# Apriori Algorithm

## **Mid Term Project**

Course: CS-634

Name: Anusha Syed

Email: as445@njit.edu

# Code:

When running in cmd, the name and path of datasets should be given according to your system's location where project is kept.

## Libraries used:

Permutations and combinations libraries are used to create combinations for in pairs of 2,2 or 3,3 for K=2 and K=3 after elimination of items based on minimum support provided by users.

```
from itertools import permutations from itertools import combinations
```

# File (dataset) Input:

Here, put your dataset name with txt extension i.e. dataset should be saved in notepad. The fileobject will read your file as indicated by 'r' and input data will read each line.

## **Cleaning of data**

Here, final data is the list in which transformed data (in form of list) will be stored. All the unnecessary spaces in dataset will be replaced by none and I have to differentiate between each transaction by comma that is given in split

```
## Transforming the dataset
final_data = []
for lines in range(1,len(input_data)):
   input_data[lines] = input_data[lines].replace("\n","")
   final_data.append(input_data[lines].split(','))
```

## **Printing Transformed data:**

We have indicated final\_data to pick data after first element so that it doesn't pick Trans1, Trans2 and so on because it was picking each row after comma so to clean that part I wrote 1:

```
## Transforming final_data
for i in range(len(final_data)):
    final_data[i] = final_data[i][1:]

print("\n","This is the dataset you entered:", "\n", "\n", final_data, '\n')
```

# **Taking Input for confidence and support:**

```
## Declaring all the required variables
min_sup = int(input("Enter Minimum Support:"))
print("\n")
min_conf = int(input("Enter Minimum Confidence:"))
print("\n")
```

## **Count of Items in each Transaction:**

I have defined all the functions on top and then called them below for my main. First Function here is to count all unique items.

- o (set) is used to remove redundant items.
- o len here will count the items in i.
- Sorted will sort in ascending order
- o uniquelist will have count of each item present in dataset

```
## Find the count of unique items and make a dictionary
def uniquelist(final_data):
    mylist = []
    for i in final_data:
        for j in range(len(i)):
            mylist.append(i[j])
    uniquelist = set(mylist)
    uniquelist = sorted(uniquelist)
    return uniquelist, mylist
```

# **Creating a Dictionary:**

After getting count of each item in each transaction, we represent that count in key:value form and for that representation we have used dictionary data structure.

firstdict{} will have key:value for item sets created for k=1

```
## Performing Iteration 1
def iteration1(mylist, uniquelist):
    firstdict = {}
    for i in uniquelist:
        firstdict[i] = mylist.count(i)
    return firstdict
```

# **Elimination of Items based on Minimum Support:**

Support\_eliminate is the function created to store the dictionary named here as mydict2 that will have the items only that met the minimum support requirement.

- o firstdict is called as parameter that will be having the count of each item along with item itself.
- keylist will the item name and valuelist will have the item count.
- Formula for support is count of transactions for each item/ total number of transactions count so we took min\_sup from user and after getting the percentage we compared it from which limit user gave then eliminated those items that did not meet the minimum support requirement
- mydict2 will have left out item sets that meets the min\_sup.

```
## Eliminate those items not satisfying minimum support
def support eliminate(firstdict):
    keylist = list(firstdict.keys())
    valuelist = list(firstdict.values())
    mydict2 = {}
    # support_dict = {}
    count = 0
    for i in valuelist:
        support = round(i * 100 / total_trxns, 2)
        count += 1
        # print(support, min sup)
        if support >= min_sup:
            # firstdict.pop(keylist[count])
            mydict2[keylist[count - 1]] = i
            # support_dict[keylist[count-1]] = support
    return mydict2
```

# Moving forward to K=1 where we create combination of elements left out after minimum support given by user:

I have used built-in function combination to create combinations for k=1.

- o create\_combinations will have mydict2 that has all itemsets that met the min\_sup.
- I stored combinations created from mydict2 in comb variable.
- o I then stored dictionary based itemset in up comb which is a list
- Finally storing key:value that is dictionary type data in list called as up\_comb

```
## Creating combinations of items in Iteration 1
def create_combinations1(mydict2, n):
    comb = combinations(list(mydict2.keys()), n)
    up_comb = []
    for i in comb:
        up_comb.append(i)

return up_comb
```

# Implementation in Main for K=1:

- o Sup\_elim1 will have the processed data to display after elimination.
- o Iteration will have data stored in form of key: value that is dictionary

