PROJECT TITLE : MARKET BASKET INSIGHTS

DatasetLink: **https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis**

Problem Statement:

The problem at hand is to analyze a dataset and discover association rules within it. Association analysis aims to find interesting relationships or patterns in data, such as products that are frequently purchased together or symptoms that co-occur in medical records. The goal is to uncover hidden insights that can be used for decision-making and business strategies.

Design Thinking Process:

1. Empathize: Understand the problem and the data available. Identify the stakeholders' needs and expectations.

2. Define: Clearly define the objectives of the association analysis. What patterns or relationships are you trying to discover?

3. Ideate: Brainstorm potential approaches and techniques for association analysis. Consider algorithms like Apriori, FP-growth, or Eclat.

4. Prototype: Develop a plan for data preprocessing, model selection, and analysis.

5. Test: Implement the analysis and evaluate the results.

6. Implement: Translate the insights into actionable recommendations for the business.

Phases of Development:

1. Data Collection: Gather the dataset that contains the relevant information for association analysis.

2. Data Preprocessing: Clean the data by handling missing values, removing duplicates, and transforming it into a suitable format.

3. Association Analysis: Apply association rule mining algorithms to discover patterns in the data.

4. Interpretation: Interpret the discovered association rules in the context of the problem and business goals.

5. Business Implications: Use the insights to make data-driven decisions, such as optimizing product placement, cross-selling, or improving healthcare practices.

Dataset Used:

Describe the dataset, including its source, format, and size. For example, it could be a retail transaction dataset with information on products purchased by customers.

Data Preprocessing Steps:

- Handle missing values: Decide on a strategy for dealing with missing data, such as imputation or removal.
- Remove duplicates: Eliminate duplicate records if they exist.
- Data Transformation: Convert the data into a suitable format for association analysis, typically a transaction format.

Association Analysis Techniques:

- Apriori Algorithm: It finds frequent itemsets and generates association rules based on support and confidence.
- FP-growth: This algorithm uses a tree structure to mine frequent patterns efficiently.
- Eclat: Eclat is a depth-first search algorithm that focuses on transaction intersection.

Discovered Association Rules:

Explain the association rules that were discovered. For example, "Customers who buy product A are 80% likely to purchase product B in the same transaction."

Business Implications:

Discuss how the discovered association rules can be used for decision-making and their potential impact on the business. For instance, "These association rules can inform product bundling strategies, leading to increased sales and customer satisfaction."

Certainly! Below is a Python code example for performing association analysis using the Apriori algorithm on a sample retail transaction dataset. This code demonstrates the essential steps from loading the dataset to discovering association rules.

python

```
# Import necessary libraries

import pandas as pd

from mlxtend.frequent_patterns import apriori
```

```
from mlxtend.frequent_patterns import association_rules

# Load the sample retail transaction dataset (replace 'retail_dataset.csv' with your dataset)
data = pd.read_csv('retail_dataset.csv')

# Data Preprocessing
# Assuming the dataset has a column 'Transaction ID' and 'Product Name'
basket = (data.groupby(['Transaction ID', 'Product Name'])['Product Name']
          .count().unstack().reset_index().fillna(0)
          .set_index('Transaction ID'))

# Convert the data into binary format (1 for purchased, 0 for not)
def encode_units(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1

basket_sets = basket.applymap(encode_units)

# Association Analysis using Apriori
frequent_itemsets = apriori(basket_sets, min_support=0.01, use_colnames=True)

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

# Display discovered association rules
print("Discovered Association Rules:")
print(association_rules)

# Business Implications
```

You can interpret and act on the discovered rules to optimize product placement or cross-selling strategies.