Avery Thomas

010664483

Data Mining

Apriori Algorithm

Based on my results after analyzing the data from all purchases in a supermarket over 1 week I have reached several conclusions. All of my transactions required a confidence of 40% in order to be included. My results are included below.

**Conclusions**

* Based on my results I can conclude that spaghetti and ground beef are frequently bought together, this is most likely because ground beef is a very popular ingredient in spaghetti.
* Milk and frozen vegetables are also bought together, this can be due to people wanting to eat healthy and milk and vegetables are both very popular healthy options.
* Shrimp and spaghetti are frequently bought together. This is probably due to the fact that shrimp is another very popular protein that can be added to spaghetti
* Spaghetti is frequently bought alone. This can be because spaghetti is a very cheap and filling dish and some customers might not have very much money.

**Results:**

Items purchased: spaghetti & ground beef

Support: 0.004799360085321957

Confidence: 0.5714285714285714

Lift: 3.2819951870487856

=====================================

Items purchased: milk & frozen vegetables

Support: 0.004799360085321957

Confidence: 0.4235294117647058

Lift: 3.2684095860566447

=====================================

Items purchased: shrimp & spaghetti

Support: 0.005999200106652446

Confidence: 0.5232558139534884

Lift: 3.005315360233627

=====================================

Items purchased: nan & spaghetti

Support: 0.004799360085321957

Confidence: 0.5714285714285714

Lift: 3.2819951870487856

=====================================

Items purchased: milk & frozen vegetables

Support: 0.004799360085321957

Confidence: 0.4235294117647058

Lift: 3.2684095860566447

=====================================

Items purchased: nan & shrimp

Support: 0.005999200106652446

Confidence: 0.5232558139534884

Lift: 3.005315360233627

Code:

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('store\_data.csv', header=None)

numrecords= len(store\_data)

records = []

for i in range(0, numrecords):

records.append([str(store\_data.values[i,j]) for j in range(0, 20)])

association\_rules = apriori(records,min\_support=0.0045,min\_confidence=0.40,min\_lift=3,min\_length=2)

association\_results=list(association\_rules)

print(len(association\_results))

#print items which follow association rules entered

for item in association\_results:

pair = item[0]

items = [x for x in pair]

print("Items purchased: " + items[0] + " & " + items[1])

#second index of the inner list

print("Support: " + str(item[1]))

#third index of the list located at 0th

#of the third index of the inner list

print("Confidence: " + str(item[2][0][2]))

print("Lift: " + str(item[2][0][3]))

print("=====================================")