 Support\_eliminate function will have all implementation for elimination based on user input for minimum support.

```
total_trxns = len(final_data)
uniques, full_list = uniquelist(final_data)
iteration1 = iteration1(full_list, uniques)
sup_elim1 = support_eliminate(iteration1)
print("\n","This is the result after first iteration:", "\n", "\n", sup_elim1, '\n')
new_combs = create_combinations1(sup_elim1, 2)
```

Moving forward to L=2 where we create count elements in each transaction that come together in possible combination in each Transaction:

- o Visualize this step where we are counting A,C or A,C coming together in each transaction.
- Up\_comb is the list where we have eliminated items from k=1 and final data is our transformed data set then we count those items coming together in each row and store in updated\_dict.

## Implementation in Main for K=2:

```
iteration2 = iter2(final_data, new_combs)

sup_elim2 = support_eliminate(iteration2)

print("\n","This is the result after second iteration:", "\n", sup_elim2, '\n')

mylist2iter = list(sup_elim2.keys())

##used this to put unique pairs in tuple to make it in a single list
listafter2 = []
for i in range(len(mylist2iter)):
    listafter2.append(mylist2iter[i][0])
    listafter2.append(mylist2iter[i][1])

listafter2 = set(listafter2)
```

# Moving forward to K=2 where we create combinations of two:

 Create\_combinations2 will have combination of two item sets appearing together in each transaction satisfying min\_sup.

```
## Creating combinations of items in Iteration 2
def create_combinations2(targlist, n):
    comb = combinations(targlist, n)
    up_comb = []
    for i in comb:
        up_comb.append(i)
    return up_comb
```

# Moving forward to K=3 where we create combinations of three:

• We list all elements that comes in pair of three in each transaction.

```
## Performing Iteration 3
def iter3(final_data, combs_3):
    updated_dict3 = {}
    for i in range(len(final_data)):
        c2=0
        for j in range(len(combs_3)):
            if (combs_3[j][0] and combs_3[j][1] and combs_3[j][2]) in final_data[i]:
            c2= c2+1
            updated_dict3[combs_3[j]] = c2

return updated_dict3
```

# Implementation in Main for K=3:

```
combs_3 = create_combinations2(listafter2, 3)
iteration3 = iter3(final_data, combs_3)
sup_elim3 = support_eliminate(iteration3)
print("\n","This is the result after third iteration:", "\n", "\n", sup_elim3, '\n')
```

\*

## **Implementation of Association Rules:**

- Highkey is tuple created in main that is the most frequent
- Highval has the value which is count of how many times items appeared
- Since, rules are derivate from most frequent set so we take the possible combinations from most frequent set and divide by the set itself
- We make combinations of 2 from most frequent set that is why comb2s variable is storing highkey that is value of 2 pairs and combination built-in function is making all possible combinations.
- Left item is the combination created whereas right one is the element left out from created combination that is why set is used
- o If condition eliminates rules/item sets based on minimum confidence.

```
## List of Frequent itemsets and Association rules
def assoc rules(highkey, highval):
    support_dict = {}
    count = 0
    item set = set(highkey)
    comb2s = combinations(highkey, 2)
    for i in comb2s:
        print("Association Rule for Pair - ", i)
        print("__Rule__", "__Confidence__")
        i = tuple(sorted(i))
        left item = i
        right_item = tuple(item_set - set(i))
        # print(left_item, "=>" ,right_item)
        denom2 = sup elim2[i]
        conf = (highval // denom2) * 100
        if conf >= min conf:
            print(left_item, "=>", right_item, conf, "This Rule is Acceptable")
        else:
            print(left_item, "=>", right_item, conf, "This Rule is Rejected")
```

Again, association rules are taken for rules created as a->{b,c} where left item is the remaining item from item set pair created.

```
comb1s = highkey
for i in comb1s:
    print("Association Rule for Pair - ", i)
    print("__Rule__", "__Confidence__")

left_item = i
    right_item = tuple(item_set - set(i))
    denom1 = sup_elim1[i]
    conf1 = highval * 100 // denom1

# print(left_item, right_item)
    if conf1 >= min_conf:
        print(left_item, "=>", right_item, conf1, "This Rule is Acceptable")
    else:
        print(left_item, "=>", right_item, conf1, "This Rule is Rejected")
print()
```

# **Implementation in Main:**

```
highkey = tuple(sorted(list(iteration3.keys())[list(iteration3.values()).index(max(iteration3.values()))]))
highval = max(iteration3.values())
assoc_rules(highkey, highval)
```

\*

# How to Run:

→ Example: Python my\_midtermproj.py

It will ask for Dataset:

Give any dataset then you want to give in txt:

I have used 5 datasets here:

- Nike.txt
- BestBuy.txt
- Amazon.txt
- Kmart.txt
- Generic.txt (for clarity)



```
C:\Users\Administrator\AppData\Local\Programs\Python\Python39\Scripts\Project>Python my_midtermproj.py

Welcome to Anusha's Apriori Algorithm Implementation.
This program will take input date of transactions from you and
give you the most frequent association rules with your product lines in return.
This can give you an idea about what products sell in a combination following them can maximize your profit.

Enter File name with '.txt' extension:BestBuy.txt

This is the dataset you entered:

[[' Desk Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Speakers', 'Anti-Virus'], [' Lab Top', 'Flash Drive', 'Microsoft Office', 'Anti-Virus'], [' Lab Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Anti-Virus', 'External Hand-Drive', 'Lab Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Anti-Virus', 'External Hand-Drive', 'Lab Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Lab Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Lab Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Lab Top Case', 'Anti-Virus', 'External Hand-Drive', 'Anti-Virus', 'Digital Camera', 'Lab Top Case', 'Anti-Virus', 'Speakers', 'External Hand-Drive', 'Anti-Virus', 'External Hand-Drive', 'Anti-Virus', 'Speakers', '( Digital Camera', 'Lab Top Case', 'Anti-Virus', 'Speakers'), '[ Digital Camera', 'Speakers', '( Digital Camera', 'Speakers'), '( Digital Camera', 'Lab Top Case', 'Anti-Virus', 'Lab Top Case', 'Anti-Virus', 'La
```

- →Now give minimum support
- →Then, minimum confidence

## This is how it will display support:

```
Enter File name with '.txt' extension:BestBuy.txt

This is the dataset you entered:

[['Desk Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Speakers', 'Anti-Virus'], ['Lab Top', 'Flash Drive', 'Microsoft Office', 'Anti-Virus'], ['Lab Top', 'Printer', 'Flash Drive', 'Anti-Virus'], ['Lab Top', 'Printer', 'Flash Drive', 'Anti-Virus'], ['Lab Top', 'Printer', 'Flash Drive', 'Anti-Virus'], ['Desk Top', 'Printer', 'Flash Drive', 'Microsoft Office'], 'Lab Top', 'Speakers', 'Anti-Virus', 'Desk Top', 'Printer', 'Flash Drive', 'Microsoft Office'], 'Lab Top', 'Speakers', 'Lab Top', 'Desk Top', 'Desk Top', 'Printer', 'Flash Drive', 'Microsoft Office', 'Lab Top', 'Desk Top', 'D
```

```
This is the result after third iteration:

(('Lab Top', 'Lab Top Case', 'Microsoft Office'): 6, ('Lab Top', 'Anti-Virus', 'Flash Drive'): 8, ('Lab Top', 'Anti-Virus', 'Printer'): 9, ('Lab Top', 'Anti-Virus', 'Microsoft Office'): 11, ('Lab Top', 'Anti-Virus', 'Desk Top'): 18, ('Lab Top', 'Digital Camera', 'Flash Drive'): 12, ('Lab Top', 'Anti-Virus', 'Desk Top'): 18, ('Lab Top', 'Digital Camera', 'Flash Drive'): 11, ('Lab Top', 'Digital Camera', 'Printer'): 12, ('Lab Top', 'Digital Camera', 'Printer'): 13, ('Lab Top', 'Digital Camera', 'Printer'): 14, ('Lab Top', 'Desk Top'): 12, ('Lab Top', 'Printer'): 14, ('Lab Top Case', 'Anti-Virus', 'Printer'): 14, ('Lab Top Case', 'Digital Camera', 'Printer'): 12, ('Lab Top Case', 'Anti-Virus', 'Printer'): 12, ('Lab Top Case', 'Digital Camera', 'Printer'): 12, ('Anti-Virus', 'Printer'): 12, ('Anti-Virus', 'Printer'): 13, ('Lab Top Case', 'Digital Camera', 'Printer'): 13, ('Lab Top Case', 'Speakers', 'Desk Top'): 13, ('Lab Top Case', 'Speakers', 'Desk Top'): 13, ('Lab Top Case', 'Speakers', 'Desk Top'): 13, ('Anti-Virus', 'Printer'): 14, ('Anti-Virus', 'Printer'): 14, ('Anti-Virus', 'Printer'):
```

