Assignment 2 Set B1(1)

March 12, 2024

1 Download the Market basket dataset. Write a python program to read the dataset and display its information. Preprocess the data (drop null values etc.) Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets and association rules.

```python
[1]: import pandas as pd
     import numpy as np
     from apyori import apriori
     df=pd.read_csv('/home/mmcc/Desktop/DA Data Sets/Market_Basket_Optimisation.csv')
     print(df)
     print(df.info())
     print("Shape : \n",df.shape)
     print("Size : \n",df.size)
```

```
              shrimp          almonds       avocado     vegetables mix  \
0             burgers        meatballs          eggs                NaN
1             chutney              NaN           NaN                NaN
2             turkey           avocado           NaN                NaN
3       mineral water            milk     energy bar  whole wheat rice
4      low fat yogurt              NaN           NaN                NaN
...              ...              ...           ...                ...
7495           butter       light mayo   fresh bread                NaN
7496          burgers  frozen vegetables         eggs      french fries
7497          chicken              NaN           NaN                NaN
7498         escalope         green tea           NaN                NaN
7499             eggs  frozen smoothie   yogurt cake    low fat yogurt

       green grapes whole weat flour yams cottage cheese energy drink  \
0               NaN              NaN  NaN           NaN          NaN
1               NaN              NaN  NaN           NaN          NaN
2               NaN              NaN  NaN           NaN          NaN
3          green tea              NaN  NaN           NaN          NaN
4               NaN              NaN  NaN           NaN          NaN
...             ...              ...  ...           ...          ...
7495            NaN              NaN  NaN           NaN          NaN
```

```
7496      magazines      green tea  NaN           NaN          NaN
7497            NaN            NaN  NaN           NaN          NaN
7498            NaN            NaN  NaN           NaN          NaN
7499            NaN            NaN  NaN           NaN          NaN

      tomato juice low fat yogurt green tea honey salad mineral water salmon  \
0              NaN            NaN       NaN   NaN   NaN           NaN    NaN
1              NaN            NaN       NaN   NaN   NaN           NaN    NaN
2              NaN            NaN       NaN   NaN   NaN           NaN    NaN
3              NaN            NaN       NaN   NaN   NaN           NaN    NaN
4              NaN            NaN       NaN   NaN   NaN           NaN    NaN
...            ...            ...       ...   ...   ...           ...    ...
7495           NaN            NaN       NaN   NaN   NaN           NaN    NaN
7496           NaN            NaN       NaN   NaN   NaN           NaN    NaN
7497           NaN            NaN       NaN   NaN   NaN           NaN    NaN
7498           NaN            NaN       NaN   NaN   NaN           NaN    NaN
7499           NaN            NaN       NaN   NaN   NaN           NaN    NaN

      antioxydant juice frozen smoothie spinach  olive oil
0                   NaN             NaN     NaN         NaN
1                   NaN             NaN     NaN         NaN
2                   NaN             NaN     NaN         NaN
3                   NaN             NaN     NaN         NaN
4                   NaN             NaN     NaN         NaN
...                 ...             ...     ...         ...
7495                NaN             NaN     NaN         NaN
7496                NaN             NaN     NaN         NaN
7497                NaN             NaN     NaN         NaN
7498                NaN             NaN     NaN         NaN
7499                NaN             NaN     NaN         NaN

[7500 rows x 20 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7500 entries, 0 to 7499
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   shrimp            7500 non-null   object
 1   almonds           5746 non-null   object
 2   avocado           4388 non-null   object
 3   vegetables mix    3344 non-null   object
 4   green grapes      2528 non-null   object
 5   whole weat flour  1863 non-null   object
 6   yams              1368 non-null   object
 7   cottage cheese    980 non-null    object
 8   energy drink      653 non-null    object
 9   tomato juice      394 non-null    object
 10  low fat yogurt    255 non-null    object
```

