MARKET BASKET INSIGHTS

Phase 2 Submission Document

Project: Market basket Insights

**Introduction:**

1. Market basket analysis is a technique that helps retailers understand the buying patterns of their customers and optimize their sales strategies accordingly. It involves analysis large data sets of purchase history to find out which products are frequently bought together, and how likely a customer is to buy one product given that they have bought another product.
2. retailers use data mining algorithms that extract patterns and rules from large data sets. One of the most common algorithms is the Apriori algorithm.

**Content of phase2:**

In this phase, we can explore innovative techniques such as ensemble methods and deep learning architectures to improve the prediction system's accuracy and robustness.

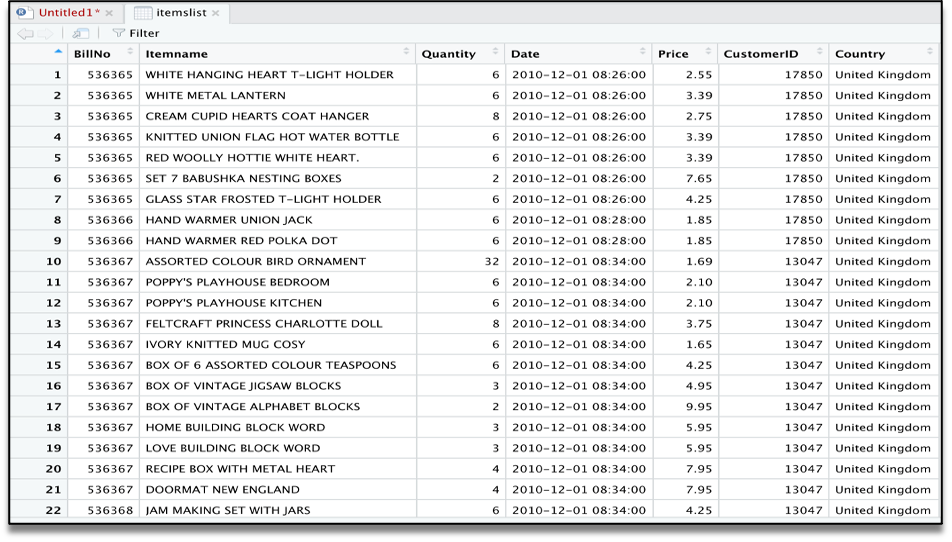
Consider exploring advanced association analysis techniques or using visualization tools for enhanced insights presentation.

**Data source:**

The data can be stored in databases, spreadsheets, or flat files, and can be structured or unstructured data. the data set should be Taccurate ,Customer Details, Unit Prices, product Quantity.

**Dataset link:**

[**https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis**](https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis)



Data set process:

1. Clean and preprocess the data, removing any irrelevant information, handling missing values, and converting the data into a suitable format for analysis. You can use tools like Excel, Python, or R to perform these tasks.
2. Apply data mining algorithms that extract patterns and rules from large data sets.

**Visualization:**

Data visualization techniques like scatter plots, heat maps, and dendrogram plots can help in visualizing the relationships between products and making the insights more understandable.

**Implementation:**

improve customer experiences and increase revenue by using these insights, businesses can create product bundles, optimize pricing strategies, and personalize recommendations for individual customers

Program

Importing the libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

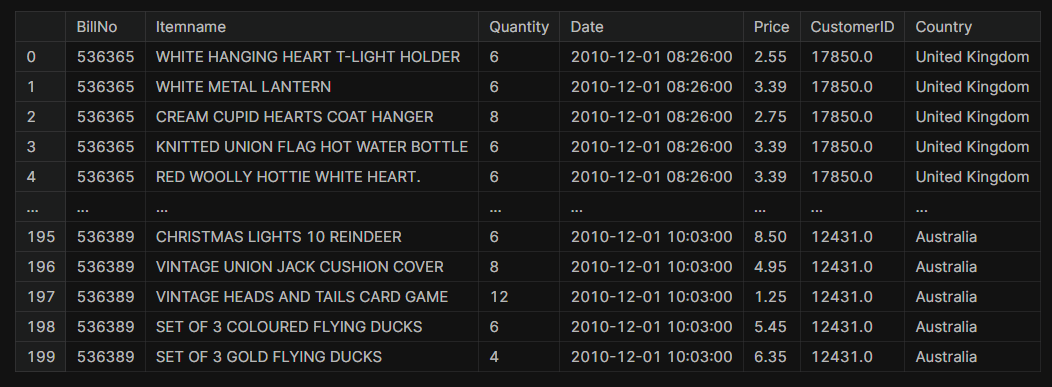
from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