## Association rules for BestBuy display:

```
Association Rule for Pair - (' Desk Top', ' External Hard-Drive')

_Rule___Confidence__
(' Desk Top', ' External Hard-Drive') => (' Microsoft Office',) 900 This Rule is Acceptable
Association Rule for Pair - (' Desk Top', ' Microsoft Office')

_Rule___Confidence__
(' Desk Top', ' Microsoft Office') => (' External Hard-Drive',) 600 This Rule is Acceptable
Association Rule for Pair - (' External Hard-Drive', ' Microsoft Office')

_Rule___Confidence_
(' External Hard-Drive', ' Microsoft Office') => (' Desk Top',) 300 This Rule is Acceptable
Association Rule for Pair - Desk Top

_Rule___Confidence_
Desk Top => (' Desk Top', ' External Hard-Drive', ' Microsoft Office') 1266 This Rule is Acceptable
Association Rule for Pair - External Hard-Drive

_Rule___Confidence_
External Hard-Drive => (' Desk Top', ' External Hard-Drive', ' Microsoft Office') 844 This Rule is Acceptable
Association Rule for Pair - Microsoft Office

_Rule___Confidence_
External Hard-Drive => (' Desk Top', ' External Hard-Drive', ' Microsoft Office') 690 This Rule is Acceptable

Microsoft Office => (' Desk Top', ' External Hard-Drive', ' Microsoft Office') 690 This Rule is Acceptable
```

## NIKE:

#### **Support:**

This is the result after third iteration:

```
::\Users\Administrator\AppData\Local\Programs\Python\Python39\Scripts\Project>Python my_midtermproj.py
Welcome to Anusha's Apriori Algorithm Implementation.
This program will take input data of transactions from you and
give you the most frequent association rules with your product lines in return.
This can give you an idea about what products sell in a combination following them can maximize your profit.
 This is the dataset you entered:
 [['Running Shoe', 'Socks', 'Sweatshirts', 'Modern Pants'], ['Running Shoe', 'Socks', 'Sweatshirts'], ['Running Shoe', 'Socks', 'Sweatshirts', 'Modern Pants'], ['Running Shoe', 'Socks', 'Sweatshirts', 'Modern Pants'], ['Running Shoe', 'Socks', 'Sweatshirts'], ['Running Shoe', 'Socks', 'Sweatshirts'], ['Running Shoe', 'Socks', 'Sweatshirts', 'Modern Pants', 'Rash Guard', 'Hoodies'], ['Swimming Shirt', 'Socks', 'Sweatshirts'], ['Swimming Shirt', 'Rash Guard', 'Dry Fit V-Nick'], ['Running Shoe', 'Swimming Shirt', 'Socks', 'Sweatshirts', 'Modern Pants', 'Soccer Shoe', 'Rash Guard', 'Hoodies', 'Tech Pants', 'Dry Fit V-Nick'], ['Running Shoe', 'Swimming Shirt', 'Socks', 'Sweatshirts', 'Modern Pants', 'Soccer Shoe', 'Rash Guard', 'Tech Pants', 'Dry Fit V-Nick'], ['Running Shoe', 'Swimming Shirt', 'Socks', 'Sweatshirts', 'Modern Pants', 'Dry Fit V-Nick'], ['Running Shoe', 'Swimming Shirt', 'Socks', 'Sweatshirts', 'Modern Pants', 'Dry Fit V-Nick', Rash Guard', 'Tech Pants', 'Dry Fit V-Nick', ['Running Shoe', 'Swimming Shirt', 'Socks', 'Sweatshirts', 'Modern Pants', 'Dry Fit V-Nick', 'Rash Guard', 'Tech Pants', 'Rash Guard'], 'Running Shoe', 'Swimming Shirt', 'Soccer Shoe', 'Hoodies', 'Dry Fit V-Nick', 'Rash Guard', 'Tech Pants', 'Rash Guard', 'Running Shoe', 'Swimming Shirt', 'Soccer Shoe', 'Hoodies', 'Rash Guard', 'Rash
```

```
nter Minimum Support:20
  nter Minimum Confidence:50
 This is the result after first iteration:
 (' Dry Fit V-Nick': 9, ' Hoodies': 8, ' Modern Pants': 10, ' Rash Guard': 12, ' Running Shoe': 14, ' Soccer Shoe': 6, ' Socks': 13, ' Sweatshirts': 13, ' Swimming Shirt': 11, ' Tech Pants': 9}
  This is the result after second iteration:
(('Dry Fit V-Nick', 'Running Shoe'): 4, ('Dry Fit V-Nick', 'Socks'): 6, ('Dry Fit V-Nick', 'Sweatshirts'): 7, ('Hoodies', 'Modern Pants'): 10, ('Hoodies', 'Indiana'): 11, ('Hoodies', 'Sweatshirts'): 12, ('Hoodies', 'Socks'): 14, ('Hoodies', 'Sweatshirts'): 15, ('Modern Pants', 'Running Shoe'): 19, ('Modern Pants', 'Socks'): 21, ('Modern Pants', 'Sweatshirts'): 22, ('Rash Guard', 'Running Shoe', 'Sweatshirts'): 22, ('Rash Guard', 'Running Shoe', 'Sweatshirts'): 28, ('Running Shoe', 'Socks'): 23, ('Running Shoe', 'Sweatshirts'): 28, ('Running Shoe', 'Socks'): 23, ('Running Shoe', 'Sweatshirts'): 40, ('Pry Fit V-Nick', 'Soccer Shoe', 'Sweatshirts'): 40, ('Pry Fit V-Nick', 'Soccer Shoe'): 26, ('Running Shoe', 'Soccer Shoe'): 31, ('Dry Fit V-Nick', 'Running Shoe', 'Sweatshirts'): 40, ('Rash Guard', 'Tech Pants'): 29, ('Running Shoe', 'Sweatshirts'): 40, ('Rash Guard', 'Tech Pants'): 29, ('Running Shoe', 'Tech Pants'): 41, ('Swimming Shirt', 'Tech Pants'): 42, ('Sweatshirts', 'Tech Pants'): 43)
```

```
This is the result after third iteration:

(('Hoodies', Tech Pants', Modern Pants'): 5, ('Hoodies', Tech Pants', Running Shoe'): 7, ('Hoodies', Sudming Shirt', Socks'): 9, ('Hoodies', Sudming Shirt', Skeatshirts'): 10, ('Hoodies', Sudming Shirt', Modern Pants'): 12, ('Hoodies', Sudming Shirt', Running Shoe'): 41, 'Hoodies', Socks', Socks',
```

#### **Association Rules for Nike:**

```
Association Rule for Pair - (' Rash Guard', ' Running Shoe')

Rule ___Confidence_
(' Rash Guard', ' Running Shoe') => (' Soccer Shoe',) 400 This Rule is Acceptable
Association Rule for Pair - (' Rash Guard', ' Soccer Shoe')

_Rule ___Confidence_
(' Rash Guard', ' Soccer Shoe') => (' Running Shoe',) 400 This Rule is Acceptable
Association Rule for Pair - (' Running Shoe', ' Soccer Shoe')

_Rule ___Confidence_
(' Running Shoe', ' Soccer Shoe') => (' Rash Guard',) 300 This Rule is Acceptable
Association Rule for Pair - Rash Guard

_Rule ___Confidence_
Rash Guard => (' Rash Guard', ' Running Shoe', ' Soccer Shoe') 1000 This Rule is Acceptable
Association Rule for Pair - Running Shoe

_Rule ___Confidence_
Running Shoe => (' Rash Guard', ' Running Shoe', ' Soccer Shoe') 857 This Rule is Acceptable
Association Rule for Pair - Soccer Shoe

_Rule ___Confidence_
Rule ___Confidence_
Soccer Shoe => (' Rash Guard', ' Running Shoe', ' Soccer Shoe') 2000 This Rule is Acceptable
```