2

```
11   green tea            153 non-null     object
12   honey                86 non-null      object
13   salad                46 non-null      object
14   mineral water        24 non-null      object
15   salmon               7 non-null       object
16   antioxydant juice    3 non-null       object
17   frozen smoothie      3 non-null       object
18   spinach              2 non-null       object
19   olive oil            0 non-null       float64
dtypes: float64(1), object(19)
memory usage: 1.1+ MB
None
Shape :
 (7500, 20)
Size :
 150000
```

[2]:
```python
#converting dataframe into list of lists
l=[]
for i in range(0,len(df)):
    l.append([str(df.values[i,j]) for j in range(0,len(df.columns))])

print(type(l))
```

```
<class 'list'>
```

[7]:
```python
association_rules = apriori(l, min_support=0.0045, min_confidence=0.2,
 →min_lift=3, min_length=2)
association_results = list(association_rules)
print("There are {} Relation derived.".format(len(association_results)))
```

```
There are 48 Relation derived.
```

[5]:
```python
for i in range(0, len(association_results)):
    print(association_results[i][0])
```

```
frozenset({'chicken', 'light cream'})
frozenset({'mushroom cream sauce', 'escalope'})
frozenset({'pasta', 'escalope'})
frozenset({'herb & pepper', 'ground beef'})
frozenset({'ground beef', 'tomato sauce'})
frozenset({'olive oil', 'whole wheat pasta'})
frozenset({'shrimp', 'pasta'})
frozenset({'chicken', 'light cream', 'nan'})
frozenset({'frozen vegetables', 'shrimp', 'chocolate'})
frozenset({'cooking oil', 'spaghetti', 'ground beef'})
frozenset({'mushroom cream sauce', 'nan', 'escalope'})
frozenset({'pasta', 'nan', 'escalope'})
```

```
frozenset({'frozen vegetables', 'spaghetti', 'ground beef'})
frozenset({'frozen vegetables', 'milk', 'olive oil'})
frozenset({'frozen vegetables', 'shrimp', 'mineral water'})
frozenset({'frozen vegetables', 'spaghetti', 'olive oil'})
frozenset({'frozen vegetables', 'spaghetti', 'shrimp'})
frozenset({'frozen vegetables', 'spaghetti', 'tomatoes'})
frozenset({'grated cheese', 'spaghetti', 'ground beef'})
frozenset({'herb & pepper', 'mineral water', 'ground beef'})
frozenset({'herb & pepper', 'nan', 'ground beef'})
frozenset({'herb & pepper', 'spaghetti', 'ground beef'})
frozenset({'milk', 'olive oil', 'ground beef'})
frozenset({'nan', 'ground beef', 'tomato sauce'})
frozenset({'spaghetti', 'shrimp', 'ground beef'})
frozenset({'spaghetti', 'milk', 'olive oil'})
frozenset({'soup', 'olive oil', 'mineral water'})
frozenset({'olive oil', 'nan', 'whole wheat pasta'})
frozenset({'shrimp', 'pasta', 'nan'})
frozenset({'spaghetti', 'olive oil', 'pancakes'})
frozenset({'frozen vegetables', 'shrimp', 'nan', 'chocolate'})
frozenset({'spaghetti', 'cooking oil', 'nan', 'ground beef'})
frozenset({'frozen vegetables', 'spaghetti', 'nan', 'ground beef'})
frozenset({'frozen vegetables', 'spaghetti', 'milk', 'mineral water'})
frozenset({'olive oil', 'frozen vegetables', 'milk', 'nan'})
frozenset({'frozen vegetables', 'shrimp', 'nan', 'mineral water'})
frozenset({'frozen vegetables', 'spaghetti', 'olive oil', 'nan'})
frozenset({'frozen vegetables', 'spaghetti', 'shrimp', 'nan'})
frozenset({'frozen vegetables', 'spaghetti', 'nan', 'tomatoes'})
frozenset({'spaghetti', 'grated cheese', 'nan', 'ground beef'})
frozenset({'herb & pepper', 'mineral water', 'nan', 'ground beef'})
frozenset({'herb & pepper', 'spaghetti', 'nan', 'ground beef'})
frozenset({'olive oil', 'milk', 'nan', 'ground beef'})
frozenset({'spaghetti', 'shrimp', 'nan', 'ground beef'})
frozenset({'olive oil', 'spaghetti', 'milk', 'nan'})
frozenset({'soup', 'olive oil', 'nan', 'mineral water'})
frozenset({'spaghetti', 'olive oil', 'nan', 'pancakes'})
frozenset({'spaghetti', 'nan', 'frozen vegetables', 'milk', 'mineral water'})
```

