

## PHASE 5 ASSIGNMENT

PROJECT TITLE: market basket insights

PROBLEM DEFINITION: Market basket insights, or market basket analysis, is the process of discovering associations and patterns in customer transaction data. The primary goal is to uncover relationships between products or items that are frequently purchased together. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

GITHUB LINK: : <https://github.com/dharshini8404/innovation>

<https://github.com/dharshini8404/market-basket-insingths>

DOCUMENT: Building the project by Feature selection, Model training, Evaluation of an dataset.

DATASET LINK ON: <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

DESIGN THINKING PROCESS:

Design thinking is a problem-solving approach that can be applied to a wide range of challenges, including gaining insights from market basket data. The design thinking process typically consists of five key stages: Empathize, Define, Ideate, Prototype, and Test. Here's how you can apply this process to gain insights from market basket data:

### **Empathize: Understand the User and the Problem**

Begin by empathizing with the users of the market basket data, which could include customers, retailers, or analysts.

Conduct interviews, surveys, and observations to understand their needs, pain points, and goals related to market basket analysis.

Gain insights into what questions or problems they are trying to address with this data.

### **Define: Reframe the Problem**

Define a clear problem statement or challenge based on the insights gathered in the empathize stage. For example, "How might we use market basket data to optimize product placement in a retail store?"

Identify the specific objectives you want to achieve through market basket analysis.

### **Ideate: Generate Creative Solutions**

Brainstorm potential ways to gain insights from market basket data that align with the defined problem statement.

Encourage diverse perspectives and creativity within your team to come up with innovative ideas.

Consider using techniques like mind mapping and brainstorming sessions to generate a wide range of ideas.

### **Prototype: Create a Visual Representation**

Develop prototypes or mock-ups of the solutions or visual representations that will help extract insights from the market basket data.

Use data visualization tools to create sample dashboards, charts, or graphs that showcase the insights you want to derive.

These prototypes should be lightweight and can evolve as you refine your approach.

### **Test: Validate and Refine Your Approach**

Share the prototypes with stakeholders and users to gather feedback.

Run test cases using real market basket data to see how well your approach extracts meaningful insights.

## **PHASE OF DEVELOPMENT:**

Developing market basket insights involves several phases, from data collection and preparation to analysis and implementation. Here are the key phases for developing market basket insights:

### **Data Collection and Gathering:**

This phase involves collecting relevant data on customer purchases. This data can come from point-of-sale (POS) systems, e-commerce transactions, or other sources.

Data may include information on products purchased, transaction timestamps, customer demographics, and more.

Ensure data quality, consistency, and security during collection.

### **Data Preprocessing:**

Prepare the raw data for analysis by cleaning and formatting it. This may include handling missing values, removing duplicates, and converting data types.

Data preprocessing also involves data transformation, such as one-hot encoding for categorical variables and scaling or normalization for numerical data.

### **Exploratory Data Analysis (EDA):**

In this phase, you explore the data to gain a preliminary understanding of market basket patterns.

Use data visualization and statistical techniques to identify trends, correlations, and potential associations between products.

EDA helps you uncover initial insights and hypotheses.

### **Market Basket Analysis:**

Conduct market basket analysis using techniques like Association Rule Mining (e.g., Apriori algorithm) or Frequent Itemset Mining.

Identify itemsets (groups of products) that tend to be purchased together, along with associated metrics like support, confidence, and lift.

This phase helps you discover patterns and relationships between products.

### **Insights Generation:**

Translate the results of market basket analysis into actionable insights.

Identify which product combinations are most frequently purchased together and the implications for marketing, sales, and inventory management.

Develop recommendations and strategies based on the insights gained.

### **Model Validation and Testing:**

Validate the market basket insights by testing them on a different dataset or a holdout sample.

Ensure that the patterns and associations found in the data are statistically significant and can be generalized to real-world scenarios.

### **Implementation:**

Implement the insights in practical ways, such as optimizing product placements in physical stores or creating personalized product recommendations for e-commerce websites.

Collaborate with relevant teams (e.g., marketing, merchandising, or inventory management) to execute the strategies derived from the insights.

### **Monitoring and Optimization:**

Continuously monitor the impact of the insights on business performance.

Use Key Performance Indicators (KPIs) to track changes in sales, customer satisfaction, or other relevant metrics.

Optimize strategies as needed based on ongoing results.

### **Documentation and Reporting:**

Maintain documentation of the entire process, including data sources, preprocessing steps, analysis techniques, and insights generated.

Create reports and visualizations to communicate findings and recommendations to stakeholders.

### **Scaling and Sustainability:**

Consider how to scale and sustain the market basket insights process as the business evolves.

