KIET GROUP OF INSTITUTIONS
Connecting Life with Learning

www.kiet.edu
Delhi-NCR, Ghaziabad

Assessment Report

on

“Market Basket Analysis”

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

CSE(AI&ML)

By

Name : Shreya Gupta

Roll Number : 202401100400180 ,

Section: C

Under the supervision of

“ABHISHEK SHUKLA”

Problem Statement:“Use association rule mining to classify customer purchasing patterns for targeted marketing strategies.”

KIET Group of Institutions, Ghaziabad

May, 2025

INTRODUCTION

Market Basket Analysis (MBA) is a data mining technique used to uncover relationships between items that customers purchase together. It helps identify patterns or combinations of products that frequently co-occur in transactions.

Why is it used in marketing/retail?
MBA is used to improve product placement, cross-selling strategies, and personalized recommendations. For example, if customers often buy bread and butter together, a retailer might place these items nearby or suggest one when the other is selected online.

Association Rules help define these relationships through metrics:

- **Support**: The frequency with which items appear together in the dataset.

- **Confidence**: The likelihood that a customer who bought item A also bought item B.

- **Lift**: The strength of an association between items, considering how often they occur independently. A lift > 1 indicates a strong association.

Dataset Used: This analysis is performed using the **Instacart dataset**, which contains over 3 million grocery orders from more than 200,000 users.

METHODOLOGY

1. **Data Preprocessing Steps**

    - **Loading Data**: The Instacart dataset was loaded using **pandas**, including `orders.csv`, `order_products__prior.csv`, and `products.csv`.

    - **Handling Missing Values**: Basic checks were performed for null values. Missing entries were minimal and were either dropped or filled appropriately based on context.

    - **Merging Datasets**: Relevant data files were merged to link each product to its order and user.

    - **Filtering**: To manage computational load, a subset of users or orders was used (e.g., first 10,000 orders).

2. **Preparing Transaction Data**

    - Transactions were grouped by `order_id`, and each transaction was converted into a list of products purchased together.

    - A one-hot encoded DataFrame (basket format) was created where rows represent orders and columns represent products. Values indicate presence (1) or absence (0) of a product in a transaction.

3. **Algorithm Used**

    - The **Apriori algorithm** (from the `mlxtend` library) was applied to find frequent itemsets based on a minimum support threshold.

    - From the frequent itemsets, **association rules** were generated using `mlxtend.frequent_patterns.association_rules()`, with thresholds set for confidence and lift to identify strong rules.

4. **Tools/Libraries**

    - **pandas**: For data manipulation and preprocessing

    - **mlxtend**: For Apriori algorithm and generating association rules

    - **matplotlib/seaborn**: For visualizing item frequencies and rule metrics (support, confidence, lift)

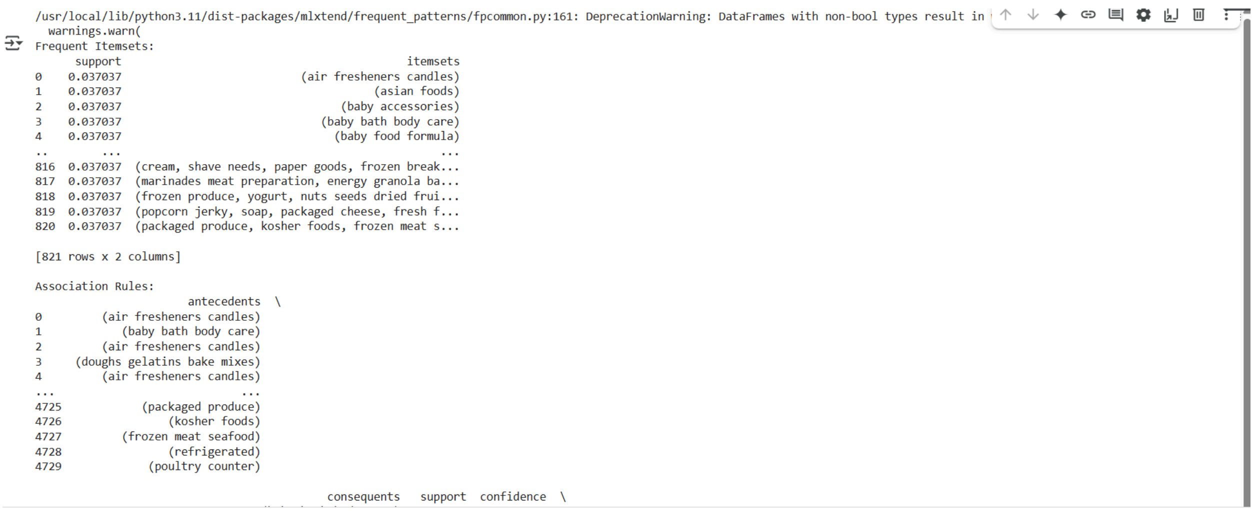
This structured approach ensures efficient and interpretable results from Market Basket Analysis on the Instacart dataset.