## **Amazon:**

#### **Support:**

```
C:\Users\Administrator\AppData\Local\Programs\Python\Python39\Scripts\Project>Python my_midtermproj.py

Welcome to Anusha's Apriori Algorithm Implementation.
This program will take input data of transactions from you and
give you the most frequent association rules with your product lines in return.
This can give you an idea about what products sell in a combination following them can maximize your profit.

Enter File name with '.txt' extension:Amazon.txt

This is the dataset you entered:

[['A Beginner's Guide', 'Java: The Complete Reference', 'Java For Dummies', 'Android Programming: The Big Nerd Ranch'], ['A Beginner's Guide', 'Java: The Complete
Reference', 'Java For Dummies'], ['Android Programming: The Big Nerd Ranch', 'Head First
Java 2nd Edition'], ['A Ageginner's Guide', 'Java: The Complete Reference', 'Java For Dummies', 'Android Programming: The Big Nerd Ranch', 'Head First
Java', 'Java Beginning Programming with Java', 'Java For Dummies', 'Android Programming with Java', 'Java Beginning Programming with Java', 'Java Beginner's Guide', 'Hada First Java 2nd Edition', 'Beginning Programming with Java', 'Java Beginner's Guide', 'Java: The Complete Reference', 'Java For Dummies', 'Android Programming: The Big Nerd Ranch', 'Head First Java 2nd Edition', 'Java: The Complete Reference', 'Java: The Complete Reference', 'Java: The Complete Reference', 'Java For Dummies', 'Android Programming: The Big Nerd Ranch', 'Hada First Java 2nd Edition', 'Java: The Complete Reference', 'Java For Dummies', 'Android Programming: The Big Nerd Ranch', 'Hada First Java 2nd Edition', 'Java: The Complete Reference', 'Java For Dummies', 'Android Programming: The Big Nerd Ran
```

```
Enter Minimum Confidence:10

This is the result after first iteration:

('A Beginner's Guide': 11, 'Android Programming: The Big Nerd Ranch': 13, 'Beginning Programming with Java': 6, 'Head First Java 2nd Edition': 8, 'Java 8 Pocket Guide': 4, 'Java For Dummies': 13, 'Java: The Complete Reference': 10)

This is the result after second iteration:

(('Android Programming: The Big Nerd Ranch', 'Java: The Complete Reference'): 4, ('Beginning Programming with Java', 'Java For Dummies'): 5, ('Beginning Programming with Java', 'Java For Dummies'): 6, 'Beginning Programming with Java', 'Java For Dummies'): 9, ('Java 8 Pocket Guide', 'Java: The Complete Reference'): 11, ('Beginning Programming with Java', 'Head First Java 2nd Edition', 'Java: The Complete Reference'): 11, ('Beginning Programming with Java', 'Head First Java 2nd Edition'): 4, ('Android Programming: The Big Nerd Ranch', 'Java: The Complete Reference'): 13, ('Java For Dummies', 'Java: The Complete Reference'): 13, ('Java For Dummies', 'Java: The Complete Reference'): 13, ('Java For Dummies', 'Java: The Complete Reference'): 14, ('Java For Dummies', 'Java: The Complete Reference'): 15, ('Java For Dummies', 'Java: The Complete Reference', Beginning Programming with Java'): 15, ('Java: The Complete Reference', 'Beginnin
```

#### **Association rules for Amazon:**

```
Association Rule for Pair - ('Android Programming: The Big Nerd Ranch', 'Beginning Programming with Java')

_Rule __Confidence_
('Android Programming: The Big Nerd Ranch', 'Beginning Programming with Java') => ('Java: The Complete Reference',) 200 This Rule is Acceptable
Association Rule for Pair - ('Android Programming: The Big Nerd Ranch', 'Java: The Complete Reference')

_Rule __Confidence_
('Android Programming: The Big Nerd Ranch', 'Java: The Complete Reference') >> ('Beginning Programming with Java',) 200 This Rule is Acceptable
Association Rule for Pair - ('Beginning Programming with Java', 'Java: The Complete Reference') => ('Android Programming: The Big Nerd Ranch',) 100 This Rule is Acceptable
Association Rule for Pair - Android Programming: The Big Nerd Ranch
Rule __Confidence_
Android Programming: The Big Nerd Ranch => ('Beginning Programming with Java', 'Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 84 This Rule is Acceptable
Association Rule for Pair - Beginning Programming with Java
_Rule __Confidence_
_Beginning Programming with Java => ('Beginning Programming with Java', 'Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 183 This Rule is Acceptable
Association Rule for Pair - Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 183 This Rule is Acceptable
Association Rule for Pair - Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 180 This Rule is Acceptable
Association Rule for Pair - Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 180 This Rule is Acceptable
Association Rule for Pair - Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 180 This Rule is Acceptable
Association Rule for Pair - Java: The Complete Reference', 'Android Programming: The Big Nerd Ranch') 110 This Rule is Acceptable
```

#### K Mart:

## **Support:**

```
Notes to Annual's Agrical Aggrithm Explosuration.

This program will take input date of transactions from you and give you the most frequent association rules with your product lines in return.

This can give you mides about with products sell in a combination following them can maximize your profit.

Enter File name with '.txt' extension:Datat.txt

This is the dataset you entered:

[[] Decorative Pillows', Oulits', 'Georgical Design of the Company of
```