Explore automation and advanced analytics techniques to enhance and refine the insights over time

### **Dataset Used:**

The dataset used for market basket analysis typically contains transaction data that records items purchased together by customers. It could come from various sources,

such as point-of-sale (POS) systems, e-commerce platforms, or customer loyalty programs. The dataset usually includes the following columns:

Transaction ID: A unique identifier for each transaction.

Date/Time: Timestamp of when the transaction occurred.

Customer ID: Identifier for the customer making the purchase.

Items: A list of items purchased in that transaction, often represented as product codes or names.

Quantity: The number of each item purchased in the transaction.

**Data Preprocessing Steps:** Data preprocessing is essential to clean and format the dataset for association analysis. Common data preprocessing steps include:

**Data Cleaning:**

Handle missing data: Fill missing values (if possible) or remove incomplete records.

Remove duplicates: Ensure each transaction is unique.

**Data Transformation:**

Convert categorical data: Encode item names or product codes as numerical values (e.g., one-hot encoding).

Normalize numerical data: Scale item quantities if necessary to ensure consistency.

**Transaction Aggregation:**

Group transactions by customer: This may involve aggregating all items purchased by each customer into a single transaction.

**Transaction Reduction:**

Reduce the dataset size if it's extensive by considering a subset of transactions or items to speed up processing.

**Market Basket Representation:**

Transform the data into a transaction-item matrix where rows represent transactions, columns represent items, and values indicate whether an item was purchased in a transaction (usually represented as binary 0/1 values).

**Association Analysis Techniques:** Association analysis aims to discover patterns and relationships between items in the dataset. Common techniques include:

### **Apriori Algorithm:**

The Apriori algorithm is a widely used technique for market basket analysis. It identifies frequent itemsets (sets of items frequently purchased together) and generates association rules.

Key metrics include support, confidence, and lift.

Support measures the frequency of an itemset, confidence quantifies the likelihood that a rule is true, and lift measures the strength of association between items.

### **FP-growth (Frequent Pattern Growth):**

FP-growth is an alternative algorithm to Apriori that is more efficient in some cases. It constructs a tree structure (FP-tree) to find frequent itemsets.

It is particularly useful when working with large datasets.

### **Eclat (Equivalence Class Transformation):**

Eclat is another approach for discovering frequent itemsets by making use of a depth-first search technique.

It focuses on vertical data format and is efficient for finding frequent itemsets with high support.

### **Rule Generation:**

Once frequent itemsets are identified, association rules are generated based on metrics like confidence and lift.

Rules are presented in the form of "If {A} then {B}" to indicate that when items A are purchased, items B are also likely to be purchased.

### **Association Rule Format:**

Association rules are typically presented in the format "If {A} then {B}" or "Antecedent => Consequent," where A and B are sets of items. The rules are generated based on metrics like support, confidence, and lift.

**Support:** The support of a rule measures the percentage of transactions that contain both A and B.

**Confidence:** Confidence indicates the likelihood that B will be purchased when A is purchased. It's calculated as the ratio of the support for both A and B to the support for A.

**Lift:** Lift measures the strength of association between A and B. A lift value greater than 1 indicates a positive association, meaning that the purchase of A is positively correlated with the purchase of B.

Now, let's discuss the business implications of these association rules:

### **Cross-Selling Opportunities:**

If {Product A} => {Product B} (with high confidence and lift):

Business Implication: Promote the purchase of Product A to increase the sales of Product B, as these two items are often bought together. Consider bundling them or running targeted promotions.

### **Product Placement Optimization:**

If {Product A} => {Product B} (with high support):

Business Implication: Place Product A and Product B near each other in physical stores or suggest them together on e-commerce platforms to increase the likelihood of joint purchases.

### **Customer Segmentation:**

If {Product A} => {Product B} (with high confidence):

Business Implication: Identify customers who have purchased Product A and create targeted marketing campaigns to encourage them to buy Product B.

### **Inventory Management:**

If {Product A} => {Product B} (with high lift):

Business Implication: Optimize inventory levels by ensuring that when Product A is restocked, there are enough units of Product B available, as they tend to be purchased together.

### **Promotion Planning:**

If {Product A, Product C} => {Product B} (with high confidence and lift):

Business Implication: Create promotion bundles that include Products A, C, and B, as they are frequently bought together. Offer discounts or incentives to drive sales.

### **Customer Loyalty Programs:**

If {Product A} => {Loyalty Program Enrollment} (with high confidence and lift):

Business Implication: Encourage customers who purchase Product A to enroll in the loyalty program, as they are more likely to do so. This can help improve customer retention.

### **Pricing Strategies:**

If {Product A} => {Product B} (with high support but low confidence):

Business Implication: Consider offering discounts on Product B when customers purchase Product A to boost the confidence level of this rule and stimulate sales.

