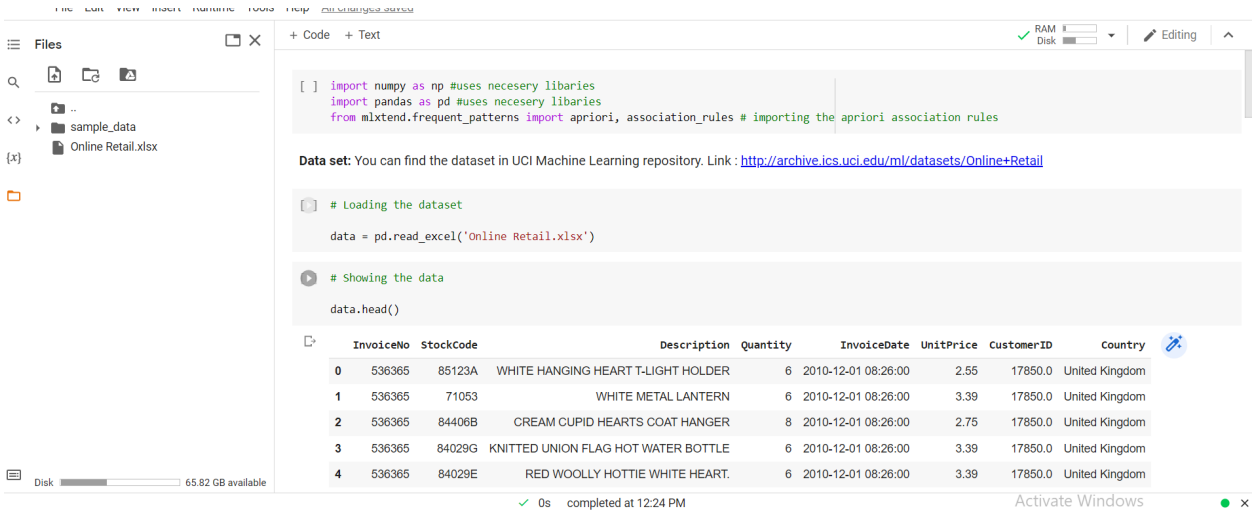
Super perform assignment

MD Mehedi Hasan Rabbe
2018200000030

COURSE CODE : CSE 4053.1

RECOMMENDATION SYSTEM USING APRIORI ALGORITHM




```
[ ]  # The columns in the data

     data.columns
```

```
Index(['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate',
       'UnitPrice', 'CustomerID', 'Country'],
      dtype='object')
```

```
[ ]  # The of the data

     data.shape #we check the number of rows and columns
```

```
(541909, 8)
```

```
[ ]  # Checkign whether there is any null values of not

     data.isnull().values.any()
```

```
True
```

```
[ ]  # As the previous cell told us that there are some null values. So, let's find them!

     data.isnull().sum()
```

```
InvoiceNo           0
StockCode           0
Description      1454
Quantity            0
InvoiceDate         0
UnitPrice           0
CustomerID     135080
Country             0
dtype: int64
```

Data Preprocessing

```
     # Stripping extra spaces in the description
     data['Description'] = data['Description'].str.strip()

     # Dropping the rows without any invoice number
     data.dropna(axis = 0, subset =['InvoiceNo'], inplace = True)
     data['InvoiceNo'] = data['InvoiceNo'].astype('str')

     # Dropping all transactions which were done on credit
     data = data[~data['InvoiceNo'].str.contains('C')]
```

```
[ ]  # Let's see the countries in our dataset

     data.Country.unique() #we check which countries are in the dayaset
```

```
array(['United Kingdom', 'France', 'Australia', 'Netherlands', 'Germany',
       'Norway', 'EIRE', 'Switzerland', 'Spain', 'Poland', 'Portugal',
       'Italy', 'Belgium', 'Lithuania', 'Japan', 'Iceland',
       'Channel Islands', 'Denmark', 'Cyprus', 'Sweden', 'Finland',
       'Austria', 'Bahrain', 'Israel', 'Greece', 'Hong Kong', 'Singapore',
       'Lebanon', 'United Arab Emirates', 'Saudi Arabia',
       'Czech Republic', 'Canada', 'Unspecified', 'Brazil', 'USA',
       'European Community', 'Malta', 'RSA'], dtype=object)
```

```
[ ]  # Splitting the data according to the region of transaction
     # Transactions done in France
     basket_France = (data[data['Country'] =="France"]
               .groupby(['InvoiceNo', 'Description'])['Quantity']
               .sum().unstack().reset_index().fillna(0)
               .set_index('InvoiceNo'))
```

```
     # Defining the hot encoding function to make the data suitable

     def hot_encode(x): #hot_encode defines essentially the representation of categorical variables as binary vectors
         if(x<= 0):
             return 0
         if(x>= 1):
```

```
[ ]         return 1
```

```
[ ] # Applying one hot encoding

    basket_encoded = basket_France.applymap(hot_encode)
    basket_France = basket_encoded
```

```
basket_France.head()
```

Description	10 COLOUR SPACEBOY PEN	12 COLOURED PARTY BALLOONS	12 EGG HOUSE PAINTED WOOD	12 MESSAGE CARDS WITH ENVELOPES	12 PENCIL SMALL TUBE WOODLAND	12 PENCILS SMALL TUBE RED RETROSPOT	12 PENCILS SMALL TUBE SKULL	12 PENCILS TALL TUBE POSY	12 PENCILS TALL TUBE RED RETROSPOT	12 PENCILS TALL TUBE WOODLAND	15CM CHRISTMAS GLASS BALL 20 LIGHTS	16 PIECE CUTLERY SET PANTRY DESIGN	18PC WOODEN CUTLERY SET DISPOSABLE	R
InvoiceNo														
536370	0	0	0	0	0	0	0	0	0	0	0	0	0	
536852	0	0	0	0	0	0	0	0	0	0	0	0	0	
536974	0	0	0	0	0	0	0	0	0	0	0	0	0	
537065	0	0	0	0	0	0	0	0	0	0	0	0	0	
537463	0	0	0	0	0	0	0	0	0	0	0	0	0	

5 rows × 1563 columns

✓ 0s completed at 12:24 PM

Activate Windows

Building the model

```
[ ] # Building the model
    frq_items = apriori(basket_France, min_support = 0.1, use_colnames = True) #for frnace country only , minimum support 0.1

    # Collecting the inferred rules in a dataframe
    rules = association_rules(frq_items, metric ="lift", min_threshold = 1) #association_rules is a function
    rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False]) #sorting them according their confidence & lift value
```

```
[ ] print(rules.head()) #printing the rules
```

```
                              antecedents  ... conviction
40          (SET/6 RED SPOTTY PAPER PLATES)  ...  21.556122
42  (SET/6 RED SPOTTY PAPER PLATES, POSTAGE)  ...  18.107143
35       (STRAWBERRY LUNCH BOX WITH CUTLERY)  ...   3.755102
27      (ROUND SNACK BOXES SET OF4 WOODLAND)  ...   3.637755
41            (SET/6 RED SPOTTY PAPER CUPS)  ...   7.852041

[5 rows x 9 columns]
```

From the above output, it can be seen that paper cups and plates are bought together in France. This is because the French have a culture of having a get-together with their friends and family atleast once a week.