C:\Users\Administrator\AppData\Local\Programs\Python\Python39\Scripts\Project>Python my\_midtermproj.py

#### Association rules for K Mart:

```
Association Rule for Pair - ('Embroidered Bedspread', 'Kids Bedding')

_Rule___Confidence_
('Embroidered Bedspread', 'Kids Bedding') => ('Shams',) 200 This Rule is Acceptable
Association Rule for Pair - ('Embroidered Bedspread', 'Shams')

_Rule___Confidence_
('Embroidered Bedspread', 'Shams') => ('Kids Bedding',) 200 This Rule is Acceptable
Association Rule for Pair - ('Kids Bedding', 'Shams')

_Rule___Confidence_
('Kids Bedding', 'Shams') => ('Embroidered Bedspread',) 100 This Rule is Acceptable
Association Rule for Pair - Embroidered Bedspread

_Rule___Confidence_
Embroidered Bedspread => ('Shams', 'Embroidered Bedspread', 'Kids Bedding') 683 This Rule is Acceptable
Association Rule for Pair - Kids Bedding

_Rule___Confidence_
Kids Bedding => ('Shams', 'Embroidered Bedspread', 'Kids Bedding') 341 This Rule is Acceptable
Association Rule for Pair - Shams

_Rule___Confidence_
Shams => ('Shams', 'Embroidered Bedspread', 'Kids Bedding') 372 This Rule is Acceptable
```

## **Generic Dataset:**

\*testing purpose\*

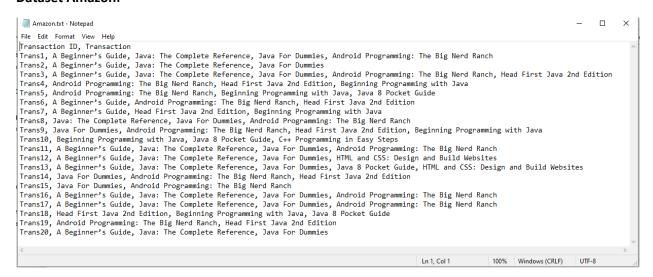
## **Support and Association rules:**

```
Enter File name with '.txt' extension:Generic.txt

This is the dataset you entered:

[[' A', ' B', ' C'], [' A', ' B', ' C'], [' A', ' B', ' C', ' D'], [' A', ' B', ' C', ' D', ' E'], [' A', ' B', ' D', ' E'], [' A', ' D', ' E'], [' A', ' E'], [' A', ' C', ' E']]
```

#### **Dataset Amazon:**



#### **Dataset Nike:**



#### **Dataset BestBuy:**

#### **Dataset K Mart:**



#### **Dataset Generic:**

## I kept this to verify my work

```
Generic.txt - Notepad

File Edit Format View Help

Transaction ID, Transaction

Trans1, A, B, C

Trans2, A, B, C

Trans3, A, B, C, D

Trans4, A, B, C, D, E

Trans5, A, B, D, E

Trans6, A, D, E

Trans7, A, E

Trans8, A, E

Trans9, A, C, E

Trans10, A, C, E

Trans11, A, C, E
```