[8]:
```python
for item in association_results:
    # first index of the inner list
    # Contains base item and add item
    pair = item[0]
    items = [x for x in pair]
    print("Rule: " + items[0] + " -> " + items[1])
    # second index of the inner list
    print("Support: " + str(item[1]))
    # third index of the list located at 0th position
    # of the third index of the inner list
```

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

Rule: chicken -> light cream
Support: 0.004533333333333334
Confidence: 0.2905982905982906
Lift: 4.843304843304844

Rule: mushroom cream sauce -> escalope
Support: 0.005733333333333333
Confidence: 0.30069930069930073
Lift: 3.7903273197390845

Rule: pasta -> escalope
Support: 0.005866666666666667
Confidence: 0.37288135593220345
Lift: 4.700185158809287

Rule: herb & pepper -> ground beef
Support: 0.016
Confidence: 0.3234501347708895
Lift: 3.2915549671393096

Rule: ground beef -> tomato sauce
Support: 0.005333333333333333
Confidence: 0.37735849056603776
Lift: 3.840147461662528

Rule: olive oil -> whole wheat pasta
Support: 0.008
Confidence: 0.2714932126696833
Lift: 4.130221288078346

Rule: shrimp -> pasta
Support: 0.005066666666666666
Confidence: 0.3220338983050848
Lift: 4.514493901473151

Rule: chicken -> light cream
Support: 0.004533333333333334
Confidence: 0.2905982905982906
Lift: 4.843304843304844

