Tarefa01

October 29, 2021

1 Tarefa 1

1.1 Aluno

- Isaque Elcio de Souza (RA: 225310)

Dataset em http://fimi.ua.ac.be/data/retail.dat que é um dataset real de compras no varejo em uma loja na Belgica

Regras de associação que tenham suporte mínimo de 0.005 e confiança minima de 0.9

Biblioteca python apyori - https://pypi.org/project/apyori/

```
[4]: import pandas as pd
     import numpy as np
     from apyori import apriori
```

```
[9]: # importar os dados e converter em lista
     transacoes = []
     with open('../MO433/dados/retail.dat', 'r') as file:
         for l in file.readlines():
             #print(l.split())
             transacoes.append(l.split())
```

```
[10]: # remover cabeçalho (primeira linha)
      transacoes.pop(0)
```

```
[10]: ['0',
       '1',
       '2',
       '3',
       '4',
       '5',
       '6',
       '7',
       '8',
```

```
    '9',
    '10',
    '11',
    '12',
    '13',
    '14',
    '15',
    '16',
    '17',
    '18',
    '19',
    '20',
    '21',
    '22',
    '23',
    '24',
    '25',
    '26',
    '27',
    '28',
    '29']
```

[18]:
```python
# suporte mínimo de 0.005 e confiança minima de 0.9
regras = apriori(transacoes, min_support = 0.005, min_confidence = 0.9)
```

[19]:
```python
# resultado e quantidade de regras geradas
resultados = list(regras)
len(resultados)
```

