Celonis with Python Integration:

AP – LATE PAYMENT PREDICTION

B.date – Due date – Clr date

Aim:

1. Predict whether if payment for particular invoice will be a late payment or not
2. Root cause analysis and summarize pattern
3. Proactively intimate related persons in order to take right actions
4. Provide weekly or monthly report saying the probable invoices and the reasons

Approach:

Include Activities(event ’-CEL\_AP\_ACTIVITIES’) table, attributes(BKPF, BSEG,LFA1, LFB1) table

1. Classify existing data into
   1. Late payment (due date<clr date)
   2. On time payment (due date=clr date)
   3. Quick payment (due date>clr date)
2. For each classification take the following features to be given to the python program
   1. Activity or event path (What activities happen in what order)
   2. Fields
      1. Payment term accepted: “BSEG”.”ZTERM”
      2. Clearing Date: “BSEG”.”AUGDT”
      3. Baseline Date for Due Date Calculation: “BSEG”.“ZFBDT”
   3. Activity names:

Hypothesis:

Timeline Goal:

Implementing Apriori algorithm in Python

**Prerequisites:** [Apriori Algorithm](https://www.geeksforgeeks.org/apriori-algorithm/)

Apriori Algorithm is a Machine Learning algorithm which is used to gain insight into the structured relationships between different items involved. The most prominent practical application of the algorithm is to recommend products based on the products already present in the user’s cart. **Walmart** especially has made great use of the algorithm in suggesting products to it’s users.

Dataset : [Groceries data](http://archive.ics.uci.edu/ml/datasets/Online+Retail)  
**Implementation of algorithm in Python:**  
**Step 1: Importing the required libraries**

filter\_none

brightness\_4

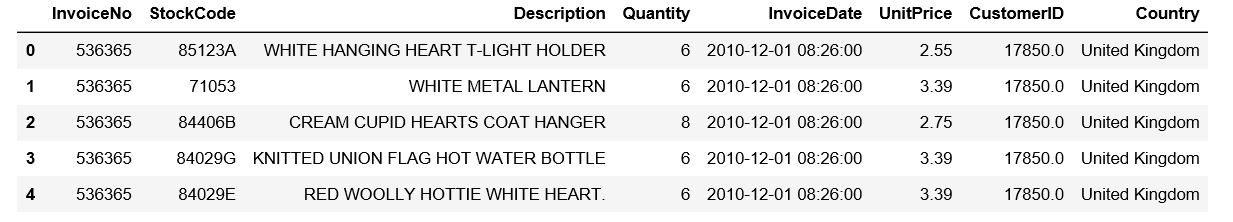
|  |
| --- |
| import numpy as np  import pandas as pd  from mlxtend.frequent\_patterns import apriori, association\_rules |

**Step 2: Loading and exploring the data**

filter\_none

brightness\_4

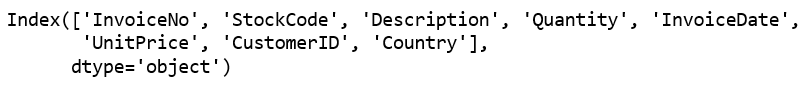
|  |
| --- |
| # Changing the working location to the location of the file  cd C:\Users\Dev\Desktop\Kaggle\Apriori Algorithm    # Loading the Data  data = pd.read\_excel('Online\_Retail.xlsx')  data.head() |



filter\_none

brightness\_4

|  |
| --- |
| # Exploring the columns of the data  data.columns |



filter\_none

brightness\_4

|  |
| --- |
| # Exploring the different regions of transactions  data.Country.unique() |



**Step 3: Cleaning the Data**

filter\_none

brightness\_4

|  |
| --- |
| # 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')] |

**Step 4: Splitting the data according to the region of transaction**

filter\_none

brightness\_4

|  |
| --- |
| # Transactions done in France  basket\_France = (data[data['Country'] =="France"]            .groupby(['InvoiceNo', 'Description'])['Quantity']            .sum().unstack().reset\_index().fillna(0)            .set\_index('InvoiceNo'))    # Transactions done in the United Kingdom  basket\_UK = (data[data['Country'] =="United Kingdom"]            .groupby(['InvoiceNo', 'Description'])['Quantity']            .sum().unstack().reset\_index().fillna(0)            .set\_index('InvoiceNo'))    # Transactions done in Portugal  basket\_Por = (data[data['Country'] =="Portugal"]            .groupby(['InvoiceNo', 'Description'])['Quantity']            .sum().unstack().reset\_index().fillna(0)            .set\_index('InvoiceNo'))    basket\_Sweden = (data[data['Country'] =="Sweden"]            .groupby(['InvoiceNo', 'Description'])['Quantity']            .sum().unstack().reset\_index().fillna(0)            .set\_index('InvoiceNo')) |