### **Seasonal Promotions:**

If {Product A} => {Product B} (with high support during certain seasons):

Business Implication: Implement seasonal marketing strategies to leverage the seasonal purchase patterns of these items.

### **Submission:**

Creating a full code implementation for market basket insights typically involves multiple steps, including data preprocessing, association rule mining, and possibly data visualization. Below is a simplified Python code example using the popular library `pandas` for data preprocessing and



`mlxtend` for association rule mining. This code assumes that you have your transaction data in a CSV file.

You'll need to install the necessary libraries if you haven't already:

```
pip install pandas mlxtend
```

Here's a sample code that includes data preprocessing and association analysis using the Apriori algorithm:

```
import pandas as pd
```

```
from mlxtend.frequent_patterns import apriori
```

```
from mlxtend.frequent_patterns import association_rules
```

```
# Load the transaction data from a CSV file
```

```
data = pd.read_csv('transaction_data.csv')
```

```
# Data Preprocessing
```

```
# Assuming that your dataset contains a 'Transaction_ID' column and an 'Item' column
```

```
basket = (data.groupby(['Transaction_ID', 'Item'])['Item']
```

```
    .count().unstack().reset_index().fillna(0)
```

```
    .set_index('Transaction_ID'))
```

```
# Convert item quantities to binary (1 or 0)
```

```
def encode_units(x):
```

```
    if x <= 0:
```

```
        return 0
```

```
    if x >= 1:
```

```
        return 1
```

```
basket_sets = basket.applymap(encode_units)
```

```
# Association Analysis
```

```
frequent_itemsets = apriori(basket_sets, min_support=0.05, use_colnames=True)
```

```
rules = association_rules(frequent_itemsets, metric='lift', min_threshold=1.0)
```

```
# Display association rules
```

```
print(rules)
```

In this code:

Load your transaction data from a CSV file, which should contain 'Transaction\_ID' and 'Item' columns.

Perform data preprocessing to create a binary transaction-item matrix.

Use the Apriori algorithm to find frequent itemsets.

Generate association rules based on metrics like lift.

Please adapt the code to your specific dataset, ensuring that you have the necessary data columns and data format. Additionally, you may need to adjust the parameters such as `min_support` and `min_threshold` based on your data and business requirements.

This is a simplified example, and in practice, you may need to handle more complex data, including data scaling, encoding, and more advanced visualization and reporting.

Creating a well-structured README file is essential for ensuring that users can understand and run your market basket insights code. Below is a template for a README file that explains how to run the code, lists the dependencies, and provides additional information:

```
# Market Basket Insights
```

## ## Overview

This code provides a Python implementation for analyzing market basket data to discover association rules. It uses the Apriori algorithm to find frequent itemsets and generate association rules. You can use this code to gain insights into customer purchase behavior and make data-driven decisions.

## ## Table of Contents

1. [Dependencies](#dependencies)
2. [Usage](#usage)
3. [Data Preparation](#data-preparation)
4. [Running the Code](#running-the-code)
5. [Adjusting Parameters](#adjusting-parameters)
6. [Results](#results)
7. [License](#license)

## ## Dependencies

This code relies on the following Python libraries:

- pandas
- mlxtend

You can install these libraries using pip:

```
```bash
```

```
pip install pandas mlxtend
```

## Usage

Clone this repository to your local machine:

```
git clone https://github.com/dharshini8404/market-basket-insights
```

Navigate to the project directory:

```
cd market-basket-insights
```

### Data Preparation

Before running the code, make sure you have your transaction data in a CSV file. The CSV file should contain at least two columns: 'Transaction\_ID' and 'Item', where 'Transaction\_ID' uniquely identifies each transaction, and 'Item' represents the purchased items.

Ensure that your CSV file is in the same directory as the code file.

### Running the Code

To run the code, execute the following command in your terminal:

```
python market_basket_analysis.py
```

The code will process the data, perform association rule mining, and display the discovered rules.

### Adjusting Parameters

You can adjust the parameters in the code to fit your specific requirements. The following parameters can be modified in the code:

**min\_support**: Set the minimum support threshold for frequent itemsets.

**min\_threshold**: Set the minimum threshold for association rules (e.g., lift).

### Results

The code will display the discovered association rules, including the antecedent, consequent, support, confidence, and lift values. These rules provide insights into item associations within your market basket data.

### License

This project is licensed under the [MIT License](#), which allows you to use and modify the code freely.

Feel free to reach out if you have any questions or need further assistance.

Happy market basket analysis!

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

SUBMITTED BY,

STUDENT REG NO: 711221104014

NAAN MUDHALVAN: au711221104014