[19]: 37

[21]:
```python
# Conjunto de regras geradas
resultados
```

[21]:
```
[RelationRecord(items=frozenset({'38', '105'}), support=0.007293474438810812,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'105'}),
items_add=frozenset({'38'}), confidence=0.9786910197869102,
lift=5.532340279266081)]),
 RelationRecord(items=frozenset({'38', '110'}), support=0.030909359013622805,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'110'}),
items_add=frozenset({'38'}), confidence=0.9753042233357194,
lift=5.513195411227261)]),
 RelationRecord(items=frozenset({'16011', '16010'}),
support=0.007384217511144384,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'16011'}),
items_add=frozenset({'16010'}), confidence=0.9730941704035876,
lift=65.18917565117833)]),
```

```
RelationRecord(items=frozenset({'38', '170'}), support=0.03438028153038192,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'170'}),
items_add=frozenset({'38'}), confidence=0.9780574378831881,
lift=5.5287587702756955)]),
 RelationRecord(items=frozenset({'38', '286'}), support=0.01265865859053323,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'286'}),
items_add=frozenset({'38'}), confidence=0.9433643279797127,
lift=5.332645711658082)]),
 RelationRecord(items=frozenset({'38', '36'}), support=0.03164664647633307,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'36'}),
items_add=frozenset({'38'}), confidence=0.9502724795640327,
lift=5.371696080459393)]),
 RelationRecord(items=frozenset({'38', '37'}), support=0.011864656707614479,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'37'}),
items_add=frozenset({'38'}), confidence=0.9739292364990689,
lift=5.505422891702643)]),
 RelationRecord(items=frozenset({'38', '371'}), support=0.008699992059981171,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'371'}),
items_add=frozenset({'38'}), confidence=0.9808184143222507,
lift=5.544366005710692)]),
 RelationRecord(items=frozenset({'38', '55'}), support=0.007452274815394562,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'55'}),
items_add=frozenset({'38'}), confidence=0.9332386363636362,
lift=5.275407246759075)]),
 RelationRecord(items=frozenset({'38', '56'}), support=0.005830242397431971,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'56'}),
items_add=frozenset({'38'}), confidence=0.9607476635514018,
lift=5.430910154293097)]),
 RelationRecord(items=frozenset({'38', '790'}), support=0.0057621850931817925,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'790'}),
items_add=frozenset({'38'}), confidence=0.9713193116634798,
lift=5.490669520105415)]),
 RelationRecord(items=frozenset({'38', '39', '105'}),
support=0.005092954934721702,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '105'}),
items_add=frozenset({'38'}), confidence=0.9868131868131867,
lift=5.578253229202191)]),
 RelationRecord(items=frozenset({'32', '110', '38'}),
support=0.0050248976304715235,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'32', '110'}),
items_add=frozenset({'38'}), confidence=0.9866369710467706,
lift=5.577257117495149)]),
 RelationRecord(items=frozenset({'38', '110', '39'}),
support=0.01973661823255181,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'110', '39'}),
items_add=frozenset({'38'}), confidence=0.9891984081864695,
lift=5.591736398058948)]),
```

```
RelationRecord(items=frozenset({'38', '110', '41'})),
support=0.00755436077176983,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'110', '41'}),
items_add=frozenset({'38'}), confidence=0.9837518463810929,
lift=5.5609480975124095)]),
 RelationRecord(items=frozenset({'38', '48', '110'})),
support=0.015437665180748857,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '110'}),
items_add=frozenset({'38'}), confidence=0.986231884057971,
lift=5.574967243551859)]),
 RelationRecord(items=frozenset({'32', '170', '38'})),
support=0.006034414310182507,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'32', '170'}),
items_add=frozenset({'38'}), confidence=0.9851851851851853,
lift=5.569050468781169)]),
 RelationRecord(items=frozenset({'38', '39', '170'})),
support=0.022901282880185114,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '170'}),
items_add=frozenset({'38'}), confidence=0.9805730937348227,
lift=5.542979258576283)]),
 RelationRecord(items=frozenset({'38', '41', '170'})),
support=0.009006249929106975,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'41', '170'}),
items_add=frozenset({'38'}), confidence=0.9863354037267081,
lift=5.5755524190786305)]),
 RelationRecord(items=frozenset({'38', '48', '170'})),
support=0.017445355656129126,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '170'}),
items_add=frozenset({'38'}), confidence=0.9877970456005137,
lift=5.5838147818150095)]),
 RelationRecord(items=frozenset({'38', '39', '286'})),
support=0.008257619582355009,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '286'}),
items_add=frozenset({'38'}), confidence=0.9706666666666665,
lift=5.486980251346498)]),
 RelationRecord(items=frozenset({'38', '48', '286'})),
support=0.006590215628225633,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '286'}),
items_add=frozenset({'38'}), confidence=0.9830795262267344,
lift=5.557147609109716)]),
 RelationRecord(items=frozenset({'32', '36', '38'})),
support=0.00535384126768072,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'32', '36'}),
items_add=frozenset({'38'}), confidence=0.9554655870445343,
lift=5.401051655516363)]),
 RelationRecord(items=frozenset({'38', '39', '36'})),
support=0.02206190946109958,
```

```
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '36'}),
items_add=frozenset({'38'}), confidence=0.9548355424644085,
lift=5.397490142293198)]),
 RelationRecord(items=frozenset({'38', '41', '36'}),
support=0.007611075191978313,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'41', '36'}),
items_add=frozenset({'38'}), confidence=0.9585714285714286,
lift=5.41860834279852)]),
 RelationRecord(items=frozenset({'38', '48', '36'}),
support=0.015426322296707162,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '36'}),
items_add=frozenset({'38'}), confidence=0.96045197740113,
lift=5.429238700927226)]),
 RelationRecord(items=frozenset({'38', '39', '37'}),
support=0.007758532684520366,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '37'}),
items_add=frozenset({'38'}), confidence=0.9674681753889675,
lift=5.468899833961705)]),
 RelationRecord(items=frozenset({'38', '48', '37'}),
support=0.006317986411224918,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '37'}),
items_add=frozenset({'38'}), confidence=0.9858407079646018,
lift=5.57275600505689)]),
 RelationRecord(items=frozenset({'38', '371', '39'}),
support=0.005966357005932329,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'371', '39'}),
items_add=frozenset({'38'}), confidence=0.9887218045112782,
lift=5.5890422549063095)]),
 RelationRecord(items=frozenset({'38', '110', '41', '39'}),
support=0.005796213745306882,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'110', '41', '39'}),
items_add=frozenset({'38'}), confidence=0.992233009708738,
lift=5.608890380157223)]),
 RelationRecord(items=frozenset({'38', '48', '110', '39'}),
support=0.011694513446989031,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '110', '39'}),
items_add=frozenset({'38'}), confidence=0.9942140790742526,
lift=5.620088960327339)]),
 RelationRecord(items=frozenset({'38', '39', '41', '170'}),
support=0.006975873685643312,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '41', '170'}),
items_add=frozenset({'38'}), confidence=0.985576923076923,
lift=5.571264883007477)]),
 RelationRecord(items=frozenset({'38', '48', '39', '170'}),
support=0.013532060661743854,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '39', '170'}),
items_add=frozenset({'38'}), confidence=0.9892205638474295,
```

```
lift=5.591861639481484)]),
 RelationRecord(items=frozenset({'38', '48', '41', '170'}),
support=0.005489955876181078,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '41', '170'}),
items_add=frozenset({'38'}), confidence=0.983739837398374,
lift=5.560880213187872)]),
 RelationRecord(items=frozenset({'38', '48', '39', '286'}),
support=0.00519504089109697,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '39', '286'}),
items_add=frozenset({'38'}), confidence=0.9870689655172413,
lift=5.579699093932131)]),
 RelationRecord(items=frozenset({'38', '39', '36', '41'}),
support=0.006272614875058132,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'39', '36', '41'}),
items_add=frozenset({'38'}), confidence=0.9667832167832167,
lift=5.465027902976735)]),
 RelationRecord(items=frozenset({'38', '48', '39', '36'}),
support=0.012250314765032157,
ordered_statistics=[OrderedStatistic(items_base=frozenset({'48', '39', '36'}),
items_add=frozenset({'38'}), confidence=0.967741935483871,
lift=5.470447343818514)])]
```

```
[28]: items_base = []
      items_add = []
      suporte = []
      confianca = []
      lift = []

      for resultado in resultados:
          #print(resultado)
          supor = resultado[1]
          #print(resultado_posicao)
          resultadoRegras = resultado[2]
          #print(resultadoRegras)
          for resultadoRegra in resultadoRegras:

              base = list(resultadoRegra[0])
              adicional  = list(resultadoRegra[1])
              conf = resultadoRegra[2]
              l = resultadoRegra[3]

              items_base.append(base)
              items_add.append(adicional)
              suporte.append(supor)
              confianca.append(conf)
              lift.append(l)
```

```
[34]: df_Regras = pd.DataFrame({'items_base': items_base, 'items_add',␣
      ↪'suporte': suporte, 'confianca': confianca, 'lift': lift})
```

```
[35]: # Visulalização das regras com suporte, confiança e lift
      # Foram encontardas 37 regras e pode ser observado que o item 38 é encontrado␣
      ↪na maioria das regras geradas
      df_Regras
```