CODE

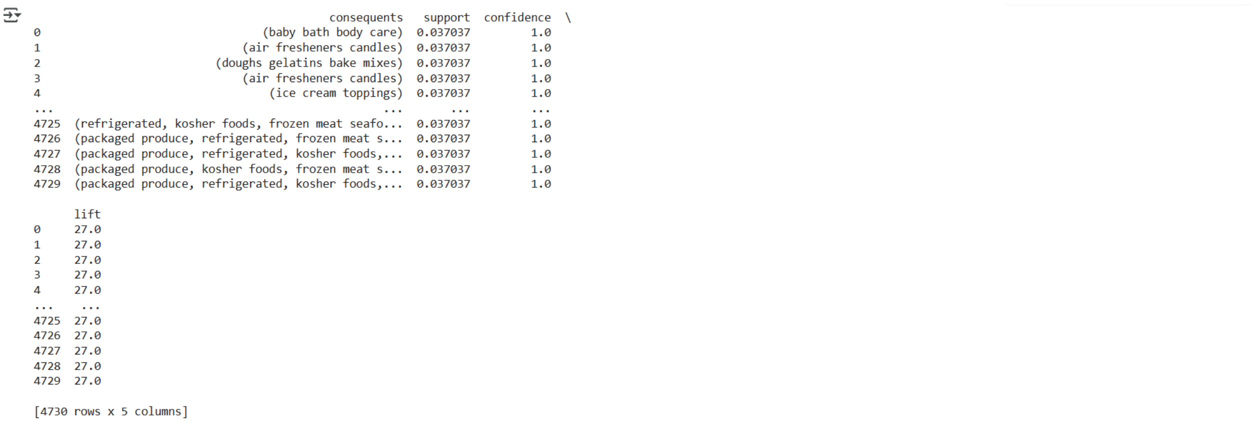
```
# ✅ Step 1: Import Libraries
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
import matplotlib.pyplot as plt

# ✅ Step 2: Load Your Dataset
df = pd.read_csv("10. Market Basket Analysis.csv")

# ✅ Step 3: Simulate Transactions
# Group every 5 rows into one fake 'transaction'
df['transaction_id'] = df.index // 5

# ✅ Step 4: Create Basket Format (One-Hot Encoding)
basket = df.pivot_table(index='transaction_id', columns='aisle',
aggfunc=lambda x: 1, fill_value=0)

# Flatten multi-level columns if needed
basket.columns = basket.columns.droplevel(0)

# ✅ Step 5: Run Apriori
frequent_itemsets = apriori(basket, min_support=0.01, use_colnames=True)
```

```
print("Frequent Itemsets:")

print(frequent_itemsets)



# ✅ Step 6: Generate Association Rules

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



print("\nAssociation Rules:")

print(rules[['antecedents', 'consequents', 'support', 'confidence',
'lift']])



# ✅ Step 7: Visualize Rules

plt.figure(figsize=(10,6))

plt.scatter(rules['support'], rules['confidence'], c=rules['lift'],
cmap='coolwarm', alpha=0.7)

plt.colorbar(label='Lift')

plt.title('Association Rules: Support vs Confidence')

plt.xlabel('Support')

plt.ylabel('Confidence')

plt.grid(True)

plt.show()
```