# Data preprocessing

dataset = pd.read\_excel('/kaggle/input/market-basket-analysis/Assignment-1\_Data.xlsx')

dataset.head(200)



dataset.isnull().sum()

OUTPUT:

BillNo 0

Itemname 1455

Quantity 0

Date 0

Price 0

CustomerID 134041

Country 0

dtype: int64

dataset['Itemname'] = dataset['Itemname'].str.strip()

dataset.dropna(axis=0, subset=['Itemname'], inplace = True)

dataset = dataset.drop(columns= ['CustomerID'])

dataset.isnull().sum()

OUTPUT:

BillNo 0

Itemname 0

Quantity 0

Date 0

Price 0

Country 0

dtype: int64

dataset.dtypes

OUTPUT:

BillNo object

Itemname object

Quantity int64

Date datetime64[ns]

Price float64

Country object

dtype: object

dataset['BillNo'] = dataset['BillNo'].astype('str')

dataset = dataset[~dataset['BillNo'].str.contains('C')]

dataset['Country'].value\_counts()

OUTPUT:

United Kingdom 486167

Germany 9042

France 8408

Spain 2485

Netherlands 2363

Belgium 2031

Switzerland 1967

Portugal 1501

Australia 1185

Norway 1072

Italy 758

Sweden 451

Unspecified 446

Austria 398

Poland 330

Japan 321

Israel 295

Hong Kong 284

Singapore 222

Iceland 182

USA 179

Greece 145

Malta 112

United Arab Emirates 68

RSA 58

Lebanon 45

Lithuania 35

Brazil 32

Bahrain 18

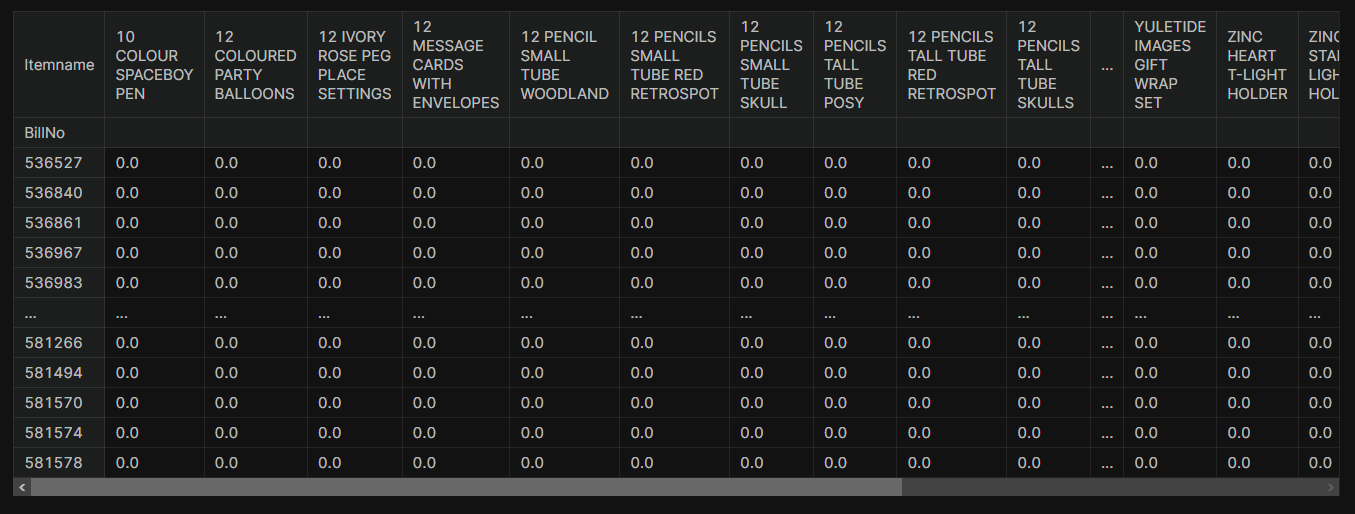
Saudi Arabia 9

Name: Country, dtype: int64

basket = (dataset[dataset['Country'] == 'Germany' ].groupby(['BillNo','Itemname'])['Quantity'].sum().unstack().fillna(0))

basket

OUTPUT:



def encode(x):

if x <= 0:

return 0

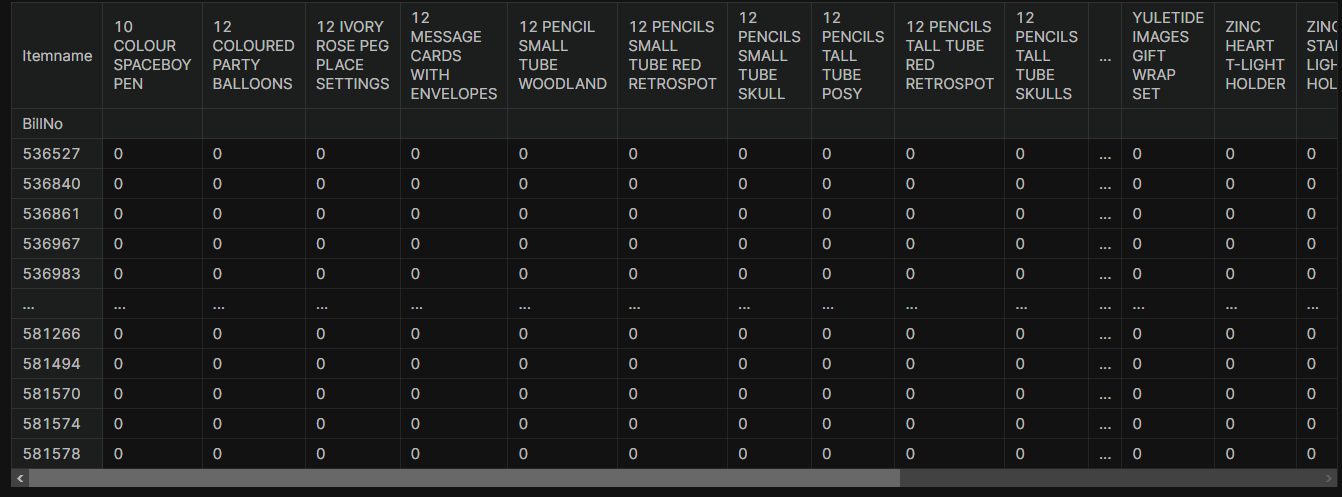
if x >= 1:

return 1

basket = basket.applymap(encode)

basket.drop('POSTAGE', inplace = True, axis = 1)