[35]:

	items_base	items_add	suporte	confianca	lift
0	[105]	[38]	0.007293	0.978691	5.532340
1	[110]	[38]	0.030909	0.975304	5.513195
2	[16011]	[16010]	0.007384	0.973094	65.189176
3	[170]	[38]	0.034380	0.978057	5.528759
4	[286]	[38]	0.012659	0.943364	5.332646
5	[36]	[38]	0.031647	0.950272	5.371696
6	[37]	[38]	0.011865	0.973929	5.505423
7	[371]	[38]	0.008700	0.980818	5.544366
8	[55]	[38]	0.007452	0.933239	5.275407
9	[56]	[38]	0.005830	0.960748	5.430910
10	[790]	[38]	0.005762	0.971319	5.490670
11	[39, 105]	[38]	0.005093	0.986813	5.578253
12	[32, 110]	[38]	0.005025	0.986637	5.577257
13	[110, 39]	[38]	0.019737	0.989198	5.591736
14	[110, 41]	[38]	0.007554	0.983752	5.560948
15	[48, 110]	[38]	0.015438	0.986232	5.574967
16	[32, 170]	[38]	0.006034	0.985185	5.569050
17	[39, 170]	[38]	0.022901	0.980573	5.542979
18	[41, 170]	[38]	0.009006	0.986335	5.575552
19	[48, 170]	[38]	0.017445	0.987797	5.583815
20	[39, 286]	[38]	0.008258	0.970667	5.486980
21	[48, 286]	[38]	0.006590	0.983080	5.557148
22	[32, 36]	[38]	0.005354	0.955466	5.401052
23	[39, 36]	[38]	0.022062	0.954836	5.397490
24	[41, 36]	[38]	0.007611	0.958571	5.418608
25	[48, 36]	[38]	0.015426	0.960452	5.429239
26	[39, 37]	[38]	0.007759	0.967468	5.468900
27	[48, 37]	[38]	0.006318	0.985841	5.572756
28	[371, 39]	[38]	0.005966	0.988722	5.589042
29	[110, 41, 39]	[38]	0.005796	0.992233	5.608890
30	[48, 110, 39]	[38]	0.011695	0.994214	5.620089
31	[39, 41, 170]	[38]	0.006976	0.985577	5.571265
32	[48, 39, 170]	[38]	0.013532	0.989221	5.591862
33	[48, 41, 170]	[38]	0.005490	0.983740	5.560880
34	[48, 39, 286]	[38]	0.005195	0.987069	5.579699
35	[39, 36, 41]	[38]	0.006273	0.966783	5.465028
36	[48, 39, 36]	[38]	0.012250	0.967742	5.470447