Rule: frozen vegetables -> shrimp
Support: 0.005333333333333333
Confidence: 0.23255813953488372

```
Lift: 3.260160834601174
--------------------------------------------------------
Rule: cooking oil -> spaghetti
Support: 0.0048
Confidence: 0.5714285714285714
Lift: 3.281557646029315
--------------------------------------------------------
Rule: mushroom cream sauce -> nan
Support: 0.005733333333333333
Confidence: 0.30069930069930073
Lift: 3.7903273197390845
--------------------------------------------------------
Rule: pasta -> nan
Support: 0.005866666666666667
Confidence: 0.37288135593220345
Lift: 4.700185158809287
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.008666666666666666
Confidence: 0.3110047846889952
Lift: 3.164906221394116
--------------------------------------------------------
Rule: frozen vegetables -> milk
Support: 0.0048
Confidence: 0.20338983050847456
Lift: 3.094165778526489
--------------------------------------------------------
Rule: frozen vegetables -> shrimp
Support: 0.0072
Confidence: 0.3068181818181818
Lift: 3.2183725365543547
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.005733333333333333
Confidence: 0.20574162679425836
Lift: 3.1299436124887174
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.006
Confidence: 0.21531100478468898
Lift: 3.0183785717479763
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.006666666666666667
Confidence: 0.23923444976076555
Lift: 3.497579674864993
--------------------------------------------------------
Rule: grated cheese -> spaghetti
```

```
Support: 0.005333333333333333
Confidence: 0.3225806451612903
Lift: 3.282706701098612
--------------------------------------------------------
Rule: herb & pepper -> mineral water
Support: 0.006666666666666667
Confidence: 0.390625
Lift: 3.975152645861601
--------------------------------------------------------
Rule: herb & pepper -> nan
Support: 0.016
Confidence: 0.3234501347708895
Lift: 3.2915549671393096
--------------------------------------------------------
Rule: herb & pepper -> spaghetti
Support: 0.0064
Confidence: 0.3934426229508197
Lift: 4.003825878061259
--------------------------------------------------------
Rule: milk -> olive oil
Support: 0.004933333333333333
Confidence: 0.22424242424242424
Lift: 3.411395906324912
--------------------------------------------------------
Rule: nan -> ground beef
Support: 0.005333333333333333
Confidence: 0.37735849056603776
Lift: 3.840147461662528
--------------------------------------------------------
Rule: spaghetti -> shrimp
Support: 0.006
Confidence: 0.5232558139534884
Lift: 3.004914704939635
--------------------------------------------------------
Rule: spaghetti -> milk
Support: 0.0072
Confidence: 0.20300751879699247
Lift: 3.0883496774390333
--------------------------------------------------------
Rule: soup -> olive oil
Support: 0.0052
Confidence: 0.2254335260115607
Lift: 3.4295161157945335
--------------------------------------------------------
Rule: olive oil -> nan
Support: 0.008
Confidence: 0.2714932126696833
Lift: 4.130221288078346
```

```
--------------------------------------------------------
Rule: shrimp -> pasta
Support: 0.005066666666666666
Confidence: 0.3220338983050848
Lift: 4.514493901473151
--------------------------------------------------------
Rule: spaghetti -> olive oil
Support: 0.005066666666666666
Confidence: 0.20105820105820105
Lift: 3.0586947422647217
--------------------------------------------------------
Rule: frozen vegetables -> shrimp
Support: 0.005333333333333333
Confidence: 0.23255813953488372
Lift: 3.260160834601174
--------------------------------------------------------
Rule: spaghetti -> cooking oil
Support: 0.0048
Confidence: 0.5714285714285714
Lift: 3.281557646029315
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.008666666666666666
Confidence: 0.3110047846889952
Lift: 3.164906221394116
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.004533333333333334
Confidence: 0.28813559322033905
Lift: 3.0224013274860737
--------------------------------------------------------
Rule: olive oil -> frozen vegetables
Support: 0.0048
Confidence: 0.20338983050847456
Lift: 3.094165778526489
--------------------------------------------------------
Rule: frozen vegetables -> shrimp
Support: 0.0072
Confidence: 0.3068181818181818
Lift: 3.2183725365543547
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.005733333333333333
Confidence: 0.20574162679425836
Lift: 3.1299436124887174
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.006
```

```
Confidence: 0.21531100478468898
Lift: 3.0183785717479763
--------------------------------------------------------
Rule: frozen vegetables -> spaghetti
Support: 0.006666666666666667
Confidence: 0.23923444976076555
Lift: 3.497579674864993
--------------------------------------------------------
Rule: spaghetti -> grated cheese
Support: 0.005333333333333333
Confidence: 0.3225806451612903
Lift: 3.282706701098612
--------------------------------------------------------
Rule: herb & pepper -> mineral water
Support: 0.006666666666666667
Confidence: 0.390625
Lift: 3.975152645861601
--------------------------------------------------------
Rule: herb & pepper -> spaghetti
Support: 0.0064
Confidence: 0.3934426229508197
Lift: 4.003825878061259
--------------------------------------------------------
Rule: olive oil -> milk
Support: 0.004933333333333333
Confidence: 0.22424242424242424
Lift: 3.411395906324912
--------------------------------------------------------
Rule: spaghetti -> shrimp
Support: 0.006
Confidence: 0.5232558139534884
Lift: 3.004914704939635
--------------------------------------------------------
Rule: olive oil -> spaghetti
Support: 0.0072
Confidence: 0.20300751879699247
Lift: 3.0883496774390333
--------------------------------------------------------
Rule: soup -> olive oil
Support: 0.0052
Confidence: 0.2254335260115607
Lift: 3.4295161157945335
--------------------------------------------------------
Rule: spaghetti -> olive oil
Support: 0.005066666666666666
Confidence: 0.20105820105820105
Lift: 3.0586947422647217
--------------------------------------------------------
```

```
Rule: spaghetti -> nan
Support: 0.004533333333333334
Confidence: 0.28813559322033905
Lift: 3.0224013274860737
-------------------------------------------------------
```

[ ]: