:** Confidence measures the probability of the consequent occurring given the antecedent. A high confidence indicates a strong association between the antecedent and consequent.

➢ **Business Implication:** High-confidence rules can be used for recommendations. If a customer buys the antecedent, you can suggest the consequent item, as there's a high probability of it being purchased.

➢ **Lift:** Lift measures the ratio of the observed support to the expected support under independence. Lift values greater than 1 suggest a positive association, while values less than 1 indicate a negative association.

➢ **Business Implication:** Lift helps identify whether the association is meaningful. Lift greater than 1 implies that the antecedent and consequent are positively associated, making it a good candidate for bundling or cross-promotion.

Data Visualisation

Data visualization plays a crucial role in gaining insights from market basket data, which typically involves analyzing the purchasing patterns of customers. Market basket analysis helps businesses understand which products are frequently purchased together, enabling them to make informed decisions about product placement, promotions, and inventory management. Here are some data visualization techniques and tools that can be used for market basket insights:

- ➢ **Bar Charts:** Create bar charts to visualize the frequency of item pairs or item sets appearing together in customer transactions. You can use these charts to identify the most common associations. Each bar represents a product, and the height of the bar represents the frequency of co-occurrence.

- ➢ **Heatmaps:** Heatmaps are great for visualizing item associations in a matrix format. Rows and columns represent products, and the cells are shaded based on the strength of association. Darker cells indicate stronger relationships. Tools like Python's Seaborn library can help create heatmaps.

- ➢ **Network Graphs:** Network graphs, also known as association graphs, can show connections between products. Each node represents a product, and the edges between nodes indicate co-purchases. The thickness or color of the edges can represent the strength of the association.

- ➢ **Sankey Diagrams:** Sankey diagrams can illustrate the flow of items from one product to another. They are useful for visualizing the paths that customers take through the purchasing process and can help identify common item sequences.

➢ **Scatter Plots:** Scatter plots can be used to display relationships between two items. Each point on the plot represents a transaction, with one item on the x-axis and another on the y-axis. Patterns and clusters in the data can be revealed through this visualization.

➢ **Word Clouds:** Word clouds can help you visualize frequently co-purchased items. The size of each item's name in the word cloud is proportional to its frequency. This can help you quickly identify popular product combinations.

➢ **Tree Maps:** Tree maps are useful for displaying hierarchical data. You can use them to represent categories of products and the relationships between items within those categories.

➢ **Time Series Plots:** If you have transaction data with timestamps, time series plots can help you visualize changes in purchasing patterns over time. Line charts can be used to show trends in item associations.

➢ **Tableau and Power BI:** Tools like Tableau and Power BI offer a range of visualization options and can connect to various data sources. You can create interactive dashboards to explore market basket insights.

➢ **R and Python:** Programming languages like R and Python offer libraries like ggplot2 and Matplotlib/Seaborn, which provide extensive data visualization capabilities for customizing your visualizations based on your specific requirements.

SOURCE CODE:

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
Item_distr=data.groupby(by="Itemname").size().reset_index(name="Frequency").
sort_values(by="Frequency", ascending=False).head(10)
bars=Item_distr["Itemname"]
height=Item_distr["Frequency"]
x_pos=np.arange(len(bars))
plt.figure(figsize=(16,9))
plt.bar(x_pos, height, color=(0.2,0.3,0.5,0.5))
plt.title("Top 10 sold item")
plt.xlabel("Item names")
plt.ylabel('numer of quantity sold')
plt.xticks(x_pos, bars)
plt.show
```

THANK YOU