**Step 5: Hot encoding the Data**

filter\_none

brightness\_4

|  |
| --- |
| # Defining the hot encoding function to make the data suitable  # for the concerned libraries  def hot\_encode(x):      if(x<= 0):          return 0      if(x>= 1):          return 1    # Encoding the datasets  basket\_encoded = basket\_France.applymap(hot\_encode)  basket\_France = basket\_encoded    basket\_encoded = basket\_UK.applymap(hot\_encode)  basket\_UK = basket\_encoded    basket\_encoded = basket\_Por.applymap(hot\_encode)  basket\_Por = basket\_encoded    basket\_encoded = basket\_Sweden.applymap(hot\_encode)  basket\_Sweden = basket\_encoded |

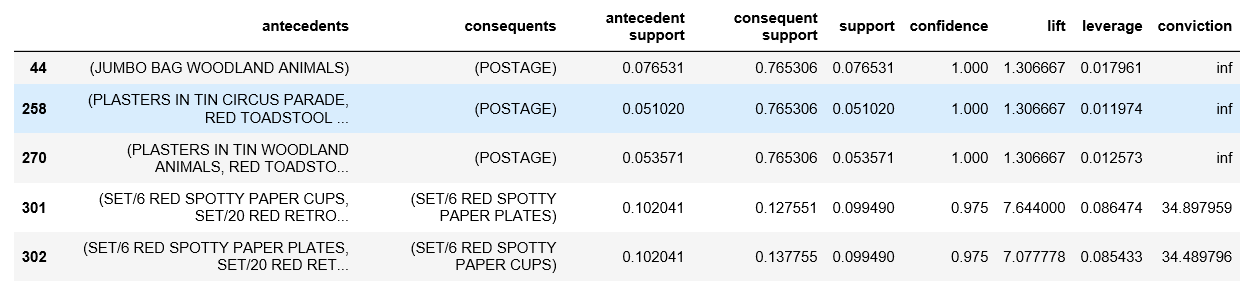
**Step 6: Buliding the models and analyzing the results**

a)**France:**

filter\_none

brightness\_4

|  |
| --- |
| # Building the model  frq\_items = apriori(basket\_France, min\_support = 0.05, use\_colnames = True)    # Collecting the inferred rules in a dataframe  rules = association\_rules(frq\_items, metric ="lift", min\_threshold = 1)  rules = rules.sort\_values(['confidence', 'lift'], ascending =[False, False])  print(rules.head()) |



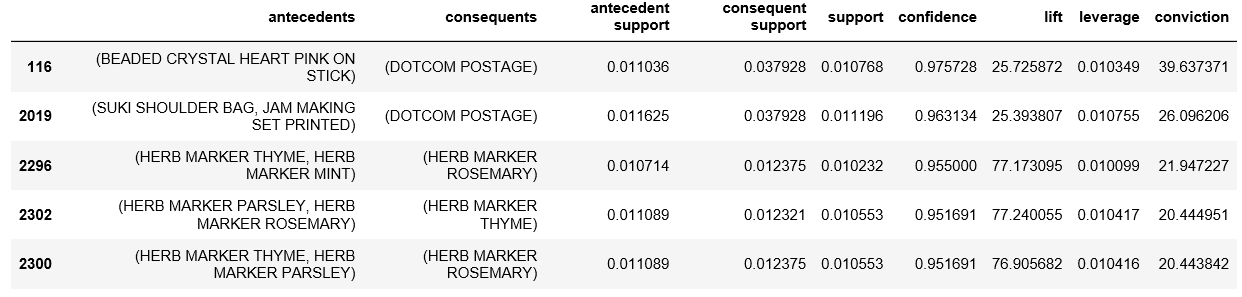
From the above output, it can be seen that paper cups and paper and plates are bought together in France. This is because the French have a culture of having a get-together with their friends and family atleast once a week. Also, since the French government has banned the use of plastic in the country, the people have to purchase the paper -based alternatives.