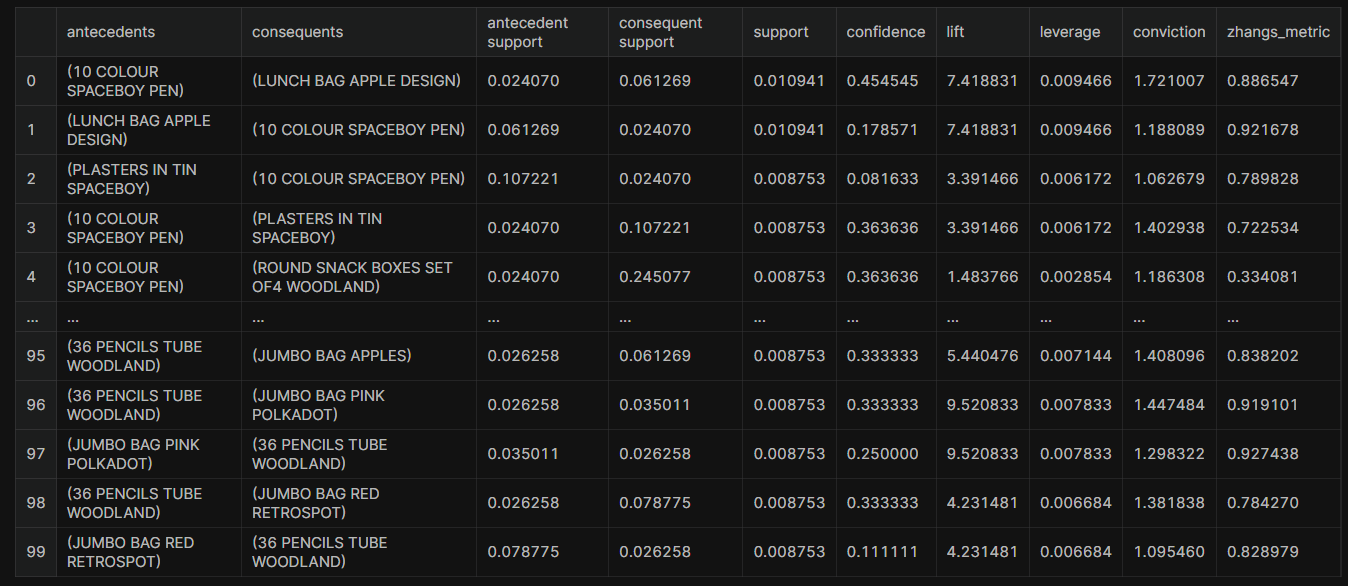
basket



frequent\_items = apriori(basket, min\_support = 0.007, use\_colnames= True)

rules = association\_rules(frequent\_items, metric = 'lift', min\_threshold = 1)

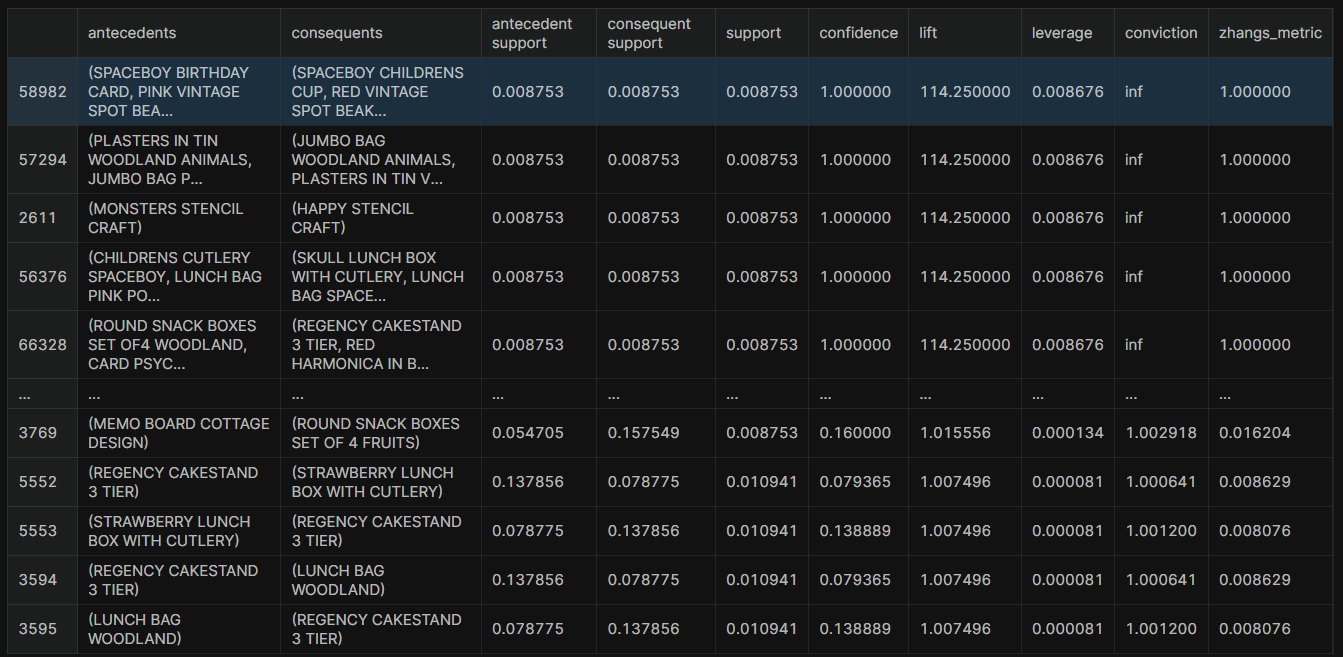
rules.head(100)



rules = rules.sort\_values(by='lift', ascending = False)

rules

OUTPUT:



**Conclusion:**

**Project conclusion:**

The phase 2 for a  exploring advanced association analysis techniques or using visualization tools for enhanced insights presentation. Over all understand the buying patterns of their customers and optimize their sales strategies accordingly.