# Final compiled code (py files):

```
Author: Anusha Syed
Project Introduction:

## Set the directory as under before running the code
## Keep the program file and the dataset in the same directory

"""

from itertools import permutations
from itertools import combinations

## Defining all the Functions

## Find the count of unique items and make a dictionary
def uniquelist(final_data):
    mylist = []
    for i in final_data:
        for j in range(len(i)):
            mylist.append(i[j])
```

```
uniquelist = set(mylist)
    uniquelist = sorted(uniquelist)
    return uniquelist, mylist
## Performing Iteration 1
def iteration1(mylist, uniquelist):
    firstdict = {}
    for i in uniquelist:
        firstdict[i] = mylist.count(i)
    return firstdict
## Eliminate those items not satisfying minimum support
def support eliminate(firstdict):
    keylist = list(firstdict.keys())
    valuelist = list(firstdict.values())
    mydict2 = {}
    # support dict = {}
    count = 0
    for i in valuelist:
        support = round(i * 100 / total trxns, 2)
        count += 1
        # print(support, min sup)
        if support >= min sup:
            # firstdict.pop(keylist[count])
            mydict2[keylist[count - 1]] = i
            # support dict[keylist[count-1]] = support
    return mydict2
## Creating combinations of items in Iteration 1
def create combinations1(mydict2, n):
    comb = combinations(list(mydict2.keys()), n)
    up comb = []
    for i in comb:
        up comb.append(i)
    return up comb
## Performing Iteration 2
def iter2(final data, up comb):
    updated dict = {}
    for i in range(len(final data)):
```

```
c2 = 0
        for j in range(len(up comb)):
            if (up comb[j][0] and up comb[j][1]) in final data[i]:
                updated dict[up comb[j]] = c2
    return updated dict
## Creating combinations of items in Iteration 2
def create combinations2(targlist, n):
    comb = combinations(targlist, n)
    up comb = []
    for i in comb:
        up comb.append(i)
    return up comb
## Performing Iteration 3
def iter3(final data, combs 3):
    updated dict3 = {}
    for i in range(len(final data)):
        c2 = 0
        for j in range(len(combs 3)):
            if (combs 3[j][0] and combs 3[j][1] and combs 3[j][2]) in fina
l data[i]:
              c2 = c2 + 1
              updated dict3[combs 3[j]] = c2
    return updated dict3
## List of Frequent itemsets and Association rules
def assoc rules(highkey, highval):
    support dict = {}
    count = 0
    item set = set(highkey)
    comb2s = combinations(highkey, 2)
    for i in comb2s:
        print("Association Rule for Pair - ", i)
        print(" Rule ", " Confidence ")
        i = tuple(sorted(i))
        left item = i
        right item = tuple(item set - set(i))
```

```
# print(left item, "=>" , right item)
        denom2 = sup elim2[i]
        conf = (highval // denom2) * 100
        if conf >= min conf:
            print(left item, "=>", right item, conf, "This Rule is Accepta
ble")
        else:
            print(left item, "=>", right item, conf, "This Rule is Rejecte
d")
    combls = highkey
    for i in comb1s:
        print("Association Rule for Pair - ", i)
        print(" Rule ", " Confidence ")
       left item = i
        right item = tuple(item set - set(i))
        denom1 = sup elim1[i]
        conf1 = highval * 100 // denom1
        # print(left item, right item)
        if conf1 >= min conf:
            print(left item, "=>", right item, conf1, "This Rule is Accept
able")
        else:
            print(left item, "=>", right item, conf1, "This Rule is Reject
ed")
   print()
## Main program to run Algorithm using the functions defined
print("\n","Welcome to Anusha's Apriori Algorithm Implementation. ", '\n',
      "This program will take input data of transactions from you and ", '
\n',
      "give you the most frequent association rules with your product line
s in return. ", '\n',
      "This can give you an idea about what products sell in a combination
 following them can maximize your profit. ", "\n")
filename = input("Enter File name with '.txt' extension:")
```

```
fileobject = open(filename, "r", encoding="UTF-
8", errors="surrogateescape")
input data = fileobject.readlines()
## Transforming the dataset
final data = []
for lines in range(1,len(input data)):
    input data[lines] = input data[lines].replace("\n","")
    final data.append(input data[lines].split(','))
## Transforming final data
for i in range(len(final data)):
  final data[i] = final data[i][1:]
print("\n","This is the dataset you entered:", "\n", "\n", final data, '\n
• )
## Declaring all the required variables
min sup = int(input("Enter Minimum Support:"))
print("\n")
min conf = int(input("Enter Minimum Confidence:"))
print("\n")
total trxns = len(final data)
uniques, full list = uniquelist(final data)
iteration1 = iteration1(full list, uniques)
sup elim1 = support eliminate(iteration1)
print("\n", "This is the result after first iteration:", "\n", "\n", sup el
im1, '\n')
new combs = create combinations1(sup elim1, 2)
iteration2 = iter2(final data, new combs)
sup elim2 = support eliminate(iteration2)
print("\n", "This is the result after second iteration:", "\n", "\n", sup e
lim2, '\n')
mylist2iter = list(sup elim2.keys())
```

```
##used this to put unique pairs in tuple to make it in a single list
listafter2 = []
for i in range(len(mylist2iter)):
    listafter2.append(mylist2iter[i][0])
    listafter2.append(mylist2iter[i][1])

listafter2 = set(listafter2)

combs_3 = create_combinations2(listafter2, 3)

iteration3 = iter3(final_data, combs_3)

sup_elim3 = support_eliminate(iteration3)

print("\n", "This is the result after third iteration:", "\n", sup_elim3, '\n')

highkey = tuple(sorted(list(iteration3.keys())[list(iteration3.values()).index(max(iteration3.values())])

highval = max(iteration3.values())

assoc rules(highkey, highval)
```