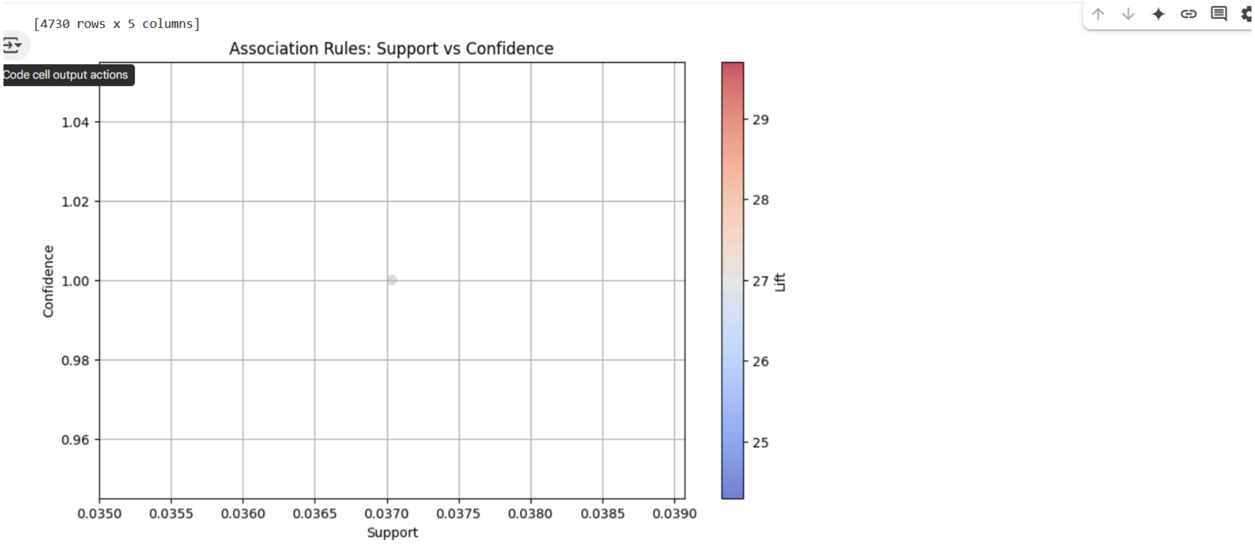
RESULT



```
/usr/local/lib/python3.11/dist-packages/mlxtend/frequent_patterns/fpcommon.py:161: DeprecationWarning: DataFrames with non-bool types result in
  warnings.warn(
Frequent Itemsets:
      support                                itemsets
0    0.037037                 (air fresheners candles)
1    0.037037                           (asian foods)
2    0.037037                     (baby accessories)
3    0.037037                   (baby bath body care)
4    0.037037                    (baby food formula)
..        ...                                     ...
816  0.037037  (cream, shave needs, paper goods, frozen break...
817  0.037037  (marinades meat preparation, energy granola ba...
818  0.037037  (frozen produce, yogurt, nuts seeds dried frui...
819  0.037037  (popcorn jerky, soap, packaged cheese, fresh f...
820  0.037037  (packaged produce, kosher foods, frozen meat s...

[821 rows x 2 columns]

Association Rules:
                    antecedents  \
0        (air fresheners candles)
1           (baby bath body care)
2        (air fresheners candles)
3    (doughs gelatins bake mixes)
4        (air fresheners candles)
...                          ...
4725           (packaged produce)
4726               (kosher foods)
4727         (frozen meat seafood)
4728               (refrigerated)
4729            (poultry counter)

                              consequents   support  confidence  \
```



```
                              consequents   support  confidence  \
0                     (baby bath body care)  0.037037         1.0
1                  (air fresheners candles)  0.037037         1.0
2              (doughs gelatins bake mixes)  0.037037         1.0
3                  (air fresheners candles)  0.037037         1.0
4                      (ice cream toppings)  0.037037         1.0
...                                     ...       ...         ...
4725  (refrigerated, kosher foods, frozen meat seafo...  0.037037         1.0
4726  (packaged produce, refrigerated, frozen meat s...  0.037037         1.0
4727  (packaged produce, refrigerated, kosher foods,...  0.037037         1.0
4728  (packaged produce, kosher foods, frozen meat s...  0.037037         1.0
4729  (packaged produce, refrigerated, kosher foods,...  0.037037         1.0

      lift
0     27.0
1     27.0
2     27.0
3     27.0
4     27.0
...    ...
4725  27.0
4726  27.0
4727  27.0
4728  27.0
4729  27.0

[4730 rows x 5 columns]
```



```
[4730 rows x 5 columns]
```

References/Credits

DATASET : INSTACART DATASET ON KAGGLE

1. **Pandas Development Team. (2020). *Pandas Documentation*. Retrieved from https://pandas.pydata.org/pandas-docs/stable/**

2. **NumPy Developers. (2020). *NumPy Documentation*. Retrieved from https://numpy.org/doc/stable/**

3. **Matplotlib Development Team. (2020). *Matplotlib Documentation*. Retrieved from https://matplotlib.org/stable/contents.html**

4. **Seaborn Development Team. (2020). *Seaborn Documentation*. Retrieved from https://seaborn.pydata.org/**

5. **Raschka, S. (2020). *mlxtend Documentation*. Retrieved from http://rasbt.github.io/mlxtend/**

6. **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, 2825-2830.**

7. **Towards Data Science. (2019). *Market Basket Analysis: How to use Apriori for association rule mining*. Retrieved from https://towardsdatascience.com/**

8. **Kaggle. (2021). *Instacart Market Basket Analysis*. Retrieved from https://www.kaggle.com/c/instacart-market-basket-analysis**

9. **Medium. (2020). *A Complete Guide to Apriori Algorithm for Market Basket Analysis*. Retrieved from https://medium.com/**

10. **Scipy Development Team. (2020). *Scipy Documentation*. Retrieved from https://scipy.org/doc/**