b)**United Kingdom:**

filter\_none

brightness\_4

|  |
| --- |
| frq\_items = apriori(basket\_UK, min\_support = 0.01, use\_colnames = True)  rules = association\_rules(frq\_items, metric ="lift", min\_threshold = 1)  rules = rules.sort\_values(['confidence', 'lift'], ascending =[False, False])  print(rules.head()) |



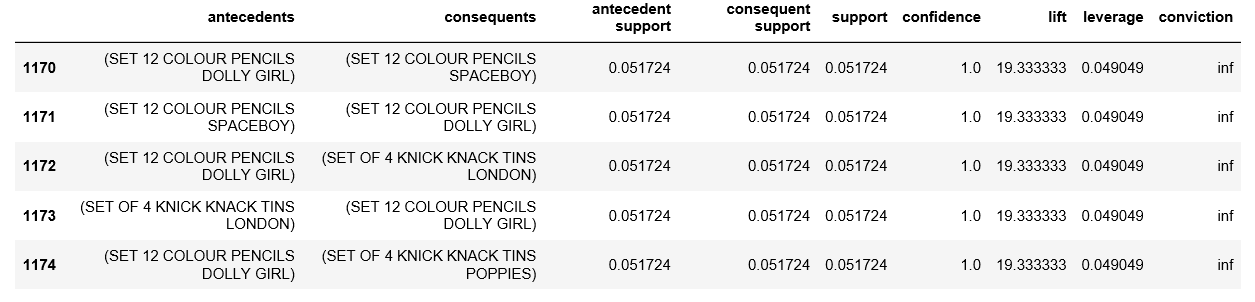
If the rules for British transactions are analyzed a little deeper, it is seen that the British people buy different coloured tea-plates together. A reason behind this may be because typically the British enjoy tea very much and often collect different coloured tea-plates for different ocassions.

c)**Portugal:**

filter\_none

brightness\_4

|  |
| --- |
| frq\_items = apriori(basket\_Por, min\_support = 0.05, use\_colnames = True)  rules = association\_rules(frq\_items, metric ="lift", min\_threshold = 1)  rules = rules.sort\_values(['confidence', 'lift'], ascending =[False, False])  print(rules.head()) |



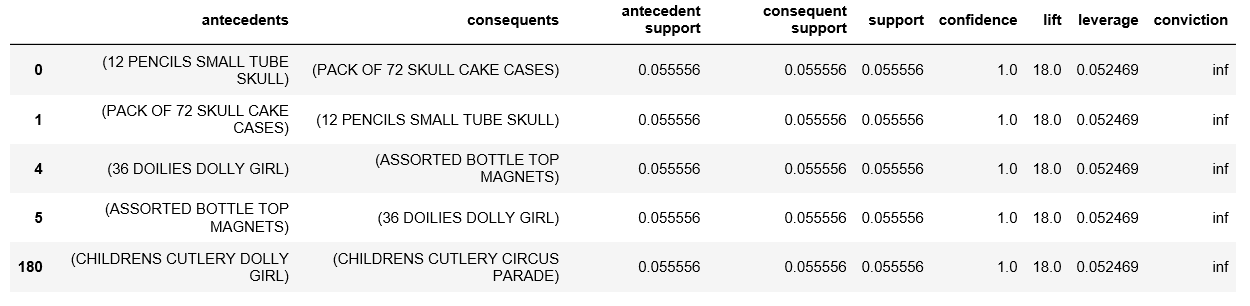
On analyzing the association rules for Portuguese transactions, it is observed that Tiffin sets (Knick Knack Tins) and colour pencils. These two products typically belong to a primary school going kid. These two products are required by children in school to carry their lunch and for creative work respectively and hence are logically make sense to be paired together.

d)**Sweden:**

filter\_none

brightness\_4

|  |
| --- |
| frq\_items = apriori(basket\_Sweden, min\_support = 0.05, use\_colnames = True)  rules = association\_rules(frq\_items, metric ="lift", min\_threshold = 1)  rules = rules.sort\_values(['confidence', 'lift'], ascending =[False, False])  print(rules.head()) |



On analyzing the above rules, it is found that boys’ and girls’ cutlery are paired together. This makes practical sense because when a parent goes shopping for cutlery for his/her children, he/she would want the product to be a little customized according to the kid’s wishes.

